EAI Endorsed Transactions
on Industrial Networks and Intelligent Systems

Research Article EAI.EU

CFM-SAR: A Class-Conditioned Flow Matching with SCM-UNet for SAR Image Generation

Nazanin Baramaki[1], Quang Nhat Le[1,*], Muhammad Fahim[2], and Tuan Thanh Nguyen[3]

[1]Memorial University, St. John's, NL A1B3X5, Canada
[2]Queen's University Belfast, Belfast BT7 1NN, UK
[3]School of Computing and Mathematical Sciences, University of Greenwich, UK

Abstract

Monitoring and analyzing sea-ice changes is critically important for navigation through polar regions, ecosystem monitoring, and environmental risk assessment, and Synthetic Aperture Radar(SAR) imagery is one of the most reliable and useful data sources for this goal. SAR images exhibit complex backscatter patterns that are generally not interpretable without technical expertise. Consequently, accurate annotation of SAR images requires domain experts to identify the corresponding sea ice classes, resulting in labelled datasets that are both limited and often imbalanced across classes.A further challenge is that SAR image channels differ from those of conventional images, with HH, HV, and incidence angle each containing important physical information. Although current image generation methodologies have strong capabilities in RGB images and have solved several limitations, including training instability and high computational cost, some challenges in SAR image generation remain. These issues, such as limited preservation of channel-specific physical information and mode collapse, motivate us to work in generative modelling and seek to overcome them. In this study, we propose a Flow Matching framework with a class-informed prior. Instead of using a standard Gaussian distribution as the initial point for learning, the Class-Conditioned Flow Matching (CFM) model estimates the mean and variance of each image class and uses them to better approximate the data distribution from a more informative starting point. Another contribution of this paper is the proposed Swin Transformer Convolutional Hybrid (SCM-UNet) model, which generates images using a modified U-Net architecture. Experiments on the chosen dataset demonstrate that the proposed model improves visual metrics in generated images (FSD=19.23, PSNR=28.0, and SSIM=0.83) and effectively enhances the downstream task. Results show that, when trained exclusively on generated data and evaluated on real SAR images, the classification model achieved an average accuracy of 91.61% and an F1-score exceeding 91.5%.

Keywords: Class-based prior, flow matching, SAR image classification, SAR image generation, SCM-UNet

Received on 17 July 2026, accepted on 12 October 2026, published on 02 November 2026

Copyright © 2026 Nazanin Baramaki *et al.*, licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetinis.133.14079

1. Introduction

Sea ice plays a vital role in understanding polar climate dynamics. Accurate monitoring of the sea ice extent, concentration, and morphological changes helps scientists to understand climate variability and predict long-term changes in the Arctic and Antarctic regions. Studying sea-ice changes reveals its broader impact on the global climate system. Analyzing these changes helps scientists to obtain information that supports maritime navigation, ecosystem monitoring, and environmental risk assessment. In recent years, as sea ice cover in polar regions has diminished, new shipping routes have been developed, altering

*Corresponding author. E-mail: qnle@mun.ca

marine habitats. These new routes, which previously were unused in polar areas, are becoming economically important [1]. Synthetic Aperture Radar (SAR) imagery is an important and reliable source of surveillance in polar regions. The SAR sensors operate in darkness, all day and night, even through dense clouds, and enable continuous observation in harsh Arctic and Antarctic areas [2]. SAR is a good source for tracking ice extent, deformation, leads, ridges and seasonal changes. SAR images are highly sensitive to physical and imaging parameters, including surface roughness, salinity, moisture, incidence angle, and sensor frequency, which all influence the backscatter intensity of sea ice. Different ice types may exhibit similar responses to changing environmental conditions, which is one of the main reasons SAR image interpretation is challenging and why deep learning methods are valuable. Deep learning methods can learn the complex relationships between these factors and the resulting backscatter patterns. SAR imagery contains types of noises and limitations, including speckle noise [3], thermal noise, clutter and background scattering, geometric distortions, and phase noise [4]. Manual labelling of these SAR images is time-consuming and requires experts, yet they are not in sufficient quantity for use by deep learning models. At the same time, sea-ice SAR imagery is a reliable remote sensing tool with robust preprocessing [5]. Sea-ice SAR images are used in various applications, such as sea ice classification [6], sea ice mapping (producing operational ice charts) [7], sea ice concentration estimation [8], sea ice thickness estimation [9], sea ice motion (drift) tracking [10], ice edge detection/extraction [11], iceberg detection and monitoring [12], ship navigation and route planning in polar waters [13], offshore operations (supporting oil, gas, and wind energy infrastructure) [14], climate change and environmental monitoring [15], search and rescue operations in ice-covered regions [16], hazard detection and risk assessment for maritime activities [2]. Sea-ice classification is among the most popular image classification-related tasks [17].

One of the most important applications of SAR images is the classification of sea ice in polar regions to facilitate navigation in these areas. Deep learning models are popular for image analysis, including image classification, and require sufficient training data. SAR datasets often have a limited number of images per class, or they are imbalanced, as in the dataset used in this paper [18]. To address these problems and due to the costs of SAR image extraction, preprocessing, and labelling, the image generation subject is introduced as an essential approach. Lack of training data or an existing imbalance in the dataset, which are the most common problems in SAR image datasets, is the most significant challenge that deep learning-based image analysis applications face. To overcome these issues, image generation techniques are widely used. These models aim to generate realistic samples and provide diverse training data through augmentation and database balancing, thereby improving the model's robustness and generalization.

Additionally, these generated images help models learn tasks when real data is sensitive or scarce, and help simulate rare scenarios for model evaluation. In recent years, generative models such as generative adversarial networks (GANs) [19], diffusion models [20], and flow-based methods [21] have improved the quality and quantity of labelled datasets. These enhanced datasets have significantly improved deep learning models' ability to classify, recognize, and reconstruct in medical, remote sensing, and other computer vision tasks [22].

It is compelling to explore these data generation models and investigate their ability and efficiency in producing SAR images, as well as the quality of these generated images in classification tasks. The SAR images are not merely visual representations, they also contain geospatial and physical information preserved within each channel. Therefore, unlike conventional optical imagery, SAR data is formed through electromagnetic backscatter mechanisms, making the generation process more challenging and physically constrained [23]. The goal of a generative model is to improve performance on downstream tasks such as classification and segmentation, and to mitigate limitations imposed by scarce labelled datasets. In this work, we present a flow-matching method with a class-aware prior that uses the proposed SCM-UNet architecture for image generation, trained to learn the distribution of each class in the dataset. This SCM-UNet is a model based on the UNet architecture, combined with ideas from spatial attention and depthwise convolutional neural networks (CNNs). After training, the generating model produces images, which are then used to train a classification model. The results were thoroughly discussed and show impressive classification performance and enhanced image quality.

Figure 1 describes the process of analyzing the SAR sea-ice images. The provided workflow begins with downloading dual-polarization SAR images from satellites such as Sentinel-1, followed by preprocessing steps including radiometric calibration, speckle filtering, terrain correction, normalization, and data annotation. Afterward, the preprocessed data are used for feature extraction and latent encoding. Subsequently, the proposed generative model is trained on the real SAR images to synthesize realistic SAR samples, which are used to augment the training dataset. Finally, the SCM-Net classification model is trained using the generated images to predict the corresponding sea-ice classes. At the last stage, some of the downstream applications are mentioned.

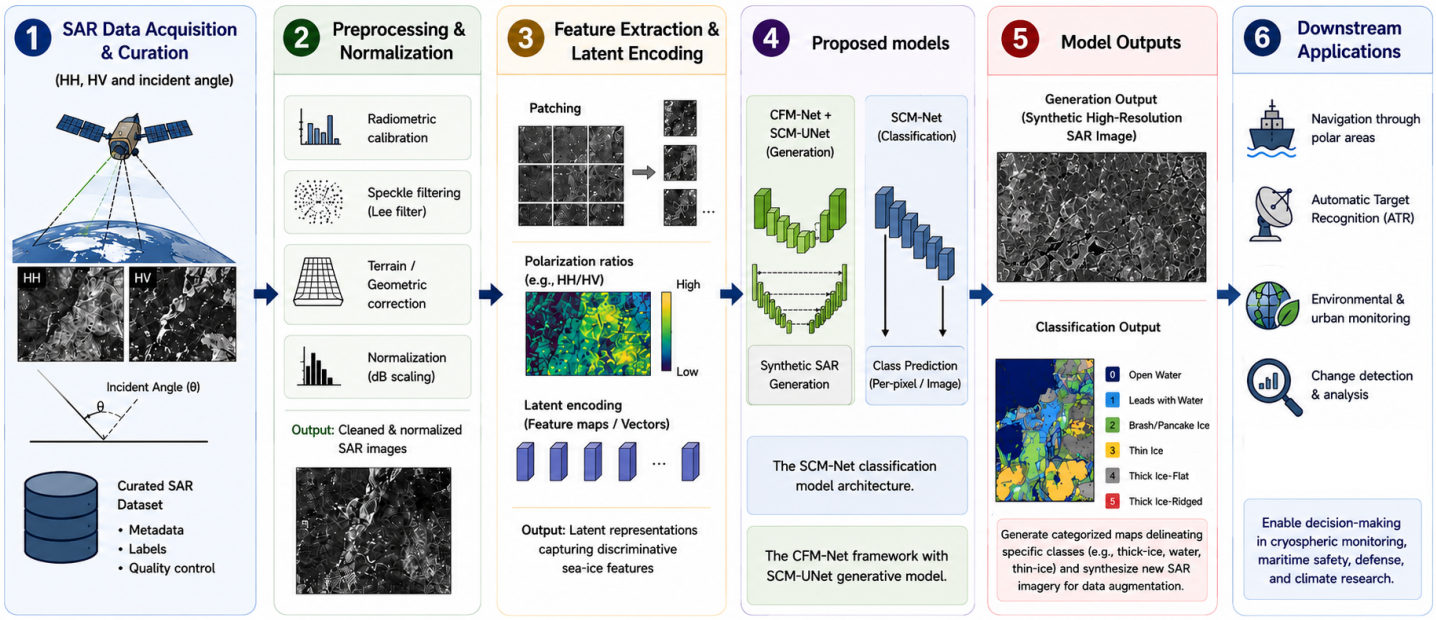


Figure 1. Overview of the proposed SAR sea-ice generation and classification framework
The pipeline consists of six stages:
(1) SAR data acquisition and curation, (2) preprocessing and normalization, (3) feature extraction and latent encoding,
(4) proposed deep learning model (CFM-Net with SCM-UNet for data generation and SCM-Net for classification),
(5) classification and synthetic SAR image generation, and (6) downstream applications

