



**Signal Processing
Systems**
Mekelweg 4,
2628 CD Delft
The Netherlands
<https://sps.ewi.tudelft.nl/>

SPS-2023-00

M.Sc. Thesis

Physics-informed Data Augmentation for Human Radar Signatures

Edoardo Focante

Physics-informed Data Augmentation for Human Radar Signatures

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

ELECTRICAL ENGINEERING

by

Edoardo Focante
born in Osimo, Italy

This work was performed in:

Signal Processing Systems Group
Department of Microelectronics
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology



Delft University of Technology

Copyright © 2023 Signal Processing Systems Group
All rights reserved.

DELFT UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF
MICROELECTRONICS

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled “**Physics-informed Data Augmentation for Human Radar Signatures**” by **Edoardo Focante** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: 24th January 2023

Chairman:

prof.dr.ir. G.J.T Leus

Advisors:

prof.dr.ir. G.J.T Leus

dr.ir. L. de Martín

Committee Members:

dr.ir. F. Fioranelli

Abstract

In recent years, neural networks (NNs) have seen a surge in popularity due to their ability to model complex patterns and relationships in data. One of the challenges of using NNs is the requirement for large amounts of labelled data to train the model effectively. In many real-world applications such as radar, labelled data may be scarce due to the high cost of acquiring measurements together with privacy and security concerns.

To overcome the lack of data, researchers have resorted to data augmentation (DA), a technique that aims to solve the problem at the root by generating new training samples by leveraging the available ones. In computer vision, image transformations and generative networks are used to perform DA. These techniques, however, may lead to the production of physically unfeasible samples that may hinder the generalization capabilities of classifiers in domains where the data has an underlying physical meaning.

Physics-informed machine learning aims to incorporate physical prior knowledge and governing equations of the target domain into the machine learning pipeline to improve the performance of NNs in fields with limited available data but with well-defined physical models. In this thesis, physics-informed DA in the radar domain is addressed to improve the task of classifying armed and unarmed walking individuals through micro-Doppler spectrograms.

To begin with, the usage of model-driven micro-Doppler radar simulations to improve the existing generative augmentations is investigated. After introducing several generative NN architectures, the quality and diversity of the produced synthetic spectrograms are evaluated together with their effect on the downstream classification task.

Next, the impact of visual transformations on micro-Doppler spectrograms is studied. Their effect on the underlying physics of the micro-Doppler spectrograms and their impact on feature extraction is assessed. Based on the results, a new DA technique is devised to improve the feature extraction process by informed segmentation of the input spectrogram, halving the size of the NN model and reducing the risk of overfitting.

Acknowledgments

I would like to express my sincere gratitude to all the people who made this thesis possible. First to my advisor, Prof. Geert Leus, for his guidance and support throughout the course of this research. His expertise and timely feedback have been invaluable in helping me navigate the complexities of this project.

This thesis would also not have been possible without the feedback of my supervisors, Dr. Lilian de Martín and Dr. Mario Coutiño. Their feedback and support throughout this journey have been invaluable in helping me navigate the complexities of this project. I am especially grateful for all their guidance and mentorship, which made me grow personally and academically beyond my expectations and always pushed me to strive for more.

I would also like to extend my thanks Dr. Francesco Fioranelli, for being part of my thesis committee and for his constructive feedback and valuable suggestions.

I am grateful to the people at the SPS department for providing a stimulating and welcoming environment, both on a personal and academic level. Special thanks also to the students and staff of the MS3 department and the colleagues from TNO, for all the joyful moments shared and the insightful discussions we had.

Finally, I would like to thank my family and all my friends for their unwavering support and encouragement throughout this journey. This work would not have been possible without you.

Edoardo Focante
Delft, The Netherlands
24th January 2023

Contents

Abstract	v
Acknowledgments	vii
1 Introduction	1
1.1 Motivation	1
2 Literature Review	3
2.1 Deep Learning	3
2.2 The Problem of Overfitting	3
2.3 Transfer Learning	5
2.4 Data Augmentation	5
2.4.1 Image Manipulation	6
2.4.2 Neural Style Transfer	8
2.4.3 Generative Neural Networks	10
2.4.4 Simulated Data	12
2.5 Physics-Informed Machine Learning	13
2.6 Use Case: Classification of Armed and Unarmed Personnel	14
2.7 Literature Summary and Thesis Objective	15
3 Dataset	17
3.1 Micro-Doppler	17
3.1.1 The Doppler effect	17
3.1.2 Micro-Doppler	17
3.1.3 STFT	18
3.2 Experimental Setup	19
3.3 Simulated Data	19
4 Generative Augmentations from Simulations	23
4.1 Evaluation Method	23
4.2 Convolutional Mapping	24
4.3 GANs	27
4.4 GANs from simulations	30
4.5 GANs from simulations plus noise	33
4.6 Multi-branch GAN	36
4.7 Effect on the downstream classification task	40
5 Image Manipulation for Radar Spectrograms	45
5.0.1 Rescaling in time and frequency	45
5.0.2 Circular time shift	47
5.1 Evaluation	48
5.2 Classification of a single step	50

5.2.1	Conclusion	53
6	Conclusions	57
6.1	Future Work	57
A	Classification	59
A.1	Non-NN Classifier	59
A.2	Transfer Learning	60
A.3	Convolutional classifier	61
B	Acquiring the Dataset	65
B.1	Radar Parameters	65
B.2	Generating the simulations	65
C	Generative Models	67
C.1	Convolutional Mapping	67
C.2	GAN	68
C.3	GAN from simulations	70
C.4	GAN from simulations plus noise	71
C.5	Multi-branch GAN	71

List of Figures

2.1	Training and validation loss curves for a neural network classifier over subsequent training epochs. (a) shows curves corresponding to a correct fitting of the training data while (b) shows signs of overfitting from epoch 60	4
2.2	(b) shows different augmentations of the original image depicted in (a). The augmentations include (first row) random cropping, (second row) rotations and (third row) changes in brightness and colour hue. The source image is taken from [39].	6
2.3	(c) The result of applying neural style transfer on (a) a visual image from [39] used for as content and (b) Composition VII by Wassily Kandinsky used as style.	8
2.4	General structure of a generative adversarial network, where the generator G takes a noise vector z as input and produces a synthetic output $G(z)$. The discriminator takes both the synthetic input $G(z)$ and true sample x as inputs and predicts whether they are real or fake. Image taken from [50]	11
2.5	Overview of the different solutions presented in the literature study. Shown in red are the areas of contribution of this work	16
3.1	Framed spectrograms from the measured datasets belonging to (a) a subject walking hands-free and (b) one carrying a rifle. The micro-Doppler response from the arm is highlighted in (a).	20
3.2	Reference points used to generate the human walking model [68]	20
3.3	Simulated spectrogram generated using the model proposed in [68]. Part of the returns from the foot, the arm and the torso are highlighted.	21
4.1	Distribution of the SSIM between every pair of experimental spectrograms belonging to the walking class	24
4.2	Overview of the training process of the convolutional mapping. The blue bounding box highlights the fact that the input simulations have to be paired with experimental samples that share the same physical parameters in order to compute the MSE loss.	25
4.3	Visual demonstration of the spectrogram generated via convolutional mapping. Synthetic spectrograms (middle row) are generated from the respective input simulation (top row). The bottom row shows the experimental spectrograms paired to the respective input simulation and act as ground truth to compute the MSE.	26
4.4	Distribution of the SSIM between each pair of synthetic spectrograms generated via the convolutional mapping and between pairs of experimental spectrograms.	27
4.5	Overview of the training process of a GAN.	28

4.6	A comparison between real experimental spectrograms (top row) and synthetic ones (bottom row). The synthetic spectrograms were produced by random sampling of the generator’s latent space. The experimental spectrograms are included here for comparison and are used only in the training phase of the GAN.	28
4.7	Distribution of the SSIM between each pair of synthetic spectrograms generated via GAN and between pairs of experimental spectrograms.	29
4.8	Comparison between the distributions of (a) gait cycle duration, (b) mean velocity separately of 500 experimental and synthetic spectrograms. (c) shows the joint distribution of the two parameters across the analysed spectrograms.	30
4.9	Overview of the structure of the GAN realised using simulations as latent vectors.	31
4.10	Visual demonstration of the spectrogram generated via a GAN with simulations as an input. Synthetic spectrograms (middle row) are generated from the respective input simulation (top row). The bottom row shows the experimental spectrograms paired to the respective input simulations and act as ground truth for comparison.	32
4.11	Distribution of the SSIM between each pair of synthetic spectrograms generated via GAN from simulations and between pairs of experimental spectrograms.	33
4.12	Overview of the structure of the GAN realised using simulations overlaid with noise as input.	34
4.13	Visual demonstration of the spectrograms generated via a GAN with simulations plus noise as an input. Synthetic spectrograms (middle row) are generated from the input simulation (top row) and 5 different instances of overlaid noise. The bottom row shows the experimental spectrogram paired to the respective input simulation and acts as ground truth for comparison.	34
4.14	Distribution of the SSIM between each pair of synthetic spectrograms generated via GAN from simulations plus noise and between pairs of experimental spectrograms.	35
4.15	Overview of the structure of the GAN realised using two generator branches responsible for processing the input simulations and a random noise vector.	36
4.16	Visual demonstration of the spectrograms generated via a GAN with multiple generator branches. Synthetic spectrograms (middle row) are generated from the input simulation (top row) and 5 different instances of input noise. The bottom row shows the experimental spectrogram paired to the respective input simulation and acts as ground truth for comparison.	37
4.17	Distribution of SSIM between each pair of synthetic spectrograms generated via multi-branch GAN and between pairs of experimental spectrograms.	38

4.18	Overview of the input and outputs at different stages of the multi-branch GAN architecture. (a) The input simulation is fed into the network to produce (b) the synthetic spectrogram. (c) The output of the simulation branch is responsible for the background of the image while (d) the output of the noise processing branch reconstructs most of the Doppler returns of the limbs.	39
4.19	Overview of the augmentation pipeline used to assess the impact of generative augmentations.	41
4.20	(a) Splitting of the dataset to account for different degrees of data scarcity, (b) when also including synthetic data. The blue and yellow portions are used for training, while the red one is used for testing. The green portion represents the experimental data that is progressively removed to simulate a data-scarce scenario. The validation set is taken from 20% of the experimental data assigned for training shown in blue.	41
4.21	Accuracies achieved in the downstream task using different generative augmentations, averaged over 10 data folds. The standard deviations are omitted for clarity.	42
5.1	Examples of rescaling in time and frequency domain.	46
5.2	Examples of circular shifts in time domain.	47
5.3	Mean and standard deviation of accuracy scores over 100 data folds using different augmentation techniques.	48
5.4	Saliency maps for different classifiers. (a) The input spectrogram of a walking person is classified. The different saliency maps are shown when using (b) a classifier trained on 80% of the available subjects, (c) 10% of the available subjects, (d) 10% of the available subjects but after applying the rescaling augmentation and (e) 10% of the available subjects after applying the circular timeshift augmentation. The red dashed lines show the separation between different steps.	49
5.5	Illustration of the process of dividing the spectrogram into separate steps	52
5.6	Mean and standard deviation of accuracy scores over 100 data folds when using a classifier trained on single steps, with independent and combined assessment.	53
5.7	(a) Three steps belonging to the same original spectrogram and (b) their respective saliency maps when using a classifier with 10% of the training data available	54
5.8	Mean and standard deviation of accuracy scores over 100 data folds using different augmentation techniques.	55
A.1	Structure of the convolutional classifier	63
A.2	Mean accuracy and standard deviation of the convolutional classifier over 10 data splits for different levels of data scarcity	64
B.1	Examples of the estimation process of the characterising parameters of the micro-Doppler simulation	66

C.1	Structure of the convolutional mapping	67
C.2	Synthetic spectrograms generated with convolutional mapping when trained with progressively decreasing amount of samples	68
C.3	Structure utilised for the proposed GAN approach	69
C.4	Synthetic spectrograms generated with GAN network when trained with progressively decreasing amount of samples	70
C.5	Structure of the generator used in the GAN from simulations approach	71
C.6	Synthetic spectrograms generated with GAN with simulations as input when trained with progressively decreasing amount of samples	72
C.7	Synthetic spectrograms generated with GAN with simulations plus noise as input when trained with progressively decreasing amount of samples	73
C.8	Structure of the generator for the simulation branch in the multi-branch GAN approach	73
C.9	Synthetic spectrograms generated with GAN with simulations plus noise as input when trained with progressively decreasing amount of samples	74

List of Tables

4.1	Mean relative error of mean velocity and gait cycle duration between synthetic outputs and input simulations	43
A.1	Performance of different non-NN classifiers	60
A.2	Training parameters of the convolutional classifier	61
A.3	Performance of different transfer learning models	61
A.4	Training parameters of the convolutional classifier	62
C.1	Training parameters of the convolutional mapping	68
C.2	Training parameters of the GAN generator and discriminator	69
C.3	Training parameters of the convolutional mapping	70

1.1 Motivation

Neural networks (NNs) [1] and data-driven approaches have surged in popularity in recent years. Due to their ability to perform feature extraction and inference directly from the data, NNs have seen application in numerous fields, such as computer vision [2, 3], healthcare [4], and natural language processing [5]. Among these, the domain of radar (radio detection and ranging) has not been an exception [6]. In fact, NNs have been employed to solve various problems in the field of radar, such as waveform design [7–9], resource allocation [10, 11], and automatic target recognition [12–14]. For the latter, in particular, NNs managed to achieve great improvements over other machine learning (ML) approaches and have been showing excellent results when classifying synthetic aperture radar (SAR) images [15–17], micro-Doppler spectrograms [13, 14, 18] and high-resolution range profiles [19–21].

Different from other domains such as computer vision, the amount of available data in the radar field may be too scarce to guarantee the reliable training of NNs. This scarcity can lead to classifiers with little generalisation capabilities, which will hence experience worsened performances when classifying unseen data samples.

The cause of the lack of available training data in the radar domain is of a dual nature. On one hand, acquiring new labelled measurements requires a lot of equipment, personnel and time, which makes the entire procedure difficult under budget and time constraints. On the other hand, samples of certain targets and situations may not be easily acquirable. These include, for example, images of vehicles and armaments belonging to foreign countries in defence scenarios.

To overcome these challenges, researchers have been exploring new techniques to improve the generalisation abilities of NNs, such as data augmentation (DA) [22]. DA aims to tackle the problem of lack of data at its root, by generating new training samples from the available ones. In computer vision, techniques such as flipping, cropping, stretching and rotating are commonly used to create new training images. However, when it comes to two-dimensional data in the radar domain, such as spectrograms or SAR images, these transformations clash with the physical nature of the data and are likely to produce physically unfeasible samples.

As a consequence, researchers have also explored the usage of generative models to produce synthetic data in radar scenarios. Generative models are particular NNs such as generative adversarial networks (GANs) [23] and variational autoencoders (VAEs) [24]. These architectures aim to learn the underlying distribution of the available data and to generate new synthetic samples by sampling such distribution. For this reason, despite being trained using real data, GANs and VAEs are still not guaranteed to produce physically feasible samples and may require additional measures to evaluate and filter

the outputs [25].

Besides the generative tools, the radar domain offers a consolidated theoretical framework and is equipped with plenty of tools for simulating radar scenarios and producing theoretical returns for specific input parameters. However, if model-driven simulations are too distant from real measured samples, these may not prove helpful to tackle the classification tasks [26]. This stems from the fact that, in order to remain computationally feasible, the theoretical models must consider several simplifications of the scenario they are trying to represent. Despite these simplifications, however, theoretical simulations may still retain sufficient physical information to contribute to the augmentation process.

In several fields, to enhance the performance of NNs and ensure the physical feasibility of the produced results, NNs have been enhanced with domain-specific knowledge regarding the physics of the area of application. This approach is known as physics-informed machine learning (PIML) [27], where physics refers to prior knowledge such as governing laws, properties and simulations of the target domain. PIML is adopted in those domains with limited data but available physical models, and aims to incorporate prior physical knowledge into the NN architectures. In the case of GANs and VAEs, the structure of the network and the loss function are commonly designed to be adapted to the physics of the problem, ensuring better physical feasibility of the output. In the scope of exploring PIML for DA in the radar domain, it would be of interest to study a way to combine the physical content of the simulations with the generative power of NNs to obtain synthetic samples that are physically plausible.

In this thesis, we answer the following main research questions:

- Can simulations generated from theoretical models be enhanced and used to improve the performance of NN classifiers in the radar domain?
- Can micro-Doppler spectrograms be augmented using image transformations that are aware and compatible with the physical content of their target?

As a way to assess the effect of the proposed augmentations, the downstream task of classifying micro-Doppler spectrograms of armed and unarmed individuals for surveillance purposes will be utilised.

In this document, chapter 2 will provide information regarding the related work and background knowledge for this thesis and chapter 3 will describe the data used for the experiments discussed in this work. Chapter 4 will go over the research done to explore augmentations via physics-informed generative NNs, while chapter 5 will cover all the experiments regarding image manipulation for DA in the radar domain. Finally, chapter 6 will summarize the conclusions and provide some ideas for future work.

2.1 Deep Learning

Deep learning (DL) [28] is a subfield of ML that involves the use of algorithms and models that are composed of multiple layers of interconnected nodes. When compared to classic feature-based methods, DL aims to solve the process of defining, selecting and evaluating the features using a single multi-layer architecture.

Neural networks (NNs) are a type of DL algorithm designed to recognize patterns and make decisions in a way that is similar to how the human brain does, by extracting features from the input data and combining them to recognize complex relationships and patterns. This is done by a stack of multiple layers that perform operations such as multiplications, convolutions, poolings, or activations. The multilayer architecture also allows the optimization of parameters such as the kernels of the convolutions and the multiplication weights. This is done via backpropagation [1] of the gradient of a defined loss function with respect to the model parameters. In such a way, NNs are able to leverage the available data to learn complex non-linear relationships between input and output data.

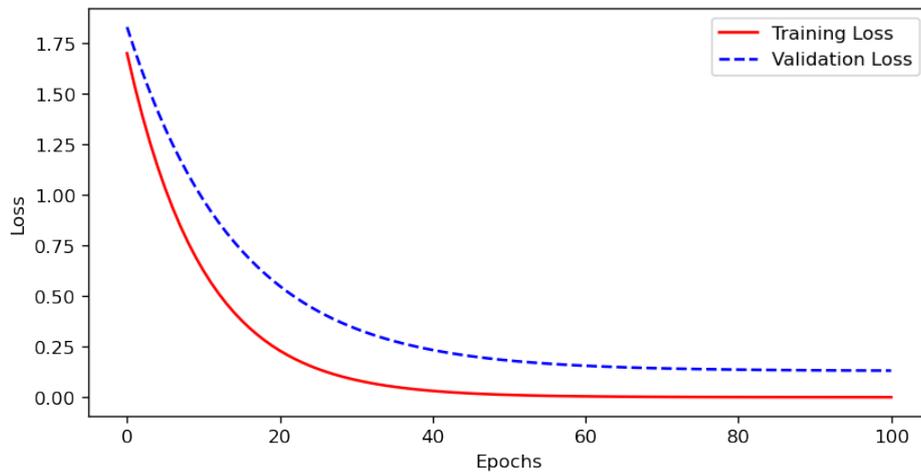
2.2 The Problem of Overfitting

Together with NNs' rise in popularity comes the growing need for large labelled datasets. In fact, when NNs are trained using only a limited amount of data, their ability to infer the relationship between inputs and outputs is heavily reduced and could lead to overfitting [29]. Overfitting occurs when a NN learns to fit perfectly the training data. In presence of overfitting, the model tends to memorize the desired output associated with each training sample rather than learning the underlying patterns and relationships between input and output data. When this happens, the NN will present an excellent level of performance on the training data but will struggle to generalize to unseen samples, leading to mediocre results in real application scenarios.

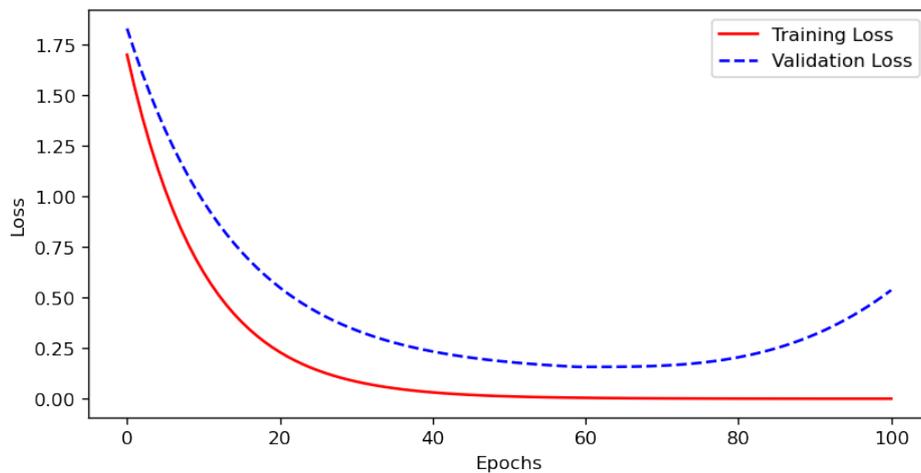
For this reason, it is crucial to monitor the performance of the trained model using unseen data in the form of a validation dataset on which the generalization capabilities of the model can be assessed during training. One way to diagnose overfitting is by looking at the output loss when passing the training data and the validation set through the model over different epochs. In ML, an epoch is a complete cycle through the dataset during the training phase.

Figure 2.1 shows plausible loss curves for a NN model. In the first 60 epochs the training and validation losses decrease jointly in both figure 2.1a and figure 2.1b, meaning that the optimization process is leading to improved performance for both the training and the unseen validation data. After the 60th epoch, however, figure 2.1b

shows an increase in the validation loss that is not matched by the training one, meaning that while the model is optimizing its weights to perform its task on the training data it is worsening its performance on the unseen validation samples. This indicates a loss in the generalization capabilities of the model, which is the main symptom of overfitting.



(a)



(b)

Figure 2.1: Training and validation loss curves for a neural network classifier over subsequent training epochs. (a) shows curves corresponding to a correct fitting of the training data while (b) shows signs of overfitting from epoch 60

To keep leveraging the powerful feature extraction capabilities of NNs, researchers have resorted to several strategies which enable good generalisation capabilities in data-scarce scenarios. These can include basic regularization techniques applied during the training as well as more sophisticated approaches that affect the whole pipeline. In recent work in the radar domain, two of the most common ways to deal with the lack of training data have been transfer learning and data augmentation.

2.3 Transfer Learning

In machine learning, transfer learning (TL) aims to leverage the volume and quality of data in a source domain to enhance the performance of NNs in a data-scarce domain referred to as the target domain.

While [30] and [31] provide extensive reviews on the topic of transfer learning and its different categorizations, in the scope of this thesis, only inductive TL will be discussed, as it has the widest range of applications in the classification of micro-Doppler radar spectrograms. In the inductive TL approach, a large volume of available labelled data in the source domain is leveraged to induce a predictive model that is applied to the target domain.

In practice, when applying inductive TL a model is trained first on data that belongs to a source domain that has an abundance of training samples. After this first training phase, where the layers of the model are initialized and are able to extract the relevant features from the data, the model is fine-tuned on the target dataset that has only limited data available. In such a way, the prior knowledge on a more NN-friendly domain can be leveraged, reducing the risk of overfitting the model to the samples of the limited target dataset.

In the task of classifying radar images, due to the visual nature of the time-Doppler spectrograms and SAR images, TL models are typically pre-trained using visual images and are then used to classify the target spectrograms. In [32, 33], for example, the authors apply the AlexNet model [34] trained on the ImageNet dataset [35] to classify SAR images and micro-Doppler spectrograms respectively. The implicit assumption in their works is that the features extracted from the visual images in the ImageNet data will be also relevant for the classification of samples in the radar target domain.

In fact, in order to apply TL successfully, the required condition is that the target domain and the domain used for pre-training must share some degree of similarity with each other. Unfortunately, there is no clear criterion for determining whether a source domain retains valuable and transferable knowledge for the target domain, as it all comes down to the use case that is being studied.

If the assumption of similarity does not hold or is weak, there is the risk that the transferred model will see a large reduction of performance in the target domain. This phenomenon is called negative transfer [36] and is particularly present for inductive TL when the relevant features of the source domain do not match the ones of the target data. While improvements have been proposed to avoid negative transfer [37], the effectiveness of TL is still completely dependent on the application case, as it is tied to the relationship between the source and target domain.

2.4 Data Augmentation

Unlike TL, DA aims to tackle data scarcity at its root. DA, in fact, is an umbrella term that covers all those techniques that aim to generate new training samples from the limited available ones. [22] and [38] provide an extensive review of DA for images in DL scenarios. In the following sections, different approaches to DA will be covered, with special attention to literature focusing on the augmentation of micro-Doppler

spectrograms in the radar domain.

2.4.1 Image Manipulation

Image manipulation is a form of DA that relies on applying visual transformations to the limited samples of the available dataset in order to generate new data to train the model with. Transformations that fall under this category include flipping, cropping and rotation of the images, as well as adding noise and changing characteristics such as contrast and brightness. More sophisticated augmentations involve the random swapping of neighbouring pixels or erasing random patches from the training images. Figure 2.2 shows different augmentations applied to a visual image.

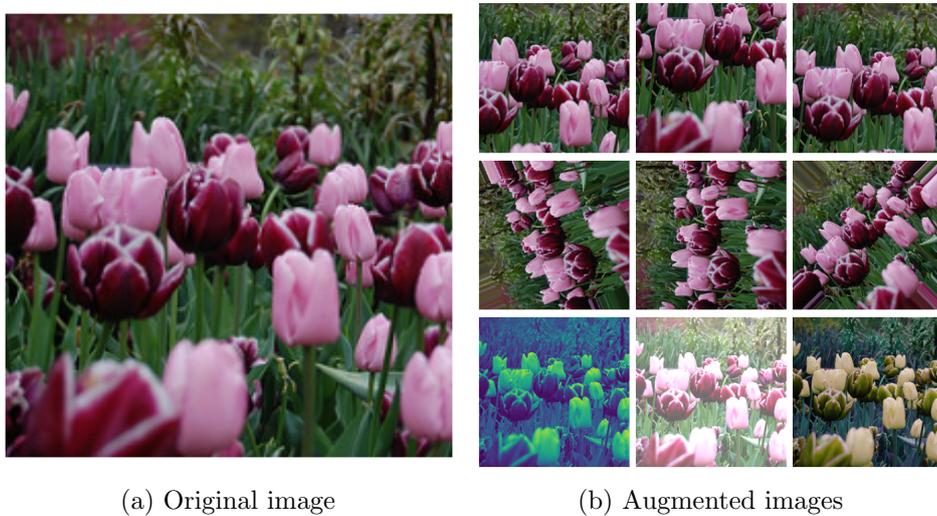


Figure 2.2: (b) shows different augmentations of the original image depicted in (a). The augmentations include (first row) random cropping, (second row) rotations and (third row) changes in brightness and colour hue. The source image is taken from [39].

The objective of image manipulation is to generate new training samples that share the core high-level features of the original ones, but present them in a different realisation. In a classification task, this is done in order to direct the focus of the classifier on relevant features rather than non-characterizing aspects of the input. For instance, rotating the training samples can aid an NN-based classifier to extract the defining feature of those samples, making the classification robust to different orientations at which the input may appear.

