ENG 4001: Research Project
Faculty of Science, Engineering and Technology

Machine learning techniques for radar target recognition
Final Report

Project ID: 2022s2-EEE-Ng-UG-63280

Author	
Name	**ID**
Duc Thang Nguyen	a1831044

May 20, 2023

Nomenclature

ISAR – Inverse Synthetic Radar
ATR – Automatic Target Recognition
CNN – Convolutional neural network
CM – Confusion matrix

Executive Summary

Radar imaging has revolutionized target detection and recognition by providing precise and high-resolution information about objects, thereby facilitating the identification and tracking of targets in diverse and challenging environments. ISAR is a powerful technique that exploits the relative motion of targets and the radar system to generate high-resolution images of targets. Automatic Target Recognition (ATR) has vastly improved surveillance and reconnaissance operations in various sectors. Extensive research in the field has investigated the application of ISAR for ATR, yielding remarkable results. Convolutional neural networks (CNNs) utilizing deep learning algorithms have demonstrated outstanding results in the recognition and classification of targets.

The project intends to improve machine learning applications in ATR by examining the correlation between CNN complexity and classification accuracy. The initiative examines and verifies the performance of three well-known CNN architectures — Lenet5, AlexNet, and GoogleNet — using a simulated dataset. These networks span a spectrum of increasing complexity, allowing for the examination of whether more intricate structures improve the classification accuracy of targets. Two verification methods were developed: one to evaluate the performance of the three networks with fewer target features and the other to assess their overall performance. The project's outcomes disclose a clear correlation between network complexity and classification precision, with only minor performance differences between LeNet5 and AlexNet. Notably, the more complex GoogleNet performs better than both other networks. In addition, to consolidate the findings, a custom CNN was developed by combining the simplicity of LeNet5 with performance-enhancing features from other networks. Understanding the effect of CNN complexity on classification precision can guide the design and selection of suitable network architectures. The developed custom CNN paves the way for future research and development to resolve particular classification issues in ATR.

Overall, this project advances machine learning in ATR by shedding light on the correlation between CNN complexity and classification accuracy. The findings provide valuable guidance for optimizing network architectures and augmenting radar-based applications' target recognition capabilities.

Contents

1. Introduction

Radar target detection is a significant obstacle in radar signal processing, with wide-ranging applications in the military, surveillance, and aerospace industries. Radar target recognition has been greatly facilitated by advances in radar imaging techniques, which offer a variety of advantages, such as all-weather capability, long-range detection, and the ability to penetrate certain materials. In addition, advancements in machine learning algorithms have played a crucial role in improving radar target recognition by facilitating automatic target recognition (ATR). Utilizing algorithms and models that autonomously learn from radar data to identify and classify radar targets constitutes the application of machine learning techniques for radar target recognition. By automatically extracting meaningful patterns and features from radar data, this method enables accurate and efficient target recognition. Notably, one of the primary benefits of machine learning is its ability to effectively manage complex and high-dimensional radar data, which can present challenges for conventional methods. By training machine learning models with labelled data, they can acquire the ability to differentiate between various target classes and accurately predict unobserved data, including instances of those classes.

Numerous machine learning algorithms and architectures have been applied to radar target recognition. Support vector machines (SVM), k-nearest neighbors (KNN), and decision trees are some examples. Moreover, deep learning models such as convolutional neural networks (CNN) and recurrent neural networks (RNN) have shown remarkable success in radar target recognition tasks. These neural networks are capable of autonomously learning hierarchical representations from raw, unprocessed radar data, resulting in enhanced target classification performance.

1.1 Background

In recent years, machine learning has granted great attention from the community. The outstanding performance of the convolutional neural network (CNN) in optical image recognition [6] brings its potential to the radar community. Because of these developments, ATR is now better equipped to handle and analyze radar data, improving accuracy in detecting and classifying targets. The use of CNNs in automatic radar target recognition has widely been studied for multiple purposes. In areas such as detection [7]-[9], data augmentation [10], and classification [9], [11], different applications of CNNs are studied and verified with high accuracy and efficiency.

Inverse synthetic radar (ISAR) is a radar technique for imaging targets that relies on rotating motion. This makes it ideal for tracking and imaging moving targets since most radar systems have to encounter real-world targets undergoing such complicated movements. Most ISAR applications are for ATR since they contain rich information about the target, such as its size, type, and position. Deep-learning approaches have also been applied to classify ISAR images. Several studies have considered using this approach for target recognition [12]-[14]. This approach works well in extreme conditions with various signal-to-noise ratios (SNR). A more complex system using 3D-ISAR for point cloud formation to enrich the information of the target is performed in [15]. The tremendous growth of the ISAR community elicits demand for data for experiments when most measured datasets from real targets are not published due to security concerns, and the generation of a simulation ISAR dataset is a resource-intensive process. Therefore, the work of N. Pandey and S. S. Ram in [13] proposes a simulation framework to generate ISAR images that are verified by measurement and then publish their generated dataset for the community. This project leverages the dataset provided by N. Pandey and S. S. Ram [13] to undertake in-depth investigations in deep learning for radar target recognition.

1.2 Project aims and scope

The project aims to enhance the radar community's knowledge by validating a previous discovery using various approaches. Firstly, the project evaluates different well-known convolutional neural networks (CNNs) on a specific dataset, employing simplified, efficiency-verified methods for ISAR image classification as experimented in [13]. Secondly, by conducting experiments on the same dataset with fewer target features provided to the deep-learning models, the project aims to assess their practical applicability in real-world scenarios where limited target information may pose challenges. The selection of CNNs is based on their design complexity to investigate the relationship between classification accuracy and model infrastructure complexity. The insights gained from these approaches will be used to develop a new CNN that maintains a simple structure while achieving moderate classification accuracy. This newly developed CNN has the potential to conserve computational resources for ATR systems while maintaining target detection and classification performance.

The study consists of four stages: a research-based stage, two implementation stages, and a CNN development stage. The research-based stage involves a comprehensive analysis to gain a solid understanding of radar principles, radar imaging techniques, and the applications of CNNs in automatic radar target recognition (ATR). This stage aims to identify areas for development and make valuable contributions to the radar community. The literature review in this document presents the findings from this stage, including relevant works in the field.

In the second stage, the dataset is utilized to experiment with one deep-learning approach, exploring various aspects of the dataset. The experimental framework established in this stage is then used for all subsequent algorithms. The following stage involves conducting experiments with other algorithms using the same framework, enabling fair comparisons. Finally, the results from the previous stage are analyzed based on CNN performance and infrastructure to develop an optimized CNN. This optimized CNN is then evaluated using the same framework to validate the improvements achieved.

1.3 Technical objectives

The SMART approach is applied in defining five technical objectives to archive project aims. By this, each objective must be Specific and Measurable with a realistic goal that can be Archivable and Relevant to the project aims. Time is also a valuable resource, so Time-bound requirements for the objective must also be guaranteed. Five technical objectives are summarized:

i. **Research and systematically document fundamental knowledge of the field and related studies of ATR**

   The first objective is to build up the fundamental knowledge of radar principles and be aware of critical studies in the area. General information on radar imaging, ATR and deep-learning techniques is acquired, and technology selection is made and reported in the literature review.

ii. **CNNs selection**

The choice of Convolutional Neural Networks (CNNs) is crucial when investigating the relationship between CNN infrastructure and classification precision. To accomplish this objective effectively, selecting CNN models with increasing complexity is necessary. This allows for a thorough investigation into factors contributing to enhanced classification accuracy. Through this investigation, we can obtain valuable insights into how different CNN configurations, such as the number of layers, filter sizes, and activation functions, affect the network's ability to recognize and classify data patterns accurately.

iii. **Build up experiment plan**

Before any experiment is conducted, a plan is needed to navigate the investigation to archive the project aims and keep track of the experiment's progress. The plan must analyze the specific aim and frame the experiment to clarify the research directions step-by-step.

