electronics

MDPI

Article

Side-Scan Sonar Image Classification Based on Style Transfer and Pre-Trained Convolutional Neural Networks

Qiang Ge [1], Fengxue Ruan [1], Baojun Qiao [1], Qian Zhang [2], Xianyu Zuo [1] and Lanxue Dang [1,*]

[1] Henan Key Laboratory of Big Data Analysis and Processing, School of Computer and Information Engineering, Henan University, Kaifeng 475004, China; gq@henu.edu.cn (Q.G.); kirito_rf@henu.edu.cn (F.R.); 10120086@vip.henu.edu.cn (B.Q.); xianyu_zuo@henu.edu.cn (X.Z.)

[2] The Institute of Acoustics of the Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100190, China; zhangqian587@mail.ioa.ac.cn

* Correspondence: danglx@vip.henu.edu.cn

Abstract: Side-scan sonar is widely used in underwater rescue and the detection of undersea targets, such as shipwrecks, aircraft crashes, etc. Automatic object classification plays an important role in the rescue process to reduce the workload of staff and subjective errors caused by visual fatigue. However, the application of automatic object classification in side-scan sonar images is still lacking, which is due to a lack of datasets and the small number of image samples containing specific target objects. Secondly, the real data of side-scan sonar images are unbalanced. Therefore, a side-scan sonar image classification method based on synthetic data and transfer learning is proposed in this paper. In this method, optical images are used as inputs and the style transfer network is employed to simulate the side-scan sonar image to generate "simulated side-scan sonar images"; meanwhile, a convolutional neural network pre-trained on ImageNet is introduced for classification. In this paper, we experimentally demonstrate that the maximum accuracy of target classification is up to 97.32% by fine-tuning the pre-trained convolutional neural network using a training set incorporating "simulated side-scan sonar images". The results show that the classification accuracy can be effectively improved by combining a pre-trained convolutional neural network and "similar side-scan sonar images".

Keywords: style transfer; target classification; side-scan sonar images; transfer learning; convolutional neural network

Citation: Ge, Q.; Ruan, F.; Qiao, B.; Zhang, Q.; Zuo, X.; Dang, L. Side-Scan Sonar Image Classification Based on Style Transfer and Pre-Trained Convolutional Neural Networks. *Electronics* **2021**, *10*, 1823. https://doi.org/10.3390/electronics10151823

Academic Editor: Manohar Das

Received: 4 June 2021
Accepted: 26 July 2021
Published: 29 July 2021

1. Introduction

As high-resolution, multi-purpose marine detection equipment, side-scan sonar is widely used in the ocean, lakes, and other bodies of water, and is currently the main technique for underwater target detection. It can quickly obtain large-area and high-resolution acoustic images of the seafloor and, combined with seafloor image data from a small number of sampling sites, researchers can distinguish the different types of objects on the seafloor based on side-scan sonar images; this is useful for activities such as mine detection, seafloor mapping, marine ecosystem monitoring, and underwater rescues [1–4]. For underwater rescues, side-scan sonar has been widely used to detect shipwrecks, aircraft, and victims on the seafloor or on the bottom of lakes.

With the increase in maritime investigation activities and the development of inland river water traffic, water accidents occur frequently, such as the lost Malaysia Airlines incident in 2014, and the 2015 Yangtze River shipwreck. For long rescue tasks, sonar rescue personnel constantly scrutinize the screen to see whether there is a target object. After working for a period of time, the staff become tired and miss the rescue targets more easily, and the efficiency is low. In order to reduce the workload of staff, decrease the number of subjective errors caused by visual fatigue, and improve work efficiency, the automatic classification of side-scan sonar seafloor images has practical significance.

Shape features or shadow shapes are generally used for classification in traditional methods of underwater target classification, and many model-based methods have been proposed. The accurate segmentation of target object shadows is achieved by using a priori knowledge of sonar images combined with a Markov random field model (MRF) to segment images into shadows, submarine reverberations, and highlight target regions as a way to improve classification accuracy [5–8]. Sinai et al. used the Chan–Vese active contour algorithm to convert images into shadows and highlight mappings and made use of geometric characteristics of the target to achieve a good recognition effect in the image [9]. Martin tried to improve the classification accuracy by fusing information [10]. Quidu et al. used dynamic segmentation and genetic algorithms to generate sets of individuals, and the Fourier decomposition of individual contours acted as potential identifiable solutions and moved contours to match shadows for dynamic classification [11]. To a certain extent, the model-based method has excellent classification performance, but it relies too much on the local feature descriptor of prior knowledge; thus, it has not been widely recognized.

As a small-sample classifier, support vector machines (SVMs) have been widely used in the field of side-scan sonar image classification, achieving excellent results. Among them, the texture features of sonar images were extracted using wavelet coefficients; then, the nearest neighbor algorithm and SVM were used to classify the sonar images [12]. Guo et al. extracted image features using a gray-level co-occurrence matrix and classified them with an SVM [13]. For the fast and efficient classification of sonar images, Zhu M. et al. proposed a classification method based on principal component analysis (PCA) and an extreme learning machine (ELM) for sonar images, which is stable and has higher classification accuracy [14]. Although the above methods are effective in improving the classification accuracy to some degree, the methods they proposed only address the specific information of sonar images and cannot make good use of all feature information. In addition, the submarine environment is complex and variable, and different angles and heights of imaging devices can produce different sonar images [15], which further limits the popularity of the above methods.

In recent years, the classification performance of machine learning has almost reached the level of humans in the field of conventional image classification. Dobeck et al. used a detection density algorithm with a stepwise feature selection strategy, combined with a k-nearest neighbor attractor neural network (KNN) and optimal discriminant filter for mine detection [16]. Similarly, deep neural networks play an important role in target recognition. Kim et al. used convolutional neural networks (CNNs) for submarine vehicle recognition [17] and Hoang et al. used CNN models for mine detection [18], both of which prove that deep neural networks can substantially improve classification accuracy compared with traditional methods. The above classification methods mainly involve the classification and detection of mines and single targets on the seabed. It cannot meet the needs of underwater rescue for the classification of multiple targets such as drowning victims, aircraft, and wrecks.

To address this problem, this paper proposes a multi-target classification based on a convolutional neural network, including drowning victims, aircraft, wrecks, and the seabed. The complexity of aircraft and shipwreck structures is much higher than that of mines and the seabed; therefore, a CNN model with powerful feature extraction capabilities was selected. Convolutional neural networks can extract sample features well and input them into the classifier to classify sonar images. However, a large amount of sample data are essential for training CNN models. Side-scan sonar data are difficult and costly to obtain, and there are fewer image samples containing specific targets such as aircraft and victims [3,4], which makes it much more difficult to train convolutional neural networks. The side-scan sonar dataset can work well as a small dataset with the proper introduction of migration learning. In addition, some easy-to-implement methods are used to further improve the classification accuracy, such as synthetic data, to efficiently solve difficult acoustic problems at a low cost. In order to overcome the deficiencies, the following procedures were conducted in this paper:

- Considering the problem of unbalanced data of side-scan sonar samples, we propose a method to generate "simulated side-scan sonar images" by combining image segmentation and style transfer networks with optical images as inputs, which are used to generate images of drowning victims and aircraft;
- We modified the image style transfer network and performed experimental comparisons, and the results showed that the improved network generates clearer and more natural images;
- By using pre-trained CNN model classification, such as VGG19, 70% of the real side-scan sonar images and "similar side-scan sonar images" were used to fine-tune the CNN model; then, 30% of the real side-scan sonar images were used to verify the model, and the final test accuracy achieved was up to 97.32%, which is better than the classification performance of the fine-tuned model merely using real side-scan sonar images.

2. Methods

2.1. Image Style Transfer Algorithm

The image style transfer algorithm based on the CNN model proposed by Gatys et al. [19] extracts the style and content feature map of the image through five convolutional layers, and the randomly generated white noise image is continuously iteratively updated to generate a simulation image that maintains both the original content of the synthesized image and the style of the side-scan sonar image.

The style image is denoted by s, where the feature maps obtained on all convolutional layers are denoted by S^l, and l denotes the number of layers. The content image is denoted by c, where the feature maps only obtained on the fourth convolutional layers are denoted by C^l. The random noise image is denoted by r, where the feature maps obtained on all convolutional layers are denoted by R^l.