1.1. Related Works

SAR is an important sensing method for sea-ice monitoring, as it is accessible in all weather conditions and does not depend on daylight. The scarcity of diverse, well-labeled SAR training data has driven growing interest in generative modeling for sample creation and deep learning-based enhancement in SAR-based sea-ice analysis. Passah et al. [24] summarized a decade of research on SAR image analysis, placing sea-ice mapping within the broader SAR computer vision landscape. Their review identified p ersistent challenges, including limited labeled data, domain shift, speckle noise, and cross-sensor variability, and highlighted the emergence of convolutional architectures as a dominant baseline for SAR image understanding tasks. In another study, Stofa et al. [25] developed a model for classification o f s ea-ice S AR i mages t hat combines group convolution, asymmetric channel weighting, and lightweight feedforward augmentation, and showed that deep learning-based methods have become central to high-resolution sea-ice concentration monitoring. However, that performance still depends strongly on dataset design, geographic transferability, and the handling of ambiguous textures in marginal ice zones. Khaleghian et al. [18] suggested a modified VGG-Net for multi-class ice-type classification u sing Sentinel-1 data, and demonstrated that learned representations can outperform more traditional approaches when sufficient labelled data are available. Different types of ice can appear similar in SAR backscatter data, and ice images from different classes can produce overlapping radar responses. Therefore, they are difficult for classification models to distinguish accurately. One of the most important parts of SAR image analysis is having datasets which contain diverse data for each class with reliable labels. Khaleghian et al. [18] suggested that improving the dataset's quality and ensuring correct labelling may be more important than simply using a deeper or more complex neural network. In SAR image classification, some works have extended beyond pure intensity-based learning and incorporated domain-relevant physical texture features. Wan et al. [26] proposed a multi-feature CNN framework that combines polarization decomposition information, spectrogram-derived descriptors, and learned spatial representations to improve classification performance on multi-polarization SAR data. Stokholm et al. [27] found that different sea-ice conditions often look identical in Sentinel-1 satellite imagery. Because of this similarity, machine learning models cannot rely solely on local patterns and must use surrounding context to identify ice types accurately. Stokholm et al. [27] showed that robustness, explainability, and deployment realism are becoming first-class design criteria, rather than accuracy, which was traditionally the most

important criterion in curated benchmarks. The limited size and diversity of labeled SAR datasets remain the most significant challenges for discriminative models, leading to growing interest in generative models for SAR image generation [28]. Current approaches of image generation begin with Goodfellow et al. [19] by introducing GANs. In SAR imagery, preserving texture, speckle characteristics, and target geometry is critical. GANs have been attractive because they can learn high-dimensional image distributions and generate accurate samples. Still, generated samples have problematic issues, such as instability, mode collapse, and sensitivity to architectural and optimization choices [29]. Oghim et al. [30] proposed the DH-GAN framework, a dual discriminator with a high-frequency pass filter, which preserves SAR-relevant detail for target recognition applications. In addition, they discussed the challenge of generating SAR images with fine-scale structure, which is physically realistic in situations with limited training data. Chen et al. [31] addressed the challenge of enhancing the texture and edge quality in SAR image generation and introduced a multi-scale fusion GAN (MSF-GAN) that fuses shallow and deep features to strengthen texture. In another study, similarly, Hwang et al. [32] proposed a modified CycleGAN framework to maintain structural fidelity and edge definition. They leveraged dual generators and a cycle-consistency loss to preserve visual details during generating optical data from SAR images.

Although GAN-based generative models can generate artificial SAR images and handle dataset imbalance, preserving physically meaningful properties remains a challenge in these augmentation models. As an alternative, diffusion models emerged. Denoising diffusion probabilistic models (DDPMs) were proposed by Ho et al. [20], in which the model learns the data generation by reversing a progressive noising process. Compared with GANs, diffusion models often offer better mode coverage and training stability, albeit at the cost of slower iterative sampling. Qosja et al. [33] presented SAR image synthesis with a diffusion-based generative model and showed that it can be transferred from natural image settings into radar imagery while preserving useful SAR characteristics. In [34], the authors proposed a single-SAR-image generation method based on the DDPM (SSDDPM) for image generation in SAR images, which uses an attention-based noise estimate generation model (ANEM). Although their model achieved promising results in generating SAR images from a single image, complex physical characteristics, such as coherent spots (speckle noise), sea clutter, and shadows still appear in the generated SAR images.

Flow Matching (FM) is a recent development introduced by Lipman et al. as a simulation-free training framework for continuous normalizing flows [21]. FM trains a model to predict a vector field (the direction in which data should move) to transport samples from a noise distribution to the target data distribution along predefined probability paths. By directly learning this vector field, FM avoids the repeated numerical simulations required by earlier continuous-flow methods, resulting in more efficient training. Tong et al. [35] used mini-batch optimal transport and showed that these optimal paths can improve and generalize flow-based generative models. Although FM-based conditional methods incur lower computational cost during sampling, they can still be used to model complex data distributions. Due to the ability of the FM-based models to learn the complex data distributions, they are widely used in remote-sensing pipelines where both realism and efficiency matter. The D-Flow (Differentiating through Flows) model enables continuous flows to avoid retraining the entire network by applying specific conditions or solving inverse problems [36]. By propagating gradients directly through the ODE paths, D-Flow overcomes the challenge of controlled generation, enabling users to apply pre-trained flow models flexibly with minimal task-specific retraining while also ensuring adherence to structural or semantic constraints. Flow-based generative models contain several approaches, including normalizing flows, continuous normalizing flows trained via flow matching, rectified flows, and related variants [37]. Pyramidal Flow Matching is another type of flow-based generative model introduced to optimize video and image generation with high clarity and computational efficiency [38]. This approach divides the vector field's trajectory into a series of hierarchical pyramid stages rather than calculating it at a single, massive resolution. The production process starts with low resolution and gradually increases it step-by-step, so that the full-resolution output is produced only in the final stage. At this point, the computational cost is significantly reduced. This model employs Diffusion Transformer (DiT) architecture and introduces a novel denoising formulation to bridge the resolution transitions across different generation stages, thereby integrating the entire generation process into a single, coherent framework [39]. It greatly reduces the training cost and memory footprint, while preserving high-quality generation. The Diffusion Transformer is a powerful backbone for video and image synthesis. DiT was initially introduced in the context of diffusion models and has subsequently been extended to other generative modeling frameworks. In fact, for diffusion models, the diffusion loss was used for training, whereas in flow-based models, the flow matching loss was used.

The Harmonizing Flow model is proposed to demonstrate the normalization capabilities of flow-based models in highly complex, physically constrained domains, such as medical MRI imaging [40]. Synthetic images from different modalities, such as SAR or MRI,

often exhibit significant range-domain variations. This method uses an unsupervised normalization mechanism to provide a precise source-domain description, and then trains a harmonizer network to map misaligned and out-of-distribution images to this learned flow. This technique helps ensure that the produced images do not require paired training data to learn the target physical description.

Flow-based models are not limited to Gaussian source distributions and can learn mappings between arbitrary source and target distributions [41]. Building on this flexibility, Feng et al. [42] proposed the first general guidance framework for flow matching, including a Monte Carlo-based, training-free, asymptotically exact guidance method. The proposed framework is able to generate samples steered toward desired properties, such as class-conditioned outputs or specific target characteristics, and to enable controlled generation across a broad range of flow-matching models, without requiring additional training.

In summary, the literature review on generative models indicates that synthesis pipelines informed by prior knowledge of the texture or other physical properties of SAR images help address data scarcity and generate high-quality SAR data. Generative models, such as GANs, diffusion models, and flow-based methods try to offer a balanced trade-off among fidelity, diversity, and efficiency [30]–[35]. These results motivate an expanded research direction at the intersection of sea-ice mapping and generative modelling. As discriminative models require richer and more transferable representations of polarimetric SAR images, a technically significant gap remains between the capabilities of sea-ice classification models and the availability of suitable SAR datasets for training. While classification models have advanced significantly, state-of-the-art approaches remain predominantly deep learning-based, whereas the availability and quality of large, balanced, labeled datasets remain limited. We can address this issue by using generative modeling and physically realistic methods that support training, augmentation, and uncertainty analysis without introducing non-realistic structures into the image. The generated SAR sea-ice images must preserve the physical behavior of SAR radar images, retain the key radar characteristics (patterns and textures) of different classes of real ice images, and be as accurate as possible so that we can use them in large operational satellite systems.

1.2. Problem Statement and Research Objectives

We need to build a classifier that achieves high accuracy in classifying water and ice, and several types of ice. This high accuracy is vital for safe navigation and weather forecasting in polar areas. One of the best image types for this kind of analysis is SAR imagery, due to its accessibility in all weather conditions and 24 hours a day. The main challenges in using SAR data are high costs, difficulty of interpretation, and thermal noise. Therefore, the goal is to develop classifiers that can leverage the limited number of available images and achieve high-accuracy predictions. Another difficulty with SAR datasets is the imbalance in the number of images per class. For example, in the dataset that is used in this paper, the number of brash/pancake images is much lower than that of open-water images, and even with the best state-of-the-art deep learning methods, there is a risk of overfitting due to this imbalance. Another reason for generating fake datasets is the limited availability of labelled data. SAR datasets from special physical places are typically small and expensive to annotate. Expert knowledge of physical scattering mechanisms and domain variability is needed. Incidence angles, polarization, i.e., co-polarization (HH) and cross-polarization (HV), sensor configuration, and environmental conditions are among the factors that make SAR images more diverse.

Generative modeling aims to improve image classification by exposing the classifier to an augmented dataset that produces diverse variations of the original dataset, preventing overfitting to specific association patterns and enhancing robustness to speckle noise. The SAR images inherently contain the multiplicative effect of speckle noise. In this type of image, noise injection and filtering variations help the model learn invariant features and more diverse samples. Therefore, more samples help smooth the optimization landscape and achieve better convergence, thereby stabilizing training.

In this research, the focus was on creating a fake balanced dataset of SAR images with peak signal-to-noise ratio (PSNR) and image quality comparable to those of the real image dataset, usable by a classifier, and achieving high accuracy. The focus was on image generation using mathematical and deep learning methods to produce high-quality images. For the learning framework, we employ FM with class-wise prior. For the data generation stage, we propose the SCM-UNet generative backbone, a model based on the U-Net architecture designed for image generation. In the classification section, we use the SCM-Net classifier, which is a lightweight classification model, proposed in our previous paper [43].

The main research objectives of our study are as follows:

- **O1: Proposing a new learning framework**

  We propose a modified FM framework that replaces the standard isotropic Gaussian base distribution with a class-conditional Gaussian prior, enabling class-aware generative modeling.

- **O2: Proposing SCM-UNet image generation architecture**

  We propose a new image generation architecture, SCM-UNet, which generates images using a FM learning algorithm, demonstrating its superiority compared to other state-of-the-art models.

- **O3: Image generation performance benchmarking**

  We evaluate and compare the proposed image generation framework and generative model, SCM-UNet, against widely used state-of-the-art deep learning models (including GANs, diffusion models, variational autoencoders (VAEs), and other flow matching variants) for sea ice multi-class classification.

- **O4: Image classification performance benchmarking**

  We evaluate and compare the generated SAR image dataset using the SCM-Net classifier by training solely on synthetic data and testing on real data, demonstrating its effectiveness in improving classification performance on highly imbalanced multi-class sea ice datasets.

1.3. Contributions and Paper Organization

The main contributions of this paper are summarized as follows:

- The paper proposes a class-based FM learning methodology that modifies the standard FM algorithm. This novel approach achieves a high PSNR for pixel-level fidelity, as well as strong structural similarity index measure (SSIM) and Fréchet SAR distance (FSD) scores for perceptual realism.
- SCM-UNet is benchmarked against standard U-Net–based image generation models and other state-of-the-art architectures, including GANs, diffusion models, VAEs, and FM variants, for SAR image generation.
- We evaluate and compare the generated SAR image dataset using the SCM-Net classifier by training exclusively on synthetic data and testing on real data, demonstrating its effectiveness in improving classification performance on highly imbalanced multi-class sea-ice datasets.

