


ADAPTATION ART IMAGE STYLE TRANSFER BY INTEGRATING CSDA-FD ALGORITHM AND OSDA-DS ALGORITHM

Peng Wang* 

Liaoning Police College, Dalian, China

**Corresponding author: Peng Wang (13478475359@163.com)*

Submitted: 21 May 2025 Accepted: 04 Aug 2025 Published: 04 Dec 2025

License: CC BY-NC 4.0 

Abstract Traditional domain adaptation learning methods have a strong dependence on data labels. The transfer process can easily lead to a decrease in training set performance, affecting the effectiveness of transfer learning. Therefore, this study proposes a domain adaptation model that combines feature disentangling and disentangling subspaces. The model separates the content and style features of images through disentangling, effectively improving the quality of image transfer. From the results, the proposed feature disentangling algorithm achieved pixel accuracy of over 84% for semantic segmentation of 14 categories, including roads, sidewalks, and buildings, with an average pixel accuracy of 85.2%. On the ImageNet, the precision, recall, F_1 score, and overall accuracy of the research algorithm were 0.942, 0.898, 0.854, and 0.841, respectively. Compared with the One-Class Support Vector Machine, the precision, recall, F_1 , and overall accuracy were improved by 8.4%, 10.3%, 27.8%, and 10.9%, respectively. The proposed model can accurately recognize and classify images, providing effective technical support for image transfer.

Keywords: image style transfer; deep domain adaptation; feature disentangling; domain shift.

1. Introduction

Image Style Transfer (IST) is a crucial research direction in computer vision and image processing, which originates from in-depth exploration of artistic creation and image processing techniques [23]. Driven by computer technology, people have begun to try to combine computer technology with artistic creation, using algorithms to simulate and implement image rendering of different art styles, thereby creating new works [18]. In recent years, deep learning has shone brightly in IST. Convolutional Neural Networks (CNN) can precisely extract content and style features, and combined with Generative Adversarial Networks (GAN) to generate realistic images. Deep learning techniques have achieved seamless transfer of style from one image to another while preserving the original content [7]. This technology is not only widely used in fields such as art creation, film special effects, and game design, but also promotes the deep integration of technology and art, bringing unprecedented creativity and possibilities to image processing [30]. However, the deep learning method has a strong dependence on data labels, and requires training and test samples to meet the same spatial distribution, which makes it have poor generalization ability when facing massive Internet data. Therefore, some scholars have proposed the Deep Domain Adaptation (DDA) method, which combines domain adaptation and deep learning theories. This method can effectively solve distribution differences between the Source Domain (training dataset, SD) and the Target

Domain (testing or application dataset, TD). It is widely used in fields such as natural language processing and image transfer [10]. However, current domain adaptation learning still faces problems such as performance degradation in the training set due to transfer processes and difficulty in separating unknown categories in open domains [19]. The research considers two aspects: close set domain and open set domain. A domain adaptation model that integrates Close Set Domain Adaptation by Feature Disentangling (CSDA-FD) and Open Set Domain Adaptation by Disentangling Subspace (OSDA-DS) is proposed. This model separates content and style features through learnable weights and introduces a Domain Shift (DS) to make the model lighter, thereby improving the style transfer accuracy. The research aims to enhance the accuracy and adaptability of adaptation learning methods, improve the transferability of adaptation models, and provide innovative and more effective solutions for artistic IST. There are two main innovations. The first is to integrate CSDA-FD and OSDA-DS, and construct a domain adaptive model from both closed domain and open domain perspectives, providing a new solution for artistic IST. The second is to introduce a DS in the CSDA-FD algorithm, which preserves important information of the source domain image during the transfer learning process, effectively solving the decreased training set performance caused by the transfer process in traditional domain adaptive learning.

The remaining part of this research is structured in four sections. The Section 2 introduces the current research on IST and DDA methods worldwide. The Section 3 introduces the construction process of the proposed deep domain adaptation model. In Section 4 the experiments are conducted to verify its feasibility. The last Section 5 summarizes and discusses the paper, pointing out the shortcomings and future prospects.

2. Related work

IST is a computer vision technique that allows the style of one image to be applied to another one while preserving the main features and details of the content image [3]. With the rapid development of computer technology, deep learning techniques have been extensively applied to IST and have made significant progress. Liao and Huang [12] built a semantic guided IST based on matching regions to address the semantic region matching caused by the mismatch between content and style image object categories. This method achieved semantic aware style transfer by performing semantic context matching and combining it with a hierarchical local to global network architecture. Lin et al. [13] proposed an IST based on semantic segmentation to address the semantic mismatch in IST. The algorithm automatically extracted semantic information from images, and used this information to guide style transfer, effectively solving the semantic mismatch. Li et al. [11] built a GAN to address the diverse types and inconsistent distribution of low-dose scanning image noise generated by different commercial scanners. The network extracted noise patterns by performing noise encoding and fusing it into the generator, effectively improving the feasibility and denoising performance of low-dose scanning

image synthesis. Ma et al. [16] proposed a parental and force embedding network to address the semantic alignment between style and content in IST. This network achieved semantic embedding of local style patterns by jointly modeling feature associations and semantic correspondences, improving the visual quality and computational efficiency of style transfer.

DDA is a technique in deep learning aimed at solving the inconsistent data distribution between the SD and TD. Deep domain adaptation methods have wide applications in transfer learning, image recognition, natural language processing, and other fields, especially in situations where data annotation costs are high or data acquisition is difficult. These methods can significantly optimize the generalization ability. Therefore, some scholars have explored. Wu et al. [26] proposed an enhanced adaptation network to address scarce TD labels in partial domain adaptation. The network optimized the source data selection strategy through a deep reinforcement learning model and combined domain adaptation techniques to automatically filter out irrelevant source data, thereby effectively improving the accuracy and generalization ability of domain adaptation. Liu et al. [14] built a three-stage unsupervised domain adaptation strategy to address the difficulty and uneven distribution of pixel annotation datasets in remote sensing image semantic segmentation. This method enhanced the correlation between feature map channels through covariance channel attention modules, significantly improving the accuracy image semantic segmentation. Shermin et al. [21] built an adversarial domain adaptation model to address the knowledge transfer problem from a finite class SD to a multi-class TD in Open Set Domain Adaptation (OSDA). The model introduced a multi-classifier structure and weighting module to distinguish between known and unknown target samples, improving the accuracy and adaptability of OSDA. In response to the unavailability of TD labels and neglected class information in traditional methods for unsupervised domain adaptation, Kang et al. [9] built a comparative adaptation network, which designed an alternating update strategy and class aware sampling method. By optimizing new indicators, the network effectively simulated intra-class and inter-class domain differences, achieving unsupervised domain adaptation optimization.

In summary, some scholars have explored the style transfer of artistic images from the perspective of deep learning and have made meaningful progress. However, despite the significant theoretical advantages of deep domain adaptation methods in image transfer, there are still not enough practical cases to combine the two. Domain adaptation learning still faces problems such as decreased training set performance during the transfer process and difficulty in classifying unknown categories in open domains. This study innovatively proposes a domain adaptation model that integrates the CSDA-FD and the OSDA-DS. From the perspectives of close set domain and open set domain, a new domain adaptation learning method is designed and applied to IST to improve the accuracy and applicability of image transfer.

