Received: 30 May 2022 | Revised: 27 July 2022 | Accepted: 2 September 2022

DOI: 10.1049/cvi2.12138

IET Computer Vision

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

ACGAN: Age-compensated makeup transfer based on homologous continuity generative adversarial network model

Guoqiang Wu[1,2] | Feng He[3] | Yuan Zhou[4] | Yimai Jing[5] | Xin Ning[6] | Chen Wang[7] | Bo Jin[8]

[1]Beijing Institute of Technology, School of Mechatronical Engineering, Beijing, China

[2]China Aerospace Science and Technology Corporation The 9th Academy Beijing Aerospace Institute of Unmanned System Engineering, Beijing, China

[3]School of Computer Science, Yangtze University, Jingzhou, China

[4]School of Information and Mathematics, Yangtze University, Jingzhou, China

[5]School of Geosciences, Yangtze University, Jingzhou, China

[6]Institute of Semiconductors, Chinese Academy of Sciences, Beijing, China

[7]School of Computer Science and Engineering, Beihang University, Beijing, China

[8]Institute of Systems and Robotics (ISR), University of Coimbra, Coimbra, Portugal

Correspondence

Feng He, School of Computer Science, Yangtze University, Jingzhou 430023, China.
Email: hefengcs@gmail.com

Funding information

National Natural Science Foundation, Grant/Award Numbers: 61901436, 51974036, 62006028

Abstract

The authors focus on the makeup transformation problem, which refers to the transfer of makeup from a reference face to a source face image while maintaining the source makeup-free face image. In recent years, makeup transformation has become a hot issue and a lot of research has been conducted on this basis, but there are some limitations in the existing methods, mainly due to the lack of consideration of age factor, which makes the final generated face makeup images appear not natural and lack appearance attractiveness. In order to further solve this problem, an age-compensated makeup transformation framework based on homology continuity is proposed. In order to achieve a stable and controllable age-compensation effect, the authors design a new coding module that can map the face makeup semantic vector into the higher feature space and achieve age compensation by adjusting the direction of the semantic vector. Finally, in order to comprehensively evaluate the effectiveness of the authors' proposed method, a large number of qualitative and quantitative experiments have been conducted, and the experimental results show that the authors' proposed framework outperforms existing methods.

1 | INTRODUCTION

It has become a major means to improve the attractiveness of one's appearance by means of makeup. Finding the right makeup for you is not an easy task and requires trying a large number of makeup looks, which consumes a lot of resources each time a new makeup look is tried. Therefore, the use of computers for complete makeup transformation is a more feasible and sustainable method of virtual makeup transformation. Makeup transformation refers to the migration of any makeup from reference image to source image. The current makeup transformation methods can be broadly divided into two categories: one is the traditional makeup transformation method. Some authors [1] proposed a face makeup migration method similar to physical makeup, which is more complex, slower to process and takes longer time. Some other authors [2] accomplish the task of makeup transformation by calculating the colour and lighting changes before and after makeup. However, the method has a high requirement for images and therefore has a low practicality. The final results of such methods are not ideal and require high datasets with before and after aligned face makeup images. In addition, it is all a simple combination of different components, and the overall image does not look natural and the effect is average. The other category is based on deep learning methods. With the continuous development of deep learning, it has a wide range of applications in various fields [3–7], and good results have been obtained in the field of makeup transformation. Some authors proposed a model based on deep local makeup migration; although this method can achieve the intensity of makeup control, the effect is not ideal and the overall effect is not natural enough.

The current approach of makeup migration based on generative adversarial networks [8] has become mainstream. CycleGAN [9] opens a new era of image-to-image translation. Image-to-image translation is accomplished in the absence of paired datasets. Most of the makeup transformation methods are inspired by CycleGAN. BeautyGAN [10] proposes a dual-input and output generative adversarial network framework that implements the makeup transformation by combining global domain-level loss and local instance loss into the same network. PairedCycleGAN [11] proposes a makeup transformation framework oriented to makeup application and removal without pairwise training data, while introducing two asymmetric functions, one for makeup application and the other for makeup removal. LADN [12] proposes a local adversarial separated network structure using multiple overlapping local adversarial discriminators as a way to implement the facial makeup transformation.

A deep neural network consists of a large number of neurons. Deep neural networks are theoretically capable of fitting any function due to the large number of complex and linear connections that precede them. In the past few decades, deep learning made amazing progress, such as applications in classification, object detection, action recognition, semantic segmentation etc., and achieved good results. In addition, due to the existence of a certain distribution of popular structures in the data in nature, it makes deep learning able to learn this part of the flow structure very well. Brain-like science has contributed to the development of neural networks to some extent. Since 1943, some authors [13] proposed the M-P linear neuron, and the M-P model was the first of all artificial neurons to be built, which shows the basic properties possessed by biological neurons in several aspects. Since then, some other authors [14] proposed RBF neurons, initially explored the different characteristics of RBF for neural network design and applied to the traditional interpolation domain, and then proposed a three-layer structured RBF neural network. Other authors [15] proposed a neural network containing local response characteristics, which is actually consistent with the RBF neural network proposed by Broomhead and Lowe, and they also proposed a training method for RBF neural networks.

The makeup transformation problem is a process of age change, not just age classification, when age compensation is performed. Homology continuity means that for any mode, there is a strong correlation between its different states of existence and there is an obvious continuity process by which the switching of different states is achieved. For example, if two modalities are homologous in nature, then there is a pathway whereby the two can be converted into each other. And in the process of this path, the intermediate state and the two modes before and after the whole path are homologous. For example, a large sheep and a small sheep, both of which belong to sheep, are homologous samples. In the process of turning a small sheep into a large sheep, it is still a sheep, that is, in this asymptotic process, this pattern is still the same as the other two patterns. By tapping this asymptotic relationship, this is used to asymptote between homologous things. The continuity between homologous samples is the principle of homologous continuity. In recent years, homologous continuity has become an important bionic pattern recognition theory [16], which is applied to ageing of faces [17], pattern recognition [18] etc. The field of makeup transformation has been developed for decades, and there is a lot of accumulation in this problem, and it is possible to obtain good results, even in the presence of large expressions and poses, and shadows and occlusions. However, most of the existing methods ignore the influence of age factor on the final makeup transformation effect, and the lack of consideration of age factor makes it impossible to meet people's demand for appearance even after makeup transformation.

