

Generative adversarial networks for ultrasound image synthesis and analysis in nondestructive evaluation

Posilović, Luka

Doctoral thesis / Disertacija

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Electrical Engineering and Computing / Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:168:751803>

Rights / Prava: [In copyright / Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-04-19**



Repository / Repozitorij:

[FER Repository - University of Zagreb Faculty of Electrical Engineering and Computing repository](#)





University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Luka Posilović

**GENERATIVE ADVERSARIAL NETWORKS FOR
ULTRASOUND IMAGE SYNTHESIS AND
ANALYSIS IN NONDESTRUCTIVE EVALUATION**

DOCTORAL THESIS

Zagreb, 2022



University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Luka Posilović

**GENERATIVE ADVERSARIAL NETWORKS FOR
ULTRASOUND IMAGE SYNTHESIS AND
ANALYSIS IN NONDESTRUCTIVE EVALUATION**

DOCTORAL THESIS

Supervisor: Professor Sven Lončarić, PhD

Zagreb, 2022



Sveučilište u Zagrebu
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Luka Posilović

**GENERATIVNE SUPARNIČKE MREŽE ZA
SINTEZU I ANALIZU ULTRAZVUČNIH SLIKA U
NERAZORNIM ISPITIVANJIMA**

DOKTORSKI RAD

Mentor: Prof. dr. sc. Sven Lončarić

Zagreb, 2022.

The doctoral thesis was completed at the University of Zagreb Faculty of Electrical Engineering and Computing, Department of Electronic Systems and Information Processing.

Supervisor: Professor Sven Lončarić, PhD

The thesis has 93 pages.

Thesis number: _____

About the Supervisor

Sven Lončarić is a professor of electrical engineering and computer science at the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia since 2011. From 2001-2003, he was an assistant professor at the New Jersey Institute of Technology, Newark, USA. As a Fulbright scholar, he received a Ph.D. degree in the field of image processing and analysis from the University of Cincinnati, OH in 1994. He received Diploma of Engineering and Master of Science degrees in electrical engineering from the University of Zagreb Faculty of Electrical Engineering and Computing (FER) in 1985 and 1989, respectively.

His areas of research interest are image processing and computer vision. He was the principal investigator on a number of R&D projects. He is the head of the Image Processing Group and was the Chair of the Department of Electronic Systems and Information Processing at FER. He is the founder and the director of the Center for Computer Vision at the University of Zagreb. He is a co-director of the Center of Excellence in Data Science and Cooperative Systems.

Prof. Lončarić co-authored more than 250 publications in scientific journals and conferences. Prof. Lončarić was the Chair of the IEEE Croatia Section. He is a senior member of IEEE and a member of the Croatian Academy of Technical Sciences. He is a program committee member and reviewer for a number of scientific conferences and journals. Prof. Lončarić received several awards for his scientific and professional work including the National Science Award for outstanding scientific achievements from the Croatian Parliament, Golden Plaque "Josip Lončar", "Fran Bošnjaković" award, "Rikard Podhorsky" award, and the Annual award for outstanding scientific achievements. He is featured on a comprehensive list released by Stanford University of the top 1% percent world's scientists in the category "Artificial intelligence and image processing" in 2021.

O mentoru

Sven Lončarić je profesor elektrotehnike i računarstva na Fakultetu elektrotehnike i računarstva pri Sveučilištu u Zagrebu, Hrvatskoj, od 2011. godine. Od 2001. do 2003. godine bio je profesor na sveučilištu New Jersey Institute of Technology, Newark, SAD. Kao dobitnik Fulbright-ove stipendije doktorirao je 1994. godine na University of Cincinnati, SAD u području obrade i analize slike. Diplomirao je i magistrirao na Sveučilištu u Zagrebu Fakultetu elektrotehnike i računarstva 1985. odnosno 1989. godine.

Područja istraživačkog interesa su obrada slike i računalni vid. Bio je voditelj više istraživačkih i razvojnih projekata. Voditelj je Grupe za obradu i analizu slike i bio je predstojnik Zavoda za elektroničke sustave i obradbu informacija. Osnivač je i voditelj Centra za računalni vid na Sveučilištu u Zagrebu. Su-voditelj je Znanstvenog centra izvrsnosti za znanost o

podacima i kooperativne sustave.

Prof. Lončarić su-autor je više od 250 publikacija u znanstvenim časopisima i konferencijama. Prof. Lončarić bio je voditelj hrvatskog IEEE ogranka. Stariji je član IEEE i redoviti član Akademije tehničkih znanosti Hrvatske. Programski je voditelj i recenzent nekolicine znanstvenih konferencija i časopisa. Prof. Lončarić dobitnik je više priznanja svog znanstvenog i profesionalnog rada uključujući Državnu nagradu za znanost Hrvatskog sabora, zlatnu plaketu "Josip Lončar", nagradu "Fran Bošnjaković", "Rikard Podhorsky i Godišnje nagrade za znanost. Na listi je Stanford University-a spomenut kao top 1% znanstvenika u kategoriji "Umjetna inteligencija i obradba slike" u 2021. godini.

Preface

This thesis is based upon the results of the research conducted in 2018. to 2022. as a part of the project "Smart UTX" (KK.01.2.1.01.0151) co-financed by the INETEC d.o.o. and the European Union from the European Regional Development Fund and supervised by Professors Sven Lončarić, Ph.D., Marko Subašić, Ph.D. and Tomislav Petković, Ph.D.

First and foremost I would like to express my gratitude and respect to my supervisor Professor Sven Lončarić, Ph.D., the head of the Image Processing Group at the Department of Electronic Systems and Information Processing at the Faculty of Electrical Engineering and Computing, University of Zagreb. His precious time that he was always so generous with, his attention to detail and his dedication to his work were priceless to me. Having a great supervisor that always pushed me to be better, but also be my friend and my most confident believer, gave me the best tailwind I could imagine.

I would also like to thank Professor Marko Subašić, Ph.D., and Tomislav Petković, Ph.D., for all the ideas they gave me and all the conversations we had that helped me finish my degree. I thank them for both professional and private advices that helped me keep my sanity in this period.

I also wish to express my special thanks to my family, my father Željko who taught me to love science, learning and how to be curious, my mother Renata who supported me in every one of my endeavors and my sister Monika who always stood proud of me. Although, without false humility, I must congratulate myself for enduring all the extra hours and having the patience for finishing the doctoral degree, none of it would be possible without the help of my friends, specially Duje Medak (soon to also be Ph.D.), my favorite co-author who shared many great, fun and also difficult moments of this project and education with me from day one.

Finally, the biggest "Thank You" goes to my one and only, Petra. As she was there in the best and the worst, supporting me no matter what and keeping me going every day. I will forever be grateful for all the breakfasts you made me and every "you can do it" you convinced me to.

Abstract

Non-destructive ultrasound evaluation is a technique for flaw detection in materials. Among other things, it is used for monitoring key components of nuclear power plants, railways, and pipelines. Analysis of ultrasound data is performed by human inspectors manually analyzing acquired images. Such a process can be tedious and is highly dependent on the inspector's previous experience. Deep learning methods are hard to develop because of the lack of available data. The shortage also impacts training new human experts in this field, since they learn through experience. Data from real inspections can not be used because of non-disclosure agreements. On the other hand, blocks with synthetic flaws are expensive to produce. The goal of this work is to develop methods for defect detection. To improve the method's performance, deep learning methods for generating additional synthetic images need to be developed. Generated data should be realistic and of high quality even to human experts in ultrasound evaluation. Additional data should be used to improve the performance of the deep learning defect detector.

Although some methods for defect detection and classification were developed, they are mostly based on traditional image and signal processing techniques such as magnitude threshold. Such an approach does not yield satisfactory results and can not generalize well on data from various ultrasound probes and scanned materials. In the field of generating synthetic ultrasound data, yet much has to be done. Only some traditional and Finite Element Methods (FEM) have been proposed. All this means that there is much space for improvements and new ideas in the field of defect detection in ultrasound non-destructive evaluation. Guided by this need for improvements, in this thesis, several methods for non-destructive ultrasound data analysis and ultrasound image generation are proposed. These methods can be summarized as follows:

- Generative adversarial network with additional object detector discriminator for synthesis of high quality realistic ultrasound images
- Improved one-stage defect detector trained using additional ultrasound images synthesised using generative adversarial network
- Method for anomaly detection in ultrasound images.

Experimental results of the proposed methods are presented and discussed. The proposed generative method outperforms existing methods when the quality of images is assessed by human experts. The generated images can even be used to improve the performance of the state-of-the-art deep learning defect detector. Also, the analysis of different methods for anomaly detection in ultrasonic images is given alongside the improvements made to some methods.

Keywords: non-destructive testing, ultrasound image analysis, automated flaw detection, image augmentation, image generation, deep learning, generative networks

Prošireni sažetak

Generativne suparničke mreže za sintezu i analizu ultrazvučnih slika u nerazornim ispitivanjima

Kontrola kvalitete i integriteta konstrukcijskih materijala u elektranama, cjevovodima, avio, auto i svemirskoj industriji obavezna je kako bi se ograničio njihov utjecaj na okoliš, osigurala dugotrajnost i sigurnost. Jedna od skupina metoda za inspekciju i provjeru materijala su metode nerazornog ispitivanja. Takva je grupa skup metoda koje provjeravaju postojanost različitih legura i kompozitnih materijala. Te metode koriste različite tehnologije, a neke od njih su:

- vizualna metoda,
- metoda temeljena na vrtložnim strujama,
- ultrazvučna metoda.

Ultrazvučna metoda jedna je od najšire korištenih zbog mnoštva pozitivnih svojstava. Neka od njih su mogućnost dubokog prodiranja u materijal, ali i visoka razlučivost u plitkim područjima. Također, moguće je ispitivanje različitih materijala, na suhom i uronjeno u vodu. Cilj ultrazvučnih i ostalih metoda je skeniranje materijala i pronalazak pukotina i ostalih deformiteta tijekom analize. Ultrazvučno se snimanje provodi korištenjem ultrazvučnih sondi koje odašilju zvučne signale visoke frekvencije. Te frekvencije su više od 20 kHz, a mogu ići i do nekoliko stotina MHz. Što je frekvencija viša, to ultrazvučni val manje prodire u materijal, ali može detektirati nepravilnosti manjih dimenzija. Ultrazvučne sonde prislanjaju se na materijal uz pomoć nekog lubrikanta koji olakšava prodiranje valova u materijal. U slučaju da nema nepravilnosti, valovi prolaze do kraja materijala, odbijaju se od njega i vraćaju u sondu. U slučaju da je prisutan defekt, valovi se odbijaju od njega i vraćaju se u sondu. Na taj način ultrazvučnim sondama možemo otkriti defekte. Kako bi ljudski eksperti educirani za analizu ultrazvučnih podataka mogli s lakoćom percipirati rezultate snimanja materijala ultrazvučnim sondama, oni koriste nekoliko osnovnih vrsta vizualizacije podataka. Vizualizacija se dijeli na A-skenove, B-skenove i C-skenove. A-skenovi je prikaz amplitude primljenog ultrazvučnog vala postavljeni u odnosu na vrijeme. B-skenovi su slike, sačinjene od mnoštva A-skenova duž jednog prolaza sonde materijalom. Oni se mogu još promatrati kao poprečni presjek materijala. C-skenovi su sačinjeni od niza B-skenova, a mogu se promatrati kao tlocrtni pogled na ispitni materijal. Provođenje nedestruktivnog testiranja može se podijeliti u dvije faze, akviziciju i analizu. Akvizicija se uglavnom provodi automatski koristeći robotske manipulatore sa sondama. Analizu s druge strane provode inspektori, eksperti za analizu ultrazvučnih podataka. Inspekcija jedne elektrane može trajati nekoliko mjeseci, u kojem se periodu sakupi velika količina podataka. Cilj ovoga rada je automatizacija analize ultrazvučnih podataka nedestruktivnog testiranja.

U području nedestruktivnog testiranja mnogo je radova i istraživanja provedeno s ciljem automatizacije analize podataka. Najviše je pomaka napravljeno u području analiza metodama

vrtožnih struja. Metode vrtožnih struja već su dulji niz godina automatizirane. S druge strane, nema puno pomaka u području ultrazvučne analize podataka. Istražene su i objavljene metode temeljene na tradicionalom pristupu, primjerice Valične transformacije, Fourierove transformacije, metode praga i druge. Takve metode nisu dostatne za ultrazvučnu analizu zbog raznovrsnosti ultrazvučnih podataka, ovisnosti izgleda podataka o vrsti korištene sonde, geometriji i materijalu snimanog bloka. Za uspješnu analizu i detekciju defekata potrebna je metoda koja bolje generalizira sve navedene slučajeve. Takve metode kriju se u dubokom učenju i dubokim konvolucijskim neuronskim mrežama. Konvolucijske neuronske mreže prikladne su za analizu slika, a mogu se koristiti za klasifikaciju, detekciju i generiranje novih podataka. Postoje mnoge popularne arhitekture neuronskih mreža koje postižu visoke performanse na generalnim, velikim i lako dostupnim skupovima podataka poput COCO i ImageNet. Takvi skupovi podataka sadrže stotine tisuća slika sa više od milijun označenih objekata poput automobila, osoba i predmeta. Mreže razvijene na ovim skupovima podataka nisu prilagođene za analizu ultrazvučnih podataka koji imaju svoje specifičnosti.

Iako se tijekom inspekcije jedne elektrane ili cjevovoda sakupi velika količina podataka za koje su potrebni mjeseci da se adekvatno analiziraju, nasreću, te konstrukcije rijetko sadrže ikakve defekte. Nedostatak defekata iz realnih inspekcija znači nedostatak pozitivnih primjera u skupu za učenje. Kako trenutno svu analizu ultrazvučnih podataka provode eksperti manualno, rezultati analize uvelike ovise o njihovom iskustvu analize na stvarnim inspekcijama. Nedostatak podataka sa defektima sa stvarnih inspekcija produljuje edukaciju novih eksperata što predstavlja veliki izazov. Najčešće se eksperti educiraju na metalnim blokovima ili komponentama s umjetno implantiranim defektima. Postoji više metoda simuliranja stvarnih defekata, a neki od njih su uzastopno grijanje i naglo hlađenje materijala te elektromagnetsko naprezanje. Proizvodnja takvih realnih blokova je vrlo skupa i dugotrajna. Nedostatak podataka otežava i adekvatno treniranje dubokih neuronskih mreža. Konačni cilj je razvoj metode za detekciju defekata i anomalnih skenova. To se može uz nedostatak podataka napraviti modifikacijom i izradom vlastitih mreža prilagođenih za analizu male količine podataka sa svojstvima ultrazvučnih podataka. Nedostatak podataka možemo se nositi korištenjem mreža pred-naučenih na velikim javnim skupovima podataka, treniranjem samo djela mreže i augmentacijom podataka.

Metode augmentacije podataka koriste se kako bi se postojeći skup podataka proširio zbog stvaranja više primjera za učenje, ali i kako bi se mreži otežalo učenje i spriječila prenaučenost na skupu za treniranje i poboljšala generalizacija na testnom skupu podataka. Augmentacija slike može se podijeliti u dvije skupine, klasične metode manipulacije karakteristika slike i metode temeljene na dubokom učenju. Tradicionalne metode augmentacije slika bazirane su na geometrijskim transformacijama i drugim funkcijama obrade slika. Te su metode relativno jednostavne za implementaciju i ne zahtijevaju velike procesorske resurse. Metode temeljene na

dubokom učenju nedavno su postale važna tema u području istraživanja prateći razvoj dubokih neuronskih mreža kao što su autoenkoderi i generativne suparničke mreže.

Klasične metode augmentacije pokazale su se korisnim za povećanje točnosti konvolucijskih neuronskih mreža. Važno je napomenuti da nije svaka augmentacijska metoda prikladna za svaki slučaj učenja i vrstu podataka. Neki skupovi podataka poput ultrazvučnih B-skenova imaju svoje specifičnosti zbog kojih se neki tipovi augmentacije ne mogu koristiti. U nastavku su navedene neke metode klasične augmentacije slike.

- **Rotiranje slike oko osi**

Augmentacija rotacijom provodi se rotacijom slike u lijevo ili desno oko osi za određeni stupanj. Ako se slika rotira za 180 stupnjeva, ta se augmentacija zove i *flip*. Najčešće se radi po horizontalnoj i vertikalnoj osi.

- **Translacija slike**

Translacija slike za određeni broj piksela može biti vrlo korisno kako bi se izbjegla lokacijska pristranost modela. Prilikom translacije, zaostali prostor koji je prije zauzimala slika može se ispuniti nulama, 255 ili bilo kojom drugom vrijednosti.

- **Modulacija boja slike**

Uobičajeno je slike spremati kao tenzore dimenzije *visina x širina x kanali boje*. Augmentacijom boje slike može se jednostavno mijenjati svjetlina ili kontrast slike i izdvojiti neki od kanala boje slike.

- **Dodavanje šuma**

Dodavanjem šuma može pomoći konvolucijskim mrežama izvući robusnije značajke iz skupa podataka za treniranje. Može se dodati Gaussov, Poissonov ili neki drugi šum.

- **Isjecivanje slike**

Ova se augmentacija može koristiti kako bi se povećao skup podataka, ali i ako radimo sa slikama velike rezolucije ili neobičnih dimenzija kako bi se normalizirao ulaz u konvolucijsku mrežu koja ima fiksnu ulaznu dimenziju.

- **Brisanje slučajnog djela slike**

Slučajno brisanje nekog djela slike nadahnuto je regularizacijom. Provodi se tako da se vrijednosti u dijelu slike postave u vrijednost 0 kako bi se mreža naučila da mora obraćati pažnju na cijelu sliku, a ne na jedan njen dio.

- **Izrezivanje i lijepljenje dijelova slike**

Dok se tradicionalne metode augmentacije temelje na promjeni neke karakteristike postojeće slike, ova se metoda koristi kako bi se stvorili potpuno novi primjeri za učenje. Ovo je metoda kojom se izrezuju objekti s primjera slike i lijepe na druge dijelove iste slike ili na druge slike. U slučaju kada postoji puno primjera slika bez objekata, a tek nekoliko primjera slika s objektima ova je metoda posebno korisna, što je slučaj u ultrazvučnim podacima. Izrezujući defekte sa pozitivnih primjera slike i lijepeći ih u skenove bez

defekata možemo poboljšati preciznost mreža jer možemo povećati broj slika sa problematičnim defektima. Problem kod ove augmentacije je da se ne smije prečesto koristiti u skupu za učenje jer se modeli lako mogu prenaučiti na anomalijama u tako stvorenim slikama.

Metode augmentacije temeljene na dubokom učenju koristeći autoenkodere ili generativne suparničke mreže nisu još dovoljno istražene. Do sada, nije istražena upotreba takvih vrsta mreža za primjenu na ultrazvučnim podacima. Generativne metode u području analize ultrazvučnih slika nedestruktivnog testiranja mogu imati višestruku primjenu. Te se metode mogu koristiti za proširenje skupa podataka za treniranje neuronskih mreža, ali mogu se koristiti i za učenje i pripremu eksperata za analizu ultrazvučnih podataka. Razvojem metode koja bi mogla generirati neograničen broj ultrazvučnih B-skenova s defektima olakšala bi se i značajno pojeftinila edukacija eksperata, a metode detekcije defekata bi povećale svoju točnost.

Generiranje realnih ultrazvučnih podataka ima nekoliko zahtjeva koje je potrebno ispuniti. Jedno od njih je generiranje različitih tipova geometrije i šuma koji se obično nalazi u ultrazvučnim podacima. Drugi je mogućnost određivanja položaja generiranog defekta na ultrazvučnoj slici. A treći je generiranje podataka koji vizualno odgovaraju stvarnim podacima i koje eksperti ne mogu razlikovati od stvarnih podataka. Da bi se riješio problem podataka i razvila generativna mreža, prvo je izrađen skup podataka za treniranje. Na skupu ultrazvučnih podataka nastalog skeniranjem šest metalnih blokova označeni su svi defekti. Svaki blok sadrži između šest i 34 defekata. Blokovi su skenirani koristeći INETEC ultrazvučne sonde. Konačan skup podataka sadrži više od 4000 slika i skoro 7000 označenih defekata. Nakon što su svi defekti označeni, izrađeni su parovi slika/maska kako bi se mogao trenirati generativni model. Maske su izrađene tako da su bile vrijednosti nula na području gdje nije bilo defekata, a na mjestu oznake defekta postavljena je bila vrijednost 255. Razvijena generativna mreža temelji se na konceptu uvjetne generativne suparničke mreže koja kao ulaz ima sliku (masku), a kao izlaz generiranu ultrazvučnu sliku s defektom na definiranom mjestu. Temeljena je na Pix2pixGAN-u. Razvijena mreža sastoji se od generatora u formi U-net mreže sa preskočnim vezama, dva PatchGAN diskriminatora koji rade na dvije različite veličine ulazne slike i dodatnim diskriminatorom, pretreniranim detektorom defekata. Problem s baznom pix2pix mrežom je nedovoljno precizno pozicioniranje defekta na željeno mjesto, te artefakti koji zbunjuju detektor defekata. Pomoću razvijene mreže generirano je 200.000 dodatnih ultrazvučnih slika. Te se slike mogu koristiti za treniranje detektora defekata. Pokazano je kako detektor treniran samo na generiranim slikama daje veću točnost za otprilike 2%, dok detektor treniran na generiranim i stvarnim podacima daje točnost veću za skoro 6% prema onome treniranom samo na pravim podacima. Također, provedeno je testiranje kvalitete generiranih slika uz pomoć eksperata za ultrazvučne podatke. Uspoređene su tri metode, naša razvijena generativna metoda, *Copy/Paste* metoda, te modificirani SPADEGAN. Ekspertima su slučajnim raspod-

jelom pokazane generirane i prave slike, a njihov zadatak je bio odrediti vjerojatnost da je prikazana slika pravi ultrazvučni B-sken. Statistička analiza pokazala je da inspektori mogu vidjeti razliku između pravih i slika generiranih Copy/Paste metodom i SPADEGAN-om, no ne mogu vidjeti razliku između slika generiranih našom metodom i pravih slika. Također, kako bi metoda bila korisna za edukaciju inspektora potrebno je generirati ultrazvučne podatke cijelog bloka, a ne samo jedan B-sken. Taj smo zadatak također ispunili modulacijom maski i generiranjem sličnih skenova koji se spajaju u 3D prikaz skeniranog virtualnog bloka.

Generativne mreže osim za generiranje ultrazvučnih podataka mogu se koristiti za detekciju anomalija. Detekcija anomalija je pristup nenadziranog učenja u kojemu se neronska mreža, primjerice autoenkoder, trenira na slikama bez defekata, a kasnije ona može prepoznati razliku između anomalne i normalne slike. Taj je pristup također istražen i razvijena je metoda za detekciju anomalija na ultrazvučnim slikama nedestruktivnog testiranja.

Ključne riječi: nerazorno ispitivanje, analiza ultrazvučnih slika, automatska detekcija defekata, augmentacija slike, generiranje slika, duboko učenje, generativne mreže

Contents

1. Introduction	1
1.1. Background and motivation	.1
1.2. Problem statement	.2
1.2.1. Defect detector	.2
1.2.2. Generative model	.3
1.2.3. Anomaly detector	.3
1.3. Scientific contributions	.4
1.4. Thesis structure	.4
2. Related work	5
2.1. Defect detection	.6
2.2. Synthetic data generation	.7
2.2.1. Autoencoder	.8
2.2.2. Generative Adversarial Networks	.10
2.3. Anomaly detection	.14
3. Non-destructive evaluation	16
3.1. Ultrasonic data acquisition	.16
3.2. Ultrasonic data evaluation	.17
4. Assisted analysis	19
4.1. Data processing algorithms	.19
4.2. Evaluation of developed algorithms	.21
5. The main scientific contributions of the thesis	22
5.1. A new deep-learning generative adversarial network	.22
5.2. A deep-learning based defect detector method	.23
5.3. A deep-learning based anomaly detection method	.23

6. Conclusion and future work	25
6.1. The main conclusions of the thesis	.25
6.2. Future work	.27
7. List of publications	28
8. Author's contribution to the publications	29
Bibliography	31
Publications	40
Pub 1: Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans	.41
Pub 2: Generating Ultrasonic Images Indistinguishable from Real Images Using Generative Adversarial Networks	.53
Pub 3: Synthetic 3D Ultrasonic Scan Generation Using Optical Flow and Generative Adversarial Networks	.62
Pub 4: Deep Learning-Based Anomaly Detection From Ultrasonic Images	.69
Pub 5: DefectDet: a deep learning architecture for detection of defects with extreme aspect ratios in ultrasonic images	.81
Biography	91
Životopis	93

Chapter 1

Introduction

Non-destructive evaluation is used in various industry and science applications to detect defects in the inspected material and prevent further damage to the system. Since time and reliability are the critical factors in ensuring the safety of the systems, it is important to develop a fast and accurate automatic or assisted analysis. For this reason, the research of a deep-learning-based ultrasonic image generation and analysis is the topic of this thesis. In this chapter first, the background and the motivation for the thesis are presented. Following that, the formal problem statement is given. Finally, the main scientific contributions of the thesis are listed, followed by the description of the thesis structure.

1.1 Background and motivation

Ensuring the structural integrity of the critical components of the nuclear power plants, airplanes, pipelines and such is the foundation for a safe environment and safe society. Most of the components cannot be frequently replaced due to the high component price and high out-of-operation cost. Therefore, they need to be inspected every once in a while in a non-destructive manner. Non-destructive evaluation (NDE) is widely used in science and industry to evaluate the properties of materials, components, or systems without causing damage [1]. Many different methods are available including visual examination, eddy current, and ultrasonic inspection. Among them, ultrasonic inspection stands out due to its versatility. High sensitivity on most materials [2], extraction of defects location and type [2], high signal-to-noise ratio [3] and high penetrating power are some of its advantages. Ultrasonic inspection can be divided into two phases, the acquisition and the analysis of acquired data. Although most of the acquisition today is carried out automatically utilizing robots, analysis is still done manually, with human experts spending hours, days, and sometimes months analyzing the data. Such evaluation of the data is tiresome, highly dependable on the inspector's previous experience, and man-hour expensive. Deep convolutional neural networks have already been proven many times to out-

perform traditional methods in image processing. For the development of automatic NDE, the deep learning solution is therefore the most promising approach. However, for the development of deep convolutional neural networks, one should have a lot of data available for training. In the field of NDE, there are not a lot of defects in material found in the inspections. Furthermore, the anomalous data that we have is subjected to different non-disclosure agreements and cannot be used to train the networks. This lack of positive examples does not necessarily affect only the development of the deep learning networks and automated analysis, but also the education and training of new human experts in this field. By generating synthetic ultrasonic data, one would jump across these obstacles and be able to develop a proper automated analysis. In order to approach the development of these methods, clear problem statements have to be made before proceeding further.

1.2 Problem statement

Ultrasonic data from a non-destructive inspection can be presented in many different formats, including the A-scan, B-scan and C-scan [4]. An A-scan is a signal's amplitude as a function of time, a B-scan displays a cross-sectional view of the inspected material, and a C-scan provides a top view of its projected features. An example of the A, B, and C scan is shown in Figure 1.1. The shown B-scan and the C-scan are artificially colored, and defects are marked with the blue bounding box on the B-scan. The main goal is to discriminate the data without defects from the data with defects, and furthermore locate the defects. Defects and some of its surrounding is best seen in a B-scan. We can achieve the defect detection by using deep-learning object detectors on images (B-scans). However, the lack of data compels us to search for the methods to increase the training dataset or decrease the amount of data needed for training such detectors. The development of special object detectors and anomaly detectors dedicated to ultrasonic data analysis and generative methods for generating synthetic ultrasonic data is the goal of this work.

1.2.1 Defect detector

Object detection is a highly researched deep-learning area with high-performance methods already developed. Since training an object detector is a supervised problem, we need a lot of labeled data for training. These methods are often trained and tested on big open-source datasets such as ImageNet [5] or Common Objects in Context (COCO) [6] which contain thousands, even millions of images. For most of the real-world challenges, such datasets are unimaginable and unattainable. Defect detection on ultrasonic images is one of such challenges. To be able to train state-of-the-art object detection models, we need to make some modifications to the architecture and training procedure. Deep learning convolutional neural networks for object de-

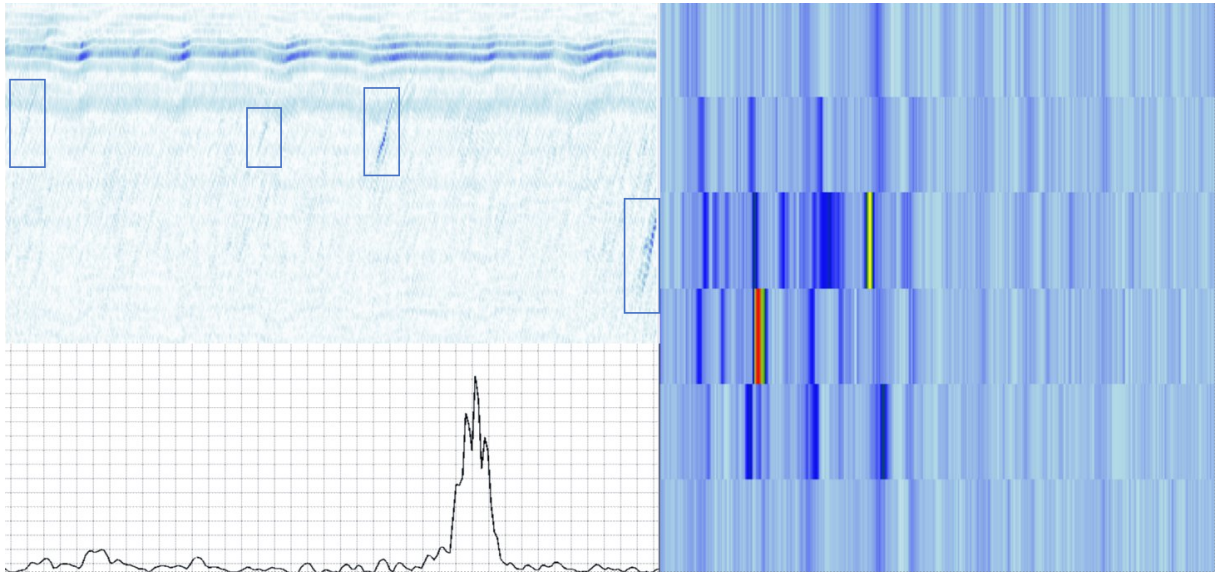


Figure 1.1: An example of the A-scan (bottom-left), artificially colored B-scan (top-left) with marked defects and artificially colored C-scan (right)

tection can be divided into two groups, the one-stage, and two-stage object detectors. One-stage object detectors are much faster, but carry their downsides. Such object detectors are made of two parts, the feature extractor, and the detection head. Improvements should be made in both parts to get the best possible defect detector which is able to be trained on a small available dataset.

1.2.2 Generative model

The main goal of a generative network here is to generate additional synthetic data which can be used for training the defect detector for the ultrasonic evaluation as well as human experts. Generative methods can be divided into traditional and deep-learning methods. In this work, we discuss both approaches to the problem. While traditional methods are based on the simple augmentation of the existing dataset, deep learning methods capture the distribution of the data. With the approximation of this distribution, we are able to generate an unlimited amount of new data. A special requirement for generating synthetic data for object detection is to generate defects in exact predefined locations, which is an additional challenge for synthetic data generation. Generative methods are trained by an unsupervised approach.

1.2.3 Anomaly detector

Finally, our goal is to be able to help the inspectors with the analysis of ultrasonic data. To speed up the analysis process, one of the approaches is to filter out the data without any visible defects and show the inspectors only anomalous data. Inspectors would then have to manually analyze only a small portion of the data, significantly reducing the time needed for the analysis.

However, this is not an easy task since defects can often be hidden in the background noise and resemble inspected component geometry reflections. Anomaly detection methods are trained using only normal images, which we have plenty of. This approach utilizes self-supervised learning.

1.3 Scientific contributions

The emphasis of this thesis is on novel methods for defect detection and ultrasonic data generation. This is achieved by developing a novel deep learning generative neural network and using this additionally generated data for defect detection performance improvement. Also, in this thesis, the quality of the generated data is assessed by both neural networks and human experts. In the end, anomaly detection in ultrasonic data, as well as the novel object detector for defect detection, is presented. The scientific contributions of this thesis are the following:

- Generative adversarial network with additional object detector discriminator for synthesis of high quality realistic ultrasound images.
- Improved one-stage defect detector trained using additional ultrasound images synthesised via generative adversarial network.
- Self-supervised anomaly detection model for ultrasound images

1.4 Thesis structure

The thesis is structured as follows. First, an overview of existing methods for defect detection, synthetic data generation, and anomaly detection in ultrasonic data analysis and similar fields is given in Chapter 2. In Chapter 2 an introduction to generative methods is given as well as some of the implementation details of the deep learning generative adversarial networks. In Chapter 3 an introduction to non-destructive testing and evaluation is given. Chapter 4 describes the challenges in the development of the assisted non-destructive testing data analysis. The main scientific contributions of the thesis are presented in Chapter 5. Chapter 6 gives the conclusion and future research directions are given. In Chapter 7 the list of publications that fully describe the main scientific contributions are given. The author's contributions to the included publications are summarized in Chapter 8.

Chapter 2

Related work

Automated analysis of non-destructive evaluation data has long been used in many NDT systems. However, so far it has been limited to classical decision-making algorithms such as signal amplitude threshold [7]. Complex data such as the one from ultrasonic inspection makes it hard to develop an automated analysis. All ultrasonic analysis is, to the best of my knowledge, done manually by a trained human expert, the inspector. The automated analysis could make the process much faster and more reliable. Plenty of efforts have been made to develop automated ultrasonic analysis [8, 9, 10]. Developed algorithms for automated defect detection can be divided into three groups related to data representations being used: A-scans [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19], B-scans [7, 20, 21, 22, 23, 24, 25] and C-scans [26, 27, 28].

Regarding A-scans, a popular approach for automatic flaw detection is the usage of discrete wavelet transform (DWT) for feature extraction. Wavelet coefficients are used as an input to classifiers such as artificial neural networks (ANN) [11] [12], support vector machines (SVM) [13, 14, 15, 16], or a combination of these two [10]. DWT was shown to be a better feature extractor than discrete Fourier transform (DFT) using an ANN classifier in [8]. On the other hand in [17] DWT, DFT and discrete cosine transform (DCT) were used and further classified using two ANNs, DFT achieved better results. Directly feeding the A-scan into an ANN was shown in [9] and [19]. In [9] a superiority of deep neural networks over single-layer networks was shown.

B-scans keep the geometrical coherence of the defect, which leads to a better noise immunity [20]. In [20] authors have used DWT to reduce noise prior to detecting flaws with Radon transform. In [21] DWT was shown to be a better approach than Gabor filter banks for ultrasonic B-scans segmentation. In [2] ultrasonic images obtained by using pulse laser illumination were classified as defect or non-defect using various methods including deep learning which achieved the best result. Two popular deep learning object detection models, YOLO and SSD, were used in [22] for defect detection. In [7] data copy/pasting data augmentation was used to enlarge the

dataset for training a deep learning detector.

Regarding C-scans, in [26] a method based on the comparison of the scan with a reconstructed reference image has been made. The method was able to detect all defects in their dataset, but with a high number of false-positive detection. There have also been some attempts in estimating defects from noisy measurements using Bayesian analysis [27].

Very few algorithms involve deep learning and modern deep learning architectures. The problem with deep learning in many fields is the lack of available data for training such networks. This problem besides ultrasonic data analysis can be found in medical image processing and other fields.

In this work we research three connected fields of ultrasonic data analysis, all with the same goal of developing automatic defect detection:

- defect detection
- synthetic data generation
- anomaly detection

Even though automated non-destructive evaluation captured a lot of researchers' attention recently, there are many works from related areas such as medical image analysis that are relevant and will be mentioned in this thesis. In this chapter, I will give an overview of the work related to this thesis. While image analysis and machine learning have been researched for decades now, only did the recent development of high-speed parallel processing units allowed deep learning to show its capabilities. Very deep neural networks have therefore been developed for purposes such as image classification, segmentation, or object detection. Many other interesting concepts of deep learning have been introduced, such as autoencoders [29] and generative adversarial networks [30].

2.1 Defect detection

A lot of research has been done on the deep learning object detectors development. They can be divided into two groups, two-stage, and one-stage object detectors. Two-stage detectors consist of a region proposal network and the other for identifying objects in proposed regions. The first of its kind is the Region - convolutional neural network (R-CNN) [31]. Commonly, for region proposals, a selective search algorithm [31] is used. Proposed regions are then propagated and fed into a convolutional neural network (CNN) that acts as a feature extractor. Extracted features are fed into a support vector machine (SVM) to classify the presence of an object within the candidate region. This process has many drawbacks, it takes a lot of time and data to train such CNN, and it cannot be implemented in real-time. Many improvements have been made to fasten the process of detection, therefore Fast R-CNN [32] and Faster R-CNN [33] have emerged. Faster R-CNN can even be used for real-time object detection. All of the above-

mentioned networks achieve great accuracy.

One stage detectors have a single CNN that predicts the bounding boxes and class probabilities for these boxes. They don't have a region proposal network, but split the image into $S \times S$ grid. Within each cell, the CNN makes m object predictions with class probabilities. Many advances have been made from the first such models, YOLO [34] and SSD [35]. YOLO has since developed in three iterations [36, 37, 38] and many more networks such as RetinaNet [39] and EfficientDet [40] have shown great precision with much faster inference time than two-stage detectors.

Usually, both one-stage and two-stage detectors are trained, and their performance is tested on large public datasets such as *Common Objects in Context* [6] (COCO), *Pattern Analysis, Statistical Modelling and Computational Learning Visual Object Classes* (Pascal VOC) [41], Canadian Institute For Advanced Research 10 (CIFAR-10) [42] or *ImageNet* [5]. COCO dataset consists of 330k images with 1.5 million object instances. ImageNet has over 1 million annotated images. These large datasets make those huge CNNs with millions or 10s of millions of parameters trainable. It remains a challenge and is almost impossible to train modern deep CNNs from scratch on more usual, smaller datasets with a few thousand or hundreds of images. When training on smaller datasets, models tend to overfit the training dataset and achieve poor results on the test dataset. Overfitting refers to the phenomenon when a network achieves great results on the training dataset, but fails to generalize on other data [43].

2.2 Synthetic data generation

As previously mentioned, ultrasonic inspection is made to ensure no cracks appear due to thermal or electromagnetic material fatigue. Therefore, many ultrasonic inspections are done with the goal of early intervention of cracks, but usually, no defects are found. As defects rarely appear in real inspections, inspectors for training use artificially created cracks in metal blocks. However, it is rather expensive to realistically create such defects in materials. On the other hand, all data from ultrasonic inspections of nuclear power plants or oil piping are protected by non-disclosure agreements and cannot be used after the inspection.

Data is the biggest throwback in the development of proper automated/assisted ultrasonic analysis. With little data, it is hard to train the before-mentioned best-performing deep CNNs. Not only in ultrasonic analysis but this challenge can also be found in many medical image processing tasks [43] where, due to the rarity of some pathology and patient privacy issues, data availability is very modest. There are some ways of training state-of-the-art CNNs on small datasets. For example, transfer learning [44] is often used in combination with freezing the backend CNN layers [45] and is shown to enhance the accuracy of models. All of the methods not only aid with small datasets but show significant improvements when training on

large above-mentioned datasets. In this section, we are focusing on methods for enhancing the performance of deep CNNs using data augmentation.

2.2.1 Autoencoder

Autoencoders (AE) are a deep learning method very useful in performing feature space augmentation on data. They consist of two networks, an encoder that maps the images into a low-dimensional vector representation, a latent space, and a decoder that reconstructs these vectors back into the original image.

$$\begin{aligned}\phi : \chi &\rightarrow F \\ \psi : F &\rightarrow \chi \\ \phi, \psi &= \operatorname{argmin} \| X - (\psi \circ \phi)X \| \end{aligned} \tag{2.1}$$

where: ϕ = the encoder

ψ = the decoder

χ = original image

F = latent space

The goal of training the autoencoder is to perfectly reconstruct the input image. It is done by some special loss functions and model architecture. There are many versions of autoencoders, and recently they have proven to be able to generate and reconstruct high-resolution images. After training, the autoencoder learns the distribution of the dataset. By removing the encoder part of the network and randomly sampling from the decoder's input, we can generate a series of different images from the same feature space.

Since the input and the desired output are the same images, autoencoders are often classified as self-supervised learning approaches. Networks are specially designed such as to impose a bottleneck structure that forces a compressed knowledge of the input image. This architecture can be seen in Figure 2.1. Autoencoders are often used for dimensionality reduction, anomaly detection, or image processing. Here, we will discuss the usage of autoencoders as an image augmentation method.