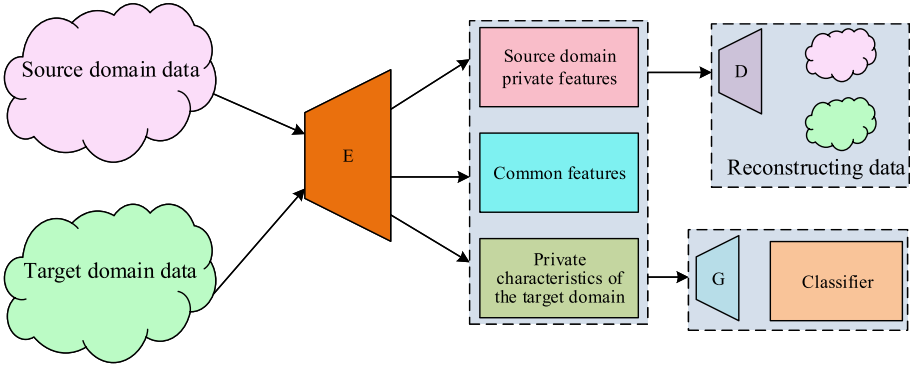


Fig. 1. The process of feature disentangling.

3. Domain adaptation model integrating feature disentangling and disentangling subspace

The domain adaptation method includes two types: close set domain and open set domain. Firstly, a detailed introduction is given to the CSDA-FD. Then, for the adaptation learning problem in the open set domain, the OSDA-DS algorithm is built.

3.1. Close set domain adaptation algorithm based on feature disentangling

Close Set Domain Adaptation (CSDA) is a subproblem in transfer learning that assumes that all categories in the SD and TD are known and identical [27]. This means that there are no new or disappearing categories between the SD and the TD during domain adaptation. In domain adaptation learning, Feature Disentangling (FD) is a commonly used method that can optimize the generalization ability [2]. However, although the category labels in CSDA are the same as those of the SD, current feature disentangling methods still suffer from performance degradation in the training set in transfer learning [5]. Therefore, the research is conducted to optimize feature disentangling. The novel CSDA-FD is proposed. Feature disentangling refers to decomposing features into simpler components, typically including task related features and irrelevant features [15]. Feature disentangling is to separate features that remain unchanged in both the SD and TD (domain invariant features) from features that only change in a specific domain (domain specific features) [25]. In this way, the model can better understand and adapt to data distributions in different fields, thereby improving the generalization ability.

The process of feature disentangling is shown in Figure 1. Firstly, it obtains the universal attributes of data by identifying cross-domain shared features. Secondly, it focuses on extracting features that only appear in a specific domain to capture unique

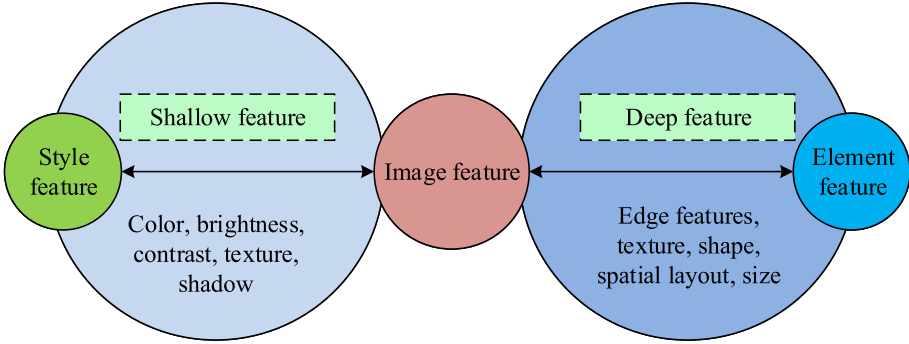


Fig. 2. Style features and content features.

information in that domain. This processing method helps the model to effectively transfer and apply knowledge between different fields. E signifies Feature Extractor, which is responsible for extracting features from both SD data and TD data. D is a feature Disentangler, which decomposes the extracted features into private features and common features. G is a Generator used to generate new data samples for training classifiers. Through feature disentangling, the model can learn features that are useful for both domains while reducing negative impacts caused by inter domain differences.

To better apply feature disentangling to IST, the concepts of *style features* and *content features* are introduced to decompose the features, as shown in Figure 2. Style features refer to the features related to artistic expression techniques, color tones, color distribution, and overall visual perception in an image. These features define the artistic style of the image, including brushstrokes, color usage, brightness, and contrast. Content features refer to the information directly related to the entity or scene represented by the image in the image, which usually includes the basic visual elements of the image, such as edges, textures, shapes, and recognition information of objects. Content features are the semantic core of an image, which helps identify the main objects and scenes in the image.

The image disentangling process based on the style and content features is shown in Figure 3. The image disentangling process based on style and content features is similar to the conventional feature disentangling process. The difference is that the disentangler decomposes the extracted features into style features and content features. Style features involve the visual style information of the image, while content features contain the structural and semantic information. The generator receives disentangled features and generates new images, which can be created based on the style features of the TD and the content features of the SD. The reconstructed image of the TD is the final output, which has the style features of the TD and the content features of

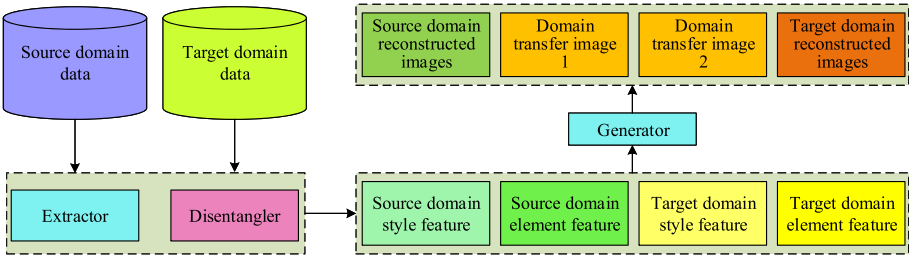


Fig. 3. Feature disentangling into processes for style and content features.

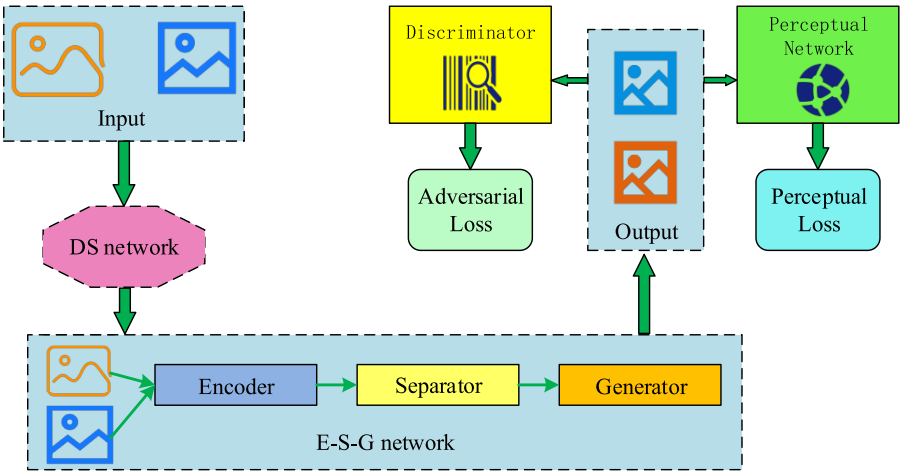


Fig. 4. Structure of close set domain adaptation algorithm based on feature disentangling.

the SD. The entire process is to separate and recombine image features through feature disentangling, so that the image can exhibit a new style while maintaining its original content.

