Assessing Eye Aesthetics for Automatic Multi-Reference Eye In-Painting

Bo Yan *, Qing Lin, Weimin Tan, Shili Zhou
Shanghai Key Laboratory
of Intelligent Information Processing,
School of Computer Science, Fudan University
{byan, 18210240028, wmtan14, 15307130270}@fudan.edu.cn

Abstract

With the wide use of artistic images, aesthetic quality assessment has been widely concerned. How to integrate aesthetics into image editing is still a problem worthy of discussion. In this paper, aesthetic assessment is introduced into eye in-painting task for the first time. We construct an eye aesthetic dataset, and train the eye aesthetic assessment network on this basis. Then we propose a novel eye aesthetic and face semantic guided multi-reference eye in-painting GAN approach (AesGAN), which automatically selects the best reference under the guidance of eye aesthetics. A new aesthetic loss has also been introduced into the network to learn the eye aesthetic features and generate high-quality eyes. We prove the effectiveness of eye aesthetic assessment in our experiments, which may inspire more applications of aesthetics assessment. Both qualitative and quantitative experimental results show that the proposed AesGAN can produce more natural and visually attractive eyes compared with state-of-the-art methods.

1. Introduction

Aesthetic quality assessment [14, 15] has been gaining increasing demands with the wide applications of digital images in social, communication, entertainment, and shopping, *etc*. The aesthetic quality of an image largely determines its using possibility due to the nature of human loving aesthetic things. Assessing image aesthetic quality is essential for screening beautiful pictures from massive online images, recommending beautiful pictures that users enjoy, and understanding image attributes such as image composition, contrast, and lighting. Eye aesthetic assessment, a branch of image aesthetic assessment, aims at using computational methods to evaluate the "aesthetic feeling" of face images by simulating human perception and cognition of beauty. Eye aesthetic quality has a great influence on users' satisfaction with photos. Knowing eye aesthetic quality is also useful for selecting better face images. In addition, it can be used as a guide to restore poor-quality eyes to high-quality ones, *i.e.*, eye in-painting.

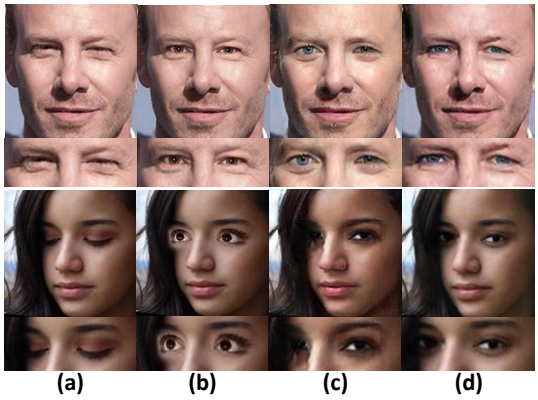

Figure 1. Eye in-painting results. Columns represent: (a) Image to in-paint, (b) The commercial state-of-the-art eye opening algorithm in Adobe Photoshop Elements 2019, (c) ExGAN result [7] and (d) Our AesGAN result.

Eye aesthetic quality assessment is a challenging task due to its extremely subjective nature. Assessing eye aesthetic quality cannot simply use objective quality assessment methods such as PSNR, MSE, and SSIM [25], which are commonly used to assess the distortions of image. In contrast, the assessment of eye aesthetic quality needs a lot of manual marking due to people's preferences. Currently, there has been no research on the assessment of eye aesthetic quality. In addition, the eye aesthetic quality assessment can help us know the quality of the eyes in a face image. Not satisfied with just knowing the aesthetic quality of the eyes, we also hope that we can use this as a guide to make the poor eyes more realistic. As a result, eye in-painting is produced, which can be seen as an application of eye aesthetics.

There are few researches on eye in-painting. Figure 1

*This work was supported by NSFC (Grant No.: 61772137).

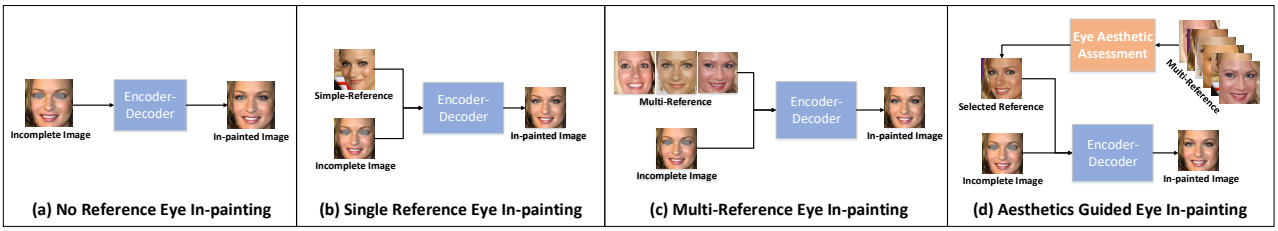

Figure 2. Comparison of different eye in-painting frameworks. (a) Traditional eye in-painting methods do not use references, which only use incomplete images as input and finally output repaired images through encoders and decoders. (b) Single reference based eye in-painting methods use one reference image to assist in-painting, while (c) multi-reference based eye in-painting methods take multiple references. (d) Our proposed aesthetic guided eye in-painting method takes aesthetics as the criterion of reference selection, and uses the final selected image as the input.