iv. **Data preprocessing**

To fulfil this objective, a comprehensive analysis of the general dataset is undertaken, leading to thorough descriptions and a profound comprehension of the dataset. Given that the project revolves around the deep learning-based classification of ISAR data generated in the frequency domain (spectrum of radar echoed signals amplitude), it becomes imperative to transform the data into a suitable image format that the network can process before training it for classification purposes. As a result, this objective holds the utmost significance in laying the groundwork for conducting experiments by ensuring the data is appropriately prepared for subsequent stages.

v. **Conduct experiments on the dataset by one CNN**

The initial CNN is implemented on the dataset in accordance with the experiment plan that was previously conducted. Each experiment outcome is meticulously recorded and documented, along with the corresponding observations made during the investigation. The algorithm's efficacy of the algorithm will be evaluated by analyzing the experiment's outcomes.

vi. **Survey of different CNN performances**

The previous objective is revisited but expanded to a larger scale, involving the application of various subsequent CNNs to the provided dataset. This process allows for a comparative evaluation of each algorithm's effectiveness on the Inverse Synthetic Aperture Radar (ISAR) images. The performance disparities between more complex CNN designs and simpler ones are also assessed.

vii. **Custom network development**

Ultimately, the project's final objective is to create a tailored Convolutional Neural Network (CNN) focusing on enhancing classification accuracy and simplicity in design, derived from the findings throughout the project's course. The choice of design will be left open-ended, pending the investigation results. This custom CNN might be an evolution of one of the analyzed CNNs, or it could represent an entirely new design if none of the previously tested designs is deemed suitable.

1.4 Document Overview

This document is the project's final report which outlines the project flow from the initial stage. Section 2 outlines the relevant research evidence and summarizes potential research gaps. The fundamental theories are provided in section 3 to prepare for the methodology detailed in section 4, and experiment results are reported in section 5 of the document. Section 6 illustrates the path for further research to extend the project before the document's brief conclusion is summarized in section 7.

2. Literature review

ATR's success is due to numerous vital reasons. ATR systems use advanced classification algorithms, frequently based on machine learning or deep learning, to capture target properties and recognize them accurately. High-quality training data from varied target classes and environmental conditions improve performance. Adaptive and robust methods handle target appearance and context changes. Sensor fusion provides complete target comprehension. Real-time processing allows quick decision-making, while ongoing refinement and adaption perfect the system. ATR systems in defense, surveillance, robotics, and autonomy use these elements to recognize targets accurately and reliably.

Research advancements in radar imaging have significantly improved ATR. Radar imaging techniques with a high resolution, such as SAR [1],[2] and ISAR [3], [4], provide detailed target information for enhanced classification and recognition. Integrating advanced signal processing algorithms and machine learning improves target detection and tracking. These advancements have substantially enhanced the performance and situational awareness of ATRs.

Automatic target recognition has been widely studied for decades. Traditionally, target recognition is based on geometric features, such as length, and width, derived from the high-resolution range profile of the target. In recent years, applications of deep-learning techniques in target detection [7]-[9], classification [12]-[15] and data augmentation [10] have made tremendous progress. The reliable performance of deep learning in ATR has widened its community within the radar industry. For the above reasons, the analysis of potential deep-learning applications in general and radar imaging techniques has been conducted by reviewing relevant literature.

2.1 Radar imaging

2.1.1 Synthetic Aperture Radar (SAR) Imaging

Due to its all-weather, day-and-night imaging capabilities, Synthetic Aperture Radar (SAR) imaging has become an indispensable instrument in remote sensing. Understanding and advancement of this technology have been substantially aided by two fundamental works: "A Tutorial on synthetic aperture radar" by Moreira et al. (2013) [1] and "Understanding Synthetic-Aperture Radar Images" by C J Oliver (1989) [2].

In their tutorial, Moreira et al. [1] explain the fundamentals, methods, and applications of SAR, providing a comprehensive overview of the field. They trace the transition of SAR from being technology-driven to user-demand driven, emphasizing the role of technological advances such as digital beamforming, Multiple-Input, Multiple-Output (MIMO), and bi- and multi-static configurations in this shift. In addition, they present a vision for the future of SAR technology in which a constellation of radar satellites provides systematic, high-resolution monitoring of the Earth's surface.

In contrast, Oliver's [2] work investigates the physical foundations of SAR imaging. It provides a nuanced comprehension of the imaging function under all operating conditions and discusses at length the issues that arise as a result of the varying aircraft dynamics. Oliver addresses common SAR challenges such as image blurring and distortion by proposing a self-consistent scheme for producing focused images that incorporates motion compensation from inertial navigation systems, data-dependent autofocus, and phase correction techniques. This work has had a substantial impact on the evolution of techniques for enhancing SAR image quality.

These two works provide a comprehensive overview of the evolution and contemporary state of SAR technology when viewed together. Oliver's [2] research addresses practical challenges in SAR imaging and provides solutions. In contrast, Moreira et al.'s [1] tutorial traces the evolution of SAR, describes its current state, and suggests future directions. However, while these works have made substantial contributions to the field, they do not thoroughly address potential obstacles and limitations of SAR technology, such as high costs, technical complexities, and data management and analysis challenges. Future research could benefit from a more balanced perspective that thoroughly addresses these issues.

2.1.2 Inver Synthetic Aperture Radar (ISAR) Imaging

Since its inception in the late 1950s, Inverse Synthetic Aperture Radar (ISAR) imaging has undergone significant development and transformation. Among the numerous studies and algorithms that have contributed to the field, Vehmas and Neuberger (2021) [3] and Xu et al. (2011) [4] stand out for their detailed insights and innovative methodologies.

[3] Vehmas and Neuberger (2021) provide a comprehensive historical perspective and state-of-the-art review of ISAR imaging. They trace the development of ISAR from the 1960s experiments at Willow Run Laboratories to the present day. Notably, they describe advancements in imaging ground-moving targets, the introduction of motion estimation and compensation techniques, and the explosion of ISAR applications. The authors also emphasize the difficulty posed by non-linearly moving targets, which has led to the development of time-frequency motion compensation and imaging techniques. The study also discusses the computational cost of the one-step approach to translational motion compensation. This method has not been widely adopted despite its robustness due to its high computational costs, although this may change as image resolution improves [3].

Innovative algorithms, such as the Bayesian ISAR imaging strategy presented by Xu et al. (2011) [4], complement the development and expansion of ISAR imaging. This study exploits the expanding interest in sparse signal representation in radar signal processing and applies Bayesian formalism to improve ISAR imaging. By considering phase errors as model errors and utilizing sparsity-driven optimization, the proposed Bayesian ISAR imaging algorithm exhibits high denoising capability and high image generation precision [4].

Experiments conducted by Xu et al. [4] demonstrate the algorithm's robustness under various conditions, including various phase errors and signal-to-noise ratios (SNR). Their comparison with conventional autofocusing techniques such as Weighted Least Squares (WLS) and Minimum Entropy Algorithm (MEA) demonstrates the superior denoising and enhanced recovery precision of their proposed method, particularly under low SNR conditions.

Vehmas and Neuberger (2021) outline prospective future directions for ISAR imaging, such as the use of Compressive Sensing (CS) and Machine Learning (ML) techniques [3]. The rapid development of inexpensive high-resolution radar systems has prompted a shift towards spatially diverse systems with higher spatial resolution.

In summary, the development of ISAR imaging illustrates a journey of constant innovation and refinement. ISAR imaging has been and will continue to be a crucial field in radar technology, from the initial emphasis on ground-moving targets and the establishment of motion compensation techniques to the most recent advancements using Bayesian formalism and the prospective incorporation of CS and ML techniques. The contributions of Vehmas and Neuberger (2021) and Xu et al. (2011) demonstrate the ingenuity and commitment of researchers to stretch the limits of what is possible in ISAR imaging.

