A Method for Face Image Inpainting Based on Autoencoder and Adversarial Generative Network

Xinyi Gao, Minh Nguyen, and Wei Qi Yan

Auckland University of Technology, Auckland 1010 New Zealand

Abstract. Face image inpainting has great value in the fields of computer vision and digital image processing. In this paper, we propose a face image inpainting method based on autoencoder and adversarial neural network (GAN). The neural network for image inpainting consists of two parts, a generator and a discriminator. The autoencoder is used twice in the discriminator part, after the final inpainted image is generated by local discriminator and global discriminator. The final loss function is obtained by combining Generative Adversarial Loss and Mean Squared Error (MSE) Loss [20]. We improve and implement an image inpainting model with two evaluation metrics, namely, Peak Signal-to-noise Ratio (PSNR) and Structural similarity index measure (SSIM) [27], respectively. The proposed model for image inpainting is much more suitable for face image inpainting.

Keywords: Autoencoder · Generative adversarial network · Face image inpainting · Convolutional neural network.

1 Introduction

Image inpainting is a process of inference from existing pixels. The method is applied mainly to fill in missing parts of an image. In 2000, Marcelo's paper [2] firstly proposed image inpainting as a professional technology through computer algorithms, so as to achieve the purpose of image repair. In this algorithm, after a user selects an image region that needs to be repaired, the algorithm will automatically repair the area through pixel information surrounding the missing region.

The purpose of image inpainting [3] is to reconstruct missing areas. Generally speaking, this process is often to repair an image based on existing information. The reconstructed image should be realistic and natural. The indistinguishable image is treated as a high-quality inpainted image [43]. A high-quality face image inpainted can be employed for unlocking mobiles using face recognition.

In recent years, the world has been affected by COVID-19 [28]. To protect us against the virus, we need to wear face masks so as to protect ourselves. While wearing masks protects our health, it also brings new problems. We need a full face to help us if we want to make quick payments or verify our identity using our faces. However, in this case, we have to take off our masks. Therefore, a fast and accurate face image inpainting method in this epidemic time is very needed.

In this paper, a deep learning method [16] for face image inpainting is proposed. Deep learning is a machine learning method [47] based on data representation learning. Convolutional neural networks (CNN) [14][23], as a kind of method in deep learning, have been utilized to produce reliable results in the field of image inpainting [31]. Autoencoders are deep nets in unsupervised learning [9]. The purpose of an autoencoder is to learn a representation of a set of data. An adversarial generative network (GAN) [36, 35] is also an unsupervised learning method whose main structure is constructed by two deep nets which play an adversarial game. The recently proposed model consists of two autoencoders and a GAN net [26].

In the rest of this paper, we will present related work on face image reconstruction in Section 2. In the third section, we introduce our method and model. In Section 4, we analyze the results of our experiments. Finally, we summarize our work and draw conclusions in Section 5.

2 Literature Review

With the development of deep learning, multiple methods for human face [29] image reconstruction are becoming more and more mature. A number of methods and frameworks for face image inpainting are emerging gradually. The typical methods include CNN, GAN, etc., which are available to attain higher-quality inpainting results. In this section, a brief review of the existing face reconstruction methods is depicted and the existing methods are summarized based on advantages and disadvantages.

Convolutional neural network (CNN or ConvNet) is also one of the important methods in deep learning [17][30]. Although CNN can generate a reasonable structure in image restoration, the image generated by CNN [18] has structural inconsistency or fuzzy texture in the relevant regions. A new method [40] was proposed and the reason for this problem is identified, especially, when the deep net borrows textures from the surrounding areas. Therefore, the method was derived from a generative model based on traditional texture and patch synthesis. This model is essentially a feedforward, fully connected network. The net can synthesize new image structures during inpainting, it has been verified to better use surrounding image features as references. The experiments have proved that the proposed model is effective to repair images from multiple datasets including human faces with higher quality than those existing methods. Later on, a new system was proposed to learn from millions of images. The basic principle of this system is based on gated convolution, which eliminates extra marks. In the specific operation of the system, partial convolution is summarized by providing a dynamic feature selection mechanism for each channel of spatial position of all layers.

In addition, the randomness of the mask was considered in the image and a new GAN loss [25] function was proposed, which is called SN-PatchGAN [41].

