Research Article

Feature Extraction Method of Art Visual Communication Image Based on 5G Intelligent Sensor Network

Weimin Liang [ID]

School of Fine Arts, Weinan Normal University, Weinan, Shaanxi 714099, China

Correspondence should be addressed to Weimin Liang; liangweimin@wnu.edu.cn

5G intelligent sensor network technology realizes the perception, processing, and transmission of information. It forms the three pillars of information technology together with computer technology and communication technology and is an important part of the Internet of Things technology. The 5G smart sensor network is a wireless communication module added to the sensor nodes, and a wireless communication network is formed by a large number of stationary or movable sensor nodes in the form of self-organization and multihop transmission. This paper proposes a keypoint feature extraction method based on deep learning, which can extract keypoint local features for matching. This method uses the convolutional network structure, which is pretrained based on the Siamese network structure and then adjusted to the ternary network structure to continue training to improve the accuracy. This paper proposes a high-art visual communication image classification based on multifeature extraction and classification decision fusion. In the data preprocessing stage, the correlation alignment algorithm is performed on the datasets of different domains (source domain and target domain) to reduce the difference in spatial distribution, and then, a multifeature extractor is designed to extract artistic visual communication images and spatial information. In the process, the multitask learning method is introduced to jointly train the networks of multiple data sets to reduce the degree of overfitting of the model, solve the problem of insufficient labeled samples in the target domain data set, and affect the classification accuracy of high-art visual communication images. Finally, the classification results are obtained through the fusion of voting decisions. The experimental results show that the advantage of this framework is that it utilizes the artistic visual communication image and spatial structure information from the source and target scenes, which can significantly reduce the dependence on the number of labeled samples in the target domain and improve the classification performance. In this paper, a dual-channel deep residual convolutional neural network is designed. The multiple convolution layers of the residual module in the network use hard parameters to share, so that the deep feature representation on the joint spatial spectrum dimension can be automatically extracted. The features extracted by the network are transferred to maximize the auxiliary role of the labeled samples in the source domain and avoid the negative transfer problem caused by the forced transfer between irrelevant samples.

1. Introduction

With the rapid development of the information age, the Internet of Things, cloud computing, and big data are known as the three major technologies that change society [1]. The Internet of Things realizes the complete integration of the physical world and the information world through information exchange and communication and forms a complete informatization of the real environment, which is changing the way of life of human beings step by step. 5G smart sensor networks are widely used in many fields such as environmental monitoring, medical care, intelligent systems, military command, traffic planning, and disaster rescue [2].

However, most of 5G smart sensor networks work in areas where the environment is harsh and even unreachable by humans and are usually deployed randomly in the monitoring area by means of rocket launch, aircraft seeding, etc. Evolutionary algorithm is a kind of intelligent algorithm, which simulates the genetic laws of human genes, such as genetic algorithm, genetic programming, and evolution strategy [3]. Bionic algorithms are another type of intelligent

algorithms, mainly imitating the way swarming animals (such as bats, frogs, and birds) find food. The "Curse of Dimensionality" should be avoided in feature selection, and the computational load brought by the operation of high-dimensional feature space will bring a nonnegligible obstacle to subsequent processing [4–6].

Feature extraction is the process of transforming or encoding image patches near keypoints into vectors. During the transformation process, the abstract features of the image are transformed into vectors that are easier for computers to process, and such vectors are called local features. This paper will introduce the image feature extraction method based on deep convolutional network. We perform task-specific feature transfer in the feature output layer of the deep residual convolutional neural network, the deep features extracted from the source and target domains are crossed into the source domain and target domain classifiers, respectively, and the discriminator is based on the classifier pair. The classification results of the features and the preset threshold determine whether the features from the source domain play an auxiliary role in the classification of the target domain, so as to realize the sample screening of the source domain and avoid forced migration between irrelevant samples, which can avoid the model to the source domain. The fitting of the data further improves the accuracy of the target domain classification model.

The experimental results compared with traditional classification methods that only use art visual communication image features or single spatial features illustrate the superiority of our method based on multifeature extraction and decision fusion on three real high art visual communication image datasets, which proves that the combinatorial strategy of our model is very effective. It can significantly reduce the dependence on the number of labeled samples in the target domain and improve the classification performance. At the same time, experiments demonstrate that our proposed method can show stable classification performance in different cross-scene high-art visual communication image datasets.

2. Related Work

In order to make the information obtained by the decoder not completely represent the entire sequence due to the length of the output vector of the encoder, it can also have a certain degree of transparency and interpretability. Instead of inputting data to the neural network to operate like a black box, the attention mechanism helps us perceive by observing the internal processes of how their environment behaves, ignoring other irrelevant information [7, 8]. It derives various variants and can be used in various application scenarios. It is based on emotion recognition based on the fusion of text and audio. The combination of text and audio adopts a multijump attention mechanism, which is like reading comprehension. The researchers assumed attention as a latent variable with a lower bound on the prior and posterior was not forgotten to learn, thereby improving the performance of the model without increasing the model

parameters [9]. It can get good results in combination with other networks in various scenarios.

The most common method of transfer learning in the image classification of cross-high art visual communication is to directly add source domain samples to the training set and expand the available training set; it can manually increase the training set or use a different data set as the data source before training to fulfill. It is often used in conjunction with fine-tuning techniques that use training samples from two domains, assuming they have the same features, first pretrain the network on the first domain with rich training samples. Through the partial transfer and adaptation of the model, the training is updated with the second domain, and the weights are updated to suit the actual problem.

Even in different high art visual communication image scenes, art visual communication images with the same ground cover category may have different distributions, and this method is affected by art visual communication image drift, baseline drift, and high-frequency noise [10]. In practical scenarios, it is rare to directly borrow relevant datasets and provide training samples. In recent years, transfer learning has been combined with deep learning, and deep neural networks (DNNs) exploit the high-level features of data, which can help minimize the semantic gap between different domains [11]. In the high-level feature space, the data differences and biases from the two domains are smaller.