The texture features of style images are represented by the Gram matrix [19]:

$$G_{ij}^l = \sum_k R_{ik}^l R_{jk}^l \tag{1}$$

where G_{ij}^l is the inner product between feature maps i and j in the l layer, and R_{ik}^l is the activation of the ith filter at position k in layer l. The formula finds the inner product of the two feature maps. There is no relationship between the texture features and the position of the image; therefore, it is the result of the inner product calculated by the Gram matrix and the position of the two feature maps, which can be used to measure the texture feature, which is the most common way to express texture features of the images.

The mean square error E_l is calculated by the Gram matrix G^l of r and the S^l of style image s, and the style loss function L_{style} is used to describe the difference in style, which is denoted as [19]:

$$L_{style}(s, r) = \sum_{l=0}^{L} w_l E_l \tag{2}$$

where w_l are weighting factors of the contribution of each layer to the total loss, L is the maximum layer index in the CNN, so the value of l ranges from 0 to L. E_l is denoted as:

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} \left(G_{ij}^l - S_{ij}^l \right)^2 \tag{3}$$

where i and j denote the ith feature map and jth feature map of the l layer, respectively.

The loss function of content image calculates the mean square error $L_{content}$ by R^l of the fourth layer and C^l, which is denoted as:

$$L_{content}(c, r, l) = \frac{1}{2} \sum_{i,j} \left(R_{ij}^l - C_{ij}^l \right)^2 \tag{4}$$

where i and j denote the ith feature map and jth feature map of the l layer, respectively.

The total loss, L_{total}, is denoted as:

$$L_{total} = \alpha L_{style}(s, r) + \beta L_{content}(c, r) \quad (5)$$

where α and β denote the weights of the style loss function and the content loss function, respectively.

The real side-scan sonar image and processed optical image are input into the convolutional neural network, and the random white noise image is constantly updated and iterated through the gradient descent; finally, the simulation image with the original content of the synthesized image and the style of the side-scan sonar image is output.

2.2. Hybrid Dilated Convolution and K-Means Algorithm

As shown in Figure 1, hybrid dilated convolution (HDC) adds cyclic cavities into the convolution filter at an interval rate of [1,1,2,2,5,5], which increases the perceptual field without losing resolution and without introducing additional parameters, and contains a larger range of information in each convolution output [20]. Better results than traditional convolution can be obtained in various problems that require global information dependence, such as style transfer and semantic segmentation. However, hybrid dilated convolution must be satisfied with the following equation:

$$M_i = max[M_{i+1} - 2r_i, M_{i+1} - 2(M_{i+1} - r_i), r_i] \quad (6)$$

where r_i is the interval rate of layer i, and M_i is the maximum interval rate of layer i. Then, suppose there are n layers, and the default is $M_n = r_n$. Assume that the convolution kernel is $k \times k$, and the target is $M_2 \leq k$. In this way, all the holes can be covered by ordinary convolution.

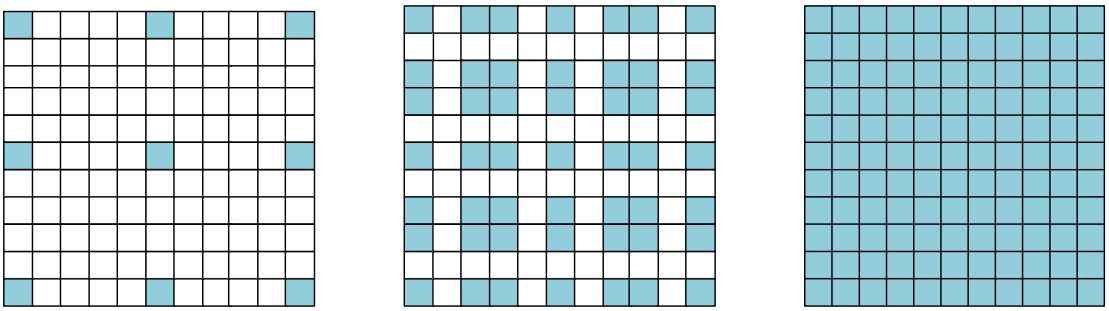


Figure 1. Schematic diagram of hybrid dilated convolution.

The K-means algorithm is an unsupervised clustering algorithm. The main idea of the algorithm is to use the mean value of the objects in each cluster as the cluster center, and the objects in the dataset are divided into k classes according to the principle of minimum distance from the cluster center through iteration, where k is the number of clusters, and the clustering performance evaluation function is optimized so that each cluster itself is as similar as possible, and each cluster is as different as possible [21].

The K-means algorithm uses distance as the similarity index. In this paper, Euclidean distance was used to calculate the distance:

$$\mathrm{dist}(X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (7)$$

where the smaller the distance, the higher the sample similarity. The larger the distance, the worse the similarity.

2.3. Pre-Trained Convolutional Neural Network

Pre-trained convolutional neural networks are also called transfer learning. In the case of small samples, transfer learning can be introduced appropriately. Transfer learning applies knowledge learned in one field to similar fields. This is due to the generality of the underlying image features, so as to achieve the learning effect of transferring labeled data or knowledge structures from related domains and completing or improving the target domain or task as a way to overcome the problem of data deficiency [22].

If a domain is represented by $D = \{X, P(X)\}$, X denotes the feature space and $P(X)$ denotes the marginal probability distribution. A task is indicated by $T = \{y, f(.)\}$, where y denotes the label space and $f(.)$ denotes the target prediction function. Then, the definition of transfer learning can be expressed as follows: given a domain D_y and a task T_y, we can obtain data information from another domain, D_x, and task, T_x [22]. The purpose of transfer learning is to improve the predictive performance $f(.)$ of the task T_y by means of this transferred knowledge. Here, domain D_y and task T_y are two different but related domains from domain D_x and task T_x. Figure 2 shows the process of deep transfer learning.

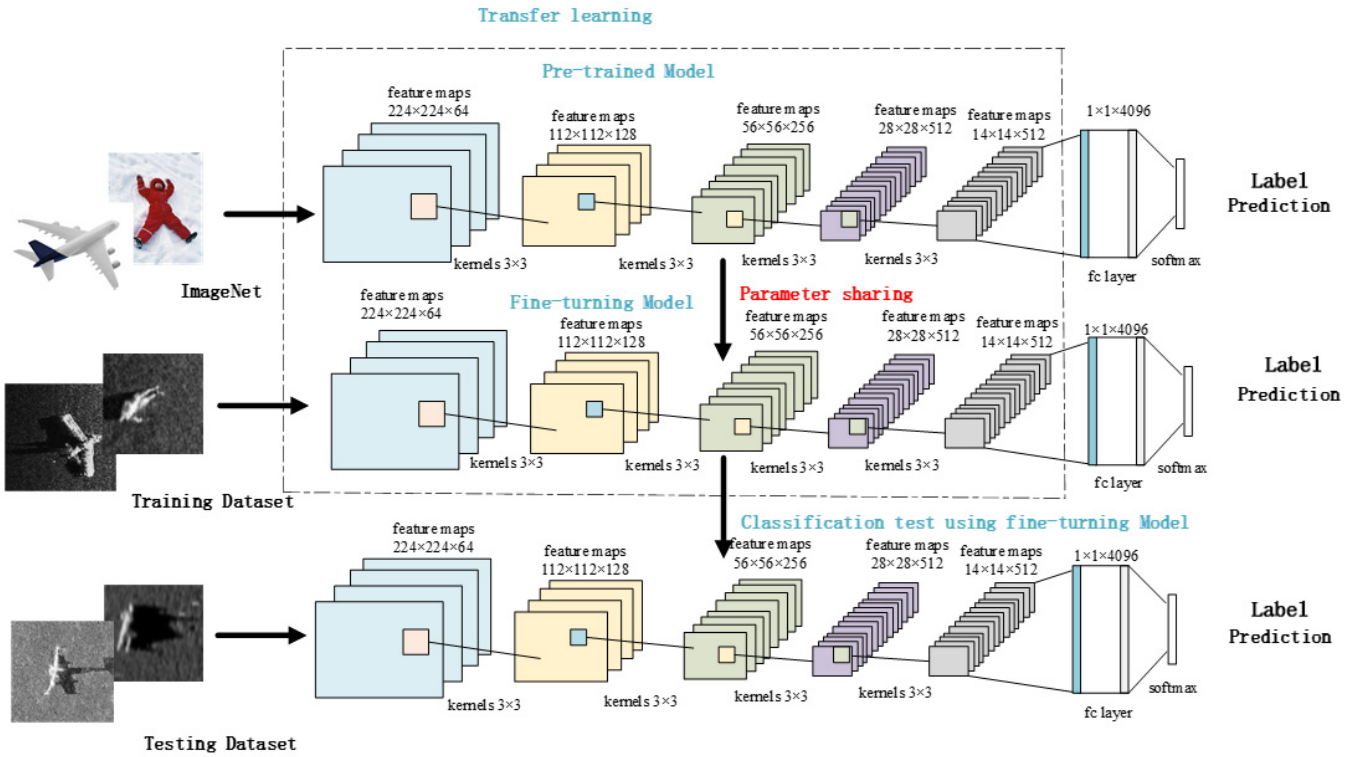


Figure 2. Deep transfer learning used in this paper. A CNN model is first pre-trained on the ImageNet dataset; then, the trained weights are transferred and the CNN model is fine-tuned by the side-scan sonar dataset; finally, the model is tested on the side-scan sonar image validation set.

