NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY POLITEHNICA BUCHAREST

Doctoral School of Electronics, Telecommunications and Information Technology

Decision No. 47 **from** 20-05-2024

Ph.D. THESIS
SUMMARY

Nicolae-Cătălin RISTEA

ÎNVĂŢARE AUTOMATĂ BAZATĂ PE MODELE FIZICE ÎN TELEDETECŢIE

PHYSICS-AWARE MACHINE LEARNING IN REMOTE SENSING

THESIS COMMITTEE

Prof. Dr. Ing. Mihai CIUC NUST Politehnica Bucharest	President
Prof. Dr. Ing. Andrei ANGHEL NUST Politehnica Bucharest	PhD Supervisor
Prof. Dr. Ing. Cosmin ANCUŢI University Politehnica Timisoara	Referee
Dr. Bertrand CHAPRON Ifremer	Referee
Prof. Dr. Ing. Mihai DATCU NUST Politehnica Bucharest	Referee

BUCHAREST 2024

Table of contents

Chapter 1

Introduction

The field of signal processing has a rich history, with its roots extending back several decades. Signal processing involves the manipulation, analysis, and interpretation of signals, which are representations of data in various forms. The historical development of signal processing has witnessed significant advancements in both theory and practical applications. Signal processing techniques have played a pivotal role in extracting meaningful information from raw data, enhancing the quality of signals, and enabling efficient communication systems. The filed have evolved even further with the development of big data sets across diverse industries, employing machine learning (ML) and deep learning (DL) algorithms. Those new techniques have emerged as transformative paradigms within the signal processing landscape.

1.1 Presentation of the field of the doctoral thesis

Automotive Radar Interference. Addressing autonomous driving and road safety stands as a crucial topic for mitigating the frequency of traffic accidents and fatalities. Automotive companies are actively addressing this concern by advancing autonomous vehicle technologies. A prevalent approach involves employing radar sensors to meticulously scan the surrounding environment. Within the automotive industry, the predominant choice for radar sensors is the deployment of frequency modulated continuous wave (FMCW) or chirp sequence (CS) radars, which transmit sequences of linear chirp signals. The signals transmitted and received by such sensors provide the means to estimate the distance and the velocity of nearby targets (e.g., vehicles, pedestrians or other obstacles). However, the growing adoption of radar sensors [17] increases the probability of interference among sensors from different vehicles, generating corrupted and unusable signals. Radio frequency interference could significantly elevate the noise floor, reaching a point where discerning potential targets becomes challenging or even impossible. In such instances, the interference may obscure targets entirely, thereby diminishing the

sensitivity of target detection methods [3]. To identify these obscured targets, it is imperative to mitigate the radar interference.

Earth observation. Earth observation involves the comprehensive integration of the physical, chemical, and biological systems of our planet. This is achieved through remote sensing (RS) technologies employed in earth surveillance techniques. The process involves the collection, analysis, and presentation of data to enhance our understanding of Earth's diverse systems [13]. Considering that the ocean accounts for about 71% of Earth's surface, the ocean observation has increasingly drawn the attention of the research community over the last decades. Humans had minimal ocean observations before 1978, when Seasat, the first Earth-orbiting satellite designed for remote sensing of Earth's oceans was launched [43]. Although Seasat only operated for about 100 days, the mission acquired more data about the ocean than all previous sensors combined. This event stimulated the fast development of ocean-satellite, leading to a growing number of satellites carrying different sensors (e.g., microwave, visible, infrared) being launched to improve our understanding about the ocean.

In addition to ocean observation, another crucial aspect has gained significant attention, the diminishing sea ice, particularly in the Arctic. This effect is a prominent consequence of global climate change. Consequently, the cover and concentration of sea ice play pivotal roles in both climate change studies and navigation within polar regions. Addressing the logistical demands of the transportation industry in these areas requires intricate, high-resolution data on local Arctic marine conditions. Such information is indispensable for strategic operational planning, route determination, and fostering the sustainable development of Northern regions [6]. Once again, considering the high amount of daily acquired data, the need of automated systems is of utmost importance.

1.2 Scope of the doctoral thesis

Considering the presented challenges in Earth observation and automotive radar industry, the rapid evolution of sensors, coupled with the substantial volume of available data, has resulted in processing algorithms that barely meet the current requirements. To this end, our objective is to develop robust and accurate deep learning networks, assisted by physics-aware processing techniques.

The main directions followed are:

1. To identify, develop and analyse the optimal deep neural network based framework for automotive radar interference mitigation in complex scenarios (e.g., multiple sources of interference).

2. To test the generalization capacity of the proposed radar interference mitigation algorithms on multiple sensors.

3. To identify, develop and analyse the optimal deep neural network based framework for ocean phenomena analysis and classification on SAR vignettes.

4. To identify, develop and analyse the optimal deep neural network based framework for sea ice segmentation in the Arctic areas.

5. To provide physics-based motivations and interpretation methods for the developed architectures.

1.3   Content of the doctoral thesis

According to the main research directions followed during the PhD, this thesis is structured according to the following chapters:

Chapter 2 (Theoretical Background and Concepts) introduces the main concepts that have been explored and used in the current work, such as physical-based concepts presents the FMCW solutions, SAR systems and neural networks.

Chapter 3 (State of the Art) provides an overview of the research conducted in our research directions. Our focus has been directed towards three primary domains: automotive radar interference mitigation, algorithms employed in ocean SAR imagery analysis, and methods for segmenting sea ice in Arctic regions.

Chapter 4 (Automotive Radar Interference Mitigation) succinctly introduces the field of automotive radar and the need of radar interference mitigation algorithms. The section delves into the description, concept, and results of our proposed FCN architecture. The chapter concludes by detailing and discussing the advantages and disadvantages offered by our framework.

Chapter 5 (Ocean Patterns Analysis) introduces the filed of ocean SAR imagery, specifically focusing on vignette classification and image retrieval within oceanic areas. The discussion extends to the description and concepts of developed algorithms, underscored by physical-based motivations. The chapter concludes in a detailed analysis and discussion of the key elements introduced by our framework, emphasizing its potential deployment in real-world scenarios.

Chapter 6 (Sea-Ice Segmentation for SAR Imagery) introduces the filed of sea ice segmentation in Arctic regions, using SAR data. Subsequently, it describes the development of deep neural network architectures aimed at achieving an optimal balance between speed and performance. Moreover, the IW-EW SAR data adaptation is described, along with the practical importance and the used algorithms. The chapter concludes in a comprehensive analysis and discussion of the innovative elements introduced in our work.

Chapter 7 (Conclusions) comprises several sections describing the findings derived from the outcomes and contributions of the thesis. Additionally, it includes a list of published papers, followed by potentially future research perspectives.

Chapter 2

Theoretical Background and Concepts

2.1 Physics-based concepts

2.1.1 Frequency Modulated Continuous Wave Radar

In FMCW radar solutions, the transmitted signal $s_{TX}(t)$ is a chirp sequence, whose frequency usually follows a sawtooth pattern. In the presence of mutual interferences, the receiving antenna collects a mix from two signals, the reflected signal and the interference signal. Consequently, the received signal is formally defined below.

$$s_{RX}(t) = \sum_{i=0}^{N_t-1} \underline{A}_i \cdot s_{TX}(t - \tau_i) + \sum_{l=0}^{N_{int}-1} s_{RFI,l}(t), \qquad (2.1)$$

where $\underline{A}_i = A_i \cdot e^{j\phi_i}$ is the complex amplitude, τ_i is the propagation delay of target i, N_t is the number of targets, and N_{int} is the number of interferers. The receive antenna collects the reflected signal $s_{RX}(t)$, which is further mixed with the transmitted signal and low-pass filtered, resulting in the beat signal $s_b(t)$. Upon combining the signal reflected by a point-like target with the transmitted signal, a signal with a constant frequency is obtained. Thus, mixing an uncorrelated interference with the transmitted chirp results in a baseband chirp signal.

2.1.2 Synthetic Aperture Processing

Synthetic Aperture Radar (SAR) is a modern radar technology that overcomes traditional limitations and achieve unparalleled imaging capabilities. Unlike conventional radars that rely on physically large antennas to achieve high resolution, SAR employs a method to synthesize a virtual aperture by exploiting the motion of the radar platform. By utilizing this motion and processing the reflected radar signals coherently, SAR can achieve azimuth resolutions on the order of meters or even centimeters, far surpassing the capabilities of traditional radars.
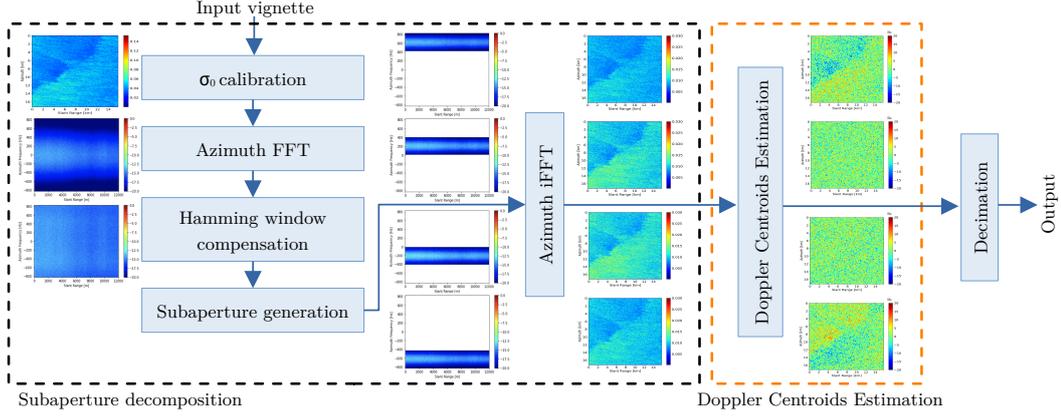
Fig. 2.1 The preprocessing subaperture decomposition pipeline. The DCE block is an additional algorithm, which can be omitted in accordance with the experiment performed.

At its core, SAR operates on the principle of coherent radar imaging, where a series of radar pulses are transmitted towards the target area while the radar platform moves along a predefined trajectory. As the radar waves interact with the land area, the waves are scattered and reflected back towards the radar sensor. By measuring the phase and amplitude of these reflected signals across the aperture, SAR creates a high-resolution image of the target scene. By exploiting the phase information contained in the radar echoes, SAR can effectively mitigate unwanted contributions, resulting in sharp and clear images even in adverse environmental conditions.