The organization of this paper is as follows: Section 2 describes the dataset and preprocessing pipeline, then provides the formal problem formulation. Section 3 introduces the proposed model's architecture, providing a detailed breakdown of the learning methodology, the generation model's core components, and an overview of other models used for comparison. Section 4 details the experimental setup and evaluation metrics. Section 5 illustrates the results of the proposed generative models in addition to performance analysis using key performance metrics, as well as comparison with other models. Moreover, all generated images are passed through SCM-Net, the image classification model used, to evaluate its performance when tested on real data. In section 6, an ablation study was conducted to evaluate the effectiveness of the proposed model and the learning framework, comparing them with other existing models. Finally, Section 7 concludes the paper by summarizing the main findings and highlighting directions for future research.

2. Dataset and Problem Formulation

2.1. Dataset Description

To analyze the efficacy of the proposed learning algorithm and our image generation model in generating realistic data, we use a SAR-based ice types/ice edge dataset as the target distribution for image generation and final classification [18]. The dataset was created from 31 Sentinel-1A Extended Wide (EW) Level-1 Ground Range Detected (GRD) scenes acquired north of the Svalbard archipelago during the winter months between September and March from 2015 to 2018. The scenes have a spatial resolution of approximately 40m × 40m per pixel and contain different sea-ice types, including (a) thick ice (ridged), (b) leads with water, (c) open water, and (d) thin ice. In practice, the model uses a three-channel input comprising HH polarization, HV polarization, and the incidence angle. The original SAR scenes were subsequently divided into smaller image patches for model training and evaluation. To construct the dataset, the original images taken from the Sentinel-1 satellite were divided into smaller patches, and then these images were manually annotated into six classes to generate labeled training and evaluation samples. For more detail, the dataset contains six classes: open water, leads with water, brash/pancake ice, thin ice, thick ice-flat, and thick ice-ridged. The data records, called patches, were extracted from each polygon with a stride of 10, in sizes 10×10, 20×20, 32×32, 36×36, and 46×46 pixels per class. In this study, we used only 32×32 patches for simplicity and comparability with our previous work. Each patch contains three channels, HH, HV, and incidence angle, in the shape of (3, patch_size, patch_size). The ranges for the HH, HV, and incidence angle are -30 to 0 dB, -35 to -5 dB, and 19 to 46 degrees, respectively. The dataset distribution reveals a severe class imbalance. For a more precise description, the "Open Water" and "Thick Ice-Ridged" classes contain 8,394 and 4,855 samples, respectively. In contrast, the "Leads with Water" and "Brash/Pancake Ice" classes contain 924 and 159 samples, respectively.

The remaining two classes, "Thin Ice" and "Thick Ice-Flat," are also underrepresented, with only 202 and 213 samples, respectively.

2.2. Problem Formulation

SAR images are among the best choices for analyzing polar data (e.g., for navigating through polar areas). Identifying the type of ice or finding a lead and/or a way through the Arctic is a vital need in those areas. This data often suffers from limited labelled data, class imbalance, and strong variability caused by speckle or thermal noise, acquisition geometry, and backscatter characteristics. These challenges reduce the generalization ability of the deep learning-based classifier, especially when some classes are underrepresented. To address this issue, SAR image augmentation was proposed as a class-conditional generative model. The goal is to generate a balanced dataset by learning the distribution of SAR images for each class, generating realistic synthetic samples, and enriching the dataset, which leads to a high-accuracy classifier.

Let the training dataset be

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}, \tag{1}$$

where $x_i \in \mathbb{R}^{C\times H\times W}$ denotes a SAR image with C channels and spatial size $H \times W$, and $y_i \in \{1, 2, \ldots, K\}$ denotes its class label among K classes. The objective of image augmentation is to generate additional synthetic samples $\tilde{x}$ such that, for each class k,

$$\tilde{x} \sim p(x \mid y = k). \tag{2}$$

The goal is to generate samples that closely match the true class-specific data distributions. These samples are then added to the original training set to improve classifier performance, especially for minority or difficult classes. In some cases, if the generated samples are of high quality, they could be used solely for training, and the real dataset can be used for testing the classifier (as in this paper).

3. Methods

3.1. Class-Mean-Based Flow Matching

In this paper, a class-wise FM framework for SAR image augmentation was proposed. The FM framework [21] formulates generative modelling as learning a continuous transformation between a simple base distribution and the data distribution. In the FM image augmentation method, instead of directly learning a fixed mapping from noise to image (as it is in diffusion models), the model learns a continuous transformation from a simple prior distribution (usually a simple Gaussian $\mathcal{N}(0, 1)$) to the target class-conditioned SAR data distribution. Typically, the base distribution is chosen as a Gaussian as follows:

$$x_0 \sim \mathcal{N}(0, 1). \tag{3}$$

$(x_0 \sim p_0(x))$ denotes a latent variable sampled from a simple base distribution, which is typically a Gaussian distribution. The ultimate goal of the model is to learn a vector field that transports samples from this base distribution to the target data distribution $(q(x_1))$. Traditional diffusion models rely on a stochastic forward-reverse process to model the transformation from noise to data. In contrast, Flow Matching (FM) directly learns a time-dependent vector field that defines this transformation.

Flow Matching (FM) is a training framework for continuous normalizing flows (CNFs), in which a neural network is trained to approximate the velocity field governing the evolution of samples over time. The FM generative process defines a time-dependent trajectory x_t for $t \in [0, 1]$, where $x_0 \sim p_0$ and $x_1 \sim p(x \mid y)$.

The generative process is defined by a time-dependent velocity field $v_t(x)$, which specifies the instantaneous direction and speed at which a sample at position x and time t must move to follow the probability path from the prior to the data distribution. Integrating this field from $t = 0$ to $t = 1$ transports a prior sample to a data sample. Since $v_t(x)$ is not available in closed form, flow matching trains a neural network $v_\theta(x_t, t, y, \eta)$ to approximate it by regressing onto the conditional target velocity u_t.

The FM objective function minimizes the difference between the predicted vector field and the target conditional velocity field along the following probability path:

$$\mathcal{L}_{FM}(\theta) = \mathbb{E}_{x_0,x_1,t}\left[\|v_\theta(x_t, t, y) - u_t(x_0, x_1)\|^2\right], \tag{4}$$

where $u_t(x_0, x_1)$ is the prescribed target velocity along the interpolation path between the noise sample x_0 and real sample x_1. This update process allows the model to progressively learn to transform noise into realistic SAR images belonging to a given class by learning the true distribution of that class.

We changed the starting point distribution from a simple Gaussian $\mathcal{N}(0, 1)$ to a class-mean-based Gaussian distribution. As can be seen from Figure 2, starting from $p_0(x \mid y) = \mathcal{N}(\mu_y, \sigma_y^2 I)$ substantially shortens the transport path to the target distribution $p_1(x \mid y)$, giving the network an easier trajectory to learn.

In other words, instead of using a single global Gaussian prior, the model adopts a class-conditional Gaussian prior. Specifically, the initial sample is drawn as

$$x_0 \sim \mathcal{N}(\mu_c, \sigma_c^2 I), \tag{5}$$

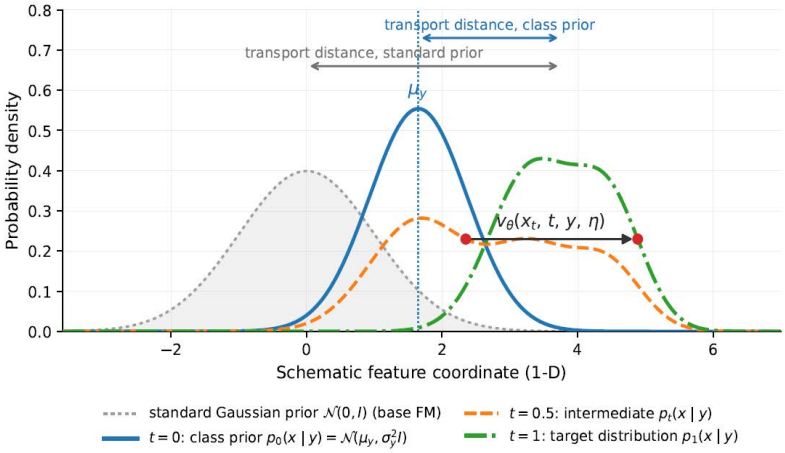


Figure 2. Schematic illustration of class-conditional flow matching The model transports samples from the class prior $p_0(x \mid y) = \mathcal{N}(\mu_y, \sigma_y^2 I)$ at $t = 0$ to the class-specific SAR data distribution $p_1(x \mid y)$ at $t = 1$, passing through intermediate distributions $p_t(x \mid y)$. The network v_θ predicts the velocity v_{pred} that moves each sample along this path. Anchoring the prior at the class mean μ_y shortens the transport distance compared with a standard Gaussian prior.

where each class c has its related mean μ_c and variance σ_c^2. In the flow matching standard formulation, there is a global prior, whose parameters are initialized to $\mu = 0$ and $\sigma^2 = 1$, used similarly across all classes and independent of the class labels. The informed prior makes the model dependent on the class label explicitly and injects prior knowledge about class-specific data distributions into the generative process.

Instead of starting the flow learning process from a blind distribution such as a standard Gaussian, this paper conducted an ablation study that showed that when the model starts with a class-dependent distribution, the learning process becomes faster as it starts closer to the target distribution for each class, reducing transport complexity and accelerating convergence during training. Choosing this starting point helps the learned model become more stable and support image generation with more class consistency. The proposed class-aware prior-guided model converges faster and helps the model learn more structured and discriminative flows. This guidance is particularly beneficial in datasets with strong inter-class variability, such as SAR imagery.

Accordingly, in the proposed FM framework, the sampling of x_0 is conditioned on class labels. Each sample that is generated in each class uses its corresponding class-specific parameters μ_c and σ_c^2. The prior distribution becomes label-aware during both training and testing as a result.

Our implementation uses an optimal transport (OT) FM update rule to ensure efficient training and stable convergence. Unlike traditional diffusion models that rely on score matching to estimate and update the gradient of the data density at each step, our proposed model uses the FM with an OT path. This technique is used to train generative models and learns a velocity field $v_\theta(\mathbf{x}_t, t, y)$, which is essentially the direction and speed to move a particle along a straight line in probability space. This direction is the shortest path between the class-conditional Gaussian prior $\mathbf{x}_0$ and the target SAR data $\mathbf{x}_1$. This time-dependent vector field $v_\theta(x_t, t, y)$ aims to match the default Gaussian distribution to a constant-velocity target $u_t(x_t \mid x_0, x_1)$ which is defined as

$$u_t(x_t \mid x_0, x_1) = x_1 - (1 - \sigma_{\min})x_0. \tag{6}$$

The model learns the aforementioned objective function, which ensures that the probability mass is transported along a straight line, that is the shortest possible path, from the class-conditional Gaussian prior to the target SAR data distribution. The objective function in OT is given by

$$\min_\theta \mathbb{E}_{t,\mathbf{x}_0,\mathbf{x}_1} \|v_\theta(\mathbf{x}_t, t, y) - (\mathbf{x}_1 - (1 - \sigma_{\min})\mathbf{x}_0)\|^2. \tag{7}$$

An ordinary differential equation (ODE) solver is a computational algorithm that computes the step-by-step path (trajectory) of a system given only its rate of change (velocity or vector field) and its initial state. The OT formulation enables the model to benefit from straight-line trajectories, which especially simplify ODE integration during inference and allow high-quality SAR image generation with fewer sampling steps. The OT-based FM objective provides a more stable and computationally efficient training signal than other flow-based models, such as score matching. Having high gradients in low-density regions can lead to numerical instability in traditional score matching. By using a deterministic ODE solver (e.g., the midpoint method) during inference, we exploit the model's straight-line transport trajectories and enable high-fidelity SAR image synthesis in significantly fewer steps than score-based diffusion methods.