Transfer learning is divided into four major categories based on the specific implementation method: instance-based transfer, feature-based transfer, transfer based on a shared parameter/model, and relationship-based transfer. Currently, model-based transfer learning is one of the most popular transfer learning methods, especially in the image domain; here, pre-training ImageNet was chosen to initialize the model. Experiments have demonstrated that the learning effect of pre-training a model based on ImageNet and then fine-tuning it on a small dataset is better than that of direct transfer learning with a fixed convolutional layer [23]. The reason why model-based transfer learning works well is the generality of the low- and mid-level features of the image data.

3. Synthesis of "Simulated Side-Scan Sonar Images"

3.1. Image Synthesis Method

There are few images containing specific target objects in side-scan sonar, which are prone to overfitting problems when using convolutional neural networks for classification. Although the problem of insufficient data can be compensated to a large extent by using the transfer learning method, less real data results in fewer samples of high-level features extracted by the convolutional neural network, which will lead to more classification errors. Therefore, it becomes particularly important to simulate more side-scan sonar images containing targets such as victims and aircraft.

Images are stylized and processed by traditional style transfer through building statistical models and textures, etc. [24–27]. Gatys et al. [19,28,29] proposed a convolutional-neural-network-based image style transfer algorithm to implement image style transfer. After training the multilayer CNN, the artistic style is recognized and extracted, and then applied to ordinary photographs, so that the generated images have the original content and artistic style at the same time. Additionally, this image content exhibits the detail loss phenomena.

In this paper, we propose a side-scan sonar image synthesis method based on convolutional neural networks; the synthesis process is shown in Figure 3. The method first clusters the optical images to highlight the target objects, then extracts clustered target contours using binary threshold segmentation, fuses them with the clustered images to form images containing shadow regions, and finally uses the real side-scan sonar images as style images and the fused images as content images for style migration. The detailed process is as follows:

- The input optical images are clustered to separate the front and back backgrounds and highlight the target objects. In this study, the K-means algorithm is used to cluster the optical images into two categories, namely, background and target object, and the detailed features are removed;
- A digital morphology opening operation is used on the image to eliminate small and meaningless target objects, fill some holes, and eliminate small particle noise in the target region;
- The background color of the clustered image is changed to gray and the color of the target object is changed to white; then, the target object is extracted using binary threshold segmentation. Likewise, the background color of the clustered image is changed to gray and expanded by 1.2 times along the x-axis or y-axis as the shadow region;
- The extracted target object is fused with the expanded image to obtain an image with a shadow region. This image used as the content image and the real side-scan sonar image used as the style image are simultaneously input into the modified style transfer network to generate the "simulated side-scan sonar images", as shown in Figure 3.

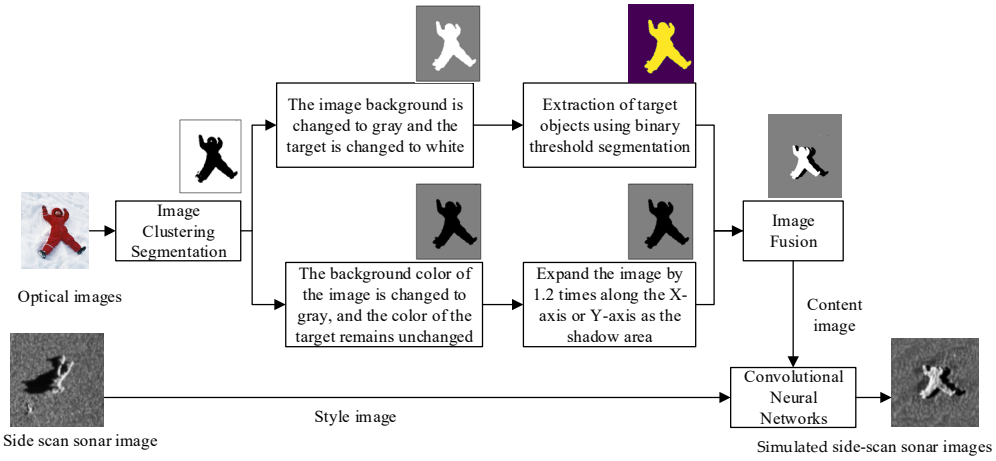


Figure 3. Synthesis process of "simulated side-scan images".

3.2. Improved Style Transfer Network

The output image obtained by the style transfer algorithm of the convolutional neural network can learn the texture information and color of the style image very well. However, the phenomenon of edge detail loss exists. Based on this problem, this paper introduces the K-means algorithm and hybrid dilated convolution on the basis of the original algorithm to make the image content and style more realistic.

As shown in Figure 4, s is the style image, c is the fused optical image, and r is the randomly generated white noise image. The style features and content features are extracted by five hybrid cavity convolutions and the obtained feature maps are labeled as S^l and C^l, respectively, maintaining S^l and C^l obtained by the fourth hybrid dilated convolution to ensure that the random white noise, r, is not disturbed by the content image while obtaining enough style features.