shows some results. Eye in-painting is a branch of face restoration problem, which is mainly applied in closed eyes and squint eyes situations, in order to produce real and natural new eyes. The current approaches can be summarized as the three types of frameworks shown in Figure 2. The previous three frameworks pose attention on whether to use reference samples or how many reference samples to use. ExGAN [7] method has well shown that the identity preserved eye in-painting result can be obtained by referring to the same identity example. Nevertheless, these three frameworks do not consider the selection of reference examples, which is a practical problem. In addition, eye aesthetic attributes that are crucial to eye in-painting, have not been considered in these frameworks. These observations motivate us to develop a new framework: Aesthetics Guided Eye In-painting, which addresses the selection problem of multiple reference examples based on our proposed eye aesthetic assessment.

Due to the lack of research on eye aesthetic, there is no dataset that can be directly used in eye aesthetic assessment. On the basis of the existing face dataset, we cut out the eye area and invite 22 volunteers to assess the eye quality aesthetically. The marked dataset contains 1,040 eye images, all of which are divided into two categories according to the average level of manual tagging: aesthetically pleasing or not. Based on this dataset, we train the eye aesthetic assessment network. We introduce the eye aesthetic assessment system into the work of eye in-painting. First, it can automatically select the appropriate reference for the eye generation. Then we introduce aesthetic loss to force the in-painting network to learn the eye aesthetic features, and promote the generated eyes to be more realistic.

This paper mainly has the following contributions:

1. This paper first shows the effectiveness of aesthetic assessment in the multi-reference eye in-painting task, which may inspire the introduction of aesthetic assessment to other work in the future.

2. We annotate a new eye aesthetic dataset and construct an eye aesthetic assessment network. To the best of our knowledge, we are the first to introduce the branch of image reconstruction into the quality assessment network, and verify its effectiveness in maintaining the sample uniqueness and improving the network performance.

3. Through the eye in-painting task, the effectiveness of the eye aesthetic assessment network is proved. We propose a novel eye aesthetic and face semantic guided multi-reference eye in-painting GAN approach (AesGAN). By using the high quality reference and aesthetic features provided by the eye aesthetic assessment network, the performance of our eye in-painting method is better than those of state-of-the-art methods in both qualitative and quantitative results.

2. Related Work

2.1. Aesthetic Assessment

The image aesthetic assessment [8, 12, 6] aims to use the computer to simulate human perception and cognition of aesthetic, which has important application prospects in clothing design, beauty makeup, face editing, and pictures beautifying [20, 3], *etc*. In addition to some objectivity, image aesthetic quality assessment has a strong subjective, which is more difficult than other image processing tasks.

Recently, the image aesthetic assessment is regarded as an independent task. However, the deep network for extracting the aesthetic features may not well explain the aesthetics. For different aesthetic tasks, the definition of aesthetics changes accordingly. Facial images are different from ordinary natural images which have more specific aesthetic characteristics [21], especially the eye area. Therefore, we consider that the aesthetic assessment can be applied to more specific image processing tasks, such as eye in-painting. This not only helps to produce more realistic results, but also makes the aesthetic features extracted by aesthetic network more explanatory.

Compared with other computer vision tasks, the data acquisition of image aesthetics is more difficult and the overall data size is smaller. Taking image recognition task as an ex-

ample, this task has a large number of research results and large datasets, such as ImageNet [5] with more than 14 million of images and tagging data. Only a few datasets are available for image aesthetic quality assessment, of which the largest ones, AROD [24] and AVA [19], only have 380K and 250K images, respectively. The tagging data of these images are obtained by users on online image-sharing sites. Most of these images come from camera photography and can not be directly used for the eye aesthetic assessment.

Due to the lack of research on eye aesthetic, we mark a dataset containing 1,040 eye images. Based on this dataset, we train the eye aesthetic assessment network. Then we introduce the eye aesthetic assessment system into the work of eye in-painting.

2.2. Eye In-Painting

Recently more people tend to record their lives with selfies. Plenty of photos appear on social media every day, especially portraits, some of which need to be repaired for blemish [29, 4, 26]. The blinking and closing eyes in the photos are big problems that bother people, which promote the task of eye in-painting.

With the wide use of GAN [10], image restoration work can obtain more real results [22, 11, 28]. The non-reference GAN can only generate the eyes according to the experience, not to the given identity in the photo. However, different people's appearance and face structure are diverse. The exemplar of a person is necessary, which can make GANs generate new eyes that are more consistent with the identity of people [18].

The previous eye in-painting methods combine the reference image directly with the image to be repaired [1, 2, 23]. These methods do not take into account the semantic and structural information around the eye, so they show a poor in-painting performance when the light or the facial posture is different, such as the commercial state-of-the-art eye opening algorithm shown in Figure 1(b). In addition, some of the methods rely on automatic eye recognition, but the eye parts of many closed-eye photographs are not well detected.

