Remote Sensing Image Harmonization Method for Fine Grained Ship Classification

Jingpu Zhang , Ziyan Zhong , Xingzhuo Wei , Xianyun Wu * , Yunsong Li

Posted Date: 28 May 2024

doi: 10.20944/preprints202405.1810.v1

Keywords: Remote Sensing; Fine-grained Ship Classification; Transfer Learning; Image Synthesis

 doi:10.20944/preprints202405.1810.v1

Article

Remote Sensing Image Harmonization Method for Fine Grained Ship Classification

Jingpu Zhang [1], Ziyan Zhong [2] , Xingzhuo Wei [3], Xianyun Wu [1,2,*] and Yunsong Li [1]

[1] State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China; zhangjingpu1012@163.com (J.Z.); ysli@xidian.edu.cn (Y.L.);
[2] Guangzhou institute of technology, Xidian University, Guangzhou 510555, China; 1005457898@qq.com;
[3] Suzhou institute of technology, Xi'an Jiaotong-Liverpool University, Suzhou 215123; 3473752344@qq.com;
* Correspondence: xywu@mail.xidian.edu.cn

Abstract: Target recognition and fine-grained ship classification in remote sensing face the challenges of high inter-class similarity and sample scarcity. A transfer fusion-based ship image harmonization algorithm is proposed to overcome these challenges. This algorithm designs a feature transfer fusion strategy based on the combination of region aware instantiation and attention mechanism. Implement adversarial learning through image harmony generator and discriminator module to generate realistic remote sensing ship harmony images. Furthermore, the do-main encoder and domain discriminator modules are responsible for extracting feature representations of foreground and background, and further aligning ship foreground with remote sensing ocean background features through feature discrimination. Compared with other advanced image conversion techniques, our algorithm delivers more realistic visuals, improving classification accuracy for six ship types by 3% and twelve types by 2.94%, outperforming Sim2RealNet. Finally, a mixed dataset containing data augmentation and harmonizing samples and real data was proposed for the fine-grained classification task of remote sensing ships. Evaluation experiments were conducted on 8 typical fine-grained classification algorithms, and the accuracy of fine-grained classification for all categories of ships was analyzed. The experimental results show that the mixed dataset proposed in this paper effectively alleviates the long tail problem in real datasets, and the proposed remote sensing ship data augmentation framework performs better than state-of-the-art data augmentation methods in fine-grained ship classification tasks.

Keywords: remote sensing; fine-grained ship classification; transfer learning; image synthesis

1. Introduction

Advancements in remote sensing technology have expanded its applicability, yet image uniformity is influenced by system instabilities and environmental variables. In particular, ship imaging in remote sensing is significantly affected by changes in illumination and weather conditions [1].

In recent years, multi-style and arbitrary-style transfer learning has gained research attention. Dumoulin et al. achieved multi-style learning with conditional instance normalization (CIN) [2] and an improved texture network[3], but with limited styles due to training costs. Ye et al.[4] combined CIN with a convolutional attention module, enabling multi-style image transfer while preserving semantic information. Meanwhile, GDWCT[5] excelled in arbitrary style transfer by regularizing group-based style features, reducing computational expense. Also, LDA-GAN successfully transformed key object features for high-quality images. Luan et al. introduced comprehensive style transfer to match texture and color, maintaining spatial consistency using a two-pass algorithm for seamless object synthesis[7]. Similar to this, RainNet[8] approached style coordination as a transfer problem, offering a region-aware normalization (RAIN) module. Finally, Jiang et al.[9] proposed SSH, a self-supervised coordination framework that bypasses manual user annotation and overcomes shortcomings in image synthesis and subsequent coordination tasks, facilitating the coordination of any photographic composite image.

 doi:10.20944/preprints202405.1810.v1

Guo et al.[10.] segmented the synthetic image into reflection and illumination parts, introducing an autoencoder designed to coordinate these elements individually. They believe that inconsistencies in reflectivity and lighting between foreground objects and background can lead to a jarring appearance in the composite image. The lighting part is coordinated through migration technology, and the reflective part is coordinated through material consistency, thereby achieving coordination of the overall image. In the same year, Guo et al.[11.] proposed the Disentangled-Harmonization Transformer (D-HT) framework, which exploits the context dependence of the Transformer to ensure structural and semantic stability while enhancing the lighting and background harmony of foreground objects, thereby making the synthesized image more realistic. In addition to the above algorithms, image coordination technology based on deep learning in recent years also includes pixel-to-pixel conversion[12.][13.]. This method facilitates dense pixel-to-pixel conversion on low-resolution images and is increasingly used for image synthesis and coordination, heralding new trends in this field.

Drawing from the previous research, this paper adopts a CycleGAN-based unsupervised model for image-to-image conversion. It introduces a local-aware progressive method for transferring attributes between domains, enhancing the synthesis of remote sensing ship images with realistic and consistent features. The model's effectiveness is confirmed through comparative studies and detailed classification experiments on remote sensing ship imagery.

2. Methods

2.1. Simulated Remote Sensing Ship Image Construction

In this study, the generated simulated ship remote sensing image mainly includes two aspects: the image foreground characterized by ship targets and the ocean remote sensing background. For the foreground part, this study uses 3DMax modeling software to build high-fidelity three-dimensional models of various ship types, and then uses projection to achieve multi-pose imaging of these objects. For the background part, the ocean scenes in the "War Thunder" game show the dynamic undulations of waves, clearly visible, and realistic rendering effects. These scenes are similar to the spatial distribution characteristics of actual remote sensing images. Therefore, the study selected a perfect ocean image with an altitude of 1.5km-2km, a visibility distance of 20km, and no ships in the Pacific battle scene in "War Thunder" as the background template. During the training of the U-Net model, network parameters are carefully fine-tuned via backpropagation to generate accurate segmentation masks aligned with the input image. Figure 1 is a flow chart of the initial stages of image data synthesis discussed in this section. After 2D projection, the 3D model is fused with the simulated remote sensing ocean background to produce an initial version of the simulated remote sensing ship image, as well as its corresponding image foreground mask, laying the foundation for the upcoming image transmission and merging tasks.

