Heritage Science

RESEARCH Open Access

Styled and characteristic Peking opera facial makeup synthesis with co-training and transfer conditional styleGAN2

Yinghua Shen[1], Oran Duan[1], Xiaoyu Xin[1], Ming Yan[1,2] and Zhe Li[3*]

Abstract

Against the backdrop of the deep integration of culture and technology, research and practice in digitization of intangible cultural heritage has continued to deepen. However, due to the lack of data and training, it is still very difficult to apply artificial intelligence to the field of cultural heritage protection. This article integrates image generation technology into the digital protection of Peking opera facial makeup, using a self-built Peking opera facial makeup dataset. Based on the StyleGAN2 network, we propose a style generative cooperative training network Co-StyleGAN2, which integrates the adaptive data augmentation (ADA) to alleviate the problem of discriminator overfitting and introduces the idea of cooperative training to stabilize the training process. We design a Peking opera facial makeup image transform conditional generation network TC-StyleGAN2 which is transferred from unconditional generation network. The weights of the unconditional pre-training model are fixed, and an adaptive filtering modulation module is added to modulate the category parameters to complete the conversion from unconditional to conditional StyleGAN2 to deal with the training difficulty of conditional GANs on limited data, which suffer from severe mode collapse. The experimental results show that the proposed training strategy is better than the comparison algorithms, and the image generation quality and diversity have been improved.

Keywords Digital protection, Peking opera facial makeup, Image generation, StyleGAN2, Co-training, Transfer learning

Introduction

Peking opera facial makeup, as a traditional Chinese art form, has primarily been studied in China. In the early stages, the research on Peking opera facial makeup mostly focused on analyzing the artistic aesthetic features of facial makeup, building a digital resource library of facial makeup, and elucidating the relationship between facial makeup colors, patterns, shapes and character personalities [1, 2]. With the support of traditional research, it focused on exploring the personality of the corresponding roles in Peking opera facial makeup to achieve the task of personality trait recognition. On the other hand, it started from the visual symbols of Peking opera facial makeup to explore the work of digital modeling. Based on the Big Five personality traits, a subjective questionnaire on the personality traits of Peking opera facial makeup was designed, and it realized the transformation from the semantic space of facial makeup personality traits to the factor space of facial makeup personality [3]. Based on the artistic features of Peking opera facial makeup representing different categories, the database images were divided into categories to train the model, and the model was eventually used to identify the category of

*Correspondence:
Zhe Li
lizhe@hus.osaka-u.ac.jp
[1] School of Information and Communication Engineering, Communication University of China, Beijing 100024, China
[2] Key Laboratory of Acoustic Visual Technology and Intelligent Control System, Ministry of Culture and Tourism, Beijing 100024, China
[3] Graduate School of Human Sciences, Osaka University, Osaka 5650871, Japan

the target Peking opera facial makeup image [4]. A facial makeup pattern library and a facial makeup auxiliary synthesis system were established. Among them, the Bezier curve was used to construct the vector pattern library of facial makeup, and the Free-Form Deformation (FFD) was used to realize the pattern deformation [5, 6]. Regarding the research on the two-dimensional projection of Peking opera facial makeup, video face recognition and video face tracking were used to project and cover the facial makeup on the face, and finally the Peking opera facial makeup was deformed according to different face features, while maintaining the shape of the facial makeup pattern and matching the key parts of the face [7]. Based on the depth camera, the Unity3D engine is used to conduct the research on the three-dimensional personalized modeling of Peking opera facial makeup, which is mainly using the registration and stitching of multi-frame face point clouds, the detection and tracking of face data, and the Laplacian mesh deformation algorithm to realize the three-dimensional modeling and deformation of facial makeup [8, 9].

All the above studies have contributed to the digital development and cultural inheritance of Peking opera facial makeup, but with the deepening of the research, we find that the incompleteness and ambiguity of image resources are still the pain points that cannot be avoided in the development process of Peking opera facial makeup. Few studies have integrated the digital modeling of Peking opera facial makeup with the personality traits implied in Peking opera facial makeup. Only by studying the regularity of color and shape application in facial makeup drawing and the inevitable connection between facial makeup characters' personality and role can we promote the development of Peking opera culture. With the advancement of deep learning in image generation, applying image generation and image translation to Peking opera facial makeup, generating facial makeup patterns that match different character personalities and roles, can generate facial makeup images that inherit cultural genes, which is conducive to the preservation and development of Peking opera culture.

In our previous work, we completed the Peking opera facial makeup generation model, created a Peking opera facial makeup dataset, studied the image generation task under limited datasets, and used different data augmentation methods to train the model, including explicit augmentation methods: geometric transformation, color transformation, and data expansion using pseudo-samples generated by the generator; differentiable augmentation methods: Randomly transforming data using a differentiable forward diffusion process, injecting instance noise, and learning facial makeup features based on the style-based GAN architecture 2nd (StyleGan2) network [10, 11], making the generated Peking opera facial makeup have good local randomness and visual quality. However, there are still some difficulties in using image generation models to generate Peking opera facial makeup. When the dataset is insufficient, the trained network is prone to overfitting, resulting in unstable training in the whole dynamic training process. The effect of using a small amount of facial makeup data on the network model training is not ideal at this stage [12]. Since most image generation models only sample the style of instance images, these networks have poor scalability and portability, and do not consider spatial relevance [13]. In addition, the generation model cannot learn the cultural connotation of Peking opera facial makeup well during training, such as ignoring the personality and role information implied by the facial makeup, resulting in chaotic colors of the generated Peking opera facial makeup, which do not conform to the rules of Peking opera facial makeup.

To solve these problems, we start from the perspective of multiple discriminators and introduce the idea of cooperative training into StyleGAN2 to improve the discriminative ability of the discriminator. We design an adaptive filtering modulation to adjust the category parameters to deal with the training difficulty of conditional Generative Adversarial Network (GAN) on limited data, which suffer from severe mode collapse. We finally integrate the generation model to form a complete unconditional Peking opera facial makeup generation model, and transfer the unconditional StyleGAN2 to the conditional StyleGAN2 of Multi-target Domains, resulting in a conditional generation model that can generate seven types of Peking opera facial makeup images. The evaluation indicators in the objective experiments are better than the current mainstream facial makeup generation model baseline. Our main contributions are as follows:

(1) We construct a style generative cooperative training network Co-StyleGAN2. We use adaptive data augmentation to alleviate the overfitting of the discriminator at the input end. On the other hand, we introduce the idea of cooperative training, adopt a dual discriminator structure, and improve the generation performance by enhancing the feature extraction ability of the discriminator. We train and test on our self-built Peking opera facial makeup dataset, and compare the generation effect with the current excellent image generation models, verifying the effectiveness of our algorithm.

(2) We design a Peking opera facial makeup image conditional generation network TC- StyleGAN2 based on StyleGAN2. We freeze the weights of the pre-

trained unconditional generation model, design an adaptive filtering modulation to adjust the category parameters, and use the modulation to transform from an unconditional network to a conditional network. We train and test on our self-built category Peking opera facial makeup dataset, which not only solves the problem of mode collapse, but also generates facial makeup with accurate category attributes.

(3) We demonstrate the practicality of our method in the objective evaluation experiments. The experimental results show that our model is superior to the current mainstream models in the facial makeup generation task, and significantly improves the indicators such as style attribute consistency and visual fidelity of the facial makeup.

Related work

Generative models The current mainstream generative models include Variational Auto-Encoder (VAE) [14], GAN [15], and Diffusion Model (DM) [16]. All three have advantages and disadvantages. VAE suffers from the problems of blurry generated images and insufficient randomness of target images; GAN network inevitably loses the diversity of generated images while ensuring the realism of generated images; Diffusion model is slow in the sampling process. In recent years, the most influential model with realistic generation quality is the continuously improved GANs. However, a study proposed in May 2021 showed that the image samples generated by the diffusion model are of higher quality than the current state-of-the-art generative models based on GANs [17]. Taking the synthetic images of face generation task as an example, compared with GANs, the images generated by the diffusion model are more diverse and realistic in details, such as face pose, race, and whether wearing glasses, and the single network constructed is more stable than the adversarial network of GANs. Although the diffusion model has great advantages in image generation, its sampling speed is slow during training and inference, resulting in very high training costs. Therefore, considering factors such as training cost and image generation quality, we prefer image generation models based on GANs.

GANs, StyleGAN and its variants In 2014, GAN [15] introduced a novel data generation approach inspired by game theory, where the generator and discriminator are trained to compete with each other to achieve a Nash equilibrium. DCGAN [18] proposed in 2016 is the pioneering work of face generation task, which uses convolutional neural network (CNN) to replace the multilayer perceptron in the original GAN, stabilizing the model training process. Progressive Growing of GANs (PGGAN) [19] uses progressive growing network to achieve high-resolution image generation. But PGGAN is prone to feature entanglement, that is, it has very limited ability to control specific features to generate images. StyleGAN [20] proposed in 2019 adds a mapping network to disentangle the input vector on the basis of PGGAN, and uses adaptive instance normalization module to precisely control the style information. Therefore, StyleGAN can not only generate high-resolution and realistic images such as faces, but also can better control the generated images. In addition, the variants of StyleGAN include StyleGAN2 [21] and StyleGAN3 [22], which achieve excellent generation effects in the field of unconditional image generation. More and more studies [23–26] are taking advantage of the powerful image generation capabilities of GANs models to expand small sample datasets by generating more images, which largely solves the problem of image category imbalance in image classification training and lack of data in target detection tasks [27, 28].