2.1.3 Subaperture Decomposition

In classical SAR systems, backscatter from irradiated targets is acquired across various positions and azimuth angles along the radar trajectory. Given a highly non-stationary scene, e.g., ocean surface, observing it from different angles can yield additional information about the illuminated area. Therefore, by decomposing the SAR image into subapertures, the scene could be artificially seen from different observation angles, capturing additional information. Each subaperture corresponds to an image formed using only a portion of the total azimuth angle, as illustrated in Figure 2.1 after the subaperture generation block.

2.1.4 Doppler Centroids Estimation

Let $X_i \in \mathbb{R}^{m \times n}$ be the ith subaperture for a SAR image, where $m, n \in \mathbb{N}$ are the azimuth and range dimensions. Let $Y_i \in \mathbb{R}^{m \times n}$ be the delayed version with 1 sample in the azimuth axis of X_i. We estimate the Doppler centroids for the ith subaperture as follows:

$$D_i = -PRF \cdot \frac{angle(Z_i)}{2\pi}, \tag{2.2}$$

where $Z_i = filt(X_i \cdot Y_i^*)$, Y_i^* is the complex conjugate of Y_i, PRF is the pulse repetition frequency, $angle()$ returns the angle of the complex input and $filt()$ is a two dimensional mean filter with $d_1 \times d_2$ kernel size. In the right part of the Figure 2.1 we illustrated the DCE result for a SAR input image.

2.2 Deep Learning Techniques

2.2.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a class of deep learning models specifically designed for processing structured grid data, such as images, audio, and time series. They have achieved remarkable success in various tasks, such as image classification [16, 29, 42, 24], object detection [7, 30, 5], and semantic segmentation [22, 28, 12]. The basic building blocks of CNNs include convolutional layers, pooling layers, and fully connected layers. Next, we briefly describe the first two blocks.

Convolutional layer. A convolutional layer applies a set of learnable filters to the input data to produce feature maps. Without loosing the generality, we will formally describe the two-dimensional scenario, the generalization being trivial. Let $X \in \mathbb{R}^2$ be the input tensor, $F \in \mathbb{R}^2$ be the filter tensor, and $Y \in \mathbb{R}^2$ be the output feature map. The convolution operation is defined as:

$$Y_{i,j} = (X * F)_{i,j} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X_{i+m,j+n} \cdot F_{m,n}, \tag{2.3}$$

where $M \in \mathbb{N}$ and $N \in \mathbb{N}$ are the dimensions of the filter F, and $*$ denotes the convolution operation.

Pooling layer. Pooling layers downsample the input feature maps, reducing the spatial dimensions while retaining the most important information, in accordance with a given mathematical operation, e.g., maximum, minimum, average. The layer is applied on each channel apart, thus only the spatial dimension of the input data is affected, while the number of channels remaining unchanged.

2.2.2 Transformer Architectures

Transformer block. In the left side of the Figure 2.2 the overall transformer architecture is illustrated. Let $X = [x_1, x_2, \ldots, x_n]$ be an input sequence, where $x_i \in \mathbb{R}^m$ is the feature vector for the ith input. The sequence is projected into tokens $T = [t_1, t_2, \ldots, t_n]$, where each token $t_i \in \mathbb{R}^d$ corresponds to the x_i input data and $d \in \mathbb{N}$ is the token embedding size. Afterwards, the positional embeddings are added for all tokens, such that each token t_i could be uniquely identified by the given positional embedding. For batch processing,
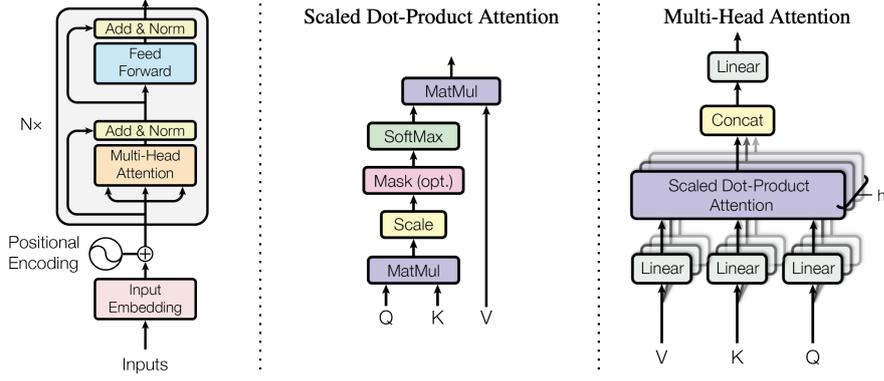
Fig. 2.2 (left) The transformer architecture [45]. (center) The scale dot-product attention. (right) The multi-head attention.

all sequences are grouped in matrices, e.g., $T \in \mathbb{R}^{n \times d}$. Next, the token sequence T is processed by the transformer block as follows:

$$Z = \text{Norm}(\text{MultiHeadAttention}(T) + T) \tag{2.4}$$

$$Y = \text{Norm}(\text{FeedForward}(Z) + Z), \tag{2.5}$$

where $Z \in \mathbb{R}^{n \times d}$, $Y \in \mathbb{R}^{n \times d}$, Norm function is a normalization layer [1] and the Feed Forward module is typically a multi layer perceptron. The block is applied $N \in \mathbb{N}$ times, outputting the deep attention-based features for the input sequence.

Scale dot-product attention. The building blocks of the scale dot-product attention are highlighted in the center of the Figure 2.2. The input consists of queries $Q \in \mathbb{R}^{n \times d_k}$, keys $K \in \mathbb{R}^{n \times d_k}$ and values $V \in \mathbb{R}^{n \times d}$. We compute the dot products of the query with all keys, divide the result with $\sqrt{d_k}$ and apply a softmax function to obtain the weights (attention matrix) of the values. The block is formally described as:

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{2.6}$$

Multi-head attention. Multi-head attention in the transformer architecture lets the model focus on different parts of the input simultaneously, improving its ability to capture complex relationships. It applies multiple linear projections called heads to input Q, K, V, computes independent attention, and concatenates outputs for the final result. This enhances feature extraction and improves performance on tasks needing long-range dependencies and contextual understanding. The block is visually described in the right part of the Figure 2.2.

Chapter 3

Automotive Radar Interference Mitigation

3.1 Introduction

The most common radar senors used in the automotive industry are frequency modulated continuous wave (FMCW) / chirp sequence (CS) radars, which transmit sequences of linear chirp signals. However, the growing adoption of radar sensors [17] increases the probability of interference among sensors from different vehicles, generating corrupted and unusable signals. In Figure 3.1, we present a range profile of a radar signal with (red) and without (blue) interference, highlighting that certain targets become absorbed by the noise floor generated by multiple interference sources. To successfully identify these targets, it is imperative to implement effective solutions for mitigating radar interference. To address this problem, researchers have proposed various techniques ranging from conventional approaches [18, 44, 14, 48, 2, 19, 27, 26] to deep learning methods [35, 37, 36, 40, 9, 25, 39].

We extend our prior work [35] by designing a novel fully convolutional network (FCN) [22] capable of retrieving both the phase and magnitude of radar beat signals and can manage multiple non-coherent RFI sources. Our network takes as input the real and imaginary parts, as well as, the magnitude of the Short-Time Fourier Transform (STFT) of the beat signal with interference. The network outputs the real part, imaginary part and magnitude of the clean range profile. While most deep learning approaches studied radar interference mitigation with a single interference source [35, 9, 26], we aim to address the RFI task under multiple interference sources. To attain this objective, we create an extensive data set that closely mimics real-world automotive scenarios involving multiple interference sources. Our training set incorporates up to three interference sources, while for the testing set, we consider scenarios with up to six interference sources. Moreover, we publish two data sets as open source for future efforts in the radar interference mitigation field.
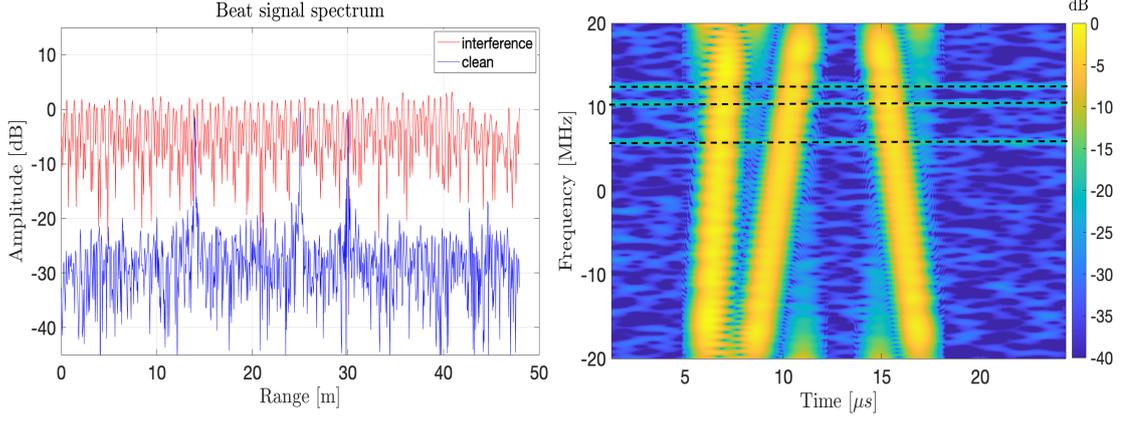
Fig. 3.1 (left) The magnitude of the range profile of an FMCW radar sensor is depicted. (right) The magnitude of the STFT corresponding to the above range profile is presented. The targets are represented by thin horizontal lines, while the interference sources manifest as thicker, more prominent diagonal lines.

3.2 Proposed Methods

3.2.1 Radar Signal Model

In FMCW radar systems, the transmitted signal $s_{TX}(t)$ is a chirp sequence. In the presence of mutual interferences, the receiving antenna collects a mix from two signals, the reflected signal and the interference signal. Following the 2.1.1, the analytical beat signal $s_b(t)$ in the presence of interferences is expressed as:

$$s_b(t) = \left\{ \sum_{i=0}^{N_t-1} \underline{A}_i \cdot \exp(j2\pi k\tau_i t) + \sum_{l=0}^{N_{int}-1} \underline{A}_{RFI,l} \cdot \exp\left[j\pi(k-k_{RFI,l})(t-t_{RFI,l})^2\right] \cdot \right.$$
$$\left. \cdot p\left(\frac{t-t_{RFI,l}}{T_{AAF,l}}\right) \right\} \cdot p\left(\frac{t-\frac{T}{2}}{T}\right), \tag{3.1}$$

where $\underline{A}_{RFI,l}$ is the complex amplitude of interference signal l and $p(t)$ is the window function. As a result, $s_b(t)$ comprises a combination of complex exponentials representing the targets and interfering signals in the form of baseband chirps.