Traditional autoencoders are simple neural networks closely related to Principal Component Analysis (PCA), another dimensionality reduction method. In fact, an autoencoder with linear activation functions will map the same subspace as the one PCA would [46]. Generally, activation functions used in autoencoders are non-linear, a sigmoid is the most commonly used

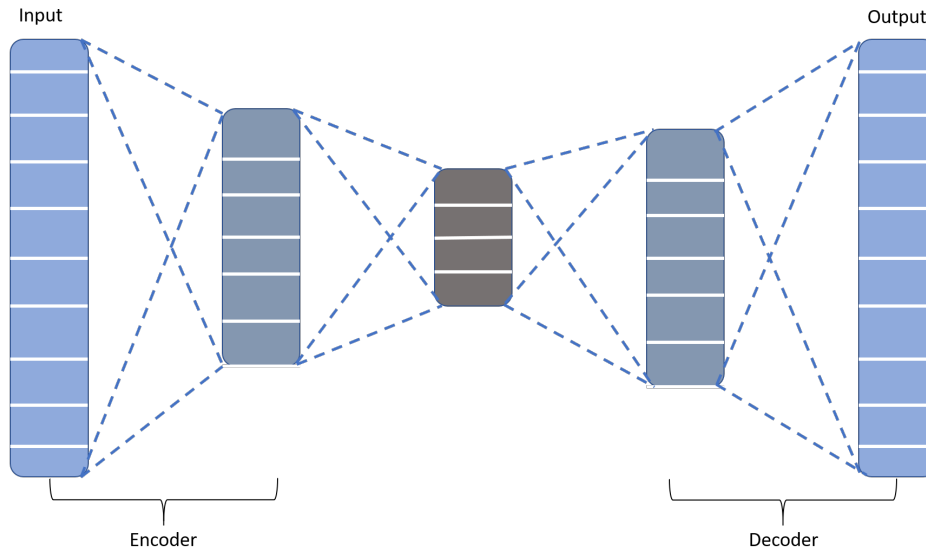


Figure 2.1: Architecture of a simple autoencoder

activation function in Autoencoders [47]. Usually Kullback–Leibler (KL) divergence [48] or l1 loss is used.

There are several types of autoencoders, but the most commonly used are Variational Autoencoders (VAE) [49]. They inherit the architecture from traditional AEs to learn a general distribution, which allows us to pick random samples from the latent space. Variational autoencoder uses variational inference to generate the approximation of a posterior distribution. This distribution separates latent variable centroids for different images. By randomly sampling from the distribution or adding random noise to the mentioned centroids, new images are generated.

There are many different versions of VAEs. One example is the Vector Quantised-Variational AutoEncoder (VQ-VAE) [50]. Typically, we assume the prior (latent vectors) and the posterior (images) to be normally distributed. The encoder then predicts the mean and variance of the posterior. In VQ-VAE authors have used discrete latent variables which is more natural to the problem we are trying to solve. For example, if we have categories like "boat" and "dog" it doesn't make sense to interpolate between them. Moreover, the authors claim their approach keeps the AE from posterior collapse, which usually prevents making use of larger and more complex networks. From the traditional VAE, VQ-VAE differs on the latent space layer called *Vector Quantization Layer*. In VQ-VAE we have a dictionary with vectors for different categories in the prior. The encoder calculates the vectors and passes them to the vector quantization layer which calculates the distances from each vector in the dictionary, swaps them for the ones in the dictionary, and passes them to the decoder. Since a minimum distance is calculated in the vector quantization layer, backpropagation is not possible, therefore input and the output of this layer are compared for training. Three losses are used for training this autoencoder, a reconstruction loss for the encoder and the decoder and l2 and commitment loss for the vector quantization layer. Commitment loss makes sure the number of latent vectors does not grow

arbitrarily. VQ-VAE exceeds in generating images, videos, and even audio.

Another example is the Information Maximizing Variational Autoencoder (InfoVAE) [51] family of autoencoders. In InfoVAE authors propose a solution to overfitting by introducing new training objectives where it is possible to weigh the preference between correct inference and fitting data distribution. A special variation of the InfoVae family is the Maximum Mean Discrepancy VAE (MMD-VAE) which uses MMD [52] loss. MMD-VAE performs better than traditional VAEs with KL-divergence loss, especially on small datasets.

Autoencoders often suffer from generating blurry images and not being able to generate a broad distribution of different images. However, autoencoders are rather stable and simple to train. Autoencoders also have the advantage of being able to precisely generate certain types of images, as one can pick the sample from the learned distribution. In the next section, we will present a deep learning method that usually performs better than autoencoders.

2.2.2 Generative Adversarial Networks

Generative adversarial networks (GAN) are a deep learning method of unsupervised learning for generating high-quality data. It can be used to generate images, video, audio, text, and much more. It was first proposed in [53] and is often called the most interesting concept in deep learning.

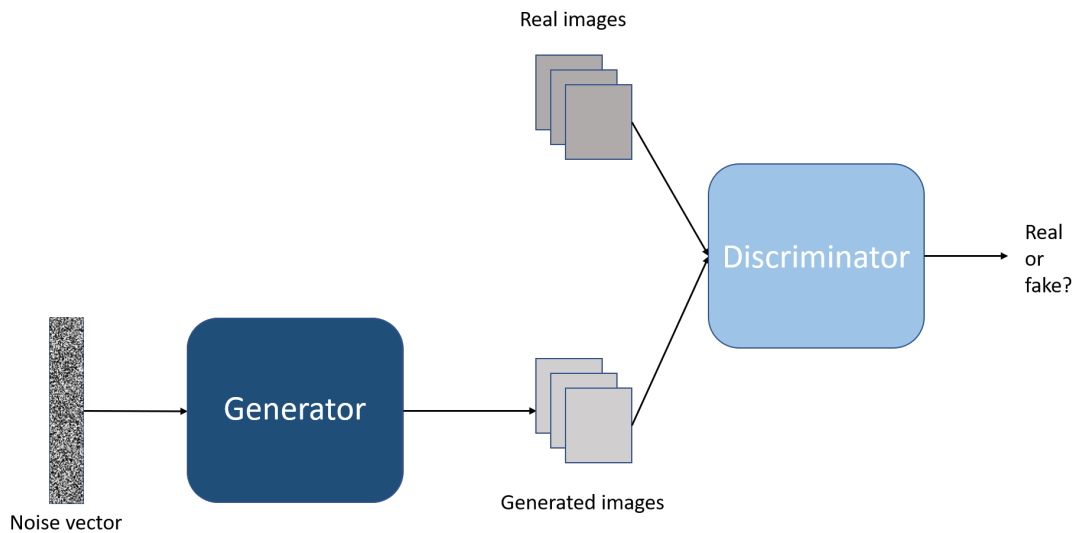


Figure 2.2: Architecture of a GAN

The basic idea behind GAN is very simple. The architecture consists of two neural networks, a generator, and a discriminator as can be seen in Figure 2.2. Generator, the neural network that generates data, usually has some random noise vector as an input. Noisy input helps generate a wide selection of images from the learned distribution. The generator's objective is to fool the discriminator, which tries to differentiate between generated and real data. The constant

ongoing rivalry between the generator and the discriminator is what makes GANs adversarial. Mathematically, discriminator and generator play a minimax game with the following function [54]:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [1 - \log(D(G(z)))] \quad (2.2)$$

where: G = the generator

D = the discriminator

x = training dataset

z = random variable

p_{data} = posterior distribution

p_z = prior distribution

The goal of the generator is to maximize the probability of discriminator labeling generated images as real samples, and the discriminator has the goal of minimizing that probability while being able to label real data as such. For image generating purposes it is convenient to use convolution operations in GAN which is presented in Deep Convolutional GAN (DCGAN) [55].

GANs usually outperform autoencoders because of their adversarial part, which enables them to generate non-blurry images. Generative adversarial networks are an interesting concept that has caught the eye of many scientists. It has been the center of extensive research in recent years. Many different families of GANs have emerged and proved to be state-of-the-art in generating artificial data.

An interesting approach with GANs are image to image translation models. They are used for style transfer between images [56], image inpainting [57] and even generating images from masks [58]. One of the examples of those models is the pix2pixGAN [58]. Pix2pix uses a conditional GAN (cGAN) with a U-net [59] based generator with skip connections and a PatchGAN discriminator. PatchGAN discriminator penalizes structure at the scale of image patches, it outputs a matrix that classifies each patch of an image as being real or not. The generator is trained with a l1 loss function and an adversarial loss. Pix2pixGAN can solve many different challenges, it can generate images with an image mask as an input to the generator or perform inpainting. Pix2pix GAN can generate realistic images even with smaller datasets such as fa-

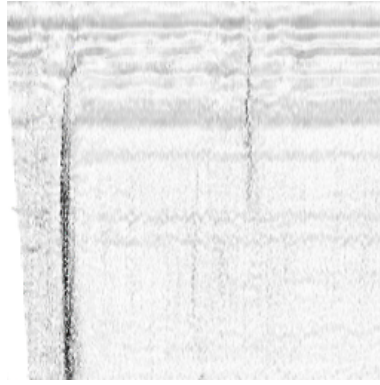


Figure 2.3: Ultrasonic B-scan generated with pix2pixGAN

grades dataset [60]. An example of an ultrasonic image generated with pix2pixGAN can be seen in Figure 2.4. While pix2pix generated images of size 256x256, a newer pix2pixHD [61] is able to generate high resolution, 2048x1024, realistic images. In pix2pixHD, authors present several improvements. It uses a revised generator which consists of two parts. First, they train a smaller network that generates low-resolution images, then they concatenate the trained small network to the bigger one, which then generates high-resolution images. They also incorporated a feature-matching loss based on the discriminator. They also use multiscale discriminators, in fact, three PatchGAN discriminators that operate on different patch sizes.

A CycleGAN [56] is a very interesting concept of Generative adversarial networks. It is made for unpaired image-to-image translation. The CycleGAN is two GANs with two generators and two discriminators wired together. The idea is to have two generators that have some image as an input, and output an image of a different domain. Then that image is run through the other generator, which has to generate an image exactly like the original one. This is called cycle consistency. The two discriminators have to determine if the images put as their input are real or generated. Therefore, there are two cycle consistency loss functions, one for each domain mapping, and an adversarial loss for each of the generators and their corresponding discriminators. CycleGAN produces images with comparable quality to pix2pixGAN, but in a fully unsupervised manner without using paired data. It has been successfully used to generate medical CT images from MR [62].

DetectorGAN [63] deals with the challenge of detecting a small object with a limited number of annotated bounding boxes. It jointly optimizes the generator model and a detector to improve the performance of a detector. Existing generative networks can generate visually appealing images, but they don't necessarily improve the ability of the detector. Authors show the proposed GAN improves the detector's average precision by 20% on an NIH X-ray dataset [64] with over a hundred thousand images. The architecture is based on a CycleGAN, the generator generates images that are then fed into a discriminator and a detector. The discriminator helps in generating realistic images, and the detector gives feedback on whether generated images im-

prove the detection. The detector, generator, and discriminator are trained during the process. The Discriminator-generator pair and the detector are pretrained for faster convergence. Both generator and discriminator are trained on adversarial losses, while the generator is also trained with cycle consistency loss. The discriminator is trained on both real and generated images.

Generating multiple objects at certain locations in an image has been a topic of [65]. Usually when generating images with GAN models use a semantic map image as an input to determine the position of the object in a generated image. Input to the generator is the text description of the generated scene and objects bounding boxes, while the discriminator gets a generated image with bounding box locations.

In [66] authors propose GAN-based methods to generate realistic synthetic skin lesion images. They evaluate the performance of a pix2pixGAN, DCGAN and a conditional version of Progressive Growing of GANs (PGAN) [67]. They have also trained a classifier on real and data generated by each of these generative networks. They have shown that a pix2pix GAN is able to generate realistic melanoma images with similar performance of the classifier as on real images. With such generative data augmentation, they have reached an improvement of 1 percentage point of a classifier when compared to training only with real images.

In [68] authors developed a 3D approach to data generation. They developed a generative network called CT-GAN which is able to generate and change 3D CT scans. CT-GAN is able to inject and remove tumors from a 3D scan in order to make a false medical reports.

A lot of work has been done for enlarging data sets in medical imagery. In [69] authors developed a multi-channel GAN (M-GAN) to generate PET images from CT scans. Similar approach with a cGAN has been made in [70, 71]. Their generated data performed only 2.8% lower in terms of recall when training a detection model with them. Generating MR images from CT scans with paired and unpaired data has been researched in [72]. An MR-GAN with the concept similar to CycleGAN has been developed for the purpose. In [73] a DCGAN has been employed to generate realistic brain MR images. Data augmentation using non-convolutional GAN was tested on three different non image datasets [74]. Generated data performed even better than real data when classifying using a Decision Tree (DT) classifier.

In [63] authors developed a GAN with the integrated detector as a discriminator for generating small objects in the non-anomalous chest X-ray images. They used the generated data to improve the performance of the RetinaNet [39]. In [75] generate realistic 3D lung nodule CT-scans which even expert physicians fail to distinguish from the real ones.

Generative adversarial networks are a great concept for generating realistic data, but are often hard to train in a stable manner. GANs tend to fall into a mode collapse where the generator finds a way to fool the discriminator by generating a single, non-realistic image. There are some ways to stabilize the training procedure, but they depend on the data and the architecture of a GAN. Also, training a GAN demands large datasets, which is opposite to the problem we are

trying to solve. Using pretrained weights while training a GAN showed to speed up the process of achieving convergence in the training [76]. A stable method to train a GAN and a proper multi-purpose architecture is yet to be found.

2.3 Anomaly detection

So far, we have introduced the concept of defect detection and localization and synthetic ultrasonic data generation. In order to train a proper assisted evaluation of ultrasonic data, we need to overcome the challenge of not having enough data. We can do it by pruning the architecture of a deep learning object detector, or we can generate enough high quality synthetic data to improve the performance of the detectors. However, while having a lot of non-defected ultrasonic data, one can not but think about going the unsupervised or self-supervised path. This is why the anomaly detection on ultrasonic data is being researched in this work. Anomaly detection in the ultrasonic imaging domain is a very difficult task, since it is hard to distinguish between the reflections of the geometry of the block being scanned and the defects. An example of ultrasonic images with and without defects can be seen in Figure 1.1.

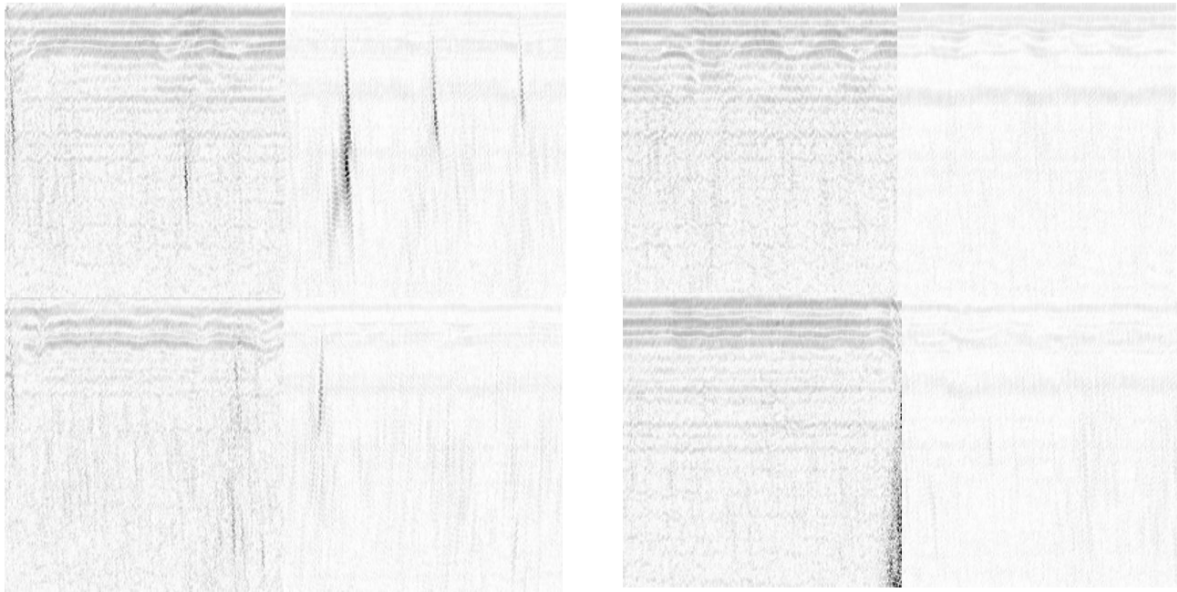


Figure 2.4: An example of anomalous (left) and normal ultrasonic images (B-scans).

Anomaly detection in non-destructive testing using the self-supervised approach is a field that yet needs to be explored. Only in [77] a form of a Variational Autoencoder (VAE), which is a go-to for anomaly detection, has been studied for ultrasonic anomaly detection. In [78] three deep learning solutions based on Generative Adversarial Networks (GANs) are compared. The authors show that results of anomaly detection yet have to be improved for real-world medical datasets. A convolutional neural network for anomaly detection in brain MRI images has been

developed in [79]. However, recently there is a good deal of research on anomaly detection in production industry. One of the popular public datasets is the MVTec AD dataset [80, 81]. The PaDiM [82] uses a ResNet [83] or any other classifier network pretrained on the ImageNet dataset. Features are extracted from the normal dataset and compared in the test time with the input image. A newer model, the PatchCore [84], reports even better results on the MVTec AD dataset. PatchCore uses the similar approach with the pretrained encoder. It also implements a *memory bank* of neighbourhood-aware patch-level features. A novel concept of normalizing flow is used in the FastFlow anomaly detection model [85]. It is currently the best performing model on the MVTec AD dataset. Normalizing flow is a set of bidirectional transformations that are able to compute the distribution of normal samples to obtain the likelihood to recognize anomalies in inference phase.

Chapter 3

Non-destructive evaluation

Non-destructive testing (NDT) is a group of analysis techniques used in science and technology to evaluate properties of materials, components, or systems without causing damage [1]. There are a variety of NDT methods used including visual testing (VT), 3D computed tomography, eddy current testing (ECT), or ultrasonic testing (UT). Among them, ultrasonic testing stands out due to its superiority in different aspects including high sensitivity to most materials' damage [2], extraction of defects location and type [2], and higher signal-to-noise ratio [3]. Ultrasonic inspection can be used for constant monitoring of various defects, as for the detection of new defects in various components of some systems. The most important task of UT is to detect all existing defects in the system and assess their size, shape, and orientation. Early detection of defects is important for the safety of the system as well as for planning the maintenance of the system accordingly. Ultrasonic testing can be divided into two parts, the acquisition and the evaluation of the acquired data.

3.1 Ultrasonic data acquisition

Acquisition of ultrasonic data is a complex procedure that is very important for the results of the inspection. It combines robotics, physics, and electronics for decoding the ultrasonic signals from the probe. An example of an ultrasonic acquisition system can be seen in Figure 3.1. Ultrasonic inspections are usually done in harsh environments with high radiation, humidity, temperature, and pressure. Due to these conditions, human experts cannot be present at the site of inspection. Furthermore, locations of the critical components being inspected are hardly accessible by a human. That is why most of the inspections are done using a robotic manipulator, as shown in Figure 3.1. Manipulators are usually remotely controlled and are connected to a computer. These manipulators are specifically designed for each inspection. Manipulators have an ultrasonic probe attached to them that they lean on the surface of the inspected block. Ultrasonic probes have to have an agent between them and the surface of the inspected material.

One of the best agents is water, and it is why both probes and the manipulators often have to be completely submerged in the water. There are different types of ultrasonic probes used that we will discuss below.

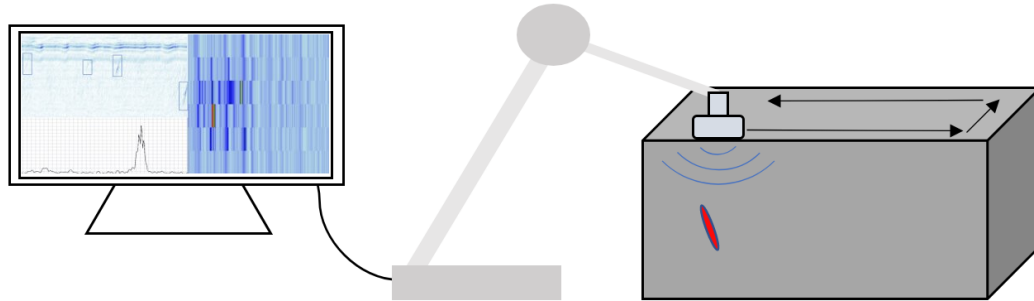


Figure 3.1: System for ultrasonic testing

There are many different divisions of ultrasonic probes. They can be divided by the angle at which they emit the ultrasonic waves, flat and angled ultrasonic probes [86]. They can also be divided into double and focused ultrasonic probes [86] or by the number of piezoelectric crystals that emit the ultrasonic wave to single element and multiple-element probes [86]. Another grouping of probes is by the frequency of emitted ultrasonic signals. Most ultrasonic probes work at frequencies between 1MHz and 10MHz, however, there are probes that work at frequencies lower than 50 kHz and higher than 200MHz. The usage of different probes depends on the inspection. Lower frequencies are lower in resolution but have a greater depth of penetration and vice versa. Ultrasonic probes are used for emitting and receiving the reflected ultrasonic waves. There are probes that have separate elements and the ones where the same element is used for generating and receiving the ultrasonic signal. The received ultrasonic signal is converted into the electric signal, which is amplified and sent to the ultrasonic receiver device. The signal is then sent to a PC application which serves to visualize the signals. Signals are then converted into more suitable formats for the ultrasonic inspection. To conclude, ultrasonic inspection is a process that can last for months, and it is very important to have high-quality ultrasonic data in order to get a proper evaluation.

3.2 Ultrasonic data evaluation

Ultrasonic data evaluation is the second phase of the ultrasonic inspection. Depending on the amount of data to be inspected, the evaluation can last anywhere from a week to a few months. Ultrasonic data evaluation is currently done by human experts (inspectors) analyzing ultrasonic data and looking for imperfections in the inspected material. In order to properly assess the data, several inspectors have to look at all of the data and document every defect, or an indication of the defect they find. To help the inspectors make the task slightly easier, inspectors have a

series of different ultrasonic data representations to look at. They include A-scans, B-scans, and C-scans. Usually, the process of the evaluation goes as follows.

First, the inspectors gather all of the available documentation of the size, shape, and material of the inspected element. With this documentation, inspectors can rule out the reflections of ultrasonic waves from the geometry of the inspected element as possible defects. In this way, they also get a better understanding of the inspected material. The path of the further inspection then depends on the element being inspected. There are different protocols for inspecting metal blocks, pipes, bolts, or welds. Below, we will focus mainly on the inspection of metal blocks. In this procedure, inspectors first look at the C-scan, the top view. There they modulate the upper and lower limit of the depth projection to roughly locate all of the defects in all depths of the material. After locating each defect, they assess them in more detail. They look at each B-scan and document the defect's depth, distance from the surface, and length across B-scans (visible on the C-scan). It is very important to properly assess found defects and their threat to the safety of the system. Not all defects are critical, and not all of them require immediate reconstruction. Some of the newly found defects require constant monitoring in order to detect when they become critical.

As the ultrasonic data evaluation can cost as much as a man-year amount of work time, a form of assisted ultrasonic inspection would be of great influence to this field. However, developing an automated defect detection from ultrasonic data is a highly complicated task. Reflections from geometry, without the use of the documentation of the inspected block, are hard to distinguish from the defects. Every inspection is different and since many different ultrasonic probes could be used, a single assisted analysis can not be developed. Also, data from the real ultrasonic inspection is hard to get by, so datasets are small and often not sufficient for deep learning solutions. Furthermore, even if the assisted analysis would be developed, certification of such software is still not officially set up. All these subjects will be addressed in the following section of the thesis.

Chapter 4

Assisted analysis

Assisted analysis of ultrasonic data is a system that combines data processing algorithms and the expert knowledge of inspectors to perform a rapid and accurate non-destructive inspection. Since non-destructive testing is a crucial part of many production and maintenance segments of the automotive, space, nuclear and other industries, assisted analysis is a topic of great research. It has already been developed for some forms of non-destructive testing, however, assisted ultrasonic testing still has to be properly developed. The development of the assisted ultrasonic evaluation has many hurdles that have to be overcome.

4.1 Data processing algorithms

The main part of the assisted analysis are the data processing algorithms. Since ultrasonic data is a group of ultrasonic signals, it is logical to use signal processing for the assisted analysis. However, since human experts analyze data by visualizing it and looking at the images of the scanned material, we use computer vision methods for the development of assisted analysis. Analyzing B-scans and C-scans instead of signals, A-scans, have the advantage of keeping the spatial information of the flaws which simplifies the detection and helps in differentiating them from the noise or from other geometry produced echos [22]. Computer vision is a highly researched topic, with many state-of-the-art convolutional deep-learning models. However, in order to develop the assisted analysis, new models have to be developed that are suitable for this specific task. As mentioned earlier, a challenge of small datasets needs to be acknowledged. We solve this challenge with three approaches. First, by developing a generative model that is able to generate high-quality synthetic ultrasonic B-scans. We then use the additional data to improve the performance of the state-of-the-art defect detectors. We also develop a novel defect detection model with a small number of parameters and a novel anchor architecture suitable for detecting objects with extreme aspect ratios. While a small number of defects is a challenge for developing supervised learning object detectors, it does not pose a problem for the self-

supervised approach. That is why we research self-supervised anomaly detection deep-learning models.

Data on which the algorithms are developed highly depend on the type of inspection being carried out. Data from the inspection of the reaction vessel of the nuclear power plant is completely different from the inspection of pipes or bolts. This is why we need different deep-learning models for each inspection. A scheme of the development of deep-learning assisted analysis algorithms is shown in Figure 4.1.

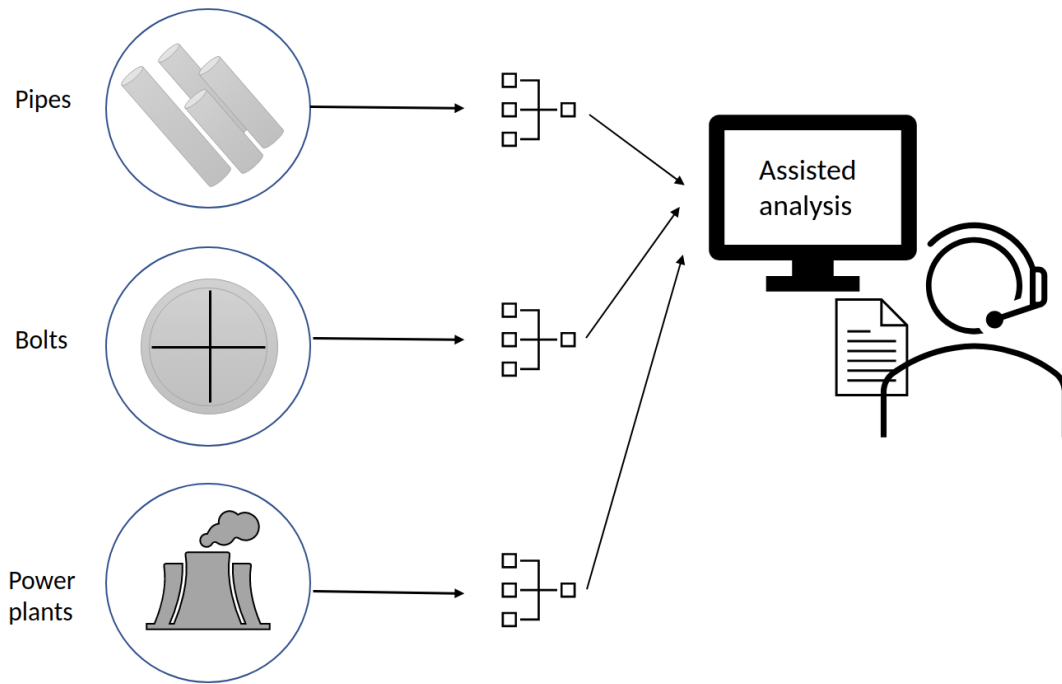


Figure 4.1: Deep-learning models for assisted analysis

The most important task of data processing algorithms in assisted analysis is to detect all defects. Further evaluation of the defect's size, shape, and importance to the structure can be left to the inspectors to evaluate. However, if inspectors could be assisted in the detection of all defects in the block, it would significantly shorten the time needed for the overall analysis. Developed algorithms need to be reliable so that inspectors could trust they would not miss a single defect. As to date, there are no certification protocols for deep-learning algorithms for non-destructive evaluation. This is why it is important to thoroughly test the algorithms. Usage of the developed deep-learning algorithms should always be taken with a grain of salt and with good knowledge of their internal work. The best-case scenario would be using the deep-learning algorithms on the same type of data from the same inspection that it was trained on. This would pose a significant improvement in the present data evaluation. For example, the first inspection of some nuclear power plants should be done by hand, data should be gathered and labeled. Every next inspection of the same power plant could be done using deep learning, since the data would be similar and algorithms would be properly tested.

4.2 Evaluation of developed algorithms

Proper evaluation of the developed algorithms is the key to safe assisted analysis and non-destructive evaluation. This is why in this thesis and all of the published works the evaluation is given great thoughts.

For evaluating the synthetic data generation, we made two experiments. We trained a deep learning defect detection network with the real data and with the synthetic data, and compared the results by testing the network on a separate real dataset. Generative networks were trained on the same dataset the defect detection network was trained on. We achieved higher performance when the defect detection network was trained using the synthetic data [24]. This shows that synthetic data is highly realistic, but even more diverse than the real dataset. It also shows the importance of the big training dataset. We also tested the synthetic data with the help of inspectors. Human experts could not discriminate the synthetic data from the real ultrasonic data. We concluded that by doing a statistical analysis of the blind test with the inspectors [87].

Evaluation of the defect detection methods, both supervised [88] and self-supervised [89] is similar. We have to prove that both methods are able to successfully detect all defects. In order to get a better evaluation of the algorithms, we made sure that we tested on each separate block, and report the results.

Chapter 5

The main scientific contributions of the thesis

The main scientific contributions of the thesis are the following: first and foremost, a deep learning-based synthetic generation method for generating high quality realistic ultrasonic data has been developed [Pub1, Pub2, Pub3]; second, a deep-learning-based defect detector method with custom parameters, trained using additional synthetic ultrasonic data generated with the deep learning-based method [Pub1]; and third, a method for defect detection in ultrasonic images using the self-supervised deep-learning approach and a new model for localization of defects with extreme aspect ratios using the supervised deep-learning approach [Pub4, Pub5].

5.1 A new deep-learning generative adversarial network

Generative adversarial networks (GANs) are one of popular the approaches to generate realistic, synthetic data. Since the initially proposed concept in 2014, many variations have been made. These state-of-the-art models are able to generate realistic images of various human portraits or streets as seen from the car. However, not much work has been done to generate images with objects in predefined locations. Furthermore, many researchers compete in generating the most realistic image possible, however, the practical usage of such technology is not clear. In the field of ultrasonic non-destructive evaluation, a dataset with a lot of images containing defects is very valuable. This data can be used for the training of deep-learning defect detectors or to train new human experts in this field.

Successfully developing the GAN network that is stable in training and doesn't suffer from mode collapse, non-convergence or diminishing gradient is a feat of its own. The main contribution of this thesis is the development of a new state-of-the-art generative network that performs better when compared to other state-of-the-art methods. We bring some innovations to the field with the new generative model.

Generating objects on predefined locations on images is a task important in order to create a whole dataset with bounding-box annotations. Using the traditional approach to GANs it is not possible to know where the objects will be in the generated images. By using an image-to-image translation problem with our GAN we make it possible. By inputting the binary image with the visible bounding box location of the object to be generated, we predetermine the desired position of the defect, but not its shape. This adds to the variety of generated defects, but also removes the need of segmenting the training dataset (labeling bounding boxes is much easier for the object detection task). However, there is no way for the generator to know if he did a good job in placing the defect in the right location. That is why we add a discriminator with spatial awareness. In this new model, we added an object detector pretrained on the same dataset as the training dataset for the GAN which guides the generator in the defect placement.

By using this new model, we generate a set of images and compare their quality to other state-of-the-art models in the field. We develop an experiment that involved human experts on ultrasonic non-destructive evaluation and a statistical evaluation of the acquired data. Our novel model outperforms other models by a significant margin.

5.2 A deep-learning based defect detector method

There have been some attempts to develop an efficient and accurate automatic or assisted defect detection algorithm. However, as mentioned before, the challenge is to acquire enough data for the training of state-of-the-art modern object detectors on the ultrasonic dataset. With our new method, we generated 200,000 labeled synthetic images and trained a new state-of-the-art model for defect detection. As the base detection model, we used the YOLOv3 model, adapted its hyperparameters like anchor parameters, and trained it using both real and synthetic labeled data. This way we once again proved the quality of our synthetic dataset and a novel GAN.

Furthermore, we explored the possibility of creating a custom architecture fit for the task of ultrasonic defect detection. This custom defect detector, which is at the moment state-of-the-art, is based on the EfficientDet detector. It has a novel, more efficient feature extractor that is trainable with a smaller training dataset. It also features a detection head adapted for the task of ultrasonic defect detection. Since defects usually have extreme aspect ratios, standard object detector anchors are not able to cover all of the image locations. With these custom anchor architecture, the challenge is solved.

5.3 A deep-learning based anomaly detection method

The third approach to the lack of labeled positive data in the training dataset is to not use it at all. By utilizing self-supervised learning, we can avoid needing the labeled defected data in the

training dataset. A lot of work has been done recently in the field of unsupervised and self-supervised learning. Anomaly detection has seen high attention from researchers. Many new anomaly detection models with state-of-the-art performance have been developed. However, hardly any research on this field has been done in the field of ultrasonic testing.

We trained a series of state-of-the-art anomaly detection models and compared their performances on our ultrasonic non-destructive testing dataset. We also compared their performance with the supervised-learning approach using the popular classifier networks. We also proposed our improvements to the best-performing Patch Distribution Modeling Framework.

Chapter 6

Conclusion and future work

6.1 The main conclusions of the thesis

These thesis conclusions can be divided into three parts that bring us to the one greater epilogue. In the first part of the thesis, we discuss the value of the data. We recognize the challenge of having to develop an accurate deep-learning solution for defect detection with only a small amount of data. As a solution to the challenge, we provide three contributions, each with its own conclusion.

First, by developing the deep-learning synthetic data generation network we find a solution to the main bottleneck of the deep-learning, the data. Our network generates high-resolution high-quality ultrasonic images with defects on predefined locations. It is the first such network for generating synthetic ultrasonic data. With multiple experiments, we demonstrate the performance of the network. First, using another deep-learning model we gain a performance boost of the defect detector using the synthetic images, which proves that these images are not only realistic but more diverse than the initial dataset of real images. Furthermore, with the help of human experts, we compared a few approaches to generating synthetic ultrasonic data. Statistical analysis of the blind test conclude that experts were not able to distinguish ultrasonic images generated with our generative network from the real images. However, they successfully distinguished between real data and other approaches to generating synthetic data. With this, we conclude that generating synthetic data in the field of ultrasonic non-destructive testing is a highly valued solution to many of the field's challenges. With this additional data, new human experts could be taught the task of the analysis of ultrasonic data. Also, this data can be used to help in training new deep-learning networks for ultrasonic data analysis.

Deep-learning with all of the research that has been done in recent years looks like a solution to all of the real-world challenges. However, modern state-of-the-art object detectors are only tested on large, common, public datasets. They perform great on these datasets, but we show that they have some deficiencies when put to test in the real-world applications. Most

objects that are found in these public datasets are common objects like vehicles, animals and people and their bounding boxes' aspect ratios are close to one. However, defects in ultrasonic data are mostly elongated and have extreme aspect ratios. Furthermore, reflections of ultrasonic waves from defects highly resemble reflections from the geometry of the inspected block, which makes it hard to train a self-supervised model on this data. Even in synthetic data generation, ultrasonic data stands out from the large public datasets. Ultrasonic data background noise is hard to generate realistically, such noise is not common in other large public datasets. By solving these deficiencies, we develop a novel deep-learning object detector suitable for detecting objects with extreme aspect ratios. Our new feature extractor for the object detector has a smaller number of parameters, so it requires a smaller training dataset. This way, we are able to enhance the performance of the defect detector while making it faster. Our second conclusion states the importance of developing usage-specific deep-learning networks for real-world applications. Although state-of-the-art popular deep-learning architectures are able to produce satisfying results, to get the extra performance that is critical for the specific field, challenges and deficiencies have to be acknowledged.

As for the third approach to solving the ultrasonic non-destructive testing data analysis challenge, we presented the self-supervised approach. Since the anomaly detection approach does not use any defective data for training, it is useful for the challenges with a small number of defective samples in the training set and a large variety of such. Many different anomaly detection models have been developed in recent times, however, none of them were tested on a real-world ultrasonic non-destructive testing dataset. We tested a series of state-of-the-art anomaly detection models and compared the results. Using the prior knowledge of the ultrasonic testing, we introduced our improvements to the models. Anomaly detection outperforms standard supervised learning classifiers when the share of positive samples in the training dataset is around a few percent of the whole training dataset. We conclude that there are circumstances where anomaly detection on ultrasonic images outperforms supervised learning, however much yet need to be done to improve the performance of these models.

We conclude that the automated or assisted analysis of the data from ultrasonic non-destructive testing is possible and has been developed as a part of this thesis. The performance of such systems should not be based on standard computer vision evaluation metrics. Systems should be tested and evaluated the same as a human expert would be, with rigorous tests to see the ability to detect all defects with a small percentage of false positives. We believe that research work and contributions that can be found in this thesis will serve as a foundation to the further development of such systems, not only in non-destructive testing but other similar science and industry fields.

6.2 Future work

When considering future research directions, one possibility is to extend the research into other types of ultrasonic inspection. In this thesis, we focused mainly on the testing of different metal blocks, but there are many other types of inspections being done. Such as, inspection of the pipelines and welds, which are two other common inspections. Different inspections include different type of data that requires a different approach than is demonstrated in this thesis.

In terms of improving the results of the synthetic data generation and defect detection, future research could be looking more into the full 3D ultrasonic data examination. Data being processed could be transformed into a full 3D view of the examined specimen, e.g. metal block. With this type of data, new deep-learning models could be developed with improved performance. However, this approach would require even more data than the one presented in this thesis, and as such would have its own set of challenges.

For assisted analysis to become more reliable and to be proven in the field of non-destructive testing, further tests and evaluations should be done. All of the developed models should be certified in order for them to be used in real inspections. A formal evaluation of the developed algorithms should be established for all of the existing algorithms and the ones that are yet to be developed. However, a human expert should always remain present at the inspection and evaluation.

Chapter 7

List of publications

- Pub 1 **L. Posilović**, D. Medak, M. Subašić, M. Budimir, S. Lončarić, "Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans", *Neurocomputing*, Vol. 459, 2021, pp. 361–369., doi: 10.1109/ISPA.2019.8868929
- Pub 2 **L. Posilović**, D. Medak, M. Subašić, M. Budimir, S. Lončarić, "Generating Ultrasonic Images Indistinguishable from Real Images Using Generative Adversarial Networks", *Ultrasonics*, Vol. 119, 2022, 106610., doi:10.1016/j.ultras.2021.106610
- Pub 3 **L. Posilović**, D. Medak, M. Subašić, M. Budimir, S. Lončarić, "Synthetic 3D Ultrasonic Scan Generation Using Optical Flow and Generative Adversarial Networks", *2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 213–218, doi:10.1109/ispa52656.2021.9552069
- Pub 4 **L. Posilović**, D. Medak, F. Milković, M. Subašić, M. Budimir, S. Lončarić, "Deep Learning-Based Anomaly Detection From Ultrasonic Images", in *Ultrasonics*
- Pub 5 D. Medak, **L. Posilović**, M. Subašić, M. Budimir, S. Lončarić, "DefectDet: a deep learning architecture for detection of defects with extreme aspect ratios in ultrasonic images", *Neurocomputing*, vol. 473, Feb. 2022, pp. 107-115, doi: 10.1016/j.neucom.2021.12.008.

Chapter 8

Author's contribution to the publications

The results presented in this thesis are based on the research carried out during the period of 2018-2022 at the University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, HR-10000 Zagreb, Croatia, as a part of the research project KK.01.2.1.01.0151 "Smart UTX", which was financially partially supported by the European Union from the European Regional Development Fund.

The thesis includes five publications written in collaboration with coauthors of the published papers. The author's contribution to each paper consists of the experiment idea, software implementation, performing the required experiments, results analysis and text writing.

[Pub1] In the paper **"Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans"** the author has proposed a novel generative adversarial network with the ability of generating objects at predefined locations. The method is used to improve the state-of-the-art deep-learning based defect detector. Using the additional synthetic data for the training set, the performance of the defect detector is improved.

[Pub2] In the paper **"Generating Ultrasonic Images Indistinguishable from Real Images Using Generative Adversarial Networks"** three approaches to generating synthetic ultrasonic non-destructive testing images are evaluated, including our own deep-learning based generative network. A test is conducted with human experts on ultrasonic data evaluation, where they assess the quality of each shown image without knowing if it is generated or not. Using the statistical student t-test analysis, results show that our generative network outperforms all other approaches by a significant margin.

