IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 12, 2020, accepted January 23, 2020, date of publication January 29, 2020, date of current version February 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2970169

Image Inpainting Based on Generative Adversarial Networks

YI JIANG[1,3], JIAJIE XU[1], BAOQING YANG[1], JING XU[1], AND JUNWU ZHU[1,2]

[1]College of Information Engineering, Yangzhou University, Yangzhou 225009, China
[2]Department of Computer Science and Technology, University of Guelph, Guelph, ON N1G 2W1, Canada
[3]State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Corresponding authors: Jing Xu (xujing@yzu.edu.cn) and Junwu Zhu (jwzhu@yzu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61872313, in part by the Key Research Projects in Education Informatization in Jiangsu Province under Grant 20180012, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX18_2366, in part by the Yangzhou Science and Technology under Grant YZ2018209 and Grant YZ2019133, in part by the Yangzhou University Jiangdu High-End Equipment Engineering Technology Research Institute Open Project under Grant YDJD201707, and in part by the Open Project in the State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University under Grant 1907.

ABSTRACT Image inpainting aims to fill missing regions of a damaged image with plausibly synthesized content. Existing methods for image inpainting either fill the missing regions by borrowing information from surrounding areas or generating semantically coherent content from region context. They often produce ambiguous or semantically incoherent content when the missing region is large or with complex structures. In this paper, we present an approach for image inpainting. The completion model based on our proposed algorithm contains one generator, one global discriminator, and one local discriminator. The generator is responsible for inpainting the missing area, the global discriminator aims evaluating whether the repair result has global consistency, and the local discriminator is responsible for identifying whether the repair area is correct. The architecture of the generator is an auto-encoder. We use the skip-connection in the generator to improve the prediction power of the model. Also, we use Wasserstein GAN loss to ensure the stability of training. Experiments on CelebA dataset and LFW dataset demonstrate that our proposed model can deal with large-scale missing pixels and generate realistic completion results.

INDEX TERMS AutoEncoder, image inpainting, skip-connection, stable training, wasserstein GAN.

I. INTRODUCTION

Image inpainting [4] is a research field of image processing. It aims to fill the missing or masked regions of the image with generated content and make the repaired image visually realistic. Image inpainting technology has been widely applicated in many fields, including ancient book restoration, medical image processing, and PhotoShop processing. Therefore, the research of image inpainting is worth studying. Due to the complexity of the natural images, there will be obvious fuzzy phenomena in the region of the repaired image and the boundary between the original region and the repaired region, which is a main difficult issue in the work of image inpainting. Also, how to ensure the semantic correctness of the repaired region is one of the difficulties in the task of image inpainting. In order to address these problems of image inpainting, existing methods categorize into two types: one category is texture synthesis methods based on the patch, the main idea is to find the boundary of the missing region to fill in the missing part of the image. The other is methods based on the Convolution Neural Networks(CNNs) [20], the main idea is to extract the features of the image through the deep convolution neural network to understand the image, and then to fill the missing region.

A typical patch-based method is a Patch-Match method proposed by Barnes *et al.* [3], which searches for the matching patch from the rest part of the image to fill in the missing region, resulting in more reasonable texture information. This method has an excellent effect on background inpainting. However, it does not perform well in the face of complex images(face, natural images) inpainting, and the result of inpainting will be very vague. Similarly, other patch-based methods [10] [11] and exemplar-based methods [7] [31] [35] are weak in inpainting the missing regions with complex

The associate editor coordinating the review of this manuscript and approving it for publication was Huimin Lu.

structures. The reason is that the texture synthesis method based on the patch is still not enough to obtain the high-level characteristics of the image.

With the rapid development of deep learning, the appearance of the feature learning-based image inpainting method exactly fills the defect of the traditional image inpainting methods, which is lack of high-level coherence and difficult to deal with the problem of large areas or complex structures missing. Neural networks are more powerful to learn high-level semantic information of images and CNNs are effective tools for image processing [18]. Neural networks do more and more image inpainting tasks. Pathak *et al.* [30] proposed the model Context Encoder, which combines encoder-decoder and Generative Adversarial Network(GAN) [12] to train in an unsupervised method. They use the adversarial loss to make the repaired image as real as possible and generated realistic results. But context encoder has drawbacks: the fully connected layer cannot save accurate spatial information and context encoder sometimes creates blurry textures inconsistent with surrounding areas of the image. After this, Yang *et al.* [37] combined the idea of style transfer with the context encoder, proposed a new method to repair high-resolution images, but the model is not powerful enough to fill the missing region with complex structures. Similarly, Yeh *et al.* [39] use DCGAN [32] for image inpainting, which can successfully generate the missing parts of the image and fill. However, the blurring situation remains exists around the border. Iizuka *et al.* [16] propose a new generative model with one generator and two discriminators. Results show that this model can refine the details of inpainting. Similar to their work, Li *et al.* [23] also propose a generative model for face completion, which consists of a generator and two discriminators. The generator is an encoding-decoding architecture, and the image filled by this method looks more realistic and semantically coherent. However, there are shortcomings in this approach: the result is not very good when dealing with some unaligned faces, and the model does not fully explore the spatial dependencies between adjacent pixels. Kamyar et al. propose a new GAN-based image inpainting model EdgeConnect [29]. Unlike other GAN-based methods, EdgeConnect first generates the edge information of the image to be repaired, and then fills the color, so that the complement results in the edge will not appear fuzzy or distorted phenomenon. However, the EdgeConnect model sometimes fails to depict the edges of highly textured areas accurately, and when most areas of the image are missing, the results of the model complement become poorer. Chuanxia Zheng et al. combine Variational Auto-Encoders(VAEs) [17] with Long Short Term Memory(LSTM) [15] and proposed PICNet [43], which can generate multiple repaired results. In addition, many other methods [9] [26] [38] [39] [19] [33] can also get realistic results, improved the blurring of the repaired images.