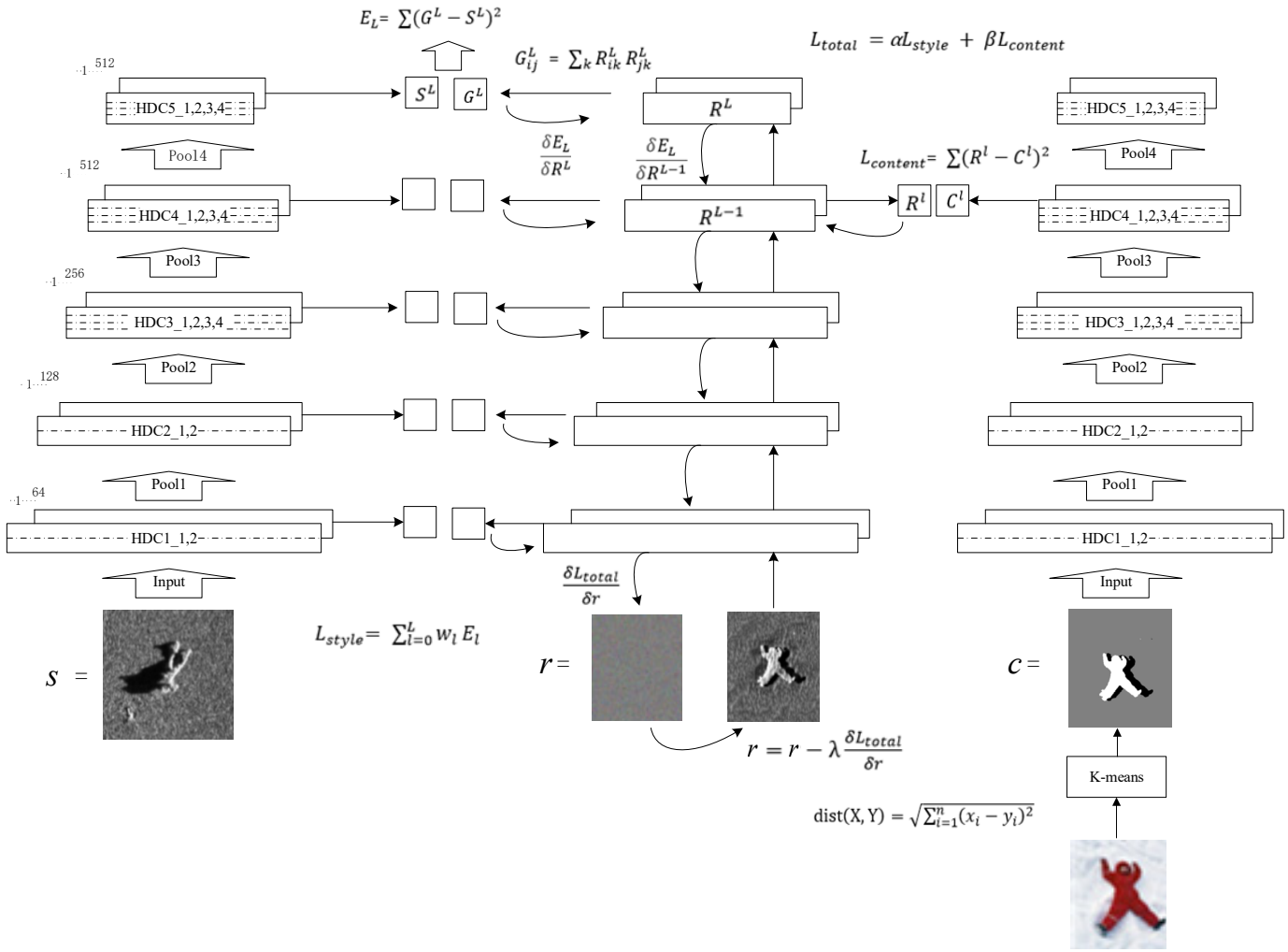


Figure 4. Improved style transfer network architecture in which the normal convolution is replaced by a hybrid dilated convolution in the network architecture, and a series of operations shown in Figure 3 are performed before inputting the image.

4. Experiment

4.1. Synthetic Data Ablation Experiment

To verify the effect of the K-means algorithm and hybrid dilated convolution on the model performance, an ablation experiment was designed in this study. The model proposed in this paper was tested for transfer learning in the same experimental environment as the model proposed by Gatys et al. [19]. The image set was composed of an optical

image, a fused optical image and two side-scan sonar images (as shown in Figure 5); the first column was the optical image as the content image and the fused optical image, and the second column was the side-scan sonar image as the style image. Three style transfer experiments were performed on the content images, and the experimental results are shown in Figure 5. The third column shows the effect of the model proposed by Gatys et al., and the fourth column shows the effect of the model proposed in this paper. From the first and second rows, it can be seen that the introduction of the K-means algorithm effectively solves the loss of image content edge details after style transfer, and the images are more similar to the real side-scan sonar images. From the second and third rows, it can be seen that the hybrid dilated convolution captures more detailed local features without increasing the parameters, which makes the graphics clearer and more natural.

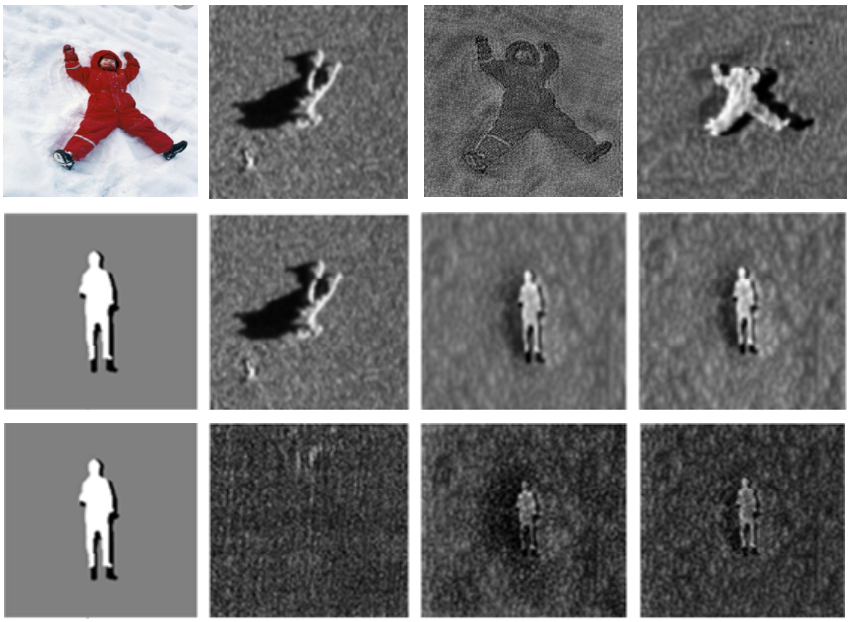


Figure 5. Results of the ablation experiments: the first row is the direct style transfer of the optical image; the second and third rows are the style transfer after optical image processing.

4.2. Experiment Based on Transfer Learning and "Simulated Side-Scan Sonar Images"

4.2.1. Dataset

The data in this paper were derived from a portion of the Seabed Objects-KLSG dataset established by Huo G. et al. [30], and also includes a portion of the data collected in this paper. The dataset was divided into four categories, i.e., shipwreck, drowning victim, aircraft, and seafloor. Table 1 shows the number of images per category in the dataset. Some of the images are shown in Figure 6.

Table 1. The number of images per category in the dataset.

Categories	Drowning Victim	Aircraft	Seafloor	Shipwreck
Numbers	18	62	289	385

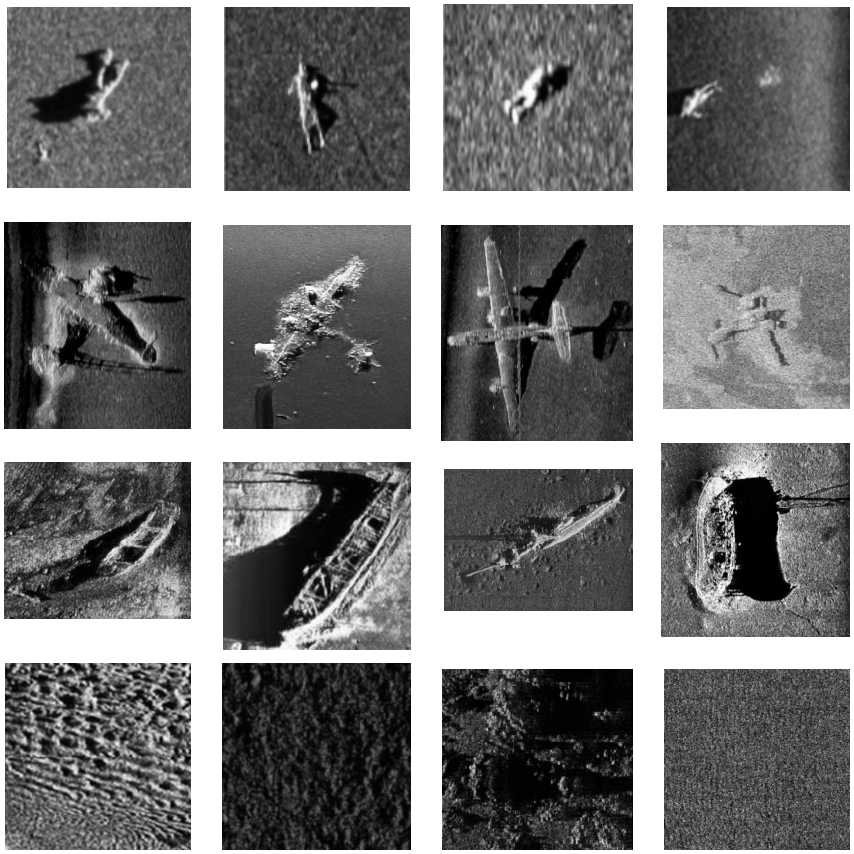


Figure 6. Real side-scan sonar images: the first row are drowning victim images; the second row are the aircraft images.

Due to the small number of images containing specific targets in the side-scan sonar images, the model could not obtain enough features to train the weights. In this paper, 54 images of victims and 60 images of aircraft as "simulated side-scan sonar images" were generated by the style transfer network. Some of the images are shown in Figure 7.