2.2 Deep learning approaches

ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is a challenge to evaluate algorithms' performance in object detection and image classification for optical images. In recent years, deep learning algorithms, specifically CNN, dominated the challenge with high performance, such as Alex net for first place in 2012, Google net and Res Net respectively won the 2014 and 2015 challenges.

2.2.1 GoogleNet

"Going Deeper with Convolutions" by Szegedy et al. (2015) [16] introduces GoogLeNet, a deep learning architecture that uses the "Inception Module" to drastically reduce deep neural network computational resource requirements. GoogLeNet, inspired by 'Network-in-Network' (NiN), has 22 layers (27 if aggregating layers are added). This architecture is based on the Inception Module, which introduces sparser architecture within a convolutional layer. After applying various filter sizes to each layer, the Inception Module concatenates the results. In Fig. 1, 1x1 convolutions are utilized to compute reductions before the more expensive 3x3 and 5x5 convolutions. The network included nine Inception modules and employed average pooling rather than fully linked layers for classification.

Szegedy and his team believe this strategy approximates the most optimal sparse structure with a significant quality gain and minimal processing needs compared to shallower and narrower designs. The model won the 2014 ILSVRC classification and detection challenges with a top-five error rate of 6.67 per cent.

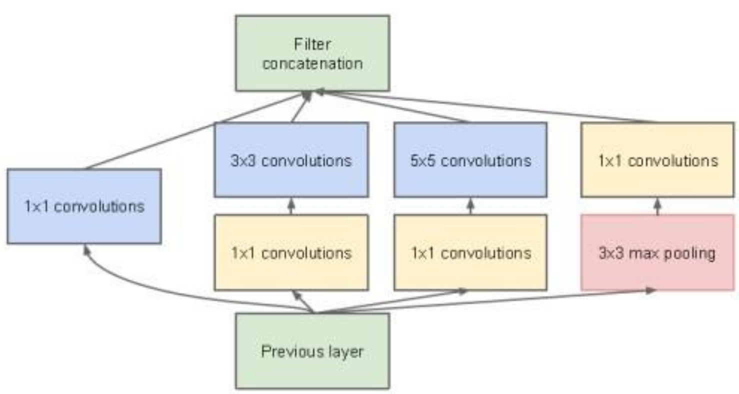

Figure 1 Inception module with dimensionality reduction [16]

2.2.2 AlexNet

AlexNet algorithm is presented in [17] with the classification error of 16.4% in ILSVRC-2012. The model uses the dataset from ILSVRC- 2010 contest for testing and validating the model, and the error rate of top-1 and top-5 in ILSVRC- 2010 are archived before being submitted for ILSVRC2012. The network consists of eight weighted layers, five convolutional layers, three fully connected layers, a final 1000-way SoftMax, and three max-pooling layers with a total of 60 million parameters. The applications of ReLu non-linearity are available in the output of every convolutional and fully connected layer. The first convolutional layer includes 96 kernels of size 11x11 with a stride of 4 pixels. The output of the first convolutional layer is filtered with 256 kernels of size 5x5, which is the architecture of the second convolutional layer. The third, fourth and fifth convolutional have the same 3x3 size kernels with the number of kernels 384, 384 and 256, respectively. Figure 2 below shows the architecture of the Alex Net CNN.

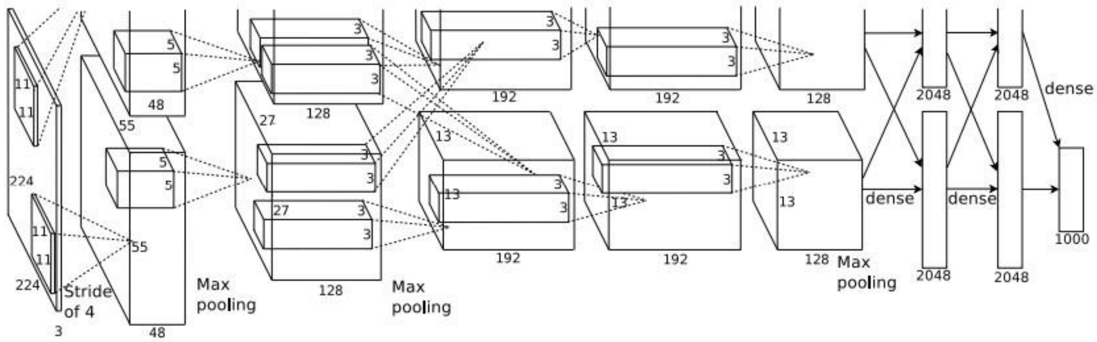

Figure 2 Alex Net architecture [17]

Krizhevsky et al. [17] paid special attention to the common issue of overfitting in models with a large number of parameters. They implemented measures to mitigate overfitting and effectively manage the network's complexity. This practical consideration was not only essential to the performance of their model, but it also established a precedent for future work in the field by underscoring the need for a balance between model complexity and generalization capability. The paper also includes qualitative evaluations of the model's performance, including incisive visualizations of the network's learned kernels and demonstrations of the model's ability to recognize objects under a variety of conditions. Not only do these visualizations make the results more interpretable, but they also provide a deeper understanding of how the network learns and recognizes patterns.

The network depth is also an important factor that is underlined in the paper when it is justified that the removal of any convolutional layer leads to degradation in classification performance (reduce 2% for each layer removal). This crucial result demonstrated the efficacy of deeper neural network topologies and spurred research towards deeper and more complicated models, which is now a cornerstone of deep learning. However, the contributions of each layer to the overall performance have not been clarified and whether just simply adding a convolutional layer to the network can improve the performance.

2.2.3 ResNet

Microsoft's Residual Networks (ResNet) [18] archived an incredible error rate of 3.6% in ILSVRC2015 competition. Resnet is an extremely deep network with 152 layers depth which is 8 times deeper than VGG Net [19] but lower complexity than VGG net. The model is evaluated the dataset that contains 1000 classes with 1.28 million images used for training and 50 thousand images for validating and finally tested with 100 thousand test images. Fig.3 below depicts a building block of residual learning. The outputs of the shortcut connections are simply added to the outputs of the stacked layers after identity mapping is performed.

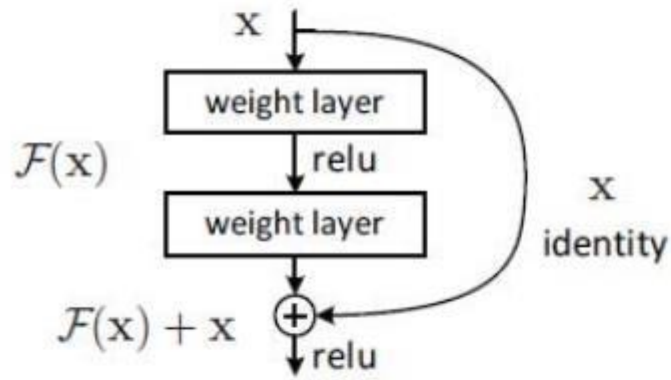

Figure 3 Residual learning building block [18]

The plain network of Res Net is inspired by the philosophy of VGG Net. Most of the convolutional network have kernels size 3x3 ended by a global average pooling layers and 1000-way fully connected layer with SoftMax. A residual network in ResNet also has same architecture but a short cut connection between each pair of 3x3 kernels is added. They employed a modified building block as a bottleneck design for the deeper nets as show in Fig 4.

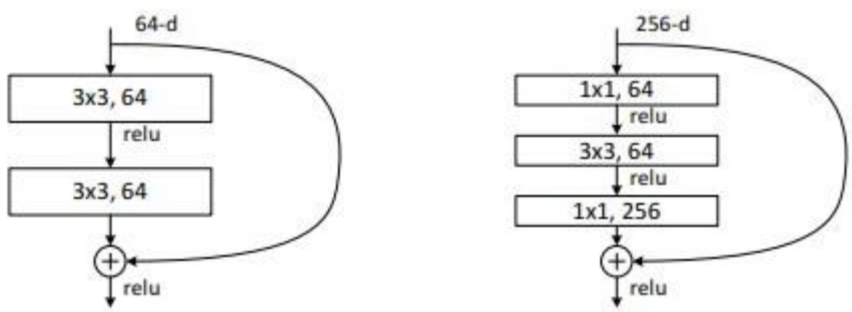

Figure 4 A deeper residual function [10]. Left: a building block for ResNet-34.