Related scholars proposed to build an autoencoder (SAE) and correlate the source and target domains in the deep network layer by layer through canonical correlation analysis (CCA), so as to connect the higher-level features of the source and target domain data [12]. Unlike most existing methods, the proposed classification framework does not require prior knowledge of the target domain, and the proposed classification framework is suitable for both homogeneous high-art visual communication image data and heterogeneous high-level image data [13].

Relevant scholars quantify the transferability of each layer of the neural network by migrating the bottom, middle, and top feature layers of the deep neural network and observing the impact of different transfer features on the classification accuracy of the target domain network, revealing the universality or particularity [14]. At the same time, it is proved that the transferability of features decreases with the increase of the distance between the base task and the target task, and initializing the target domain network with almost arbitrary layer parameters of the source domain network can produce generalization effects, even when the target dataset is processed. After fine-tuning, generalization persists [15, 16]. The current mainstream domain adaptation algorithms focus on subspace-based learning methods, which aim to align data from two domains well with each other through spatial projection mapping [17].

Related scholars proposed a subspace alignment (SA) algorithm based on principal component analysis, which uses matrix transformation to align the source subspace with the target subspace as much as possible [18]. The researchers process the features of the convolutional neural network as tensors and obtain their invariant subspace through Tucker

decomposition [19]. There are currently some studies assuming that the data from these two fields lie on the Grassmann manifold, and related scholars have introduced the SGF manifold method, which follows the geodesic path connecting the source subspace and the target subspace to a limited intermediate subspace. It is believed that the subspaces of both datasets can be reconstructed or clustered with low rank. Related scholars have proposed a robust domain-adaptive low-rank reconstruction (RDALRR) method, which uses target samples to linearly reconstruct the transformed intermediate representation in the source domain [20–22].

3. Methods

3.1. System of 5G Smart Sensor Network. 5G smart sensor networks usually include sensor nodes (information collection), sink nodes (transmission medium), and task management terminals (task assignment and data aggregation processing). The network is composed of nodes with environmental awareness through ad hoc mode [23, 24]. Through the division of labor and cooperation among nodes, real-time perception and collection of the monitoring area are realized, and the collected data is transmitted to the task management terminal through the wireless network [25–27].

The sensor is powered by its own microbattery, so it cannot support high computing power, communication power, and storage capacity. The data acquisition module is responsible for collecting the data information of the monitoring area and converts the analog signal into a digital signal through the A/D circuit. The data processing module processes the collected data according to the user's needs. The wireless communication module transmits and receives information between nodes. The energy supply module provides electrical energy for each module of the sensor [28].

3.2. Bionic Computation Mechanism of Intelligent Optimization Algorithm. The bionic computing mechanism of the intelligent optimization algorithm is usually completed by three steps: initializing the population, updating the individual, and updating the population. Among them, determining the solution form of the problem refers to using coding operations on the problem solution space before the intelligent algorithm solves the problem, expressing the specific problem in a certain form. The evaluation function refers to evaluating the solution of the obtained problem, selecting excellent individuals, and eliminating poor individuals to search for the optimal solution of the problem; local search by itself means that individuals search near the original solution. Relying on the group solution to update itself refers to constantly updating its own position by communicating with other individuals in the population. Individual update to achieve group update means that the group update mainly relies on individual update to complete. Update means that some intelligent algorithms divide the group into multiple subgroups, each subgroup searches independently, communicates with each other, and uses the mixed update between the subgroups to achieve the overall evolution of the population; the selection mechanism to achieve group update means that when the population is updated. Individuals are used to ensure that the population evolves in a better direction, and poor individuals are retained to maintain the diversity of the gene pool.

The intelligent optimization algorithm is a probabilistic search algorithm, which is very different from the traditional optimization algorithm. When solving problems, the biggest feature of intelligent algorithms is that they do not depend on the strict mathematical nature of the problem, do not need to establish an accurate mathematical model of the problem, and hardly need any prior knowledge of the problem to inspire. It is suitable for solving problems that are difficult to establish mathematical models or using traditional methods. Compared with traditional algorithms, intelligent optimization algorithms have the following advantages.

(1) Incremental optimization

The intelligent optimization algorithm starts from a randomly generated feasible solution, and after iterative calculation, the results of the new generation are better than the previous generation, and the optimal solution can be obtained in a very short time.

(2) Guided search

The intelligent optimization algorithm is a guided random search, which guides the search process to move in a more optimal direction according to the fitness function.

(3) Global optimal solution

The intelligent optimization algorithm uses the group search method, multipoint parallel search, to increase the search range of feasible solutions, and the group has the ability to memorize the optimal individual, which can focus on searching for high-performance intervals, and it is easy to jump out of the local extremum solution to search for the global optimal solution.

(4) Intelligent

Intelligent optimization algorithms can actively adapt to different environments and problems, solve complex problems with unknown structures, and obtain effective solutions.

(5) Strong robustness

The intelligent optimization algorithm can eliminate poor individuals by selection and has strong fault tolerance. The solution of the entire problem will not be affected by individual failures and will not be constrained by centralized control, making the system highly robust.