ExGANs [7] use one of the different images of the same identity as the reference for generator training, which can provide additional information to the generation network. Different from previous GANs, these additional identity information can be inserted into the network at multiple points to help it have better expression ability. Although ExGANs can produce real eye in-painting results, there are also some limitations. The random selection of exemplar can only provide the basic reference information, but does not take into account the quality and fitness of the eyes. When confronted with occluded eyes and side faces, ExGANs do not perform satisfactorily. Therefore, we add the constraints of eye aesthetic assessment and face semantic parsing, and replace the original square masks with the elliptical ones to solve these limitations. Figure 1 also shows the results of our eye in-painting networks compared with ExGANs.

Figure 3. The labeled eye aesthetics assessment dataset according to manual scoring. The dataset has a total of 1,040 eye images divided into two categories. The first line shows low-quality eye images, and the second line shows high-quality ones.

3. Eye Aesthetic Assessment

3.1. Building Aesthetic Dataset

The eye aesthetics assessment is still a new topic with few studies. Unlike the traditional aesthetic assessment task, we need to build a specific dataset for eye aesthetic. In view of the difficulty of obtaining aesthetic dataset, we choose to label the traditional face dataset. The CAS-PEAL[9] Face Dataset contains 1,040 face-frontal images and standardizes the remaining objective factors, making the training of the network more focused on extracting effective features of the eyes. So we choose to use the CAS-PEAL dataset for tagging and training.

Based on the CAS-PEAL Dataset, we annotate a new eye dataset with 1,040 eye images. 22 volunteers are invited to make the eye aesthetic assessment. According to the average level of manual tagging, the dataset is divided into two grades: high quality and low quality. Figure 3 shows part of the eye aesthetic dataset. The number before the eye images represent the score that they are rated. The eyes rated as 2 are generally more aesthetically pleasing.

3.2. Aesthetic Assessment Network

Based on the eye aesthetic dataset, we propose an eye aesthetics assessment network (AesNet). The traditional quality assessment network has only one branch, which directly trains a classifier to output the corresponding image quality level in an end-to-end manner. For aesthetic quality, especially the beauty of the eyes, each sample has its own unique features. In order to maintain the uniqueness of the samples, we add a reconstruction branch to the image quality assessment task for the first time. As shown in Figure 4, our AesNet consists of three parts: eye aesthetic feature extraction, eye scoring and eye reconstruction. The aesthetic feature extraction module contains an encoder and nine residual blocks. The encoder has three convolution modules, which consists of one convolution layer, one normalized layer, one relu activation layer and one max-pooling layer.

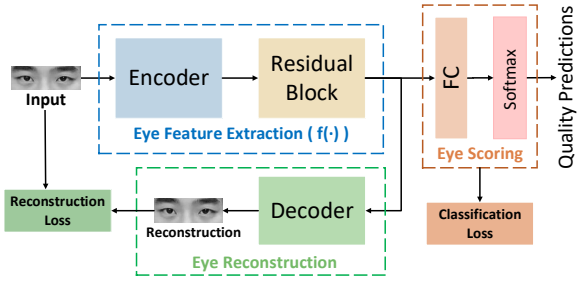

Figure 4. The architecture of our eye aesthetic assessment network. We first introduce the reconstruction branch into the image quality assessment task to maintain the uniqueness of eye aesthetic. Only the eye aesthetic feature extraction module and the eye scoring module are needed during testing.

We send the extracted features into the eye scoring module and the eye reconstruction module at the same time. The eye scoring module outputs the prediction result of the eye assessment. The eye reconstruction module is composed of a decoder and the output is a generated eye image. We use reconstruction loss to constrain the generated eyes to be similar to the input ones. Only the eye aesthetic feature extraction module and the eye scoring module are needed during testing.

Assume that for each image e_i input to the AesNet, we have its corresponding aesthetic label y. We use the softmax cross-entropy loss as the classification loss defined as

$$\mathcal{L}_{Classification} = -\sum_{j=1}^{T} y_j \log s_j \tag{1}$$

where T is the number of categories, and s_j is the j-th value of the softmax output vector, which represents the probability that the sample belongs to category j. We use the MSE loss as the reconstruction loss defined as

$$\mathcal{L}_{Reconstruction} = \frac{1}{n}\sum_{i=1}^{n}(g_i - e_i)^2 \tag{2}$$

where g_i is the generated eye image, and n is the pixel number. The overall loss function is defined as

$$\mathcal{L}_{EyeAes} = \mathcal{L}_{Classification} + \lambda_{rec}\mathcal{L}_{Reconstruction} \tag{3}$$

where λ_{rec} is the weight to balance the effects of different losses and takes 0.01 when training. We divide the dataset into five and cross-validate. The trained eye aesthetic assessment network can achieve the accuracy of 0.84.

