

# A Survey on SAR ship classification using Deep Learning

Ch Muhammad Awais\*<sup>†‡</sup>, Marco Reggiannini<sup>†‡</sup>, Davide Moroni<sup>†</sup>, and Emanuele Salerno<sup>†</sup>

\*PhD School in Computer Science, University of Pisa, 56126 Pisa, Italy

<sup>†</sup>Institute of Information Science and Technologies, National Research Council of Italy, 56124 Pisa, Italy

<sup>‡</sup>National Biodiversity Future Center - NBFC, Palermo, Italy

**Abstract**—Deep learning (DL) has become a central approach for ship classification using synthetic aperture radar (SAR) imagery. This survey reviews 74 representative studies selected from 187 publications, categorizing them into a taxonomy with four main dimensions: (i) DL architectures, (ii) datasets, (iii) image augmentation, and (iv) learning techniques. We analyze how approaches such as handcrafted feature integration, data augmentation, fine-tuning, and transfer learning influence classification performance, and summarize the use of public benchmarks including OpenSARShip and FUSARShip. This survey highlights key challenges: limited data availability, class imbalance, lack of standardized metrics, and limited interpretability of DL models. Future research directions include the development of SAR-specific DL architectures, advanced augmentation and generative approaches, integration of handcrafted and deep features, interpretable DL, and stronger interdisciplinary collaboration. By addressing these challenges, DL-based SAR ship classification can achieve greater robustness, accuracy, and transparency, ultimately strengthening maritime surveillance and operational monitoring.

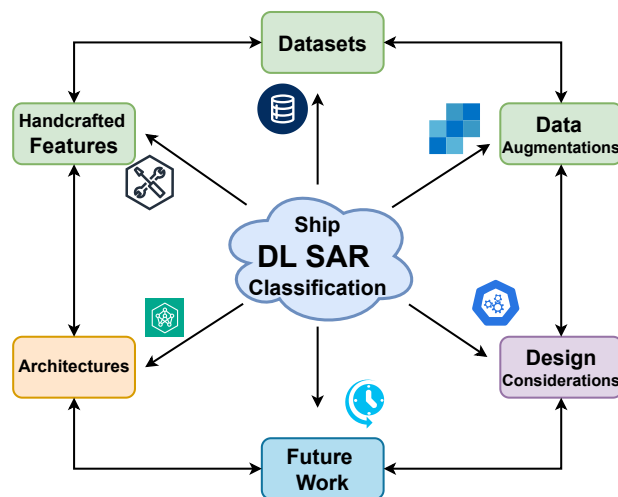
**Index Terms**—SAR ship classification, Deep learning, Synthetic Aperture Radar

## I. INTRODUCTION

Our planet is covered by approximately 71% water, which facilitates over 80% of global trade. This vast commercial activity necessitates maritime surveillance, which is a complex research field due to its extensive coverage and diverse monitoring requirements. These requirements include irregular migration, piracy, fisheries management, and traffic monitoring. Maritime surveillance also plays a crucial role in promoting environmental sustainability by providing real-time data and insights to tackle pollution, reduce emissions, conserve marine ecosystems, and ensure the long-term health of our oceans.

Many common applications of maritime monitoring exploit SAR and Automatic Identification System (AIS) data. SAR data can be collected from satellites and airplanes, with the ability to acquire information about objects on the sea surface, such as ships. The benefits of SAR are that it can operate in any light and weather conditions, it can provide images of a vast area and revisit the same area repeatedly. AIS data are transmitted from ships, received and broadcasted by ground or satellite-based stations, and contain information about the vessel, such as its unique ID, position, speed, course, and other relevant details. These data are frequently transmitted for applications such as maritime traffic management, collision avoidance, search and rescue. AIS and SAR data can be combined to create datasets that enhance maritime monitoring applications.

A feature specific to SAR technology is polarization. SAR systems operate in various polarization modes, such as single-polarization (VV or HH), dual-polarization (VV and VH, or HH and HV), and fully polarized modes (VV, VH, HH, HV). Here, “V” stands for vertical polarization and “H” for horizontal polarization, with the first letter indicating the transmitted polarization and the second the received polarization. These



different modes provide different levels of detail and information about the target, influencing the type of analysis and the performance of specific DL algorithms. Different acquisition modes, such as single-look or multi-look, further influence image characteristics. Understanding these fundamental aspects of SAR imagery is essential for an effective utilization of the information contained within these datasets. Due to their broad coverage, availability, and specific benefits, SAR data can be used for several applications, such as oil spill detection [1]–[3], ship detection [4]–[9], ship velocity estimation [10]–[13], and ship classification [14]–[17]. In SAR images, the area represented by each pixel varies depending on the specifics of the sensing platform. For example, the Sentinel-1 C-band SAR satellite constellation provides images of Earth’s surface with a spatial resolution of approximately 20 meters. This limited resolution can render vessels as only a few pixels, making object identification challenging without additional information. To address this limitation and achieve accurate target labeling, when possible, SAR-derived objects are integrated with AIS data [18].

To properly generate training sets for classification, SAR data can be matched with AIS data using several techniques [19] such as time-based matching, range-based matching and feature-based matching. Usually, SAR images are cropped so that each image chip represents a single detected ship, which is then labeled according to the matched AIS data. Then, different machine learning models are trained to detect and classify ships from unseen SAR images. The output is usually a set of bounding boxes in the large SAR image, indicating the presence of the detected ships and their associated categories. This approach to classification has shown good results using classical machine learning models, but DL models are also being explored for the same purpose.

Ships exhibit a remarkable diversity in types. For example, the FUSAR dataset [20] encompasses 15 primary ship cat-

Manuscript received September 29, 2025; revised March 13, 2026. Corresponding author: Ch M. Awais (email: chmuhammad.awais@phd.unipi.it).

egories and 98 subcategories, whereas this extensive dataset proves valuable, ship classification remains a challenging task due to several factors. The limited resolution of SAR images often obscures the unique characteristics of different ships. Additionally, their single-channel nature, as opposed to the multichannel images commonly available from optical sensors, prevents the use of color-based image processing. Moreover, the prevalence of cargo ships, the frequent type, introduces a significant class imbalance, making cargo ships easier to be classified compared to less represented vessels. Machine learning models often struggle with ship classification due to the absence of a clear understanding of which features are most relevant for accurate classification.

Conversely, DL approaches offer a promising solution for ship classification challenges because they can autonomously extract relevant features from satellite imagery, thus aiding vessel identification. However, the effectiveness of DL models relies heavily on crucial decisions regarding data selection, augmentation techniques, learning algorithms, and model architectures. This survey provides a comprehensive review of various techniques used to train DL models for ship classification.

Although several surveys review deep-learning-based SAR analysis, they primarily focus on ship detection [4], [5] or provide a broad SAR-based automatic target recognition (ATR) perspective across targets and tasks [21], where ship classification is treated only briefly. As a result, a dedicated, classification-centric survey of DL methods for SAR ship type discrimination remains lacking. Specifically, [5] is a systematic literature review (SLR) of DL-based *ship detection* that restricts its corpus to journal articles from 2016–2022 and organizes contributions around detection-related aspects such as DL types, improved detection techniques, parameters, applications, and detection challenges, without providing a classification-centric synthesis of ship type discrimination. Similarly, [4] surveys 177 papers devoted to ship detection and emphasizes detector categories and benchmark evolution, including two-stage and single-stage pipelines, anchor-free and oriented bounding-box detectors, multi-scale and real-time detection, and future directions for standardization and large detection benchmarks, which are not sufficient to guide fine-grained ship *classification* under long-tail distributions. In contrast, [21] provides a broad SAR ATR perspective across targets and tasks, covering datasets and metrics, pre-DL ATR, deep-learning-era methods, and issues such as limited samples, imbalance, polarimetric SAR, and adversarial robustness, but ship classification remains a small part of this multi-category ATR scope. Our survey therefore delineates a clear innovation boundary by being explicitly classification-centric and by materializing two contributions that earlier reviews do not provide for ship type discrimination: a four-dimensional taxonomy spanning DL architectures, datasets, data augmentation, and learning techniques that is both complete and operational for practitioner decision-making, and a standardized performance compilation that consolidates Overall Accuracy (OA) and F1 results organized by class count and annotated with dataset identity and available split information, enabling controlled within-group comparisons and transparent filtering when cross-study protocols differ.

The remainder of this paper is organized as follows. Section II details the survey methodology and introduces the proposed taxonomy. Section III reviews the relevant literature. Section IV synthesizes key insights from the reviewed studies and includes a comparative discussion that links technique families to data constraints. Section V provides design considerations for DL-based SAR ship classification, including recommendations

under class imbalance and limited labels. Section VI discusses key issues and open challenges. Section VII outlines possible future developments, and concludes the paper.

### Contributions and Scope

*Contributions:* This survey gives the following contributions:

- A systematic review of 187 papers, the most pertinent 74 thereof are analyzed in depth. These references are listed in Table II, and their key features are summarized in Section III.
- A clear taxonomy (Figure 2) covering (i) DL architectures, (ii) datasets, (iii) image augmentation, and (iv) learning techniques (including transfer learning and fine-tuning).
- A standardized performance table, reporting scores, classes, and datasets found in the reviewed works, with emphasis on benchmarks such as OpenSARShip and FUSARShip. The performance indices are consistently reported using F1 score and OA.
- A synthesis of empirical trends, including the effectiveness of shallow CNNs under data scarcity, the benefits of data augmentation and fine-tuning, strategies for class imbalance (such as data splitting and tailored loss functions), and the impact of feature/polarization fusion, metric learning, multi-scale/pyramid learning, learning rate scheduling, and architecture search (NAS).
- Practical steps to mitigate data scarcity, including creating high-quality datasets from open-source SAR imagery, establishing AIS–SAR mapping methodologies for reliable labels, and encouraging standardization and data sharing across the community.
- A selection of open issues, such as data scarcity, class imbalance, standardized reporting, interpretability, and reproducibility.
- Concrete directions to guide future work in SAR ship classification.

*Scope:* The focus of this survey is SAR-based ship classification (category-level recognition on cropped ship chips/SAR patches) using DL. We include studies that (i) operate on SAR imagery from spaceborne or airborne platforms, (ii) consider single-, dual-, or fully-polarized data, and (iii) report F1 score and/or OA, which serve as the basis for comparison in this survey<sup>1</sup>. We include a discussion of handcrafted feature baselines where they inform DL design and performance. We exclude studies limited to detection or segmentation without a classification component, optical-only methods without a SAR pathway, and non-maritime ATR unless directly comparable to ship classification.

## II. SURVEY METHODOLOGY

### A. Research Questions

The aim of this survey is to find and highlight techniques that are being used in improving ship classification through DL. The research questions are focused on finding the concerning literature:

- **RQ1:** How has the use of specific DL techniques evolved over time in the field of SAR ship classification?

<sup>1</sup>When available, F1 is treated as macro-F1; if a source reports a different F1 variant, it is indicated in the performance table. Likewise, any form of accuracy reported in the source (e.g., top-1 accuracy, classification accuracy, recognition rate) is consistently referred to as overall accuracy (OA) in this survey.

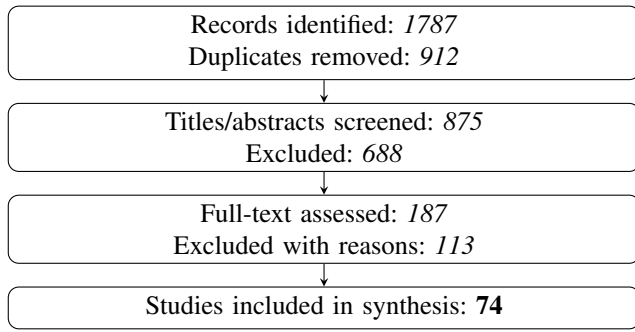


Fig. 1: Screening flow.

- **RQ2:** What are the performance characteristics (e.g., accuracy, F1) achieved by different DL techniques in SAR ship classification?
- **RQ3:** What is the impact of incorporating advanced techniques (such as pyramid features, feature fusion, and generative methods) on the performance of SAR ship classification using DL?
- **RQ4:** What are the commonly employed network architectures (e.g., convolutional neural network (CNN), ResNet) for SAR ship classification using DL, and how do they compare in terms of performance?
- **RQ5:** What are the limitations and challenges associated with the existing DL techniques for SAR ship classification, as identified in the literature?
- **RQ6:** Are there any specific datasets or preprocessing techniques commonly used in the studies on SAR ship classification using DL?

These research questions aim to provide insights into the utilization, performance, challenges, and future directions of DL methods for SAR ship classification.

### B. Search Query

Search strings are crucial tools to filter through vast amounts of information and identify relevant and accurate sources. The search string formulated to capture studies that are specifically part of our scope was

(Deep learning OR Neural network) AND (Ship classification OR Vessel classification OR Ship categorization) AND (Synthetic Aperture Radar OR SAR)

Apparently, this string addresses the use of DL approaches in SAR ship classification, including variations of terms such as DL, neural network, ship classification, vessel classification, ship categorization, synthetic aperture radar, and SAR.

### C. Study Selection and Data Synthesis

To identify a relevant set of papers for the systematic literature review, the study selection process (Figure 1) involved an initial search (using the query in Section II-B) across multiple databases, followed by title/abstract screening and full-text assessment by two independent reviewers, with disagreements resolved through discussion and consensus. Relevant information was systematically extracted from the included papers, and the data were synthesized and analyzed to answer the research questions.

The study selection process was conducted meticulously to ensure the inclusion of high-quality and relevant papers that contribute to the body of knowledge on SAR ship classification using DL. The inclusion criteria applied during the selection

TABLE I: Search results before/after deduplication and screening.

Source	Initial	After dedup	Included
Web of Science	52	33	13
IEEE Xplore	55	43	18
Scopus	63	19	7
ScienceDirect	87	2	1
SpringerLink	30	-	-
Google Scholar	1460	90	35
ResearchGate <sup>3</sup>	100	-	-
<b>Total (unique)</b>	<b>1787</b>	<b>187</b>	<b>74</b>

TABLE II: Chronological Summary of SAR Ship Classification Papers

Year	Published Studies
2017	[24]
2018	[25], [26], [27]
2019	[28], [29], [30], [31], [32]
2020	[33], [34]
2021	[35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48]
2022	[49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59]
2023	[60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71]
2024	[72], [73], [74], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84], [85], [86], [87]
2025	[88], [89], [90], [91], [92], [93], [94], [95], [96], [97]

process were inspired by [22], which derives them from the PICOC approach (Population, Intervention, Comparison, Outcome, Context) proposed by [23] to ensure a focused and comprehensive selection.

1) *Inclusion Criteria:* This survey includes papers focused specifically on ship classification from satellite SAR images using DL techniques, algorithms or models, and appearing in peer-reviewed journals, conference proceedings or other reputable scholarly sources. Additional criteria to be satisfied were:

- 1) Studies reporting performance metrics, evaluation results, or quantitative measures of DL model effectiveness.
- 2) Studies published in English to ensure accessibility and facilitate data extraction and analysis.
- 3) Studies using publicly available datasets.
- 4) Studies published after 2010.

All the papers not complying with these requirements were excluded from any further analysis.<sup>2</sup>

Per-database counts *after deduplication* are reported in Table I. The initial screening reduced the corpus from 1,847 records to 187 by removing duplicates, off-topic items, and studies that did not meet the inclusion criteria.

During the full-text assessment stage, the remaining 187 papers were examined against the research questions and inclusion criteria to identify the most pertinent studies on DL-based SAR ship classification; this evaluation resulted in 74 included papers (see Table II), forming the basis for the subsequent synthesis and analysis.

After the meticulous and thorough evaluation of the full texts described above, we derived the taxonomy illustrated in Figure 2, which shows models, datasets, augmentations and

<sup>2</sup>*Threats to Validity:* **Selection bias:** despite multi-database search, some works may be missed. **Metric comparability:** heterogeneous splits limit strict cross-paper comparisons. **Reporting variance:** inconsistent metric definitions (F1 variants) and missing details can affect synthesis. We mitigate these by transparent inclusion criteria, deduplication, and explicit caveats in tables.

*Reproducibility Checklist:* We provide: (i) the performance table (tables-VII, VIII); (ii) search strings (section-II-B); (iii) deduplication logic summary (table-I). Each entry records dataset, split notes, and metrics (F1/OA).

<sup>3</sup>Used primarily for discovery; entries were verified against peer-reviewed sources.

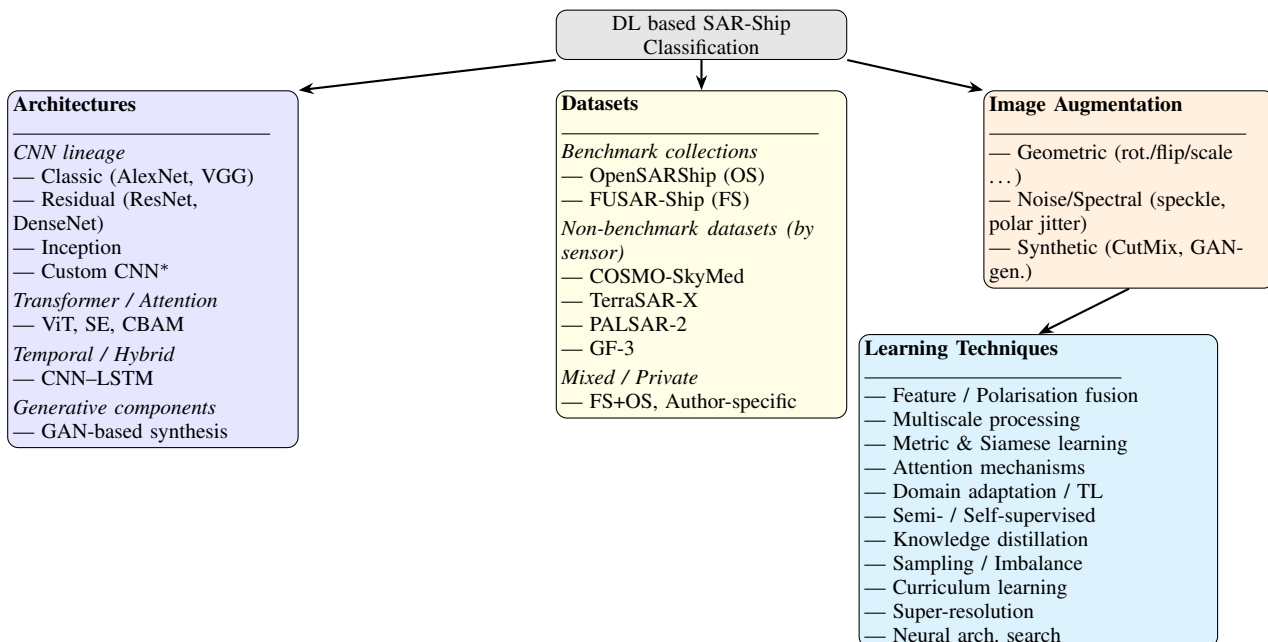


Fig. 2: Taxonomy of DL methods for SAR ship classification. The taxonomy is organized along four principal factors: (i) backbone architectures, (ii) datasets, (iii) image augmentation, and (iv) learning techniques. These dimensions are not mutually exclusive: a single study can combine multiple learning techniques with a given backbone, dataset, and augmentation setup. Here, *Architectures* denotes the primary feature-extraction backbone, whereas *Learning Techniques* denotes training strategies and architectural add-ons used to improve robustness under limited data and class imbalance. Dataset entries are listed by benchmark name when publicly released, otherwise grouped by sensor family or described as author-specific collections. \* denotes CNN variants specifically designed by the respective authors.

learning techniques used in the literature for DL-based SAR ship classification. The taxonomy emerged inductively from systematic analysis of methodological patterns across the 74 papers, reflecting the fundamental design choices researchers make: model architecture, training data, data expansion strategies, and learning techniques. This structure achieves a complete coverage of the reviewed literature while enabling the comparative analysis presented in Section IV.