Without this kind of augmentation, there is the risk for a classifier to tie non-defining properties of the training samples to the desired output label. For example, a classifier tasked with the identification of different objects and trained only with images of red cars may struggle to correctly classify cars of different colours, because the training process resulted in a dependency between a non-characterizing feature such as the colour and the output label. With data augmentation, instead, the input training data is used to condition the classifier to be robust to a certain property by indirectly guiding the models' attention to relevant features using artificially diversified training samples.

Image manipulation has the advantage of being simple to implement and cheap in terms of computational effort, but they also need to be applied consciously with respect to the labels and the meaning of target data. For example, in the common ML problem of MNIST [40] digit classification, rotating a '6' digit could turn it into a '9' digit, effectively changing its classification label. In addition to that, the axes of images that are related to physical representations typically have a physical meaning that cannot be ignored. In the particular case of radar spectrograms, for example, they represent time and frequency. It is clear that applying transformations such as rotations will completely destroy the underlying physical meaning of these axes and with that the spectrogram's representation.

While the impact of these transformations on the classification task cannot be known a priori, it is clear that applying image manipulation must be done consciously with respect to the meaning of the target image. In the radar literature, several examples of DA by means of image manipulation can be found.

In their work, Kern and Waldschmidt [41] perform image manipulation on micro-Doppler spectrograms trying to be mindful of the physics of the problem. In particular, they apply a scaling in both frequency and time to generate new spectrograms from the available data, achieving an increase of up to 4% in the achieved accuracy in data-scarce scenarios. In the frequency domain, they apply a rescaling of the Doppler components related to the outer limbs but keep the frequencies related to slower-moving body parts unaltered to obtain more realistic spectrograms. They also apply compression and expansion in time, which for spectrograms of walking people, corresponds to a shorter and longer duration of the gait cycle.

According to the nature of the human walking motion [42], a reduction in the duration of the gait cycle of a subject should imply an increase in the mean velocity along the frequency axis. In [41], however, this factor is not accounted for, meaning that the only way to physically justify the gait speed change is in a variation of the target's height. While the approach doesn't necessarily make the new samples physically defunct, it is very likely to produce spectrograms of individuals with unrealistic heights that will never appear in experimental recordings.

She et al. [43] apply time shifts, frequency shifts and disturbances in frequencies to augment spectrograms of different human activities. Random disturbances can help model the random components that can vary between different measurements of the same activity, such as slight differences in the body shape or clothing. Timeshifts, on the other hand, change the starting point on the time axis of the activity and apply a horizontal translation to the micro-Doppler signatures.

A frequency shift, however, is used to consider the different velocities at which the activities can be carried out, which would as well affect the nature of the spectrogram in the time dimension. Similarly to the observations made for [41], a change in the velocity should, for example, also affect the duration of the activity, which is not accounted for in [43].

In order to improve the classification of micro-Doppler signatures belonging to different targets, Al Hadhrami et al. [44] apply flipping along the time axis and circular time shifts to generate new training spectrograms. Flipping a spectrogram along the temporal axis, however, is very likely to produce a micro-Doppler signature that com-

pletely lacks any physical feasibility, as the sequential nature of any motion is altered without any regard for the frequency.

The circular shift, on the other hand, can be a more feasible augmentation for micro-Doppler signatures that present a repetitive nature, such as the ones belonging to walking humans or rotating drone blades. In this case, a circular shift would produce a feasible sample at the price of introducing a discontinuity in the point where the spectrogram wraps around itself.

While an improvement is achieved in classification accuracy using the reviewed augmentations, it is still unclear whether including more physically plausible augmentations could bring further improvement in generalization capabilities.

In this thesis, two of the reviewed augmentations, namely the rescaling in time and frequency and the circular timeshift, will be implemented to assess their impact on the classification task and on the feature extraction performed by the classifier.

2.4.2 Neural Style Transfer

Neural style transfer (NST) is a technique that aims to synthesise a stylised image from two different input images, one providing the semantic content and high-level features, and the other providing the overall style of the image [45]. The final product is an image that shares the subject of the first input but is presented with the texture and low-level visual features of the second input. An example of NST applied to a visual image is shown in figure 2.3.



Figure 2.3: (c) The result of applying neural style transfer on (a) a visual image from [39] used for as content and (b) Composition VII by Wassily Kandinsky used as style.

In the scope of DA, NST can be applied to help a NN classifier separate the high-level content of the input from the low-level context information, by augmenting the same content image in the training set with multiple styles. In image classification tasks, it is not unusual for samples of the same class in a dataset to be acquired in similar scenarios. For example, different pictures of camels could happen to always be taken in a desert landscape. This could hurt generalization, as the NN classifier may tie context information such as the background or lighting condition to the class output, and fail to classify the same object if presented in a different environment or with a different illumination.

In [46] and [47], the authors apply several styles to the same target images, effectively augmenting their dataset by reiterating the same high-level content overlaid with different low-level style features. By being presented with the same content under different contexts, the classifier learns to untangle the information on the target from its style and is able to achieve better results, especially when classifying images that belong to domains with different textures and low-level features.

While NST has been widely applied to visual images, the leap to physically characterised domains is not straightforward. In fact, in this case, there is first a need to determine what can be considered the content of a sample and what is instead part of the style. One possible assumption is to consider physical prior information and simulated data as the high-level content input, while experimental realisations would carry the style in the form of detailed physics, non-idealities and artefacts that are not modelled in the theoretical framework.

The problem with this approach lies in the process of extracting the style and content information. In the visual domain, this is done using the features computed by the layers of a NN trained on an ample dataset, such as Visual Geometry Group (VGG) [48]. Similarly to what happens in transfer learning, however, there is no guarantee that a network trained on visual images will be able to extract the proper content and style information from a sample belonging to the target domain, especially if the expectation is for the content to be related to the physical characteristics of the input image. For these reasons, applying style transfer to physically structured images is not immediate and may require developing a proper and application-specific way of acquiring relevant features to compute the style and content information.

Vishwakarma et al. [49] utilise NST to add the texture of measured micro-Doppler spectrograms to simulated ones to improve the recognition of different human activities. In their work, the authors utilise the simulations as the source of content and the layers of a pre-trained VGG to extract the style information from the real measurements. The goal of this method is to transfer the contributions of effects such as target-dependent multipath, occlusions, noise and other artefacts that appear in the measurement to the clean simulated spectrogram. This leads to an increased fidelity of the simulations and hence a better accuracy in the downstream classification task when training the classifier with both augmented and measured spectrograms.

While the results show an improvement for the specific application, by relying on a pre-trained network to infer the style there is no guarantee that the content and style information extracted will match the physical features considered relevant for the task. Similarly to what happens in transfer learning, if the source domain where the feature extraction network is pre-trained doesn't match the target domain, there may be the risk of negative transfer. In the NST scenario, this would imply that the content and style information may be incorrectly inferred, leading to the transfer of irrelevant features to the content image and an overall worsening in the quality of the output.

In addition to that, the NST approach described in [49] is a 1-to-1 optimization process, meaning that each augmented spectrogram is generated by augmenting a single simulation with the information coming from a single measurement. However, as in general more than one measurement is available, an advanced augmentation technique may consider computing the style information by combining multiple available

measurements.

In view of these limitations, this thesis will not include further research on the topic of NST.

2.4.3 Generative Neural Networks

Generative NNs are a particular subset of ML models that are designed to generate new samples that are similar to a given set of training data. These models are typically trained to learn a mapping from an input random distribution (typically uniform or Gaussian) to the underlying posterior distribution of the training data. By randomly sampling the input distribution it is then possible to generate new output samples that will be different from the training ones but will show similar features, and can hence be used to effectively augment the original available data. Generative NN typically includes architectures such as VAEs [24] and GANs [23]. In the scope of this work, GANs will be the model of choice for performing DA with generative networks.

The original formulation of the GAN structure, by Goodfellow et al. [23], comprises two separate models. The first is a generator G that takes a white Gaussian noise vector z as input and produces a synthetic sample $G(z)$ as output. The second is a discriminator D that takes an image sample x as input and outputs the probability $D(x)$ of the image being a real sample from the training set or a fake sample produced by G .

The two models G and D are then trained in an adversarial manner. The discriminator D is presented with real and synthetic samples and is trained to maximize the probability of labelling the inputs from the training set as real and the ones produced by G as fake. On the other hand, G is trained to maximize the probability that its generated samples are labelled as real by D . Mathematically, this can be written as the following minimax problem:

$$\min_G \max_D E_x[\log(D(x))] + E_z[\log(1 - D(G(z)))]$$

where E is the expectation with respect to the subscripted variable. The minimax cost of the GAN problem is also referred to as critic loss. Ideally, after a successful training phase, the two networks will reach a state of balance where the D will have to guess the origin of x randomly with an accuracy of 50%, as the generated images $G(z)$ will look exactly like samples from the training set. Figure 2.4 illustrates the architecture of a GAN.

Despite their potential, the training process for GANs is known to be unstable [51]. Typical issues that appear when training GANs are ones such as vanishing gradients and mode collapse. The first issue arises when the gradients of the parameters with respect to the loss function become very small, resulting in slow or stalled learning. The second, instead, arises when the generator learns to produce an image $G(z)$ that is able to fool D , and learns to map all the z sampled from the Gaussian noise to that image. This leads to the whole framework getting stuck in an “optimal” point and losing all of its generative properties. To mitigate these effects, new training strategies and architectures have been developed such as the Wasserstein-GAN (WGAN) [52] and the WGAN with gradient penalty (WGAN-GP) [53].

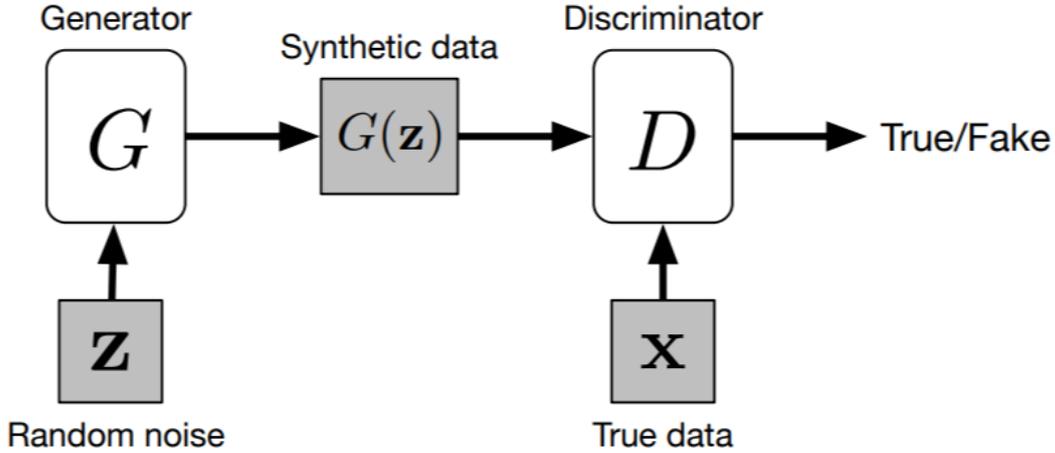


Figure 2.4: General structure of a generative adversarial network, where the generator G takes a noise vector z as input and produces a synthetic output $G(z)$. The discriminator takes both the synthetic input $G(z)$ and true sample x as inputs and predicts whether they are real or fake. Image taken from [50]

In the radar domain, [54–56] use GANs to generate micro-Doppler spectrograms belonging to individuals performing different activities. While the final tasks may vary slightly, the approach relies on training the GAN via the available training set and producing the output by randomly sampling the latent space. The generated synthetic spectrograms are then used to train a classifier together with the available data to improve the performance in the downstream task.

While GANs hold very powerful generative capabilities, there is usually little to no control over the physical feasibility of the generated samples. In fact, the statistical similarity between the real and synthetic samples does not imply conformity with the physical properties of the target domain.

Erol et al. [25] produce several augmented samples using GANs and VAEs. After the generation of the synthetic spectrograms, the authors sieve the generated samples by using different criteria. First, they establish a series of high-level physical rules and exclude all the samples that do not comply with them. Secondly, they compute the bounds of the principal components obtained via principal component analysis. They then remove all the generated samples that lie outside of these bounds, obtaining a set of synthetic spectrograms that are overall more consistent with the kinematics of the real measurements. By relying only on the synthetic spectrograms that comply with the sieving criteria the authors are able to achieve better performance in classifying human activities using augmented data compared to the accuracy obtained using the unfiltered generated samples. The results show indeed that physical feasibility is important for generalization purposes and that GANs without proper supervision may produce kinematically unfeasible samples.

While this approach leads to improved performance in the downstream classification

task, there could be greater benefits in enforcing physical knowledge directly in the training process and design of the ML model. For this reason, this thesis will further employ GANs to explore ways of embedding physical information in generative models.

2.4.4 Simulated Data

The thorough study of the radar domain has yielded well-characterized governing laws that can be used to model different sensing scenarios. If an accurate model of the target is provided, the backscattering of the radar waves can be computed through the available equations. The spectrograms constructed by processing the results of such computations are what are referred to as simulated spectrograms.

When constructing a simulation, several assumptions can be made, which determine the degree of complexity and the fidelity of the simulation with respect to a real experimental spectrogram. These can range from simplified radar returns of approximated kinematic models of the targets to a full-wave study of the electromagnetic scenario. As more complex targets with non-rigid bodies are considered, the computational complexity grows considerably due to factors such as the evolution in time and the interaction between different moving parts.

In the case of a walking individual examined in this thesis, the multiple reflections between different body parts, the different reflectivity of clothing and skin, the shape of the limbs and the shadowing that they can cause during the movement are all factors that make accurately simulating such scenario a difficult task. For these reasons, simpler conditions are usually assumed. For example, the human body can be assumed to be made of limited scatterers with known reflectivity or multi-path and shadowing effects between different limbs. While these assumptions succeed in limiting the computational complexity, they also reduce the fidelity of the simulations, which, depending on the context, may not be sufficient for the target application.

Huizing et al. [26] consider theoretical simulations of micro-Doppler spectrograms of drones, obtained by using simple kinematic models. In their work, the authors show that by only relying on such simulations to train a downstream classifier the achieved accuracy becomes mediocre, when compared to the one achieved using experimental spectrograms.

On the other hand, by using simulated data Tang et al. [57] are able to improve the recognition of different human activities in data-scarce scenarios. To achieve this, they utilise radar simulations generated from motion-captured kinematic data. Motion capture uses cameras to track and record the movement of several points in space. In this case, 25 LED markers were used to track the movement of body parts of individuals performing several activities [58]. The movement of these points in space is then used to generate the radar returns for the specific activity and radar parameters, such as the carrier frequency, pulse repetition interval or the aspect angle with respect to the subject.

While these techniques can yield high-quality simulations when compared to theoretical kinematic models, they still require real motion-captured measurements of different subjects. Moreover, measuring the motion-capture data allows for the simulation of different parameters related to the radar but doesn't provide a way to change the

attributes of the subjects such as height or speed at which the activity is performed other than by manually altering the position of the captured markers.

Simulations made from theoretical models, on the other hand, allow more freedom when choosing the target parameters but yield spectrograms that are too distant from the real measurements and would need further enhancement to be beneficial in the downstream classification task.

2.5 Physics-Informed Machine Learning

Physics-informed machine learning is a term that covers all those techniques aimed at incorporating prior physical knowledge into ML models [27, 59, 60]. In this context, the word physics indicates information such as domain-specific properties, simulations and governing equations. By embedding the proper physical prior knowledge into the NN framework, researchers look to develop physics-inform neural networks (PINNs), which have already been employed to tackle problems such as solving partial differential equations [61], fluid modelling [62] and geophysical image inversion [63].

PINNs bring several advantages over regular NNs. Despite their capabilities, regular NN-based approaches struggle to extract easily interpretable features from the available data. By incorporating physical information in the architecture of the NN, the model can be guided into learning how to gather and analyze specific characteristics of the input data that are physically meaningful for the domain of the application.

Additionally, fully data-driven approaches may overfit the available data, leading to poor generalization performance and biased predictions. The physical prior information that is provided to PINNs, on the other hand, will result in theoretical constraints to aid ensure that the model is able to learn proper underlying relationships and that the produced output complies with the governing laws of the domain of application.

This is particularly relevant in data-scarce scenarios, where the training of a model on a limited amount of observations can be guided and constrained to improve the generalization capabilities and to focus the attention on physically meaningful features.

The analysis presented in [27] defines three approaches to embed physics in ML models and refers to them as introducing observational, inductive and learning biases:

- **Observational biases** are provided through the input training data of the model. While many physical and engineering sciences don't have access to large labelled datasets, these may be integrated with simulations or data that is produced by domain-specific augmentations, in order to preserve the physical feasibility of the original data.
- **Inductive biases** are introduced by making tailored decisions on the architecture of an ML model. In the case of NN architectures, by appropriate design of the input, the layers and the graph of the network, the extraction of domain-specific features with appropriate properties can be enforced. While the inductive bias approach is robust and guaranteed to enforce the prior information, it is often difficult to implement for complex physical laws and is usually used to enforce more simple geometrical properties such as symmetries.

- **Learning biases** are enforced by ad-hoc choices in the learning process. Generally, this involves a domain-specific design of an appropriate loss function and constraint on the weights. These approaches do not strictly enforce physics, but by adding penalties based on the domain’s governing laws the convergence towards physically-compliant solutions is favoured.

In the radar domain, observational bias has been already employed in the form of DA using simulated data.

On the other hand, it can be argued whether image manipulations fit under the label of physics-informed techniques. In fact, these augmentations do not directly introduce physical prior knowledge into the training data, but rather exploit invariances of the target data without interfering with its physical content. Since the reviewed approaches and the image manipulation proposed in this work do not directly leverage physical information but are designed to be compliant with the physical feasibility of the micro-Doppler spectrograms, they will be referred to as physics-aware techniques, rather than physics-informed.

In the scope of physics-informed techniques, Rahman et al. [64] propose a multi-branch GAN architecture to embed physical information in the process of generating spectrograms for the classification of different human activities. In particular, two branches are added to the discriminator of the adversarial architecture and are responsible for extracting the upper and lower envelope of the spectrograms, which are then used in the discrimination process. This approach falls under the category of both inductive and learning bias, as the architecture is modified to extract a physically meaningful feature in the form of the envelope, but also the critic loss is adapted to take into account the envelope of the generated spectrogram.

The results show that the physical information contained in the envelope of the spectrograms can help produce more physically compliant samples, which can also lead to better performances when used to train NN-based classifiers. However, the shape of the envelope can only be applied to the specific domain of activity recognition, where it can be used to characterise different human actions.

In the scope of this work, the embedding of domain-specific physics in the task of classifying armed and unarmed walking individuals will be investigated. In particular, the focus will be on inducing an observational bias in generative NNs via the training data by using the available theoretical simulations. In addition to that, the effect of using physics-aware image transformations and the observational bias that is induced by them in the downstream classifier will be assessed.

2.6 Use Case: Classification of Armed and Unarmed Personnel

In order to assess the effectiveness of the different data augmentation techniques, the downstream task of classifying armed and unarmed walking personnel [65, 66] will be utilised. As of today, cameras are widely used for surveillance in the security domain, due to their limited cost and the fact that visual images can be easily interpreted by a human operator without any special training. Regardless of this, surveillance radars

hold many benefits, especially in the defence domain. First of all, a radar detector can work regardless of the weather, in conditions of low visibility or at night, contrary to typical camera systems. Moreover, radars can also be employed in situations where privacy is a concern. Despite these benefits, radar spectrograms and images cannot be directly interpreted by an untrained operator, which gives the motivation to research an automated way to detect signatures belonging to individuals carrying weapons.

While the classification task holds important practical applications, the focus of the work presented in this thesis will be on the augmentation process. In this regard, the expectation is that the approaches proposed in the scope of this work may be successfully adapted to other use cases involving spectrograms of walking humans or different targets such as drones.

2.7 Literature Summary and Thesis Objective

After reviewing several techniques to address the deployment of NNs in domains with limited available training data, it has emerged that multiple directions exist from which the problem can be tackled. In particular, DA has seen a lot of applications because of its ability to deal with the problem at its source and hence being compatible with most of the DL solutions and architectures.

However, in the radar domain, visual transformations are hindered by the difficulty of implementing such approaches while preserving the physical feasibility of the target images. To the best of the author’s knowledge, most of the examples of image manipulation applied to micro-Doppler spectrograms do not consider the effect of the augmentation on the underlying physical model and do not assess the effect of the feature extraction process.

When considering automatic target recognition and human activity classification through micro-Doppler spectrograms, most of the available DA literature revolves around generative NNs in the form of GANs and VAEs. While these approaches ensure some degree of similarity between the generated samples and the original ones, there is no guarantee of the physical feasibility of the augmented data. To the best of the author’s knowledge, only a few examples exist of trying to improve the physical feasibility of GAN-generated spectrograms, none of which rely on the theoretical simulations available in the radar domain.

The goal of this work is to explore physics-informed data augmentation for radar micro-Doppler spectrograms. This will be done by studying ways to enhance the available theoretical simulations in order for them to be employed for DA in a downstream classification task. In addition to that, this thesis will also tackle the issue of physics-aware image transformations for the radar domain, by reviewing two of the existing techniques and proposing new alternatives for the classification of micro-Doppler human signatures.

Summarising the main contributions of this work:

- To provide insight into the usage of theoretical simulations to aid generative DA approaches. In particular, several generative augmentation techniques are proposed and their limitations with respect to their generative potential will be dis-

cussed, by evaluating the quality and diversity of the produced spectrograms and their impact on the downstream classification task.

- To analyse recent work on visual transformations for DA in the radar domain, with respect to their impact on the achieved accuracy in the downstream classification task and on the feature extraction process. A new augmentation is then proposed for the task of classifying spectrograms of armed and unarmed individuals by leveraging the cyclic nature of the walking motion.

Figure 2.5 shows a scheme of the review literature and in red the areas of contribution of this thesis.

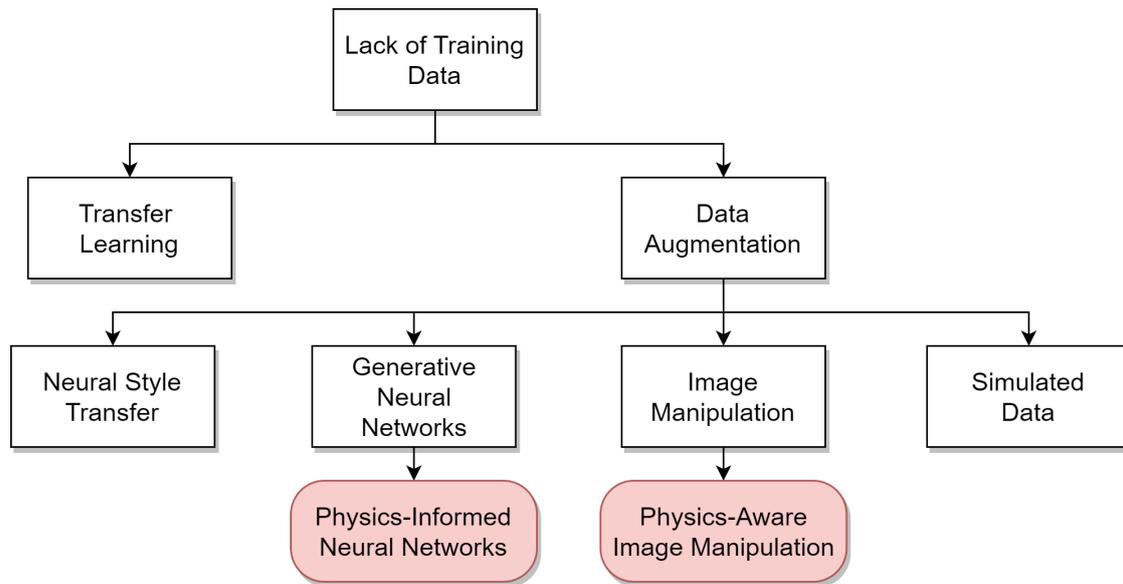


Figure 2.5: Overview of the different solutions presented in the literature study. Shown in red are the areas of contribution of this work

3.1 Micro-Doppler

3.1.1 The Doppler effect

A monostatic radar is a device used to both transmit an electromagnetic signal to an object and receive the reflected response from the object itself. The time elapsed between transmission and reception of the reflected signal can be used to measure the radial distance of the object from the radar. For a non-static object, the received signal will present a frequency shift if compared to the transmitted signal, due to a phenomenon known as the Doppler effect.

The magnitude of the Doppler frequency shift is proportional to the radial velocity of the target with respect to the radar transceiver in the direction of the radar line of sight (LOS) and can be used to determine the speed at which an object is approaching the transceiver.

In the case of a static radar and a target moving away in a radial direction from the source with velocity v , the frequency of the received signal f_r can be expressed as

$$f_r = \frac{1 + v/c}{1 - v/c} f \quad (3.1)$$

where f is the transmitted frequency of the signal and c is the speed of light. Note that as a convention the velocity will be assumed positive when the target is moving away from the radar and negative in the opposite scenario.

If the velocity of the target is small compared to the speed of light c , then the relativistic effects can be ignored and the Doppler frequency shift $f_d = f_r - f$ can be expressed in Hz as

$$f_d = -\frac{2v}{c} f \quad (3.2)$$

The Doppler frequency shift is commonly observed in the frequency domain, by looking at the Fourier transform (FT) of the transmitted and received signal. In the frequency domain, the peak components are used to find the different Doppler frequency shifts of the individual targets.

3.1.2 Micro-Doppler

The Doppler frequency shift described in Section 3.1.1 is commonly referred to as bulk Doppler, as it is associated with the motion of the entire target in the radar LOS direction.

Real targets will usually present an extra layer of modulation on top of the bulk Doppler shift, called micro-Doppler effect [67], which is caused by any component of the target that has an oscillatory or vibrating motion. Examples of this effect include the modulation induced by the rotation of the blades of a helicopter, the swinging of the limbs of a walking person or the vibration that an engine induces on the body of a vehicle.

As the micro-Doppler effect is tightly associated with the structure and nature of the observed object, it can be used to identify and classify targets seen by radar systems. In this context, the cumulative micro-Doppler effect presented by a target, represented in the joint time and Doppler frequency domain, is referred to as its micro-Doppler signature.

3.1.3 STFT

Due to the non-stationary nature of the micro-Doppler effect, good visualization of the phenomenon cannot be achieved by direct application of the Fourier transform on the received signal. To better represent signals with time-varying frequency content the Short-Time Fourier Transform (STFT) is introduced. The STFT consists of applying the classical FT to a time-limited window of the original signal, in which ideally the frequency content of the signal can be considered constant. In such a way, by taking the FT of consecutive windows of the original signal, one may construct a representation of the evolution of the spectral density over time. Mathematically, the discrete STFT of a digitised signal $x[n]$ is defined as:

$$STFT(m, k) = \sum_{n=0}^{N-1} x[n]w[n - mH] \exp\left(\frac{-2\pi jk(n - mH)}{N}\right) \quad (3.3)$$

where w is a window function of length N and H is the size of the hop in samples between subsequent windows. The properties of the STFT are determined by three parameters related to the window choice. These are namely the window duration, the window type and the overlap between consecutive windows.