3.2.2 Data Preprocessing

As depicted in Figure 3.1, the computation of the discrete STFT is essential to separate the targets from the interference sources. The equation below outlines the transformation of a time-domain signal into an image-alike representation using the discrete STFT, using a *hamming* window. Additionally, to bring the input data approximately within the range of $[-1,1]$, we scale the STFT by dividing it with $\alpha = 40$, a value obtained statistically across the entire training set. To obtain clean range profiles from the STFT

9
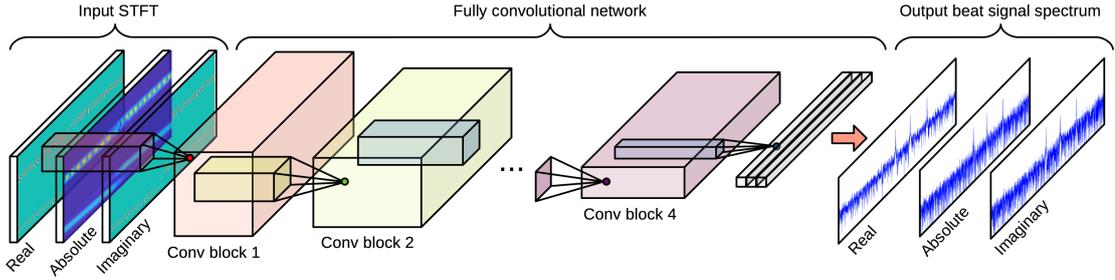
Fig. 3.2 The structure of our FCN model involves the processing of the input STFT through a sequence of four convolutional blocks (comprising convolutional and pooling layers). This processing continues until the vertical dimension is reduced to 1, obtaining the spectrum of the beat signal.

of the beat signal, we conduct a Fast Fourier Transform (FFT) on our time-domain labels to obtain the ground-truth clean range profiles.

3.2.3   Neural Network Model

Our aim is to develop a neural network model able to effectively address RFI by precisely mapping a noisy STFT input to a clean FFT of the beat signal for any given signal, considering both magnitude and phase aspects. To fulfill this goal, we introduced a FCN architecture, visually represented in Figure 3.2. While prior works, such as [27], have explored transformations from STFT to FFT, these approaches do not employ the capabilities of deep learning techniques.

The novelty of our neural architecture primarily lies in the structure of the input and output, each comprising a representation composed of three distinct channels. The first and third channels of the input correspond to the real and imaginary parts of the STFT, while the middle channel represents the magnitude of the STFT. The magnitude of an STFT captures the most significant visual information and can be viewed as an attention map [47, 46], which rather than being computed by the network, it is provided as an input channel. The output follows to a similar channel-based design, with the primary distinction lying in its spatial dimension, as described below. Although our network does not explicitly calculate the phase, it can be derived from the real and imaginary parts.

Our neural model is structured to handle an input tensor of dimensions $154 \times 2048 \times 3$ and produce an output tensor of dimensions $1 \times 2048 \times 3$. The network systematically reduces the dimension along the vertical axis (154), representing the number of time bins for STFT computation, until it reaches the size of 1. Meanwhile, it maintains the dimensions along the other axes constant, specifically the number of FFT points (N_x) and the number of channels, respectively.

Our architecture, depicted in Figure 3.2, consists of 10 convolutional layers organized into 4 blocks. Each of the first two blocks includes 3 conv layers followed by max-pooling, while the third and fourth blocks have 2 conv layers each, with the latter lacking pooling. Leaky ReLU [23] follows each conv layer except the last two. The number of

Table 3.1 Minimum and maximum values for each parameter in our joint uniform distribution used for generating the samples in our database.

Parameter	Minimum	Maximum	Step
Interference sources	1	3	1
SNR [dB]	5	40	5
SIR [dB]	-5	40	-
Relative interference signal slope	0	1.5	-
Number of targets	1	4	-
Target amplitude	0.01	1	-
Target distance [m]	2	95	-
Target phase [rad]	$-\pi$	π	-

filters starts at 32 and increases by 32 in subsequent blocks, reaching 128 in the final block. The kernel size decreases from 13×13 to 5×5 along the network, the exception being the last conv layer with a kernel of 1×1. Conv filters have a stride of 1×1 with circular padding, and pooling filters are 2×1, halving the size vertically. Zero padding ensures even-sized input activation maps for max-pooling.

3.2.4 Weight Pruning

We have developed a weight pruning method [11, 21], with two phases. The first phase starts with a standard training phase. Subsequently, a noise-constrained training phase is executed, specifically designed to prune the inner network noise and enhance its signal-to-noise ratio. It is noteworthy that the network architecture remains unchanged during testing. In the second stage, the pruning method reduce the inner noise of the network by putting to zero the weights smaller than a certain threshold. The threshold is determined by sorting the weights in accordance with the value and put to zero the smallest *r* percentage of the weights.

3.3 Data Sets

We expand the ARIM data set [35] to include various real-world automotive scenarios with multiple sources of interference. In our recent publication [36], we introduced a novel and extensive database named ARIM-v2, comprising 144,000 synthetically generated samples that faithfully replicate real-world automotive scenarios. The generation of each sample involves the random selection of values from a set of realistic parameters detailed in Table 3.1.

Synthetically generated data grants full control over the process, providing comprehensive information for superior solutions. Clean signals serve as ground-truth labels during machine learning model training, facilitating objective performance assessment by comparing predicted output with expected output.

(a) $f_0 = 76.25GHz, B = 1GHz$. (b) $f_0 = 77GHz, B = 1.6GHz$.
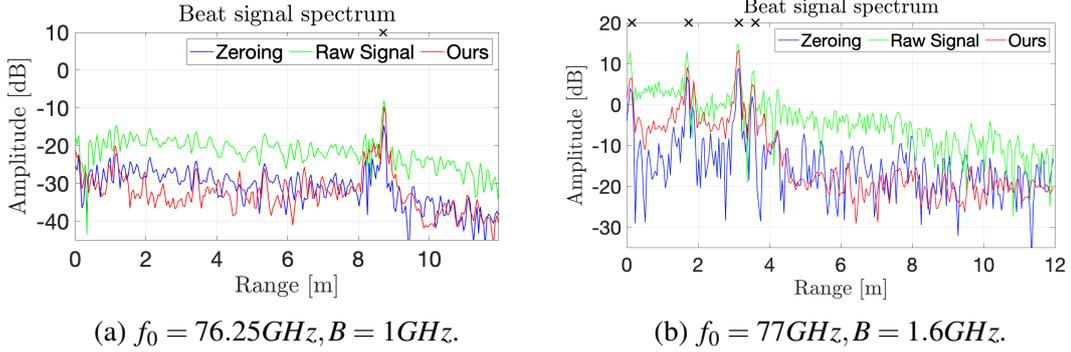
Fig. 3.3 Our FCN+pruning model's qualitative outputs. The real signals are obtained using the following setups: (a) with a radar sensor from FAU, and (b) with the NXP TEF810X 77 GHz radar transceiver. Here, the parameter f_0 denotes the interference central frequency, and B signifies the chirp bandwidth.

Table 3.2 The results on the ARIM-v2 validation and test sets, which contains up to three sources of interference per range profile. Amplitude is measured in decibels (dB), while phase is measured in degrees. The symbol $\uparrow$ indicates higher values are preferable, while $\downarrow$ signifies lower values are better.

Method	Validation set				Test set			
	$\overline{\Delta SNR}\uparrow$	AUC$\uparrow$	MAE$\downarrow$ Ampl.	MAE$\downarrow$ Phase	$\overline{\Delta SNR}\uparrow$	AUC$\uparrow$	MAE$\downarrow$ Ampl.	MAE$\downarrow$ Phase
Oracle (true labels)	16.87	0.971	0	0	17.15	0.970	0	0
Zeroing	8.64	0.930	2.11	12.63	8.94	0.929	2.13	12.55
CNN [40]	12.64	0.952	2.14	8.25	12.94	0.953	2.17	8.13
FCN (ours)	15.02	0.961	1.39	7.32	15.28	0.959	1.40	7.44
FCN + pruning (ours)	**15.09**	**0.963**	**1.25**	**6.41**	**15.36**	**0.961**	**1.27**	**6.58**

3.4 Results

3.4.1 Evaluation Metrics

Our evaluation metrics are AUC, amplitude MAE, phase MAE, and mean Signal-to-Noise Ratio improvement ($\overline{\Delta SNR}$). The mean SNR improvement is calculated for the target with the highest amplitude, representing the difference between the SNR before and after interference mitigation in the range profile.

3.4.2 Hyper-Parameters Tuning

In all experiments, we utilized mini-batches of 16 samples, employing the Adam optimizer [15] with a learning rate of $5 \cdot 10^{-5}$ and a weight decay of 10^{-5}. The noise reduction ratio (r) is validated across the set of values $\{0.15, 0.3, 0.45\}$, with the optimal performance observed at $r = 0.3$. The network is optimized using the mean absolute error for all three considered channels.

3.4.3 Results on ARIM-v2

In the ARIM-v2 data set, we compared our FCN under both conventional and weight pruning regimes with the oracle, zeroing baseline, and CNN by Rock *et al.*[40]. Results in Table 3.2 shows that our approach obtained superior performance across all metrics, close to the oracle levels. In challenging scenarios with multiple interference sources, our FCN achieves significantly lower errors compared to zeroing. Our method outperforms the state of the art methods, being a robust method even for multiple interferene sources.