To solve this problem, this paper proposes an age-compensated makeup transfer based on homologous continuity generative adversarial network model (ACGAN). Two main network branches are included: the age-compensated network branch and the makeup transfer network branch. In the age-compensation branch, we design a new encoder based on homologous continuity. After inputting a makeup image of a face without makeup, the corresponding semantic information will be extracted and then mapped to a high-dimensional feature space to achieve the compensation of age by adjusting the vector direction. In addition, since the adjustment of the semantic vector direction is controllable, the compensation of age is also stable and controllable. In the makeup transformation branch, we introduce the attention mechanism to deal with the shortage of encoder–decoder structure. We also design a style-based encoder. At the same time, because it is a two-way encoder–decoder structure, it has better parallelism, which can make good use of GPU and achieve faster model training and inference. ACGAN is a robust method with strong generality. It is able to generate makeup of different intensities and also has good compatibility with the gender of characters. Inputting female and male character images, we can eventually get more natural makeup effects; unlike the simple combination of previous makeup, ACGAN is generated with a holistic idea. Also ACGAN is robust to expressions and even for large expressions, it can eventually generate natural face makeup images. The superiority of our proposed model over existing methods is demonstrated through qualitative and quantitative experiments. The contributions of this paper can be summarised in the following three main aspects:

- We find that the existing makeup transformation methods lack the consideration of age factor. To solve this problem we propose an age-compensated makeup transformation framework based on homology continuity.
- Based on the homology continuity, we designed a new coding module, the homology continuity-based encoder (HCEncoder). It can extract the input face makeup image and then map it to the corresponding high-level feature space to achieve age compensation by adjusting the semantic vector direction. In addition, the final age-compensation effect is stable and controllable because the whole adjustment process can control the intensity. The final age-compensation effect is stable and controllable because the whole adjustment process can control the intensity.
- In order to comprehensively evaluate and diagnose the effectiveness of the proposed framework, we conducted a large number of experiments and compared the output results of existing methods. The qualitative and quantitative assessments illustrate the superiority of our proposed framework over existing models.

2 | RELATED WORK

2.1 | Makeup transfer

Makeup transformation refers to the migration of arbitrary makeup from source image to source image. In recent years, the study of makeup transformation has attracted a lot of attention from authors. Currently, GAN-based methods have become the mainstream approach to solve the makeup transformation task. Large pose and expression makeup transformation has been a challenging problem until now. To address this problem, the authors proposed the pose and expression robust spatially aware GAN (PSGAN) [19] although, PSGAN has partial chromatic aberration when performing makeup transformation. However, it is able to achieve partial and shade-controllable makeup transfer in most cases. To address the problem that most models have difficulty in controlling the intensity of makeup, the authors propose a spatially aware GAN for pose and expression robustness (abbreviated as PSGAN++) [20]. PSGAN++ is able to perform detail-preserving makeup transformation and effective makeup removal. CPM [21] proposes that makeup transformation should include colour transfer and pattern addition. To achieve this functionality, the authors propose a holistic approach that converts colours and patterns from the reference image to the source image. For most current methods, the image quality is not good enough, the authors in Ref. [22] propose a GAN-based makeup migration method, called RAMT-GAN. This is an unsupervised method to achieve realistic and accurate makeup migration. The makeup migration is accomplished while preserving the background information and face identity. However, none of these methods are optimised for the elderly, and most of them are inspired by StyleGAN, using its two-dimensional framework to accomplish makeup transformation. In this paper, we notice for the first time the influence of age factor on the effect of makeup transformation and provide an effective solution to this problem by substantially improving the appearance attractiveness after makeup transformation through age-compensated network branches and makeup transformation network branches.

2.2 | Style transfer

Style transfer [23–26] has been widely studied by a large number of authors and has become a hot issue. Style transfer has also been considered as a subproblem of texture synthesis. Some authors [27] have proposed image-based methods for generating new visual appearances, which can quickly re-render the styles of different images. Some other authors [28] have proposed an example-based synthesis technique, which is implemented by separating the style and content of image fragments, and the method can be applied to various arts and sketches. The authors in Ref. [29] proposed an unsupervised style transformation method, and some authors [30] introduced an artistic style neural algorithm for separating and reorganising the picture content and style of images. There is currently less data available for training. Also, there may be multiple output images for a single input image. To address these problems, authors in Ref. [31] have proposed a method based on non-entangled representations. By introducing a recurrent loss, it is possible to generate diverse and realistic images. Some authors [32] explored new painting styles by arbitrarily combining styles learnt from individual paintings, paving the way for a structure of academic representations of artistic styles. The authors in Ref. [33] propose a new adaptive instance normalisation (AdaIN) layer. It enables content averaging. It is a way to align the variance with the corresponding feature mean and variance. This approach also enables content style tradeoffs, style interpolation, and colour and space control.

3 | ACGAN

Note that the entire training process does not require paired pre- and post-makeup face images or aligned cross-age face images. We compare existing makeup transformation methods in Table 1, mainly from four aspects: robustness, controllability, detail processing, and age compensation, and the results show that ACGAN has good performance in all four aspects.

3.1 | Overview

The network structure of our proposed method is shown in Figure 1, where we adopt a two-way encoder–decoder framework to design the network, which is a classical “representation” based framework. Note that the “representation” here covers all the information of the input face image to facilitate further information processing. The decoder is able to process this representation into a face image, and the core of this process is the processing of this representation. Inspired by traditional GAN networks, we introduce a generator-discriminator structure, where the generator is responsible for generating face images with makeup. The discriminator needs to identify whether the image is a natural makeup face image or an image generated by the generator as much as possible. The two will continuously play a game, and through optimisation iterations, eventually, the discriminator cannot determine the image authenticity. Specifically, the network framework is composed of a discriminator-distinguisher, two encoders and a decoder, and ACGAN is able to perform both makeup transformation and age supplementation.

TABLE 1 Comparison with existing methods

Method	Robust	Controllable	Detail	Age − Compensation
DMT		✓		
BeautyGAN			✓	
CPM	✓			
BeautyGlow		✓		
PSGAN	✓	✓	✓	
ACGAN(Ours)	✓	✓	✓	✓

Note: We use "Robust" to represent the robustness, "Degree" to represent the controllable makeup transformation, and we use "Detail" to represent the makeup transformation that preserves the details better. Among them, we use "Age-Compensation" to represent the compensation of age, which is used to achieve a significant improvement in appearance attractiveness.

For ACGAN, the face image without makeup will be fed into two encoders. At the same time, the homology-based continuity encoder and the multi-style encoder, in which the HCEncoder is mainly responsible for extracting the age-related semantics. Then mapping it to the higher space to achieve the purpose of age supplementation by shifting the vector direction, and can achieve the intensity controls the process. MSEncoder is mainly responsible for extracting style and identity information; on the one hand, it extracts the original identity information from the face image without makeup and sends it to the generator to retain this information, so as to complete the retention of face identity information. On the other hand, style information is extracted from the face image with makeup, where the style mainly refers to makeup information. 68 key points are used to locate the face location and extract the corresponding makeup information, including the blush style, eye shadow style, lip colour style etc. This part of information will be used to build a new makeup face image to complete the task of makeup transformation. As shown in Figure 2, it is the effect of ACGAN makeup transformation, where u represents the age-compensation effect of different intensities.