The experimental results show that the results produced by the system have higher quality and much flexible results. This allows the system to assist users to quickly remove distracting objects, modify image layout, wipe off watermarks [37][6][1], etc.

The existing image reconstruction methods are required to pay special attention to the resolution of the target image. In order to solve this problem, Zeng, et al. proposed a different high-quality image reconstruction method [43], which is called Pyramid-context Encoder Network (PEN-Net). In the method, the structure of context encoder is increased, pyramid context encoder, multiscale decoder, and adversarial training loss are established. In this method, U-Net is employed as the backbone, the pyramid context encoder is applied to gradually fill in the missing content and ensure the consistency of the visual effects of image reconstruction. Then, a multiscale decoder with deep supervision function is harnessed to calculate the loss. The use of this method allows the process to converge quickly during training time, and a large number of experiments have proved the reliability and excellent performance of the proposed net.

In order to solve the problem of blurred or missing face images collected due to acquisition method in the process of face recognition, conventional face image restoration models often solve this problem from image viewpoint. The classic structure-based image restoration methods are CNN [32] and generative adversarial network. Wei, et al. put forward a face image reconstruction method based on generative confrontation network from a new perspective [34]. This method locates plane position of a face by determining two parallel lines of a vector. The different planes of the face are determined according to the given parallel lines, edge curve is fitted through straight line segment to make facial contour clearer, and make final facial features quite obvious [5]. Compared with the previous structure-based methods, this method can achieve better visual effects based on edges. The performance of face inpainting is greatly improved. It is worth mentioning that this method is only suitable for small-scale reconstruction processes. If there are too many missing parts in an image, the restoration effect will become vague from the original image.

Image information missing is one of the most popular damages in image damage. The existing image reconstruction algorithms still have shortcomings, such as blurred details and poor visual [39] perception in terms of visual effects and algorithm efficiency after the reconstruction. In order to solve this problem, Heshu et al. [44] proposed a new semantic restoration method for facial images in 2020. On the basis of generating a confrontation network, this method fuses multiscale features of the given face to obtain more details without increasing the parameters. By expanding the receptive field in the deep net [21], the problem of insufficient edge information of the generated image is made up. In addition, the learning ability of the generative network and the discriminative network is justified, which further improved the final performance of the proposed method.

Deep learning is the mainstream method for image restoration. The use of deep learning methods [24][42][45] in image reconstruction can better restore the image texture of human faces and obtain abstract features in the image. Therefore, in the same way, Han et al. [10] also employed generative adversarial network as the basis for image inpainting process. However, they proposed a different method to solve vanishing gradient problem in the training process of GAN model. Evolutionary concepts were adopted to create a Generative Adversarial Network (EG-GAN) with an evolutionary generator for face image restoration. In the training process, EG-GAN updates the parameters of the generative network by combining two cost functions, generates offspring generators through crossover, and adds a matcher-assisted discriminator to criticize the generated images. Through the conception, the generative network continues to be evolved. This not only helps EG-GAN successfully overcome the vanishing gradient problem, but also improves the quality of image reconstruction and generated images that are in line with human vision.

In the following sections, we will briefly describe the method we developed in this paper. The method is slightly different from other work. Our method is use of deep learning [13][33] and adds an autoencoder to enhance our method, so that this method can much effectively resolve the problem of face image inpainting.

3 Methodology

In this research project, we take use of CNN in deep learning [12] as the net base and the GAN model to build a complete network for face image reconstruction, train our image inpainting model through discriminator and generator in the GAN model to achieve higher quality.

As a typical method for face image reconstruction, GAN [4] is also one of the most important methods in this paper, which was firstly proposed in 2014. The basic idea of this method is to deploy two contrastive neural networks against each other. This is a way to get better results by playing against two networks. GAN generally consists of a generative network and a discriminative network. The generator needs to generate more realistic images, the discriminator judges how "real" the input is.

Fig. 1. GAN training process. This network consists of a generator and a discriminator. The discriminator part is composed of a global discriminator and a local discriminator. The discriminator is used to judge whether the image is real or not.