Based on the work of Li *et al.* [23], we propose a new model for image inpainting. Combined with the characteristics of skip-connection [13] and Auto-Encoders, the model consists of an encoder-decoder as the generator to synthesize the missing regions from random noise, two adversarial discriminators judge whether the image generated by the generator is true or false. We add skip-connection in the generator, which can help us use the underlying network to enhance the prediction ability of the decoding process, and prevent the gradient vanishing caused by the deep neural network. Similar to the architecture proposed by Iizuka *et al.* [16], we use the architecture of dual discriminators: global discriminator and local discriminator. The difference is that we use Wasserstein GAN loss [2] to train our model, which can ensure our model's stable training. Also, Hua *et al.* [48] use Wasserstein GAN to do the task of image inpainting. Different from their method, we use the original WGAN loss to train our model instead of WGAN-GP [49]. Another difference is that we use l_2 norm loss function instead of l_1 norm loss function. We demonstrate that the model we proposed is capable of generating realistic and semantically coherent images when inpainting images.

In this paper, we make the following contributions

- Skip-connection [13] is used in the generator to strengthen the predictive ability of the generator and to prevent the gradient vanishing caused by the deep network. The result shows that the image completed by the Encoder-Decoder with skip-connection is more realistic.
- In order to solve the problem of training adversarial networks and improve the accuracy of completed regions, Wasserstein GAN loss [2] is used to train our model instead of original cross-entropy loss.

The rest of the paper is structured as follows: In Section II, related works including multiple image inpainting methods are discussed. This is followed in Section III by details about our methods. Section IV describes our experiments in details. Section V analyzes the results, conclusions and future works are given in Section VI.

II. RELATED WORK

Since this paper is based on the deep generation model for image inpainting, some of the deep learning techniques of the framework used in this paper are briefly below.

A. GENERATIVE ADVERSARIAL NETWORKS (GAN)

Generative Adversarial Networks (GAN) [12] is a method of training generation model proposed by Ian Goodfellow in 2014. The core idea is a zero-sum game. The GAN consists of two parts: a generator and a discriminator. The generator is used to fit the distribution of the training data. The discriminator is used to determine whether the input into the discriminator is the actual data of the training set or the data generated by the generator. In the generative adversarial network, a random noise z obeying the $P_z(z)$ distribution generates x (subject to the $P_G(x)$ distribution) via the generator and then inputs the generated result and the real data x (subject to the $P_{data}(x)$ distribution) into the Discriminator. The discriminator discriminates the input x is a real sample or a sample generated by the generator. The goal of the network is to achieve a Nash Equilibrium [28], making it impossible for the discriminator to distinguish whether the input is a real sample or a generated sample. The model trains the generator and the discriminator simultaneously by optimizing the loss function below:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[D(x)] - \mathbb{E}_{z\sim p_z(z)}[log(1 - D(G(z)))] \quad (1)$$

However, original GAN is challenging to train, and the loss of generator and discriminator could not indicate the training process. Wasserstein GAN [2] was proposed to solve these problems, and it works well. In this paper, we use the Wasserstein GAN loss for stable training.

At present, many research works on image inpainting have applied the idea of GAN [41] [9] [19] [23], and these methods have achieved better completion results. Also, adversarial game application is a hot topic and many researchers [45] [46] [40] [21] [22] [46] [44] [14] [24] [25] [8] [5] focus on this area.

B. SKIP-CONNECTION

Skip-connection is a technique proposed by Kaiming He in ResNet [13] to solve the problem of gradient vanishing. The traditional convolutional neural network model increases the depth of the network by stacking convolutional layers, thereby improving the recognition accuracy of the model. When the network level is increased to a certain number, the accuracy of the model will decrease because the neural network is back-propagating. The process needs to propagate the gradient continuously, and when the number of network layers is deepened, the gradient will gradually disappear, resulting in the inability to adjust the weight of the previous network layer. In order to solve this problem, Kaiming He et al. proposed the idea of taking shortcuts so that the gradient from the deep layer can be unimpededly propagated to the upper layer so that the shallow network layer parameters can be effectively trained. In this paper, we use the technique of skip-connection to improve the predictive ability of the generator and to enhance the quality of the repaired images.

C. AUTOENCODER

AutoEncoder [34] is an unsupervised learning method proposed by Junbo Zhao et al. The model consists of two parts: an encoder and a decoder. The input picture is encoded by the encoder to generate the code, and then input the code into the decoder to obtain the output. The purpose of the model is to make the picture entered into the encoder and the picture output from the decoder as similar as possible. In this paper, we add the skip-connection between the encoder and the decoder. The above [16] [23] is to use an auto-encoder architecture as the model's generative network for image inpainting. Similarly, we use the auto-encoder architecture as the generator in the model we proposed.