Cooperative Training Cooperative Training was originally proposed to improve the classification recognition performance in the case of limited data. Ning X et al. [29] and Multimodal co-learning [30] use multiple classifiers to learn and capture complementary information about limited data from different views, thus effectively alleviating data constraints. Based on this assumption, Cooperative Training [31] proposes a dual-view training algorithm, and further shows that under the assumption that the two views of each instance are conditionally independent given the category, Cooperative Training has a similar Probably Approximately Correct (PAC) [32] guarantee for semi-supervised learning. In recent years, the idea of Cooperative Training has been introduced into various deep network training tasks, mainly applied to image classification and semantic segmentation. For example, to avoid the ideal conditions required by the dual-view training algorithm, Peng J et al. [33] proposes a deep Cooperative Training technique that transforms from multiple views to multiple learners, and encourages view diversity by training multiple deep neural networks in semi-supervised image recognition tasks. Ma Y et al. [34] and Shahbazi M et al. [35] use the idea of cooperative training in semantic segmentation task to align the feature categories between source domain and target domain, and use the divergence of two classifiers in the prediction of target domain to optimize the classification decision boundary. In our work, we have referred to the cooperative training idea to build and train the network, which helps us adopt a dual discriminator structure and enhance the feature extraction ability of the discriminator.

Method

Unconditional generation

Adaptive data augmentation module

If data augmentation is directly applied to the training task of StyleGAN2, it may cause the gain to leak to the generator. For example, if the data augmentation adds rotation, the generator will also generate rotated images. AmbientCycleGAN [36] conducted data perturbation experiments on GAN and concluded that if the data transformation is reversible on the probability distribution, then the model training will bypass the data perturbation and find the correct distribution. Therefore, it does not affect the generator training, and cause the gain leakage problem. Specifically, the strength of data augmentation is defined as a scalar$p$, and controls $p \in [0,1)$, then the data augmentation can become reversible in the training of StyleGAN2. The following will introduce the adaptive control method of data augmentation strength p in detail.

Since StyleGAN2 training is in a dynamic equilibrium state, the enhancement strength can be dynamically controlled according to the overfitting degree of the discriminator, which can avoid the complexity and computational cost brought by manual adjustment. To quantify the degree of overfitting, we analyze the application of different data augmentation methods in GANs, focusing on a series of reasonable heuristics derived from the original output logic of the discriminator, and adapt the enhancement strength according to the heuristic matching a suitable target value.

ADA [37] uses standard geometric and color transformations as data augmentation methods, and introduces a probability-based adaptive strategy to stabilize the training process. Specifically, a heuristic algorithm is defined to estimate the part of the training set that obtains positive discriminator output. The overfitting heuristic is as follows:

$$r_t = \mathbb{E}[\mathrm{sign}(D_{train})] \tag{1}$$

where D_{train} represents the discriminator output of the training set, and $\mathbb{E}[\cdot]$ represents the average of N consecutive mini-batch discriminator outputs. In practice, N is set to 4, *Batch_size*=16, that is the average of 64 images. For this heuristic, r_t= 0 means no overfitting, and r_t= 1 means complete overfitting.

The strategy for using r_t to adjust p is as follows: first, set a threshold t, and initialize p to zero. If r_t indicates too much or too little overfitting about t (i.e., greater or less than t), then the probability p will increase or decrease by a fixed step. The experiment is set up to adjust p every four mini-batches, and clamps p to 0 after each adjustment. In this way, the data augmentation strength can be adaptively controlled according to the degree of overfitting.

Since GAN itself has a powerful image generation ability, using the images generated by the generator in GAN is also a natural and feasible data augmentation scheme. APA [38] extends ADA by integrating the generated data with the standard data transformation methods as a new data augmentation method, and defines a new heuristic to estimate the part of the real images that obtain positive logit predictions from the discriminator. The formula is as follows:

$$\lambda_r = \mathbb{E}\big(\mathrm{sign}(D_{real})\big) = \mathbb{E}\big(\mathrm{sign}(\mathrm{logit}(D(\mathbf{x})))\big) \tag{2}$$

where λ_r is consistent with the strategy of r_t adjusting p mentioned above, except for the symbol difference. Since the generated data is treated as real data for training and passed to the discriminator, the optimization objective is updated as follows:

$$\begin{aligned}\min_G \max_D V(G,D) =&(1-\alpha)\mathbb{E}_{\mathbf{x}\sim p_r}[\log D_\theta(\mathbf{x})]\\ &+\alpha\mathbb{E}_{\mathbf{z}\sim p_z}\left[\log D_\theta(G_\phi(\mathbf{z})))\right]\\ &+\mathbb{E}_{\mathbf{z}\sim p_z}\left[\log(1-D_\theta(G_\phi(\mathbf{z})))\right]\end{aligned} \tag{3}$$

where α is the expected strength of the dynamic adjustment effect in the whole training process. Since $p \in [0,1)$, it is stipulated that $0 \leq \alpha < p_{max} < 1$, where p_{max} is the maximum probability of adding generated data in the whole training process.

In addition to the data augmentation methods mentioned above, Diffusion GAN [39] studied adding noise to the input data of the discriminator, and proposed a new adaptive differentiable data augmentation method to alleviate the overfitting problem of the discriminator. It combines the diffusion model and injects noise from the Gaussian mixture distribution to achieve data transformation (the Gaussian mixture distribution is composed of weighted diffusion samples from clean images at different time steps).

$\forall t \in \{1,\cdots,\mathcal{T}\}, \mathcal{T} = \mathcal{T} + \mathrm{sign}\big(r_t - d_{target}\big) * C$. As the step length t increases, the noise-to-data ratio increases, making the task of the discriminator more and more difficult. Diffusion GAN designed the adaptive control of diffusion intensity, and modified:

$$r_d = \mathbb{E}_{\mathbf{y},t\sim p(\mathbf{y},t)}\big[\mathrm{sign}\big(D\big(\mathbf{y},t\big) - 0.5\big)\big] \tag{4}$$

where r_t is as shown in Eq. (1), the hyperparameter d_{target} is a threshold for identifying whether the current discriminator is overfitting (referring to the threshold t set by ADA [37] and APA [38], which is set to 0.6 in the experiment) C is a constant.

To encourage the discriminator to observe the newly added diffusion samples when $\mathcal{T}$ increases, the distribution of the step length t is also defined:

$$t \sim p_\pi := Discrete\left(\frac{1}{\sum_{t=1}^{T} t}, \frac{2}{\sum_{t=1}^{T} t}, \cdots, \frac{T}{\sum_{t=1}^{T} t}\right) \tag{5}$$

We match a suitable target value by heuristic to achieve the purpose of adaptively adjusting the enhancement strength of different data augmentation methods. This study uses an adaptive data augmentation module to process the input data, which avoids the generator leakage problem in an adaptive way, and uses data augmentation to alleviate the overfitting problem of the discriminator.

Style generative cooperative training network

We design a style generative cooperative training network (Co-training StyleGAN2, Co-StyleGAN2), as shown in Fig. 1, which applies the cooperative training idea to the image generation with limited data, and learns from multiple different but complementary views of the limited data to balance the training process.

We mainly design two Co-StyleGAN2 instances, which enable the discriminators to learn from unique and comprehensive views. As shown in Fig. 1, the architecture consists of four modules: Image Sampling, Image Generation, Weight-variance Co-training StyleGAN2 (WCSG), and Data-variance Co-training StyleGAN2 (DCSG). Image Sampling samples images x from the limited training data, and Image Generation module uses generator G to generate images G(z). x and G(z) are fed into WCSG and DCSG to jointly train the discriminators D_1 and D_3, where different views of x and G(z) are generated by random frequency component suppression module R and fed into D_3.

The first one is Weight-Variance Co-training WCSG, which is designed to jointly train two different discriminators by reducing the weight difference and diversifying the parameters. The second one is Data-Variance Co-training DCSG, which is designed to jointly train different discriminators by inputting different views of the images. Specifically, different frequency components of each training image are extracted to generate different views.

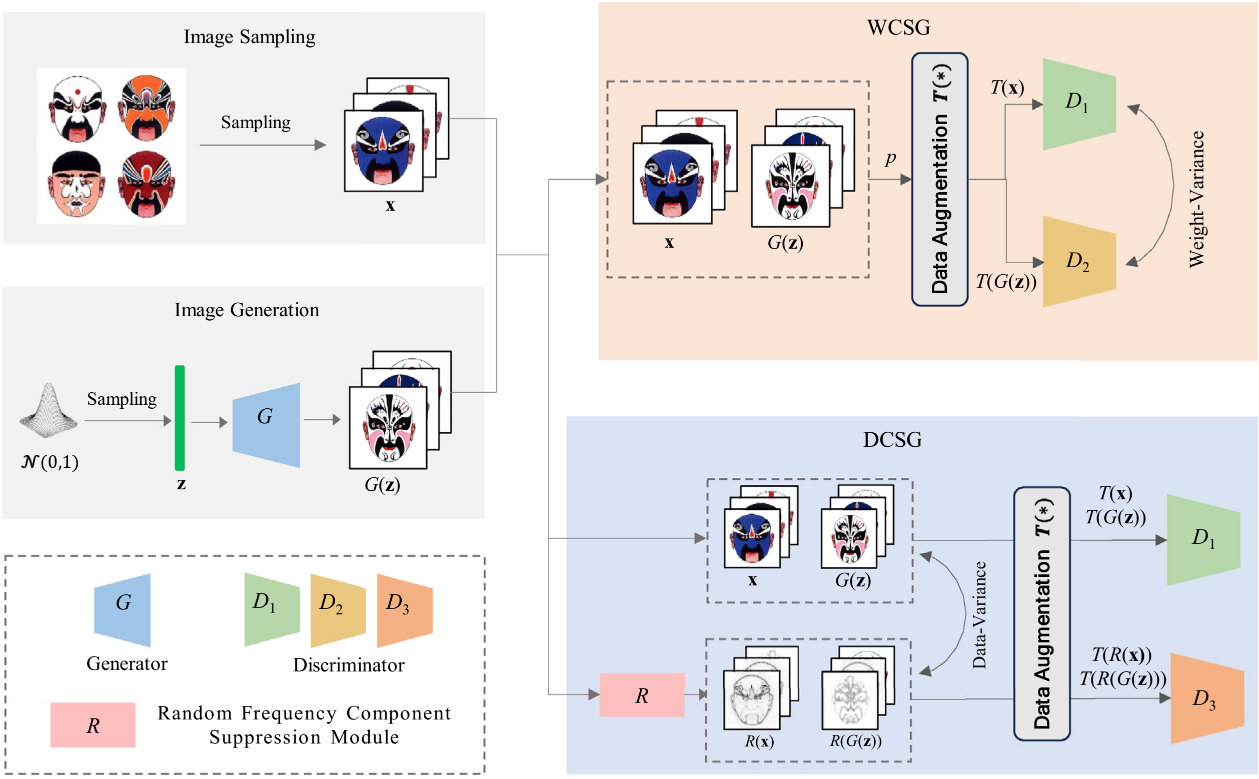


Fig. 1 Style generative cooperative training network architecture. We design two instances: Weight-variance Co-training StyleGAN2(WCSG) and Data-variance Co-training StyleGAN2(DCSG), discriminators are trained separately by diversifying the parameters and inputting images from different perspectives