When varying window length, shorter windows will provide better resolution in time but a worse resolution in frequency when compared to longer windows, as they are able to better capture variations in the frequency content over time. On the contrary, longer windows allow for better frequency resolution but make tracking changes in time more difficult, leading to a worse resolution in time. In addition to that, the type of window used can offer different properties in terms of resolution depending on the chosen shape, and an overlap between different windows can be introduced to smoothen the changes in frequency over time, and usually varies between 30% and 80%.

Finally, after calculating the complex STFT, the variation of the spectral content as a joint function of time and frequency is usually shown using the spectrogram, defined as $|STFT(m, k)|^2$.

3.2 Experimental Setup

The experimental radar measurements were acquired in the scope of the work described in [65]. The measuring campaign was performed using the AMBER radar system, a frequency-modulated continuous wave radar that operates in the X-band. More information about the radar parameters is provided in section B.1.

Measurements were made of people walking towards the radar with their hands free and people walking with a metal pipe in both of their hands to mimic a rifle. For the experiments, a total of 35 test subjects were recorded performing both activities twice.

After removing some faulty measurements, 931 spectrograms of 1.5s are extracted from the original measurements. The frames are taken with matching initial gait phases and contain around 1.5 gait cycles, or three steps, varying based on the height and velocity of the subjects. Out of the 931 spectrograms, 445 belong to the walking class and 486 to the rifle one, making them 48% and 52% of the whole dataset, respectively. Despite the slight difference in the number of available samples, the classification problem will be treated as balanced as no significant bias is observed towards the class with the most spectrograms available.

In addition, the number of spectrograms per subject can also be considered balanced. On average, 27 experimental spectrograms are available for each tested individual.

In the processing phase, each spectrogram is first converted to dB scale and then scaled again such that every bin in time and frequency has a magnitude between 0 and 1 by using the value of the greatest and smallest bin of the spectrogram. This achieves the goal of making the spectrogram bound, which improves the training of NN models, by removing any potential bias that could occur by having unbounded values of the inputs.

Figure 3.1 shows spectrograms from the two classes. The vertical axis represents the velocity corresponding to the received Doppler shift. The velocity is considered negative when directed towards the radar, and positive when away.

Visually, the two classes share similar features. Intuitively, we would expect a person walking hands-free to show a more prominent return related to the movement of the arms. Fig 3.1a highlights the return associated with the movement of the arms, which is indeed slightly more evident in the case of a person walking without carrying an object. However, not all subjects show such features, as there could easily be walking individuals that keep their arms by their sides or in their pockets.

3.3 Simulated Data

The simulations used in this thesis are generated according to the theoretical model described in [68]. To characterize the motion of walking individuals, the global human walking model proposed by Boulic et al. [42] is used. In [42], the model of human motion is described as a combination of translations and rotation of several reference points over the time of a gait cycle, which is the term used to refer to the cyclic pattern of movement that occurs while walking. Fig 3.2 shows the reference points used to describe the human walking model.

To compute the radar backscattering of the walking individual, each body segment

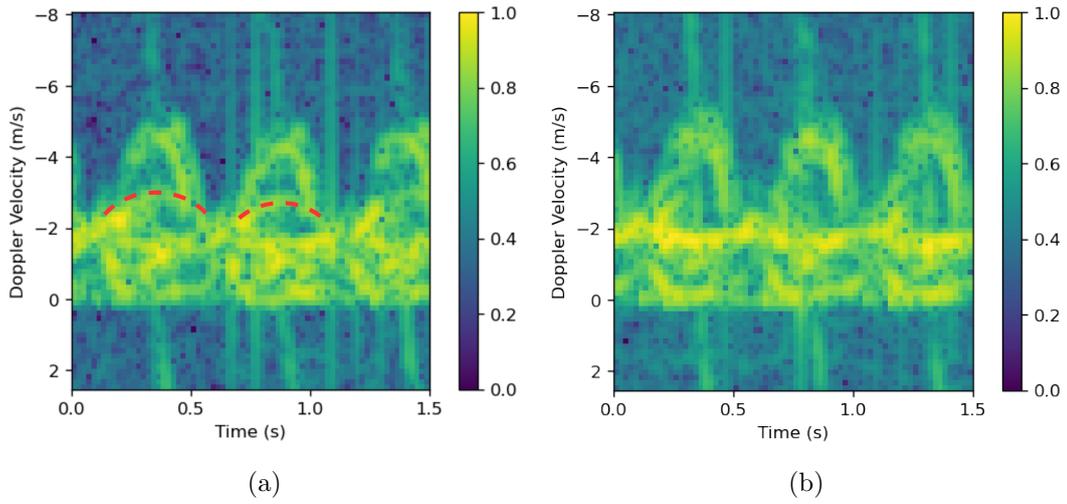


Figure 3.1: Framed spectrograms from the measured datasets belonging to (a) a subject walking hands-free and (b) one carrying a rifle. The micro-Doppler response from the arm is highlighted in (a).

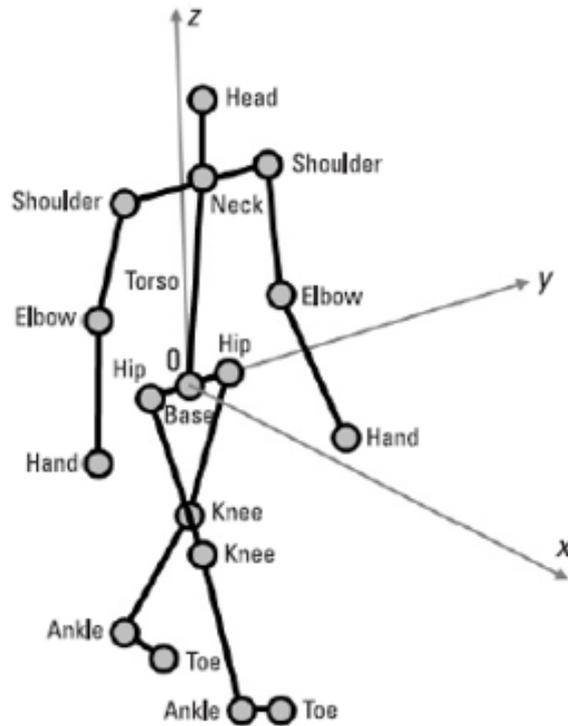


Figure 3.2: Reference points used to generate the human walking model [68]

is modelled as a perfectly electrically conducting ellipsoid of appropriate size. Using the theoretical RCS of each ellipsoid and its position and rotation over time, the cumulative

backscattering from all the moving body parts is then obtained and used to compute the spectrogram. Figure 3.3 shows a simulated spectrogram where the contribution of the different limbs is highlighted.

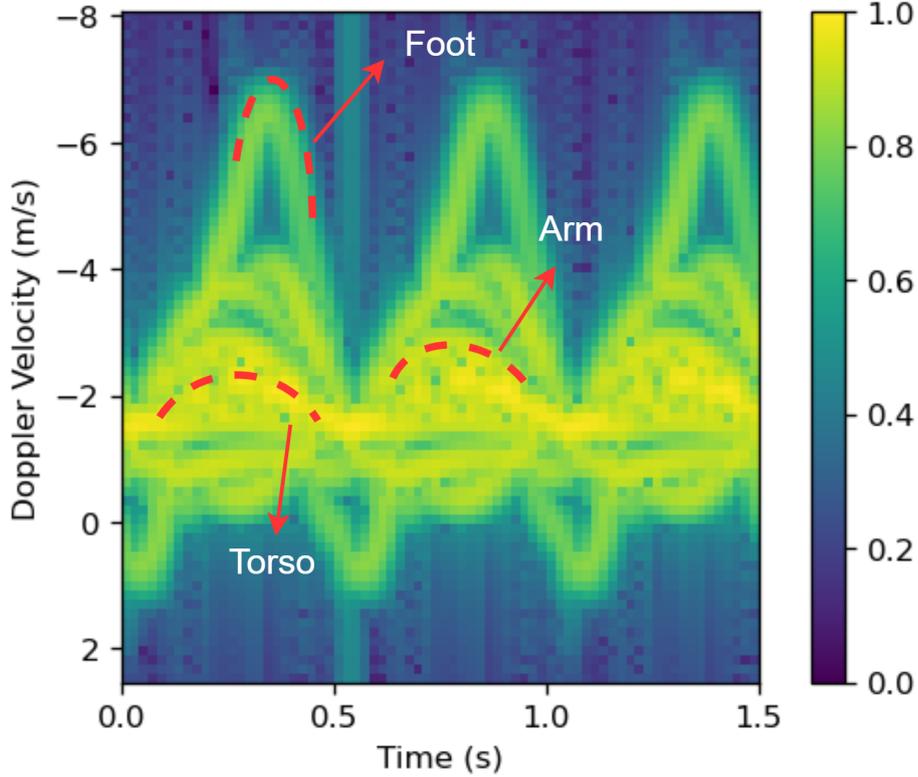


Figure 3.3: Simulated spectrogram generated using the model proposed in [68]. Part of the returns from the foot, the arm and the torso are highlighted.

The simulations include no random component and are completely determined by a set of two kinematic parameters. These are namely the mean velocity at which the person moves and the gait cycle duration, which is the duration of the cyclic pattern of movement that occurs while walking. To obtain the final dataset, these two parameters are extracted from each experimental spectrogram and used to generate a paired simulated one. The process to extract the kinematic parameters is described in Appendix B.2. As no theoretical simulation is available for the rifle class, the pairing is still done using simulation produced using the global human walking model. Paired simulated spectrograms are also generated for a third class of spectrogram of individuals carrying a backpack that was recorded in [65]. While the spectrograms belonging to the third class are not used for classification purposes in this thesis, their kinematic parameters are extracted to generate simulations that will be used to test the proposed augmentations. In such a way, the parameters used to generate the simulation are ensured to belong to an experimental distribution.

The assumptions made to generate simulated spectrograms according to [68] show several gaps when compared to a real scenario. First of all, a perfectly reflecting

ellipsoid may not be an accurate model for human joints, as it ignores the actual shape and reflectivity of the limbs. Secondly, the shadowing of one body part by means of another one is ignored, as well as second-order reflections, which are indeed affecting real experimental spectrograms. Thirdly, the simulated spectrograms repeat perfectly during each gait cycle, while spectrograms belonging to real subjects inevitably show variations between different cycles, due to subtle changes in the motion of the subjects from step to step.

Despite these simplifications, simulated spectrograms still show much of the frequency content that appears in experimental ones and could prove to be a reasonable starting point for applying augmentation techniques in the radar domain.

Generative Augmentations from Simulations

4

In this chapter, the focus will be on studying different approaches in which theoretical simulations can be enhanced for DA purposes. The goal of these augmentations is to leverage the information contained in the simulations and the available experimental measurements to produce physically feasible augmented spectrograms.

After introducing the tools for assessing the proposed techniques, the different approaches will be presented and a final comparison will be shown.

To keep this dissertation lighter, only a brief overview of the training and augmentation process is given for each proposed method. More details on the architectures and the training parameters can be found in appendix C.

4.1 Evaluation Method

The first evaluation criterion for the proposed augmentations will be the visual quality of the generated synthetic samples. The expectation is that a good synthetic spectrogram shall show similar visual features when compared to the experimental spectrograms, as the classifier will potentially rely on those to infer the class label.

In the experimental work discussed in this section, both classes will be augmented using separate generative networks. To keep the dissertation concise, only samples from the walking class will be shown and evaluated, but the same remarks and conclusions regarding the visual quality of the synthetic spectrograms could be reached by looking at the rifle class.

When developing generative models, a common issue is mode collapse, which happens when only a single synthetic sample that optimizes the task is produced and all the variability in the generated output is lost. To evaluate the possibility of this happening, the structural similarity index metric (SSIM) is utilised [69]. The SSIM measures the similarity between images by taking into account factors that are perceived by the human eye such as luminance, contrast and structure of the image, and used in [25] to also determine the similarity between micro-Doppler spectrograms. The SSIM between two images X and Y is defined as

$$SSIM(X, Y) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)}$$

where μ and σ^2 denote the mean and variance of the pixels of the subscripted images, σ_{XY} is the covariance of X and Y , and c_1 and c_2 are two small constants to stabilize the division. To assess the variability of a models' output, the SSIM between all the possible pairs of generated synthetic spectrograms will be compared to the SSIM scores between the pairs of experimental samples. In addition to that, to better evaluate the similarity of the micro-Doppler signatures, only the frequency content between -6 m/s and -1

m/s will be considered, as it happens to comprise most of the returns coming from the limbs of the walking individuals. The distribution of the SSIM between every pair of experimental spectrograms belonging to the walking class is shown in figure 4.1. As can be seen, the SSIM between different experimental spectrograms lies between 0.2 and 0.7. Good synthetic samples will have a similar distribution of the SSIM. A distribution with higher SSIM scores will indicate that the generative model is producing samples that are too similar to each other up to the extreme of mode collapse, where only one sample is produced.

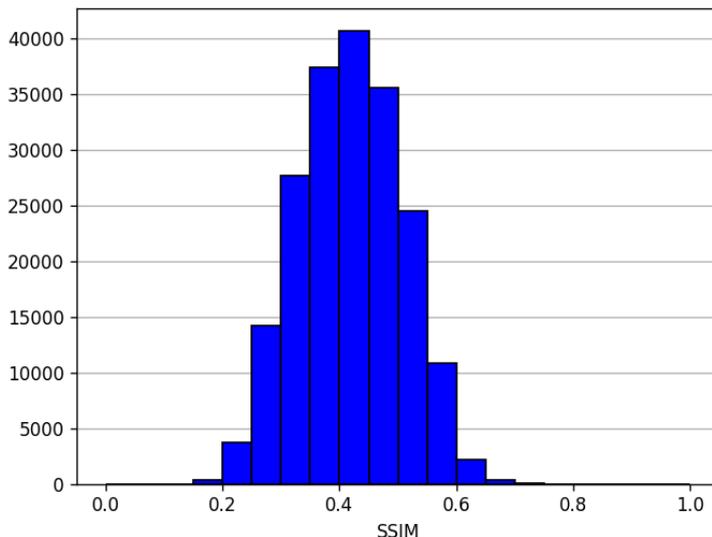


Figure 4.1: Distribution of the SSIM between every pair of experimental spectrograms belonging to the walking class

Finally, as a final metric of evaluation, the quality of the proposed generative augmentations will be assessed by their impact on the downstream classification task in section 4.7

4.2 Convolutional Mapping

In order to augment the simulations, a NN mapping is designed by stacking a series of convolutional layers. The goal of the mapping is to augment the simulations generated by using theoretical models with the information available in the form of experimental spectrograms.

A deterministic NN with a fixed input cannot generate the random noise and artefacts of the measurements. However, the systematic interactions between different limbs that are not considered in the simulation scenario could potentially be learned by such a model using the available experimental data.

To achieve this, the convolutional mapping will be trained on the matching pairs of simulations and measurements provided in the dataset. To train the mapping, the

simulations belonging to the training set are fed into the convolutional network. The synthetic output is then compared to the paired experimental spectrogram which is used as the ground truth for the training. The loss computed is the mean-squared error (MSE) between the pixels of the two spectrograms, defined as

$$MSE(S, M) = \frac{1}{N_f N_t} \sum_{i=1}^{N_f} \sum_{j=1}^{N_t} (S_{i,j} - M_{i,j})^2$$

where N_f is the number of frequency bins, N_t is the number of time bins, S and M are the synthetic and measured spectrograms respectively and the i, j subscripts denote the entry corresponding to the i -th frequency bin and the j -th time bin.

After computing the MSE, the loss is backpropagated through the network and the training step is concluded. Figure 4.2 shows a schematic of the process just described.

Once the network is properly trained, new synthetic spectrograms will be generated by enhancing new simulations that possess a different set of mean velocities and gait cycle durations.

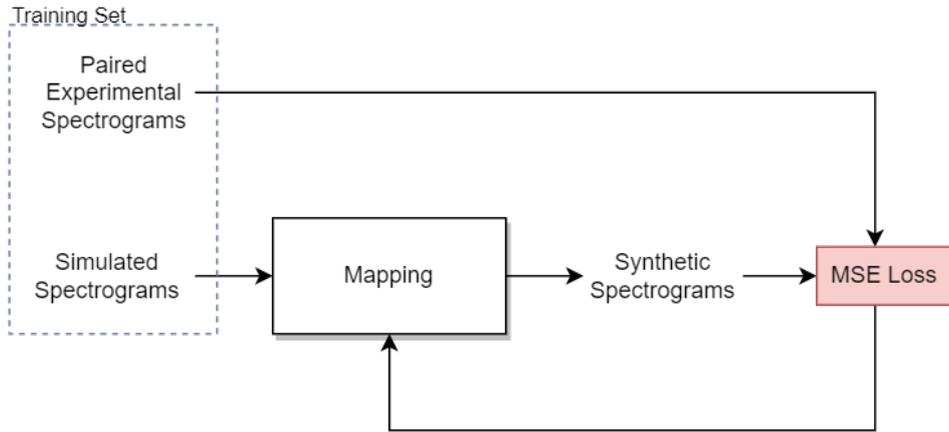


Figure 4.2: Overview of the training process of the convolutional mapping. The blue bounding box highlights the fact that the input simulations have to be paired with experimental samples that share the same physical parameters in order to compute the MSE loss.

This approach can be seen as the decoder part of an autoencoder [70], where the simulations act as a regularized latent vector in the parameter space of mean velocity and gait cycle duration. In this way, the synthetic output will also show the same parameters as the input simulation, ensuring its physical feasibility together with more realistic features given by the trained convolutional layers.

To assess the result, the first metric will be the visual quality of the spectrogram. Figure 4.3 shows an example of augmented spectrograms against their input simulation and the paired experimental spectrogram used as ground truth. As can be seen, the synthetic spectrograms appear blurry when compared to the experimental ground truths. This is most likely due to the MSE loss, as is also pointed out in [25]. By minimizing the error across different samples during the training process, the convolutional mapping learns to model an average style of the reference measurements, which

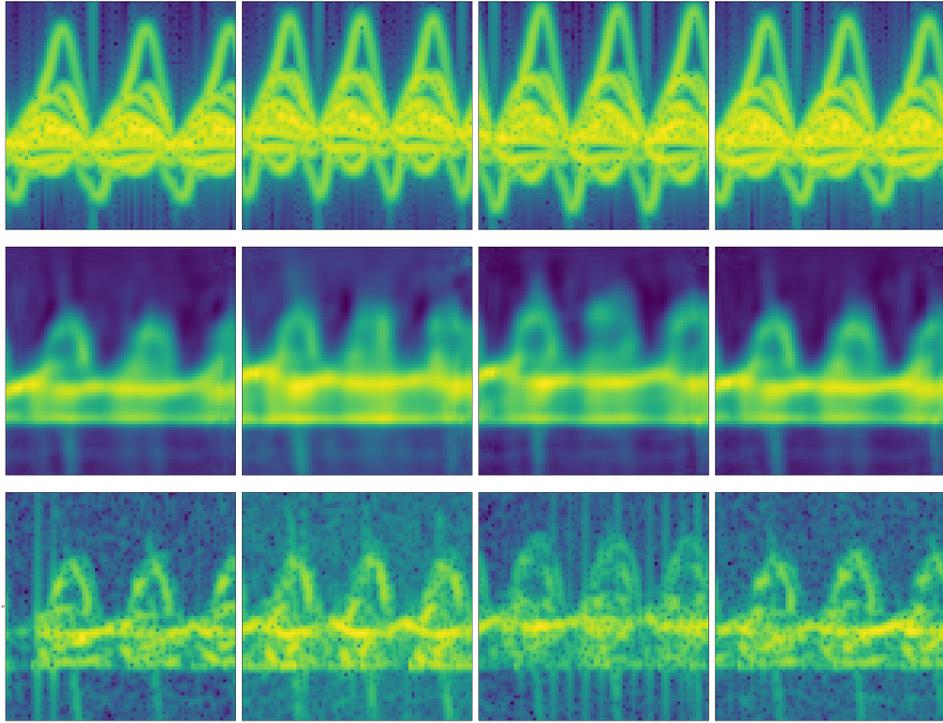


Figure 4.3: Visual demonstration of the spectrogram generated via convolutional mapping. Synthetic spectrograms (middle row) are generated from the respective input simulation (top row). The bottom row shows the experimental spectrograms paired to the respective input simulation and act as ground truth to compute the MSE.

results in a visually blurred texture in the final output. In addition to that, due to the limited diversity that can be achieved by the simulations, very similar inputs are mapped to experimental outputs that instead show a greater degree of diversity, up to the limit where the same simulation is mapped to multiple experiments. Together with the MSE loss, this could lead the model to produce synthetic spectrograms with a visually blurred style.

In addition to that, the generative potential of the simulations comes only from varying the mean velocity and gait cycle duration of the target. While these two parameters are enough to fully determine the theoretical simulation, they cannot fully grasp the variability of the experimental measurements, which show slight differences between different gait cycles. Using the proposed mapping, in fact, augmenting the same simulation twice would lead to two identical synthetic spectrograms. In theory, a perfect augmentation would be able to recreate the variations in the gait that happen when a person executes the same walk twice. It is clear, however, that a fully deterministic mapping cannot model this effect and hence some kind of controlled randomness has to be added during the generation process.

When looking at the differences between the synthetic spectrograms generated by the convolutional mapping, figure 4.4 shows that the similarity between pairs of synthetic spectrograms is significantly higher than the one between experimental ones. This may mean that the convolutional mapping is indeed producing an averaged syn-

thetic output in order to minimise the MSE loss. As a consequence, the simulations are mapped to very similar synthetic spectrograms, which lack resolute enough visual features to be distinguishable from each other.

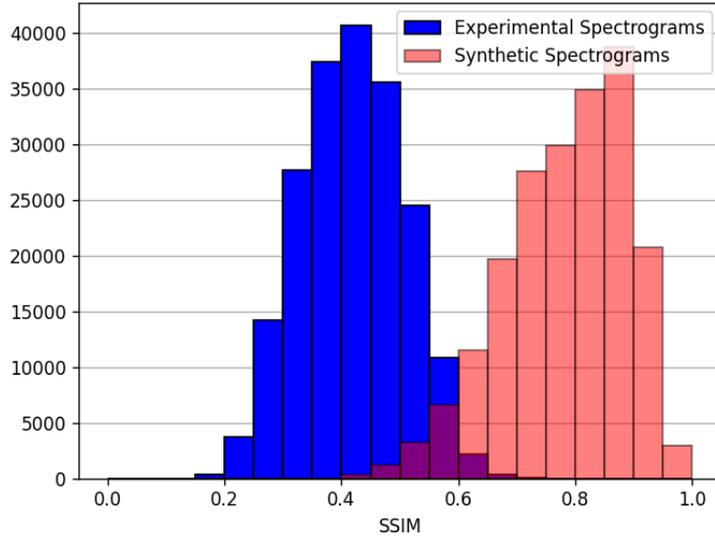


Figure 4.4: Distribution of the SSIM between each pair of synthetic spectrograms generated via the convolutional mapping and between pairs of experimental spectrograms.

To address these gaps there is a need to change the nature of the model used to augment the input simulation network. In view of its generative power, the adversarial training scheme will be explored, where instead of a reconstruction loss a discriminator network is used to determine the quality of the synthetic spectrograms during the training phase.

4.3 GANs

As the first step in analysing adversarial training for DA purposes, the classical GAN approach is explored. For this initial case, only the experimental data will be used, without any support from the theoretical simulations.

In the scope of this thesis, the WGAN-GP architecture will be employed [53]. A diagram of the training process is shown in figure 4.5. The input latent space made of a vector of 100 independent normally distributed variables is sampled and then fed through the generator network, which produces the synthetic outputs. These are then sent to the discriminator, which tries to distinguish between the artificial spectrograms and the real experimental ones. After that, the critic loss associated with the discriminator’s ability to distinguish real and synthetic spectrograms is computed and backpropagated to update the weights of both networks accordingly. The discriminator is trained to maximize its ability to detect synthetic and real spectrograms, while the generator is optimized in an adversarial fashion to better fool the discriminator.

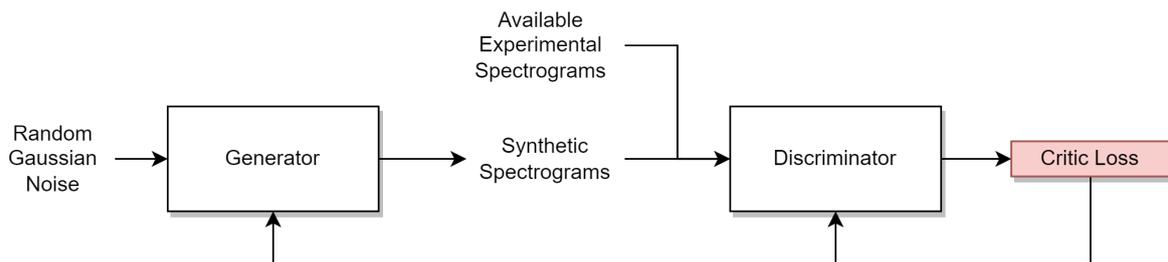


Figure 4.5: Overview of the training process of a GAN.

Figure 4.6 shows samples of the spectrograms generated with a GAN. As can be seen, the synthetic spectrograms resemble closely the experimental ones in terms of visual features. In particular, the random nature of the latent vector used as input allows the generator to successfully model the noise and artefacts present in the experimental spectrograms, as well as the meaningful frequency content of the micro-Doppler signature.

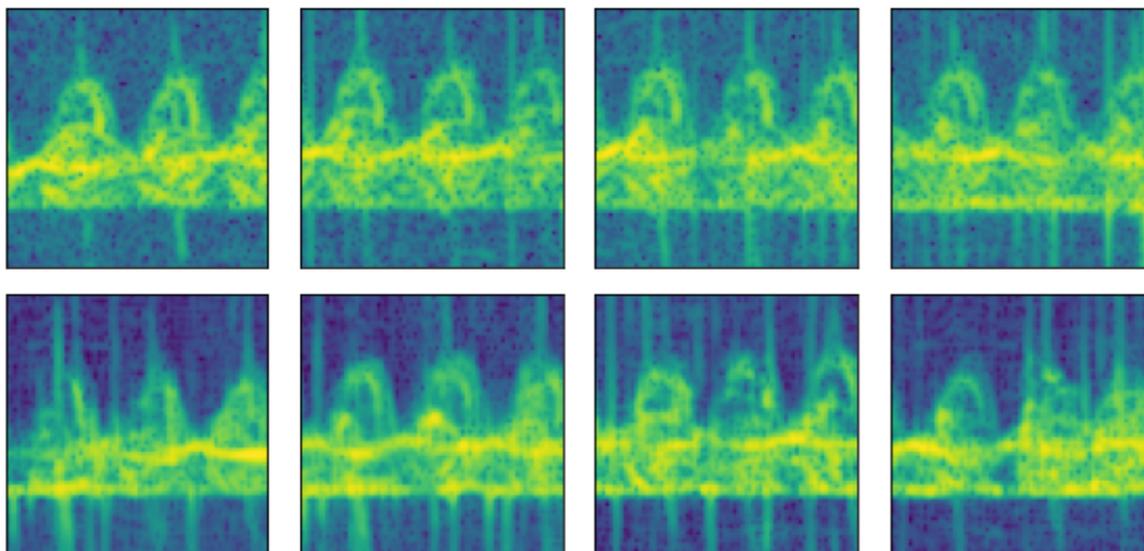


Figure 4.6: A comparison between real experimental spectrograms (top row) and synthetic ones (bottom row). The synthetic spectrograms were produced by random sampling of the generator’s latent space. The experimental spectrograms are included here for comparison and are used only in the training phase of the GAN.