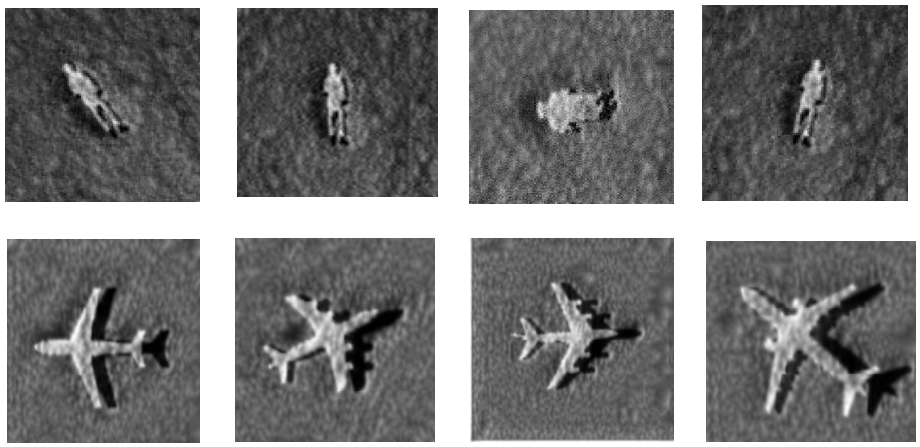


Figure 7. Similar side-scan sonar images: the first row are images of the drowning victim; the second row are images of the airplane.

4.2.2. Experimental Environment

In order to verify that the methods using transfer learning and "simulated side-scan images" are effective, this paper presents a comparative analysis of VGG16 [31], VGG19 [32], and resnet18 [33].

To save the training time of the CNN model, a deep learning model that had been pre-trained on ImageNet was downloaded. For each category in the side-scan sonar dataset, the training set accounted for 70% of the total, while the remaining 30% were in the test set. In order to test the effectiveness of the "simulated side-scan sonar images", they were put into the training samples for training. The initialization of model parameters had certain randomness; therefore, it affected the classification results to a certain extent. The training process was repeated five times, and the average value was used as the classification accuracy.

The choice of hyper-parameters is also particularly important when training networks in order to achieve better results. In this paper, an adaptive learning rate adjustment (Adadelta) optimizer was used: the initial learning rate was set to 0.01 and the remaining parameters were default parameters; 32 batches were used. The experiment was terminated when the training had completed 100 epochs.

4.2.3. Results

In this study, two methods were used to train the model: one was to only use the real side-scan sonar image training set for training; the other was to train a mixed training set of real side-scan sonar images and "simulated side-scan sonar images", using the real images for verification. The "simulated side-scan sonar images" in the training set consisted of 54 images of victims and 60 images of aircraft. Figure 8 shows a comparison of the five classification results. The performance of all methods was evaluated by testing the overall accuracy (OA). The classification accuracy values are shown in Table 1.

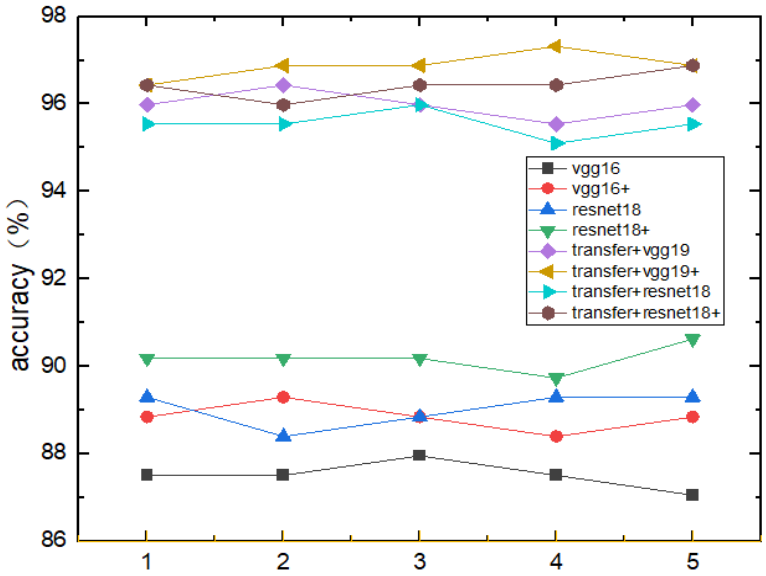


Figure 8. Plots of the five classification results. Transfer + denotes transfer learning, and + denotes using "simulated side-scan sonar images" in the training set.

As shown in Figure 8 and Table 2, after five repeated experiments, vgg16 and resnet18 could achieve certain results in sonar image classification, with accuracies of 87.50% and 89.02%, respectively. It was proven that CNN is suitable for the classification of side-scan sonar images. However, there were a large number of parameters in the neural network. When small sample datasets are classified, the network parameters cannot be fully trained, and it is easy to overfit them, which affects the final classification accuracy. However, by adding "simulated side-scan sonar images" to expand the data sample, the data shortages can be made up to a certain extent. Therefore, the CNN can extract enough features to train the parameters, avoid falling into overfitting, and improve the classification accuracy. When vgg16 and resnet18 used "simulated side-scan sonar images" for classification, the accuracy values were 88.84% and 90.18%, respectively. The classification accuracy improved slightly,

which proves the effectiveness of the "simulated to side scan sonar images" in classification. Although there were some differences between the "simulated side-scan sonar images" and real side-scan sonar images, the "simulated side-scan sonar images" could still be used for training. This is mainly because a "simulated side-scan sonar image" retains the main contour of the target object and conforms to the feature distribution of a real side-scan sonar images.

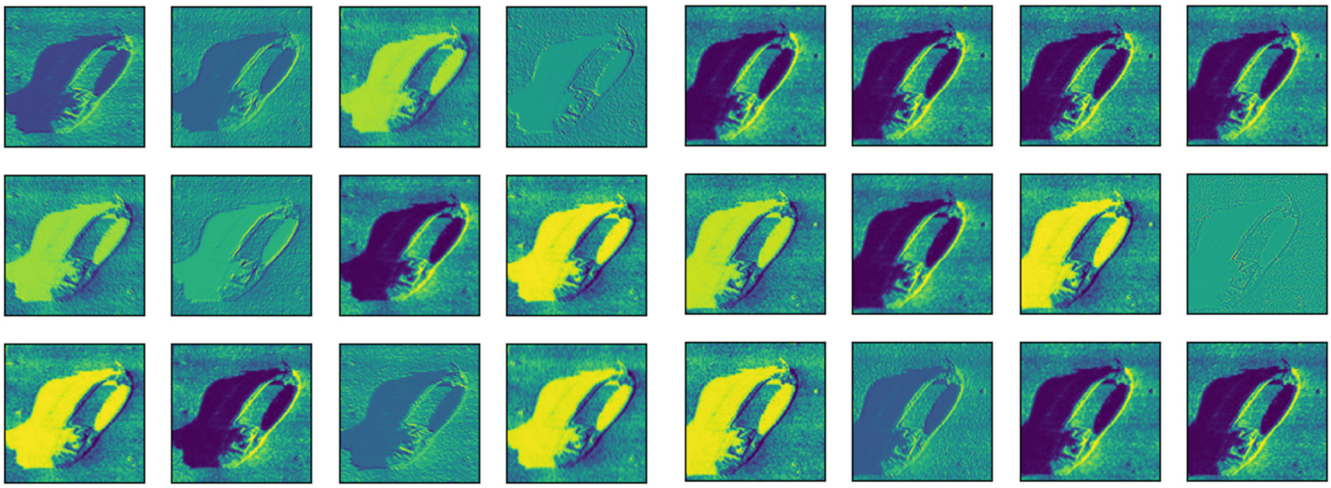
Table 2. Comparison of the average results of different models using the real data training set, and real data and the synthetic data training set, from five repetitions.

Methods	OA (%) Using Real Data Only	OA (%) Using Real Data and Synthetic Data
VGG16	87.50%	88.84%
Resnet18	89.02%	90.18%
Transferred VGG19	95.98%	96.88%
Transferred Resnet18	95.54%	96.43%

At the same time, as shown in Figure 8 and Table 2, the accuracy of using pre-trained vgg19 and pre-trained resnet18 networks in a classification task was 95.98% and 95.54%, respectively, and the accuracy of the vgg16 and resnet18 networks was 87.50% and 89.02%, respectively. The accuracy of classification using a pre-trained neural network was much higher than that without a pre-trained neural network. It can be proved that the classification accuracy in the task of side-scan sonar image classification can be significantly improved by transfer learning. The reason is that the CNN was trained on an ImageNet dataset. Due to the large number of data samples, the network parameters were fully trained, and a large number of middle- and low-level features were learned. These features are universal in images. When the side-scan sonar images are classified by a pre-trained network, the features of the side-scan sonar images can be extracted only by fine-tuning the parameters of a small number of side-scan sonar images, so as to complete the classification task quickly and efficiently. For the deep neural network without pre-training, there are many different weights to learn, and the weights cannot be fully trained by the small-sample dataset, resulting in poor feature extraction ability; therefore, it cannot distinguish some complex-shaped images. In order to verify that features can be better extracted by using pre-trained CNNs, the features of the first 12 channels of vgg19 and the first convolution layer of pre-trained vgg19 are visualized in Figure 9. In comparison, more feature information is obtained from the channel in Figure 9b, and the texture structure is clearer. The experimental results show that the pre-trained CNN has better feature extraction ability in the side-scan sonar image.