To solve the performance degradation of the training set during the migration process, a DS network is introduced to optimize the accuracy of the SD and TD. The CSDA-FD is displayed in Figure 4. The CSDA-FD consists of a DS, Encoder (E), feature Separators (S), Generator, Discriminator (D), and Perceptual Network (PN). After the original images are input into the network, DS processes them and preserves important information to maintain the model’s classification ability for the SD. Next, the encoder extracts the features of these images, and the feature separator divides the image features into content features and style features. The generator maps features into the image space. PN is applied to extract perceptual features of images and constrain content and

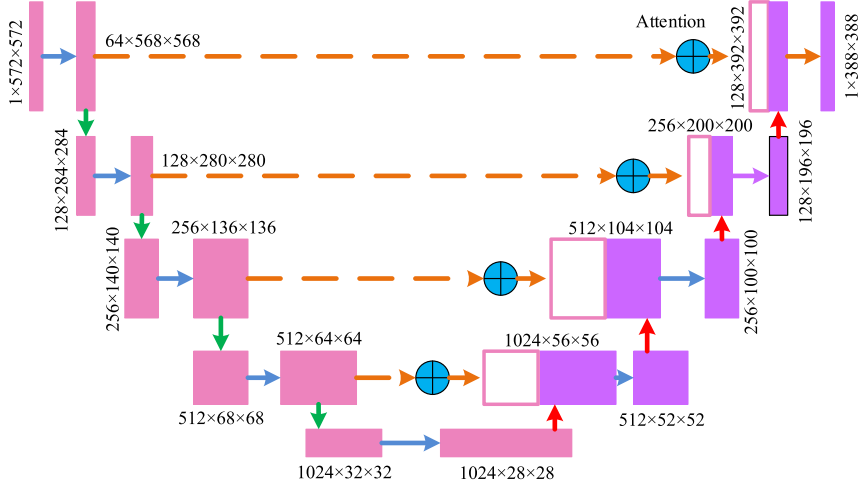


Fig. 5. Structure of domain shift.

style similarity. This step is achieved through pre-training. Adversarial losses during training are applied through discriminators.

The DS is displayed in Figure 5. The DS adopts a U-shaped network structure, which is similar to a “U” shape in architecture. Therefore, it is named *U-net*. It has a contraction path (encoder) and an expansion path (decoder), which undergo convolution and nonlinear activation operations at each stage [4]. The input size of U-net is $1 \times 572 \times 572$, and the output is $1 \times 388 \times 388$. The encoder section consists of five convolutional blocks, the first four of which are composed of convolutional layers, ReLU activation functions, and a 2×2 max pooling layer with a stride of 2. The convolutional layer use a 3×3 filter. The fifth convolutional block consists of convolutional layers and ReLU activation functions. The decoder section consists of four up-sampling blocks, each consisting of an up-sampling layer, a convolutional layer, a ReLU activation function, and skip connections. The convolutional layer uses a 3×3 filter. The output layer uses 1×1 convolution to convert the $64 \times 392 \times 392$ feature map into an output image of $1 \times 388 \times 388$. The condition satisfied by the DS is displayed in equation (1)

$$Y(I_{\text{DSB}}) \sim Y(I_A), Y(I_{\text{DSA}}) \sim Y(I_A), \quad (1)$$

where $Y(I_{\text{DSB}})$ signifies the output of the TD image I_{DSB} processed by DS. $Y(I_A)$ signifies the output of the SD image I_A . Similarly, $Y(I_{\text{DSA}})$ represents the output of the SD image I_{DSA} processed by the DS. DS is constrained, as shown in equation (2)

$$I_{\text{DSA}} \sim I_A, I_{\text{DSB}} \sim I_A. \quad (2)$$

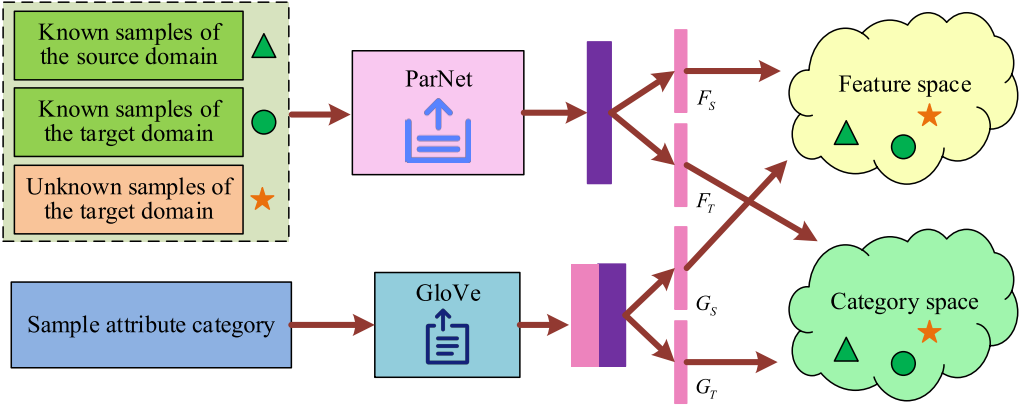


Fig. 6. Structure of OSDA-DS.

The study takes a loss function to deceive the discriminator in the model, as shown in equation (3)

$$L_{DS}^d = E[y_a \log(D(x_a)) + y_b \log(D(x_b))], \quad (3)$$

where L_{DS}^d represents the domain adaptation loss function. E represents the expected value. x_a and y_a are the sample data and label data of the SD, respectively. x_b and y_b are the same. $D(x_a)$ and $D(x_b)$ represent the outputs of the discriminator for the input of SD and TD. To make the TD image closer to the SD image, DS retains the perturbation changes in the image during training, as shown in equation (4)

$$I_{DSB} = I_B + j, \quad (4)$$

where j represents the perturbation change of the TD image.

3.2. Disentangling the subspace adaptation learning algorithm in open set domain

CSDA-FD effectively separates the content and style features through nonlinear disentangling, significantly improving the adaptability of the model in close set domain scenarios. However, domain adaptation problems in the real world are often more complex, especially when the TD contains unknown categories, that is, open set domains, which make feature disentangling more difficult to separate the TD [8]. To address this challenge, the OSDA-DS is proposed. This algorithm achieves precise separation of known and unknown categories by constructing a disentangled subspace of features and categories, making a breakthrough in the field of OSDA learning.

The OSDA-DS structure is shown in Figure 6. OSDA-DS performs feature extraction

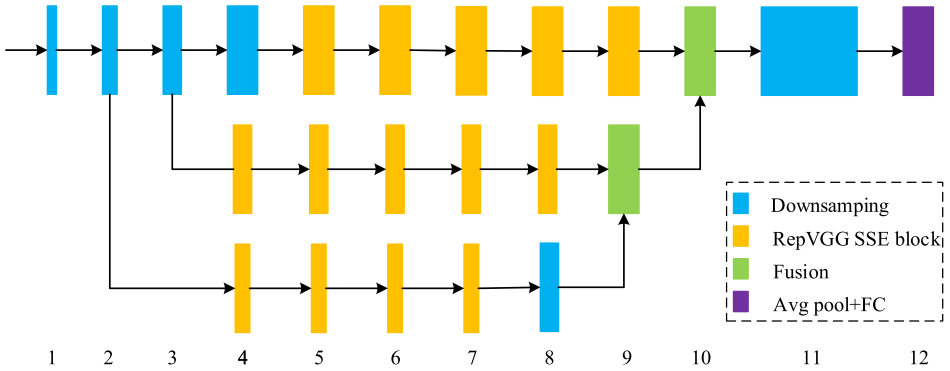


Fig. 7. Structure of ParNet.

through a Parallel Network structure (ParNet), and then maps it to a feature subspace to generate feature vectors F_s and F_t . The Global Global Vectors for Word Representation (GloVe) model is used to encode the textual descriptions of the sample category and map them to semantic feature vectors G_s and G_t in the category subspace. The study takes ParNet for feature extraction, which employs a parallel subNet structure to effectively reduce depth while maintaining high performance, effectively preventing the gradient explosion caused by excessive depth [6].