The backbone that we used to parameterize this velocity field is the SCM-UNet architecture. The SCM-UNet is designed to capture both global semantic structure and fine spatial details in SAR images through a multi-scale encoder-decoder framework with skip connections. Given an input interpolated sample $\mathbf{x}_t$, time embedding t, class condition y, and incidence angle condition η, the SCM-UNet predicts the following deterministic velocity field:

$$\mathbf{v}_\theta(\mathbf{x}_t, t, y, \eta) \approx \frac{d\mathbf{x}_t}{dt}. \tag{8}$$

The dataset, which is finally augmented, is constructed as

$$\mathcal{D}_{aug} = \mathcal{D} \cup \{(\tilde{x}_j, y_j)\}_{j=1}^{M}, \tag{9}$$

where $\tilde{x}_j$ are generated class-conditioned SAR samples.

Overall, the goal is for the proposed class-conditioned generative model to learn to produce realistic, diverse SAR images for each category. This model will use the FM framework, with an OT path, to achieve this goal. The SCM-UNet architecture utilizes this learning framework to predict the velocity field (or flow field) by optimizing the model to find the linear path between the class-conditional Gaussian priors and the target SAR distribution. By capturing the global structure and fine-grained SAR texture details through this deterministic transport mechanism, the model can generate high-fidelity SAR images. The algorithm of the proposed class-wise FM methodology is described in Algorithm 1.

Algorithm 1. Class-Conditional Flow Matching for SAR Generation (CFM-SAR)

Input: Training dataset $\mathcal{D}$, class statistics $\{(\mu_c, \sigma_c)\}$
Output: Trained SCM-UNet model and generated image

/* Phase 1: Training */
repeat
  Sample real image $\mathbf{x}_1$, class c, and angle η from $\mathcal{D}$
  Sample random time $t \sim \mathcal{U}(0, 1)$
  Create class-specific starting noise $\mathbf{x}_0$ using (μ_c, σ_c)
  Create intermediate image $\mathbf{x}_t$ by blending $\mathbf{x}_0$ and $\mathbf{x}_1$ at time t
  Calculate the exact target direction $\mathbf{u}_t$ from $\mathbf{x}_0$ to $\mathbf{x}_1$
  Predict direction $\mathbf{v}_{pred}$ using SCM-UNet$(\mathbf{x}_t, t, c, \eta)$
  Train model by minimizing the difference between $\mathbf{v}_{pred}$ and $\mathbf{u}_t$
until model converges

/* Phase 2: Generation (Inference) */
Choose a target class c and angle η
Initialize $\mathbf{x}_0$ as class-specific noise using (μ_c, σ_c)
for $t = 0$ to 1 taking $N = 50$ small steps **do**
  Predict the velocity at the current point using the EMA model
  Predict the velocity again at the midpoint of the step
  Update the image using the midpoint velocity
end for
return Final generated SAR image $\mathbf{x}_1$

This algorithm shows that, by using the OT formulation in the FM methodology, the model is trained to find the shortest path between the per-class Gaussian priors and the actual images.

3.2. SCM-UNet Architecture for Image Generation

SCM-UNet is a UNet architecture proposed to learn the time-dependent vector field required by the FM framework. Figure 3 illustrates the architecture of the proposed SCM-UNet. The input consists of multi-channel SAR image patches from a six-class dataset, each with a spatial size of 32×32. The three input channels correspond to the HH polarization, HV polarization, and the incidence angle. The example shown in the figure belongs to the Thick Ice Ridged class. The model generates a multi-channel SAR image with the same dimensions (3, 32, 32) and aims to learn to produce samples that closely match the distribution of real SAR data. The model follows an encoder-decoder architecture that performs denoising in the latent space. The encoder progressively extracts multi-scale feature representations through three downsampling stages. The first two stages utilize DWSepResBlocks with embedding injection, enabling efficient feature extraction while incorporating time and class conditioning. In the third stage, a shifted window multi-head self-attention (ShiftedWinMHSA) module is used to capture long-range spatial dependencies and improve the modeling of global contextual relationships.

There are several benefits of using SCM-UNet as a backbone with the FM framework. The SCM-UNet predicts vector fields and maps out trajectories. The UNet backbone helps multi-scale feature learning by capturing global structures at lower resolutions, while skip connections help preserve fine textures. This backbone effectively captures both global contextual information and fine-grained, high-frequency details, making it well suited to the straight-line transport trajectories employed by Flow Matching. It is essential for SAR imagery, where both large-scale structures and fine speckle patterns are important. On the other hand, the Depthwise Separable Residual Block (DWSepResBock) improves computational efficiency and is well-suited for modeling speckle-dominated SAR textures, allowing deeper architectures without high computational cost. In addition, when a Swin Transformer attention module integrates with windowed attention, it enables the model to capture long-range spatial dependencies and global context. This ability is critical for maintaining structural consistency in generated SAR images.

Figure 4 represents the three fundamental building blocks of SCM-UNet. The DWSepResBock performs efficient local feature refinement using depthwise separable convolutions within a residual structure. Group normalization (GN), Sigmoid Linear Unit (SiLU) activation, dropout, and depthwise separable convolution are used in sequence, with a residual skip connection to facilitate feature propagation. The time and class embeddings are then combined with the final hidden representation. The ShiftedWinMHSA

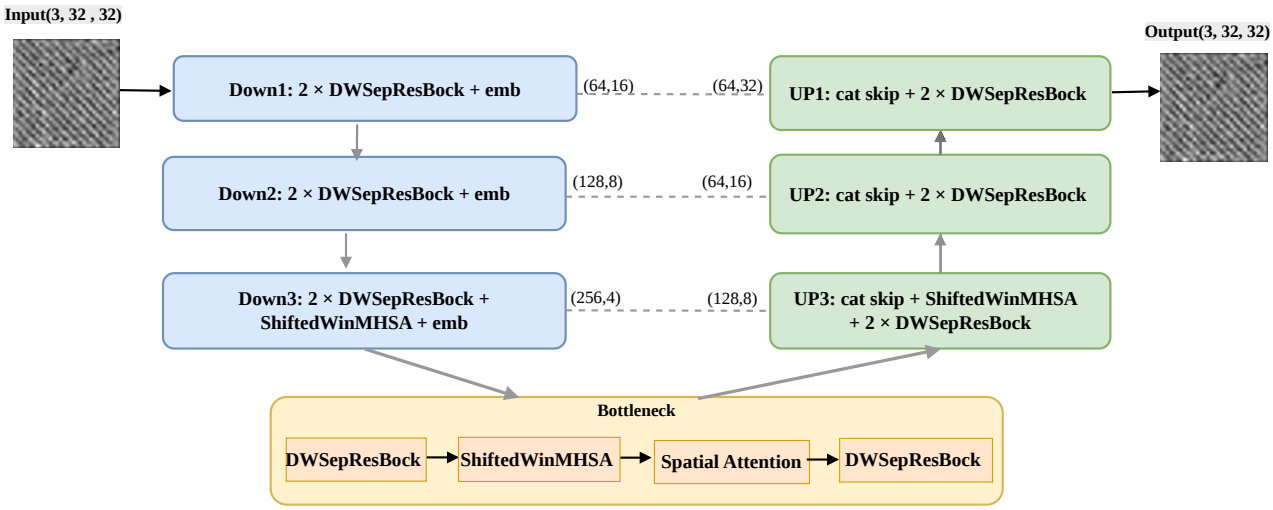


Figure 3. SCM-UNet model architecture.

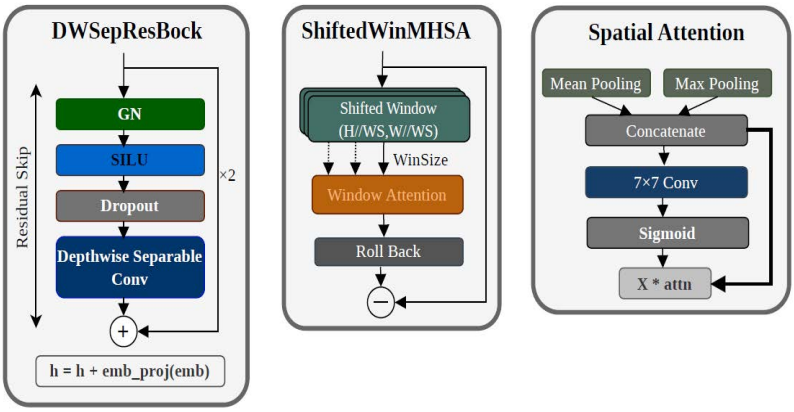


Figure 4. SCM-UNet model components

module is the second module used to capture long-range spatial dependencies by partitioning the feature map into shifted local windows, applying self-attention within each window, and restoring the original spatial arrangement via a rollback operation, all completed via a residual connection. The spatial attention module enhances informative spatial regions by combining mean-pooled and max-pooled feature maps, followed by a 7×7 convolution and a sigmoid activation to generate an attention map, which is then multiplied by the input feature map.

In addition to the above-mentioned reasoning, the ShiftedWinMHSA combined with the spatial attention module enhances feature representation by applying channel and spatial attention, enabling the model to focus on physically meaningful regions and important SAR channels.

The SCM-UNet architecture aligns with FM requirements. It supports conditioning on time and class information, helps the model have sufficient ca pacity to learn smooth, continuous vector fields, improves training stability, accelerates convergence, and yields higher-quality, class-consistent generated samples.

3.3. Classification Model (The SCM-Net)

For classification, SCM-Net model is used, which is a state-of-the-art architecture for SAR image classification [43]. The model is designed to be lightweight and computationally efficient, enabling effective deployment on standard computing platforms without the need for high-performance computing resources. For an overview of the model, the architecture is as follows:

- a convolutional stem,
- depthwise separable convolution blocks,
- shifted window multi-head self-attention,
- a multi-layer perceptron (MLP)-style channel mixer,
- spatial attention,
- and a final classification head.

The SCM-Net model is a lightweight hybrid CNN-transformer architecture for image classification. There is a convolutional stem that projects the input SAR image into a 64-channel feature space while reducing spatial resolution at its beginning. A depthwise separable convolution block performs local feature extraction, improving efficiency and reducing the parameter count. The network incorporates a shifted-window multi-head self-attention module, followed by an MLP-style channel-mixing block to model non-local dependencies, and both of them are equipped with residual connections. The model applies a second depthwise separable convolutional block to further refine the features. At the end, a spatial attention module highlights informative regions. The resulting features are averaged globally and passed

to a fully connected layer for classification. The combination of depthwise and pointwise convolutions with the transformer architecture makes this model fast, lightweight, and effective as an image classifier [43].

4. Experimental Setup and Evaluation Metrics

This section describes the experimental protocol used to compare all models and the evaluation metrics reported in the results.

4.1. Experimental Setup

In all experiments, we implement everything in Python using the PyTorch framework. We implemented the preprocessing pipeline and classical baseline components using the scikit-learn library. The evaluation consists of two main parts: image generation and image classification. During image generation, models were trained for 500 epochs. After training, each model generated 4,000 synthetic images per class, resulting in a balanced augmented dataset. The classification model was trained for 100 epochs.

