HiFC-GAN: Hierarchical Feature-Constrained GAN for Optical-to-SAR Transfer in SAR Target Classification

Hao Zheng[1,2], Meiguang Zheng[1], Zhigang Hu[1], Liu Yang[1], Aikun Xu[1*], Tingxuan Chen[1], Rongchang Zhao[1], Boyu Wang[2]

[1]School of Computer Science and Engineering, Central South University, Changsha, China
[2]Department of Computer Science, University of Western Ontario, London, Canada
{zhenghao, zhengmeiguang, zghu, yangliu, aikunxu, chentingxuan, zhaorc}@csu.edu.cn; bwang@csd.uwo.ca

Abstract

The limited availability of high-quality training data poses a persistent challenge for synthetic aperture radar (SAR) target classification. Existing data augmentation methods mainly adopt a simplistic application of GAN-based style transfer techniques to directly synthesize pseudo-SAR images from optical images. However, our in-depth analysis of this cross-modal conversion reveals that such straightforward strategies primarily focus on transferring high-level semantic information (e.g., target shapes), thus failing to adequately capture the essential low-level features unique to SAR imagery (e.g., scattering textures). To address this inherent trade-off between high-level semantic preservation and low-level feature authenticity, we propose a Hierarchical Feature-Constrained GAN (HiFC-GAN) tailored for optical-to-SAR style transfer. Specifically, HiFC-GAN enhances the representation of low-level SAR features by introducing local texture contrast constraints at shallow layers, while introducing explicit feature mapping constraints at deeper layers to maintain high-level semantic consistency throughout the reconstruction process. Experimental results demonstrate that HiFC-GAN significantly outperforms existing GAN-based techniques in image generation quality, particularly improving the low-level feature authenticity of pseudo-SAR images. Moreover, the generated pseudo-SAR images further improve the performance of downstream target classification tasks, yielding accuracy gains ranging from 3.56% to 5.90% on average with mainstream CNN-based models.

Introduction

With the rapid advancement of remote sensing technology, SAR has been widely utilized in military, environmental monitoring, and disaster management fields due to its all-weather, all-time imaging capabilities (Zheng et al. 2023a, 2022; Feng et al. 2024). Unlike optical imagery, which relies on visible light reflection to capture rich texture details and illumination variations (Cheng et al. 2025; Li et al. 2025), SAR imaging is based on radar wave scattering signals, resulting in sparse textures often accompanied by characteristic speckle noise (Xu et al. 2024). This fundamental difference in imaging mechanisms makes SAR valuable for penetrating clouds, smoke, and darkness. However, the performance of deep learning-based SAR classifiers heavily depends on large-scale, high-quality labeled training datasets, which are scarce due to the complexity and cost of SAR image acquisition and annotation (Yin et al. 2024; Lin et al. 2024; Wang et al. 2025). This data limitation poses a persistent challenge to the development of robust SAR classification models (Lang et al. 2022).

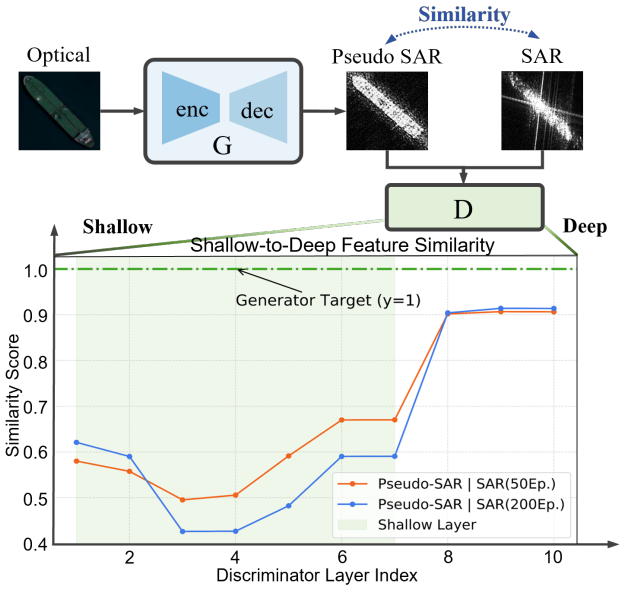


Figure 1: Shallow-to-deep feature discrepancy between generated Pseudo-SAR and real SAR Images. The upper part shows the generation process from optical to pseudo-SAR. The graph below reveals that pseudo-SAR images excessively retain optical image details in shallow features and fail to incorporate SAR low-level characteristics.

Current SAR data augmentation strategies have increasingly turned to Generative Adversarial Networks (GANs) through cross-modal style transfer, particularly by generating pseudo-SAR images from more abundant optical imagery (Song et al. 2022; Gao et al. 2023; Shi et al. 2022; Manocha and Afaq 2023). These methods often directly apply existing optical-to-optical GAN frameworks, such as CycleGAN (Dwarkani et al. 2021) and Pix2Pix (Qu et al.

*Corresponding authors: Aikun Xu

2019). The former utilizes cycle consistency loss to maintain structural integrity during unpaired image translation, while the latter employs conditional adversarial networks for paired image-to-image translation with direct supervision. However, due to the fundamental differences in imaging mechanisms between optical and SAR modalities, this direct application causes models to primarily focus on high-level semantics (such as target shapes and structures) while neglecting SAR-specific low-level features scattering structures and radar-induced texture patterns, which are critical for accurate SAR interpretation.

As a further illustrative example, Fig. 1 provides a detailed analysis of the cross-modal conversion from optical to SAR. The upper part shows the generation pipeline from optical to pseudo-SAR, while the lower graph quantifies the feature similarity across network depths between generated pseudo-SAR and real SAR. A high degree of similarity is observed in the deeper layers (Layers 8–10), suggesting that existing methods primarily capture high-level semantic structures. In contrast, the low similarity in the shallow layers (Layers 1–7) indicates a failure to model SAR-specific low-level features. This pattern reflects a critical limitation of existing GAN-based style transfer methods in balancing semantic preservation with SAR-specific feature fidelity.