The ParNet is displayed in Figure 7. The network depth is 12, and the initial layer consists of a series of down-sampling blocks. The outputs of down-sampling blocks 2, 3, and 4 are sent to streams 1, 2, and 3, respectively [29]. 3 is the optimal number of streams for a given parameter budget. Each stream is composed of some representative visual blocks with visual attention mechanisms. These blocks process these features at different resolutions, and then use Fusion connections to fuse features from different streams [17]. Finally, the output is passed to the down-sampling block at depth 11. A disentangled subspace based on the feature distribution obtained from ParNet is constructed. The mapping network includes linear layers and activation layers, with output and subspace dimensions set to 200. The existence and learnability of mapping functions are crucial in domain adaptive learning. According to the general approximation theorem, neural networks such as multi-layer perceptrons can approximate any continuous function with arbitrary accuracy, providing theoretical support for mapping functions. Under the reasonable assumption of continuity between feature space and subspace, and continuity of mapping function, there exists a mapping function that can effectively capture the complex relationships between features and achieve accurate mapping from features to subspace. The mapping function maps the features extracted by ParNet to the feature space and the category space, as shown in equation (5)

$$\begin{cases} \alpha_a^S = F_S(x_a), \alpha_b^S = F_S(x_b), \alpha_c^S = F_S(x_c), \\ \alpha_a^T = F_T(x_a), \alpha_b^T = F_T(x_b), \alpha_c^T = F_T(x_c), \end{cases} \quad (5)$$

where $\alpha_a^S, \alpha_b^S, \alpha_c^S$ and $\alpha_a^T, \alpha_b^T, \alpha_c^T$ are both mapped visual features, with the former in the feature space and the latter in the category space. After mapping the function, the distribution of visual features is readjusted to reduce the differences in feature space caused by categories, making the model more sensitive and enhancing the discriminative ability. In terms of learnability, appropriate loss functions are designed and optimization algorithms such as gradient descent are used to optimize the parameters of the mapping function, which can gradually approach the true feature mapping relationship. Under certain conditions, optimization algorithms can ensure that the learning process of the mapping function converges to a local or global optimal solution, thereby ensuring the learnability of the mapping function.

To achieve the disentangling function of mapping, a distributed loss function is constructed, as shown in equation (6)

$$\begin{cases} L_{\text{dist}} = L_{\text{dist}}^S + L_{\text{dist}}^T, \\ L_{\text{dist}}^S = \max(0, d(\alpha_a^S, \alpha_b^S) - d(\alpha_b^S, \alpha_c^S) + m), \\ L_{\text{dist}}^T = \max(0, d(\alpha_a^T, \alpha_c^T) - d(\alpha_a^T, \alpha_b^T) + m), \end{cases} \quad (6)$$

where L_{dist} represents the total domain distance loss, which consists of the domain distance loss L_{dist}^S in the feature space and the domain distance loss L_{dist}^T in the category space. $d(\cdot, \cdot)$ is a distance function used to calculate the distance between features. m is a margin used to promote greater separation of features between different categories. The loss function further enhances the discriminative ability and optimizes the ability to distinguish differences between different categories. GloVe is used to extract the semantic features and can work synchronously with mapping networks. This model is a word embedding model based on global statistical information, proposed by researchers at Stanford University [20]. It learns the vector representation of vocabulary by analyzing the co-occurrence information of vocabulary in large-scale text corpora [22]. The GloVe is to use the co-occurrence matrix of vocabulary and obtain the low dimensional vector representation of vocabulary through matrix decomposition [1]. These vectors capture the semantic and syntactic relationships between words, allowing semantically similar words to approach each other in the vector space [28]. The training process of GloVe is relatively easy to parallelize, with fast training speed and the ability to utilize global information, which makes it perform well on small datasets [24]. GloVe projects the mapping onto the disentangled subspace to further enhance the model performance, as shown in equation (7)

$$\begin{cases} \beta_a^S = G_S(z_a), \beta_b^S = G_S(z_b), \beta_c^S = G_S(z_c), \\ \beta_a^T = G_T(z_a), \beta_b^T = G_T(z_b), \beta_c^T = G_T(z_c), \end{cases} \quad (7)$$

where $\beta_a^S, \beta_b^S, \beta_c^S$ represent the semantic features of the feature space obtained by transforming sample z_a, z_b, z_c in the SD through the mapping function G_S of the SD. Similarly, $\beta_a^T, \beta_b^T, \beta_c^T$ represent the semantic features of the category space obtained by transforming the same samples in the TD through the mapping function G_T of the TD. This study constructs a new correlation loss function, aiming to align features and semantic information in the subspace. The correlation loss function is shown in equation (8)

$$\begin{cases} L_{\text{dist}}^S = L_{\text{con}}^S + L_{\text{con}}^T, \\ L_{\text{con}}^S = \max(0, d(\alpha_a^S, \beta_a^S) - d(\alpha_a^S, \beta_b^S) + m), \\ L_{\text{con}}^T = \max(0, d(\alpha_a^T, \beta_a^T) - d(\alpha_a^T, \beta_b^T) + m), \end{cases} \quad (8)$$

where L_{con}^S represents the loss function of the feature space, and L_{con}^T represents the loss function of the category space. The association between these two functions establishes a connection between visual and semantic features, and enables objects of the same type to have a common mapping area. By constructing an association loss function, it is possible to effectively connect visual and semantic features, thereby more accurately identifying the features and categories of unknown test samples in two different subspaces. The model cleverly projects the semantic features of the labels onto the disentangled subspace through two mapping functions. This process not only achieves effective separation between visual features and categories, but also deeply explores the intrinsic connections between the two. In the disentangled subspace, two subspaces are carefully constructed, which comprehensively optimize visual features and categories respectively, and deeply understand the relationship between the two. This design classifies features and categories more accurately on the test set when dealing with unknown samples, significantly improving performance. The collaborative optimization of the two subspaces further enhances the accuracy and generalization ability, making it more adept at handling complex tasks.

4. Experimental verification and analysis of deep domain adaptation model

In Section 2, this study provides a detailed introduction to the construction process of the domain adaptive model that integrates feature decoupling and decoupling subspaces, including the detailed structure and implementation mechanisms of CSDA-FD and OSDA-DS. Next, to verify the effectiveness and superiority of the proposed model, a series of experiments are conducted to validate and analyze the model. To verify the effectiveness of the domain adaptation model fusing the CSDA-FD and the OSDA-DS, the performance of the domain adaptation algorithm under close set domain and open set domain is verified. Finally, the research model is applied to artistic IST to further verify its effectiveness in transfer applications. In data preprocessing, all input images are scaled to 256×256 pixels and normalized to map pixel values to the range of $[0, 1]$.

Tab. 1. Experimental environment and parameters.

Experimental environment	
Configuration item	Configuration details
Processor	Interl Corei7-8750H
Graphics processing unit	NVIDIA Tesla K80
Internal memory	16 G
Hard disk	500 G
Operating system	Ubuntu 20.04
Deep Learning Framework	Pytorch 2.2.2
Programming Language	Python 3.7
Experimental parameter	
Configuration item	Configuration details
Training rounds	100
Batch size	64
Learning rate	0.003
Dropout	0.2

For labeled data, the labels are converted into a single hot encoding form for model training. This study uses the Adam optimizer and adds operations such as random horizontal flipping, random rotation, and random color jitter during the training process to increase data diversity. The initial learning rate is set to 0.003 and adjusted using cosine annealing strategy. The batch size is 64 and the epoch is 100. In addition, the Dropout technique with a dropout rate of 0.2 is used to prevent overfitting. When the loss values on the training and validation sets change by less than 0.001 within 10 epochs, the model is considered to have converged. The early stop condition is triggered when the loss value on the validation set rises continuously for 15 epochs to avoid overfitting. The experimental platform selects a deep learning framework based on Pytorch. The experimental equipment and related parameters are described in detail, as displayed in Table 1.

