Face Image Inpainting Based on Generative Adversarial Network

Xinyi Gao, Minh Nguyen, Wei Qi Yan

Auckland University of Technology
Auckland 1010 New Zealand

Abstract—**Face image inpainting is essential in the fields such as protection and preservation of human images with patterns. Using image inpainting to remove facial masks on human faces is one of the challenging tasks. In this paper, we propose a face image inpainting method based on an adversarial neural network. In general, face image inpainting is composed of generators and discriminators in deep nets. The loss function combines the losses from Mean Square Error (MSE) and Generative Adversarial Networks (GANs). In this paper, we have designed and implemented a new model for face image inpainting with up to half of the given image (50% of the area). The average of the evaluation metrics PSNR and SSIM are 31.86dB and 0.89, respectively. We improved image inpainting with a new model that is much suitable for face images.**

Keywords—Generative adversarial network, face image inpainting, convolutional neural network

I. Introduction

Image inpainting refers to the process of supplementing the missed parts of digital images or videos by using photography or image coherence [1]. The inpainted image should be natural, has fewer differences from the original image. If the inpainted image does not achieve this outcome, the repaired part looks weird, which also means that it is not a perfect image reconstruction. Therefore, high-quality image inpainting requires that the missed parts of this image must be seamlessly connected to the connected regions. The textures of the inpainted image should not have artefacts.

The early image inpainting methods met challenges due to a lack of training samples which resulted in image inpainting not being widespread. Previously, image inpainting was only used for recovering files or printed images. With the continuous development of deep learning and data science, image inpainting, at present, has made significant progress. The current image inpainting has taken benefits from deep learning and big visual data. The image inpainting methods based on deep learning and big data uplift the quality of inpainted images [2]. Using the state-of-the-art methods to generate the missed parts of the given images, the reconstructed images often show a decent appearance. The inpainted textures are also kept clear locally. The inpainting methods based on deep learning include GANs, Convolutional Neural Networks (CNNs) [3]. These image inpainting methods often do not require manual intervention. Deep neural networks are repeatedly trained and optimized through a large-scale image dataset. If the image needs to be inpainted, the image is only needed to be input into the deep train net. The inpainted image will be generated automatically.

The image inpainting methods have been broadly developed and benefited from the latest development of deep learning and big data. For example, we mark and recover the cracks, scratches, and blemish regions on old and damaged photos due to various reasons; we are able to erase the visiable watermark on the photos or digital videos; we remove the unwanted content from the images and fill it up with semantic content [4]. In this paper, we will focus on the inpainting methods for human face images.

In this paper, we will mainly use CNNs to construct a complete network for image inpainting. In addition, a GAN discriminator will be constructed for evaluating the quality of inpainted images. These two parts will constitute a complete image inpainting net structure [5]. In this project, we mainly conduct our work based on human faces as a pattern. The novelty of this paper is to develop a new method for human face inpainting.

CNN is a deep learning algorithm, which is distinct from conventional neural networks [6]. It comprises an input layer, convolutional layer, pooling layer, and fully connected layer [7]. These layers construct a complete CNN model. In the design and implementation of GAN discriminator, a local discriminator and a global discriminator are applied to ensure that the inpainted part of the generated image conforms to the overall appearance of the given image, the reconstructed part is natural, and the texture is smooth and clear [8]. These two discriminators will discriminate whether the image is inpainted with the completed network or has any differences with the actual image. Combining the results of these two discriminators will result in a much accurate output of image artefact removal.

Due to the impact of the global epidemic in recent years, facial masks are usually worn to protect ourselves if we need to go out. However, this also brings in new problems, such as using mobile phones in public places, using face recognition to unlock their phones, and using face recognition for auto-payment, etc. Thus, we need a way to unlock our phones with facial masks for face recognition [9]. Therefore, a fast and accurate face image inpainting method is needed.

Fig. 1. Unlocking iPhone with face mask on is not possible

In the remaining part of this paper, we introduce the existing image inpainting methods in Section II. In Section III,

978-1-6654-0645-1/21/$31.00 ©2021 IEEE

we address our method and computational algorithms. In Section IV, we analyze the experimental results generated. We summarize our work and draw the conclusion in Section V.