3.4.4 Generalization to Real Data

The primary consideration when training a neural network on synthetically generated samples is its ability to generalize to real-world data. Consequently, we assess the generalization capacity of our FCN by testing it on real samples obtained from two distinct radar sensors. In Figure 3.3, we present qualitative results for two real samples featuring interference, comparing our method with the zeroing approach. The first plot, illustrated in Figure 3.3a, are generated using real data sourced from FAU [10]. Upon inspection of the results, it becomes evident that our network succeed to provide more accurate estimations of target amplitudes, effectively mitigating interference and reducing the noise floor. The second plot, represented in Figure 3.3b, contains data supplied by the NXP company, captured using the NXP TEF810X 77 GHz radar transceiver during a couple of outdoor experiments on a two-lane road. In these experiments, the victim radar was affixed to the bumper of a car, while the interfering radar was positioned on a tripod at a fixed location outside the roadway. Our approach successfully mitigates the interference, yielding superior results in terms of target amplitude compared to the zeroing algorithm.

3.5 Discussion

In this work, we present a significant contribution to the field of automotive radar signal processing through the introduction of a novel fully convolutional network capable of accurately estimating both the magnitude and phase of radar signals, particularly in the presence of multiple interference sources. Complementing this advancement is the creation of a comprehensive database comprising radar signals simulated within realistic and complex scenarios. Moreover, our model is able to generalize to two distinct radar sensors without any fine-tuning.

Chapter 4

Ocean Patterns Analysis

4.1 Introduction

In our research, we build upon [33, 34] to develop the unsupervised ocean image retrieval task. More exactly, we integrated the subaperture decomposition (SD) algorithm with unsupervised deep learning (UDL). By employing UDL for image retrieval, we eliminate the need for labeled data. Furthermore, by coupling it with a preprocessing algorithm based on SD, we significantly enhance retrieval accuracy, approaching levels of accuracy attained by supervised learning approaches. Additionally, we evaluate the capability of our model to accommodate physics-guided remote sensing algorithms, such as Doppler centroids images. Utilizing the described physics-aware techniques, we developed an efficient query-by-image algorithm tailored to assist experts in identifying similar phenomena on the ocean surface. Each vignette is characterized by an embedding vector computed using a pretrained deep neural network, trained in an unsupervised fashion. Furthermore, we have expanded the application of query-by-image to a more complex approach named query-by-physical-parameters. Specifically, we have estimated the Doppler centroids images of the subaperture single-look complex (SLC) vignettes and utilized them for the query system.

4.2 Proposed Method

4.2.1 Subaperture Decomposition

Given the highly non-stationary nature of the ocean surface, observing it from different angles can yield additional information about the illuminated area. Hence, we adopt a strategy of decomposing the vignette into subapertures, each corresponding to an image formed using only a portion of the total azimuth angle. This decomposition allows us to simulate different observation angles of the same scene, thereby gathering more comprehensive information. The SD algorithm is depicted in the first part of Figure 2.1.
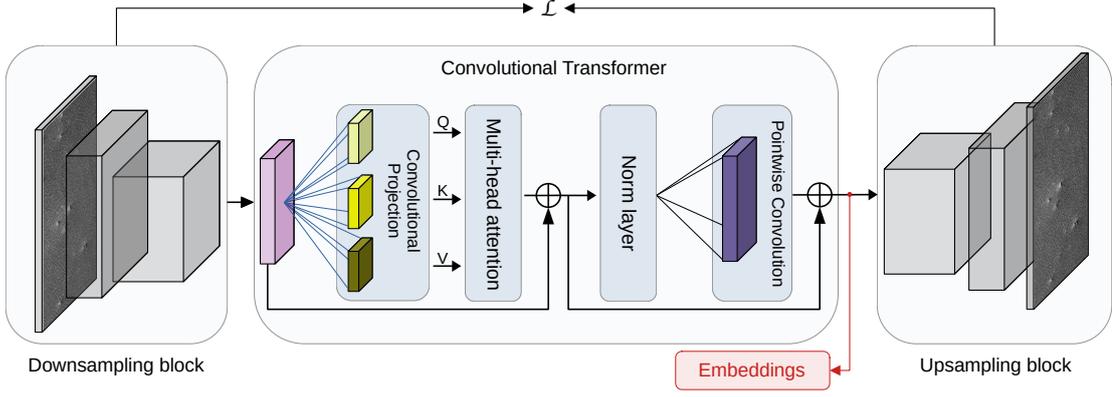
Fig. 4.1 The auto-encoding CyTran architecture. The red arrow highlights the place in the network where the embeddings are considered.

4.2.2 Doppler Centroids Estimation

We proposed a physics-based preprocessing technique on the ocean surface based on the DCE. Each vignette, or subaperture generated by SD, could be further processed in order to obtain the Doppler centroids. This processing stage transforms the SAR data into physical information, related to the movement of the illuminated area. The processing is formally described in Section 2.1.4. An illustration of this process is depicted in the orange box in Figure 2.1. When the DCE is integrated into the pipeline, each Doppler image is estimated on a subaperture and subsequently decimated.

4.2.3 Unsupervised Neural Network

In our study, we employed the CyTran generative architecture, consisting of a convolutional downsampling block, a convolutional transformer block, and a deconvolutional upsampling block, as depicted in Figure 4.1. Inspired by the block proposed in our related work [38], the convolutional transformer block is employed as the bottleneck for the architecture. Let $X \in \mathbb{R}^{c \times m \times n}$ be the input tensor for the transformer block, where c represents the number of channels, and m and n denote the width and height, respectively. The spatial dimensions of the visual tokens are determined by the receptive field of the filters in the convolutional projection layer, as illustrated in Figure 4.1. The convolutional projection comprises three nearly identical projection blocks, each with separate parameters. Let W_Q, W_K, and W_V represent the learnable parameters of the three projection layers. The query, key, and value embeddings are computed as follows:

$$
\begin{aligned}
Q &= \text{conv\_projection}\,(X, W_Q)\,, \\
K &= \text{conv\_projection}\,(X, W_K)\,, \\
V &= \text{conv\_projection}\,(X, W_V)\,,
\end{aligned}
\tag{4.1}
$$

where $Q \in \mathbb{R}^{n_q \times d_q}$, $K \in \mathbb{R}^{n_k \times d_k}$ and $V \in \mathbb{R}^{n_v \times d_v}$. For the subsequent operation involving matrix multiplications, we need $d_q = d_k$, and $n_k = n_v$. Due to the equal number of filters in the pointwise convolution in all three blocks, $d_q = d_k = d_v$. The output query, keys

Table 4.1 Accuracy results for a ResNet18 model on the TenGeoP-SARwv test set. We denote by "Subaperture (1)" that the input is only the first subaperture. The significantly better results (level 0.01), according to a paired McNemar's test, are marked with †.

Vignette	98.0
Subaperture (1)	94.0
Subapertures	98.9†
DCE Vignette	78.6
DCE Subapertures (1)	75.3
DCE Subapertures	93.3†

Table 4.2 Retrieval results for SD and DCE inputs. We included the results for the supervised frameworks (S-) and unsupervised frameworks (U-) based on the proposed CyTran architecture.

	S-Vig	S-Subap	U-Vig	U-Subap	S-Dop Vig	S-Dop Subap	U-Dop Vig	U-Dop Subap
P@5	98.1	98.9	47.4	72.6	76.9	91.3	51.1	66.7
P@50	97.4	97.2	32.6	57.3	69.8	88.0	36.0	52.0

and values are passed to a multi-head attention layer, with the goal of capturing the interaction among all tokens by encoding each entity in terms of the global contextual information.

We employ the CyTran architecture in an unsupervised manner, aiming to approximate the identity function through input auto-encoding. We considered the mean squared error as the objective function and extracted the embeddings as illustrated in Figure 4.1.

4.2.4 Content-Based Image Retrieval

Considering a very large database with ocean SAR images, we introduce an unsupervised algorithm designed to identify similar vignettes. We emphasize that our algorithm is versatile, requiring no labels and not being constrained to any particular input data type. To showcase the versatility of our approach, we consider two distinct data types as input: SAR subapertures and Doppler centroids estimated from subapertures. Consequently, we conduct content-based image retrieval using both amplitude SAR data and physics-aware representations. Leveraging the latter data type, the algorithm can construct a more sophisticated search engine capable of identifying phenomena based on specific physical features, such as ocean currents with particular speeds.

Technically, the retrieval algorithm associate an embedding vector for all images in the database. The embedding vector is obtained by encoding the image with the CyTran network. Afterwards, having a query image, we obtain the embedding descriptor, compute the cosine distance between the query and entire database and retrieve the most N similar examples.

4.3 Results

4.3.1 Data Set

The TenGeoP-SARwv dataset comprises over 37,000 ocean vignettes representing 10 distinct geophysical phenomena on the ocean surface. Following the methodology outlined in [33, 34], we randomly divided the data into training (70%), validation (15%), and test (15%) sets. Additionally, for experiments involving Doppler-based analyses, we followed the pipeline depicted in Figure 2.1. For brevity, we will employ the following abbreviations for the data set classes: POW - Pure Ocean Waves, WS - Wind Streaks, MCC - Micro Convective Cells, RC - Rain Cells, BS - Biological Slicks, SI - Sea Ice, Ic - Iceberg, LWA - Low Wind Area, AF - Atmospheric Front, OF - Oceanic Front.

4.3.2 Hyper-Parameters Tuning

In the classification experiment, we calibrated the hyper-parameters similar to those specified in [33]. As for the CyTran model, we adhered to the identical network hyper-parameters outlined in [34], making adjustments solely to the input and output channel numbers to align with the input type. Training the model for 100 epochs, we employed the Adam optimizer and a mini-batch size of 16. For DCE, we set $d_1 = d_2 = 32$ for the mean filter.

4.3.3 Evaluation Metrics

We presented the accuracy for the classification task and conducted McNemar statistic tests to demonstrate the statistical significance of our findings. As for the retrieval task, given our focus on big data streams, we provided precision scores for 5 ($P@5$) and 50 ($P@50$) examples. Each score was averaged over 100 queries. More precisely, we computed $P@5$ and $P@50$ for 100 query samples and then averaged the results.

4.3.4 Classification Results

We expand upon the findings presented in [33] in Table 4.1, where we present the classification accuracy achieved by the ResNet18 model on the TenGeoP-SARwv test set, considering various input data types. We observe that the SD algorithm improves the accuracy for both amplitude SAR data and DCE.