3.1 Generative Network

CNN is the core part of the entire inpainting algorithm, so the convolutional layer becomes the key to our entire CNN algorithm. This is because most of the operations in the entire inpainting process are generated in convolutional layers. The generative network consists of two autoencoders connected. Each autoencoder consists of 12 layers of convolution operations, 4 layers of dilated convolutions, and 2 layers of deconvolutions [8] [7]. The face image [38] generated by two autoencoders is closer to the "real" image than the image generated by only using one autoencoder.

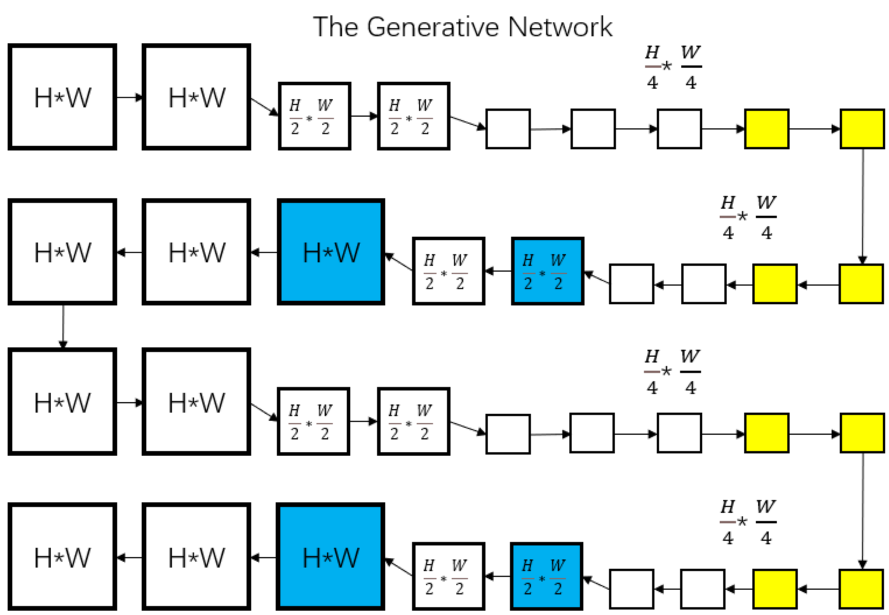

Fig. 2. The generative network consists of a total of 36 layers of convolutional neural networks. There are 24 convolutional layers in the generative network, represented by rectangles with white background color. There are 8 yellow rectangles in total, which are dilated convolution. There are also 4 additional deconvolution layers, represented by blue squares.

3.2 Discriminator Network

The discriminative network consists of a global discriminator network and a local discriminator network [11]. Its purpose is to judge whether the image generated by the generative network is close to the "real" image and identify whether a part of the image is "real" through a local discriminator. The global discriminator

Table 1. The generative network.

Convolution types	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
Dilated Convolution1	3×3	2	256
Dilated Convolution2	3×3	4	256
Dilated Convolution3	3×3	8	256
Dilated Convolution4	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
Convolution13	5×5	None	64
Convolution14	3×3	None	64
Convolution15	3×3	None	128
Convolution16	3×3	None	128
Convolution17	3×3	None	256
Convolution18	3×3	None	256
Convolution19	3×3	None	256
Dilated Convolution5	3×3	2	256
Dilated Convolution6	3×3	4	256
Dilated Convolution7	3×3	8	256
Dilated Convolution8	3×3	16	256
Convolution20	3×3	None	256
Convolution21	3×3	None	256
Deconvolution3	4×4	None	128
Convolution22	3×3	None	128
Deconvolution4	4×4	None	64
Convolution23	3×3	None	32
Convolution24	3×3	None	3

is responsible for the plausibility of overall image inpainting. A more accurate and realistic inpainting result is achieved by combining the results of the two discriminators. These two parts of the network are still composed of CNN. The global discriminative network is composed of 5 convolutional layers. The details are demonstrated in Table 2. The local discriminative network consists of 4 convolutional layers. The details are shown in Table 3.

Table 2. The global discriminator network.

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

Table 3. The local discriminator network.

Convolution types	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

3.3 Algorithms

Loss function is an important element in deep learning [46][19]. It is generally used to measure the degree of inconsistency between the predicted value of the model and the true value. A loss function can provide a lot of practical flexibility to a neural network, it will define how the output of the network is connected to the rest of the network. Two loss functions are employed in our experiments. One is GAN loss and the other is MSE loss.