II. Literature Review

The emergence of deep learning and big data in computer science is the key to developing new image inpainting methods. Deep learning is one part of machine learning [5]. Since its appearance, it has been used in computer vision, natural language processing, image analysis, and a lot of fields [10]. Image inpainting is the tip of iceberg for deep learning applications.

Nowadays, image inpainting is used in personal images or damaged videos. A plenty of precious but accidentally damaged photos and videos are able to be recovered. In photography, image inpainting was applied to remove visual watermarks or logos as well as transparent and obstacles or tiny details. In the field of public safety and media security, image inpainting methods thus play an essential role. In recent years, affected by epidemics, people often wear facial masks while travelling. This makes them hard to be recognized. Thus, image inpainting methods are able to solve this problem. Image inpainting can completely reconstruct the occluded part, treat it as a missed object and compare it with the existing ones.

Current image inpainting methods are roughly grouped into 2-fold: Traditional methods and deep learning methods. Sample-based texture synthesis methods, example-based structure synthesis methods, diffusion-based methods, sparse representation methods, and hybrid methods are all traditional image inpainting methods [11,25]. Deep learning methods mainly encapsulate CNNs and GANs. Most of the current image inpainting methods have been transitioned from only using CNN models alone to the combination of CNNs and GNNs together. CNN models have been employed as a feature extractor, GANs are employed to enhance the quality of the given images [12].

Super-resolution methods have been recommended as a solution to deal with low-resolution problems. A denoising autoencoder trained by using a deep neural network is able to deal with image denoising tasks. The use of stacked sparse denoising autoencoders often requires a consortium of supervised learning work, which often only tackle images with small-size objects for denoising purposes [11].

Although there have been many examples of GANs applied to digital images before, the proper use of GANS in image inpainting was introduced in 2016. A CNN, called context encoder, was proffered; After training the network, two different methods, the standard pixel reconstruction method and the additional GAN, were reconstructed. The results show that the latter yields better performance. In addition, this context encoder was also proffered to the task of semantic repairing. This connection layer is consistent with the standard connection layer. By using this connection layer, the neural network is able to obtain a sufficient understanding of the internal relationship between all features, making the network training much comprehensive [13].

Modern CNNs are combined to create a new shift connection layer based on the advantages of the traditional copy-and-paste method. This net architecture of the two new methods that are different from the previous ones. The characteristics are applied to estimate the missed part, which makes the filled regions much reasonable. The second method is to propose a use of the shift-connected connection layer. By using this new connection layer, the details of transparent and local textures were reconstructed [14].

A more powerful deep learning mode Shift-Net was proposed, which take advantage of the idea of copying and pasting with a contextual attention layer. This connection layer is not much different from the previously displaced connection layer. It compares the difference between the visual features of missed parts and the features of non-missing parts. The features of the non-missed parts are used to supply the missed parts [15].

A simple convolutional autoencoder (CAE) architecture was propounded by using the primary network components, namely, the convolutional and skip connection layers. Nevertheless, the performance of such a straightforward architecture surpasses those of nets that take use of adversarial training. The most critical point of this is to utilize evolutionary algorithms to search for suitable frames. In addition, Adam optimization was employed to minimize the loss between the restored image and the original image. This is the reason why the simple architecture is better than advanced methods [16].

The idea of using multiple image inpainting methods was put forward. According to this idea, innovative usage of a framework with two basic frameworks and the same probability distribution, one is the reconstruction framework, and the other is the generating framework. Both of them ultilize GAN as the basis net. In addition, the overall layout of the reconstructed images is improved by introducing a new LSTM attention layer. This new idea provides a variety of solutions. The experimental results show that quality of the output images is able to be guaranteed in the reconstructed images having large missed regions [17].

The image reconstruction methods, such as GAN, have significant chromatic aberrations after the artefacts are removed. Therefore, postprocessing is needed to reduce the occurrence of these artefacts. However, due to the high cost of postprocessing, the idea of using context-content convolutions instead of postprocessing is problematic. The core idea of this work is to use a standard convolutional layer plus a sigmoid activation function to update the mask. In addition, in order to improve training stability, spectral normalization (SN) was proposed in convolutional layer of the discriminator. The results show that after the gated convolutions, the quality of reconstructed images will not deteriorate with the growth of missing regions [18].