4.3.5 Unsupervised Training Results

We trained the CyTran auto-encoder model [38] on the TenGeoP-SARwv training set and selected the best-performing model based on the reconstruction loss on the evaluation set. It is important to note that despite trying multiple models (such as ResNet auto-encoder and U-Net), they failed to converge to optimal reconstruction results, thus were excluded from consideration.
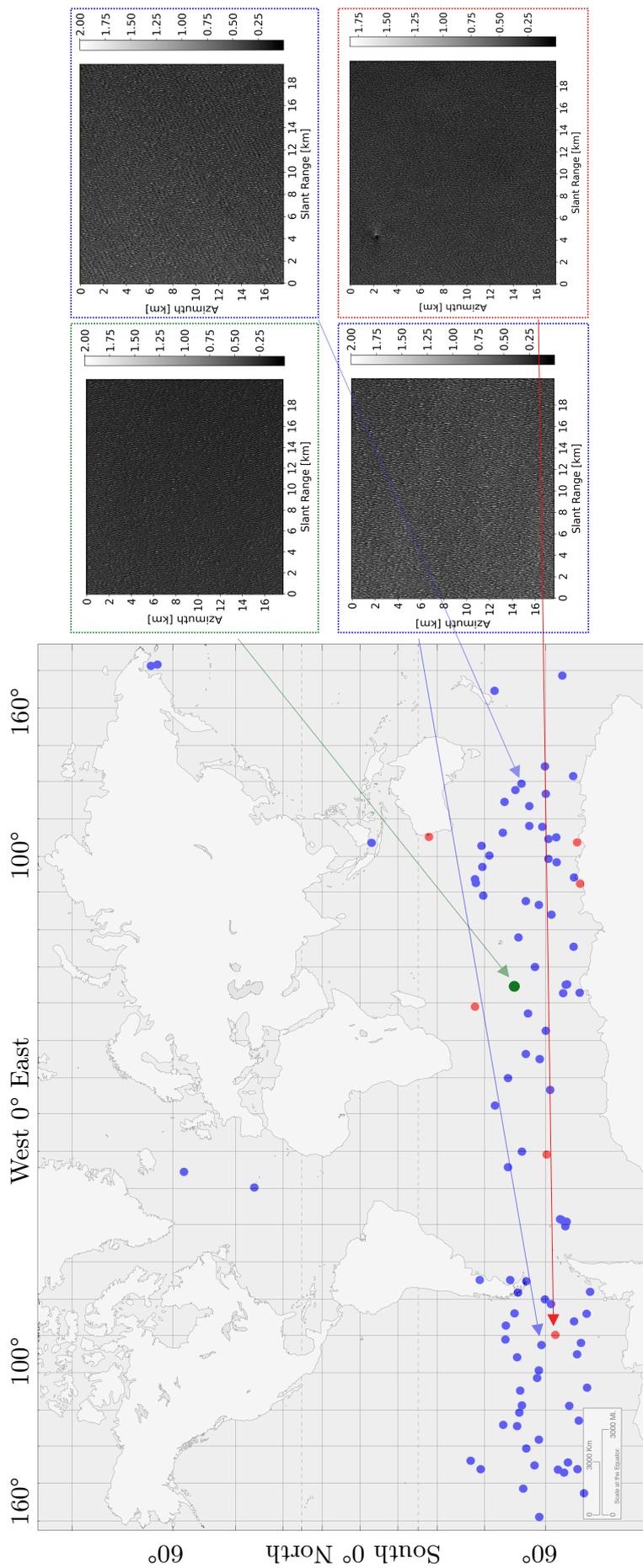
Fig. 4.2 Retrieval results are depicted based on embeddings from the CyTran model trained on all subapertures from the original vignette. We showcase the most similar $N_{max} = 100$ samples along with their localization information. In the visualization, the query image is denoted in green, images from the same class are in blue, and images from different classes are in red. On the right side, original vignettes are displayed for selected samples: green and blue represent Pure Ocean Waves, while red represents Iceberg images.

4.3.6 Retrieval Results

In Table 4.2, we present the retrieval performance using embeddings generated by the CyTran network trained on both original vignettes and subapertures, comparing them against embeddings produced by a ResNet18 model trained in a supervised manner. When comparing the supervised embeddings on original vignettes (S-Vig) and subapertures (S-Subap), the results exhibit negligible differences, with overall variances smaller than 1% for both $P@5$ and $P@50$. However, the introduction of the SD algorithm consistently enhances precision in the retrieval results with unsupervised embeddings. This reduction in the retrieval performance gap between supervised and unsupervised approaches underscores the significant impact of the SD algorithm.

Moreover, in Table 4.2, we present the retrieval performance using embeddings generated by the CyTran network trained on both DCE on original vignettes and subapertures, comparing them against embeddings produced by a ResNet18 model trained in a supervised manner. As anticipated from the classification experiment, the retrieval performance sees significant improvement when utilizing supervised embeddings based on subapertures.

We illustrated the retrieval results for the unsupervised embeddings trained on subapertures for one query images in Figure 4.2 (pure ocean waves class). We notice that the retrieved images, representing the most similar samples, are scattered across the geographical area where the phenomena could potentially occur. This suggests that the unsupervised model does not exhibit overfitting concerning the geographical location. Additionally, we observed structural similarities among the images retrieved with incorrect labels (highlighted by the red points in Figure 4.2). This observation could indicate the presence of two phenomena in the same location or other intrinsic similarities.

4.4 Discussion

In this study, we expanded upon prior research conducted in [33] by integrating the SD algorithm into unsupervised feature learning pipelines with transformer networks. Our objective was to leverage SD as a pretraining stage for models aimed at SAR retrieval tasks on the ocean surface. Our investigation revealed the substantial impact of the SD method on retrieval performance, particularly when incorporating more physics-based algorithms such as Doppler Centroids Estimation for ocean retrieval. This experimental paradigm enabled the development of more sophisticated search engines capable of identifying similar physical parameters rather than merely analogous structures, such as ocean current speeds. In essence, our approach centered around a data-centric methodology aimed at augmenting the efficacy of both supervised and unsupervised classification and retrieval algorithms.

Chapter 5

Sea Ice Segmentation for SAR Imagery

5.1 Introduction

Sea ice is a crucial component of the Earth's climate system, highly responsive to changes in temperature and atmospheric conditions. Accurately assessing sea ice parameters is vital for understanding and predicting climate variations. However, manually analyzing the extensive satellite data covering ice-covered areas is impractical. Therefore, automated algorithms are essential to fully utilize the continuous satellite data streams. In this study, we developed two architectures, one based on the previously described CyTran and the second called UT-MHTA, a novel architecture based on the UNet transformer, designed for sea ice segmentation using SAR satellite imagery. UT-MHTA replaces the conventional multi-head attention (MHA) block with a multi-head transposed attention (MHTA), enabling the capture of long-range pixel interactions while maintaining suitability for large images. Our approaches outperforms state-of-the-art methods, delivering superior results without significantly increasing computational complexity. Specifically, our best network, UT-MHTA achieves a mean intersection over union (mIoU) of 68.76% on the AI4Arctic [4] data set, with an inference time of 865ms for a 400 km^2 product.

5.2 Proposed Methods

5.2.1 Data Preprocessing

Considering the extensive dimensions of the products within the AI4Arctic data set [4], training neural networks on the entire data set samples becomes impractical. Instead, we adopt a strategy of extracting fixed-length windows sized at 512×512. Each training sample, sized at $512 \times 512 \times 3$, includes both HH and HV polarizations alongside the incidence angle.
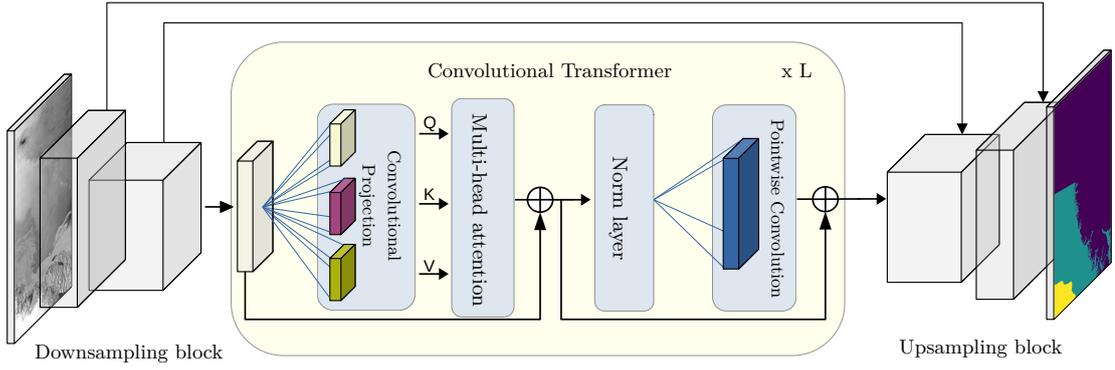
Fig. 5.1 U-CyTran segmentation architecture. The models contains upsampling and downsampling blocks, with a convolutional transformer backbone.
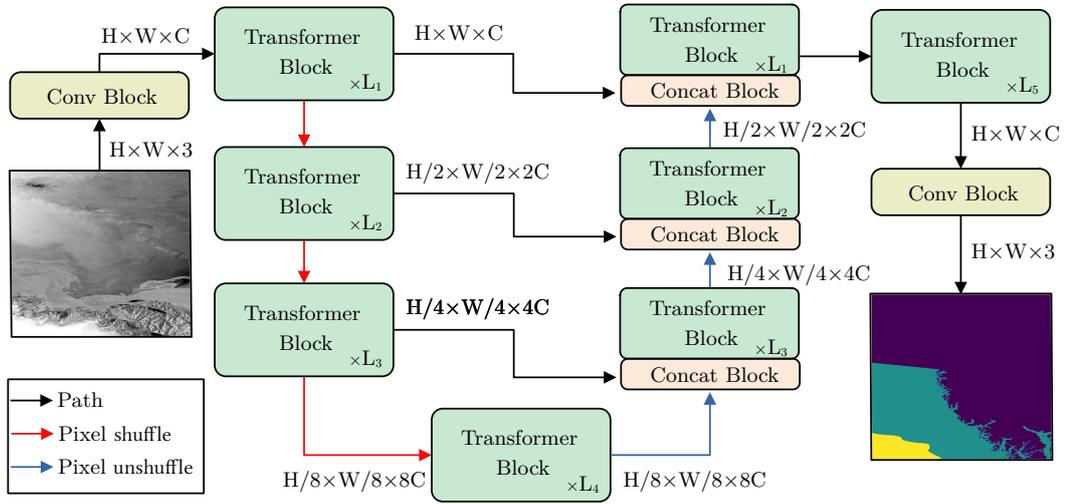
Fig. 5.2 UT-MHTA segmentation architecture comprises a sequence of transformer blocks, flanked by two convolutional blocks at both the beginning and the end of the architecture. The significance of each arrow is illustrated in the lower-left corner.