WCSG aims to learn two different discriminators D_1 and D_2 by diversifying the parameters. This study achieves different parameter learning by defining a weight difference loss that minimizes the cosine distance between the weights of D_1 and D_2:

$$L_{wd}(D_1, D_2) = \frac{\vec{W_{D_1}}\vec{W_{D_2}}}{\left|\vec{W_{D_1}}\right|\left|\vec{W_{D_2}}\right|} \tag{6}$$

where $\overrightarrow{W_{D_1}}$ and $\overrightarrow{W_{D_2}}$ are the weights of D_1 and D_2, and L_{wd} controls the distance between the two sets of parameters by calculating the cosine value of the angle between the two sets of weight vectors. The smaller L_{wd} is, the more diverse the weight parameters become, which suppresses the overfitting of the discriminator to some extent. Because when one tends to overfit, the weight difference loss will slow down the overfitting of the other discriminator. Therefore, the losses of D_1 and D_2 can be formulated:

$$L_{D_1}^{WCSG} = L_{uc}^{D}(D_1; \mathbf{x}, G(\mathbf{z})) \tag{7}$$

$$L_{D_2}^{WCSG} = L_{uc}^{D}(D_2; \mathbf{x}, G(\mathbf{z})) + L_{wd}(D_1, D_2) \tag{8}$$

where L_{uc}^{D} is the general discriminator loss, and L_{wd} is the weight difference loss defined in Eq. (6). The weight difference loss is applied on D_2 to ensure that the two discriminators learn different parameter information. Therefore, the total loss of WCSG is as follows:

$$L_{\mathrm{WCSG}} = L_{D_1}^{WCSG} + L_{D_2}^{WCSG} \tag{9}$$

DCSG designs to feed different views of the input image to two different discriminators D_1 and D_3 for data-variance co-training. Specifically, D_1 is fed with the original input image, while D_3 takes the input image processed by the Random Frequency Component Suppression Module as input.

The Random Frequency Component Suppression Module consists of three processes and shown in Fig. 2. First, the image transformation and decomposition process R_t, which transforms the image from spatial representation to frequency representation, and decomposes the transformed image into multiple frequency components (FCs). Then, the random frequency component suppression process R_r, which randomly suppresses some FCs. Finally, the image reconstruction process $R_{t^{-1}}$, which connects the remaining FCs and transforms the image back to spatial representation.

Specifically, R_t first uses Fast Fourier Transform (FFT) to transform the image x, from spatial representation $x \in \mathcal{R}^{H\times W\times C}$ to frequency domain representation $x_f \in \mathbb{C}^{H\times W\times C}$. In this experiment, x contains real image $\mathbb{X}$ or generated image $G(z)$. Then, it uses band-pass filter $\mathcal{B}_p$ to decompose $x_f \in \mathbb{C}^{H\times W\times C}$ into multiple frequency components $x_{fc} \in \mathbb{C}^{H\times W\times C\times N}$, where $\mathbb{C}$ represents complex numbers. The formula of $R_t(\cdot)$ is as follows:

$$x_{fc} = R_t(x) = \mathcal{B}_p(FFT(x)) = \{x_{fc}(0), x_{fc}(1), \ldots, x_{fc}(N)\} \tag{10}$$

where $x_{fc}(i) \in \mathbb{C}^{H\times W\times C}$ represents each FC, and N is the number of decomposed FCs (default is 64). This paper constructs the band-pass filter $\mathcal{B}_p$ according to the literature [40].

The random FCs suppression process R_r uses a band-stop filter to randomly suppress some FCs in x_{fc}, while the rest of the FCs remain intact. The band-stop filter $\mathcal{B}_r(\cdot)$ is defined as follows:

$$x\prime_{fc} = B_r\left(x_{fc}, I\right) = \begin{cases} 0, & if\ Ii = 0, \\ x_{fc}(i), & if\ Ii = 1, \end{cases} \tag{11}$$

where $x\prime_{fc} \in \mathbb{C}^{H\times W\times C\times N}$ is the FC representation after randomly suppressing frequency components, and I is a binary mask, where ’0’ indicates suppression and ’1’ indicates retention of the corresponding FC. In this process, the values in I are randomly generated, and the

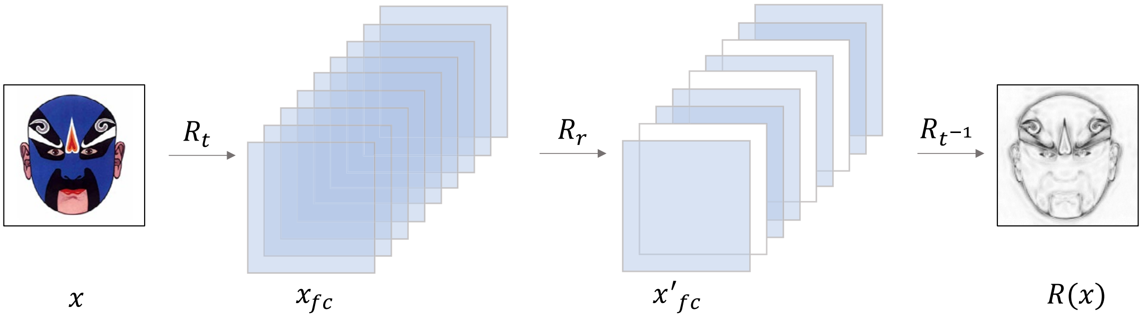


Fig. 2 The random frequency component suppression module

percentage of '0' is controlled by a hyperparameter P, which is set to 0.2 according to experience.

The image reconstruction process R_{t-1} is the inverse process of R_t, which concatenates FCs in $x\prime_{f_c}$ and converts the image back to the spatial representation. Therefore, the loss functions of D_1 and D_3 can be defined by the following formula:

$$L_{D_1}^{DCSG} = L_D(D_1; \mathbf{x}, R(G(\boldsymbol{z}))) \tag{12}$$

$$L_{D_3}^{DCSG} = L_D(D_3; R(\mathbf{x}), R(G(\mathbf{z}))) \tag{13}$$

The difference between the losses of D_1 and D_3 is mainly controlled by the different inputs from the random frequency component suppression module. The overall loss of DCSG is as follows:

$$L_{\mathrm{DCSG}} = L_{D_1}^{DCSG} + L_{D_3}^{DCSG} \tag{14}$$

A single discriminator may overfit and focus on learning simple structures and patterns. This paper proposes a dual-discriminator training method for StyleGAN2 by introducing the idea of cooperative training. The new discriminator has the same internal structure as the original discriminator, but is encouraged to learn different parameters or input data. D_2 in WCSG and D_3 in DCSG learn different information, complementing the original discriminator D_1, to pay attention to different types of information, improve the discriminator's discrimination ability, and stabilize the training process.

Conditional generation

Unconditional StyleGAN2 can generate high-fidelity and diverse images through a stable training process on the same data. However, conditional StyleGAN2 trained on the limited data suffers from severe mode collapse. When mode collapse occurs, it generates images with weak diversity and similar features at best, and extremely similar and distorted images at worst. This part aims to alleviate the mode collapse problem and achieve image conditional generation using a single generative model.

The strategy of transformation from unconditional to conditional generation

To transition from unconditional $G(\boldsymbol{z})$ to conditional $G(\boldsymbol{z}, \boldsymbol{c})$, the generator needs to change the discrete architecture during the training process. By defining a transition function λ_t (the subscript t indicates the number of iterations during training), the generator is modulated to $G(\boldsymbol{z}, \boldsymbol{c}, \lambda_t)$ to avoid tedious system modifications. λ_t is defined as follows:

$$\lambda_t = \min\left(\max\left(\frac{t - T_s}{T_e - T_s}, 0\right), 1\right) \tag{15}$$

where, T_s and T_e respectively represent the time steps when the transition starts and ends. The transition function $\lambda_t \geq 0$ controls the transition from unconditional training to conditional training. $\lambda_t = 0$ means a purely unconditional learning, that is, the conditional information does not affect the generator and discriminator networks. $\lambda_t = 1$ means that the current training has completely become a conditional generation state.

During the training, as λ_t increases, the conditional information is gradually merged by using the following form of generator:

$$G(\mathbf{z}, \mathbf{c}, \lambda_t) = G(S(\mathbf{z}) + \lambda_t \cdot E(\mathbf{c})) \tag{16}$$

where S and E are neural network modules that transform the latent vector and the conditional vector respectively.