III. Our Methodology

In this paper, we take advantage of CNNs in deep learning as the basis net, the weighted mean squared error (MSE) loss and the GAN loss functions are combined to improve the stability of model training, the complete network of the entire structure of convolutional nets is employed for the image inpainting [5]. In addition, global and local discriminator nets are also put forward to improve the quality of image inpainting. In fact, the ultimate goal of complete network training is to fool judgment of the discriminator network and let the identification network output the actual image. The image is not the one reconstructed by using the completion network. Therefore, it is able to achieve this goal by training all nets simultaneously during training time, thereby uplifting the quality of final output images.

Our overall idea is different from the design proposed by Satoshi in 2017. Firstly, the dataset used in this paper is different. The image resolution in the dataset for image reconstruction is relatively small, namely, 178 × 218. This further leads to a different design of our discriminator. In addition, there are also differences in our algorithms, which will be explained in subsequent sections.

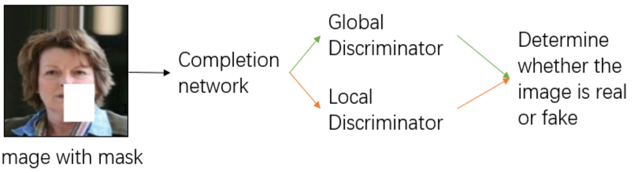


Fig. 2. Network training process. The whole network is composed of the complete network with the global discriminator and the local discriminator. These two discriminators are trained to determine the inpainted image. The completion network is trained fully for these two discriminators.

A. Completion Network

In our inpainting algorithm, convolutional layer is the core part of constructing the entire CNN network. Most of the calculations are generated by using the convolutional layers. The convolutional layers are essential in a filter; its primary role is to extract visual features from the input image after convolution operations and pooling operations. The parameters in the convolutional layer encapsulate the size of convolution kernel, net depth, and convolution step length, etc. Among them, the convolution kernels often determines the number of neurons corresponding to the convolution layer.

In addition, each neuron in the convolution layer will select a local region in the input data to corresponding neuron connections. The spatial resolution of this connection is called the receptive field of this net. In reality, this connection is always closely related to the net depth. The input of the network is often equal to the connections in the depth direction. The length of convolution steps determines the distance each time the filter sweeps the feature map [19]. For example, if the step length is 1, the filter will scan every element of the feature map. If the step length is at 2, the filter will skip one element after scanning the elements of the feature map. That is to say, if the step length is *n*, it will skip *n*-1 elements in the following operations.

The pooling layer is also an essential component in general CNNs. There are only two types of pooling layers: Max pooling and average pooling [20]. Generally speaking, the use of pooling layers is more often than that of the convolutional layers. If CNN has consecutive convolutional layers, it will insert a pooling layer. The role of pooling layers should not be underestimated. Using the pooling layers can assist CNNs to control the size of convolutional features, gradually reduce the size of the data volume, thereby taking over the number of parameters in the network and cutting down the number of calculations. It reduces the computational resources required by using dimensionality reduction methods. This avoids overfitting. In addition, it also benefits the network extract essential features. Consequently, the model training process is very effective.

Although the role of pooling layers is crucial, the CNN in this paper does not intend to use too many pooling layers. We will use dilated convolution to replace the role of pooling operations. In addition, we have to use deconvolution in the completion part. As a convolution method, literally, dilated convolution adds holes in the standard convolution process to increase the receptive field and reduce the number of calculations. In practice, dilated convolution is broadly applied to visual object detection. Deconvolution is also called transposed convolution. This convolution method is the inverse process of convolution operations [5].