In the classification part, we use the generated images from the previous section as the training dataset, while reserving real images for testing. For the two majority classes, a stratified random 20% subset was used in order to obtain a more balanced test set, as detailed in Section 5. This formulation lets us evaluate the fidelity of the generated data and, through a downstream classification task, assess how well the synthetic samples capture the true data distribution.

Because the generative model is fitted on the real dataset, the real images used to test the classifier also contributed to training the generator. The classifier is trained from scratch on synthetic samples only and receives no real images as training input, so no image appears in both the classifier's training and test sets. In fact, the protocol measures whether the synthetic data carry sufficient class-discriminative structure to train a usable classifier, rather than providing a fully held-out generalization estimate.

All experiments were conducted on a laptop with an NVIDIA GeForce RTX 3070 GPU and an AMD Ryzen 9 5900HX processor.

4.2. Evaluation Metrics

For evaluation, we used two complementary perspectives: first, metrics that assess image generation quality and demonstrate how well the proposed image generation models perform; and second, metrics that assess the classification performance of generated images. For image quality assessment, we use three standard metrics: PSNR for pixel-level fidelity, SSIM for structural preservation, and FSD for perceptual realism. For classification performance, we used the F1 score, which is the most important metric for imbalanced datasets. In addition, we present the ROC and AUPRC curves, along with the confusion matrix for the classification model.

4.2.1. Image Quality Metrics

Peak Signal-to-Noise Ratio (PSNR):

PSNR measures the pixel-level fidelity between a generated image $\hat{x}$ and a reference image x. It is defined as

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{MAX^2}{\mathrm{MSE}}\right), \tag{10}$$

where MAX denotes the maximum possible pixel value,

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2, \tag{11}$$

where N is the total number of pixels, x_i is the original pixel value, and $\hat{x}_i$ is the corresponding generated or reconstructed pixel value.

Structural Similarity Index Measure (SSIM):

Structural Similarity Index Measure (SSIM) is used in image generation for evaluating the generated images from the aspect of structural similarity between them and the real data by comparing luminance, contrast, and structure, and it is defined as follows:

$$\mathrm{SSIM}(x,\hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)}, \tag{12}$$

where μ, σ^2, and $\sigma_{x\hat{x}}$ denote mean, variance, and covariance, respectively, and $C_1 = (k_1 L)^2$ and $C_2 = (k_2 L)^2$ are stabilizing constants that prevent instability when the denominator approaches zero, with L denoting the dynamic range of the pixel values and $k_1 = 0.01$ and $k_2 = 0.03$ as conventionally used.

Fréchet SAR Distance (FSD):

It is the Fréchet distance [44] for SAR images and it is a metric that estimates the distance between the feature distributions of real and generated SAR images:

$$FSD = \|\mu_r - \mu_g\|_2^2 + \mathrm{Tr}\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{1/2}\right), \tag{13}$$

where (μ_r, Σ_r) and (μ_g, Σ_g) are the mean and covariance of real and generated features, respectively. The lower the score, the more similar the two sets of images (0 means the two sets of images are the same).

KL Divergence (Kullback–Leibler Divergence):

If $P(x)$ = true distribution and $Q(x)$ = approximated/generated distribution, then

$$D_{KL}(P \parallel Q) = \sum_x P(x)\log\frac{P(x)}{Q(x)}. \tag{14}$$

For continuous distributions

$$D_{KL}(P \parallel Q) = \int P(x) \log \frac{P(x)}{Q(x)}\, dx. \tag{15}$$

KL divergence is a measure that calculates how different two probability distributions are. It is a well-understood metric for image generation and measures the amount of information lost when using Q to approximate P [45]. Smaller KL values indicate distributions are more similar, and if KL is 0, the distributions are identical. The larger KL usually indicates that the distributions are very different.

Earth Mover's Distance (EMD)/Wasserstein Distance: The Wasserstein distance measures the minimum cost required to transform one distribution into another. The Wasserstein distance accounts for geometric structure, spatial relationships, and transport cost. So it behaves more smoothly when distributions do not overlap. Unlike the KL divergence, the Wasserstein distance works well even when the two distributions barely overlap [45].

The commonly used form in image generation is given by

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} \mathbb{E}_{(x,y)\sim\gamma} \left[\|x - y\| \right], \tag{16}$$

where P_r is the real data distribution, P_g is the generated data distribution, $\Pi(P_r, P_g)$ denotes the set of all joint distributions (couplings) whose marginals are P_r and P_g, $\gamma \in \Pi(P_r, P_g)$ is one such transport plan, and $\|x - y\|$ is the cost of moving a generated sample y toward a real sample x. The infimum (inf) selects the coupling γ that minimizes the expected transport cost between the two distributions.

4.2.2. Classification Metrics

F1 Score:

The F1 score is the harmonic mean of two of the most important measures in classification problems: precision and recall. It is particularly suitable for imbalanced datasets and is defined as follows, where TP and TN denote true positives and true negatives, respectively, while FN denotes false negatives and FP denotes false positives:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{17}$$

where

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}. \tag{18}$$

Receiver Operating Characteristic (ROC):
The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR):

$$\text{TPR} = \frac{TP}{TP + FN}, \quad \text{FPR} = \frac{FP}{FP + TN}. \tag{19}$$

Area Under Precision-Recall Curve (AUPRC):
AUPRC summarizes the Precision-Recall curve, which is more informative for imbalanced datasets:

$$\text{AUPRC} = \int_0^1 \text{Precision}(r)\, dr, \tag{20}$$

where r denotes recall.

Confusion Matrix:
The confusion matrix is a tabular representation of classification results, summarizing counts of

$$\{TP, TN, FP, FN\}, \tag{21}$$

which provide detailed insight into class-wise prediction performance.

4.3. Hyperparameter Settings

The proposed generative model uses an optimal transport flow matching (OT-FM) objective with the SCM-UNet architecture. The model incorporates depthwise separable residual blocks and shifted window multi-head self-attention (ShiftedWinMHSA). We used a base channel width of 64 during training and optimized the model for 500 epochs. The proposed model applied the Adam optimizer with a learning rate of 1×10^{-4} and a cosine learning-rate scheduler. We applied an exponential moving average (EMA) to the model weights with a decay rate of 0.9999 to stabilize training and enhance sample quality. The input images consisted of 32×32 patches and were processed using a batch size of 64. The minimum noise level of the flow-matching trajectory was set to 1×10^{-4}. During inference, we used a midpoint ordinary differential equation (ODE) solver with 50 integration steps to generate SAR imagery for the six sea-ice classes.

In the image-generation pipeline, we separate the dataset using a stratified 90/10 split. The generator was trained using 90% of the samples from each class to learn the class-conditional image distribution, while the remaining 10% was held out and used only to monitor the flow-matching loss during training. The held-out samples contributed neither to gradient updates nor to estimating the class statistics μ_c and σ_c.

The model treats each data sample as a multi-channel structure consisting of three channels. The first two channels correspond to the HH and HV SAR polarizations, which the model learns to generate. Before feeding them into the model, these channels were normalized to $[0, 1]$ and then scaled to $[-1, 1]$ to facilitate training with the flow-matching objective. The third channel represents the incidence angle (IA) map and serves as a spatial conditioning input.

For the downstream classification task, the classification model was trained to classify 32×32 pixel input patches into six distinct sea-ice classes. We used the

generated data only to train the classification model, with a fixed random seed to ensure reproducibility, while reserving the real dataset exclusively for testing. We trained the classification model with a batch size of 32 for 100 epochs. The SCM-Net classification model uses the AdamW optimizer with a learning rate of 1×10^{-3} and a weight decay of 1×10^{-1} for regularization. We used the standard cross-entropy loss function for training.

In the hybrid architecture, we set the initial channel dimension to 64. The ShiftedWinMHSA blocks used four attention heads, a window size of 4, and a window shift of 2 to capture both local and global spatial dependencies.

5. Results

In this section, the quality of the proposed image generation model is first investigated and compared with that of other state-of-the-art image generation models. Then, the ability of a standard image classification model trained solely on these generated images was tested to assess its performance in practice. In this paper, a class-aware FM methodology with a modified UNet model for generating high-fidelity images is applied to the selected SAR sea ice dataset, and the performance and the quality of the generated images were compared with several state-of-the-art generative approaches, including standard diffusion models and the baseline FM framework. A comprehensive comparative analysis of these models was conducted, with particular emphasis on the class-aware FM generation framework and the proposed SCM-UNet architecture.

Alongside the above-mentioned innovation, we introduce the SCM-UNet model and evaluate its image generation performance within a class-conditioned FM framework. Next, we investigate and analyze the benefits and drawbacks of the proposed framework in detail. Furthermore, the effectiveness of the generative models for multi-class SAR image synthesis is demonstrated through quantitative evaluation using structural and perceptual quality metrics, including SSIM, PSNR, and FSD. In addition, for the classification model, several evaluation metrics were used to assess the utility of the generated images. Specifically, an SCM-Net classifier, which was proposed by Baramaki et al. [43] and previously applied to the chosen dataset, was trained on the generated dataset and evaluated using real SAR data from the chosen dataset as the test set. The results were then compared with those obtained from state-of-the-art approaches trained and tested entirely on real data. The comparative analysis evaluates the capability of the generated samples to support downstream classification tasks, using performance metrics such as accuracy, F1-score, confusion matrices, ROC and AUPRC curves.

5.1. Overall Image Generation Model Performance

Generative Model Result Comparison

In Table 1, a comparison was conducted between the selected generative models and the proposed model. Selected generative models are as below: The *Base UNet with Base FM* corresponds to the original FM model proposed by [21]. It uses the exact FM framework with OT and the ODE solver. Moreover, the UNet model is a standard UNet model. After that, we used the **SCM-UNet with Base FM**, which is the basic FM framework with the proposed SCM-UNet backbone instead of the basic UNet. In the next model, we try to replace the base FM with the class-wise method and use it with the base UNet to see how this change will improve the quality of generative images. As shown in Table 1, the class-wise FM with SCM-UNet method achieves the best results by combining a class-wise flow matching framework with the proposed generative image backbone. In addition to these models, we try to check other changes, like changing the starting point distribution with the Beta distribution instead of the Gaussian distribution in **Beta-based FM with the SCM-UNet model** or channel-wise FM that gets the mean of two HH and HV channels, separates incidence angle, tries to start a Gaussian distribution with learning these two parameters of each image, and adds the incidence angle in the last step. The results show that the class-aware model outperforms all the other proposed models on pixel-level quality metrics and the downstream classification task.

The other comparison was between existing generative models, such as GANs, DDPMs, and the diffusion score matching with the SCM-UNet model, which are widely used techniques in image generation. The results showed that although these generative models perform well, the proposed model benchmarks against these state-of-the-art deep learning models in the aforementioned metrics. It is important to note that all reported results are based on generating 4000 images.