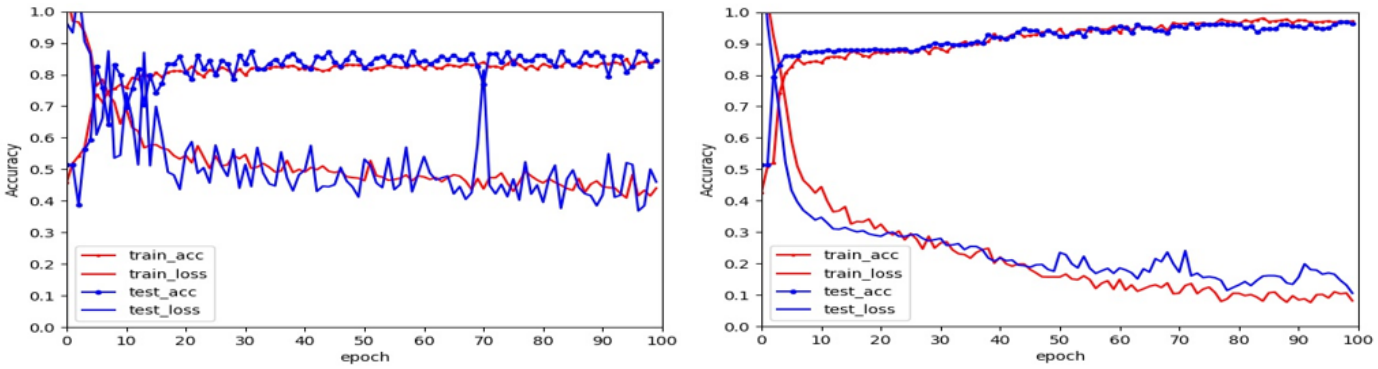
In order to further explore the classification ability of "simulated side-scan sonar images" and pre-trained CNN on side-scan sonar images, the processes of training vgg16 and pre-training vgg19 using real side-scan sonar images are given in Figure 10, respectively. From the comparison of Figure 10a,b, it can be seen that the vgg16 network trained with real side-scan sonar images and the network model oscillated and failed to converge. On the contrary, the vgg19 network had a deeper network, more parameters, and was more difficult to train. Through the use of transfer learning, the network is more stable, and its accuracy is substantially improved. Therefore, a pre-trained CNN can not only improve the classification accuracy, but also complete classification tasks quickly and stably. However, imbalances of the real side-scan sonar image samples may still lead to more classification errors, which can be improved by using "simulated side-scan sonar images". Figure 11 shows the processes of training resnet18, pre-trained resnet18 with real side-scan sonar images, and pre-trained resnet18 with real side-scan sonar images and "simulated side-scan sonar images". From the comparison of Figure 11a–c, it can be seen that Figure 11c shows

a more stable network performance and a substantial improvement in accuracy compared with Figure 11a. The convergence rate of the network is faster compared with b. This further verifies that the "simulated side-scan sonar images" combined with the pre-trained CNN has better accuracy and effectiveness. At the same time, it can highlight the target contour, and the extracted features are clearer, which is more conducive to classification, as shown in Figure 9. Therefore, the proposed pre-trained CNN combined with the "simulated side-scan sonar images" can solve the problem of overfitting caused by insufficient data samples, improve the classification accuracy, and have certain practicability and effectiveness in the classification of side-scan sonar images.



(**a**) (**b**)

Figure 9. The first 12 channel images in the first convolution layer of (**a**) vgg19 and (**b**) pre-trained vgg19.



(**a**) (**b**)

Figure 10. (**a**) Accuracy and loss function of training sets and validation sets of the vgg16 network model trained with real side-scan sonar images. (**b**) Accuracy and loss function of the training sets and validation sets of the pre-trained vgg19 network model fine-tuned with real side-scan sonar images.

In order to verify that "simulated side-scan sonar images" can improve the classification accuracy when training the model, rather than the accuracy improvement caused by the correct recognition of other categories, this study used real side-scan sonar images and "simulated side-scan sonar images" to fine-tune the vgg19 model, and then derive the confusion matrix of the training set; the results are shown in Tables 3 and 4.

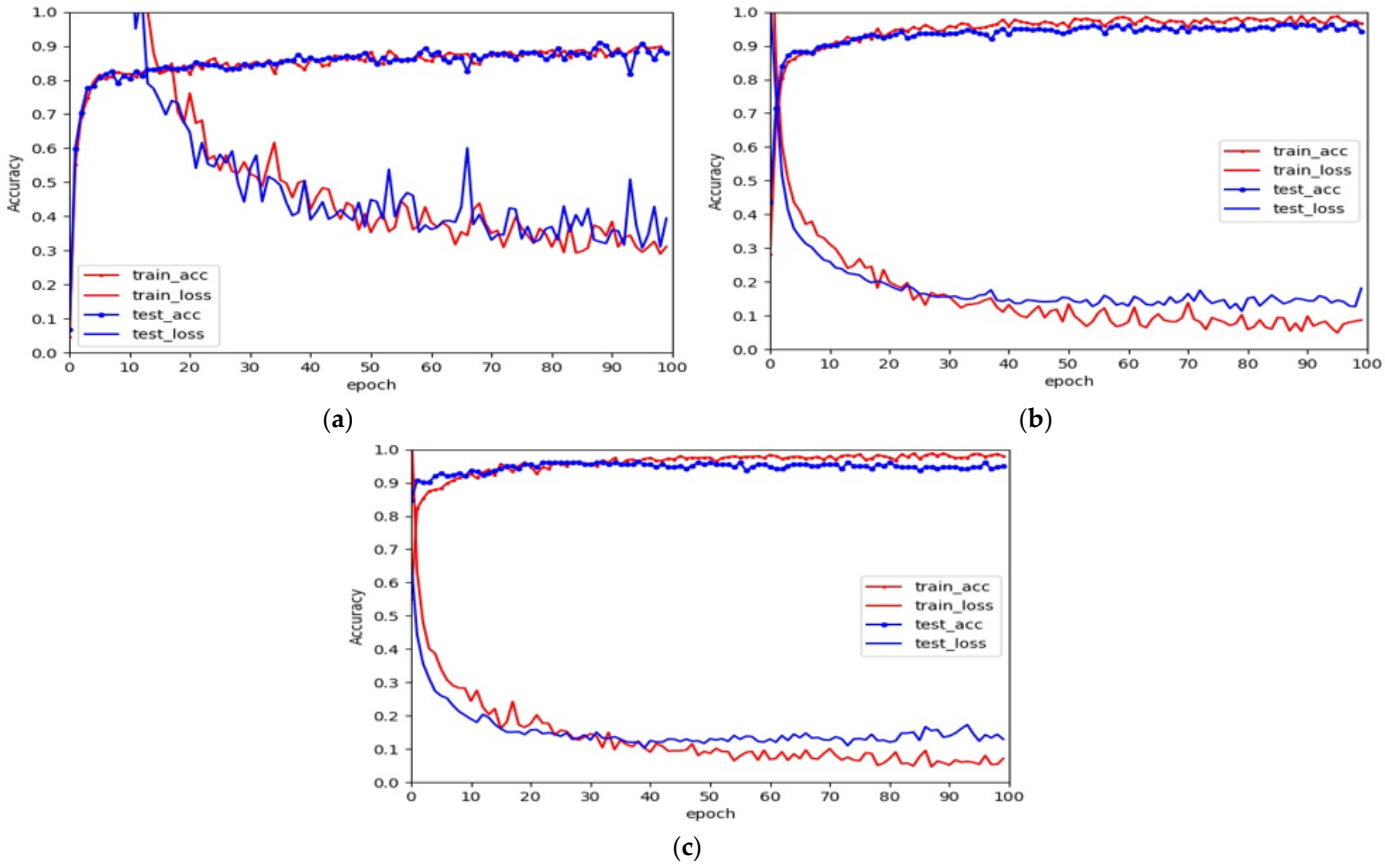


Figure 11. (**a**) Accuracy and loss function of the training sets and validation sets of the resnet18 network model trained with real side-scan sonar images. (**b**) Accuracy and loss function of the training and validation sets of the pre-trained resnet18 network model fine-tuned with real side-scan sonar images. (**c**) Accuracy and loss function of the training sets and validation sets of the pre-trained resnet18 network model fine-tuned with real side-scan sonar images and "simulated side-scan sonar images".