The study first evaluates the performance of the CSDA-FD algorithm in digital classification under close set domains. The research model is compared with multiple classical domain adaptation methods. Three classical digital datasets are selected: MINST handwritten digit dataset, SVHN dataset, and EMNIST dataset.

Each method is evaluated 10 times on each task to ensure the stability of the results. The evaluation metric is numerical recognition accuracy, and three domain adaptation tasks are performed, namely: MINST to SVHN, SVHN to MINST, and EMNIST to MINST, as displayed in Table 2. In the three domain adaptation tasks of MINST to SVHN, SVHN to MINST, and EMNIST to MINST, the numerical recognition accuracy of the proposed CSDA-FD algorithm was 97.9%, 92.5%, and 98.1%, respectively, which

Tab. 2. Experimental results of numerical recognition accuracy [%].

Method	MINST to SVHN	SVHN to MINST	EMNIST to MINST
Source Only	81.1	67.8	83.2
DANN	83.0	75.1	85.5
DDC	82.1	77.4	86.2
MK-MMD	85.7	84.1	85.1
GAN	88.2	85.1	89.4
ADDA	90.2	88.3	91.6
DSN	87.9	86.5	90.7
CyCADA	92.1	88.8	92.5
DRANet	96.5	87.4	96.9
CDA	95.1	89.4	97.8
CSDA-FD	97.9	92.5	98.1
Target Only	98.5	92.3	99.4

was superior to that of other algorithms. To quantify the performance difference between CSDA-FD and other methods, this study takes Cohen's d to calculate the effect size and reports the 95% confidence interval. Taking the MINST to SVHN task as an example, the Source Only method is selected as the baseline (accuracy 81.1% and standard deviation 2.2%) and compared with the CSDA-FD method (accuracy 97.9% and standard deviation 1.6%). The calculated Cohen's d value was 8.75, with a 95% confidence interval of [7.37, 10.13]. This indicates that the performance of the CSDA-FD method is significantly better than that of the Source Only method, and the effect size is statistically significant. Overall, in all three domain adaptation scenarios, the research model demonstrated good numerical recognition accuracy, and the research algorithm was close to the accuracy of training directly using the labeled TD. Especially in the SVHN to MINST task, the accuracy of the research algorithm even exceeded that of training directly using the TD by 0.2%. The semantic segmentation performance of the CSDA-FD algorithm was tested in this study.

The experimental dataset was selected from the GTA5 road scene dataset, and 14 classic categories were selected for training. The evaluation metric is Pixel Accuracy (PA), as displayed in Table 3. The semantic segmentation results for 14 categories including roads, sidewalks, and buildings showed that the average PA of the CyCADA algorithm was 80.0%, the DRANet algorithm was 81.0%, the CDA algorithm was 83.1%, and the proposed CSDA-FD had a PA of over 84%, with an average PA of 85.2%. Overall, the semantic segmentation performance of the research algorithm is excellent, with high PA.

The memory usage of the algorithm in MINST to SVHN and SVHN to MINST transfer learning is tested, as displayed in Figure 8. According to Figure 8a, in the

Tab. 3. Experimental results on semantic segmentation accuracy [%].

Method	Road	Sidewalk	Building	Fence	Tree	Traffic light	Vehicle
CyCADA	80.6	79.9	79.4	79.1	79.1	80.0	80.4
DRANet	81.3	81.7	81.5	80.1	81.4	81.3	81.6
CDA	83.1	82.6	83.1	83.6	83.9	83.0	83.5
CSDA-FD	84.4	85.7	85.3	85.7	85.4	85.5	85.2

	Bicycle	Motorcycle	Pedestrian	Animal	Sky	Trash can	Street light
CyCADA	80.9	80.4	79.5	80.3	79.7	79.5	81.0
DRANet	80.1	81.1	81.4	80.2	80.5	81.9	80.5
CDA	82.5	83.3	83.2	82.2	82.7	82.1	83.9
CSDA-FD	85.7	85.2	84.1	85.7	85.6	84.0	85.8

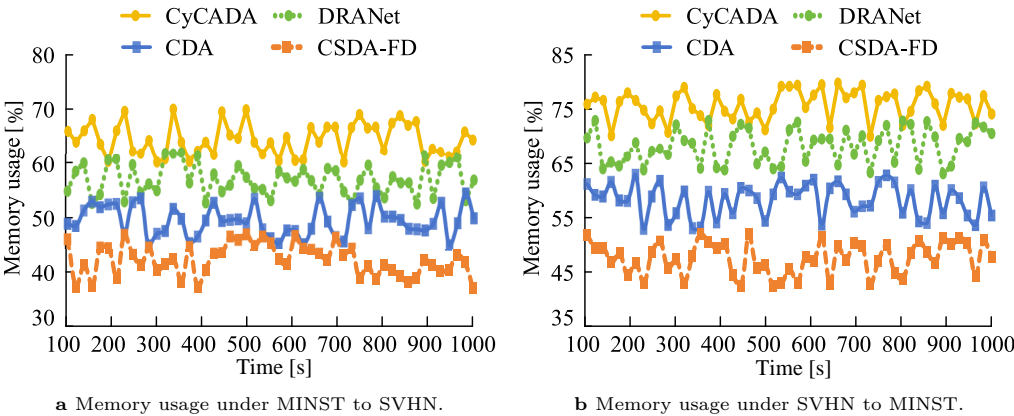


Fig. 8. Comparison result of memory usage.

MINST to SVHN transfer learning, the average memory usage of the CyCADA algorithm was 65.2%, the DRANet algorithm was 57.5%, the CDA algorithm was 49.9%, and the proposed CSDA-FD had an average memory usage of 42.6%. In SVHN to MINST transfer learning (see Fig. 8b), the average memory usage of the CyCADA algorithm was 75.7%, the DRANet algorithm was 68.2%, the CDA algorithm was 58.3%, and the proposed CSDA-FD algorithm was 47.5%. Overall, the research algorithm has the lowest memory usage, which is beneficial for improving the efficiency of algorithm operation.

The study tested the OSDA-DS algorithm in an open set domain using the Office-31 as the experimental dataset. The Office-31 is a benchmark dataset used for domain adaptation research, which includes images from different office environments such as Amazon, Webcam, and Digital Single Lens Reflex. The testing indicators are average accuracy and shared accuracy. Average accuracy is one of the most frequently applied

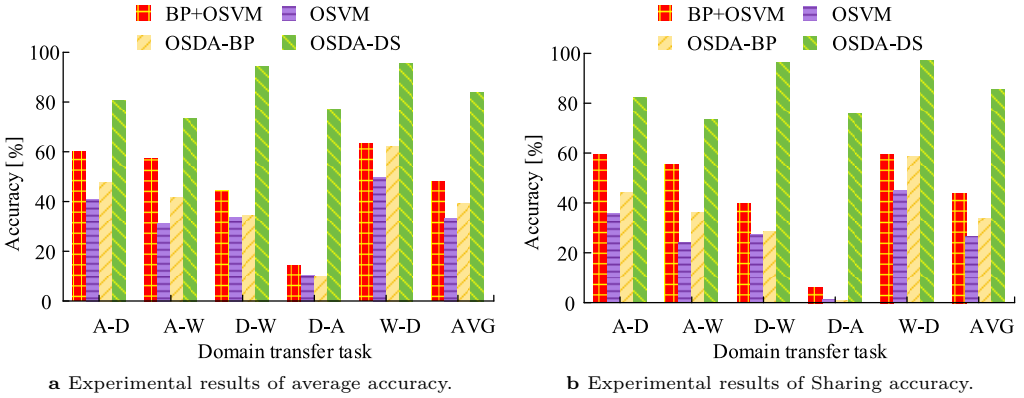


Fig. 9. Classification accuracy test results under open set domain.