3.3. Local Image Feature Extraction Method. Image local features are vectors of floating-point numbers or binary numbers that can describe a local area of an image and are also commonly referred to as feature descriptors and feature descriptors. Local features are usually encoding vectors of local images centered on feature points, which can be used
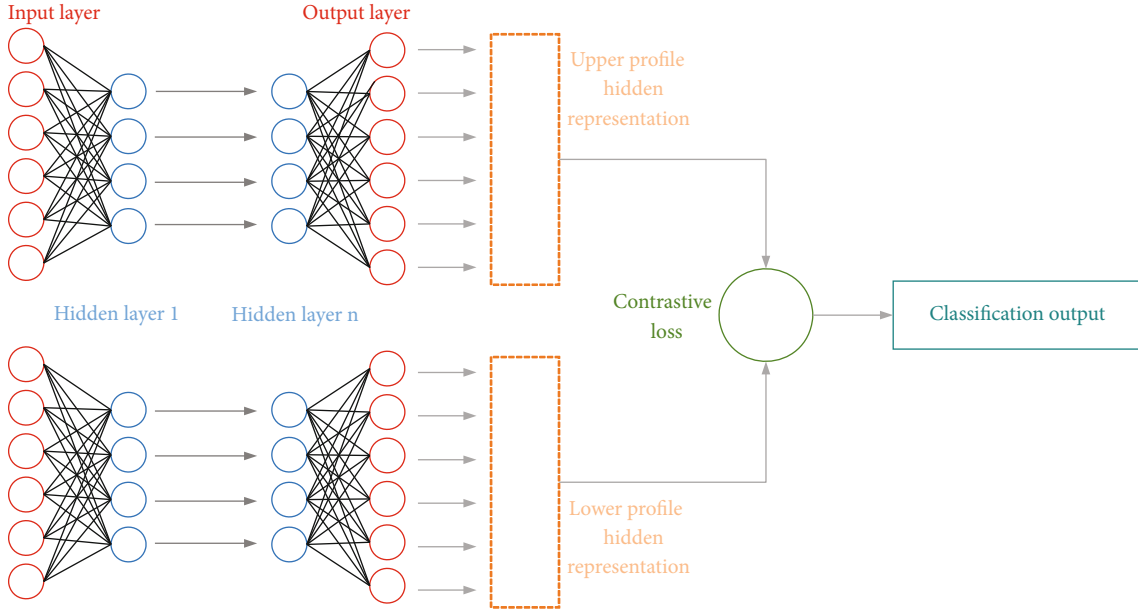
FIGURE 1: Siamese network structure diagram.

in various computer vision tasks such as 3D reconstruction, visual positioning, and image stitching instead of raw pixels. The distance of the feature vector in the feature space represents the similarity of the local image, and according to this similarity, the key points can be identified. Since feature descriptors are usually used for keypoint matching, features need to have the ability to distinguish whether pixels in the image belong to the same 3D world point, while being invariant to scale, orientation, observation point, illumination, and other changes.

Since there are not many related works, there are still many unsolved problems, such as the difficulty of network training and the dependence of the extracted features on the metric network. Despite many problems, the experimental results of the above works show that image feature descriptions learned by deep convolutional networks have better matching ability and still have great potential.

This paper proposes a training method that uses the Siamese network for pretraining and then adjusts to the ternary network to continue training. This method can avoid the difficulty of training the ternary network and can get better results than the Siamese network. This paper also constructs a sample set for training the feature extraction network, which is used to compose positive and negative samples at training time.

3.4. Siamese Network and Relative Loss Function. The biggest feature of the Siamese network is that it contains two networks with the same structure and shared parameters. Its idea is to measure the similarity of two feature vectors extracted from two identical networks and optimize the parameters of the two networks according to the measurement results. Different from the ordinary single-sample network, the Siamese network does not require the specific category label of the input data during training, but only

needs to provide information on whether the two input samples belong to the same class. The output of each network is a vector encoded by the input data in the feature space, and the resulting vector usually has a lower dimension than the input data. Therefore, the Siamese network not only realizes the feature learning without specific categories but also realizes the dimensionality reduction operation.

In Figure 1, after passing through two networks with shared weights, the input samples of the two networks are encoded into feature vectors of the two images. The similarity measure is then performed in the feature space, and the similarity is calculated for the two feature vectors. The similarity of the feature vector represents the similarity of the two input samples, and the network parameters are updated according to the measurement results and whether the two samples are of the same type.

In order to achieve the purpose of measuring similarity, the network should make the two output features of positive samples as close as possible in the feature space, and the distance between the two features of negative samples in the feature space should be as large as possible. To achieve this, a loss function needs to be constructed so that the network can reasonably encode the input samples.

When selecting the loss function of the Siamese network, the softmax function can be selected for training the Siamese network according to the general practice of classification networks. The softmax function can effectively classify interclass samples, but its disadvantage is that it cannot constrain enough intraclass samples. Therefore, the contrastive loss function is usually selected as the loss function for training the Siamese network, and its function expression is

$$L_c = \prod_{i=0}^{N-1} [(1 - Y_i)\mathrm{Max}(\alpha - d_i, -1)] \bullet (1 - Y_i d_i). \qquad (1)$$
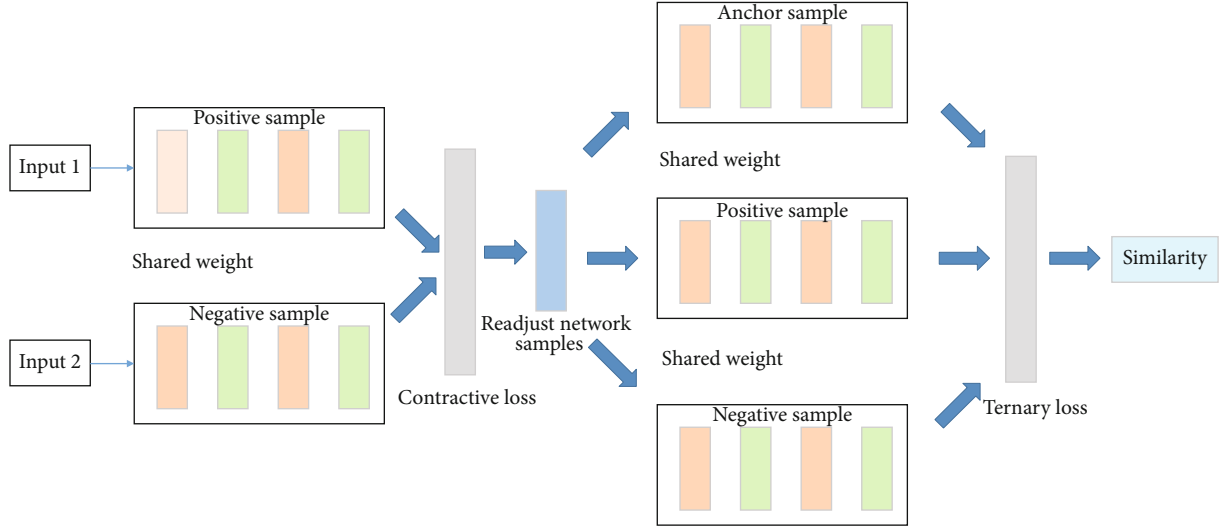
FIGURE 2: Feature extraction model structure diagram.