#### D. Survey Findings

The analysis of relevant literature revealed several key insights into the use of DL techniques for ship classification from SAR images. Here is a list thereof, ordered by research questions (see Section II):

**RQ1 Evolution of DL Techniques:** The adoption of CNN models initially stemmed from their success in other image processing tasks. However, as the demand for enhanced performance intensified, researchers integrated various image processing techniques and handcrafted features to augment the effectiveness of DL models [35], [67], [74], [75], [85], [98].

**RQ2 Performance Characteristics:** Accuracy is the most commonly reported performance metric in the literature. However, additional metrics should be considered, particularly when addressing long-tailed or imbalanced datasets [87]. Relying solely on accuracy can be misleading, as it may not adequately reflect performance across different classes, especially minority categories. To address this, we extracted the reported accuracy and F1 scores from the studies and summarized them in Tables VII and VIII. It is important to note that these scores are not directly comparable across studies due to differences in dataset selection: subsets of test data vary, and even when the same number of classes is reported, the specific class composition may differ.

**RQ3 Impact of Advanced Techniques:** Techniques like pyramid features, feature fusion, curriculum learning, and transfer learning were employed to bolster ship classification performance on SAR images [29], [44], [53], [65], [66], [71], [91], [92], [97]. Handcrafted features and pyramid features demonstrated particularly promising results. However, the research in this area remains in its nascent stages, warranting further investigation.

**RQ4 Commonly Employed Network Architectures:** Specific CNN architectures, particularly deeper models such as VGG and ResNet, proved to be promising in SAR ship classification due to their ability in extracting complex features. Interestingly, simpler and shallow CNN models, when combined with effective image processing such as noise reduction and image enhancement, have also achieved commendable accuracy [59], [78].

**RQ5 Limitations and Challenges:** SAR ship classification faces several critical challenges. The scarcity and imbalance of labeled SAR image data, coupled with limitations in SAR image processing techniques, hinder model performance. Additionally, relying solely on DL networks designed for optical imagery or choosing inappropriate SAR-specific architectures can lead to suboptimal results. Furthermore, SAR-AIS matching, while valuable, has limitations due to its inherent unreliability (e.g. lack of transmission, noisy and inconsistent messages [99]). The development of robust, automated SAR-AIS matching algorithms is crucial. Finally, SAR processing models can benefit significantly from tailored loss functions [27], [38], [55] that effectively filter out irrelevant information and focus on the significant features of target objects. Research on these tailored loss functions is an active area of development.

**RQ6 Specific Datasets and preprocessing Techniques:** Studies in the literature commonly use two publicly available

datasets: FUSAR-Ship [20] and OpenSARShip [100]. Geometric preprocessing techniques, such as cropping, rotation, flipping and scaling, are frequently employed to improve the generalizability of ship classification models. However, the necessity for these techniques arises from the challenges associated with small and low-resolution vessel representations in the data. Super-resolution techniques offer a potential avenue for addressing these limitations [93], [94], [101]–[103].

### III. BACKGROUND

Several techniques are employed for SAR ship classification using DL, as illustrated in Figure 2. This section is divided into four parts. The first part explores the various DL techniques used. The second part examines the impact of handcrafted features on the performance of DL models. As SAR images possess unique attributes, the third part discusses how these attributes are leveraged. Finally, the fourth part investigates the use of fine-tuning.

#### A. DL Techniques

As apparent from Table II, we found the origin of SAR-based ship classification using DL in 2017 [24], when four different ship types detected in six VV-polarized COSMO-SkyMed scenes were classified using a simple CNN model with two convolution layers and two fully connected layers. The images of the detected targets were cropped to  $64 \times 64$  chips, four scenes were then used for training, one for testing and one for validation. The results showed that the model was able to learn the cargo ships better than the others, because of their high prevalence in the training set. The advancement of SAR platforms over time has resulted in images with better and better spatial resolutions. In [25], CNNs are used to treat images from TerraSAR-X, with a resolution of 3 meters per pixel. Four augmentation techniques are applied to form the training set: reflections (horizontal and vertical), rotations (-10 to 10 degrees), translation (-5 to 5 pixels) and noise addition (randomly changing pixel values to zero), achieving an F1 score of 94% by using multiple-resolution images.

Two important aspects of a DL model are explored in [27]: an attention-based CNN model and a novel loss function for class-imbalanced datasets, named “weighted distance measure”. A VH-polarized dataset with Ground Range Detected (GRD) images is divided into 4 classes, cargo, tanker, tug and others. The results show that the model is capable of learning class representation, achieving a recognition accuracy of 84%.

In [31] a convolution network for detection and classification is evaluated. The input data are drawn from the PALSAR-2 L-band SAR with 10 m resolution in HV polarization. The results indicate that the CNN-based models can detect ships, but the classification performance is affected by ships that look alike.

Ship classification can also be achieved using a sequence-based approach. In [33] a sequence input is constructed by capturing ten consecutive images of the same target at different times and rotation degrees. The authors employ a combination of CNN and long short-term memory (LSTM) with four categories of OpenSARShip data (boat, cargo, container, and oil tanker). Their findings suggest that the model learns improved class representations when using image sequences. Similarly, the authors of [78] first optimize the CNN design by reducing filter redundancy and network size, and then integrate LSTM layers to capture sequential dependencies and enhance discrimination. Comparative experiments against state-of-the-art CNNs (e.g., VGG-16, ResNet50, DenseNet121) show that

the hybrid network achieves competitive accuracy while significantly reducing training time and complexity. These results highlight the potential of lightweight CNN–LSTM frameworks for efficient ship classification in resource-constrained settings.

A CNN, when not fully connected, can be used as a feature extractor. The extracted features can then be fed to a classifier or further processed using methods like metric learning. Metric learning involves training a model to measure the similarity or distance between data points, often to cluster extracted features in a way that makes them distinguishable, thereby enhancing the model’s performance. In [34], the performance of a CNN-only approach is compared to a CNN combined with metric learning. The OpenSARship dataset is used, restricted to bulk carrier, container, and oil tanker classes, focusing on data with larger sample sizes. The features extracted using the CNN are then enhanced by concatenating additional features learned using metric learning. The results show that the CNN combined with metric learning outperforms the simple CNN.

In [64], a novel Hierarchical Embedding Network with Center Calibration (HENC) is proposed for few-shot ship classification. HENC uses a CNN to extract hierarchical features at different scales, achieving an accuracy of 81.03% on a six-category OpenSARShip dataset (including cargo, tanker, tug, fishing, dredging and others) with few-shot training, thus demonstrating the effectiveness of deep models in learning class representations even with limited data.

In [65], a deep metric learning approach is discussed, which leverages a pairwise representation fusion for multi-proxy learning and introduces a novel distribution loss to refine the sample distribution. This allows for both clustering similar classes together and training the model to make a clear distinction between all classes. The effectiveness of this three-pronged approach is evaluated on two optical ship datasets (FGSC-23 and FGSCR-42) and one SAR dataset (5-class FUSARShip).

In [71], the authors introduce HDSS-Net, a novel DL architecture for ship classification in SAR images. This addresses challenges such as large variations in ship size, the elongated shapes of ships in SAR images, and intra-class diversity alongside interclass similarity. HDSS-Net employs a hierarchical design with three key modules. The first, a feature aggregation module (FAM), creates a multi-scale feature pyramid. The second, a Feature Boost module (FBM), captures the elongated features of ships through rectangular convolutions. Finally, a spherical space classifier (SSC) processes features in a spherical space to enhance class separability. Evaluations on OpenSARShip (three and six classes) and FUSAR-Ship (seven classes) demonstrate that HDSS-Net outperforms traditional CNN methods, highlighting its potential value for maritime surveillance applications.

There exists a trade-off between spatial resolution and the cost of SAR imagery. Developing algorithms that achieve accurate results using medium-resolution data is essential for broader applications. DenseNet and metric learning, with a novel loss function incorporating deep metric learning [38], are used in [104] to perform SAR ship classification on three-class (tanker, container, bulk carrier) and five-class (tanker, container, bulk carrier, cargo, general cargo) datasets from OpenSARShip. To help the model learn accurate class representations, image triplets (an anchor image, a positive image, and a negative image) are used, and data augmentation techniques such as flipping, rotation, translation and Gaussian noise addition are applied to address overfitting. The results demonstrate superior performance in both the three-class and five-class experiments.

Each CNN layer stores important information about features,

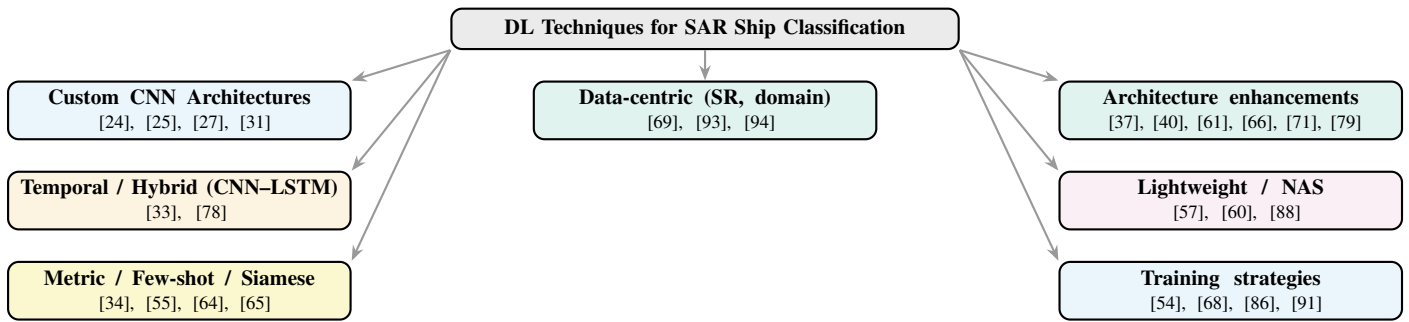


Fig. 3: Overview of DL techniques applied to SAR ship classification. The methods are grouped into major categories, including core Custom CNN Architectures, architecture enhancements, temporal and hybrid approaches, lightweight and NAS-based models, metric and few-shot learning, training strategies, and data-centric methods. Representative studies for each category are indicated by citation keys.

but some information may be lost when features pass from one layer to another. In [37], a novel multi-scale CNN (MS-CNN) is proposed to utilize all the information from each convolutional layer. Using three classes from the OpenSARship dataset (bulk carrier, container ship and tanker), the MS-CNN, integrated with information from each convolutional layer, achieves a 5% better performance than the same CNN without integration. This improvement suggests that information loss occurs within the network.

In [61], a novel deep network is proposed, with an inception-residual controller (IRC) module to improve the performance. The authors train the model on a five-category dataset from FUSARShip. The data undergo homogenization, normalization, and augmentation techniques such as flipping and rotation and, after 50 training epochs, the model with IRC achieves an accuracy of 98.71%, indicating that deep models trained on pre-processed data can significantly improve the classification performance.

Lightweight CNNs have been explored as means to reduce the computational burden of ship classification from high-resolution SAR data while maintaining competitive accuracy. A compact architecture called LN-SCNet is proposed in [88]. It integrates a preprocessing module to suppress speckle noise, dilated convolutions for initial feature extraction, and incorporates a feature fusion module that combines high-resolution features for small ships with deeper semantic features for large ships. To mitigate class imbalance, a seesaw loss is introduced, dynamically adjusting gradients across classes. Evaluated on 8 classes of the FUSAR-Ship dataset, LN-SCNet achieves 76.1% accuracy and an F1 score of 56.7%, outperforming other lightweight designs such as MobileNet and ShuffleNet while requiring only a fraction of their parameters and floating-point operations. These results suggest that carefully designed lightweight networks can offer an effective trade-off between efficiency and performance in SAR ship classification.

A plug-and-play Multi-scale Feature Spatial Attention Fusion (MFSAF) module is introduced in [79] to enhance CNN performance in SAR ship classification. This module integrates spatial attention with feature alignment, enabling CNNs to better capture ship features across multiple scales. Applied to several backbones (e.g., VGG-11, ResNet-18, DenseNet-121, EfficientNetV2, ConvNeXt), the MFSAF module consistently outperforms the baseline models, achieving an 81.9% accuracy on OpenSARShip2.0 when applied to ConvNeXt. These results highlight the adaptability of MFSAF and its effectiveness in improving spatial attention and feature fusion for robust ship classification in SAR imagery.

A multi-scale dual-attention CNN that leverages information

from convolutional layers is proposed in [40]. The authors employ a VGG-Net architecture for feature extraction and utilize a three-class dataset (cargo, container ship, and tanker) from OpenSARShip. Their methodology involves modifying the VGG network to extract features from SAR images, followed by the application of squeeze-and-excitation (SE) blocks to focus on both channel-wise and positional contexts. SE blocks learn the importance of feature maps, promoting informative features and suppressing less useful channels. The results demonstrate the effectiveness of this method, achieving an accuracy of 92.66%. This success suggests the potential of attention mechanisms in this field. An enhanced vision transformer incorporating feature fusion modules to capture local information followed by SE attention is proposed in [66]. The effectiveness of this approach is evaluated on a three-class OpenSARship dataset, achieving an accuracy of 76.25%.

In the training stage, the convergence of a DL model is controlled by several parameters, among which the learning rate is of paramount relevance. It determines how the model adjusts its internal parameters during training, and can either be fixed throughout the process or adjusted dynamically. Three learning rate management strategies and their impact on transfer learning are explored in [54]. The results suggest that a smooth learning rate change can improve the model's ability to understand the data, leading to enhanced classification performance. The authors also investigate the effect of freezing different layers during transfer learning on a CNN, emphasizing the importance of layer analysis before fine-tuning. The models are first trained on the ten-category MSTAR dataset (SAR images of tanks, [105]) and then fine-tuned on a three-category OpenSARShip dataset (cargo, container ship, and bulk carrier). The results demonstrate that incorporating learning rate scheduling and thoughtful layer freezing significantly improves the model performance. This finding underscores the importance of employing effective training strategies for optimal SAR ship classification.

Selecting optimal DL architectures for specific tasks can be challenging. Traditionally, researchers rely on exploring existing architectures and their applications. However, this approach often lacks a systematic methodology, leading to difficulties in finding the ideal architecture for a given classification problem. To address this challenge, a lightweight searched binary neural network (SBNN) is proposed in [57], obtained through network architecture search (NAS), quantized for weight reduction and incorporating patch shift processing to remove noise. The authors employ three (bulk carrier, container and tanker) and six-class (bulk carrier, container ship, tanker, cargo, general cargo and fishing) datasets from OpenSARShip, containing both VV and VH polarizations. They apply data augmentation

methods such as random cropping and horizontal flipping. The SBNN achieves accuracy scores of 80.03% and 56.73% for three and six classes, respectively. NAS not only results in superior performance but also leads to a lightweight model. A novel mini-size searched convolutional metaformer (SCM) is introduced in [60], leveraging an updated NAS algorithm with progressive data augmentation (crop, flip, rotate, resize, cutout) to discover an efficient convolutional network baseline. This baseline is then enhanced with a transformer classifier to improve spatial awareness, and further augmented with a ConvFormer cell, combining the searched convolutional cell within a Metaformer block for superior feature extraction. Evaluated on VV- and VH-polarized datasets from OpenSARShip (3 categories, i.e. bulk carrier, container, and tanker) and FUSAR-Ship (4 categories, i.e. bulk carrier, fishing, cargo, and tanker), SCM achieves accuracies of 82.06% and 63.9%, respectively, with a lightweight design (about half a million weights). In [55], a one-shot learning approach using a siamese network is proposed. The approach employs a 7-layer, fully connected feature fusion that enables the network to classify 16 categories within the OpenSARShip dataset. This methodology involves three stages: (i) preprocessing the images and converting them to three dimensions by adding two additional channels (actual, filtered, and inverted SAR Image), (ii) creating image triplets containing an anchor image, a positive image, and a negative image, and (iii) training the model using a loss function that ensures the anchor image to be closer to the positive image and farther from the negative image. The results demonstrate the model's effectiveness for ship classification.

The challenge of handling the long-tailed distribution problem in imbalanced SAR data is addressed in [68]. The authors propose a novel architecture that incorporates a multi-branch expert network with a dual-environment sampling strategy. This network consists of three classifiers trained to handle different distribution scenarios: long-tailed, balanced, and inverse-long-tailed (to focus on classes with lower samples). Intra-class dual-environment sampling adjusts sampling rates within each class based on classification confidence, creating varied distributions. In [91], curriculum learning is used to address long-tailed distributions by introducing two curriculum methodologies for training, and also explores the importance of loss functions. Combined with center loss [106], this method enhances robustness and accuracy, outperforming existing techniques on the MSTAR and FUSAR datasets. Experiments confirmed significant improvements in recognizing both majority and minority classes, thus demonstrating the effectiveness of the proposed approach in balancing feature learning and improving overall and average accuracy.

Overfitting remains a key limitation in SAR ship classification, particularly in few-shot and noisy data scenarios. To address this, [86] presents a double reverse regularization strategy that leverages both offline and online self-distillation in a complementary manner. An Adaptive Weight Assignment (AWA) module is used to dynamically adjust the balance between the two sources of soft targets according to network performance, rather than relying on fixed schedules. Experiments on OpenSARShip and FUSAR-Ship show that this strategy consistently improves the performance of classical CNN backbones, achieving accuracy gains of up to 7% over baselines and surpassing other distillation-based regularization approaches. The results suggest that adaptive distillation can provide an effective path toward improved generalization under limited training data.