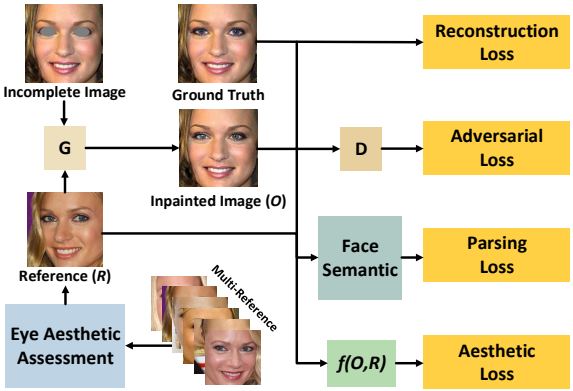

Figure 5. The architecture of our eye in-painting network (AesGAN) based on eye aesthetic and face semantic, containing a generator, two discriminators, an eye aesthetic assessment network and a parsing network. The function *f(O,R)* is the eye aesthetic feature extraction module in Figure 4.

4. Eye In-Painting with Eye Aesthetic Assessment

4.1. Overview

We introduce the eye aesthetic assessment into the eye in-painting task, and propose the eye aesthetic and face semantic guided multi-reference eye in-painting method (AesGAN). Given an incomplete image, our goal is to produce natural and attractive eyes that are both semantically consistent with the whole object and visually realistic. Figure 5 shows the proposed network that consists of one generator, two discriminators, an eye aesthetic assessment network and a parsing network.

We use the eye aesthetics assessment network and structural similarity index (SSIM) to automatically select the best reference. In order to highlight the role of eye aesthetic assessment, we introduced a new aesthetic loss. At the same time, a parsing loss is added to ensure the fidelity and semantic consistency of pixels. The parameters of parsing network and eye assessment network are fixed when training.

4.2. Guidance of Eye Aesthetic Assessment

People's growing pursuit of beauty gives a new idea to the image restoration task. It is instructive to introduce eye aesthetic assessment to eye in-painting task. The guidance of eye aesthetic assessment is manifested in three aspects.

Firstly, based on eye aesthetics, we propose a multi-reference selection mechanism. We use the AesNet to score the eyes of the references. Then we calculate the SSIM values of each image except for the eye region in order to select the image which is most similar to the structure of the image to be in-painted in shape and motion. The selected reference

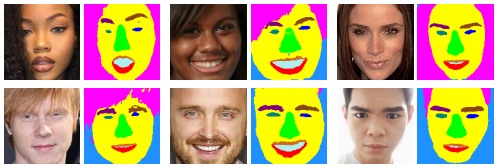

Figure 6. The segmentation result of face semantic parsing network testing on Celeb-ID Dataset. Different colors of pixels represent different face components.

can provide additional eye information for the generator, especially eye aesthetic features.

Secondly, in addition to providing aesthetic prior knowledge, we also need to constrain beauty in the training of the network. We use the eye feature extraction module of AesNet to extract the eye aesthetic features of the generated eyes and the reference. The aesthetic loss calculates the $\mathcal{L}2$ distance between these two features, making the generator learn the concept of eye aesthetic better.

Last but not least, we can use the eye aesthetic assessment network to compare our results with those of other advanced methods to verify the effectiveness of introducing eye aesthetics.

4.3. Face Semantic Embedding

Traditional GAN models independently generate facial components which may not be suitable for the original face. As mentioned in [7], if a part of the eyes is obscured, the new eyes take on strange shapes or have blurred eye details. Therefore, inspired by [16], we introduce a parsing network, which is implemented by changing the last layer of object contour detection network proposed in [27] to 11 outputs.

We train a segmentation model on the CelebA [17] dataset, which achieves the f-score of 0.822. Figure 6 shows the segmentation result of face semantic parsing network testing on Celeb-ID Dataset. The parsing result of the generated image is compared with that of the original image, and softmax cross-entropy loss is used as the parsing loss of the in-painting network to make the generated eye details more consistent with the overall coherence of the image.

4.4. Loss Functions

The global loss function of the network is defined as

$$\mathcal{L} = \mathcal{L}_{GAN} + \lambda_r \mathcal{L}_r + \lambda_p \mathcal{L}_p + \lambda_{aes} \mathcal{L}_{aes} \tag{4}$$

where $\mathcal{L}_{GAN}$ is the adversarial loss, and $\mathcal{L}_r$ is the reconstruction loss used in [7]. $\mathcal{L}_p$ is the parsing loss, which is the softmax cross-entropy loss. $\mathcal{L}_{aes}$ is the aesthetic loss, which is the activation of the residual blocks' final layer defined as

$$\mathcal{L}_{aes} = \|\mathcal{F}(g_i) - \mathcal{F}(e_i)\|_2 \tag{5}$$

where $\mathcal{F}(g_i)$ and $\mathcal{F}(e_i)$ are the eye feature layers of the generated image and the reference, respectively. By shortening the aesthetic distance between the generated eyes and the reference eyes, we make the generated eyes look more aesthetically pleasing. λ_r, λ_p and λ_{aes} are the weights to balance the effects of different losses.