To resolve this issue, we propose a Hierarchical Feature-Constrained GAN (HiFC-GAN) framework for Optical-to-SAR style transfer, which jointly enhances low-level SAR-specific feature fidelity and high-level semantic consistency. Specifically, at shallow layers of the generator, we introduce a local texture contrast (LTC) constraint to highlight SAR-specific structural patterns while suppressing redundant optical textures. At deeper layers, a prediction-based semantic feature mapping (SFM) constraint is applied to establish semantic alignment with SAR images in intermediate feature spaces, thereby reducing reconstruction information loss and enhancing semantic coherence. This hierarchical feature constraint mechanism enables HiFC-GAN to generate modality-consistent pseudo-SAR images that significantly improve downstream classification performance. Our main contributions are summarized as follows:

- We are the first to reveal that existing GAN-based optical-to-SAR style transfer methods primarily focus on transferring high-level semantic information while neglecting low-level features unique to SAR imagery.
- We propose a Hierarchical Feature-Constrained GAN (HiFC-GAN) that jointly employs local texture contrast at shallow layers and semantic feature mapping at deeper layers to bridge the modality gap from both structural and semantic perspectives.
- Extensive experiments demonstrate that HiFC-GAN significantly reduces the feature discrepancy between pseudo-SAR and real SAR images and improves the accuracy of downstream target classification tasks.

Related Work

Optical-to-SAR Image Generation

Transferring knowledge from well-annotated optical images to SAR images has attracted increasing attention recently (Zhao and Lang 2022; Lang et al. 2022; Tai et al. 2022), showing great potential in improving SAR target recognition accuracy and alleviating the scarcity of labeled samples. Two main approaches have been explored: domain-adversarial learning and GAN-based image translation, both offering new perspectives for cross-modal data augmentation and knowledge transfer. Domain-adversarial methods aim to align feature distributions between modalities. For instance, (Shi et al. 2024) proposed an unsupervised domain adaptation framework that integrates adversarial learning, consistency constraints, and contrastive learning to transfer knowledge from labeled optical to unlabeled SAR images, achieving fine-grained SAR ship classification. GAN-based methods, which mainly rely on basic architectures like Pix2Pix and CycleGAN, focus on generating pseudo-SAR images to expand training data. (Gao et al. 2023) introduced an Attention-Dense CycleGAN that incorporates CBAM to enhance the fine-grained representational quality of generated SAR images, thereby enhancing the performance in downstream tasks. To enhance transfer effectiveness, it is crucial to develop SAR-specific generative models that can better capture the unique modality properties.

Contrastive Representation Learning

Contrastive learning as an unsupervised representation learning method enables model to learn meaningful feature by pulling positive sample pairs closer and pushing negative pairs apart in the embedding space. The core of contrastive learning lies in the construction of positive and negative pairs and the design of pretext tasks that provide supervision signals derived from the data itself. For example, context prediction tasks (Zheng et al. 2023b; Wang et al. 2023) model spatial or temporal relationships between image patches or video frames, enabling the network to capture structural properties of data. Image colorization (Bourriez et al. 2024) requires model to recover color information from grayscale inputs, thus learning semantic and textural details. Collectively, these approaches leverage intrinsic data properties to drive the progress of unsupervised representation learning. Recent advancements have notably driven the progress of contrastive representation learning. For instance, SimCLR (Chen et al. 2020) enhances feature robustness through data augmentation and large batch training; SwAV (Caron et al. 2020) uses clustering to enforce multi-view consistency; and BYOL (Grill et al. 2020) and SimSiam (Chen and He 2021) eliminate the need for negative samples by maximizing the consistency between augmented views through prediction. Building on these foundations, a growing number of studies have explored the integration of contrastive learning into I2I translation tasks. Representative works (Park et al. 2020; Han et al. 2021; Lee et al. 2024) have introduced contrastive Learning methods into the generator or discriminator to improve image fidelity, semantic consistency, and training stability. Motivated by these efforts, we further investigates how contrastive learning can be leveraged across multiple feature levels to mitigate feature transfer imbalance in cross-modal translation settings.

Methodology

Preliminaries

Given a set of optical images from the source domain $\mathcal{S} = \{s|s \in \mathbb{R}^{H\times W\times 1}\}$ and a set of SAR images from the target domain $\mathcal{T} = \{t|t \in \mathbb{R}^{H\times W\times 1}\}$, we aim to learn a specific mapping $\mathcal{G} : \mathcal{S} \to \mathcal{T}$ from the source domain to the target domain to achieve a single-direction conversion from optical to SAR images. The sample size of the target domain N_t is typically smaller than that of the source domain N_s (i.e. $N_s \gg N_t$). HiFC-GAN consists of a generator G, a discriminator D, and two feature extractors E_x and E_y. The generator adopts a symmetric architecture with an encoder G_{enc} for feature compression and a decoder G_{dec} for reconstruction. Each component includes L layers of subnetworks. The input image is processed by the encoder and decoder of the generator, producing multi-level feature maps $\{f^l_{\text{enc}}\}^L_{l=1}$ and $\{f^l_{\text{dec}}\}^L_{l=1}$, which are then utilized by the constraint modules to guide hierarchical feature alignment:

$$f_L = \text{Encoder}(s) : f_1 \to f_2 \to \cdots \to f_L \tag{1}$$

$$t = \text{Decoder}(f_L) : f_L \to \tilde{f}_{L-1} \to \cdots \to \tilde{f}_0 \tag{2}$$

the discriminator differentiates real SAR from pseudo-SAR images to encourage realistic synthesis. The extractors are designed to respectively capture low-level and high-level feature representations from the generator's output.