Due to the resolution of SAR images, research on small-

sized fishing vessels has not been given sufficient importance. However, as the resolution of SAR systems increases, there is a growing need to address the classification of fishing vessels. A novel architecture to this purpose is proposed in [69], consisting of three modules: (i) the multi-feature extraction (MUL) module, which extracts features using DenseBlocks and channel-wise concatenation, (ii) the feature fusion (FF) module, which aggregates features from different layers, and (iii) the parallel channel and spatial attention (PCSA) module, which applies channel and spatial attention. The paper also proposes a method for SAR and AIS integration to create a fishing vessel dataset. This methodology was tested on three fishing vessel classes, and the model achieved an accuracy of 89.7%.

The importance of super resolution in SAR ship classification is discussed in [94], whereas [93] emphasizes the importance of considering classification loss when enhancing the image resolutions of imbalanced datasets.

### B. Handcrafted Features

Traditional features like histogram of oriented gradients (HOG) can be combined with features extracted by a CNN to enhance the performance of DL models that require a large amount of features. HOG features capture the distribution of edge orientations within an image region, describing its local shape and appearance. A VGG network is used in [35] as the backbone for feature extraction and concatenate these features with HOG features on three categories (bulk carrier, container ship and tanker) from the OpenSARship dataset. The results indicate that the model that incorporates HOG features achieves 7% better performance than the model without HOG features.

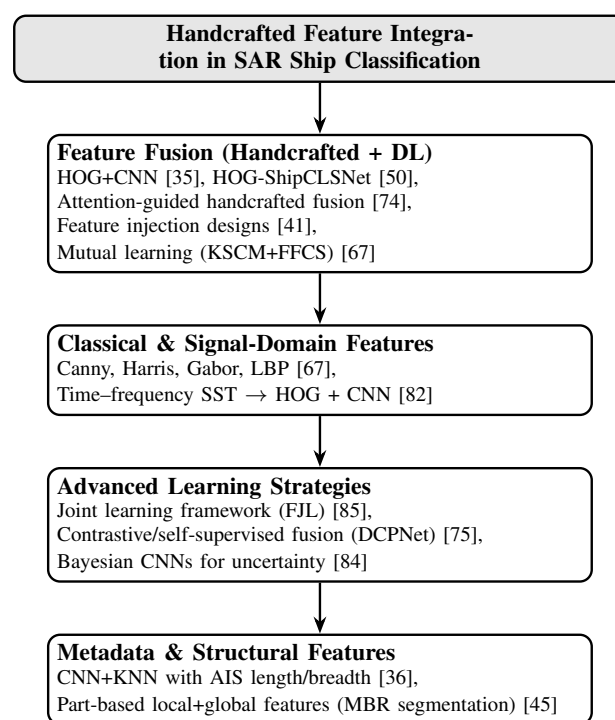


Fig. 4: Overview of handcrafted feature integration approaches in SAR ship classification. The methods are organized into four categories: (i) fusion of handcrafted descriptors with deep features, (ii) use of classical and signal-domain descriptors, (iii) advanced learning strategies incorporating handcrafted cues, and (iv) metadata- and structure-driven feature designs. Representative studies for each category are indicated by citation keys.

A method called HOG-ShipCLSNet is proposed in [50]. This method first extracts HOG features and then learns deep features

using a CNN called the multiscale classification mechanism (MS-CLS). MS-CLS extracts features from each convolutional layer. Since the features from different layers have varying dimensions, the network employs a global self-attention mechanism to focus on important features within each layer. To ensure stable training, a fully connected balance mechanism resizes the features to a uniform size and then averages them. Finally, the HOG feature fusion mechanism merges the handcrafted HOG features with the DL features before performing the classification. The authors conducted experiments on two datasets: three categories from OpenSARShip (bulk carrier, container ship, tanker) and seven categories from FUSARShip (bulk carrier, container ship, fishing, tanker, general cargo, other cargo, other). The results indicate that HOG-ShipCLSNet is effective for ship classification. Another attention-guided neural network, using OpenSARship, is proposed in [74]. This network incorporates 14 handcrafted features processed by a feedforward network and combines them with features extracted by the attention-guided network. These combined features are then employed to train a classification model. The model, trained on 4 categories from OpenSARship, is evaluated on a dataset of SAR-AIS integrated Sentinel-1 SAR images. This approach achieves an accuracy of 86.3%, demonstrating the potential effectiveness of combining handcrafted features with an attention mechanism for ship classification tasks.

Choosing effective features plays a vital role in classification performance: indeed, meaningful features help the model to be trained optimally. Four handcrafted features (Canny, Harris, Gabor, and local binary patterns (LBP)) are compared in [67], where an optimal method for feature fusion is proposed. The method consists of two blocks: (1) a knowledge supervision and collaboration module (KSCM), which facilitates mutual learning between handcrafted and deep features, and (2) a feature fusion and contribution assignment module (FFCS), which enhances channel attention. The proposed model has been tested on three OpenSARShip categories and seven FUSARShip categories using both deep and shallow versions of AlexNet, VGGNet, ResNet and DenseNet. The results show that handcrafted features improve the performance of all models, with Gabor features showing the most significant average gain of 3.4%. Shallow networks (VGG-11 and ResNet-19) exhibit the largest improvement over their baseline counterparts following feature fusion.

A ship classification framework that incorporates non-stationary information of SAR targets into DL models is introduced in [82]. By this approach, the data undergo the so-called “second-order synchrosqueezing transform” (SST), which allows the ship characteristics to be represented in the time–frequency domain. Then, the HOG features are extracted, thus reducing the dimensionality and providing discriminative inputs. These features are then fused with CNN representations, with ResNet identified as the most effective backbone. Experiments on 3 classes from OpenSARShip and FUSARShip show clear performance gains, with accuracies of 82% and 90%, respectively. This study highlights the value of non-stationary signal characteristics for improving robustness and generalization in SAR ship classification.

In [85], a multidimensional feature joint learning framework (FJL-Framework) is introduced, designed to overcome the limitations of simple feature concatenation. The framework consists of three modules: (1) a subspace-level feature selection model with geometric constraints (SGC-GA) to refine pattern features, (2) a global–local interaction model (GFIC) to enhance deep feature learning, and (3) a joint learning model (MFJL) to

fuse both types of features. This approach consistently achieves accuracies of 79.0% and 87.6% for 3-class OpenSARShip and 7-class FuSARShip respectively. By selectively integrating geometric, pattern, and deep features, the FJL-Framework demonstrates the importance of conflict-aware joint learning for robust SAR target recognition.

In [84], the use of Bayesian CNNs (BCNNs) for SAR ship classification is investigated, comparing them to deterministic CNNs. Using three classes from OpenSARShip, the study shows that a Bayesian variant of LeNet-5 (BLeNet-5) improves accuracy by 1.6% over LeNet-5, nearly matching the performance of VGG-16 while requiring far less computation. Beyond accuracy, BCNNs provide calibrated uncertainty estimates, decomposing them into aleatoric and epistemic components. The results suggest that Bayesian models can deliver more reliable predictions and offer useful confidence measures, highlighting their potential for robust SAR ship classification in limited-data settings.

The effect of combining handcrafted features with DL features is investigated in [41] by considering four handcrafted features (HOG, naive geometric features (NGF), local radar cross section (LRCS), and principal axis features (PAFs)) in conjunction with four DL models (AlexNet, VGGNet, ResNet, and DenseNet). The authors analyze different feature injection locations and techniques, recommending injection in the final fully-connected layer with dimension unification, weighted concatenation and feature normalization (DUW-Cat-FN) for optimal results. They use three classes (bulk carrier, container ship, and tanker) from OpenSARShip and seven classes (bulk carrier, container ship, fishing, tanker, general cargo, other cargo, and other) from FUSARShip. The results reveal that injecting handcrafted features into a DL model has the potential to improve performance by up to 6%, thus highlighting the importance of handcrafted features. In [75], a novel contrastive learning-based method is proposed, leveraging the fusion of handcrafted features with SAR images. It employs data augmentation techniques on SAR images and handcrafted features, applying distinct augmentations to each data type. Subsequently, the model calculates three types of losses during deep training to facilitate learning of two primary pretext tasks and a cluster-level task. The proposed model, named DCPNet, achieves accuracies of 73.66% with three OpenSARShip classes and 87.94% with seven FUSARShip classes. These results suggest the potential effectiveness of contrastive learning for ship classification tasks.

Images in OpenSARship contain information such as the length and breadth of the vessel. The ship classification performance of a DL model is enhanced in [36] by combining the outputs of a CNN and a K-nearest neighbor model (KNN), where the latter classifies new data based on the majority class of its nearest neighbors. KNN is trained to classify ships based on length and breadth taken from AIS data, while the CNN uses the image data. Three classes (cargo, tanker, and others) from OpenSARship dataset are used. The model combined with KNN showed superior performance compared to the CNN model alone.

Since a ship comprises various sections, the authors of [45] first employ a minimum bounding rectangle (MBR) to locate the ship in the image, then subdivide it into the bow, midship, and stern sections and extract the features using two models based on the VGG network. One model leverages the data from the three sections to capture local features, while the other analyzes the entire MBR to capture global features. Finally, feature fusion is employed to classify the ships into

two categories: military and civilian. The method, applied to high-resolution images captured by China's X-band GaoFen-3 (GF-3) satellite, achieves an accuracy of 96%.

### C. Use of SAR properties

Maritime monitoring through SAR imagery represents a unique challenge in case of moving ships. Indeed, motion can cause the ship to appear shifted within the SAR image with respect to its actual geographical position. This information can be exploited proficiently for classification purposes. A complex-valued CNN (CV-MotionNet) that considers both amplitude and phase information from single-look complex SAR images is proposed in [39]. The experiments were conducted on two datasets: a five-class simulated dataset and a three-class real-world GaoFen-3 dataset containing cargo ships, tankers, and other categories. The results demonstrate the model's effectiveness in classifying SAR targets.

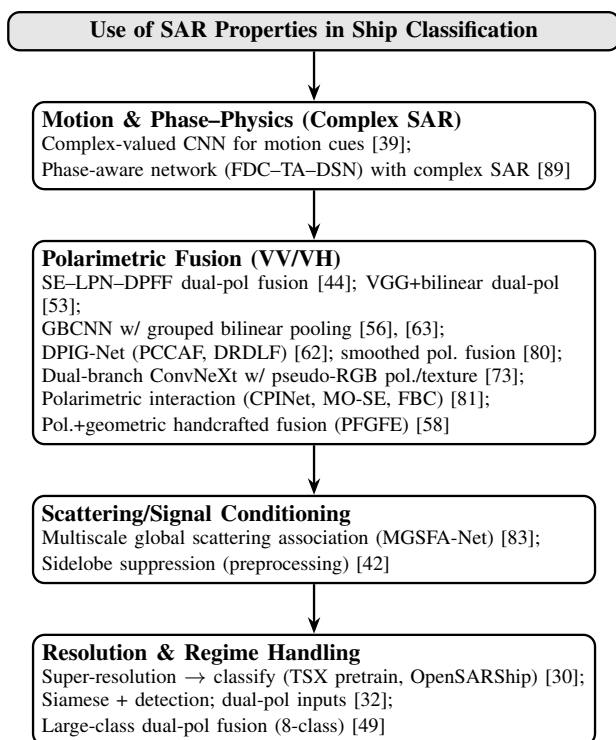


Fig. 5: Overview of methods that explicitly exploit SAR imaging properties for ship classification. Approaches are grouped into four categories: (i) motion and phase information from complex-valued data, (ii) polarimetric fusion of VV/VH channels, (iii) scattering- and signal-conditioning techniques, and (iv) resolution- and regime-specific strategies. Representative studies for each category are indicated by citation keys.

Sidelobes, faint echoes in SAR images caused by the finite length of the radar pulse, can hinder the performance of DL models for ship classification. The effect of sidelobe suppression is investigated in [42] using a simple CNN model on ten classes from FUSARShip (anti-pollution, bulk carrier, container, dredger, fishing, cargo, law enforce, passenger, oil tanker, and port tender). The results show that suppressing the sidelobes improves the model performance, achieving up to 98% accuracy.

The authors of [44] investigate whether using information from two different polarizations, VV and VH, can improve the model performance by capturing more useful features. They propose a novel network called squeeze-and-excitation Laplacian pyramid network with dual-polarization feature fusion (SE-

LPN-DPFF). This first merges features extracted from both polarizations, then uses squeeze and excitation blocks to enhance feature representation by selectively emphasizing informative features. Finally, the features are fed into a Laplacian pyramid network based on residual blocks, which extracts information from different resolutions and performs classification ensuring that the model is trained on this fine-grained information. Two datasets from OpenSARship are used, one with three categories (bulk carriers, container ships, and tankers) and another with six categories (bulk carriers, cargo, container ships, fishing, general cargo and tankers). To ensure a balanced dataset, the authors use 70% of the data from the category with the fewest samples as the training set and the remaining data for testing. They compare the performance of SE-LPN-DPFF to 35 other DL classifiers. The results show that SE-LPN-DPFF outperforms all other models.

A different approach is proposed in [53], using a dual-polarized network based on a pretrained VGG architecture. This network learns representations from both VV and VH polarizations and then applies bilinear pooling. The authors train the model on a three-class dataset from OpenSARShip (tanker, container ship, and bulk carrier), where VV and VH images are paired for training. The results show that the model outperforms state-of-the-art models, achieving an accuracy of 88%. A novel methodology, named the group bilinear convolutional neural network (GBCNN), is proposed in [56], [63], combining a loss function and bilinear pooling, based on dense blocks. Grouped VV and VH images are used to train two parallel models. The extracted features are then passed through bilinear pooling. The resulting features from both pooling layers are sent to the loss function, along with a third feature created by merging the pooling outputs. Using three- and five-class datasets from OpenSARShip (tanker, container ship, bulk carrier, cargo, general cargo), with data augmentation techniques such as flipping, rotation, translation, and Gaussian noise, the GBCNN achieved accuracy scores of 88% and 67% for three and five classes, respectively.

Recent work has begun to explore the use of complex-valued SAR data for ship classification, leveraging both amplitude and phase information rather than relying on intensity alone. The authors of [89] introduce FDC-TA-DSN, a deep SAR network that combines four-dimensional dynamic convolution (FDC) to suppress speckle noise with a triple-attention (TA) mechanism to filter sea clutter and emphasize discriminative features. Phase information is incorporated through joint time-frequency analysis and autoencoder-based feature learning, and fused with amplitude features to enhance classification. To support this effort, a new high-resolution dataset, ComplexSAR\_Ship, was constructed from GF-3 imagery, covering 17 ship categories. Experiments show that FDC-TA-DSN consistently outperforms standard CNNs and DSN baselines, confirming the value of phase information and complex-valued features for more robust ship classification.

In [58], polarization is investigated in conjunction with handcrafted features. A novel polarization fusion-based geometric features network (PFGFE) is proposed, utilizing geometric features alongside VH-polarized, VV-polarized and a combined VV-VH image. PFGFE employs three VGG network-based CNNs for feature extraction. A global feature self-attention mechanism is then applied to the output of each CNN. Subsequently, the extracted features are merged, and their classification probabilities are concatenated with corresponding coefficients. Geometric features are processed using LSTM to capture the importance of long-term dependencies. Finally,

the outputs from the CNNs and LSTM are concatenated, and this final representation is used for classification. The authors evaluate PFGFE on three-class and six-class datasets from OpenSARShip (respectively, bulk carrier, container ship, and tanker, and bulk carrier, cargo, container ship, fishing, general cargo, and tanker). The results show accuracies of 79.84% and 56.83% respectively.

Building on the idea that exploiting dual polarization information can enhance performance, the authors of [62] propose a novel dual-polarization information-guided network (DPIG-Net) that leverages a polarization channel cross-attention framework (PCCAF) and a dilated residual dense learning framework (DRDLF) for effective feature extraction and fusion. It evaluates DPIG-Net on the OpenSARShip dataset using VH and VV polarizations for three (bulk carrier, container ship, tanker) and six classes (adding cargo, fishing, and general cargo), achieving accuracies of 81.28% and 58.68%, respectively. Smoothed polarization fusion, coupled with attention modules such as multiscale feature fusion and dynamic gating feature fusion, has been shown to improve model accuracy for ship classification. This approach is adopted in [80] on OpenSARShip, achieving accuracies of 87.13% and 65.97% for three and six classes, respectively. These results suggest that smoothed polarization fusion can be more effective than coherence fusion [62]. In [73], a novel dual branch deep network for ship classification is proposed and tested on OpenSARship. The network leverages VH and VV polarizations and comprises three modules: (1) an image construction module that generates two pseudo-RGB images representing polarimetric and texture features, (2) a feature extraction module that employs two pretrained ConvNeXt networks to extract features from the constructed images independently and (3) a feature fusion and classification module that combines the extracted features and performs classification. This network achieves accuracies of 87.62% and 64.26% for 3-class and 6-class classification tasks, respectively.

A Multiscale Global Scattering Feature Association Network (MGSFA-Net) is proposed in [83]. It integrates scattering features with deep features for SAR ship classification. After fine segmentation to isolate ship targets, scattering centers (SCs) are extracted and modeled as local graphs, which are then associated through the SCFA module and enhanced with a multiscale fusion module. In parallel, deep features are extracted using a multikernel module, and both representations are fused by weighted integration. Experiments on FUSAR-Ship and OpenSARShip demonstrate that MGSFA-Net achieves accuracies of 85.3% (three-class FuSARShip) and 85.7% (three-class OpenSARShip), outperforming other feature-fusion methods. The results show that incorporating scattering characteristics improves robustness and interpretability, particularly in few-shot scenarios.

A Cross-Polarimetric Interaction Network (CPINet) is proposed in [81] to exploit the complementary information in dual-polarized SAR images. The network employs a multiscale deep feature extraction framework and introduces a mixed-order squeeze-excitation (MO-SE) attention module, where first- and second-order statistics from one polarization guide the learning of the other. To further enhance feature fusion, factorized bilinear coding (FBC) is adopted, which reduces the parameter size while suppressing noise. An adaptive loss balancing strategy based on GradNorm is also applied to optimize multitask training. Evaluated on three- and five-class subsets of OpenSARShip, CPINet consistently outperforms both single-pol baselines and prior dual-pol methods, demonstrating the effectiveness of polarimetric feature interaction for robust ship

classification.

The limited spatial resolution of SAR imagery presents a significant challenge for ship classification tasks, particularly for small vessels. These ships often occupy only a few pixels within the image, making it difficult to extract the features needed for an accurate classification. To address this, the authors of [30] propose a joint convolutional network to enhance the spatial resolution before classification. It uses TerraSAR-X images for pretraining and three classes from OpenSARship for training (bulk carrier, container, tanker). The data are first fed into a generative model for upscaling, and then classified using a dense blocks-based convolutional network. It also compares the performance of models trained on VV and VH polarizations. The results show an increase in performance compared to simpler networks, achieving classification accuracies of 81.02% and 79.12% for VH and VV, respectively.