3.2 | Principle of homologous continuity

Homologous continuity means that for any mode, the different states of existence are strongly correlated and there is a clear

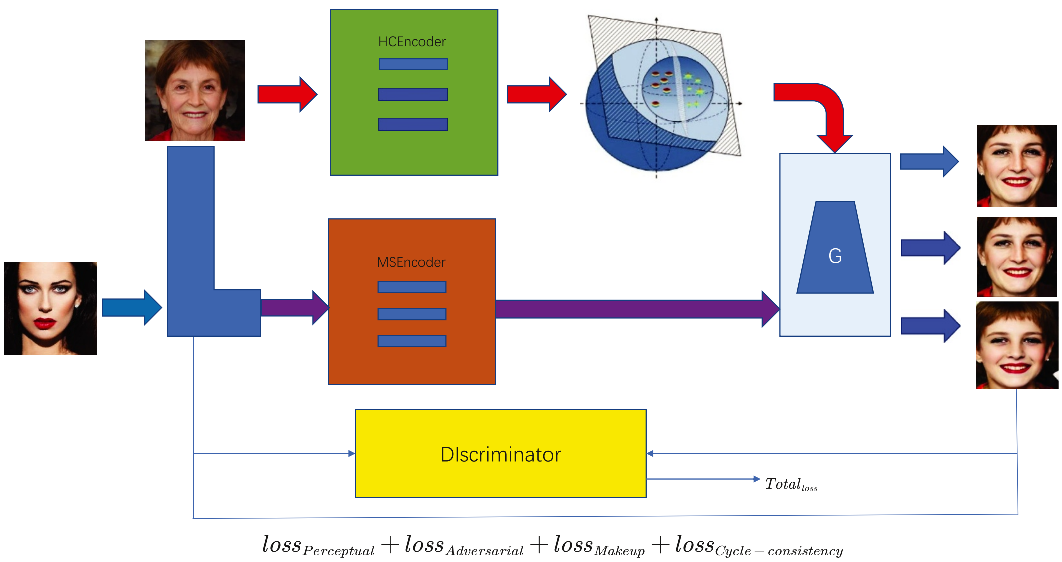

FIGURE 1 A diagram of the network framework of ACGAN. ACGAN has two branches. Use blue to indicate shared paths. The red one represents the age-compensated branch. The purple branch represents the makeup transformation. ACGAN uses a dual encoder–decoder. It has good parallelism when training and inference are performed on GPU. Inspired by GAN, ACGAN adopts a generator-discriminator structure, using a generator to generate makeup-transformed face images, and a discriminator to train the generator and discriminator simultaneously by gaming the generated makeup face images and the real makeup images. ACGAN can be roughly divided into two parts: age-compensated branch and makeup-transformed branch. In the age-compensation framework, we design a new encoder based on homology continuity to map the input face images into the corresponding high-dimensional feature space and achieve age compensation by adjusting the direction of the semantic vectors. In the makeup transformation branch, we use a style-based encoder to accomplish the makeup transformation task, which can better maintain the face identity information and background information before and after the makeup transformation

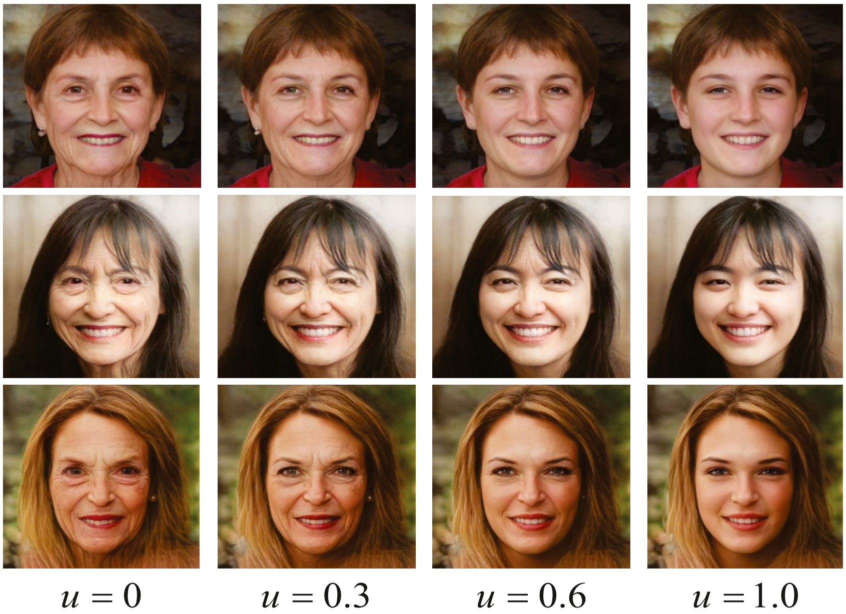

FIGURE 2 The makeup transformation effect of ACGAN's different age-compensation strengths. u indicates the intensity of age compensation, and a larger u indicates a higher corresponding compensation intensity. Age compensation is done by the age-compensation framework of ACGAN. The implementation principle is that the input face image is mapped into the corresponding high-dimensional feature space to obtain the corresponding semantic vector, and the age compensation is achieved by adjusting the direction of the semantic vector. Since we design a new encoder (MSEncoder) based on homology continuity in the age-compensation framework, it makes the final age-compensation effect more natural. At the same time, there is a good preservation of other background information and the identity information of the face

continuum through which the switching of the different states is achieved.

Homology continuity is one of the most fundamental properties, which is reflected in many aspects and has been rapidly developed in recent years, and has applications in many aspects. For example, a newborn lamb, after several years of development, becomes an adult sheep. The adult sheep differs greatly in appearance from the previous lamb, and the whole process is continuous and slow as the lamb ages. For example, for people, their facial expressions are variable, and the change of facial expressions is achieved by the movement of facial muscles.

The traditional pattern recognition method is based on the best segmentation of the classified samples in the feature space to accomplish the classification. In contrast, homology continuum-based bionic pattern recognition is achieved by, after preprocessing the sample points, mapping them into a high-dimensional feature space to achieve the classification effect through the best coverage of the sample distribution. Compared with traditional pattern recognition methods, coverage learning can obtain more accurate results with fewer parameters involved, making the network more lightweight as a way to achieve further model deployment and grounded applications.

3.3 | Covered learning theory

For the real world, 1D is a line, 2D is a surface, and 3D is a spatial body, while for multidimensionality, it is difficult to visualise. Therefore, for high-dimensional space, due to its existence being challenging to observe, in its corresponding high-dimensional feature space. The distribution shape of the corresponding patterns generally cannot be simple regular geometries. In order to solve this problem, existing simple geometries need to be combined as basic units to build more complex geometric distributions. Moreover, through the constructed complex geometry to complete the specified pattern, the best coverage of the specified pattern is accomplished by the constructed complex geometry.

