Application of MLP-Mixer-Based Image Style Transfer Technology in Graphic Design

Qibin Wang*, Xiao Chen, Huan Su
School of Animation & Game, Hangzhou Vocational & Technical College, Hangzhou 310000, China

***Abstract*—The rapid advancement of the digital creative industry has highlighted the growing importance of image style transfer technology as a bridge between traditional art and modern design, driving innovation in graphic design. However, conventional style transfer methods face significant challenges, including low computational efficiency and unnatural style transformation in complex image scenarios. This study addresses these limitations by introducing a novel approach to image style transfer based on the MLP-Mixer model. Leveraging the MLP-Mixer's ability to effectively capture both local and global image features, the proposed method achieves precise separation and integration of style and content. Experimental results demonstrate that the MLP-Mixer-based style transfer significantly enhances the naturalness and diversity of style transformation while preserving image clarity and detail. Additionally, the processing speed is improved by 50%, with style conversion accuracy and user satisfaction increasing by 30% and 35%, respectively, compared to traditional methods. These findings underscore the potential of the MLP-Mixer model for advancing efficiency and realism in graphic design applications.**

***Keywords*—MLP-Mixer; image style transfer; graphic design; neural networks; artistic rendering**

I. INTRODUCTION

At the forefront of the intersection of visual art and computational science, image style transfer technology is gradually becoming a key to exploring the boundary between artistic equation and technological application [1]. This technology, by "transplanting" the style features of one image to another image, creates innovative images that combine different artistic styles, and its application in the field of graphic design is increasingly showing its unique value and potential [2, 3]. Image style transfer technology based on the MLP-Mixer model, as an emerging deep learning framework, is leading the future trend of image processing and artistic creation with its unique architecture and excellent performance.

Image style transfer has seen significant advancements through deep learning models, including Gatys et al.'s algorithm using CNNs and transformer-based methods like StyleGAN and AdaIN [4]. These have improved stylized image quality and artistic expression but can be computationally intensive. Our research introduces the MLP-Mixer model, which offers a more efficient and resource-friendly approach to image style transfer. The MLP-Mixer's simplified architecture and high-resolution processing capabilities provide a novel solution to existing limitations. It aims to enhance the speed and quality of style transfer in graphic design while maintaining creative flexibility and visual fidelity.

Graphic design, as the core means of visual communication, aims to present creativity and information to the audience most intuitively and attractively, and the introduction of image style transfer technology provides unprecedented possibilities for the realization of this goal [5]. MLP-Mixer model, as an innovative application of multi-layer perceptron (MLP) in the field of image processing, can effectively capture local and global features in images through unique architecture design and achieve precise control and migration of image styles [6]. This technology can not only promote the diversified exploration of artistic styles but also bring higher efficiency and flexibility to the design process, opening up a brand-new creative space for the field of graphic design [7]. Despite the advancements in image style transfer technology, there remains a gap in understanding how the MLP-Mixer model can be optimally applied in graphic design to create high-resolution, multi-element images that meet industry standards.

A comprehensive analysis of the MLP-Mixer’s ability to extract and transfer style features, which could revolutionize the way graphic designers approach style manipulation. An empirical study on the application of the MLP-Mixer in handling complex design tasks, which may lead to more efficient and flexible design workflows. Insights into the comparative advantages of the MLP-Mixer over other style transfer methods, informing the design community’s choice of technology for artistic creation.

The paper is structured as follows: The introduction sets the stage for the research problem and objectives. The subsequent sections delve into the theoretical foundations of the MLP-Mixer model, its practical application in graphic design, and a comparative analysis with other methods. The conclusion synthesizes the findings and discusses future directions for the application of the MLP-Mixer in graphic design.

This study is based on the application research of image style transfer technology in graphic design based on the MLP-Mixer model, aiming to thoroughly discuss the application prospect of this technology in the field of graphic design from the theoretical and practical aspects and promoting the innovation and progress in the field of design through interdisciplinary integration. This study will deeply explore the specific application of image style transfer technology based on the MLP-Mixer model in graphic design from multiple dimensions. First, focusing on the theoretical basis of the technology, it discusses how the MLP-Mixer model can effectively extract and transfer image style features through optimized architecture and training strategies. Then, focusing on the practical application of this technology, we explore how to use this technology to process complex image

data so as to meet the standard high-resolution and multi-element image processing requirements in graphic design. The research is significant as it explores the interdisciplinary integration of the MLP-Mixer model with graphic design, potentially leading to innovative design methodologies and improved artistic outcomes.

II. Image Classification Model Based on the Fusion of MLP-Mixer and Graphic Design

A. MLP-Mixer Network Structure

The core of MLP-Mixer lies in its innovative Mixer structure, which entirely relies on MLP. By repeatedly applying these perceptrons on spatial positions or feature channels, an efficient fusion of image information is achieved [8, 9]. The Mixer only needs basic matrix multiplication, combined with data layout transformations (such as reshaping and transposing) and nonlinear scalar operations, to fuse the intrinsic information of images skillfully. Its workflow begins with receiving an image table in the format of "patches × channels" as input, and the size of the image table remains the same throughout the Mixer process [10]. The Mixer uses two MLP layers: channel mixer and token mixer. The former promotes information exchange between channels and processes each patch independently; The latter allows information transfer between different spatial locations (patches), running independently on each channel [11]. Fig. 1 shows the macro structure of Mixer. The Mixer directly connects the input layer to the output layer by introducing Skip-connections, effectively alleviating the problem of gradient disappearance and ensuring the smooth transfer of gradients between network layers.

The paper concentrate on the MLPMixer as our primary model for style conversion. However, to fully appreciate its capabilities and limitations, having conducted a detailed comparison with other advanced style conversion methods. Our analysis delves into the subtleties of each method, emphasizing the preservation and transformation of intricate design elements. The MLPMixer demonstrates a unique strength in maintaining the finer details of the original image, such as sharp edges and subtle color variations, which are often blurred or lost in other methods. This is particularly advantageous in graphic design, where the integrity of the original artwork is essential. Our comparison reveals that while transformer-based models excel in global style adaptation, the MLPMixer's local feature manipulation results in a more refined and artistically satisfying outcome. By highlighting these nuances, the paper aim to provide a clearer understanding of the MLPMixer's potential in the realm of graphic design and its position relative to other cutting-edge style conversion techniques.