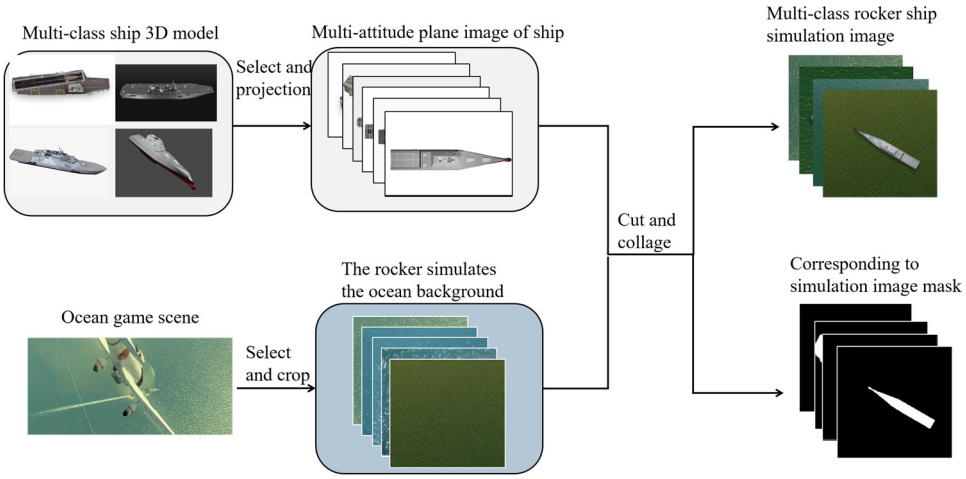

Figure 1. Simulated imaging module architecture.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 May 2024 doi:10.20944/preprints202405.1810.v1

2.2. Data Augmentation Model Based on Transfer Learning

As shown in Figure 2, the remote sensing ship target fine-grained recognition data augmentation model presented in this paper consists of three main modules: the Simulated Image Generating (SIG) module, the Foreground Feature Translation Aligning (FFTA) module, and the Background Feature Translation Aligning (BFTA) module. Both the FFTA and BFTA modules are based on the Local-Aware Progressive Image Conversion Network, LA-CycleGAN.

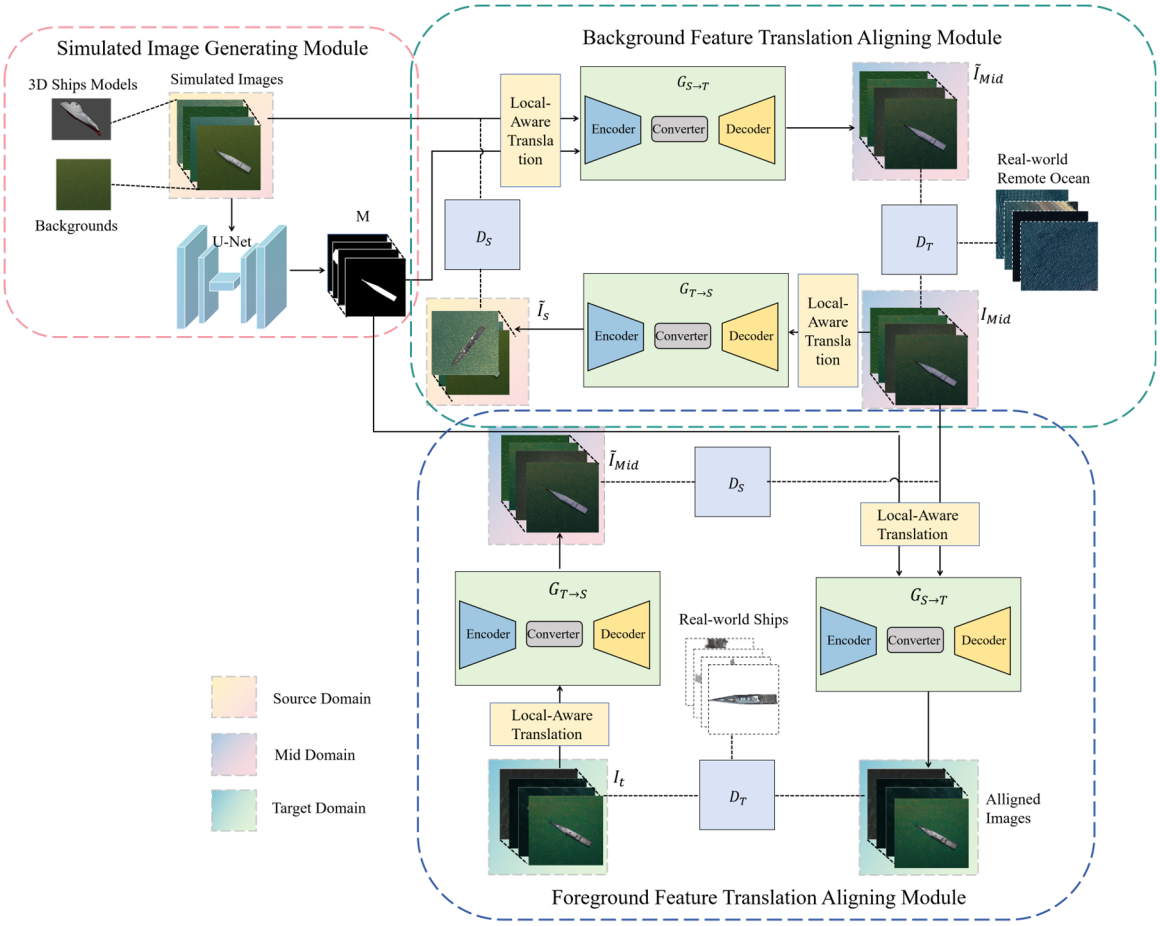

Figure 2. Framework of Data Augmentation Model.