Overlay learning differs from traditional learning methods in that it is based on the principle of homologous continuity. In the age-compensation network branch, we use overlay learning to compensate the age by mapping the face image into the high-dimensional feature space and by regressing the original face makeup image. The whole procedure can be roughly summarised as follows: first, the input face makeup image is mapped into the high-dimensional feature space, and its stream shape distribution in the high-dimensional feature space is mined. Then, the flow distribution is constructed by self-organising mapping, and then in the flow distribution subspace, coverage learning is performed, and finally age compensation is completed by adjusting the direction of the corresponding semantic vectors. The visualisation process of coverage learning is shown in Figure 3. Here, unlike the traditional parameter learning method, this coverage learning of spherical regions is more intuitive and transforms the training of parameters into a coverage problem of the sample point set.

3.4 | Super sausage neuron model

Inspired by the bionic pattern recognition, a lightweight and malleable super sausage neuron model is introduced. An example of super sausage neuron for classification is shown in Figure 4. As the basic computation unit in the makeup transformation network, the neuron directly determines the effect of the generated face image makeup. The super sausage neuron is introduced to achieve the face image identity and background information consistency before and after the makeup transformation. The super sausage neurons have good plasticity and can change the corresponding coverage parameters adaptively according to the sample distribution characteristics of the dataset. At the same time, it is able to maximise the presence of distribution of the face image dataset and complete the update and adjustment of the parameters. Therefore, its natural plasticity has a good match with different styles of makeup in the makeup transformation task, as a way to apply different degrees of makeup styles and obtain better face makeup transformation images.

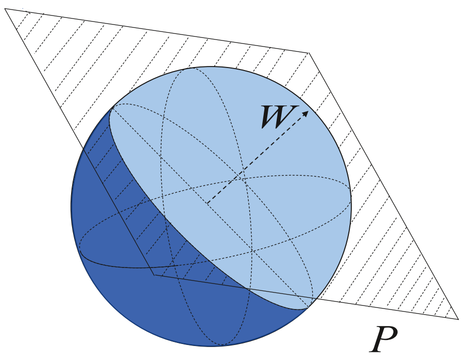

FIGURE 3 Visualisation of the covering learning process, where P denotes a hyperplane and W is a weight vector. Overlay learning is a new network learning algorithm that adaptively determines the optimal hyperparameters and is able to achieve end-to-end learning

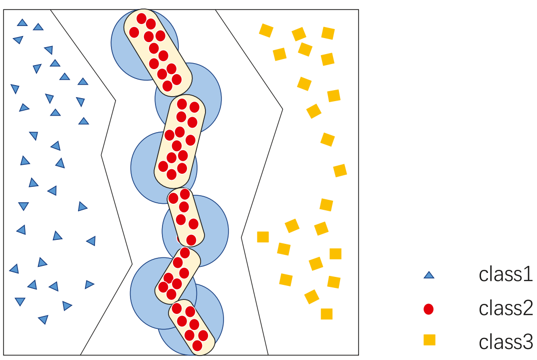

FIGURE 4 Here different colours and shapes represent different categories. An example of classification based on homology continuity. Traditional pattern recognition is to perform the best segmentation of different samples in the feature space, whereas homologous continuity-based affine pattern recognition is the best coverage of different samples in the feature space. With the introduction of homologous continuity, the coverage is smaller compared to the optimal segmentation, due to the optimal coverage, that is, better classification results and lower correct rates can be obtained. In the age branch of ACGAN, a new encoder is designed based on homology continuity, which is able to achieve an accurate complement of age and a stable and controllable age-compensation effect

3.5 | Multi-style encoder

Multi-style encoder (MSEncoder) is mainly responsible for extracting the makeup style and the face image's background and identity information. Keep the face's identity and background information consistent before and after the makeup transformation. The MSEncoder obtains the contextual semantic vectors of different styles of makeup from the face images with makeup. A dual parallel encoder–decoder architecture is used here, which is compatible with the characteristics of GPU and can optimise the speed of the network from hardware. At the same time, 68 face feature points are used to locate the relative position of the face, to further perceive the makeup information and to embed this part of the semantics into a new 'representation'. As shown in Figure 5, it is the effect of the age branch on the compensation of different intensities of age.

3.6 | Attentional mechanism

Authors in Ref. [34] draw attention to the first application of attentional mechanisms in the discipline of computing. Attention mechanisms have been widely used in recent years for many problems, such as pattern recognition [35], image translation [36], scene understanding [37, 38], and so on. Attention mechanism has become a hot topic in deep learning nowadays. Initially, the attention mechanism refers to the human visual system, which tends to focus on specific parts and ignore other parts when processing a large amount of information. This mechanism allows humans to quickly extract a large amount of useful information in a short period, which greatly improves the overall processing efficiency of information. We introduce an attention mechanism in the makeup transformation branch to enable it to focus on both identity and makeup information. The encoder–decoder structure, which leads to a large amount of information that prevents the network from processing quickly, is optimised by introducing the attention mechanism.

$$\text{Attention} = f(g(x), x) \tag{1}$$

$g(x)$, attending to the discriminative regions; $f(g(x), x)$, processing input x based on the attention $g(x)$.

3.7 | Objective function

3.7.1 | Cycle-consistent loss

The cycle-consistency loss was originally proposed by the authors of CycleGAN and has been used in a large number of