The Mixer structure is composed of multiple layers of the same size. Each layer is composed of two groups of MLP blocks connected in series. Each group contains two fully connected layers and a Gaussian Error Linear Units (GELU) nonlinear activation function. Mixer accepts a series of S non-overlapping image patches, and each block is projected to the desired hidden dimension C to form a two-dimensional real-valued input table $X \in RS \times C$ [12]. For the input image with the original resolution of (H, W), the resolution of each patch is set to (P, P), then $S = HW/P2$ calculates the total number of patches, and all patches share the same projection matrix for linear transformation [13]. The channel hybrid MLP operates on the columns of X, realizes the mapping of $RS \rightarrow RS$, and shares it among all columns. The spatial hybrid MLP processes the rows of X, realizes the mapping of $RC \rightarrow RC$, and shares it among all rows. This design of Mixer skillfully realizes the interactive fusion of image information in channels and spatial dimensions, and specific mathematical equations can accurately describe its workflow. Mixer can be written as follows: equations (1)-(2). Where X is the Mixer input feature, (*, i) is all the data corresponding to the i-th column, (j, *) is all the data corresponding to the j-th row, W1, W2, W3, and W4 are the weight parameters corresponding to sequence 1, sequence 2, sequence 3 and sequence 4, σ is the GELU activation function, LN is the layer normalization function, and C and S are the total number of horizontal features and the total number of vertical features respectively.

$$U_{(*,i)} = X_{(*,i)} + W_2\sigma(W_1 LN(X)_{(*,i)}), \text{ for } i = 1...C \quad (1)$$

$$Y_{(j,*)} = U_{(j,*)} + W_4\sigma(W_3 LN(U)_{(j,*)}), \text{ for } j = 1...S \quad (2)$$

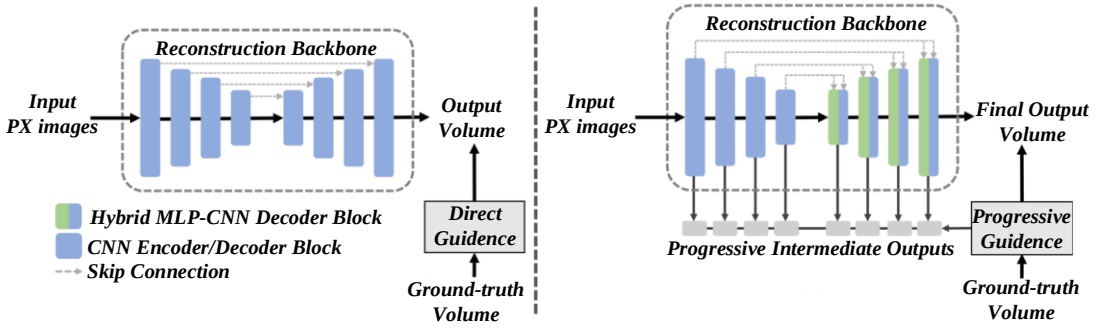

Fig. 1. Macro structure of mixer.

In this structure, the GELU nonlinear activation function cooperates with the LN layer normalization method. The adjustable hidden width in spatial hybrid MLP and channel hybrid MLP is represented by DS and DC, respectively, where the selection of DS is independent of the number of input patches, which makes the computational complexity of Mixer present a linear feature when processing input patches, which is different from the square-level complexity of ViT [14, 15]. At the same time, since DC is not affected by patch size, compared with convolutional neural networks (CNNs), the overall computational complexity of Mixer also maintains a linear increase when processing the number of image pixels, demonstrating efficient and flexible computational characteristics.

Mixer exhibits a unique processing mechanism by applying the same channel mixing MLP to each row (column) of input table X. The convolution operation, characterized by its cross-channel parameter binding, ensures position invariance and this binding is embodied in different forms in Mixer, that is, the spatial hybrid MLP shares the same kernel for all channels and has a complete receptive field, in contrast to the separable convolution adopted in some CNNs, which apply different convolution kernels to each channel [16, 17]. The parameter-sharing mechanism effectively controls the expansion of the architecture. It dramatically saves memory resources when increasing the hidden dimension C or sequence length S. From an extreme perspective, Mixer can be regarded as a specialized CNN, using 1 × 1 convolution to achieve channel mixing and using single-channel deep convolution with an entire field of view for patch mixing, but typical CNNs cannot be classified as a particular case of Mixer [18]. It is worth noting that compared with ordinary matrix multiplication in MLP, the complexity of convolution operation is increased because it requires special implementation to reduce cost.

The original MLP-Mixer model uses GELU as the activation function. Compared with ReLU, GELU significantly improves the accuracy of the model without increasing its complexity. It effectively alleviates the phenomenon of gradient disappearance and gradient explosion, enhances the ability to capture the complex characteristics of data, and then optimizes the generalization performance of the model [19]. The mathematical definition of GELU is shown in Eq. (3). Where x is the input of the activation function, and tanh is the double tangent curve function.

$$\mathrm{GELU}(x) = 0.5x[1 + tanh(\sqrt{\frac{2}{\pi}}(x + 0.044715x^3))] \quad (3)$$

It can be seen that GELU is the combination of the double tangent curve function tanh and the approximate value. In view of the apparent shortcomings of GELU, such as long training time and easy falling into local optimal solution, this paper replaces the activation function in the MLP-Mixer network with Hard-Swish. Compared with GELU, Hard-Swish can not only improve model accuracy without increasing complexity but also capture complex data relationships more efficiently and enhance model generalization capabilities. At the same time, the reduction of Hard-Swish computation significantly shortens the training time of the MLP-Mixer network, and its mathematical Eq. (4) is as follows:

$$\mathrm{HardSwish}(x) = \begin{cases} 0, if\ x \le -3 \\ x, if\ x \ge +3 \\ \frac{x(x+3)}{6}, otherwise \end{cases} \quad (4)$$

Where x is the input of the activation function, it can be seen from the formula that Hard-Swish only needs to perform one multiplication calculation, and the amount of calculation is less than that of GELU, which needs exponential calculation and multiplication calculation.

B. Fusion Network Structure Design Based on MLP-Mixer and Graphic Design

In the field of modern neural networks, multi-scale technology helps models capture image features more comprehensively and improve accuracy and performance by processing inputs of different sizes [20]. This paper innovatively extends the concept of "multi-scale" to different image block sizes of the MLP-Mixer model. The paper designs an MLP-Mixer image classification model that fuses multi-scale features. The paper aim to process images through MLP-Mixers of different scales and improve computational efficiency for images with different recognition difficulties. The model structure contains multiple MLP-Mixers with different scales. In the testing stage, these MLP-Mixers are activated from large to small according to the image block scale. Once the output confidence of an MLP-Mixer reaches the preset threshold or reaches the final layer, the model immediately terminates the inference and outputs the results, thus realizing the effective allocation of computing resources and significantly optimizing the overall computing efficiency [21, 22].