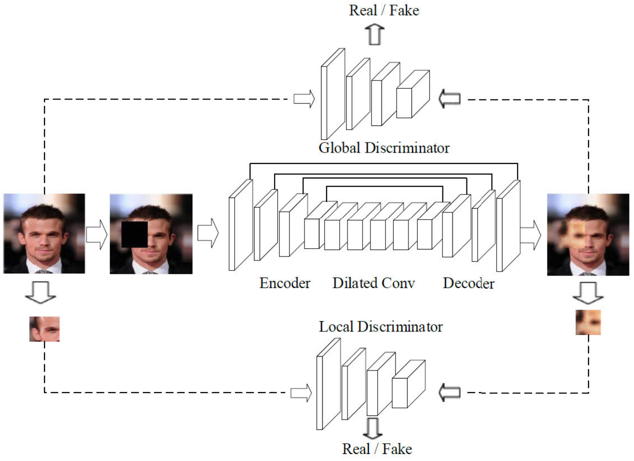


FIGURE 1. Network architecture.

III. METHODOLOGY

In this section, we first introduce the problem formulation and some basic symbols. Then we detail the framework proposed in this paper and the objective function. Finally, we propose the algorithm of training our model.

A. PROBLEM FORMULATION

The task of image inpainting is formalized as follows: Given an image x and partially mask it to get an incomplete input image x_m, with r representing the real region. We find a function f to produce an image $f(x_m)$. The purpose of image inpainting is to ensure that the generated image $f(x_m)$ as close as possible to the original complete image x. Formal representation is as follows:

$$f = \arg\min_f \| f(x_m) - x \|_2^2 \quad (2)$$

B. FRAMEWORK OVERVIEW

This paragraph details the model framework for image inpainting. Given an incomplete image, the purpose of our model is to fill the missing region of the image so that the entire image is visually and semantically plausibly realistic. Our model consists of a generator and double discriminators, as shown in FIGURE.1.

C. GENERATOR

Inpainting is part of a large set of image generation problems. To solve this problem, we use an auto-encoder as the generator of our model. The auto-encoder contains two networks: an encoder and a decoder. In this paper, we input the image to be repaired into the encoder and encode it into code, and then reconstruct and generate the repaired image via decoder decoding.

Different from the typical AutoEncoder architecture, we add skip-connection between the corresponding layers of the encoder and decoder sections to prevent the network layer from deteriorating due to the deepening of the network layers. Skip-connection can make sure that the decoding stage can utilize the output of the low-level coding stage of the corresponding resolution to supplement the decoder with part of the structural feature information lost during the encoder downsampling phase, and enhance the structure prediction capability of the generator.

In this paper, the encoder uses a multi-layer convolution layer architecture. Similarly, the architecture of the decoder is symmetric to the encoder with transposed convolution layers. Between the encoder and the decoder, we employ four layers of dilated convolution instead of fully connected layers.

D. DISCRIMINATOR

The generator is responsible for inpainting the missing or masked regions of the image. However, the generator can not guarantee that the generated regions are accurate or consistent with the original image. In order to ensure that the generated image is much more realistic, this paper uses the discriminator as a binary classifier to distinguish whether the image comes from real data distribution or generated by the generator. Also, the discriminator helps to improve the generator's ability to generate more realistic images to fool the discriminator.

We use two CNN architectures as the local discriminator and the global discriminator in the discriminative network. First, the local discriminator mainly identifies whether the result of the missing part is semantically accurate. For example, if the missing part is the nose, then the local discriminator needs to identify whether the completed part is the nose. The input of the local discriminator is the part of the original image that is missing or occluded, and the part generated by the generator. We input them into the local discriminator through channel splicing. Since we use the CNN architecture, the output of the local discriminator is a scalar, which indicates the generated region is true (from real data distribution) or false (generated by the generator).

However, only the local discriminator is not enough. Although the result is partially correct, the overall coherence is considered successful. So we use the global discriminator to identify the degree of coherence between the generated region and the original image. Similar to the local discriminator, the global discriminator's inputs are also in two categories: the original image as ground truth and the entire repaired image generated by the generator. Similarly, the channel stitches two pictures into the global discriminator, and the output is also a scalar, indicating the degree of trust of the global discriminator for the entire image after completion, and whether it is semantically coherent.

E. THE JOINT LOSS FUNCTION

Since this article uses one generator and two discriminators, we will analyze the loss function of each module.

- In this paper, we train the generator by minimizing the reconstruction loss L_r. Between the L_1 norm loss function and the L_2 norm loss function, the L_2 norm loss function is chosen in this paper. Because the L_2 norm penalizes the outliers, it is suitable for the inpainting tasks, but the disadvantage is that the robustness of L_2 norm is not strong enough. The reconstruction loss is defined as

$$L_r = \| G(x_m) - x \|_2^2 \tag{3}$$

- Adversarial loss is necessary for training GANs and becomes common in many creation tasks, and low adversarial loss means that the generator has stronger power to fill the holes. For stable training, we apply Wasserstein GAN loss [2] and use global and local discriminators. The Wasserstein GAN loss is defined as

$$\begin{aligned}\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[D(x)] \\ - \mathbb{E}_{z\sim p_z(z)}[D(G(z))]\end{aligned} \tag{4}$$

The global discriminate loss and local discriminate loss are defined as

$$L_{global} = \mathbb{E}_{x_c\sim p_g}[D_g(x_c)] - \mathbb{E}_{x\sim p_{data}}[D_g(x)] \tag{5}$$

$$L_{local} = \mathbb{E}_{m_c\sim p_g}[D_l(m_c)] - \mathbb{E}_{m\sim p_{data}}[D_l(r)] \tag{6}$$

where L_{global} and L_{local} represent the losses of the global discriminator and the local discriminator. D_g and D_l represent the function of the global discriminator and the local discriminator. x_c is the whole image with generated region and m_c is region generated by the generator. x and r are real image and region from real data distribution.