The FSD in Table 1 is the Fréchet SAR distance, which is used to compare the generative model's ability to generate SAR images. In conventional RGB image generation tasks, the Fréchet distance metric is commonly referred to as the Fréchet Inception Distance(FID) because it uses the Inception-v3 network for feature extraction [44]. When we pass the real and generated datasets through the model (e.g., Inception-v3) and strip off the final classification layer, we are left with a dense vector, called an embedding. This embedding serves as a compressed, highly informative summary of an image's style, geometry, and texture. It is typically used to compare generative images and assess how well models generate them. Inception-v3 requires fixed 299×299 inputs, making it unsuitable for small SAR patches, as extreme upscaling introduces interpolation artifacts and destroys important radar

Table 1. Image augmentation result comparison

Method	FSD ↓	EMD ↓	KL ↓	PSNR ↑	SSIM ↑
Base UNet with Base FM	76.9	84.316	5.52	8.032	0.69
SCM-UNet with Base FM	91.84	72.31	5.527	12.09	0.70
SCM-UNet With FM JointConditional	86.2	61.65	4.19	8.80	0.65
Base UNet with Class-Wise FM	58.1	5.76	8.04	15.03	0.76
Basic Generative Adversarial Network (GAN)	89.99	3.615	0.95	18.823	0.76
Denoising Diffusion Probabilistic Model (DDPM)	70.69	0.98	0.71	21.819	0.79
Diffusion Score Matching with SCM-UNet model	38.6	20.7	0.95	19.311	0.65
Beta-based FM with SCM-UNet model	22.05	11.46	0.86	20.16	0.78
Channel-Mean based FM with SCM-UNet model	68.65	20.55	0.60	16.34	0.68
CFM-SAR model (proposed)	**19.23**	**0.63**	**0.41**	**28.0**	**0.83**

textures. In addition, Inception-v3, ResNet-50, and VGG-16 are pretrained on ImageNet, which provides flexibility in handling varying image sizes. However, their pretraining on RGB optical images may introduce a domain bias, as these models are optimized for optical image characteristics rather than SAR-specific properties, such as speckle patterns, polarization information, and geometric scattering effects [46]. To calculate the Fréchet distance for SAR images, a ResNet-50 classifier was trained from scratch on the real SAR dataset. Then the classifier's weights are frozen, and the model is used as the feature extractor for Fréchet distance computation. Since the network learns radar-specific representations directly from the SAR real image dataset, the resulting feature embeddings provide a more reliable evaluation of generated SAR images than those from ImageNet-pretrained models.

Convergence Plots

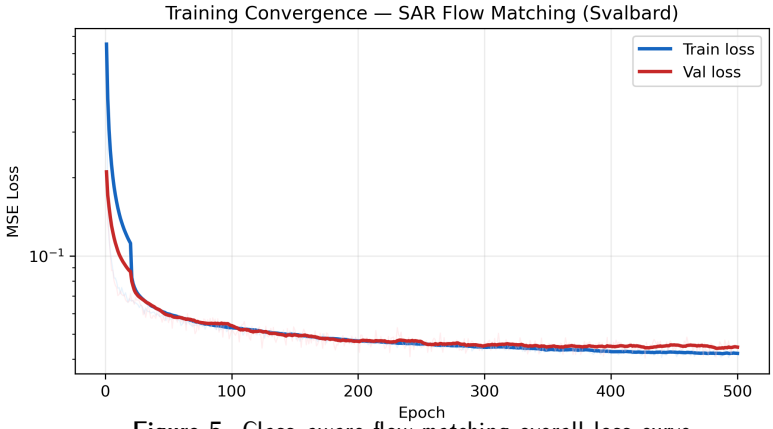
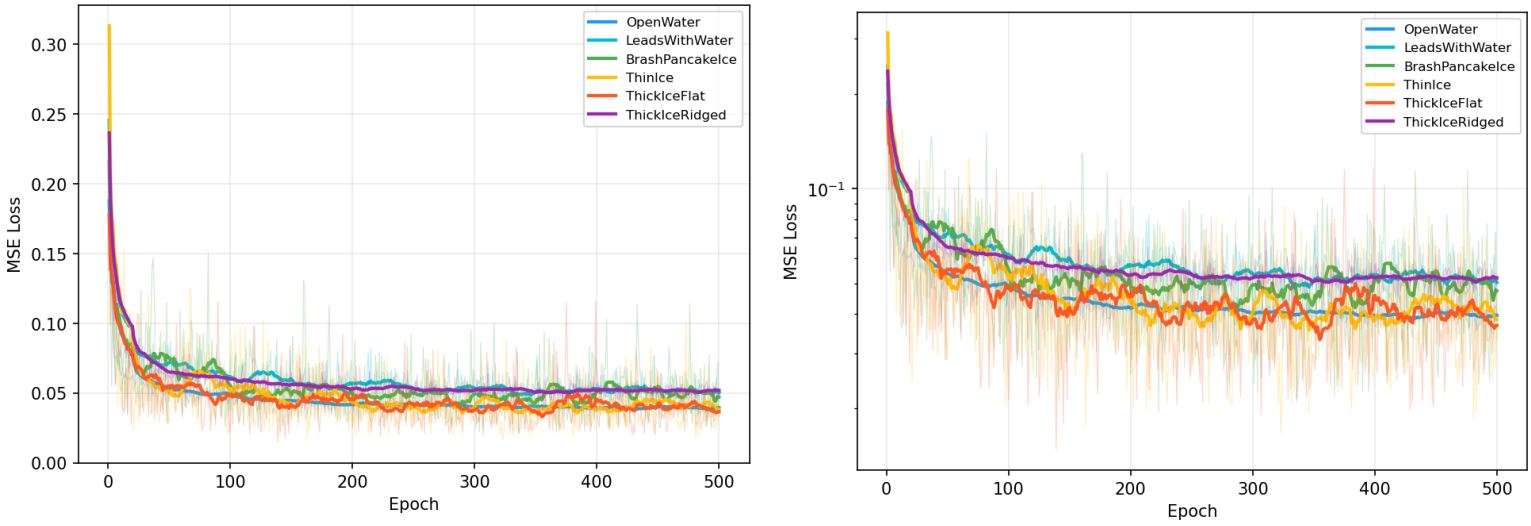


Figure 5. Class-aware flow matching overall loss curve

The validation loss overlay shows that the class-mean FM approach achieves a lower global minimum and superior convergence stability than traditional diffusion and standard UNet architectures. Figure 5 shows the training and validation loss of the proposed class-mean FM model. It not only achieves a lower error rate, but also does it so faster and remains smoother than any other test. Another important point is that the sharp drop in the first 20 epochs indicates that the model is quickly identifying the primary features of the sea-ice data. The validation loss (red line) closely follows the train loss (blue line). This indicates that the model does not exhibit significant overfitting and demonstrates good generalization to unseen data. It is a learning pattern that works on new, unseen SAR images. It can be observed that after epoch 60, the curves become nearly flat, indicating that the model had largely converged by this point. Nevertheless, training was continued for 500 epochs to ensure stable convergence and consistent model performance.

Figure 6 shows model convergence per class for each class in the SAR sea-ice image dataset. This image shows a validation loss overlay that tracks how well the proposed model learns to generate sea-ice images in each class across 500 training epochs. By breaking down the mean squared error (MSE) validation loss by class, we can find how the model learns each class of images. The model shows convergence across all classes. There is a sharp initial drop in loss during the first 50 epochs, followed by a long "tail" where the model fine-tunes. Figure 6b shows a log-scale plot, which is particularly useful here, as it reveals that while the linear plots (part a of the figure) look flat, the model is still making marginal gains or experiencing slight oscillations well into epoch 500. In class-specific performance analyses, we can see two distinct performance tiers: the first group is the high performers (lower loss), including ThickIceFlat (orange) and ThinIce (yellow), which consistently maintain the lowest MSE. Figure 6 suggests these classes have more uniform textures or higher internal consistency in the SAR imagery, making them easier for the model to reconstruct or generate accurately. In addition to these classes, the open water (blue line) also performs well, likely due to the relatively low backscatter variance in calm water in



(a) Validation loss (linear scale)

(b) Validation loss (logarithmic scale)

Figure 6. Per-class validation loss for all sea ice classes

SAR. The second group are the challenging classes (that have higher loss). These are LeadsWithWater (cyan), BrashPancakeIce (green) and ThickIceRidged (purple). LeadsWithWater remains at a higher loss plateau. "Leads" are linear features that often contain a mix of water and ice edges. This structural complexity likely increases the MSE, as the model struggles to place edges precisely. BrashPancakeIce and ThickIceRidged show higher variance and higher final loss. The "ridging" and "brash" ice are characterized by high backscatter, "noisy" textures, and chaotic physical structures.

Classification Result Using Generated Data

Table 2 demonstrates the result of the SCM-Net model in downstream classification task. For the classification process, the training and test data are used as follows: all generated data are used for training purposes (4000 per class), and the real data (the chosen SAR dataset) is used as the test data. The results are shown in the aforementioned graphs and tables. It should be noted that the final Test Accuracy is 91.61%.

Table 2. Class-mean based FM results

Class	Precision	Recall	F1-score
0	0.98	0.88	0.92
1	0.95	0.76	0.84
2	0.94	0.81	0.87
3	0.95	0.94	0.95
4	0.97	0.94	0.95
5	0.96	0.96	0.96

To ensure a fair comparison, the SCM-Net classifier was trained using the same hyperparameters under an identical experimental setup. Specifically, each image generation model generated 4,000 images, and the classifier was trained on this data and evaluated on the same real data used as the test dataset. The classification results obtained using SCM-Net under all experimental settings are presented in Table 3.

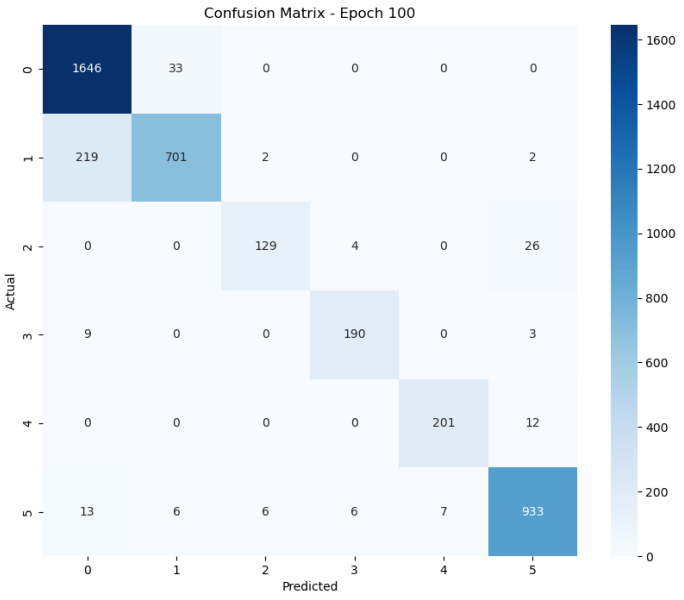


Figure 7. Confusion matrix of the SCM-Net classifier

Another important result to examine is the confusion matrix, which provides a detailed view of the model's classification p erformance b y d istinguishing between true positive, true negative, false positive, and false negative predictions for each class. As shown in Figure 7, the classification m odel (SCM-Net) can effectively c lassify i mages f rom t he r eal d ataset with only training on the generated dataset. The confusion matrix demonstrates high classification performance, as most predictions are correctly aligned along the diagonal. This result demonstrates that the quality of the generated model is high enough to serve as training data for a classification task.