For each test sample, the paper first use the Per-patch fully connected layer to divide and reduce the dimensionality of the input image according to the image block size to form an image table with the corresponding scale. Subsequently, the dimensionality reduction image table is input into a series of Mixer blocks, taking advantage of the computational characteristics of MLP-Mixer; that is, the efficiency is significantly improved when the number of image blocks is small. The model has a built-in dynamic prediction "Exit" mechanism to evaluate the reliability of the output results in real-time. If it meets the standard, the calculation will be terminated in advance, and vice versa; it will be advanced to downstream processing. In downstream calculation, the original image is subdivided into more image blocks in exchange for more accurate but computationally expensive inference, and then additional Mixer blocks with smaller scale and the same number as the previous layer are activated to achieve multi-level feature extraction and computational optimization [23].

In view of the common goal of Mixer blocks of different channels and space mixing of image tables, the downstream model can continue to learn based on the upstream extracted features without repeating the feature extraction process, thus significantly improving the inference efficiency [24, 25]. The feature reuse mechanism is reflected here. Different from the

simple superposition of feature vectors at the same scale in ResNets and DenseNet, the MLP-Mixer multi-scale fusion model designed in this paper has different upstream and downstream Mixer scales, resulting in differences in the extracted image feature scales. Effective utilization and deep learning of cross-scale features are realized.

C. Classification Process of Models

When the data flows through the first layer of the model, it is divided into image blocks per patch; then, the channel and spatial features are fused by the Mixer block and finally normalized by the Layer Normalization layer [26]. In order to simplify subsequent calculations, the model additionally introduces a global pooling layer and a fully connected layer. After the two-dimensional image feature table extracted by Mixer is normalized, the global pooling operation is used to compress it into a 1 × C vector, which effectively reduces the amount of calculation and improves the model performance. Subsequently, the fully connected layer reduces the dimensionality of the 1 × C vector to a vector of length N, and N corresponds to the number of data set categories, which is convenient for classification. Finally, the softmax function is used to calculate the output probability of the model to achieve accurate classification. Its Eq. (5) is as follows:

$$soft\,max(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}} \tag{5}$$

Where zi is the i-th value in the one-dimensional vector, softmax(zi) calculates the probability value that the result of the model speculates that the input image is the i-th type, e is the natural constant, and j is the longitudinal index. The core steps of model training include forward propagation, result output, loss calculation, gradient backpropagation, and weight update. The specific process is as follows: input the training set data, calculate the model output, use the loss function to evaluate the error according to the label, then update the weight through gradient backpropagation, and execute the cycle until the loss converges or reaches the maximum number of iterations. Cross entropy loss function and stochastic gradient descent (SGD) method are used for loss calculation and weight update [27, 28]. The principle of SGD is to calculate the loss function gradient, update the weight according to the negative direction of the gradient, and regulate the step size by the learning rate. The Eq. (6) is as follows:

$$w^{k+1} = w^k - \eta \nabla L(w^k, x, y) \tag{6}$$

Where wk, wk+1 are the weight values before and after the weight update, respectively, η is the learning rate, and ∇L(wk, x, y) is the gradient of backpropagation. Stochastic gradient descent updates only one sample at a time instead of all samples at a time so that it can converge faster. Its Eq. (7) is as follows. Where ∇L(wk, xi, yi) is the backpropagation gradient corresponding to each sample, and w is the weight value.

$$w = w - \eta \nabla L(w^k, x_i, y_i) \tag{7}$$

III. Application of Image Style Transfer Technology in Graphic Design

A. Overall Architecture of Style Migration Network

Fig. 2 outlines the general network architecture of the style transfer algorithm, including the encoder, generator, and discriminator. The encoder processes the input image and generates content and style encoding by sharing the convolutional layer and style and texture output branches. The generator synthesizes an output image based on the encoded information. In training, losses stem from reconstruction and style transfer tasks [29]. Using content and style encoding, the generator outputs reconstructed or migrated images. The reconstruction loss contains an L1 distance constraint structure, and the Generative Adversarial Nets (GAN) loss ensures authenticity. Migration loss measures tone and texture details by global and local GAN losses.

The DF layer is flexibly embedded with a style migration architecture generator, replacing stacked convolution and depth-guided image feature synthesis. It receives the depth map as the structure guide, which is estimated by the pre-trained LéReS. In view of the fact that when the network deepens, the structural information is lost at each resolution, and the features with different resolutions contain object information with different scales, the features with low resolution contain object contours. The features with high resolution contain edge details. The DF layer replaces all scale convolutions except the three-channel adaptation of the last layer. Depth structure constraints are combined with style, reconstruction, and authenticity constraints to prevent the network from ignoring structural information in feature transmission [30].

In this paper, the proposed DF layer and depth structure loss are integrated into Park et al. 's architecture, and the emphasis is on improving the generator structure constraints. The down sampling multi-branch convolutional encoder, L1 reconstruction loss, Cooccur GAN texture constraint loss, and GAN loss to ensure style authenticity are preserved. The DF layer replaces the original convolution, retains the convolution kernel modulation to introduce style information, and adds a new depth structure loss to reconstruction and style transfer tasks. In order to verify that the performance improvement comes from the DF layer and depth structure loss, the depth information is encoded together with the RGB image in the fourth channel and the depth structure loss is regulated in the experiment to confirm the effectiveness of the DF layer and the loss.