While GAN may suffer from mode collapse and resort to the generation of a single synthetic output, the visual examples seem to prove otherwise. Even the distribution of the SSIM scores illustrated in figure 4.7, shows that the degree of similarity between the synthetic spectrograms matches the one between the experimental ones, meaning that the GAN-generated samples possess a realistic variability.

To assess the physical feasibility of the generated spectrogram, figure 4.8 shows the

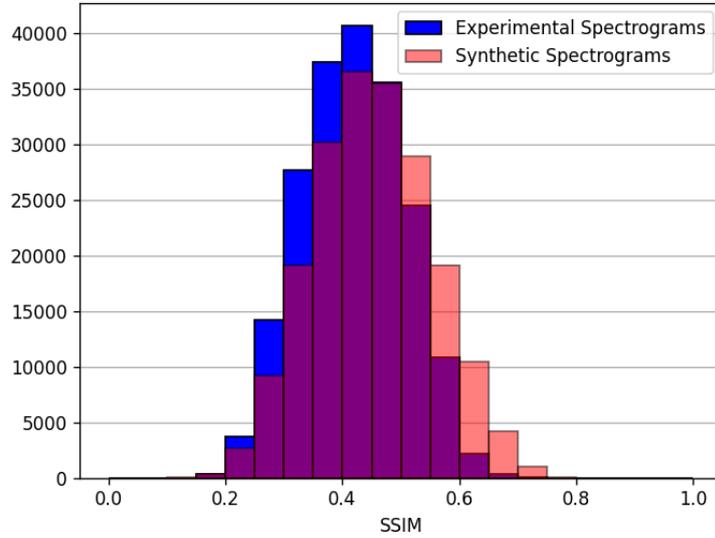


Figure 4.7: Distribution of the SSIM between each pair of synthetic spectrograms generated via GAN and between pairs of experimental spectrograms.

separate and joint distribution of the two kinematic parameters of the walking individuals for both synthetic and experimental spectrograms. Apart from a few outliers, the distributions are almost overlapping, meaning that while no particular measure to enforce physical feasibility was adopted, the pairs of mean velocities and gait cycle durations of the generated spectrograms do not seem to blatantly contradict the underlying kinematic model.

Despite this, the usage of simulations could still prove useful in making the training phase more stable and producing better synthetic samples, with control over the output’s physical parameters. For this reason, the rest of this chapter will focus on a physics-informed generative approach to produce synthetic spectrograms from theoretical simulations while leveraging the proven benefit of using a critic loss instead of the MSE.

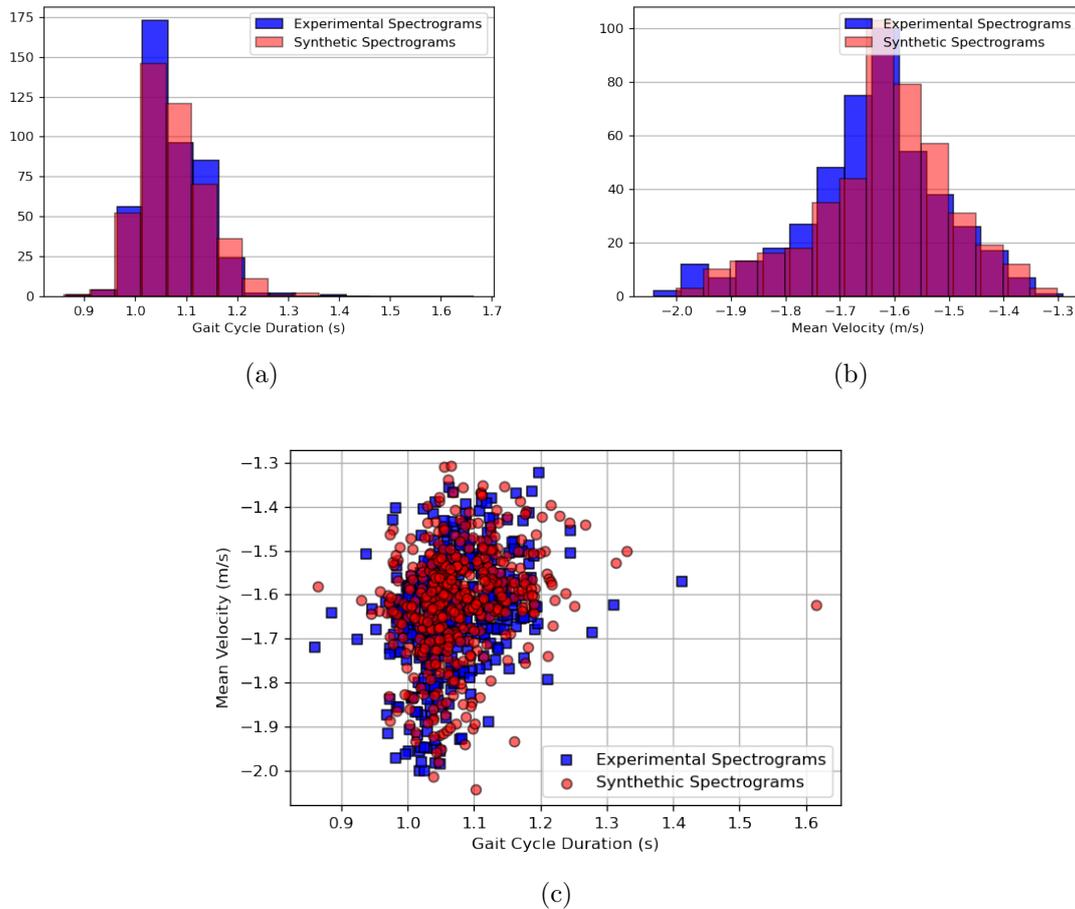


Figure 4.8: Comparison between the distributions of (a) gait cycle duration, (b) mean velocity separately of 500 experimental and synthetic spectrograms. (c) shows the joint distribution of the two parameters across the analysed spectrograms.

4.4 GANs from simulations

To incorporate the theoretical simulations in the process of generating synthetic spectrograms, the classic GAN architecture is modified to use the simulations as a starting point for the generator.

This approach resembles the work done by image translation networks such as Pix2Pix [71] and CycleGANs [72], which have been used to tackle problems where an input image has to be mapped to its respective representation in another domain. Examples of this include translating images from black and white to coloured or transforming pictures taken during the day into the same picture taken at night.

Despite the similarity of these translation tasks to the one of transforming simulated spectrograms into experimental ones, across several experiments the architectures described in [71] and [72] experienced severe mode collapse, and were not able to produce any reasonable output for the use case presented in this thesis.

For this reason, a custom architecture was designed to perform the task of augment-

ing simulated spectrograms into experimental ones. This follows the same principle as the classical GAN, but instead of utilising random noise as the latent vector, considers the simulated spectrogram as its input space, as was done for the convolutional mapping. This allows the generator to start from a point that is already closer to the final solution, which may help by making the generation process more stable and by propagating the input physical parameters to the output. The discriminator is left unchanged from the classic GAN approach, and in the training phase, it will learn to discriminate between samples from a batch of experimental data and a batch of paired synthetic ones. In principle, the pairing that was needed for the convolutional mapping is not required when using adversarial training. For the models proposed in this thesis, the matching between simulations and measurements will however be kept to ensure a better propagation of physical parameters from the input to the output. The study of the case where no matching between experiments and simulations is available is left for future research. Figure 4.9 shows the structure of the proposed model.

Contrary to the convolutional mapping using the MSE loss, this model employs an adversarial architecture to tackle the problem of the blurry output caused by the MSE loss. Note that an MSE loss could be introduced between the synthetic output and the experiment paired with the simulation. This is done to increase the fidelity of the output in the Pix2Pix architecture. In the experiments carried out in the scope of this thesis, this additional MSE loss resulted in additional blurriness of the generated spectrogram and was hence not employed for these studies on GANs using simulations.

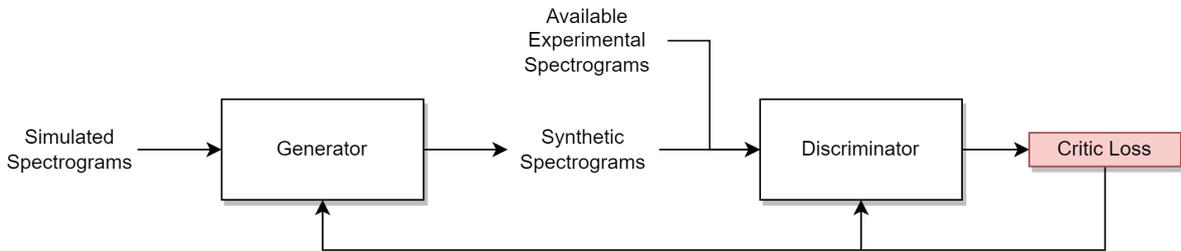


Figure 4.9: Overview of the structure of the GAN realised using simulations as latent vectors.

A few examples of synthetic spectrograms generated using the GAN approach from simulations are shown in figure 4.10. As expected, the critic loss employed for the adversarial architecture is successful in removing the blurriness that appears in the output when using the MSE loss. However, the generated spectrograms still appear visually distant from the experimental ground truth. In particular, while the envelope of the spectrogram can be clearly distinguished, the components related to the limbs of the subjects cannot be separated from the rest of the torso, and this could be an issue for improving the performance in the downstream task.

Figure 4.11 shows that by using a discriminator to compute the loss, a lower similarity between the synthetic outputs can be achieved when compared to the convolutional mapping case.

As was happening in the convolutional mapping case, however, using only the simulations as an input limits the generative properties of the model, as augmenting the same

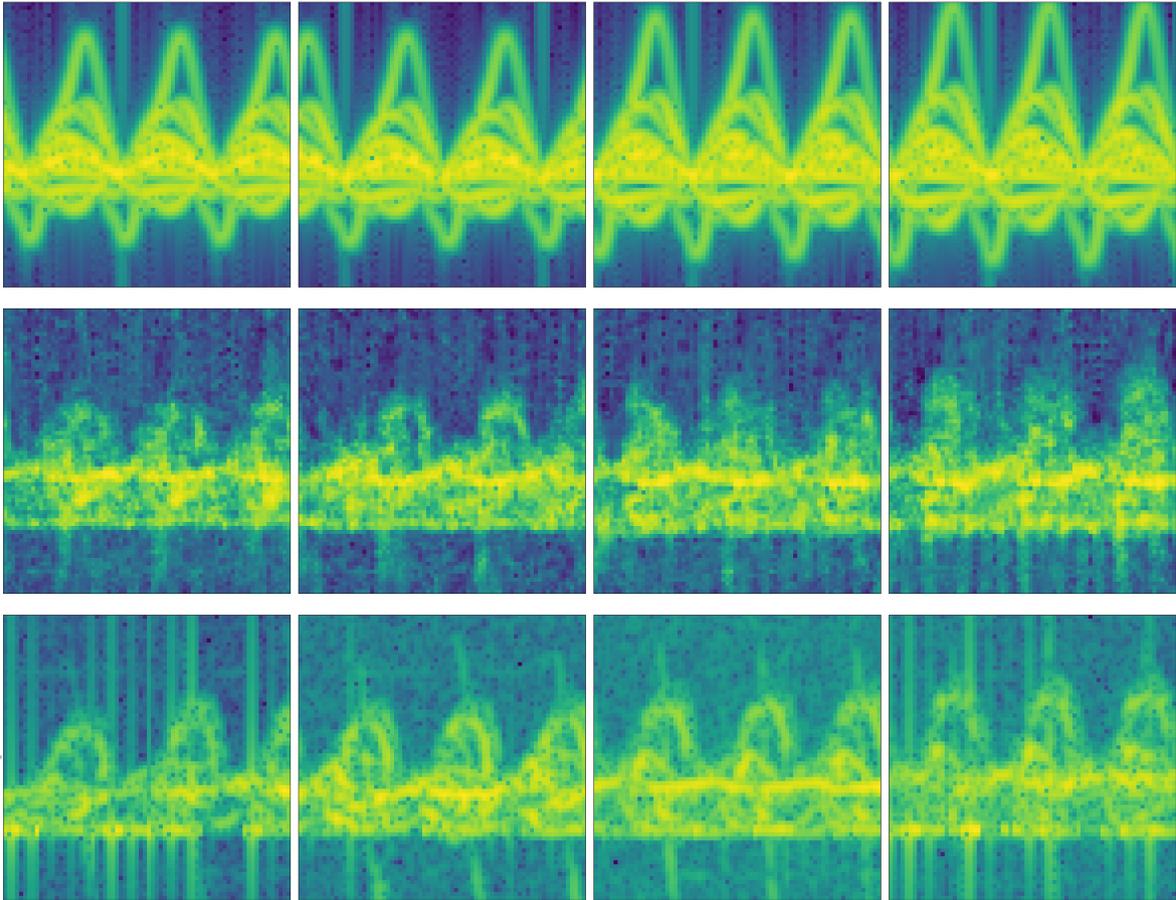


Figure 4.10: Visual demonstration of the spectrogram generated via a GAN with simulations as an input. Synthetic spectrograms (middle row) are generated from the respective input simulation (top row). The bottom row shows the experimental spectrograms paired to the respective input simulations and act as ground truth for comparison.

simulation twice will produce exactly identical synthetic spectrograms. To increase the ability of the GAN learning to reproduce imperfections and variations between the gait cycles, there is still the need to investigate possible ways to add generative potential to the augmentation model.

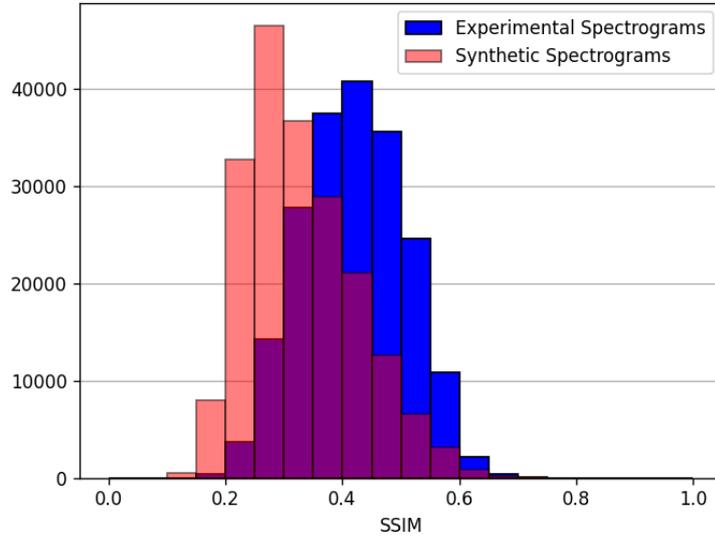


Figure 4.11: Distribution of the SSIM between each pair of synthetic spectrograms generated via GAN from simulations and between pairs of experimental spectrograms.

4.5 GANs from simulations plus noise

The most straightforward way to add generative properties to the network responsible for augmenting the simulations is to add noise to the input. To do this, the training step of the network is modified to include the addition of white Gaussian noise to the spectrograms every time an input is passed forward into the network, to mimic the sampling done in the classical GAN approach. Figure 4.12 shows a diagram of the training process.

The idea of adding noise to the input simulations can be seen as a hybrid approach between the simulation-only case and the classical GAN approach. When the variance of the overlaid noise is 0, then the network will behave as the previously illustrated simulation augmentation, while when the noise power overwhelms the input image, then the generator will behave as a GAN with an input latent vector with as many variables as the number of bins in the spectrogram. For this approach, the variance of the noise was chosen to be 0.1, as it was observed to be a good balance between the quality of the generated spectrograms and their variability for different instances of noise.

The visual quality analysis, different from the previous cases, is performed on the visual differences of different synthetic spectrograms generated from the same simulated one. This is done to highlight the effect of introducing the addition of noise to the input in the training step. Figure 4.13 shows five synthetic spectrograms generated by feeding the input simulation into the generator after adding different instances of white Gaussian noise.

While the synthetic spectrograms do not seem to differ substantially in nature, the addition of noise generates slight discrepancies in the envelope across different steps and

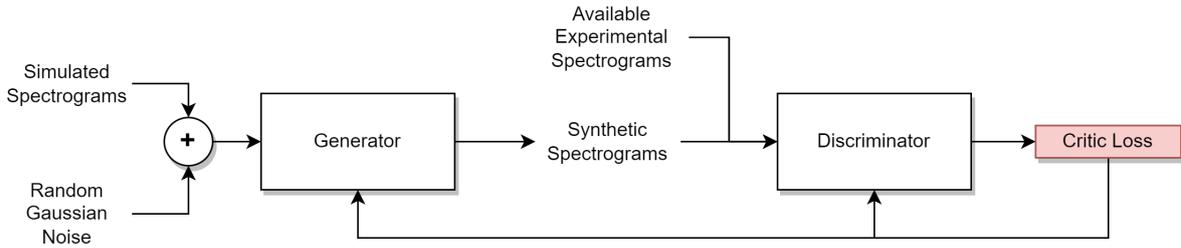


Figure 4.12: Overview of the structure of the GAN realised using simulations overlayed with noise as input.

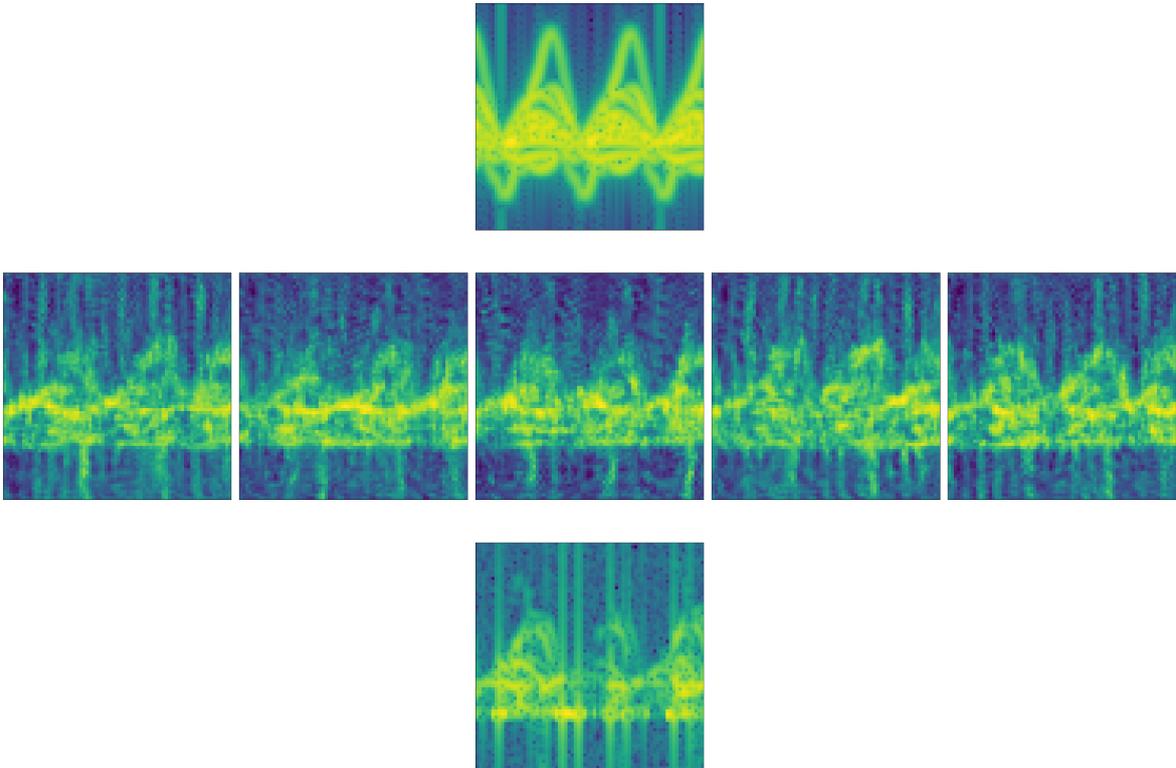


Figure 4.13: Visual demonstration of the spectrograms generated via a GAN with simulations plus noise as an input. Synthetic spectrograms (middle row) are generated from the input simulation (top row) and 5 different instances of overlaid noise. The bottom row shows the experimental spectrogram paired to the respective input simulation and acts as ground truth for comparison.

in the Doppler velocity of the limbs, which however don't seem to bring a considerable improvement in quality when compared to the experimental reference.

Figure 4.14 shows that indeed the overlaid noise contributes to decreasing the similarity of the generated spectrograms significantly below the similarity between experimental ones. While this may be an advantage in terms of mode collapse, it could also mean that the noise is impacting the quality of the synthetic spectrograms neg-

atively. While typically the dimension of the noise latent vector in a GAN is limited to a few hundred, the noise overlayed to the spectrogram has almost 5000 variables. This can result in the latent space containing several irrelevant or redundant dimensions, leading to difficulties in learning the appropriate mapping. As a consequence, the quality of the output may be worsened when compared to architectures that use smaller random input vectors.

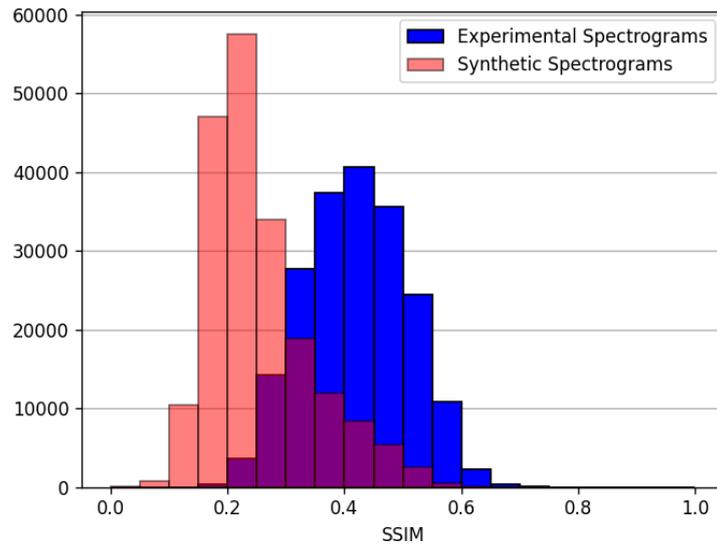


Figure 4.14: Distribution of the SSIM between each pair of synthetic spectrograms generated via GAN from simulations plus noise and between pairs of experimental spectrograms.

4.6 Multi-branch GAN

To reduce the disadvantages of a too large latent space, the architecture can be modified to include two different processing branches. The first will focus on augmenting the theoretical simulation input, while the second one will process a vector of 100 normally distributed random variables, such as the one utilised in the classical GAN approach. The output of the two branches will be then summed to obtain the generated synthetic spectrogram. Figure 4.15 illustrates the architecture of the proposed generative model.

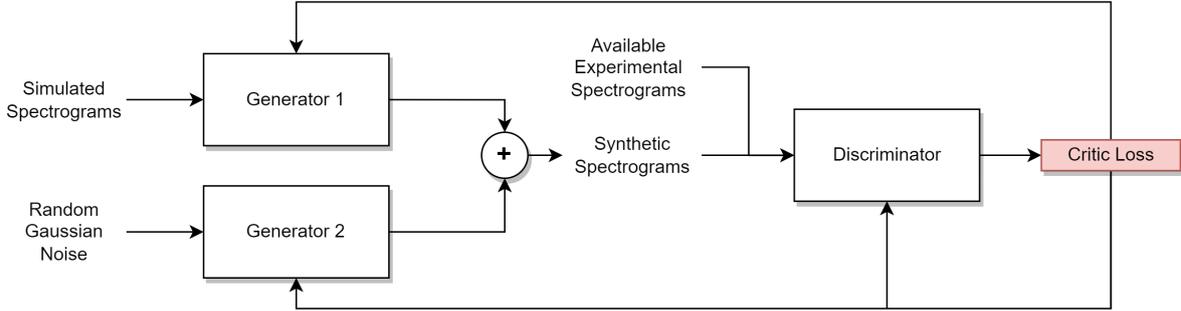


Figure 4.15: Overview of the structure of the GAN realised using two generator branches responsible for processing the input simulations and a random noise vector.

Intuitively, this approach tries to bias the network towards separating a contribution to the physical content from the input simulation and a contribution related to random effects which are modelled via the noise input. Note, however, that this bias is not enforced apart from the changed architecture, and the generative process is left free to choose what kind of contribution shall the simulation and the random noise input bring.

Figure 4.16 shows different augmentations of the same input theoretical simulation using different instances of the noise vector. The different Doppler components are distinguishable, especially the one associated with arm movement. In addition to that, the variation between different realisations of the augmentation is noticeable, both in the shape of the envelope and in the returns of the different limbs.

The similarity between generated spectrograms is also again closer to the one between experimental ones, as figure 4.17 clearly shows.

To assess the processing operated by the two branches, the generated spectrogram can be decomposed into the contribution from coming from the two different inputs. Figure 4.18c and 4.18d show the output of the simulation branch and noise branch respectively. As can be seen, the simulation is used to reconstruct the background of the image, both in the frequency bins without any micro-Doppler signature and in the area around the return of the torso. Most of the visual features of the final output, instead, are modelled by the noise input, reinforcing the hypothesis that a random noise input is fundamental to modelling the variations in the experimental spectrograms.