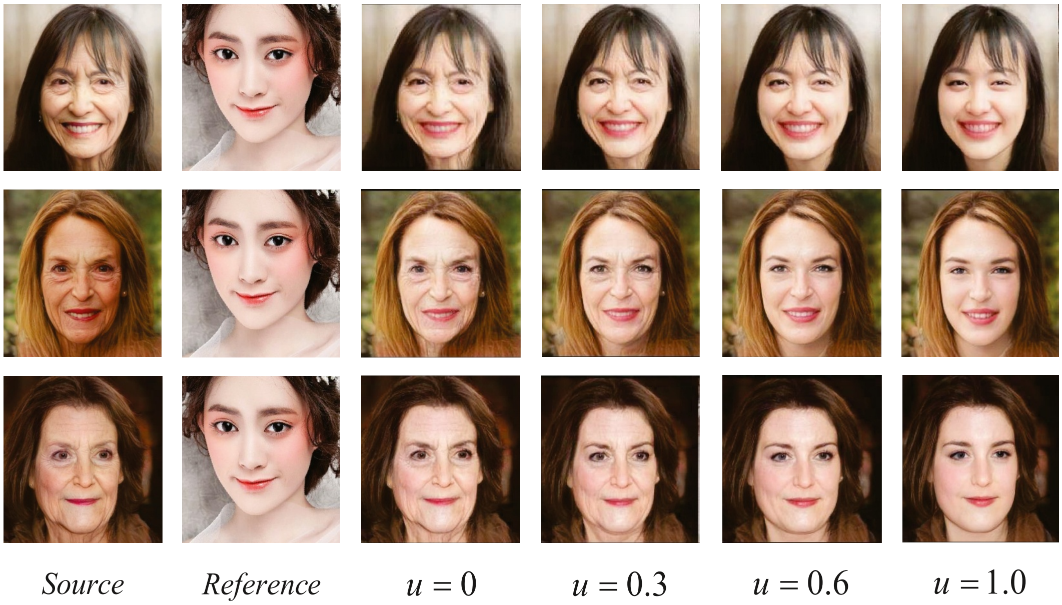

FIGURE 5 Compensation effect of age branch on face makeup images with different intensities. Here, we used the age-compensated effects of ACGAN with different intensities corresponding to the makeup transformed face images, where u represents the effect of age compensation, and the larger u indicates the higher intensity of age compensation. ACGAN has a makeup transform branch and an age-compensation branch, and after inputting the face image, the corresponding age semantic vector is obtained in the age-compensation framework, and after performing compensation, the corresponding semantic vector after age compensation is obtained. Meanwhile, in the makeup transformation framework, the face identity and background information vectors and the corresponding makeup vectors are extracted and transferred to the generator to generate the face image after makeup transformation. In addition, there exists a discriminator, and through the iteration of the generator-discriminator game, it is finally able to generate the overall natural and highly appearance-attractive makeup transformed image

image translation tasks. When dealing with the makeup transformation problem, the introduction of cycle-consistency loss ensures the consistency of face identity information and background information before and after the makeup transformation. In the research process of makeup transformation, it is necessary to map from the source domain X of unpaired images to the target domain Y, as shown in Figure 6. To ensure the consistency of identity and background information before and after the makeup transformation, we introduce the cyclic consistent loss, where the distance $G(G(x, y), x)$ between x and the reconstructed image is taken as the perceptual distance, and we use the L_c loss for constrained reconstructed images and define the cyclic consistency loss L_G as

$$L_G^{cyc} = \mathbb{E}_{x\sim\mathcal{P}_{X,y\sim\mathcal{P}}}[\|G(G(x,y),x) - x\|_1] + \mathbb{E}_{x\sim\mathcal{P}_X}, y \sim \mathcal{P}_Y[\|G(G(y,x),y) - y\|_1]. \quad (2)$$

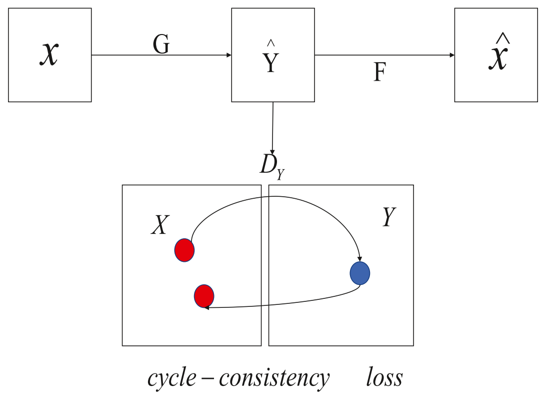

FIGURE 6 Schematic diagram of the process of cycle-consistent loss. The cyclic consistency loss is introduced to constrain the newly generated makeup-transformed face images by introducing cyclic consistency since we do not use the paired face images before and after the makeup transformation throughout the training process. Among them, X and Y are training samples, G and F are two mapping functions, and D_Y is the discriminator

3.7.2 | Makeup loss

When doing makeup transfer task, the main ones that have the greatest impact on the overall effect of the makeup are lipstick, eye shadow, and foundation. Therefore, it is necessary to focus on this part when dealing with makeup transformations. In ACGAN's makeup transformation branch, makeup loss is introduced. In the makeup transformation study, instead of using the missing histogram of the whole image, we divided the makeup style into three important parts: lipstick, eye shadow and foundation. During the study, it was found that the latent vector of the reference image contained makeup features, but the image representing the makeup style could not be found; therefore, the makeup features could not be obtained. As a result, we trained a makeup style with multiple latent features and thus extracted the corresponding attributes from the latent features. By calculating the average latent vectors of all images with makeup and all images of faces without makeup, which are noted as $\overline{L}^{Y}, \overline{L}^{X}$, respectively, their differences represent the direction from the non-makeup latent vector to the makeup latent vector, and thus we define the makeup loss L_m as

$$\mathcal{L}_m = \left\| M_r^Y - \left(\overline{L}^{Y} - \overline{L}^{X}\right) \right\|_2 \tag{3}$$

3.7.3 | Perceptual loss

In the field of makeup transformation, BeautyGlow is the first to use perceptual loss for facial feature extraction. Inspired by BeautyGlow, we introduce perceptual loss in the objective function. The makeup transformation problem also belongs to the task of image translation. Most of the methods use supervised methods to constrain images and keep the identity information and background information of the source face image as consistent as possible. If the pixel-level difference is directly measured, it will result in a large amount of computation and slow convergence of the final network. Therefore, we introduce a perceptual loss to complete the preservation of face identity information before makeup transformation. Perceptual loss, in the field of makeup transformation, does not directly measure the Euclidean distance between different pixel levels but calculates the difference between high-level features extracted by deep convolutional networks. In ACGAN, we use the VGG-16 model pre-trained on ImageNet. After the features are extracted, the Euclidean distance between the features is calculated, which is expressed in Equation (4). During computation, the semantic vector processed by the age-compensation branch is used for feature extraction. In this paper, we assume an image without makeup as the source image, and in order to extract facial features, we introduce a perceptual loss to make the images before and after makeup consistent. We define the perceptual loss L_p as

$$\mathcal{L}_p = \left\| F_s^X - L_s^X \right\|_2 \tag{4}$$

thus achieving the constraint on the distance between the original image L_s^X and the facial potential features F_s^X.

3.7.4 | Adversarial loss

To complete the training of iterative games for ACGAN, we introduced adversarial loss to handle this problem. The essence of the adversarial loss is to make the generated image consistent with the real image so that the generated image is more realistic. In the field of makeup transformation, we use two discriminators D_X and D_Y to distinguish between the source image domain X and the reference image Y so that the generator outputs a more realistic image. The adversarial loss L_D^{adv} between the discriminator and the generator is further calculated as

$$\begin{aligned} L_D^{adv} &= -\mathbb{E}_{x\sim\mathcal{P}_X}[\log D_X(x)] - \mathbb{E}_{y\sim\mathcal{P}_Y}[\log D_Y(y)] \\ &\sim -\mathbb{E}_{x\sim\mathcal{P}_X}, y \sim \mathcal{P}_Y[\log(1 - D_X(G(y,x)))] \\ &- \mathbb{E}_{x\sim\mathcal{P}_X}, y \sim \mathcal{P}_Y[\log(1 - D_Y(G(x,y)))]\ L_G^{adv} \\ &= -\mathbb{E}_{x\sim\mathcal{P}_X}, y \sim \mathcal{P}_Y[\log(D_X(G(y,x)))] \\ &- \mathbb{E}_{x\sim\mathcal{P}_X}, y \sim \mathcal{P}_Y[\log(D_Y(G(x,y))] \end{aligned} \tag{5}$$

3.7.5 | Total loss

For the total losses, we apply a weighting factor to each loss, thus defining the total loss objective function as

$$\begin{aligned} L_D &= L_{adv}^D + L_p^D + L_m^D \\ L_G &= L_{adv}^G + \lambda_{\mathrm{p}} L_{\mathrm{p}}^G + \lambda_m L_m^G + \lambda_{cycle} L_{cycle} \end{aligned} \tag{6}$$

where λ_p, λ_m, and λ_{cycle} are the weight factors corresponding to the perceptual loss, makeup loss, and cyclic consistency loss, respectively.