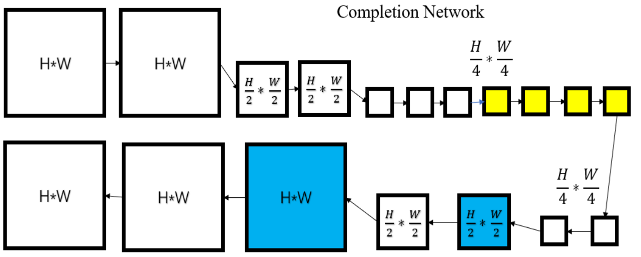

Completion Network
H*W
H*W
$\frac{H}{2}*\frac{W}{2}$
$\frac{H}{2}*\frac{W}{2}$
$\frac{H}{4}*\frac{W}{4}$
H*W
H*W
H*W
$\frac{H}{2}*\frac{W}{2}$
$\frac{H}{2}*\frac{W}{2}$
$\frac{H}{4}*\frac{W}{4}$

Fig. 3. The completion network consists of 18 layers of convolutional neural networks in total. It includes 12 layers of convolution operations, four layers of dilated convolution, and two layers of deconvolution, where white blocks indicate convolution layers, yellow rectangles show dilated layers, blue squares reflect the deconvolution layers.

TABLE I. THE COMPLETION NETWORK

Convolution types	Parameters		
	Kernels	*Dilations*	*Outputs*
Convolution1	5×5	None	64
Convolution2	3×3	None	64
Convolution3	3×3	None	128
Convolution4	3×3	None	128
Convolution5	3×3	None	256
Convolution6	3×3	None	256
Convolution7	3×3	None	256
Dilated1	3×3	2	256
Dilated2	3×3	4	256
Dilated3	3×3	8	256
Dilated4	3×3	16	256
Convolution8	3×3	None	256
Convolution9	3×3	None	256
Deconvolution1	4×4	None	128
Convolution10	3×3	None	128
Deconvolution2	4×4	None	64
Convolution11	3×3	None	32
Convolution12	3×3	None	3

TABLE II. THE LOCAL DISCRIMINATOR

Convolution types	Parameters		
	Kernels	*Dilations*	*Outputs*
Convolution1	5×5	None	64
Convolution2	5×5	None	128
Convolution3	5×5	None	256
Convolution4	5×5	None	512

TABLE III. GLOBAL DISCRIMINATOR

Convolution types	Parameters		
	Kernels	*Dilations*	*Outputs*
Convolution1	5×5	None	64
Convolution2	5×5	None	128
Convolution3	5×5	None	256
Convolution4	5×5	None	512
Convolution5	5×5	None	512

B. Discriminator Network

The discriminator network is composed of a global discriminant network and a local discriminant network. The purpose is to determine whether the image that has completed the network training is good enough. The foundation of these two networks is still CNN.

The global discriminant network consists of 5 convolutional layers. The stride of each convolution is 2×2. Meanwhile, the local discriminant network is composed of 4 convolutional layers. The stride of each convolution is also 2×2. Pertaining to settings of the local discriminator and the global discriminator, we reduce one convolutional layer each. Because of images in the dataset with a resolution 178×218, there is no requirement for too many convolutions. One of the two discriminators is working on the local region; the other is on the global region. Both are to ensure the quality of the exported images. Finally, the output of two discriminators will be combined by using a fully connected layer.

C. Algorithms

The training process is the core work for improving deep network performance. In our experiments, the loss training from Iizuka's net is still adopted. Due to the need to train the three nets simultaneously, the stability of model training is required. It is essential to ensure the nets work well. In the training process, we choose the weighted MSE to ensure the stability of the model training. The total loss function takes use of a combination of MSE loss and GAN loss to produce better training outcomes.

MSE Loss

The MSE has been efficiently applied to a loss function. This function is always employed to calculate the difference between the trained model and the actual one. The calculation process takes the average of squared errors between the predicted values and the actual values [21]. The MSE loss function is shown as eq.(1).

$$L_{mse} = ||M_i \odot (C(I, M_i) - I)||2 \qquad (1)$$

where $C(I, M_i)$ represents the completion of the network, I is the input image. $\odot$ is the pixel-wise multiplication [22]. M_i is a mask image with the exact size of the input image. The value is '1' in the regions that needs to be inpainted, and '0' in those that do not need to be inpainted. The smaller the value of this loss function, the more realistic the model is. After each round of model training is fulfilled in the complete network, we have to evaluate whether the model is working well by using this loss function. As the number of net training increases, the output of the loss function becomes smaller and smaller, which means that the performance of the proposed network is steadily converged. This is the reason why we chose this loss function in the generating network of GANs.