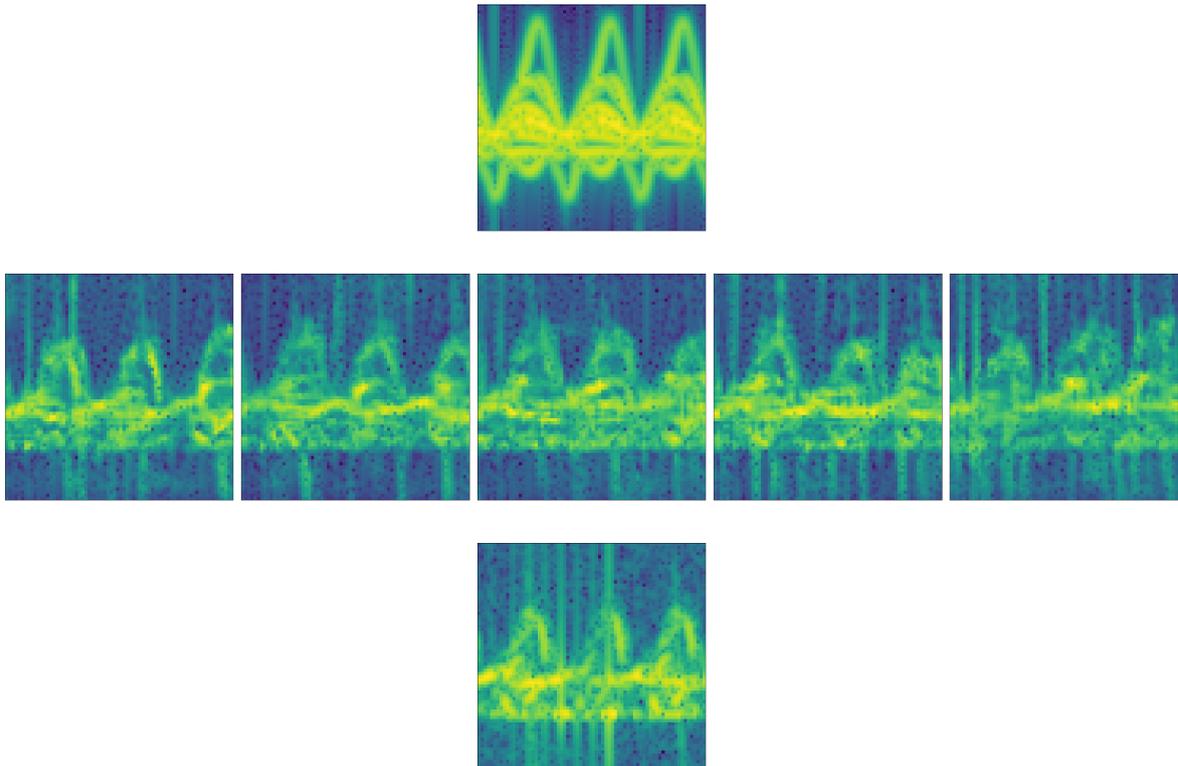


Figure 4.16: Visual demonstration of the spectrograms generated via a GAN with multiple generator branches. Synthetic spectrograms (middle row) are generated from the input simulation (top row) and 5 different instances of input noise. The bottom row shows the experimental spectrogram paired to the respective input simulation and acts as ground truth for comparison.

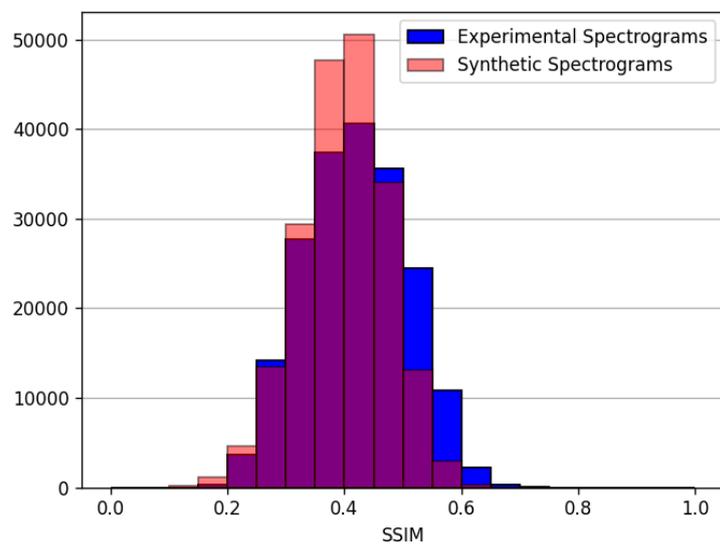


Figure 4.17: Distribution of SSIM between each pair of synthetic spectrograms generated via multi-branch GAN and between pairs of experimental spectrograms.

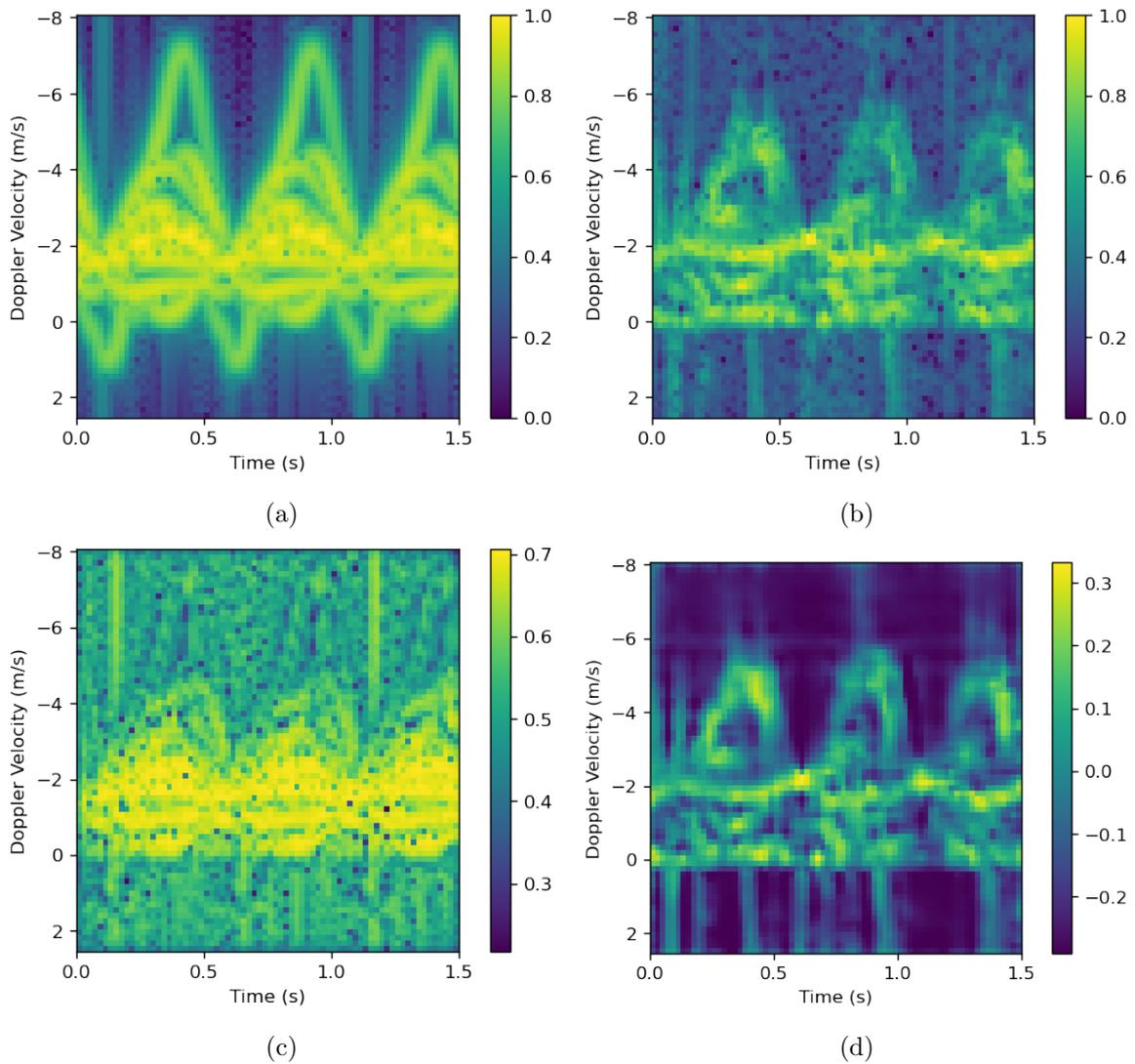


Figure 4.18: Overview of the input and outputs at different stages of the multi-branch GAN architecture. (a) The input simulation is fed into the network to produce (b) the synthetic spectrogram. (c) The output of the simulation branch is responsible for the background of the image while (d) the output of the noise processing branch reconstructs most of the Doppler returns of the limbs.

4.7 Effect on the downstream classification task

The final evaluation will be done by assessing the performance of a classifier trained with augmented data in the downstream classification task. Using the available experimental spectrograms a generative network is trained to produce several synthetic ones. As the augmentation needs to be class dependent, one generative network per class will be trained, with the purpose of generating synthetic samples belonging to each class separately. Note that available simulations only belong to walking individuals, hence when generating synthetic samples for the rifle class from simulations the augmentation process will also have to enforce the specific class label on top of the experimental measurements' visual features. While architectures such as conditional GANs (C-GANs) [73] and auxiliary-classifier GANs (AC-GANs) [74] exist to tackle multi-class generation, the separate augmentation approach is chosen. Due to the greater complexity and training instability of these architectures, the usage of C-GANs and AC-GANs is left for future research.

A convolutional NN is chosen as the model to perform the classification task and assess the improvement achieved with the different augmentations. Appendix A gives more insight into the process of designing a classifier for the downstream task adopted in this thesis.

To properly evaluate the classification, 20% of the total amount of data will be reserved for testing purposes only. Out of the remaining experimental data, 20% will be used for validation and the remaining part will be used for training the classifier. In order to assess the impact of the generative augmentations on the generalization capabilities of the downstream classifier, the split between the training, validation and test set will be done subject-wise, to ensure the maximum degree of independence between the training and test data. To achieve more significant results, the experiments will be run over 10 different random splits of the dataset. The achieved accuracy for each split is then averaged to obtain the final scores.

In addition to that, to evaluate the effect of the augmentations under various degrees of data scarcity, the training dataset for both the classifier and the generative augmentation will be progressively reduced by removing subjects.

For each size of the training set, the classifier is first trained with the available experimental data to establish a baseline. After that, the same data is used to train the generative augmentation under test. The generative network is then used to produce new synthetic samples that, on top of the available experimental ones, are used to train another instance of the downstream classifier. The newly achieved accuracy using the synthetic data is then compared with the baseline value. Figure 4.19 shows the augmentation and classifier training pipeline.

In the case of augmentations that employ theoretical simulations, the synthetic spectrograms are generated using a set of 500 input simulations. For the ones that employ noise in the generation process, two noise instances are sampled per input simulation, leading to the production of 1000 synthetic spectrograms. In the case of the classic GAN approach, 1000 spectrograms are generated by sampling the random input variables.

Figure 4.20 shows diagrams of the splitting of the dataset for different levels of

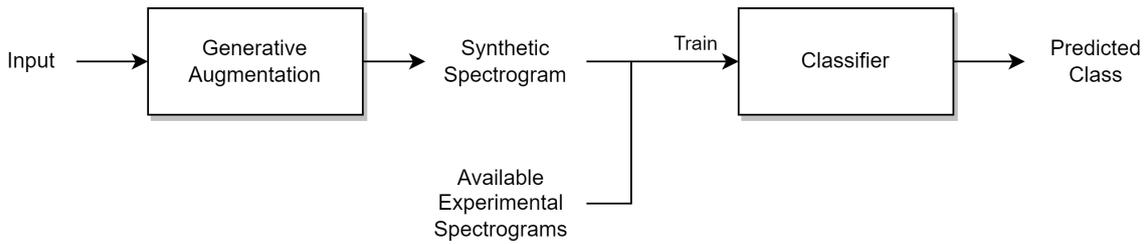


Figure 4.19: Overview of the augmentation pipeline used to assess the impact of generative augmentations.

data scarcity, without augmentation in 4.20a and with augmentation in 4.20b. Note that the reduction in the training set applies both to the classifier and the generative network, which is done to correctly evaluate how the effectiveness of the generative augmentations scales with the amount of available data.

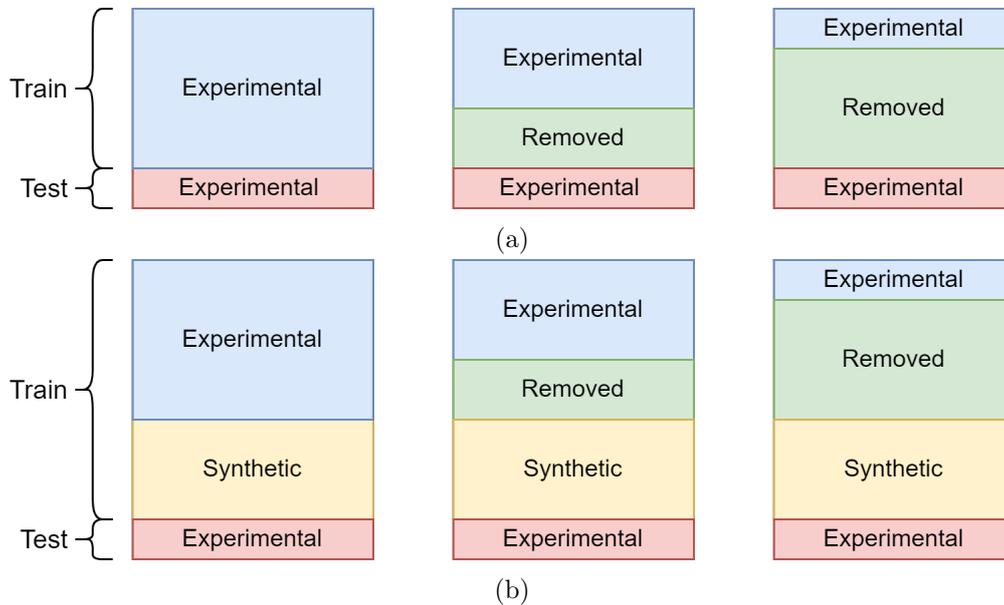


Figure 4.20: (a) Splitting of the dataset to account for different degrees of data scarcity, (b) when also including synthetic data. The blue and yellow portions are used for training, while the red one is used for testing. The green portion represents the experimental data that is progressively removed to simulate a data-scarce scenario. The validation set is taken from 20% of the experimental data assigned for training shown in blue.

In the scope of this work, accuracy will be used as a metric for the classification task. In a practical scenario, it could be argued that detecting an armed individual is more crucial than correctly classifying unarmed ones, and hence metrics such as precision, recall or the F1-score should be considered.

While this argument is true, in view of the balanced nature of the two classes and to

allow for better comparison between this thesis and work carried out in other scenarios in the literature, accuracy was picked as the sole metric of assessment.

Figure 4.21 shows the accuracy achieved in the classification task for different levels of scarcity when the proposed augmentations are utilised to augment the training set.

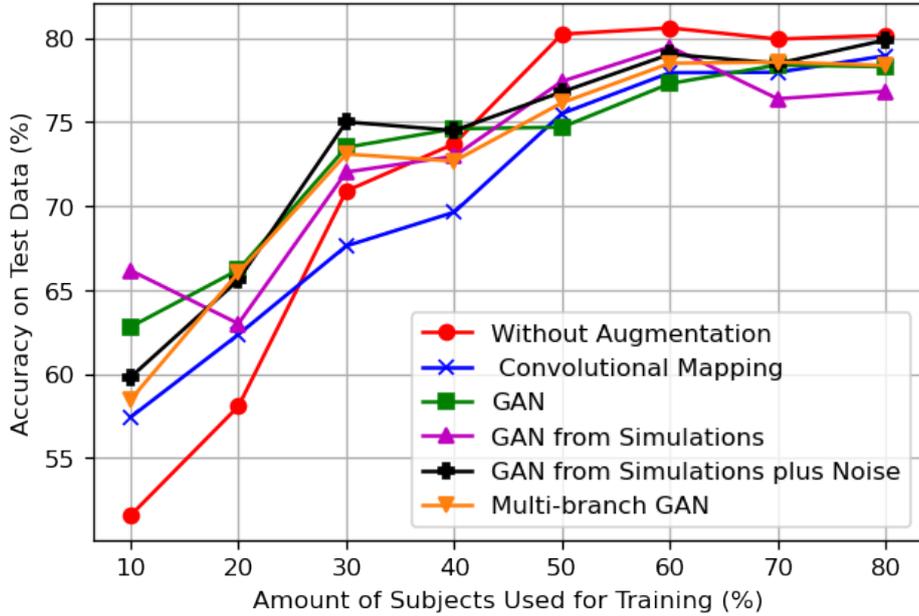


Figure 4.21: Accuracies achieved in the downstream task using different generative augmentations, averaged over 10 data folds. The standard deviations are omitted for clarity.

As can be seen when training with more than 40% of the available data the augmentations do not bring any improvement in the downstream task. Despite their visual quality, the synthetic spectrogram may fail to capture and reproduce useful features for the classification task and contributing to adding more confusion in the classification process.

On the other hand, when utilising less than 40% of the data the synthetic spectrograms seem to be beneficial to the classification task. Note that the training data available for classification is the same as the training data used for the augmentations. As the amount of available subjects decreases, so does the accuracy of the classifier and the quality of the spectrograms produced via the generative networks. Visual examples of the quality of synthetic spectrograms with respect to the amount of training data are provided in appendix C.

Despite this phenomenon, the augmented samples still prove useful in regularizing the learning in data-scarce situations and preventing overfitting. When it comes to the benefit of using simulations, however, no real advantage is found over the classic GAN approach, except for a small increase in accuracy when using 10% of the available data and augmenting with a GAN from simulations without any introduction of noise.

This lack of benefit from using the simulation may be due to the too-abstract nature of the model used. Instead of providing a physically reasonable starting point for the generative networks, the simulations may be constraining the augmentation process

too much by limiting its generative potential and requiring the introduction of noise to achieve a better output quality.

In addition to that, the observational bias introduced by using the simulations doesn't strictly ensure that the physical parameters of the theoretical input will be propagated to the output.

Table 4.1 shows the error between the input parameters of the simulation and the output synthetic spectrogram. Apart from the error that is introduced by the parameter estimation process described in B.2, both the mean velocity and gait cycle duration of the output are slightly different from the ones of the input.

Model	Mean velocity error	Gait cycle duration error
Convolutional Mapping	4.4%	2.4%
GAN from simulations	7.8%	7.1%
GAN from simulations and noise	8.5%	9.6%
Multi-branch GAN	7.7%	6.6%

Table 4.1: Mean relative error of mean velocity and gait cycle duration between synthetic outputs and input simulations

In particular, convolutional mapping seems to perform better in this sense, as there the MSE loss enforces reconstruction between simulations and experimental spectrograms with matched parameters. On the other hand, augmentations that rely on adversarial training seem to have a greater error for the produced parameters.

This can be related to multiple reasons. First, the visual properties of the synthetic spectrograms may affect the parameter estimation process, as the peaks and the torso component may not always be as clear as in the experimental spectrograms.

Secondly, for approaches that rely on noise, the error could stem from the fact that the components used for the estimation of the parameters, namely the envelope and the torso, are influenced more by the noise input rather than the simulation. For this reason, a mismatch between the input and output parameters could be introduced in the augmentation process.

To improve on this, future work could focus on adding a loss term or an adaptation in the architecture to penalise such mismatch in the input and output parameters, to ensure better reconstruction and enable better-targeted generation for synthetic spectrograms with specific parameters. Further research could also investigate adversarial training without matched pairs of experiments and simulations, to assess whether the proposed approaches could manage to reliably propagate the physical parameters of the input when discriminating the synthetic spectrograms from measurements with different mean velocity and gait cycle duration.

Furthermore, it would be of interest to assess whether an augmentation model trained on experimental data with only a subset of the kinematic parameters would be able to produce reasonable outputs when augmenting simulations with parameters outside of the training set. This could be useful in the case where measurements are only available for targets with determined parameters, and synthetic spectrograms of targets with a different set of parameters need to be generated. An example of this

could include a defence scenario where only measurements of allied drones are available, and knowing the physical specification of enemy unmanned aerial vehicles (UAV) could allow the generation of synthetic spectrograms that could be used to train a classifier responsible for target detection.

Another important factor to point out is that for the specific problem of classifying armed and unarmed individuals, the physical parameters are not correlated with the class of the target. This means that even if the generative augmentations are able to produce physically feasible spectrograms, these may not contribute to improving the downstream classification task unless they properly capture the relevant features for determining the class of the spectrograms.

On top of that, both of the classes are augmented using theoretical simulations of walking individuals, which in the rifle case introduces the extra difficulty for the augmentation to also learn how to enforce the class-specific features in the generated spectrograms.

For these reasons, generative augmentations, even if physically informed by the use of simulations, struggle to provide any concrete benefit in the classification task.

To obviate most of the problems of generative augmentations, image manipulation will be investigated, with particular attention to their role with respect to the underlying physics of the micro-Doppler spectrograms.

Image Manipulation for Radar Spectrograms

5

In this chapter, image manipulations will be explored as an alternative DA technique to generative networks. Compared to the latter, image transformations are computationally cheaper and can be applied without regard to the amount of available data. However, their application can easily compromise the underlying physical content of the spectrograms and has to be carefully studied.

With this in mind, while not directly leveraging the governing laws of the target domain, these visual augmentations can be designed to be physics-aware and coherent with the underlying physical content of the data representation.

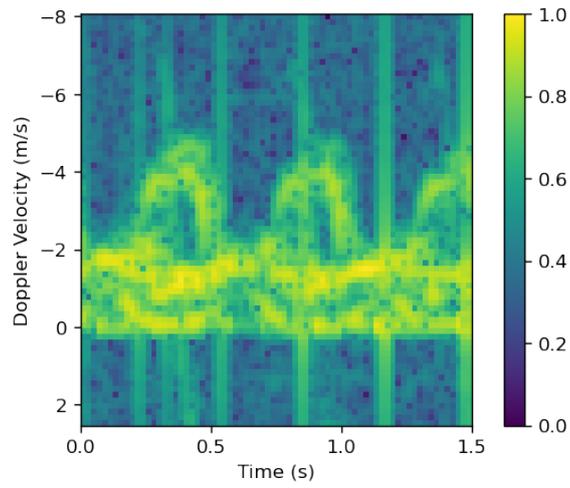
In this thesis, two techniques from the literature, namely rescaling in time and frequency and circular time shift will be compared with a third technique that exploits the cyclic nature of the walking motion.

5.0.1 Rescaling in time and frequency

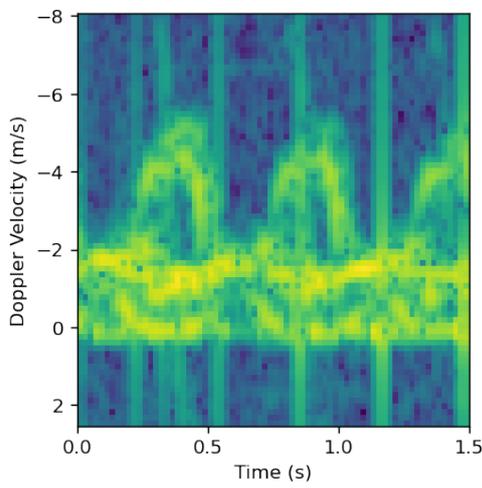
The rescaling in the time and frequency domain applied in [41] aims to generate new samples by considering variations in the gait of the subjects. The time rescaling mimics a variation in the pace and the duration of the gait cycle of the subject. The rescaling in frequency, on the other hand, implies a faster swing of the outer limbs such as legs and arms. These transformations modify the arm component that is considered relevant to classification by stretching and compressing it in time and frequency, but in doing so they do not alter the original class label.

Figure 5.1 shows examples of rescaling in the frequency and time domain. For both the time domain and frequency domain, the rescaling factors chosen are 0.8, 1 and 1.2. To preserve more physical feasibility, the augmentation in the frequency domain is applied only to the frequencies outside the range that comprises the return from the torso, which was chosen to be between -2.1 m/s and -0.9 m/s.

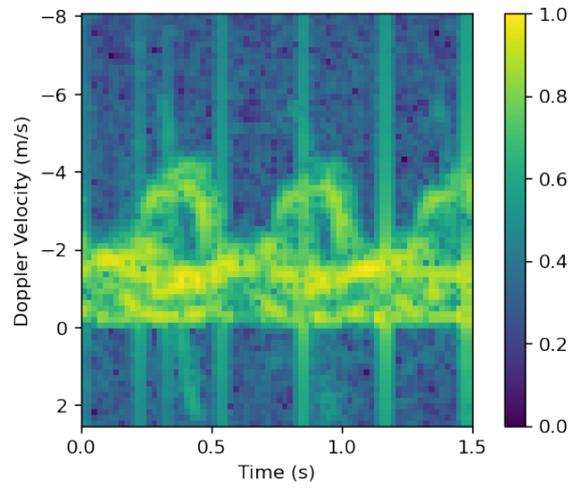
It is important to note that different from the work presented in [41], in the scope of this thesis the rescalings are directly applied to the magnitude of the final spectrogram rather than to the raw IQ radar returns. While the latter approach is more accurate and doesn't require any pixel interpolation, the former is closer to the visual transformations applied in the computer vision domain. Nonetheless, the motivation behind the two is the same and the outcome is comparable.



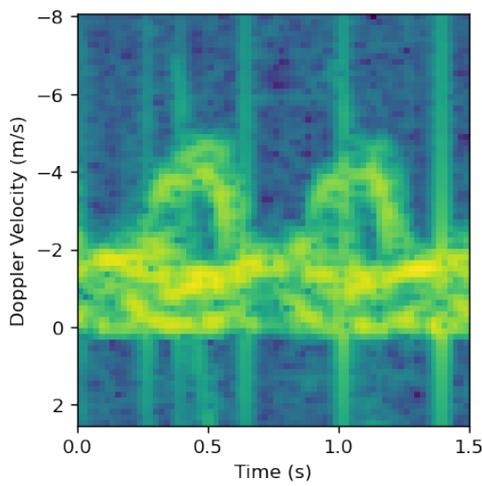
(a) Original



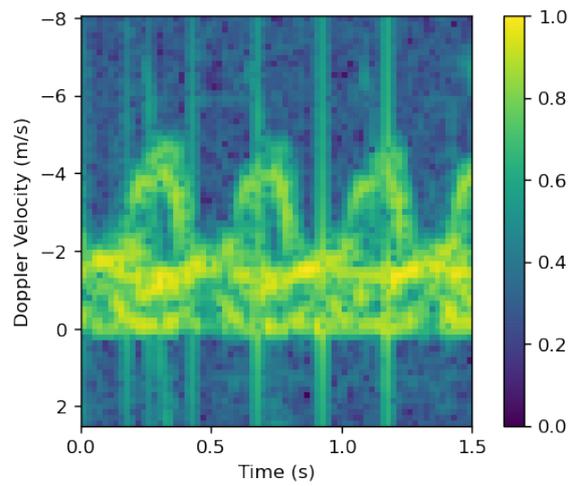
(b) Stretch in time domain



(c) Compression in the frequency domain



(d) Stretch in the frequency domain



(e) Compression in the time domain

Figure 5.1: Examples of rescaling in time and frequency domain.

5.0.2 Circular time shift

A circular shift in time has the benefit of preserving the features and characteristics of the target's signature. By changing the initial phase and moving the different gait cycles around, this augmentation can make the classifier less prone to fitting the entire spectrogram and focus more on the relevant features of the two classes. A circular shift in time will also introduce a discontinuity, which however may not necessarily negatively impact classification.