5. Experiment

This section provides a detailed assessment of the eye aesthetic and its effectiveness for eye in-painting. Specifically, we first analyze the influence of different modules of AesNet on network performance. Then we conduct the ablation study to analyze the effectiveness of different designs of our AesGAN, including the different settings of loss functions and eye aesthetic assessment. We also demonstrate the experiment between *Single Example VS. Aesthetic Assessment Guided Eye In-painting* and *Multi Examples VS. Aesthetic Assessment Guided Eye In-painting*. Finally, we compare the latest and representative eye in-painting methods.

For the eye in-painting task, we use the Celeb-ID [7] dataset to train and test our model, which contains about 17k personal identities and a total of 100k photos. Each celebrity has at least 3 photos. We split the dataset according to the following criteria: for any celebrity, if there is a closed-eye photo in his samples, all his photos will be classified as the test set, otherwise classified as the training set. So every image in the training set contains a person with eyes opened, forcing the network to produce open eyes. Each training image has a reference of identity.

All experiments are conducted on a machine with an Nvidia GTX 1080Ti GPU with learning rate 1e-4. The parameters were optimized by ADAM [13] with parameters $\beta_1 = 0.5, \beta_2 = 0.999$. To balance the effects of different losses, we use $\lambda_r = 1$, $\lambda_p = 0.03$ and $\lambda_{aes} = 1$ in our experiments. Further results are shown in the supplementary file to provide a more detailed understanding of the performance advantages of our method.

5.1. Discussion of the AesNet modules

The traditional image quality assessment network is an end-to-end mode with only one classification branch. The network architecture is simple and intuitive, which can achieve good classification effect with a large number of training samples. Different from the existing quality assessment task, the eye aesthetic assessment is more subjective and self-employed. However, due to the small number of ocular aesthetic samples, the lack of sufficient learning knowledge in the network results in unstable performance. Thus we add the image reconstruction branch to assist the deep network to study the concept of eye aesthetic. The two branches common one eye aesthetic feature extraction module, and the eye reconstruction module helps the network to

Framework	Accuracy	Recall	Precision	F1
Baseline	0.7	**0.84**	0.656	0.737
Baseline+(a)	0.73	0.78	0.709	0.743
Baseline+(a)+(b)	**0.84**	0.76	**0.905**	**0.826**

Table 1. Comparison of AesNet performance under different modules. The baseline model consists of a encoder and the eye scoring module. Module(a) represents the residual blocks. Module(b) represents the eye reconstruction module.

Network	Ref-Select	$Mean\mathcal{L}_1^-$	$PSNR^+$	$MS-SSIM^+$	IS^+	FID^-
ExGAN	Random	7.15E-3	38.57dB	0.9344	3.56	15.66
Our baseline	SSIM	4.82E-3	42.56dB	0.9708	4.10	6.74
Our baseline+$\mathcal{L}_p$	SSIM	4.78E-3	42.57dB	0.9720	4.11	6.47
Our baseline	Aesthetic	4.76E-3	42.56dB	0.9720	4.04	6.99
Our baseline+$\mathcal{L}_p$	Aesthetic	4.71E-3	42.57dB	0.9728	4.09	6.55
Our baseline+$\mathcal{L}_{aes}$(a)	Aesthetic	4.74E-3	42.56dB	0.9722	4.03	6.57
Our baseline+$\mathcal{L}_{aes}$(b)	Aesthetic	4.70E-3	42.59dB	**0.9730**	4.08	6.78
Our baseline+$\mathcal{L}_p$+$\mathcal{L}_{aes}$(a)	Aesthetic	**4.67E-3**	**42.60dB**	0.9729	4.10	6.66
Our baseline+$\mathcal{L}_p$+$\mathcal{L}_{aes}$(b)	Aesthetic	4.68E-3	42.58dB	**0.9730**	**4.15**	**6.43**

Table 2. Quantitative results of AesGAN with different structures. The second column represents the way the network selects the reference. We use one-branch AesNet in $\mathcal{L}_{aes}$(a) and two-branch AesNet in $\mathcal{L}_{aes}$(b). $^-$Lower is better. $^+$Higher is better. IS means the inception score.

visualize the learned concept of eye aesthetic.

The comparison of AesNet performance under different modules is shown in Table 1. The baseline network is shallow with simple structure resulting in an undesirable accuracy. After 9 residual blocks are added, the accuracy is improved, and the network's judgment on positive and negative samples also tends to be balanced. After the reconstruction branch is added, the accuracy of the network is greatly improved, which proves the effectiveness of the module in aesthetic assessment.

5.2. Ablation Study on the Eye In-painting Network

Effectiveness of Eye Aesthetic Assessment. As described in section 3, we trained an eye aesthetic assessment network, and then used this mechanism to select the best reference. Because AesNet can only output two categories, when the references' eye quality all belong to the high grade, we choose the one closest to the input. We consider the pose, angle of the face by measuring the structural similarity (SSIM) between the reference and the input image, so that the eye aesthetic features can be given more to the generator. Figure 7 shows our algorithm of how to select the best reference and the corresponding in-painting results. Table 2 shows the different effects of network with different reference selecting metrics. SSIM can only select the face with similar position and pose without considering the eye aesthetics. By using the eye aesthetics as an index, the generator can learn the most suitable eyes, so as to improve the in-painting effect.