[Pub3] In the paper **"Synthetic 3D Ultrasonic Scan Generation Using Optical Flow and Generative Adversarial Networks"** a model for generating 3D ultrasonic scans is proposed. Two approaches were researched, one using optical flow and the other using the GAN. Both methods were able to generate a 3D ultrasonic block by generating a series of correlated B-scans. The evaluation of the scans is

[Pub4] In the paper **"Deep Learning-Based Anomaly Detection From Ultrasonic Im-**

ages" the author has proposed a self-supervised learning approach to ultrasonic non-destructive data evaluation. Since the self-supervised approach uses only normal images, of which there is plenty, it is able to outperform state-of-the-art supervised learning classifiers when the share of positive examples is around a few percent of the whole dataset.

[Pub5] In the paper "**DefectDet: a deep learning architecture for detection of defects with extreme aspect ratios in ultrasonic images**" the author proposed a novel object detection deep-learning model for defect detection. It features a novel feature extractor and the detection head adapted for defects with extreme aspect ratios by changing the anchor distribution on the image.

Bibliography

- [1]Cartz, L., Nondestructive testing: radiography, ultrasonics, liquid penetrant, magnetic particle, eddy current. ASM International, 1995, available at: <https://books.google.hr/books?id=0spRAAAAMAAJ>
- [2]Ye, J., Ito, S., Toyama, N., “Computerized ultrasonic imaging inspection: From shallow to deep learning”, *Sensors*, Vol. 18, No. 11, 3820 Nov 2018, available at: <https://doi.org/10.3390/s18113820>
- [3]Broberg, P., “Imaging and analysis methods for automated weld inspection”, Luleå tekniska universitet, 2014.
- [4]Krautkrämer, J., Krautkrämer, H., *Ultrasonic Testing of Materials*. Springer-Verlag, 1983, available at: <https://books.google.hr/books?id=AvwrAAAIAAJ>
- [5]Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., “Imagenet: A large-scale hierarchical image database”, in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [6]Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., “Microsoft coco: Common objects in context”, in *European conference on computer vision*. Springer, 2014, pp. 740–755.
- [7]Virkkunen, I., Koskinen, T., Jessen-Juhler, O., Rinta-Aho, J., “Augmented ultrasonic data for machine learning”, *arXiv preprint arXiv:1903.11399*, 2019.
- [8]Souza, I. S., Albuquerque, M. C., de SIMAS FILHO, E. F., FARIAS, C. T., “Signal processing techniques for ultrasound automatic identification of flaws in steel welded joints—a comparative analysis”, in *18th World Conference on Nondestructive Testing*, 2012, pp. 16–20.
- [9]Munir, N., Kim, H.-J., Song, S.-J., Kang, S.-S., “Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments”, *Journal of Mechanical Science and Technology*, Vol. 32, No. 7, Jul 2018, pp. 3073–3080, available at: <https://doi.org/10.1007/s12206-018-0610-1>

- [10]Meng, M., Chua, Y. J., Wouterson, E., Ong, C. P. K., “Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks”, *Neurocomputing*, Vol. 257, 2017, pp. 128 - 135, machine Learning and Signal Processing for Big Multimedia Analysis, available at: <https://doi.org/10.1016/j.neucom.2016.11.066>
- [11]Bettayeb, F., Rachedi, T., Benbartaoui, H., “An improved automated ultrasonic nde system by wavelet and neuron networks”, *Ultrasonics*, Vol. 42, No. 1, 2004, pp. 853 - 858, proceedings of Ultrasonics International 2003, available at: <https://doi.org/10.1016/j.ultras.2004.01.064>
- [12]Sambath, S., Nagaraj, P., Selvakumar, N., “Automatic defect classification in ultrasonic ndt using artificial intelligence”, *Journal of Nondestructive Evaluation*, Vol. 30, No. 1, Mar 2011, pp. 20–28, available at: <https://doi.org/10.1007/s10921-010-0086-0>
- [13]Chen, Y., Ma, H.-W., Zhang, G.-M., “A support vector machine approach for classification of welding defects from ultrasonic signals”, *Nondestructive Testing and Evaluation*, Vol. 29, No. 3, 2014, pp. 243-254, available at: <https://doi.org/10.1080/10589759.2014.914210>
- [14]Al-Ataby, A., Al-Nuaimy, W., Brett, C., Zahran, O., “Automatic detection and classification of weld flaws in tofd data using wavelet transform and support vector machines”, *Insight - Non-Destructive Testing and Condition Monitoring*, Vol. 52, 11 2010, pp. 597-602, available at: <https://doi.org/10.1784/insi.2010.52.11.597>
- [15]Matz, V., Kreidl, M., Smid, R., “Classification of ultrasonic signals”, *International Journal of Materials*, Vol. 27, 10 2006, pp. 145-, available at: <https://doi.org/10.1504/IJMPT.2006.011267>
- [16]Khelil, M., Boudraa, M., Kechida, A., Draï, R., “Classification of Defects by the SVM Method and the Principal Component Analysis (PCA)”, available at: <https://doi.org/10.5281/zenodo.1060751> 09 2007.
- [17]Cruz, F., Filho, E. S., Albuquerque, M., Silva, I., Farias, C., Gouvêa, L., “Efficient feature selection for neural network based detection of flaws in steel welded joints using ultrasound testing”, *Ultrasonics*, Vol. 73, 2017, pp. 1 - 8, available at: <https://doi.org/10.1016/j.ultras.2016.08.017>
- [18]Guarneri, G. A., Junior, F. N., de Arruda, L., “Weld discontinuities classification using principal component analysis and support vector machine”, *XI Simpósio Brasileiro de Automação Inteligente*, 2013, pp. 2358–4483.

- [19]Veiga, J., A. de Carvalho, A., Silva, I., M. A. Rebello, J., “The use of artificial neural network in the classification of pulse-echo and tofd ultra-sonic signals”, Journal of The Brazilian Society of Mechanical Sciences and Engineering - J BRAZ SOC MECH SCI ENG, Vol. 27, 10 2005, available at: <https://doi.org/10.1590/S1678-58782005000400007>
- [20]Cygan, H., Girardi, L., Aknin, P., Simard, P., “B-scan ultrasonic image analysis for internal rail defect detection”, in World Congress on Railway Research, 10 2003.
- [21]Kechida, A., Draï, R., Guessoum, A., “Texture analysis for flaw detection in ultrasonic images”, Journal of Nondestructive Evaluation, Vol. 31, No. 2, Jun 2012, pp. 108–116, available at: <https://doi.org/10.1007/s10921-011-0126-4>
- [22]Posilović, L., Medak, D., Subašić, M., Petković, T., Budimir, M., Lončarić, S., “Flaw detection from ultrasonic images using yolo and ssd”, in 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA). IEEE, 2019, pp. 163–168.
- [23]Medak, D., Posilović, L., Subašić, M., Budimir, M., Lončarić, S., “Automated defect detection from ultrasonic images using deep learning”, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 68, No. 10, 2021, pp. 3126-3134.
- [24]Posilović, L., Medak, D., Subašić, M., Budimir, M., Lončarić, S., “Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic b-scans”, Neurocomputing, Vol. 459, 2021, pp. 361-369, available at: <https://www.sciencedirect.com/science/article/pii/S0925231221010304>
- [25]Medak, D., Posilović, L., Subašić, M., Petković, T., Budimir, M., Lončarić, S., “Rapid defect detection by merging ultrasound b-scans from different scanning angles”, in 2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA), 2021, pp. 219-224.
- [26]Kieckhefer, H., Baan, J., Mast, A., Volker, W. A., “Image processing techniques for ultrasonic inspection”, in Proc. 17th World Conference on Nondestructive Testing, Shanghai, China, 2008.
- [27]Dogandžić, A., Zhang, B., “Bayesian nde defect signal analysis”, Signal Processing, IEEE Transactions on, Vol. 55, 02 2007, pp. 372 - 378, available at: <https://doi.org/10.1109/TSP.2006.882064>
- [28]Filipović, B., Milković, F., Subašić, M., Lončarić, S., Petković, T., Budimir, M., “Automated ultrasonic testing of materials based on c-scan flaw classification”, in 2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA), 2021, pp. 230-234.

- [29]Ballard, D. H., “Modular learning in neural networks.”, in AAI, Vol. 647, 1987, pp. 279–284.
- [30]Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., “Generative adversarial networks”, *Communications of the ACM*, Vol. 63, No. 11, 2020, pp. 139–144.
- [31]Girshick, R., Donahue, J., Darrell, T., Malik, J., “Rich feature hierarchies for accurate object detection and semantic segmentation”, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.
- [32]Girshick, R., “Fast r-cnn”, in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.
- [33]Ren, S., He, K., Girshick, R., Sun, J., “Faster r-cnn: Towards real-time object detection with region proposal networks”, in *Advances in neural information processing systems*, 2015, pp. 91–99.
- [34]Redmon, J., Divvala, S., Girshick, R., Farhadi, A., “You only look once: Unified, real-time object detection”, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [35]Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A. C., “Ssd: Single shot multibox detector”, in *European conference on computer vision*. Springer, 2016, pp. 21–37.
- [36]Redmon, J., Farhadi, A., “Yolo9000: better, faster, stronger”, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.
- [37]Redmon, J., Farhadi, A., “Yolov3: An incremental improvement”, *arXiv preprint arXiv:1804.02767*, 2018.
- [38]Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y. M., “Yolov4: Optimal speed and accuracy of object detection”, *arXiv preprint arXiv:2004.10934*, 2020.
- [39]Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., “Focal loss for dense object detection”, in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [40]Tan, M., Pang, R., Le, Q. V., “Efficientdet: Scalable and efficient object detection”, *arXiv preprint arXiv:1911.09070*, 2019.

- [41]Everingham, M., Van Gool, L., Williams, C. K., Winn, J., Zisserman, A., “The pascal visual object classes (voc) challenge”, *International journal of computer vision*, Vol. 88, No. 2, 2010, pp. 303–338.
- [42]Krizhevsky, A., Hinton, G. *et al.*, “Learning multiple layers of features from tiny images”, 2009.
- [43]Shorten, C., Khoshgoftaar, T. M., “A survey on image data augmentation for deep learning”, *Journal of Big Data*, Vol. 6, No. 1, 60 2019.
- [44]Ventura, D., Warnick, S., “A theoretical foundation for inductive transfer”, Brigham Young University, College of Physical and Mathematical Sciences, 2007.
- [45]Soekhoe, D., Van Der Putten, P., Plaat, A., “On the impact of data set size in transfer learning using deep neural networks”, in *International Symposium on Intelligent Data Analysis*. Springer, 2016, pp. 50–60.
- [46]Plaut, E., “From principal subspaces to principal components with linear autoencoders”, *arXiv preprint arXiv:1804.10253*, 2018.
- [47]Charte, D., Charte, F., García, S., del Jesus, M. J., Herrera, F., “A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines”, *Information Fusion*, Vol. 44, 2018, pp. 78–96.
- [48]Kullback, S., Leibler, R. A., “On information and sufficiency”, *The annals of mathematical statistics*, Vol. 22, No. 1, 1951, pp. 79–86.
- [49]Kingma, D. P., Welling, M., “Auto-encoding variational bayes”, *arXiv preprint arXiv:1312.6114*, 2013.
- [50]van den Oord, A., Vinyals, O. *et al.*, “Neural discrete representation learning”, in *Advances in Neural Information Processing Systems*, 2017, pp. 6306–6315.
- [51]Zhao, S., Song, J., Ermon, S., “Infovae: Information maximizing variational autoencoders”, *arXiv preprint arXiv:1706.02262*, 2017.
- [52]Gretton, A., Borgwardt, K., Rasch, M., Schölkopf, B., Smola, A. J., “A kernel method for the two-sample-problem”, in *Advances in neural information processing systems*, 2007, pp. 513–520.
- [53]Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., “Generative adversarial nets”, in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

- [54]Kazeminia, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., Mukhopadhyay, A., “Gans for medical image analysis”, arXiv preprint arXiv:1809.06222, 2018.
- [55]Radford, A., Metz, L., Chintala, S., “Unsupervised representation learning with deep convolutional generative adversarial networks”, arXiv preprint arXiv:1511.06434, 2015.
- [56]Zhu, J.-Y., Park, T., Isola, P., Efros, A. A., “Unpaired image-to-image translation using cycle-consistent adversarial networks”, in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2223–2232.
- [57]Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., Catanzaro, B., “Image inpainting for irregular holes using partial convolutions”, in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 85–100.
- [58]Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., “Image-to-image translation with conditional adversarial networks”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.
- [59]Ronneberger, O., Fischer, P., Brox, T., “U-net: Convolutional networks for biomedical image segmentation”, in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.
- [60]Tyleček, R., Šára, R., “Spatial pattern templates for recognition of objects with regular structure”, in Proc. GCPR, Saarbrücken, Germany, 2013.
- [61]Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B., “High-resolution image synthesis and semantic manipulation with conditional gans”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8798–8807.
- [62]Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Xu, Z., Prince, J., “Unpaired brain mr-to-ct synthesis using a structure-constrained cyclegan”, in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, 2018, pp. 174–182.
- [63]Liu, L., Muelly, M., Deng, J., Pfister, T., Li, L.-J., “Generative modeling for small-data object detection”, in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 6073–6081.
- [64]Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R., “Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases”, in IEEE CVPR, 2017.

- [65]Hinz, T., Heinrich, S., Wermter, S., “Generating multiple objects at spatially distinct locations”, arXiv preprint arXiv:1901.00686, 2019.
- [66]Bissoto, A., Perez, F., Valle, E., Avila, S., “Skin lesion synthesis with generative adversarial networks”, in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, pp. 294–302.
- [67]Karras, T., Aila, T., Laine, S., Lehtinen, J., “Progressive growing of gans for improved quality, stability, and variation”, arXiv preprint arXiv:1710.10196, 2017.
- [68]Mirsky, Y., Mahler, T., Shelef, I., Elovici, Y., “Ct-gan: Malicious tampering of 3d medical imagery using deep learning”, in *28th {USENIX} Security Symposium ({USENIX} Security 19)*, 2019, pp. 461–478.
- [69]Bi, L., Kim, J., Kumar, A., Feng, D., Fulham, M., “Synthesis of positron emission tomography (pet) images via multi-channel generative adversarial networks (gans)”, in *molecular imaging, reconstruction and analysis of moving body organs, and stroke imaging and treatment*. Springer, 2017, pp. 43–51.
- [70]Ben-Cohen, A., Klang, E., Raskin, S. P., Amitai, M. M., Greenspan, H., “Virtual pet images from ct data using deep convolutional networks: initial results”, in *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, 2017, pp. 49–57.
- [71]Ben-Cohen, A., Klang, E., Raskin, S. P., Soffer, S., Ben-Haim, S., Konen, E., Amitai, M. M., Greenspan, H., “Cross-modality synthesis from ct to pet using fcn and gan networks for improved automated lesion detection”, *Engineering Applications of Artificial Intelligence*, Vol. 78, 2019, pp. 186–194.
- [72]Jin, C.-B., Kim, H., Liu, M., Jung, W., Joo, S., Park, E., Ahn, Y. S., Han, I. H., Lee, J. I., Cui, X., “Deep ct to mr synthesis using paired and unpaired data”, *Sensors*, Vol. 19, No. 10, 2361 2019.
- [73]Kazuhiro, K., Werner, R. A., Toriumi, F., Javadi, M. S., Pomper, M. G., Solnes, L. B., Verde, F., Higuchi, T., Rowe, S. P., “Generative adversarial networks for the creation of realistic artificial brain magnetic resonance images”, *Tomography*, Vol. 4, No. 4, 159 2018.
- [74]Tanaka, F. H. K. d. S., Aranha, C., “Data augmentation using gans”, arXiv preprint arXiv:1904.09135, 2019.
- [75]Han, C., Kitamura, Y., Kudo, A., Ichinose, A., Rundo, L., Furukawa, Y., Umemoto, K., Nakayama, H., Li, Y., “Synthesizing diverse lung nodules wherever massively: 3d multi-

- conditional gan-based ct image augmentation for object detection”, 2019 International Conference on 3D Vision (3DV), 2019, pp. 729-737.
- [76]Wang, Y., Wu, C., Herranz, L., van de Weijer, J., Gonzalez-Garcia, A., Raducanu, B., “Transferring gans: generating images from limited data”, in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 218–234.
- [77]Milković, F., Filipović, B., Subašić, M., Petković, T., Lončarić, S., Budimir, M., “Ultra-sound anomaly detection based on variational autoencoders”, in 2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA), 2021, pp. 225-229.
- [78]Počević, M., Eilertsen, G., Lundström, C., “Unsupervised anomaly detection in digital pathology using gans”, in 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021, pp. 1878-1882.
- [79]van Hespen, K. M., Zwanenburg, J. J., Dankbaar, J. W., Geerlings, M. I., Hendrikse, J., Kuijf, H. J., “An anomaly detection approach to identify chronic brain infarcts on mri”, Scientific Reports, Vol. 11, No. 1, 2021, pp. 1–10.
- [80]Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., Steger, C., “The mvtec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection”, International Journal of Computer Vision, Vol. 129, No. 4, 2021, pp. 1038–1059.
- [81]Bergmann, P., Fauser, M., Sattlegger, D., Steger, C., “Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection”, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9592–9600.
- [82]Defard, T., Setkov, A., Loesch, A., Audigier, R., “Padim: a patch distribution modeling framework for anomaly detection and localization”, in International Conference on Pattern Recognition. Springer, 2021, pp. 475–489.
- [83]He, K., Zhang, X., Ren, S., Sun, J., “Deep residual learning for image recognition”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- [84]Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P., “Towards total recall in industrial anomaly detection”, arXiv preprint arXiv:2106.08265, 2021.
- [85]Yu, J., Zheng, Y., Wang, X., Li, W., Wu, Y., Zhao, R., Wu, L., “Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows”, arXiv preprint arXiv:2111.07677, 2021.

- [86] Krstelj, V., *Ultrazvu čna kontrola: odabrana poglavlja*. Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje, 2003.
- [87] Posilović, L., Medak, D., Subašić, M., Budimir, M., Lončarić, S., “Generating ultrasonic images indistinguishable from real images using generative adversarial networks”, *Ultrasonics*, Vol. 119, 106610 2022, available at: <https://www.sciencedirect.com/science/article/pii/S0041624X21002298>
- [88] Medak, D., Posilović, L., Subašić, M., Budimir, M., Lončarić, S., “Defectdet: A deep learning architecture for detection of defects with extreme aspect ratios in ultrasonic images”, *Neurocomputing*, Vol. 473, 2022, pp. 107-115, available at: <https://www.sciencedirect.com/science/article/pii/S0925231221018464>
- [89] Posilović, L., Medak, D., Milković, F., Subašić, M., Budimir, M., Lončarić, S., “Deep learning-based anomaly detection from ultrasonic images”, *Ultrasonics*, Vol. 124, 106737 2022, available at: <https://www.sciencedirect.com/science/article/pii/S0041624X2200049X>

Publications

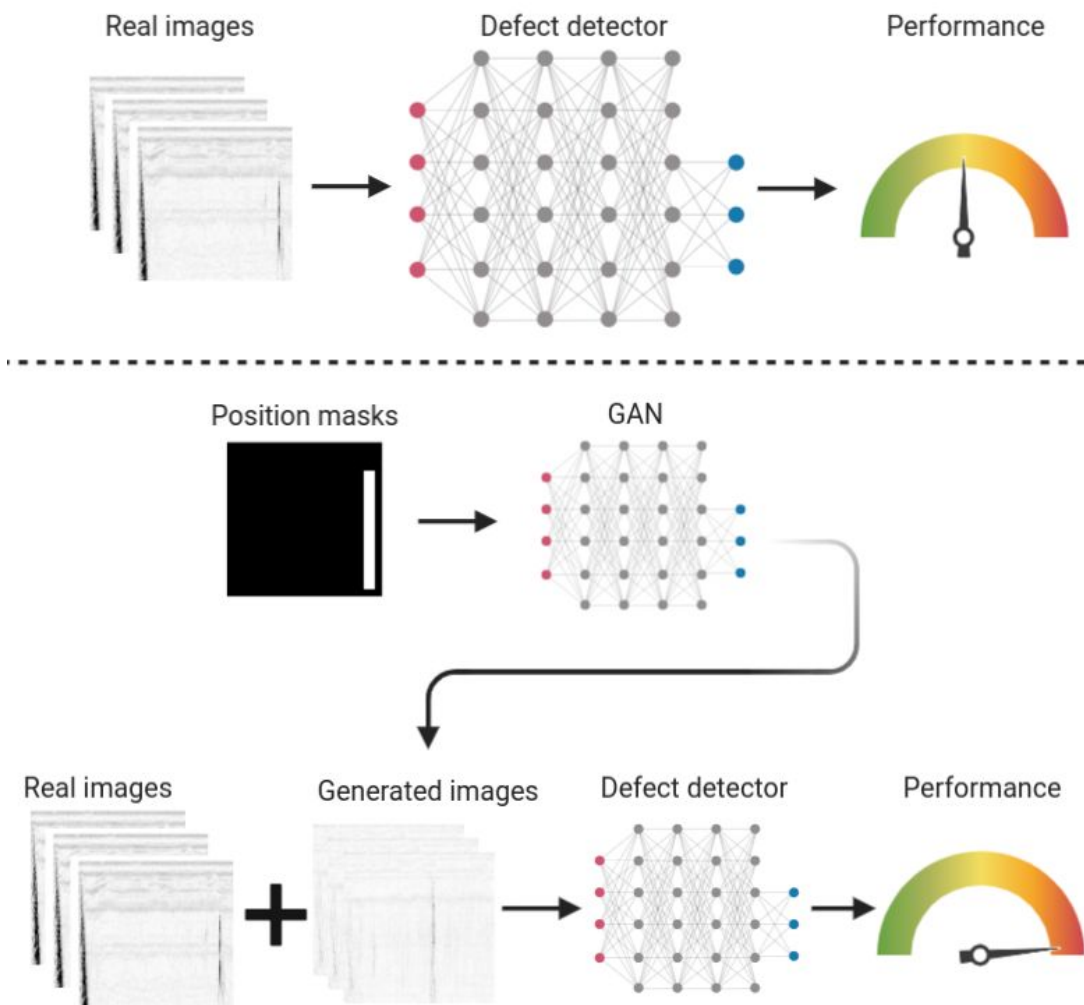
Publication 1

L. Posilović, D. Medak, M. Subašić, M. Budimir, S. Lončarić, "Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans", *Neurocomputing*, Vol. 459, 2021, pp. 361–369.

Graphical Abstract

Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans

Luka Posilović, Duje Medak, Marko Subašić, Marko Budimir, Sven Lončarić

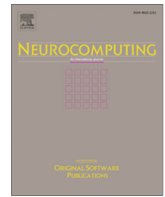


Highlights

Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans

Luka Posilović, Duje Medak, Marko Subašić, Marko Budimir, Sven Lončarić

- A novel GAN architecture for generating high quality images with objects at precise locations
- Our proposed GAN is able to generate highly realistic data that can improve the object detector's performance
- An improvement of almost 6% of average precision was achieved when training on a combined dataset of real and images generated with our GAN



Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans



Luka Posilović^{a,*}, Duje Medak^a, Marko Subašić^a, Marko Budimir^b, Sven Lončarić^a

^a University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia

^b Institute for Nuclear Technologies (INETEC), Zagreb, Croatia

ARTICLE INFO

Article history:

Received 19 November 2020

Revised 29 April 2021

Accepted 28 June 2021

Available online 2 July 2021

Communicated by Zidong Wang

Keywords:

Non-destructive testing

Ultrasonic B-scan

Automated defect detection

Image generation

Generative adversarial networks

ABSTRACT

Non-destructive testing is a set of techniques for defect detection in materials. While the set of imaging techniques is manifold, ultrasonic imaging is the one used the most. The analysis is mainly performed by human inspectors manually analyzing the acquired images. A low number of defects in real ultrasonic inspections and legal issues concerning data from such inspections make it difficult to obtain proper results from automatic ultrasonic image (B-scan) analysis. The goal of presented research is to obtain an improvement of the detection results by expanding the training data set with realistic synthetic samples. In this paper, we present a novel deep learning Generative Adversarial Network model for generating realistic ultrasonic B-scans with defects in distinct locations. Furthermore, we show that generated B-scans can be used for synthetic data augmentation, and can improve the performances of deep convolutional neural object detection networks. Our novel method was developed on a dataset with almost 4000 images and more than 6000 annotated defects. When trained only on real data, detector can achieve an average precision of 70%. By training only on generated data the results increased to 72%, and by mixing generated and real data we achieve almost 76% average precision. We believe that synthetic data generation can generalize to other tasks with limited data. It could also be used for training human personnel.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Non-destructive testing (NDT) is widely used in science and industry to evaluate properties of materials, components, or systems without causing damage [1]. Many different methods are available such as visual examination, ultrasonic, eddy current, to name a few. Among them, ultrasonic testing (UT) stands out due to its versatility. Some of the advantages of UT include high sensitivity for most of the materials [2], high signal to noise ratio [3] and the ability to precisely determine the location and the type of the defect [2]. Ultrasonic data can be represented in several different formats suitable for analysis including A, B, or C-scans [4]. An A-scan shows signal's amplitude as a function of time, B-scan displays a cross-sectional view of the inspected material, and a C-scan provides a top view of its projected features [5]. During analysis, inspectors simultaneously use multiple data representations in order to make a decision and evaluate the data.

Automated analysis has long been used in many NDT systems. However, so far it has been limited to classical decision-making algorithms such as amplitude thresholding [6]. Complex data such as the one from ultrasonic inspection makes it hard to develop an automated analysis. All ultrasonic analysis is, to the best of our knowledge, done manually by a trained human inspector. It makes ultrasonic analysis highly reliant on the inspector's experience. The automated analysis could make the process much faster and more reliable. There have been some attempts in developing an automated UT analysis [5–9], but very few of them involve using deep learning and modern deep convolutional neural networks (CNNs) on B-scans. The prerequisite for using deep learning is a large, annotated dataset. Due to a low number of flaws in real ultrasonic inspections and legal issues considering data from such inspections available data is limited. Data is the biggest drawback in the development of proper automated/assisted ultrasonic analysis. This challenge can also be found in many medical image analysis tasks [10] where, due to the rarity of some pathology and patient privacy issues, data availability is very modest. Furthermore, unlike medical datasets, there are no publicly available UT datasets.

Researchers attempt to overcome this problem by using transfer learning [11] in combination with freezing the backend CNN layers [12] which is shown to enhance the accuracy of models.

* Corresponding author.

E-mail addresses: luka.posilovic@fer.hr (L. Posilović), duje.medak@fer.hr (D. Medak), marko.subasic@fer.hr (M. Subašić), marko.budimir@inetec.hr (M. Budimir), sven.loncarić@fer.hr (S. Lončarić).

Using data augmentation is also the standard procedure for network training. However, data augmentation methods are limited and only slightly change some aspects of existing images (e.g. brightness modulation). Very limited additional information can be gained by such modifications. Synthetic data generation of high-quality images is a new type of state-of-the-art data augmentation [13]. Generative models such as generative adversarial networks (GANs) offer more variability and enrich the dataset to further improve the training process.

In this work, we present a novel GAN architecture for generating high-quality and realistic UT B-scans. Afterwards, we demonstrate that generated images can be used to train an object detection neural network to detect defects in real images. We show that images generated using our method improve the detector's average precision by more points than previous state-of-the-art augmentation techniques.

1.1. Contributions

The main contributions of this work are the following:

- a novel GAN architecture for generating high-quality ultrasonic images with objects at precise locations,
- experimental demonstration that expanding the ultrasonic dataset with generated synthetic data increases the performance of the defect detector,
- to the best of our knowledge, this is the first time a GAN is trained on ultrasonic NDT images.

1.2. Related work

Data availability is a major problem when using deep learning for defect detection. B-scans are the ideal data representation for accurately detecting defects and further estimating their depth and size. However, most authors focus on developing methods for A-scan analysis because it is easier to gather enough data. Developed algorithms for defect detection can be divided into three groups related to data representation being used; A-scans [7,8,14–23], B-scans [5,6,24,25] and C-scans [26,27]. The A-scan analysis is the most researched group of all which is also related to the data problem. Developed algorithms mostly include a combination of wavelet transform [14–19,9], discrete Fourier transform [7,21] or discrete cosine transform [21] and a support vector machine or artificial neural network classifier. B-scans keep the geometrical coherence of the defect, as can be seen in Fig. 1, which leads to a better noise invariance [24]. However, the analysis of B-scans can only be seen in a few works [5,6]. In [5] two popular deep learning object detection models, YOLOv3 [28] and SSD [29], have been used for defect detection. In [6] a deep learning classifier has been tested on augmented images, but with only three defects in the specimen block. Regarding C-scans, in [26] a method based on the comparison of the scan with a reconstructed reference image has been made. The method was able to detect all defects in their dataset, but with a high number of false-positive detections. There have also been some attempts in estimating defects from noisy measurements using Bayesian analysis [27].

There have been some attempts in using data augmentation to enlarge existing datasets. As mentioned, in [6], although only three defects were present in the test block, a copy/pasting data augmentation has been used to enlarge the dataset for training a deep learning detector. There are many variations on pasting and blending objects on the background in order to make the images look as realistic as possible. For instance, it can be done using Gaussian blur or Poisson blending [30] to smooth the edges. In [31] a comparison between different merging techniques has been made, using a combination of blending methods performed the best for

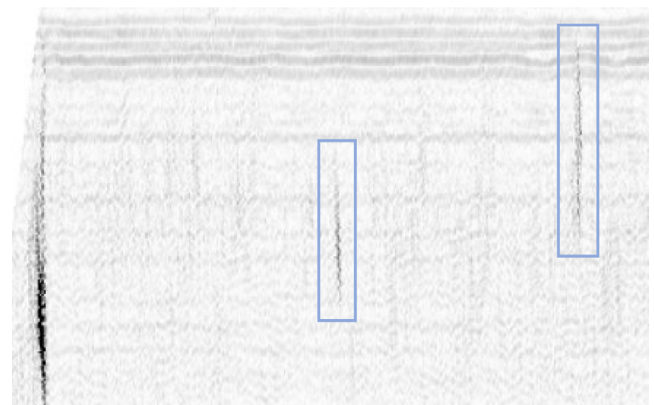


Fig. 1. Example of an ultrasonic B-scan with defects. The defects are indicated by bounding boxes.

most objects. On the other hand in [32] authors have pasted objects on random backgrounds and achieved improvements without any blending. Finally, generative adversarial networks (GANs) have recently become a popular choice for synthetic data generation and augmentation. GANs were first conceptualized in [33] in 2014. They can be used to generate images, video, audio, text, and much more. The development of the GAN came a long way in a short period of time. There are many different GAN architectures. Interesting approaches to GANs are image-to-image translation models. They are used for style transfer between images [34], image inpainting [35] and even generating images from masks [36]. One of the examples of those models is the Pix2pixGAN [36] and its successor pix2pixHD [37]. GANs show promising results in generating realistic images for human faces from noise with StyleGAN2 [38] or converting position mask images to street-view with Pix2pixHD. A lot of work has been done for enlarging data sets in medical imagery. Pix2pixHD has proved to be useful in generating skin lesion images using semantic label maps [39]. An Inception-v4-based classifier [40] has been trained using real and combined real and data generated with the Pix2pixHD. Training the classifier on a combined real and generated data achieved a 1% improvement of the area under the ROC curve. In [13] authors have applied the GAN framework to synthesize high-quality liver lesion images for improved classification. In [41] authors have developed a multi-channel GAN (M-GAN) to generate PET images from CT scans. A similar approach with a cGAN has been made in [42,43]. Using generated data, they have achieved a 28% reduction in average false positive per case. Generating MR images from CT scans with paired and unpaired data has been researched in [44]. An MR-GAN with a concept inspired by CycleGAN [34] has been developed for this purpose. In [45] a DCGAN has been employed to generate realistic brain MR images. Data augmentation using non-convolutional GAN has been tested on three different non-image datasets [46]. Generated data has performed even better than real data when classifying using a Decision Tree (DT) classifier.

1.3. Outline

The rest of the paper is organized as follows. Section 2 gives a detailed description of the used dataset. Section 3 describes the experimental procedure. Proposed GAN architecture and copy/pasting method are presented in Section 4. Results are shown in Section 5 follow by the conclusion in Section 6.

2. Dataset

The dataset was obtained by scanning six steel blocks containing artificially created defects in the internal structure. Blocks varied in size and contained between six to 34 defects. In total there were 68 defects. Blocks were scanned using INETEC Dolphin scanner with a phased array probe. An INETEC phased array ultrasound transducer with a central frequency of 2.25 MHz was used. Angles ranging from 45 degrees up to 79 degrees with a 2-degree increment were acquired during the scanning. Blocks were also scanned with a skew of zero and 180 degrees. INETEC SignyOne data acquisition and analysis software was used to process the data and create B-scans (further noted as images) that were used in the dataset. Data were converted to B-scans as-is, without pseudo-coloring, as grayscale images. All images were converted into patches of size 256x256 pixels and annotated by multiple human experts. There were in total 3825 images with a total of 6238 annotations. We split the dataset into subsets for training, validation and testing. Details of the train, validation and test subsets can be seen in Table 1. Each subset contains unique defects that do not appear in other subsets. Our dataset is highly realistic and finding all of the defects is challenging even for the human inspectors. As for the copy/pasting method, we copied the defects from the training subset in the form of the rectangle patches from the annotations and pasted them on the empty UT backgrounds. There were 3400 empty UT backgrounds from all of the blocks and 4283 defect patches from the training subset.

3. Synthetic data generation

The acquired dataset is not large enough to properly train an object detector to detect defects. For this reason, we propose two methods to expand our dataset with synthetic data.

In this section, we have described the procedure of the experiment in this work. We developed two methods for synthetic image generation and use a state-of-the-art object detector for defect detection to test the quality of the generated data. We start by describing the current state-of-the-art method for generating images and proceed to describe a deep learning approach with our GAN. We then explain the usage of the object detector in the experiment.

Our first generative method is a copy/pasting (C/P) technique. Copy/pasting is a very logical method for enlarging the ultrasonic dataset because of the large number of B-scans without defects. We call these images canvases because we paste extracted defects on them. We extracted all of the defects from the training set and pasted them on canvases in random locations. The exact method is explained in the next section. An example of an empty image canvas, extracted defect, its pseudo mask, and the resulting image can be seen in Fig. 2.

The second method we propose is our own GAN architecture for the purpose of generating UT B-scans. Our GAN is an image-to-image GAN. This means that the position mask used as the network's input is translated to a realistic B-scan with defects at specified positions. We make position masks from all annotated images in the training set. An example of an input–output pair is shown in Fig. 3. Position masks on the input of the generator serve as a location label for the desired position of the defect on the generated

image. The main novelty of our GAN is the usage of a pre-trained object detector for training the GAN. We use the object detector as an additional discriminator to provide information on the quality of the defect on the generated image when compared to the real image. It is important that the defect is positioned accurately as drawn in the position mask and that it is merged well with the background. After training the GAN we generated new position masks used for the generation of synthetic data. We determine the sizes and shapes of the defects on the position masks by extracting the aspect ratios of all annotations from the training set. Our generated images contain between one and four defects per image.

To estimate the quality of the generated images we used a popular object detector YOLOv3 [28]. This detector was already proven to work well for the task of defect detection from UT images in [5]. It is currently the state-of-the-art in defect detection. We first trained the detector using only real images and some traditional augmentations explained in the next section. We then tried training the object detector with images generated using the copy/paste method. We also trained the detector with a combination of real and generated images. Finally, we generated synthetic data with our GAN and again trained the object detector with generated images and a combination of real and generated data. Each of the trained versions of the object detector was tested on the same test dataset described in the previous section. Also, the same validation set was used in all three training variations.

4. Methods

In this section, a detailed explanation of developed methods is given. First, the copy/pasting method is described. Then the architecture of our proposed GAN is described with all of its special features. In the end, a short overview of the used object detector is given.

4.1. Copy/paste method

We used copy/paste method as a baseline to illustrate the complexity of generating synthetic data. While these images might look visually appealing, they are not of the same quality as the ones generated by the GAN.

As mentioned in Section 2 we have previously extracted defects from images in the training set. We paste them on random locations on images without visible defects. The process goes as follows. First, we randomly pick a canvas and randomly select the defect that will be pasted on it. We then put a threshold on a defect image. We make a binary pseudo mask by creating a binary image from the thresholded image and dilate it for two iterations with a 5x5 kernel. We then use the mask to extract only the defect from the initial defect patch image. We randomly select the position where we will paste the defect and calculate the compatibility of the selected defect background and the canvas on that location. We calculate the compatibility by calculating an intensity value of the background of the canvas and the defect. If these two values do not differ by more than 5%, we accept the proposed location. If these two values differ by more than 5% we try to select another location. We then select another image/canvas pair and repeat the process. For each new image, we set the limit of 100 attempts after which we just move on to generate another image. Usually, this limit is rarely reached since the right pair of canvas/defect nad location is usually found quickly. When the right pair is found, we proceed to paste the defect on the canvas. We first adapt the brightness of the defect to even further match the one from the canvas. We calculate the brightness of the location on the

Table 1
Number of images and annotations in train, validation and test subsets

	TRAIN	VALIDATION	TEST
Number of images	2278	379	1168
Number of annotations	4283	745	1210

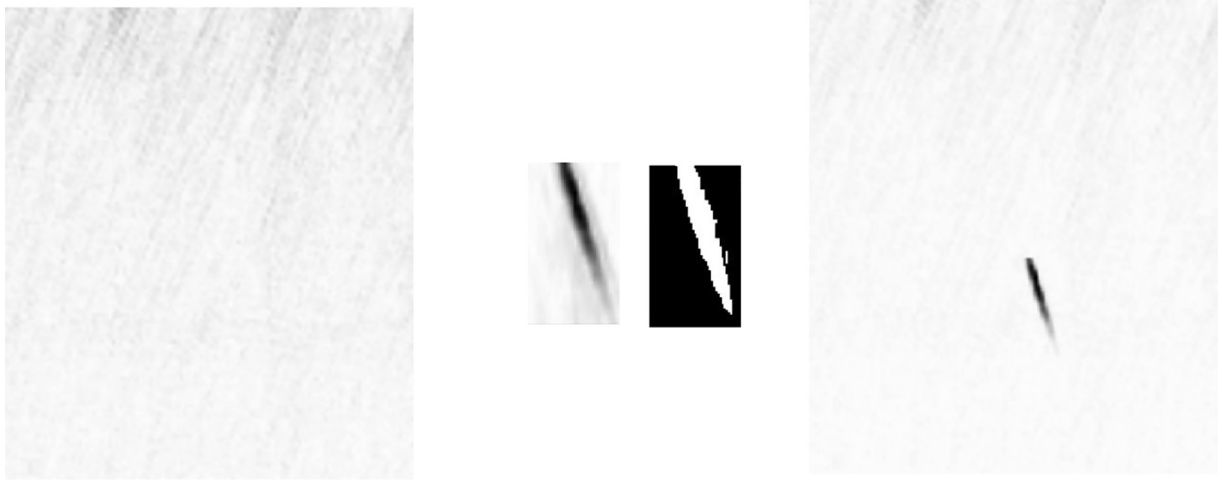


Fig. 2. Example of (from left to right) an empty canvas (B-scan without defects), an extracted defect and its binary pseudo mask, and resulting generated image.

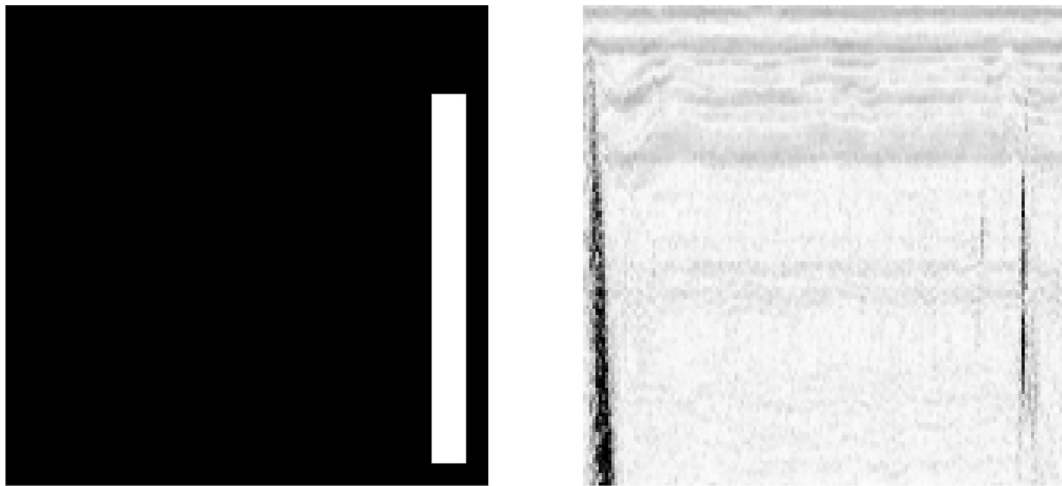


Fig. 3. Example of GAN input position mask (left) and corresponding desired output with the defect (right).

canvas and adapt the brightness of the defect to it. All that is left after that is to paste the defect. We concatenate the canvas and the defect, calculate the per pixel minimum of the two and merge it into one resulting image.