Methods such as the siamese networks can be helpful in overcoming the challenge of scarce and imbalanced data. Siamese networks focus on learning a similarity function by training two networks that share weights, allowing for better learning of data representations. In [32], a novel loss function with a siamese network architecture is proposed. Two inception blocks are used on different polarizations (VV and VH) for four classes from OpenSARship (cargo, tanker, tug, and other). The images are first passed through a target recognition algorithm (RCNN) to crop the targets. Then, the cropped images are fed into the inception blocks, the features are extracted and the loss is calculated. The results indicate that siamese networks with target detection outperform models without target detection, achieving an accuracy of 87.04%.

In ship classification tasks, researchers typically use datasets with three or four categories. This is because DL models require a large amount of data for training, and maritime data are inherently imbalanced, making it challenging to achieve good performance with more classes. However, a study in [49] demonstrates the potential of feature fusion for handling a larger number of categories. It employs an eight-category dataset from OpenSARShip (cargo, tanker, carrier, container, fishing, dredger, tug, and passenger). A VGG-based feature extractor is used to simultaneously extract feature maps from both VV and VH polarizations. These feature maps are then merged. Finally, a novel hybrid loss function is applied to reduce intra-class distances and increase inter-class distances, leading to improved classification performance. The results, with a mean average precision of 82%, support the effectiveness of this approach.

#### D. Transfer Learning and Domain Adaptation

High resolution images (e.g. from CosmoSkyMed mission, level 1A HIMAGE single-look complex slant (SCS) in the X-band, single HH or VV polarization) can be exploited to test the effectiveness of pre-trained models on datasets that have a limited number of classes (e.g. three). In [26] the authors follow this approach and conclude that fine-tuning a VGG-16 network can lead to F1-scores of up to 97.8%. This shows the usefulness of DL techniques for small-sized datasets (like usual SAR imagery datasets).

In [28], the authors demonstrate that leveraging transfer learning with pre-trained models (based on the DenseNet-121 architecture), combined with effective data augmentation strategies, can significantly improve the classification accuracy, even when working with small-sized training datasets. Stripmap input images were sourced from TerraSAR-X and manually annotated by experts using AIS data. Only three classes are employed (bulk carrier, container ship, and oil tanker), and

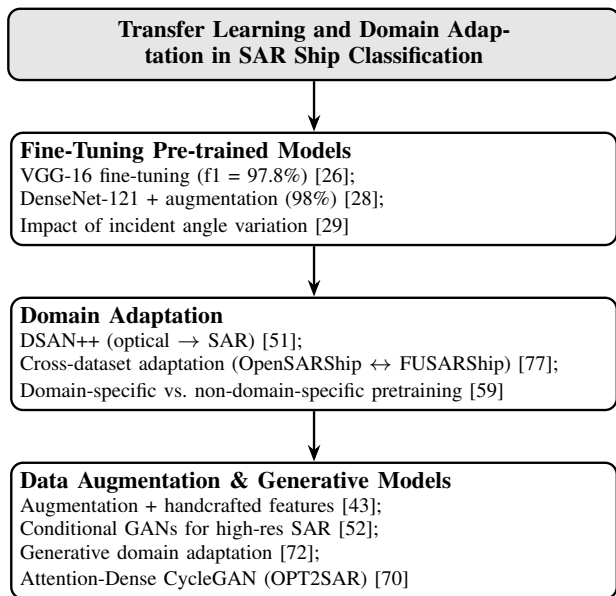


Fig. 6: Overview of transfer learning strategies for SAR ship classification. The approaches are grouped into three categories: (i) fine-tuning of pre-trained models, (ii) domain adaptation across modalities or datasets, and (iii) data augmentation and generative models to mitigate data scarcity. Representative studies are indicated by citation keys.

data augmentation techniques are applied (flipping, brightening, sharpening, random cropping, and rotating). Despite the tiny size of the training dataset, a 98% accuracy was achieved.

In [29], the authors examine the effect of different incident angles on model performance. They use a dataset containing three classes of images taken from Cosmo-SkyMed and GaoFen-3, encompassing incident angles ranging from 20 to 60 degrees. The data were divided into three overlapping subsets: one containing the entire dataset, a second one containing data with small incident angles (less than 48 degrees), and a third one containing data with large incident angles (more than 45 degrees). All three datasets were tested on four pre-trained DL models: VGG-16, VGG-19, ResNet-50, and ResNet-101. The results indicate that incorporating data from all available angles enhances the models' predictive accuracy.

Data scarcity can be addressed by training the model on domain-specific data, such as optical images of ships, then applying domain adaptation to SAR data. In [51], DSAN++, a novel method for domain adaptation, is proposed. This first involves training on optical images and then transferring the learned knowledge to SAR data. DSAN++ utilizes two ResNet networks for feature extraction. The extracted features undergo feature fusion, followed by the application of a subdomain adaptation network to learn transferable representations. To capture both local and global distributions, shallow and deep ResNet models (ResNet-18 and ResNet-50) are integrated with the convolutional block attention module. The experiments were conducted by training the model on an optical remote sensing (ORS) dataset obtained from Google Earth, with four classes (cargo, bulk carrier, container ship, and oil tanker). Subsequently, domain adaptation was performed on two datasets: (1) GF-SAR with three classes (bulk carrier, container ship, and oil tanker) collected from FUSAR and GaoFen, and (2) a three-class dataset (cargo, container ship, and oil tanker) derived from TerraSAR-X. The results demonstrate that DSAN++ leads to improved performance in subdomain adaptation. In [77], the authors apply domain adaptation from OpenSARship to

FuSARShip and vice versa on multiple scales resulting in better performance for both datasets as compared to the one obtained by training and testing on a single dataset. In [59], the authors compare the fine-tuning performance of models pre-trained on domain-specific versus non-domain-specific datasets using CNNs. The methodology was evaluated on three-class targets from VV polarized TerraSAR-X (carrier, container, and oil tanker), and four-class targets from FUSARShip (cargo, bulk carrier, container, and oil tanker). The results suggest that shallow models with feature refinement can achieve better performance when fine-tuned on domain-specific models. The authors investigate the pre-training effects of ImageNet and an ORS dataset with data augmentation techniques (scaling, flipping, and rotating). Whereas ImageNet, due to its large size, yields better performances on deep networks, the ORS dataset enhanced the performance of shallow networks, indicating that shallow models are more effective with less data.

### E. Data Scarcity Mitigation

Techniques like data augmentation, when combined with transfer learning, can address data-related problems. In [43], the authors demonstrate that using transfer learning with data augmentation and manual features can improve the performance of a DL model. They use ground range detected SAR data acquired in interferometric wide-swath mode from OpenSARShip. The data consist of two channels, with VH and VV polarizations, stacked together. The study focuses on classifying ships into four categories: bulk carrier, container ship, fishing vessel, and tanker. The authors use five data augmentation techniques (horizontal flipping, vertical flipping, rotation by 90°, rotation by 270°, and Gaussian noise injection) and incorporate 14 handcrafted features. The effectiveness of these handcrafted features combined with DL features are compared across three machine learning and four DL models. The ML models are Radial Basis Function (RBF)-SVM with two stage feature selection, SVM classifier combination, and SVM one-vs-rest; the DL models are CNN, AlexNet, VGG-19, and ResNet50. Additionally, the authors present a methodology for matching AIS and SAR data and test the performance on real data collected from three different locations. Their results show that the pretrained ResNet50 model, when fine-tuned on SAR images and handcrafted features, achieves an accuracy of 82.5%.

One of the possible DL-based solutions to the data scarcity problem in optical images is represented by generative models and can also be applied to SAR images. These models learn data representations and produce images that resemble the originals, effectively increasing the available data. In [52], the authors propose a novel conditional generative adversarial network method (GAN) that generates high-quality SAR images, improving ship classification performance. The method consists of two parts: (1) a residual convolutional block is integrated with the generator to enhance the detail of generated images, and (2) the generated images are merged with real images and sent to a discriminator for two tasks: distinguishing real from fake images and classifying them into three classes. The authors use a three-class dataset (cargo, container, tanker) from GaoFen-3. The results indicate that the GAN is capable of generating high-resolution images that, coupled with the real images, enhance the classification performance.

A generative model-based domain adaptation framework for ship classification on a 5-class FUSARShip dataset is proposed in [72]. This framework leverages domain adaptation by transferring knowledge from the optical images of the FGSCR42 dataset. In [70], the Attention-Dense CycleGAN

(ADCG) is introduced to improve recognition by leveraging transfer learning from optical to SAR domains (OPT2SAR). ADCG enhances CycleGAN with a dense connection module (DCM) and a convolutional block attention module (CBAM). Using FUSARShip and the FGSCR optical dataset, ADCG achieves superior performance, evaluated via Fréchet inception distance (FID) and kernel inception distance (KID). It demonstrates a 6% improvement in recognition accuracy for SAR ship classification networks, thus validating its effectiveness in addressing class imbalance and sample scarcity through effective transfer learning.

#### IV. FINDINGS

##### A. Datasets

The literature explores two public datasets, FUSARShip and OpenSARShip, and some additional, distinct data sources for SAR ship classification. More detailed information is provided below.

1) *OpenSARShip*: The OpenSARship dataset is a publicly available resource designed to support research in ship classification using SAR imagery. This dataset provides a collection of pre-processed SAR image chips specifically focused on ship detection and classification tasks. The current version contains 34528 vessel instances (Table III). The data are collected using two products from ESA's Sentinel-1 repository (single look complex, and ground range detected) in interferometric wide swath mode. The dataset consists of 16 ship typologies, among which classes like Cargo, Tanker, Fishing, Tug, and Dredging can be found. However, the dataset suffers from class imbalance, with 61% data belonging to Cargo category. Three classes represent 95% of data, while the remaining 13 classes only represent a 5% of the data provided. This imbalance can lead any model to bias toward the majority class. The literature shows that researchers often address this problem by only using the 3 to 5 most populated classes, e.g., cargo, tanker, tug, fishing, and dredging [27], [33], [38], [40], [50], [64], [67], or by utilizing subclasses within the cargo category, e.g., general cargo, container ship, bulk carrier [34], [37], [54]. OpenSARship data incorporate polarization information, VH and VV, which can be leveraged to create fusion networks [44], [53], [56]–[58], [60], [62], [63], [73]. The dataset's main weaknesses such as low spatial resolution (20 m/pixel) and class imbalance, limit the performance of any DL model. Nevertheless, if these limitations are addressed, OpenSARship serves as a valuable starting point for researchers to explore their methodologies before acquiring commercial SAR imagery.

2) *FUSARShip*: FUSARShip provides a valuable resource for researchers in SAR ship classification. This dataset uses high-resolution GF-3 SAR images, specifically designed for ocean remote sensing and marine monitoring tasks. FUSARShip is constructed from a collection of image chips extracted from over 126 GF-3 scenes, encompassing 15 primary ship categories and 98 subcategories (Table IV). Each ship instance has a size of  $512 \times 512$  pixels, with a pixel size of 0.5 meters. Similar to OpenSARship, FUSARShip presents a challenge due to class imbalance. Three sub-categories constitute approximately an 81% of the data, while the remaining 95 categories collectively only represent a 19% thereof. This significant imbalance can lead to models exhibiting bias towards the overrepresented classes. As observed for the OpenSARship dataset, researchers often address this issue by focusing on the 3-5 most populated classes, e.g., cargo, fishing, tanker, bulk carrier [51], [59]–[61], [65], [70], [72].

TABLE III: OpenSARShip 2.0: key dataset facts. *Notes: Type levels reflect label-information granularity (rough: AIS-only; elaborate: AIS plus MarineTraffic when available and verified). Interference tags flag low-quality chips (e.g., multiple targets, near-shore clutter, or land contamination). Label reliability summarizes expected label and geolocation accuracy after AIS-SAR matching and any verification.*

Item	Value
Platform	Sentinel-1
Products	52 GRD + 35 SLC (87 total)
Ship chips	34,528
Resolution / spacing	GRD: $20 \times 22$ m, $10 \times 10$ m; SLC: $2.7\text{--}3.5 \times 22$ m, $2.3 \times 17.4$ m
Polarization	VV/VH
Labels	AIS; MarineTraffic verification when available
Type levels	Rough; elaborate when available
Interference tags	Dense ships; inshore chips
AIS-SAR matching	Semi-automatic; manual correction
Label reliability	73% ship chips are within a 100 m distance error range
Formats	TIFF (orig./calib.); grayscale/pseudo-color visualizations
Pros / cons	Large-scale, publically available, interference labels / imbalance, medium resolution
Use cases	Sentinel-1 ship-type classification under interference, prompt maritime traffic monitoring [107]

TABLE IV: FUSAR-Ship: key dataset facts. *Notes: Labels are derived from AIS; label reliability summarizes expected correctness (especially near shore and in dense traffic, where manual correction may be required).*

Item	Value
Platform	GaoFen-3
Mode / scenes	UFS; 126 scenes
Ship chips	>5,000 with AIS
Resolution	Az.: 1.124 m; Rg.: 1.700–1.754 m
Polarization	HH/VV
Taxonomy	15 main categories; 98 subcategories
Chip size	$512 \times 512$
Labels	AIS-based category/subcategory;
Label reliability	Near-shore dense cases need manual correction; accuracy 94-70%
Pros / cons	High-res, publically available, diverse scenes / Moderate Imbalance
Use cases	Ship recognition and related maritime monitoring tasks

3) *Other Data Sources*: Large datasets are crucial for training a DL model. However, data imbalance can hinder model performance. Researchers address this challenge using two techniques. Firstly, they leverage high-quality data from various satellite missions, such as TerraSAR-X [25], [28], [30], [51], [59], CosmoSkyMed [26], PALSAR-2 [31], and GaoFen-3 [29], [39], [45], [52], see Table V. These missions not only provide high quality data, but also allow researchers to exploit AIS to strategically select images based on specific locations and timeframes. For instance, if a researcher aims to develop a model for detecting fishing activity during a particular season in a specific region, AIS data can be analyzed to identify relevant satellite images for purchase.

Secondly, due to the cost associated with acquiring high-resolution commercial imagery, researchers often employ transfer learning techniques. This involves training a model on a large, general-purpose dataset (e.g., ImageNet), or domain

TABLE V: SAR Platforms and their resolutions

Platform	Band	Image Modes	Resolution(s)
Sentinel-1	C	StripMap, Interferometric Wide Swath (IW), Extra Wide Swath (EW)	- StripMap: $5 \times 5$ m - IW: $5 \times 25$ m - EW: $20 \times 40$ m
Cosmo-SkyMed	X	Spotlight, StripMap, Wide	- Spotlight: Up to 1 m - StripMap: Up to 3 m - Wide: Up to 30 m
TerraSAR-X	X	StripMap, Spotlight, High Resolution Spotlight (HRS)	- StripMap: Up to 3 m - Spotlight: Up to 0.3 m - HRS: Up to 0.3 m
Gaofen-3	X	StripMap Spotlight	- StripMap: Up to 2.5 m - Spotlight: Up to 0.9 m
PALSAR-2	L	StripMap Spotlight	- StripMap: $3 \times 3$ m - Spotlight: $1 \times 3$ m

specific dataset, e.g., MSTAR [105], optical remote sensing images of ships taken from Google Earth [108] and then on a smaller, task-specific dataset (e.g., FUSARShip). Pretraining on a comprehensive dataset equips the model with the ability to extract generic features from images. Subsequent fine-tuning on the target dataset refines these features for the specific classification task.

### B. Data Augmentation

Due to the limited availability of SAR images and the inherent data requirements of DL models, data augmentation techniques have been proposed in the literature to address this challenge. These techniques can be employed in an attempt to improve the performance of DL models for SAR ship classification.

1) *Rotation*: in the literature, rotation is widely employed to enhance the robustness of DL models for SAR ship classification [25], [38], [43], [56], [60], [61], [63]. It involves artificially rotating the training images by various angles to simulate real-world scenarios where objects might be oriented differently relative to the SAR sensor. A common approach is to rotate images within a range of  $+45$  to  $-45$  degrees, as this range captures a significant portion of the potential ship orientations. This augmentation strategy helps the model learn features that are invariant to rotation, improving its generalization ability and reducing the risk of overfitting to the specific orientations present in the original training data.

2) *Flipping*: the literature commonly employs horizontal and vertical flipping data augmentation techniques [28], [43], [57], [59]–[61], [63]. These techniques generate new images by mirroring the ships along the horizontal or vertical axes. This process simulates scenarios where ships are viewed from opposite directions, effectively augmenting the dataset and potentially improving the model's ability to generalize to unseen orientations.

3) *Scaling*: DL networks typically require a fixed-size input. Image scaling, a data augmentation technique that resizes images to a standard size, addresses this requirement. Scaling can be performed by increasing or decreasing image size, with larger images demanding significantly more computational resources for processing [59], [60]. In SAR ship classification, a common practice is to scale images to a size of  $128 \times 128$  pixels, balancing accuracy and computational efficiency.

4) *Translation*: in image processing, translation refers to the spatial transformation of an image by a specific offset in both horizontal and vertical directions. This essentially moves the entire image content without any scaling or rotation. In SAR ship classification, translation augmentation can be employed to artificially introduce variations in the training data [25], [38], [56], [63]. By exposing the DL model to images with slightly

shifted ship positions, translation augmentation can potentially improve the model's generalization ability. This is because the model learns to recognize ships irrespective of their small location variations within the image, making it more robust to real-world scenarios.

5) *Gaussian Filtering*: SAR images of marine scenes are often corrupted by a strong noise component. To address this challenge, Gaussian filtering is commonly employed [38], [43], [56], [63]. This technique works by averaging the intensity values of pixels within a localized neighborhood around each pixel. This process reduces the random variations in pixel intensity caused by noise, enhancing the distinction of ship features from the background noise.

6) *Cropping*: maritime SAR imagery contains a lot of information that is useless for ship classification purposes. In fact, the representation of the ship target is usually way smaller than the surrounding background, and this leads the model to learn non-informative details. As a possible countermeasure to reduce useless information, cropping, i.e. the process of extracting a region of interest containing the ship target only, is adopted. Cropping is commonly employed in literature as a data augmentation technique [24], [28], [32], [60].

### C. Handcrafted Features

While data augmentation is often used to increase training diversity, it can also lead to overfitting, especially when applied to low-quality SAR images. In such cases, augmentation may generate additional noisy samples that confuse the model rather than improve it. Handcrafted features provide an alternative source of information that can be fused with DL features to strengthen classification robustness (see Fig. 9).

DL models excel at automatically extracting features, yet their performance in SAR ship classification is constrained by data scarcity, imbalance, and strong speckle noise. To address these issues, researchers have integrated handcrafted features that embed prior knowledge about ship structure and radar scattering. Below, we review the most commonly used handcrafted features in the literature.