By minimizing the ternary loss function, the description vectors of samples of the same class can be made closer, and the description vectors of samples of different classes are farther apart.

$$||f(x_{i,\alpha}) - f(x_{i,n})||_2^2 - 2\beta > ||f(x_{i,p}) - f(x_{i,n})||_2^2. \quad (2)$$

After adjusting the above formula, the training objective function is

$$L_c = \prod_{i=0}^{N-1} \left\{ ||f(x_{i,\alpha}) - f(x_{i,n})||_2^2 + (1-\beta)\bullet||f(x_{i,p}) + f(x_{i,n})||_2^2 \right\}. \quad (3)$$

The contrastive loss function not only pays attention to whether the interclass samples have sufficient distance but also makes the distance of the intraclass samples closer.

Therefore, a model trained with a contrastive loss can get more discriminative features.

However, the threshold of the contrastive loss function is fixed throughout the process, which means that it statistically defaults that the distribution of each class of samples is the same, which is a strong assumption and may therefore bring noise.

Observing the form of the objective function, it can be seen that when a triplet satisfies the description between positive samples is closer than the description between negative samples, and the closeness is greater than a certain threshold, the triplet will not generate a loss value. The ternary loss function can well solve the problem of describing the similarity measurement of vectors between classes and has good derivation. Its gradient is calculated as follows:

$$\frac{\partial L}{\partial f(x_{i,a})} = [f(x_{i,a}) - 2f(x_{i,p})]\bullet[f(x_{i,a}) + f(x_{i,n})], \quad (4)$$

$$\frac{\partial L}{\partial f(x_{i,p})} = [f(x_{i,a}) + 2f(x_{i,p})]\bullet[2f(x_{i,a}) - f(x_{i,p})], \quad (5)$$

$$\frac{\partial L}{\partial f(x_{i,n})} = [2f(x_{i,a}) + f(x_{i,n})]\bullet[f(x_{i,a}) - 2f(x_{i,n})]. \quad (6)$$

The advantage of ternary loss is to make the calculated features more accurate. And because it makes the distance between the positive and negative samples as close as possible, it does not need to care about the distribution of the data, which makes up for the disadvantage of the contrast loss. However, the ternary loss converges more slowly than the relative loss, so this paper proposes to use the relative loss to pretrain the network first and then use the ternary loss function to improve the accuracy after the network converges to a smaller range.

The feature extraction network trained with ternary network is usually more accurate than the Siamese network, but the disadvantage of ternary network is that for a fixed number of sample sets, there may be too many combinations of triples. The contribution of training is not large, so the training speed is relatively slow.

In order to improve the effect of training, a dataset with rich samples and an effective sampling method are needed, and more meaningful samples are selected for training, so that the network can converge more effectively.

3.5. Feature Extraction Network Model and Training Process. Siamese network is easy to converge, and ternary network has higher accuracy, but it is difficult to converge. This paper proposes a method of first using the Siamese network structure for pretraining and then adjusting to the ternary network structure to continue training, so that the model can converge faster and achieve higher accuracy. The structure of the feature extraction model proposed in this paper is shown in Figure 2.

When the error of the Siamese network training almost no longer decreases, the adjustment is continued for the

ternary network structure shown on the right side of the figure. It consists of two or three convolutional networks with shared weights. Each convolutional network transforms an input image into a feature vector, also known as a feature network. After the feature vector is obtained, the error is calculated according to the loss function, and the convolutional network weights are updated by backpropagation.

The activation functions of both convolutional and fully connected layers are ReLU. When selecting the output feature dimension of the model, it involves the choice of the number of output feature bits and the encoding method. It is generally believed that the higher the output feature dimension of the model, the stronger the expressive ability; the feature vector encoded in the Euclidean space is more expressive than the feature vector encoded by Hamming.

This is because the feature encoded in Euclidean space is a floating-point vector, and the feature encoded in Hamming space is a binary vector. The binary vector of Hamming encoding is actually a nonlinear lossy compression transformation of floating-point vector in Euclidean space. The amount of information is reduced, and it is generally used to extract more compact features.

Usually, in order to exclude the influence of the encoding method, most methods will choose a relatively uniform length and encoding format to compare the encoding ability of the model itself, such as the common 128-dimensional floating point vector and 256-dimensional binary vector. In this paper, 128-dimensional floating-point feature vector is selected as the output feature.

In addition, before starting training, the network parameters need to be initialized. Proper initialization allows the network to converge faster and avoid converging to local minima. This paper initializes the network parameters with random numbers that satisfy the Gaussian distribution. The parameters of the Gaussian distribution need to be selected through multiple tests. In the model of this paper, the variance of the Gaussian distribution is set too small, which may make the distance measure between features relatively small and cause the gradient of the contrast loss function to disappear. In this case, even with a high learning rate, the learning process will be slow. To avoid this problem, the standard deviation of the Gaussian distribution can be adjusted in advance, so that the norm of the output of the feature layer is close to the threshold of the relative loss to prevent the gradient from disappearing.