Overall, the total loss function is defined as follows:

$$L = L_r + \lambda_1 L_{local} + \lambda_2 L_{global} \tag{7}$$

where λ_1 and λ_2 are the weights that balance the effects of different losses.

F. TRAINING ALGORITHM

We propose an image inpainting algorithm for training in this paragraph. The mini-batch training method is used to occlude the image of the dataset in each iteration. Firstly, we sample a mini-batch of images x from training data and mask them with random holes. Then we get a mini-batch of masked images z, real regions before being masked r and masks m. $z = x \odot m$ where $\odot$ represents element-wise multiplication. Then we train the generator s times with L_r loss. After training the generator, we fix the generator and train discriminators t times with L_{global} and L_{local}. Finally, we train the joint model with joint loss L. Input z into the model and output the predicted images c. Combining the masked regions of c with z, we get the final inpainting images $x_i = z + c \odot (1-m)$. The training procedure can be represented in Algorithm 1.

IV. EXPERIMENTS

We use CNN architecture to implement our proposed method. Leaky Rectified Linear Unit(LeakyReLU) with α of 0.2 is employed as the activation function. In this paper, the values

Algorithm 1 Algorithm Training Model

```
while iterations $k < T_{train}$ do
    if $k < s$ then
        • Sample mini-batch of $m$ images $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$ from training set $p_{\mathrm{data}}(x)$.Noise occlusion processing for each image to generate a dataset of $m$ masked images $\{z^{(1)}, z^{(2)}, \ldots, z^{(m)}\}$ and a dataset of $m$ real regions before being masked $\{r^{(1)}, r^{(2)}, \ldots, r^{(m)}\}$..
        • Enter mini-batch of $m$ masked images $\{z^{(1)}, z^{(2)}, \ldots, z^{(m)}\}$ to the generator and obtain $m$ generated images $\{\tilde{x}^{(1)}, \tilde{x}^{(2)}, \ldots, \tilde{x}^{(m)}\}$, $\tilde{x}^{(i)} = G(z^{(i)})$. Update the generator with reconstruction loss (Eq.3) using $(G(z^i), x^i)$.
    else
        • Enter mini-batch of $m$ generated images $\{\tilde{x}^{(1)}, \tilde{x}^{(2)}, \ldots, \tilde{x}^{(m)}\}$ and $m$ real images $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$ to the global discriminator. Enter mini-batch of $m$ completed regions $\{c^{(1)}, c^{(2)}, \ldots, c^{(m)}\}$ and $m$ real regions $\{r^{(1)}, r^{(2)}, \ldots, r^{(m)}\}$ to the local discriminator.
        • Update local discriminator and global discriminator by Wasserstein GAN loss (Eq.5 and Eq.6 ).
        if $k > s + t$ then
            • Enter minibatch of $m$ masked images $\{z^{(1)}, z^{(2)}, \ldots, z^{(m)}\}$ to the generator and obtain $m$ generated images $\{\tilde{x}^{(1)}, \tilde{x}^{(2)}, \ldots, \tilde{x}^{(m)}\}$, $\tilde{x}^{(i)} = G(z^{(i)})$.
            • Discriminate generated images $\{\tilde{x}^{(1)}, \tilde{x}^{(2)}, \ldots, \tilde{x}^{(m)}\}$ by local discriminator and discriminate completed regions $\{c^{(1)}, c^{(2)}, \ldots, c^{(m)}\}$ by global discriminator. Update the joint model with joint loss(Eq. 7 ).
        end if
    end if
end while
```

of parameters λ_1 and λ_2 are set to 100. Parameters were updated using the Adadelta optimization algorithm [42]. Our implementation is with Python v3.5, TensorFlow v1.10.0 [1], Keras v2.2.4 [6], CUDNN v9.2 and CUDA v9.2. We use the same experimental settings to train and test our model.

A. DATASETS

We use CelebA dataset [27] and LFW dataset [47] to learn and to evaluate our model. CelebA dataset [27] contains 202599 RGB color facial image. We use 100k images for training and 1000 images for testing. The LFW dataset [47] contains 13233 facial images of 5749 individuals. We use 12k images for training and 1k images for testing. In training, we use images of resolution 256 × 256 with the largest hole size 128 × 128 and irregular masks in random positions.

B. POST PROCESSING

In this paper, we directly feed the generator image to be repaired, which will reconstruct the areas outside the repaired region. Although the generated image and the original image are consistent in content and structure, there are still small differences at the pixel level. So we only choose the filling part to fuse with the image to be repaired.

V. RESULTS

In order to analyze the effect of image restoration, the model is compared with GLCIC [16], Fast Marching Method (FMM) [3] and DIP [50].