2.2.2 Lenet-5

Yann LeCun, Leon Bottou, Yosuha Bengio, and Patrick Haffner devised LeNet-5, a seminal convolutional neural network (CNN) architecture, in 1998. As one of the earliest CNNs, it was intended to recognize handwritten and machine-printed numbers. The model was effectively implemented in practical systems in the United States for reading postal codes, numbers, etc. on mail. LeNet-5 marked a significant milestone in the field of deep learning by laying the groundwork for the design of contemporary convolutional neural networks.

Seven layers comprise LeNet-5, including two convolutional layers, two average pooling layers, and three fully connected layers. The first layer accepts 32x32 pixel grayscale input images. Each convolutional layer uses a 5x5 convolutional kernel to generate multiple feature maps. The pooling layers, also known as subsampling layers, are intended to gradually reduce the spatial scale, number of parameters, and computational complexity to prevent overfitting. The final layers of the network are fully connected layers, including a unique 'radial basis function' layer that employs a softmax activation function to generate the probabilistic output of the network. Fig 5 below shows the flow in Lenet5 design.

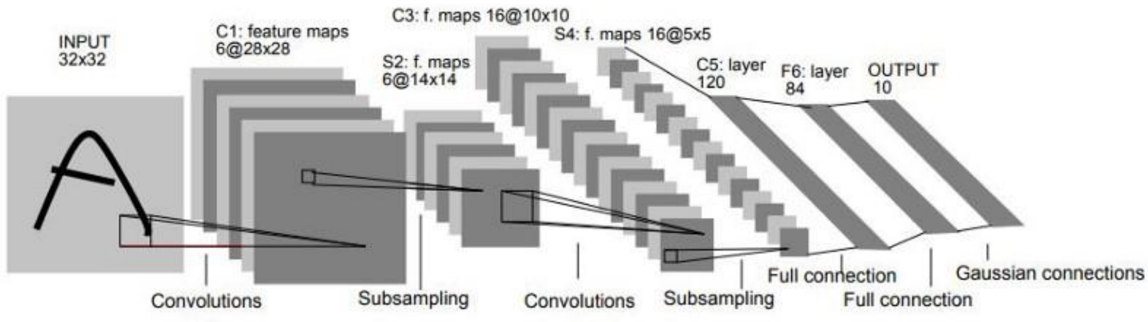

Figure 5. Lenet-5 Architecture

2.3 Deep learning in automatic target recognition

Recent research in the field of automatic target recognition (ATR) has shown a rise in the use of deep learning techniques to address problems in diverse domains, such as maritime surveillance, Inverse Synthetic Aperture Radar (ISAR) image recognition, automotive target classification, and ISAR ship imaging. This exhaustive review concentrates on four influential studies that have significantly contributed to the comprehension and development of deep learning techniques in ATR.

Liu et al. [11] propose a multi-task convolutional neural network (CNN) for maritime target detection, a recently gaining application of interest. This study demonstrates an innovative method of utilizing cross-layer connections, which substantially improves the CNN's ability to represent data. The multi-task CNN model developed by the authors outperforms other detection models, such as Faster-RCNN, R-FCN, and MNC, demonstrating the potential of multi-task models in ATR.

In the context of ISAR image recognition, Xue et al. [12] present a novel approach in their paper titled "Complex ISAR Target Recognition Using Deep Adaptive Learning." They combine deep learning methods with handcrafted techniques, reducing the data dependence of deep learning models while capitalizing on the interpretability advantages of handcrafted techniques. Their DD RoI (Deformable Detection Region of Interest) pooling method brings a high degree of flexibility to ISAR target recognition, addressing numerous real-world challenges effectively.

Using ISAR images, Pandey and Ram [13] examine the classification of various automotive targets. This study presents a comprehensive simulation methodology that models the radar signals dispersed by a variety of automotive targets. For the classification task, the authors employ both traditional machine learning techniques and deep learning algorithms effectively. Their research demonstrates that deep learning algorithms are superior in classifying automotive targets, even under challenging conditions.

Musman et al.'s [14] study "Automatic recognition of ISAR ship images" is a seminal contribution to the field of ISAR imaging. Using databases of ship characteristics, this study proposes automated techniques that replicate the computer-assisted human recognition process. The authors present multi-frame processing techniques that enhance feature extraction and target classification.

In summary, these studies demonstrate the capability and adaptability of deep learning methodologies in various automated target recognition tasks. They demonstrate how deep learning models can be supplemented with handcrafted techniques, propose innovative multi-task models, and highlight the significance of multi-frame processing techniques. Future research should focus on further refining the integration mechanisms of deep learning and handcrafted features, expanding the availability of realistic and diverse training data, and developing more advanced models for robust target classification under varying conditions, despite the substantial progress made in the field.

2.4 Summary and potential gaps

The literature review emphasizes the significance of deep learning techniques in Automatic Target Recognition (ATR) and advances in radar imaging, with a particular emphasis on Inverse Synthetic Aperture Radar (ISAR) imaging. Vehmas and Neuberger provide a thorough overview of ISAR imaging, discussing its evolution, motion compensation techniques, and expanding applications. Xu et al. present a Bayesian ISAR imaging algorithm with high denoising capability and precise image generation. These studies highlight the potential of ISAR images for investigating the relationship between the infrastructure of Convolutional Neural Networks (CNNs) and the classification accuracy of radar images.

However, the literature identifies two significant omissions. The relationship between CNN's infrastructure and its effectiveness in radar image classification has not been investigated. The vast majority of deep learning techniques in ATR have predominantly focused on algorithms developed for optical images, ignoring the distinctive challenges and characteristics of radar images. This deficiency highlights the need to investigate the efficacy of CNNs for radar image classification tasks, ensuring that CNN architectures are tailored to manage the complexities of radar data.

Second, CNN models developed specifically for radar image classification problems are lacking. CNNs have exhibited remarkable performance in numerous domains, including the classification of optical images, but their application to radar images has been limited. To address this deficiency, it is essential to develop CNN architectures that are specifically designed to handle radar data, taking into consideration the unique characteristics, noise sources, and challenges posed by radar images.

ISAR imaging's unique benefits justify its selection over Synthetic Aperture Radar (SAR) for research involving these gaps. ISAR provides high-resolution images of moving targets, capturing information such as the geometry, rotation, and internal structure of the target. This level of specificity is necessary for precise target recognition and classification. By utilizing ISAR images, researchers can gain a better comprehension of how CNN infrastructures process and interpret radar data, leading to the development of CNN architectures specialized for image classification tasks involving radar data. Moreover, ISAR images provide a more suitable platform for examining the relationship between CNN infrastructure and radar image classification accuracy, given the specific challenges and complexities of radar imagery.

3. Theories

3.1 ISAR imaging

In Synthetic Aperture Radar (SAR) imaging, the radar system is carried by an airborne or spaceborne platform that follows a predetermined path. The radar system emits radio wave pulses towards the target scene, which are reflected and received by the system as it advances. SAR can generate high-resolution images regardless of the weather or lighting conditions by analyzing the received signals and their respective travel times. This has multiple applications, including terrain mapping, surveillance, weather prediction, and environmental monitoring.

Inverse Synthetic Aperture Radar (ISAR), on the other hand, is predominantly used to image noncooperative moving targets, such as aircraft or ships, as opposed to static ground scenes. ISAR uses the rotation or motion of the target to generate the synthetic aperture, rather than the movement of the radar as in SAR. This allows for two-dimensional imaging of the target with high resolution. Here are some theoretical ISAR considerations:

First is the Doppler Effect theory. ISAR imaging is fundamentally based on the Doppler Effect, which describes the change in frequency and wavelength of a wave when an observer moves relative to the wave source. In ISAR, the rotation or other relative motion of the target generates a Doppler frequency shift in the returned radar signals. Due to their differing relative velocities, each point on the target will have its own Doppler shift. ISAR separates various scatterers on the target and enables high-resolution, cross-range imaging by classifying received echoes based on their Doppler frequencies.