5.2.2 Residual CyTran-Based Architecture

The U-CyTran architecture is based on the success of the CyTran model [38] of being both performance and speed optimized. Since our goal is to develop a systems which is able to process the data near real-time (in milliseconds), we found suitable the CyTran architecture, which speed-up the processing time by downsampling and upsampling the input data for the transformer block. The architecture is built by downsampling and upsampling blocks, with a bottleneck of a convolutional transformer module. The overall architecture is visually described in Figure 5.1.
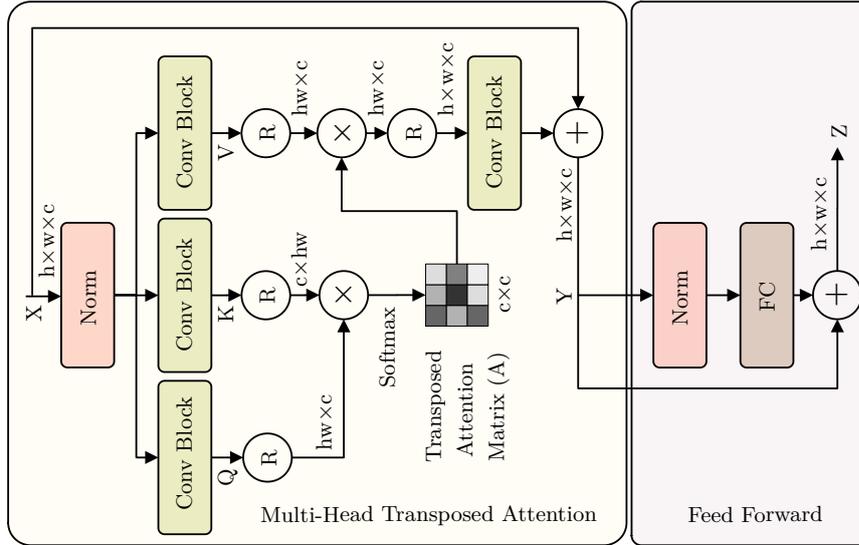
Fig. 5.3 Transformer block in UT-MHTA. The figure describes the processing flow for the transformer block, the input being processed by the multi-head transposed attention and the feed forward blocks.

5.2.3 Multi-Head Transposed Attention Transformer

5.2.4 UT-MHTA Architecture

Given the input data $I \in \mathbb{R}^{H \times W \times 3}$, UT-MHTA applies firstly a conv block to obtain low-level features $F \in \mathbb{R}^{H \times W \times C}$, where $H, W, C \in \mathbb{N}^+$. Both input and output conv blocks are built by a convolutional layer, having the kernel size 3 and padding and stride 1. The convolutional layer is followed by batch-norm and ReLU activation function. Next, F passes through a four level UNet-alike architecture, where the processing modules are Transformer blocks, as depicted in Figure 5.3. The layers on the downsampling part of the architecture are composed solely of MHTA-based transformer blocks, the downsampling operation being performed by pixel shuffle blocks (highlighted by the red arrows in Figure 5.3). Regarding the upsampling part of the network, besides the identical transformer block, we also employed a concatenation block, to aggregate the information coming from early layers. Similarly, for the upsampling procedure we employed pixel unshuffle operation. The pixel shuffle and pixel unshuffle operations are used to prevent information loss along the network. Lastly, the output of the last Transformer block from the UNet system is passed into a final Transformer block, followed by the output conv block, which predicts the segmentation maps. Similar to the conventional MHA [8], we split the number of channels into heads and learn separate attention maps in parallel.
Transformer block. In Figure 5.3, we provide a detailed overview of the transformer block employed in UT-MHTA. The initial component of this block is the multi-head transposed attention module, designed to facilitate channel-wise attention while main-

taining linear time and memory complexity concerning the input dimensions $h \times w$. Notably, all convolutional blocks consist of a standard convolutional layer with a kernel size of 3, along with padding and stride set to 1. We highlight that the attention matrix $A \in \mathbb{R}^{c \times c}$ exclusively operates on the channel dimension, ensuring low complexity. The Feed Forward module employed here mirrors the structure from vanilla transformers [8]. This design choice allows for consistent performance and facilitates seamless integration within the broader architecture.

Concat block. The Concat block performs a regular concatenation on the channel axis, followed by a point-wise convolutional layer. The point-wise convolution keeps the same number of maps and is intended to aggregate information from encoding and decoding branches.

5.3 Results

5.3.1 Data Set

The AI4Arctic data set [4] was published for the AI4EO sea ice competition initiated by the European Space Agency. This data set consists of Sentinel-1 active microwave SAR data paired with passive Microwave Radiometer data from the AMSR2 satellite sensor. The GRDM products cover a 400 km^2 region, with a resolution of 90 meters and a pixel spacing of 40 meters. We utilized the same data splits as outlined in [32].

5.3.2 Hyper-Parameters Tuning

U-CyTran. U-CyTran employs the Adam optimizer with the focal loss function [20] to optimize model performance. We initiate training with an initial learning rate set to 10^{-3}, accompanied by a decay factor of 0.1 applied after every 10 epochs. Each model undergoes training for 50 epochs, with mini-batches comprising 32 samples.

UT-MHTA. The UT-MHTA is optimized using the AdamW optimizer in conjunction with the focal loss function [20]. Training starts with an initial learning rate of 10^{-4}, followed by a decay factor of 0.5 applied after every 10 epochs. The model is trained 50 epochs, utilizing mini-batches of 8 samples. We configure the number of blocks across different stages, with $L_1 = L_2 = L_3 = L_4 = L_5 = 3$, and 5 attention heads.

5.3.3 Evaluation Metrics

In our semantic segmentation task involving three distinct classes, we have identified the mean intersection over union (mIoU) as the most informative metric for evaluating model performance. This metric proves particularly valuable in scenarios with imbalanced class distributions, as it provides insights into the model's ability to accurately segment each class while accounting for variations in class sizes. Moreover, we report the inference

Table 5.1 Segmentation and inference time results on the AI4Arctic test set. We included for ablation the UT-MHTA against a vanilla Transformer and UT without MHTA block.

Method	mIoU (%)	Inference time (ms)
ResNet AE	53.04	87
UNet [31]	56.43	92
AE	51.23	47
Transformer	64.12	673
CyTran [32]	63.68	120
U-CyTran	63.98	121
UT	62.13	182
UT-MHTA	68.76	865

time for processing a full-resolution scene with spatial dimensions of approximately 1100×1100.

5.3.4 Experimental Results

U-CyTran. Table 5.1 presents a comparative analysis of the performance of U-CyTran against two baseline methods: a ResNet-based auto-encoder and UNet [41], as utilized in [31]. We observe that U-CyTran outperforms both baseline methods by more than 7% in terms of accuracy. Notably, we also investigate the impact of different components of the network. When solely employing the downsampling and upsampling blocks (AE), we achieve the highest inference speed. However, this approach significantly compromises performance. Conversely, utilizing only the transformer block results in a notable decrease in speed, with a marginal improvement in accuracy of about 0.14%.

UT-MHTA. Table 5.1 showcases the comparative results of UT-MHTA against three baseline methods: a ResNet-based auto-encoder, UNet [31], and CyTran [32]. UT-MHTA outperforms all baseline methods by more than 5% in terms of mIoU. Specifically, compared to the pure transformer architecture, UT-MHTA achieves a notable improvement of 4.64% while experiencing a modest increase in inference time of less than 200ms. This enhancement underscores the efficacy of integrating the MHTA block into the UT architecture, which significantly boosts performance without highly compromising computational efficiency.

5.3.5 Applicability to IW Data

Considering that or AI4Arctic [4] data set contains only EW GRDM data, we need to further expand our processing techniques to handle IW SLC data. We avoid a laborious manual labelling performed by experts, by proposing a preprocessing technique that transforms the IW SLC data into a EW GRD-alike image, called IW-EW GRD. In Figure 5.4, we illustrate the processing framework from IW SLC image to multi-look detected
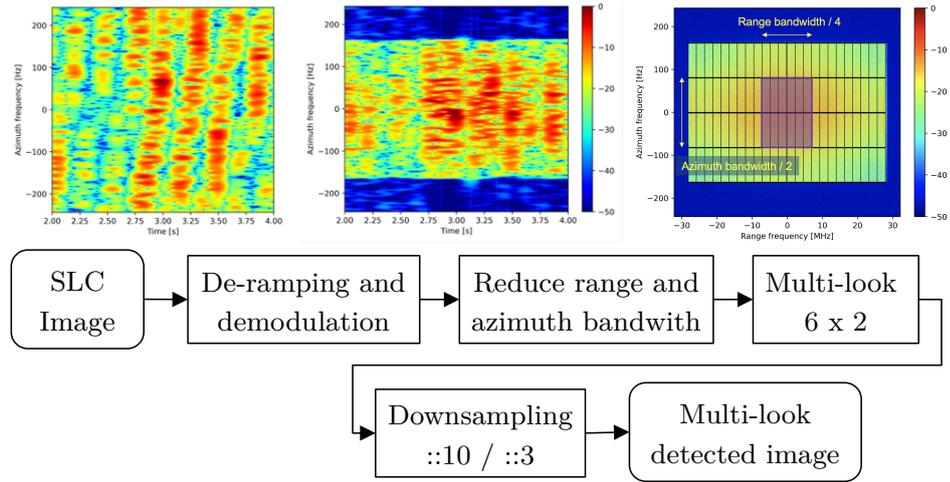
Fig. 5.4 The preprocessing framework for IW SLC adaptation to EW GRD-alike image.

image (IW-EW GRD), aligning its resolution and pixel spacing closely with the EW GRDM training data.

To subjectively evaluate the preprocessing techniques and the UT-MHTA model, we have included a figures showcasing the obtained results. Figure 5.5 present HV mode for both IW GRDH and IW SLC within the same geographic area. In the bottom of the figure, we overlapped the labels predicted by the UT-MHTA network onto the HH products. Clearly, the network demonstrates robust generalization to unseen data, with similar predictions between SLC and GRDH products.