Considering the large amount of training data required for conditional training and the instability of the performance during the process, we propose a method of transferring pre-trained unconditional StyleGAN2 to conditional StyleGAN2 network (Transfer Conditional StyleGAN2, TC-StyleGAN2). Compared to transforming into multiple unconditional StyleGAN2 generation tasks, transforming into a multi-condition StyleGAN2 can not only reduce the time and resource cost of training multiple models, but also the advantage of generating multiple categories with one model is that multiple classes share weights during training, thereby using the similarity between different categories to improve the image generation quality.

First, we define the unconditional dataset as the source domain D_s and the conditional dataset as the multi-class target domain D_t. Given the pre-trained generative model $f_0(\cdot)$ for the source domain, $f_0(\cdot)$ is chosen as the unconditional Peking opera facial makeup generation model with the best generation quality. The goal of this study is to use transfer learning based on $f_0(\cdot)$ to generate weight networks $f_t(\cdot)$ for all classes in the target domain. To transform unconditional StyleGAN2 into conditional StyleGAN2, this study introduces class-specific parameters, which are modulated by the forward pass of the generator, to push the unconditional generative model towards the distribution of each target class. Next, we explain the basic principle of the adaptive filtering modulation (mAdaFM) [41] applied to the class parameters.

Adaptive filtering modulation first removes the source style encoded by μ and σ, and then applies the style learned from γ_i and β_i to model the statistics of the target distribution generation process. In this study, we apply this modulation to solve the problem of transfer learning across multiple domains. The network weights

W and b are shared among all transfer classes, while the modulation parameters γ_i, β_i and b_i are the only changing parameters. By adjusting the modulation parameters, the unconditional basis network can be transformed into a conditional network.

Specifically, given the source domain—the fully connected layer $h_s(\mathbf{x}) = W\mathbf{x} + b$ of the unconditional pre-trained generative model, the pre-trained model's weights $W \in \mathbb{R}^{d_{out} \times d_{in}}$ and input $\mathbf{x} \in \mathbb{R}^{d_{in}}$. To transfer to the target domain—the conditional generative model, this study readjusts its statistics to form different layers, as follows in Eq. (17):

$$\hat{W}_i = \gamma_i \odot \frac{W - \mu}{\sigma} + \beta_i, \hat{b}_i = b + b_i \tag{17}$$

where, $\gamma_i, \beta_i \in \mathbb{R}^{d_{out} \times d_{in}}$ are both learned parameters, $i = 1, \cdots, N_c$ represents the class label, N_c is the number of classes (in the experiments of this study, $N_c = 7$), μ and σ are respectively the mean and standard deviation of W.

Hypernetwork design

The purpose of the modulation class parameters is to optimize for each class in the target domain, yet these modulation parameters are not shared across different classes. To solve this problem, this study uses Hypernetworks [42], which facilitates information sharing and reduces memory consumption by accumulating knowledge in a newly introduced module.

In this study, we apply the hypernetwork H to conditionally predict the modulation parameters for each target generative model. The input of the hypernetwork H is from the class embedding network $\mathcal{C}(i) = \mathbf{c} \in V$, where V is the class embedding space. The hypernetwork H takes the embedding vector $\mathbf{c}$ and maps it to the modulation parameters, as demonstrated by:

$$\begin{aligned} &\gamma_{\mathbf{c}}, \beta_{\mathbf{c}} = \mathrm{H}(\mathbf{c}), \\ &b_{\mathbf{c}} = \mathrm{H}_b(\mathbf{c}). \end{aligned} \tag{18}$$

where H is an affine transformation in the space V, and each modulation layer is set to have an H module, to achieve the purpose of sharing parameters between target classes.

Thus, the modulation formula for generating target-specific activations $h_s(\mathbf{x}) = W\mathbf{x} + b$ from the source domain—the fully connected layer $h_{\mathbf{c}}(\mathbf{x}) = \widehat{W}_{\mathbf{c}}\mathbf{x} + \widehat{b}_{\mathbf{c}}$ of the unconditional pre-trained model is as follows:

$$\begin{aligned} &\hat{W}_{\mathbf{c}} = \gamma_{\mathbf{c}} \odot \frac{W - \mu}{\sigma} + \beta_{\mathbf{c}}, \\ &\hat{b}_{\mathbf{c}} = b + b_{\mathbf{c}}. \end{aligned} \tag{19}$$

where W and b are the frozen source weights. Finally, the generative weight networks f_t for all classes in the target domain are assigned with the embedded class $\mathbf{c}$ and the normalized source weights $\widetilde{w}$, in order to produce the desired target weights.

$$f_t(\cdot) = f_{\widetilde{w}}(\mathbf{c}) = \gamma_{\mathbf{c}} \odot \widetilde{w} + \beta_{\mathbf{c}} = \widehat{w}_{\mathbf{c}}.$$

We propose a hypernetwork architecture TC-StyleGAN2 that transforms unconditional StyleGAN2 to conditional StyleGAN2, as shown in Fig. 3, which improves the generator part based on the original conditional StyleGAN2. Combining the good generation performance of the unconditional StyleGAN2 generative model and the training method of StyleGAN2 that focuses on the whole first and then supplements the details, and the discriminator's sensitivity to contours and colors, we propose to first use an unlabeled dataset to train a good unconditional generator; then, combining the idea of knowledge transfer, we freeze the learned source weights and transfer them to an untrained conditional generative model (corresponding to the A and B modules in the dashed box in Fig. 3, representing the source domain weight freezing), and at the same time, the class embedding network $\mathcal{C}(i)$ controls the class information through the affine H and modulation (corresponding to the H and Mod modules in the solid box in Fig. 3, representing the class modulation parameters trainable), thus obtaining an initial conditional StyleGAN2 model that generates well in details but has no class control; finally, we perform conditional generation training on this model, achieving the knowledge transfer from unconditional StyleGAN2 to conditional StyleGAN2, which can achieve good conditional generation results.

Experiments and results

The dataset of Peking opera facial makeup

In our previous studies [10, 11], we constructed a Peking opera facial makeup dataset for image generation research. The dataset mainly uses the method of cutting pictures from electronic scanned books, obtaining 1286 hand-drawn facial makeup images, and also obtaining 494 machine-drawn vector images through web crawling. The image data was cleaned and filtered to remove duplicate facial makeup images, resulting in 1780 original data. Then, only two methods of mirroring and hue change were used to augment the original data, and finally a dataset containing 7120 Peking opera facial makeup images with a size of 256×256×3 was created. As far as we know, this dataset is the richest data resource for Peking opera facial makeup images.

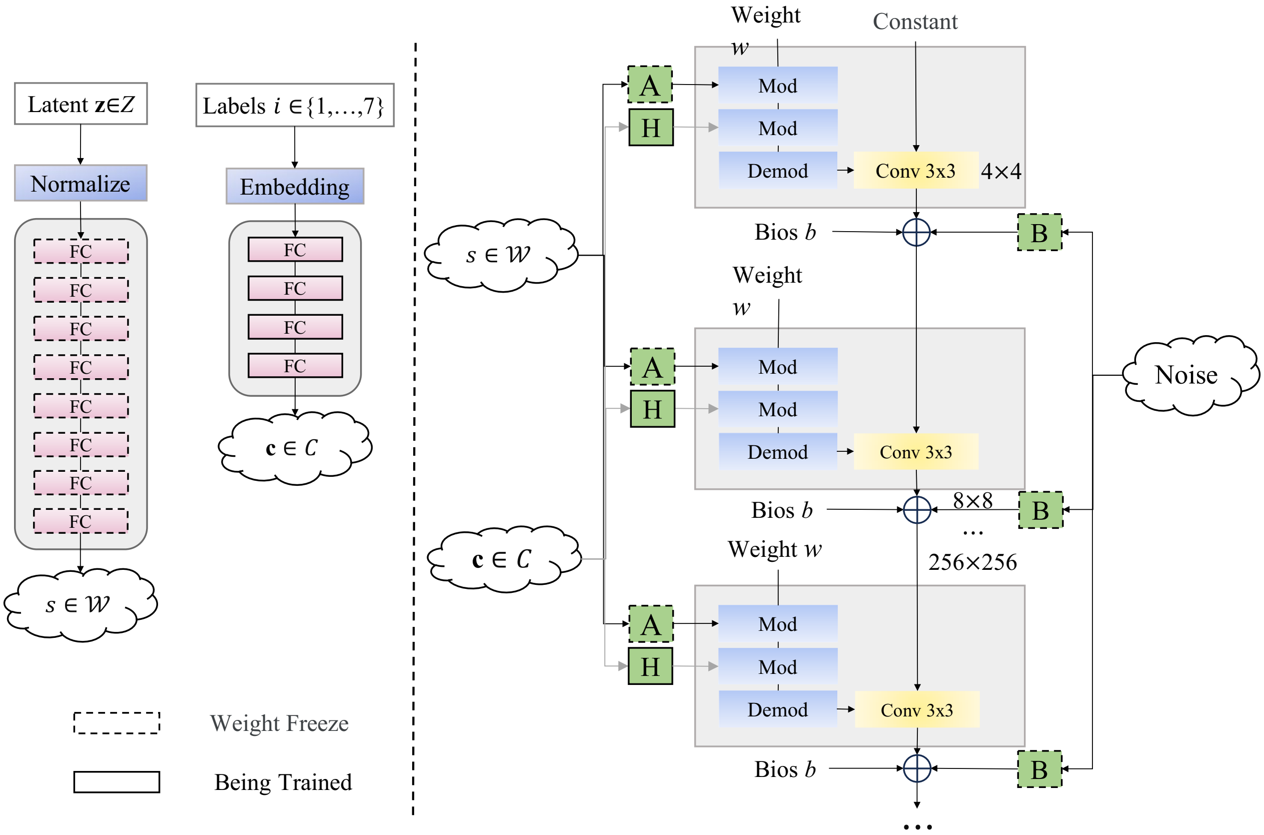


Fig. 3 Hypernetwork architecture TC-StyleGAN2: Transforming from unconditional StyleGAN2 to conditional StyleGAN2

Experiments

To quantify the quality of generated images, the Fréchet Inception Distance (FID) [43] and Kernel Inception Distance (KID) [44] are often used as objective evaluation metrics for image generation. FID is one of the most commonly used metrics to compare the similarity between real and synthetic images. Its core idea is to embed real and generated images into a visually relevant feature space and compute the distance between the two distributions. KID computes the squared Maximum Mean Discrepancy (MMD) between the feature representations of real and generated images, using a polynomial kernel. Lower FID and KID values indicate higher quality of GAN-generated images.