Table 3. Comprehensive per-class F1-score comparison across all models

Methodology	Generative Model	Cl. 0	Cl. 1	Cl. 2	Cl. 3	Cl. 4	Cl. 5
Raw Data	No Generation	0.90	0.83	0.58	0.79	0.93	0.89
Base FM	Base UNet	0.77	0.11	0.56	0.75	0.61	0.89
Base FM	SCM-UNet	0.91	0.29	0.55	0.69	0.73	0.91
GAN Model	-	0.65	0.05	0.35	0.56	0.65	0.70
Diffusion Model (DDPM)	Base U-Net	0.78	0.09	0.55	0.83	0.94	0.90
SCM-UNet With JointConditional FM	0.78	0.34	0.22	0.31	0.26	0.76	
Diffusion Score Matching	Base UNet	0.77	0.31	0.21	0.12	0.15	0.84
DDPM-Advanced Model	Base UNet	0.75	0.50	0.30	0.43	0.11	0.80
Channel-Mean Gaussian FM	SCM-UNet	0.81	0.45	0.38	0.42	0.59	0.79
Beta-based Distribution FM	SCM-UNet	0.90	0.83	0.59	0.83	0.93	0.87
Class-Mean FM Gaussian	SCM-UNet	**0.92**	**0.84**	**0.87**	**0.95**	**0.95**	**0.96**

Class Key: 0: Open Water, 1: Leads with Water, 2: Brash/Pancake Ice, 3: Thin Ice, 4: Thick Ice Flat, 5: Thick Ice Ridged.

It is important to note that there is a significant imbalance in the real SAR dataset we used as a test set. As mentioned in the database section (Section 2), the number of class 0 (Open water) images is 8,394, and for class 5 (thick ice), the number of images is 4,855. To fix this imbalance issue in the test dataset, and just for these two classes, we test on only 20 percent of the real data and aim to create a more balanced test set. Then the test set amounts are precisely as follows: the "Open water" and "Thick ice" classes are 1,678 and 971 samples, respectively, and other classes, exactly as described in section 2.

5.2. Discriminative Ability via ROC and AUPRC Curves

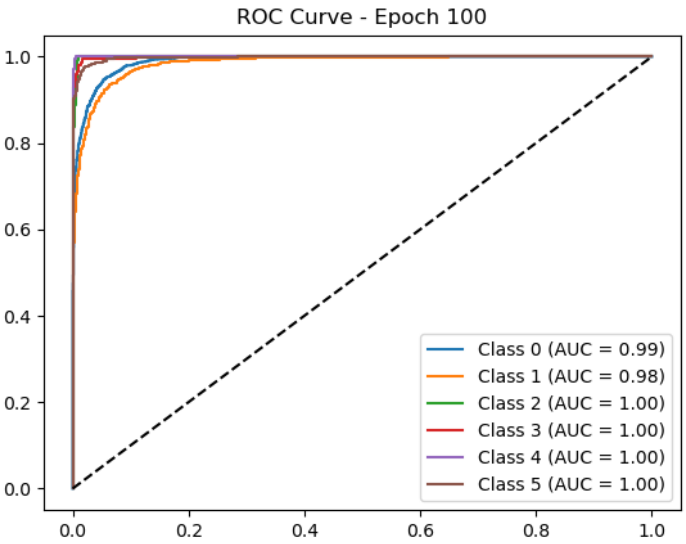


Figure 8. ROC curve of the SCM-Net classifier

Figure 8 shows the Receiver Operating Characteristic (ROC) curves of the SCM-Net classifier model for all six classes, and the area under the curve (AUC) values across all classes are consistently high, indicating strong discriminative capability. The ROC curves of all classes are generally concentrated in the top-left corner, demonstrating strong discriminative performance of the classifier and characterized by high True Positive Rate (TPR) and low False Positive Rate (FPR). Although Class 1 exhibits a comparatively lower AUC (0.98), its performance remains strong, while Classes 2, 3, 4, and 5 achieve very high discrimination. Overall, these results demonstrate the robustness and reliability of the trained classification model in multi-class classification with only training on the generated images, indicating the high quality of the generated images.

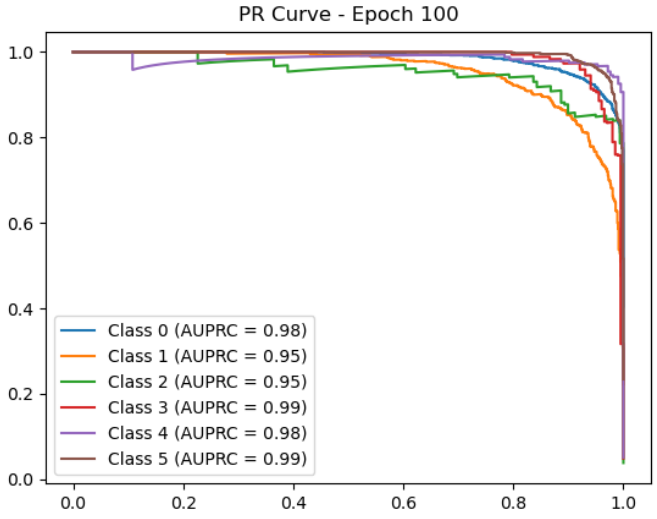


Figure 9. Precision-recall curve of the classification model

We used the area under the precision-recall curve (AUPRC), which is defined here as the macro-averaged precision-recall ranking metric, and it was computed across all six classes using a one-vs-rest strategy. Figure 9 shows the precision–recall (PR) curve for a multi-class SAR image classification model. The figure

shows that all curves stay very close to the top-right corner, indicating high precision and recall across most thresholds. Classes 3 and 5 perform the best (AUPRC=0.99), while Classes 1 and 2 are slightly weaker but still strong (AUPRC=0.95). There is a sharp drop near recall = 1.0 for most classes, which is typical when trying to capture all positives.

6. Ablation Studies

We conducted a systematic ablation study to evaluate the individual contributions of the core components of our proposed framework. The primary parts of the proposed framework are the SCM-UNet generative architecture and the Class-aware prior FM, which we separately compare against standard baselines (Standard UNet and Standard FM model) across image generation fidelity (Table 1) and downstream per-class F1-scores (Table 3).

- Impact of the Generative Backbone: Standard UNet vs. SCM-UNet

  As a first step, it is necessary to evaluate the impact of upgrading the generative backbone from a standard UNet to our proposed SCM-UNet while keeping the underlying FM methodology constant.

  - Fidelity dynamics of the images (Table 1): When the FM methodology is held constant, replacing the Base UNet with the SCM-UNet improves pixel-level reconstruction quality: the PSNR increases from 8.032 to 12.09, the SSIM from 0.69 to 0.70, and the EMD decreases from 84.316 to 72.31. The improvement is therefore concentrated in reconstruction fidelity rather than in distributional alignment; the FSD in fact rises from 76.9 to 91.84. This indicates that the SCM-UNet backbone provides greater structural capacity for modelling complex image distributions, but that this capacity alone is not sufficient to align the generated distribution with the real one while the prior remains uninformed. The peak values reported later (PSNR of 28.0 and SSIM of 0.83) are obtained only when this backbone is combined with the class-conditional prior.
  - In the downstream classification task (Table 3): Under the base FM setups, the standard UNet model struggles with minority classes, such as class 1 (Leads with Water), which has a significantly low F1-score of 0.11. Integrating base FM with the SCM-UNet elevates the F1-score related to class 1 to 0.29 and class 0 to 0.91. However, because the prior remains unconditioned, it experiences distribution collapses on some other categories (e.g., class 3 drops from 0.75 to 0.69). The incremental results show that although the structural architectural enhancements alone are beneficial, adding informed prior conditioning is even more beneficial.

- Impact of the Class-Wise Prior: Standard Gaussian vs. Class-wise FM

  In the next step, we isolate the impact of our proposed class-conditioned prior by comparing it to the standard Gaussian (Base FM) initialization.

  - Image fidelity dynamics (Table 1): Holding the Base UNet backbone constant, switching from the standard Base FM to the class-wise FM results in a significant improvement in distributional alignment. The SSIM increases from 0.69 to 0.76, the FSD drops from 76.9 to 58.1, and the EMD decreases sharply from 84.316 to 5.76. This shows that anchoring the trajectory endpoints to class-specific means drastically simplifies the transport mapping task for the model. The KL divergence rises from 5.52 to 8.04 in this pairing, which suggests that the class prior alone sharpens each class-conditional mode before the backbone has sufficient capacity to cover the full intra-class variability; this effect disappears once the two components are combined, where the KL divergence reaches its lowest value of 0.41.
  - Downstream classification (Table 3): For this comparison, we use paired comparisons with our SCM-UNet models, which demonstrate how the class-specific prior is manifested in the generated samples. The transition from Base FM to class-mean FM with the SCM-UNet produces a substantial increase in F1-scores across all sea-ice categories. Most notably, the highly challenging Cl. 1 (Leads with Water) jumps from 0.29 to 0.84. Similarly, Cl. 2 (Brash/Pancake Ice) category increases from 0.55 to an impressive 0.87. This confirms that the class-mean prior successfully preserves critical semantic boundaries necessary to generate high-utility augmentations for highly imbalanced classes.

- Structural Synergy: The Full Proposed Configuration

  The complete integration of our proposed framework, FM with class-aware prior, which uses the

SCM-UNet backbone, generates a powerful synergistic effect that outperforms all alternative pairings and standalone generative models like GANs and standard DDPMs.

- Peak generation metrics: The results in Table 1 demonstrate that the proposed image generation model achieves the best scores across all metrics, with an FSD of 19.23, EMD of 0.63, KL divergence of 0.41, PSNR of 28.0, and SSIM of 0.83. Neither component reaches these values on its own: the SCM-UNet backbone under the base FM attains a PSNR of 12.09 and an SSIM of 0.70, while the class-wise prior with the Base UNet attains an FSD of 58.1 and an SSIM of 0.76. The proposed configuration also clearly outperforms both the standard DDPM (PSNR: 21.819, SSIM: 0.79) and the basic GAN (PSNR: 18.823, SSIM: 0.76). The results show that pairing a class-mean trajectory with an SCM-UNet backbone yields the cleanest probability paths.
- Downstream data enhancement: As shown in Table 3, using our full model's generated images for data augmentation significantly outperforms the raw data baseline. It achieves an F1-score of 0.92 for Cl. 0, 0.84 for Cl. 1, 0.87 for Cl. 2, and 0.95 or above for Cl. 3, 4, and 5. When trained on the raw data, the model achieves an overall F1-score of 0.90, whereas training exclusively on images generated by the proposed class-informed Flow Matching model with the SCM-UNet backbone yields an overall F1-score of 0.915.
- Other proposed and testing models: During testing the model's ability, we proposed and tested other models as well. The Beta-based distribution FM is the model that uses the Beta distribution as the starting distribution rather than the Gaussian distribution. The channel-mean Gaussian FM, as its name shows, is a model that starts with a channel mean Gaussian prior, as the used SAR images have 2 channels, HH and HV, and the incidence angle gets fixed and added to the model in the last stage. The other model evaluated was the DDPM-Advanced model, in which the incidence-angle conditioning was modified to improve the model's performance. This model introduces dual, domain-specific conditioning, directly injecting semantic class labels and concatenating a physical, spatial incidence angle map to the input. Another model that we tested was the joint-conditional FM, which tries to transform random noise into realistic data by satisfying multiple conditions simultaneously. Across all the proposed models, the best result was achieved with the class-aware prior FM and the SCM-UNet backbone, as evaluated on image generation and downstream classification tasks.