The LA-CycleGAN framework comprises two generators $(G_{S\to T}, G_{T\to S})$and two discriminators (D_S, D_T). Specifically, $G_{S\to T}$ maps from the source to the target domain $(S \to T)$, while $G_{T\to S}$ reverses this process $(T \to S)$; The discriminator D_S discerns if an image originates from the source domain, while D_T assesses its belonging to the target domain. Upon model convergence, the aim is to maximize and align the migration area distribution $P_{target}(\hat{I}_t)$ in the output target image I_t with its feature distribution, denoted as $P_{target}(\hat{I}_t) = P_{target}(I_t)$. Nevertheless, discriminator D_T occasionally encounters difficulties in accurately identifying target domain images. A local perception strategy is thus invoked to alleviate blurring and augment the accurate capture of discrete regional features and textures, augmenting the exactness and detailed representation capabilities of image transference. This local-aware progressive transfer approach employs a binary mask M to preserve the invariant region R_{s_i} through element-wise multiplication with the source domain image. Post-migration feature mapping by the network, the unaltered area R1 is reapplied to the corresponding location in the generated image, ensuring $R_{s_i} = R_{o_i}$. This method neglects the transfer operation's effects on the invariant region R_{s_i}, circumventing irrelevant feature interference and maintaining the autonomous integrity of R_{s_i}.

In our system, two principal components work in tandem to achieve feature transfer alignment: the background feature transfer alignment module and the foreground feature transfer alignment

 doi:10.20944/preprints202405.1810.v1

module. Each module is embedded with an LA-CycleGAN network, and they are configured to operate in a sequential manner. The primary objective is to synchronize the simulated image's representational distribution with that of the authentic domain while ensuring the preservation of essential textural and structural attributes.

For the background component, we endeavored to enhance the variety of oceanic backdrop style representations by amassing over 1,000 distinct remote sensing images of the ocean and harbors characterized by varied surface hues and wave patterns from multiple regions via Google Earth. During the background feature alignment phase, simulated remote sensing images define the source domain, whereas authentic oceanic backdrops form the target domain. The transitional area R_{s_t} corresponds to the expansive scenic background of the synthetic image, and the ship target R_{s_i} remains unchanged. Inputs for the background feature transfer alignment module encompass the genuine ocean background imagery, the synthetic remote sensing images of the source domain, and the image mask pertinent to the transitional area R_{s_t}. It is pivotal for the background feature transfer alignment module to preserve the constancy of the ship target while aligning the background features.

The foreground feature transfer alignment module receives as its input the ship target image from the real-world domain, an intermediate image representative of the intermediate domain, and the corresponding foreground mask depicting the ship target's migration region. Within the realm of foreground alignment, the intermediate image serves as the source domain, and the real-world remote sensing ship image constitutes the target domain. Generator $G_{S\to T}$ facilitates the conversion from the intermediate to the real-world domain. During this phase, the ship target represents the migration area R_{s_t}, while the unchanged area R_{s_i} corresponds to the intermediate image's background. Throughout this operation, the focus is placed exclusively on the ship target, as the ocean background is disregarded, and cycle consistency loss is computed solely for the ship target, effectively the foreground object.

2.3. Remote Sensing Ship Image Harmonization Algorithm

Figure 3 presents the overall architecture of our proposed remote sensing ship image harmonization algorithm, which employs a transfer fusion stratagem. Embracing the principles of adversarial learning, this algorithm dissects the network into four strategic components: the harmonization image generator, harmonization image discriminator, domain encoder, and domain discriminator. Specifically, the harmonization image generator, referred to as G_H, is entrusted with generating the harmonized image I_H from the provided input. The role of the harmonization image discriminator is to evaluate both the generator's output and the actual remote sensing ship imagery, enabling adversarial learning through the propagation and updating of adversarial loss. The domain encoder, utilizing the produced harmonized image I_H, the genuine remote sensing image, and relevant masks for both the foreground and background, generates four distinct categories of features. Subsequently, the domain discriminator identifies the correlation between the foreground and background features and communicates the corresponding adversarial loss. This results in the ultimate harmonization of the composite remote sensing ship image.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 May 2024 doi:10.20944/preprints202405.1810.v1

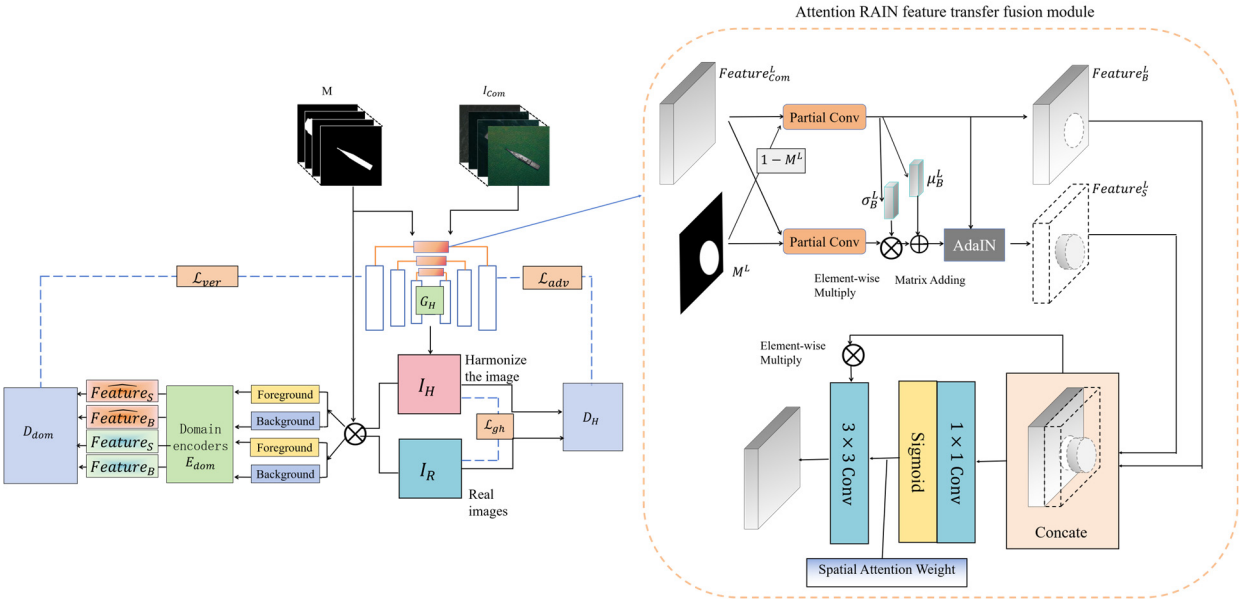

Figure 3. Overall Framework of Remote Sensing Ship Image Harmonization Algorithm.