As shown in Table 1, ADA achieves the best generation results by using standard data augmentation methods (geometric transformation and tonal transformation) to train StyleGAN2, producing face painting images with smooth lines and uniform colors. In theory, using generated data to expand the training set is a way to avoid model overfitting, which can stabilize the training process and enable the generative model to learn more details and features of face painting and improve the quality of synthetic images. However, according to the experimental results analysis, the reason why APA has poor improvement effect may be due to the interference of the discriminator's normal training by the addition of fake images. Simply put, when a fake image generated at the beginning is judged as fake by the discriminator,

Table 1 Objective evaluation results on self-built Peking opera facial makeup dataset

Method	Data augmentation	Baseline Network	FID ↓	KID × 10^3 ↓
StyleGAN2	None	–	21.63	14.94
ADA	Geometric and tonal transformations	StyleGAN2	**15.71**	**6.97**
APA	Image generation	StyleGAN2	18.87	10.81
Diffusion-GAN	Noise injection	StyleGAN2	22.58	11.65

Bold values represent the best performance and the best score in the experiment

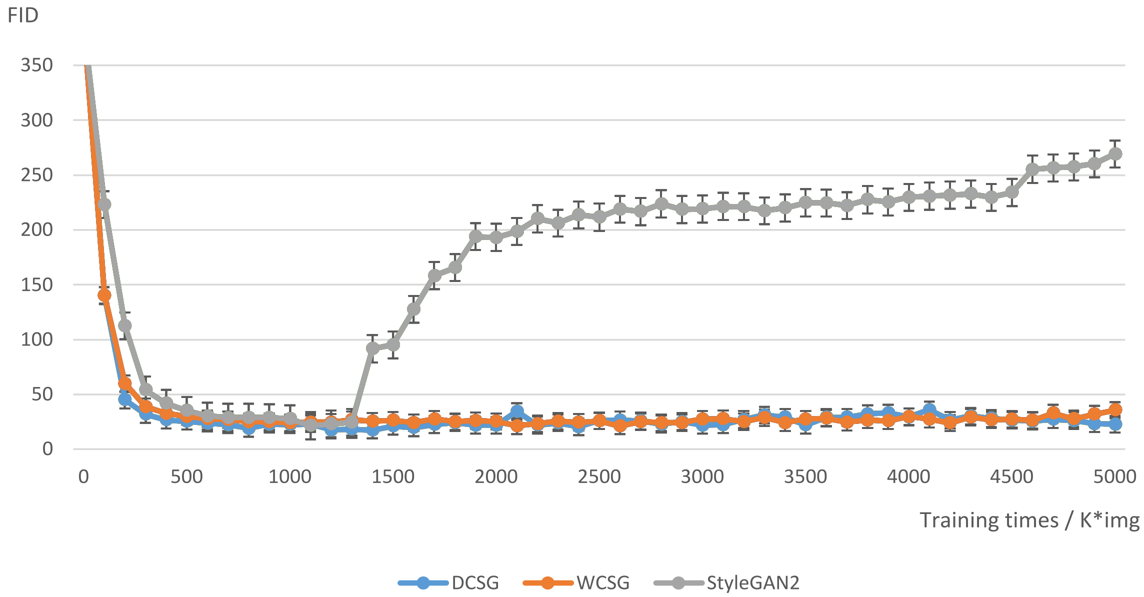


Fig. 4 Training process stability detection

and then this image is used as training data input to the discriminator, it will cause confusion in the true–false discrimination. ADA and APA explicitly enhance the images, but simply perturbing the real data distribution in GAN training may disturb the generated data distribution. Diffusion-GAN, under the same number of training times, has worse generation effect. This paper speculates that the main reason is that the discriminator learning not only needs to distinguish between real and fake data, but also needs to distinguish between diffusion-generated samples and diffusion-real samples. The injection of instance noise increases the learning task of model training. Although it avoids the catastrophic forgetting of the discriminator to some extent, it prolongs the training time. This paper hopes to find a way to reduce the training time cost but substantially improve the training effect.

The experiments in this paper show that data augmentation is still the simplest and effective way to avoid discriminator overfitting and improve image generation quality. When lacking training data, applying data augmentation to StyleGAN2 network can reliably stabilize the training and effectively improve the image generation quality. Of course, data augmentation cannot replace real data. Training StyleGAN2 should first try to collect a large amount of high-quality training set, and then use data augmentation to fill the gap. This paper applies different data augmentation methods in StyleGAN2 network in an adaptive way, and explores that ADA using standard data augmentation is the best data processing method. This will lay a good foundation for the later work, including improving the network structure and training strategy.

In the research work on unconditional image generation networks, we design ablation experiments for different improvement modules, to demonstrate the feasibility of our research scheme. By leveraging the co-training concept, we learn complementary information from multiple perspectives, enhancing the discriminator's discrimination ability, which in turn provides more effective feedback to the generator. This ultimately improves the quality of image generation, allowing the discriminator and generator to maintain a more stable game state.

As shown in Fig. 4, using the FID metric to monitor the network training process, the FID value of StyleGAN2 network exhibits a sudden surge at a certain point, leading to gradual divergence in training. By incorporating the co-training idea into the baseline network, the DCSG and WCSG networks proposed in this paper, without

Table 2 Objective evaluation results on the dataset

Method	Baseline network	FID↓	KID × 10^3 ↓
StyleGAN2	–	21.63	14.94
FastGAN	–	53.24	30.71
InsGen	StyleGAN2	20.46	10.60
Projected GAN	StyleGAN2	40.24	7.09
Projected GAN	FastGAN	**13.07**	**2.13**
WCSG	StyleGAN2	**15.38**	7.33
DCSG	StyleGAN2	**13.43**	**6.47**

Bold values represent the best performance and the best score in the experiment adding the adaptive data augmentation module, exhibit a consistent downward trend in FID values throughout the entire training process, thereby enhancing the stability of the training.

Combining the above two modules, this research designs comparative experiments from the perspective of improving the discriminator, to verify the effectiveness of the Co-StyleGAN2 algorithm proposed in this paper. As shown in Table 2, the FID value of InsGen is better than the baseline network StyleGAN2, which proves that the discriminator's discrimination ability and the generator's generation ability are proportional. The FID value of FastGAN is far worse than StyleGAN2. The comparative experiment Projected GAN proposes to use the pre-trained representation ability to improve the discriminator, and Projected GAN based on FastGAN confirms that designing multiple discriminators to improve the discrimination ability can provide better feedback to the generator. And Projected GAN based on StyleGAN2 has very poor generation effect as expected, because using a very strong pre-trained network will make the discriminator too strong, and naturally it can easily discriminate between generated data and real data, resulting in serious training imbalance. And the WCSG or DCSG proposed in this paper can train stably and generate better results than the baseline network, which proves the effectiveness of introducing the co-training idea into StyleGAN2. This idea alleviates the overfitting of the discriminator by learning from multiple different views, thereby stabilizing the training process. The generator can better learn and fit the real image distribution, improving the quality of generation.

As shown in Fig. 5, the face painting images generated by InsGen and FastGAN have uneven color blocks; the Projected GAN based on StyleGAN2 has poor generation effect on the facial details of the face painting, and the patterns are messy. Under the condition that the FID indicators are not much different, although the quantitative analysis data of Projected GAN based on FastGAN is the best, the face painting images generated by the proposed DCSG are better in visual effect, and Projected GAN tends to produce messy circular spots on the pattern details. The proposed DCSG and WCSG are obviously better than other models in evaluation results.