Samples of the real image, image generated with our GAN, and image generated with a copy/paste method are shown in Fig. 5.

4.2. GAN

The basic architecture of GAN is a combination of two neural networks, a generator, and a discriminator. The generator generates high-quality images from random noise. Noisy input helps generate a wide selection of images from the learnt distribution. Discriminator on the other hand tries to distinguish between generated images and the real ones. The constant rivalry between the generator and the discriminator is what makes GANs adversarial. Mathematically, discriminator and generator play a minimax game with the following function [47]:

$$\min_G \max_D V(D, G) = E_s [\log(D(s))] + E_z [1 - \log(D(G(z)))] \quad (1)$$

where : G = the generator
 D = the discriminator
 s = training sample
 z = random variable

The goal of the generator is to maximize the probability of discriminator labeling generated images as real samples and the discriminator has the goal of minimizing that probability while being able to label real data as such. This neural network configuration enables unsupervised learning of both generator and discriminator. For image generating purposes it is convenient to use convolution operations in GAN which is presented in Deep Convolutional GAN (DCGAN) [48].

We call our GAN the DetectionGAN (DetGAN) for its specific architecture. We base it on the Pix2pixHD implementing some of the features from it. Our DetectionGAN consists of a U-net generator with skip connections, two PatchGAN discriminators [36] that work on different scales and a pre-trained object detector which serves as an additional discriminator. We train the proposed GAN with image pairs of real images and their position masks. Position masks can be viewed as a conditional input of the generator and the discriminator. This version of GAN is called a conditional GAN and its objective can be express as:

$$\min_G \max_D V(D, G) = E_x [\log(D(x, s))] + E_z [1 - \log(D(x, G(x, z)))] \quad (2)$$

where: x = conditional variable

Input to the generator is the position mask defining the position of the defect. Proposed GAN does not have an input noise. The output of the generator is connected to the discriminators and the object detector. There is a total of 54,409,603 parameters in the generator. All of them are randomly initialized and trained. Unlike in Pix2pixHD, we do not use a two-stage generator nor do we upscale the position mask to generate a higher resolution image. We use skip connections with concatenation in the generator.

Discriminator has a position mask concatenated to the generated or real image as an input. Image and mask are concatenated across the channels axis. The goal of the concatenation of the position mask and the image is to provide information on the position of defects in the image for the discriminator. This concatenation leads to an improvement as shown in Section 5. Discriminator gets the real and the generated image during each step as an input. In order to discriminate images on two different scales, we use two discriminators. This way we can generate more realistic images with both coarse and fine details. Both discriminators have 1,391,554 parameters that are randomly initialized.

For the additional discriminator we use a YOLO object detector during this experiment, but any other object detector could be used. Usage of the YOLO discriminator helps the GAN with the generation of highly realistic images with defects in precise, desired locations. We input the generated image and then the real image and compare the outputs. We want these two outputs to be the same so that there is no difference between the generated and real image for the detector. This way we ensure defects are placed on the exact locations and without any artifacts. Using an object detector as a discriminator provides a significant improvement as shown in Section 5. To the best of our knowledge, this is the first time an object detector has been used as a discriminator in a GAN in order to enhance the quality of generated images.

An illustration of the proposed GAN is shown in Fig. 4. Filter sizes of each layer of the generator and discriminators are noted

in the figure. Overall, the forward pass of our model can be explained as:

$$G(x) = g$$

$$D_1(\text{concatenate}(x, r)) = d_{11}, fm_{11}$$

$$D_1(\text{concatenate}(x, g)) = d_{12}, fm_{12}$$

$$D_2(\text{downsample}(\text{concatenate}(x, r))) = d_{21}, fm_{21}$$

$$(3) \quad D_2(\text{downsample}(\text{concatenate}(x, g))) = d_{22}, fm_{22}$$

$$Y(\text{upsample}(r)) = y_1$$

$$Y(\text{upsample}(g)) = y_2$$

where: G = the generator

D_1 = the discriminator 1

D_2 = the discriminator 2

Y = the YOLO discriminator

g = generated image

x = positional mask

r = real image

d_{ij} = output of the discriminator

fm_{ij} = second to last layer of discriminator

$\text{downsample}()$ = downsampling by a factor of 2

$\text{upsample}()$ = upsampling to 416x416 px

We train our GAN using a set of loss functions. For the generator, we use four different losses. At the output of the generator, we calculate the L1 loss on the generated image and the paired real image:

$$G_{\text{loss}} = |g - r| \quad (4)$$

For propagating discriminator output to the generator we use the mean squared error loss:

$$G_{\text{dloss}} = \frac{1}{2} \sum_{k=1}^2 (d_{k1} - d_{k2})^2 \quad (5)$$

We also use the feature matching loss for training the generator, similar to the one in [37]:

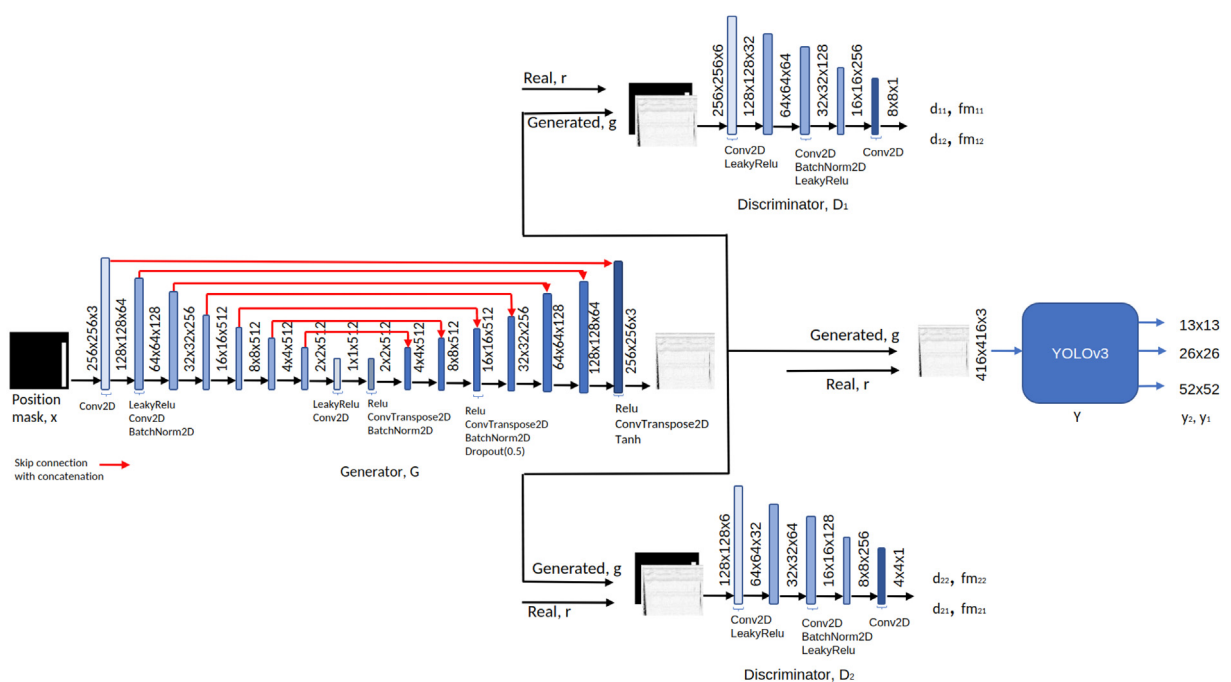


Fig. 4. Simplified architecture of our DetectionGAN.

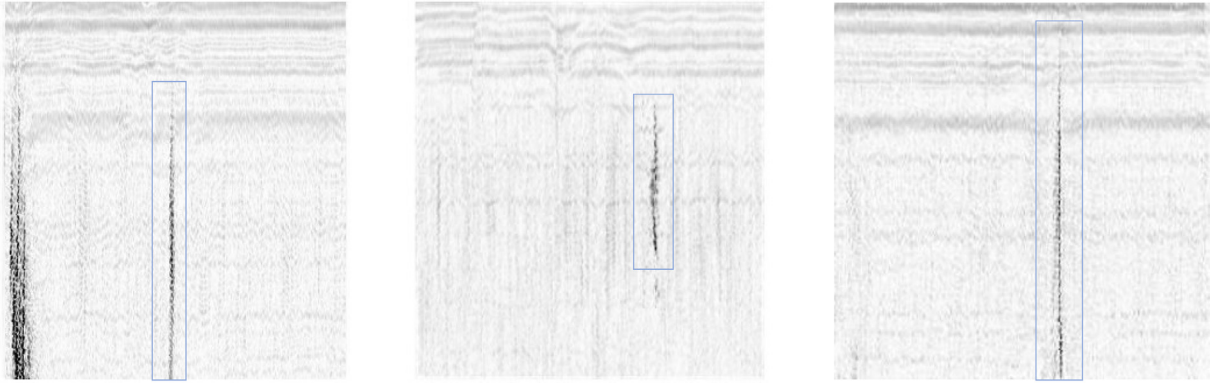


Fig. 5. Samples of (from left to right): real image, an image generated with copy/paste method, an image generated with DetectionGAN. In each image, the defects are indicated by a bounding box.

$$G_{f_{\text{loss}}} = \sum_{k=1}^2 \frac{1}{2} |f_{m_{k1}} - f_{m_{k2}}| \quad (6)$$

We also compare the output of all three scales of the object detector when inputting the real image and the generated one. We again use the L1 loss to propagate the error down to the generator:

$$G_{y_{\text{loss}}} = |y_1 - y_2| \quad (7)$$

For training the discriminator we use the mean square error loss, just like it is used in Pix2pixHD. Since we have two discriminators we have a multi-task learning problem of

$$\min_G \max_{D_1, D_2} \sum_{k=1,2} \mathcal{L}_{\text{GAN}}(G, D_k) \quad (8)$$

4.3. Object detector

Our object detector, You Only Look Once (YOLO) version 3 is taken from [5] where it proved to be able to detect defects with high average precision.

YOLOv3 is an object detector that belongs to the one-stage detector family. This means that the model directly searches for objects' presence at predefined places without performing a region proposal step. The detector consists of a backbone network, Darknet-53 [28], used to extract useful features from the image and the detection head. The detection head is used for the localization and classification of the objects. To improve invariability to objects sizes, the detection process is performed at three different scales using feature maps with resolutions: 13x13, 26x26, and 52x52. Each value of the feature map is used to perform three predictions (for objects of 3 different aspect ratios). The coordinates of the bounding boxes can then be determined by decoding the predictions. Non-maximum-suppression and object threshold are also performed after the decoding to limit the number of predicted bounding boxes and keep only the boxes that encapsulate the object the best. For the training of the GAN, we use outputs of the three mentioned feature maps from the YOLO.

The aim of this work is to improve the performance of the object detector using synthetic data. We train the object detector on real, generated data, and a combination of those two. We train the networks with the same hyperparameters in order to have a fair comparison. We first trained the detector on real data, tuned the hyperparameters to achieve the best possible performance and used the same hyperparameters for training with other data combinations. We input images of size 416x416 pixels. We used a pre-trained backbone and froze its parameters while training. Hence,

we trained only 20,974,518 of a total number of 61,576,342 parameters.

5. Experimental setup and results

5.1. Experimental setup

In this section, we describe the experiment and hyperparameters used to train our GAN and the object detector. Our experiment goes as follows. We first trained an object detector with real data. This trained network is used as the YOLO discriminator of our GAN. We generated synthetic images using the copy/pasting method and the GAN method. We generated 200,000 synthetic images with both the copy/paste and DetectionGAN method. We also trained the DetectionGAN without using the object detector discriminator and concatenation in the discriminators to compare the effectiveness of each proposed modification. For each version of the proposed GAN, we use the same position masks to generate synthetic images. We again train the object detector using the generated data and compare results.

For training the object detector we used the following configuration. We use batch size eight and Adam optimizer with a learning rate of 1e-3. Anchor hyperparameters were calculated using the K-means with Jaccard distance as proposed in [49]. Custom anchors were calculated on the training set for all of the training combinations. We slightly changed only the ignore threshold hyperparameter to 0.6 from the original YOLO implementation. We used checkpoints while training the model. An early stopping callback was used to stop the training after the validation loss didn't improve for over eight epochs. We reduced the learning rate after every two epochs with no improvement on the validation set. We also used some basic augmentations while training all of the models. Those augmentations include horizontal image flipping, random cropping, and HSV space modulation. We also tried training the object detector without augmentations. It took us around 30 min to train the object detector. For testing the object detector we used the following hyperparameters. The object threshold of YOLO was 0.001 while the non-maximum suppression threshold was 0.5 and the intersection over union threshold was 0.5. These hyperparameters are a standard for evaluating object detection challenges and were used in [5].

We train our GAN as follows. Position masks for the input of the GAN are of size 256x256 pixels, as well as the generated images. For training the generator we use an Adam optimizer with a first-moment term of 0.5, the second one of 0.999, and a learning rate of 0.0002. One of our discriminators has an input image of 256x256 pixels, while the other one has a downsampled image of

128x128 pixels as an input. We train discriminators using Adam optimizer with the same parameters as the generator.

We use the pre-trained object detector and do not train it during training the GAN. We implement a set of simple data augmentations for training the GAN including horizontal flipping, brightness modulation, and random cropping. These data augmentations enable us to achieve great results in training the GAN and generating high-quality images. With data augmentation, we expand the training subset to over 9000 images. We train the GAN for 800 epochs and for the last 100 epochs we linearly reduce the learning rate to zero. Each epoch corresponds to one pass through all the images in the training dataset. We trained with a batch size eight. The training takes around 96 h using a single NVIDIA RTX 2080Ti graphics card. Although this may seem like a long time it is similar to the training time of pix2pix [36]. Taking into consideration the number of images in the training subset, the number of epochs, and the complexity of the whole GAN it is the expected training duration.

5.2. Results and discussion

The performance of the proposed approach was tested on a test subset. As described in Section 3 we test the quality of generated images by training an object detector on real and generated images. We used an average precision (AP) metric for assessing the performance of an object detector on a test set. Each experiment with the object detector was run three times. In tables, we present the mean value and the standard deviation for each result. Generated images used in this test were not handpicked but randomly generated.

Detailed results can be seen in Table 2. We ran the training of the object detector on real data with and without data augmentation. Using the C/P method for image synthesis did not provide any improvements in the detection. When training only on C/P images we acquire a result of only 51% AP on the same test set. When training on the combination of both real and C/P images we again do not get any improvements. The reason could be that this data has some artifacts when compared to the real images. Although visually, both images generated with DetectionGAN and with the copy/paste method look realistic, the object detector tends to learn wrong features and can not converge to a better model than the one trained on real images. However, we achieved an improvement with DetectionGAN-generated images when opposed to training the object detector with only real images. An improvement of 2% has been achieved while training only on DetectionGAN-generated images, and an improvement of almost 6% of AP was achieved when training on a combined dataset of real and images generated with our GAN. As a reference, experiments with two versions of the DetectionGAN without the object detector discriminator and without position mask and image concatenation in the

discriminator were made. Both versions perform worse than our DetectionGAN.

The final score for defect detection is almost 76%, which seems rather low, but this is due to a very difficult dataset used, which is problematic even for human inspectors. Such problematic datasets are the primary target for result improvements. The presented results demonstrate that we have obtained our initial goal of improving the detection results using realistic synthetic samples produced by our generation method. The results also show that the existing generative methods are not capable of generating images of sufficient quality to improve the performance of the defect detector.

This experiment indicates that it is important to have the most realistic data as it is possible to achieve an improvement. We illustrated the complexity of the problem of generating synthetic data for training the object detector. Our proposed GAN can generate highly realistic data that can improve the object detector's performance.

6. Conclusion and future work

In this paper, we propose a novel generative adversarial network for generating highly realistic B-scans (images) from position mask images. Our DetectionGAN generates highly realistic ultrasonic images from position masks that can be used to train an object detector. We achieved an improvement of almost 6% while training on a combination of generated and real data. We also developed a copy/pasting method based on the previous state-of-the-art approach for data augmentation to compare it to our proposed method. As we didn't cherry-pick DetectionGAN-generated images all of the generated images were proven to be of high quality.

With the increasing problem of lack of data and advances in generating high-quality synthetic data, networks such as our DetectionGAN could be used in many science and industry fields. In future work, DetectionGAN should be tested using different object detectors as discriminators. Also, other state-of-the-art object detectors should be tested on this dataset.

CRediT authorship contribution statement

Luka Posilović: Conceptualization, Methodology, Software, Validation, Data curation, Writing - original draft. **Duje Medak:** Software, Data curation, Writing - review & editing. **Marko Subašić:** Conceptualization, Resources, Writing - review & editing, Supervision. **Marko Budimir:** Resources, Data curation, Writing - review & editing, Funding acquisition. **Sven Lončarić:** Resources, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was co-funded by the European Union through the European Regional Development Fund, under the grant KK.01.2.1.01.0151 (Smart UTX).

References

- [1] L. Cartz, Nondestructive testing: radiography, ultrasonics, liquid penetrant, magnetic particle, eddy current, ASM International (1995), url: <https://books.google.hr/books?id=0spRAAAAMAAJ>.

Table 2
Results of training the object detector on different training datasets

TRAINING DATA	AP (%)	# IMAGES
Real w/ augm.	70.36 σ 1.11	2,278
Real w/o augm.	47.64 σ 6.16	2,278
C/P	51.13 σ 2.72	200,000
C/P + real	68.07 σ 1.57	202,278
DetGAN w/o conc.	40.78 σ 2.35	200,000
DetGAN w/o conc. + real	67.78 σ 3.14	202,278
DetGAN w/o yolo	62.60 σ 0.85	200,000
DetGAN w/o yolo + real	69.40 σ 2.83	202,278
DetectionGAN	72.17 σ 0.16	200,000
DetectionGAN + real	75.91 σ 0.95	202,278

- [2] J. Ye, S. Ito, N. Toyama, Computerized ultrasonic imaging inspection: From shallow to deep learning, *Sensors* 18 (11) (2018) 3820, <https://doi.org/10.3390/s18113820>.
- [3] P. Broberg, Imaging and analysis methods for automated weld inspection, Ph. D. thesis, Luleå tekniska universitet (2014)..
- [4] J. Krautkrämer, H. Krautkrämer, *Ultrasonic Testing of Materials*, Springer-Verlag (1983), [url:https://books.google.hr/books?id=AvwrAAAAIAAJ](https://books.google.hr/books?id=AvwrAAAAIAAJ).
- [5] L. Posilović, D. Medak, M. Subašić, T. Petković, M. Budimir, S. Lončarić, Flaw detection from ultrasonic images using yolo and ssd, in: 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), IEEE, 2019, pp. 163–168.
- [6] I. Virkkunen, T. Koskinen, O. Jessen-Juhler, J. Rinta-Aho, Augmented ultrasonic data for machine learning, *Journal of Nondestructive Evaluation* 40 (1) (2021) 1–11.
- [7] I.S. Souza, M.C. Albuquerque, E.F. de SIMAS FILHO, C.T. FARIAS, Signal processing techniques for ultrasound automatic identification of flaws in steel welded joints—a comparative analysis, in: 18th World Conference on Nondestructive Testing, 2012, pp. 16–20..
- [8] N. Munir, H.-J. Kim, S.-J. Song, S.-S. Kang, Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments, *Journal of Mechanical Science and Technology* 32 (7) (2018) 3073–3080, <https://doi.org/10.1007/s12206-018-0610-1>.
- [9] M. Meng, Y.J. Chua, E. Wouterson, C.P.K. Ong, Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks, *Neurocomputing* 257 (2017) 128–135.
- [10] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, *Journal of Big Data* 6 (1) (2019) 60.
- [11] D. Ventura, S. Warnick, A theoretical foundation for inductive transfer, Brigham Young University, College of Physical and Mathematical Sciences..
- [12] D. Soekhoe, P. Van Der Putten, A. Plaat, On the impact of data set size in transfer learning using deep neural networks, in: International Symposium on Intelligent Data Analysis, Springer, 2016, pp. 50–60.
- [13] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification, *Neurocomputing* 321 (2018) 321–331.
- [14] F. Bettayeb, T. Rachedi, H. Benbartaoui, An improved automated ultrasonic nde system by wavelet and neuron networks, *Ultrasonics* 42 (1) (2004) 853–858, *Proceedings of Ultrasonics International* 2003..
- [15] S. Sambath, P. Nagaraj, N. Selvakumar, Automatic defect classification in ultrasonic ndt using artificial intelligence, *Journal of Nondestructive Evaluation* 30 (1) (2011) 20–28.
- [16] Y. Chen, H.-W. Ma, G.-M. Zhang, A support vector machine approach for classification of welding defects from ultrasonic signals, *Nondestructive Testing and Evaluation* 29 (3) (2014) 243–254, [arXiv:https://doi.org/10.1080/10589759.2014.914210](https://doi.org/10.1080/10589759.2014.914210).
- [17] A. Al-Ataby, W. Al-Nuaimy, C. Brett, O. Zahran, Automatic detection and classification of weld flaws in tofd data using wavelet transform and support vector machines, *Insight – Non-Destructive Testing and Condition Monitoring* 52 (2010) 597–602, <https://doi.org/10.1784/insi.2010.52.11.597>.
- [18] V. Matz, M. Kreidl, R. Smid, Classification of ultrasonic signals, *International Journal of Materials* 27 (2006) 145, <https://doi.org/10.1504/IJMT.2006.011267>.
- [19] M. Khelil, M. Boudraa, A. Kechida, R. Draï, Classification of Defects by the SVM Method and the Principal Component Analysis (PCA) (09 2007). doi:10.5281/zenodo.1060751..
- [20] M. Meng, Y.J. Chua, E. Wouterson, C.P.K. Ong, Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks, *Neurocomputing* 257 (2017) 128–135, *machine Learning and Signal Processing for Big Multimedia Analysis*. doi: 10.1016/j.neucom.2016.11.066..
- [21] F. Cruz, E.S. Filho, M. Albuquerque, I. Silva, C. Farias, L. Gouvêa, Efficient feature selection for neural network based detection of flaws in steel welded joints using ultrasound testing, *Ultrasonics* 73 (2017) 1–8, <https://doi.org/10.1016/j.ultras.2016.08.017>.
- [22] G.A. Guarneri, F.N. Junior, L. de Arruda, Weld discontinuities classification using principal component analysis and support vector machine, *XI Simpósio Brasileiro de Automação Inteligente* (2013) 2358–4483.
- [23] J. Veiga, A.A. de Carvalho, I. Silva, J.M.A. Rebello, The use of artificial neural network in the classification of pulse-echo and tofd ultra-sonic signals, *Journal of The Brazilian Society of Mechanical Sciences and Engineering – J BRAZ SOC MECH SCI ENG* 27. doi:10.1590/S1678-58782005000400007..
- [24] H. Cygan, L. Girardi, P. Akin, P. Simard, B-scan ultrasonic image analysis for internal rail defect detection, in: World Congress on Railway Research, 2003..
- [25] A. Kechida, R. Draï, A. Guessoum, Texture analysis for flaw detection in ultrasonic images, *Journal of Nondestructive Evaluation* 31 (2) (2012) 108–116, <https://doi.org/10.1007/s10921-011-0126-4>.
- [26] H. Kieckhefer, J. Baan, A. Mast, W.A. Volker, Image processing techniques for ultrasonic inspection, in: Proc. 17th World Conference on Nondestructive Testing, Shanghai, China, 2008..
- [27] A. Dogandzic, B. Zhang, Bayesian nde defect signal analysis, *Signal Processing, IEEE Transactions on* 55 (2007) 372–378, <https://doi.org/10.1109/TSP.2006.882064>.
- [28] J. Redmon, A. Farhadi, Yolov3: An incremental improvementCite arxiv:1804.02767Comment: Tech Report. url:https://arxiv.org/abs/1804.02767.
- [29] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S.E. Reed, C. Fu, A.C. Berg, SSD: single shot multibox detector, *CoRR* abs/1512.02325. arXiv:1512.02325, doi:10.1007/978-3-319-46448-0_2..
- [30] P. Pérez, M. Gangnet, A. Blake, Poisson image editing, *ACM SIGGRAPH 2003 Papers* (2003) 313–318.
- [31] D. Dwibedi, I. Misra, M. Hebert, Cut, paste and learn: Surprisingly easy synthesis for instance detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 1301–1310.
- [32] J. Rao, J. Zhang, Cut and paste: Generate artificial labels for object detection, in: Proceedings of the International Conference on Video and Image Processing, 2017, pp. 29–33.
- [33] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680..
- [34] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.
- [35] G. Liu, F.A. Reda, K.J. Shih, T.-C. Wang, A. Tao, B. Catanzaro, Image inpainting for irregular holes using partial convolutions, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 85–100.
- [36] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.
- [37] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, B. Catanzaro, High-resolution image synthesis and semantic manipulation with conditional gans, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8798–8807.
- [38] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, T. Aila, Analyzing and improving the image quality of StyleGAN, in: Proc. CVPR, 2020..
- [39] A. Bissoto, F. Perez, E. Valle, S. Avila, Skin lesion synthesis with generative adversarial networks, in: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, Springer, 2018, pp. 294–302..
- [40] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, *arXiv preprint arXiv:1602.07261*..
- [41] L. Bi, J. Kim, A. Kumar, D. Feng, M. Fulham, Synthesis of positron emission tomography (pet) images via multi-channel generative adversarial networks (gans), in: molecular imaging, reconstruction and analysis of moving body organs, and stroke imaging and treatment, Springer, 2017, pp. 43–51..
- [42] A. Ben-Cohen, E. Klang, S.P. Raskin, M.M. Amitai, H. Greenspan, Virtual pet images from ct data using deep convolutional networks: initial results, in: International Workshop on Simulation and Synthesis in Medical Imaging, Springer, 2017, pp. 49–57.
- [43] A. Ben-Cohen, E. Klang, S.P. Raskin, S. Soffer, S. Ben-Haim, E. Konen, M.M. Amitai, H. Greenspan, Cross-modality synthesis from ct to pet using fcnet and gan networks for improved automated lesion detection, *Engineering Applications of Artificial Intelligence* 78 (2019) 186–194.
- [44] C.-B. Jin, H. Kim, M. Liu, W. Jung, S. Joo, E. Park, Y.S. Ahn, I.H. Han, J.I. Lee, X. Cui, Deep ct to mr synthesis using paired and unpaired data, *Sensors* 19 (10) (2019) 2361.
- [45] K. Kazuhiro, R.A. Werner, F. Toriumi, M.S. Javadi, M.G. Pomper, L.B. Solnes, F. Verde, T. Higuchi, S.P. Rowe, Generative adversarial networks for the creation of realistic artificial brain magnetic resonance images, *Tomography* 4 (4) (2018) 159.
- [46] F.H.K. d. S. Tanaka, C. Aranha, Data augmentation using gans, *arXiv preprint arXiv:1904.09135*..
- [47] S. Kazemina, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, A. Mukhopadhyay, Gans for medical image analysis, *arXiv preprint arXiv:1809.06222*..
- [48] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, *arXiv preprint arXiv:1511.06434*..
- [49] J. Redmon, A. Farhadi, Yolov3: Better, faster, stronger, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517–6525, <https://doi.org/10.1109/CVPR.2017.690>.



Luka Posilović received his M.Sc. from the University of Zagreb, Faculty of Electrical Engineering and Computing in 2019. He is currently working as a young researcher in an Image Processing Group in the Department of Electronic Systems and Information Processing and working on his Ph.D. at the same University. His research interests include visual quality control, deep learning object detection, and synthetic image generation.



Duje Medak received his M.Sc. from the University of Zagreb, Faculty of Electrical Engineering and Computing in 2019. He is currently pursuing a Ph.D. degree in the same faculty while working as a researcher in the Image Processing Group in the Department of Electronic Systems and Information Processing. His research interests include image processing, image analysis, machine learning, and deep learning. His current research interest includes deep learning object detection methods and their application in the non-destructive testing (NDT) domain.



Marko Subašić is an associate professor at the Department for Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, the University of Zagreb, and has been working there since 1999. He received his Ph.D. degree from the Faculty of Electrical Engineering and Computing, University of Zagreb, in 2007. His field of research is image processing and analysis and neural networks with a particular interest in image segmentation, detection techniques, and deep learning.



Marko Budimir received his M.Sc. of physics at the University of Zagreb, Faculty of Science in 2000., and his Ph.D. at Ecole Polytechnique Federale de Lausanne in Switzerland in 2006. He has worked at EPFL since 2006. till 2008. From 2008. he is working at the Institute of Nuclear Technology (INETEC). He coordinated many key projects at INETEC and although he is a key person in a company in the industry sector he is still working close to the field of science.



Dr. Sven Lončarić is a professor of electrical engineering and computer science at the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia. As a Fulbright scholar, he received a Ph.D. degree in electrical engineering from the University of Cincinnati, OH in 1994. From 2001–2003, he was an assistant professor at the New Jersey Institute of Technology, USA. His areas of research interest are image processing and computer vision. He was the principal investigator on a number of R&D projects. Prof. Lončarić co-authored more than 250 publications in scientific journals and conferences. He is the director of the Center for Computer Vision at the University of Zagreb and the head of the Image Processing Group. He is a co-director of the Center of Excellence in Data Science and Cooperative Systems. Prof. Lončarić was the Chair of the IEEE Croatia Section. He is a senior member of IEEE and a member of the Croatian Academy of Technical Sciences. Prof. Lončarić received several awards for his scientific and professional work.

Publication 2

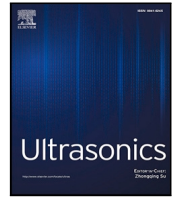
L. Posilović, D. Medak, M. Subašić, M. Budimir, S. Lončarić, "Generating Ultrasonic Images Indistinguishable from Real Images Using Generative Adversarial Networks", *Ultrasonics*, Vol. 119, 2022, 106610.

Highlights

Generating Ultrasonic Images Indistinguishable from Real Images Using Generative Adversarial Networks

Luka Posilović, Duje Medak, Marko Subašić, Marko Budimir, Sven Lončarić

- Our GAN is able to generate images indistinguishable from real ultrasonic images.
- The most thorough statistical quality analysis to-date is shown in this work.
- Evaluation of generated images has been conducted with the participation of experts.
- Inspectors have not been able to distinguish between real and generated images.
- Generated synthetic ultrasonic images can be used for education of inspectors.



Generating ultrasonic images indistinguishable from real images using Generative Adversarial Networks

Luka Posilović^{a,*}, Duje Medak^a, Marko Subašić^a, Marko Budimir^b, Sven Lončarić^a

^a University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia

^b INETEC Institute for Nuclear Technology, Zagreb, Croatia

ARTICLE INFO

Keywords:

Non-destructive testing
Ultrasonic testing
Synthetic data generation
Generative Adversarial Network
Deep learning

ABSTRACT

Ultrasonic imaging is widely used for non-destructive evaluation in various industry applications. Early detection of defects in materials is the key to keeping the integrity of inspected structures. Currently, there have been some attempts to develop models for automated defect detection on ultrasonic data. To push the performance of these models even further more data is needed to train deep convolutional neural networks. A lot of data is also needed for training human experts. However, gathering a sufficient amount of data for training is a challenge due to the rare occurrence of defects in real inspection scenarios. This is why inspection results heavily depend on the inspector's previous experience. To overcome these challenges, we propose the use of Generative Adversarial Networks for generating realistic ultrasonic images. To the best of our knowledge, this work is the first one to show that a Generative Adversarial Network is able to generate images indistinguishable from real ultrasonic images. The most thorough statistical quality analysis to date of generated ultrasonic images has been conducted with the participation of human expert inspectors. The experimental results show that images generated using our Generative Adversarial Network provide the highest quality images compared to other published methods.

1. Introduction

Defect detection is the crucial part of non-destructive evaluation (NDE) required for properly assessing the integrity of the inspected structure. However, detecting defects and estimating their position, shape, and size is a tiring task for every human inspector. The inspector's ability to detect defects is highly correlated with the inspector's previous experience in inspections of such materials. NDE is a set of techniques for the evaluation of the integrity of inspected materials without causing any damage to them [1]. Popular methods include visual inspection, eddy current, and ultrasonic to name a few. Ultrasonic inspection stands out due to its ability to accurately determine the size and position of defects both shallow and deep in the material [2]. It can be used to inspect large facilities such as nuclear power plants, pipelines, and railway tracks.

In order to improve the speed and reliability of inspections and therefore reduce their price, an automated analysis should be developed. Even more than human experts, automated systems require large amounts of data for training. This can be a challenge since ultrasonic NDE datasets are not widely available. Furthermore, data from field inspections are often protected under non-disclosure agreements and

should be destroyed after the inspection. Only available data used for training novice human inspectors and automated defect detection systems are the ones from calibration and test blocks with artificial defects generated in them. These blocks come with a high production complexity and price, therefore there is always a limited supply of them. Even with many such blocks, it is only possible to gather a limited number of images, which limits the deep learning approach. That is why there have only been a few attempts to develop the automated ultrasonic analysis using deep learning [3,4]. It would be of great value if there would be a way to generate an unlimited amount of ultrasonic data with defects on previously defined locations. To tackle similar challenges some generative methods were already researched, but not much work was done in the field of NDE.

In [5] authors used a finite element simulation method to generate synthetic ultrasonic data. Extracting defect signals and implanting them into other scan data without defects was researched in [4]. Mentioned methods describe a traditional approach to generating synthetic data. In recent times, the introduction of Generative Adversarial Networks (GANs) revolutionized the field of generating synthetic data. First published in [6] it was described as a system of two networks, a generator,

* Corresponding author.

E-mail addresses: luka.posilovic@fer.hr (L. Posilović), duje.medak@fer.hr (D. Medak), marko.subasic@fer.hr (M. Subašić), marko.budimir@inetec.hr (M. Budimir), sven.loncagic@fer.hr (S. Lončarić).

<https://doi.org/10.1016/j.ultras.2021.106610>

Received 15 July 2021; Received in revised form 30 September 2021; Accepted 30 September 2021

Available online 27 October 2021

0041-624X/© 2021 Elsevier B.V. All rights reserved.



Fig. 1. An example of a B-scan (left) from which we extracted the defect patch (with a blue bounding box), a magnified extracted defect (middle), and an example of a UT B-scan without a defect.

and discriminator, where the generator produces realistic images from input noise, and the discriminator tries to distinguish between generated and real images. Many different architectures and changes were added to the first concept of GAN during years of research. GANs are very successful in generating realistic synthetic images, videos, audio, etc. They were proven to be a huge hit in generating medical images. Since medical image analysis tackles similar problems to NDE methods, like detection, segmentation, and synthetic data generation, developed methods can be inter-exchangeable and comparable. In [7] authors compared a series of GANs in a skin lesion generation task. They have shown that a combination of real and data generated with PGAN [8] and pix2pixHD [9] in a training set leads to an increase of the area under the ROC curve when classifying data lesions as malignant or non-malignant. In [10] authors used CycleGAN [11] to generate B-mode musculoskeletal ultrasound images. Generating synthetic MR images using GAN was researched in [12], an expert physician was unable to distinguish generated images from the real ones. However, generated images were of only 64x64 and 128x128 px resolution. The generation of CT-like 3D visualization from biplanar X-rays using a GAN was done in [13]. Vox2Vox GAN [14] is able to generate 3D segmentation maps of MR brain images.

As mentioned, there are many different concepts of GANs for generating synthetic images. An interesting concept is an image to image translation models. They are used for style transfer between images [11], image inpainting [15] and even generating images from masks [16]. One of the examples of those models is the Pix2pixGAN [16] and its successor pix2pixHD [9]. A further modification to the pix2pixHD model is the SPADE GAN [17]. When training GANs using Wasserstein [18] or Hinge [19] loss generated images are more diverse than in the case of using classic GAN losses in training the generator [20]. An approach to generating 3D objects is made in [21].

In this paper, we present three methods for generating high-quality ultrasonic B-scans with defects based on real data. The first method is a traditional approach to generating images, we call it the Copy/Paste method. A method similar to this one was seen in [4] where generated images were used to improve the classifier. The other two methods are two different deep learning Generative Adversarial Networks. One of them is presented in [22] where it showed to improve the performance of the deep learning defect detector and the other is a newer GAN built upon the first one. In this work human experts in non-destructive evaluation of ultrasonic data assess the quality of the generated images using all three methods. We bring the comparison between the quality of generated images with these three different methods and conclude by highlighting the best performing method.

1.1. Outline

This paper is organized as follows. Section 2.1 gives a detailed description of the dataset to develop generative methods. In the following

Section 2.2 we explain the developed generative methods. Section 3 describes the experimental procedure. Results are shown in Section 4 and a conclusion is given in Section 5.

2. Materials and methods

2.1. Dataset

We obtained our dataset by scanning six steel blocks containing defects in the internal structure. These defects were artificially created. Blocks varied in size and contained between six and 34 defects. There were 68 different defects in total. The ultrasonic acquisition was performed using INETEC Dolphin scanned with a phased array probe. The used probe is the INETEC phased array ultrasonic transducer with a central frequency of 2.25MHz. The phased array transducer simultaneously scans from an angle of 45 to 79 degrees with a two-degree increment. Blocks were also scanned in two orthogonal directions. Also, scanning was performed with a probe skew of zero and 180 degrees. Ultrasonic analysis was done using a compatible INETEC SignyOne data acquisition and analysis software. It was used to process the data and create Volume-corrected B-scans (VC-B scans), further noted as images. Ultrasonic data were converted to images without the use of pseudo-coloring. These grayscale images were then annotated by multiple human experts. Defects were sometimes visible on multiple B-scans taken from different angles and sometimes skews. All visible appearances of defects in images were annotated. Annotated images were cut into patches of size 256x256 px. Finally, there were 3825 images with a total of 6238 annotations. These images formed our dataset.

The Copy/Pasting method on the other hand requires a different form of data. From the same dataset, we copied the defects in the form of the rectangle patches from the annotations. We used these extracted defects to paste them on UT B-scans without defects. There were 3400 empty UT images from all of the blocks. We manually filtered out the extracted defects that were not extracted well due to high background noise or were hardly visible. Finally, we obtained 4283 defect patches.

2.2. Methods

In this work, we performed a series of experiments on generating synthetic ultrasonic images with defects. The goal of this work is to generate realistic images that could serve for training new human experts for conducting ultrasonic NDE inspections. We tested three different methods for generating synthetic images. Two of them involve a deep learning approach using GANs and one describes a more traditional approach.

The first method, a traditional Copy/Paste one [22], uses ultrasonic images without defects for generating new images. We extracted defects

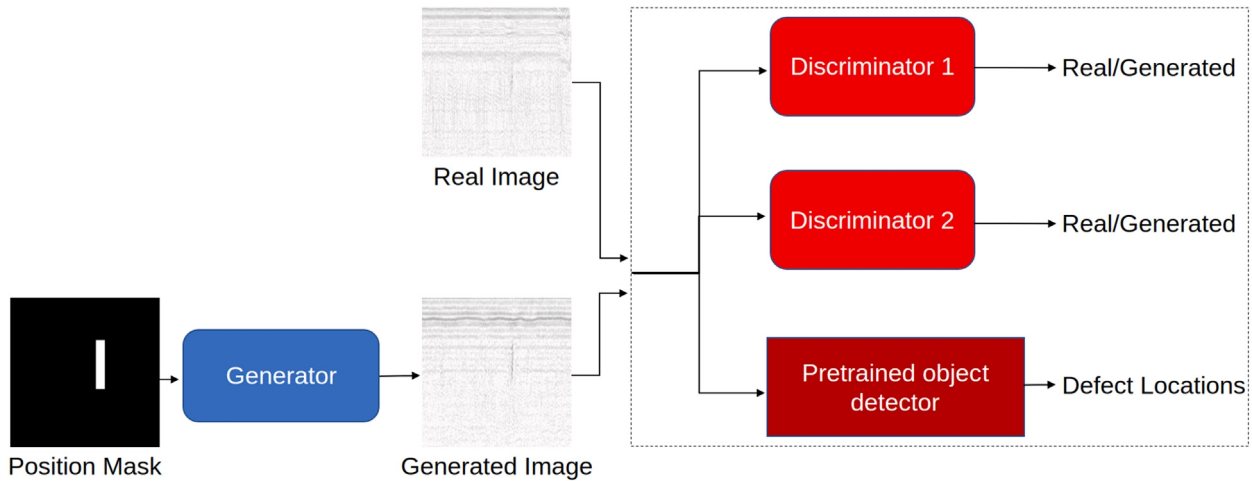


Fig. 2. A scheme of the main concept of the proposed Deep learning Generative networks.

from the images as patches whose locations were defined in the annotations as bounding boxes. These cut-out patches make our dataset of extracted defects. We also collected a dataset of images without defects. In real inspections, there are many such images. We paste the extracted defect onto empty images we call canvases. This way we can create multiple different images by permuting all of the possible indications, canvases, and positions of the indication on the canvases. An example of the base B-scan from which we extracted a defect, an extracted and processed defect, and a canvas on which we could paste the extracted defect is shown in Fig. 1.