GAN Loss

In GAN, another loss function was employed in our experiments. The basic idea of this loss function is determined by GANs. The entire GAN is related to two networks. One of them is a generator network; the other is a discriminator network [23]. Both networks are playing a game following the min-max decision rule and use mathematical expectation as the loss output. The function is also called min-max loss function, as shown in eq.(2). The function $D(I, M_d)$ represents the discriminator network, M_d is a randomly generated mask, $C(I, M_i)$ refers to the generator network, I is the input image.

$$L_{GAN} = \min_{C} \max_{D} \mathbb{E}\left[\log D(I, M_d) + \log\left(1 - D(C(I, M_i), M_i)\right)\right] \quad (2)$$

Joint Loss

The joint loss is further subdivided into discriminator loss and generator loss, which is determined by the structure of GAN as shown in eq.(3). The generator makes the value of this function as small as possible, while the discriminator boosts the value as significant as possible. It is worth noting that this function is not used alone in this experiment which is only one part of a total loss function. The total loss is based on a combination of MSE loss L_{mse} and GAN loss so as to obtain better training results, where α refers to weight of our nets.

$$L_{joint} = \min_{C} \max_{D} \mathbb{E}\left[L_{mse} + \log D(I, M_d) + \alpha \log\left(1 - D(C(I, M_i), M_i)\right)\right] (3)$$

Although the same algorithm is supplied for neural network training, the training steps in this paper are quite different from previous work. In the first step, we do not change the use of MSE loss pair alone till we complete the network training; we no longer offer GAN to train the discriminant network but directly combine the two parts for training in the second step. Our training method has a fast speed and aims at face image inpainting accurately and precisely.

IV. OUR RESULTS AND DISCUSSION

We make use of 202,599 face images from the CelebA dataset to train our proposed model. The CelebA dataset includes 2,000 images in total [24]. In this project, 1,000 images in the dataset were randomly selected and trained 500 times. Firstly, we resize the images to the resolution 128×128. While testing image restoration, we randomly select a face image from the test dataset as the input. We randomly add a binary mask image to it. The size of randomly generated masks ranges from 24×24 to 48×48. Then, we make use of two image quality metrics: PSNR and SSIM, which are usually offered to evaluate the quality of the inpainted images. PSNR is an evaluation of the human perception of reconstruction quality. PSNR is often applied to measure the reconstruction quality of lossy compression codecs, which it is also accommodated to measure the quality of image inpainting. Generally, a higher PSNR indicates a better quality of image inpainting. SSIM is a model based on perception. It is employed to measure the similarity between two images. SSIM ranges from 0 to 1.00, 1.00 means that the inpainted image perfectly matches the original image.

It is worth mentioning that the input image in the dataset is small. Compared with Iizuka's net, we added a layer of the convolutional network in the completion net, and reduced the convolution operations of each discriminator network by using one layer. Therefore, the time required for training is shortened. Only 700 epochs are needed to obtain a well-trained net from the training images. The four groups of images in Fig. 4 show the outputs of face image inpainting by using a well-trained algorithm: Input image, output image, and the ground truth. The four sets of images in Fig. 5 are the results obtained by using the same dataset. As a comparison, the same dataset was applied to conduct 700 epochs and the results were obtained as shown in Fig. 6.