Another theoretical consideration is algorithms for Fourier Transform and Range-Doppler. After data collection, ISAR data processing significantly depends on the Fourier Transform, a mathematical tool that decomposes a waveform into its sinusoids of various frequencies. Using Fourier Transform, frequency information is extracted from radar emissions. ISAR image processing typically employs the Range-Doppler algorithm, which entails two Fourier Transforms. The first Fourier Transform is used to determine the radial distance (range) of scatterers from the radar, while the second transform is used to separate scatterers based on their Doppler shifts, thereby providing cross-range resolution. The target is rendered as a two-dimensional image.

A final consideration is Integration coherence and image formation. ISAR imaging relies heavily on coherence as well. Multiple radar pulses are combined to improve the signal-to-noise ratio (SNR) in this instance. Due to the extended range of radar imaging and the small dimensions of typical radar scatterers, the target's reflected radar echoes are typically quite weak. After coherent integration, a two-dimensional image of the target is generated, with one axis representing the radial distance of the scatterers (range) and the other their Doppler shifts (cross-range). Each point's intensity on the image represents the radar reflectivity of the corresponding target point.

3.2 Deep learning

Deep learning is a subset of artificial intelligence and machine learning that employs artificial neural networks for data analysis and interpretation. Using neural networks, this method attempts to simulate the functionality of the human brain, albeit on a much smaller scale. Each neuron in an artificial neural network (ANN) receives an input, processes it based on internal parameters and a specific set of principles known as the activation function, and produces an output. These neurons are interconnected, and their synapses bear weights that are adjusted throughout the learning process.

Deep Neural Networks (DNNs), the primary architecture employed in deep learning, are ANNs with multiple hidden layers placed between the input and output layers. These layers are responsible for extracting increasingly complex features from the input data. In an image recognition task, for instance, the initial layer of a deep neural network may be used to identify basic features like edges. The subsequent layer could then detect more intricate shapes, with each successive layer distinguishing increasingly intricate objects and patterns. Consequently, deep learning enables the modelling of complex data abstractions via a hierarchical layer structure.

4. Methodology

This section outlines the experimental process used during the project with three stages. Firstly, we delve into a thorough exploration of the dataset to establish an initial comprehension of it and to identify how it will be scrutinized in subsequent experiments. Following this, we preprocess the dataset in preparation for model training, as its current state isn't suitable for direct model input. The third step involves the formulation of an experimental plan to guarantee consistent alignment with project objectives. In the final stage, we meticulously detail the experimental procedures and result-reporting methods. This provides insight into the nature of the project experiments and the evaluation criteria for the results obtained.

4.1 Dataset description

In this work, we make use of the simulation dataset released by [13]. The dataset consists of over 30000 realistic ISAR images validated with measurement data provided by Texas Instrument's AWR 1843 77GHz automotive radar. The scatterers of five commonly found vehicles, a bicycle, an Autorickshaw, a truck, a mid-size car, and a full-size car, are simulated. All vehicles are assumed to undergo four types of traffic trajectories at an intersection. The ground is aligned with the XY plane of the system with the axis aligned with lanes coming from the North (N) – South (S) direction and West (W) – EAST (E) direction, respectively, as shown in Fig 1. The height is along the Z-axis, and an ego radar is assumed to be held at the point (0, 0,0.5) m along the south road. 16 possible trajectories could be performed in this intersection (shown in Fig 7) with four left turns, four right turns, four U-turns, and four straight paths. All trajectories are simulated at a specified speed with different clutter conditions. The noise and clutter models are presented in [13] with the ground-based clutter in the range dimension and wind-based clutter in the Doppler dimension of the ISAR images realistically simulated. ISAR images of all trajectories are generated with 5 different signal-to-noise ratios (SNR) varying from -5dB to 10dB and 4 different wind velocities, 2.5 Meter Per Second (MPs), 5 MPs, 7.5 MPs, and 10 MPs, for Doppler clutter simulation.

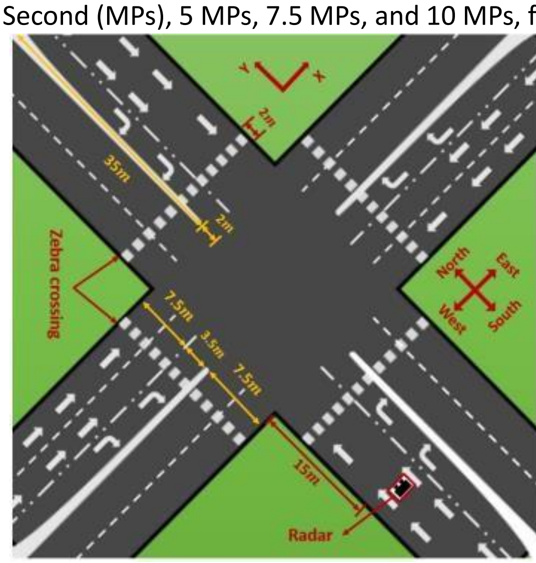

Figure 6 Road geometry of the intersection

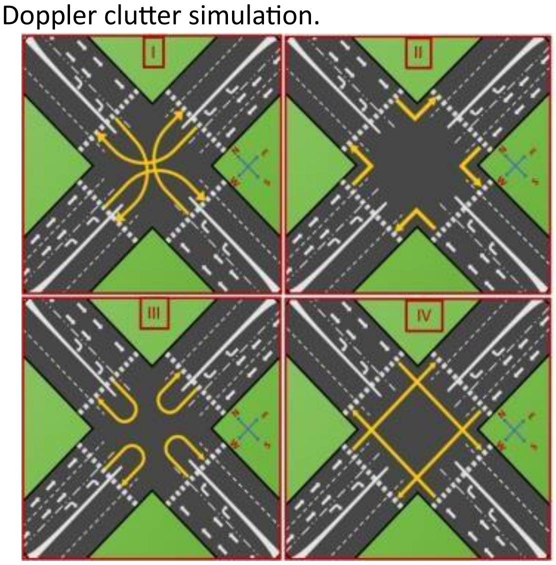

Figure 7 Trajectories undertaken by the automotive target in a four-way junction - (i) Right turn, (ii) Left turn, (iii) U-turn and (iv) Straight through.

4.2 Data preprocessing

The ISAR dataset provided by [13] is the targets' complex range and cross-range profile, which cannot be directly fed into the deep-learning model. The model requires the data to be in image format before the classification can be accomplished; therefore, the conversion step is needed. The ISAR data is converted to the 8-bit greyscale image with the decibel dynamic range before being rescaled from 0 to 255 to fit the 8-bit greyscale format, with each pixel being the normalized dB scale magnitude of the aligned range - cross-range profile following the formula below:

$$P_i = 255 * \frac{20 * \log(|x|) - \min(P_i)}{\max(P_i) - \min(P_i)}$$

P_i : the value of each pixel

$|x|$: the magnitude of each ISAR complex resolution cell

4.3 Experiment plan

The experimental plan is meticulously devised to align with the project's objectives. The focus is on discerning the relationship between the structure of Convolutional Neural Networks (CNNs) and the effectiveness of radar image classification. The initial phase involves selecting CNNs of varying design complexity, with the final choices being Lenet, AlexNet, and Googlenet, known for their incremental structural complexity.

Subsequently, the correct partitioning of the dataset for training, validation, and testing the model is crucial to provide sufficient data for training and validation while preserving adequate data for model evaluation. We're adopting the dataset from [13], so the splitting ratio used there – 70% for training, 10% for validation, and 20% for testing – is deemed most suitable.