In Tables 3 and 4, the diagonal lines are the predicted categories that are the same as the true categories, which are marked in blue, and the categories marked in red are the prediction errors.

Table 3. Confusion matrix derived from fine-tuning the pre-trained vgg19 network using real side-scan sonar images.

True Class	Predicted Class			
	Drowning Victim	**Aircraft**	**Seafloor**	**Shipwreck**
Downing Victim	5	0	0	0
Aircraft	0	11	0	7
Seafloor	0	0	85	1
Shipwreck	0	1	0	114

Table 4. Confusion matrix derived from fine-tuning the pre-trained vgg19 network using real side-scan sonar images and "simulated side-scan sonar images".

True Class	Predicted Class			
	Drowning Victim	**Aircraft**	**Seafloor**	**Shipwreck**
Downing Victim	5	0	0	0
Aircraft	0	13	0	5
Seafloor	0	0	85	1
Shipwreck	0	1	0	114

According to Table 2, the accuracy of classification without using "simulated side-scan sonar images" and using "simulated side-scan sonar images" was 95.98% and 96.88%, respectively. It can be seen from Tables 3 and 4 that when the "simulated side-scan sonar images" were unused or used, the numbers of correct and incorrect predictions of shipwreck and the seafloor did not change, whereas the number of correct predictions of airplanes increased, which indicates that the improvement in accuracy is not affected by shipwreck and seafloor images. The improvement in accuracy comes from an increase in the number of correct classifications of airplane images. However, there was no change in the number of drowning victim images. The reason is that the structure of drowning images is relatively simple and only has five prediction samples. When the side-scan sonar images were classified by a pre-trained network, the whole network model could be fine-tuned by a small number of training samples, which shows the effectiveness of the pre-trained network on small-sample datasets. The above experiments show that the neural network model trained by the training set containing "simulated side-scan sonar images" can improve the classification accuracy to a certain extent, which proves the practicability and effectiveness of "simulated side-scan sonar images".

However, there were still five airplane images and one seafloor image that were misclassified as a wreck image, as shown in Figure 12.

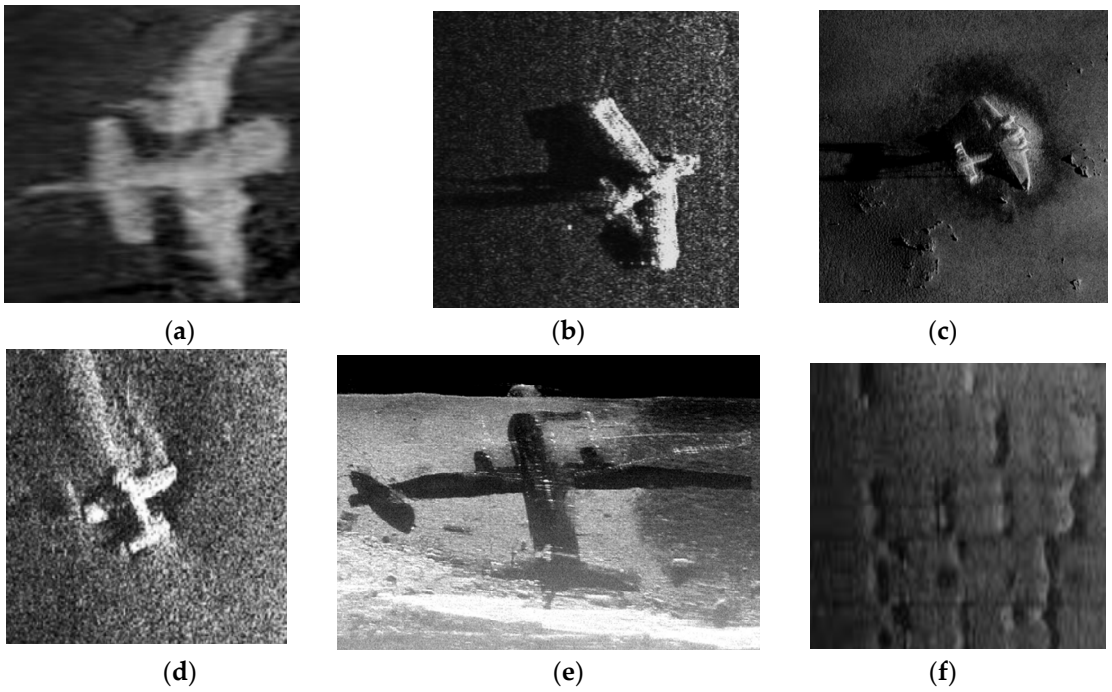
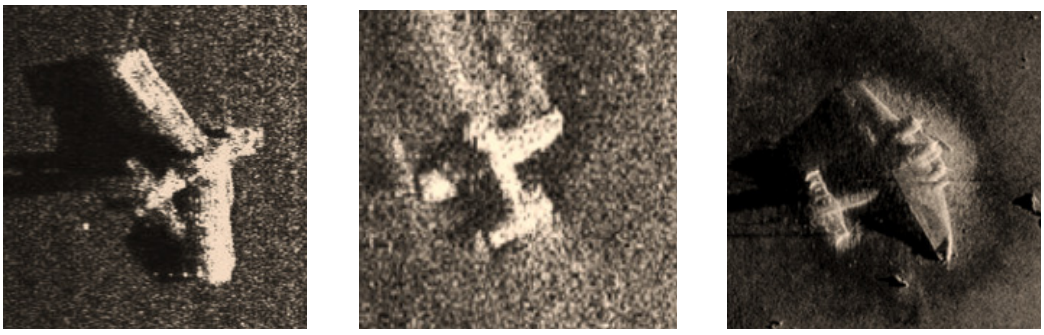


Figure 12. (**a**) Aircraft with fuzzy features. (**b**–**d**) Aircraft with smaller target object. (**e**) Aircraft with missing tail. (**f**) Pure seabed image.

In Figure 12a, the airplane features are fuzzy, whereas in Figure 12e, the tail part of the airplane is missing, which is not conducive to feature extraction and causes classification errors. In Figure 12f, the shapes of sea bottom rocks are similar to the wreck, and it could easily be misclassified as a wreck in classification. Although the airplane features in Figure 12b–d are clear, there were also classification errors. In order to further study the cause of the classification error, we redefined the size of the target object in the airplane images, as shown in Figure 13. In this study, the weights trained in Tables 3 and 4 were used to predict in Figure 13a–c, and the results are shown in Table 5.



(**a**) (**b**) (**c**)

Figure 13. (**a**–**c**) Images after redefining the size of the aircraft.

Table 5. Using the trained weights in Tables 3 and 4 to predict Figure 12a–c.

Airplane	Predicted Class	
	Table 3	Table 4
a	ship	airplane
b	ship	airplane
c	ship	airplane

It can be seen from Table 5 that the network used in Table 4 classified the airplane correctly, whereas the network used in Table 3 classified it incorrectly. This further proves the effectiveness of the "simulated side-scan sonar images". At the same time, it proved that the ability to extract features is insufficient when using the "simulated side-scan sonar images" and the pre-trained CNN to classify the small target with a complex structure and the side-scan sonar image with fuzzy image features, which eventually leads to a classification error.

In summary, the proposed method of generating side-scan acoustic data based on a style transfer network is not strict, but it can still effectively solve the problem of inaccurate classification due to insufficient training samples, mainly because the "simulated side-scan sonar images" retain the main contours of the target object and conform to the feature distribution of real side-scan sonar images. The "simulated side-scan sonar images" have also been well received by peers in the field. However, there is still a problem of insufficient feature extraction in the classification of small targets with complex structures and side-scan sonar images with fuzzy features.