Unidirectional Generative Model

Generator and Discriminator Loss. Based on the core concept of Generative Adversarial Networks (GANs), the generator $G_{S\to T}$ learns the mapping from the optical domain to the SAR domain, generating pseudo-SAR images $\hat{t} = G_{S\to T}(s)$. The discriminator D_T distinguishes between real SAR images t and the generated pseudo-SAR images $\hat{t}$. Both the generator and discriminator are trained alternately to improve the quality of the generated images and the discriminative ability of D_T. The optimization function for updating the discriminator is given by:

$$\max_{D_T} \mathcal{L}_G(D_T, G_{S\to T}) = \mathbb{E}_{t\sim T}[\log D_T(t)] + \mathbb{E}_{s\sim S}[\log(1 - D_T(G_{S\to T}(s)))] \tag{3}$$

where $D_T(t)$ indicates the discriminator's output for a real SAR image t, whose value should be closed to 1 to correctly identify the real sample; $D_T(G_{S\to T}(s))$ represents the discriminator's output for a pseudo-SAR image generated by $G_{S\to T}$, whose value should be close to 0 to accurately mark the fake sample. Thus, the discriminator aims to maximize $L(D_T, G_{S\to T})$ to improve its discrimination ability. During training, the discriminator is updated first, followed by the generator. In the generator update stage, its optimization function is defined as:

$$\min_{G_{S\to T}} \mathcal{L}_D(D_T, G_{S\to T}) = \mathbb{E}_{s\sim S}[\log(1 - D_T(G_{S\to T}(s)))] \tag{4}$$

At this point, $\mathbb{E}_{t\sim T}[\log D_T(t)]$ is independent of the generator and does not affect the optimization process. It is essential to minimize the $D_T(G_{S\to T}(s))$ to make the discriminator believe the generated samples are real, thus enabling the generator to update its parameters based on the discriminator's feedback. Combining these two update stages, the overall adversarial optimization objective is defined as:

$$\min_{G_{S\to T}} \max_{D_T} \mathcal{L}_{adv} = \mathcal{L}_G + \mathcal{L}_D \tag{5}$$

Identity Consistency Loss. In order to further stabilize the training process and ensure that the generator $G_{S\to T}$ does not deviate from the target SAR image distribution, we introduce the identity consistency loss in CycleGAN, thereby improving the alignment of the generated image with the target. The mathematical expression is shown below:

$$\mathcal{L}_{\text{idt}}(G_{S\to T}) = \mathbb{E}_{t\sim T}[\|G_{S\to T}(t) - t\|_1] \tag{6}$$

where $\|\cdot\|_1$ represents the L1 norm, which is used to measure the pixel-level difference between the generator output and the input. By minimizing $\mathcal{L}_{\text{idt}}$ during training, the generator is encouraged to maintain the original features and avoid unnecessary modifications when inputting SAR images.

The unidirectional generative model produces target-aligned SAR images but struggles to preserve precise spatial correspondence between input and output regions, leading to structural deviations of the optical image content in the conversion. To address this issue, we propose a hierarchical feature constraint mechanism that integrates multi-layer contrastive learning, replacing cycle-consistency constraints to minimize optical feature redundancy. The constraints are described in the following two subsections.

Local Texture Contrast Constraint

In shallow local texture contrast (Shallow-LTC) constraint, we extract intermediate feature maps $\{f^l_{\text{enc}}\}^L_{l=1}$ and $\{f^l_{\text{dec}}\}^L_{l=1}$ from the encoder and decoder, and sample P patches from each feature map, which are projected into a shared embedding space through a feature extractor $H_x(E_x(\cdot))$. As shown in Fig.2, decoder features (dark orange) serve as anchor points, while encoder features at identical spatial positions (light orange) are treated as positive samples. Features from distinct spatial locations (yellow) are negative samples. The alignment mechanism optimizes feature distances through:

$$\mathcal{L}_{\text{PatchNCE}} = -\frac{1}{P}\sum_{p=1}^{P} \log\left[\frac{\exp\left(\langle f^{\text{dec}}_p, f^{\text{enc+}}_p\rangle/\tau\right)}{\sum_{k=1}^{K}\exp\left(\langle f^{\text{dec}}_p, f^{\text{enc-}}_k\rangle/\tau\right)}\right] \tag{7}$$

where f^{dec}_p denotes the p-th patch feature from the decoder, $f^{\text{enc+}}_p$ its positive counterpart from the encoder, $f^{\text{enc-}}_k$ negative samples, P the total patches, and τ the temperature hyperparameter. For shallow layers, the composite loss aggregates layer-wise contrasts:

$$\mathcal{L}_{\text{LTC}} = \frac{1}{N_{\text{s}}}\sum_{l=1}^{N_{\text{s}}} \mathcal{L}^{(l)}_{\text{PatchNCE}} \tag{8}$$

where N_{s} denotes shallow layer count. This shallow-layer constraint avoids the reconstruction of redundant optical textures and ensures early-stage alignment between the encoded high-level semantics and the decoder-reconstructed SAR domain structures.