We construct C-StyleGAN2 [20] based on StyleGAN2 to embed the category information. Conditional image generation is controlled by the hyperparameter *cond*. When $cond = 0$, it represents unconditional generation during training, and $cond = 1$ signifies conditional generation. Shahbazi [35] utilizes a transitional training strategy, starting with unconditional StyleGAN2 and gradually injecting category conditions into the generator

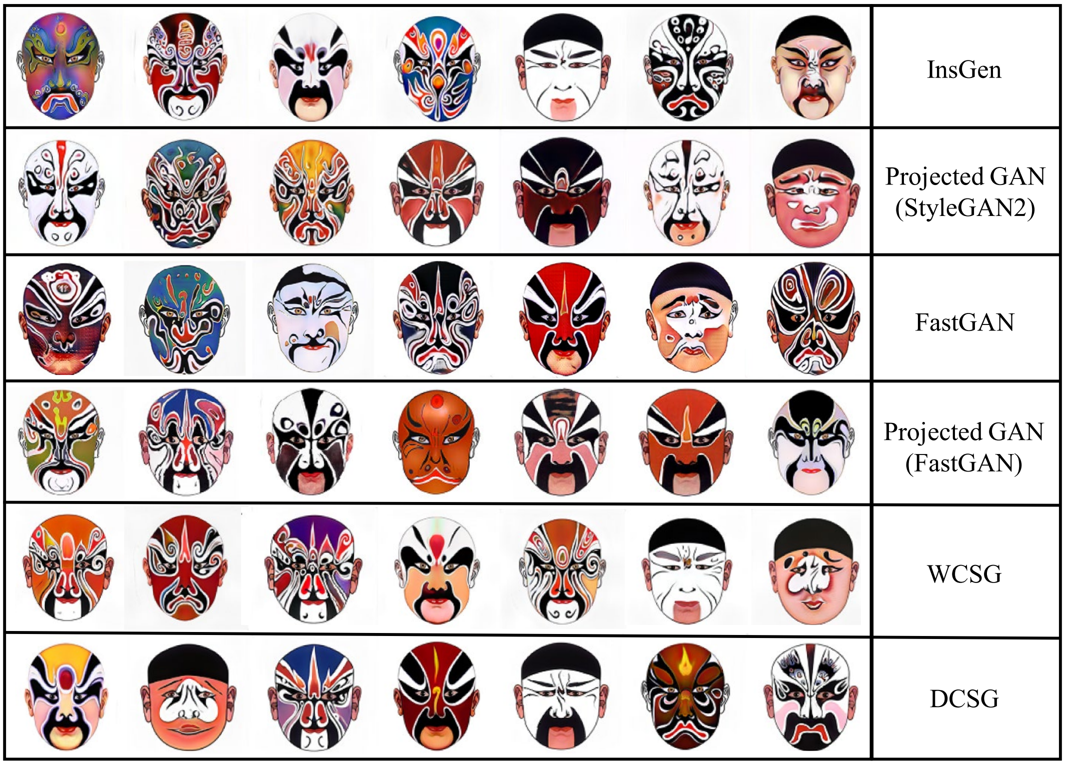


Fig. 5 Qualitative results of different improved discriminator studies

Table 3 The FID result on the dataset

Method	Personality category	Spectrum category
C-StyleGAN2	88.62	91.17
Shahbazi	20.82	19.62
TC- StyleGAN2	**13.43**	**18.59**

Bold values represent the best performance and the best score in the experiment

and objective function, ultimately achieving conditional StyleGAN2. In Shahbazi's training, the official parameter settings are used: initially, *cond* is set to 1, with $t\_start\_kimg = 2000$ indicating that class information is injected into the network starting from $2000kimg$, and $t\_end\_kimg = 4000$ indicating that the model completes the transition from unconditional StyleGAN2 to conditional StyleGAN2 by $4000kimg$.

The TC-StyleGAN2 proposed in this paper transitions from a high-quality pre-trained unconditional StyleGAN2 to a conditional StyleGAN2. The pre-trained unconditional StyleGAN2 network uses the DCSG network, constructed in this study, with the best generative quality. The source weights learned by the unconditional StyleGAN2 network are first frozen as the initialized weights for training the conditional StyleGAN2. Subsequently, $cond = 1$ is set for conditional StyleGAN2 training.

In the research work on conditional image generation networks, we design a comparison of image generation effects with comparative experiments on two different datasets. We use quantitative analysis methods to evaluate the performance of our algorithm and other algorithms. We show the advantages of our algorithm in image quality and image diversity.

As shown in Table 3, Figs. 6, 7, we compare the visual effects of the conditional images generated by different methods. We point out that our method takes advantage of the drawback of the discriminator, which focuses more on the category than the details under the guidance of the category information, by continuing to train the conditional model on the effectively learned generative model, which is equivalent to playing a classification effect, achieving the detail characteristics and diversity of the intra-class generated images. We also demonstrate that our method allows effective transfer learning, where the frozen pre-trained weights are conditionally modulated to produce outputs specific to the target category. Surprisingly, the C-StyleGAN2 yielded the worst FID scores on both conditional generation tasks. The qualitative results of the generated images also indicated that

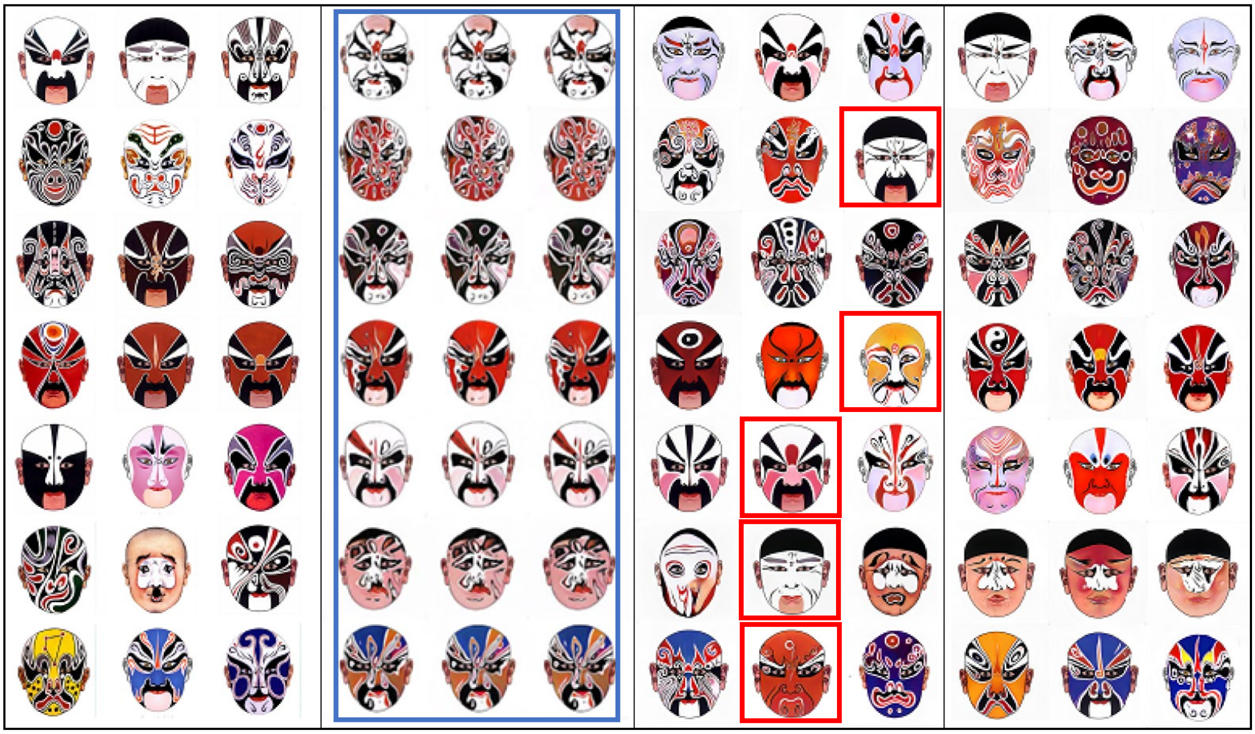


Fig. 6 Qualitative results for Personality Category. (From left to right: real facial makeup, C-StyleGAN2, Shahbazi, TC-StyleGAN2; From top to bottom: treacherous, demonic, integrity, loyalty, old, ugly, and reckless)

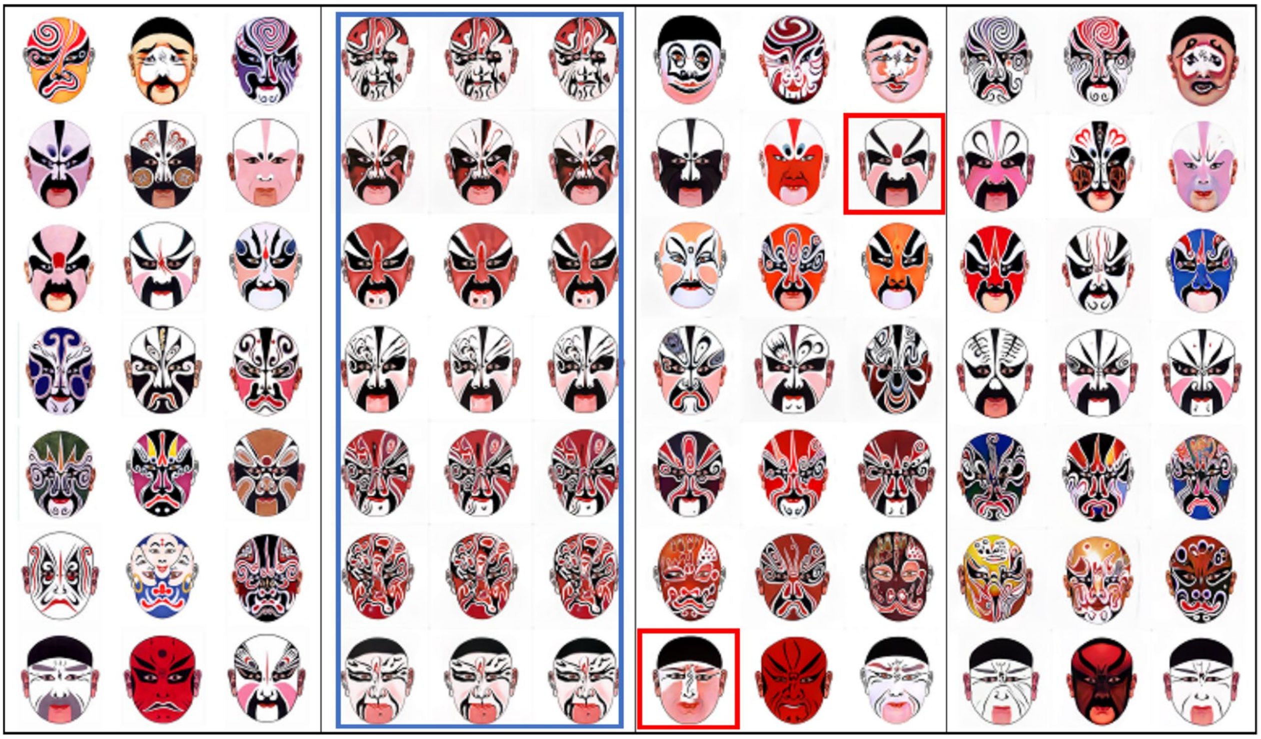


Fig. 7 Qualitative results for Spectrum Category. (From left to right: real facial makeup, C-StyleGAN2, Shahbazi, TC-StyleGAN2; From top to bottom: ugly, six-pointed, three-piece, cross-shaped, floral, demonic, whole)

the complexity of the original conditional generation network in StyleGAN2 was insufficient to generalize effectively to the two proposed facial makeup conditional generation tasks.