4 | EXPERIMENT

In this section, the experimental validation part of the model is presented in detail, including the dataset used, the training configuration used, and the results of qualitative and quantitative experiments. We also introduced the details of the experiment and the number of training parameters in detail. In addition, we also introduced the Baseline, which is the current state-of-the-art makeup model. Through quantitative and qualitative experiments, we illustrate that the model proposed is attractive in makeup. There is a certain degree of improvement in strength, especially as the age of the face image increases, showing a greater advantage. The coverage learning method is used to further improve the accuracy of age supplementation. In the mapped high-dimensional space, the direction adjustment of the relevant semantic vector can obtain a better age-compensation effect and achieve stability and controllability. As shown in Figure 7, it is the result of face parsing by ACGAN.

4.1 | Implementation details

4.1.1 | Dataset introduction

In the training process of the model, two main datasets are included, the FlickrFaces HQ (FFHQ) [39] and Makeup Transfer (MT) [10] datasets, where the FFHQ dataset is mainly used to train on the age-compensation branch and the MT dataset is mainly used to train on the Makeup Transfer branch. The FFHQ dataset consists of 70,000 high-resolution

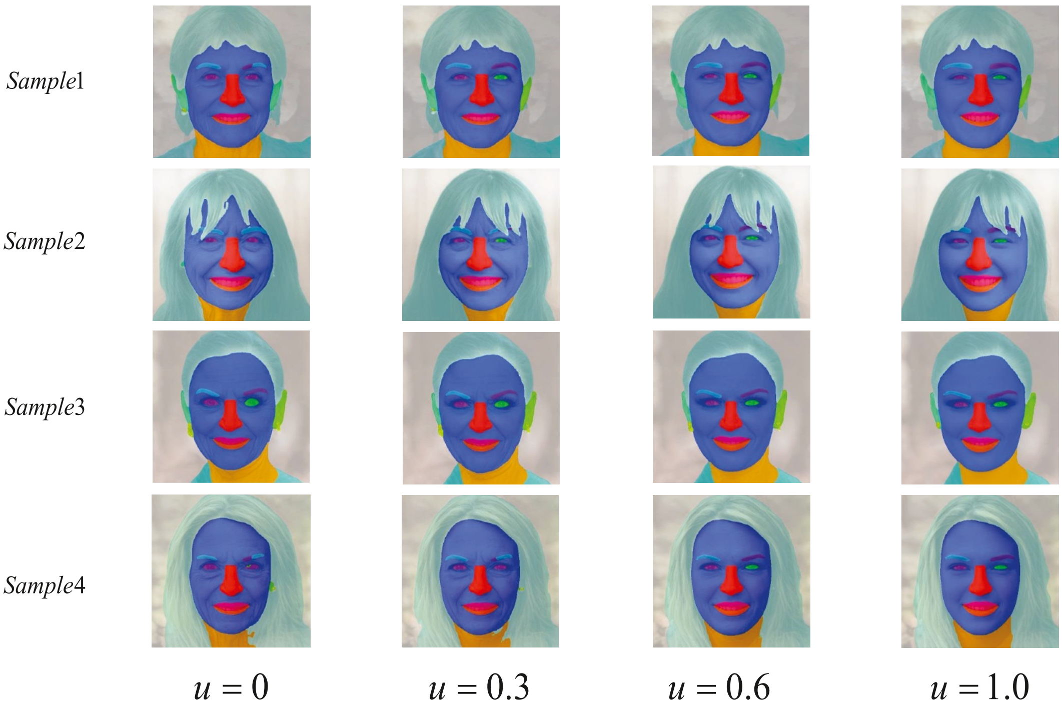

FIGURE 7 Rendering of ACGAN for face analysis. Face parsing is a kind of semantic segmentation, where face parsing calculates different semantic components in a face image and then gives a pixel-level label mapping. For example, after inputting a face image, the face parsing can finally get many semantic components, such as hair, skin, eyes, nose, mouth, ears etc. The human head is parsed at pixel level to get the corresponding pixel-level labels

(1024 × 1024) face images. The face images in this dataset are diverse, mainly in terms of background, ethnicity, age distribution etc. In addition, there are many decorations such as glasses, hats etc. For the FFHQ dataset, 10% of them are used as training test dataset, and the remaining 90% are used as training and validation data. The MT dataset, which is the most commonly used dataset in the field of makeup transformation, contains 3834 face images with a resolution of 361 × 316, of which 2718 are face images with makeup and 1115 are face images without makeup. For the MT dataset, 100 of the non-makeup face images and 200 face images with makeup are used for testing the model, and the remaining face images are used for training and validation of the model.

4.1.2 | Training details

We implemented ACGAN using the Pytorch [40] scientific computing framework. We used the FFHQ dataset to train the age-compensation branch, while we used the MT dataset to train the makeup transformation branch. We used a four-way NVIDIA TITAN V to train and infer the model. We used Adam optimiser [41] with a learning rate of 0.0001 to train ACGAN, and the model gradually converged after about 500 epochs. Limited by the size of video memory, we use batch-size = 16 and mainly use the sigmoid function as the activation function. For better results, we use the Dlib library to crop and deflate the face images.

4.1.3 | Face parsing

In the process of makeup transformation, face analysis is critical. Pixel-level semantic labels need to be assigned to essential parts. In ACGAN, we use BiSeNet [42] for face parsing, pre-train on the CelebMask-HQ database, and GPU for inference. This global approach predicts semantic labels for essential parts of a face image, such as eyes, nose, mouth, eyebrows, hair etc. In the age-compensation branch, we compensate the age and perform face analysis on the compensated face image. After obtaining the corresponding label, we can obtain the face image after makeup transformation through the generator. In Figure 7, we used four face images as example images. The parameter u represents the strength of the age compensation. Correspondingly, the larger the u, the higher the strength of the age compensation. That is, the larger the u, the smaller the age from the visual point of view.