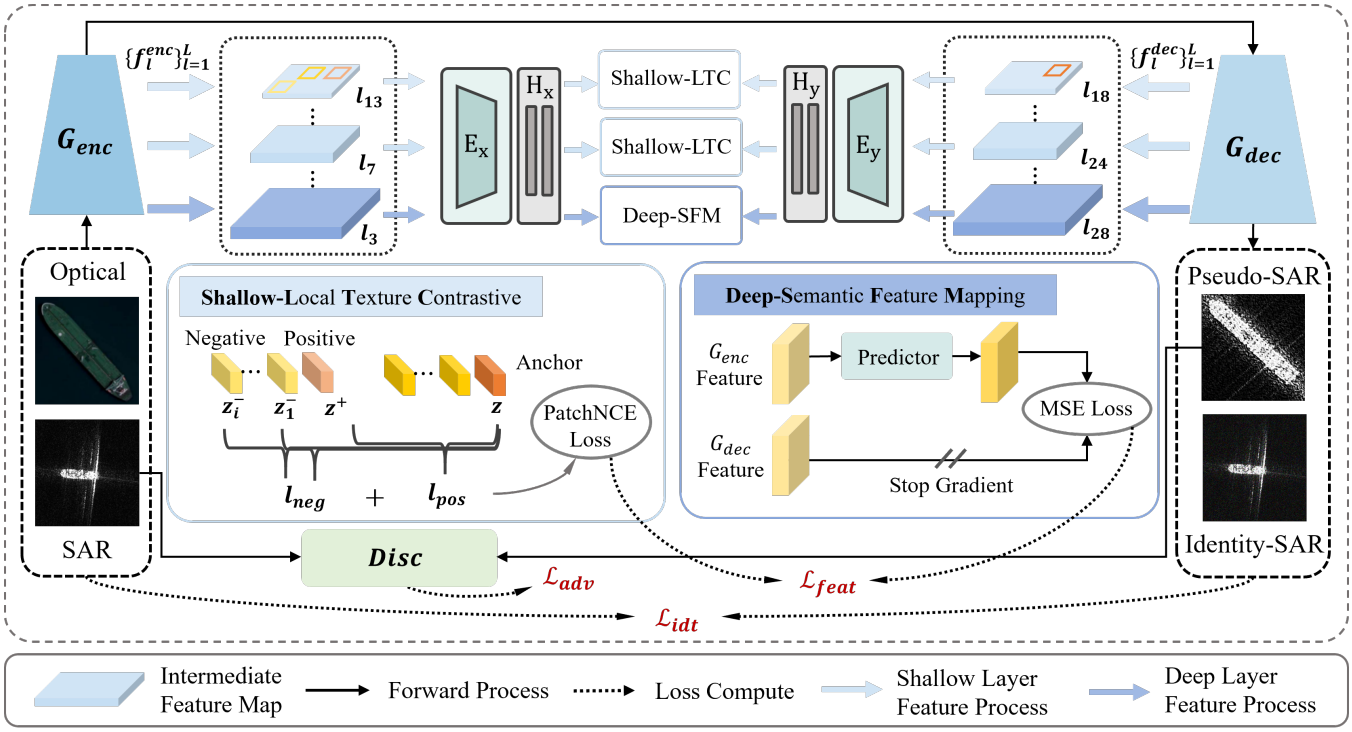


Figure 2: Illustration of the proposed HiFC-GAN method. HiFC-GAN integrates three losses: (1) global adversarial loss to facilitate GAN-based adversarial training; (2) an identity consistency loss to preserve content for real SAR inputs; (3) a hierarchical feature constraint combining shallow-LTC and deep-SFM for effective multi-level feature guidance.

Semantic Feature Mapping Constraint

For high-level features, the deep semantic feature mapping (Deep-SFM) constraint is proposed to ensure global semantic consistency in image translation tasks. In contrast to Shallow-LTC's local detail matching, Deep-SFM employs a prediction-based mechanism to align features across branches, promoting consistent high-level semantics and greater flexibility in style transfer by avoiding excessive constraints. high-level feature maps are first processed by the feature extractor $H_y(E_y(\cdot))$. The decoder output G_{dec} serves as the target representation z_{dec}, while the encoder output G_{enc} is transformed by a prediction head $P(\cdot)$ into the predicted representation z_{enc}. The Deep-SFM loss is:

$$\mathcal{L}_{\text{SFM}} = \frac{1}{N_d} \sum_{n=1}^{N_d} \|z_{enc} - \text{stopgrad}(z_{dec})\|_2^2 \tag{9}$$

where N_d denotes deep layer count, and both z_{dec} and z_{enc} are L2-normalized. The stopgrad prevents gradient updates to z_{dec} during backpropagation. This global perspective enhances the semantic integrity of the generated images, while the prediction mechanism facilitates mutual information transfer across feature spaces.

Through the synergy of Shallow-LTC and Deep-SFM, the model achieves precise feature alignment during image translation. Shallow-LTC ensures local detail consistency at shallow layers, while Deep-SFM optimizes global semantics at deeper layers. The overall hierarchical feature constraint loss is defined as:

$$\mathcal{L}_{\text{feat}} = \alpha \mathcal{L}_{\text{LTC}} + (1-\alpha)\mathcal{L}_{\text{SFM}} \tag{10}$$

where α is a weight that balances the degree of constraints between shallow and deep layers.

In summary, the adversarial loss $\mathcal{L}_{\text{adv}}$, the identity consistency loss $\mathcal{L}_{\text{idt}}$, and the feature constraints loss $\mathcal{L}_{\text{feat}}$ are integrated to formulate the total loss function for the optical-to-SAR generation task, denoted as follow:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{adv}}(G_{S\to T}, D_T, S, T) + \lambda_{\text{idt}}\mathcal{L}_{\text{idt}} + \lambda_{\text{feat}}\mathcal{L}_{\text{feat}} \tag{11}$$

where λ_{idt} and λ_{feat} are weights balancing the identity loss and the feature constraints loss, respectively.

Experiments

In this section, we evaluate our proposed HiFC-GAN from two perspectives: the quality of the generated pseudo-SAR images and their effectiveness in improving the performance of downstream classification tasks.

Experimental Setup

SAR Image Datasets. We use two benchmark SAR datasets for target classification, both serving as the target domain in the optical-to-SAR translation task. The OpenSARShip dataset (Huang et al. 2017), collected by the Sentinel-1A satellite, contains three types of ships. The FUSAR-Ship dataset (Hou et al. 2020) is derived from the high-resolution GF-3 SAR imagery and includes seven ship categories.