As shown in the blue boxes in Figs. 6, 7, the intra-class variation of C-StyleGAN2 is very small, mainly limited to the color change of the face painting, while maintaining the same structure and pose. In addition, the images lack realism and contain obvious artifacts. Compared with C-StyleGAN2, our TC-StyleGAN2 does not suffer from mode collapse, and the face painting images generated are not limited to one style, and the face painting details are more diverse. Compared with Shahbazi's research, our TC-StyleGAN2 does not have the problem of category information leakage. Shahbazi proposed a transition training strategy that has the problem of incomplete transition, for example, as shown in the red boxes in Fig. 6, the whole face category face painting will appear in multiple category face painting; as shown in the red boxes in Fig. 7, the three-tile face category face painting will appear in the generated six-part face category face painting images, or the clown face category face painting will appear in the generated whole face category face painting images.

The above evaluation methods for the generated images are all calculated from a macro perspective. Generally speaking, the score improvement represents that the generative model learns more details and the generated samples are more realistic and natural. In addition to generating high-fidelity images, more attention should be paid to the situation where the generative model generates non-realistic images during the evaluation. This study designs to independently generate 5000 samples by each model and find the worst samples compared with the real image distribution.

Specifically, first use the Inception [45] feature space of the real images to fit the Gaussian model. Then calculate the log-likelihood of each sample given the Gaussian prior, and display the image with the smallest log-likelihood (the largest Mahalanobis distance), that is, the sample with the worst image quality in the generated samples. Mahalanobis Distance (MD) [46] represents the covariance distance of the data, which is an effective method to calculate the similarity between two unknown sample sets. The formula is in Eq. (20):

$$d = \sqrt{(x-y)^T \sum\nolimits^{-1} (x-y)} \tag{20}$$

where x and y are the real and generated samples respectively, $\sum$ is the covariance matrix of the real image

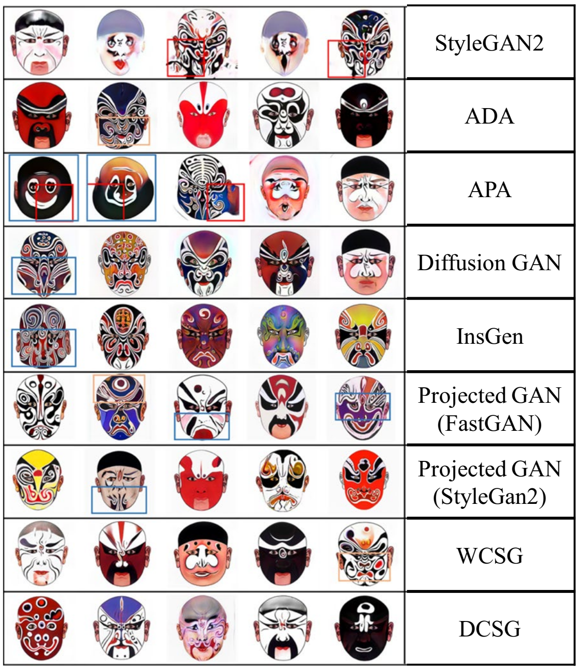


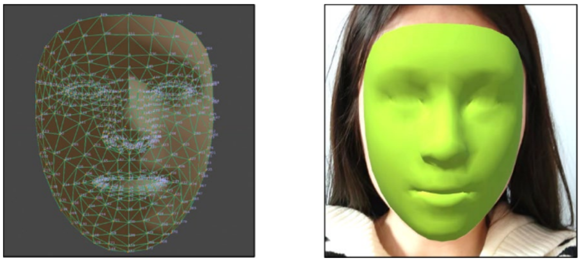
Fig. 8 The worst Peking opera facial makeup sample example

dataset. The worst samples obtained by the calculation of each model are shown in Fig. 8:

It is observed that the most prominent problem of the baseline network StyleGAN2 is that the generated facial makeup images have strange artifacts (as shown by the red box in Fig. 8), which is similar to the problem of APA; the second is the deformation of the facial makeup head or facial features, as shown by the blue box in Fig. 8, the facial makeup head of APA, Diffusion GAN and InsGen are deformed, and the facial features of Projected GAN are lost or distorted in multiple facial makeup images; the last is the distortion of the facial makeup lines, as shown by the yellow box in Fig. 8, the most obvious is that the lines in the facial makeup generated by Projected GAN based on FastGAN are fine and messy, which also exist in ADA and the WCSG and DCSG proposed in this paper. In contrast, the method of this study can not only generate high-quality Peking opera facial makeup, but also avoid the problems of artifacts, head and facial features deformation, etc. that exist in other models.

AR visualization

To give full play to the application value of the Peking opera facial makeup images generated by this study, and to display the generated facial makeup images to the public and popularize the knowledge of facial makeup briefly, this part adopts AR technology and designs a Peking opera facial makeup AR face-changing system based on mobile devices. The basic module is based on the Augmented Faces API of AR Core, which uses the ARFace component to present the default material on the detected face mesh model, as shown in Fig. 9, which contains a dense 3D face mesh model with 468 points and the default material on the face mesh. With the help of the face mesh model, different details of texture maps can be drawn.



(a) Face mesh model

(b) Mesh assigned default material

Fig. 9 3D face mesh

Figure 10 shows the different face-changing effects. Users can complete the free face-changing function, and the Peking opera facial makeup fits the face normally. Users can see the unique Peking opera facial makeup images generated by this paper in the face-changing process. The Peking opera facial makeup generated by the generative model has high application and research value.

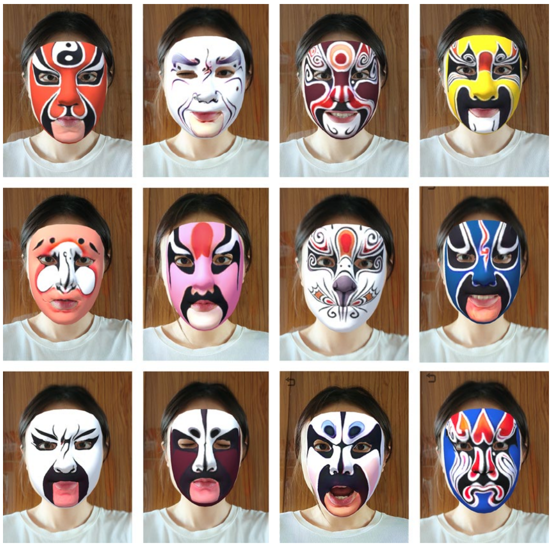


Fig. 10 Face changing effect example

Conclusions

Image generation technology can generate unique and novel images by mimicking human creative rules, which is irreplaceable for artistic creation, graphic design and other fields where data resources are precious. In this paper, based on deep learning methods, we improve the StyleGAN2 network structure in generative adversarial networks, and propose two models: Co-StyleGAN2 and TC-StyleGAN2, to achieve unconditional and conditional generation tasks of Peking opera facial makeup images. By comparing with the advanced algorithms in the field of image generation, our algorithm can not only guarantee high FID and KID values, but also generate Peking opera facial makeup images with better visual effects than other algorithms. Finally, we design a module to display the generation effects of Peking opera facial makeup based on Unity.

For future work, there are several research goals including: (1) Dataset Usage and Expansion: This paper focuses on the task of face mask generation, using our self-constructed Peking Opera face mask dataset. However, we did not use our proposed adaptive data augmentation module for experiments on other image datasets of different scales. Future work should apply this data augmentation module to other image datasets to evaluate its generality and effectiveness; (2) Dynamic Adjustment of the Discriminator: The discriminator plays a crucial role in the training of GANs. The data distribution of generated images continuously changes due to the evolving generator, impacting the discriminator's task of distinguishing real from fake images. Therefore, future research will explore dynamically adjusting the capacity of the discriminator to better adapt to this time-varying task; (3) Evaluation Metrics for Image Generation Quality: Experimental results show that commonly used objective evaluation metrics for image generation quality, such as FID and KID, differ from human subjective evaluations. FID is primarily concerned with a few features, aiming to assist ImageNet classification [47] rather than providing a thorough analysis of the entire image. Future research will investigate replacing the feature space of FID with models such as CLIP [48] and self-supervised SwAV [49] to reduce the influence of ImageNet classification on the effectiveness of FID.

Acknowledgements
We would like to thank the Science Research Department of Communication University of China for its strong support for this research.

Author contributions
Yinghua Shen: Conceptualization, Methodology, Formal Analysis, Writing—Original Draft Preparation, Writing—Review and Editing, Funding Acquisition, Supervision. Oran Duan: Data Curation, Software Development, Validation, Writing—Original Draft Preparation, Visualization. Xiaoyu Xin: Investigation, Data Curation, Resources, Software Development. Ming Yan: Methodology, Validation, Formal Analysis, Writing—Review and Editing. Zhe Li: Supervision, Project Administration, Writing—Review & Editing.