Minimax Loss A GAN can have two loss functions: One for generator training and the other for discriminator training. The generative and discriminative losses look different from one equation [10]. In this experiment, we mainly take use of minimax loss function in GAN model. In the minimax loss function, $D(I, M)$ represents the discriminative network and $G(I, M_i)$ shows the generative network, M is a randomly generated mask, and I is the input image. M_i is the mask image having the exact size of the input image.

$$L_{MinMax} = \min_G \max_D \mathbf{E}[logD(I, M) + log(1 - D(G(I, M_i), M_i))]. \qquad (1)$$

MSE Loss Mean Squared Error (MSE) is a popular loss function. You need to square the difference between the prediction and the ground truth [48]. Then, it takes average over the entire dataset. Finally, the MSE loss value can be obtained. In the MSE loss function [48], G represents the generative network and I is the input image, $\odot$ is pixel-wise multiplication. The value of MSE loss is always greater than 1.00. The closer this value is to 1.00, the more realistic the training results are

$$L_{MSE} = ||M_i \odot (G(I, M_i) - I)||^2. \tag{2}$$

Joint Loss The joint loss function is to combine together MSE loss and minimax loss to obtain better training results. In training the network, we firstly employed only the minimax loss. Then we take use of the joint loss for training in later training. Such training methods can effectively improve training results.

$$L_{joint} = \min \lim_{G} \max \lim_{D} \mathbf{E}[L_{MSE} + log(I, M) + log(1 - D(G(I, M_i), M_i))]. \tag{3}$$

In the following sections, we will present our experimental results, which further demonstrate the superiority of our algorithm design.

4 Result Analysis

In this research project, we took use of the CelebA dataset with 202,599 face images for training [15]. CelebA is a large-scale dataset dedicated to face experiments, containing more than 200,000 face images. In addition, the backgrounds of the various face images in this dataset are often complex, which makes this dataset a very suitable training dataset for this study. An example image of the dataset is shown in Figure 3. There are 2,000 face images which were randomly selected for training and testing.

The images are firstly preprocessed, and the pixels of the images to be used are resized to 128×128. During the model training, we randomly add a mask image whose size ranges from 24×24 to 48×48. The inpainting result is shown in Figure 5, we see that when we randomly add a mask to the image, the inpainting is accomplished by using the autoencoder. The repaired image is very close to the result of the unbroken image. From a naked eye point of view, this is undoubtedly a very successful restoration. However, we still need data to support the conclusions we see, here we choose PSNR and SSIM as our data support. We took use of PSNR and SSIM as evaluation metrics in the testing phase. PSNR is the most popular objective measure for evaluating image quality. Usually, the larger the PSNR value, the higher the quality of image inpainting. SSIM is a measure of how similar two images are [22]. The result is between 0 and 1.00, and the closer the result is to 1.00, the higher the quality of image inpainting.

Since our generative network is use of two autoencoder networks, it is slower to train than a model with just one autoencoder. But at the same training 1,000

Fig. 3. The example of CelebA dataset.

epochs, the results of our inpainting method are better than that of just using an autoencoder network.

Table 4. Comparison of Image Inpainting Methods.

Model Name	PSNRs	SSIMs
Glcic(One Autoencoder)	31.86	0.89
Our(Two Autoencoder)	36.74	0.91

The test results are shown in Table 4. The average PSNR of the network with only one autoencoder in the generative network is 31.86, and the SSIM is 0.89. Our test result has an average PSNR 36.74, and SSIM is 0.91. We see that our model has higher PSNR compared to the model using one autoencoder, our model is about 5% higher based on this metric. The same is true for another metric, SSIM is 0.02 higher than the model using one autoencoder. By comparison, it is found that our modified model works better. Furthermore, our SSIM is up to 0.91, which is very close to 1.00. The PSNR is also high, which not only proves that our results are very good, but also further shows the reliability of these two indicators. Overall, the deep learning model we used can accurately inpaint missing parts of face images. With a number of iterations, the more