Model	P/m	T/h	FID↓		KID↓	
			M-T	S-T	M-T	S-T
CycleGAN	21.20	26	112.95		0.076	
DualGAN	17.47	20	128.41		0.079	
DCLGAN	28.81	18	91.94	289.66	0.065	0.287
CUT	**14.41**	**9**	88.56		0.061	
HiFC-GAN	14.65	12	**69.98**		**0.049**	

Table 1: Comparison of image generation quality with GAN-based models on OpenSARShip.

Optical Image Datasets. The FGSCR-42 dataset (Di, Jiang, and Zhang 2021) contains 9,320 ship instances across 42 ship types, with image sizes ranging from 50×50 to 1500×1500 pixels. For the source domain, four types that overlap with the target domain are selected: bulk carriers, container ships, tankers, and cargo, totaling 8,340 images.

Implementation details. The experiment consists of two stages. In the intermediate domain generation stage, the backbone network of the generator is selected as ResNet-9blocks. The learning rates for generator and feature extractor are set to 5×10^{-4}, respectively, while the discriminator's is set to 2×10^{-4}. Intermediate feature maps from the selected layer are used as input, and 256 patches are sampled for the shallow-LTC constraint. λ_{idt} and λ_{feat} are set to 2 and 1. In evaluation stage, the number of training epochs for each model is 100, the batch size is 32, and the learning rate is 2×10^{-3}.

Evaluation metrics. We evaluate the quality of generated images using Fréchet Inception Distance (FID) and Kernel Inception Distance (KID). FID extracts features from both generated and real images using the Inception network and computes the Fréchet distance to assess distribution differences in feature space. KID, based on Maximum Mean Discrepancy (MMD) and a polynomial kernel, calculates the distance between the feature representations of generated and real images. Smaller values of both metrics indicate higher similarity between generated and real images, reflecting better generative performance. Additionally, we use accuracy to evaluate SAR target classification performance.

Comparison with the GAN-based Models

HiFC-GAN produces the best-reconstructed SAR-style images. As shown in Fig. 3, bidirectional GANs enhance structural consistency between generated and source images, effectively preserving details such as ship structures, clouds, and noise. Specifically, CycleGAN ensures the reconstruction of optical details, but may introduce color shifts; DualGAN emphasizes adversarial loss, producing darker images while maintaining ship contours; DCLGAN improves image clarity by adding contrastive loss to better align features; unidirectional CUT learns local contrastive information but cannot guarantee full reconstruction. In contrast, the proposed HiFC-GAN preserves semantic information and effectively reconstructs SAR-style scattering characteristics by introducing a hierarchical feature constraint.

HiFC-GAN achieves the optimal KID and FID metrics. We quantify the performance of HiFC-GAN using metrics including FID, KID, model parameters, and training duration. Tab.1 presents the pseudo-SAR image generation results from optical images to OpenSARShip. M, S, and T represent the pseudo-SAR domain, the optical domain, and the SAR domain, respectively. The FID and KID values from S to T are higher than those from M to T. Bidirectional GAN models (top three rows) reduce the domain gap by an average of 61% (FID) and 74% (KID), while unidirectional GAN models (bottom two rows) achieve reductions of 70% (FID) and 79% (KID), demonstrating that GAN-based models can effectively reduce the domain gap. HiFC-GAN outperforms other GAN-based models on OpenSARShip, achieving an FID of 69.98 and a KID of 0.049. It also achieves higher training efficiency, with a parameter count of 14.65M and a training time of 12 hours. Despite being slightly less efficient than the CUT model, HiFC-GAN delivers superior performance, overcoming both the parameter limitations of bidirectional models and the shortcomings of unidirectional models in SAR style transfer.

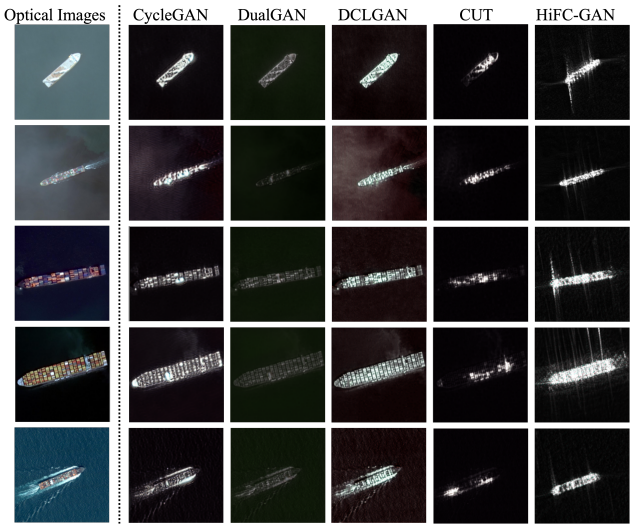


Figure 3: Comparison of intermediate domain generation results on the Optical-to-SAR task. We compare HiFC-GAN with existing methods and HiFC-GAN shows visually satisfactory results. The first column shows the input to the generator.

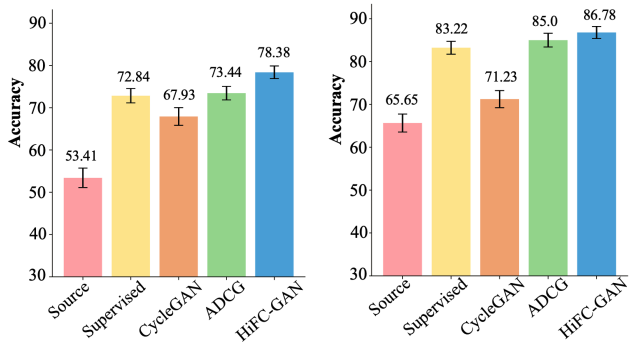
As shown in Tab.2, despite the increase in the training times of the larger FUSAR-Ship dataset, HiFC-GAN maintains high efficiency, achieving the best values of FID (89.76) and KID (0.067). Given the significant improvements in generative performance, we believe this lightweight computational overhead is acceptable.