Funding
Not applicable.

Availability of data and materials
The data and materials used in this study are available from the corresponding author upon reasonable request. To protect the privacy of the participants, some objective evaluation data may only be provided in compliance with privacy protection regulations.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All participants have consented to the publication of the results of this study in academic journals and have signed the corresponding consent forms.

Competing interests
Not applicable.

Received: 6 June 2024 Accepted: 25 September 2024
Published online: 14 October 2024

References

1. Zheng Z. Evolution, symbolism, artistry: a study on the colors of Peking opera facial makeup. Art Perform Lett. 2023;4(12):36–42.
2. Shmagalo R, Hu X. The art of the mask and make-up in the traditions of the East and West: artistic features, stylistics, interrelationship. Herança. 2024;7(1):100–12.
3. Wu G, He F, Zhou Y, Jing Y, Ning X, Wang C, Jin B. ACGAN: age-compensated makeup transfer based on homologous continuity generative adversarial network model. IET Comput Vision. 2023;17(5):537–48.
4. Ma J, Han J, Li Z, Liu Y, Guo H. Msaff: A multi-scale attention feature fusion classification model and Colp-Id. 2024. https://doi.org/10.2139/ssrn.4824785
5. Zhou E, Li N, Liu B, Chen Y. Watching opera at your own ease—A virtual character experience system for intelligent opera facial makeup. Proceedings of the Eleventh International Symposium of Chinese CHI. 2023; 443–448.
6. Yan M, Xiong R, Wang Y, Li C. Edge computing task offloading optimization for a UAV-assisted internet of vehicles via deep reinforcement learning. IEEE Trans Veh Technol. 2024;73(4):5647–58.
7. Gao M, Wang P. Personalized facial makeup transfer based on outline correspondence. Comput Anim Vir World. 2024;35(1): e2199.
8. Shi H, Li J, Xue L, Song Y. OperAR: Using an augmented reality agent to enhance children's interactive intangible cultural heritage experience of the Peking Opera. Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. 2023; 1–3.
9. Chen J, Liufu C, Zhang W, Luo C, Fu K, Lin J, et al. Preparation and efficacy verification of three-dimensional printed partitioned multi-effect precision-care gel facial mask. Int J Cosmet Sci. 2024;46(2):209–27.
10. Yan M, Xiong R, Shen Y, Jin C, Wang Y. Intelligent generation of Peking opera facial masks with deep learning frameworks. Heritage Science. 2023;11(1):20.
11. Xin X, Shen Y, Xiong R, Lin X, Yan M, Jiang W. Automatic image generation of Peking opera face using styleGAN2. 2022 International Conference on Culture-Oriented Science and Technology (CoST). IEEE. 2022; 99–103.
12. Huynh N, Deshpande G. A review of the applications of generative adversarial networks to structural and functional MRI based diagnostic classification of brain disorders. Front Neurosci. 2024;18:1333712.
13. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Adv Neural Inf Process Syst. 2020;33:6840–51.

14. Chen Y, Liu J, Peng L, Wu Y, Xu Y, Zhang Z. Auto-encoding variational Bayes. Cambridge Explor Arts Sci. 2024. https://doi.org/10.61603/ceas.v2i1.33.
15. Dewi C. Generative adversarial network for synthetic image generation method: review, analysis, and perspective. In: Dewi C, editor. Applications of generative AI. Cham: Springer; 2024. p. 91–116.
16. Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. Adv Neural Inf Process Syst. 2021;34:8780–94.
17. Onakpojeruo E, Mustapha M, Ozsahin D, Ozsahin I. A comparative analysis of the novel conditional deep convolutional neural network model, using conditional deep convolutional generative adversarial network-generated synthetic and augmented brain tumor datasets for image classification. Brain Sci. 2024;14(6):559.
18. Liang J, Yang X, Huang Y, Li H, He S, Hu X, et al. Sketch guided and progressive growing GAN for realistic and editable ultrasound image synthesis. Med Image Anal. 2022;79: 102461.
19. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019; 4401–4410.
20. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T. Analyzing and improving the image quality of StyleGAN. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020; 8110–8119.
21. Karras T, Aittala M, Laine S, Härkönen E, Hellsten J, Lehtinen J, Aila T. Alias-free generative adversarial networks. Adv Neural Inf Process Syst. 2021;34:852–63.
22. Che A, Mohd T, Hilmi M, Mohd K. Assessing the efficacy of StyleGAN 3 in generating realistic medical images with limited data availability. Proceedings of the 2024 13th International Conference on Software and Computer Applications. 2024;192–197.
23. Pavez V, Hermosilla G, Pizarro F, Fingerhuth S. Thermal image generation for robust face recognition. Appl Sci. 2022;12(1):497.
24. Situ Z, Teng S, Liu H, Luo J, Zhou Q. Automated sewer defects detection using style-based generative adversarial networks and fine-tuned well-known CNN classifier. IEEE Access. 2021;9:59498–507.
25. Zhang Y, Wang Y, Jiang Z, Liao F, Zheng L, Tan D, et al. Diversifying tire-defect image generation based on generative adversarial network. IEEE Trans Instrum Meas. 2022;71:1–12.
26. Zhao C, Shuai R, Ma L, Liu W, Hu D, Wu M. Dermoscopy image classification based on StyleGAN and denseNet201. IEEE Access. 2021;9:8659–79.
27. Chen F, Zhu F, Wu Q, Hao Y, Wang E. Infrared image data augmentation based on generative adversarial network. J Comput Appl. 2020;40(7):2084.
28. Yan M, Luo M, Chan CA, Gygax AF, Li C, Chih-Lin I. Energy-efficient content fetching strategies in cache-enabled D2D networks via an Actor-Critic reinforcement learning structure. IEEE Trans Vehicul Technol Early access. 2024. https://doi.org/10.1109/TVT.2024.3419012.
29. Ning X, Wang X, Xu S, Cai W, Zhang L, Yu L, Li W. A review of research on co-training. Concurr Comput Pract Exp. 2023. https://doi.org/10.1002/cpe.6276.
30. Rahate A, Walambe R, Ramanna S, Kotecha K. Multimodal co-learning: challenges, applications with datasets, recent advances and future directions. Inf Fusion. 2022;81:203–39.
31. Cui K, Huang J, Luo Z, Zhang G, Zhan F, Lu S. Genco: Generative co-training for generative adversarial networks with limited data. Proc AAAI Conf Artif Intell. 2022;36(1):499–507.
32. Gong Y, Wu Q, Cheng D. A co-training method based on parameter-free and single-step unlabeled data selection strategy with natural neighbors. Int J Mach Learn Cybern. 2023;14(8):2887–902.
33. Peng J, Estrada G, Pedersoli M, Desrosiers C. Deep co-training for semi-supervised image segmentation. Pattern Recogn. 2020;107: 107269.
34. Ma Y, Yang Z, Zhang Z. Multisource maximum predictor discrepancy for unsupervised domain adaptation on corn yield prediction. IEEE Trans Geosci Remote Sens. 2023;61:1–15.
35. Shahbazi M, Danelljan M, Paudel DP, Van Gool L. Collapse by conditioning: Training class-conditional GANs with limited data. 2022. https://doi.org/10.48550/arXiv.2201.06578
36. Xu X, Chen W, Zhou W. AmbientCycleGAN for establishing interpretable stochastic object models based on mathematical phantoms and medical imaging measurements. Medical imaging 2024: image perception, observer performance, and technology assessment. SPIE. 2024;12929:234–40.
37. Tseng HY, Jiang L, Liu C, Yang MH, Yang W. Regularizing generative adversarial networks under limited data. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021; 7921–7931.
38. Jiang L, Dai B, Wu W, Loy CC. Deceive D: Adaptive pseudo augmentation for GAN training with limited data. Adv Neural Inf Process Syst. 2021;34:21655–67.
39. Wang Z, Zheng H, He P, Chen W, Zhou M. Diffusion-GAN: training GANs with diffusion. arXiv preprint arXiv:2206.02262. 2022. https://doi.org/10.48550/arXiv.2206.02262
40. Huang J, Guan D, Xiao A, Lu S. RDA: Robust domain adaptation via Fourier adversarial attacking. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021; 8988–8999.
41. Cong Y, Zhao M, Li J, Wang S, Carin L. GAN memory with no forgetting. Adv Neural Inf Process Syst. 2020;33:16481–94.
42. Chauhan V, Zhou J, Lu P, Molaei S, Clifton D. A brief review of hypernetworks in deep learning. arXiv preprint arXiv:2306.06955. 2023; 57(9): 1–29.
43. Kynkäänniemi T, Karras T, Aittala M, Aila T, Lehtinen J. The role of ImageNet classes in Fréchet inception distance. arXiv preprint arXiv:2203.06026. 2022. https://doi.org/10.48550/arXiv.2203.06026
44. Bińkowski M, Sutherland DJ, Arbel M, Gretton A. Demystifying MMD GANs. arXiv preprint arXiv:1801.01401. 2018. https://doi.org/10.48550/arXiv.1801.01401
45. Kelbert M. Survey of distances between the most popular distributions. Analytics. 2023;2(1):225–45.
46. Dashdondov K, Kim M. Mahalanobis distance-based multivariate outlier detection to improve performance of hypertension prediction. Neural Process Lett. 2023;55(1):265–77.
47. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90.
48. Radford A, Kim J, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. Learning transferable visual models from natural language supervision. Proceedings of the International Conference on Machine Learning. PMLR. 2021; 139: 8748–8763.
49. Caron M, Misra I, Mairal J, Goyal P, Bojanowski P, Joulin A. Unsupervised learning of visual features by contrasting cluster assignments. Adv Neural Inf Process Syst. 2020;33:9912–24.

Publisher's Note