We then use the eye aesthetic assessment network to assess the generated images and find that our method does improve eye quality. Figure 8 compares the eye assessment

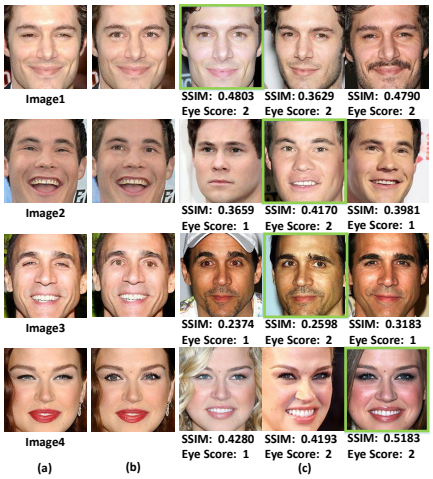

Figure 7. Our algorithm of how to select the best reference. First of all, the network assesses the eye aesthetic grade for each reference. In all references with a rating of 2, the algorithm calculates SSIM between the input and the references. Considering these two factors, we finally choose the most suitable reference map. Columns represent: (a) Original image, (b) AesGAN's in-painting results and (c) The selected best reference is marked with a green box.

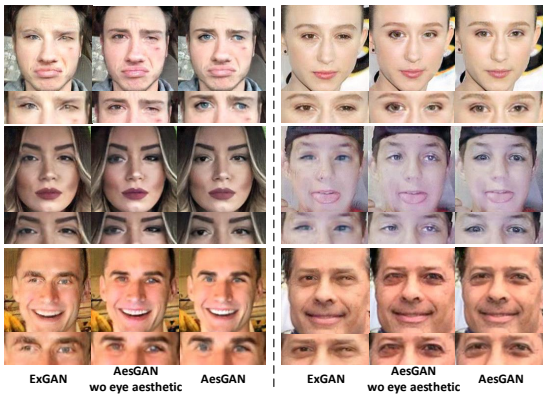

Figure 8. Comparison between ExGAN, AesGAN without aesthetic and AesGAN. Network with aesthetic constraints can produce more satisfactory results.

results. We find that **10.1**% of the test images changed from low quality to high quality with eye scoring constraints. Experimental results show that AesNet greatly improves the quality of eye in-painting, which also suggest that our network has a good generalization ability without domain shift problem.

Discussion of Loss Functions. In order to compare the influence of different network structures, we add several innovation points to the network separately. Table 2 shows the quantitative results of different network structures. The smaller $\mathcal{L}_1$ is and the larger SSIM is, the closer the generated image is to the input. This means that the in-paintings do not change the individual information of the original im-

Figure 9. Visual comparison of AesGAN with different structures. The second column are ExGAN's results. The third column are the results of our AesGAN without using aesthetic selection and aesthetic loss. The fourth column are the results of our AesGAN without using parsing loss. The last column are the results of our AesGAN with all modules. Better zoom in.

age. Larger value of PSNR means less distortion and inception score(IS) is a quota of image richness which higher is better.

From Table 2, we can find that after adding the parsing loss constraint into the network, the inception score of the network has been greatly improved. This shows that segmentation constraints can better preserve the individual's feature information. We use one-branch AesNet in AesGAN(a). After adding the eye aesthetic constraint, most of the quantitative results of the network reach the best. However, the inception score decreases slightly, which may be due to the aesthetic learning leading to a decline in the richness of eye samples. We can see that the inception score is significantly improved after using the two-branch AesNet in AesGAN(b). This also demonstrates the effectiveness of the proposed eye reconstruction module on maintaining sample features. As mentioned in [7], the FID score is closely related to perceived quality compared to several other metrics, which increases with the fuzziness of the image. We also measure the FID scores of the eye area, and the results show that our final model can achieve the best performance.

Figure 9 shows a comparison of the visual effects of different network structures. It can be observed that the network with parsing constraint and eye aesthetic constraint can effectively solve the limitations such as the eye detail blur and generate more realistic eyes.

5.3. Comparison with State-of-the-arts

Single Reference VS. Aesthetic Assessment Guided Eye In-Painting: ExGAN [7] selects a single reference randomly which may have some limitations. Figure 10 shows different in-painting results of the same photo with different references, which suggests that the eye quality of the reference greatly influences the final in-painting result. Aesthetic assessment guided eye in-painting method can effectively utilize the individual's original eye information and make the generated eye meet the requirements of eye quality in vision. The visual contrast results are shown in Figure 11.

Figure 10. Different in-painting results of the same photo with different references. Row (a) represent different references and row (b) represent the corresponding in-painting results with ExGAN.