The following two methods are based on a deep learning concept of Generative Adversarial Network (GAN). These methods consist of a neural network called the *Generator* which generates realistic images and *Discriminators* that try to distinguish between the generated and the sample of a real image. A concept of the proposed deep generative networks is shown in Fig. 2.

The second method is the DetectionGAN [22] model based upon the pix2pixHD. This model already proved to perform well in generating synthetic ultrasonic scans and enhancing the performance of the object detector for detecting defects. It is an image-to-image GAN. It means that at the input of the generator we put a binary position mask that serves to determine the desired position of the defect in the synthetic image. This position mask carries the information about the size of the generated defects and their position on the scan. By altering the aspect ratio and the position of the bounding boxes in the position mask we can generate different kinds of defects. This information can be used as a ground truth for the evaluation of the automated defect detectors or for testing the human experts. An example of the real image and a co-responding position mask can be seen in Fig. 3. The position mask is then translated to a realistic B-scan. DetectionGAN consists of a U-net generator with skip connections and two PatchGAN discriminators [16] that work on different scales. It utilizes a YOLOv3 [23] object detector pretrained on our UT dataset as an additional discriminator. YOLOv3 has been shown to perform well on defect detection from ultrasonic images [24].

The third method is SPADE GAN [17]. It is based on pix2pixHD with the addition of the encoder for generating noise for the input to the generator. It utilizes the spatially adaptive normalization in the generator. We again modified the architecture by adding the pretrained YOLOv3 discriminator. For the input, we also have the binary position mask and expect an imagined B-scan as an output. With this GAN we can determine the style of the generated image and get a non-deterministic output of the model. This way, even with the same position mask, we can get a different B-scan, which enables us to generate a more diverse set of images. The main difference between the SPADE GAN and DetectionGAN is the generator. The generator in

SPADE GAN is fed with a noise generated by an encoder. The encoder generates noise that tries to capture the style of the reference image. The generator with the noise as an input and position mask to each SPADE layer generates a highly realistic image with a certain style and predefined position of defects. This way we can generate images with some background noise style and visible block geometry. The encoder is built of six conv2D layers with stride two. The generator is made of several ResNet blocks with upsampling layers. The modulation parameters of each normalization layer are learned using SPADE. In each normalization block, we feed the downsampled position mask. The generator and the encoder together make a Variational Autoencoder (VAE).

For testing the quality of each generation method we perform a variation of a Visual Turing Test (VTT) [25]. The test is performed in three stages and each serves to assess the quality of synthetic images. The tests are developed in the form of a website in which each human expert (inspector) logs in with hers/his credentials, reads the instructions, and starts the test. In the first stage of the test, inspectors are shown randomly picked series of generated and real images, inspector's task is to evaluate each image's probability of being real assigning a value from one to five with an increment of one. Real and generated images are shown randomly to the inspector. Inspectors are not given any other information except the shown images. The goal is that inspectors cannot conclude whether the image is real or generated from any other factor except the visual quality of an image. In the second test, a series of only generated images are shown. The inspector's task is to assign a grade of quality for the generated image. In this test, the inspector knows that all of the shown images are generated. The goal of the third test is to evaluate the quality and position of generated defects in images. The inspectors are shown a series of randomly picked real and generated images and had to put bounding boxes around defects shown in the image. We perform a statistical t-student test [26] to evaluate the differences between a group of real and generated images. We perform this test for each generation method.

3. Experimental setup

In this section, a description of the training of deep learning models is given. We train two GANs and describe the Copy/Paste method. We also describe the statistical analysis of the generated image quality assessment.

3.1. Copy/paste method

As mentioned, the idea for the C/P method is partially taken from [4], but the exact implementation differs and is thoroughly

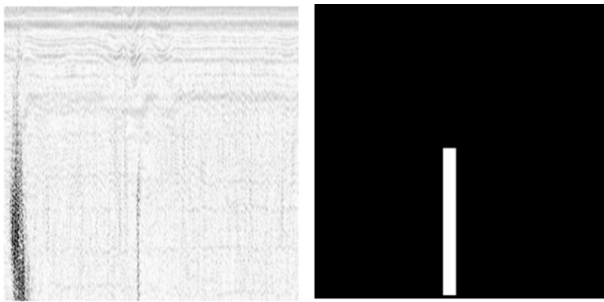


Fig. 3. A sample of a real UT B-scan (left) and a corresponding position mask (right).

explained in [22]. The procedure goes as follows. We randomly pick an empty image (canvas) and randomly select the defect we would like to paste on it. We then threshold the defect patch image. We make a binary pseudo mask and dilate it for two iterations in order to extract the whole defect from the initial patch image. We randomly select a position on the canvas to paste the selected and extracted defect on, and calculate the compatibility of the chosen position and selected defect. We calculate the average intensity value of the pixels at the destination position on the canvas and the background of the defect. If these two values do not differ by more than 5% we accept the proposed location, otherwise select another image/canvas pair and repeat the process. If we found the right pair, we proceed to paste the defect on the canvas. We first adapt the brightness of the defect to the position on canvas to even further match the pair. Finally, we concatenate the canvas and the defect, calculate the per pixel minimum of the two and merge it into the resulting image.

3.2. DetectionGAN

The DetectionGAN is based on the Pix2pixHD implementing some of the features from it. It is described in [22]. We train this model with image pairs of real images and their position masks. Position masks can be viewed as a conditional input of the generator and the discriminator. Both generated images and position masks are sizes of 256x256 px. There is a total of 54,409,603 parameters in the generator. All of them are randomly initialized and trained. Two discriminators are used, one for input images of the original size, and the other for input images downscaled by a factor of two. Discriminator's inputs are real and generated images in each step during training. Both discriminators have 1,391,554 parameters and all are randomly initialized. A pretrained YOLO object detector is used as an additional discriminator in the DetectionGAN. A pair of generated and real images are inputted to the object detector and outputs are compared.

DetectionGAN is trained using a set of loss functions. At the output of the generator, the L1 loss is calculated on the generated image and the paired real image. Also, the feature matching loss for training the generator is used. The output of all three scales of the object detector is compared when inputting the real image and the generated one. Again the L1 loss is used to propagate the error down to the generator. For training the discriminator the mean square error loss is used. Object detector is pretrained and its weights are not updated during training this GAN.

3.3. SPADE GAN

SPADE GAN is also based on pix2pixHD. We input the SPADE GAN with the same position masks as in the DetectionGAN. Both input position masks and generated images are 256x256 px resolution. The encoder of this GAN has 10,463,424 parameters, and all are randomly initialized. Generator has 114,119,875 trainable parameters. Since the same discriminators are used as in DetectionGAN, they both have 1,391,554 parameters.

We train the encoder using KL-divergence loss. We train the generator the same way as in DetectionGAN except using hinge loss instead of L1 loss. For discriminator, we again use two PatchGAN discriminators and train them the same way as in DetectionGAN. Regarding object detection discriminator we use the same pretrained YOLO object detector with the same loss functions.

3.4. Model's training

Both of the GANs are trained as follows. Position masks for the input of the GANs are of size 256x256 pixels, as well as the generated images. For training the generator we use an Adam optimizer with a first-moment term of 0.5, the second one of 0.999, and a learning rate of $2e-3$. One of our discriminators has an input image of 256x256 pixels, while the other one has a downscaled image of 128x128 pixels as an input. We train discriminators using Adam optimizer with the same parameters as the generator.

We use the pre-trained object detector and do not train it during training the GAN. We implement a set of simple data augmentations for training the GAN including horizontal flipping, brightness modulation, and random cropping. These data augmentations enable us to achieve great results in training the GAN and generating high-quality images. With data augmentation, we expand the training subset to over 9000 images. We train the GAN for 100 epochs and for the last 10 epochs we linearly reduce the learning rate to zero. Each epoch corresponds to one pass through all the images in the training dataset. We trained the DetectionGAN with a batch size eight, and SPADE GAN with a batch size six. We trained the DetectionGAN on a single NVIDIA RTX 2080Ti graphics card, and for the SPADE GAN, we use three of such GPUs.

3.5. Statistical evaluation

In this work we evaluate proposed methods by performing a test with highly trained human experts. To the best of our knowledge this is the first time such thorough statistical quality analysis of synthetic ultrasonic data generation methods has been done. Six experts were involved in this work. One with ultrasonic level one [27], three with ultrasonic level two [27], and two with the ultrasonic level 3 [27,28]. The test is performed in three phases. Experts are shown a series of random real and generated images one after the other and are asked to vote if the currently shown image is a real scan or a generated one. The scale of possible votes for the realism of the shown image range from one to five with an increment of one, where one marks a totally unrealistic, and five marks a completely realistic image. In the second phase, experts are again shown randomly real and generated images and are asked to draw bounding boxes around defects. In the third and final phase, experts are shown only generated scans with prior knowledge that all shown images are generated. They are asked to grade the quality of the shown image in the range from one to five. In this final phase the goal is to get the subjective opinion of the human experts of the quality of generated images. In the first phase we show 50 images, 25 real and 25 generated. In the second phase, we show 10 generated and 10 real images, and in the final third phase, we show 20 synthetic images. This test is performed for each of the proposed generation methods. We generated 200,000 images with each method and use these images as a pool from which we pick an image and show it to the inspector. Since so many images were generated we could say that to each expert different real and generated images are shown. Also, since we randomly pick from our pool of generated images for each proposed method, we did not handpick the best-generated images which could affect the evaluation. Although our models generate grayscale images, inspectors are used to seeing artificially colored data in real inspections. Because of that we artificially colored all images, real and generated with all three methods, using the same custom color map. Examples of artificially colored real and images generated with all three methods are shown in Fig. 5.

Table 1

Results of the statistical analysis, first two columns are the results of the first test phase, and the third column is for the third test phase. Average grades for real and images generated with each method are shown. Also, p -value and an average subjective grade is shown.

Method	Avg grade for generated/real images (1-5)	p -value (%)	Avg subjective grade (1-5)
Copy/Paste	3.56/3.85	2	3.7
DetectionGAN	3.90/3.93	41	3.6
Modified SPADE GAN	3.69/4.23	0.4	3.3

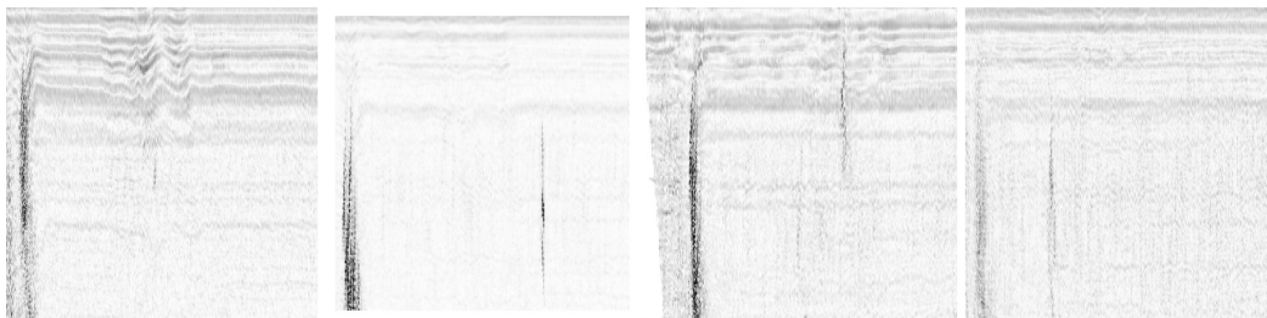


Fig. 4. Samples of (from left to right) real ultrasonic image, synthetic image generated using the copy/pasting method, DetectionGAN and SPADEGAN.

After completing the test, we perform the student t -test [26] on the results in order to assess if there is a difference between the votes on real and generated images. If there would be significant differences we could claim that the shown generation method is not of sufficient quality. We calculate the t -value and the p -value of significance. We also present the average subjective realism grade from the third phase of the test.

4. Results and discussion

All three of our developed methods are able to generate high-quality images. Samples of the real image, image generated with the copy/paste method, and the two GANs are shown in Fig. 4. Generated ultrasonic images highly resemble real B-scans. There are no artifacts in these images. Generated images picture similar geometry, similar noise, and the same defects shown in the real data the models were trained on. With these methods, we can simulate the materials and geometry of blocks similar to the original data. With the C/P method, we can not achieve high diversity of generated images and the resemblance from the original data since this method uses original real data to generate a new one. However, with DetectionGAN and SPADE GAN, we can generate a diverse set of images. Even more, with SPADE GAN we can predetermine not only the position of the defect in the scan, but we can also determine the style of the generated image.

Even though images generated by all methods look appealing, results of the evaluation of the generated data quality with human experts vary from method to method. The performance of each of the proposed methods was assessed by the p -value metric. A higher p -value means a higher probability of the two sets of images, real and generated, being of the same kind, or from the same set. Results for each of the methods are shown in Table 1.

Human experts can easier distinguish images generated with the Copy/Paste method from the real images. However, they tend to give high-quality grades to the images generated with the Copy/Paste method. This might be because they are familiar with the images and defects used in this method to generate new samples. Since deep learning methods can generate a rather diverse set of images they sometimes do not resemble any of the images human experts have previously seen. This could also be a good sign that these images could be used to train the experts. Images generated with deep learning methods are generated from scratch and still achieve great results. The DetectionGAN performs the best of all tested methods. With the

highest p -value, we conclude that inspectors are not able to detect any differences between real images and the ones generated using the DetectionGAN model.

5. Conclusions and future work

In this paper, we tested three methods for ultrasonic image generation. Two of them were custom deep generative networks, and one is a traditional approach to image generation. We performed an evaluation of the three methods with human experts. We confirmed that images generated with our custom deep learning generative network can not be distinguished from the real ultrasonic images.

With the ever-increasing need for more data in image analysis and the high price of training human experts using physical metal blocks with implemented defects, methods for generating new ultrasonic data are highly valuable. Using the proposed approach, images can be generated for the education of human experts and for training deep neural networks for ultrasonic image analysis.

CRedit authorship contribution statement

Luka Posilović: Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft. **Duje Medak:** Software, Data curation, Writing – review and editing. **Marko Subašić:** Conceptualization, Resources, Writing – review and editing, Supervision. **Marko Budimir:** Resources, Data curation, Writing - review and editing, Funding acquisition. **Sven Lončarić:** Resources, Writing - review and editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was co-funded by the European Union through the European Regional Development Fund, under the grant KK.01.2.1.01.0151 (Smart UTX). We would like to offer our special thanks to the inspectors, Nikola Babić, Pierre-Adrien Itty, Dražen novak, Leonardo Trupinić and Alojzije Matoković from INETEC Ltd. involved in this work. We thank Domagoj Čevd from ETH Zurich for helping us with the statistical evaluation.

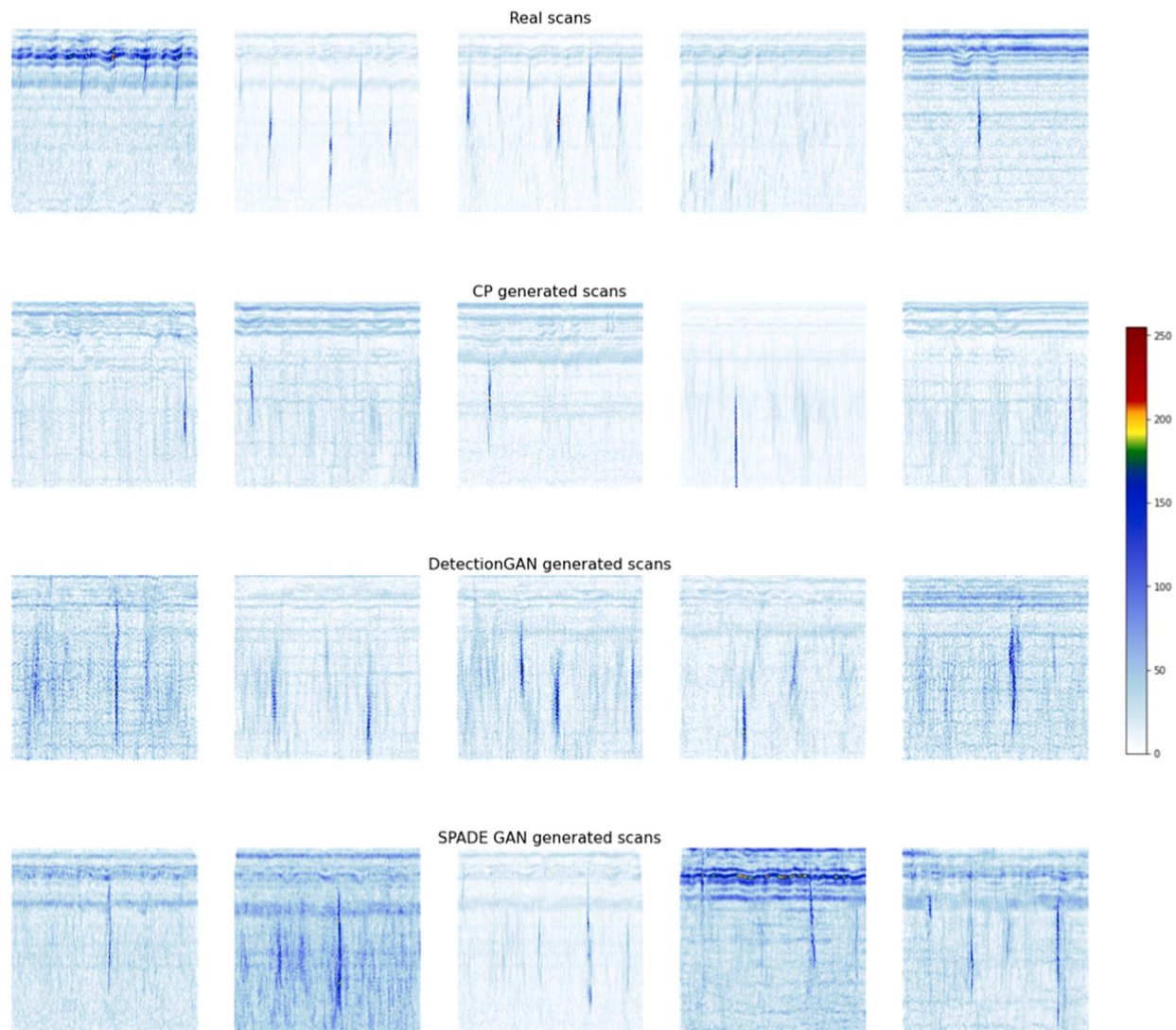


Fig. 5. Examples of artificially colored images shown to the inspectors.

References

- [1] L. Cartz, *Nondestructive Testing: Radiography, Ultrasonics, Liquid Penetrant, Magnetic Particle, Eddy Current*, ASM International, 1995, URL <https://books.google.hr/books?id=OspRAAAAMAAJ>.
- [2] J. Ye, S. Ito, N. Toyama, Computerized ultrasonic imaging inspection: From shallow to deep learning, *Sensors* 18 (11) (2018) 3820, <https://doi.org/10.3390/s18113820>, URL <https://doi.org/10.3390/s18113820>.
- [3] D. Medak, L. Posilović, M. Subašić, M. Budimir, S. Lončarić, Automated defect detection from ultrasonic images using deep learning, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* (2021).
- [4] I. Virkkunen, T. Koskinen, O. Jessen-Juhler, J. Rinta-Aho, Augmented ultrasonic data for machine learning, *J. Nondestruct. Eval.* 40 (1) (2021) 1–11.
- [5] S. Liu, J.H. Huang, J. Sung, C. Lee, Detection of cracks using neural networks and computational mechanics, *Comput. Methods Appl. Mech. Engrg.* 191 (25) (2002) 2831–2845, [https://doi.org/10.1016/S0045-7825\(02\)00221-9](https://doi.org/10.1016/S0045-7825(02)00221-9), URL <http://www.sciencedirect.com/science/article/pii/S0045782502002219>.
- [6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.
- [7] A. Bissoto, F. Perez, E. Valle, S. Avila, Skin lesion synthesis with generative adversarial networks, in: *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*, Springer, 2018, pp. 294–302.
- [8] T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of gans for improved quality, stability, and variation, 2017, arXiv preprint [arXiv:1710.10196](https://arxiv.org/abs/1710.10196).
- [9] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, B. Catanzaro, High-resolution image synthesis and semantic manipulation with conditional gans, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8798–8807.
- [10] N.J. Cronin, T. Finni, O. Seynnes, Using deep learning to generate synthetic B-mode musculoskeletal ultrasound images, *Comput. Methods Programs Biomed.* 196 (2020) 105583, <https://doi.org/10.1016/j.cmpb.2020.105583>, URL <http://www.sciencedirect.com/science/article/pii/S0169260720314164>.
- [11] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.
- [12] C. Han, H. Hayashi, L. Rundo, R. Araki, W. Shimoda, S. Muramatsu, Y. Furukawa, G. Mauri, H. Nakayama, GAN-based synthetic brain MR image generation, in: *2018 IEEE 15th International Symposium on Biomedical Imaging, ISBI 2018*, 2018, pp. 734–738, <https://doi.org/10.1109/ISBI.2018.8363678>.
- [13] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, Y. Zheng, X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10619–10628.
- [14] M.D. Cirillo, D. Abramian, A. Eklund, Vox2Vox: 3D-GAN for brain tumour segmentation, 2020, arXiv preprint [arXiv:2003.13653](https://arxiv.org/abs/2003.13653).
- [15] G. Liu, F.A. Reda, K.J. Shih, T.-C. Wang, A. Tao, B. Catanzaro, Image inpainting for irregular holes using partial convolutions, in: *Proceedings of the European Conference on Computer Vision, ECCV*, 2018, pp. 85–100.
- [16] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.
- [17] T. Park, M.-Y. Liu, T.-C. Wang, J.-Y. Zhu, Semantic image synthesis with spatially-adaptive normalization, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

- [18]M. Arjovsky, S. Chintala, L. Bottou, Wasserstein gan, 2017, arXiv preprint [arXiv:1701.07875](https://arxiv.org/abs/1701.07875).
- [19]J.H. Lim, J.C. Ye, Geometric gan, 2017, arXiv preprint [arXiv:1705.02894](https://arxiv.org/abs/1705.02894).
- [20]H.-W. Dong, Y.-H. Yang, Towards a deeper understanding of adversarial losses, 2019, arXiv preprint [arXiv:1901.08753](https://arxiv.org/abs/1901.08753).
- [21]J. Wu, C. Zhang, T. Xue, W.T. Freeman, J.B. Tenenbaum, Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling, in: *Advances in Neural Information Processing Systems*, 2016, pp. 82–90.
- [22]L. Posilović, D. Medak, M.S. sić, M. Budimir, S. Lončarić, Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans, *Neurocomputing* (2021) <https://doi.org/10.1016/j.neucom.2021.06.094>, URL <https://www.sciencedirect.com/science/article/pii/S0925231221010304>.
- [23]J. Redmon, A. Farhadi, YOLOv3: An Incremental Improvement, cite arxiv:1804.02767Comment: Tech Report, 2018, URL <https://arxiv.org/abs/1804.02767>.
- [24]L. Posilović, D. Medak, M. Subašić, T. Petković, M. Budimir, S. Lončarić, Flaw detection from ultrasonic images using YOLO and SSD, in: *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, IEEE, 2019, pp. 163–168.
- [25]D. Geman, S. Geman, N. Hallonquist, L. Younes, Visual turing test for computer vision systems, *Proc. Natl. Acad. Sci.* 112 (12) (2015) 3618–3623.
- [26]Student, The probable error of a mean, *Biometrika* (1908) 1–25.
- [27]American Society For Nondestructive Testing, Standard for Qualification and Certification of Nondestructive Testing Personnel, Standard, (CP-189) American National Standards Institute, 2020.
- [28]ISO Central Secretary, Non-Destructive Testing — Qualification and Certification of NDT Personnel, Standard, (ISO 9712:2012) International Organization for Standardization, Geneva, CH, 2012.



Luka Posilović received his M.Sc. from the University of Zagreb, Faculty of Electrical Engineering and Computing in 2019. He is currently working as a young researcher in an Image Processing Group in the Department of Electronic Systems and Information Processing and working on his Ph.D. at the same University. His research interests include visual quality control, deep learning object detection, and synthetic image generation.



Duje Medak received his M.Sc. from the University of Zagreb, Faculty of Electrical Engineering and Computing in 2019. He is currently pursuing a Ph.D. degree in the same faculty while working as a researcher in the Image Processing Group in the Department of Electronic Systems and Information Processing. His research interests include image processing, image analysis, machine learning, and deep learning. His current research interest includes deep learning object detection methods and their application in the non-destructive evaluation (NDE) domain.



Marko Subašić received the Ph.D. degree from the Faculty of Electrical Engineering and Computing, University of Zagreb, in 2007. Since 1999, he has been working at the Department for Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb, where he is currently working as an Associate Professor. He teaches several courses at the graduate and undergraduate levels. His research interests include image processing and analysis and neural networks, with a particular interest in image segmentation, detection techniques, and deep learning. He is a member of the IEEE of Computer Society, the Croatian Center for Computer Vision, the Croatian Society for Biomedical Engineering and Medical Physics, and the Centre of Research Excellence for Data Science and Advanced Cooperative Systems.



Marko Budimir received his M.Sc. of physics at the University of Zagreb, Faculty of Science in 2000., and his Ph.D. at Ecole Polytechnique Federale de Lausanne in Switzerland in 2006. He worked at EPFL from 2006. till 2008. From 2008. he is working at the Institute of Nuclear Technology (INETEC). He coordinated many key projects at INETEC and although he is a key person in a company of industry sector he is still working close to the field of science.



Dr. Sven Lončarić is a professor of electrical engineering and computer science at the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia. As a Fulbright scholar, he received a Ph.D. degree in electrical engineering from the University of Cincinnati, OH in 1994. From 2001–2003, he was an assistant professor at the New Jersey Institute of Technology, USA. His areas of research interest are image processing and computer vision. He was the principal investigator on a number of R&D projects. Prof. Lončarić co-authored more than 250 publications in scientific journals and conferences. He is the director of the Center for Computer Vision at the University of Zagreb and the head of the Image Processing Group. He is a co-director of the Center of Excellence in Data Science and Cooperative Systems. Prof. Lončarić was the Chair of the IEEE Croatia Section. He is a senior member of IEEE and a member of the Croatian Academy of Technical Sciences. Prof. Lončarić received several awards for his scientific and professional work.

Publication 3

L. Posilović, D. Medak, M. Subašić, M. Budimir, S. Lončarić, "Synthetic 3D Ultrasonic Scan Generation Using Optical Flow and Generative Adversarial Networks", *2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 213–218

Synthetic 3D Ultrasonic Scan Generation Using Optical Flow and Generative Adversarial Networks

Luka Posilović¹, Duje Medak¹, Marko Subašić¹, Tomislav Petković¹, Marko Budimir², Sven Lončarić¹

¹University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia

²INETEC Institute for Nuclear Technology, Zagreb, Croatia

Email: luka.posilovic@fer.hr

Abstract—Non-destructive ultrasonic analysis of materials is a method for assessing the integrity of the inspected components. It is commonly used in monitoring critical parts of the power plants, in aeronautics, oil and gas, and the automotive industry. Since most ultrasonic inspections rely on expert's previous experience they must constantly practice on new, unseen data. Acquiring enough data for training human experts on non-destructive ultrasonic scan analysis can be an expensive and time-consuming task. The only possibility to get new data for practicing is to implant synthetic defects in real metal blocks. Artificial defects are made by temperature strain, electrical discharge, and physical damage. All of those methods are very complicated and expensive to perform. Also metal blocks have to be taken from the components of the power plants to have the same structure and be realistic. In this work, some attempts have been made to generate 3D ultrasonic scans using computer vision and deep learning methods.

Index Terms—image processing, image generation, optical flow, generative adversarial networks, ultrasonic imaging, non-destructive evaluation

I. INTRODUCTION

Non-destructive testing (NDT) is a set of techniques that are heavily used in industry to evaluate the integrity of materials in production systems and products. These techniques offer a variety of approaches and one of them is ultrasonic testing. It is extensively used to find defects in all sorts of materials, but mostly in various steel and aluminum blocks. Although the ultrasonic acquisition is fairly simple and straightforward, analysis tends to be tiresome and complicated. It is mainly done by human experts manually analyzing all the acquired data. In many cases, it is very important that all possible defects are found, but that only depends on one's previous experience and training on such data.

Human experts can only gain experience by performing constant analysis on new ultrasonic testing (UT) data. Obtaining enough of such novel examples of UT data is hard and expensive. Usually, they practice on data acquired on metal blocks with artificially implanted defects. Defect creation in such metal blocks is very expensive and complicated. Therefore, there is not a wide variety of such blocks available for training. This can be a problem since that is the only way of training the skilled personnel on which the safety of systems like nuclear power plants depend.

For proper analysis of ultrasonic data, inspectors use a variety of data visualizations. Some of them include A-scans,

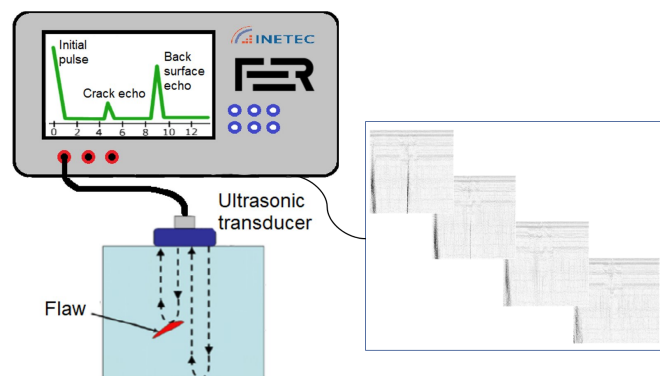


Fig. 1: Drawing of the process of ultrasonic acquisition. A-scan representation is shown in the device, B-scans are shown in the window.

B-scans, and C-scans. An A-scan is a signal's amplitude as a function of time, a B-scan displays a cross-sectional view of the inspected material, and a C-scan provides a top view of its projected features.

In Figure 1 the process of ultrasonic acquisition and samples of a sequence of B-scans are shown. With the ultrasonic transducer or probe, ultrasonic signals are acquired and sent to the ultrasonic device that enables us to see the data in a more human-friendly form. In the shown B-scans that are acquired by such installation, one can see a defect that is the most visible in the first B-scan. In the second B-scan, it fades away and continues to do so until it is not visible by the ultrasonic transducer and therefore not shown in a B-scan.

To ease and make the process of training the inspectors more efficient one could generate images, B-scans, that resemble real data. There have been some attempts in generating ultrasonic data using finite element simulation method [1]. In [2] authors utilized scans without defects and pasted extracted defects on them. Generation of high-quality images could be done using some deep learning neural network such as autoencoders or generative adversarial networks (GANs). In the medical imaging field, one can find similar problems of lack of data and attempts to generate more images using deep learning. In [3] authors compared several GANs in a skin lesion generation task. Generating ultrasonic B-scan images of musculoskeletal system was done in [4] using CycleGAN

[5]. Generating MR and CT images is the most similar task to generating ultrasonic data. MR images of the low resolution of only 64x64 and 128x128 px with GAN were done in [6].

However, since ultrasonic data consists of multiple B-scans it remains a challenge to generate correlated B-scans that could form the whole block or 3D data. There have been some attempts in generating 3D medical brain MRI scans [7], but the resulting data is of low quality and resolution.

In this paper, a method using optical flow for predicting the following B-scans from the single B-scan is used. We also investigate the possibility of generating whole blocks, a set of correlated B-scans, by using a generative adversarial network.

In section II a detailed description of the dataset used in this work is given. In section III methods based on optical flow and used generative adversarial networks are presented. In section IV the proposed methods are applied and the results are discussed. Section V concludes the paper.

II. DATASET

The dataset was obtained by scanning six steel blocks containing artificially created defects in the internal structure. Blocks varied in size and contained between six to 34 defects. In total there were 68 defects. Blocks were scanned using INETEC Dolphin scanner with phased array probes. An INETEC phased array ultrasound transducer of a central frequency of 2.25MHz was used. Data were acquired by scanning the blocks angles from 45 degrees to 79 degrees with a 2-degree increment. Blocks were also scanned with a skew of zero and 180 degrees. INETEC SignyOne data acquisition and analysis software was used to process the data and create B-scans (further noted as images) that were used in the dataset. Data were converted to B-scans as-is, without pseudo-coloring, as grayscale images. All images were then converted into patches of size 256x256 pixels and annotated by multiple human experts. There were in total 9188 images, 3825 images with defects and 5363 images without defects. There was a total of 6238 bounding box annotations.

This dataset was used to train the GAN for generating ultrasonic B-scan images. This dataset was also used to calculate optical flow between each neighboring B-scan in the block.

III. PROPOSED METHODS

In this work, we performed a series of experiments on generating 3D ultrasonic data. Our goal is to generate consecutive B-scans to form a whole block. We developed two methods for that purpose. First one generates an optical flow matrix using a generative adversarial network. This GAN inputs a single B-scan and predicts the movement of the geometry and defects depicted on the following B-scan. Using the predicted optical flow from the base scan, we can reconstruct the following B-scan. The other method uses a different GAN to directly generate realistic consecutive B-scans, one by one. In this section we describe both methods and all prerequisites needed to develop these methods.

A. Method 1 - Optical flow, GAN and reconstruction

1) *Optical flow*: Optical flow is the pattern of motion of objects and surfaces caused by the relative motion of the scene or the observer [8]. It was first mentioned in [9] by describing the visual stimulus provided to animals moving through the world. It can be used to detect the movement of the objects in a video [10] and predict the trajectory of the moving objects [11]. It can also be used to predict the appearance of the following scene in a video [12]. In this work, we are using optical flow to predict the following B-scan in an ultrasonic scan. Since we do not have the ground truth of the motion in two consecutive images, we calculate pseudo ground truth using traditional optical flow calculation methods.

There are multiple ways of calculating optical flow. In this work, we evaluated some of them. Optical flow calculation methods can be divided into traditional methods like the ones we used and deep learning methods such as RAFT [13].

The most famous method for calculating optical flow is the Lucas-Kanade [14] algorithm. It can be used to track the motion of some predefined points in the image. The problem with this method is it calculates movement only of those certain points. For prediction of the next image from the current image and calculated optical flow we need to have the motion of each pixel in an image. That is why we used dense optical flow algorithms in this work.

Dense optical flow computes the optical flow vector for every pixel of the frame. These methods are much slower but lead to a more accurate result. In this work we tested Farneback method [15], Dual TV-L1 [16], [17] and Dense Inverse Search (DIS) algorithms [18]. We calculate optical flow with each of these methods. We decided to use Farneback and DIS methods to generate optical flow as pseudo ground truth of our optical flow. These two methods were capable to detect motion in ultrasonic images. Methods for calculating our pseudo ground truth optical flows are shown in Figure 2.

We calculated optical flow using both methods in both directions for all blocks and scans in the dataset. We used these optical flows as pseudo ground truth for training the GAN for generating optical flow from a single input image.

2) *Generative adversarial network for generating optical flow*: In this work, we wanted to develop the algorithm for fantasizing possible optical flow matrix from a single B-scan ultrasonic image. In this way, by generating diverse optical flows we are able to predict possible following B-scans from the input one.

For generating optical flow from an input image we use a custom GAN. GANs can be used to predict optical flow and are shown to produce good results. In [19] authors used a GAN network to calculate optical flow from two consecutive images in a video.

Our GAN consists of a generator and two discriminators. The generator has an encoder-decoder-like architecture with skip connections. The encoder has seven sets of layers of 2D convolutions, Batch normalization, and Leaky ReLU. Each convolutional layer has a kernel size of four and strides two. There are seven decoder groups of layers each with a 2D

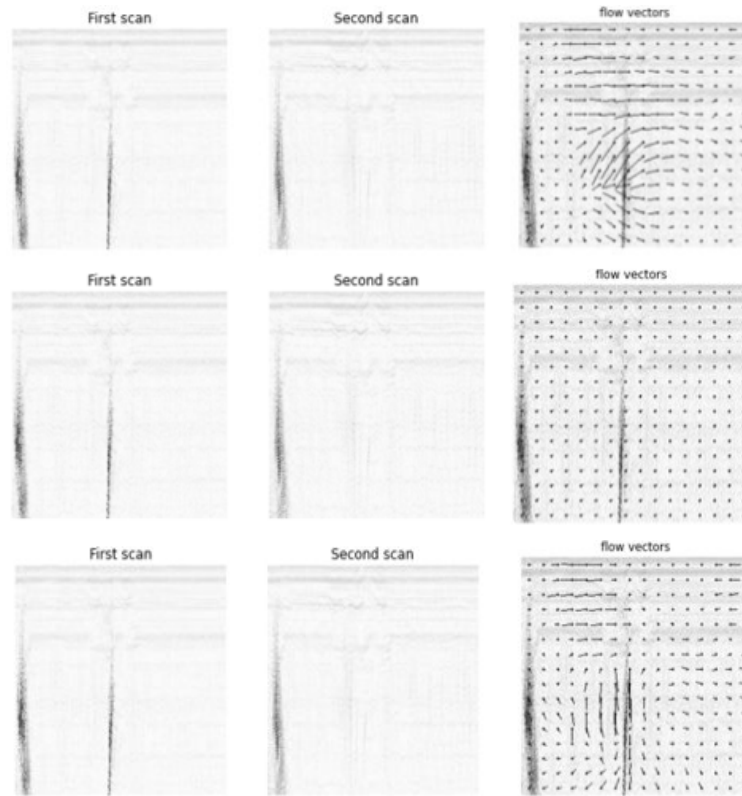


Fig. 2: From up to down, left to right are shown two consecutive B-scans and flow vectors of the computed optical flow using Farneback, Dual TV-L1 and DIS method

transposed convolutional layer, Batch normalization, Dropout, and ReLU layer. The last layer of the decoder has tanh as an activation instead of ReLU because optical flow values can be negative. It also does not use Batch normalization or Dropout. There is a skip connection on each of the encoder layers to the corresponding decoder layer except for the bottleneck layer. Input to the generator is a single B-scan and the goal of the generator is to create a possible optical flow based on that image. Input and output resolution of images are 256x256px. Input images have one channel, and output optical flow matrices have two channels, one for the movement in each x and y-axis.

There are two PatchGAN [20] discriminators whose goal is to distinguish between generated and ground truth optical flow. One discriminator takes an image of original resolution, 256x256px, and the other one inputs a downscaled image by factor two. Goal of the discriminators is to help the generator generate more realistic optical flows. Usage of the two discriminators enables us to have the highest quality of both coarse and fine details.

We use Wasserstein [21] loss to train this GAN since we want the network to learn the complex distribution not the exact translation from image to a certain optical flow. We also trained a model using L1 loss.

3) *Reconstruction:* Once we generate a possible optical flow of the base scan, we can reconstruct the following scan

from the optical flow and the base image. As previously mentioned, generated optical flow matrices have two channels, one for the movement in the x-axis and one for the movement in y-axis. To reconstruct the movement of each pixel in the input image, for each pixel we perform this process. First from generated optical flow we create the following matrix:

$$M = \begin{bmatrix} 1 & 0 & x_{i,jdistance} \\ 0 & 1 & y_{i,jdistance} \end{bmatrix} \quad (1)$$

where x, y are the values of optical flow matrix at the position i,j

We then calculate a product of matrix M and the following matrix:

$$P_{i,j} = M * \begin{bmatrix} i & j & 1 \end{bmatrix} \quad (2)$$

The result is the position of the pixel on the resulting image. We calculate this for each pixel in the input image to reconstruct the next B-scan from the generated optical flow.

B. Method 2 - Generative adversarial network for generating realistic B-scans

The second method we developed uses a GAN to directly generate a series of consecutive images, B-scans. This generative network is slightly more complex than the previous one since it has a harder task of creating new complex images from simpler position masks. It is based on SPADE GAN



Fig. 3: A pair of input position mask (left) and the corresponding generated image (right).

[22]. SPADE GAN utilizes spatially-adaptive normalization (SPADE) which enhances the quality of generated images.

Input to this GAN is a binary position mask, an image of the same resolution as an output image, with all values of zero except at the position of the bounding box. With a position mask, we can define the position of the defect in the generated image. An example of the input position mask and corresponding generated image can be seen in Figure 3. While generating images we can create bounding boxes of random sizes and at random positions and use them to generate new images. For generating neighboring scans we create a position mask and modulate the sizes and positions of bounding boxes by a few pixels. This way we can simulate change of appearance of the defects throughout scans.