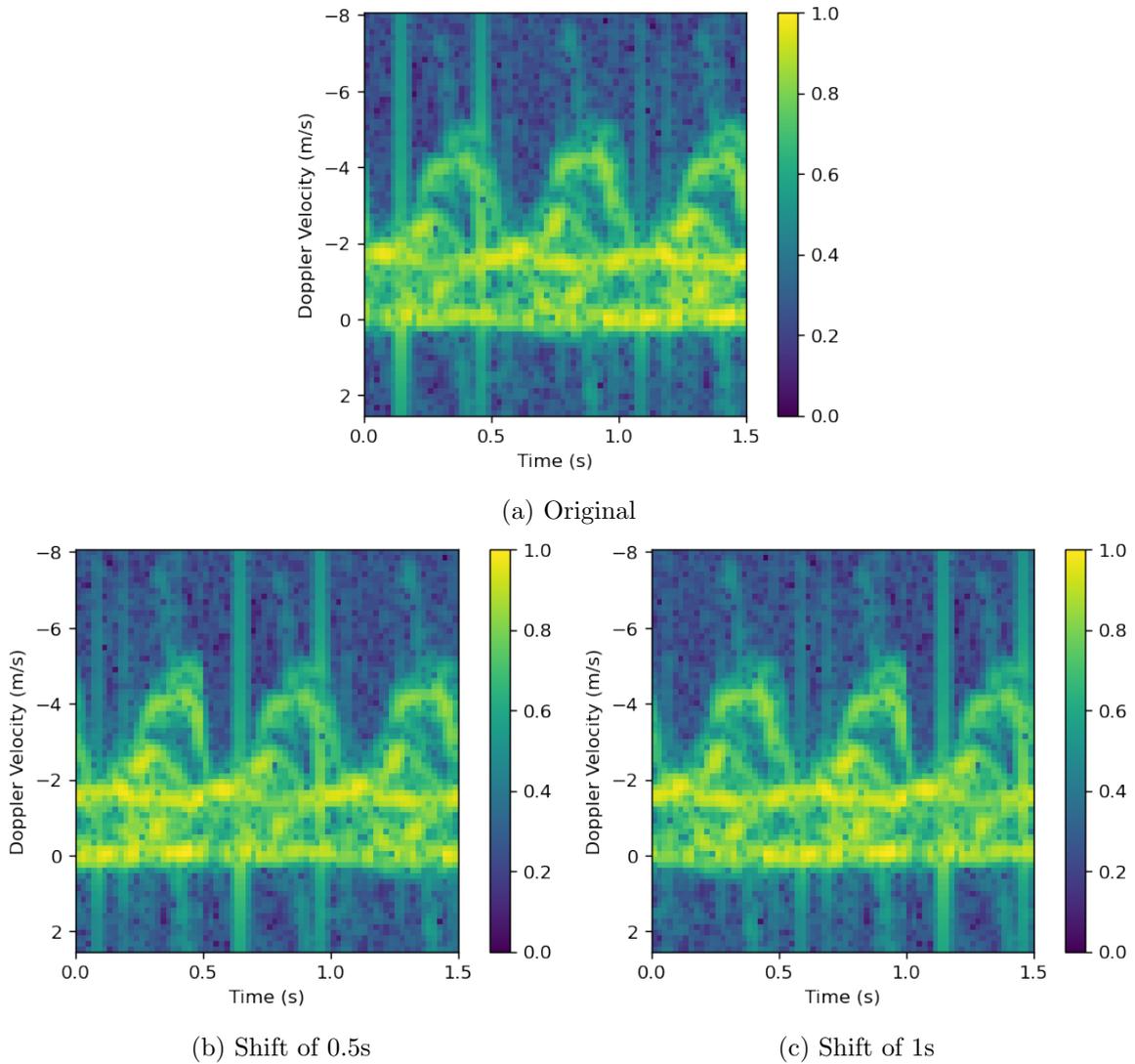


Figure 5.2: Examples of circular shifts in time domain.

Figure 5.2 illustrates an example of a circular timeshift applied to an experimental spectrogram. In the scope of this thesis, two circular shifts of 0.5s and 1s will be used to augment the training dataset.

5.1 Evaluation

To evaluate the impact of visual augmentation a similar process to the one utilised for generative augmentation is employed.

A classifier is trained using the available experimental data and the classification accuracy is then compared to the one achieved by the same classifier trained on the experimental spectrograms and their augmented versions. The experiment is repeated by progressively removing part of the training data, and for each size of the training sets, it is repeated for 100 different subject-wise splits of the data.

Figure 5.3 shows the classification accuracy when training with the available experimental samples and their augmentations. As can be seen, the rescaling and timeshift augmentations have a similar impact on classification, apart from marginal differences. For both of them, a considerable gain is achieved when using limited training data, which goes up to 13% when using 10% of the subjects for training. The gain from the augmentations saturates as the available training data increases, leading only to marginal improvements when using more than 60% of the data for training.

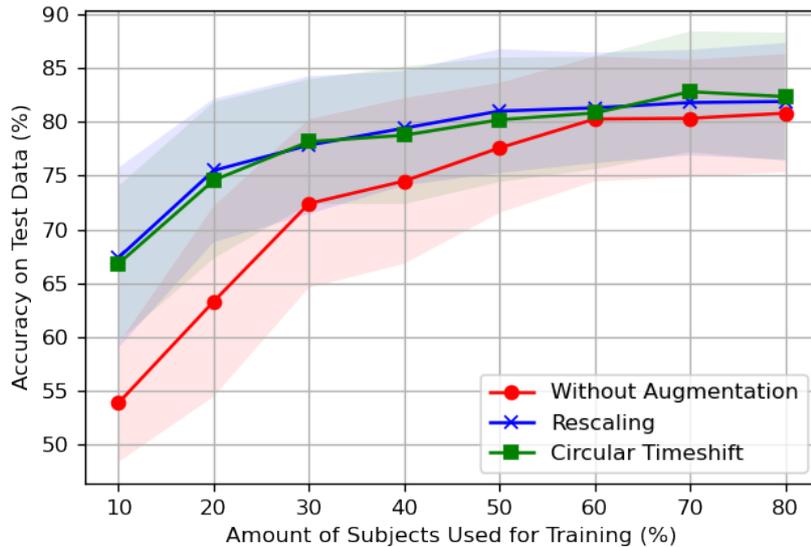


Figure 5.3: Mean and standard deviation of accuracy scores over 100 data folds using different augmentation techniques.

While altering the underlying parameters of the walking person or introducing discontinuities in the spectrogram, these augmentations are successful in increasing the final classification accuracy with all degrees of data scarcity.

This may be due to the fact that these augmentations do not alter the defining features of the two classes, and by leveraging invariances such as the duration of the steps, velocity of the limbs and the position and order of different steps in time, they regularize the training process of the model and help it better identify the relevant features to separate the two classes.

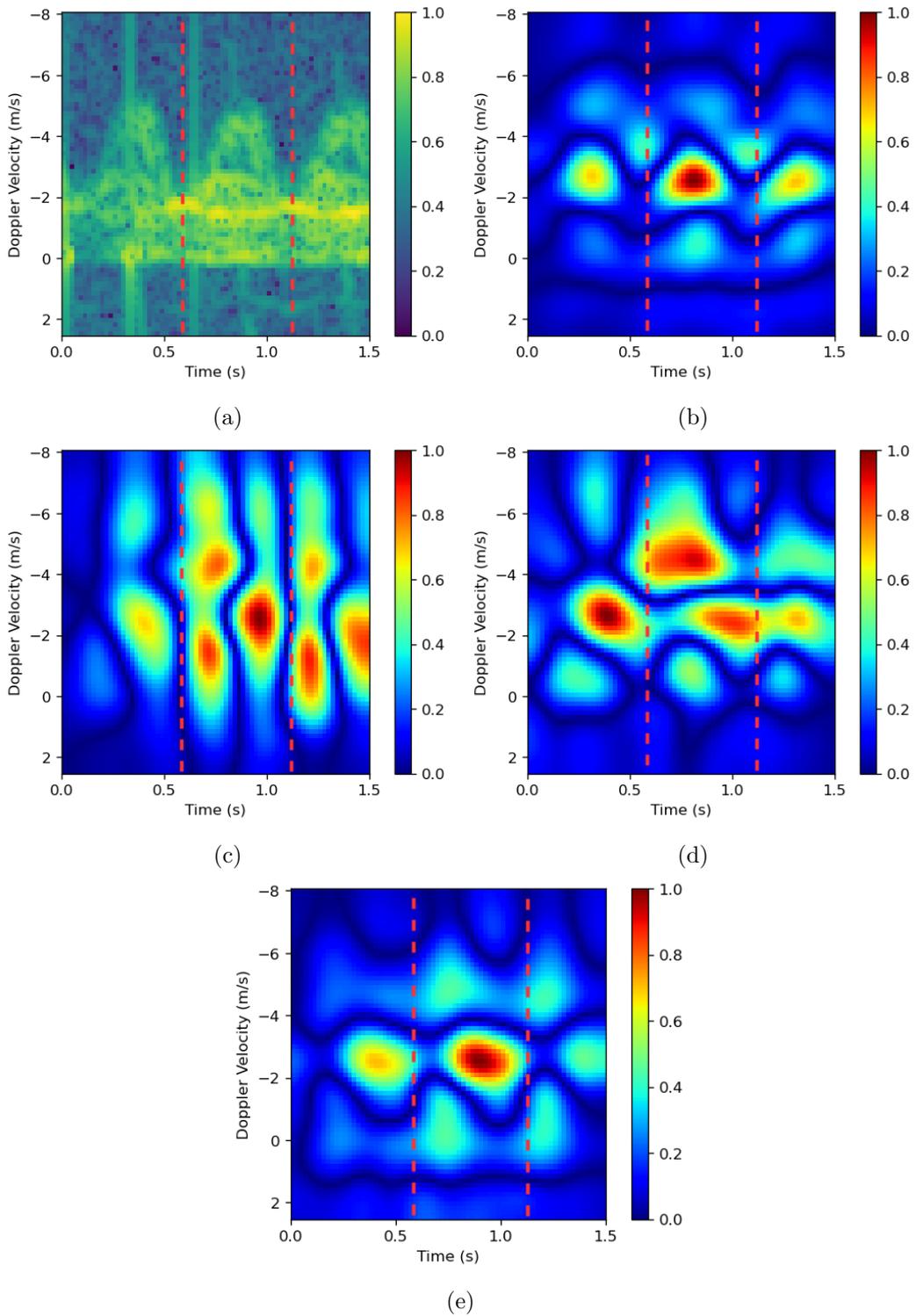


Figure 5.4: Saliency maps for different classifiers. (a) The input spectrogram of a walking person is classified. The different saliency maps are shown when using (b) a classifier trained on 80% of the available subjects, (c) 10% of the available subjects, (d) 10% of the available subjects but after applying the rescaling augmentation and (e) 10% of the available subjects after applying the circular timeshift augmentation. The red dashed lines show the separation between different steps.

To better understand the effects of these augmentations, the saliency maps of the different classifying scenarios can be examined. Saliency maps show the magnitude of the gradients of the class score with respect to every pixel of the input spectrogram and are a way to measure the contribution of different areas of the input sample towards the predicted label [75].

Figure 5.4 shows the saliency maps of the classifier in different classification scenarios when using the experimental spectrogram shown in figure 5.4a as an input. Each of the saliency maps represents the absolute value of the gradient of the output label with respect to every input pixel, which is scaled in the $[0,1]$ range based on the maximum and minimum gradient for each image. To smoothen the saliency map, a Gaussian filter is applied.

Figure 5.4b shows how when training with all the available subjects the classifier is clearly able to recognize the distinctive feature of the arm for a walking person. When restricting the number of training subjects to 10% of the total available amount, figure 5.4c shows that the classifier is unable to pinpoint the same features, but relies on the returns of multiple limbs as well as the background between different steps, which leads to the original input being classified as belonging to a rifle carrier.

When including the rescaled data in the training set, however, figure 5.4d shows that the feature extraction resembles again the case when most of the data is used and the output label is corrected, meaning that the augmentation successfully mitigates the lack of training data. An important contribution to the gradient can be observed coming from the envelope of the central step of the spectrogram. This may be due to the fact that when augmenting the data by stretching the spectrograms in frequency, the component of the arm may be moved to higher frequencies and be confused with the return of the leg in the original non-augmented spectrogram. However, this phenomenon doesn't seem to affect the improvement in classification, not in this example nor in the general trend.

When applying the circular time shift augmentation, similar conclusions can be drawn. As shown in 5.4e, the augmentation again brings the focus back to the micro-Doppler component of the arm, hence correcting the output class label.

Despite these augmentations succeeding in improving the downstream classification task in data-scarce scenarios, the question remains open of whether a visual transformation that better complies with the physical model of the walking subject may bring an ulterior improvement.

5.2 Classification of a single step

Since the classifier seems to already focus on the features of a single step, a new augmentation scheme can be realised by segmenting the input in order to classify each gait element separately. This approach has several benefits:

- The risk of the classifier learning dependencies between different steps is removed. As one gait element may hold enough information to characterize the target as being unarmed or not, splitting the input into different segments prevents the classifier from overfitting and memorising the whole spectrogram. From the per-

spective of physics-informed machine learning, this can be seen as applying an inductive bias through which the classifier is informed of the cyclic nature of the walking motion and its attention is artificially pointed to the features of the single element of the gait.

- By reducing the size of the input, the number of parameters of the NN classifier decreases and the amount of training samples grows, leading to less risk of overfitting.
- Separating the different steps also opens up new possibilities for the label prediction task, as the information from different steps belonging to the same original spectrogram can be combined to reach a unanimous decision.

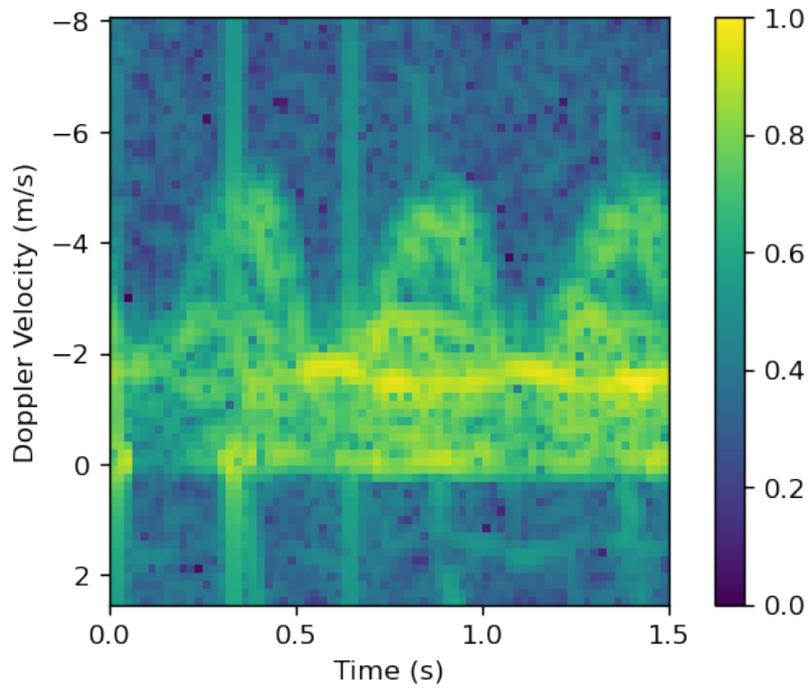
To implement the segmentation, the spectrograms are split by means of a window centred around the peaks of a sinusoidal curve that is fitted to the spectrogram’s envelope, using the procedure described in section B.2. Figure 5.5 shows the visual result of framing individual steps. To assess the performance in the downstream task, the structure of the classifier is not changed and is kept the same as the one shown in section A.1 used for the previously introduced augmentations.

When classifying the individual steps separately, figure 5.6 shows that there is a gain in accuracy in data-scarce scenarios, which can be related to the reduction in the model’s number of parameters and the increase in the number of training samples that comes from splitting the original spectrograms. However, when using 60% of the available data or more, the average accuracy is slightly lower. This can be explained because, in the presence of enough training data, the classifier that evaluates the whole frame can rely on multiple steps at the same time to make its final decision.

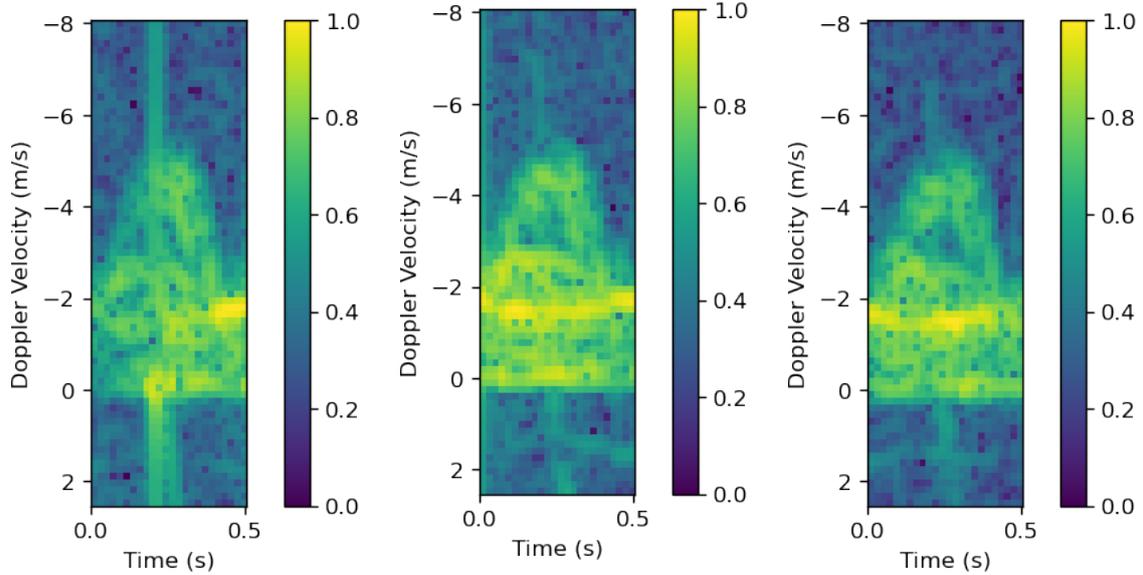
When looking at the saliency maps of the single steps, figure 5.7 shows that just applying the inductive bias to artificially inform the classifier of the cyclic nature of the input is enough to direct the attention to the relevant feature of the spectrogram, explaining the gain in accuracy for data-scarce scenarios observed in figure 5.6. When classifying the single steps there is also a discrete amount of relevancy found in the upper and lower envelope of the gait element. While no clear explanation was found for this phenomenon, it is not to exclude that the NN classifier may find features that are useful to separate the two classes that are not evident to a human observer.

While the inductive bias is already enough to produce improvements in data-scarce regimes, an ulterior benefit can be achieved when the information regarding the fact that different steps belong to the same experimental spectrogram is utilised. In particular, the softmax output of the classifier of different steps can be combined so that the final classification decision is made using the information belonging to several steps, as it is done implicitly by the layers of the classifier when classifying the entire spectrogram frame. To achieve this, the steps belonging to the same test spectrogram are passed through the classifier to obtain the classification output of the softmax layer. After that, these are averaged together and a threshold is applied to make the final decision on the target’s class. The results are shown in figure 5.6, where this approach shows an improvement in accuracy even when using the full training dataset.

The improvement comes from the fact that while one step may be classified wrongly, a correct classification of the other steps belonging to the same frame may compen-



(a) Original



(b) Three separate steps

Figure 5.5: Illustration of the process of dividing the spectrogram into separate steps

sate for it, leading to an overall increased accuracy in the test phase. In such a way, the softmax output of the classifier for each step is used as a confidence measure to determine the label of the entire spectrogram. By summing the contribution of each step before thresholding, the information from the different steps is combined equally

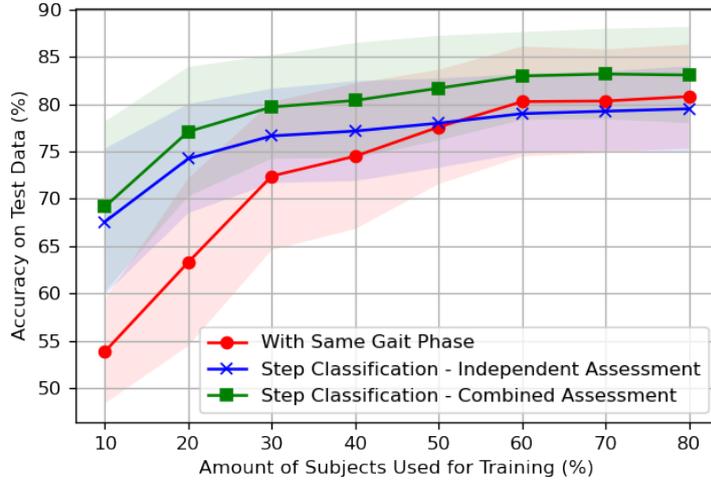


Figure 5.6: Mean and standard deviation of accuracy scores over 100 data folds when using a classifier trained on single steps, with independent and combined assessment.

but is not to exclude that weights could be applied to the sum to account for different reliabilities, that could be due, for example, to the amount of noise and clutter present in the single step.

Figure 5.8 compares this approach to the rescaling and the circular timeshift augmentations. The gain in accuracy between the first two and the classification of the single steps is marginal but consistent between 1% and 2%. Augmenting the single steps, however, has the benefit of reducing the complexity of the model by shrinking the input size, achieving the same results with less than half of the parameters required to classify the full spectrograms.

In addition to that, classifying the single steps has the benefit of being more flexible when it comes to adjusting the input size. While when classifying the whole spectrogram changing the input size to account for more observation time would require retraining the classifier from scratch, classifying extra steps would only imply considering additional inputs, without any need for retraining nor any increase in the number of parameters of the classifier.

5.2.1 Conclusion

Image manipulations such as the one discussed in this thesis proved a computationally-efficient solution for improving the performance in the downstream classification task, especially in scenarios with a low amount of training data. Despite meddling with the underlying physical nature of the micro-Doppler spectrograms, these transformations do not alter the features that make the classification possible and hence can help the classifier focus on the important micro-Doppler components. In particular, classifiers trained with few samples will tend to overfit the data and try to learn how to classify the whole spectrogram without isolating the underlying features of the data. This can be done by using augmentations which provide new samples with unvaried class-

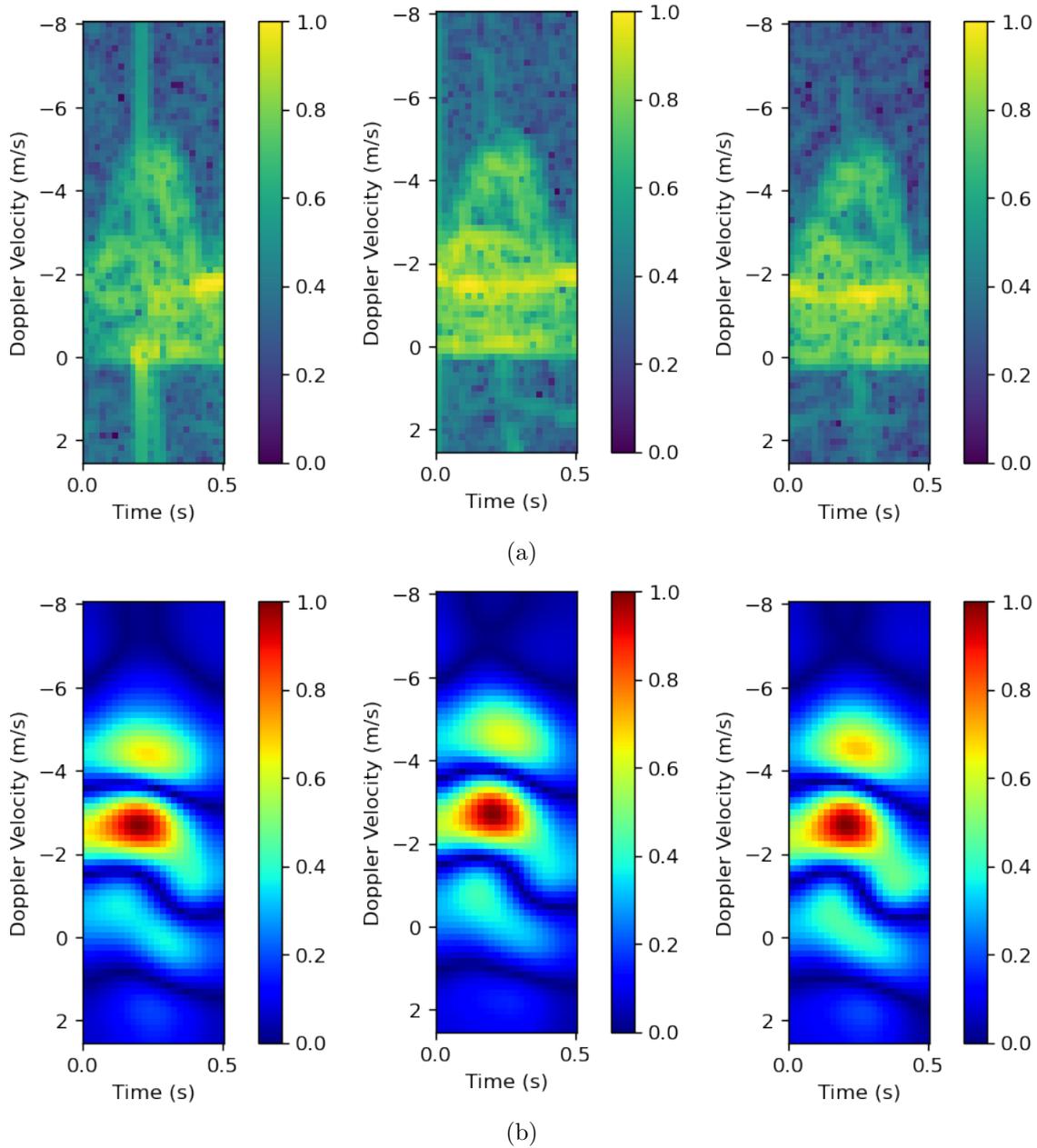


Figure 5.7: (a) Three steps belonging to the same original spectrogram and (b) their respective saliency maps when using a classifier with 10% of the training data available

related features. To improve on the physically reasonable augmentations provided in the literature, an inductive bias is applied to the classifier by separating the single steps in the input in order to direct the classifier’s attention to the feature of the single step.

This also comes with a reduced number of parameters and an increased amount of available training data, which can in general help protect the classifier from overfitting.

Finally, when reaching the final classification decision by combining the steps belonging to the same original frame, the physics-informed inductive bias leads to a marginal

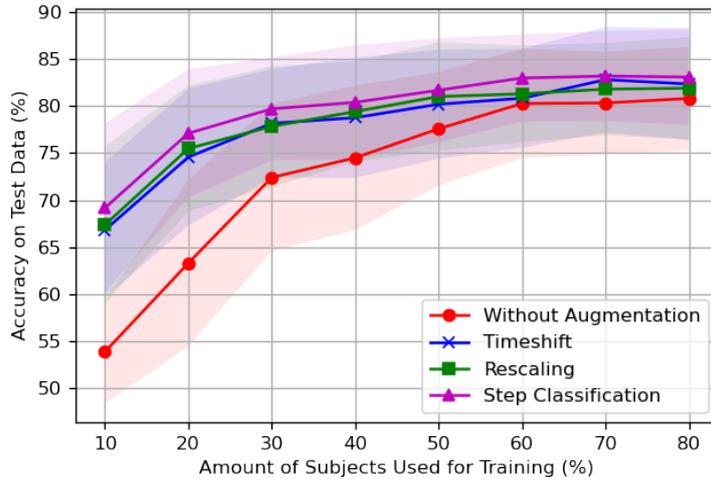


Figure 5.8: Mean and standard deviation of accuracy scores over 100 data folds using different augmentation techniques.

but consistent gain in the accuracy achieved in the downstream task.

Despite all these benefits, this type of augmentation can only be applied in the case of a cyclic micro-Doppler signature, which is not the case for many problems of human activity recognition and automatic target recognition. However, further research could attempt to adapt the proposed approach for applications that involve micro-Doppler signatures of UAVs or helicopters, where the rotating blades could show a cyclic motion as do the limbs of walking humans.