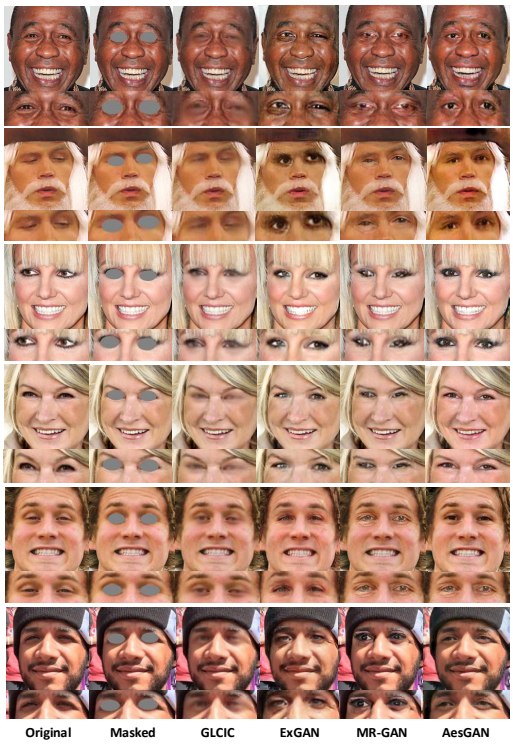

Figure 11. Comparison with state-of-the-arts. GLCIC [11] is a non-exemplar method. ExGAN [7] is the first exemplar based eye in-painting GAN. MR-GAN is the multi-reference model we trained. AesGAN is the eye aesthetic based eye in-painting method we proposed. Better zoom in.

Multi-Reference VS. Aesthetic Assessment Guided Eye In-Painting: Due to the limitations of single reference, a direct idea is to select multiple references. Since there is no existing multi-reference method, we trained a model called MR-GAN. The experimental results show that

Method	$Mean\mathcal{L}_1^-$	$PSNR^+$	$MS-SSIM^+$	IS^+	FID^-
GLCIC	7.36E-3	28.94dB	0.7261	3.72	15.30
ExGAN	7.15E-3	38.57dB	0.9344	3.56	15.66
MR-GAN	11.77E-3	38.37dB	0.9277	3.90	13.61
AesGAN	**4.68E-3**	**42.58dB**	**0.9730**	**4.15**	**6.43**

Table 3. Quantitative results of different methods. $^-$Lower is better. $^+$Higher is better. IS means the inception score.

although multi-reference can solve the problem of posture inadaptability, more references provide more in-painting directions, resulting in uncontrollable eye quality of the results. AesGAN can encourage the generator to learn the eye aesthetic features and produce better in-painting results as shown in Figure 11.

Qualitative and Quantitative Results: Figure 11 and Table 3 show the qualitative and quantitative results compared with the state-of-the-art methods. It is obvious that the exemplar-based methods have a better performance than the non-exemplar method. And our AesGAN can generate the most realistic and natural eyes. From the quantitative results, the numerical values of our method are much higher than those of state-of-the-art methods.

6. Discussion

6.1. Challenge Cases and Real-world Examples

It is mentioned in [7] that ExGAN can not deal well with occluded eyes and some iris colors of the new eyes may be inconsistent with the original image. As shown in Figure 12, AesGAN with parsing constraint and eye aesthetic constraint can address these limitations well. We also test closed-eye images of several celebrities from the Internet with a selected reference. These photos are taken in reality, without any pre-processing, matching the actually closed eyes situation. Figure 13 shows our test results, proving that our method can produce realistic eyes.

6.2. Limitation and Future Work

The effectiveness of eye aesthetic in improving the quality of eye in-painting is shown above. However, due to the small number of samples used for training, the performance of the eye aesthetic assessment network remains to be further improved. The experimental results show that the variousness of the network-generated eyes decreases after the addition of aesthetic loss. This may be due to the simple structure of the aesthetic assessment network, which leads to the singleness of the aesthetic feature extraction. We demonstrate in the experiment that the addition of an eye reconstruction branch to the aesthetic network is useful for increasing the sample diversity. However, due to the different learning efficiency of the high-level and lower-level features, the network performance is not stable. When the two-branch AesNet is used in the eye in-painting task, some metrics are reduced. Therefore, how to extract and apply

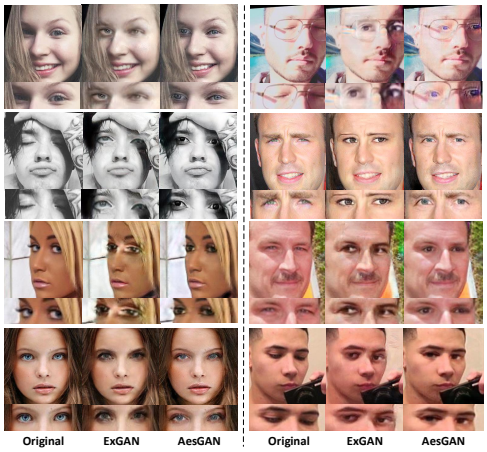

Figure 12. Improvement results to ExGAN's limitations. We can better handle occluded eyes and side faces, and also improve the color change of the iris mentioned in ExGAN.