SPADE GAN utilizes an encoder that inputs a scan which depicts the desired style of the generated image. By sampling from the output of the encoder, like in the case of an autoencoder, we get a distribution which, hopefully, describes the desired style. This distribution is then passed on to the generator. This procedure can help us to generate images with or without visible geometry and with certain noise and defect shapes. It also enables us to have a non-deterministic output of the model. The generator of this GAN is in form of a decoder. It is made of several ResNet blocks with upsampling layers. We feed the downsampled position mask in each normalization block. Generator and encoder together form a Variational Autoencoder (VAE).

Our GAN has two PatchGAN discriminators. Just like the previously explained GAN one discriminator's input is a downsampled image. Our modification of this GAN introduces the object detection network, YOLOv3 [23], as an additional discriminator. We train the object detector on our dataset to detect defects and then use it as the discriminator by learning the same output of the detector with the generated and ground truth image.

Our GAN generates images of 256x256 pixels resolution. We train the encoder using KL-divergence loss. We train the generator with four different losses. At the output of the generator, we calculate hinge loss on the generated image and the paired real image. We tried using wasserstein loss like in the previous GAN but didn't get the satisfactory results. For propagating discriminator loss to the generator we use the mean squared error loss. We also use feature matching loss.

Comparing the output of three scales in the object detector with a real and generated image as output makes the fourth loss. We use the L1 loss function for that. For training the discriminator we use the mean squared error loss.

By slightly modifying the size and position of the defect with the position mask and by keeping the same style of the generated image we can generate consecutive B-scans. This way we can generate realistic blocks by generating consecutive B-scans of the same style.

IV. EXPERIMENTAL SETUP

In this section the process of training the developed generative models is described. In the first method, we train a GAN to generate optical flow from a single input image. We then use the generated optical flow and reconstruct the following scan. With the second method, we train a SPADE GAN to generate realistic B-scan images from input position mask.

For training the optical flow generation GAN we use Adam optimizer with the learning rate of $2e-3$, first-moment term of 0.5, and second-moment term of 0.999. We trained the network with a batch size of eight for 100 epochs on a single NVIDIA Titan Xp graphics card. We use the generated optical flow and the input image to reconstruct the following scan. We train this single GAN with a mixed dataset of pseudo ground truth optical flows calculated using both Farneback and DIS method.

We train the SPADE GAN using the Adam optimizer with a first-moment term of 0.5 the second one of 0.999 and a learning rate of $2e-3$. We implemented a set of simple data augmentations for training this GAN. Augmentations include horizontal flip, brightness modulation, and random cropping. We train the GAN for 100 epochs and for the last 10 epochs we linearly reduce the learning rate to zero. We train the network with a batch size of 6. We train it using three NVIDIA Titan Xp graphics cards.

For evaluating the quality of generated images we made a test with four human experts on non-destructive ultrasonic image processing. The test was performed in three stages. In the first stage, experts were shown 10 real and images generated using optical flow side by side. They were asked to compare the images and select the better looking one. In the second stage experts were shown the combinations of real and images generated with the GAN. In the third phase, experts were shown combinations of images generated with the optical flow method, and the GAN. During testing, experts did not know which image is from which group. In each test, experts were shown two consecutive images for each of the tested method, to evaluate not only the appearance of the images, but also how realistically are motions between two consecutive scans depicted.

V. RESULTS AND DISCUSSION

In this section, both methods are tested and their results are shown. As can be seen in Figure 2, the Dual TV-L1 method can barely capture motion between two consecutive

TABLE I: Percentages of shown images of each generation method experts declared more realistic.

Real/Flow	Real/SPADE	Flow/SPADE
73% / 27%	63% / 37%	30% / 70%

scans, therefore we chose the Farneback and DIS method for calculating our pseudo ground truth optical flow matrices.

The performances of our developed methods are visually evaluated. We tested the optical flow generation algorithm, generated optical flow matrices, and compared the reconstructed images with the reconstructions from the pseudo ground truth optical flows. In Figure 4 examples of optical flow and reconstructed images can be seen.

As can be seen in the examples, images reconstructed using ground truth optical flows calculated with Farneback and DIS algorithms tend to produce deformations of the defect and a lot of anomalies and noise. However, reconstructions from the GAN generated optical flow are of high quality and without anomalies. Reconstructions are similar and slightly changed from the original image. Noise is unchanged, but defects are eroded and geometry is translated. These transformations are similar to the ones found in real ultrasonic scans.

Our second method trained successfully and is able to generate highly realistic sequence of B-scan images. With those images, we are able to generate whole blocks. An example of the generated ultrasonic data can be seen in Figure 5. As can be seen in the figure, most left defect fades away and gets distorted from the first to the third scan. The defect on the most right of the first scan gets smaller on the other two scans. These movements and modulations in size resemble real situation with ultrasonic scans. In the end, generated images are of high quality and highly resemble real ultrasonic scans.

The test with the human experts evaluating the quality of the generated images concluded with the following results. In the first test, in only 27% of the cases experts voted in favor of the images generated with the optical flow method being better quality than the real images. The SPADE GAN method outperformed the optical flow method with 37% of the cases experts voted in favor of the SPADE GAN than the real images. SPADE GAN method greatly outperformed the optical flow method when comparing those two in the third test. In 70% of the time, experts voted in favor of the SPADE GAN method, than the optical flow method. Results of the test can be seen in Table I. Results of this test has shown that generated images still can not mislead the experts into thinking they are real images. However, the SPADE GAN method has come close to it, and outperforms the Optical flow method.

VI. CONCLUSION

In this work, we developed two methods for generating consecutive ultrasonic B-scan images. Our first method generated optical flow from the single input image. With this optical flow and the input image we can then generate the following B-scan.

In the second developed method using our modified SPADE GAN we are able to generate a sequence of ultrasonic B-scans. We achieved great results using both methods and validated the concept of generating 3D ultrasonic scans. We have shown that the SPADE GAN method outperforms the optical flow method.

With increasing requirements for data in the field of ultrasonic testing, a method for a fast and efficient way of generating such data is highly valuable. Generated data, if of sufficient quality, could be used for developing state-of-the-art deep learning networks or training human experts for UT analysis. In future work, a GAN capable of generating 3D data and the quality of generated images should be evaluated.

ACKNOWLEDGMENT

This research was co-funded by the European Union through the European Regional Development Fund, under the grant KK.01.2.1.01.0151 (Smart UTX). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

REFERENCES

- [1] S. Liu, J. H. Huang, J. Sung, and C. Lee, "Detection of cracks using neural networks and computational mechanics," *Computer Methods in Applied Mechanics and Engineering*, vol. 191, no. 25, pp. 2831 – 2845, 2002. [Online]. Available: [https://doi.org/10.1016/S0045-7825\(02\)00221-9](https://doi.org/10.1016/S0045-7825(02)00221-9)
- [2] J. Virkkunen, T. Koskinen, O. Jessen-Juhler, and J. Rinta-Aho, "Augmented ultrasonic data for machine learning," *Journal of Nondestructive Evaluation*, vol. 40, no. 1, pp. 1–11, 2021. [Online]. Available: <https://doi.org/10.1007/s10921-020-00739-5>
- [3] A. Bissoto, F. Perez, E. Valle, and S. Avila, "Skin lesion synthesis with generative adversarial networks," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, pp. 294–302. [Online]. Available: https://doi.org/10.1007/978-3-030-01201-4_32
- [4] N. J. Cronin, T. Finni, and O. Seynnes, "Using deep learning to generate synthetic b-mode musculoskeletal ultrasound images," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105583, 2020. [Online]. Available: <https://doi.org/10.1016/j.cmpb.2020.105583>
- [5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232. [Online]. Available: <https://doi.org/10.1109/ICCV.2017.244>
- [6] C. Han, H. Hayashi, L. Rundo, R. Araki, W. Shimoda, S. Muramatsu, Y. Furukawa, G. Mauri, and H. Nakayama, "Gan-based synthetic brain mr image generation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 734–738. [Online]. Available: <https://doi.org/10.1109/ISBI.2018.8363678>
- [7] G. Kwon, C. Han, and D.-s. Kim, "Generation of 3d brain mri using auto-encoding generative adversarial networks," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds. Cham: Springer International Publishing, 2019, pp. 118–126. [Online]. Available: https://doi.org/10.1007/978-3-030-32248-9_14
- [8] D. Warren and E. Strelow, *Electronic Spatial Sensing for the Blind: Contributions from Perception, Rehabilitation, and Computer Vision*, ser. Nato Science Series E. Springer Netherlands, 1985. [Online]. Available: https://books.google.hr/books?id=-I_Hazgqx8QC
- [9] J. J. Gibson, "The perception of the visual world." 1950. [Online]. Available: <https://doi.org/10.1126/science.113.2940.535>
- [10] A. A. Shafie, F. Hafiz, M. Ali et al., "Motion detection techniques using optical flow," *World Academy of Science, Engineering and Technology*, vol. 56, pp. 559–561, 2009.

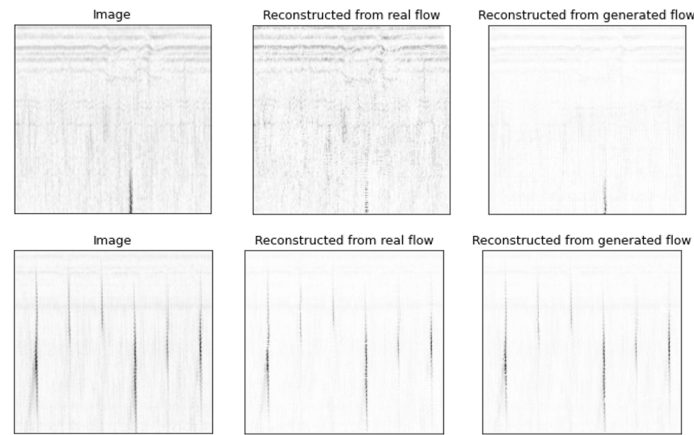


Fig. 4: Base input image used for generating optical flow is shown in the first column. A visualisation of reconstruction from pseudo ground truth (DIS method - second column) and from optical flow generated with our GAN (third column) trained using Wasserstein loss (first row) and L1 loss (second row).

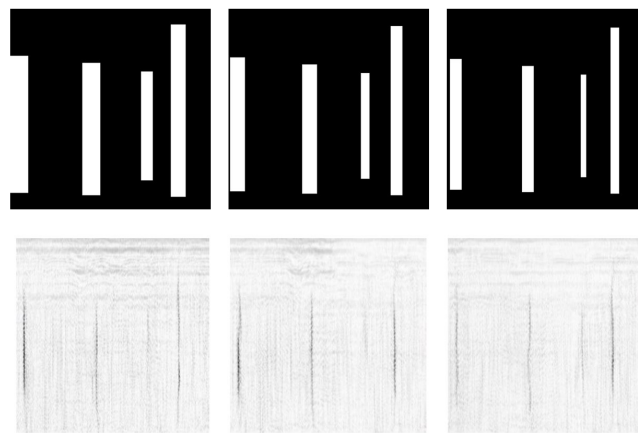
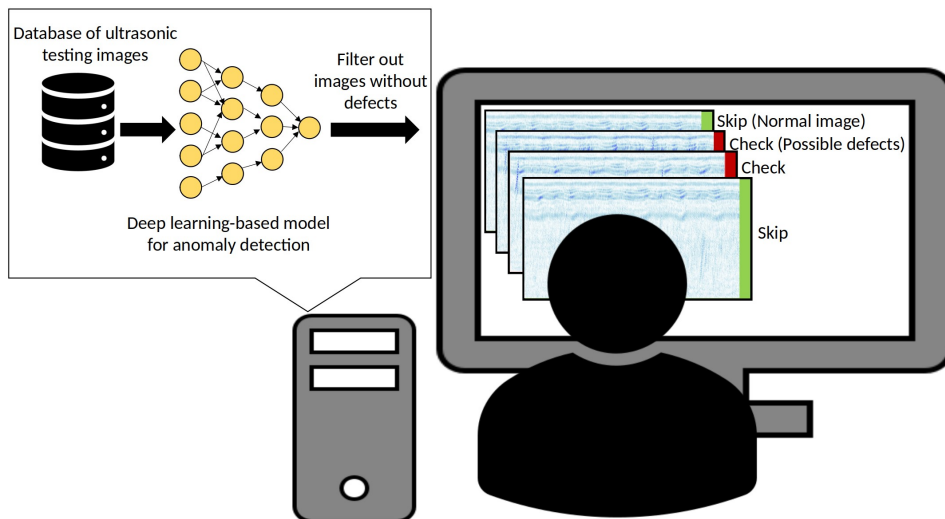


Fig. 5: An example of three consecutive B-scans generated using our modified SPADE GAN. First row shows position masks with bounding box modulation across scans. Second row are resulting generated images.

- [11] P. Wood-Bradley, J. Zapata, J. Pye *et al.*, "Cloud tracking with optical flow for short-term solar forecasting," in *Proceedings of the 50th Conference of the Australian Solar Energy Society, Melbourne*. Citeseer, 2012.
- [12] J. Saric, M. Orsic, T. Antunovic, S. Vrazic, and S. Segvic, "Warp to the future: Joint forecasting of features and feature motion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. [Online]. Available: <https://doi.org/10.1109/CVPR42600.2020.01066>
- [13] Z. Teed and J. Deng, "Raft: Recurrent all-pairs field transforms for optical flow," in *European Conference on Computer Vision*. Springer, 2020, pp. 402–419. [Online]. Available: https://doi.org/10.1007/978-3-030-58536-5_24
- [14] B. D. Lucas, T. Kanade *et al.*, "An iterative image registration technique with an application to stereo vision." Vancouver, British Columbia, 1981. [Online]. Available: <https://doi.org/10.5555/1623264.1623280>
- [15] G. Farneback, "Two-frame motion estimation based on polynomial expansion," in *Scandinavian conference on Image analysis*. Springer, 2003, pp. 363–370.
- [16] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime tv-l 1 optical flow," in *Joint pattern recognition symposium*. Springer, 2007, pp. 214–223. [Online]. Available: <https://doi.org/10.5555/1771530.1771554>
- [17] J. S. Perez, E. Meinhardt-Llopis, and G. Facciolo, "Tv-l1 optical flow estimation," *Image Processing On Line*, vol. 2013, pp. 137–150, 2013.
- [18] T. Kroeger, R. Timofte, D. Dai, and L. Van Gool, "Fast optical flow using dense inverse search," in *European Conference on Computer Vision*. Springer, 2016, pp. 471–488.
- [19] W.-S. Lai, J.-B. Huang, and M.-H. Yang, "Semi-supervised learning for optical flow with generative adversarial networks," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 353–363. [Online]. Available: <https://doi.org/10.5555/3294771.3294805>
- [20] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134. [Online]. Available: <https://doi.org/10.1109/CVPR.2017.632>
- [21] J. Adler and S. Lunz, "Banach wasserstein gan," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018. [Online]. Available: <https://proceedings.neurips.cc/paper/2018/file/91d0dbfd38d950cb716c4dd26c5da08a-Paper.pdf>
- [22] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019. [Online]. Available: <https://doi.org/10.1109/CVPR.2019.00244>
- [23] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018, cite arxiv:1804.02767Comment: Tech Report. [Online]. Available: <https://arxiv.org/abs/1804.02767>

Publication 4

L. Posilović, D. Medak, F. Milković, M. Subašić, M. Budimir, S. Lončarić, "Deep Learning-Based Anomaly Detection From Ultrasonic Images", in *Ultrasonics*

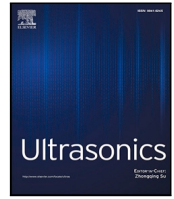


Highlights

Deep Learning-Based Anomaly Detection From Ultrasonic Images

Luka Posilović, Duje Medak, Fran Milković, Marko Subašić, Marko Budimir, Sven Lončarić

- The first comparison of multiple anomaly detection methods on an UT dataset
- Experiments show significant advantage in using the self-supervised anomaly detection
- Presented methods can be trained using only images without defects
- The best method achieves almost 82% of the AUC ROC metric
- The best method successfully detects all defects



Deep learning-based anomaly detection from ultrasonic images

Luka Posilović^{a,*}, Duje Medak^{a,1}, Fran Milković^a, Marko Subašić^a, Marko Budimir^b, Sven Lončarić^a

^a University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia

^b INETEC Institute for Nuclear Technology, Zagreb, Croatia

ARTICLE INFO

Keywords:

Non-destructive testing
Ultrasonic testing
Anomaly detection
Generative Adversarial Network
Deep learning

ABSTRACT

Non-destructive testing is a group of methods for evaluating the integrity of components. Among them, ultrasonic inspection stands out due to its ability to visualize both shallow and deep sections of the material in the search for flaws. Testing of the critical components can be a tiring and time-consuming task. Therefore, human experts in analyzing inspection data could use a hand in discarding anomaly-free data and reviewing only suspicious data. Using such a tool, errors would be less common, inspection times would shorten and non-destructive testing would be more efficient. In this work, we evaluate multiple state-of-the-art deep-learning anomaly detection methods on the ultrasonic non-destructive testing dataset. We achieved an average performance of almost 82% of ROC AUC. We discuss in detail the advantages and disadvantages of the presented methods.

1. Introduction

Automated defect detection in some metal blocks is a highly researched problem in the field of non-destructive evaluation (NDE). Detection and proper evaluation of every defect in the inspected block is a crucial segment of every inspection. This is why the task of evaluating inspection data is, to the best of our knowledge, mostly assigned to human experts. Development of the automated defect detection is a challenging task with the most important part of collecting enough data being sometimes impossible. Lack of real inspection data with enough visible defects is a challenge that researchers tried to tackle using augmented data [1] or by generating additional synthetic data using deep learning [2]. Positive samples of data with visible defects are sparsely found, but scans of blocks without defects are plentiful, especially in real-world inspections. It would be very helpful and time-saving if the assisted analysis module would filter out these scans and leave an evaluation of the potentially anomalous scans to the inspector. Inspectors would then have to analyze only anomalous data and classify defects. It would shorten the time needed for the data evaluation, but also would not exclude the inspectors from the process. However, in order to develop such a module for the detection of anomalous scans containing defects again a representative training set is needed. Without data containing defects, one can go a self-supervised-learning direction and develop a deep-learning anomaly detection model.

Anomaly detection challenges are not only present in non-destructive testing, but also in medical image analysis [3] and product industry quality control [4]. Although some work has been done in these areas, no state-of-the-art self-supervised anomaly detection models have been researched for the field of non-destructive testing. A semi-supervised approach to anomaly detection is the best approach for non-destructive testing due to a huge number of normal data without defects in real inspections. Since there is only a small number of examples of defects in real ultrasonic inspections, usual supervised learning models for classification require at least a few percent shares of positive samples in the training dataset to achieve state-of-the-art results. On the other hand, semi-supervised anomaly detection is built upon a dataset of only normal images. Our main goal in this work is to set a baseline of state-of-the-art anomaly detection models on the ultrasonic non-destructive testing dataset. To the best of our knowledge, this is the first such work in analyzing different models on a real-world and very complex non-destructive evaluation dataset.

The development of algorithms and deep learning models for automatic or assisted ultrasonic non-destructive evaluation has recently seen a big growth in published works. Most of the work is done using traditional approaches and analyzing only ultrasonic data as one-dimensional signals. In [5] the authors used discrete wavelet transform

* Corresponding author.

E-mail addresses: luka.posilovic@fer.hr (L. Posilović), duje.medak@fer.hr (D. Medak), fran.milkovic@fer.hr (F. Milković), marko.subasic@fer.hr (M. Subašić), marko.budimir@inetec.hr (M. Budimir), sven.loncari@fer.hr (S. Lončarić).

¹ Equal contribution.

<https://doi.org/10.1016/j.ultras.2022.106737>

Received 19 November 2021; Received in revised form 18 March 2022; Accepted 19 March 2022

Available online 2 April 2022

0041-624X/© 2022 Elsevier B.V. All rights reserved.

for ultrasonic flaw detection in signals. A more recent article [6] introduces Artificial Neural Networks (ANNs) for defect detection in ultrasonic non-destructive testing (NDT). However, the authors also use the wavelet transform for feature extraction and ANN only for classification. Using such traditional approaches results in bad generalization on a broad specter of data structure and flaw types. In [7] authors classified ultrasonic NDT images. They compared various methods and a deep learning model achieved the best results. The authors of [8] investigated a deep learning network with drop-out regularization that outperformed other defect classification methods with prior feature extraction. The lack of data containing defects in developing the automatic ultrasonic analysis has been a burden to researchers. They tried to solve it by generating new examples [1,2,9] and using the generated data to enhance the performance of deep learning models. Deep learning solutions for defect detection have been researched in [10,11]. Researchers worked with ultrasonic images and used the popular architectures YOLO and SSD to detect defects in [11]. Analyzing multiple ultrasonic images at a time to speed up the process of defect detection is done in [12]. In [13] authors developed a deep convolutional neural network for estimating flaws in ultrasonic phased array scans.

Following previous work, we researched the deep learning approach to anomaly detection on ultrasonic images (B-scans). In this work, we develop a method for detecting ultrasonic images containing defects using the anomaly detection approach. All of the above-mentioned articles are dealing with defect localization or classification. Anomaly detection has not yet been researched well in this field. Only in [14] a form of a Variational Autoencoder (VAE), which is a go-to for anomaly detection, has been studied for ultrasonic anomaly detection. In [15] a reference C-scan image was used to find anomalies on defective scans. In [16] authors worked with ultrasonic signals and various machine learning algorithms such as Isolation forest and SVM to detect defects in the aluminum and carbon-fiber-reinforced-plastic (CFRP) plates. Since there are hardly any papers on ultrasonic NDT anomaly detection we look at the related area of medical image analysis. In [17] three deep learning solutions based on Generative Adversarial Networks (GANs) are compared. The authors show that results of anomaly detection yet have to be improved for real-world medical datasets. A convolutional neural network for anomaly detection in brain MRI images has been developed in [18].

In recent times, anomaly detection has seen a huge rise in researchers' attention. However, most of the work has been done on small and relatively simple publicly available datasets such as MVTec AD [4,19], CIFAR [20] or MNIST [21]. Models on such datasets with low-resolution images or a small number of samples with low variance between defects perform very well. However, performance on these datasets does not necessarily reflect real-world situations. This is why it is very important to test the current state-of-the-art methods on real-world application datasets. There are many methods and deep learning approaches to anomaly detection. In this work, we tested some of the most popular models. To the best of our knowledge, this is the first time state-of-the-art anomaly detection models have been tested on realistic ultrasonic non-destructive testing images. The first one is the Ganomaly [22] model. It is based on a GAN architecture utilizing the generator/discriminator network. More recent method is the PaDiM [23], a Patch Distribution Modeling model. It performs very well on a public dataset such as MVTec. In [24] authors developed a new semi-supervised model with normalizing flows for anomaly detection called DifferNet. In this work, we use the Ganomaly, PaDiM, and DifferNet models to determine the baseline anomaly detection score on our ultrasonic non-destructive testing dataset. We bring some modifications to these networks to improve their performance on such specific challenges. Furthermore, we train a series of classifiers to conclude when self-supervised learning outperforms the supervised.

1.1. Outline

This paper is organized as follows. Section 2.1 gives a detailed description of our ultrasonic non-destructive testing dataset. In the following Section 2.2 state-of-the-art anomaly detection methods are introduced. Section 3 describes our modifications to the models and the training procedures for the presented models. Detailed results are shown and a discussion of the performances of each model is given in Section 4. We conclude our work in Section 5.

2. Materials and methods

2.1. Dataset

One of the hurdles of the development of automatic ultrasonic inspection algorithms is the unavailability of public ultrasonic datasets. One of the few publicly available datasets is found in [1], but it is made using an augmentation technique and is not necessarily realistic. This is why a proper description of the used dataset is required for the repeatability of the published experiments. Our dataset is also a proprietary one but is a real-world-inspection ultrasonic dataset. Some work using this dataset can be seen in [2,25]. We obtained our dataset by scanning six steel blocks with artificially created defects in the internal structure. These defects include, but are not limited to, thermal fatigue cracks, mechanical fatigue cracks, electric discharge machined notches, and solidification cracks. Blocks are of different sizes and contain between six and 34 defects. In total, there were 68 different defects. The distribution of the defects' sizes across the six blocks is shown in Fig. 2.

Blocks were scanned using specialized equipment for NDT. Ultrasonic INETEC Dolphin device with INETEC phased array ultrasonic transducer was used. The phased array transducer with a central frequency of 2.25 MHz was used. It enabled us to scan for defects in the shallow and mid-depths of the blocks. It simultaneously scans angles from 45 to 79 degrees with a two-degree increment. Blocks were scanned in two orthogonal, x and y , directions. Also, the probe was rotated by 180 degrees resulting in data with skew zero and 180. Ultrasonic analysis was done using the INETEC SignyOne data acquisition and analysis software. The software was used to convert the raw ultrasonic data and create Volume-corrected B-scans (VC-B scans). These B-scans, or images, were used in the original grayscale color scheme, without pseudo-coloring. Images are one-channel, 8-bit data, the same as the ones that inspectors look at during the inspection. We cropped patches of 256×256 px resolution from original images with various dimensions for our dataset. Images were not only grouped in the normal/anomalous classes, they were annotated on every single B-scan. Although there were in total 68 different defects, they could be seen from different angles and skews. Such images are different, so every appearance of the defect was annotated. In total there were 5715 anomalous images. In the dataset, there were also 11709 images without defects. Examples of images from the dataset with and without defects can be seen in Fig. 1. In a real-world inspection, this amount of anomalous images are usually not available, therefore we focus on developing semi-supervised methods trained without the use of anomalous images. In this dataset, the number of anomalous images and defects are such because we scanned the blocks with artificially implanted defects. Our ultrasonic dataset is difficult for the deep-learning networks to analyze because of the different geometries shown in the images. Geometries are reflections of the ultrasonic waves that appear because of the irregular shape of the block being scanned. They often resemble defects and as such make it hard for neural networks to distinguish between geometries and defects.

Since semi-supervised anomaly detection models are trained using only normal images and tested on a set of normal and anomalous images we created an appropriate dataset. Because the same defects are found multiple times across the single block it would not be fair if

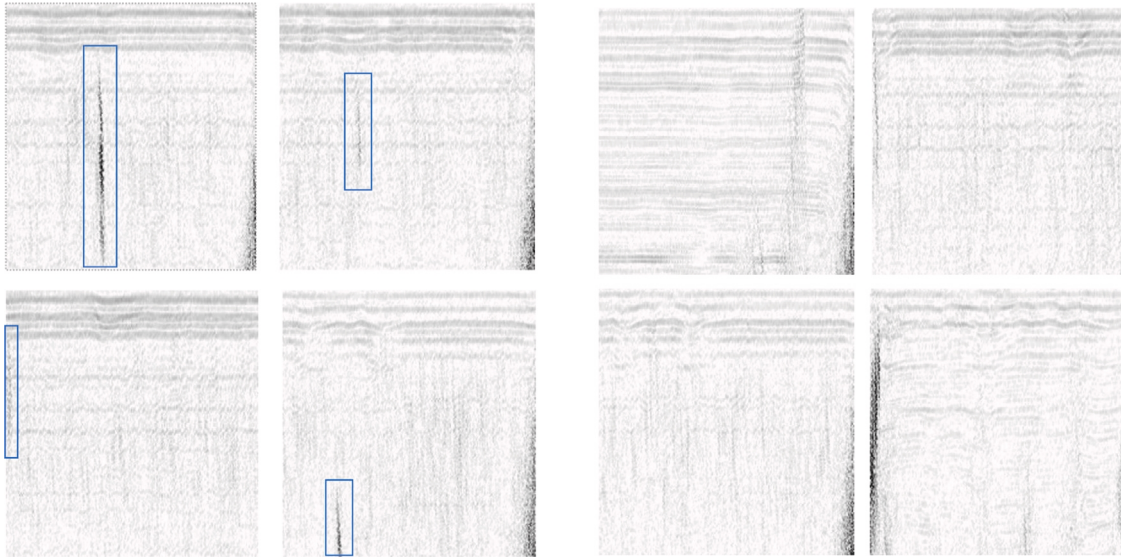


Fig. 1. An example of anomalous (left) and normal (right) ultrasonic images (B-scans) from our dataset. Defects are marked with blue bounding boxes.

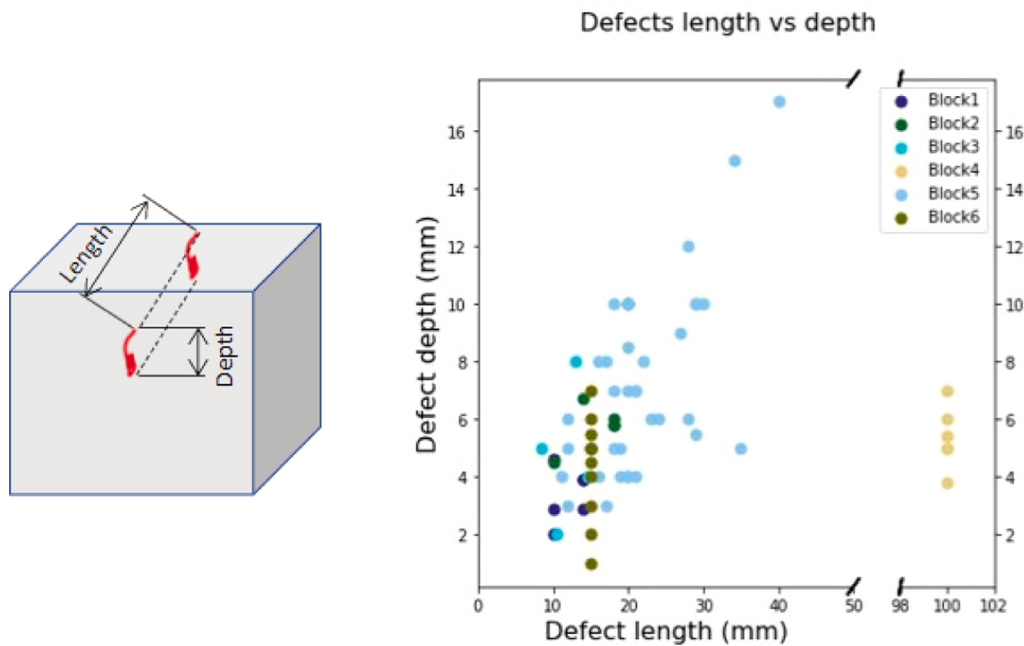


Fig. 2. Distribution of defect dimensions.

training and evaluation were done on the images from the same metal block. Also, to carry out a proper evaluation of the proposed algorithms we divided the dataset into 8 different train/test subsets we named Fold0 to Fold7. Since the biggest metal block in the dataset is several times bigger than others and contains much more defects, we divided this block into data scanned from two orthogonal, x and y , directions. Furthermore, we split the collection of consecutive scans from each direction by half such that different defects are visible in each half. We placed these four disjunctive parts of the block into four different test folds (Fold0–Fold3). Folds 4–7 contain data from only different blocks in the test subset.

2.2. Methods

In this work, we performed a series of experiments using different state-of-the-art anomaly detection models. The goal of this work is to

provide the baseline performance of state-of-the-art semi-supervised anomaly detection models on an industrial real-world challenge. We tested three methods that have different approaches to anomaly detection and semi-supervised learning. One of them is a GAN-based method, and two are based on feature extraction from images. To the proposed methods we provide our modifications that improve the performance. These methods were chosen for multiple reasons. The first method is one of the most popular anomaly detection methods that serve as a baseline for most of the anomaly detection datasets. The other two methods are the ones that perform the best on the MVTec AD dataset. We also give a detailed analysis of each method's performance. We also trained several classifiers and tested their performance for different percentages of positive samples (anomalous B-scans) in the training dataset.

The first method is the GAN-based method, Ganomaly [22]. The main idea behind it is to use the autoencoder to reconstruct the input

image. The autoencoder is trained only on normal images, therefore the hope is that when given the anomalous image it will not be able to reconstruct the abnormality properly. It will then result in the bigger reconstruction loss which is our anomaly metric. Ganomaly improves this concept by adding a discriminator to the model, and another encoder to the autoencoder-generator. The reconstruction of the anomalous images should hopefully result in a higher reconstruction loss due to the additional encoder as opposed to the reconstruction of the normal images. Also, the use of the discriminator should result in a better reconstruction of normal images. Our modification of the Ganomaly is explained in Section 3.

The second method is the Patch Distribution Modeling model [23]. PaDiM utilizes some popular classifiers pretrained on a big public dataset such as ImageNet [26]. The classifier extracts features from the normal images and calculates the distribution of the embeddings on each pixel of the input image. During the training, each patch of the normal image is associated with its corresponding activation vector of the pretrained feature extractor activation map [23]. In this way, the model is able to incorporate the spatial information of the input image to better assess the anomalies of the image. Finally, in the testing phase, the model assesses the abnormality of the image by calculating the distance of the extracted features from the distribution calculated on the normal dataset. Such a model requires no training, except for the pretraining on some big classification datasets. It is also able to localize the defect in the anomalous image. This approach outperformed existing methods on the MVTec anomaly dataset.

The third and final tested model is the DifferNet [24]. It uses likelihoods from a normalizing flow on multi-scale image features. Normalizing flows [27] are neural networks that can learn transformations between data distributions. Since their mapping is bijective they can be used to generate the samples from the modeled distribution and assign likelihood values to the input sample. DifferNet uses the calculated likelihoods to classify the input image as anomalous or normal. Before inputting the image to the feature extractor, a transformation and scaling are performed on the input image. It is what makes the model multi-scale. It also means the model does not need a lot of data for training. DifferNet achieves a state-of-the-art detection performance on MVTec anomaly dataset.

3. Experimental setup

In this section, a description of the training and configuration of deep learning models is given. Our improvements to the standard models are thoroughly explained.

The Ganomaly model is made of a generator and a discriminator. The generator has an encoder-decoder-encoder architecture. The first and the second encoder have the same architecture. They have an initial convolutional layer with 64 filters and a pyramid structure with three convolutional layers with the number of filters ranging from 128 to 512 filters. In the pyramid structure, there is a Batch normalization and a Leaky-ReLU layer after each convolutional layer. The decoder has the same number of convolutional layers, batch normalization, and ReLU instead of the Leaky-Relu. The single discriminator has the same architecture as the encoder with a sigmoid activation as the last layer. We train it using the adversarial loss, image reconstruction loss, and loss between outputs of each of the two encoder layers. The adversarial loss and the encoder loss are the l2 loss, while the reconstruction loss is the l1 loss. The discriminator is trained using the binary cross-entropy loss. We also trained the Ganomaly with the addition of skip-connections in the generator, but with no performance improvement. To make the problem of reconstruction a bit harder for the generator we also tried randomly cutting the patches in the input image and setting the values of the patch to zero which resulted in modest improvement of results. We used the 256×256 px resolution for the input image to the Ganomaly. We trained the Ganomaly model for 100 epochs.

The second model, PaDiM, does not need training, but it needs to extract features from the training dataset. For the feature extraction we used three different feature extraction models, ResNet18 [28], MobileNet-large [29] and AlexNet [30]. They were all pretrained on the ImageNet dataset. We chose these three models as they are the popular classifier choices. The Resnet18 as a feature extractor of the PaDiM achieves state-of-the-art results on the MVTec dataset [23]. Since PaDiM model requires a lot of memory and processing power to calculate the distribution of the activation vectors we choose the MobileNet feature extractor to reduce the time needed for training and inference. We also choose the AlexNet since it is used as a feature extractor in the following model, DifferNet. In the ResNet model, features were extracted from three of its layers. From the MobileNet, features were extracted from the output of the model. In the AlexNet model, we extracted features from four of its layers. Extracted features were then transformed into embedding vectors from which we calculated the mean and the covariance at each pixel position of the input image. During testing embedding from the input image was then compared to the distribution from the training dataset. Distance between the training dataset distribution and the input image embedding were calculated using Mahalanobis distance [31]. In the first paper, PaDiM features are randomly selected, but we used all of the dimensions both from ResNet, MobileNet and AlexNet. The used input image size was 256×256 , but we also used the MobileNet with the input image resolution of 512×512 .

Our work with DifferNet was motivated by the similarity with the PaDiM model. Since both of them work on the same concept of feature extraction using a pretrained neural network and both work at the spatial level of the image features we thought we should get similar results. DifferNet was trained as-is, following the guidelines of the initial work. Images were augmented by rotating and scaling them before being inputted to the feature extractor. The goal of these augmentations is to get a robust anomaly score. We used the AlexNet [30] as a feature extractor. We again used the same 256×256 px input image size.

Since in PaDiM model we used the MobileNet, Resnet and AlexNet classifier models and in the DifferNet we also used the AlexNet model we choose these classifiers and trained them on the same eight-fold dataset split, with the only difference of using the positive samples in the training set. In [1] authors used the VGG model to classify scans as defect/not defect and achieved great results, so additionally we also tested the VGG model. While training the classifier we gradually reduced the number of positive images containing defects in the training set to assess when the performance of the classifier reaches the performance of the self-supervised anomaly detection models. However, to insure the fair training of the classifiers we over-sample the positive class of images in the training dataset after the appropriate subset of positive samples was selected. Otherwise, the classifier might be trained to always predict the negative class. The input resolution to the classifier models is the standard 224×224 px. All of the classifiers were pretrained on the ImageNet dataset.

4. Results and discussion

We tested three different methods with the modifications mentioned in the previous section. All methods were tested on our eight-fold dataset split. For evaluating and comparing models we used the Area-Under-the-Curve (AUC) of the Receiver-Operating-Characteristic (ROC) curve. We also calculated the mean performance across all folds for each of the models. The results of the eight-fold evaluation can be seen in Table 1. We also show the attention maps for the PaDiM and DifferNet models to show how the improvements could be made and explain their performance.

The Ganomaly model trained successfully and was able to completely reconstruct the input images. However, the reconstruction was sometimes even too good to differentiate the normal images from the anomalous ones. Our Ganomaly network learned the reconstruction of

Table 1
AUC ROC results of different anomaly detection models per data fold in %.

Method	Fold0	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Mean
Ganomaly	73.5	62.0	61.0	69.0	74.0	80.0	89.0	75.6	73.0
Ganomaly w/crop patch	78.0	60.2	62.0	66.0	72.5	82.0	89.0	76.0	73.2
PaDiM (Resnet18)	82.7	61.9	65.1	69.8	96.4	89.4	99.6	89.9	81.9
PaDiM (MobileNet-large)	84.5	63.9	67.1	68	97.3	82.7	99.1	88.7	81.4
PaDiM (MobileNet-large-512 × 512)	88.3	64.4	65.5	68.8	97.3	84.1	96.9	89.8	81.9
PaDiM (AlexNet)	83.4	62.4	64.6	66.1	93.2	92.4	99.5	89.5	81.4
DifferNet	79.6	60.6	62.0	65.9	78.0	83.9	94.1	74.7	74.8

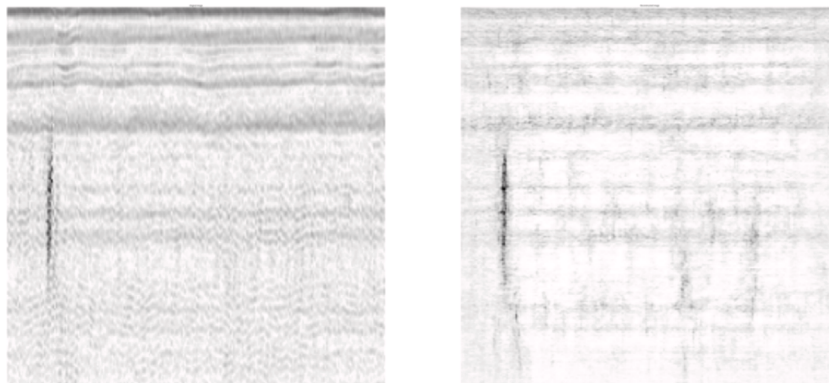


Fig. 3. An example of the anomalous image with a visible defect and the image reconstructed by the Ganomaly model.

the defects from reconstructing the geometry and noise. As can be seen in Table 1 it only achieved the 73% of AUC ROC. An example of the reconstructed anomalous image can be seen in Fig. 3. As can be seen in the figure, the model successfully reconstructed the defect, which poses a problem in this anomaly detection challenge. Since the defect is well reconstructed there is not a major difference between the input image and the reconstructed one. Therefore, it is impossible to adjust the reconstruction metric to detect all anomalous images in the dataset.