evaluation indicators in classification problems, which represents the proportion of correctly classified samples to the total sample size. High accuracy indicates good performance. Shared accuracy only considers the accuracy of known samples. The comparison methods include the One-Class Support Vector Machine (OSVM) algorithm, the OSDA optimized by Back Propagation (OSDA-BP) algorithm, and the domain adaptation algorithm BP+OSVM, which combines back propagation and OSVM. The results are shown in Figure 9, where in the horizontal axis, A-D signify the accuracy from the Amazon domain to the scanner domain, A-W represents the accuracy from the Amazon domain to the camera domain, D-W represents the accuracy from the scanner domain to the camera domain, D-A represents the accuracy from the scanner domain to the Amazon domain, and W-D represents the accuracy from the camera domain to the scanner domain. According to Fig. 9a, under five transfer learning tasks, the average classification accuracy of OSVM was 33.5%, the OSDA-BP algorithm was 38.9%, the BP+OSVM was 48.0%, and the OSDA-DS algorithm was 83.6%. According to Fig. 9b, the average shared accuracy of the research algorithm was 85.6%, which was better than that of comparison algorithms. Overall, the research algorithm has the highest average accuracy and shared accuracy in image classification in transfer learning, and can effectively perform semantic recognition and classification in open set domains.

The Precision (P), Recall (R), F_1 score, and Overall Accuracy (OA) of the algorithm are tested, as displayed in Table 4. According to this Table, on the Office-31 dataset, the P , R , F_1 , and OA values of the OSVM algorithm were 0.928, 0.844, 0.788, and 0.775, respectively. The P , R , F_1 , and OA values of the OSDA-BP algorithm were 0.951, 0.879, 0.864, and 0.886, respectively. The P , R , F_1 , and OA values of the BP+OSVM algorithm were 0.961, 0.907, 0.933, and 0.925, respectively. The P , R , F_1 , and OA values of the proposed OSDA-DS were 0.978, 0.943, 0.960, and 0.955, respectively, which were

Tab. 4. Test results for P , R , F_1 , and OA.

On the Office-31 dataset				
Model	P	R	F_1	OA
OSVM	0.928	0.844	0.788	0.775
OSDA-BP	0.951	0.879	0.864	0.886
BP+OSVM	0.961	0.907	0.933	0.925
OSDA-DS	0.978	0.943	0.960	0.955

On the ImageNet dataset				
Model	P	R	F_1	OA
OSVM	0.869	0.814	0.668	0.732
OSDA-BP	0.860	0.821	0.712	0.755
BP+OSVM	0.887	0.856	0.780	0.794
OSDA-DS	0.942	0.898	0.854	0.841

Tab. 5. Results of the ablation experiment.

Model	Average accuracy	P	R	F_1
Without DS	0.782	0.942	0.908	0.928
Without GloVe	0.805	0.957	0.919	0.937
Without ParNet	0.768	0.934	0.903	0.922
Complete model	0.836	0.978	0.943	0.960

superior to those of other algorithms. On the ImageNet, the P , R , F_1 , and OA values of the research algorithm were 0.942, 0.898, 0.854, and 0.841, respectively. Compared with the OSVM algorithm, the P , R , F_1 , and OA values improved by 8.4%, 10.3%, 27.8%, and 10.9%, respectively. Overall, the research algorithm showed good detection results on both datasets, with superior performance and good transferability. To evaluate the independent contributions of each component in the OSDA-DS algorithm, this study conducts ablation experiments using the Office-31 dataset.

The complete model is compared with models without DS, GloVe, and ParNet. The results of the ablation experiment are shown in Table 5. According to this Table, the average accuracy OA, P , R , and F_1 score of the complete model were the highest, at 83.6%, 97.8%, 94.3%, and 96.0%, respectively. The model performance was the worst after removing ParNet, indicating that ParNet played a crucial role in feature extraction and model performance. The results indicate that the three components of DS, GloVe, and ParNet all play important roles in the OSDA-DS algorithm, jointly improving the classification performance and transfer ability of the model in open set domains.

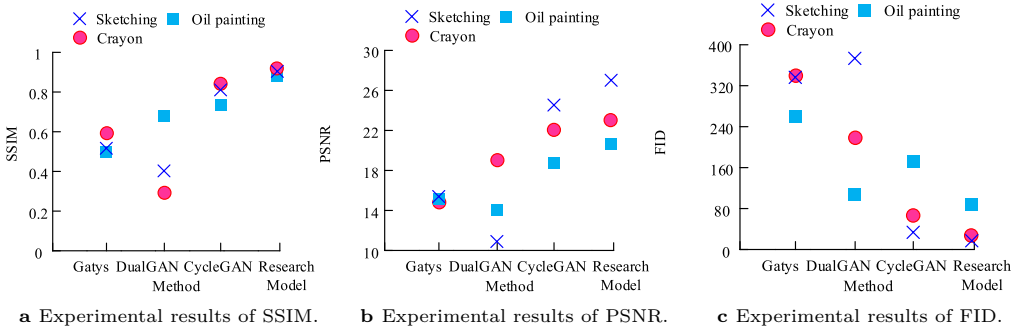


Fig. 10. Experiment results for SSIM, PSNR and FID.

Furthermore, the study tested the artistic image transfer effect of the domain adaptation model by selecting images of painting styles such as crayon drawing, oil painting, and sketching as the training dataset. The image size is set to 256×256 , and photographic images of landscapes, buildings, etc. are selected as the test data. A comparative experiment is conducted to evaluate the quality of art images generated by transfer learning. The comparison methods selected are Gatys algorithm, Dual Generative Adversarial Networks (DualGAN), and Cycle-Consistent Generative Adversarial Networks (CycleGAN). The evaluation metrics are Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Fréchet Inception Distance (FID), as displayed in Figure 10. According to Fig. 10a, the average SSIM of the Gatys algorithm was 0.53, the average SSIM of the DualGAN algorithm was 0.48, the average SSIM of the CycleGAN was 0.81, and the average SSIM of the research model was 0.88. According to Fig. 10b, the average PSNR of the Gatys algorithm was 14.51, the average PSNR of the DualGAN algorithm was 15.11, the average PSNR of the CycleGAN algorithm was 22.91, and the average PSNR of the research model was 22.90. According to Fig. 10c, the average FID index of Gatys algorithm was 308.55, the average FID of DualGAN algorithm was 233.85, the average FID of CycleGAN was 87.19, and the average FID of the research model was 0.88. Overall, the research model generates images with minimal noise and can produce high-quality artistic images of different styles.

The study randomly selects three photographs for style transfer and generates images with sketching, oil painting, and crayon drawing styles, as shown in Figure 11. The research model was able to learn from images of different painting types, with good style transfer effects. The generated images retained the structural and semantic information of the original photographic images, with minimal distortion and reasonable color filling. Overall, the new images generated by the research model have vibrant colors and distinct lines, which can effectively facilitate the transfer learning of different artistic styles.