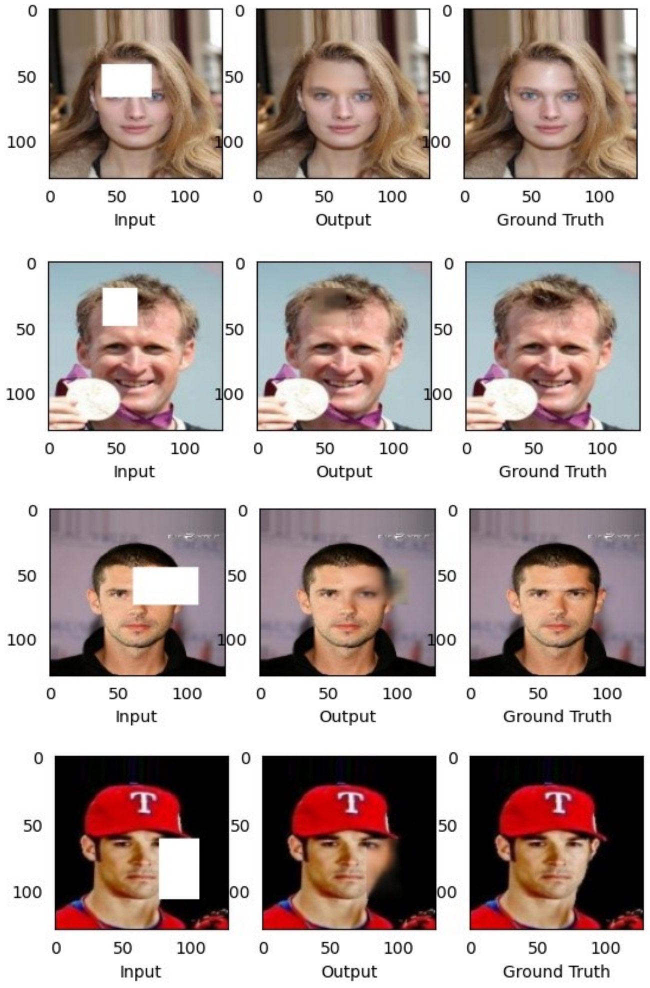
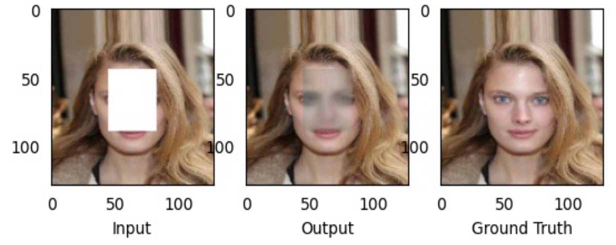


Fig. 4. Four masked positions and the corresponding results of face image inpainting. The first column is the input images with binary marks, the second column is the output result, and the third column is the ground truth.



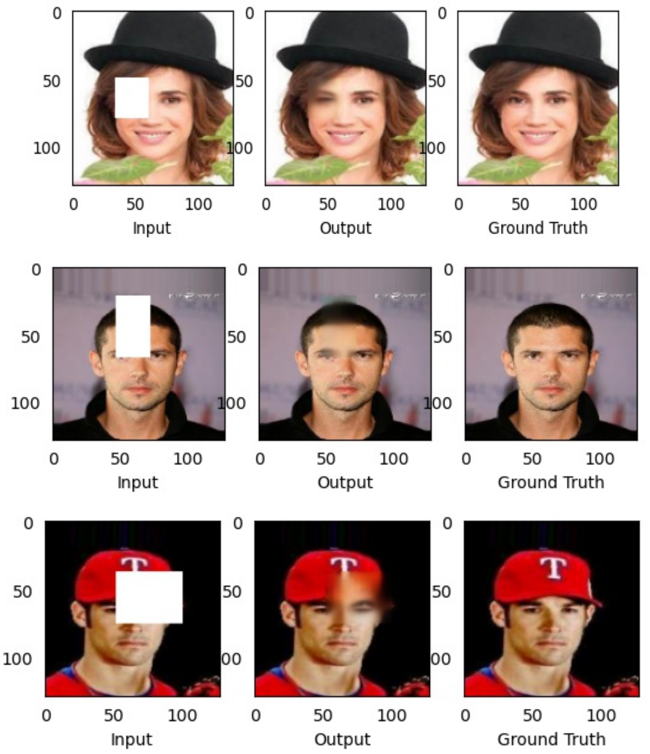


Fig. 5. Four masked positions and the corresponding results of face image inpainting by using the trained model of Iizuka's net. The first column is the input images with binary marks, the second column is the output result, and the third column is the ground truth.