1) *Histogram of oriented gradients*: Histogram of Oriented Gradients (HOG, see Fig. 7) captures edge orientations and local gradient structures [35], [41], [50]. In SAR images, where ship targets typically exhibit strong geometric edges, HOG can highlight hull and superstructure outlines. However, speckle noise can degrade gradient estimation, making HOG alone unreliable. Recent works therefore combine HOG with DL features, where it contributes complementary edge information that can guide the network's learning of ship contours.

2) *Local Radar Cross-Section*: Local Radar Cross-Section (RCS) features describe scattering intensities within specific image regions. These features directly capture radar-domain

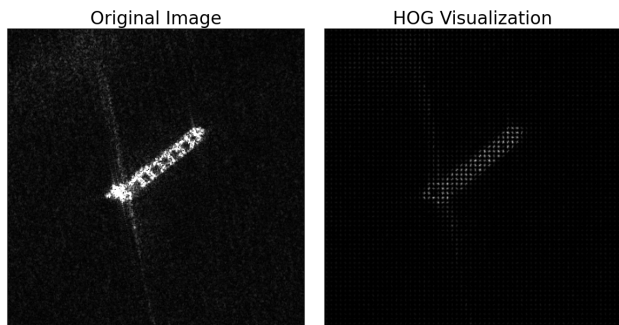


Fig. 7: HOG features of a SAR image

priors that are not always learned effectively by DL models. When fused with learned features, local RCS provides physical interpretability and helps reduce confusion between ships with similar shapes but different scattering patterns [41]. Nonetheless, RCS features are sensitive to resolution and viewing geometry, which may limit generalization across datasets.

3) *Geometric Features*: Geometric features (such as length, width or related combinations) can be proficiently exploited for classification. During training with annotated datasets, these features can be drawn from the AIS information [36]. During operational classification, image processing algorithms can be applied to segment ship targets (e.g. through bounding box fit) and, provided that the resolution is known, estimate the geometric variables of interest [45], as shown in Fig. 8. Length and width can be used to extract additional features [41], such as area, perimeter, aspect ratio, shape etc. Adding this information can help DL models to improve the ship classes prediction.

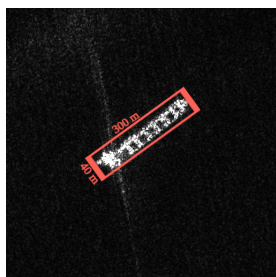


Fig. 8: Bounding Box of a ship target in a SAR image chip.

Overall, handcrafted features act as complementary signals rather than stand-alone solutions. HOG captures edge structure but is noise-sensitive, RCS injects radar-specific priors but depends on resolution, and geometric descriptors are highly discriminative yet prone to segmentation errors. Their effectiveness lies in feature fusion with DL models, especially under data scarcity and imbalance, where handcrafted cues can provide stability and interpretability.

#### D. Training Techniques

Although handcrafted features can be helpful, they require high-resolution images for effective extraction, which is often not the case, as seen in the OpenSARShip dataset. To address this challenge, researchers have explored novel DL techniques, further explained below.

A DL model takes input in the form of an image, learns the features, and then predicts the class of the image. However, different training strategies can be attempted to achieve promising classification performances. The techniques used in the literature for better classification are described in this section.

1) *Feature Fusion*: a DL model trains on the features that it extracts, but when it comes to SAR ship classification, the challenge of data scarcity becomes a hurdle on the way to better results. To help the model understand the data well, several researchers propose to add handcrafted features to the DL model [35], [50], [67], [74], [75]. Deep and handcrafted features are merged and the model classifies on this merged information, thus obtaining a better classification performance [41]. A merging strategy is schematized in Fig. 9.

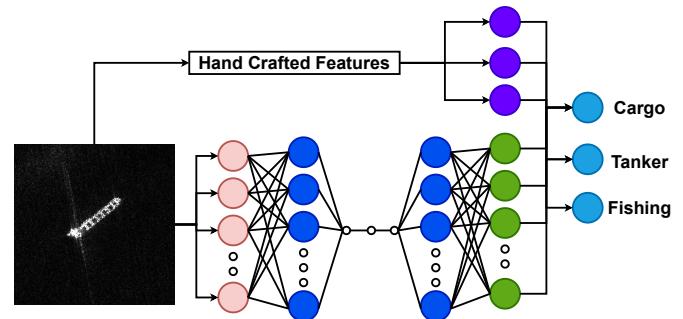


Fig. 9: Feature Fusion

Feature fusion is a useful technique; however, it requires manual effort, as there is no existing pipeline to extract these features automatically. Additionally, there are no standardized features that consistently enhance classification performance. It is important to note that the choice of features is resource- and time-intensive and does not always lead to better results.

2) *Polarization fusion*: SAR images can be acquired in four transmit-receive polarizations: VV, VH, HH, and HV. While most studies in the literature have focused solely on VV or VH polarizations, it is important to note that each polarization provides distinct information about the target. Utilizing multiple polarizations has been shown to improve classification performance. This is because combining information from different polarizations allows the model to learn more discriminative features, leading to enhanced classification accuracy. A common approach in the literature involves creating image pairs and training two separate DL networks simultaneously, one for each polarization [32], [44], [53], [58], [62]. As illustrated in Fig. 10, the features extracted from these networks are then merged for final classification.

Exploiting the polarization information is indeed a useful technique that provides the model with more data to learn from; however, this is not always readily available. From a cost perspective, researchers must purchase two images instead of one, significantly increasing the acquisition cost. For public datasets, polarization is available only in the OpenSARShip

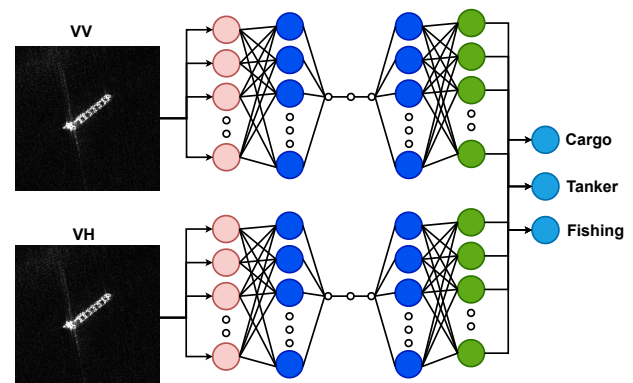


Fig. 10: Polarization Fusion (simplified version from [53])

dataset and not in the FUSARShip dataset, limiting experimentation. Furthermore, training two models doubles the training cost and increases resource requirements.

3) *Metric Learning*: DL models rely on feature extraction, which involves iteratively adjusting the model's weights to optimize the extracted features. Ideally, for effective classification, the features corresponding to each class should cluster tightly in a low-dimensional space, while remaining well-separated from features of other classes. However, standard training procedures may not inherently prioritize this inter-class separation. In such cases, metric learning techniques are employed to explicitly address this challenge [34], [38], [65]. Metric learning aims to embed the extracted features into a low-dimensional space while maximizing the similarity between features belonging to the same class and minimizing the similarity between features from different classes. This is often achieved by utilizing distance metrics like cosine similarity or Euclidean distance, as illustrated in Fig. 11. Metric learning helps the model clearly distinguish between features of different classes, leading to superior performance. However, since the data is often naturally imbalanced, metric learning may be prone to overfitting. Therefore, it is crucial to balance the data beforehand or account for the imbalance when applying metric learning.

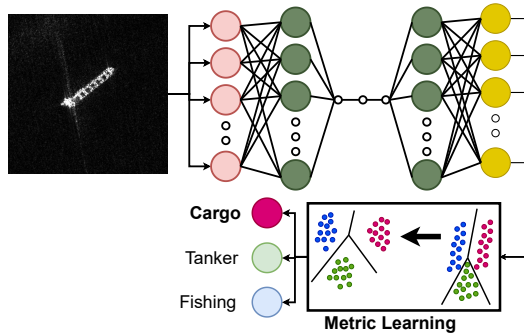


Fig. 11: Metric Learning

4) *Multi-Scale/Pyramid Training*: A DL model consists of several layers; each layer learns different scaled features, and the model makes the prediction based on the features of the last convolution layers [37], [40], [44], [68], [69], [71], [92]. In pyramid training, see Fig. 12, the features from all the layers are directly used in decision-making by merging them using concatenation or other statistical measures (e.g. average, addition etc.). Although the multiscale technique is effective at introducing features from different layers, providing the model with more data to learn from, deciding which layers to use is a critical aspect. Including all layers may introduce irrelevant

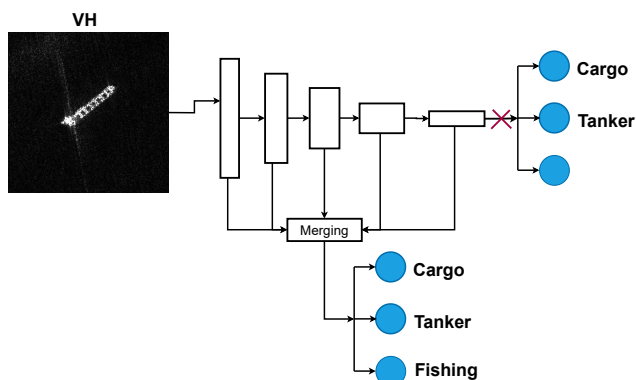


Fig. 12: Pyramid networks

information, leading the model to learn unhelpful patterns. Additionally, the method of merging features significantly impacts performance. For instance, concatenating features increases the feature size, while summing features may result in the loss of important information. Features from different scales often have varying sizes, which must be upsampled or downsampled for merging. Therefore, adopting an appropriate approach to align feature sizes is crucial for achieving better classification performance.

5) *Transfer-Learning and Domain Adaptation*: SAR data are expensive and difficult to access, while DL models require large amounts of data. A deep neural network learns how to extract features and then applies classification to these learned features. Transfer learning can be beneficial in such scenarios. It allows leveraging other datasets to help a network learn feature extraction and then fine-tune the network with the SAR data. Transfer learning can be applied in various ways, see Fig. 13:

- Using an optical image dataset, such as ImageNet, to pre-train a network, then fine-tuning on the SAR dataset [26], [28], [51], [59].
- Utilizing open-access SAR datasets for pre-training before fine-tuning [54], [68].
- Merging optical and SAR imagery to see whether the model performs better at feature extraction, thus leading to improved overall performance [52], [65], [70], [72], [90].

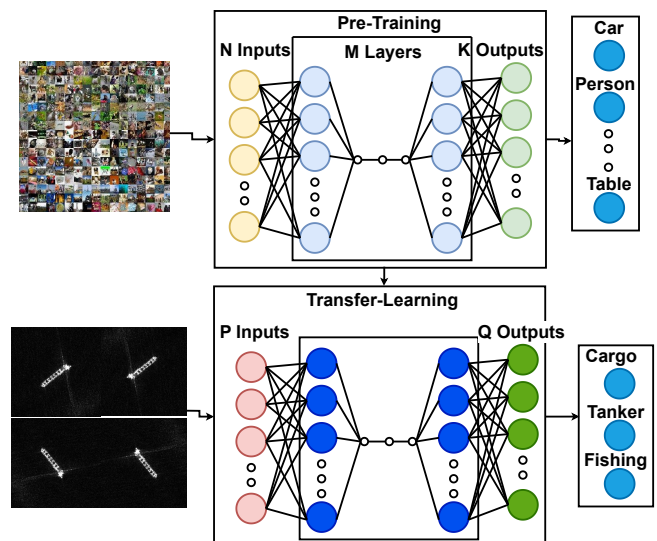


Fig. 13: Transfer Learning

Transfer learning comes with the challenge of selecting the appropriate model, as numerous architectures are pretrained on large datasets. However, these models may not perform well on SAR ship classification due to the inherent imbalance in SAR data. Additionally, the depth of the networks can lead to varying classification performances. Furthermore, the nature of the data on which the model was pretrained can significantly impact its performance. Therefore, it is crucial to carefully evaluate the pretraining dataset before utilizing any pretrained network.

6) *Learning Rate management*: One crucial component of DL is the learning rate, which helps the network to update its weights efficiently. Traditionally, the learning rate is kept fixed during the entire training. However, as a deep neural network progresses through training iterations, it gradually learns information from the data. A constant learning rate in this scenario can slow down training or potentially cause the model to miss the point of lowest error. To address this issue, a strategy was introduced in [54] that changes the learning

rate based on the number of iterations. This approach reduces the learning rate as training progresses, facilitating convergence towards the optimal weights. This technique not only reduces the training time but can also improve the overall performance of deep neural networks. In tasks like SAR ship classification, where data availability is limited, such dynamic learning rate adjustments can be particularly beneficial for achieving better performances.

7) *Searched Networks*: The architecture of a DL model is traditionally determined through the researcher's intuition or by replicating existing architectures. However, Neural Architecture Search (NAS) offers a more objective approach to identify the optimal architecture for a specific task. NAS is a search algorithm that explores a range of possible network configurations to identify the one that performs best for the given task. This approach empowers researchers to discover the most suitable architecture for their data, potentially leading to significant advancements in maritime monitoring [57], [60]. While searched networks are an excellent option, they come with a significant computational cost. The more the layers included in the search process, the higher the computational demand. Therefore, it is crucial to carefully consider the computational cost before applying NAS.

### E. Comparative Analysis of Technique Families

The reviewed literature reveals distinct performance characteristics across major technique families, with effectiveness varying significantly under different data constraints.

1) *Handcrafted Feature Integration*: Handcrafted features (HOG, RCS, geometric descriptors) demonstrate consistent value under data scarcity. Studies report performance gains of 3–7% when fusing handcrafted features with deep features [35], [41], [50], [67]. However, effectiveness depends critically on image resolution: geometric features require sufficient pixel resolution for accurate extraction; their effectiveness is thus reduced when dealing with medium- and low-resolution datasets such as OpenSARShip.

Several methodological concerns emerge from the reviewed studies. First, many papers report overall accuracy improvements without providing class-wise performance breakdowns, potentially masking overfitting to majority classes while handcrafted features contribute minimally to minority class recognition. Second, the choice of which handcrafted features to integrate often lacks principled justification, studies typically select features based on prior work in optical ship recognition without validating their discriminative power on SAR imagery. Third, feature fusion strategies vary widely (early fusion, late fusion, attention-weighted fusion) with inconsistent reporting of fusion architecture details, making it difficult to determine whether performance gains stem from the features themselves or from increased model capacity. The finding that handcrafted features prove most valuable with shallow networks [67], where VGG-11 and ResNet-19 exhibit larger gains than their deeper counterparts, suggests that their primary value may be compensating for limited representational capacity rather than providing SAR-specific information that deep networks cannot learn. This raises the question of whether properly regularized deeper networks might achieve similar performance without engineered features.

2) *Polarization Fusion*: Dual-polarization fusion (VV/VH) consistently improves classification accuracy when polarimetric data are available. Methods achieve 86–89% F1 for three-class problems [44], [53], [56], representing substantial improvement over single-polarization baselines. However, polarization fusion

shows diminishing returns with increased class complexity: five-class F1 scores drop to 57–58% [56], [63], and six-class performance ranges only 51–58% [62], [80]. While polarization provides complementary scattering information for ship type discrimination, it does not remediate class imbalance, which arises from insufficient minority-class training samples. The sharp performance decline beyond five classes demonstrates this fundamental limitation: additional polarization channels cannot compensate for inadequate sample coverage in minority classes.

Critical examination reveals competing explanations for polarization fusion effectiveness that remain unresolved in the literature. Some studies attribute improvements to physically meaningful scattering differences between ship types, while others achieve similar gains through simple channel concatenation without explicit physical modeling, suggesting that performance may derive primarily from increased feature dimensionality rather than polarimetric information per se. The reviewed studies exhibit inconsistent polarization preprocessing: some apply calibration and speckle filtering before fusion, others fuse raw intensities, and few report ablation studies isolating the contribution of preprocessing choices. The sharp performance decline beyond five classes reveals that dual-channel fusion provides marginal gains that are overwhelmed by fundamental data scarcity in fine-grained regimes. The technique requires dual-channel acquisition, doubling data cost and limiting applicability to datasets offering both polarizations, yet cost-benefit analysis comparing dual-pol acquisition against simply collecting more single-pol training samples is absent from the literature.

3) *Metric Learning and Siamese Networks*: Metric learning approaches excel in few-shot and highly imbalanced scenarios by learning discriminative feature embeddings rather than direct classification boundaries. Siamese networks achieve remarkable F1 scores on 16-class OpenSARShip data [55], substantially outperforming standard CNNs. Metric learning combined with CNNs shows consistent gains over CNN-only approaches in three-class scenarios [34], [38]. These methods prove particularly effective in imbalanced settings, as they leverage similarity comparisons rather than requiring dense class coverage.

Studies claiming metric learning superiority often compare against baseline CNNs without equivalent regularization, data augmentation, or hyperparameter tuning, confounding the comparison. When controlled comparisons are performed with properly tuned baselines, metric learning advantages narrow substantially [38]. The synthesis of competing approaches suggests that metric learning provides genuine value primarily in extreme imbalance (>100:1 class ratios) or true few-shot scenarios (<10 samples per class), but for moderately imbalanced datasets with hundreds of samples per class, well-regularized standard classifiers with appropriate loss functions may offer equivalent performance at lower computational cost.

4) *Multi-Scale and Pyramid Architectures*: Multi-scale methods demonstrate the most consistent cross-dataset generalization. They achieve similar performance on OpenSARShip and FUSARShip (84% vs 86% F1 for three classes) [83], and sustain strong results even at seven classes (86–90% F1) [50], [71]. The approach proves robust across varying resolutions, with effectiveness maintained from Sentinel-1 (20 m/pixel) to GaoFen-3 (0.5 m/pixel) imagery. Multi-scale feature aggregation compensates for information loss between network layers, yielding 5% improvement over single-scale baselines [37].

*Synthesis and Critical Recommendations*: Under severe data scarcity ( $\leq 50$  samples per class), handcrafted feature fusion or metric learning provides reliable gains, though practition-

ers should be aware that these approaches may primarily compensate for insufficient model capacity or regularization rather than providing irreplaceable SAR-specific information. When dual-polarization data are available, polarization fusion improves ship type discrimination for up to five classes, though it does not remediate class imbalance and collecting additional single-polarization samples may prove more cost-effective in imbalanced scenarios. Multi-scale architectures emerge as the most versatile choice, maintaining effectiveness across varying class counts, resolutions, and datasets, though their advantage may be partially attributable to increased model capacity rather than architectural innovation per se.