Inspired by the concept of U-Net, the generator framework proposed in this study employs a feature transfer fusion strategy and embraces a simple, symmetric encoder-decoder structure devoid of a feature normalization layer. The encoder module, when fed with remote sensing ship simulation pictures along with their corresponding image masks, processes them through a convolution layer followed by three fundamental blocks, each structured in a LeakyReLU-Conv-IN pattern. LeakyReLU, besides being instrumental in feature extraction through convolutional layers, exhibits superior resistance to saturation when compared to ReLU. This aids in mitigating the vanishing gradient problem and catalyzes network convergence. The method utilizes intra-normalization IN[15.] to handle internal covariate shifts thereby enhancing the model's generalization capabilities. Owing to the symmetric structure, every fundamental block in the decoder section incorporates an Attention RAIN module. This facilitates the transfer of statistical data from background features to normalized foreground features, uninfluenced by foreground objects. It empowers the decoder to understand correlations between spatially varied features, thereby magnifying feature significance and minimizing informational loss or blurring. Lastly, through upsampling operations, the deconvolution layer reinstates the dimensionality of the feature map to match the original input image size, which not only restores detailed information but also promotes feature fusion.

The Attention RAIN module synergistically integrates RAIN with an attention mechanism to dynamically adjust ship foreground features while preserving background elements, thereby achieving complete style harmonization and facilitating holistic feature transfer and fusion within the image. In the encoder segment, it is possible to derive the feature map $Feature_{Com}$ of the remote sensing composite image along with the mask M of the ship's foreground target. Taking the Lth layer within the module as an instance, H^L, W^L, and C^L symbolize the height, width, and channel count of features at layer L, respectively, with $Feature^L_{Com}$ denoting the remote sensing composite image's feature map at layer L; and M^L represents a mask of resized ship foreground objects at layer L. Initially, in the adaptive calibration phase for ship foreground features, the process diverges from RAIN's approach of straightforward multiplication of input module's feature $Feature^L_{Com}$ by its foreground and background masks followed by normalization via IN. Instead, this module uses partial convolution on resized foreground feature $Feature^L_S$ and background feature $Feature^L_B$, using AdaIN to accomplish precise alignment and calibration between ship foreground features and remotely sensed ocean background features. Moreover, to counteract the potential for positional information loss induced by partial convolution in foreground and background features, a spatial attention mechanism is introduced. The technique involves processing the remote sensing ship synthetic image feature $Feature^L_{Com}$ through a 1x1 convolution kernel and activation function to

 doi:10.20944/preprints202405.1810.v1

derive the spatial attention weight w_{Com}^{SA}. Subsequently, the weight w_{Com}^{SA} is multiplied and combined with $Feature_{Com}^{L}$ to complete the fusion of ship foreground and remote sensing ocean background details. In the final stage, a 3x3 convolution kernel is used for information fusion and dimensionality reduction to generate the image feature $Feature_{Com}^{L}$ after feature splicing. This methodology significantly enhances the interconnectedness between foreground and background features, serving to augment feature extraction and dimensional reduction while conservatively retaining positional data.

$$Feature_{Com}^{L} = Conv(w_{Com}^{SA} \otimes Feature_{Com}^{L}) \quad (1)$$

The remote sensing ship harmonization algorithm, grounded in the transfer fusion strategy, aims to utilize the generator G_H to construct a harmonized image $G_H(I_{Com}, M)$, given synthetic image I_{Com} and its corresponding binary mask M for the foreground. The loss function of this algorithm is tripartite:

Primarily, to acknowledge the domain discrepancy between the input remote sensing synthetic ship image and the actual remote sensing ship image, the global fusion loss function L_{gh} is introduced to gradually approximate the input synthetic image I_{Com} to the real-world remote sensing image I_R.Subsequently, the harmonization image discriminator D_H adopts an adversarial learning approach. Its adversarial loss function, expressed as L_{adv}, inputs genuine remote sensing ship samples along with harmonized synthetic ship samples to ascertain the authenticity of the image. Concurrently, a domain discriminator D_{dom} is incorporated within the image harmonization operations to determine whether alignment in distribution is achieved between the foreground and background. Lastly, the algorithm invokes a corresponding loss function L_{ver} for the domain transfer procedure transitioning from the foreground to the background of the harmonized image sample.

$$\begin{aligned} L(D_H, D_{dom}, I_R, I_H, M) &= \lambda_1 L_{adv}(D_H, I_R, I_H) + \lambda_2 L_{ver}(D_{dom}, I_R, I_H, M) \\ &= \lambda_1 \left(\mathbb{E}_{I_R} \left[\max(0, 1 - D_H(I_R)) \right] + \mathbb{E}_{I_H} \left[\max(0, 1 + D_H(I_H)) \right] \right) \\ &\quad + \lambda_2 \left(\mathbb{E}_{I_R} \left[\max(0, 1 - D_{dom}(I_R, M)) \right] + \mathbb{E}_{I_H} \left[\max(0, 1 + D_{dom}(I_H, M)) \right] \right) \end{aligned} \quad (2)$$

$$\begin{aligned} L(G_H, I_{Com}, M) &= \lambda_1 L_{adv}(G_H, I_{Com}, M) + \lambda_2 L_{ver}(G_H, I_{Com}, M) + \lambda_3 L_{gh}(G_H, I_{Com}, M) \\ &= -\lambda_1 \cdot \mathbb{E}_{I_{Com}} \left[D_H(G_H(I_{Com}, M)) \right] - \lambda_2 \cdot \mathbb{E}_{I_{Com}} \left[D_{Dom}(G_H(I_{Com}, M), M) \right] \\ &\quad + \lambda_3 \cdot \left\| G_H(I_{Com}, M) - I_R \right\|_1 \end{aligned} \quad (3)$$