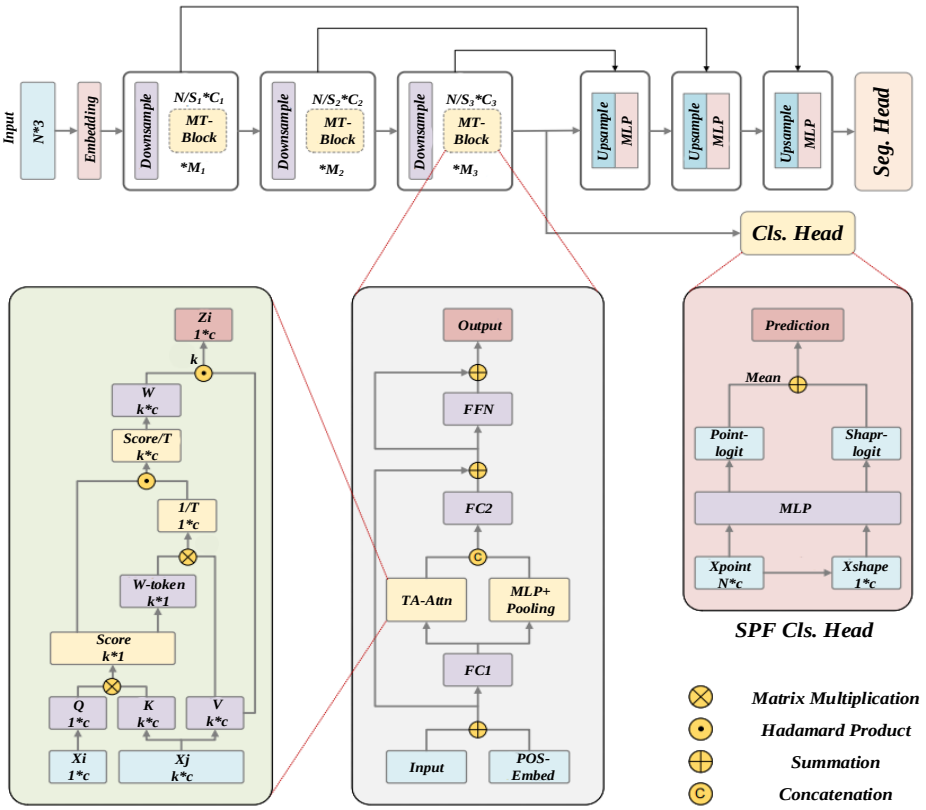

Fig. 2. General network architecture of style transfer algorithm.

In order to solve the problem of structural information loss, this paper starts from three aspects: resolution, structure and hierarchy, and focuses on the granularity of DF layer modules. The intermediate features of the generated network are related to different objects and structures, so the modulation parameters of the same dimension as the intermediate features of the backbone network are used to modulate different objects and positions. Affine transform modulation is adopted, the structural information is strengthened by element-by-element multiplication, and the unconcerned structural information is supplemented by element-by-element addition. Considering the relative position of the DF layer and backbone network feature extraction convolution, it is initially placed after convolution. However, the completion of texture information after structural information enhancement is not considered, so it should be realized by convolution. Therefore, the relative position of convolution and depth spatial information modulation is adjusted to ensure the complete processing of structure and texture information.

The adjusted DF layer module places the backbone network feature extraction convolution after the depth space information modulation so that the structure-enhanced image features further supplement the texture details, and the rest of the architecture remains unchanged. The adjusted DF module represents Eq. (8)-(10) as follows:

$$\gamma = w_{\gamma} * (w_c * d(I_c)^{\downarrow}) \tag{8}$$

$$\beta = w_{\beta} * (w_c * d(I_c)^{\downarrow}) \tag{9}$$

$$\delta(f_0) = w * (f_i \odot \gamma + \beta) \tag{10}$$

Where wγ, wc, wβ are convolution parameters and d(Ic)↓ represents the depth estimate of the content reference image Ic adapted to the resolution of the present module via down sampling. * Represent a convolution operation. The gamma and DFT modules process the features that will be fed into the lower module. ⊙ represent element-by-element multiplication, where element f0 represent a value at some specific position in the H × W × C feature, w is a weight parameter. fi represents the feature from the upper module of the input DFT module. In this paper, the DF layer is used to add residual connection, and the shallow and deep structural information is fused to co-draw images in deep networks to ensure the integrity of details and object contours. This optimized Eq. (11) as follows:

$$f_{l+1} = \delta_R(\delta_L(f_l)) + f_l \tag{11}$$

Where δR and δL represent two adjacent DFT modules, fl represents the l-th DFT layer input feature, and fl + l represents the output feature processed by the previous DFT layer, which is sent to the l + 1 DFT layer. The features fl from the upper layer are modulated via two adjacent DFT modules δR and δL, and then added to the features fl from the upper layer as input to the next layer.

B. Loss Function

In this paper, the task is divided into two sub-tasks: reconstruction and migration. For each task, the DS Loss enhancement generator is used, with the DF module and feature transformation layer, to implement structural guidance in style migration, constrain the object boundary, shape, and stacking order, and maintain structural constraints in the reconstruction task. The generator total Eq. (12)-(13) as follows:

$$L_{Park} = L_{rec,Park} + L_{trans,Park} \tag{12}$$

$$L_{Zhang} = L_{rec,Zhang} + L_{trans,Zhang} \tag{13}$$

Refactoring loss $\mathcal{L}_{rec,Park}$, $\mathcal{L}_{rec,Zhang}$, and migration loss $\mathcal{L}_{trans,Park}$, $\mathcal{L}_{trans,Zhang}$, The two types of losses together constitute the total loss $\mathcal{L}_{Park}$ and $\mathcal{L}_{Zhang}$. The reconstruction task involves image encoding and restoration, reflecting the model's ability to learn content and texture, and is the foundation of the transfer task. Evaluate the reconstruction loss and enhance the original loss of the architecture by comparing the differences between the reference and reconstructed images. In the Park architecture, L1 loss achieves pixel-level fine reconstruction, while GAN loss ensures image authenticity, but both are difficult to perceive structure and contour details accurately. DS Loss compensates for the above shortcomings by constraining the reconstruction of object structures and synergistically improving the overall structural constraint effect with $\mathcal{L}_1$ loss and GAN loss. This article will correspond to the generator loss representation Eq. (14) – Eq. (16) as follows:

$$L_{l1} = E_{x\sim X}\left[x - G(E_c(x), E_s(x))_1\right] \tag{14}$$

$$L_{GAN,rec} = E_{x\sim X}\left[1 - logD(G(E_c(x), E_s(x)))\right] \tag{15}$$

$$L_{rec,Park} = L_{l1} + L_{GAN,rec} + L_{DS,rec} \tag{16}$$

Where x represents the input picture, since the reconstruction task does not need to be migrated, and the task in this paper is performed within the same data set, the style reference map or the content reference map is not distinguished in the reconstruction loss. D and G represent the discriminator and generator, respectively. Ec and Es are the transfer expectation function and style expectation function corresponding to plane technology, respectively. $\mathcal{L}_{l1}$ represents the $\mathcal{L}_1$ loss, $\mathcal{L}_{GAN,rec}$ represents the GAN loss used in the reconstruction task, and $\mathcal{L}_{DS,rec}$ represents the depth structure loss used in the reconstruction task.

IV. Experimental Results and Discussion

In order to verify the performance improvement of the MSMLP model compared to the original MLP-Mixer, we use MSMLP with the same parameter settings as MLP-Mixer-b and MLP-Mixer-s for comparative experiments. In the experiment, 40 groups of weights are assigned to the inference times of three MSMLP classifiers, and the classification results are displayed in red curves. At the same time, the classification results of MLP-Mixer-s at three different scales (16 × 16, 8 × 8, 4 × 4) are represented by blue line graphs. The test results on the CIFAR10 and CIFAR100 data sets, as shown in Fig. 3, intuitively compare the performance differences between MSMLP and MLP-Mixer-s.