4.2 | Quantitative experiment

For the field of makeup transformation, the comparison of its generated images is difficult to have a real quantitative analysis, so the most makeup transformation effect should be evaluated from the consistency of the identity information and background information of the images before and after makeup, the attractiveness of the appearance of the images after makeup, the naturalness of the images, the consistency of the makeup etc.

Then, these indicators are difficult to be quantified, so in order to evaluate the model objectively, we use the form of expert scoring to evaluate the images. We invited 20 experts in the field as judges to score the face images, and in order to eliminate the preference for different makeup styles between men and women, 10 female experts and 10 male experts were finally assigned to score the images. There are five indicators: identity information of the image before and after the makeup transformation, consistency of background information before and after the makeup transformation, attractiveness of the image, overall naturalness of the image, and consistency of the makeup, each of which scores 20 points, and the total score is 100 points. The results of the scoring are shown in Table 2.

4.3 | Qualitative experiment

Figure 8 shows the comparison of ACGAN with the current state-of-the-art makeup transformation methods. It is obvious that ACGAN has good results with the help of age-compensation mechanism. In particular, it has a clear advantage in terms of appearance attractiveness. In the future makeup transformation will likely join the metaverse ecology; virtual makeup transformation methods are the future trend, and when ACGAN is used for makeup transformation, better appearance attractiveness can be obtained for users. In addition, ACGAN has a more natural effect, because ACGAN is considered for the whole when performing makeup transformations, so it can obtain a natural effect. We list the current state-of-the-art makeup transformation methods and their corresponding code addresses in the Table 3 and the reproduction results in the text, from the corresponding code addresses.

5 | CONCLUSION AND FUTURE WORK

In this paper, we propose a new homology-continuous-based makeup transformation method, which can be roughly divided into two network branches: the age-compensation branch and the makeup transformation branch. Specifically,

TABLE 2 Comparison with state-of-the-art makeup transformation methods

Method	Identity information consistency	Background consistency	Attractive appearance	Natural	Makeup consistency	Total score
DMT	16	16	16	15	17	80
BeautyGAN	17	17	14	14	16	78
CPM	16	16	15	16	18	81
ACGAN(Ours)	18	17	19	18	19	91

Note: We conducted a quantitative comparison with the current state-of-the-art makeup transformation methods. First, we gathered 20 experts in the image field and introduced 10 males and 10 females in order to eliminate the different preferences for makeup between men and women. The makeup-transformed face images were scored in five aspects: identity information consistency, background information consistency, appearance attractiveness, naturalness, and makeup consistency, each with 20 points, for a total of 100 points. From the table, we can see that ACGAN has a clear advantage in appearance attractiveness, because most of the current methods ignore the influence of age, and ACGAN can obtain a clear improvement in appearance attractiveness after supplementing age. In addition, it also outperforms the existing models in other aspects.

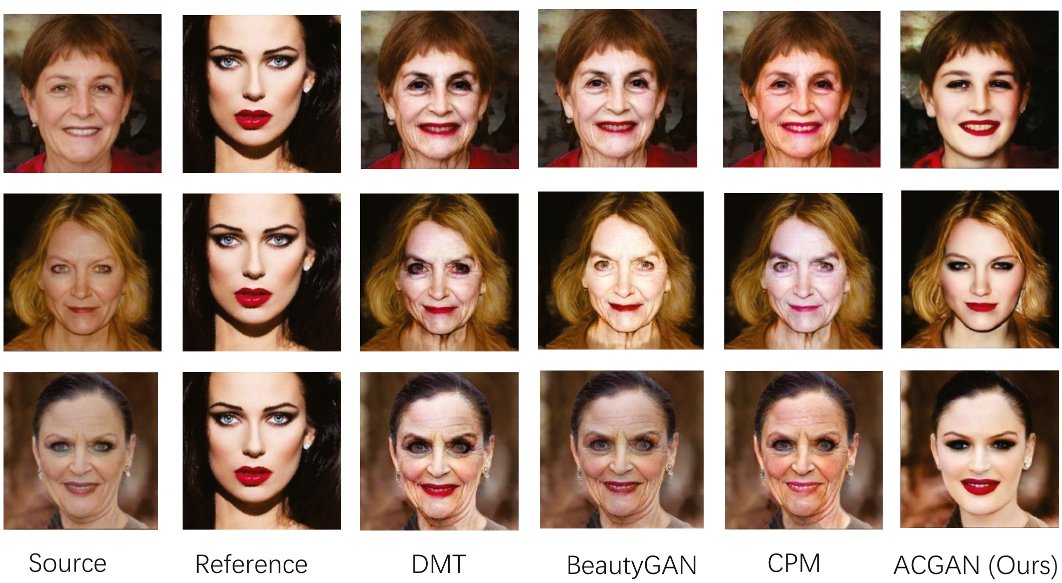

FIGURE 8 The comparison of ACGAN with the current state-of-the-art makeup transformation methods. ACGAN has been qualitatively compared with the prior state-of-the-art makeup transformation method, and from the figure, it is obvious that the images generated by ACGAN, due to age compensation, make the final generated face makeup images, have a better naturalness and overall look more harmonious. In addition, the improvement in appearance attractiveness is also very obvious. Not only that, ACGAN can also better retain the consistency of face identity information and the consistency of background information

TABLE 3 State-of-the-art makeup transformation methods and their corresponding code addresses

Method	Github	Note
DMT [43]	https://github.com/Honlan/DMT	Official
BeautyGAN [10]	https://github.com/wtjiang98/BeautyGAN_pytorch	Official
CPM [21]	https://github.com/VinAIResearch/CPM	Official

in the age-compensation branch, based on the same Source Continuity designed a new encoding module, which can map the face vector into the corresponding high-dimensional vector space and realise the compensation for age by adjusting the vector direction. In the makeup transformation branch, we designed a multi-style encoder to handle different types of makeup, such as Chinese makeup, Japanese makeup, Korean makeup etc. At the same time, since the face makeup image will be converted into a large number of semantic vectors during the encoding process, to further improve the processing speed, we introduce an attention mechanism to speed up the information processing speed. In addition, our proposed network structure is a two-pass encoder–decoder structure, which has good parallelism and can achieve better results when training and inference on GPU. In the work of ACGAN, we try our best to deal with the existing problems, such as ignoring the age factor, but due to some reasons, there are still some problems. For example, the current makeup transformation is 2-dimensional and in the real world it is 3-dimensional, so there is a big difference between 2-dimensional makeup and 3-dimensional makeup. Next, we will focus on multi-view consistent makeup transformation. In addition, for another example, the currently existing methods use relatively large network structures when performing makeup transformation. In the future, we will also focus on lightweight network design to complete the makeup transformation task.

ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61901436, 51974036 and 62006028.

CONFLICT OF INTEREST

The authors declared that they have no conflicts of interest to this work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in reference number [10, 39].