Comparison with the Supervised Baseline

HiFC-GAN Improves the Performance of All Supervised Baseline Models. To address the challenge of limited labeled SAR images, we use HiFC-GAN to generate intermediate domain images, enabling the direct transfer of labels from optical images. Table 3 presents the comparative results of training the supervised baseline model before and

Model	P/m	T/h	FID↓		KID↓	
			M-T	S-T	M-T	S-T
CycleGAN	21.20	34	148.98		0.104	
DualGAN	17.47	30	132.77		0.084	
DCLGAN	28.81	24	114.29	281.38	0.078	0.257
CUT	**14.41**	**13**	101.23		0.071	
HiFC-GAN	14.65	15	**89.76**		**0.067**	

Table 2: Comparison of image generation quality with GAN-based models on FUSAR-Ship



(a) OpenSARShip (b) FUSAR-Ship

Figure 4: Accuracy performance of different comparison methods on two datasets of SAR images(AVG ± STD).

after data expansion. The performance of the models showed significant improvement on both datasets after data expansion. On the OpenSARShip dataset, VGG16 achieved the highest classification accuracy for container ships, bulk carriers, and tankers, with accuracy gains of 2.07%, 12.87%, and 3.44%, respectively. The overall accuracy, averaged over five training runs, increased from 72.84% to 78.38%, reflecting a gain of 5.90%. On the FUSAR-Ship dataset, expansion was applied to four categories: container ships, bulk carriers, tankers, and cargo ships. After expansion, VGG11 and VGG16 both achieved 100% accuracy for container ships and bulk carriers, while tankers and cargo ships achieved their best performance with VGG11. The Acc(7) score, which measures the average classification accuracy across all seven ship categories, increased from 83.22% to 86.78%, demonstrating a 3.56% performance gain.

HiFC-GAN outperforms SOTA methods in the optical-to-SAR translation task. In Fig. 4, the Source model, which directly applies optical classifiers to SAR data without any domain adaptation, performs poorly (53.41% on OpenSARShip, 65.65% on FUSAR-Ship), highlighting the severe domain gap. CycleGAN introduces cycle-consistency constraints for unpaired translation, but fails to capture SAR-specific low-level patterns, leading to suboptimal performance (67.93% / 71.23%), and even negative transfer in some cases. The ADCG model improves upon CycleGAN by incorporating a Dense Connection Module that alleviates feature redundancy, resulting in modest gains (73.44% / 85.00%). However, its reliance on cycle-consistency still limits its ability to adapt structural and semantic features jointly. In contrast, HiFC-GAN leverages a hierarchical constraint mechanism to achieve more effective feature alignment across domains. This design leads to the highest classification accuracy among all methods, reaching 78.38% on OpenSARShip and 86.78% on FUSAR-Ship, clearly demonstrating the advantage of joint low-level and high-level feature adaptation.

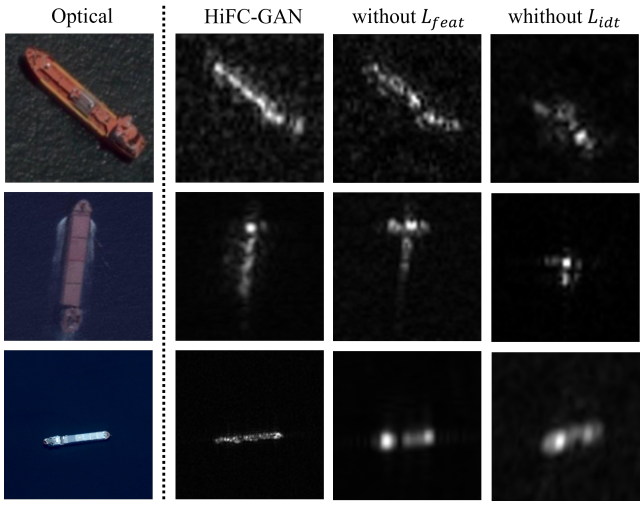


Figure 5: Ablation study results for the loss in HiFC-GAN.

Ablation Study

Impact of the two major losses in HiFC-GAN. As illustrated in Fig. 5, removing the feature constraint loss $\mathcal{L}_{feat}$ degrades image clarity and detail, highlighting its role in preserving spatial structures. Excluding the identity consistency loss $\mathcal{L}_{idt}$ results in images that retain basic shapes but lack accurate identity features, leading to category ambiguity and diminished augmentation effects. The full HiFC-GAN model achieves visually coherent and semantically reliable translations only when both losses are integrated.

Impact of hierarchical feature loss $\mathcal{L}_{feat}$ in HiFC-GAN. Fig. 6 shows that SFM-Only, relying solely on high-level semantic features, exhibits the slowest FID decline, indicating limited effectiveness for rapid domain adaptation. In contrast, LTC-Only, based on low-level texture cues, achieves faster FID reduction during the early training stage (within the first 50 epochs), but its performance plateaus due to the lack of global semantic consistency, resulting in suboptimal alignment at higher feature levels. HiFC-GAN achieves the lowest FID, demonstrating the advantage of joint hierarchical feature modeling for robust adaptation and improved image quality. HiFC-GAN consistently achieves the lowest FID throughout training. This demonstrates the complementary nature of the hierarchical constraint mechanism, where shallow layers capture modality-specific scattering structures and deep layers ensure semantic integrity.