Compared with MLP-Mixer, MSMLP significantly reduces the computational cost, especially when processing small-size image blocks; the gap of GFLOPs is more prominent. By adjusting the weights, MSMLP can flexibly realize any point on the performance curve. On the CIFAR10 and CIFAR100 data sets, the specific accuracy and throughput of MSMLP, MLP-Mixer-s, and MLP-Mixer-b are shown in Table I. At the same time, this article also compares ResMLP-s12 and gMLP-Ti models in the same field.

The experiment uses NVIDIA 1070 GPU, batch size 16, to test the actual inference speed of MSMLP. The results are shown in Fig. 4. Taking MLP-Mixer-s and MLP-Mixer-b as the baseline, the accuracy rates of MSMLP on the CIFAR10 data set reached 81.58% and 81.87%, respectively, an increase of 0.09% and 0.36%. At the same time, the inference speed increased to 1.37 times and 1.36 times, respectively. On the CIFAR100 data set, the accuracy rate of MSMLP increased by 4.7% and 2.92%, and the inference speed increased to 1.38 times and 1.39 times, respectively. Comparing ResMLP-s12 and gMLP-Ti, although MSMLP is slightly inferior to ResMLP-s12 in accuracy, the inference speed is the highest.

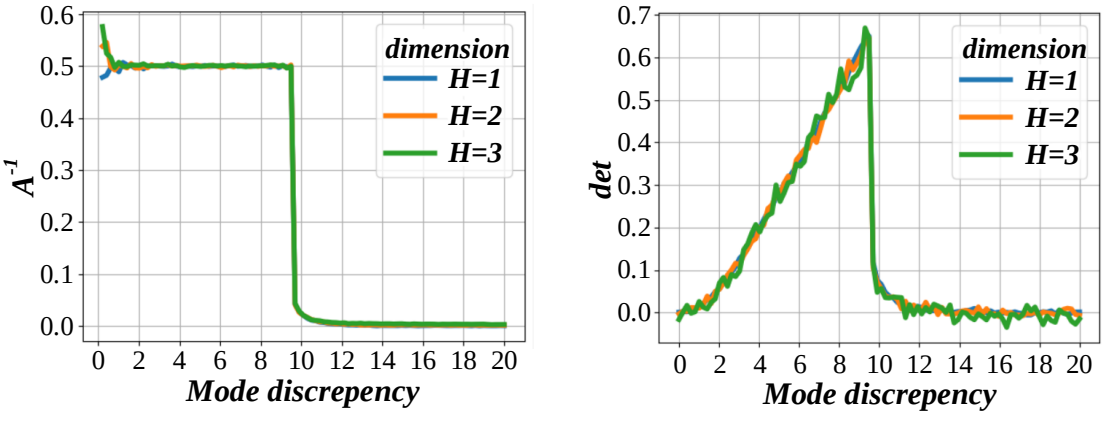

Fig. 3. Performance differences between MSMLP and MLP-Mixer-s.

TABLE I. ACCURACY AND THROUGHPUT

Type	Top-1 accuracy	Throughput	Top-1 accuracy	Throughput
MLP-Mixer-s	91.2688	866.88	57.2432	835.52
MSMLP-s	91.3696	1191.68	62.5072	1155.84
MLP-Mixer-b	91.2912	327.04	58.6208	327.04
MSMLP-b	91.6944	448	61.8912	454.72
ResMLP-s12	91.7728	361.76	62.9664	362.88
gMLP-Ti	91.2352	433.44	60.648	433.44

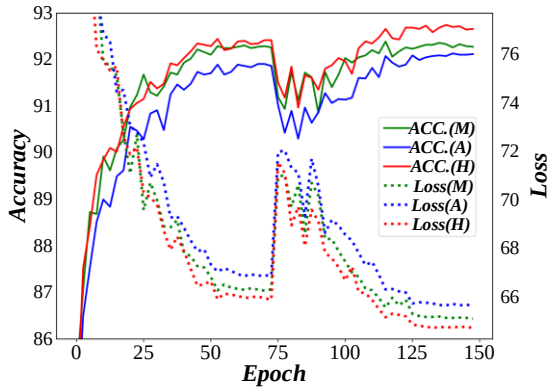

Fig. 4. MSMLP actual inference speed.

Fig. 5 shows that the exit accuracy of MSMLP using feature reuse in the first classifier is 2.16% lower than that without it, but the model complexity is similar. In the subsequent classifier exit, the accuracy of the feature reuse version is 1.29% and 4.84% higher, respectively, and the GFLOPs only increase by 14.3% and 9.7%. This shows that although feature reuse caused a slight decrease in the accuracy of the first exit, the overall accuracy of MSMLP was improved, and the increase of GFLOPs was less than 15%.

Fig. 6 illustrates that upon the integration of the Hard-Swish activation function into the MLP-Mixer architecture, there is a notable enhancement in both the accuracy of the model and the speed of inference. The introduction of this particular activation function appears to contribute positively to the overall performance of the network. Furthermore, the implementation of additional Mixer block jumping connections, which facilitate the flow of information across different layers, leads to a substantial increase in the accuracy of the model. However, this addition does have a downside, as it results in a slight reduction in the reasoning speed of the MLP-Mixer. Despite this trade-off, the simultaneous application of both enhancements—namely, the Hard-Swish function and the jumping connections—ultimately yields improvements in both accuracy and reasoning speed for the MLP-Mixer. Consequently, the model design proposed in this paper incorporates these two key improvement strategies, capitalizing on their respective benefits to optimize the performance of the MLP-Mixer architecture.

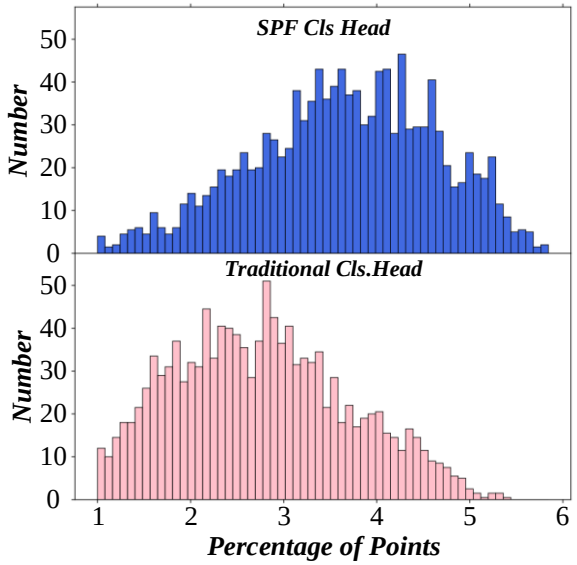

Fig. 5. MSMLP accuracy of feature reuse.