Among them, $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = 100$. Due to the unavailability of a public remote sensing ship IH dataset, this chapter harnesses the general benchmark synthetic dataset iHarmony4[16.] to train the image harmonization model. To differentiate images at varying processing stages, images produced by the remote sensing ship image harmonization algorithm, predicated on the transfer fusion strategy proposed in this section, are designated as harmonized images. The algorithm creates a harmonious linkage between two independent operational phases: the background feature transfer alignment module and the foreground feature transfer alignment module. By adjusting features such as foreground brightness representation, it ensures that the entire aligned image visually resembles actual images optimally.

3. Results

3.1. Dataset

The FGSR-42 dataset[17.], publicly available for fine-grained ship classification in remote sensing images, comprises approximately 9,320 images across 42 categories. It aggregates data from Google Earth and key remote sensing repositories like DOTA, HRSC2016, and NWPUVHR-10, featuring diverse warships and civilian vessels.However, FGSR-42's[17.] long-tail distribution leads to varied model performance across its categories. For simplicity and direct comparison, this study narrows down to 12 ship types from our lab. Table 1 shows the actual images in our experimental dataset.

 doi:10.20944/preprints202405.1810.v1

Table 1. Experimental Dataset.

Ship category	Detailed name	Inclusion of Generated Samples	Training Set Size	Test Set Size
Aircraft_ carrier	**Charles_de_Gaulle_aircraft_carrier**	Y	34	34
	Kuznetsov-class_aircraft_carrier	Y	34	34
	Nimitz-class_aircraft_carrier	F	388	165
	Midway-class aircraft_carrier	F	146	62
Landing_ship	Whitby_island-class_dock_landing_ship	F	195	83
Destroyer	Arleigh_Burke-class_destroyer	F	407	174
	Atago-class_destroyer	Y	35	35
	Murasame-class_destroyer	F	407	174
	Type_45_destroyer	Y	112	48
	Zumwalt-class-destroyer	Y	25	25
Combat_ship	Independence-class_combat_ship	F	148	62
	Freedom-class_combat_ship	Y	123	53

3.2. Experimental Environment

All experimentation conducted in this study was performed utilizing the Ubuntu 20.04 LTS operating system, bolstered by an NVIDIA GeForce GTX 2080 GPU equipped with 12GB memory capacity. The software environment was powered by Pytorch-2.13.0 and CUDA 11.6.

In this chapter, eight exemplary fine-grained classification algorithms were selected to substantiate the robustness of the experimental conclusions. These include ResNet-110[18.], ResNext[19.], DenseNet-121[20.], PyramidNet[21.], Wide Residual Network (WRN)[22.], ShuffleNet-v2[23.], EfficientNet-v2[24.], and Swin-Transformer[25.]. Uniformity in experimental conditions was maintained by setting identical epochs and hyperparameters for training these models. Specifically, ResNet-110 underwent training for 200 epochs, while the remaining classification models were trained for 100 epochs each. Stochastic gradient descent (SGD) was employed during the training phase with a momentum of 0.9 and a weight decay parameter set to 0.0001.

3.3. Ablation Experiment

This section divides the comprehensive model into a foreground feature transfer alignment module, a background feature transfer alignment module and an image coordination module based on the transfer fusion strategy, and combines them to complete the ablation experiment. The efficacy of these modules was assessed both visually and in terms of algorithm performance.

Figure 4 presents illustrative examples from each pivotal phase. As the training cycles iterate within the foreground feature transfer alignment module, the stylistic features of the sea surface area in the simulated image progressively align with those of the actual domain. Importantly, this transformation preserves other dimensional aspects of the background, such as sea surface brightness variations, wave texture formations, and so forth. Subsequently, the target is transferred from the simulated domain to the real domain through training iterations within the background feature transfer alignment module.

 doi:10.20944/preprints202405.1810.v1

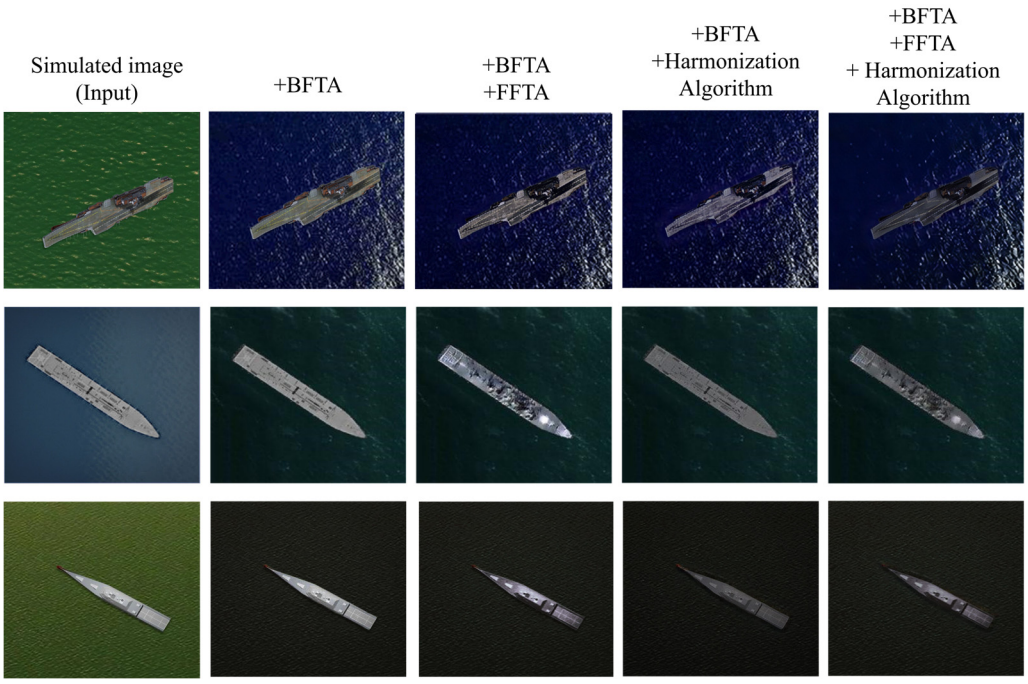

Figure 4. Results of Ablation Studies Across Different Modules.