Influence of the parameters of α in HiFC-GAN. Ablation I in Tab. 4 shows the effect of varying the parameter α on the balance between two major losses. As α increases,

Expand	Model	OpenSARShip				FUSAR-Ship					
		Categories			Acc(3)	Categories				Acc(4)	Acc(7)
		Container	Bulk	Tanker		Container	Bulk	Tanker	Cargo		
Before	AlexNet	0.6726	0.7133	0.6988	69.97±0.36	0.8260	0.9884	0.8695	0.8279	0.8780	79.55±0.53
	VGG11	0.6985	0.7461	0.7022	72.56±0.68	0.9637	0.9961	0.5393	0.9737	0.8682	81.10±0.62
	VGG16	0.7091	0.7445	0.7182	72.84±0.74	0.9924	0.9942	0.9328	0.9211	0.9601	83.22±0.23
	ResNet18	0.7033	0.7341	0.7012	71.70±0.94	0.9503	0.9207	0.9520	0.8421	0.9163	79.71±0.73
After	AlexNet	0.6372	0.7512	0.7671	72.35±0.41	0.9962	0.9942	0.8983	0.9190	0.9519	83.05±0.62
	VGG11	0.7134	0.7772	0.7260	76.24±0.26	**1.0000**	**1.0000**	**0.9827**	**0.9534**	**0.9840**	86.46±0.89
	VGG16	**0.7238**	**0.8403**	**0.7429**	**78.38±0.67**	**1.0000**	**1.0000**	0.9770	0.9514	0.9821	**86.78±0.71**
	ResNet18	0.7195	0.7884	0.7192	75.95±0.35	**1.0000**	0.9429	0.9770	0.8806	0.9501	83.34±0.31

Table 3: Classification accuracy before and after pseudo-SAR expansion by HiFC-GAN. Acc(4) and Acc(7) denote the average classification accuracy over the four augmented categories and all seven categories in the FUSAR-Ship dataset, respectively.

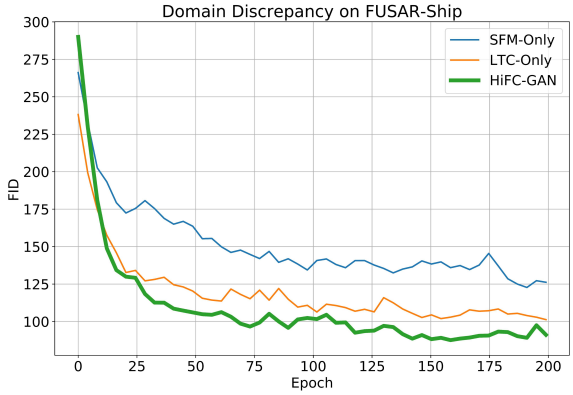


Figure 6: Comparison of domain discrepancy with different method settings on the FUSAR-Ship dataset.

	Settings	Optical→OpenSAR		Optical→FUSAR	
		FID↓	Acc	FID↓	Acc
I	$\alpha = 0.2$	**67.3**	77.46%	92.8	84.34%
	$\boldsymbol{\alpha = 0.4}$	69.9	**78.38%**	92.2	84.29%
	$\boldsymbol{\alpha = 0.5}$	70.5	76.63%	89.8	**86.78%**
	$\alpha = 0.8$	76.2	72.51%	**87.1**	85.27%
II	$\boldsymbol{L_{map} = 6}$	**69.9**	**78.38%**	**89.8**	**86.78%**
	$L_{map} = 8$	70.3	78.12%	90.4	86.10%

Table 4: Ablations on the effect of hyperparameters.

the model focuses more on the $\mathcal{L}_{LTC}$ loss; conversely, decreasing α shifts the focus towards the $\mathcal{L}_{SFM}$ loss. On the OpenSARShip dataset, decreasing α emphasizes on high-level semantics, which better aligns with the dataset's characteristics. Due to its lower resolution, relaxing detailed constraints helps enhance structural fidelity in generated images and improves classification performance. On the FUSAR-Ship dataset, a higher α enhances the preservation of low-level texture details, which is essential for high-resolution SAR imagery. When $\alpha = 0.5$, the model achieves an optimal trade-off between fine textures and semantic consistency, yielding the best classification accuracy.

Influence of the parameters of L_{map} in HiFC-GAN. The results of Ablation II in Tab. 4 show that $L_{map} = 6$ (with 3 feature maps for both the encoder and decoder) strikes a better balance between FID and classification accuracy. Fewer feature maps allow the model to focus on the most relevant features, avoiding redundancy and improving overall performance. In contrast, an excessive number of feature maps introduces unnecessary complexity and noise, negatively impacting model performance. In the symmetric architecture of HiFC-GAN, intermediate feature maps progress from shallow to deep layers and back to shallow layers. For $L_{map} = 6$, the involved layers are [3, 7, 13, 18, 24], with shallow constraints applied to pairs like $(3, 24)$ and $(7, 18)$, and deep constraints to pairs such as $(13, 18)$. This strategy leverages the principle that convolutional layers capture essential features for learning, with shallow layers focusing on low-level textures details and deeper layers capturing high-level semantics. Limiting the number of feature maps helps the model avoid overfitting and facilitates more effective learning.

Conclusion

We propose HiFC-GAN to address the trade-off between high-level semantic preservation and low-level feature authenticity in the Optical-to-SAR image translation. HiFC-GAN suppresses redundant optical features via local texture contrast constraints at shallow layers, while maintaining high-level semantic consistency through explicit feature mapping constraints at deeper layers. Experimental results demonstrate that HiFC-GAN not only enhances the visual fidelity of the generated pseudo-SAR images, but also boosts classification accuracy when applied to downstream classification tasks, outperforming existing SOTA methods. These results validate the effectiveness of hierarchical constraints in cross-modal generation and the practical utility of our method for SAR applications.

Acknowledgements

We appreciate constructive feedback from anonymous reviewers and meta-reviewers. This work was supported by the National Natural Science Foundation of China (62172442, 62172451), China Scholarship Council, and High Performance Computing Center of Central South University.

References

Bourriez, N.; Bendidi, I.; Cohen, E.; Watkinson, G.; Sanchez, M.; Bollot, G.; and Genovesio, A. 2024. Chadavit: Channel adaptive attention for joint representation learning of heterogeneous microscopy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11556–11565.

Caron, M.; Misra, I.; Mairal, J.; Goyal, P.; Bojanowski, P.; and Joulin, A. 2020. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems*, 33: 9912–9924.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 1597–1607. PmLR.