To further validate the superiority of the proposed model, this study takes the

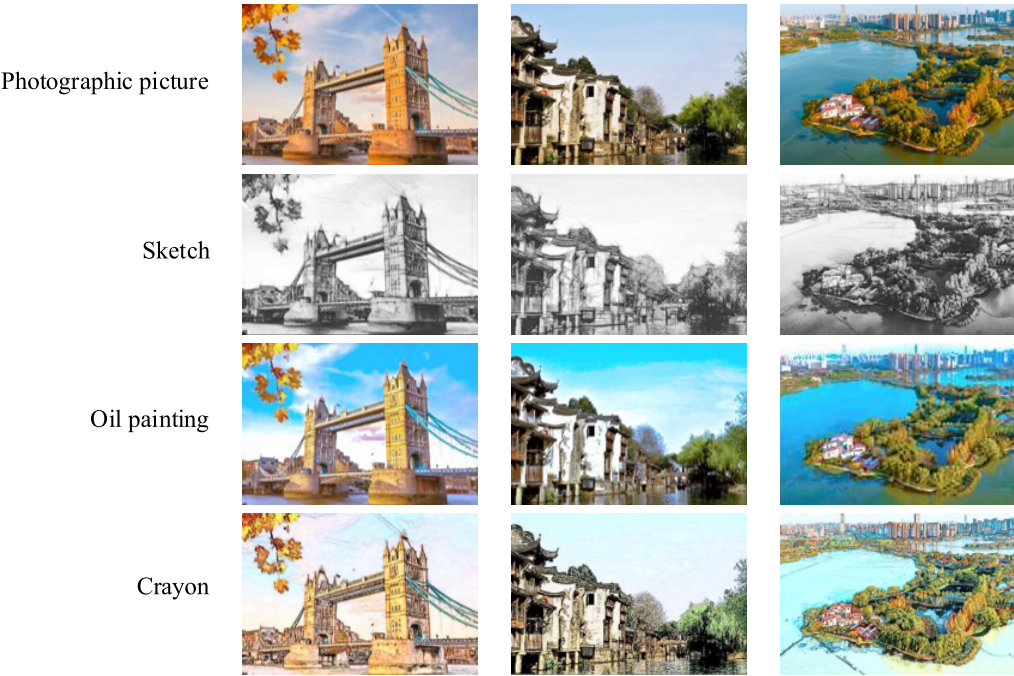


Fig. 11. Image style transfer rendering.

Tab. 6. Performance comparison results of five models.

Model	SSIM	PSNR	FID
Neural style conversion	0.65	18.3	120.2
AdaIN	0.71	20.0	94.6
WCT	0.74	20.8	85.1
StyleGAN2	0.79	21.9	74.7
Research model	0.90	23.2	0.89

WikiArt dataset for testing, which includes 55 art styles. The proposed model is compared with four state-of-the-art style conversion baseline models: neural style conversion, AdaIN, WCT, and StyleGAN2. The performance comparison results of these five models are shown in Table 6. The image quality generated by the proposed model in the WikiArt dataset was still good, with SSIM, PSNR, and FID of 0.90, 23.2, and 0.89, respectively, all higher than those of baseline models. The results indicate that the proposed model has stronger style transfer ability and higher generation quality when processing complex art style datasets, demonstrating certain superiority.

5. Discussion and conclusion

In response to the performance degradation of the training set and the difficulty in classifying unknown categories in domain adaptation learning during the transfer of artistic images, this study considered both close and open set domains to propose a domain adaptation model that integrated the CSDA-FD algorithm and the OSDA-DS algorithm. The model separated image features through feature disentangling and established a disentangling subspace. A DS was introduced to make the model lighter, improving the accuracy and efficiency of style transfer. Finally, the effectiveness and superiority were verified through experiments. The performance evaluation results showed that the research algorithm achieved an average numerical recognition accuracy of 96.2% in three domain adaptation tasks, which was superior to that of other algorithms. The results of the semantic segmentation task showed that the PA of the research algorithm was above 84%, with an average PA of 85.2%. The semantic segmentation performance was excellent and the PA was high. The test results under the open set domain showed that the OSDA-DS algorithm had an average classification accuracy of 83.6% and an average sharing accuracy of 85.6%, which could accurately classify images and prepare for subsequent image transfer. The quality of art images generated by transfer learning was evaluated. The average SSIM was 0.88, the average PSNR was 22.90, and the average FID was 0.88. The research model was able to learn from images of different painting types, with good style transfer effect, low generated image noise, and high quality. Overall, the domain adaptation model fused the CSDA-FD algorithm and the OSDA-DS algorithm for image transfer under different art styles can accurately recognize and classify images, thereby achieving style transfer. The proposed model has broad application prospects in real-world fields such as business design and educational platforms. For example, in the advertising industry, this model can quickly convert ordinary product images provided by customers into images with specific artistic styles, thereby enhancing the attractiveness of advertisements. Online art education platforms can also integrate this model to provide students with an interactive learning tool, assisting in creation and enhancing their art appreciation abilities. However, in practical operation, it may not be possible to completely and accurately decouple image content and style features, thereby affecting the accuracy and effectiveness of style transfer. Moreover, the model requires high computational resources and time during the training phase, especially when dealing with large-scale datasets, which limits its applicability in practical applications. Therefore, in future research, multi-scale analysis tools such as wavelet transform can be further combined to decompose images into subbands of different scales and perform feature decoupling operations separately. The network structure of the model is simplified and optimized to reduce redundant computational layers and parameters, and improve the computational efficiency of the model.