PERMISSION TO REPRODUCE MATERIALS FROM OTHER SOURCES

Please mention whether you have used any copyrighted materials in your work and also inform whether you have obtained permission to use the material. We have not used any copyrighted materials in our work.

ORCID

Feng He https://orcid.org/0000-0001-5059-4423
Bo Jin https://orcid.org/0000-0001-9255-5772

REFERENCES

1. Guo, D., Sim, T.: Digital face makeup by example. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 73–79. IEEE (2009)
2. Tong, W.-S., et al.: Example-based cosmetic transfer. In: 15th Pacific Conference on Computer Graphics and Applications (PG'07), pp. 211–218. IEEE (2007)
3. Wang, C., et al.: Learning discriminative features by covering local geometric space for point cloud analysis. IEEE Trans. Geosci. Rem. Sens. 60, 1–15 (2022). https://doi.org/10.1109/tgrs.2022.3170493
4. Jia, D., et al.: Multi-stream densely connected network for semantic segmentation. IET Comput. Vis. 16(2), 180–191 (2022). https://doi.org/10.1049/cvi2.12078
5. Ren, Y., Huang, Z.: Distribution Probability-Based Self-Adaption Metric Learning for Person Re-identification. IET Computer Vision (2022)
6. Zhang, H., et al.: Object detection via structural feature selection and shape model. IEEE Trans. Image Process. 22(12), 4984–4995 (2013). https://doi.org/10.1109/tip.2013.2281406
7. Ning, X., et al.: Face editing based on facial recognition features. IEEE Trans. Cogn. Dev. Syst., 1 (2022). https://doi.org/10.1109/tcds.2022.3182650
8. Goodfellow, I.: Nips 2016 Tutorial: Generative Adversarial Networks (2016). arXiv preprint arXiv: 1701.00160
9. Zhu, J.-Y., et al.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232. (2017)
10. Li, T., et al.: BeautyGAN: instance-level facial makeup transfer with deep generative adversarial network. In: Proceedings of the 26th ACM International Conference on Multimedia, pp. 645–653. (2018)
11. Chang, H., et al.: PairedCycleGAN: asymmetric style transfer for applying and removing makeup. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 40–48. (2018)
12. Gu, Q., et al.: LADN: local adversarial disentangling network for facial makeup and de-makeup. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10481–10490 (2019)
13. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5(4), 115–133 (1943). https://doi.org/10.1007/bf02478259
14. Broomhead, D.S., Lowe, D.: Multivariable functional interpolation and adaptive networks. Complex Syst. 2 (1988)
15. Moody, J., Darken, C.J.: Fast learning in networks of locally-tuned processing units. Neural Comput. 1(2), 281–294 (1989). https://doi.org/10.1162/neco.1989.1.2.281
16. Ning, X., Li, W., Jiang, Xu: The principle of homology continuity and geometrical covering learning for pattern recognition. Int. J. Pattern Recogn. Artif. Intell. 32(12), 1850042 (2018). https://doi.org/10.1142/s0218001418500428
17. Gou, D., et al.: A face aging network based on conditional cycle loss and the principle of homology continuity. In: 2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), pp. 264–268. IEEE (2019)
18. Wang, C., et al.: Uncertainty estimation for stereo matching based on evidential deep learning. Pattern Recogn. 124, 108498 (2022). https://doi.org/10.1016/j.patcog.2021.108498
19. Jiang, W., et al.: PSGAN: pose and expression robust spatial-aware GAN for customizable makeup transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5194–5202. (2020)
20. Liu, Si, et al.: PSGAN++: robust detail-preserving makeup transfer and removal. IEEE Trans. Pattern Anal. Mach. Intell., 1 (2021). https://doi.org/10.1109/tpami.2021.3083484
21. Nguyen, T., Tran, A.T., Hoai, M.: Lipstick ain't enough: beyond color matching for in-the-wild makeup transfer. In: Proceedings of the IEEE/

CVF Conference on Computer Vision and Pattern Recognition, pp. 13305–13314. (2021)
22. Yuan, Q.-L., Zhang, H.-L.: RAMT-GAN: realistic and accurate makeup transfer with generative adversarial network. Image Vis Comput. 120, 104400 (2022). https://doi.org/10.1016/j.imavis.2022.104400
23. FitzGerald, J.D., et al.: Impact of the medicare short stay transfer policy on patients undergoing major orthopedic surgery. Health Serv. Res. 42(1p1), 25–44 (2007). https://doi.org/10.1111/j.1475-6773.2006.00606.x
24. Li, Y., et al.: Universal style transfer via feature transforms. Adv. Neural Inf. Process. Syst. 30 (2017)
25. Gatys, L.A., et al.: Controlling perceptual factors in neural style transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3985–3993. (2017)
26. Huang, H., et al.: Real-time neural style transfer for videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 783–791. (2017)
27. Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, pp. 341–346. (2001)
28. Drori, I., Cohen-Or, D., Yeshurun, H.: Example-based style synthesis. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings, vol. 2, pp. II–143. IEEE (2003)
29. Frigo, O., et al.: Split and match: example-based adaptive patch sampling for unsupervised style transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 553–561. (2016)
30. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414–2423. (2016)
31. Lee, H.-Y., et al.: Diverse image-to-image translation via disentangled representations. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 35–51. (2018)
32. Vincent, D., Shlens, J., Kudlur, M.: A Learned Representation for Artistic Style (2016). arXiv preprint arXiv: 1610.07629
33. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1501–1510. (2017)
34. Bahdanau, D., Cho, K., Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate (2014). arXiv preprint arXiv: 1409.0473
35. Li, X., et al.: Without Detection: Two-step Clustering Features with Local–Global Attention for Image Captioning. IET Computer Vision (2022)
36. Fukui, H., et al.: Attention branch network: learning of attention mechanism for visual explanation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10705–10714. (2019)
37. Yan, C., et al.: Stat: spatial-temporal attention mechanism for video captioning. IEEE Trans. Multimed. 22(1), 229–241 (2019). https://doi.org/10.1109/tmm.2019.2924576
38. Liu, G., Guo, J.: Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing. 337, 325–338 (2019). https://doi.org/10.1016/j.neucom.2019.01.078
39. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4401–4410. (2019)
40. Adam, P., et al.: PyTorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32 (2019)
41. Kingma, D.P., Jimmy, Ba: Adam: A Method for Stochastic Optimization (2014). arXiv preprint arXiv: 1412.6980
42. Yu, C., et al.: BiSeNet: bilateral segmentation network for real-time semantic segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 325–341 (2018)
43. Zhang, H., et al.: Disentangled Makeup Transfer with Generative Adversarial Network (2019). arXiv preprint arXiv: 1907.01144

How to cite this article: Wu, G., et al.: ACGAN: Age-compensated makeup transfer based on homologous continuity generative adversarial network model. IET Comput. Vis. 17(5), 537–548 (2023). https://doi.org/10.1049/cvi2.12138