When class counts exceed six and training data remain limited, specialized approaches such as curriculum learning [91] or knowledge distillation [86] become necessary to achieve acceptable performance. However, practitioners should recognize that the reviewed literature suffers from pervasive methodological limitations: inconsistent baseline comparisons, absent ablation studies, failure to report class-wise performance and confidence intervals, and inadequate control for confounding factors such as model capacity and hyperparameter tuning effort. Until these issues are addressed through more rigorous experimental standards, technique selection should be guided by robustness across multiple studies rather than peak performance in individual papers, and practitioners are advised to implement multiple competing approaches and evaluate them on their specific data rather than relying solely on published benchmarks.

## V. DESIGN CONSIDERATIONS

### A. Model Architectures

For designing custom CNNs, researchers have proposed shallow DL architectures. These architectures typically consist of 3-5 convolutional layers followed by pooling and activation functions, ultimately connecting to 2-3 fully connected linear layers. NAS can be a valuable tool to identify the optimal architecture for a specific task. In contrast, other proposed architectures leverage established pre-trained models:

- **AlexNet** [109] is a pioneering deep CNN with 8 layers: 5 convolutional layers followed by 3 fully connected layers. It introduced the use of ReLU activation functions and dropout for regularization.
- **VGGNet** [110] features a 16-layer version, VGGNet-16, a deep CNN architecture including 13 convolutional layers and 3 fully connected layers. The convolutional layers use small 3x3 receptive fields and are stacked on top of each other to increase the depth of the network. The network also includes five max pooling layers that follow some of the convolutional layers. Its 19-layer version, VGGNet-19, includes three additional convolutional layers. Like VGGNet-16, VGGNet-19 uses small 3x3 receptive fields and includes max pooling layers.
- **ResNet** [111] features a 50-layer and a 34-layer versions, ResNet-50 and Resnet-34, respectively, utilizing residual learning to mitigate the vanishing gradient problem commonly encountered in deep networks. This architecture is composed of convolutional layers organized into residual blocks, each containing shortcut connections that bypass one or more layers. These shortcut connections enable the network to learn identity mappings, thus facilitating the training of much deeper networks by improving gradient flow and convergence.
- **DenseNet-101**, or Dense Convolutional Network [112], comprises 101 layers with a unique dense connectivity

pattern. Each layer receives inputs from all preceding layers and passes its own feature maps to all subsequent layers, promoting feature reuse and mitigating the vanishing gradient problem. This dense connectivity reduces the number of parameters and encourages the network to learn more compact and efficient representations.

- **InceptionNet**, also known as GoogLeNet [113], introduces the inception module, which allows for multi-scale processing by applying parallel convolutional filters of different sizes ( $1 \times 1$ ,  $3 \times 3$  and  $5 \times 5$ ) within the same layer. This architecture comprises 22 layers, with inception modules stacked sequentially to capture diverse spatial features. This design efficiently manages computational resources while enhancing the network's ability to learn complex and varied patterns in the input data.

These models are often adapted for SAR data processing through transfer learning approaches. Additionally, generative adversarial networks have also been explored for potential applications in SAR ship classification [52], [70].

### B. Train-Test Split

The literature commonly uses an 80-20 split for training and testing, respectively. However, unbalanced data poses a challenge for model learning. An effective approach, as seen in the literature [60], [62], involves using the class with the fewest samples to determine the size of the training set. The remaining data are used for testing. An example of train-test ratio for class imbalanced dataset is visualized in Table VI. This approach ensures that the model learns balanced information from all classes. However, in cases of extreme imbalance, the training set becomes very limited. Therefore, it is crucial to understand the data complexity before applying such an approach.

TABLE VI: Numbers of ship-type occurrences for imbalanced and balanced train-test splits (example)

Without considering Class Imbalance			
Class	Total 100%	Train 80%	Test 20%
Cargo	1500	1200	300
Tanker	900	720	180
Fishing	600	480	120
After considering Class Imbalance			
Class	Total	Train	Test
Cargo	1500	480	1020
Tanker	900	480	420
Fishing	600	480	120

### C. Hyperparameters

1) **Epochs**: In DL, an epoch is a fundamental unit of training, defined as one complete iteration over the entire training set. During each epoch, the neural network processes every sample in the dataset, enabling the model to learn and adjust its parameters—such as weights and biases—based on the error between its predictions and the actual target values. The role of epochs is critical in the iterative optimization process, as each epoch provides an opportunity for the model to refine its internal representations and progressively minimize the loss function.

The importance of epochs lies in balancing the trade-off between underfitting and overfitting; a sufficient number of epochs ensures that the model captures the underlying patterns in the data, while avoiding excessive training that might lead to overfitting, where the model performs well on the training data but poorly on new, unseen data. Properly tuning the

number of epochs, often through techniques such as early stopping, is essential to achieve optimal model performance and generalization. The literature suggests a common range of 50–200 epochs for DL model convergence, but the optimal number ultimately depends on the chosen learning strategy [65], [67]–[70], [73], [80].

2) **Batch Size:** In DL, batch size refers to the number of training samples processed before updating the model parameters. It affects training stability and efficiency: smaller batch sizes offer more frequent updates and can improve generalization, but may introduce noise, while larger batch sizes provide more stable and accurate gradient estimates but require more memory and can slow down the training per epoch. Thus, selecting an appropriate batch size is crucial for balancing training speed, computational resources, and model performance. In the literature, batch sizes of 64, 32 and 16 are commonly used [56], [67]–[69], [71], [74], [80].

3) **Image Size:** Selecting image size plays a crucial role in determining the network computational requirements, memory usage, and ability to capture detailed features. Larger image sizes can provide more detailed spatial information, enhancing the model's ability to detect fine-grained patterns, but they also significantly increase computational load and memory consumption. Conversely, smaller image sizes reduce computational demands and memory usage but may result in the loss of important visual details, potentially degrading the model performance. The literature commonly utilizes image sizes of  $224 \times 224$ ,  $64 \times 64$  and  $128 \times 128$  [70], [71], [73], [74].

4) **Learning Rate:** The Learning Rate (see also Section IV-D6) determines the step size at each iteration while moving towards a minimum of a loss function. It controls the weights of a DL network with respect to the loss gradient. A high learning rate can accelerate training by making large updates to the parameters, but risks overshooting the optimal values, leading to divergence or instability. Conversely, a low learning rate ensures more precise adjustments, promoting stable convergence, but can make the training process exceedingly slow and may get stuck in local minima. Therefore, selecting an appropriate learning rate is critical for achieving efficient and effective training. In the literature, the common learning rates are usually set in the range 0.0001 to 0.001 [65], [67], [70], [71], [73], [74]; moreover, learning rate management can be employed to train the model faster and more precisely [68], [69].

#### D. Number of Classes

Due to the imbalanced dataset problem in SAR ship data, an increase in the number of classes results in fewer samples per class. For instance, in both the public datasets considered here (FUSARShip and OpenSARShip), over 80% of the data are concentrated in just five classes on average, with the remaining data sparsely distributed among other classes. This imbalance causes classes with fewer samples to have negligible impact during training, leading the model to overfit towards the more prevalent classes. Consequently, most research focuses on 3–5 classes (see tables-VII, VIII); although a few studies have explored the usage of more than five classes, it is common experience that the model performance significantly deteriorates as the number of classes increases. Tables VII and VIII compile performance results across 74 studies, enabling systematic analysis of technique effectiveness and performance trends.

#### E. Performance Analysis

: Tables VII and VIII reveal clear performance degradation as the number of classes increases, with a particularly sharp decline observed at the 6-class threshold. For 3-class problems, F1 scores typically range from 75% to 90%, with the best result reaching 97.80%. However, 6-class classification on OpenSARShip show substantially lower performances, with most methods achieving only 51–65% F1, indicating that class imbalance and interclass similarity become critical barriers beyond 5 classes. Notably, some specialized approaches maintain high performance even with many classes: the Siamese network achieves 97.97% F1 on 16-class OpenSARShip, and the novel CNN reaches 97.69% F1 on 5-class FUSARShip, demonstrating that targeted architectural innovations can overcome data constraints.

Regarding technique family effectiveness, multi-scale (MS) methods and polarization fusion (PF) demonstrate the most consistent performance across different class counts and datasets. For 3-class problems, PF achieves 86–89% F1 consistently, while MS methods maintain 80–86% F1. In more challenging scenarios (7 classes), MS methods sustain strong performance (86–90% F1), whereas PF effectiveness diminishes in 5-class problems (57–58% F1). Feature fusion (FF) shows high variance (75–89% F1 range), suggesting that implementation details significantly impact results. Cross-dataset comparisons reveal that methods generalizing well include MS approaches, which achieve similar F1 scores on both OpenSARShip and FUSARShip (84% vs 86%). However, these scores are not directly comparable because the train-test splits, class definitions, and datasets differ across studies. Commercial high-resolution SAR platforms (TerraSAR-X, GaoFen-3, COSMO-SkyMed) generally enable higher performance (95–98% range) compared to Sentinel-1 based datasets, confirming that spatial resolution remains a fundamental factor. Interestingly, recent methods (2024–2025) do not consistently outperform earlier work, with the best 3-class F1 score dating from 2018, suggesting that data quality and proper technique selection may be more critical than architectural novelty alone.

We additionally visualize within-dataset performance distributions for the two most widely used benchmarks. Figure 14 shows box plots of OA and  $F_1$  grouped by class count for OpenSARShip and FUSARShip. Unlike highlighting single best values, these plots reveal the median performance, spread, and outliers across studies. The distributions confirm a systematic degradation as the number of classes increases, particularly beyond five classes, while also exposing substantial variability within the same dataset and class-count setting.

## VI. OUTLOOK

This survey provides a comprehensive review of various techniques employed to train DL models for ship classification using SAR imagery. Through a systematic literature review, we assessed the content of 187 papers, rigorously narrowing down our selection to the most pertinent 74 studies. Their references are collected in Table II and their features are summarized in Section III, offering an overview of their key elements and contributions to the field.

The literature reveals a broad range of DL techniques employed for SAR ship classification, as illustrated in Figure 2. These techniques include CNNs, in the earliest proposals, and state-of-the-art architectures such as ResNet, VGG, DenseNet etc., each one tailored to leverage the unique attributes of SAR images.

TABLE VII: SAR ship-classification results on sar ship datasets with  $\leq 3$  target classes. Yellow cells mark the best Overall Accuracy (OA) and  $F_1$  within each class-count block. *Method abbreviations*: CNN—Convolutional Neural Network; DA—Domain Adaptation; FF—Feature Fusion; GAN—Generative Adversarial Network; HC—Hand-Crafted features; KD—Knowledge Distillation; ML—Metric Learning; MS—Multi-Scale; NAS—Neural Architecture Search; PF—Polarization Fusion; *Dataset abbreviations*: CSK—COSMO-SkyMed; FS—FUSAR-Ship; GF-3—Gaofen-3; OS—OpenSARShip; TX—TerraSAR-X.

Method	Dataset	#Cls.	OA [%]	$F_1$ [%]
PF (2021) [45]	GF-3	2	96.55	93.88
HC FF (2023) [98]	OS	2	88.87	83.49
CNN (2018) [26]	CSK	3	97.66	97.80
DA+TL (2019) [26]	TX	3	98.52	97.15
Inc. Angle (2019) [29]	CSK	3	98.13	–
Upsampling (2019) [30]	OS	3	81.56	–
ML (2020) [34]	OS	3	83.67	–
GAN (2020) [52]	GF-3	3	95.00	–
FF (2021) [35]	OS	3	77.69	–
AIS+CNN (2021) [36]	OS	3	93.30	93.73
ML (2021) [38]	OS	3	88.97	–
Attn. FF (2021) [40]	OS	3	92.66	–
FF (2021) [41]	OS	3	78.00	–
FF (2021) [41]	FS	3	86.86	–
MS+PF (2021) [44]	OS	3	79.25	77.62
FF (2021) [48]	OS	3	77.69	–
MS+HC FF (2022) [50]	OS	3	78.15	75.04
FF (2022) [51]	TX	3	88.89	–
FF (2022) [51]	FS	3	89.11	–
PF (2022) [53]	OS	3	88.00	–
PF (2022) [56]	OS	3	88.80	86.81
NAS (2022) [57]	OS	3	88.02	–
PF+HC (2022) [58]	OS	3	79.84	78.00
DA (2022) [59]	TX	3	87.67	–
NAS (2023) [60]	OS	3	82.06	–
PF (2023) [62]	OS	3	81.28	–
MS+PF (2023) [63]	OS	3	88.93	86.54
Novel CNN (2023) [69]	GF-3	3	89.79	89.71
Attn. Dense GAN (2023) [70]	FS	3	92.00	–
MS (2023) [71]	OS	3	82.23	80.91
Dual-Pol (2024) [73]	OS	3	87.62	86.74
DA (2024) [77]	FS	3	75.94	–
Novel Arch. (2024) [79]	OS	3	81.86	80.75
Novel Arch. (2024) [79]	FS	3	75.18	73.05
PF (2024) [80]	OS	3	87.13	86.27
PF (2024) [81]	OS	3	89.40	86.89
FF (2024) [82]	OS	3	82.00	82.00
FF (2024) [82]	FS	3	90.00	89.00
MS (2024) [83]	OS	3	85.66	84.90
MS (2024) [83]	FS	3	85.26	86.17
B-CNN (2024) [84]	OS	3	76.30	76.30
FM (2024) [85]	OS	3	79.02	–
KD (2024) [86]	OS	3	80.03	–
Imbal. (2024) [87]	FS+OS	3	74.60	67.50
CV-DL (2025) [89]	GF-3	3	72.22	–
DA (2025) [90]	FS	3	95.33	–
DA (2025) [90]	TX	3	89.33	–
MS FF (2025) [92]	OS	3	80.09	79.27
Super-Res (2025) [93]	OS	3	72.94	61.53
Super-Res (2025) [94]	OS	3	87.68	–

TABLE VIII: SAR ship-classification results on SAR Ship datasets with &gt;3 target classes. Abbreviations as in Table VII.

Method	Dataset	#Cls.	OA. [%]	F <sub>1</sub> [%]
CNN + Spatial Attention (2018) [27]	OS	4	84.00	–
CNN (2019) [31]	P2	4	65.70	–
Deep FF (2019) [32]	OS	4	87.04	–
FF (2021) [43]	OS	4	82.35	82.19
DA (2022) [59]	FS	4	82.33	–
NAS (2023) [60]	FS	4	63.90	–
HC (2024) [74]	OS	4	87.50	86.00
LSTM + CNN (2024) [78]	FS	4	65.58	–
Complex-Valued DL (2025) [89]	GF-3	4	69.01	–
<hr/>				
CNN (2018) [25]	TX	5	–	94.00
Metric Learning (2021) [38]	OS	5	68.16	–
PF (2022) [56]	OS	5	66.90	57.54
Novel CNN (2023) [61]	FS	5	98.71	97.69
MS + PF (2023) [63]	OS	5	67.03	57.89
Data Augmentation (2024) [76]	FS	5	72.19	71.97
PF (2024) [81]	OS	5	67.83	58.01
DA (2024) [72]	FS	5	91.22	–
MS (2024) [83]	FS	5	84.24	84.36
Metric Learning (2024) [65]	FS	5	78.57	–
<hr/>				
MS + PF (2021) [44]	OS	6	56.66	51.38
Sampling (2021) [47]	OS	6	80.58	–
NAS (2022) [58]	OS	6	56.73	–
PF + HC (2022) [58]	OS	6	56.83	51.82
PF (2023) [62]	OS	6	58.68	–
MS (2023) [64]	OS	6	81.83	–
MS (2023) [71]	OS	6	60.76	59.89
Dual-Pol (2024) [73]	OS	6	64.26	58.44
PF (2024) [80]	OS	6	65.97	58.63
MS FF (2025) [92]	OS	6	60.63	58.29
Super-Res (2025) [94]	OS	6	69.71	–
Super-Res (2025) [96]	OS	6	–	65.40
<hr/>				
FF (2021) [41]	FS	7	86.86	–
MS + HC FF (2022) [50]	FS	7	86.69	86.58
HC FF (2023) [67]	FS	7	87.23	86.69
MS (2023) [71]	FS	7	90.08	90.12
Feature Manip. (2024) [85]	FS	7	87.59	–
Knowledge Distill. (2024) [86]	FS	7	86.07	–
HC (2024) [75]	FS	7	87.94	–
<hr/>				
Deep FF (2025) [88]	FS	8	76.12	56.67
Curriculum Learning (2025) [91]	FS	9	68.34	55.49
Over-Sampling (2025) [97]	FS	9	–	52.87
Side-Lobe Elimin. (2021) [42]	FS	10	98.40	97.96
Siamese Net (2021) [55]	OS	16	–	97.97

Our analysis indicates that the choice of the DL technique significantly impacts classification performance. For instance, shallow CNNs have been particularly effective to learn spatial hierarchies from SAR images, as the data are scarce.

The FUSARShip and OpenSARShip datasets, along with other distinct data sources, are extensively used in the literature. These datasets are crucial for training and evaluating DL models. However, the limited availability of SAR images poses a significant challenge. To address this, data augmentation techniques have been proposed and widely adopted. These techniques artificially increase the diversity and quantity of training data, which is essential for improving the performance and robustness of DL models. Fine-tuning pre-trained DL models on

SAR-specific datasets has emerged as a successful strategy. This approach leverages the knowledge from pre-trained models and adapts it to the specific characteristics of SAR data, resulting in improved classification accuracy.

The issue of data imbalance has been addressed in the literature through two main solutions: adjusting train-test datasets ratios and employing tailored loss functions. Additionally, some researchers propose the use of architecture search techniques like NAS for selecting the optimal DL architecture. This survey also highlights the significance of various training techniques, including feature fusion, polarization fusion, metric learning, multi-scale/pyramid learning, and learning rate management.

The findings of our survey underscore the importance of data

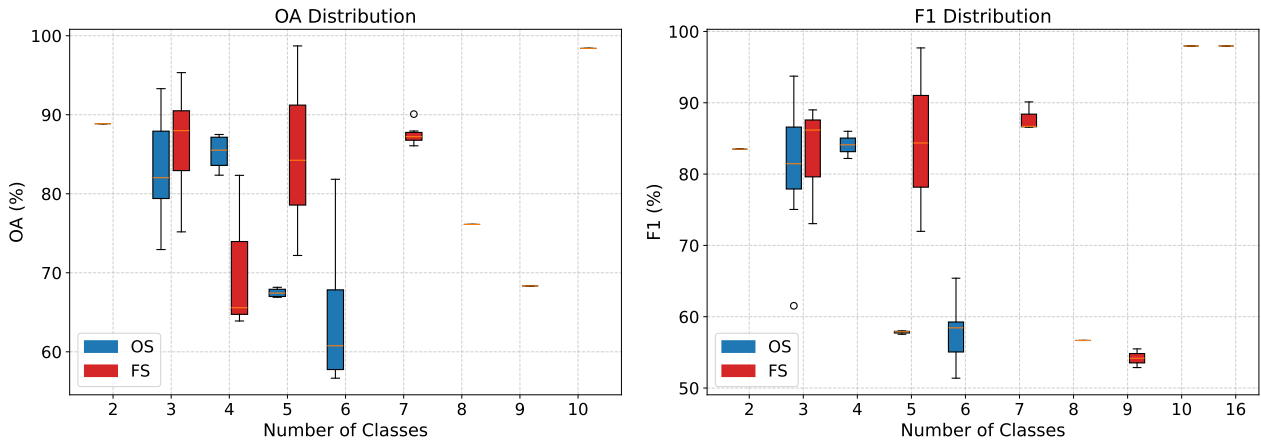


Fig. 14: OA and  $F_1$  distributions across class counts for OpenSARShip (OS) and FUSAR-Ship (FS).