The next step is to scrutinize overall classification performance. This helps understand the correlation between network complexity and classification accuracy and also validates the research presented in [13] by applying similar methodologies on the same dataset.

After that, a more complex problem is posed, simulating a scenario where a single motion type from the targets is known, but the classification is needed for targets undergoing varied motions – turning left, turning right, going straight, or executing a U-turn. Models trained with one motion type are used to classify each different motion type, testing the ability to recognize targets from unfamiliar perspectives.

The results from this phase are then analyzed and employed to create a new CNN, striving for improved ISAR image classification accuracy, while maintaining design simplicity. The efficacy of the new CNN is verified by reiterating the same experimental procedure applied to other networks.

In essence, the steps in the experimental plan are as follows:

i. Selection of suitable CNNs for the experiment.
ii. Defining the dataset's splitting proportion for training, validation, and model testing.
iii. Execution of classification on the entire dataset to assess overall performance and confirm the findings in [13].
iv. Training models with a single target motion type, followed by classification of targets in each distinct motion.
v. Analysis of results to construct a new CNN.
vi. Evaluation of the new CNN following the established procedure.

5. Results

In this section, three convolutional neural networks (CNN), namely Lenet5, Alex Net and Google Net are examined to evaluate and compare their performances on the same dataset provided by [13]. All three CNNs are applied to classify ISAR images of five types of vehicles: an autorickshaw, a bicycle, a mid-size car, a full-size car and a truck.

All CNNs are trained and evaluated using the same simulation framework to assure a fair comparison of their outputs. The framework consists of two classification problems, one to examine the overall performance of each network, in which the network is trained with ISAR images of all vehicles undergoing all trajectories (turning left, turning right, going straight, and making U-turns) and then validated and tested with other images from the dataset, with 70%, 10%, and 20% of the dataset used for training, validating, and testing, respectively. Test the cross-motion classification performance of CNNs as a second classification issue. At this stage, each CNN is only trained with ISAR images of vehicles undergoing a single type of trajectory and is evaluated using images of vehicles executing all trajectories. For instance, a CNN is trained with ISAR images of vehicles turning left; it must then classify these vehicles as turning right or remaining straight. The objective is to evaluate the performance of each network when fewer features of the targets are supplied, as well as the network's ability to accurately recognize the target from a different angle than that for which it was trained.

5.1 Overall Performances

a. GoogleNet

In this setting, GoogleNet is trained with 70% of the given dataset and used to classify 20% of the remaining images by applying the simulation framework. The result is that Google Net accurately classifies 97.38% of the test samples in the test dataset, consistent with the results of [13] with the same classification setting (over 90%).

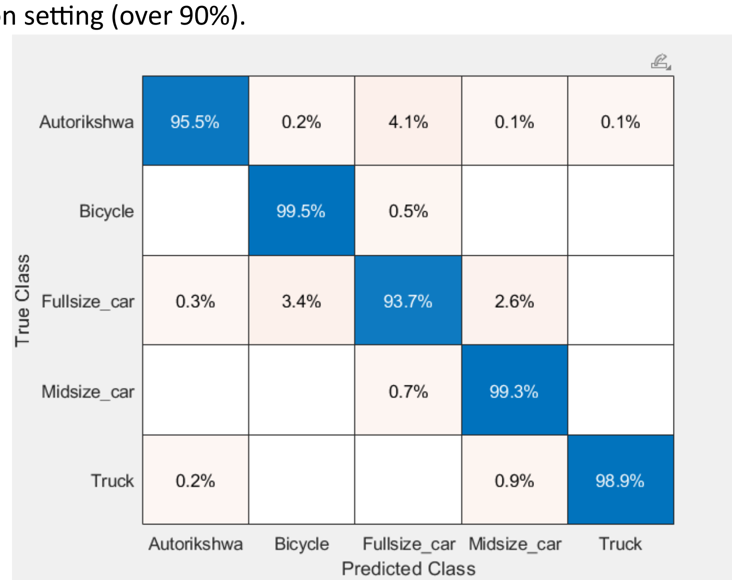

Figure 8 Confusion matrix – GoogleNet overall classification

GoogleNet performs exceptionally well in classifying the five vehicle types under consideration. Bicycles have the lowest classification error of all vehicle types, presumably due to their unique size compared to other vehicle types. Occasionally, however, there are misclassifications between the two car classes and between the auto-rickshaw and truck classes. The most prevalent misclassification occurs when 4.1% of auto-rickshaw images are incorrectly classified as full-size cars. These findings emphasize the strengths and limitations of the GoogleNet architecture for classifying vehicle images and could inform the design and development of other models or approaches for image classification tasks with similar requirements.

b. AlexNet

Following the same procedure as Google Net, Alex Net correctly predicts 91.24% of the test images and takes less than 27 times the training duration of Google Net. This result is even better than what is stated in [13] for Alex Net (Over 80%)

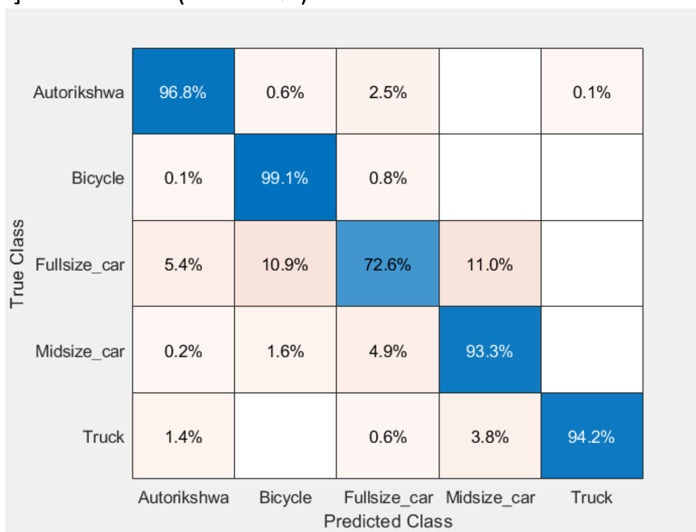

Figure 9 Confusion matrix – AlexNet overall classification

AlexNet produces similar outcomes as GoogleNet with lower accuracy. Bicycles continue to have the lowest misclassification rate, while both kinds of cars are frequently misclassified with one another. In addition, the model's classification error is most significant when 11% of Full-size car images are misclassified as Mid-size cars. In addition, AlexNet has a remarkable error rate, identifying bicycles in 10.9% of full-size images. Overall, AlexNet performs marginally worse than GoogleNet, but its implementation requires substantially fewer resources.

c. LeNet

As anticipated, LeNet performs comparatively worse than GoogleNet and AlexNet when trained under the same conditions, with a classification accuracy of only 80.62 %. However, the patterns of misclassification between vehicles remain consistent, with two categories of cars frequently misclassified with each other at a high rate of 12 percent, and Trucks and Autorickshaws occasionally misclassified as cars. Similar to GoogleNet and AlexNet, 19.43% of Full-size automobile images are incorrectly classified as Bicycles by LeNet.