A. VISUAL COMPARISONS

Comparisons with FMM, GLCIC and DIP inpainting. We compare our method with FMM [36] with random regular and irregular masks. Results show that FMM based methods cannot recover enough image details and generate blurry and noisy results. Then we compare our results with those obtained from the GLCIC [16] and DIP [50]. In the same experimental environment, we use the same number of CelebA [27] images and LFW [47] facial images to train GLCIC and our model separately. The results show that our method is slightly better than the GLCIC model in image authenticity and our model performs much better than DIP model.

FIGURE.2, FIGURE.3 and FIGURE.4 show comparisons of results obtained using our proposed method, FMM based method, GLCIC model and DIP model. Column (a) represents the ground truth images and column (b) is the masked images. Columns (c), (d) and (e) show the results generated by FMM [36], GLCIC [16] and DIP [50] respectively. The last column shows our results in the figure. The result shows

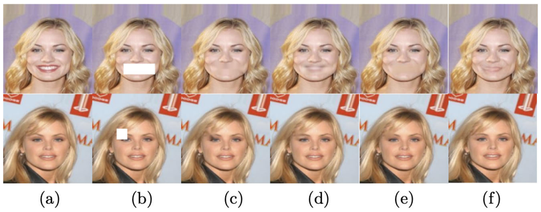


FIGURE 2. Comparisons of inpainting results on CelebA test dataset with regular missing region.

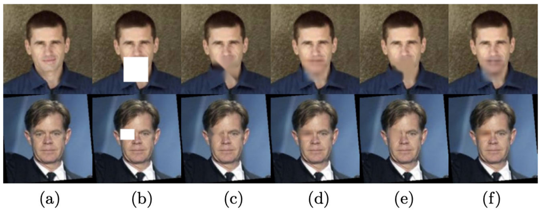


FIGURE 3. Comparisons of inpainting results on LFW test dataset with regular missing regions.

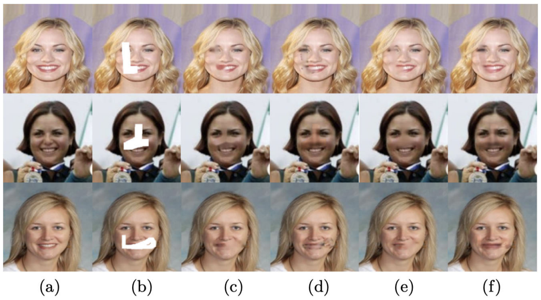


FIGURE 4. Comparisons of inpainting results on CelebA test dataset and LFW test dataset with irregular missing regions.

that the images generated by our model are more similar to the ground truth images compared to other methods.

Since there are some symmetrical contents on the face, we have done some experiments for it. As is shown in FIGURE. 4, the first row shows the result of image inpainting with one eye masked. The image of column (f) is generated by the model proposed in this paper. Our model can generate the semantic information of the missing region but failed to repair the texture information. In the first row of FIGURE.4, compared with FMM [36], GLCIC [16] and DIP [50], our model can generate more realistic results when one eye is masked with an irregular mask. FMM and DIP fail to generate the missing eye. In the second row of FIGURE.4, only our model and GLCIC succeed in filling the missing nose, while the other two models are unable to fill in the missing regions with correct semantic information. In the last row, our model remains to perform better than other models.

We also compare our model with the model without skip-connection, as is shown in FIGURE.5. We train our model and the model without skip-connection in the same environment.

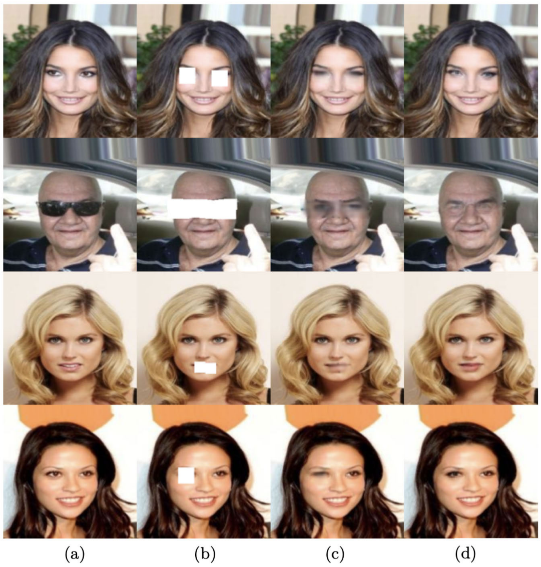


FIGURE 5. Comparisons of the model with skip-connection and the model without skip-connection in inpainting details.

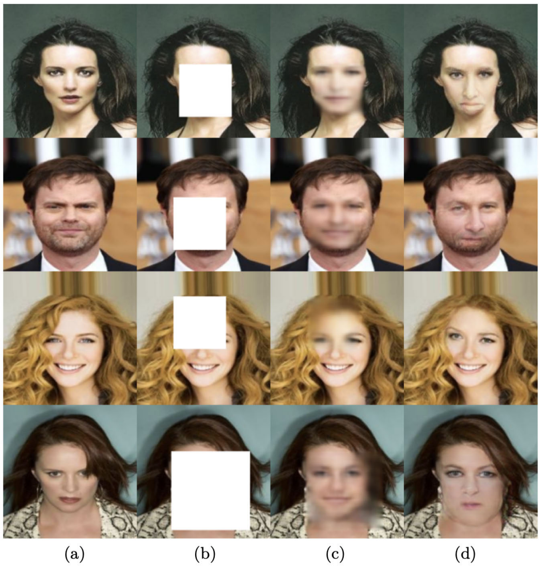


FIGURE 6. Results of filling big holes on CelebA dataset.