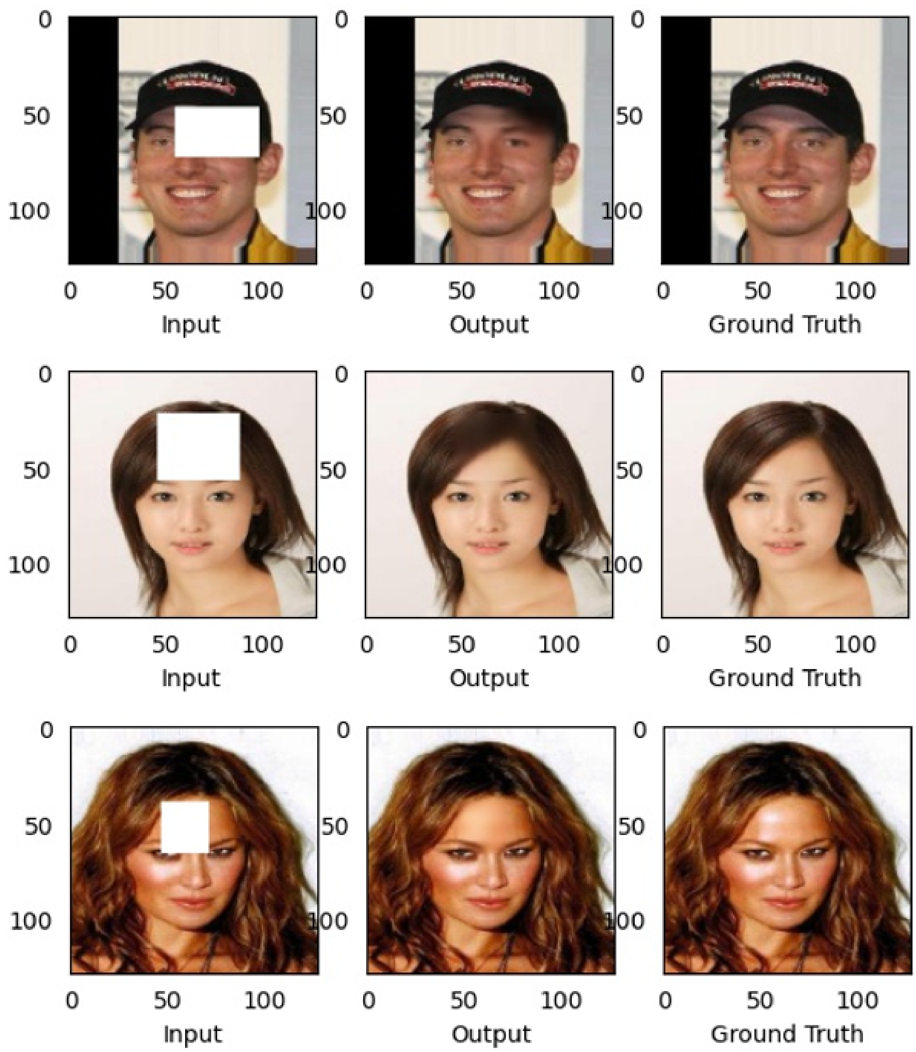

Fig. 4. Three face image inpainting results. The first column is the input image with Mask, the second is the output result, and the third is the ground truth.

iterations, the better the visual effect. Although the model suffers from slow inpainting speed, it is still an efficient model. In addition, the results of SSIM and PSNR are also excellent, the overall reconstruction results are very realistic. These are the great contribution of this paper.

5 Conclusion

This research project investigates face image inpainting based on autoencoder and GAN. We propose a new deep learning model. Two autoencoder networks are adopted in the generative part. We modified the training method. Firstly, we calculated the minimax loss, and then the joint loss of the combination of minimax loss and MSE loss. Due to the increased number of layers of the generative network, the training speed has decreased. But the improved method uplifts image inpainting quality and increases the SSIM of the inpainting result from 0.89 to 0.91.

In the future, based on this result, we will make use of other test methods to further demonstrate the superiority of our algorithm. In addition, we will continue to optimize the algorithm and achieve better results. We will improve the inpainting speed without reducing the visual effect. Furthermore, we will test more datasets and check the results of our proposed method. We will inpaint face images with large missing parts to improve the generality of the model.