5. Conclusions

In this paper, a transfer-learning-based method for the automatic classification of side-scan sonar images has been proposed. Experiments show that the use of a pre-trained CNN network model can effectively solve the overfitting problem caused by small data samples and greatly improve the classification accuracy. In addition, this paper also proposes a convolutional-neural-network-based method of side-scan sonar image generation to solve difficult acoustic problems effectively at a lower cost. It has been verified that training a classification network model using simulated side-scan sonar images can effectively

improve the classification accuracy, and the model converges rapidly and stably. Moreover, the simulated side-scan sonar images are highly similar to the real side-scan sonar images. At the same time, it also shows that for datasets with uneven distributions of data samples, transfer learning can still classify the targets correctly, and the classification accuracy of the targets can be further improved by adding "simulated side-scan sonar images". However, there is still a problem of insufficient feature extraction ability for the classification of side-scan sonar images with small targets and complex structures. In future studies, we will further test side-scan sonar images by combining the strong feature-extraction network GoogleNet and deeper ResNet.

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

Funding: This work was supported by grants from the National Basic Research Program of China [grant number 2019YFE0126600], the Major Project of Science and Technology of Henan Province [grant number 201400210300], the National Natural Science Foundation of China [grant numbers U1704122], and Key Research and Promotion Projects of Henan Province [grant numbers 212102210393, 202102110121].

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

CNN	convolutional neural networks
MRF	Markov random field
SVM	support vector machine
PCA	principal component analysis
ELM	extreme learning machine
KNN	k-nearest neighbor attractor neural network

References

1. Key, W.H. Side scan sonar technology. In Proceedings of the OCEANS 2000 MTS/IEEE Conference and Exhibition, Providence, RI, USA, 11–14 September 2000; IEEE: New York, NY, USA, 2000; Volume 2, pp. 1029–1033.
2. Klein, M. Side Scan Sonar. In *International Handbook of Underwater Archaeology*; Springer: Boston, MA, USA, 2002; pp. 667–678.
3. Klaucke, I. Side Scan Sonar. In *Submarine Geomorphology*; Springer: Cham, Switzerland, 2018; pp. 13–24.
4. Sadjadi, F.A. Studies in Adaptive Automated Underwater Sonar Mine Detection and Classification—Part 1: Exploitation Methods. In *Automatic Target Recognition XXV*; International Society for Optics and Photonics: Bellingham, WA, USA, 2015; Volume 9476, p. 94760K.
5. Zhai, H.; Jiang, Z.; Zhang, P.; Tian, J.; Liu, J. Underwater object highlight segmentation in SAS image using Rayleigh mixture model. In Proceedings of the 2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 27–29 November 2015; pp. 418–423.
6. Reed, S.; Petillot, Y.; Bell, J. An automatic approach to the detection and extraction of mine features in side scan sonar. *IEEE J. Ocean. Eng.* **2003**, *28*, 90–105. [CrossRef]
7. Ye, X.-F.; Zhang, Z.-H.; Liu, P.X.; Guan, H.-L. Sonar image segmentation based on GMRF and level-set models. *Ocean Eng.* **2010**, *37*, 891–901. [CrossRef]
8. Kumar, N.; Mitra, U.; Narayanan, S.S. Robust object classification in underwater side scan sonar images by using reliability-aware fusion of shadow features. *IEEE J. Ocean. Eng.* **2014**, *40*, 592–606. [CrossRef]
9. Sinai, A.; Amar, A.; Gilboa, G. Mine-like objects detection in side-scan sonar images using a shadows-highlights geometrical features space. In Proceedings of the OCEANS 2016 MTS/IEEE Monterey, Monterey, CA, USA, 19–23 September 2016; pp. 1–6.
10. Martin, A. Comparative study of information fusion methods for sonar images classification. In Proceedings of the 2005 7th International Conference on Information Fusion, Philadelphia, PA, USA, 25–28 July 2005; Volume 2, p. 7.

11. Quidu, I.; Malkasse, J.P.; Burel, G.; Vilbe, P. Mine classification based on raw sonar data: An approach combining Fourier descriptors, statistical models and genetic algorithms. In Proceedings of the OCEANS 2000 MTS/IEEE Conference and Exhibition. Conference Proceedings (Cat. No. 00CH37158), Providence, RI, USA, 11–14 September 2000; Volume 1, pp. 285–290.
12. Karine, A.; Lasmar, N.; Baussard, A.; El Hassouni, M. Sonar image segmentation based on statistical modeling of wavelet subbands. In Proceedings of the 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA), Marrakech, Morocco, 17–20 November 2015; pp. 1–5.
13. Guo, J.; Jin-Feng, M.A.; Wang, A.X. Study of Side Scan Sonar Image Classification Based on SVM and Gray Level Co-Cccurrence Matrix. *Geomat. Spat. Inf. Technol.* **2015**, *68*, 60–63.
14. Zhu, M.; Song, Y.; Guo, J.; Feng, C.; Li, G.; Yan, T.; He, B. PCA and kernel-based extreme learning machine for side-scan sonar image classification. In Proceedings of the 2017 IEEE Underwater Technology (UT), Busan, Korea, 21–24 February 2017; pp. 1–4.
15. Williams, D.P. Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 2497–2502.
16. Dobeck, G.J.; Hyland, J.C. Automated detection and classification of sea mines in sonar imagery. Detection and Remediation Technologies for Mines and Minelike Targets II. *Int. Soc. Opt. Photonics* **1997**, *3079*, 90–110.
17. Kim, J.; Cho, H.; Pyo, J.; Kim, B.; Yu, S.-C. The convolution neural network based agent vehicle detection using forward-looking sonar image. In Proceedings of the OCEANS 2016 MTS/IEEE Monterey, Monterey, CA, USA, 19–23 September 2016; pp. 1–5.
18. Phung, S.L.; Nguyen, T.N.A.; Le, H.T.; Chapple, P.B.; Ritz, C.H.; Bouzerdoum, A.; Tran, L.C. Mine-like object sensing in sonar imagery with a compact deep learning architecture for scarce data. In Proceedings of the 2019 Digital Image Computing: Techniques and Applications (DICTA), Perth, WA, Australia, 2–4 December 2019; pp. 1–7.
19. Gatys, L.A.; Ecker, A.S.; Bethge, M. A neural algorithm of artistic style. *arXiv* **2015**, arXiv:1508.06576. [CrossRef]
20. Wang, P.; Chen, P.; Yuan, Y.; Liu, D.; Huang, Z.; Hou, X.; Cottrell, G. Understanding convolution for semantic segmentation. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1451–1460.
21. Zaki, M.J.; Meira, W., Jr.; Meira, W. *Data Mining and Analysis: Fundamental Concepts and Algorithms*; Cambridge University Press: Cambridge, UK, 2014.
22. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
23. Oquab, M.; Bottou, L.; Laptev, I.; Sivic, J. Learning and transferring mid-level image representations using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1717–1724.
24. Winnemöller, H.; Olsen, S.C.; Gooch, B. Real-time video abstraction. *ACM Trans. Graph. (TOG)* **2006**, *25*, 1221–1226. [CrossRef]
25. Hertzmann, A.; Jacobs, C.E.; Oliver, N.; Curless, B.; Salesin, D.H. Image analogies. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, 12–17 August 2001; pp. 327–340.
26. Ashikhmin, N. Fast texture transfer. *IEEE Comput. Graph. Appl.* **2003**, *23*, 38–43. [CrossRef]
27. Lee, H.; Seo, S.; Ryoo, S.; Yoon, K. Directional texture transfer. In Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering, Annecy, France, 7 June 2010; pp. 43–48.
28. Gatys, L.A.; Ecker, A.S.; Bethge, M. Texture synthesis using convolutional neural networks. *arXiv* **2015**, arXiv:1505.07376.
29. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
30. Huo, G.; Wu, Z.; Li, J. Underwater Object Classification in Sidescan Sonar Images Using Deep Transfer Learning and Semisynthetic Training Data. *IEEE Access* **2020**, *8*, 47407–47418. [CrossRef]
31. Liu, B.; Zhang, X.; Gao, Z.; Chen, L. Weld Defect Images Classification with Vgg16-Based Neural Network. In *International Forum on Digital TV and Wireless Multimedia Communications*; Springer: Singapore, 2017; pp. 215–223.
32. Subetha, T.; Khilar, R.; Christo, M.S. A comparative analysis on plant pathology classification using deep learning architecture–Resnet and VGG19. *Mater. Today Proc.* **2021**. [CrossRef]
33. Su, F.; Sun, Y.; Hu, Y.; Yuan, P.; Wang, X.; Wang, Q.; Li, J. Development and validation of a deep learning system for ascites cytopathology interpretation. *Gastric Cancer* **2020**, *23*, 1041–1050. [CrossRef] [PubMed]