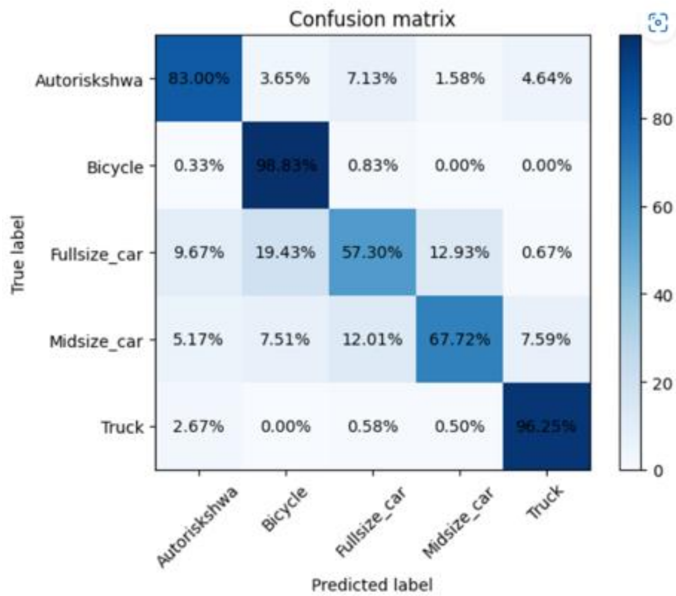

Figure 10 Confusion matrix – Lenet overall classification

Discussion

The analysis of the CNN architectures' overall performance on the vehicle image dataset revealed interesting outcomes. GoogleNet showed exceptional accuracy (97.38%) but experienced minor issues distinguishing between different car classes and between auto-rickshaws and trucks. AlexNet, while exhibiting slightly lower accuracy (91.24%), proved to be resource-efficient by accomplishing its training in significantly less time compared to GoogleNet. Yet, it encountered similar challenges to GoogleNet in classifying different types of cars and still had a notably low error rate like GoogleNet with bicycle identification. LeNet's performance, while predictably lower (80.62%), showed consistency in the types of misclassifications seen with GoogleNet and AlexNet, highlighting persistent challenges in distinguishing between certain vehicle types. These results provide valuable insights into the strengths and limitations of these models, contributing to more informed decisions in selecting or developing CNN architectures for specific image classification tasks in the future. Table 1 below shows the overall performances of the three CNNs.

Model	Accuracy
Google net	97.38%
Alex net	91.24%
Lenet5	80.94%

Table 1. Overall performances of 3 CNNs

5.2 Cross-motion performances

Due to the large number of classifications conducted resulting in an abundance of confusion matrices, the F1 score has been used to evaluate model performance in this section. The F1 score, a harmonic mean of precision and recall, is especially useful in situations involving unbalanced data or where misclassifications have significant consequences. It provides a comprehensive comprehension of a model's performance by encapsulating the balance between minimizing both false positives and false negatives. It is defined as:

$$F1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

Precision: the number of true positive results divided by the number of all positive results
Recall: the number of true positive results divided by the number of all samples that should have been identified as positive

a. GoogleNet

Even when trained on a single type of motion, Google Net demonstrated notable performance in classifying vehicles undergoing various motions. The results presented in Table 2 show that the F1 score model's classification varied depending on the direction of motion relative to the training data. For instance, models trained on images of vehicles turning left and right exhibited the best performance when classifying vehicles moving straight. Similarly, the best classification outcomes for straight and U-turns were achieved when the model was trained on images of the same motion. The complexity of the motion trajectory was found to be another factor affecting the classification results. Specifically, the model trained on simple trajectories of vehicles moving straight exhibited the lowest F1 score. Nevertheless, although turning left and turning right motions are thought the same the model trained with a Left turn has the best classification performance.

Train\Test	Left turn	Right turn	Straight	U-turn	Average
Left turn	0.89	0.83	0.92	0.89	0.88
Right turn	0.83	0.87	0.90	0.83	0.86
Straight	0.83	0.80	0.91	0.71	0.81
U-turn	0.86	0.87	0.84	0.88	0.86
Average	0.85	0.84	0.89	0.83	0.85

Table 2. F1 scores of GoogleNet cross-motion check models

b. AlexNet

Alex Net's cross-motion classification accuracy is lower than GoogleNet but still higher than Lenet. However, the model trained with Left turn trajectories still has the most outstanding performance and a significantly higher accuracy rate than models trained with other trajectories which is also seen by GoogleNet. Moreover, another similar patterns between these two CNNs is that model trained with images of vehicles going straight has the worst performance. This suggests that Alex Net's performance in classifying vehicles is widely influenced by the type of motion it is trained on. The model's accuracy may vary significantly depending on the type of motion. Classification of ISAR images of vehicles going straight seems to be the simplest task since most of the models of AlexNet as well as GoogleNEt have highest performance when tested with it.

Train\Test	Left turn	Right turn	Straight	U-turn	Average
Left turn	0.78	0.80	0.85	0.80	0.81
Right turn	0.71	0.81	0.82	0.73	0.77
Straight	0.67	0.71	0.78	0.57	0.68
U-turn	0.75	0.77	0.80	0.79	0.78
Average	0.73	0.77	0.81	0.72	0.76

Table 3. F1 scores of AlexNet cross-motion check models

c. Lenet

As expected, Lenet has the lowest cross-motion F1 score in the 3 CNNs. However, Lenet5 has a slightly greater error rate. In addition, Lenet5 performs optimally when trained with a straight trajectory approach, as opposed to the Left turn method favored by Google Net and AlexNet.

Train\Test	Left turn	Right turn	Straight	U-turn	Average
Left turn	0.56	0.58	0.57	0.53	0.56
Right turn	0.61	0.53	0.44	0.48	0.52
Straight	0.57	0.63	0.56	0.51	0.57
U-turn	0.66	0.53	0.48	0.52	0.55
Average	0.60	0.57	0.51	0.51	0.55

Table 4. F1 scores of LeNet cross-motion check models

5.3 Custom-developed CNN

The previous experiments conclusively demonstrate a correlation between the intricacy of a Convolutional Neural Network (CNN) and its capability to accurately classify radar data. It's observed that the more layered the network, the superior its performance. Nonetheless, we are yet to uncover how each layer within these CNNs influences the classification efficacy on Inverse Synthetic Aperture Radar (ISAR) images. Therefore, the decision to develop a custom CNN was influenced by these findings, beginning with the incorporation of additional layers to the simplest CNN architecture, Lenet5. Given that GoogleNet has shown the best overall and cross-motion ISAR image classification performances, it would be a rational choice to enhance the custom CNN's performance by integrating features from GoogleNet. Notably, GoogleNet's unique characteristic is its Inception module, hence the first development choice for the custom CNN involves the embedding of Inception layers.

Let's revisit the Lenet infrastructure, it contains:

i. Input Layer: The input for LeNet-5 is a 32x32 grayscale image.

ii. Convolution Layer C1: This is the first convolutional layer with six feature maps. Each unit in each feature map is connected to a 5x5 neighborhood in the input. This results in feature maps of size 28x28.

iii. Subsampling Layer S2 (also called Pooling Layer): This layer follows the first convolutional layer and contains six feature maps of size 14x14. It is a down-sampling layer that reduces the dimensionality of the previous layer, reducing the computational complexity and helping with the problem of overfitting. This is usually achieved through average pooling or max pooling. In the case of LeNet-5, average pooling was used.

iv. Convolution Layer C3: This is the second convolutional layer with 16 feature maps. Each unit in each feature map is connected to several 5x5 neighborhoods at identical locations in a subset of S2's feature maps.

v. Subsampling Layer S4: This is the second pooling layer, and it works in a similar way to S2, further reducing the dimensionality of the previous layer.

vi. Convolution Layer C5: This is the third convolutional layer with 120 feature maps. Each unit is connected to a 5x5 neighborhood in all 16 of S4's feature maps.

vii. Fully-Connected Layer F6: This layer is a fully connected layer with 84 units.

viii. Output Layer: This layer is also fully connected, with a number of units equal to the number of classes to be predicted. The Euclidean Radial Basis Function is applied at this layer for the output. This was used for digit classification in the original

LeNet-5 architecture, and it contained 10 output units corresponding to digits from 0-9.

The initial step will involve integrating the first two Inception Modules from the Googlenet architecture into the Lenet design. This will be installed between the Convolution Layer C5 and the Fully Connected Layer F6. The experimental procedure will follow the same approach as other CNNs, beginning with a test for overall performance, followed by a cross-motion check. Fig 11 below shows confusion matrix of the overall performance of the custom CNN.