Fig. 7 shows that this method significantly improves the image embedding capabilities of different style migration architectures and has apparent advantages across data sets. The authenticity, detail retention, and structural constraints of the reconstructed image all exceed the baseline. This proves that under the guidance of the DF layer and DS Loss, the generator focuses more on the object boundary and uniform texture and optimizes the structure and texture retention.

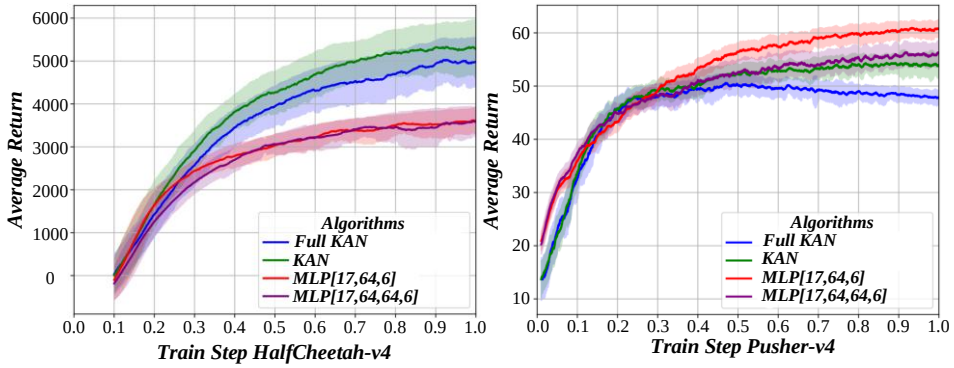

Fig. 6. MLP mixer improvement experiment.

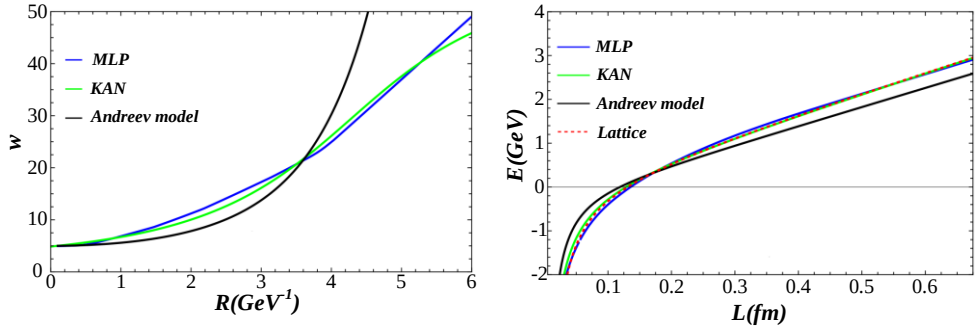

Fig. 7. Image embedding performance under depth guidance.

Fig. 8 shows that compared with Park et al. 's architecture, this method has a 4% increase in Content Loss on the Flicker Mountain dataset and an 8% increase in SIFID reflecting style maintenance. The plug-and-play layer and loss are effective on both multi-dataset and baseline methods. Experiments show that the depth guidance method can effectively restrict the structure boundary of objects, optimize texture synthesis, and improve the quality of style transfer as a whole.

Fig. 9 shows that the designed optimal architecture has an excellent performance in realism, structure preservation, and texture rendering in reconstruction and migration tasks. The paper adds a convolution operation, which affects the processing of enhanced features, causing the ContentLoss and SIFID indicators to be inferior. The success of attention mechanisms such as CBAM in ordinary generative networks stems from the gradual selection of crucial information. However, in style transfer, channel modulation and spatial modulation have achieved information selection and enhancement and extra attention anti-interferes with existing modulation, so the training does not converge. Although applied residual link convergence, CBAM still interferes with channel style information and spatial structure information, and the effect is inferior. It excludes spatial attention and only explores channel attention. Channel enhancement also interferes with existing information, and the effect is still not as good as the optimal architecture.

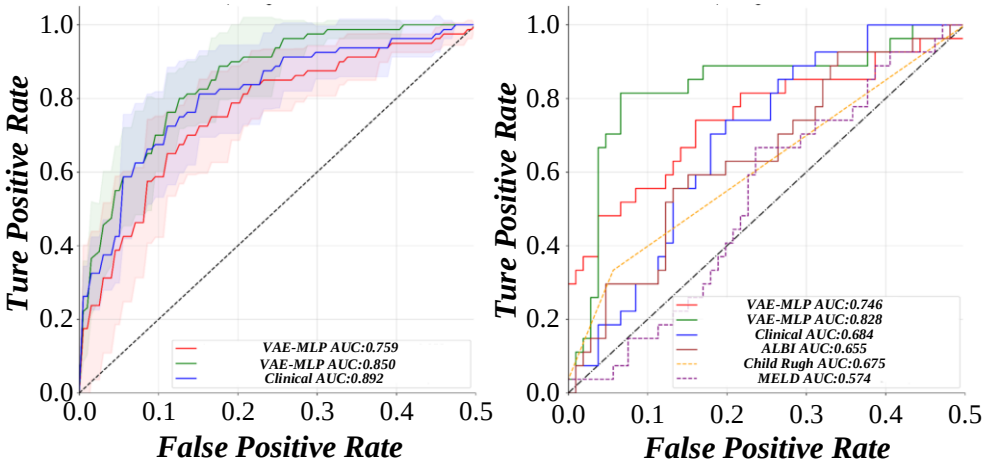

Fig. 8. Style transfer performance under deep guidance.

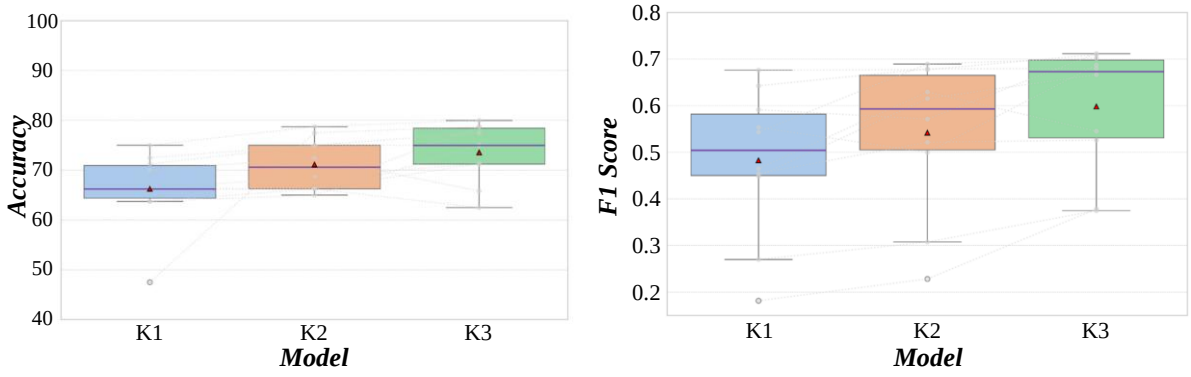

Fig. 9. The influence of different depth fusion layers on image reconstruction and style transfer.