Conclusions

Data augmentation is a widely-used technique for increasing the size and variety of a dataset, which can be particularly useful in data-scarce scenarios. It has the potential to improve the performance and generalization capabilities of ML models in situations where real data is difficult to collect, such as in the radar domain.

Physics-informed machine learning aims to leverage the physical knowledge and the governing laws of the problem at hand to design tailored models that are more efficient than traditional ones in domains with limited available data and well defined physical frameworks. The role of physical information in the process of augmenting micro-Doppler spectrograms has been explored in this thesis.

In the first part of this work, several generative networks that make use of the theoretical simulations available in the radar domain were devised. While capable of producing synthetic spectrograms with physics matching the input simulations, the proposed generative networks did not achieve consistent improvements over the classic GAN approach when comparing the achieved accuracy in the downstream classification task used as a benchmark. This may be due to the fidelity of the simulations, that despite being physically feasible are still too distant from real experimental spectrograms to be beneficial for the generation of synthetic micro-Doppler signatures.

In the second part of this thesis, the application of image manipulation to micro-Doppler spectrograms was discussed. More insight was provided on two augmentations from the literature and their effect on the downstream classifier. It was seen that by providing augmented samples to the network, its attention could be brought to the relevant features for discerning the two classes in data-scarce scenarios. Based on these results, a new augmentation scheme was proposed, where the input is segmented to artificially bias the focus of the network. By using this new input, a small but consistent improvement in the classification accuracy was achieved, with the benefit of a considerable reduction in the number of parameters and increased flexibility in the final class assessment.

6.1 Future Work

To expand on the work presented in this thesis, several steps could be taken.

First, the extrapolation capabilities of the generative augmentations with respect to the parameters that characterize the theoretical simulations should be assessed. This study could provide insight into how the physical parameters of the simulations are seen and extracted by the augmentation network. In addition to that, it could prove useful in the scenario where experimental data of certain situations with specific physical parameters are lacking, as the simulations for that particular case could be used to generate physically compatible synthetic spectrograms.

For this purpose, the loss and the structure of the classifiers could be adapted to include terms based on the walking person’s parameters, to increase the fidelity and matching between the parameters of the theoretical simulation inputs and the synthetic outputs.

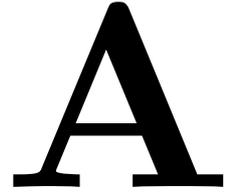
Next, the impact of the generative augmentations schemes proposed in this work could be re-assessed using a classification task where the physical parameters of the simulations are relevant for the classification label, such as for the identification of micro-Doppler signatures of UAVs. In this case, having access to new synthetic spectrograms with different physical parameters could have a different effect on the final accuracy achieved in the downstream classification task.

Finally, the use of C-GANs and AC-GANs could be considered as an alternative to having to train separate models for augmenting different classes. In particular, there is the need to verify that the architecture is capable of extracting the characterizing features of both classes and is able to produce synthetic samples accordingly.

When it comes to visual augmentation, there are a few directions in which the classification of different steps could be improved.

First, it would be of interest to research how the different approaches to combine the decision made by the classifier on different steps can be used to improve the prediction of the class. In this thesis, a simple sum was used, but different combination techniques could be adopted to make sure that the right relevance is given to each step, for example by estimating the noise and clutter present in the spectrogram of the single gait element.

Finally, in this thesis, the combination of different augmentation was not explored. In general, the gain from visual augmentations tends to saturate when increasing the number of augmented samples in the training set. Still, it would be of interest to test the boundaries of how much the proposed augmentations can push the average accuracy achieved in the classification task.



In order to assess the impact of the developed augmentations, there is a need of defining the specifics of the downstream classification task. In particular, the design and parameters of the classifier have to be addressed.

In sections A.1, A.2 and A.3 three different approaches to classification will be discussed.

A.1 Non-NN Classifier

The classical machine learning approach to classification is to design multiple features from the data that can be used to infer the target class. The designed features should be relevant to the task at hand and should also be as independent as possible, to avoid increasing the dimensionality of the problem without adding extra information.

To tackle the task of detecting armed and unarmed individuals, 13 features are extracted from the available experimental data. These comprise the mean, variance and range of the unnormalized spectrogram's magnitude and the mean and variance of the first left and right singular vectors obtained by performing singular value decomposition on the magnitude of the spectrogram [76]. In addition to that, the mean and variance of the Doppler centroid and bandwidth are computed [77]. For each time bin j , the Doppler centroid and bandwidth are defined respectively as

$$f_c(j) = \frac{\sum_i v(i)S(i, j)}{\sum_i S(i, j)}$$

and

$$B_c(j) = \sqrt{\frac{\sum_i (v(i) - f_c(j))^2 S(i, j)}{\sum_i S(i, j)}}$$

respectively, with $v(i)$ being the radial velocity associated with the i -th Doppler bin, and $S(i, j)$ the value of the spectrogram for the i -th Doppler bin and j -th time bin. These represent the centre of mass and the spread of the spectrogram's amplitude over time and could prove helpful in distinguishing signatures where the arms are moving freely and ones where the movement of the arms is constrained by the weapon.

The extracted features are standardized to have zero mean and unitary variance and are then classified using algorithms from the radar literature [77, 78]. Linear support vector machine (SVM) ([79], naive Bayes [80], linear discriminant analysis (LDA) [81] and K-nearest neighbours (KNN) [82] algorithm with K equal to 5 will be used for this analysis. The algorithms are implemented using the scikit-learn [83] open-source library. The average classification score achieved on the test set over 20 different splits of the data is reported in Table A.1. As can be seen, linearly combining the features

via LDA and linear SVM seem to give the best accuracy, which is however still quite limited.

Model	Average Accuracy Score
Linear SVM	64%
Naive Bayes	58%
LDA	63%
KNN	56%

Table A.1: Performance of different non-NN classifiers

The classification score could be improved by designing more ad-hoc features, by better selecting the available ones, or by employing NNs to implement the feature extraction and classification pipeline. Due to their potential and usage in the field of micro-Doppler classification, the option of employing NN classifiers will be explored.

A.2 Transfer Learning

Among NN approaches, TL can prove effective to tackle classification in domains that suffer from the lack of training data.

For the downstream classification task analysed in this work, six widely-used classifiers for TL purposes, trained on the ImageNet dataset [35], will be fine-tuned using spectrograms of armed and unarmed walking personnel. The six models will include the VGG16 and VGG19 models [48], Resnet50V2 and 152V2 [84], InceptionV3 [85] and Xception [86]. All the models are implemented using the TensorFlow [87] open-source library.

Fine-tuning is achieved by first freezing the weights of the convolutional layers trained on ImageNet. Subsequently, two final dense layers with ReLU activation functions and one output layer with Softmax activation are appended to the pre-trained convolutional layers and fine-tuned on the target spectrograms. Each one of the six models is tested using three different input shapes, the original size of the spectrogram (72x72). 128x128 and 224x224. The last two are obtained by zero-padding the input samples. In addition to that, three shape configurations of the dense layers are considered, 512 and 256 nodes, 1024 and 512, and 2048 and 1024. The last dense layer with the softmax activation is always kept with two nodes, equal to the number of classes in the target dataset. Table A.2 shows the parameters used for training the TL models. The learning rate determines the step size for each iteration of the optimization process while the batch size determines the number of training samples on which the gradient is computed at every iteration.

For all the classifiers and the augmentations proposed in this thesis the Adam [88] optimizer will be used to perform stochastic gradient descent and train the weights. The parameters β_1 and β_2 control the exponential decay of the moving average of the gradient and squared gradient that is used to adjust the learning rate.

To train the TL models, an early stopping strategy with a patience of 20 epochs is employed. If no improvement in validation accuracy is achieved after 20 epochs, the

training is interrupted and the best-performing model is restored.

Binary cross-entropy is used as a loss function. This is defined as

$$\mathcal{L}_{CE}(y, p) = -y \log(p) - (1 - y) \log(1 - p)$$

where y is the class label (0 for the rifle class and 1 for the walking class) and p is the softmax output related to the rifle class.

Parameter	Value
Optimizer	Adam with $\beta_1 = 0.9$ and $\beta_2 = 0.999$
Learning Rate	2e-5
Epochs	200 with early stopping
Batch size	32

Table A.2: Training parameters of the convolutional classifier

Table A.3 shows the best classification accuracy across all the scenarios for every model. All the scores lie between 60% and 65% with Resnet152V2 being the best-performing model. Even for the best-performing model, when comparing the target domain score with the top-1 accuracy of 79% [84] achieved on Image-net, there is a considerable loss in terms of test accuracy which could be due to negative transfer. This is likely a consequence of the differences between the visual images contained

Model	Best Accuracy Score
Resnet50V2	64%
Resnet152V2	65%
InceptionV3	64%
Xception	64%
VGG16	62%
VGG19	62%

Table A.3: Performance of different transfer learning models

in ImageNet and the micro-Doppler spectrograms. By freezing the weights of the convolutional layers of the pre-trained models, it is implicitly assumed that the features can be extracted from the source and the target data alike. The poor classification performance, however, shows that it may not be the case, justifying the possibility of designing and training a classifier directly on the target domain.

A.3 Convolutional classifier

As a third way of tackling the downstream classification task, an ad-hoc convolutional NN classifier is designed. This approach has the benefit of automatically performing feature extraction without having to rely on data from different domains. There is the risk, however, that due to the scarce amount of available spectrograms, the classifier

may overfit the training data and show worsened accuracy when classifying unseen data.

Figure A.1 shows the structure of the classifier. In the proposed architecture, every convolutional and dense layer is followed by a ReLU activation. In addition, the dense layers also employ dropout [89] of 0.1 to prevent overfitting.

Table A.4 shows the parameters used for training the convolutional classifier. Cross-entropy is chosen as the loss function for this architecture.

The classifier is implemented using the TensorFlow [87] open-source library.

To assess the classifier, the same scheme utilised in the experimental work of this thesis is adopted: 20% of the data is reserved for training purposes, while of the remaining spectrograms 20% is used for validation and the rest for end-to-end training. To assess the robustness of the classifier in data-scarce scenarios, 10% of the data is progressively removed and new training is performed. At each data point, the downstream accuracy over 10 subject-wise splits of the data is averaged to get the final result. Note that the structure and training parameters are kept the same for every amount of available training data.

Figure A.2 shows the final average accuracy and standard deviation for different amounts of training data. When using 80% of the data for training and validation the classifier achieves an average accuracy of 80%, which surpasses the performance of the feature-based and TL approaches. Even when removing part of the training set, the convolutional classifier still outperforms other classification methods when using only 30% of the available data.

This is due to the ad-hoc classifier being able to automatically extract the relevant feature that can separate the two classes. While the proposed hand-crafted features and the ones inferred from the visual domain can be used only to limitedly separate the data, the convolutional classifier is able to properly identify the characterising aspects of the walking and the rifle class.

For this reason, this classifier will be chosen as the one used for assessing the impact of the proposed augmentations in the downstream classification task.

Parameter	Value
Optimizer	Adam with $\beta_1 = 0.9$ and $\beta_2 = 0.999$
Learning Rate	2e-5
Epochs	60
Batch size	32

Table A.4: Training parameters of the convolutional classifier

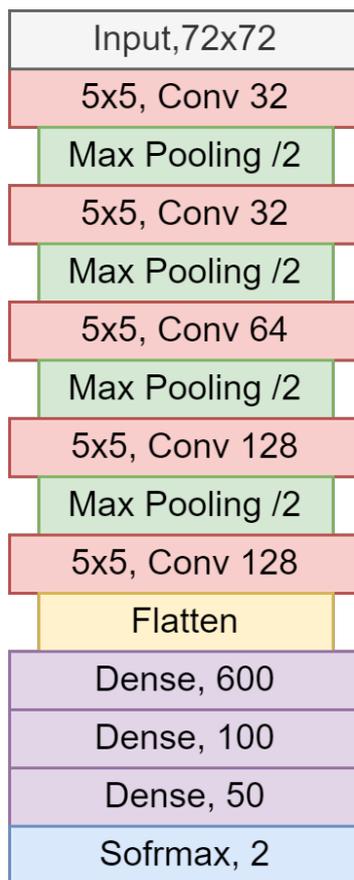


Figure A.1: Structure of the convolutional classifier

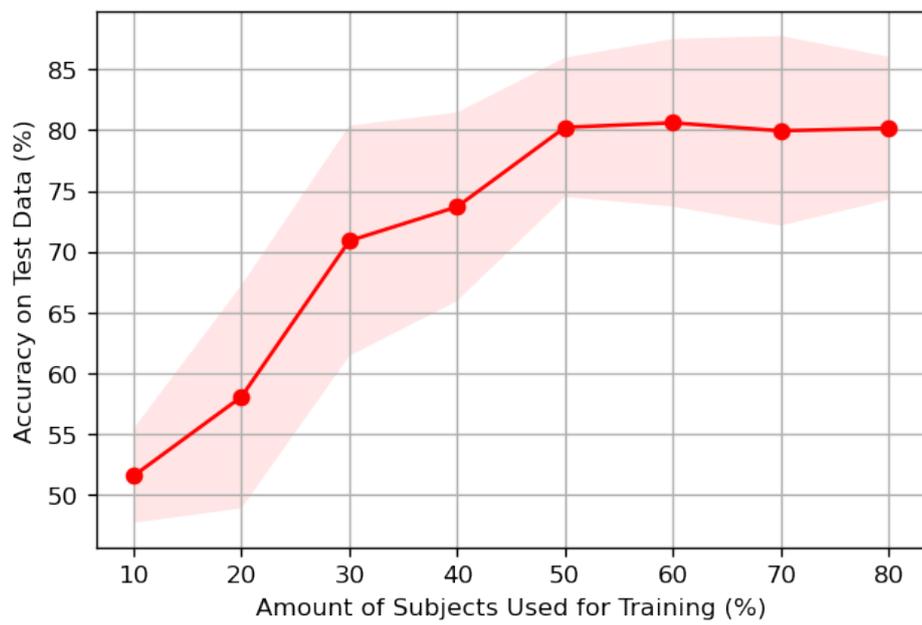


Figure A.2: Mean accuracy and standard deviation of the convolutional classifier over 10 data splits for different levels of data scarcity

B

Acquiring the Dataset

B.1 Radar Parameters

This section describes the parameters of the radar used to acquire the experimental data in [65]. The measurements were performed using the AMBER radar system, which is a frequency-modulated continuous-wave radar that operates in the X band. For this set of experiments, the carrier was set to 9.8 GHz.

The waveform used was a frequency chirp with a sweep repetition frequency of 2.5 kHz over a bandwidth of 100 MHz. To compute the radar spectrograms, a Blackman window of 0.1 seconds with an overlap of 80% was applied to the radar sequence.

B.2 Generating the simulations

This section will cover the In order to generate simulations that match the physics of the problem there is a need to understand what are the physical parameters that are required to uniquely determine a simulation.

[42] defines the relations between the different parameters of the simulation, normalized by the height of the thigh H_t . As the model is intended to be representative of the average walking human being, this normalization ensures a framework that can be easily scaled on the height of the subject, as H_t is linearly related to the subject's height [42].

In this view, the normalized velocity is defined as

$$v_r = \frac{v}{H_t} \quad (\text{B.1})$$

where v is the average velocity of the motion. v_r is empirically related to the relative length of a cycle, as

$$L_{r,c} = 1.346\sqrt{v_r} \quad (\text{B.2})$$

The absolute temporal duration of a gait cycle D_c can then be found by dividing the relative spatial length of a cycle by the relative velocity as

$$D_c = \frac{L_{r,c}}{v_r}. \quad (\text{B.3})$$

Combining B.2 and B.3 yields the relationship between the relative velocity and the cycle duration

$$v_r = \left(\frac{1.346}{D_c} \right)^2 \quad (\text{B.4})$$

If the average walking speed is known, then H_t and all the other parameters can be computed with the only knowledge of the gait cycle duration D_c . Thus, the two parameters that need to be extracted from the measurements to have matching simulations are the cycle duration D_c and the average walking speed v .

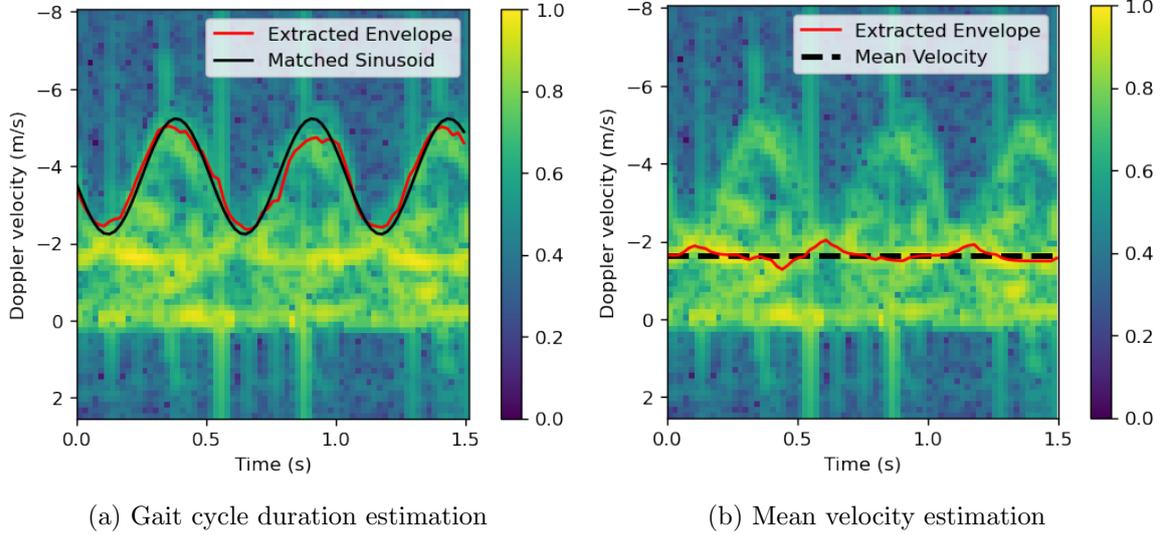


Figure B.1: Examples of the estimation process of the characterising parameters of the micro-Doppler simulation

The two parameters can be extracted directly by analysing the real spectrograms. To obtain the gait cycle duration D_c , a threshold is applied to the normalized spectrogram to isolate the Doppler response coming from the subject from the background noise. Secondly, the envelope of the response is extracted, and a sinusoid is fitted to it, as shown in figure B.1b. The gait cycle duration is then assumed to be equal to twice the period of the sinusoidal envelope.

To estimate the average walking speed, the average velocity of the torso across time is considered. First, the envelope of the return of the torso is extracted, which is assumed to be the strongest backscatter, around -2 m/s. The average velocity is then estimated as the temporal average over the frame window, as shown in figure B.1b.

Once the mean velocity v and the gait cycle D_c are known, using B.1 and B.4, the characterising parameters v_r and H_t can be obtained to produce the matched theoretical simulation.

Generative Models

This chapter will provide additional information and analysis on the generative models presented in chapter 4. In particular, more details on the structure and the training parameters will be given. In addition to that, plots showing the decrease in visual quality when removing training data will be provided.

All the proposed generative models have been implemented using the TensorFlow [87] open-source library.

C.1 Convolutional Mapping

Figure C.1 shows the structure of the convolutional mapping. Which was obtained by stacking successive convolutional layers with ReLU activations. Over different experiments, the depth of 22 layers was found to be achieving good visual results. A smaller amount of layers has been found to lead to worse visual results, while with more layers the mapping learns to map similar simulations to few outputs, further worsening the variability of the output.

The maximum amount of channels for the convolutional layers was fixed to 32, as wider layers left most of the extra output features not utilised.

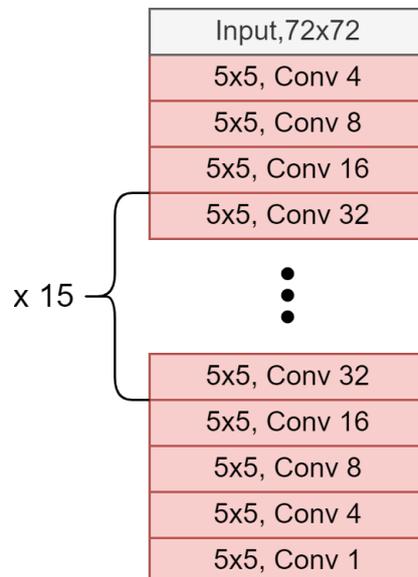


Figure C.1: Structure of the convolutional mapping

Table C.1 shows the training parameters for convolutional mapping.

Parameter	Value
Optimizer	Adam with $\beta_1 = 0.9$ and $\beta_2 = 0.999$
Learning Rate	1e-4
Epochs	1500
Batch size	32

Table C.1: Training parameters of the convolutional mapping

Figure C.2 shows visual examples of the generated spectrograms when the size of the training set decreases.

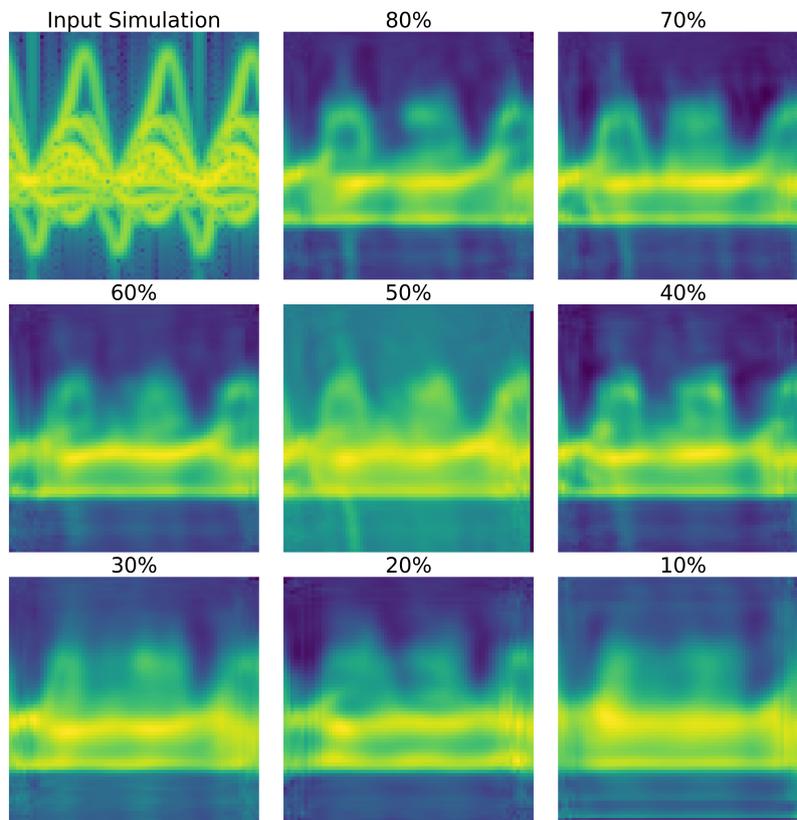


Figure C.2: Synthetic spectrograms generated with convolutional mapping when trained with progressively decreasing amount of samples

C.2 GAN

Figure C.3 shows an overview of the structure of the GAN network, divided into generator and discriminator. Each convolutional layer, unless differently specified by the activation layers coloured in blue, employs a leaky ReLU activation function with slope set to 0.2 for negative values.

In addition, every layer of the generator apart from the last one also uses batch normalization [90] with a momentum of 0.8, to improve the training process as suggested in [52]. A dropout with a factor of 0.15 is also added after every layer of the generator. For the discriminator, only a dropout of 0.25 is utilised after each layer.

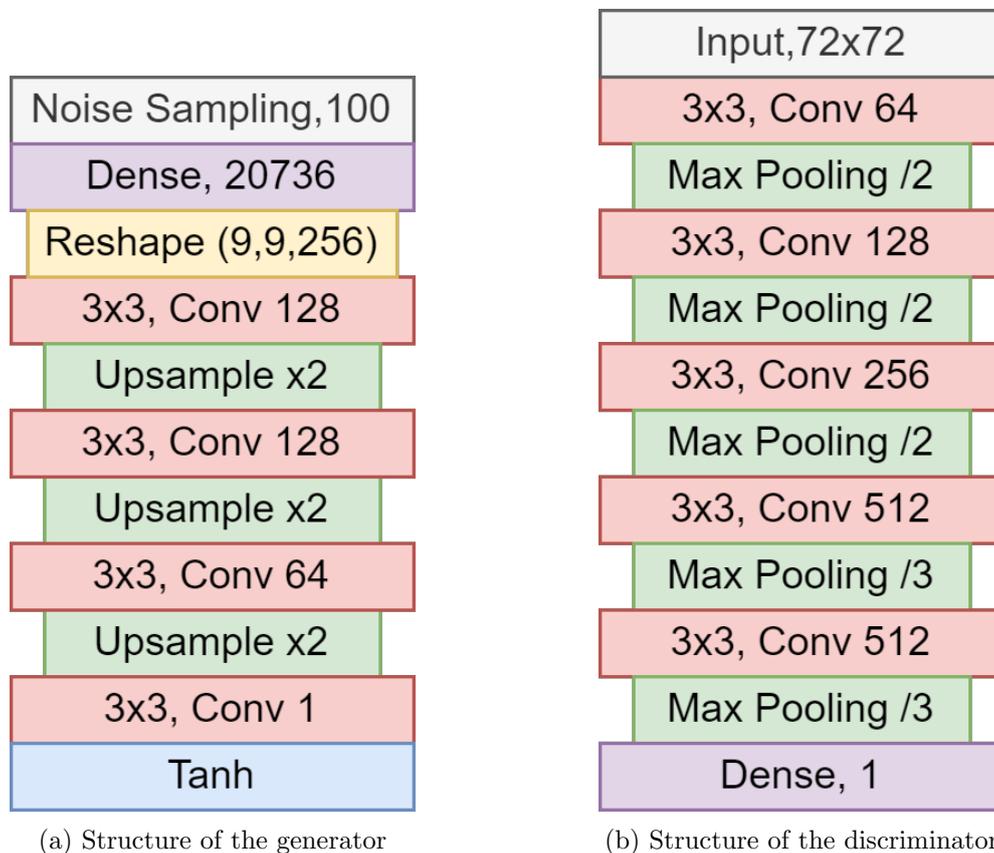


Figure C.3: Structure utilised for the proposed GAN approach

Table C.2 shows the training for both the generator and discriminator of the GAN network.

Parameter	Value
Optimizer	Adam with $\beta_1 = 0.5$ and $\beta_2 = 0.9$
Learning Rate	1e-4
Epochs	1000
Batch size	32

Table C.2: Training parameters of the GAN generator and discriminator

Figure C.4 shows visual examples of the generated spectrograms when the size of the training set decreases.