Chen, X.; and He, K. 2021. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 15750–15758.

Cheng, J.; Zheng, H.; Zheng, M.; Wang, L.; Wu, H.; and Zhang, J. 2025. ElimPCL: Eliminating Noise Accumulation with Progressive Curriculum Labeling for Source-Free Domain Adaptation. *arXiv preprint arXiv:2503.23712*.

Di, Y.; Jiang, Z.; and Zhang, H. 2021. A public dataset for fine-grained ship classification in optical remote sensing images. *Remote Sensing*, 13(4): 747.

Dwarkani, A.; Jain, M.; Thakkar, J.; and Kottursamy, K. 2021. Unpaired image-to-image translation using cycle generative adversarial networks. *International Journal of Engineering and Advanced Technology*, 9(6): 380–385.

Feng, X.; Zheng, H.; Hu, Z.; Yang, L.; and Zheng, M. 2024. Dual-stream contrastive predictive network with joint handcrafted feature view for SAR ship classification. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7810–7814. IEEE.

Gao, G.; Dai, Y.; Zhang, X.; Duan, D.; and Guo, F. 2023. ADCG: A cross-modality domain transfer learning method for synthetic aperture radar in ship automatic target recognition. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–14.

Grill, J.-B.; Strub, F.; Altché, F.; Tallec, C.; Richemond, P.; Buchatskaya, E.; Doersch, C.; Avila Pires, B.; Guo, Z.; Gheshlaghi Azar, M.; et al. 2020. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33: 21271–21284.

Han, J.; Shoeiby, M.; Petersson, L.; and Armin, M. A. 2021. Dual contrastive learning for unsupervised image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 746–755.

Hou, X.; Ao, W.; Song, Q.; Lai, J.; Wang, H.; and Xu, F. 2020. FUSAR-Ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition. *Science China Information Sciences*, 63: 1–19.

Huang, L.; Liu, B.; Li, B.; Guo, W.; Yu, W.; Zhang, Z.; and Yu, W. 2017. OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(1): 195–208.

Lang, H.; Yang, G.; Li, C.; and Xu, J. 2022. Multisource heterogeneous transfer learning via feature augmentation for ship classification in SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–14.

Lee, H.; Seol, J.; Lee, S.-g.; Park, J.; and Shim, J. 2024. Contrastive learning for unsupervised image-to-image translation. *Applied Soft Computing*, 151: 111170.

Li, Z.; Cai, J.; Xu, G.; Zheng, H.; Li, Q.; Zhou, F.; Yang, S.; Ling, C.; and Wang, B. 2025. Versatile Transferable Unlearnable Example Generator. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.

Lin, W.; Zheng, H.; Hu, Z.; Zheng, M.; and Yang, L. 2024. Coarse-to-fine granularity in multiscale featurefusion network for sar ship classification. In *International Conference on Artificial Neural Networks*, 31–45. Springer.

Manocha, A.; and Afaq, Y. 2023. Optical and SAR images-based image translation for change detection using generative adversarial network (GAN). *Multimedia Tools and Applications*, 82(17): 26289–26315.

Park, T.; Efros, A. A.; Zhang, R.; and Zhu, J.-Y. 2020. Contrastive learning for unpaired image-to-image translation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, 319–345. Springer.

Qu, Y.; Chen, Y.; Huang, J.; and Xie, Y. 2019. Enhanced pix2pix dehazing network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8160–8168.

Shi, Y.; Du, L.; Guo, Y.; and Du, Y. 2022. Unsupervised domain adaptation based on progressive transfer for ship detection: From optical to SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–17.

Shi, Y.; Du, L.; Guo, Y.; Du, Y.; and Li, Y. 2024. Unsupervised Domain Adaptation for Ship Classification via Progressive Feature Alignment: From Optical to SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*.

Song, Y.; Li, J.; Gao, P.; Li, L.; Tian, T.; and Tian, J. 2022. Two-stage cross-modality transfer learning method for military-civilian SAR ship recognition. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.

Tai, Y.; Tan, Y.; Xiong, S.; and Tian, J. 2022. Cross-domain few-shot learning between different imaging modals for fine-grained target recognition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 9186–9197.

Wang, F.; Wang, Y.; Li, D.; Gu, H.; Lu, T.; Zhang, P.; and Gu, N. 2023. Cl4ctr: A contrastive learning framework for

ctr prediction. In *Proceedings of the sixteenth ACM international conference on web search and data mining*, 805–813.

Wang, P.; Zheng, H.; Hu, Z.; Xu, A.; Zheng, M.; and Yang, L. 2025. PCM-SAR: Physics-Driven Contrastive Mutual Learning for SAR Classification. *arXiv preprint arXiv:2504.09502*.

Xu, B.; Zheng, H.; Hu, Z.; Yang, L.; Zheng, M.; Feng, X.; and Lin, W. 2024. Double reverse regularization network based on self-knowledge distillation for sar object classification. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7800–7804. IEEE.

Yin, J.; Duan, C.; Wang, H.; and Yang, J. 2024. A review on the few-shot SAR target recognition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Zhao, S.; and Lang, H. 2022. Improving deep subdomain adaptation by dual-branch network embedding attention module for SAR ship classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 8038–8048.

Zheng, H.; Hu, Z.; Liu, J.; Huang, Y.; and Zheng, M. 2022. MetaBoost: A novel heterogeneous DCNNs ensemble network with two-stage filtration for SAR ship classification. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.

Zheng, H.; Hu, Z.; Yang, L.; Xu, A.; Zheng, M.; Zhang, C.; and Li, K. 2023a. Multifeature collaborative fusion network with deep supervision for SAR ship classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–14.

Zheng, X.; Chen, X.; Schürch, M.; Mollaysa, A.; Allam, A.; and Krauthammer, M. 2023b. Simts: Rethinking contrastive representation learning for time series forecasting. *arXiv preprint arXiv:2303.18205*.