References

1. Bansal, M., Yan, W.Q., Kankanhalli, M.S.: Dynamic watermarking of images. In: International Conference on Information, Communications and Signal Processing and the Fourth Pacific Rim Conference on Multimedia. vol. 2, pp. 965–969. IEEE (2003)
2. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: Annual Conference on Computer Graphics and Interactive Techniques. pp. 417–424 (2000)
3. Chen, Y., Liu, L., Tao, J., Xia, R., Zhang, Q., Yang, K., Xiong, J., Chen, X.: The improved image inpainting algorithm via encoder and similarity constraint. The Visual Computer **37**(7), 1691–1705 (2021)
4. Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., Bharath, A.A.: Generative adversarial networks: An overview. IEEE Signal Processing Magazine **35**(1), 53–65 (2018)
5. Cui, W., Yan, W.Q.: A scheme for face recognition in complex environments. International Journal of Digital Crime and Forensics (IJDCF) **8**(1), 26–36 (2016)
6. Ding, W., Yan, W.Q., Qi, D.X.: Digital image scrambling and digital watermarking technology based on conway's game. Journal of North China University of technology **12**(1), 1–5 (2000)
7. Gao, X.: A Method for Face Image Inpainting Based on Generative Adversarial Networks (Masters Thesis). Auckland University of Technology, New Zealand (2022)
8. Gao, X., Nguyen, M., Yan, W.Q.: Face image inpainting based on generative adversarial network. In: International Conference on Image and Vision Computing New Zealand (IVCNZ). pp. 1–6. IEEE (2021)

9. Givkashi, M.H., Hadipour, M., PariZanganeh, A., Nabizadeh, Z., Karimi, N., Samavi, S.: Image inpainting using AutoEncoder and guided selection of predicted pixels. arXiv preprint arXiv:2112.09262 (2021)
10. Han, C., Wang, J.: Face image inpainting with evolutionary generators. IEEE Signal Processing Letters **28**, 190–193 (2021)
11. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Transactions on Graphics (ToG) **36**(4), 1–14 (2017)
12. Le, H., Nguyen, M., Nguyen, Q., Nguyen, H., Yan, W.Q.: Automatic data generation for deep learning model training of image classification used for augmented reality on pre-school books. In: International Conference on Multimedia Analysis and Pattern Recognition (MAPR). pp. 1–5 (2020)
13. Liang, C., Lu, J., Yan, W.: Human action recognition from digital videos based on deep learning. ACM ICCCV 2022 (2022)
14. Liu, Z., Yan, W.Q., Yang, M.L.: Image denoising based on a CNN model. In: International Conference on Control, Automation and Robotics (ICCAR). pp. 389–393. IEEE (2018)
15. Liu, Z., Luo, P., Wang, X., Tang, X.: Large-scale celebfaces attributes (CelebA) dataset. Retrieved August **15**(2018), 11 (2018)
16. Lu, J., Nguyen, M., Yan, W.Q.: Sign language recognition from digital videos using deep learning methods. In: International Symposium on Geometry and Vision. pp. 108–118. Springer (2021)
17. Lu, J., Yan, W.Q., Nguyen, M.: Human behaviour recognition using deep learning. In: IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). pp. 1–6. IEEE (2018)
18. Mehtab, S., Yan, W.Q.: Flexible neural network for fast and accurate road scene perception. Multimedia Tools and Applications **81**(5), 7169–7181 (2022)
19. Nguyen, M., Yan, W.Q.: Temporal colour-coded facial-expression recognition using convolutional neural network. In: International Summit Smart City 360°. pp. 41–54. Springer (2022)
20. Pajot, A., de Bezenac, E., Gallinari, P.: Unsupervised adversarial image inpainting. arXiv preprint arXiv:1912.12164 (2019)
21. Qin, Z., Yan, W.Q.: Traffic-sign recognition using deep learning. In: International Symposium on Geometry and Vision. pp. 13–25. Springer (2021)
22. Sara, U., Akter, M., Uddin, M.S.: Image quality assessment through FSIM, SSIM, MSE and PSNR— a comparative study. Journal of Computer and Communications **7**(3), 8–18 (2019)
23. Shen, D., Nguyen, M., Yan, W.Q.: Flame detection using deep learning. In: International Conference on Control, Automation and Robotics (ICCAR). pp. 389–393. IEEE (2018)
24. Shen, Y., Yan, W.Q.: Blind spot monitoring using deep learning. In: International Conference on Image and Vision Computing New Zealand (IVCNZ). pp. 1–5. IEEE (2018)
25. Siavelis, P.R., Lamprinou, N., Psarakis, E.Z.: An improved gan semantic image inpainting. In: International Conference on Advanced Concepts for Intelligent Vision Systems. pp. 443–454. Springer (2020)
26. Siddavatam, I., Dalvi, A., Pawade, D., Bhatt, A., Vartak, J., Gupta, A.: A novel approach for video inpainting using autoencoders. I.J. Information Engineering and Electronic Business **6** (2021)
27. Tiefenbacher, P., Bogischef, V., Merget, D., Rigoll, G.: Subjective and objective evaluation of image inpainting quality. In: 2015 IEEE International Conference on Image Processing (ICIP). pp. 447–451. IEEE (2015)