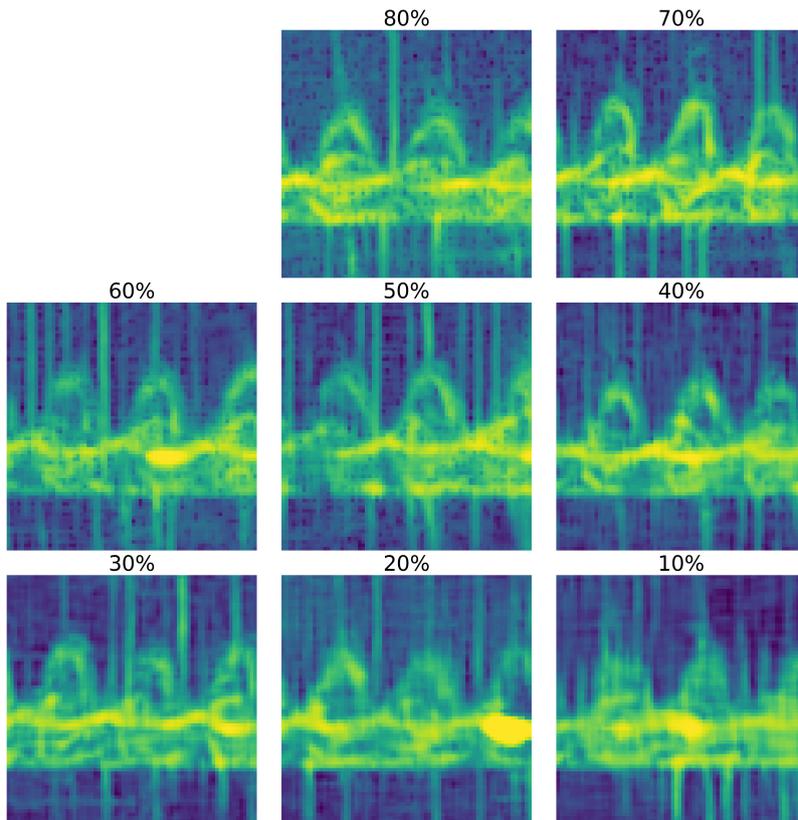


Figure C.4: Synthetic spectrograms generated with GAN network when trained with progressively decreasing amount of samples

C.3 GAN from simulations

When generating synthetic spectrograms using the adversarial training scheme and simulation as an input, the architecture is kept similar to the regular GAN approach. In particular, the same discriminator shown in figure C.3b is used for this approach. The generator is adapted to process the input simulation and its structure is shown in figure C.5. Again each convolutional layer is followed by a leaky ReLU with slope 0.2 for negative values. Batch normalization with a momentum of 0.8 and dropout with a rate of 0.15 is used after every convolutional layer except for the last.

Table C.3 shows the training parameters for the proposed architecture.

Parameter	Value
Optimizer	Adam with $\beta_1 = 0.5$ and $\beta_2 = 0.999$
Learning Rate	8e-5
Epochs	600
Batch size	32

Table C.3: Training parameters of the convolutional mapping

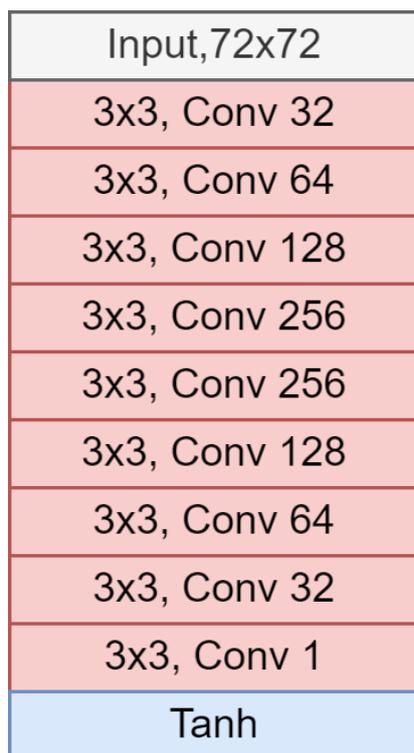


Figure C.5: Structure of the generator used in the GAN from simulations approach

Figure C.6 shows visual examples of the generated spectrograms when the size of the training set decreases.

C.4 GAN from simulations plus noise

For the GAN from simulations plus noise, the same structure and parameters as the noiseless case are used. The only difference between the two approaches stands in the fact that during the training and testing, the input is overlaid with white gaussian noise with a variance of 0.1.

Figure C.7 shows visual examples of the generated spectrograms when the size of the training set decreases.

C.5 Multi-branch GAN

In the multi-branch GAN approach, both the branch processing the noise and the branch processing the input simulation have to be designed. For the noise processing, the same structure used for the regular GAN shown in C.3a is adopted. For the simulation processing, a new architecture illustrated in figure C.8 is employed.

Note that this structure is different from the one used to process the simulations in the previous adversarial approaches involving simulations. As this new network

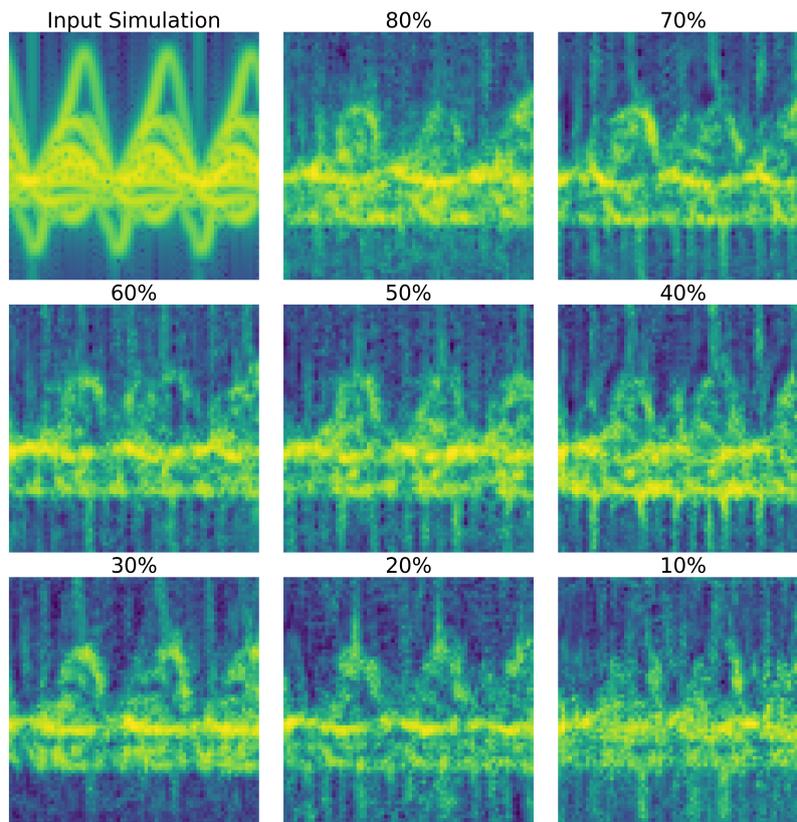


Figure C.6: Synthetic spectrograms generated with GAN with simulations as input when trained with progressively decreasing amount of samples

includes a branch responsible for processing a noise input, using a shallower structure to process the simulation yielded better visual results.

Figure C.9 shows visual examples of the generated spectrograms when the size of the training set decreases.

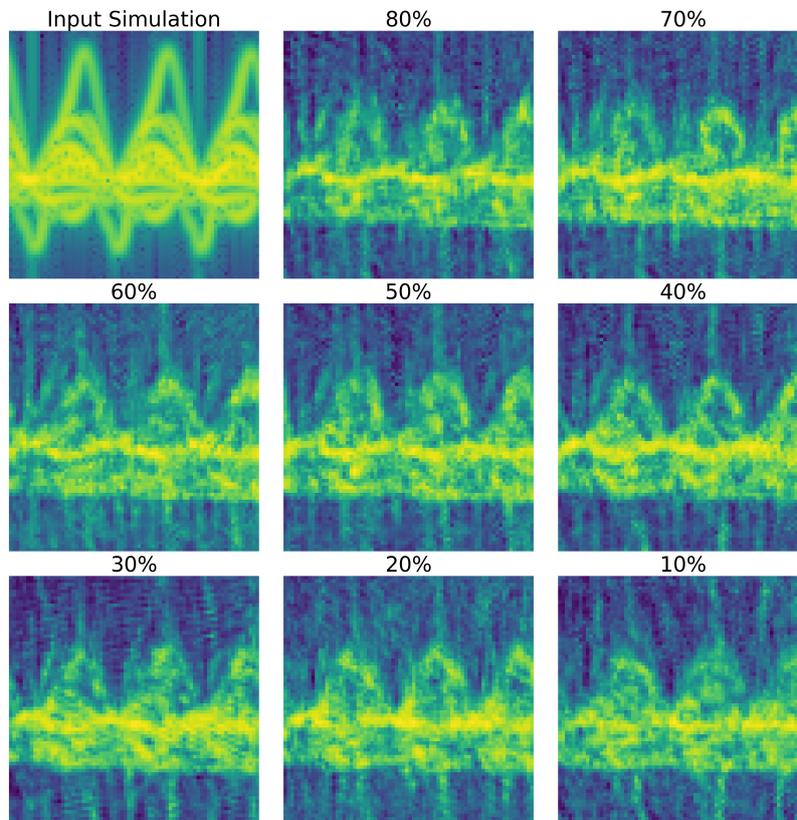


Figure C.7: Synthetic spectrograms generated with GAN with simulations plus noise as input when trained with progressively decreasing amount of samples

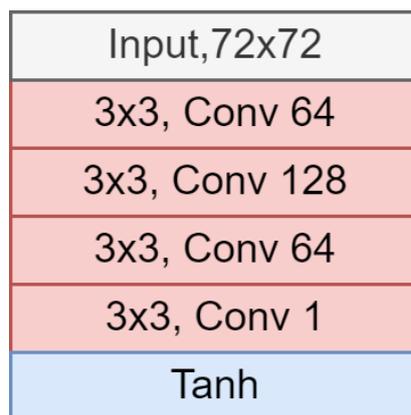


Figure C.8: Structure of the generator for the simulation branch in the multi-branch GAN approach

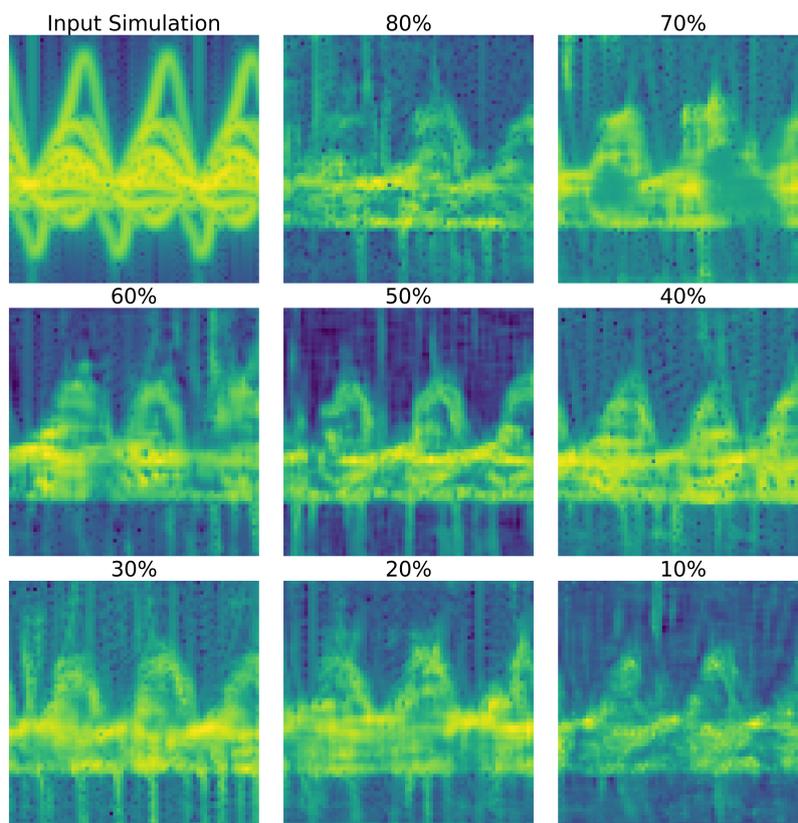


Figure C.9: Synthetic spectrograms generated with GAN with simulations plus noise as input when trained with progressively decreasing amount of samples

Bibliography

- [1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–44, 05 2015.
- [2] J. Chai, H. Zeng, A. Li, and E. W. Ngai, “Deep learning in computer vision: A critical review of emerging techniques and application scenarios,” *Machine Learning with Applications*, vol. 6, p. 100134, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666827021000670>
- [3] E. Nishani and B. Çiço, “Computer vision approaches based on deep learning and neural networks: Deep neural networks for video analysis of human pose estimation,” in *2017 6th Mediterranean Conference on Embedded Computing (MECO)*, 2017, pp. 1–4.
- [4] N. Shahid, T. Rappon, and W. Berta, “Applications of artificial neural networks in health care organizational decision-making: A scoping review,” *PLOS ONE*, vol. 14, p. e0212356, 02 2019.
- [5] S. A. Fahad and A. E. Yahya, “Inflectional review of deep learning on natural language processing,” in *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, 2018, pp. 1–4.
- [6] Z. Geng, H. Yan, J. Zhang, and D. Zhu, “Deep-learning for radar: A survey,” *IEEE Access*, vol. 9, pp. 141 800–141 818, 2021.
- [7] W. Zhang, J. Hu, Z. Wei, H. Ma, X. Yu, and H. Li, “Constant modulus waveform design for mimo radar transmit beampattern with residual network,” *Signal Processing*, vol. 177, p. 107735, 08 2020.
- [8] A. Baietto, J. Boubin, P. Farr, T. J. Bihl, A. M. Jones, and C. Stewart, “Lean neural networks for autonomous radar waveform design,” *Sensors*, vol. 22, no. 4, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/4/1317>
- [9] K. Zhong, W. Zhang, Q. Zhang, J. Hu, P. Wang, X. Yu, and Q. Zhou, “Mimo radar waveform design via deep learning,” in *2021 IEEE Radar Conference (RadarConf21)*, 2021, pp. 1–5.
- [10] M. Shaghghi and R. S. Adve, “Machine learning based cognitive radar resource management,” in *2018 IEEE Radar Conference (RadarConf18)*, 2018, pp. 1433–1438.
- [11] S. Durst and S. Bruggenwirth, “Quality of service based radar resource management using deep reinforcement learning,” in *2021 IEEE Radar Conference (RadarConf21)*. IEEE, may 2021. [Online]. Available: <https://doi.org/10.1109%2Fradarconf2147009.2021.9455234>

- [12] I. M. Gorovyi and D. S. Sharapov, “Comparative analysis of convolutional neural networks and support vector machines for automatic target recognition,” in *2017 IEEE Microwaves, Radar and Remote Sensing Symposium (MRRS)*, 2017, pp. 63–66.
- [13] Y. Sun, S. Abeywickrama, L. Jayasinghe, C. Yuen, J. Chen, and M. Zhang, “Micro-doppler signature-based detection, classification, and localization of small uav with long short-term memory neural network,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, pp. 1–16, 10 2020.
- [14] G. Mendis, T. Randeny, J. Wei, and A. Madanayake, “Deep learning based doppler radar for micro uas detection and classification,” *MILCOM 2016 - 2016 IEEE Military Communications Conference*, pp. 924–929, 2016.
- [15] Y. Zhang, X. Sun, H. Sun, Z. Zhang, W. Diao, and K. Fu, “High resolution sar image classification with deeper convolutional neural network,” in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 2374–2377.
- [16] L. Gu, F. He, and S. Yang, “Crop classification based on deep learning in northeast china using sar and optical imagery,” in *2019 SAR in Big Data Era (BIGSAR-DATA)*, 2019, pp. 1–4.
- [17] Y. Ren, H. Xu, B. Liu, and X. Li, “Sea ice and open water classification of sar images using a deep learning model,” in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 3051–3054.
- [18] S. Z. Gurbuz and M. G. Amin, “Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring,” *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16–28, 2019.
- [19] O. Karabayır, O. M. Yücedağ, M. Z. Kartal, and H. A. Serim, “Convolutional neural networks-based ship target recognition using high resolution range profiles,” *2017 18th International Radar Symposium (IRS)*, pp. 1–9, 2017.
- [20] Z. Fu, S. Li, X. Li, B. Dan, and X. Wang, “A performance analysis of neural network models in hrrp target recognition,” in *2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, 2019, pp. 1–4.
- [21] B. Ding and P. Chen, “Hrrp feature extraction and recognition method of radar ground target using convolutional neural network,” in *2019 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, 2019, pp. 0658–0661.
- [22] C. Shorten and T. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, 07 2019.
- [23] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” 2014. [Online]. Available: <https://arxiv.org/abs/1406.2661>

- [24] D. P. Kingma and M. Welling, “An introduction to variational autoencoders,” *Foundations and Trends® in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019. [Online]. Available: <https://doi.org/10.1561%2F22000000056>
- [25] B. Erol, S. Z. Gurbuz, and M. G. Amin, “Motion classification using kinematically sifted acgan-synthesized radar micro-doppler signatures,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 4, pp. 3197–3213, 2020.
- [26] A. Huizing, M. Heiligers, B. Dekker, J. de Wit, L. Cifola, and R. Harmanny, “Deep learning for classification of mini-uavs using micro-doppler spectrograms in cognitive radar,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 34, no. 11, pp. 46–56, 2019.
- [27] “Physics-informed machine learning,” *Nature Reviews Physics 2021 3:6*, vol. 3, pp. 422–440, 5 2021. [Online]. Available: <https://www.nature.com/articles/s42254-021-00314-5>
- [28] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [29] X. Ying, “An overview of overfitting and its solutions,” *Journal of Physics: Conference Series*, vol. 1168, p. 022022, 02 2019.
- [30] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [31] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *Journal of Big data*, vol. 3, no. 1, pp. 1–40, 2016.
- [32] M. Al Mufti, E. Al Hadhrami, B. Taha, and N. Werghi, “Sar automatic target recognition using transfer learning approach,” in *2018 International Conference on Intelligent Autonomous Systems (ICoIAS)*, 2018, pp. 1–4.
- [33] E. A. Hadhrami, M. A. Mufti, B. Taha, and N. Werghi, “Transfer learning with convolutional neural networks for moving target classification with micro-doppler radar spectrograms,” in *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 2018, pp. 148–154.
- [34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
- [35] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
- [36] W. Zhang, L. Deng, L. Zhang, and D. Wu, “A survey on negative transfer,” 2020. [Online]. Available: <https://arxiv.org/abs/2009.00909>

- [37] Z. Wang, Z. Dai, B. Póczos, and J. Carbonell, “Characterizing and avoiding negative transfer,” 2018. [Online]. Available: <https://arxiv.org/abs/1811.09751>
- [38] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, “Image data augmentation for deep learning: A survey,” 2022. [Online]. Available: <https://arxiv.org/abs/2204.08610>
- [39] T. T. Team, “Flowers,” jan 2019. [Online]. Available: http://download.tensorflow.org/example_images/flower_photos.tgz
- [40] L. Deng, “The mnist database of handwritten digit images for machine learning research [best of the web],” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.
- [41] N. Kern and C. Waldschmidt, “Data augmentation in time and doppler frequency domain for radar-based gesture recognition,” in *2021 18th European Radar Conference (EuRAD)*, 2022, pp. 33–36.
- [42] R. Boulic, N. Thalmann, and D. Thalmann, “A global human walking model with real-time kinematic personification,” *The Visual Computer*, vol. 6, 05 2002.
- [43] D. She, X. Lou, and W. Ye, “Radarspecaugment: A simple data augmentation method for radar-based human activity recognition,” *IEEE Sensors Letters*, vol. 5, no. 4, pp. 1–4, 2021.
- [44] E. Alhadhrami, M. Al Mufti, B. Taha, and N. Werghi, “Ground moving radar targets classification based on spectrogram images using convolutional neural networks,” 06 2018, pp. 1–9.
- [45] L. A. Gatys, A. S. Ecker, and M. Bethge, “A neural algorithm of artistic style,” 2015. [Online]. Available: <https://arxiv.org/abs/1508.06576>
- [46] X. Zheng, T. Chalasani, K. Ghosal, S. Lutz, and A. Smolic, “STaDA: Style transfer as data augmentation,” in *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications, 2019. [Online]. Available: <https://doi.org/10.5220%2F0007353401070114>
- [47] P. T. Jackson, A. Atapour-Abarghouei, S. Bonner, T. Breckon, and B. Obara, “Style augmentation: Data augmentation via style randomization,” 2018. [Online]. Available: <https://arxiv.org/abs/1809.05375>
- [48] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [49] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R. Adve, and K. Chetty, “Neural style transfer enhanced training support for human activity recognition,” 2021. [Online]. Available: <https://arxiv.org/abs/2107.12821>

- [50] H. Huang, P. S. Yu, and C. Wang, “An introduction to image synthesis with generative adversarial nets,” 2018. [Online]. Available: <https://arxiv.org/abs/1803.04469>
- [51] M. Arjovsky and L. Bottou, “Towards principled methods for training generative adversarial networks,” 2017. [Online]. Available: <https://arxiv.org/abs/1701.04862>
- [52] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein gan,” 2017. [Online]. Available: <https://arxiv.org/abs/1701.07875>
- [53] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, “Improved training of wasserstein gans,” 2017. [Online]. Available: <https://arxiv.org/abs/1704.00028>
- [54] B. Erol, S. Z. Gurbuz, and M. G. Amin, “Gan-based synthetic radar micro-doppler augmentations for improved human activity recognition,” in *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1–5.
- [55] C. Wu and W. Ye, “Generative adversarial network for radar-based human activities classification with low training data support,” in *2021 IEEE 4th International Conference on Electronic Information and Communication Technology (ICEICT)*, 2021, pp. 415–419.
- [56] I. Alnujaim and Y. Kim, “Augmentation of doppler radar data using generative adversarial network for human motion analysis,” *Healthcare Informatics Research*, vol. 25, p. 344, 10 2019.
- [57] C. Tang, S. Vishwakarma, W. Li, R. Adve, S. Julier, and K. Chetty, “Augmenting experimental data with simulations to improve activity classification in healthcare monitoring,” in *2021 IEEE Radar Conference (RadarConf21)*, 2021, pp. 1–6.
- [58] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R. Adve, and K. Chetty, “Simhumalator: An open source wifi based passive radar human simulator for activity recognition,” 2021. [Online]. Available: <https://arxiv.org/abs/2103.01677>
- [59] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, “Integrating scientific knowledge with machine learning for engineering and environmental systems,” 2020. [Online]. Available: <https://arxiv.org/abs/2003.04919>
- [60] L. von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, M. Walczak, J. Pfrommer, A. Pick, R. Ramamurthy, J. Garcke, C. Bauckhage, and J. Schuecker, “Informed machine learning - a taxonomy and survey of integrating prior knowledge into learning systems,” *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2021. [Online]. Available: <https://doi.org/10.1109%2Ftkde.2021.3079836>
- [61] Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli, and A. Anandkumar, “Physics-informed neural operator for learning partial differential equations,” 2021. [Online]. Available: <https://arxiv.org/abs/2111.03794>

- [62] L. Sun, H. Gao, S. Pan, and J.-X. Wang, “Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data,” *Computer Methods in Applied Mechanics and Engineering*, vol. 361, p. 112732, apr 2020. [Online]. Available: <https://doi.org/10.1016%2Fj.cma.2019.112732>
- [63] Y. Yang, X. Zhang, Q. Guan, and Y. Lin, “Making invisible visible: Data-driven seismic inversion with spatio-temporally constrained data augmentation,” *IEEE Transactions on Geoscience and Remote Sensing*, 6 2021. [Online]. Available: <http://arxiv.org/abs/2106.11892http://dx.doi.org/10.1109/TGRS.2022.3144636>
- [64] M. M. Rahman, S. Z. Gurbuz, and M. G. Amin, “Physics-aware design of multi-branch gan for human rf micro-doppler signature synthesis,” *IEEE National Radar Conference - Proceedings*, vol. 2021-May, 5 2021.
- [65] R. J. de Jong, J. J. de Wit, and F. Uysal, “Classification of human activity using radar and video multimodal learning,” *IET Radar, Sonar & Navigation*, vol. 15, no. 8, pp. 902–914, 2021. [Online]. Available: <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/rsn2.12064>
- [66] Q. Chen, Y. Liu, F. Fioranelli, M. Ritchie, B. Tan, and K. Chetty, “Dopnet: A deep convolutional neural network to recognize armed and unarmed human targets,” *IEEE Sensors Journal*, vol. 19, no. 11, pp. 4160–4172, 2019.
- [67] V. Chen, F. Li, S.-S. Ho, and H. Wechsler, “Micro-doppler effect in radar: phenomenon, model, and simulation study,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 2–21, 2006.
- [68] —, “Micro-doppler effect in radar: phenomenon, model, and simulation study,” pp. 2–21, 2006.
- [69] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [70] D. Bank, N. Koenigstein, and R. Giryes, “Autoencoders,” 2021.
- [71] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” 2016. [Online]. Available: <https://arxiv.org/abs/1611.07004>
- [72] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” 2017. [Online]. Available: <https://arxiv.org/abs/1703.10593>
- [73] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” 2014. [Online]. Available: <https://arxiv.org/abs/1411.1784>
- [74] A. Odena, C. Olah, and J. Shlens, “Conditional image synthesis with auxiliary classifier gans,” 2016. [Online]. Available: <https://arxiv.org/abs/1610.09585>

- [75] K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep inside convolutional networks: Visualising image classification models and saliency maps,” 2013. [Online]. Available: <https://arxiv.org/abs/1312.6034>
- [76] J. J. M. de Wit, R. I. A. Harmanny, and P. Molchanov, “Radar micro-doppler feature extraction using the singular value decomposition,” in *2014 International Radar Conference*, 2014, pp. 1–6.
- [77] F. Fioranelli, M. Ritchie, and H. Griffiths, “Centroid features for classification of armed/unarmed multiple personnel using multistatic human micro-doppler,” *IET Radar, Sonar & Navigation*, vol. 10, no. 9, pp. 1702–1710, 2016. [Online]. Available: <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-rsn.2015.0493>
- [78] M. G. Amin, Z. Zeng, and T. Shan, “Hand gesture recognition based on radar micro-doppler signature envelopes,” in *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1–6.
- [79] V. Vapnik, *The Support Vector Method of Function Estimation*. Boston, MA: Springer US, 1998, pp. 55–85. [Online]. Available: https://doi.org/10.1007/978-1-4615-5703-6_3
- [80] T. Vikramkumar, Vijaykumar, “Bayes and naive bayes classifier,” 2014. [Online]. Available: <https://arxiv.org/abs/1404.0933>
- [81] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, “Linear discriminant analysis: A detailed tutorial,” *Ai Communications*, vol. 30, pp. 169–190,, 05 2017.
- [82] P. Cunningham and S. Delany, “k-nearest neighbour classifiers,” *Mult Classif Syst*, vol. 54, 04 2007.
- [83] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [84] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” 2016. [Online]. Available: <https://arxiv.org/abs/1603.05027>
- [85] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” 2014. [Online]. Available: <https://arxiv.org/abs/1409.4842>
- [86] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” 2016. [Online]. Available: <https://arxiv.org/abs/1610.02357>
- [87] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens,

- B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, software available from tensorflow.org. [Online]. Available: <https://www.tensorflow.org/>
- [88] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [89] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. [Online]. Available: <http://jmlr.org/papers/v15/srivastava14a.html>
- [90] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” 2015. [Online]. Available: <https://arxiv.org/abs/1502.03167>