The application of the image harmonization algorithm brings significant improvements to the previously discordant visual effects observed between the ship foreground and the ocean background within the respective foreground and background feature transfer alignment modules. This progress is vividly depicted in the fourth and fifth columns, from left to right, in Figure 4. The cumulative effects of each implemented stage serve to underscore the intricate features of the ship target, yielding a more cohesive and unified appearance across various regions. This enhancement in the richness of the visual portrayal of images augments the diversity of multi-dimensional information, thereby facilitating the elevation of accuracy in fine-grained ship classification.

Given that this study consists of a multi-stage, gradual feature alignment process, it's vital to scrupulously analyze each module's impact. Consequently, the foreground feature transfer alignment module, background feature transfer alignment module, and transfer fusion harmonization module were deconstructed and reassembled in this section. This reorganization facilitated a fine-grained identification experiment on remote sensing ships under three distinct classification algorithms, with results as shown in Table 2. Among them, Pyramid refers to PyramidNet[21.], EffiN-v2 refers to EfficientNet-v2[24.], and Swin-T refers to Swin-Transformer[25.]. In the table header, "HA" stands for Image Harmonization Algorithm. In the classification performance of five different classification algorithms (ResNet-11、ResNext、PyramidNet、EfficientNet-v2 and Swin-Transformer) using various module combinations, the transfer fusion harmonization module optimized the five models' classification performance by an average of 3.04% (average classification accuracy for six types of ships) and 2.67% (average classification accuracy for twelve types of ships). The results presented above indicate that the transfer fusion and harmonization module, as proposed in this paper, positively impacts the task of fine-grained ship classification, leading to a significant performance enhancement of the model over that achieved by utilizing the background feature transfer alignment module alone. The image harmonization algorithm adeptly integrates the foreground and background within the synthetic image, paving the way for the extraction of distinctive features using a deep learning model. This significantly enhances both the scope and fidelity of how image information is depicted. The detailed evaluation of the classification effect of the algorithm shows that the remote sensing ship image coordination technology based on the transfer fusion method detailed in this chapter is both influential and beneficial. It markedly improves the interpretation of image content as well as the precision in the fine-grained classification of ships.

 doi:10.20944/preprints202405.1810.v1

Table 2. Ablation Experiments for Fine-Grained Ship Classification Across Different Modules.

Classification AR		Baseline	SIG	+BFTA	+BFTA +FFTA	+BFTA +FFTA +HA	BFTA Gain	FFTA Gain	HA Gain
$\overline{AR}_{6class}$	ResNet	68.60	76.98	79.03	82.81	**86.00**	+2.05	**+3.78**	+3.13
	ResNext	74.64	79.66	76.31	82.03	**85.51**	-3.35	**+5.72**	+3.48
	Pyramid	76.57	78.47	81.63	85.01	**87.45**	+3.16	**+3.38**	+2.44
	EffiN-v2	83.68	86.16	87.23	88.90	**91.69**	+1.07	+1.67	**+2.79**
	Swin-T	87.32	88.14	87.37	91.48	**94.85**	-0.77	**+4.11**	+3.37
$\overline{AR}_{12class}$	ResNet	79.44	84.73	87.00	89.13	**92.05**	+2.27	+2.13	**+2.92**
	ResNext	83.95	84.91	85.55	89.02	**91.26**	+0.64	**+3.47**	+2.24
	Pyramid	84.48	85.33	87.22	89.30	**92.38**	+1.89	+2.08	**+3.08**
	EffiN-v2	89.10	90.56	90.70	92.31	**94.79**	+0.14	+1.61	**+2.48**
	Swin-T	91.53	92.99	92.24	94.48	**97.12**	-0.75	+2.24	**+2.64**

3.4. Hybrid Dataset Experiment

During the experimental process, it was discovered that datasets such as DOTA[26.], NWPU VHR-10[27.], and DSCR[28.] contain comparatively rare ship types, rendering them insufficient for fine-grained classification tasks. The FGSR-42[17.] dataset, due to its long-tail distribution, lacks authentic images in certain specific ship categories. In contrast to these datasets, this paper combines remote sensing ship harmonized images with real-world remote sensing ship imagery at a ratio of 7:1, enhancing the amalgamated remote sensing images with a series of high-precision category labels to create a new mixed dataset comprising 12 types of remote sensing ship images. Within this selection, the images of six ship types consist of composite data, specifically including Zumwalt-class destroyers, Charles de Gaulle aircraft carriers, Atago-class destroyers, Type 45 destroyers, Kuznetsov aircraft carriers, and Freedom class combat ships. The original images are sourced from the FGSCR-42 dataset. To address the scarcity of real images for these six ship types, 1,092 high-quality harmonized images were created as supplementary samples. These were then amalgamated with real-world images to establish a comprehensive hybrid dataset. Further experiments are conducted based on this meticulously formulated hybrid dataset.

This section evaluates the impact of this dataset on classification performance by conducting benchmark tests using seven common CNN classification networks and one Transformer network. These eight networks are ResNet-110, ResNext, DenseNet-121, PyramidNet, ShuffleNet-v2, WRN, EfficientNet-v2, and Swin-Transformer, with Table 3 listing the accuracy results of this dataset on each classification network.

 doi:10.20944/preprints202405.1810.v1

Table 3. Accuracy of Multiclass Classification Algorithms on Mixed Datasets.