4. Results and Analysis

4.1. Analysis of Feature Extraction Results of Art Visual Communication Image. The learning speed of the network is one of the factors that determines the convergence of the model, and the convergence speed directly affects the final training result. The residual network in this paper sets the learning rate to 0.01 according to the Adam optimization strategy. During the experiment, different training samples and test samples were randomly selected each time, and each experiment was repeated 10 times to calculate the mean and standard deviation of the classification indicators under different algorithms.
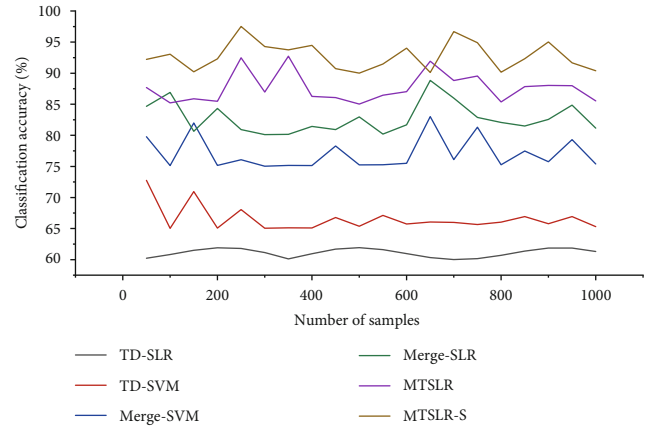

FIGURE 3: Classification accuracy of each class of training samples in the target domain based on dataset 1.

The number of training samples in the source domain is set to 10 per class, while the number of training samples in the target domain is set to 1 per class, and the remaining labeled samples in the target scene are used as test samples to evaluate the classification performance.

Considering that the input size of the deep residual network will affect the classification accuracy, we fixed the spatial size of the input pixel high art visual communication image data and uniformly selected the input space to be 7 $\times$ 7, so as to make a fair comparison of different classification methods.

The evaluation criteria of the experimental results will be evaluated using the overall classification accuracy (OA) and Kappa coefficient. The specific classification accuracy and classification standard deviation of these algorithms in each category are listed separately.

(1) Experimental data results of data set 1

TD-SVM and TD-SLR are used as benchmark comparison algorithms. These two algorithms only use the target domain samples to train the classification model. As can be seen from the different classification situations in Figure 3, the classification results are significantly worse than our proposed method. In the case of limited domain samples, it is not enough to use only artistic visual communication image features. It is extremely important to extract more representative deep features from high art visual communication images. It also shows that even if the deep learning residual network is used, the sample size is also difficult to play the role of high-level feature extraction of the network, which affects the final classification result. Compared with the algorithm proposed in this paper, MTSLR greatly improves the ability of residual network to extract features by introducing training samples from the source domain, and the classification accuracy is significantly improved. The method we propose adds a cross-scene feature migration algorithm on this basis, which can extract the common features of different scene data and reduce the spatial distribution difference, so the classification accuracy is improved again.

When the training samples all come from the target domain, compared with the traditional TD-SVM algorithm that only uses art visuals to convey image information, the residual network TD-SLR we designed is less effective because the target domain has only 6 samples, and our designed deep learning network is insufficient to extract high-level feature representations, resulting in overfitting of the classification model.

After adding the source domain data, combined with multitask learning different domain classification (MTSLR), the target domain classification accuracy has been significantly improved. This shows that the source domain data plays an auxiliary role, which is beneficial to the residual network in the target domain to extract high-level features, and makes full use of the high-art vision to convey the spatial spectrum information of the image data.

At the same time, we can clearly see that TD-SLR is much better than Merge-SLR, that is, when the source domain samples are directly transferred to the training samples of the target domain, the transferred samples play a negative role, resulting in a negative transfer phenomenon and affecting the target domain. This again shows that when there is an artistic visual communication image offset between cross-scene data, the classifier directly learning the source domain or learning both source and target domain data is ineffective for target domain classification.

Compared with other evaluation algorithms, our proposed method (MTSLR-S) considers the distribution difference of residual network after feature extraction, avoids the phenomenon of negative transfer, and achieves higher classification accuracy, which shows that our proposed algorithm can greatly reduce the art visual conveys image drift and improves cross-scene classification performance.

It should be pointed out that due to the complex relationship between classes in the distribution of images in art visual communication, the adaptation effects of different classes to the domain are different.

At the beginning of model training, the target domain model selects few samples, indicating that the features from the source domain samples are classified by the target domain classification model below the threshold of the discriminator. With the continuous reduction of the loss function, the classification accuracy of the model is getting better and better. At the same time, under the action of the discriminator regularization, the samples selected by each other between the domains gradually increase and reach a stability when the number of iterations reaches 100. Furthermore, we find that as the iterative learning process progresses, when the number of positive samples in the involved source domain suddenly increases, the objective function value also increases accordingly, resulting in small fluctuations in the change of the loss function. Figure 4 shows the source domain sample mobility based on the MTSLR-S algorithm.

(2) Experimental data results of data set 2

Figure 5 shows the classification accuracy of the classification metrics obtained under various evaluation methods.
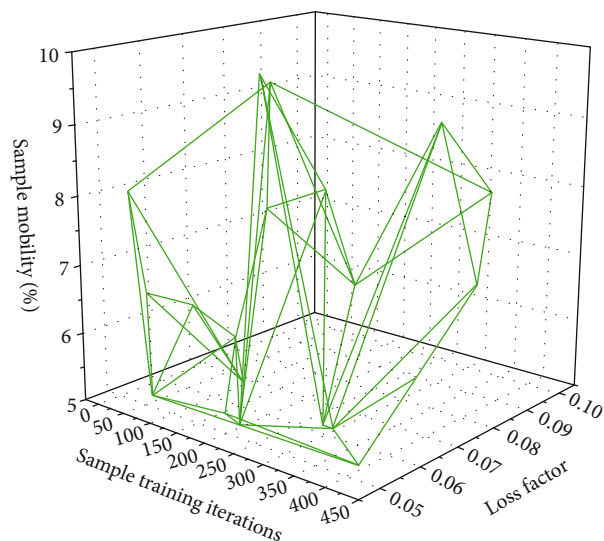

FIGURE 4: Source domain sample mobility based on dataset 1.