28. Velavan, T.P., Meyer, C.G.: The COVID-19 epidemic. Tropical Medicine & International Health **25**(3), 278 (2020)
29. Wang, H., Yan, W.Q.: Face detection and recognition from distance based on deep learning. In: Aiding Forensic Investigation Through Deep Learning and Machine Learning Frameworks, pp. 144–160. IGI Global (2022)
30. Wang, L., Yan, W.Q.: Tree leaves detection based on deep learning. In: International Symposium on Geometry and Vision. pp. 26–38. Springer (2021)
31. Wang, S.Y., Wang, O., Zhang, R., Owens, A., Efros, A.A.: CNN-generated images are surprisingly easy to spot... for now. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8695–8704 (2020)
32. Wang, X., Zhang, J., Yan, W.Q.: Gait recognition using multichannel convolution neural networks. Neural computing and applications **32**(18), 14275–14285 (2020)
33. Wang, Y., Yan, W.Q.: Colorizing grayscale ct images of human lungs using deep learning methods. Multimedia Tools and Applications pp. 1–15 (2022)
34. Wei, T., Li, Q., Liu, J., Zhang, P., Chen, Z.: 3D face image inpainting with generative adversarial nets. Mathematical Problems in Engineering (2020)
35. Yan, W.Q.: Introduction to Intelligent Surveillance. Springer (2019)
36. Yan, W.Q.: Computational Methods for Deep Learning: Theoretic, Practice and Applications. Springer (2021)
37. Yan, W.Q., Kankanhalli, M.S.: Erasing video logos based on image inpainting. In: IEEE International Conference on Multimedia and Expo. vol. 2, pp. 521–524. IEEE (2002)
38. Yan, W.Q., Kankanhalli, M.S.: Face search in encrypted domain. In: Image and Video Technology. pp. 775–790. Springer (2015)
39. Yan, W., Kieran, D.F., Rafatirad, S., Jain, R.: A comprehensive study of visual event computing. Multimedia Tools and Applications **55**(3), 443–481 (2011)
40. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 5505–5514 (2018)
41. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Free-form image inpainting with gated convolution. In: IEEE/CVF International Conference on Computer Vision. pp. 4471–4480 (2019)
42. Yu, Z., Yan, W.Q.: Human action recognition using deep learning methods. In: International Conference on Image and Vision Computing New Zealand (IVCNZ). pp. 1–6. IEEE (2020)
43. Zeng, Y., Fu, J., Chao, H., Guo, B.: Learning pyramid-context encoder network for high-quality image inpainting. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1486–1494 (2019)
44. Zhang, H., Li, T.: Semantic face image inpainting based on generative adversarial network. In: Youth Academic Annual Conference of Chinese Association of Automation (YAC). pp. 530–535. IEEE (2020)
45. Zhang, Q., Yan, W.Q.: Currency detection and recognition based on deep learning. In: IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). pp. 1–6. IEEE (2018)
46. Zhang, Q., Yan, W.Q., Kankanhalli, M.: Overview of currency recognition using deep learning. Journal of Banking and Financial Technology **3**(1), 59–69 (2019)
47. Zhang, Z., Cui, P., Zhu, W.: Deep learning on graphs: A survey. IEEE Transactions on Knowledge and Data Engineering (2020)
48. Zhao, G., Liu, J., Jiang, J., Wang, W.: A deep cascade of neural networks for image inpainting, deblurring and denoising. Multimedia Tools and Applications **77**(22), 29589–29604 (2018)