References

- [1] V. Belcamino, A. Carfi, and F. Mastrogiovanni. A systematic review on custom data gloves. *IEEE Transactions on Human-Machine Systems* 54(5):520–535, 2024. doi:[10.1109/THMS.2024.3394674](https://doi.org/10.1109/THMS.2024.3394674).
- [2] H. Cheng, Y. Wang, H. Li, A. C. Kot, and B. Wen. Disentangled feature representation for few-shot image classification. *IEEE Transactions on Neural Networks and Learning Systems* 35(8):10422–10435, 2024. doi:[10.1109/TNNLS.2023.3241919](https://doi.org/10.1109/TNNLS.2023.3241919).
- [3] M. Cotogni, M. Arazzi, and C. Cusano. PhotoStyle60: A photographic style dataset for photo authorship attribution and photographic style transfer. *IEEE Transactions on Multimedia* 26(6):10573–10584, 2024. doi:[10.1109/TMM.2024.3408683](https://doi.org/10.1109/TMM.2024.3408683).
- [4] Y. Feng, J. Chen, S. He, T. Pan, and Z. Zhou. Globally localized multisource domain adaptation for cross-domain fault diagnosis with category shift. *IEEE Transactions on Neural Networks and Learning Systems* 34(6):3082–3096, 2023. doi:[10.1109/TNNLS.2021.3111732](https://doi.org/10.1109/TNNLS.2021.3111732).
- [5] Y. Gao, S. Ma, and J. Liu. DCDR-GAN: A densely connected disentangled representation generative adversarial network for infrared and visible image fusion. *IEEE Transactions on Circuits and Systems for Video Technology* 33(2):549–561, 2023. doi:[10.1109/TCSVT.2022.3206807](https://doi.org/10.1109/TCSVT.2022.3206807).
- [6] T. He, C. Shen, and A. van den Hengel. Dynamic convolution for 3D point cloud instance segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45(5):5697–5711, 2023. doi:[10.1109/TPAMI.2022.3216926](https://doi.org/10.1109/TPAMI.2022.3216926).
- [7] W. Hu, H. Song, F. Zhang, Y. Zhao, and X. Shi. Style transfer of Thangka images highlighting style attributes. *IEEE Access* 11(9):104817–104829, 2023. doi:[10.1109/ACCESS.2023.3318258](https://doi.org/10.1109/ACCESS.2023.3318258).
- [8] J. Huang, W. Yan, G. Li, T. Li, and S. Liu. Learning disentangled representation for multi-view 3D object recognition. *IEEE Transactions on Circuits and Systems for Video Technology* 32(2):646–659, 2022. doi:[10.1109/TCSVT.2021.3062190](https://doi.org/10.1109/TCSVT.2021.3062190).
- [9] G. Kang, L. Jiang, Y. Wei, Y. Yang, and A. Hauptmann. Contrastive adaptation network for single- and multi-source domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44(4):1793–1804, 2022. doi:[10.1109/TPAMI.2020.3029948](https://doi.org/10.1109/TPAMI.2020.3029948).
- [10] J. Li, Y. Xiang, H. Wu, S. Yao, and D. Xu. Optimal transport-based patch matching for image style transfer. *IEEE Transactions on Multimedia* 25(9):5927–5940, 2023. doi:[10.1109/TMM.2022.3201387](https://doi.org/10.1109/TMM.2022.3201387).
- [11] M. Li, J. Wang, Y. Chen, Y. Tang, Z. Wu, et al. Low-dose CT image synthesis for domain adaptation imaging using a generative adversarial network with noise encoding transfer learning. *IEEE Transactions on Medical Imaging* 42(9):2616–2630, 2023. doi:[10.1109/TMI.2023.3261822](https://doi.org/10.1109/TMI.2023.3261822).
- [12] Y. S. Liao and C. R. Huang. Semantic context-aware image style transfer. *IEEE Transactions on Image Processing* 31:1911–1923, 2022. doi:[10.1109/TIP.2022.3149237](https://doi.org/10.1109/TIP.2022.3149237).
- [13] Z. Lin, Z. Wang, H. Chen, X. Ma, C. Xie, et al. Image style transfer algorithm based on semantic segmentation. *IEEE Access* 9(1):54518–54529, 2021. doi:[10.1109/ACCESS.2021.3054969](https://doi.org/10.1109/ACCESS.2021.3054969).
- [14] Y. Liu, X. Kang, Y. Huang, K. Wang, and G. Yang. Unsupervised domain adaptation semantic segmentation for remote-sensing images via covariance attention. *IEEE Geoscience and Remote Sensing Letters* 19(7):6513205, 2022. doi:[10.1109/LGRS.2022.3189044](https://doi.org/10.1109/LGRS.2022.3189044).
- [15] Z. Liu, G. Chen, Z. Li, S. Qu, A. Knoll, et al. D2IFLN: Disentangled domain-invariant feature learning networks for domain generalization. *IEEE Transactions on Cognitive and Developmental Systems* 15(4):2269–2281, 2023. doi:[10.1109/TCDS.2023.3264615](https://doi.org/10.1109/TCDS.2023.3264615).
- [16] Z. Ma, T. Lin, X. Li, F. Li, D. He, et al. Dual-affinity style embedding network for semantic-aligned image style transfer. *IEEE Transactions on Neural Networks and Learning Systems* 34(10):7404–7417, 2023. doi:[10.1109/TNNLS.2022.3143356](https://doi.org/10.1109/TNNLS.2022.3143356).

- [17] A. Mao, C. Dai, Q. Liu, J. Yang, L. Gao, et al. STD-Net: Structure-preserving and topology-adaptive deformation network for single-view 3D reconstruction. *IEEE Transactions on Visualization and Computer Graphics* 29(3):1785–1798, 2023. doi:[10.1109/TVCG.2021.3131712](https://doi.org/10.1109/TVCG.2021.3131712).
- [18] W. Mao, S. Yang, H. Shi, J. Liu, and Z. Wang. Intelligent typography: Artistic text style transfer for complex texture and structure. *IEEE Transactions on Multimedia* 25:6485–6498, 2023. doi:[10.1109/TMM.2022.3209870](https://doi.org/10.1109/TMM.2022.3209870).
- [19] H. Mun, G. J. Yoon, J. Song, and S. M. Yoon. Texture preserving photo style transfer network. *IEEE Transactions on Multimedia* 24(8):3823–3834, 2022. doi:[10.1109/TMM.2021.3108401](https://doi.org/10.1109/TMM.2021.3108401).
- [20] M. Pan, Y. Tang, and H. Li. State-of-the-art in data gloves: A review of hardware, algorithms, and applications. *IEEE Transactions on Instrumentation and Measurement* 72(2):4002515, 2023. doi:[10.1109/TIM.2023.3243614](https://doi.org/10.1109/TIM.2023.3243614).
- [21] T. Shermin, G. Lu, S. W. Teng, M. Murshed, and F. Sohel. Adversarial network with multiple classifiers for open set domain adaptation. *IEEE Transactions on Multimedia* 23:2732–2744, 2021. doi:[10.1109/TMM.2020.3016126](https://doi.org/10.1109/TMM.2020.3016126).
- [22] Y. Tang, M. Pan, H. Li, and X. Cao. A convolutional-transformer-based approach for dynamic gesture recognition of data gloves. *IEEE Transactions on Instrumentation and Measurement* 73(5):2518813, 2024. doi:[10.1109/TIM.2024.3400361](https://doi.org/10.1109/TIM.2024.3400361).
- [23] Q. Wang, S. Li, Z. Wang, X. Zhang, and G. Feng. Multi-source style transfer via style disentanglement network. *IEEE Transactions on Multimedia* 26:1373–1383, 2024. doi:[10.1109/TMM.2023.3281087](https://doi.org/10.1109/TMM.2023.3281087).
- [24] Z. Wang, X. Zhou, Z. Zhou, Y. Zhang, Y. Zhang, et al. MateJam: Multi-material teeth-clutching layer jamming actuation for soft haptic glove. *IEEE Transactions on Haptics* 16(2):276–286, 2023. doi:[10.1109/TOH.2023.3269063](https://doi.org/10.1109/TOH.2023.3269063).
- [25] H. Wu, Y. Han, Q. Zhu, and Z. Geng. Novel feature-disentangled autoencoder integrating residual network for industrial soft sensor. *IEEE Transactions on Industrial Informatics* 19(10):10299–10308, 2023. doi:[10.1109/TII.2023.3240923](https://doi.org/10.1109/TII.2023.3240923).
- [26] K. Wu, M. Wu, Z. Chen, R. Jin, W. Cui, et al. Reinforced adaptation network for partial domain adaptation. *IEEE Transactions on Circuits and Systems for Video Technology* 33(5):2370–2380, 2023. doi:[10.1109/TCSVT.2022.3223950](https://doi.org/10.1109/TCSVT.2022.3223950).
- [27] X. Wu, J. Chen, F. Yu, M. Yao, and J. Luo. Joint learning of multiple latent domains and deep representations for domain adaptation. *IEEE Transactions on Cybernetics* 51(5):2676–2687, 2021. doi:[10.1109/TCYB.2019.2921559](https://doi.org/10.1109/TCYB.2019.2921559).
- [28] J. Zhang, X. Li, H. Li, H. Wang, J. Zhang, et al. Leader-follower control of rehabilitative soft glove based on collaborative sensing and fine motion recognition. *IEEE Sensors Journal* 24(19):30329–30339, 2024. doi:[10.1109/JSEN.2024.3435491](https://doi.org/10.1109/JSEN.2024.3435491).
- [29] R. Zhang, T. Kong, W. Wang, X. Han, and M. You. 3D part assembly generation with instance encoded transformer. *IEEE Robotics and Automation Letters* 7(4):9051–9058, 2022. doi:[10.1109/LRA.2022.3188098](https://doi.org/10.1109/LRA.2022.3188098).
- [30] Z. Zhou, Y. Wu, X. Yang, and Y. Zhou. Neural style transfer with adaptive auto-correlation alignment loss. *IEEE Signal Processing Letters* 29:1027–1031, 2022. doi:[10.1109/LSP.2022.3165758](https://doi.org/10.1109/LSP.2022.3165758).