AR (%)	RN-110	ResNext	DenseNet	PyramidNet	WRN	ShuffleNet-v2	EfficientNet-v2	Swin-T
$\overline{AR}_{6class}$	86.20	86.08	83.44	87.27	92.63	90.41	93.10	95.02
$\overline{AR}_{12class}$	92.07	92.20	87.89	93.02	95.82	84.93	95.88	97.43

Furthermore, to discern the detailed classification accuracy for each ship type within the dataset, this study employed the EfficientNet-v2 algorithm. Its performance in classifying all 12 types of ships is enumerated in Table 4.

Table 4. Classification Accuracy for 12 Types of Ships in EfficientNet-v2.

Ship Categories	AR(%)
Charles_de_Gaulle_aircraft_carrier	97.14
Kuznetsov-class_aircraft_carrier	100.00
Atago-class_destroyer	93.39
Type_45_destroyer	74.98
Zumwalt-class-destroyer	96.09
Freedom-class_combat_ship	97.00
Nimitz-class_aircraft_carrier	99.86
Midway-class aircraft_carrier	100.00
Whitby_island-class_dock_landing_ship	98.77
Arleigh_Burke-class_destroyer	97.26
Murasame-class_destroyer	97.78
Independence-class_combat_ship	98.89

As illustrated in Table 4, the classification accuracy for destroyers registers lower compared to that of aircraft carriers. Comparing the accuracy of different types of destroyers reveals that when the appearance of a destroyer is more similar to that of an aircraft carrier (such as the Atago-class destroyer), the corresponding classification accuracy tends to be higher. Moreover, ships that are larger in spatial dimensions and possess distinct shape attributes tend to achieve more favorable classification accuracy. This observation holds true across the accuracy results for various ship categories by other classification algorithms as well. The paper suggests that the smaller scale of ship targets makes their features relatively more difficult to discern, hence, it is also more challenging to extract detailed feature information during the image conversion and harmonization process.

4. Discussion

The image conversion network and the remote sensing ship image harmonization algorithm advocated in this article are benchmarked against other conversion techniques rooted in deep learning, to discern the effectiveness of this algorithm. Both CycleGAN[14.] and Sim2RealNet[1.] are acclaimed for their embodiment of global perceptual feature learning and classic neural style transfer, respectively. Simultaneously, CUT[29.] and SemI2I[30.] embrace the vanguard in contrasting learning and image conversion within the domain of remote sensing image processing, respectively.

 doi:10.20944/preprints202405.1810.v1

Implementing the transformation of identical remote sensing ship simulation images, this algorithm yields harmonized images, the results of which are juxtaposed against those generated by these representative models, as depicted in Figure 5.

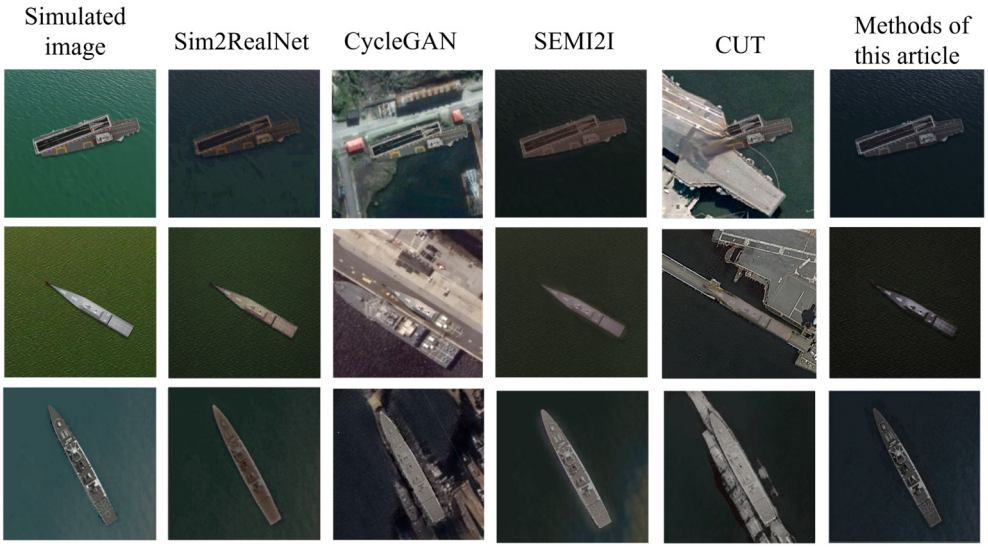

Figure 5. Visualization of Comparative Experiments for Image Conversion Methods.

As demonstrated in the second and fourth columns from left to right in Figure 5, it is apparent that neither Sim2RealNet[1.] nor SemI2I[29.] are capable of entirely transforming the full stylistic appearance of the image. Additionally, given the ocean background occupies a significant portion of the image, the deep learning model tends to favor the style feature distribution of the ocean backdrop during the conversion process over the visual expression features of the ship target. CycleGAN[14.] and CUT[29.], which employ global perception feature alignment for image conversion and style transfer, manage to achieve a positive visual alignment in the background area of the image. However, interference from features in other areas can cause the images generated through global perception to appear deformed or distorted, as illustrated in the third and fifth columns from left to right in the figure.

5. Conclusions

Addressing ship classification challenges in remote sensing, this paper introduces a transfer learning-based data enhancement framework, which simulates ship images and employs an image migration and fusion model for cross-domain mapping. To harmonize the foreground-background discordance, a remote sensing ship image harmonization algorithm was developed. These techniques generated a rich dataset, improving fine-grained ship classification. Experimental results show our methodology outshines traditional data augmentation methods, indicating substantial classification accuracy boosts up to 14.89% across different algorithms. Moreover, through ablation studies, each component of our feature transfer fusion strategy substantially boosted performance, enhancing model accuracy by 3.04% for six ship types and 2.67% for twelve types, validating our local-aware progressive image conversion and harmonization approach.