Column (a) represents the ground truth images and column (b) is the masked images. Column (c) and column (d) show the result generated by the model without skip-connection and our model respectively. In the first and last rows of the FIGURE.5, we cover the eyes. It can be seen from the results that our model can produce more realistic details than the model without skip-connection. It seems that our model can enhance the ability to generate details and structure information by skip-connection.

In addition, we also do experiments on filling big holes. As is shown in FIGURE.6, column (a) represents the ground truth images and column (b) shows the masked images with a big mask. Column (c) is the inpainting results of GLCIC model, and column (d) shows the results generated by our model. Since FMM(traditional method) [36] and DIP [50] are not good at filling big holes, FIGURE.6 only shows the comparisons of GLCIC model and our model. Although the results of our model repair may seem strange, our model is able to produce results consistent with the surrounding parts without significant blurring. All the experiments were done in the same environment.

B. QUANTITATIVE COMPARISON

In addition to the visual comparison, the experimental results in this paper are selected in comparison with GLCIC [16], Fast Marching Method(FMM) [36] and DIP [50] with PSNR and SSIM indexes.

1) COMPARISONS WITH PSNR

PSNR(Peak Signal to Noise Ratio) is the most common and widely used image objective evaluation index, which is based on the error between corresponding pixels, that is, based on error-sensitive image quality evaluation. The larger the PSNR, the closer the repaired image is to the ground truth.

TABLE 1. Quantitative results of different methods on CelebA test dataset and LFW test dataset.

	PSNR		SSIM	
	CelebA	LFW	CelebA	LFW
FMM	23.52	21.36	0.84	0.79
GLCIC	27.81	26.45	0.87	0.82
DIP	25.46	24.75	0.85	0.80
Ours	**27.90**	**26.92**	**0.88**	**0.84**

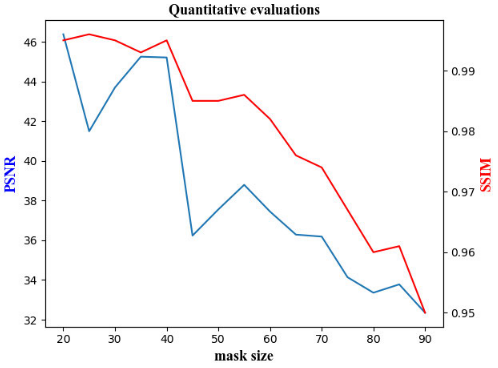


FIGURE 7. Quantitative evaluations in terms of PSNR and SSIM at different mask size.

We calculate PSNR by:

$$PSNR = 10 \cdot \log_{10} \frac{255^2}{\varepsilon} \tag{8}$$

where ε is the mean square error between the repaired image and the ground truth image.

2) COMPARISONS WITH SSIM

SSIM(Structural Similarity Index) evaluates the similarity of the two pictures as a whole from three aspects of brightness, contrast, and structure. The SSIM value range is [0, 1]. The larger the SSIM value, the higher the image similarity. We calculate SSIM by:

$$SSIM = \frac{(2\mu_X\mu_Y + C_1)\,(2\sigma_{XY} + C_2)}{\left(\mu_X^2 + \mu_Y^2 + C_1\right)\left(\sigma_X^2 + \sigma_Y^2 + C_2\right)} \tag{9}$$

where μ_X and μ_Y represent the average gray-scale value of image X and image Y. σ_X and σY are the standard deviations of image X and image Y, σ_{XY} is the covariance between image X and image Y. C_1 and C_2 are very small constants set to prevent the denominator from being zero.

We use the same training and test datasets and minimize mean squared error (MSE) in order to directly compare our results with the GLCIC model [16]. We use the same masks and the same testing images of CelebA dataset and LFW dataset to compute the values of index PSNR and SSIM. TABLE.1 shows the PSNR and SSIM results of our method and the other methods on CelebA test dataset and LFW test dataset. As shown in TABLE.1, our proposed method performs a little better than GLCIC [16].

We also evaluate the generalization ability by using different sizes of masks from 20 to 90 with stride 5. As is shown in FIGURE.7, our model performs better on small mask size inpainting. As the mask size increases, the performance of our model is getting worse, which is expected as the larger mask size indicates more structural information and semantic information. With a small mask size like 40, our model has good PSNR and SSIM indicators. It is because that small mask may only occlude part of a facial organ.

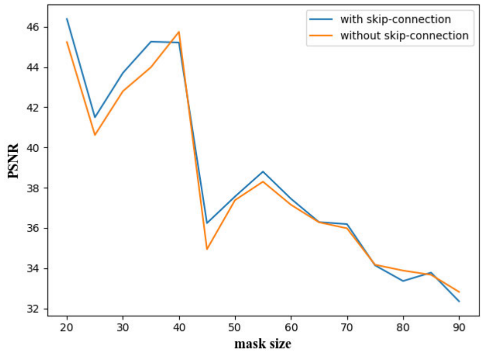


FIGURE 8. Quantitative evaluations in terms of PSNR of the model with skip-connection or not.