TABLE IV. COMPARISONS OF IMAGE RESTORATION METHODS

Our results			***Iizuka's results***		
Names	***PSNRs***	***SSIMs***	***Names***	***PSNRs***	***SSIMs***
Result1	58.13	0.98	Result1	17.76	0.80
Result2	18.40	0.77	Result2	19.54	0.80
Result3	31.44	0.95	Result3	16.64	0.83
Result4	19.47	0.86	Result4	14.39	0.81
Average	**31.86**	**0.89**	**Average**	**17.09**	**0.84**

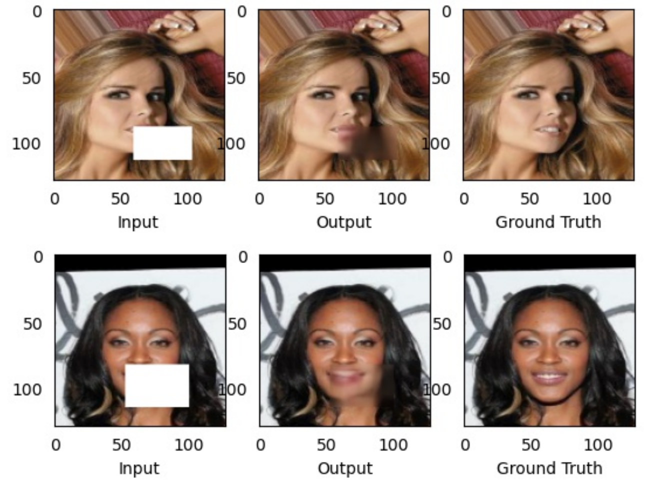


Fig. 6. Inpainted face images that obscure the part of human teeth

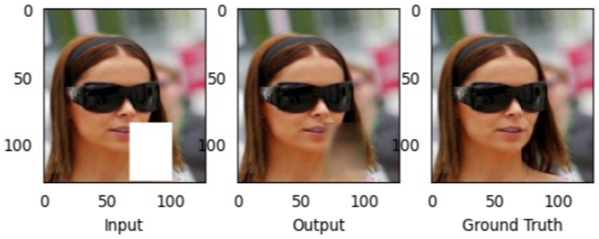


Fig. 7. The result of face image inpainting with the missed edges

We see that the average of PSNRs in the test results obtained by our proposed net is 31.86, and the average of SSIMs is 0.89. After trained Iizuka's network with 700 epochs, the average of PSNRs in the test results is 17.09, the average of SSIMs is 0.84. By comparing the results obtained by other training models, we find that the image quality of our modified model is improved.

In our results, there are also artefacts in the inpainted images. The worst case is the inpainted teeth on human face images. Using this network to reconstruct an image with masked teeth, the teeth on the images in the resultant image will be reconstructed as lips, as shown in Fig. 6. The differences between the inpainted images and ground truths show that the teeth of human face images have been blocked, but the images in the inpainted outputs only have lips. The teeth in the reconstructed face images have been lost.

V. Our Conclusion and Future Work

This research project is based on the image inpainting methods along with the deep neural networks. We propose the local and global discriminators. We modified the training method to minimize the MSE and two discriminators at the same time. In the completion net, we added convolution operations with a 3×3 kernel. The global discriminator is designed as a 5-layer convolutional network, the local discriminator is designed as a 4-layer convolutional network. The improved method considerably shortens the training time required. Currently, this work is able to inpaint up to 50.00% face area occluded in the given images.

In the experimental outcomes, the human face image reconstruction is excellent, especially for the regions of eyes and hair. In the long run, using facial masks to cover faces will be the norm. Therefore, the combination of face image inpainting and face recognition is possible to reduce the inconvenience caused by facial masks. We expect that our future work will solve the problems we found in this paper [25, 26]. One of the disadvantages of face image inpainting is that if the teeth are partially masked, the lips will be generated and output to replace our teeth after being inpainted. Moreover, if the edges on a face image are missing, the inpainted result is also not perfect, as shown in Fig. 7.

References

[1] C. Guillemot and O. Le Meur, “Image inpainting: Overview and recent advances,” IEEE Signal Processing Magazine, 2013, vol. 31, no. 1, pp. 127–144.

[2] Y. Jiang, J. Xu, B. Yang, J. Xu, and J. Zhu, “Image inpainting based on generative adversarial networks,” IEEE Access, 2020, vol. 8, pp. 22 884–22 892.

[3] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, “GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification,” Neurocomputing, 2018, vol. 321, pp. 321–331.

