Overview of Image Defect Inpainting Methods Based on U-Net

Kai Xue

***Abstract*—The research of image defect inpainting aims to automatically repair the defect content in the image through the computer. In recent years, the emergence of deep neural network technology has effectively promoted the development of related research. This article systematically sorts out and comprehensively introduces U-Net inpainting methods. We specifically analyzed the ideas, characteristics, advantages and disadvantages of each type of method, and based on systematic experiments, objectively compared and evaluated the accuracy and performance of each type of method on a public large-scale data set. Finally, the current problems and challenges in related work are elaborated and introduced.**

***Index Terms*—Computer vision, Deep neural network, Image inpainting, Image processing**

I. Introduction

Images are one of the most important ways of expressing and transmitting information for human beings, and they play an irreplaceable and prominent role in human communication. However, in the process of information transmission and storage, losses will inevitably occur, affecting the quality of information presentation. In the most commonly used digital images at present, the main manifestation of information deficiency is pixel loss, as shown in Fig. 1.

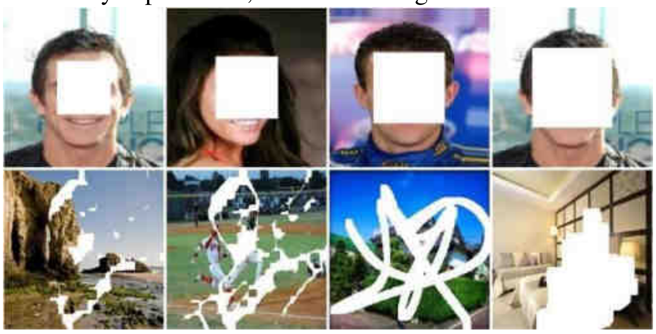


Fig. 1 Image defect. In the first row, the proportion of defects from left to right was 10%, 15%, 20%, and 25%, respectively; in the second line, the proportion of defects from left to right was 10%, 20%, 30% and 40%, respectively.

It can be seen from the figure that the lack of information has a serious impact on the presentation of image content. Moreover, because the image is non-structural information, the relative independence of each part of the content is poor, even a small proportion of defects (such as the first column, the image is only missing 10% of the content), will bring the whole image Intuitive feeling of poor quality. Obviously, images with missing content cannot be directly presented to readers and users. Today, with the continuous popularity of image acquisition equipment, the number of various types of images has increased exponentially every day, and it is imperative to research and design automated methods and technologies that can intelligently repair image defects.

Kai Xue, Computer Science and Technology, Tiangong University, City Name, China

Because the missing information can not be made out of nothing, the key work to deal with the image content defect is how to obtain the compensation information. As we all know, images can express a wide range of contents, including color, texture, structure, composition and other aspects contain a lot of variability. Therefore, the compensation knowledge modeling of general images belongs to the high-level structural abstraction of super large-scale information. Under the premise of limited storage space and computing capacity, it is extremely difficult to carry out such work. Therefore, in the past quite a long time, the problem of image defect automatic inpainting developed slowly until the emergence of deep neural network [1] technology.

The main contributions of this paper are:

(1) The most effective state of the art method for solving image defect inpainting problems, that is, U-Net [2] methods, has been systematically analyzed and summarized;

(2) The open data sets, accuracy and performance evaluation indexes of current image defect inpainting methods are systematically summarized and introduced;

(3) In this paper, we carry out analysis and evaluation experiments on large-scale open data sets, analyze and compare the actual accuracy and performance of the relevant methods within and between classes, and explore the core causes behind them;

(4) Systematic analysis lists the challenging problems and key technical difficulties faced by the current research.