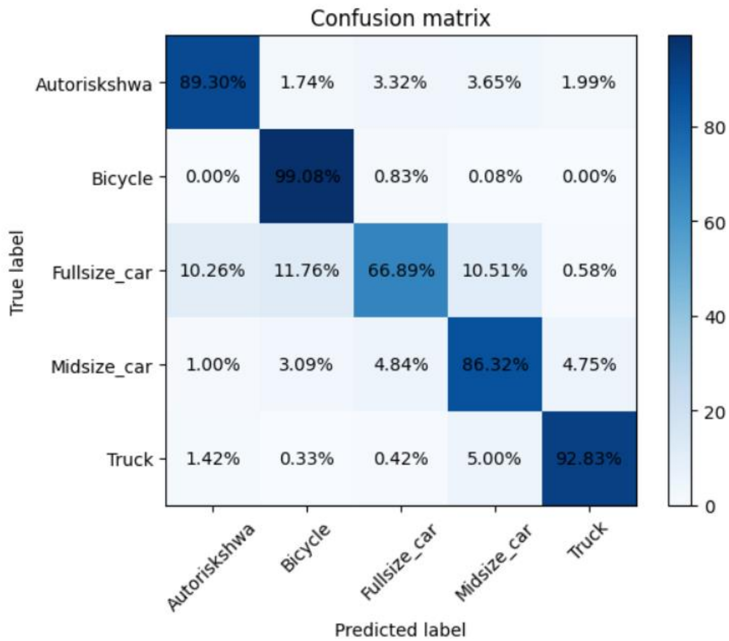

Figure 11 Confusion matrix – CustomNet overall classification

Post-implementation of two Inception modules, the CustomNet boasts an average accuracy of 86.89%, representing a notable increase of approximately 6% in comparison to Lenet. Yet, despite this improvement, the design remains significantly less complex than that of AlexNet and Googlenet. The integration of two Inception modules has decidedly enhanced the network's overall classification performance, without adding significant complexity.

Furthermore, there has been a slight enhancement in the cross-motion classification performance of the CustomNet compared to Lenet, evidenced by a marginally improved F1 score.

Train\Test	Left turn	Right turn	Straight	U-turn	Average
Left turn	0.63	0.49	0.53	0.56	0.55
Right turn	0.65	0.61	0.59	0.50	0.59
Straight	0.59	0.60	0.62	0.49	0.58
U-turn	0.68	0.57	0.43	0.50	0.55
Average	0.64	0.57	0.54	0.51	0.57

Table 4. F1 scores of Custom cross-motion check models

6. Further research

The experimental findings support the hypothesis that network depth positively correlates with performance. This is evident by the enhanced performance of the custom CNN following the incorporation of the Inception module from GoogleNet, particularly in terms of ISAR image

classification. In turn, this paves the way for extensive research concentrated on the Inception module, which could yield invaluable insights for the greater community.
This development poses several intriguing experimental queries. For example, at what point does the addition of additional Inception modules no longer affect performance? Different performance outcomes may result from repositioning the Inception modules within the network architecture. These queries suggest a plethora of potential research avenues that could be pursued to further optimize CNN's effectiveness and versatility in image classification tasks.
It is essential to keep in mind that the placement and number of Inception modules in the network architecture can affect the network's learning capacity and its ability to extract pertinent features for classification. As a result, methodical and nuanced experimentation will be required to completely comprehend the dynamics of these factors and how to optimally exploit them to improve performance. In essence, this necessitates a broader discussion and further examination of the impact of these modules on the overall network architecture, from both a theoretical and practical standpoint.

7. Conclusion

In conclusion, through methodical exploration of the field and critical appraisal of existing works, this project has been successful in pinpointing areas of research that were hitherto unexplored. To ensure the validity and dependability of the findings, these research gaps were scrutinized within the confines of a carefully crafted framework. This project not only confirms aspects of the research conducted in [13], but it also derives and explains the correlation between the complexity of a Convolutional Neural Network (CNN) and its ability to classify Inverse Synthetic Aperture Radar (ISAR) images.
A custom CNN was developed to demonstrate the effects of incorporating GoogleNet's Inception modules. The performance improvement emphasizes the need for further research into the optimal use of these modules within a network. In essence, our results demonstrate the significance of Inception modules in the design of more effective CNNs for ISAR image classification, paving the way for future research in this promising field.

References

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," in IEEE Geoscience and Remote Sensing Magazine, vol. 1, no. 1, pp. 6-43, March 2013, doi: 10.1109/MGRS.2013.2248301.R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun, "Deep Image: Scaling up Image Recognition," ArXiv Preprint ArXiv:1501.02876, vol. 7, no. 8, 2015

[2] C J Oliver 1989 J. Phys. D: Appl. Phys. 22 871, "Synthetic-aperture radar imaging" IOPscience, https://iopscience.iop.org/article/10.1088/0022-3727/22/7/001

[3] R. Vehmas and N. Neuberger, "Inverse Synthetic Aperture Radar Imaging: A Historical Perspective and State-of-the-Art Survey," in IEEE Access, vol. 9, pp. 113917-113943, 2021, doi: 10.1109/ACCESS.2021.3104799.

[4] G. Xu, M. Xing, L. Zhang, Y. Liu and Y. Li, "Bayesian Inverse Synthetic Aperture Radar Imaging," in IEEE Geoscience and Remote Sensing Letters, vol. 8, no. 6, pp. 1150-1154, Nov. 2011, doi: 10.1109/LGRS.2011.2158797.

[5] D. P. Jorgensen, P. H. Hildebrand, and C. L. Frush, "Feasibility test of an airborne pulse-doppler meteorological radar," AMETSOC, https://journals.ametsoc.org/view/journals/apme/22/5/1520-0450_1983_022_0744_ftoaap_2_0_co_2.xml

[6] M. Pak and S. Kim, "A Review of Deep Learning in Image Recognition," in 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT). IEEE, 2017, pp. 1– 3

[7] L. Wang, J. Tang and Q. Liao, "A Study on Radar Target Detection Based on Deep Neural Networks," in IEEE Sensors Letters, vol. 3, no. 3, pp. 1-4, March 2019, Art no. 7000504, doi: 10.1109/LSENS.2019.2896072.

[8] J. Williams, L. Rosenberg, V. Stamatescu, and T.-T. Cao, "Maritime Radar Target Detection Using Convolutional Neural Networks," in IEEE Radar Conference.

[9] M. Upadhyay, S. K. Murthy and A. A. B. Raj, "Intelligent System for Real time detection and classification of Aerial Targets using CNN," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1676-1681, doi: 10.1109/ICICCS51141.2021.9432136

[10] J. Ding, B. Chen, H. Liu and M. Huang, "Convolutional Neural Network With Data Augmentation for SAR Target Recognition," in IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 3, pp. 364-368, March 2016, DOI: 10.1109/LGRS.2015.2513754.

[11] Z. Liu, M. Waqas, J. Yang, A. Rashid and Z. Han, "A Multi-Task CNN for Maritime Target Detection," in IEEE Signal Processing Letters, vol. 28, pp. 434-438, 2021, doi: 10.1109/LSP.2021.3056901.

[12] B. Xue, W. Yi, F. Jing, and S. Wu, "Complex Isar target recognition using deep adaptive learning," Engineering Applications of Artificial Intelligence, vol. 97, p. 104025, 2021.

[13] N. Pandey and S. S. Ram, "Classification of automotive targets using inverse synthetic aperture radar images," IEEE Transactions on Intelligent Vehicles, pp. 1–1, 2022.

[14] S. Musman, D. Kerr, and C. Bachmann, "Automatic recognition of isar ship images," IEEE Transactions on Aerospace and Electronic Systems, vol. 32, no. 4, pp. 1392–1404, 1996.

[15] C. Pui, B. Ng, L. Rosenberg and T. Cao, "Target Classification for 3D-ISAR using CNNs".

[16] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[17] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097-1105, 2012.

[18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.
[19] MuhammadRizwan, “Lenet-5 - a classic CNN architecture,” Data Science Central, https://www.datasciencecentral.com/lenet-5-a-classic-cnn-architecture/