[4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, “Image in-painting,” in Annual Conference on Computer Graphics and Interactive Techniques, 2000, pp. 417–424.

[5] S. Iizuka, E. Simo-Serra, and H. Ishikawa, “Globally and locally consistent image completion,” ACM Transactions on Graphics (ToG), 2017, vol. 36, no. 4, pp. 1–14.

[6] M. Azman and A. Sarlan, “Aedes larvae classification and detection (ALCD) system by using deep learning,” in International Conference on Computational Intelligence, 2020, pp. 179–184.

[7] Q. Wang, Y. Chen, N. Zhang, and Y. Gu, “Medical image inpainting with edge and structure priors,” Measurement, 2021, vol. 185, pp. 110027.

[8] X. He, X. Cui, and Q. Li, “Image inpainting based on inside–outside attention and wavelet decomposition,” IEEE Access, 2020 vol. 8, pp. 62 343–62 355.

[9] M. M. Boulos, “Facial recognition and face mask detection using machine learning techniques,” Masters Thesis, Montclair State University, USA, 2021.

[10] W. Yan, M. Kankanhalli, “Erasing video logos based on image inpainting,” IEEE ICME, 2002: 521-524.

[11] J. Jam, C. Kendrick, K. Walker, V. Drouard, J. G.-S. Hsu, and M. H. Yap,“A comprehensive review of past and present image inpainting methods,” Computer Vision and Image Understanding, 2020, pp. 103-147.

[12] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in International Conference on Engineering and Technology, 2017, pp. 1–6.

[13] J. Xie, L. Xu, and E. Chen, “Image denoising and inpainting with deep neural networks,” in Advances in Neural Information Processing Systems, 2012, pp. 341–34.

[14] Z. Yan, X. Li, M. Li, W. Zuo, and S. Shan, “Shift-Net: Image inpainting via deep feature rearrangement,” in European Conference on Computer Vision (ECCV), 2018, pp. 1–17.

[15] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Generative image inpainting with contextual attention,” in IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5505–5514.

[16] M. Suganuma, M. Ozay, and T. Okatani, “Exploiting the potential ofstandard convolutional autoencoders for image restoration by evolutionary search,” in International Conference on Machine Learning. PMLR, 2018, pp. 4771–478.

[17] W. Yan, J. Wang, M. Kankanhalli, “Automatic video logo detection and removal,” Multimedia System, 2005, 10(5): 379-391.

[18] 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 European Conference on Computer Vision (ECCV), 2018, pp. 85–100.

[19] J. M.-T. Wu, Z. Li, N. Herencsar, B. Vo, and J. C.-W. Lin, “A graph-based CNN-LSTM stock price prediction algorithm with leading indicators,” Multimedia Systems, 2021, pp. 1–20.

[20] J. Vargas, J. Esquivel, and O. Tickoo, “Introducing regionpooling learning,” in International Conference on Pattern Recognition. Springer, 2021, pp. 714–724.

[21] Y. Yang, Z. Cheng, H. Yu, Y. Zhang, X. Cheng, Z. Zhang, and G. Xie,“MSE-Net: generative image inpainting with multi-scale encoder,” The Visual Computer, 2021, pp. 1–13.

[22] R. Zhang, W. Quan, B. Wu, Z. Li, and D. Yan, “Pixel-wise densedetector for image inpainting,” in Computer Graphics Forum, vol. 39, no. 7, 2020, pp. 471-482.

[23] Y. Chen, H. Zhang, L. Liu, X. Chen, Q. Zhang, K. Yang, R. Xia, and J. Xie, “Research on image inpainting algorithm of improved GAN basedon two-discrimination networks,” Applied Intelligence, 2021, vol. 51, no. 6, pp. 3460–3474.

[24] Z. Liu, P. Luo, X. Wang, and X. Tang, “Large-scale CelebFaces Attributes (CelebA) dataset,” Multimedia Laboratory, The Chinese University of Hong Kong China.

[25] W. Yan, Computational Methods for Deep Learning Theoretic, Practice and Applications. Springer, 2021.

[26] W. Yan, Introduction to Intelligent Surveillance Surveillance Data Capture, Transmission, and Analytics. Springer, 2019.