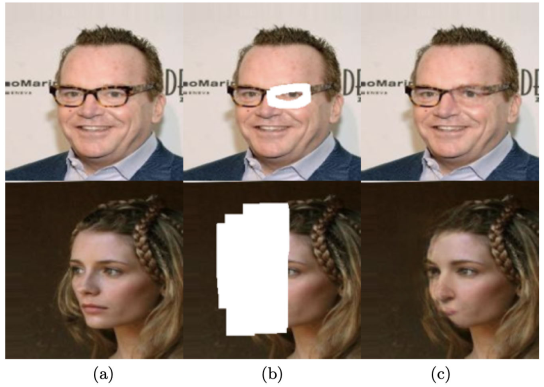


FIGURE 9. Repaired results where our model fails to generative relevant information.

Also, we compare the PSNR indicator of the model adding skip-connection or not. We train the model with skip-connection and the model without skip-connection in the same environment. As is shown in FIGURE.8, our model with skip-connection performs better than the model without skip-connection. It is because that skip-connection can help the generator learn more high-level semantic information.

C. FAILURE CASE

Although our model can generate some realistic inpainting results, it has some limitations. Our model is not powerful enough to generate relevant information. As is shown in the first row in FIGURE.9, when the missing part is a part of the glasses, our model cannot obtain information from the right half of the glasses, only from around the eyes to fill the missing region. Also, our model fails to generate missing regions of the images with unaligned faces. One failure inpainting result is shown in the second rows of FIGURE.9, when we mask a half part of the face, the result can not generate the right structure information of the face. This problem can be solved by data augmentation.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we propose a novel approach based on Wasserstein GAN for image inpainting. Skip-connection is added to the generative model of autoencoder architecture to enhance the prediction ability of the generated model. The experimental results demonstrate that the model of adversarial network proposed in this paper plays an important role in image inpainting. By comparing our model with GLCIC, FMM and DIP by visual effects, PSNR and SSIM, the results show that our proposed model can generate better and more realistic results.

In the future, we plan to extend our model to deal with the task of image inpainting with complex structure information missing and compare our model with more state-of-the-art methods. In addition, the issues proposed in the last paragraph of Section V will be addressed in our future work. Also, the inpainting framework we proposed can be applied to the task of image super-resolution and image denoising.

REFERENCES

[1] M. Abadi *et al.* (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: http://tensorflow.org/

[2] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: https://arxiv.org/abs/1701.07875

[3] C. Barnes, E. Shechtman, A. Finkelstein, and B. G. Dan, "PatchMatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–11, 2009.

[4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," *Siggraph*, vol. 4, no. 9, pp. 417–424, 2005.

[5] Q. Shen, P. Shi, J. Zhu, and L. Zhang, "Adaptive consensus control of leader-following systems with transmission nonlinearities," *Int. J. Control*, vol. 92, no. 2, pp. 317–328, Feb. 2019.

[6] F. Chollet. (2015). *Keras*. [Online]. Available: https://github.com/keras-team/keras

[7] A. Criminisi, P. Perez, and K. Toyama, "Object removal by exemplar-based inpainting," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Nov. 2003, pp. 1–8.

[8] X. Sun, B. Li, H. Leung, B. Li, and J. Zhu, "Static change impact analysis techniques: A comparative study," *J. Syst. Softw.* vol. 109, pp. 137–149, Nov. 2015.

[9] U. Demir and G. Unal, "Patch-based image inpainting with generative adversarial networks," 2018, *arXiv:1803.07422*. [Online]. Available: https://arxiv.org/abs/1803.07422

[10] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, pp. 1033–1038.

[11] F. Farahnakian, P. Liljeberg, and J. Plosila, "LiRCUP: Linear regression based CPU usage prediction algorithm for live migration of virtual machines in data centers," in *Proc. 39th Eur. Conf. Softw. Eng. Adv. Appl.*, Sep. 2013, pp. 357–364.

[12] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, X. Bing, and Y. Bengio, "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, 2014, pp. 2672–2680.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[14] H. Lu, B. Li, J. Zhu, Y. Li, Y. Li, X. Xu, L. He, X. Li, J. Li, and S. Serikawa, "Wound intensity correction and segmentation with convolutional neural networks," *Concurrency Comput., Pract. Exper.*, vol. 29, no. 6, p. e3927, Mar. 2017.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *TOGACM Trans. Graph.*, vol. 36, no. 4, pp. 1–14, Jul. 2017.

[17] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*. [Online]. Available: https://arxiv.org/abs/1312.6114

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[19] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5505–5514.

[20] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[21] J. Zhu, H. Song, Y. Jiang, and B. Li, "On truthful auction mechanisms for electricity allocation with multiple time slots," *Multimedia Tools Appl.*, vol. 77, no. 9, pp. 10753–10772, 2018.

[22] J. Zhu, H. Song, Y. Jiang, and B. Li, "On complex tasks scheduling scheme in cloud market based on coalition formation," *Comput. Elect. Eng.*, vol. 58, pp. 465–476, Feb. 2017.

[23] Y. Li, S. Liu, J. Yang, and M.-H. Yang, "Generative face completion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3911–3919.

[24] X. Zhang, B. Li, and J. Zhu, "A combinatorial auction-based collaborative cloud services platform," *Tsinghua Sci. Technol.*, vol. 20, no. 1, pp. 50–61 2015.