Author Contributions: Conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the China Postdoctoral Science Foundation (2013M540735); by the National Nature Science Foundation of China under Grant 61901388, 61301291, 61701360; by the 111 Project under Grant B08038; by the Shaanxi Provincial Science and Technology Innovation Team; by the Fundamental Research Funds for the Central Universities; by the Youth Innovation Team of Shaanxi Universities.

 doi:10.20944/preprints202405.1810.v1

Data Availability Statement: Data are available at https://github.com/Phoebe30/IHMFSC.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xiao Q, Liu B, Li Z, et al. Progressive data augmentation method for remote sensing ship image classification based on imaging simulation system and neural style transfer[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 9176-9186.
2. Dumoulin V , Shlens J , Kudlur M .A Learned Representation For Artistic Style[J]. 2016.DOI:10.48550/arXiv.1610.07629.
3. Ulyanov D, Vedaldi A, Lempitsky V. Improved texture networks: Maximizing quality and diversity in feed-forward styliza-tion and texture synthesis[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 6924-6932.
4. Ye W , Chen Y , Liu Y ,et al.Multi-style transfer and fusion of image's regions based on attention mechanism and instance segmentation[J].Signal Processing. Image Communication: A Publication of the the European Association for Signal Processing, 2023.
5. Cho W , Choi S , Park D ,et al.Image-to-Image Translation via Group-wise Deep Whitening and Coloring Transformation[J]. 2018.DOI:10.48550/arXiv.1812.09912.
6. Zhao J , Lee F , Hu C ,et al.LDA-GAN: Lightweight domain-attention GAN for unpaired image-to-image transla-tion[J].Neurocomputing, 2022.
7. Luan F , Paris S , Shechtman E ,et al.Deep Painterly Harmonization[J].Computer Graphics Forum, 2018, 37(4):95-106.DOI:10.1111/cgf.13478.
8. Ling J , Xue H , Song L ,et al.Region-aware Adaptive Instance Normalization for Image Harmonization[J]. 2021.DOI:10.48550/arXiv.2106.02853.
9. Jiang Y , Zhang H , Zhang J ,et al.SSH: A Self-Supervised Framework for Image Harmonization[J]. 2021.DOI:10.48550/arXiv.2108.06805.
10. Guo Z , Zheng H , Jiang Y ,et al.Intrinsic Image Harmonization[C]//Computer Vision and Pattern Recognition.IEEE, 2021.DOI:10.1109/CVPR46437.2021.01610.
11. Guo Z , Guo D , Zheng H ,et al.Image Harmonization With Transformer[C]//International Conference on Computer Vi-sion.2021.DOI:10.1109/ICCV48922.2021.01460.
12. Cong W , Tao X , Niu L ,et al.High-Resolution Image Harmonization via Collaborative Dual Transformations[J]. 2021.DOI:10.48550/arXiv.2109.06671.
13. Zhu Z , Zhang Z , Lin Z ,et al.Image Harmonization by Matching Regional References[J]. 2022.DOI:10.48550/arXiv.2204.04715
14. Zhu J Y , Park T , Isola P ,et al.Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[J].IEEE, 2017.DOI:10.1109/ICCV.2017.244.
15. Ulyanov D , Vedaldi A , Lempitsky V .Instance Normalization: The Missing Ingredient for Fast Stylization[J]. 2016.DOI:10.48550/arXiv.1607.08022.
16. Cong W , Zhang J , Niu L ,et al.Image Harmonization Dataset iHarmony4: HCOCO, HAdobe5k, HFlickr, and Hday2night.2019[2024-03-03].
17. Di Y , Jiang Z , Zhang H .A Public Dataset for Fine-Grained Ship Classification in Optical Remote Sensing Images[J]. 2021.DOI:10.3390/rs13040747.
18. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
19. Xie S , Girshick R ,Dollár, Piotr,et al.Aggregated Residual Transformations for Deep Neural Networks[J].IEEE, 2016.DOI:10.1109/CVPR.2017.634.
20. Huang G , Liu Z , Laurens V D M ,et al.Densely Connected Convolutional Networks[J].IEEE Computer Society, 2016.DOI:10.1109/CVPR.2017.243.
21. Han D , Kim J , Kim J .Deep Pyramidal Residual Networks[J]. 2016.DOI:10.1109/cvpr.2017.668.
22. Devries T , Taylor G W .Improved Regularization of Convolutional Neural Networks with Cutout[J]. 2017.DOI:10.48550/arXiv.1708.04552.
23. Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 116-131.
24. Tan M , Le Q V .EfficientNetV2: Smaller Models and Faster Training[J]. 2021.DOI:10.48550/arXiv.2104.00298.
25. Liu Z , Lin Y , Cao Y ,et al.Swin-Transformer: Hierarchical Vision Transformer using Shifted Windows[J]. 2021.DOI:10.48550/arXiv.2103.14030.
26. Xia G S , Bai X , Ding J ,et al.DOTA: A Large-scale Dataset for Object Detection in Aerial Images[J]. 2017.DOI:10.48550/arXiv.1711.10398.

 doi:10.20944/preprints202405.1810.v1

27. Cheng G , Han J , Zhou P ,et al.Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J].ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98(dec.):119-132.DOI:10.1016/j.isprsjprs.2014.10.002.
28. Di Y, Jiang Z, Zhang H, et al. A public dataset for ship classification in remote sensing images[C]//Image and Signal Processing for Remote Sensing XXV. SPIE, 2019, 11155: 515-521.
29. Park T, Efros A A, Zhang R, et al. Contrastive learning for unpaired image-to-image translation[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16. Springer International Publishing, 2020: 319-345.
30. Tasar O , Happy S L , Tarabalka Y ,et al.SemI2I: Semantically Consistent Image-to-Image Translation for Domain Adaptation of Remote Sensing Data[J].arXiv e-prints, 2020.DOI:10.1109/IGARSS39084.2020.9323711.