5.4   Discussion

In this work we introduced U-CyTran and UT-MHTA, two novel sea ice segmentation architectures based on the UNet transformer framework, tailored for analyzing SAR satellite imagery. In the U-CyTran architecture, we iterated over the work presented in [32] by adding the skip connections, allowing the network to easily pass features between input and output parts. In UT-MHTA, we innovate by replacing the conventional MHA mechanism with our Multi-Head Transposed Attention. This adaptation enables the capture of long-range pixel interactions, a crucial capability for accurately delineating sea ice boundaries in large-scale images. Furthermore, UT-MHTA exhibits robustness and scalability, making it well-suited for processing extensive datasets typical of satellite imagery. Moreover, we extended the area of usage of our networks by applying a preprocessing framework from IW SLC data to EW GRD-alike data. Considering this additional stage, we observed in multiple products that the predictions are similar, validating the usability of our network for directly Level 1 SLC products.

SLC Product name
S1A_IW_SLC__1SDH_20220309T083922_20220309T083949_042242_0508E1_1A70
GRDH Product name
S1A_IW_GRDH_1SDH_20220309T083924_20220309T083949_042242_0508E1_E3A0
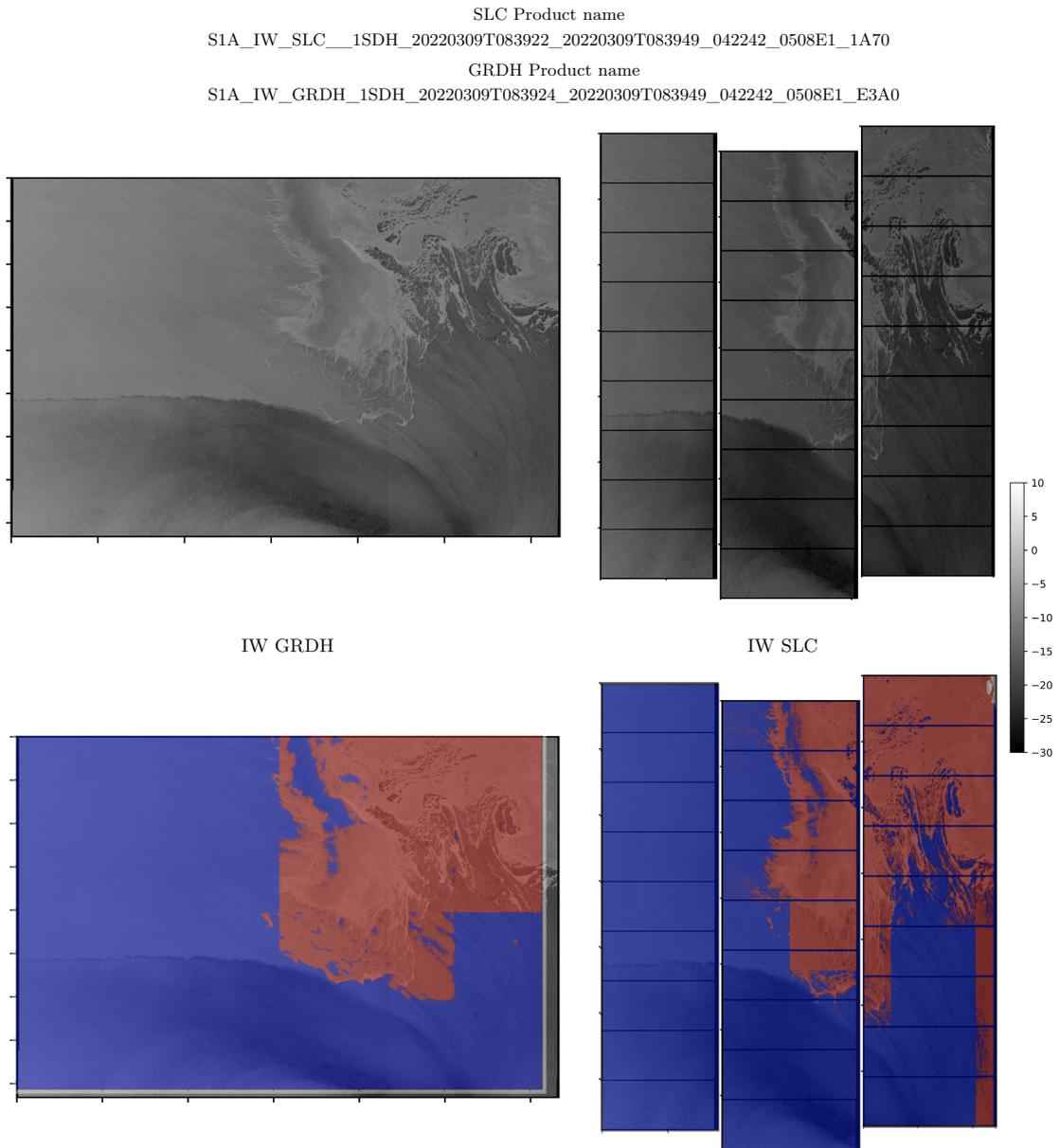
IW GRDH

IW SLC

Fig. 5.5 In the top left part of the figure we illustrate the IW GRDH image after down-sampling. In the top right part of the figure we illustrate the IW SLC image after the our preprocessing framework. In the bottom part of the figure we illustrate the corresponding outputs of our UT-MHTA network downsampling for the above images. All products are in HH mode.

Chapter 6

Conclusions

In our work we aimed to further develop in the field of signal processing by conducting an application-based research and applying innovative deep learning techniques to overcame existing gaps. We focused our research in two major areas, automotive interference radar mitigation and earth observation. The conclusions are detailed below in accordance with the specific domain.

6.1 Automotive Radar Interference Mitigation

Our contribution in this domain is as follows:

- We designed a novel fully convolutional neural network for automotive radar interference mitigation in complex scenarios, such as multiple sources of interference. The designed FCN is able to estimate both the magnitude and phase of the clean beat signal spectrum, being compatible with upcoming radar processing blocks (e.g., speed estimation), which needs the phase.

- The publicly released ARIM-v2, a large scale data set for automotive radar interference mitigation. The database contains complex real-world scenarios, as well as, multiple sources of interference.

- We conducted an extensive set of experiments, showcasing that our FCN attains state-of-the-art results in the field. Moreover, we empirically proved the FCN capability of generalization on real radar sensor data, mitigating the interference for two distinct sensors, without any further fine-tuning.

6.2 Ocean Patterns Analysis

Our contribution in this domain is as follows:

- We applied the subaperture decomposition to improve the network performance for both vignette and Doppler centroids maps.

- We developed a hybrid convolutional transformer for unsupervised image retrieval on the ocean surface.

- We developed a data centric approach by combining both SD and DCE for a higher unsupervised retrieval, as well as, a physics-based search framework. Our data centirc approach proved to be generic and performant, allowing our model to learn complex patterns, closing the gap between supervised and unsupervised image retrieval.

6.3 Sea Ice Segmentation for SAR Imagery

Our contribution in this domain is as follows:

- We evolved the CyTran architecture to include UNet-alike skip connections for more accurate sea-ice segmentation results. Moreover, we developed the UT-MHTA architecture which replaces the conventional multi-head attention module with a multi-head transposed attention mechanism, facilitating the capture of long-range pixel interactions, while being suitable for high dimensional images.

- Considering the need of both accurate and fast algorithms in the field, we optimized our networks to attain the best accuracy-speed trade-off, such that our networks are feasible candidates for near real-time processing.

- We developed a preprocessing framework for IW SLC product adaptation to EW GRD-alike images, extending the networks applicability to Level 1 SLC products.

6.4 List of Publications

Journal Articles

1. Ristea, N.C., Anghel, A. and Ionescu, R.T., 2021. Estimating the Magnitude and Phase of Automotive Radar Signals Under Multiple Interference Sources with Fully Convolutional Networks. IEEE Access, 9, pp.153491-153507.

2. Ristea, N.C., Anghel, A., Datcu, M. and Chapron, B., 2023. Guided Unsupervised Learning by Subaperture Decomposition for Ocean SAR Image Retrieval. IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-11, 2023, Art no. 5207111.

3. Ristea, N.C., Miron, A.I., Savencu, O., Georgescu, M.I., Verga, N., Khan, F.S. and Ionescu, R.T., 2023. CyTran: a Cycle-Consistent Transformer with Multi-Level Consistency for Non-Contrast to Contrast CT Translation. Neurocomputing, 538, p.126211.

Conference Proceedings

1. Ristea, N.C., Anghel, A., Datcu, M. and Chapron, B., 2022, July. Guided Deep Learning by Subaperture Decomposition: Ocean Patterns from SAR Imagery. In Proceedings of IEEE International Geoscience and Remote Sensing Symposium (pp. 6825-6828).

2. Chirtu, A., Ristea, N.C. and Radoi, A., 2022, July. Convolutional Transformers for Aerial Image Classification: a General to Specific Learning Curve. In Proceedings of IEEE International Geoscience and Remote Sensing Symposium (pp. 659-662).

3. Ristea, N.C., Anghel, A. and Datcu, M., 2023, July. Sea Ice Segmentation From SAR Data by Convolutional Transformer Networks. In Proceedings of IEEE International Geoscience and Remote Sensing Symposium (pp. 168-171).

4. Ristea, N.C., Anghel, A. and Datcu, M., 2024, July. Multi-Head Transposed Attention Transformer for Sea Ice Segmentation in SAR Imagery. In Proceedings of IEEE International Geoscience and Remote Sensing Symposium.

6.5   List of Publications Outside of the Main Scope of the Thesis

Journal Articles

1. Cutler, R., Saabas, A., Naderi, B., Ristea, N.C., Braun, S. and Branets, S., 2024. ICASSP 2023 Speech Signal Improvement Challenge. IEEE Open Journal of Signal Processing.

2. Cutler, R., Saabas, A., Pärnamaa, T., Purin, M., Indenbom, E., Ristea, N.C., Gužvin, J., Gamper, H., Braun, S. and Aichner, R., 2024. ICASSP 2023 Acoustic Echo Cancellation Challenge. IEEE Open Journal of Signal Processing.