[25] J. Zhu, J. Wang, and B. Li, "A formal method for integrating distributed ontologies and reducing the redundant relations," *Kybernetes*, vol. 38, no. 10, pp. 1870–1879, Oct. 2009.

[26] G. Liu, F. A. Reda, K. J. Shih, T. C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 85–100.

[27] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3730–3738.

[28] J. Nash, "Non-cooperative games," *Ann. Math.*, vol. 54, no. 2, pp. 286–295, 1951.

[29] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, "EdgeConnect: Generative image inpainting with adversarial edge learning," 2019, *arXiv:1901.00212*. [Online]. Available: https://arxiv.org/abs/1901.00212

[30] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2536–2544.

[31] Z. Qiang, L. He, and X. Dan, "Exemplar-based pixel by pixel inpainting based on patch shift," in *Proc. CCF Chin. Conf. Comput. Vis.*, 2017, pp. 370–382.

[32] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: https://arxiv.org/abs/1511.06434

[33] R. Gao and K. Grauman, "On-demand learning for deep image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1086–1095.

[34] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Cognit. Model.*, vol. 5, no. 3, p. 1, 1988.

[35] X. N. Tang, J. G. Chen, C. M. Shen, and G. X. Zhang, "Modified exemplar based image inpainting algorithm," *J. East China Normal Univ.*, vol. 135, no. 6, pp. 24–28, 2009.

[36] A. Telea, "An image inpainting technique based on the fast marching method," *J. Graph. Tools*, vol. 9, no. 1, pp. 23–34, Jan. 2004.

[37] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, "High-resolution image inpainting using multi-scale neural patch synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6721–6729.

[38] G. Liu, F. A. Reda, K. J. Shih, T. C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 85–100.

[39] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2016, pp. 5485–5493.

[40] J. Zhu, L. Teng, Z. Zhu, and H. Lu, "A pricing method of online group-buying for continuous price function," *Neural Comput. Appl.*, pp. 1–9, Jan. 2019, doi: 10.1007/s00521-019-04017-y.

[41] J. Yu, L. Zhe, J. Yang, X. Shen, L. Xin, and T. S. Huang, "Free-form image inpainting with gated convolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2018, pp. 4471–4480.

[42] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," 2012, *arXiv:1212.5701*. [Online]. Available: https://arxiv.org/abs/1212.5701

[43] C. Zheng, T. J. Cham, and J. Cai, "Pluralistic image completion," 2019, *arXiv:1903.04227*. [Online]. Available: https://arxiv.org/abs/1903.04227

[44] J. Zhu, H. Song, Y. Jiang, B. Li, and J. Wang, "On cloud resources consumption shifting scheme for two different geographic areas," *IEEE Syst. J.*, vol. 11, no. 4, pp. 2708–2717, Dec. 2017.

[45] J. Shi, Z. Yang, and J. Zhu, "An auction-based rescue task allocation approach for heterogeneous multi-robot system," *Multimedia Tools Appl.*, pp. 1–10, Dec. 2018, doi: 10.1007/s11042-018-7080-4.

[46] L. Teng, J. Zhu, B. Li, and Y. Jiang, "A voting aggregation algorithm for optimal social satisfaction," *Mobile Netw. Appl.*, vol. 23, no. 2, pp. 344–351, Apr. 2018.

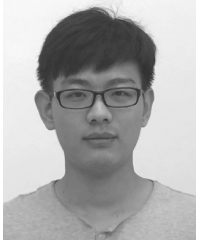
[47] G. B. Huang and E. Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep. 14–003, 2014.

[48] P. Hua, X. Liu, M. Liu, L. Dong, M. Hui, and Y. Zhao, "Image inpainting using wasserstein generative adversarial network," *Proc. SPIE*, vol. 10751, Sep. 2018, Art. no. 107510T.

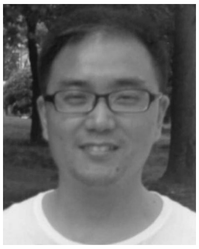
[49] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.

[50] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," 2017, *arXiv:1711.10925*. [Online]. Available: https://arxiv.org/abs/1711.10925

YI JIANG received the M.Sc. degree in management science and engineering from the Jiangsu University of Science and Technology, in 2005. She is currently an Associate Professor with the School of Information Engineering, Yangzhou University. Her research interests include ontology, management information systems, and artificial intelligence. She has published more than 20 articles in the above areas.



JIAJIE XU received the B.S. degree from Yangzhou University. He is currently pursuing the master's degree with the School of Information Engineering, Yangzhou University. His research interests include algorithmic game theory, game description language, and image inpainting.



BAOQING YANG received the Ph.D. degree from the Department of Automation, Shanghai Jiao Tong University, China. He is currently a Teacher with the School of Information Engineering, Yangzhou University. His major research interests include visual surveillance, face recognition, and pattern analysis.

JING XU received the M.Sc. degree in computer education from Yangzhou University, in 2004. She is currently an Associate Professor with the School of Information Engineering, Yangzhou University. Her research interests include management information systems, on-line education, and artificial intelligence. She has published more than ten articles in the above areas.

JUNWU ZHU received the Ph.D. degree in computer science from the Nanjing University of Aeronautics and Astronautics, in 2008. He is currently a Professor with the School of Information Engineering, Yangzhou University, and also a Visiting Professor with the School of Computer Science, University of Guelph, Canada. His research interests include knowledge engineering, ontology, mechanism design, and cloud computing.