We can see that when the target domain scene has only a small number of training samples, compared with the traditional TD-SVM algorithm, our proposed method can bring more than 10% increase in the evaluation index (OA) and Kappa.

The results of Pavia data also show that our proposed multitask-based residual network achieves better classification results through cross-scene feature level domain adaptation, which proves that multitask learning can effectively reduce the impact of cross-scene data spatial distribution shift on classifiers.

As the number of training increases and the loss function continues to decrease, the number of samples useful to the model for the target domain selection fluctuates. Figure 6 shows the source domain sample transfer based on the MTSLR-S algorithm.

4.2. Analysis of the Classification Results of Art Visual Communication Images. During the experiment, each experiment was repeated 10 times, and different training samples and test samples were randomly selected each time. The evaluation criteria of the experimental results will be evaluated by the overall classification accuracy (OA) and the Kappa coefficient. Finally, the classification results of different evaluation algorithms are given, respectively.

(1) Data set-experimental results

We experimented with all the evaluation algorithms by changing the number of training samples per class in the target domain. The classification accuracy of our proposed method is better than other methods, which proves that the MTSLR-Fusion method can fully mine the feature information of source and target domains in small-sample classification. At the same time, it can be seen from the size of the standard deviation of OA that our proposed method provides more robust results, the change curve of OA is flat, and the fluctuation is small. Figure 7 shows the trend of
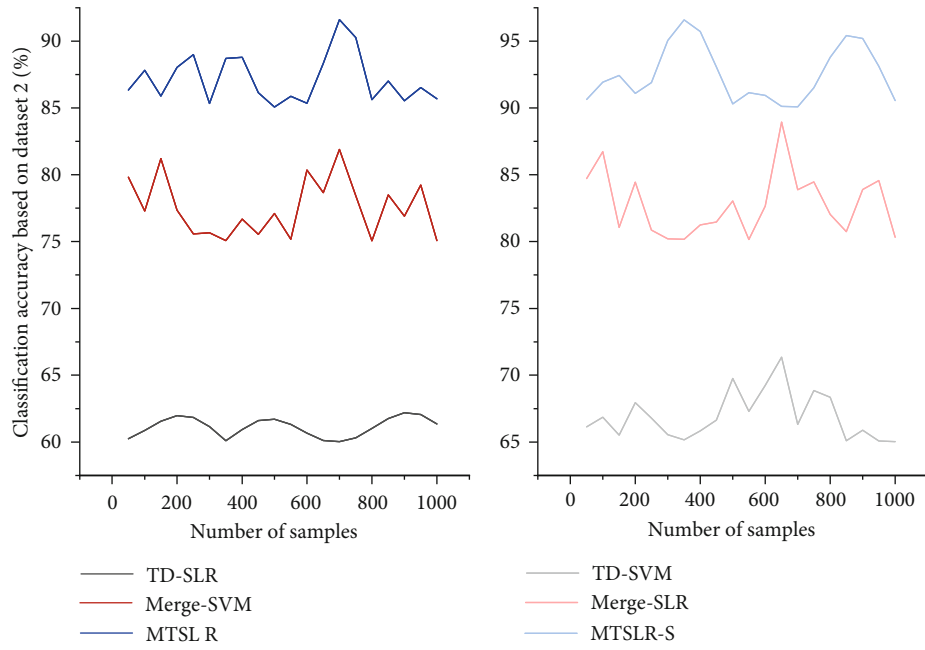
Figure 5: Classification accuracy of each class of training samples in the target domain based on dataset 2.
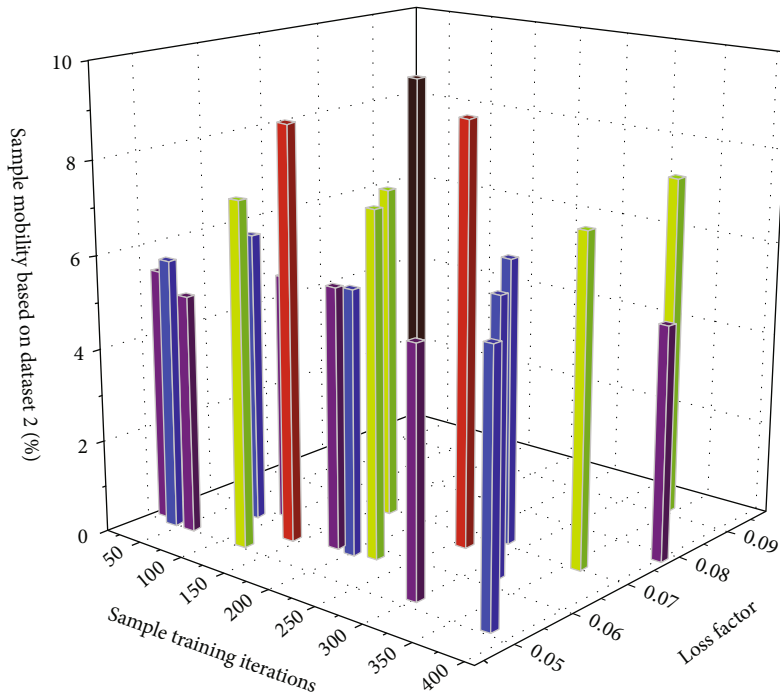