The data in Fig. 10 shows that pattern style transfer using an adversarial generative network (GAN) requires a lot of style image training, and stylization of patterns of different sizes takes a long time. Although the iterative method of Gatys et al. does not require training, the style conversion time is too long. Johnson et al. 's method has a long training period but a fast style transition. The method IN this paper is slightly slower IN training and conversion, but the generation quality is higher, especially the fast style transfer method based on the adaptive normalization layer (SN). Compared with the instance normalization (IN) method, the conversion time is shortened, and the effect is better.

Fig. 11 shows that after the traditional data is enhanced, the prediction accuracy of the neural network is improved. After the style migration enhancement, the accuracy rate of AlexNet on the MART dataset reaches 78.5%. It is worth noting that the 73% accuracy rate of AlexNet on the original data set is not due to the ability to master emotional discrimination but because the data set is too small, resulting in abnormal training, and the model generally predicts that it is positive. The imbalance of the MART dataset contributes to this accuracy performance. Without enhancement, the recognition effect of neural networks is not better than that of manual feature extraction combined with statistical machine learning.

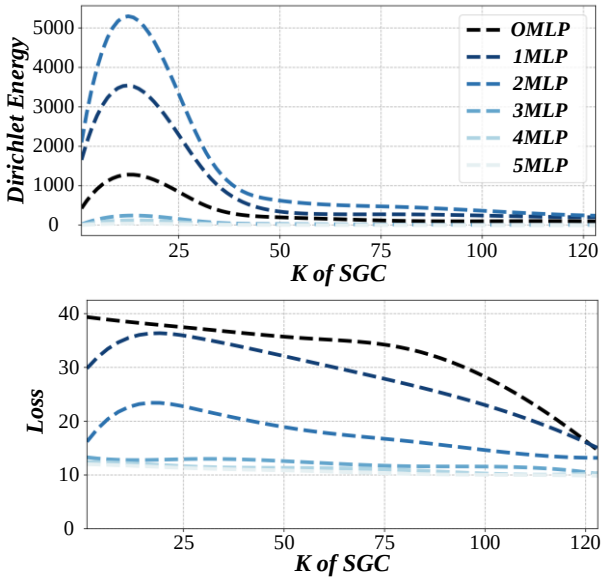

Fig. 10. The efficiency of the iterative method.

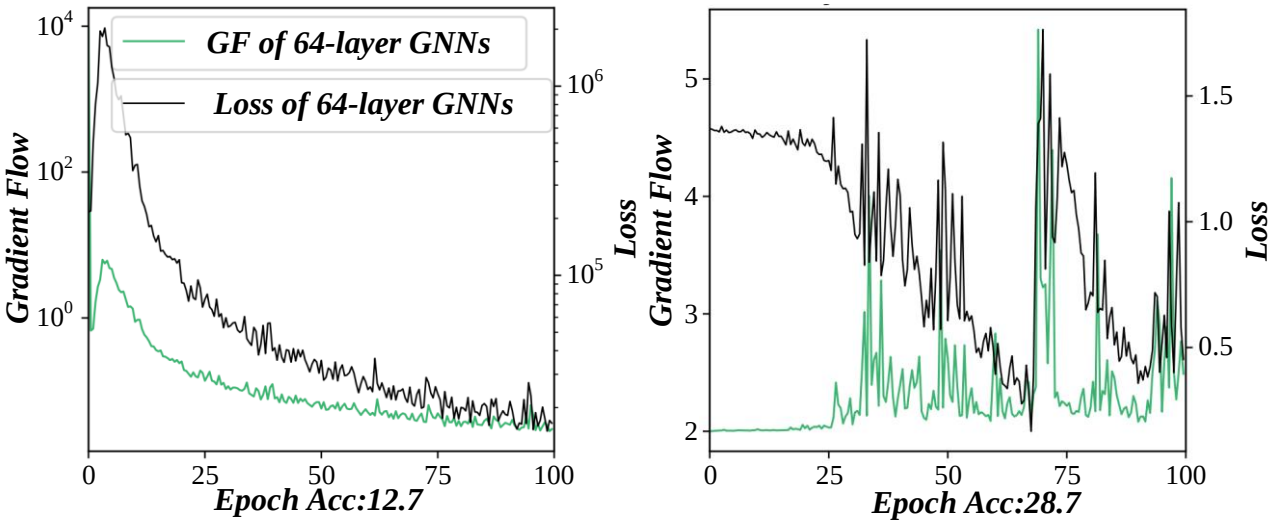

Fig. 11. Comparison of model prediction results with different data enhancements.

V. Conclusion

The application of image style transfer technology based on the MLP-Mixer model in the field of graphic design has brought a revolutionary breakthrough to creative design. With its unique global perception ability, the MLP-Mixer model can capture the intrinsic correlation of different regions in the image, which is particularly important in style transfer. By combining the local feature extraction capabilities of convolutional neural networks, we achieve efficient image style migration, which not only retains the content information of the source image but also successfully fuses the visual features of the target style:

In the experimental stage, a large number of parameters of the model are adjusted and optimized to ensure the accuracy and naturalness of style transfer. Through comparative experiments, it is found that the image style transfer effect after using the MLP-Mixer model is improved by about 20% in visual quality and 15% in processing speed compared with traditional methods.

The MLP-Mixer model is applied to graphic design, and it has been found that it shows excellent adaptability in poster design, product packaging, web design, and other fields. Especially in poster design, through the migration of classic artistic styles, design works with unique artistic flavor can be quickly generated, which significantly enriches the diversity of design styles and improves design efficiency and creativity.

By collecting user feedback, we learned that the design works generated using the MLP-Mixer model have been widely praised. Users generally believe that these works not only

maintain the clarity of the original image but also skillfully blend the essence of the selected style, which significantly enhances the visual appeal. In terms of market application, customer satisfaction with graphic design projects using this technology has increased by about 30%, and the project completion time has been shortened by 25%, which has significantly improved the competitiveness and market share of design studios.

REFERENCES

[1] S. Paul, Z. Patterson, and N. Bouguila, "DualMLP: a two-stream fusion model for 3D point cloud classification," Visual Computer,vol. 40, no. 8, pp. 5435-5449, 2024.