3. Madan, N., Ristea, N.C., Ionescu, R.T., Nasrollahi, K., Khan, F.S., Moeslund, T.B. and Shah, M., 2023. Self-Supervised Masked Convolutional Transformer Block for Anomaly Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 1, pp. 525-542.

Conference Proceedings

1. Ristea, N.C. and Ionescu, R.T., 2023, October. Cascaded Cross-Modal Transformer for Request and Complaint Detection. In Proceedings of the 31st ACM International Conference on Multimedia (pp. 9467-9471).

2. Indenbom, E., Ristea, N.C., Saabas, A., Parnamaa, T., Guzvin, J. and Cutler, R., 2023. DeepVQE: Real Time Deep Voice Quality Enhancement for Joint Acoustic Echo Cancellation, Noise Suppression and Dereverberation. In Proceedings of INTERSPEECH Conference.

3. Ristea, N.C., Ionescu, R.T. and Khan, F., September, 2022. SepTr: Separable Transformer for Audio Spectrogram Processing. In Proceedings of INTER-SPEECH Conference (pp. 4103-4107).

6.6   Future Directions

Thanks to the significant focus placed on deep learning techniques, architectures undergo continuous refinement and enhancement. Considering that the current solutions in the signal processing field are far from being perfect, further development is necessary to attain better performance or faster inference, such that the models would be seamlessly deployed in real-world applications and settings. To this end, our future research will focus to (i) integrate and adapt new models in the current described applications and (ii) explore the possibility to integrate deep learning frameworks in more signal processing-based applications. Summarising, our future research will focus on:

- Improve the proposed models in the radar interference mitigation task in terms of speed, while keeping about the same performance. This research venue is specifically intended to shrink the current models, such that the solutions would be viable to run on real-time low-end devices.

- Improve the proposed model for ocean image retrieval in terms of performance. Since the application's speed requirements are more relaxed, further improvements should be developed to surpass the current retrieval accuracy.

- Developing new unsupervised learning strategies, specifically designed for physical-based ocean image retrieval. Exploring data characteristics could lead to better unsupervised training regimes, which eventually would increase the retrieval performance.

- The current performance of sea ice segmentation on the IW SLC products (by performing IW-EW GRD transform) could be further improved by aligning even more the resolution of the obtained product to the EW GRDM resolution. This could be corrected by adapting the number of sub-looks and performing sub-looks overlapping.

- Explore new areas of signal processing where deep learning techniques could be applied, e.g., wind estimation based on SAR data for Earth observation.

References

[1] Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer Normalization.

[2] Bechter, J., Rameez, M., and Waldschmidt, C. (2017). Analytical and experimental investigations on mitigation of interference in a dbf mimo radar. *IEEE Transactions on Microwave Theory and Techniques*, 65(5):1727–1734.

[3] Brooker, G. M. (2007). Mutual interference of millimeter-wave radar systems. *IEEE Transactions on Electromagnetic Compatibility*, 49(1):170–181.

[4] Buus Hinkler, J., Wulf, T., Stokholm, A. R., Korosov, A., Saldo, R., and Pedersen, L. T. (2022). AI4Arctic Sea Ice Challenge Dataset. *Technical University of Denmark*.

[5] Cai, Z., Fan, Q., Feris, R. S., and Vasconcelos, N. (2016). A unified multi-scale deep convolutional neural network for fast object detection. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 354–370. Springer.

[6] Carter, N. A., Dawson, J., Joyce, J., Ogilvie, A., and Weber, M. (2018). Arctic Corridors and Northern Voices: Governing Marine Transportation in the Canadian Arctic (Pond Inlet, Nunavut Community Report).

[7] Dhillon, A. and Verma, G. K. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, 9(2):85–112.

[8] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *Proceedings of ICLR*.

[9] Fan, W., Zhou, F., Tao, M., Bai, X., Rong, P., Yang, S., and Tian, T. (2019). Interference Mitigation for Synthetic Aperture Radar Based on Deep Residual Network. *Remote Sensing*, 11(14):1654.

[10] Fuchs, J., Dubey, A., Lübke, M., Weigel, R., and Lurz, F. (2020). Automotive Radar Interference Mitigation using a Convolutional Autoencoder. In *Proceedings of RADAR*.

[11] Han, S., Pool, J., Tran, J., and Dally, W. (2015). Learning both weights and connections for efficient neural network. *Proceedings of NIPS*, 28:1135–1143.

[12] He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.

[13] Jiao, L., Tang, X., Hou, B., and Wang, S. (2015). Sar images retrieval based on semantic classification and region-based similarity measure for earth observation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8):3876–3891.

[14] Kim, G., Mun, J., and Lee, J. (2018). A Peer-to-Peer Interference Analysis for Automotive Chirp Sequence Radars. *IEEE Transactions on Vehicular Technology*, 67(9):8110–8117.

[15] Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of ICLR*.

[16] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of NIPS*, pages 1097–1105.

[17] Kunert, M. (2012). The EU project MOSARIM: A general overview of project objectives and conducted work. In *Proceedings of EuRAD*.

[18] Kunert, M., Bodereau, F., Goppelt, M., Fischer, C., John, A., Wixforth, T., Ossowska, A., Schipper, T., and Pietsch, R. (2010). D1.5 - Study on the state-of-the-art interference mitigation technique, MOre Safety for All by Radar Interference Mitigation (MOSARIM) project. Technical report, Robert Bosch GmbH.

[19] Laghezza, F., Jansen, F., and Overdevest, J. (2019). Enhanced Interference Detection Method in Automotive FMCW Radar Systems. In *Proceedings of IRS*, pages 1–7.

[20] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal Loss for Dense Object Detection. In *Proceedings of ICCV*, pages 2980–2988.

[21] Liu, Z., Sun, M., Zhou, T., Huang, G., and Darrell, T. (2019). Rethinking the value of network pruning. *Proceedings of ICLR*.

[22] Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of CVPR*, pages 3431–3440.

[23] Maas, A. L., Hannun, A. Y., and Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In *Proceedings of WDLASL*.

[24] Maggiori, E., Tarabalka, Y., Charpiat, G., and Alliez, P. (2016). Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):645–657.

[25] Mun, J., Ha, S., and Lee, J. (2020). Automotive Radar Signal Interference Mitigation Using RNN with Self Attention. In *Proceedings of ICASSP*, pages 3802–3806.

[26] Mun, J., Kim, H., and Lee, J. (2018). A Deep Learning Approach for Automotive Radar Interference Mitigation. In *Proceedings of VTC-Fall*.

[27] Neemat, S., Krasnov, O., and Yarovoy, A. (2018). An interference mitigation technique for FMCW radar using beat-frequencies interpolation in the STFT domain. *IEEE Transactions on Microwave Theory and Techniques*, 67(3):1207–1220.

[28] Noh, H., Hong, S., and Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1520–1528.

[29] Rawat, W. and Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9):2352–2449.

[30] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on computer vision and pattern recognition*, pages 779–788.

[31] Ren, Y., Xu, H., Liu, B., and Li, X. (2020). Sea Ice and Open Water Classification of SAR Images Using a Deep Learning Model. In *Proceedings of IGARSS*, pages 3051–3054. IEEE.

[32] Ristea, N.-C., Anghel, A., and Datcu, M. (2023a). Sea Ice Segmentation From SAR Data by Convolutional Transformer Networks. In *Proceedings of IGARSS*, pages 168–171. IEEE.

[33] Ristea, N.-C., Anghel, A., Datcu, M., and Chapron, B. (2022). Guided Deep Learning by Subaperture Decomposition: Ocean Patterns from SAR Imagery. In *Proceedings of IGARSS*, pages 6825–6828. IEEE.

[34] Ristea, N.-C., Anghel, A., Datcu, M., and Chapron, B. (2023b). Guided Unsupervised Learning by Subaperture Decomposition for Ocean SAR Image Retrieval. *Proceedings of Transactions on Geoscience and Remote Sensing*.

[35] Ristea, N.-C., Anghel, A., and Ionescu, R. T. (2020). Fully Convolutional Neural Networks for Automotive Radar Interference Mitigation. In *Proceedings of VTC-Fall*.

[36] Ristea, N.-C., Anghel, A., and Ionescu, R. T. (2021a). Estimating the magnitude and phase of automotive radar signals under multiple interference sources with fully convolutional networks. *IEEE Access*, 9:153491–153507.

[37] Ristea, N.-C., Anghel, A., Ionescu, R. T., and Eldar, Y. C. (2021b). Automotive radar interference mitigation with unfolded robust pca based on residual overcomplete auto-encoder blocks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3209–3214.

[38] Ristea, N.-C., Miron, A.-I., Savencu, O., Georgescu, M.-I., Verga, N., Khan, F. S., and Ionescu, R. T. (2023c). Cytran: A cycle-consistent transformer with multi-level consistency for non-contrast to contrast ct translation. *Neurocomputing*, 538:126211.

[39] Rock, J., Roth, W., Toth, M., Meissner, P., and Pernkopf, F. (2021). Resource-Efficient Deep Neural Networks for Automotive Radar Interference Mitigation. *IEEE Journal of Selected Topics in Signal Processing*, 15(4):927–940.

[40] Rock, J., Toth, M., Messner, E., Meissner, P., and Pernkopf, F. (2019). Complex Signal Denoising and Interference Mitigation for Automotive Radar Using Convolutional Neural Networks. In *Proceedings of FUSION*.

[41] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of MICCAI*, pages 234–241. Springer.

[42] Sharma, N., Jain, V., and Mishra, A. (2018). An analysis of convolutional neural networks for image classification. *Procedia Computer Science*, 132:377–384.

[43] Stewart, R. H. (1988). Seasat: results of the mission. *Bulletin of the American Meteorological Society*, 69(12):1441–1447.

[44] Uysal, F. (2019). Synchronous and asynchronous radar interference mitigation. *IEEE Access*, 7:5846–5852.

[45] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Proceedings of NIPS*, pages 5998–6008.

[46] Wang, W. and Shen, J. (2017). Deep visual attention prediction. *IEEE Transactions on Image Processing*, 27(5):2368–2378.

[47] Woo, S., Park, J., Lee, J.-Y., and So Kweon, I. (2018). CBAM: Convolutional Block Attention Module. In *Proceedings of ECCV*, pages 3–19.

[48] Xu, Z. and Shi, Q. (2018). Interference Mitigation for Automotive Radar Using Orthogonal Noise Waveforms. *IEEE Geoscience and Remote Sensing Letters*, 15(1):137–141.