The PaDiM model with our additional modifications works the best. It successfully distinguished between geometry, noise, and defects. It even turned out to be spatially invariant, meaning it successfully detected defects in various locations in the image. It achieved surprisingly good results with almost 82% ROC AUC score. ROC curves for Fold 1, 5, and 6 are shown in Fig. 4. We chose these folds because the performance for Fold0 is the worst of all, for Fold6 is the best, and the performance for Fold5 is similar to the average result across all folds. The most important thing to achieve with the anomaly detection model is to detect all anomalous images. As can be seen in Fig. 4, we achieve this in all three ROC curves. Next, we evaluate the advantage of using the anomaly detection model. By looking at the ROC for Fold5 it can be seen that the number of images that the inspectors need to look at would be decreased by more than 10% by using the proposed anomaly detection model. The best possible situation is Fold6, where the time needed for the evaluation of inspection is reduced by almost 90%. However, if we look at Fold1, the time for the inspection is not reduced, but no anomalous images were missed which is a very important result. To conclude, the results depend on the complexity of the data being tested, but on average we can gain a speedup of the inspection of around 10%. Although in the [24] authors claim that more complex feature extractors such as VGG [32] and ResNet perform worse than the simpler AlexNet [30], we tried the VGG, AlexNet, ResNet, and MobileNet, but VGG did not achieve performances comparable to the latter three. Although ResNet18 feature extractor achieves great results, even for a fraction of a percent better than the MobileNet, MobileNet is a much faster and smaller model. We also tried feeding the MobileNet with the images resized to the 512 × 512 px resolution and achieved comparable results to using the ResNet18. When using the AlexNet as a feature extractor on average we achieved the same results as with

the MobileNet. Although PaDiM does not need a lot of GPU memory to work, for calculating the distributions of features it is memory and processor-power-hungry. This is a major throwback for the industrial processes, but for the ultrasonic analysis, it is still faster than the human expert. Examples of input anomalous images and attention maps of the PaDiM model can be seen in Fig. 5. The figure clearly shows how the model ignores the geometry learned from the training dataset and concentrates on the defect. Geometry is visible on the left side of the top-left, bottom-left, and bottom-right images in Fig. 5. It is also visible on the top-right part of the top-right image in Fig. 5. This anomaly detection model could not only be used for filtering out the normal images but also pointing out the defects to the inspectors.

Although utilizing a similar approach as the PaDiM model, DifferNet does not achieve such performance. This is a great example of why it is important to test the models on a specific dataset and how data-specific deep learning solutions have to be for real-world solutions. Using normalizing flows to capture the distribution of the features does not work well on our dataset. Also, since ultrasonic data is almost always acquired from the same direction and ultrasonic images do not suffer from rotational differences. DifferNet augmentation method, therefore, does not improve the results. Normalizing flows are better at capturing the distribution of the large dataset consisting of mostly small images. Since our dataset is relatively big, the DifferNet cannot successfully capture the distribution. PaDiM model calculates the multi-variate Gaussian distribution directly on the features which result in better anomaly detection performance. Even with using the same feature extractor in the PaDiM as in the DifferNet, the AlexNet, the latter does not achieve comparable performance results.

To further analyze the best performing model, the PaDiM with the MobileNet feature extractor, we ran the analysis of detection of all flaws across all folds. Since flaws usually appear in more than one scan, even though the AUC ROC is not perfect does not mean all defects were not detected. We state that a defect is successfully detected if it has been detected in at least one of its occurrence. Using the best anomaly detection method we successfully detected all of the defects. Fig. 6 shows the number of appearances and detections of each defect in the dataset. When using the threshold of 0.5, the PaDiM model misses two defects, one in the Fold0 and one in the Fold3. Both of these defects

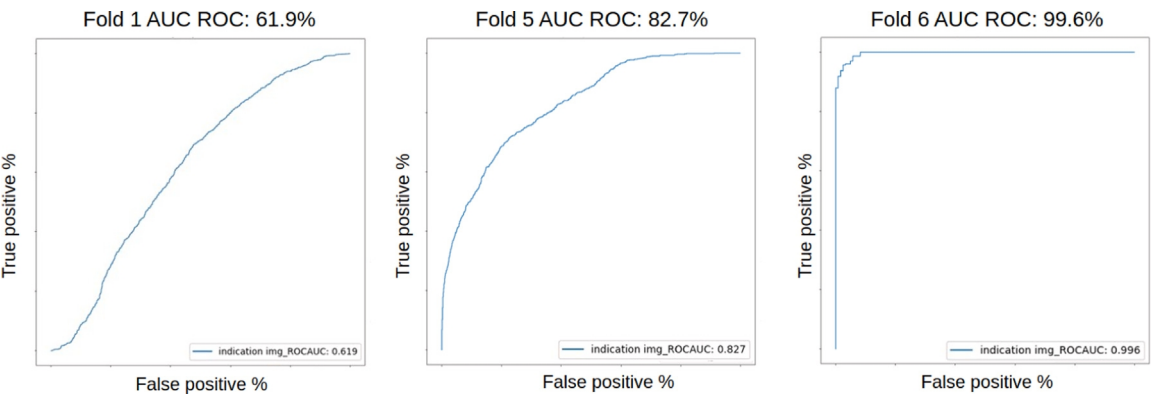


Fig. 4. ROC curves for Fold1 (left), Fold5 (middle), and Fold6 (right) with the AUC ROC result from the PaDiM model using ResNet18.

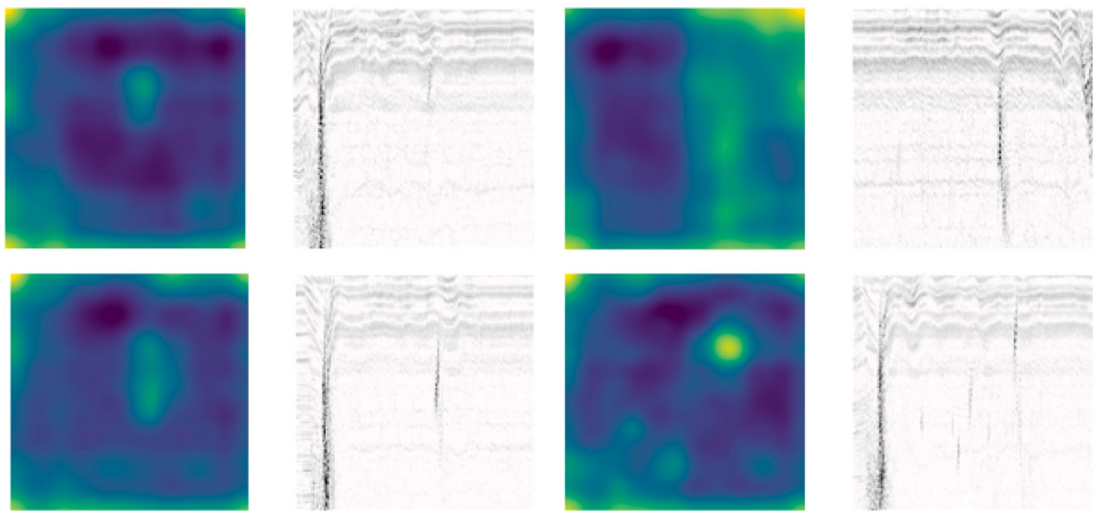


Fig. 5. Examples of anomalous images and the attention from the PaDiM model.

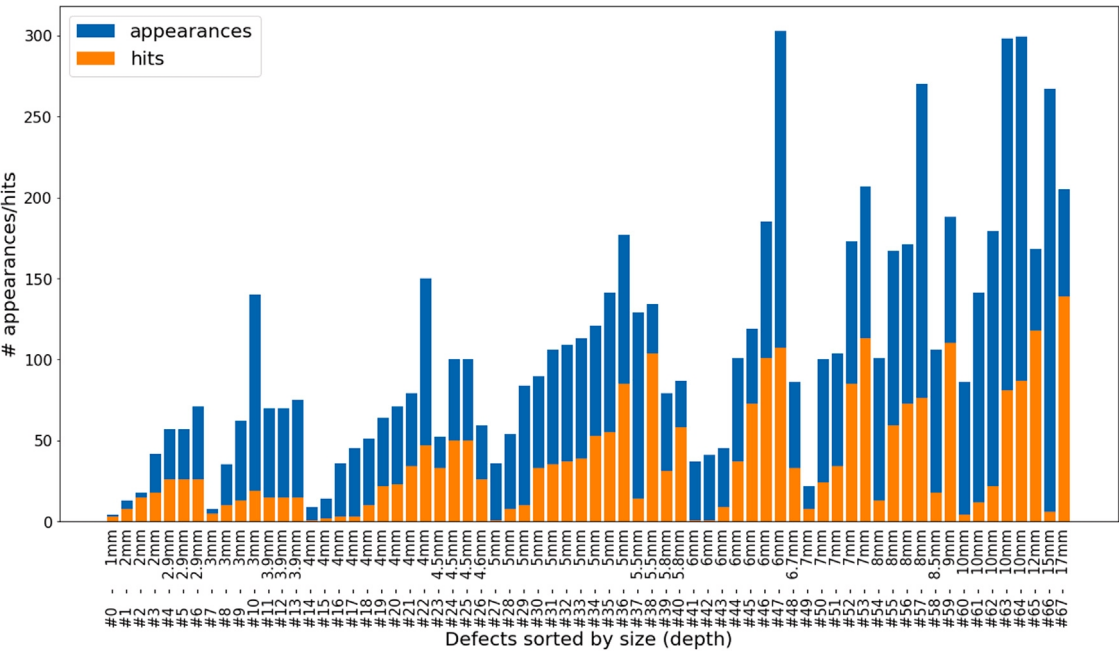


Fig. 6. The number of appearances and detections of each defect in the dataset.

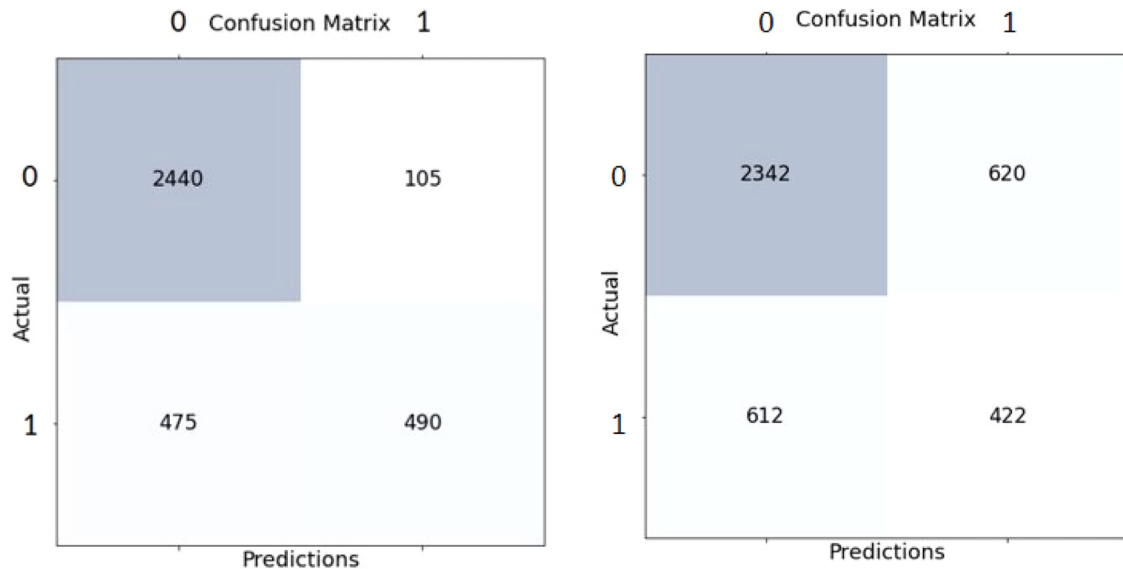


Fig. 7. Confusion matrix of using the best model (PaDiM with the MobileNet) with the anomaly threshold lowered to 0.2 for the Fold0 (left) and Fold3 (right).

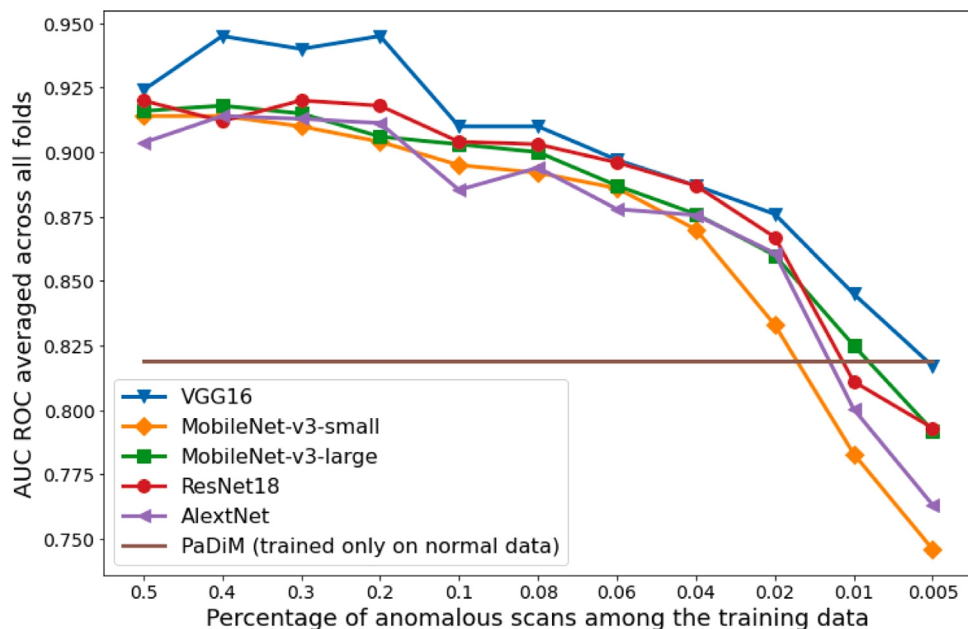


Fig. 8. Average classifier AUC ROC for various percentages of positive samples.

have a depth size of six millimeters. These defects were successfully detected with the threshold lowered to 0.2. However, as can be seen in the confusion matrices in Fig. 7 the benefit of using this model even with the lowered threshold is significant. Since some of the defects are detected only a few times it would be beneficial to improve the model and make it more robust.

As for the results of training the classifiers, experiments show a significant advantage in using the self-supervised anomaly detection approach when the share of positive samples in the training dataset approaches a few percentages of the whole training dataset. The results across all tested classifiers can be seen in Fig. 8. VGG16 architecture achieved the best results for all of the tested shares of the positive samples. If the ratio of defective B-scans in the whole training set is 40%, this model achieves an AUC ROC of 94.5%. However, it is more realistic to expect that only a few percent of the acquired UT data will be anomalous during real-life inspections. Proposed anomaly detection

approaches can still be used in such a setting, unlike the supervised image classification approaches.

5. Conclusions and future work

In this paper, we tested three state-of-the-art methods for anomaly detection in images. We proposed the improvements and evaluated them on our eight-fold real-world ultrasonic NDT dataset. For the Ganomaly model, we proposed cropping-out patches from the images to generate better reconstructions of the normal images and worse for the anomalous images which result in a shy performance improvement. We tested new feature extraction networks for PaDiM and concluded that the performance of the originally used ResNet can be matched using a lot faster MobileNet network. We used the original DifferNet model without any modifications. Taking into account the complexity of our dataset, some of the methods stood very well, and PaDiM method performed the best of all three. It successfully detects all defects in

the dataset and outperforms classifiers when the share of positive samples in the training set is very small. In this work, we highlight the importance of testing the anomaly detection methods on a real-world dataset before using them in production. In the future, the mentioned shortages in the analyzed models should be exploited to develop a more robust method for anomaly detection in ultrasonic images.

While the demand for a high-quality ultrasonic NDE is rising, the time needed to analyze complex structures of some pipelines, vehicles, and power plants is a major drawback. A system such as the one proposed in this work could assist the inspectors and make the inspection more reliable and faster.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was co-funded by the European Union through the European Regional Development Fund, under the grant KK.01.2.1.01.0151 (Smart UTX).

References

- [1] I. Virkkunen, T. Koskinen, O. Jessen-Juhler, J. Rinta-Aho, Augmented ultrasonic data for machine learning, *J. Nondestruct. Eval.* 40 (1) (2021) 1–11.
- [2] L. Posilović, D. Medak, M. Subašić, M. Budimir, S. Lončarić, Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans, *Neurocomputing* 459 (2021) 361–369, <http://dx.doi.org/10.1016/j.neucom.2021.06.094>, URL <https://www.sciencedirect.com/science/article/pii/S09252321221010304>.
- [3] M.E. Tschuchnig, M. Gadermayr, Anomaly detection in medical imaging—a mini review, 2021, arXiv preprint [arXiv:2108.11986](https://arxiv.org/abs/2108.11986).
- [4] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, C. Steger, The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection, *Int. J. Comput. Vis.* 129 (4) (2021) 1038–1059.
- [5] E. Oruklu, J. Saniie, Ultrasonic flaw detection using discrete wavelet transform for NDE applications, in: *IEEE Ultrasonics Symposium*, 2004, Vol. 2, 2004, pp. 1054–1057, <http://dx.doi.org/10.1109/ULTSYM.2004.1417956>.
- [6] S. Sambath, P. Nagaraj, N. Selvakumar, Automatic defect classification in ultrasonic NDT using artificial intelligence, *J. Nondestruct. Eval.* 30 (1) (2011) 20–28.
- [7] J. Ye, S. Ito, N. Toyama, Computerized ultrasonic imaging inspection: From shallow to deep learning, *Sensors* 18 (11) (2018) <http://dx.doi.org/10.3390/s18113820>, URL <https://www.mdpi.com/1424-8220/18/11/3820>.
- [8] N. Munir, H.-J. Kim, S.-J. Song, S.-S. Kang, Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments, *J. Mech. Sci. Technol.* 32 (7) (2018) 3073–3080.
- [9] A. Boikov, V. Payor, R. Savelev, A. Kolesnikov, Synthetic data generation for steel defect detection and classification using deep learning, *Symmetry* 13 (7) (2021) <http://dx.doi.org/10.3390/sym13071176>, URL <https://www.mdpi.com/2073-8994/13/7/1176>.
- [10] D. Medak, L. Posilović, M. Subašić, M. Budimir, S. Lončarić, Automated defect detection from ultrasonic images using deep learning, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 68 (10) (2021) 3126–3134, <http://dx.doi.org/10.1109/TUFFC.2021.3081750>.
- [11] D. Medak, L. Posilović, M. Subašić, T. Petković, M. Budimir, S. Lončarić, Rapid defect detection by merging ultrasound B-scans from different scanning angles, in: *2021 12th International Symposium on Image and Signal Processing and Analysis, ISPA*, 2021, pp. 219–224, <http://dx.doi.org/10.1109/ISPA52656.2021.9552050>.
- [12] L. Posilović, D. Medak, M. Subašić, T. Petković, M. Budimir, S. Lončarić, Flaw detection from ultrasonic images using YOLO and SSD, in: *2019 11th International Symposium on Image and Signal Processing and Analysis, ISPA*, 2019, pp. 163–168, <http://dx.doi.org/10.1109/ISPA.2019.8868929>.
- [13] O. Siljama, T. Koskinen, O. Jessen-Juhler, I. Virkkunen, Automated flaw detection in multi-channel phased array ultrasonic data using machine learning, *J. Nondestruct. Eval.* 40 (3) (2021) 1–13.
- [14] F. Milković, B. Filipović, M. Subašić, T. Petković, S. Lončarić, M. Budimir, Ultrasound anomaly detection based on variational autoencoders, in: *2021 12th International Symposium on Image and Signal Processing and Analysis, ISPA*, 2021, pp. 225–229, <http://dx.doi.org/10.1109/ISPA52656.2021.9552041>.
- [15] H. Kieckhefer, J. Baan, A. Mast, W.A. Volker, Image processing techniques for ultrasonic inspection, in: *Proc. 17th World Conference on Nondestructive Testing*, Shanghai, China, 2008.
- [16] I. Kraljevski, F. Duckhorn, C. Tschöpe, M. Wolff, Machine learning for anomaly assessment in sensor networks for NDT in aerospace, *IEEE Sens. J.* 21 (9) (2021) 11000–11008.
- [17] M. Pocevičiūtė, G. Eilertsen, C. Lundström, Unsupervised anomaly detection in digital pathology using GANs, in: *2021 IEEE 18th International Symposium on Biomedical Imaging, ISBI*, 2021, pp. 1878–1882, <http://dx.doi.org/10.1109/ISBI48211.2021.9434141>.
- [18] K.M. van Hespén, J.J. Zwanenburg, J.W. Dankbaar, M.I. Geerlings, J. Hendrikse, H.J. Kuijff, An anomaly detection approach to identify chronic brain infarcts on MRI, *Sci. Rep.* 11 (1) (2021) 1–10.
- [19] P. Bergmann, M. Fauser, D. Sattlegger, C. Steger, MVTec AD—a comprehensive real-world dataset for unsupervised anomaly detection, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9592–9600.
- [20] A. Krizhevsky, G. Hinton, et al., *Learning Multiple Layers of Features from Tiny Images*, Citeseer, 2009.
- [21] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.
- [22] S. Akcay, A. Atapour-Abarghouei, T.P. Breckon, Ganomaly: Semi-supervised anomaly detection via adversarial training, in: *Asian Conference on Computer Vision*, Springer, 2018, pp. 622–637.
- [23] T. Defard, A. Setkov, A. Loesch, R. Audigier, PaDiM: A patch distribution modeling framework for anomaly detection and localization, in: *International Conference on Pattern Recognition*, Springer, 2021, pp. 475–489.
- [24] M. Rudolph, B. Wandt, B. Rosenhahn, Same same but different: Semi-supervised defect detection with normalizing flows, in: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1907–1916.
- [25] D. Medak, L. Posilović, M. Subašić, M. Budimir, S. Lončarić, Automated defect detection from ultrasonic images using deep learning, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* (2021).
- [26] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, *Int. J. Comput. Vis.* 115 (3) (2015) 211–252.
- [27] D. Rezende, S. Mohamed, Variational inference with normalizing flows, in: *International Conference on Machine Learning, PMLR*, 2015, pp. 1530–1538.
- [28] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [29] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, et al., Searching for mobilenetv3, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1314–1324.
- [30] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* 25 (2012) 1097–1105.
- [31] R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart, The mahalanobis distance, *Chemometr. Intell. Lab. Syst.* 50 (1) (2000) 1–18.
- [32] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).



Luka Posilović received his M.Sc. from the University of Zagreb, Faculty of Electrical Engineering and Computing in 2019. He is currently working as a young researcher in an Image Processing Group in the Department of Electronic Systems and Information Processing and working on his Ph.D. at the same University. His research interests include visual quality control, deep learning object detection, and synthetic image generation.



Duje Medak received his M.Sc. from the University of Zagreb, Faculty of Electrical Engineering and Computing in 2019. He is currently pursuing a Ph.D. degree in the same faculty while working as a researcher in the Image Processing Group in the Department of Electronic Systems and Information Processing. His research interests include image processing, image analysis, machine learning, and deep learning. His current research interest includes deep learning object detection methods and their application in the non-destructive evaluation (NDE) domain.



Fran Milković received his M.Sc. from the University of Zagreb, Faculty of Electrical Engineering and Computing in 2020. He is employed as a junior researcher in the Image Processing Group at the same faculty, where he is also pursuing a Ph.D. degree. At the moment, his research is focused on deep learning anomaly detection methods and the application of such methods in the domain of non-destructive evaluation. His research interests also include image processing, image analysis, machine and deep learning in general.



Marko Budimir received his M.Sc. of physics at the University of Zagreb, Faculty of Science in 2000., and his Ph.D. at Ecole Polytechnique Federale de Lausanne in Switzerland in 2006. He worked at EPFL from 2006 till 2008. From 2008 he is working at the Institute of Nuclear Technology (INETEC). He coordinated many key projects at INETEC and although he is a key person in a company of industry sector he is still working close to the field of science.



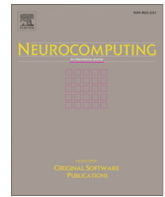
Marko Subašić received the Ph.D. degree from the Faculty of Electrical Engineering and Computing, University of Zagreb, in 2007. Since 1999, he has been working at the Department for Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb, where he is currently working as an Associate Professor. He teaches several courses at the graduate and undergraduate levels. His research interests include image processing and analysis and neural networks, with a particular interest in image segmentation, detection techniques, and deep learning. He is a member of the IEEE of Computer Society, the Croatian Center for Computer Vision, the Croatian Society for Biomedical Engineering and Medical Physics, and the Centre of Research Excellence for Data Science and Advanced Cooperative Systems.



Dr. Sven Lončarić is a professor of electrical engineering and computer science at the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia. As a Fulbright scholar, he received a Ph.D. degree in electrical engineering from the University of Cincinnati, OH in 1994. From 2001–2003, he was an assistant professor at the New Jersey Institute of Technology, USA. His areas of research interest are image processing and computer vision. He was the principal investigator on a number of R&D projects. Prof. Lončarić co-authored more than 250 publications in scientific journals and conferences. He is the director of the Center for Computer Vision at the University of Zagreb and the head of the Image Processing Group. He is a co-director of the Center of Excellence in Data Science and Cooperative Systems. Prof. Lončarić was the Chair of the IEEE Croatia Section. He is a senior member of IEEE and a member of the Croatian Academy of Technical Sciences. Prof. Lončarić received several awards for his scientific and professional work.

Publication 5

D. Medak, L. Posilović, M. Subašić, M. Budimir, S. Lončarić, "DefectDet: a deep learning architecture for detection of defects with extreme aspect ratios in ultrasonic images", *Neuro-computing*



DefectDet: A deep learning architecture for detection of defects with extreme aspect ratios in ultrasonic images

Duje Medak^{a,*}, Luka Posilović^a, Marko Subašić^a, Marko Budimir^b, Sven Lončarić^a

^a University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia

^b Institute for Nuclear Technologies (INETEC), Zagreb, Croatia

ARTICLE INFO

Article history:

Received 30 June 2021

Revised 24 November 2021

Accepted 4 December 2021

Available online 10 December 2021

Communicated by Zidong Wang

Keywords:

Image analysis

Convolutional neural networks

Non-destructive testing

Ultrasonic imaging

Defect detection

ABSTRACT

Non-destructive testing (NDT) is a set of techniques used for material inspection and detection of defects. Ultrasonic testing (UT) is one of the NDT techniques, commonly used to inspect components in the oil and gas industry, aerospace, and various types of power plants. Acquisition of the UT data is currently done automatically using robotic manipulators. This ensures the precision and uniformity of the acquired data. On the other hand, the analysis is still done manually by trained experts. Since the acquired UT data can be represented in the form of images, computer vision algorithms can be applied to analyze the content of images and localize defects. In this work, we propose a novel deep learning architecture designed specifically for defect detection from UT images. We propose a lightweight feature extractor that improves the precision and efficiency of the detector. We also modify the detection head to improve the detection of the objects with extreme aspect ratios which are common in UT images. We tested our approach on an in-house dataset with over 4000 images. The proposed architecture outperformed the previous state-of-the-art method by 1.7% (512 × 512 px input resolution) and 2.7% (384 × 384 px input resolution) while significantly decreasing the inference time.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Non-destructive testing (NDT) is a popular approach for material evaluation and defect detection [1]. It is used for continuous inspection in numerous domains but most commonly in oil and gas industries, power and energy industries, aerospace, and construction. NDT includes a variety of techniques such as ultrasonic, eddy current, thermography, and x-radiography, to name a few. Each of the methods comes with its own advantages and disadvantages and they are sometimes also used jointly in order to increase the probability of finding a defect. None of the NDT techniques cause any damage to the inspected material so the tested component can normally be used after the inspection (if no problems were found) or sometimes even during the inspection. Ultrasonic testing (UT) is one of the most used NDT methods for detection, localization and measurement of flaws present in engineering materials under inspection [2]. UT is simple to perform, yields a precise location of the defect, and in general has a high signal-to-noise ratio [3]. Inspection is performed by the generation and

detection of mechanical vibrations or waves within test objects [4]. There are several ways how this can be done. Pulse-echo (PE), time-of-flight-diffraction (TOFD), and phased array systems are the standard three implementations. A phased array system is a multi-channel ultrasonic system, which uses the principle of a time-delayed triggering of the transmitting transducer elements, combined with a time corrected receiving of detected signals [5]. Using the phased array it is possible to inspect the material from various angles at the same time, which is the main advantage compared to other types of UT probes. Inspecting the component using different angle values makes the process more reliable but it also produces huge amounts of data. Fig. 1 illustrates the principle behind phased array system inspection. Data from UT inspection can be displayed in different forms. As the probe is moved along the surface of the inspected material, at each position it transmits and receives ultrasound waves. The energy of the received ultrasound signal can be shown as a function of time in a representation called A-scan. Each A-scan can be converted into one image column so multiple A-scans can be stacked to form an image representation called B-scan. Since the ultrasound waves are often transmitted at some specific angle, A-scans can also be transferred onto the image at that angle. A view created this way is called

* Corresponding author.

E-mail addresses: duje.medak@fer.hr (D. Medak), luka.posilovic@fer.hr (L. Posilović), marko.subasic@fer.hr (M. Subašić), marko.budimir@inetec.hr (M. Budimir), sven.loncaric@fer.hr (S. Lončarić).

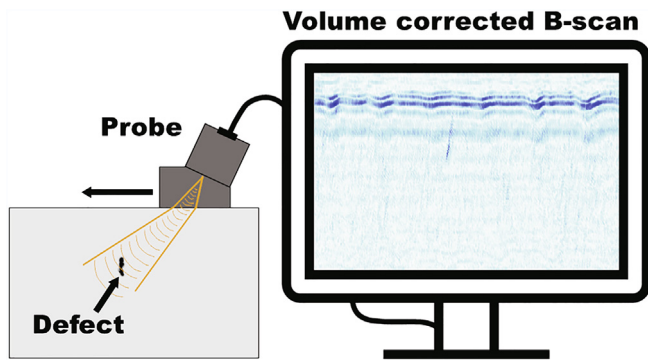


Fig. 1. Illustration of phased array system inspection. An example of volume corrected B-scan (VC-B-scan) is shown on the right side of the figure.

volume-corrected-B-scan (VC-B-scan) and it is the one used in this work.

Data acquired during the UT inspection still has to be analyzed manually by trained experts. This process is laborious and time-consuming. The number of ultrasonic inspections is increasing because most of the existing components require more inspections as time passes and the chance of defect occurrence increases. This also increases the need for optimizing the process of data analysis. Automated analysis can be used to improve the reliability when performing a manual inspection or to speed up the analysis by several orders of magnitude if used independently. The idea of automated analysis of UT data is not new, but the methods proposed so far are not reliable enough to be used in real-life situations. Some of the problems encountered when developing automated analysis of UT data include difficulty with the acquisition of large and diverse datasets, noise, and irregular signal appearances caused by odd defects' shapes and geometry of the inspected component. Recently an improvement in automated UT image analysis was made by employing deep learning approaches for classification and object detection. If an existing architecture is used for object detection, it is assumed that the shapes of the objects that need to be detected will be similar to the common objects found in PASCAL VOC [6] and COCO [7] datasets. Taking into consideration aspect ratios of the objects that need to be detected is very important and in some cases [8,9], proper design of architecture and training procedure leads to improved results. The goal of this work is to design an architecture that can precisely localize defects from B-scans obtained with a phased array probe. The usage of such probes is increasing in real-life inspections and a proper method for analysis of the collected data would be very useful. Depending on the inspected configuration and material, defects' signals can appear very elongated. This can make training difficult because popular anchor-based object detectors [10–12] have a limited number of anchors that are distanced from each other by a fixed value (stride). Having an extreme aspect ratio (>4) leads to a small overlap between the neighboring anchors thus reducing the coverage of an image. This decreases the number of sampled anchors used during the training which can have a negative impact on the detector's performance.

In this work, we propose a deep learning object detector to analyze VC-B-scans and localize all of the visible defects. We start from the state-of-the-art object detection architecture EfficientDet [12] and revise the building components of this model. We first replace the originally used EfficientNet [13] network with our custom feature extraction network. A new model is more precise and uses drastically fewer parameters leading to a faster prediction process. We then redesign the detection head in order to account for extreme aspect ratios that appear in UT images. We propose the usage of asymmetrical feature maps as inputs to the detection

head in combination with lower template anchors stride. This increases the overlap between the template anchors and the ground truth labels and leads to a better model performance with a small computational overhead. The final object detector proposed in this work achieves a mean average precision of 91.3% which is 1.7% more than the previous state-of-the-art model EfficientDet-D0 [14]. Furthermore, the proposed model reduces the needed inference time by more than 30% and has 6 times fewer parameters compared to EfficientDet-D0.

1.1. Contributions

The main contributions of this work are the following:

- A novel feature extractor for the EfficientDet that improves the precision while using six-time fewer parameters.
- A method for detection of objects with extreme aspect ratios based on a modified detection head and dense placement of the anchors.
- A novel deep learning architecture created by joining aforementioned components into a new model that outperforms the previous state-of-the-art in defect detection in ultrasonic images.

1.2. Related work

Analyzing NDT data is a time-consuming process prone to human errors since it depends solely on the experience and the knowledge of the person performing the analysis. In order to assist the experts during the analysis, various methods for defect detection were proposed throughout the years. Developed methods can work with different types of NDT data such as the data acquired during a visual inspection [15,16], thermography inspection [17–19], radiography inspection [20,21], or ultrasonic inspection [22–27]. While the exact implementation depends on the used inspection technique and material, approaches for data analysis and ideas behind them are usually similar. Most of the recent methods rely on convolutional neural networks (CNNs) [15–25,27] since this type of architecture works well with one-dimensional and two-dimensional data such as sequences and images. It was shown that CNNs outperform classical approaches based on hand-crafted features in many general computer vision challenges like PASCAL [6], COCO [7], or ImageNet [28]. The authors of several works [15,3,16,22] tested this hypothesis for NDT data and came to the same conclusion that deep learning approaches outperform classical approaches.

Acquiring the data with non-destructive testing can be a costly process. The equipment required for inspection, as well as the examples of materials containing realistic flaws, are usually very expensive. Since only a fraction of the collected data represents defect signals, collecting a large set of useful images is difficult. This drawback can be solved in three ways: (I) Analysis of A-scans instead of B-scans (II) Application of traditional methods for image analysis that do not require a large dataset (III) Generating or simulating images that can be used to develop a modern deep learning model. The main problem with the A-scan analysis is the lack of context from the surrounding area which makes the decision-making process difficult. The most popular approach for defect detection from A-scans is using the wavelet transform to calculate features and then classifying extracted features using support vector machines (SVM) [29–31] or artificial neural networks (ANN) [32,33]. This way the available data is used solely for classifier training since the feature extraction is predefined. If the available dataset of B-scans is not big enough, some traditional approaches can be used but their performance and generalization are usually not as good as in deep learning approaches. In [34], the authors used the adaptive histogram equalization technique

followed by morphological operations to separate the defective zones from the non-defective zones in the ultrasonic TOFD images. Analyzing TOFD images was also the topic in [35,36] where the authors showed how the parabola matched filter and Hough transform can be used to locate parabolas in TOFD B-scans. In [37], Radon transform was used to detect defects from B-scans that were denoised using the wavelet transform. Having several hundred images already allows for deep learning methods to be employed. In that case, a CNN can be trained with the help of transfer learning [38] and data augmentation. This approach proved to be useful for defect detection from UT images [25,14] and X-ray images [19,20]. Another approach is to use simulated [23,39] or generated [21,24,40] data. While these types of images can be useful for model training, evaluation should be performed on a real dataset to ensure the credibility of the obtained results. In [41] the authors used a generated dataset of B-scans to train a VGG-like classification model. They tested the performance on a separate dataset of real B-scans and reported results almost as good as the one achieved by the human inspectors. The equipment used for acquisition in that work is quite similar to the one used for the collection of the dataset in our work but the tested specimen is different. In [42], the authors tested several deep learning classifiers. The dataset was acquired by a pulsed laser that transmits ultrasonic waves through the material while the contact transducer which is attached to the scanned object captures a series of snapshots of the propagating waves. Among the tested classifiers DenseNet [43] achieved the best result reaching an f1 score of 95.33%.

Defect detection from images can be done on various localization granularity levels. Some of the work [15,17,16,21,24,23] only determine if an image contains a defect or not. This is usually done by employing one of the popular image classification architectures such as VGG [44], Inception [45–47], ResNet [48,49], MobileNet [50–52], or by building a custom CNN. Other works [19,20,25,14] use approaches that determine a coarse location of the defect. This can be done by using object detection architectures which are usually divided into two families: One-stage detectors [53,54,11,12] and two-stage detectors [55–57]. Finally, a fine-grained localization (pixel-wise) can also be obtained as an output [17,18] if a model for semantic segmentation such as U-net [58] is used.

Having a coarse defect location is often good enough. In that case, using an object detection model instead of a semantic segmentation model is better since the inference time for object detectors is usually smaller. In this work, we use EfficientDet [12] architecture as a starting point. This state-of-the-art one-stage object detector was proven to work well with UT images [14]. We change the building blocks of the EfficientDet model by proposing a novel feature extraction network which we use instead of the standard EfficientNet [13] backbone. We also propose a modification of the model's detection head in order to improve the detection rate of objects with extreme aspect ratios. The description of the proposed components is given in Section 3.

2. Dataset

The architecture proposed in this work is developed for defect detection from ultrasonic images. The dataset was obtained by scanning six steel blocks with a phased array probe. Some of the images had a lot bigger width compared to their height. This can cause problems after padding and resizing images to input resolution so we cropped those types of images into multiple patches. Before the cropping is performed a desired width of the patches must be determined. In our case, the desired width was equal to the image height (all of the images that required cropping had a height of 375 px) since we wanted to get patches with an aspect ratio closest to one. We then divided an image into patches such

that the obtained patches have the width as close as possible to the desired width. We also allowed the overlap of 20% between the neighboring patches. The final dataset contains more than 4000 VC-B-scans. The distribution of widths and heights of the images is shown in Fig. 2. The blocks contain 68 defects and each defect can be seen in multiple VC-B-scans (e.g. in various angles or scanning directions). All of the scans combined contain 6637 annotated defects. We do not distinguish between different types of defects so all of them are labeled with the same class. We divided the data into five folds where each fold contains unique defects. All of the appearances of a defect are placed into the same fold to ensure the credibility of the results during the cross-validation. More details about the dataset can be found in [14]. We used the same split so that we could compare results with the previous state-of-the-art approach.

3. Methodology

Deep learning object detectors are usually divided into two categories: one-stage detectors and two-stage detectors. Two-stage detectors used to be more precise but slower compared to the one-stage detectors. The accuracy gap between these two families of object detectors was decreased recently when new one-stage architectures were proposed [12,59]. In [14], it was shown that EfficientDet-D0 is the best choice among tested methods for defect detection from UT images. However, EfficientDet architecture was developed for general object detection on public datasets like [7], PASCAL VOC [6], or ImageNet [28]. Even though EfficientDet achieves good results when used for defect detection in ultrasonic images, we demonstrate that task-specific knowledge can be used to develop an even faster and more precise model.

3.1. Backbone design

General deep learning object detection architectures usually consist of three parts: feature extractor (backbone), feature network (detection neck), and detection head. The first part of the network is used to extract the features from the images. Feature extractors contain millions of parameters even if a simple network such as EfficientNet is used. Having a complex backbone ensures

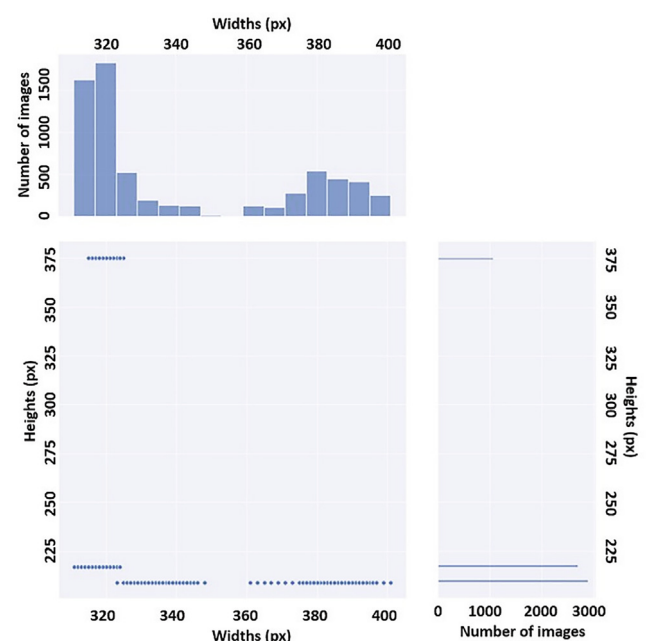


Fig. 2. Distribution of widths and heights of the images from the dataset.

that the extracted features are discriminative. This is very important when the task is to distinguish between dozens of complex objects which is a common requirement in popular object detection datasets. In this work, the goal is to detect only one class (defect). It was shown in [14] that simpler feature extractors can perform better than complex ones. This is why we decided to swap a standard EfficientNet architecture with a simpler novel network illustrated in Fig. 3. We propose an encoder-decoder type of network that looks similar to the U-net [58]. However, there are some key differences between the U-net and the architecture that we propose: (I) The proposed architecture does not use multiple blocks for each resolution level. (II) We do not increase the number of filters as the resolution is decreased. We use 32 filters in the first encoder-decoder part and 64 filters when creating feature maps used as input to the feature network. (III) Our blocks also contain batch normalization and dropout layers which help with the network regularization and improve results. (IV) Recently, to provide high performance of a deep network, several activation functions have been applied in different works [60–63]. However, they can lead to high computational costs and have been applied with different types of images rather than UT images. Instead of ReLU activation, our model uses the swish activation function (used in EfficientDet). The proposed network first downsamples the input features by performing a series of convolutions followed by the max-pooling operation. The decoding part of the network is similar to the one used in Hourglass networks [64]. The feature maps are first upsampled using the nearest-neighbor interpolation. We then perform addition with the feature map of the same resolution from the encoder and pass the resulting feature map through the activation function. We then perform 1x1 convolution followed by batch normalization and activation before upsampling the layer again. Once the original input resolution is reached the feature maps are downsampled again to create feature maps (P3–P7) that are used as an input to the feature network (bidirectional FPN used in EfficientDet [12]). Features P6 and P7 are not actually a part of the backbone. We calculated them using the same implementation used for their calculation in EfficientDet architecture.