TABLE IX: Challenges in SAR ship classification and practical remedies.

Challenge	Remedies / notes
Data scarcity	High-quality open SAR curation; AIS–SAR mapping for labels; strong augmentations; transfer learning.
Class imbalance	Rebalanced splits; focal/Class Balanced loss; mixup/cutmix/feature-space-oversampling; report $F_1$ alongside OA, curriculum learning.
Domain shift (sensor/scene)	Normalization; domain adaptation; test-time adaptation; confidence calibration.
Small targets / clutter	Multi-scale features; higher input resolution; denoising; polarization/feature fusion, super resolution.
Reproducibility	Report splits/seeds/hyperparams; share CSV of results.

quality and the choice of techniques in SAR ship classification. The integration of handcrafted features, effective data augmentation, and fine-tuning of pre-trained models are all critical factors contributing to the success of DL models in this domain. Building on the findings of this survey and acknowledging the limitations of current research, several exciting avenues for future work in SAR ship classification using DL are identified and described in the following subsections.

### A. Addressing Data scarcity

The data used by researchers for DL-based SAR ship classification is seldom made public due to privacy or digital rights reasons. Future research should focus on developing methodologies to publicize the data without such constraints. In addition, creating high-quality datasets from open-access SAR imagery and establishing methodologies for reliable AIS–SAR mapping can play a key role in alleviating the problem of data scarcity. Rigorous evaluation of AIS-SAR matching uncertainties and systematic assessment of label noise propagation represent important but underexplored directions. Quantifying how annotation errors affect DL model robustness would strengthen the empirical foundations of SAR ship classification research. Developing more robust data augmentation techniques can further contribute to the diversity of SAR datasets [76]. Finally, collaborative efforts aimed at standardizing and sharing datasets would significantly benefit the research community.

### B. Novel DL architectures

The development of DL architectures specifically tailored for SAR ship classification holds significant promise for advancing the research in this field. This includes exploring and optimizing various DL architecture components, such as loss functions, learning rates, activation functions, and even utilizing

techniques like architecture search and fine-tuning. Foundation models are increasingly shaping remote sensing, yet SAR ship classification still lacks a foundation model trained specifically for ship type discrimination. Current SAR foundation efforts target broader SAR ATR objectives and include ship classification datasets only as part of multi-category pretraining and transfer evaluation, for example SARATR-X [114] and SAR-JEPA [115]. These models demonstrate that SAR-domain self-supervised pretraining can yield transferable representations for downstream recognition tasks, but they do not directly discuss classification-specific challenges that dominate SAR ship datasets, most notably long-tail class distributions and severe class imbalance. This gap motivates SAR ship classification-oriented pretraining and adaptation protocols, where imbalance-aware sampling, loss design, and evaluation become first-class considerations. In parallel, rapid progress in earth observation and remote sensing foundation models, including Prithvi-EO [116], SatMAE [117], DOFA [118], TerraMind [119], RemoteCLIP [120], and multimodal optical–SAR representation learning such as CROMA [121], strengthens the case for systematic benchmarking on SAR ship classification, including careful transfer studies from EO and RS pretrained models to SAR.

Adapting foundation models to SAR ship classification requires addressing several domain-specific challenges. Speckle noise, inherent to SAR imaging, degrades feature consistency and may require speckle-aware augmentation during fine-tuning or adaptation of denoising strategies within self-supervised pretraining objectives. Polarimetric SAR data (VV/VH or HH/VV channels) provides complementary information but requires architecture modifications to handle dual-channel inputs effectively, either through early fusion with modified input layers or late fusion of separately encoded polarization streams. Class imbalance remains critical: even if a foundation model provides strong general-purpose embeddings, the classification head must be trained with imbalance-aware loss functions (focal loss, class-balanced loss) and sampling strategies. Finally, the resolution mismatch between optical pretraining data (high-resolution RGB) and operational SAR ship chips (often medium-resolution, single or dual polarization) may limit direct transfer effectiveness, suggesting that SAR-pretrained or SAR-optical jointly pretrained models such as SARJepa, SARDet-100k, and CROMA may offer better initialization than purely optical models.

Practically, if a pretrained model provides stable and discriminative embeddings, a lightweight pipeline becomes feasible, where embeddings are extracted once and only a small fully

connected classification head is trained, which can reduce computation while improving robustness under limited labels and imbalanced data, consistent with common linear-probing strategies in self-supervised learning [122].

Furthermore, the continuous evolution of DL architectures presents an exciting opportunity to investigate the efficacy of novel models in SAR ship classification tasks. Recent advancements in areas such as transformers and generative models (e.g., TransGAN [123]) are particularly promising for tasks such as noise reduction and data augmentation, which can be crucial for improving the performance of DL models in SAR ship classification.

### C. Handcrafted and Deep Features Integration

DL models have achieved remarkable success in feature learning for various tasks. However, handcrafted features, meticulously designed to capture specific characteristics of SAR images, remain a valuable tool. Further research is warranted to explore how these can be effectively integrated with DL models for SAR image classification. This combined approach has the potential to leverage the strengths of both machine learning and DL methods, potentially leading to significant performance improvements. Moreover, leveraging handcrafted features and intrinsic SAR properties may provide promising pathways for the development of next-generation models.

### D. Standardized Performance Metrics

The field of SAR ship classification currently lacks a standardized set of performance metrics, hindering comprehensive comparisons between different studies. Utilizing metrics like the F1 score can partially address this issue, but future research should focus on establishing a common set of metrics specifically tailored to the nuances of SAR ship classification tasks. This would enable more robust evaluations and facilitate meaningful comparisons between different DL models and training strategies.

Furthermore, inconsistencies in data utilization across studies pose another challenge to comparison. Even when researchers employ the same datasets, variations in sample selection can significantly impact the results. Moreover, among the researchers in SAR ship classification, code sharing is not common; if we want to progress further and faster, code sharing should be adopted as a practice. Standardized metrics and data utilization practices are crucial for achieving truly comparable research, ultimately accelerating the advancements in the field.

### E. Interpretability of DL models

DL models, while demonstrably powerful, are often characterized as “black boxes” due to their complex internal workings. This lack of interpretability can hinder our understanding of the rationale behind their classification decisions. Fortunately, various interpretability techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM, [124]), SHapley Additive exPlanations (SHAP, [125]) and Local Interpretable Model-Agnostic Explanations (LIME, [126]) can be employed to shed light on the internal decision-making processes of these models.

While interpretability has not always been a central focus in SAR ship classification research, it presents a valuable avenue for future exploration. By leveraging interpretability techniques, researchers can gain deeper insights into the features most critical for the model’s decisions. This enhanced understanding can lead to more informed analysis, improved trust in model predictions, and even to the identification of potential biases within the model.

### F. Interdisciplinary Collaborations

Significant advancements in SAR ship classification using DL can be achieved by fostering interdisciplinary collaborations between researchers in remote sensing, computer vision, and maritime studies. This collaborative approach has the potential to not only generate innovative solutions for the challenges faced in SAR ship classification, but also contribute to environmental protection efforts by facilitating the detection of illegal maritime activities.

By focusing on these directions, future research can build on our findings and continue to advance the field of SAR ship classification using DL, ultimately improving the accuracy and efficiency of maritime surveillance and related applications, we also present common challenges and remedies to those challenges in table-IX.

## VII. CONCLUSION

This survey has systematically reviewed the diverse techniques employed to train DL models for SAR ship classification. We identified critical trends and challenges in this domain, emphasizing the importance of integrating handcrafted features, utilizing public datasets, data augmentation, fine-tuning, explainability, and interdisciplinary collaborations for enhancing DL model performance.

We established a survey methodology with clearly defined research questions. Our literature search yielded a comprehensive dataset of relevant papers, which were then categorized into a first-of-its-kind taxonomy encompassing DL techniques, handcrafted feature use, SAR attribute utilization, and the impact of fine-tuning.

In particular, Section III provides detailed explanations of the included papers, categorized into the four sections (discussing DL techniques, impact of handcrafted features, use of SAR properties and fine-tuning). Section IV presents the methodologies employed in SAR ship classification. Finally, Section VI outlines potential avenues for addressing current challenges, such as data scarcity, novel DL architectures, interpretability, performance metrics, feature integration, and interdisciplinary collaborations.

Advancements in DL techniques for SAR ship classification hold significant promise for improving maritime surveillance and related applications. Continued exploration and innovation in this field can lead to more accurate and efficient ship classification systems, ultimately enhancing the capabilities of maritime operations.

## ACKNOWLEDGMENTS

This work was supported by National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU. Award Number: Project code CN\_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP D33C22000960007, Project title “National Biodiversity Future Center - NBFC”.

## REFERENCES

- [1] H. Espedal, “Satellite SAR oil spill detection using wind history information,” *International Journal of Remote Sensing*, vol. 20, no. 1, pp. 49–65, 1999.
- [2] A.-B. Salberg, Ø. Rudjord, and A. H. S. Solberg, “Oil spill detection in hybrid-polarimetric SAR images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6521–6533, 2014.

- [3] K. N. Topouzelis, "Oil spill detection by SAR images: Dark formation detection, feature extraction and classification algorithms," *Sensors*, vol. 8, no. 10, pp. 6642–6659, 2008.
- [4] J. Li, C. Xu, H. Su, L. Gao, and T. Wang, "Deep learning for SAR ship detection: Past, present and future," *Remote Sensing*, vol. 14, no. 11, p. 2712, 2022.
- [5] M. Yasir, W. Jianhua, X. Mingming, S. Hui, Z. Zhe, L. Shanwei, A. T. I. Colak, and M. S. Hossain, "Ship detection based on deep learning using SAR imagery: a systematic literature review," *Soft Computing*, vol. 27, no. 1, pp. 63–84, 2023.
- [6] T. Zhang, X. Zhang, J. Li, X. Xu, B. Wang, X. Zhan, Y. Xu, X. Ke, T. Zeng, H. Su *et al.*, "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sensing*, vol. 13, no. 18, p. 3690, 2021.
- [7] M. Yasir, S. Liu, X. Mingming, J. Wan, S. Pirasteh, and K. B. Dang, "Shipgeonet: Sar image-based geometric feature extraction of ships using convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024.
- [8] M. Yasir, S. Liu, S. Pirasteh, M. Xu, H. Sheng, J. Wan, F. A. de Figueiredo, F. J. Aguilar, and J. Li, "Yoloshiptracker: Tracking ships in sar images using lightweight yolov8," *International Journal of Applied Earth Observation and Geoinformation*, vol. 134, p. 104137, 2024.
- [9] M. Yasir, L. Shanwei, X. Mingming, W. Jianhua, S. Hui, S. Nazir, X. Zhang, and A. T. I. Colak, "Yolov8-byte: Ship tracking algorithm using short-time sequence sar images for disaster response leveraging geolai," *International Journal of Applied Earth Observation and Geoinformation*, vol. 128, p. 103771, 2024.
- [10] G. Zilman, A. Zapolski, and M. Marom, "The speed and beam of a ship from its wake's SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 10, pp. 2335–2343, 2004.
- [11] K.-m. Kang and D.-j. Kim, "Ship velocity estimation from ship wakes detected using convolutional neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 11, pp. 4379–4388, 2019.
- [12] M. D. Graziano, A. Renga, and A. Moccia, "Integration of automatic identification system (AIS) data and single-channel synthetic aperture radar (SAR) images by SAR-based ship velocity estimation for maritime situational awareness," *Remote Sensing*, vol. 11, no. 19, p. 2196, 2019.
- [13] P. Heiselberg, K. Sørensen, and H. Heiselberg, "Ship velocity estimation in SAR images using multitask deep learning," *Remote Sensing of Environment*, vol. 288, p. 113492, 2023.
- [14] X. Xing, K. Ji, H. Zou, W. Chen, and J. Sun, "Ship classification in TerraSAR-X images with feature space based sparse representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 6, pp. 1562–1566, 2013.
- [15] M. Jiang, X. Yang, Z. Dong, S. Fang, and J. Meng, "Ship classification based on superstructure scattering features in SAR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 5, pp. 616–620, 2016.
- [16] H. Lang, J. Zhang, X. Zhang, and J. Meng, "Ship classification in SAR image by joint feature and classifier selection," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 2, pp. 212–216, 2015.
- [17] J. Wu, Y. Zhu, Z. Wang, Z. Song, X. Liu, W. Wang, Z. Zhang, Y. Yu, Z. Xu, T. Zhang *et al.*, "A novel ship classification approach for high resolution SAR images based on the BDA-KELM classification model," *International Journal of Remote Sensing*, vol. 38, no. 23, pp. 6457–6476, 2017.
- [18] A. Galdelli, A. Mancini, C. Ferrà, and A. N. Tassetti, "Integrating AIS and SAR to monitor fisheries: A pilot study in the Adriatic Sea," in *Proceedings of the 2020 IMEKO TC-19 International Workshop on Metrology for the Sea, Naples, Italy, 2020*, pp. 5–7.
- [19] S. K. Chaturvedi, C.-S. Yang, K. Ouchi, and P. Shanmugam, "Ship recognition by integration of SAR and AIS," *The Journal of Navigation*, vol. 65, no. 2, pp. 323–337, 2012.
- [20] X. Hou, W. Ao, Q. Song, J. Lai, H. Wang, and F. Xu, "FUSAR-Ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition," *Science China Information Sciences*, vol. 63, pp. 1–19, 2020.
- [21] J. Li, Z. Yu, L. Yu, P. Cheng, J. Chen, and C. Chi, "A comprehensive survey on sar atr in deep-learning era," *Remote Sensing*, vol. 15, no. 5, p. 1454, 2023.
- [22] G. Succi, "A meta-analytical comparison of naive Bayes and random forest for software defect prediction," in *Intelligent Systems Design and Applications: 22nd International Conference on Intelligent Systems Design and Applications (ISDA 2022) Held December 12-14, 2022-Volume 3*, vol. 716. Springer Nature, 2023, p. 139.
- [23] M. Petticrew and H. Roberts, *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons, 2008.
- [24] C. Wang, H. Zhang, F. Wu, B. Zhang, and S. Tian, "Ship classification with deep learning using COSMO-SkyMed SAR data," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 558–561.
- [25] C. Bentes, D. Velotto, and B. Tings, "Ship classification in TerraSAR-X images with convolutional neural networks," *IEEE Journal of Oceanic Engineering*, vol. 43, no. 1, pp. 258–266, 2018.
- [26] Y. Wang, C. Wang, and H. Zhang, "Ship classification in high-resolution SAR images using deep learning of small datasets," *Sensors*, vol. 18, no. 9, 2018. [Online]. Available: <https://www.mdpi.com/1424-8220/18/9/2929>
- [27] J. Shao, C. Qu, J. Li, and S. Peng, "A lightweight convolutional neural network based on visual attention for SAR image target classification," *Sensors*, vol. 18, no. 9, 2018. [Online]. Available: <https://www.mdpi.com/1424-8220/18/9/3039>
- [28] C. Lu and W. Li, "Ship classification in high-resolution SAR images via transfer learning with small training dataset," *Sensors*, vol. 19, no. 1, 2019. [Online]. Available: <https://www.mdpi.com/1424-8220/19/1/63>
- [29] Y. Dong, C. Wang, H. Zhang, Y. Wang, and B. Zhang, "Impact analysis of incident angle factor on high-resolution SAR image ship classification based on deep learning," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 1358–1361.
- [30] Y. Wu, Y. Yuan, J. Guan, L. Yin, J. Chen, G. Zhang, and P. Feng, "Joint convolutional neural network for small-scale ship classification in SAR images," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 2619–2622.
- [31] S. Hashimoto, Y. Sugimoto, K. Hamamoto, and N. Ishihama, "Ship classification from SAR images based on deep learning," in *Intelligent Systems and Applications*, K. Arai, S. Kapoor, and R. Bhatia, Eds. Cham: Springer International Publishing, 2019, pp. 18–34.
- [32] Y. Xi, G. Xiong, and W. Yu, "Feature-loss double fusion siamese network for dual-polarized SAR ship classification," in *2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, 2019, pp. 1–5.
- [33] H. Zhu, N. Lin, and D. Leung, "Ship classification from SAR images based on sequence input of deep neural network," *Journal of Physics: Conference Series*, vol. 1549, no. 5, p. 052042, jun 2020. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/1549/5/052042>
- [34] Y. Li, X. Li, Q. Sun, and Q. Dong, "SAR image classification using CNN embeddings and metric learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2020.
- [35] T. Zhang, X. Zhang, J. Shi, and S. Wei, "A HOG feature fusion method to improve CNN-based SAR ship classification accuracy," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 5311–5314.
- [36] H.-K. Jeon and C.-S. Yang, "Enhancement of ship type classification from a combination of CNN and KNN," *Electronics*, vol. 10, no. 10, 2021. [Online]. Available: <https://www.mdpi.com/2079-9292/10/10/1169>
- [37] X. Xu, X. Zhang, and T. Zhang, "Multi-scale SAR ship classification with convolutional neural network," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 4284–4287.
- [38] J. He, Y. Wang, and H. Liu, "Ship classification in medium-resolution SAR images via densely connected triplet CNNs integrating Fisher discrimination regularized metric learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 3022–3039, 2021.
- [39] Y. Zhang, Q. Hua, Y. Jiang, H. Li, and D. Xu, "CV-MotionNet: Complex-valued convolutional neural network for SAR moving ship targets classification," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 4280–4283.
- [40] G. Zhou, G. Zhang, Z. Fang, and Q. Dai, "A multiscale dual-attention based convolutional neural network for ship classification in SAR image," in *2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2021, pp. 1–5.
- [41] T. Zhang and X. Zhang, "Injection of traditional hand-crafted features into modern CNN-based models for SAR ship classification: What, why, where, and how," *Remote Sensing*, vol. 13, no. 11, 2021. [Online]. Available: <https://www.mdpi.com/2072-4292/13/11/2091>
- [42] H. Zhu, "Ship classification based on sidelobe elimination of SAR images supervised by visual model," in *2021 IEEE Radar Conference (RadarConf21)*, 2021, pp. 1–6.
- [43] H. Relekar and P. Shanmugam, "Transfer learning based ship classification in Sentinel-1 images incorporating scale variant features," *Advances in Space Research*, vol. 68, no. 11, pp. 4594–4615, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S027311772100692X>
- [44] T. Zhang and X. Zhang, "Squeeze-and-excitation laplacian pyramid network with dual-polarization feature fusion for ship classification in SAR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.
- [45] Z. Sun, B. Xiong, Y. Lei, X. Leng, and K. Ji, "Ship classification in high-resolution SAR images based on CNN regional feature fusion," in *2021 CIE International Conference on Radar (Radar)*, 2021, pp. 1445–1449.
- [46] C. Wang, J. Shi, Y. Zhou, X. Yang, Z. Zhou, S. Wei, and X. Zhang, "Semisupervised learning-based SAR ATR via self-consistent augmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 4862–4873, 2021.
- [47] Y. Zhang, Z. Lei, L. Zhuang, and H. Yu, "A CNN based method to solve class imbalance problem in SAR image ship target recognition," in *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, vol. 5, 2021, pp. 229–233.