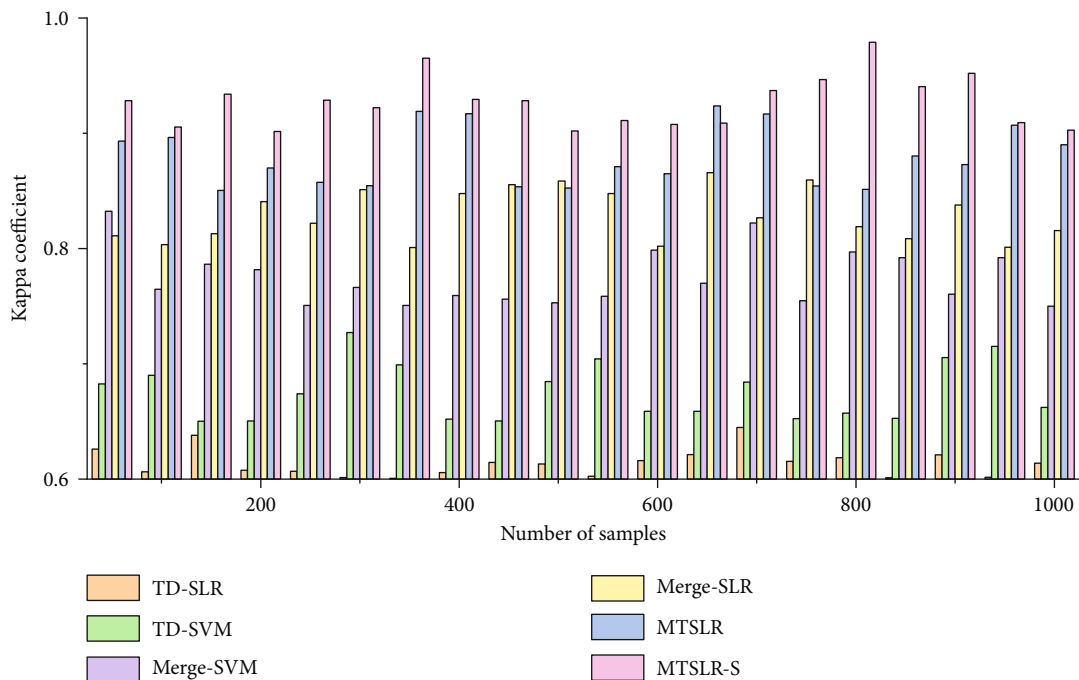Figure 6: Source domain sample mobility based on dataset 2.

FIGURE 7: The influence of the number of training samples of each class on the Kappa coefficient of the target domain based on dataset 1.
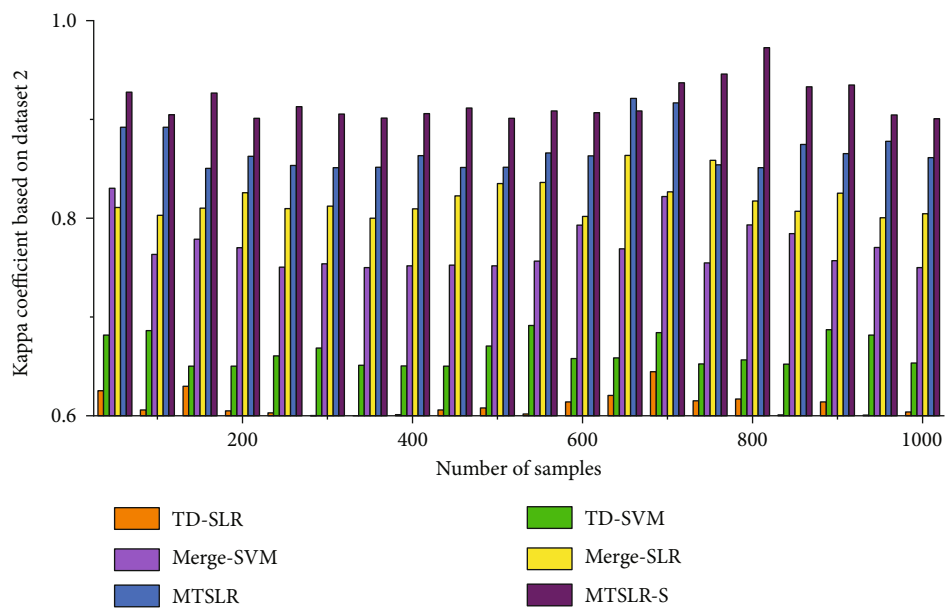

FIGURE 8: The influence of the number of training samples of each type in the target domain on the Kappa coefficient based on dataset 2.

the Kappa coefficient as the number of training samples increases gradually.

(2) Experimental results of dataset 2

The classification accuracy of our proposed method is better than other methods, which proves that the MTSLR-Fusion method can fully mine the feature information of source and target domains in small-sample classification. At the same time, it can be seen from the standard deviation of OA that with the gradual increase of the number of samples, the overall classification accuracy is correspondingly improved. Compared with other evaluation algorithms, our proposed method provides more robust results, the OA change curve is flat, and the fluctuation is small. Figure 8 shows the trend of the Kappa coefficient as the number of training samples increases gradually.

5. Conclusion

According to the related work in the field of feature extraction and matching, this paper proposes a method to extract feature descriptions in the neighborhood of key points. Aiming at the problem that there are few labeled samples in the target domain, it is not enough to use art to convey image information only to extract distinguishable feature representations. This paper proposes an image classification algorithm for high-art visual communication images based on multifeature extraction and classification decision fusion. This method aligns the high art visual communication image datasets in different domains (source domain and target domain) by correlation, reducing the spatial distribution difference of the datasets, which is beneficial to the subsequent extraction of common features, and then designs multiple feature extractors to extract art visuals separately. It conveys image spatial features and texture features, then effectively combines the extracted features through a multitask sparse logistic regression classifier and finally obtains the target domain classification results through a voting decision fusion mechanism. The algorithm makes full use of the auxiliary information of the source domain dataset. It avoids the troubles brought by traditional feature extractors and further effectively utilizes the transferable features of the source domain, while preventing the negative transfer phenomenon that often occurs in the feature transfer process. In this paper, a dual-channel deep residual convolutional neural network is designed. The multiple convolutional layers of the residual module in the network use hard parameters to share, so that the deep feature representation on the joint spatial spectrum dimension can be automatically extracted. Finally, the features extracted by the network are screened to maximize the auxiliary role of the labeled samples in the source domain and avoid the negative transfer problem caused by the forced transfer between irrelevant samples.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] D. Shi, L. Zhu, Z. Cheng, Z. Li, and H. Zhang, "Unsupervised multi-view feature extraction with dynamic graph learning," *Journal of Visual Communication and Image Representation*, vol. 56, pp. 256–264, 2018.