If the EfficientNet backbone is replaced by the architecture proposed in this section, the total number of detector parameters is reduced from 3.88 million to 0.53 million. We showed in Section 4.2 that the proposed backbone increases the accuracy while simultaneously decreasing the inference time.

3.2. Detection head design

Many state-of-the-art methods use default boxes (anchors) as a rough starting shape to encapsulate objects and then they perform

an extra step to fit the predicted boxes around the object more tightly. The shape of the anchors is determined from the hyperparameters. Since popular object detection datasets display everyday objects, there is no need for extremely shaped anchors (for example extreme aspect ratio or scale). This is why popular deep learning object detectors are designed to work only with standard anchor shapes or slightly modified ones. The problem with defect detection from ultrasonic images is the extreme aspect ratio of the objects. This can partially be solved by proper calculation of aspect ratio and scales hyperparameters as shown in [14]. In this work, we adopted the same aspect ratios and scales values used in that work. However, setting these hyperparameters does not solve another core problem that appears when using extreme aspect ratios which is the default placement of anchors. The default anchors use stride that is four times smaller than their size (for example feature map P5 uses a default size of anchor 128x128 and strides of 32x32). When common aspect ratios are used, this stride is sufficient to get proper coverage of the image meaning that the template anchors will overlap and no parts of the image will be left uncovered. However, if an extreme aspect ratio is used, a stride that is four times smaller is not sufficient.

We show in Fig. 4 how some of the used anchors with extreme aspect ratios appear once they are placed over the image. We plotted the placement for feature map P5 because it has fewer anchors compared to P3 and P4 so the image is concise and clear. It can be seen from the image that vertical gaps appear for these values of anchors which makes the detection harder. This problem can be solved by introducing more anchors with reduced horizontal spacing (stride) between them. This requires modification of the architecture. Originally used feature maps do not have sufficient resolution to increase the number of anchors so the feature maps of higher resolutions are needed. To accomplish this we shifted the input to the feature network (biFPN). This way the feature maps input to the detection head will also have a sufficient resolution. We use feature maps P2–P4 from the original network and the last two feature maps we calculated the same way that was used to calculate P6 and P7 in the original EfficientDet architecture. Since the defects from ultrasonic images are always elongated in the same direction (vertical), the reduction of stride is not needed in both directions. We inserted extra convolutional layers before the detection head to create asymmetrical feature maps. These convolutional layers downsample the height of the feature map so that the stride in vertical orientation could be left unchanged. Network which is modified to deal with detection of extreme aspect ratio objects is shown in Fig. 5. Proposed modifications can be used regardless of the chosen backbone. In Section 4.2 we showed that using the modified detection head improves the mean

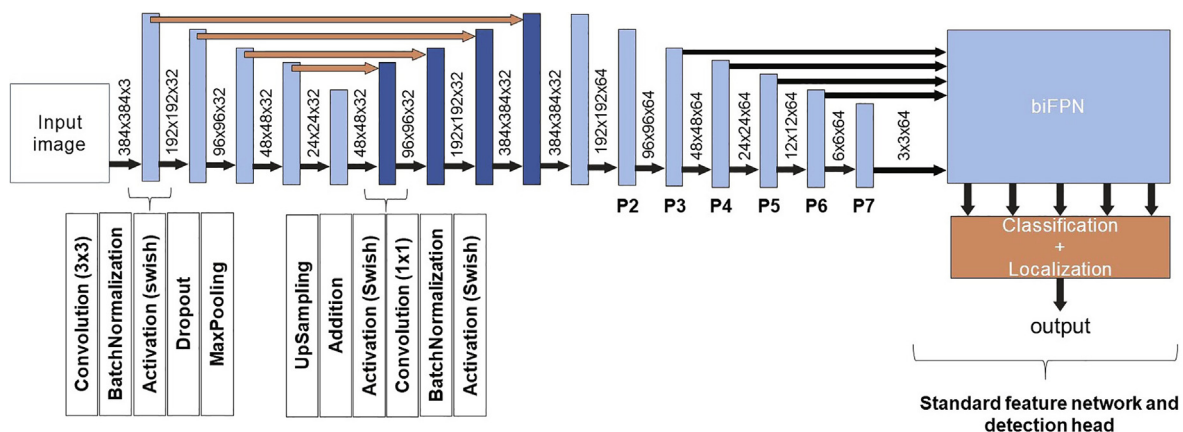


Fig. 3. The architecture of the proposed feature extraction network.

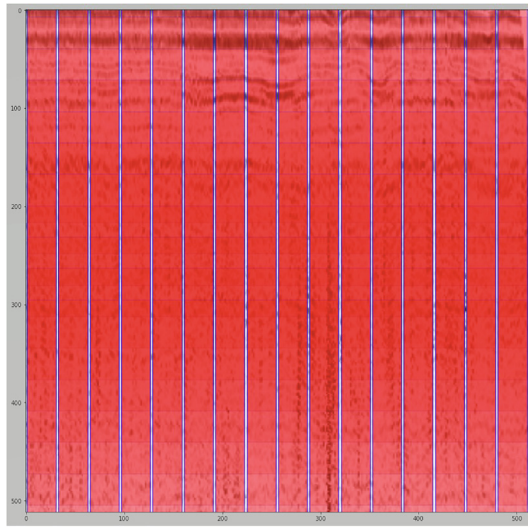


Fig. 4. Placement of anchors with extreme aspect ratios in standard detection head.

average precision on the defect detection task. Described modifications have a small computational overhead if used in combination with the custom feature extractor proposed in the previous section. If the modified detection head is used in combination with EfficientNet the inference time is actually decreased. This happens because the P5 feature map from EfficientNet does not need to be calculated so the number of parameters is reduced by almost three times.

4. Experimental setup and results

We evaluated the methods proposed in this work on our in-house dataset with over 4000 ultrasonic images containing defects. The dataset split and evaluation procedure are the same as in [14]. This allows us to compare obtained results with the top-performing method reported in that work which is EfficientDet-D0. Additionally, We compare the results of our DefectDet with the state-of-the-art family of object detectors YOLOv5 [65]. We run experiments that introduce proposed modules individually. First, we used the EfficientDet model but with the swapped back-

Table 1

Impact of design choices on custom model with input size 384x384 pixels. The first row refer to EfficientDet-D0 from [14].

Custom backbone	Custom detection head	mAP	Inference time (ms)
		0.881	57.0
✓		0.894	39.1
	✓	0.891	45.6
✓	✓	0.908	40.3

Table 2

Impact of design choices on custom model with input size 512x512 pixels. The first row refer to EfficientDet-D0 from [14].

Custom backbone	Custom detection head	mAP	Inference time (ms)
		0.896	62.8
✓		0.900	39.5
	✓	0.902	49.4
✓	✓	0.913	41.7

bone (using architecture introduced in Section 3.1 instead of EfficientNet). We then test the EfficientDet model but with the modified detection head that was proposed in Section 3.2. Finally, we join the two proposed modules into a new deep learning architecture that we named DefectDet. We train the proposed model using the focal loss [11]:

$$FL(p_t) = -(1 - p_t)^{\gamma} \log(p_t) \quad (1)$$

We test the performance of individual modules and their combination with two input image resolutions: 512x512 pixels and 384x384 pixels. The training details are given in the next section.

4.1. Experimental setup

All of the models evaluated in this work were first pretrained on the COCO [7] dataset. As standard practice when using pretrained weights as a starting point, the input RGB images were normalized by subtracting the mean values (0.485, 0.456, 0.406) and dividing them with the standard deviations (0.229, 0.224, 0.225). There was no need for additional intensity normalization which is sometimes done when dealing with unnatural images such as ultrasonic images or magnetic resonance images [66,67]. RetinaNet, EfficientDet, and DefectDet were then trained using the ADAM opti-

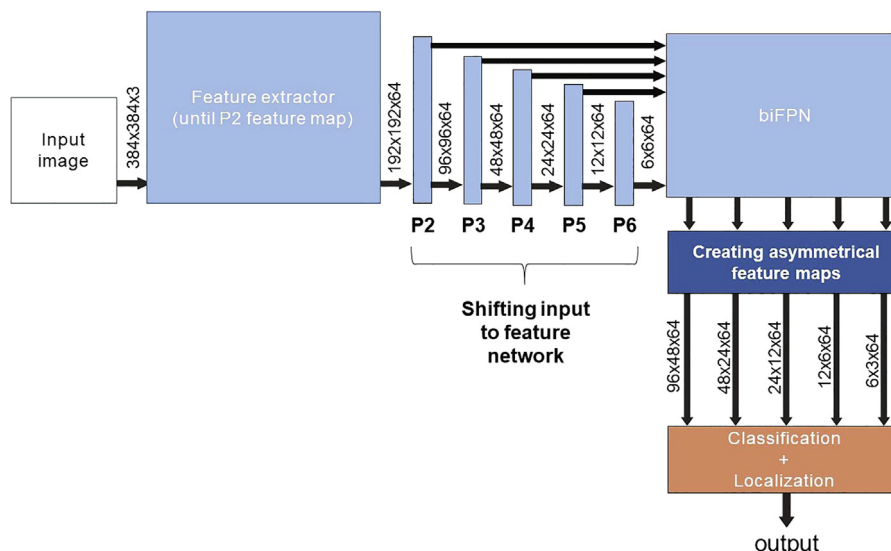


Fig. 5. Feature network and detection head that are customized to handle extreme aspect ratios.

Table 3

Mean average precision (mAP) and inference time for various architectures. Results for EfficientDet models and RetinaNet were taken from [14]. All of the models in the table were tested with input image of 512x512x3 except RetinaNet which achieves better results when the images are only padded.

Model	Fold1	Fold2	Fold3	Fold4	Fold5	Average	Inference time (ms)
EfficientDet-D0	0.937	0.829	0.879	0.943	0.893	0.896	67.2
EfficientDet-D1	0.927	0.793	0.869	0.917	0.901	0.881	75.2
EfficientDet-D2	0.936	0.780	0.826	0.920	0.895	0.871	76.4
RetinaNet	0.872	0.821	0.830	0.901	0.850	0.855	25.6
YOLOv5-s	0.926	0.853	0.827	0.946	0.830	0.876	12.3
YOLOv5-m	0.924	0.813	0.840	0.947	0.875	0.880	16.0
YOLOv5-l	0.922	0.861	0.869	0.944	0.852	0.890	19.5
YOLOv5-x	0.925	0.839	0.838	0.951	0.823	0.875	22.8
DefectDet	0.942	0.869	0.903	0.956	0.894	0.913	35.8

mizer with an initial learning rate of 10^{-3} . We left out 15% of the training subset for validation which was used to reduce the learning rate on a plateau and early stopping of the training. Smaller models (using 384x384 input) were trained with batch size 8 and 500 steps per epoch while the bigger models were trained with batch size 4 and 1000 steps per epoch. Batch size 16 was used for YOLOv5 models and the optimization was done using the SGD since it achieved better results compared to the ADAM optimizer. The rest of the hyperparameters for YOLOv5 were set to default values proposed by its creators. RetinaNet, EfficientDet, and DefectDet were implemented in the Keras library (version 2.2.5) using the Tensorflow backend (version 1.15.0). PyTorch version 1.9.0 was used when testing YOLOv5 models. Inference times from Table 1 and Table 2 were measured on a machine with Titan Xp GPU and CUDA 11.0. Inference times from Table 3 were measured on the same machine but the CUDA 11.2 version. We used mean average precision (mAP) averaged across 5 folds to evaluate the performance of the models. The results are shown and discussed in the following section.

4.2. Results and discussion

The results of the experiments are shown in Tables 1 and 2. The first row of the table corresponds to the EfficientDet-D0 that is the current state-of-the-art in the defect detection task. Swapping the EfficientNet backbone with the one proposed in this work improves the mean average precision (mAP) while simultaneously decreasing the inference time. The mAP is especially increased for the smaller model. Since the images from our datasets are all smaller than 400x400 pixels the difference between the performances of smaller and bigger models should not be big.

For the original EfficientDet architecture the difference between the models of lower (384×384) and higher (512×512) resolution was 1.5%. If the backbone proposed in this work is used this difference is reduced three times. This indicates that the proposed backbone was designed well and that more information is preserved when analyzing the images in their natural resolution. The third row shows the benefits of the modified detection head with asymmetrical feature map inputs and decreased stride. As explained in

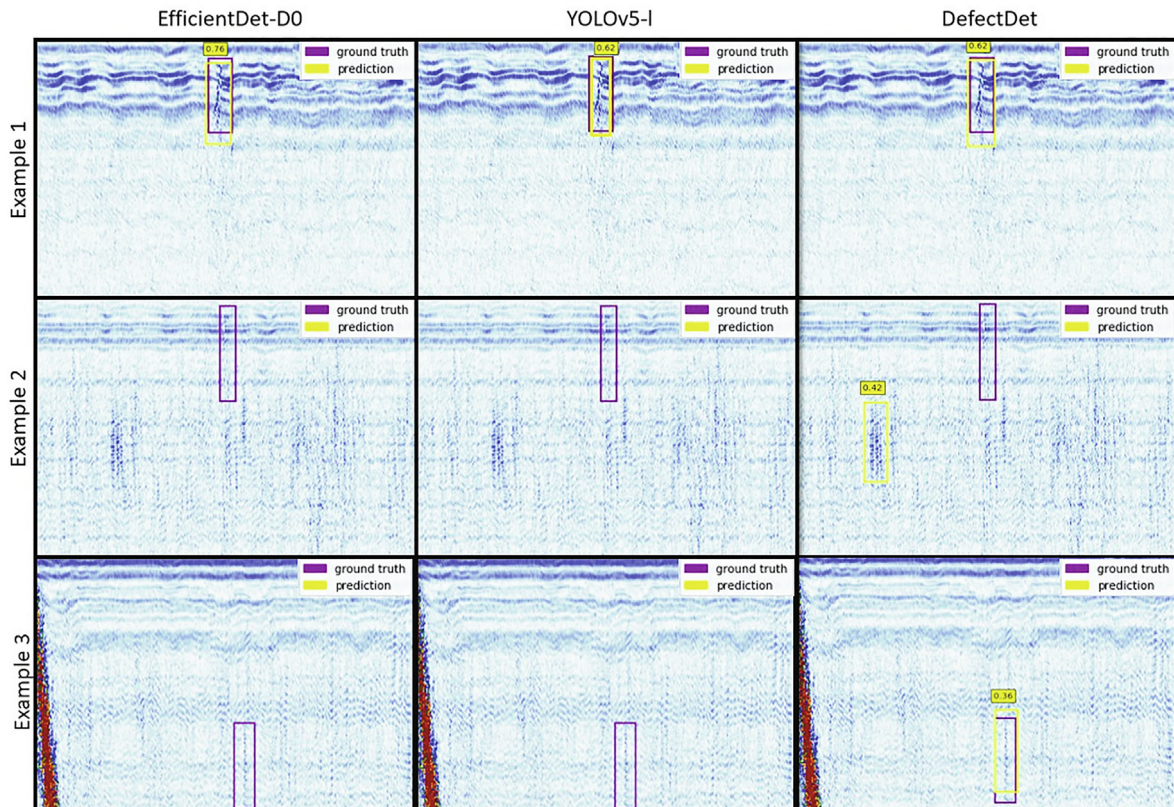


Fig. 6. A few examples of detection on the test images. The threshold for all of the models was 0.3.

Section 3.2, using the custom detection head in combination with EfficientNet actually decreases the inference time since the P5 feature map does not need to be calculated. Finally, the last rows show the performance of our DefectDet which was obtained by joining the custom backbone and custom detection head. If the custom feature extractor is used, a replacement of the standard detection head with the one proposed in this work will lead to a very small increase of inference time. However, the proposed architecture is still more than 30% faster compared to the baseline model EfficientDet-D0. At the same time, the mean average precision increases by 2.7% for the smaller model and 1.7% for the bigger model.

A comparison of the DefectDet with current state-of-the-art models is given in Table 3. A state-of-the-art model YOLOv5 achieves similar results as the previously tested EfficientDet. The proposed DefectDet architecture outperforms all the other tested models for each fold except the fourth fold. Even the DefectDet with input resolution 384×384 surpasses all of the other architectures with greater input resolution. The quickest model among the tested ones was YOLOv5-small but the mean average precision of that model is 3.7% lower than our DefectDet. We also tested the YOLOv5 family of models with a smaller input resolution ($384 \times 384 \times 3$) and all of the tested models achieved less than 87.5% of mAP which is significantly lower compared to the DefectDet with the same input resolution (90.8%). Some prediction examples can be seen in Fig. 6. Shown examples were randomly picked from the second fold test set which is the subset for which the models achieved the lowest mAP on average. None of the tested models were trained using the picked examples. All of the models successfully detected the defect on the first example image. The second example contains a signal that is barely visible and can not be detected without looking at surrounding B-scans so it is not surprising that all of the models failed to detect it. When tested on the third example, EfficientDet and YOLOv5 did not manage to detect a defect while DefectDet managed. Looking at the example images one can notice how hard it is to detect a defect in some of the images, especially when the image is noisier or when it contains geometry signals.

5. Conclusion

In this paper, we propose a novel architecture for detecting defects from ultrasonic images. We designed a simple feature extraction network that enables quicker and more precise detection of defects compared to the previously used models. The proposed feature extractor also reduces the difference in performance between models with different input image resolutions. Furthermore, we proposed a solution to improve the detection of objects with extreme aspect ratios by altering the detection head of the model. With these changes introduced, our defect detection framework outperformed the previous state-of-the-art baseline, and at the same time, required fewer parameters, thus reducing memory usage and inference time. Compared to the state-of-the-art EfficientDet-D0 model, our architecture improves mean average precision by 1.7% for the bigger model, and 2.7% for the smaller model while simultaneously decreasing the inference time by more than 30%. Even though the developed architecture was designed for a specific application, we believe that the proposed ideas can be generalized to other domains with similar problems as defect detection.

CRedit authorship contribution statement

Duje Medak: Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft. **Luka Posilović:**

Software, Data curation, Writing – review & editing. **Marko Subašić:** Conceptualization, Resources, Writing – review & editing, Supervision. **Marko Budimir:** Resources, Data curation, Writing – review & editing, Funding acquisition. **Sven Lončarić:** Resources, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was co-funded by the European Union through the European Regional Development Fund, under the grant KK.01.2.1.01.0151 (Smart UTX). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

- [1] L. Cartz, Nondestructive testing: radiography, ultrasonics, liquid penetrant, magnetic particle, eddy current, ASM International, 1995. URL: <https://books.google.hr/books?id=0spRAAAAMAAJ..>
- [2] J. Veiga, A.A. de Carvalho, I. Silva, J.M.A. Rebelo, The use of artificial neural network in the classification of pulse-echo and tofd ultra-sonic signals, *Journal of The Brazilian Society of Mechanical Sciences and Engineering - J BRAZ SOC MECH SCI ENG* 27. doi:10.1590/S1678-58782005000400007..
- [3] J. Ye, S. Ito, N. Toyama, Computerized ultrasonic imaging inspection: From shallow to deep learning, *Sensors* 18 (11) (2018) 3820, <https://doi.org/10.3390/s18113820>.
- [4] D. Forsyth, 5 - nondestructive testing of corrosion in the aerospace industry, in: S. Benavides (Ed.), *Corrosion Control in the Aerospace Industry*, Woodhead Publishing Series in Metals and Surface Engineering, Woodhead Publishing, 2009, pp. 111–130. doi:10.1533/9781845695538.2.111..
- [5] A. Bulavinov, D. Joneit, M. Kroening, L. Bernus, M. Dalichow, K. Reddy, Sampling phased array - a new technique for signal processing and ultrasonic imaging, *Fraunhofer IZFP-D* 48. doi:10.1117/12.717891..
- [6] M. Everingham, L. van Gool, C. Williams, J. Winn, A. Zisserman, The PASCAL Object Recognition Database Collection. URL: <http://host.robots.ox.ac.uk/pascal/VOC/>, [Online; accessed 1-May-2020] (2012)..
- [7] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: Common objects in context, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014*, Springer International Publishing, Cham, 2014, pp. 740–755.
- [8] D. Kim, S. Kim, S. Jeong, J.-W. Ham, S. Son, J. Moon, K.-Y. Oh, Rotational multipyramid network with bounding-box transformation for object detection, *International Journal of Intelligent Systems* 36 (9) (2021) 5307–5338. doi:10.1002/int.22513..
- [9] J.-B. Hou, X. Zhu, X.-C. Yin, Self-adaptive aspect ratio anchor for oriented object detection in remote sensing images, *Remote Sens.* 13 (7). doi:10.3390/rs13071318..
- [10] J. Redmon, A. Farhadi, Yolo9000: Better, faster, stronger, *CoRR abs/1612.08242*. url:<http://arxiv.org/abs/1612.08242..>
- [11] T. Lin, P. Goyal, R.B. Girshick, K. He, P. Dollár, Focal loss for dense object detection, *CoRR abs/1708.02002*. arXiv:1708.02002. url:<http://arxiv.org/abs/1708.02002>.
- [12] M. Tan, R. Pang, Q.V. Le, Efficientdet: Scalable and efficient object detection, *ArXiv abs/1911.09070*..
- [13] M. Tan, Q.V. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, *CoRR abs/1905.11946*. arXiv:1905.11946. url:<http://arxiv.org/abs/1905.11946>.
- [14] D. Medak, L. Posilović, M. Subašić, M. Budimir, S. Lončarić, Automated defect detection from ultrasonic images using deep learning, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* (2021), <https://doi.org/10.1109/TUFFC.2021.3081750>, 1–1.
- [15] J. Masci, U. Meier, D. Ciresan, J. Schmidhuber, G. Fricout, Steel defect classification with max-pooling convolutional neural networks, in: *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 2012, pp. 1–6, <https://doi.org/10.1109/IJCNN.2012.6252468>.
- [16] Y. Yu, H. Cao, X. Yan, T. Wang, S.S. Ge, Defect identification of wind turbine blades based on defect semantic features with transfer feature extractor, *Neurocomputing* 376 (2020) 1–9, <https://doi.org/10.1016/j.neucom.2019.09.071>.
- [17] Q. Luo, B. Gao, W. Woo, Y. Yang, Temporal and spatial deep learning network for infrared thermal defect detection, *NDT & E Int.* 108 (2019), <https://doi.org/10.1016/j.ndteint.2019.102164> 102164.

- [18] L. Ruan, B. Gao, S. Wu, W.L. Woo, Defectnet: Joint loss structured deep adversarial network for thermography defect detecting system, *Neurocomputing* 417 (2020) 441–457, <https://doi.org/10.1016/j.neucom.2020.07.093>.
- [19] H.-T. Bang, S. Park, H. Jeon, Defect identification in composite materials via thermography and deep learning techniques, *Compos. Struct.* 246 (2020), <https://doi.org/10.1016/j.compstruct.2020.112405> 112405.
- [20] W. Du, H. Shen, J. Fu, G. Zhang, Q. He, Approaches for improvement of the x-ray image defect detection of automobile casting aluminum parts based on deep learning, *NDT & E Int.* 107 (2019), <https://doi.org/10.1016/j.ndteint.2019.102144> 102144.
- [21] X. Le, J. Mei, H. Zhang, B. Zhou, J. Xi, A learning-based approach for surface defect detection using small image datasets, *Neurocomputing* 408 (2020) 112–120, <https://doi.org/10.1016/j.neucom.2019.09.107>.
- [22] M. Meng, Y.J. Chua, E. Wouterson, C.P.K. Ong, Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks, *Neurocomputing* 257 (2017) 128–135, machine Learning and Signal Processing for Big Multimedia Analysis. doi:10.1016/j.neucom.2016.11.066..
- [23] K. Virupakshappa, E. Oruklu, Multi-class classification of defect types in ultrasonic ndt signals with convolutional neural networks, in: *IEEE International Ultrasonics Symposium (IUS) 2019* (2019) 1647–1650, <https://doi.org/10.1109/ULTSYM.2019.8926027>.
- [24] I. Virkkunen, T. Koskinen, O. Jessen-Juhler, J. Rinta-aho, Augmented ultrasonic data for machine learning, *J. Nondestruct. Eval.* 40 (1) (2021) 1–11.
- [25] L. Posilović, D. Medak, M. Subašić, T. Petković, M. Budimir, S. Lončarić, Flaw detection from ultrasonic images using yolo and ssd, in: *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, IEEE, 2019, pp. 163–168..
- [26] N. Munir, H.-J. Kim, S.-J. Song, S.-S. Kang, Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments, *J. Mech. Sci. Technol.* 32 (7) (2018) 3073–3080, <https://doi.org/10.1007/s12206-018-0610-1>.
- [27] N. Munir, H.-J. Kim, J. Park, S.-J. Song, S.-S. Kang, Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions, *Ultrasonics* 94 (2019) 74–81, <https://doi.org/10.1016/j.ultras.2018.12.001>.
- [28] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, *Int. J. Comput. Vis.* 115 (3) (2015) 211–252, <https://doi.org/10.1007/s11263-015-0816-y>.
- [29] V. Matz, M. Kreidl, R. Smid, Classification of ultrasonic signals, *International Journal of Materials* 27 (2006) 145–, doi:10.1504/IJMP.2006.011267..
- [30] A. Al-Ataby, W. Al-Nuaimy, C. Brett, O. Zahran, Automatic detection and classification of weld flaws in tofd data using wavelet transform and support vector machines, *Insight - Non-Destructive Testing and Condition Monitoring* 52 (2010) 597–602, <https://doi.org/10.1784/insi.2010.52.11.597>.
- [31] Y. Chen, H.-W. Ma, G.-M. Zhang, A support vector machine approach for classification of welding defects from ultrasonic signals, *Nondestruct. Test. Eval.* 29(3) (2014) 243–254. doi:10.1080/10589759.2014.914210..
- [32] F. Bettayeb, T. Rachedi, H. Benbartoui, An improved automated ultrasonic nde system by wavelet and neuron networks, *Ultrasonics* 42(1) (2004) 853–858, proceedings of Ultrasonics International 2003. doi:10.1016/j.ultras.2004.01.064..
- [33] S. Sambath, P. Nagaraj, N. Selvakumar, Automatic defect classification in ultrasonic ndt using artificial intelligence, *J. Nondestruct. Eval.* 30 (1) (2011) 20–28, <https://doi.org/10.1007/s10921-010-0086-0>.
- [34] T. Merazi-Meksen, M. Boudraa, B. Boudraa, Ultrasonic image enhancement to internal defect detection during material inspection, in: *MATEC Web of Conferences*, vol. 208, EDP Sciences, 2018, p. 01005..
- [35] P. Petcher, S. Dixon, Parabola detection using matched filtering for ultrasound b-scans, *Ultrasonics* 52 (1) (2012) 138–144.
- [36] P. Bolland, L. Lew Yan Voon, B. Gremillet, L. Pillet, A. Diou, P. Gorria, The application of hough transform on ultrasonic images for the detection and characterization of defects in non-destructive inspection, in: *Proceedings of Third International Conference on Signal Processing (ICSP'96)*, Vol. 1, 1996, pp. 393–396 vol 1. doi:10.1109/ICSP.1996.567285..
- [37] H. Cygan, L. Girardi, P. Akin, P. Simard, B-scan ultrasonic image analysis for internal rail defect detection, *World Congress on Railway Research*, 2003.
- [38] A.S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, Cnn features off-the-shelf: An astounding baseline for recognition, *IEEE Conference on Computer Vision and Pattern Recognition Workshops 2014* (2014) 512–519, <https://doi.org/10.1109/CVPRW.2014.131>.
- [39] R.J. Pyle, R.L.T. Bevan, R.R. Hughes, R.K. Rachev, A.A.S. Ali, P.D. Wilcox, Deep learning for ultrasonic crack characterization in nde, *IEEE Trans. Ultrason. Ferroelectrics Freq. Control* (2020) 1–1 doi:10.1109/TUFFC.2020.3045847..
- [40] L. Posilović, D. Medak, M. Subašić, M. Budimir, S. Lončarić, Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic b-scans (2021). arXiv:2106.04281..
- [41] O. Siljama, T. Koskinen, O. Jessen-Juhler, I. Virkkunen, Automated flaw detection in multi-channel phased array ultrasonic data using machine learning, *J. Nondestruct. Eval.* 40(3), funding Information: Welds were contributed by Suisto Engineering. UT data scanning was contributed by DEKRA. Data augmentation was contributed by Trueflaw. Their support is gratefully acknowledged. Publisher Copyright: 2021, The Author(s). doi:10.1007/s10921-021-00796-4..
- [42] J. Ye, N. Toyama, Benchmarking deep learning models for automatic ultrasonic imaging inspection, *IEEE Access* 9 (2021) 36986–36994, <https://doi.org/10.1109/ACCESS.2021.3062860>.
- [43] G. Huang, Z. Liu, K.Q. Weinberger, Densely connected convolutional networks, *CoRR abs/1608.06993*. arXiv:1608.06993. url: <http://arxiv.org/abs/1608.06993>.
- [44] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *CoRR abs/1409.1556*. url: <https://arxiv.org/abs/1409.1556>.
- [45] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, *CoRR abs/1512.00567*. url: <http://arxiv.org/abs/1512.00567>.
- [46] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: *CVPR*, IEEE Computer Society, 2016, pp. 2818–2826. doi:10.1109/CVPR.2016.308..
- [47] C. Szegedy, S. Ioffe, V. Vanhoucke, Inception-v4, inception-resnet and the impact of residual connections on learning, *CoRR abs/1602.07261*. arXiv:1602.07261. url: <http://arxiv.org/abs/1602.07261>.
- [48] K. He, X.Z. 0006, S. Ren, J.S. 0001, Deep residual learning for image recognition, *CoRR abs/1512.03385*. url: <http://arxiv.org/abs/1512.03385>.
- [49] S. Xie, R.B. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, *CoRR abs/1611.05431*. arXiv:1611.05431. url: <http://arxiv.org/abs/1611.05431>.
- [50] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, *CoRR abs/1704.04861*. arXiv:1704.04861. url: <http://arxiv.org/abs/1704.04861>.
- [51] M. Sandler, A.G. Howard, M. Zhu, A. Zhmoginov, L. Chen, Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation, *CoRR abs/1801.04381*. arXiv:1801.04381. url: <http://arxiv.org/abs/1801.04381>.
- [52] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q.V. Le, H. Adam, Searching for mobilenetv3, in: *The IEEE International Conference on Computer Vision (ICCV)*, 2019..
- [53] J. Redmon, S.K. Divvala, R.B. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: *CVPR*, IEEE Computer Society, 2016, pp. 779–788. doi:10.1109/CVPR.2016.91..
- [54] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. Berg, Ssd: Single shot multibox detector, *Vol. 9905*, 2016, pp. 21–37. doi:10.1007/978-3-319-46448-0_2..
- [55] R. Girshick, Fast r-cnn, in: *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV '15, IEEE Computer Society, USA, 2015, p. 1440–1448. doi:10.1109/ICCV.2015.169..
- [56] S. Ren, K. He, R.B. Girshick, J.S. 0001, Faster r-cnn: Towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6) (2017) 1137–1149. doi:10.1109/TPAMI.2016.2577031..
- [57] T.-Y. Lin, P. Dollár, R.B. Girshick, K. He, B. Hariharan, S.J. Belongie, Feature pyramid networks for object detection, *CoRR abs/1612.03144*. url: <http://arxiv.org/abs/1612.03144>.
- [58] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28..
- [59] A. Bochkovskiy, C.-Y. Wang, H.-Y.M. Liao, YOLOv4: Optimal speed and accuracy of object detection, *ArXiv abs/2004.10934*.
- [60] E. Goceri, Analysis of deep networks with residual blocks and different activation functions: Classification of skin diseases, in: *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2019, pp. 1–6, <https://doi.org/10.1109/IPTA.2019.8936083>.
- [61] L. Nanni, A. Lumini, S. Ghidoni, G. Maguolo, Stochastic selection of activation layers for convolutional neural networks, *Vol. 20*, 2020. doi:10.3390/s20061626..
- [62] E. Goceri, Diagnosis of skin diseases in the era of deep learning and mobile technology, *Comput. Biol. Med.* 134 (2021), <https://doi.org/10.1016/j.combiomed.2021.104458> 104458.
- [63] E. Goceri, Deep learning based classification of facial dermatological disorders, *Comput. Biol. Med.* 128 (2021), <https://doi.org/10.1016/j.combiomed.2020.104118> 104118.
- [64] A. Newell, K. Yang, J. Deng, Stacked hourglass networks for human pose estimation, in: *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, 2016, pp. 483–499..
- [65] G. Jocher, A. Stoken, J. Borovec, NanoCode012, A. Chaurasia, TaoXie, L. Changyu, A. V. Laughing, tkianai, yxNONG, A. Hogan, lorenzomamma, AlexWang1900, J. Hajek, L. Diaconu, Marc, Y. Kwon, oleg, wangaoyang0106, Y. Defretin, A. Lohia, ml5ah, B. Milanko, B. Fineran, D. Khromov, D. Yiwei, Doug, Durgesh, F. Ingham, ultralytics/yolov5: v5.0 – YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations (Apr. 2021). doi:10.5281/zenodo.4679653..
- [66] E. Goceri, Intensity normalization in brain mr images using spatially varying distribution matching, in: *11th Int. Conf. on computer graphics, visualization, computer vision and image processing (CGVCVIP 2017)*, 2017, pp. 300–4..
- [67] E. Goceri, Fully automated and adaptive intensity normalization using statistical features for brain mr images, *Celal Bayar Univ. J. Sci.* 14 (1) (2018) 125–134.



Duje Medak received his M.Sc. from the University of Zagreb, Faculty of Electrical Engineering and Computing in 2019. He is currently pursuing a Ph.D. degree in the same faculty while working as a researcher in the Image Processing Group in the Department of Electronic Systems and Information Processing. His research interests include image processing, image analysis, machine learning, and deep learning. His current research interest includes deep learning object detection methods and their application in the non-destructive testing (NDT) domain.



Marko Budimir received his M.Sc. of physics at University of Zagreb, Faculty of Science in 2000., and his Ph.D. at Ecole Polytechnique Federale de Lausanne in Switzerland om 2006. He worked at EPFL from 2006. till 2008. From 2008. he is working at the Institute of Nuclear Technology (INETEC). He coordinated many key projects at INETEC and although he is a key person in a company of industry sector he is still working close to the field of science.



Luka Posilović received his M.Sc. from the University of Zagreb, Faculty of Electrical Engineering and Computing in 2019. He is currently working as a young researcher in an Image Processing Group in the Department of Electronic Systems and Information Processing and working on his Ph.D. at the same University. His research interests include visual quality control, deep learning object detection, and synthetic image generation.



Sven Lončarić is a professor of electrical engineering and computer science at the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia. As a Fulbright scholar, he received a Ph.D. degree in electrical engineering from the University of Cincinnati, OH in 1994. From 2001–2003, he was an assistant professor at the New Jersey Institute of Technology, USA. His areas of research interest are image processing and computer vision. He was the principal investigator on a number of R&D projects. Prof. Lončarić co-authored more than 250 publications in scientific journals and conferences. He is the director of the Center for Computer Vision at the University of Zagreb and the head of the Image Processing Group. He is a co-director of the Center of Excellence in Data Science and Cooperative Systems. Prof. Lončarić was the Chair of the IEEE Croatia Section. He is a senior member of IEEE and a member of the Croatian Academy of Technical Sciences. Prof. Lončarić received several awards for his scientific and professional work.



Marko Subašić is an associate professor at the Department for Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb, and has been working there since 1999. He received his Ph.D. degree from the Faculty of Electrical Engineering and Computing, University of Zagreb, in 2007. His field of research is image processing and analysis and neural networks with a particular interest in image segmentation, detection techniques, and deep learning

Biography

Luka Posilović was born on April 11th, 1995. in Zagreb, Croatia, where he attended primary and secondary school, XV. gimnazija Zagreb. He continued his education at the University of Zagreb, Faculty of Electrical Engineering and Computing, where he graduated in 2019.

He worked at the INETEC - Institute for Nuclear Technology as a student electronic engineering intern in 2018. Since 2018., he has been working as a Young researcher at the Department of Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb where he participated in the scientific project KK.01.2.1.01.0151 (Smart UTX) and various projects with the industry. His research interests include object detection in images, synthetic image generation and image processing in general.

He is involved in educational activities on courses Digital Logic. He was also involved in the organization of several international conferences, workshops and summer schools. He actively uses English and has basic French understanding. He is an author or co-author of six journal papers and three conference papers.

Luka is also a secretary and a volunteer of the Charity "Kolajna ljubavi".

Published articles

Journal publications

1. **Posilović, L.**, Medak, D., Subašić, M., Budimir, M., Lončarić, S. (2021) Generating ultrasonic images indistinguishable from real images using Generative Adversarial Networks. *Ultrasonics*, 119, 106610, 10 doi:10.1016/j.ultras.2021.106610.
2. **Posilović, L.**, Medak, D., Subašić, M., Budimir, M., Lončarić, S., “Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans”, *Neurocomputing*, Vol. 459, 2021, pp. 361-369., doi: 10.1109/ISPA.2019.8868929.
3. Medak, D., **Posilović, L.**, Subašić, M., Budimir, M., Lončarić, S., “Automated Defect Detection from Ultrasonic Images Using Deep Learning”, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 10, pp. 3126-3134, Oct. 2021, doi: 10.1109/TUFFC.2021.3081750.

- 4.D. Medak, **L. Posilović**, M. Subašić, M. Budimir, S. Lončarić, "DefectDet: a deep learning architecture for detection of defects with extreme aspect ratios in ultrasonic images", *Neurocomputing*, vol. 473, Feb. 2022, pp. 107-115, doi: 10.1016/j.neucom.2021.12.008.
- 5.D. Medak, **L. Posilović**, M. Subašić, M. Budimir, S. Lončarić, "Deep learning-based defect detection from sequences of ultrasonic B-scans", *IEEE Sensors*, Dec. 2021, doi: 10.1109/JSEN.2021.3134452.
6. **Posilović, L.**, Medak, D., Milković, F., Subašić, M., Budimir, M., Lončarić S., "Deep learning-based anomaly detection from ultrasonic images", *Ultrasonics*, Vol. 124, 106737 2022, doi: 10.1016/j.ultras.2022.106737

Conference publications

1. **Posilovic, L.**, Medak, D., Subasic, M., Petkovic, T., Budimir, M., Loncaric, S. (2021) Synthetic 3D Ultrasonic Scan Generation Using Optical Flow and Generative Adversarial Networks. U: 2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA) doi:10.1109/ispa52656.2021.9552069.
- 2.Medak, D., **Posilovic, L.**, Subasic, M., Petkovic, T., Budimir, M., Loncaric, S. (2021) Rapid Defect Detection by Merging Ultrasound B- scans from Different Scanning Angles. U: 2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA) doi:10.1109/ispa52656.2021.9552050.
3. **Posilović, L.**, Medak, D., Subašić, M., Petković, T., Budimir, M., Lončarić, S., "Flaw Detection from Ultrasonic Images using YOLO and SSD", 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), 2019, pp. 163-168., doi: 10.1109/ISPA.2019.8868929.

Životopis

Luka Posilović rođen je 11. travnja 1995. u Zagrebu, Hrvatskoj. Tamo je pohađao osnovnu i srednju školu, XV. gimnaziju u Zagrebu. Nastavio je obrazovanje na Sveučilištu u Zagrebu, Fakultetu elektrotehnike i računarstva, gdje je i diplomirao 2019. godine.

Radio je u tvrtci INETEC - Institut za nuklearnu tehnologiju kao pripravnik elektronički inženjer 2018. godine. Od 2018. radi kao Mlađi istraživač na Zavodu za elektroničko inženjerstvo i obradbu informacija na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu. Tamo radi u projektu KK.01.2.1.01.0151(Smart UTX) i drugim projektima s industrijom. Njegovi interesi su detekcija objekata u slikama, generiranje sintetskih slika i općenito obradba i analiza slike.

Radi kao asistent na predmetu Digitalna logika. Također, bio je uključen u organizaciju nekoliko međunarodnih konferencija, radionica i ljetnih škola. Aktivno se koristi engleskim jezikom i ima osnovno razumijevanje francuskog jezika. Autor je i ko-autor šest članaka u časopisu i tri konferencijska članka.

Luka je također tajnik i volonter Udruge "Kolajna ljubavi".