- [48] T. Zhang and X. Zhang, "Integrate traditional hand-crafted features into modern CNN-based models to further improve SAR ship classification accuracy," in *2021 7th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, 2021, pp. 1–6.
- [49] L. Zeng, Q. Zhu, D. Lu, T. Zhang, H. Wang, J. Yin, and J. Yang, "Dual-polarized SAR ship grained classification based on CNN with hybrid channel feature loss," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [50] T. Zhang, X. Zhang, X. Ke, C. Liu, X. Xu, X. Zhan, C. Wang, I. Ahmad, Y. Zhou, D. Pan, J. Li, H. Su, J. Shi, and S. Wei, "HOG-ShipCLSNet: A novel deep learning network with HOG feature fusion for SAR ship classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–22, 2022.
- [51] S. Zhao and H. Lang, "Improving deep subdomain adaptation by dual-branch network embedding attention module for SAR ship classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 8038–8048, 2022.
- [52] L. Li, C. Wang, H. Zhang, and B. Zhang, "SAR image ship object generation and classification with improved residual conditional generative adversarial network," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [53] J. He, W. Chang, F. Wang, Q. Wang, Y. Li, and Y. Gan, "Polarization matters: On bilinear convolutional neural networks for ship classification from synthetic aperture radar images," in *2022 4th International Conference on Natural Language Processing (ICNLP)*, 2022, pp. 315–319.
- [54] A. Toumi, J.-C. Cexus, and A. Khenchaf, "A proposal learning strategy on CNN architectures for targets classification," in *2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2022, pp. 1–6.
- [55] J. A. Raj, S. M. Idicula, and B. Paul, "One-shot learning-based SAR ship classification using new hybrid siamese network," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [56] J. He, W. Chang, F. Wang, Y. Liu, Y. Wang, H. Liu, Y. Li, and L. Liu, "Group bilinear CNNs for dual-polarized SAR ship classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [57] H. Zhu, S. Guo, W. Sheng, and L. Xiao, "SBNN: A searched binary neural network for SAR ship classification," *Applied Sciences*, vol. 12, no. 14, 2022. [Online]. Available: <https://www.mdpi.com/2076-3417/12/14/6866>
- [58] T. Zhang and X. Zhang, "A polarization fusion network with geometric feature embedding for SAR ship classification," *Pattern Recognition*, vol. 123, p. 108365, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320321005458>
- [59] H. Lang, R. Wang, S. Zheng, S. Wu, and J. Li, "Ship classification in SAR imagery by shallow CNN pre-trained on task-specific dataset with feature refinement," *Remote Sensing*, vol. 14, no. 23, 2022. [Online]. Available: <https://www.mdpi.com/2072-4292/14/23/5986>
- [60] H. Zhu, S. Guo, W. Sheng, and L. Xiao, "SCM: A searched convolutional metaformer for SAR ship classification," *Remote Sensing*, vol. 15, no. 11, 2023. [Online]. Available: <https://www.mdpi.com/2072-4292/15/11/2904>
- [61] H. Guo and L. Ren, "A marine small-targets classification algorithm based on improved convolutional neural networks," *Remote Sensing*, vol. 15, no. 11, 2023. [Online]. Available: <https://www.mdpi.com/2072-4292/15/11/2917>
- [62] Z. Shao, T. Zhang, and X. Ke, "A dual-polarization information-guided network for SAR ship classification," *Remote Sensing*, vol. 15, no. 8, 2023. [Online]. Available: <https://www.mdpi.com/2072-4292/15/8/2138>
- [63] J. He, W. Chang, F. Wang, Y. Liu, C. Sun, and Y. Li, "Multi-scale dense networks for ship classification using dual-polarization SAR images," in *2023 IEEE Radar Conference (RadarConf23)*, 2023, pp. 1–6.
- [64] M. Yang, X. Bai, L. Wang, and F. Zhou, "HENC: Hierarchical embedding network with center calibration for few-shot fine-grained SAR target classification," *IEEE Transactions on Image Processing*, vol. 32, pp. 3324–3337, 2023.
- [65] J. Xu and H. Lang, "A unified multiple proxy deep metric learning framework embedded with distribution optimization for fine-grained ship classification in remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 5604–5620, 2024.
- [66] Y. Qi, L. Wang, C. Zhao, N. Wang, and J. Chen, "Using squeeze-and-excitation vision transformer with local feature fusion for ship classification in SAR images," in *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 2023, pp. 7499–7502.
- [67] H. Zheng, Z. Hu, L. Yang, A. Xu, M. Zheng, C. Zhang, and K. Li, "Multifeature collaborative fusion network with deep supervision for SAR ship classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.
- [68] Y. Liu, F. Zhang, L. Ma, and F. Ma, "Long-tailed SAR target recognition based on expert network and intra-class resampling," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.
- [69] Y. Guan, X. Zhang, S. Chen, G. Liu, Y. Jia, Y. Zhang, G. Gao, J. Zhang, Z. Li, and C. Cao, "Fishing vessel classification in SAR images using a novel deep learning model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–21, 2023.
- [70] G. Gao, Y. Dai, X. Zhang, D. Duan, and F. Guo, "ADCG: A cross-modality domain transfer learning method for synthetic aperture radar in ship automatic target recognition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.
- [71] Y. Shang, W. Pu, C. Wu, D. Liao, X. Xu, C. Wang, Y. Huang, Y. Zhang, J. Wu, J. Yang, and J. Wu, "HDSS-Net: A novel hierarchically designed network with spherical space classifier for ship recognition in SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–20, 2023.
- [72] H. Zhang, Y. Lv, J. Zhang, Q. Wu, and Z. Wen, "Rendering-inspired cross-source feature disentanglement for domain adaptation-based sar ship classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.
- [73] N. Xie, T. Zhang, W. Guo, Z. Zhang, and W. Yu, "Dual branch deep network for ship classification of dual-polarized sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024.
- [74] S. Bhattacharjee, P. Shanmugam, and S. Das, "Attention-guided convolution neural network assisted with handcrafted features for ship classification in low-resolution sentinel-1 sar image data," *IEEE Access*, vol. 12, pp. 48 668–48 685, 2024.
- [75] X. Feng, H. Zheng, Z. Hu, L. Yang, and M. Zheng, "Dual-stream contrastive predictive network with joint handcrafted feature view for SAR ship classification," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 7810–7814.
- [76] L. Wang, Y. Qi, P. T. Mathiopoulos, C. Zhao, and S. Mazhar, "An improved SAR ship classification method using text-to-image generation-based data augmentation and squeeze and excitation," *Remote Sensing*, vol. 16, no. 7, 2024. [Online]. Available: <https://www.mdpi.com/2072-4292/16/7/1299>
- [77] Z. Liu, K. Li, L. Wang, and Z. Zhang, "Multi-scale alignment domain adaptation for ship classification in multi-resolution SAR images," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 4051–4062, 2025.
- [78] A. Toumi, J.-C. Cexus, A. Khenchaf, and M. Abid, "A combined CNN-LSTM network for ship classification on SAR images," *Sensors*, vol. 24, no. 24, 2024. [Online]. Available: <https://www.mdpi.com/1424-8220/24/24/7954>
- [79] N. Xie, M. Xiong, F. Wei, J. Li, T. Zhang, and W. Yu, "MFSAF: A plug-and-play module for SAR ship classification," in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 2024, pp. 9075–9079.
- [80] N. Xie, T. Zhang, F. Wei, and W. Yu, "Pfdn: A polarimetric feature-guided deep network for dual-polarized SAR ship classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.
- [81] J. He, R. Sun, Y. Kong, W. Chang, C. Sun, G. Chen, Y. Li, Z. Meng, and F. Wang, "CPINet: Towards a novel cross-polarimetric interaction network for dual-polarized SAR ship classification," *Remote Sensing*, vol. 16, no. 18, 2024. [Online]. Available: <https://www.mdpi.com/2072-4292/16/18/3479>
- [82] D. Li and Z. Zhao, "SAR ship classification based on second-order synchrosqueezing transform and HOG feature," in *2024 9th International Conference on Signal and Image Processing (ICSIP)*, 2024, pp. 585–590.
- [83] X. Zhang, S. Feng, C. Zhao, Z. Sun, S. Zhang, and K. Ji, "MGSA-Net: Multiscale global scattering feature association network for SAR ship target recognition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 4611–4625, 2024.
- [84] A. A. Al Hinai and R. Guida, "Uncertainty estimation in Bayesian convolutional neural networks for SAR ship classification," in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 2024, pp. 7604–7608.
- [85] Z. Cui, L. Mou, Z. Zhou, K. Tang, Z. Yang, Z. Cao, and J. Yang, "Feature joint learning for SAR target recognition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–20, 2024.
- [86] B. Xu, H. Zheng, Z. Hu, L. Yang, M. Zheng, X. Feng, and W. Lin, "Double reverse regularization network based on self-knowledge distillation for SAR object classification," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 7800–7804.
- [87] C. Awais and M. Reggiannini, "Deep learning for SAR ship classification: Focus on unbalanced datasets and inter-dataset generalization," in *2024 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, 2024, pp. 1–1.
- [88] S. Zhao, W. Li, F. Shen, and M. You, "Ln-scnnet: A lightweight convolutional neural network for SAR ship classification," *IEEE Access*, vol. 13, pp. 39 394–39 404, 2025.
- [89] G. Gao, Y. He, J. Zhao, S. Li, M. Wang, G. Yang, and X. Zhang, "FDC-TA-DSN ship classification model and dataset construction based on complex-valued SAR," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 7034–7047, 2025.
- [90] Z. Han and H. Lang, "Unsupervised SAR fine-grained ship classification via spherical metric refinement with deep subdomain adaptation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 16 003–16 019, 2025.

- [91] C. M. Awais, M. Reggiannini, and D. Moroni, "A framework for imbalanced SAR ship classification: Curriculum learning weighted loss functions and a novel evaluation metric," in *Proceedings of the Winter Conference on Applications of Computer Vision (WACV) Workshops*, February 2025, pp. 1570–1578.
- [92] J. Zheng, J. Cao, and X. Hu, "Msf-at: A study on ship SAR image classification based on multi-scale feature and attention mechanism," *IEEE Access*, vol. 13, pp. 55 467–55 475, 2025.
- [93] C. M. Awais, M. Reggiannini, and D. Moroni, "Image quality vs performance in super-resolution for SAR ship classification," in *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2025, pp. 1–5.
- [94] W.-L. Tseng, Y.-L. Wu, M.-C. Lee, and H.-H. Shuai, "RSMAE: Radiometric resolution and scale-aware masked autoencoder for SAR ship recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 22, pp. 1–5, 2025.
- [95] N. Xie, T. Zhang, L. Zhang, J. Chen, F. Wei, and W. Yu, "VLF-SAR: A novel vision-language framework for few-shot SAR target recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2025.
- [96] C. M. Awais, M. Reggiannini, D. Moroni, and O. Karakus, "A classification-aware super-resolution framework for ship targets in SAR imagery," *arXiv preprint arXiv:2508.06407*, 2025.
- [97] —, "Feature-space oversampling for addressing class imbalance in SAR ship classification," *arXiv preprint arXiv:2508.06420*, 2025.
- [98] E. Nehary, A. Dey, S. Rajan, B. Balaji, A. Damini, and R. Chanchlani, "Synthetic aperture radar-based ship classification using CNN and traditional handcrafted features," in *2023 IEEE Sensors Applications Symposium (SAS)*, 2023, pp. 01–06.
- [99] T. Emmens, C. Amrit, A. Abdi, and M. Ghosh, "The promises and perils of automatic identification system data," *Expert Systems with Applications*, vol. 178, p. 114975, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417421004164>
- [100] B. Li, B. Liu, L. Huang, W. Guo, Z. Zhang, and W. Yu, "OpenSARShip 2.0: A large-volume dataset for deeper interpretation of ship targets in Sentinel-1 imagery," in *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA)*. IEEE, 2017, pp. 1–5.
- [101] W. Yang, Z. Ma, and Y. Shi, "SAR image super-resolution based on artificial intelligence," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 4643–4646.
- [102] L. Bu, J. Zhang, Z. Zhang, Y. Yang, and M. Deng, "Deep learning for integrated speckle reduction and super-resolution in multi-temporal SAR," *Remote Sensing*, vol. 16, no. 1, 2024. [Online]. Available: <https://www.mdpi.com/2072-4292/16/1/18>
- [103] J. W. Smith, Y. Alimam, G. Vedula, and M. Torlak, "A vision transformer approach for efficient near-field SAR super-resolution under array perturbation," in *2022 IEEE Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)*. IEEE, Apr. 2022. [Online]. Available: <http://dx.doi.org/10.1109/WMCS55582.2022.9866326>
- [104] E. Hoffer and N. Ailon, "Deep metric learning using triplet network," in *Similarity-based pattern recognition: third international workshop, SIMBAD 2015, Copenhagen, Denmark, October 12-14, 2015. Proceedings 3*. Springer, 2015, pp. 84–92.
- [105] "The Air Force moving and stationary target recognition database. available online.;" <https://www.sdms.af.mil/index.php?collection=mstar>, [Accessed 26-06-2024].
- [106] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Computer vision—ECCV 2016: 14th European conference, amsterdam, the netherlands, October 11–14, 2016, proceedings, part VII 14*. Springer, 2016, pp. 499–515.
- [107] M. Reggiannini, M. Righi, M. Tampucci, A. Lo Duca, C. Bacciu, L. Bedini, A. D'Errico, C. Di Paola, A. Marchetti, M. Martinelli, C. Mercurio, E. Salerno, and B. Zizi, "Remote sensing for maritime prompt monitoring," *Journal of Marine Science and Engineering*, vol. 7, no. 7, 2019. [Online]. Available: <https://www.mdpi.com/2077-1312/7/7/202>
- [108] H. Lang, G. Yang, C. Li, and J. Xu, "Multisource heterogeneous transfer learning via feature augmentation for ship classification in SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.
- [109] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.
- [110] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [111] 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*, 2016, pp. 770–778.
- [112] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [113] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [114] W. Li, W. Yang, Y. Hou, L. Liu, Y. Liu, and X. Li, "Saratr-x: Towards building a foundation model for sar target recognition," *IEEE Transactions on Image Processing*, 2025.
- [115] W. Li, W. Yang, T. Liu, Y. Hou, Y. Li, Z. Liu, Y. Liu, and L. Liu, "Predicting gradient is better: Exploring self-supervised learning for sar atr with a joint-embedding predictive architecture," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 218, pp. 326–338, 2024.
- [116] D. Szwarcman, S. Roy, P. Fraccaro, O. E. Gíslason, B. Blumenstiel, R. Ghosal, P. H. De Oliveira, J. L. de Sousa Almeida, R. Sedona, Y. Kang *et al.*, "Prithvi-eo-2.0: A versatile multi-temporal foundation model for earth observation applications," *IEEE Transactions on Geoscience and Remote Sensing*, 2025.
- [117] Y. Cong, S. Khanna, C. Meng, P. Liu, E. Rozi, Y. He, M. Burke, D. Lobbell, and S. Ermon, "Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery," *Advances in Neural Information Processing Systems*, vol. 35, pp. 197–211, 2022.
- [118] Z. Xiong, Y. Wang, F. Zhang, A. J. Stewart, J. Hanna, D. Borth, I. Papoutsis, B. L. Saux, G. Camps-Valls, and X. X. Zhu, "Neural plasticity-inspired multimodal foundation model for earth observation," *arXiv preprint arXiv:2403.15356*, 2024.
- [119] J. Jakubik, F. Yang, B. Blumenstiel, E. Scheurer, R. Sedona, S. Maurogiovanni, J. Bosmans, N. Dionelis, V. Marsocci, N. Kopp *et al.*, "Terramind: Large-scale generative multimodality for earth observation," *arXiv preprint arXiv:2504.11171*, 2025.
- [120] F. Liu, D. Chen, Z. Guan, X. Zhou, J. Zhu, Q. Ye, L. Fu, and J. Zhou, "Remoteclip: A vision language foundation model for remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–16, 2024.
- [121] A. Fuller, K. Millard, and J. Green, "Croma: Remote sensing representations with contrastive radar-optical masked autoencoders," *Advances in Neural Information Processing Systems*, vol. 36, pp. 5506–5538, 2023.
- [122] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging properties in self-supervised vision transformers," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 9650–9660.
- [123] Y. Jiang, S. Chang, and Z. Wang, "Transgan: Two pure transformers can make one strong gan, and that can scale up," *Advances in Neural Information Processing Systems*, vol. 34, pp. 14 745–14 758, 2021.
- [124] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
- [125] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.
- [126] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should i trust you?' Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

## VIII. BIOGRAPHY SECTION



**Ch Muhammad Awais** is a Ph.D. student at the University of Pisa, specializing in SAR ship classification. Hailing from Pakistan, he has been engaged in DL research for the past six years and has worked on SAR ship classification for the last two years.



**Marco Reggiannini** received his M.Sc. in Physics and a Ph.D. in Automation Engineering from the University of Pisa. Currently, he is a researcher at the Signals and Imaging Lab at ISTI-CNR, with more than ten years of experience devoted to multisensor imagery analysis.



**Davide Moroni** received his M.Sc. in Mathematics (Hons.) from the University of Pisa and a Ph.D. in Mathematics from the University of Rome La Sapienza. He is a Senior Researcher and Head of the Signals and Images Lab at ISTI-CNR, specializing in image processing and computer vision.



**Emanuele Salerno** graduated in Electronic Engineering from the University of Pisa in 1985, joined CNR in 1987 and is now a Senior Research Associate at ISTI-CNR. His expertise spans SAR image processing, cultural heritage technologies, and computational biology, with a significant academic and research background in signal and image processing.