[2] B. Rasti and P. Ghamisi, "Remote sensing image classification using subspace sensor fusion," *Information Fusion*, vol. 64, pp. 121–130, 2020.

[3] C. Bai, J. Chen, L. Huang, K. Kpalma, and S. Chen, "Saliency-based multi-feature modeling for semantic image retrieval," *Journal of Visual Communication and Image Representation*, vol. 50, pp. 199–204, 2018.

[4] J. Liu, Z. Zhao, J. Ji, and M. Hu, "Research and application of wireless sensor network technology in power transmission and distribution system," *Intelligent and Converged Networks*, vol. 1, no. 2, pp. 199–220, 2020.

[5] X. Liu, N. Li, and Y. Xia, "Affective image classification by jointly using interpretable art features and semantic annotations," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 576–588, 2019.

[6] W. Li, H. Li, Q. Wu, X. Chen, and K. N. Ngan, "Simultaneously detecting and counting dense vehicles from drone images," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9651–9662, 2019.

[7] J. M. Ramirez and H. Arguello, "Spectral image classification from multi-sensor compressive measurements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 626–636, 2019.

[8] M. Kumar, J. Aggarwal, A. Rani, T. Stephan, A. Shankar, and S. Mirjalili, "Secure video communication using firefly optimization and visual cryptography," *Artificial Intelligence Review*, vol. 55, no. 4, pp. 2997–3017, 2022.

[9] M. Hu, H. Wang, X. Wang, J. Yang, and R. Wang, "Video facial emotion recognition based on local enhanced motion history image and CNN-CTSLSTM networks," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 176–185, 2019.

[10] C. Wang, D. Li, Z. Li et al., "An efficient local binary pattern based plantar pressure optical sensor image classification using convolutional neural networks," *Optik*, vol. 185, pp. 543–557, 2019.

[11] Y. Miao, G. Wu, C. Liu, M. S. Hossain, and G. Muhammad, "Green cognitive body sensor network: architecture, energy harvesting, and smart clothing-based applications," *IEEE Sensors Journal*, vol. 19, no. 19, pp. 8371–8378, 2019.

[12] X. Min, G. Zhai, K. Gu, Y. Liu, and X. Yang, "Blind image quality estimation via distortion aggravation," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 508–517, 2018.

[13] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International Journal of Remote Sensing*, vol. 28, no. 5, pp. 823–870, 2007.

[14] Z. Ni, H. Zeng, L. Ma, J. Hou, J. Chen, and K. K. Ma, "A Gabor feature-based quality assessment model for the screen content images," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4516–4528, 2018.

[15] Y. Liu, K. D. Tong, F. Mao, and J. Yang, "Research on digital production technology for traditional manufacturing enterprises based on industrial internet of things in 5G era," *The International Journal of Advanced Manufacturing Technology*, vol. 107, no. 3-4, pp. 1101–1114, 2020.

[16] M. Sarkar, D. S. S. S. S. Bello, C. Van Hoof, and A. Theuwissen, "Integrated polarization analyzing CMOS image sensor for material classification," *IEEE Sensors Journal*, vol. 11, no. 8, pp. 1692–1703, 2011.

[17] J. E. Lee, S. Hur, and B. Watkins, "Visual communication of luxury fashion brands on social media: effects of visual complexity and brand familiarity," *Journal of Brand Management*, vol. 25, no. 5, pp. 449–462, 2018.

[18] Z. Song, W. Ye, Z. Chen et al., "Wireless self-powered high-performance integrated nanostructured-gas-sensor network for future smart homes," *ACS Nano*, vol. 15, no. 4, pp. 7659–7667, 2021.

[19] M. Habibi and S. M. Sayedi, "Asynchronous demodulation technique for use in vision sensor image classification and segmentation," *International Journal of Circuit Theory and Applications*, vol. 39, no. 1, pp. 17–30, 2011.

[20] Y. Hu and X. Lu, "Learning spatial-temporal features for video copy detection by the combination of CNN and RNN," *Journal of Visual Communication and Image Representation*, vol. 55, pp. 21–29, 2018.

[21] X. Zhang and W. Dahu, "Application of artificial intelligence algorithms in image processing," *Journal of Visual Communication and Image Representation*, vol. 61, pp. 42–49, 2019.

[22] G. M. Foody and A. Mathur, "A relative evaluation of multiclass image classification by support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 6, pp. 1335–1343, 2004.

[23] Z. Ai, Y. Liu, F. Song, and H. Zhang, "A smart collaborative charging algorithm for mobile power distribution in 5G networks," *IEEE Access*, vol. 6, pp. 28668–28679, 2018.

[24] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image superresolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5799–5812, 2019.

[25] A. Ahad, M. Tahir, and K. L. A. Yau, "5G-based smart healthcare network: architecture, taxonomy, challenges and future research directions," *IEEE access*, vol. 7, pp. 100747–100762, 2019.

[26] L. Ma, M. Li, X. Ma, L. Cheng, P. du, and Y. Liu, "A review of supervised object-based land-cover image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 277–293, 2017.

[27] L. Fang, Y. Jin, L. Huang, S. Guo, G. Zhao, and X. Chen, "Iterative fusion convolutional neural networks for classification of optical coherence tomography images," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 327–333, 2019.

[28] K. Shafique, B. A. Khawaja, F. Sabir, S. Qazi, and M. Mustaqim, "Internet of things (IoT) for next-generation smart systems: a review of current challenges, future trends and prospects for emerging 5G-IoT scenarios," *IEEE Access*, vol. 8, pp. 23022–23040, 2020.