The standard FM and UNet generative backbone works as a bottlenecked baseline. The SCM-UNet restores structural fidelity (SSIM) while introducing a class-mean prior to align the semantic distribution (EMD and per-class F1). Together, Class-wise FM and the SCM-UNet form a cohesive framework optimized for high-fidelity, class-balanced image synthesis. One limitation to note when interpreting these results is as follows. Under the train-on-synthetic, test-on-real protocol used here, the generator was fitted on the same real dataset from which the classifier's test samples are drawn. The reported accuracy therefore reflects the class-discriminative content of the synthetic data rather than a fully held-out generalization estimate. Evaluating the pipeline with a generator trained on a disjoint partition is a natural next step, particularly for the minority classes, where the number of real samples is small enough that the generator may reproduce individual training patches.

7. Conclusion

In conclusion, to achieve high accuracy, we need a large amount of training data, and the need for a balanced dataset is another barrier in this field. In this study, we proposed a flow-matching-based method with class-wise priors, using an SCM-UNet architecture, for the SAR image generation problem. We tried to generate high-quality images for training a classifier and achieve the goal of a precise multi-class SAR image classifier. Several state-of-the-art approaches, including diffusion score matching and other flow matching frameworks, were implemented and evaluated, along with multiple proposed modifications tailored to the characteristics of SAR imagery. The proposed model shows superior quantitative results in FSD, SSIM, and PSNR. In particular, the class-mean-based flow-matching approach achieved the best overall balance among image fidelity, structural similarity, and reconstruction quality. These results indicate that incorporating class-aware prior distribution can significantly improve generative performance for the SAR image dataset. The generated SAR images were used as training data in the classification problem, and a test was performed on the real dataset. The result shows improved robustness and recognition accuracy in the six-class target classification task. These results also confirm the high quality of the generated SAR images and their effectiveness as a supplementary

source of training data. Future work may focus on advances in generative models for image processing, more advanced conditional guidance mechanisms, or real-time augmentation pipelines for operational SAR systems.

References

[1] Rodríguez JP, Klemm K, Duarte CM, Eguíluz VM. Shipping traffic through the Arctic Ocean: Spatial distribution, seasonal variation, and its dependence on the sea ice extent. iScience. 2024 July;27:110236.

[2] Zakhvatkina N, Smirnov V, Bychkova I. Satellite SAR Data-Based Sea Ice Classification: An Overview. Geosciences. 2019;9(4):152.

[3] Choi H, Jeong J. Speckle Noise Reduction Technique for SAR Images Using Statistical Characteristics of Speckle Noise and Discrete Wavelet Transform. Remote Sensing. 2019;11(10):1184.

[4] Soumekh M. Synthetic Aperture Radar Signal Processing with MATLAB Algorithms. New York, USA: Wiley-Interscience; 1999.

[5] Peter E, Ang LM, Seng KP, Srivastava S. Recent Advances in Deep Learning for SAR Images: Overview of Methods, Challenges, and Future Directions. Sensors. 2026;26(4):1143. Available from: https://doi.org/10.3390/s26041143.

[6] Boulze H, Korosov A, Brajard J. Classification of Sea Ice Types in Sentinel-1 SAR Data Using Convolutional Neural Networks. Remote Sensing. 2020;12(13):2165. Available from: https://doi.org/10.3390/rs12132165.

[7] Maillard P, Clausi DA, Deng H. Operational Map-Guided Classification of SAR Sea Ice Imagery. IEEE Transactions on Geoscience and Remote Sensing. 2005 Dec;43(12):2940-51.

[8] Radhakrishnan K, Scott KA, Clausi DA. Sea Ice Concentration Estimation: Using Passive Microwave and SAR Data With a U-Net and Curriculum Learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2021;14:5339-51.

[9] Ricker R, Hendricks S, Kaleschke L, Tian-Kunze X, King J, Haas C. A Weekly Arctic Sea-Ice Thickness Data Record from Merged CryoSat-2 and SMOS Satellite Data. The Cryosphere. 2017;11(4):1607-23.

[10] Komarov AS, Barber DG. Sea Ice Motion Tracking From Sequential Dual-Polarization RADARSAT-2 Images. IEEE Transactions on Geoscience and Remote Sensing. 2014;52(7):4431-41.

[11] Ressel R, Frost A, Lehner S. A Neural Network-Based Classification for Sea Ice and Open Water Detection Using Sentinel-1 SAR Data. Remote Sensing. 2016;8(12):1006.

[12] Power D, Youden J, Lane K, Randell C, Flett D. Iceberg Detection Capabilities of RADARSAT Synthetic Aperture Radar. Canadian Journal of Remote Sensing. 2001;27(5):476-86.

[13] Gascard JC, Riemann-Campe K, Gerdes R, Schyberg H, Randriamampianina R, Karcher M, et al. Future sea ice conditions and weather forecasts in the Arctic: Implications for Arctic shipping. Ambio. 2017;46(Suppl 3):355-67.

[14] Løset S, Shkhinek K, Gudmestad OT, Strass P, Michalenko E, Frederking R, et al. Comparison of the Physical Environment of Some Arctic Seas. Cold Regions Science and Technology. 1999;29(3):201-14.

[15] Kwok R. Arctic Sea Ice Thickness, Volume, and Multiyear Ice Coverage: Losses and Coupled Variability (1958–2018). Environmental Research Letters. 2018;13(10):105005.

[16] Aase J. The Ortelius Incident in the Hinlopen Strait—A Case Study on How Satellite-Based AIS Can Support Search and Rescue Operations in Remote Waters. Resources. 2017;6(3):35.

[17] Passah A, Sur SN, Paul B, Kandar D. SAR Image Classification: A Comprehensive Study and Analysis. IEEE Access. 2022;10:20385-99.

[18] Khaleghian S, Ullah H, Kræmer T, Hughes N, Eltoft T, Marinoni A. Sea Ice Classification of SAR Imagery Based on Convolution Neural Networks. Remote Sensing. 2021 Apr;13(9):1734.

[19] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Advances in Neural Information Processing Systems; 2014. p. 2672-80.

[20] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems. vol. 33; 2020. p. 6840-51.

[21] Lipman Y, Chen RTQ, Ben-Hamu H, Nickel M, Le M. Flow matching for generative modeling. In: International Conference on Learning Representations (ICLR); 2023. .

[22] Esser P, Rombach R, Ommer B. Taming Transformers for High-Resolution Image Synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021. p. 12873-83.

[23] Wang Z, Zhang Z, Shan X, Wei H, Tang P. Generative models for SAR–optical image translation: A systematic review. International Journal of Applied Earth Observation and Geoinformation. 2026 February;146:105009.

[24] Passah A, Sur SN, Abraham A, Kandar D. Synthetic aperture radar image analysis based on deep learning: A review of a decade of research. Engineering Applications of Artificial Intelligence. 2023;123:106305.

[25] Stofa MM, Abdani SR, Moubark AM, Zainuri MAAM, Ibrahim AA, Kamari NAM, et al. Recent developments of artificial intelligence methods for sea ice concentration monitoring using high-resolution imaging datasets. Ecological Informatics. 2025;87:103132.

[26] Wan H, Luo X, Wu Z, Qin X, Chen X, Li B, et al. Multi-featured sea ice classification with SAR image based on convolutional neural network. Remote Sensing. 2023;15(16):4014.

[27] Stokholm A, Wulf T, Kucik A, Saldo R, Buus-Hinkler J, Hvidegaard SM. AI4SeaIce: Toward Solving Ambiguous SAR Textures in Convolutional Neural Networks for Automatic Sea Ice Concentration Charting. IEEE Transactions on Geoscience and Remote Sensing. 2022;60:1-13.

[28] Song Q, Xu F, Jin YQ. SAR Image Representation Learning With Adversarial Autoencoder Networks. In:

2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Yokohama, Japan: IEEE; 2019. p. 9498-501.
[29] Liu K, Qiu G, Tang W, Zhou F. Spectral regularization for combating mode collapse in GANs. Image and Vision Computing. 2020;104:104019.
[30] Oghim S, Kim Y, Bang H, Lim D, Ko J. SAR image generation method using DH-GAN for automatic target recognition. Sensors. 2024;24(2):670.
[31] Chen Y, Zhu Z, Huang Y, Wang P, Huang B, Mura MD. MSF: A Multi-Scale Fusion Generative Adversarial Network for SAR-to-Optical Image Translation. In: IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium; 2024. p. 9058-61.
[32] Hwang J, Yu C, Shin Y. SAR-to-Optical Image Translation Using SSIM and Perceptual Loss Based Cycle-Consistent GAN. In: 2020 International Conference on Information and Communication Technology Convergence (ICTC); 2020. p. 191-4.
[33] Qosja D, Wagner S, O'Hagan D. SAR image synthesis with diffusion models. In: Proc. IEEE Radar Conference (RadarConf); 2024. p. 1-6.
[34] Wang J, Yang H, Liu Z, Chen H. SSDDPM: A single SAR image generation method based on denoising diffusion probabilistic model. Scientific Reports. 2025;15(10867).
[35] Tong A, Fatras K, Malkin N, Huguet G, Zhang Y, Rector-Brooks J, et al. Improving and generalizing flow-based generative models with minibatch optimal transport. Transactions on Machine Learning Research. 2024.
[36] Ben-Hamu H, Puny O, Gat I, Karrer B, Singer U, Lipman Y. D-Flow: Differentiating through Flows for Controlled Generation. In: Proceedings of the 41st International Conference on Machine Learning. vol. 235. PMLR; 2024. p. 3462-83.
[37] Kong L, Tao M, Liu Y, Wang B, Fu J, Wang CC, et al. AlignFlow: Improving Flow-Based Generative Models with Semi-Discrete Optimal Transport. In: Proceedings of the 14th International Conference on Learning Representations (ICLR). Rio de Janeiro, Brazil; 2026. .
[38] Jin Y, Sun Z, Li N, Xu K, Xu K, Jiang H, et al. Pyramidal Flow Matching for Efficient Video Generative Modeling. In: Proceedings of the 13th International Conference on Learning Representations (ICLR). Singapore; 2025. .
[39] Peebles W, Xie S. Scalable Diffusion Models with Transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France; 2023. p. 4195-205.
[40] Beizaee F, Lodygensky GA, Adamson CL, Thompson DK, Cheong JLY, Spittle AJ, et al. Harmonizing flows: Leveraging normalizing flows for unsupervised and source-free MRI harmonization. Medical Image Analysis. 2025;101:103483.
[41] Kobyzev I, Prince SJD, Brubaker MA. Normalizing Flows: An Introduction and Review of Current Methods. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2021;43(11):3964-79.
[42] Feng R, Yu C, Deng W, Hu P, Wu T. On the Guidance of Flow Matching. In: Proceedings of the 42nd International Conference on Machine Learning (ICML). Vancouver, BC, Canada; 2025. .
[43] Baramaki N, Le QN, McNiven BDE, Fahim M. SCM-Net: A Lightweight AI-Based Sea Ice Classification for Climate Change. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 2026 Jun;13(2). Available from: https://doi.org/10.4108/eetinis.132.12246.
[44] Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 30; 2017. .
[45] Betzalel E, Penso C, Fetaya E. Evaluation Metrics for Generative Models: An Empirical Study. Machine Learning and Knowledge Extraction. 2024;6(3):1531-44.
[46] Kynkäänniemi T, Karras T, Aittala M, Aila T, Lehtinen J. The Role of ImageNet Classes in Fréchet Inception Distance. In: International Conference on Learning Representations (ICLR); 2023. .