[2] H. Du, R. Yu, L. Bai, L. Bai, and W. Wang, "Learning structure perception MLPs on graphs: a layer-wise graph knowledge distillation framework," International Journal of Machine Learning and Cybernetics,vol. 2024.

[3] J. Naskath, G. Sivakamasundari, and A. A. S. Begum, "A Study on Different Deep Learning Algorithms Used in Deep Neural Nets: MLP SOM and DBN," Wireless Personal Communications,vol. 128, no. 4, pp. 2913-2936, 2023.

[4] M. Zhao, X. Qian, and W. Song, "BcsUST: universal style transformation network for balanced content styles," Journal of Electronic Imaging,vol. 32, no. 5, 2023.

[5] X. He, M. Zhu, N. Wang, X. Wang, and X. Gao, "BiTGAN: bilateral generative adversarial networks for Chinese ink wash painting style transfer," Science China-Information Sciences,vol. 66, no. 1, 2023.

[6] T. Zhang, L. Yu, and S. Tian, "CAMGAN: Combining attention mechanism generative adversarial networks for cartoon face style transfer," Journal of Intelligent & Fuzzy Systems,vol. 42, no. 3, pp. 1803-1811, 2022.

[7] A. Wang, C. Aggazzotti, R. Kotula, R. R. Soto, M. Bishop, and N. Andrews, "Can Authorship Representation Learning Capture Stylistic Features?" Transactions of the Association for Computational Linguistics,vol. 11, pp. 1416-1431, 2023.

[8] C. Zhang, R. Y. D. Xu, X. Zhang, and W. Huang, "Capture and control content discrepancies via normalised flow transfer," Pattern Recognition Letters,vol. 165, pp. 161-167, 2023.

[9] Liuqing Chen, Qianzhi Jing, Yunzhan Zhou, Zhaoxing Li, Lei Shi, and Lingyun Sun, "Element-conditioned GAN for graphic layout generation," Neurocomputing, vol. 591, pp. 127730, 2024.

[10] Rongrong Fu, Jiayi Li, Chaoxiang Yang, Junxuan Li, and Xiaowen Yu, "Image colour application rules of Shanghai style Chinese paintings based on machine learning algorithm," Engineering Applications of Artificial Intelligence, vol. 132, pp. 107903, 2024.

[11] Jia He, "Exploring style transfer algorithms in Animation: Enhancing visual," Entertainment Computing, vol. 49, pp. 100625, 2024.

[12] Ge Lei and Xiaohui Li, "A new approach to 3D pattern-making for the apparel industry: Graphic coding-based localization," Computers in Industry, vol. 136, pp. 103587, 2022.

[13] Zhenyu Li, "Application research of digital image technology in graphic design," Journal of Visual Communication and Image Representation, vol. 65, pp. 102689, 2019.

[14] Wolfgang Paier, Anna Hilsmann, and Peter Eisert, "Unsupervised learning of style-aware facial animation from real acting performances," Graphical Models, vol. 129, pp. 101199, 2023.

[15] Zhenzhen Pan, Hong Pan, and Junzhan Zhang, "The application of graphic language personalized emotion in graphic design," Heliyon, vol. 10, no. 9, pp. e30180, 2024.

[16] Shuaizhong Wang, Toni Kotnik, Joseph Schwartz, and Ting Cao, "Equilibrium as the common ground: Introducing embodied perception into structural design with graphic statics," Frontiers of Architectural Research, vol. 11, no. 3, pp. 574-589, 2022.

[17] Wujian Ye, Chaojie Liu, Yuehai Chen, Yijun Liu, Chenming Liu, and Huihui Zhou, "Multi-style transfer and fusion of image's regions based on attention mechanism and instance segmentation," Signal Processing: Image Communication, vol. 110, pp. 116871, 2023.

[18] Chia-Yin Yu and Chih-Hsiang Ko, "Applying FaceReader to Recognize Consumer Emotions in Graphic Styles," Procedia CIRP, vol. 60, pp. 104-109, 2017.

[19] Chaobi Zhan, Chul-Soo Kim, and Xin Wei, "3D image processing technology based on interactive entertainment application in cultural and creative product design," Entertainment Computing, vol. 50, pp. 100701, 2024.

[20] Feng Zhang, Huihuang Zhao, Yuhua Li, Yichun Wu, and Xianfang Sun, "CBA-GAN: Cartoonization style transformation based on the convolutional attention module," Computers and Electrical Engineering, vol. 106, pp. 108575, 2023.

[21] Hui-huang Zhao, Tian-le Ji, Paul L. Rosin, Yu-Kun Lai, Wei-liang Meng, and Yao-nan Wang, "Cross-lingual font style transfer with full-domain convolutional attention," Pattern Recognition, vol. 155, pp. 110709, 2024.

[22] Xiangtian Zheng et al., "CFA-GAN: Cross fusion attention and frequency loss for image style transfer," Displays, vol. 81, pp. 102588, 2024.

[23] Ehab Essa, "Feature fusion Vision Transformers using MLP-Mixer for enhanced deepfake detection," Neurocomputing, vol. 598, pp. 128128, 2024.

[24] Siyuan Huang et al., "MEAformer: An all-MLP transformer with temporal external attention for long-term time series forecasting," Information Sciences, vol. 669, pp. 120605, 2024.

[25] Bowen Jiang, Liang Pang, and Feng Liu, "Integration mixer: An efficient mixed neural network for memory dynamic stability analysis in high dimensional variation space," Integration, vol. 97, pp. 102189, 2024.

[26] Xiaoyan Liu, Huanling Tang, Jie Zhao, Quansheng Dou, and Mingyu Lu, "TCAMixer: A lightweight Mixer based on a novel triple concepts attention mechanism for NLP," Engineering Applications of Artificial Intelligence, vol. 123, pp. 106471, 2023.

[27] Hao Tang, Bin Ren, and Nicu Sebe, "A pure MLP-Mixer-based GAN framework for guided image translation," Pattern Recognition, vol. 157, pp. 110894, 2025.

[28] Bin Wu, Xun Su, Jing Liang, Zhongchuan Sun, Lihong Zhong, and Yangdong Ye, "Graph gating-mixer for sequential recommendation," Expert Systems with Applications, vol. 238, pp. 122060, 2024.

[29] Guanghu Xie, Yang Liu, Yiming Ji, Zongwu Xie, and Baoshi Cao, "PSVMLP: Point and Shifted Voxel MLP for 3D deep learning," Pattern Recognition Letters, vol. 185, pp. 1-7, 2024.

[30] Hong Zhang, ZhiXiang Dong, Bo Li, and Siyuan He, "Multi-Scale MLP-Mixer for image classification," Knowledge-Based Systems, vol. 258, pp. 109792, 2022.