Figure 13. The real-world test results.

the aesthetic information accurately still needs more exploration.

In fact, many image completion tasks are faced with the problem of selecting reference samples, and the selection of appropriate samples is of great significance for the final results. As an important image quality assessment index, aesthetic is worth paying more attention to in the future for selecting appropriate references and improving the quality of image restoration. This is our first attempt to prove the value of aesthetics. And how to edit the image aesthetically needs more research.

7. Conclusion

This paper shows the effectiveness of the eye aesthetic assessment for eye in-painting. This suggests that aesthetic assessment is of great value for image completion. In this paper, a new dataset is constructed to train the assessment network. A wide range of experimental results show that the proposed eye aesthetic assessment network greatly improves the quality of eye in-painting. The subjective effect and the objective quality have reached the state-of-the-art performance. In order to improve the quality of image completion, we look forward to more research on aesthetic assessment.

References

[1] Aseem Agarwala, Mira Dontcheva, Maneesh Agrawala, Steven Drucker, Alex Colburn, Brian Curless, David Salesin, and Michael Cohen. Interactive digital photomontage. In *ACM Transactions on Graphics (ToG)*, volume 23, pages 294–302. ACM, 2004.

[2] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. Patchmatch: A randomized correspondence algorithm for structural image editing. In *ACM Transactions on Graphics (ToG)*, volume 28, page 24. ACM, 2009.

[3] Subhabrata Bhattacharya, Rahul Sukthankar, and Mubarak Shah. A framework for photo-quality assessment and enhancement based on visual aesthetics. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 271–280. ACM, 2010.

[4] Matthew Brand and Patrick Pletscher. A conditional random field for automatic photo editing. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7. IEEE, 2008.

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[6] Yubin Deng, Chen Change Loy, and Xiaoou Tang. Image aesthetic assessment: An experimental survey. *IEEE Signal Processing Magazine*, 34(4):80–106, 2017.

[7] Brian Dolhansky and Cristian Canton Ferrer. Eye in-painting with exemplar generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7902–7911, 2018.

[8] R Edler, M Abd Rahim, D Wertheim, and D Greenhill. The use of facial anthropometrics in aesthetic assessment. *The Cleft palate-craniofacial journal*, 47(1):48–57, 2010.

[9] Wen Gao, Bo Cao, Shiguang Shan, Xilin Chen, Delong Zhou, Xiaohua Zhang, and Debin Zhao. The cas-peal large-scale chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(1):149–161, 2008.

[10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[11] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)*, 36(4):107, 2017.

[12] Wei Jiang, Alexander C Loui, and Cathleen Daniels Cerosaletti. Automatic aesthetic value assessment in photographic images. In *2010 IEEE International Conference on Multimedia and Expo*, pages 920–925. IEEE, 2010.

[13] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[14] Congcong Li and Tsuhan Chen. Aesthetic visual quality assessment of paintings. *IEEE Journal of selected topics in Signal Processing*, 3(2):236–252, 2009.

[15] Congcong Li, Andrew Gallagher, Alexander C Loui, and Tsuhan Chen. Aesthetic quality assessment of consumer photos with faces. In *2010 IEEE International Conference on Image Processing*, pages 3221–3224. IEEE, 2010.

[16] Yijun Li, Sifei Liu, Jimei Yang, and Ming-Hsuan Yang. Generative face completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3911–3919, 2017.

[17] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015.

[18] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[19] Naila Murray, Luca Marchesotti, and Florent Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2408–2415. IEEE, 2012.

[20] L Neumann, M Sbert, B Gooch, W Purgathofer, et al. Defining computational aesthetics. *Computational aesthetics in graphics, visualization and imaging*, pages 13–18, 2005.

[21] Ira D Papel. Quantitative facial aesthetic evaluation with computer imaging. *Facial Plastic Surgery*, 7(01):35–44, 1990.

[22] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.

[23] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. *ACM Transactions on graphics (TOG)*, 22(3):313–318, 2003.

[24] Katharina Schwarz, Patrick Wieschollek, and Hendrik PA Lensch. Will people like your image? learning the aesthetic space. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2048–2057. IEEE, 2018.

[25] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[26] Fei Yang, Jue Wang, Eli Shechtman, Lubomir Bourdev, and Dimitri Metaxas. Expression flow for 3d-aware face component transfer. *ACM Transactions on Graphics (TOG)*, 30(4):60, 2011.

[27] Jimei Yang, Brian Price, Scott Cohen, Honglak Lee, and Ming-Hsuan Yang. Object contour detection with a fully convolutional encoder-decoder network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 193–202, 2016.

[28] Raymond A Yeh, Chen Chen, Teck Yian Lim, Alexander G Schwing, Mark Hasegawa-Johnson, and Minh N Do. Semantic image inpainting with deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5485–5493, 2017.

[29] Seunghwan Yoo and Rae-Hong Park. Red-eye detection and correction using inpainting in digital photographs. *IEEE Transactions on Consumer Electronics*, 55(3):1006–1014, 2009.