II. Overview of relevant basic knowledge

A. *Evaluation index*

As it is a relatively new research topic, there is no industry-recognized proprietary quantitative evaluation index for defect image restoration. The evaluation of related methods is currently mainly based on intuitive qualitative observation or overall image quality evaluation indicators. For the latter, the most commonly used in recent years are the Peak Signal to Noise Ratio (PSNR) [3] and SSIM (Structural Similarity) [4]. Both of these indicators evaluate the effect of the method by comparing the difference between the image I_x obtained after processing and the target image I_y. Its specific definition is as follows:

$$PSNR(I_x, I_y) = 10 \cdot \log_{10}\left(\frac{MAX_I^{\,2}}{MSE}\right), \quad (1)$$

$$SSIM(I_x, I_y) = \frac{(2\mu_{I_X}\mu_{I_y}+c_1)(2\sigma_{I_xI_y}+c_2)}{(\mu_{I_x}^2+\mu_{I_y}^2+c_1)(\sigma_{I_x}^2+\sigma_{I_y}^2+c_2)} \quad (2)$$

$$MSE(I_x, I_y) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\left\|I_x(i,j) - I_y(i,j)\right\|^2 \quad (3)$$

Here, MSE is the mean square error of the image I_x and the image I_y; MAX_I indicates the upper limit of the gray value allowed by the input image format, and for grayscale images, the value is 255; μ_{I_X} means the pixel average of the image I_x and the image I_y and σ_{I_x} , σ_{I_y} respectively variance; $\sigma_{I_x I_y}$ means covariance; c_1, c_2 are constants; H and W are the height and width of the image respectively.

B. Evaluation datasets

Standard image datasets are the basis for scientific evaluation of image defect inpainting algorithms. We have classified and summarized commonly used datasets in the field of image inpainting. See Table. 1 for detailed datasets information.

Table 1 Datasets information

Datasets	Download link	Size
CelebA [5]	http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html	202k
Places2 [6]	http://places2.csail.mit.edu/	10M

III. RESEARCH STATUS AND ANALYSIS

U-Net network evolved from FCN [7] network and was initially applied in the field of image segmentation. The reason why U-Net network is applied to image defect inpainting is that it has a unique feature fusion method: U-Net can splice features of different scales together in channel dimension, and the last feature of up-sampling output is obtained by fusing the features from the first convolution module and the last up sampling output. This unique feature fusion method deeply fuses the low-level features and high-level features in the image. In the field of image defect inpainting, feature fusion is very important for image restoration. It not only needs to use low-level features such as color and edge, but also combines high-level features such as structure and semantics. Based on this, U-Net structure has been widely used in the field of image defect inpainting [8]-[10].

As shown in Fig. 2, the U-Net network structure mainly consists of two parts: the first part is mainly used for feature extraction. The network architecture is similar to the VGG network architecture [11]. Each pooling layer in the network layer is a scale; the second part is the upper sampling part. Each up-sampling, the output layer of the network is fused with the channel number corresponding to the feature extraction part. The U-Net network has no full connection layer, and the convolution process adopts the valid mode, which ensures that the results are obtained from the context features without missing.

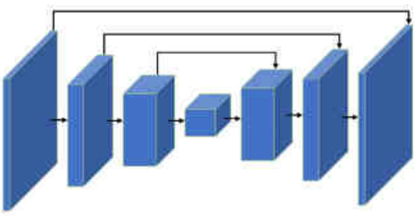


Fig. 2 U-Net network architecture

IV. EXPERIMENT ANALYSIS

Table. 2 and Table. 3 show the quantitative experimental performance of u-net class methods on the images of central region defect and random region defect respectively. It can be seen from the table that the performance of DF-Net method [12] is overall superior to that of DF-Net method (29 out of all 32 tests have achieved the best results). It can be seen that the loss of perception and style introduced by DF-Net, as well as the embedding strategy of fusion module, have a significant effect on the performance improvement of u-net class methods.

Table 2 Inpainting performance in central region defect image

Datasets	CelebA		Place2	
Index	PSNR	SSIM	PSNR	SSIM
Shift-Net[13]	22.82	0.865	20.98	0.802
DFNet[12]	25.17	0.860	21.10	0.808
PEN-Net[14]	21.70	0.817	20.22	0.729

Table 3 Inpainting performance in random region defect image

Datasets	CelebA		Place2	
Index	PSNR	SSIM	PSNR	SSIM
Shift-Net[13]	23.62	0.808	20.70	0.712
DFNet[12]	26.96	0.864	21.63	0.754
PEN-Net[14]	20.04	0.713	17.78	0.564

Fig. 3 and Fig. 4 show the corresponding qualitative experimental performance. It can be seen from the figure that the DF-Net method can achieve better inpainting effect in accordance with the quantitative performance.

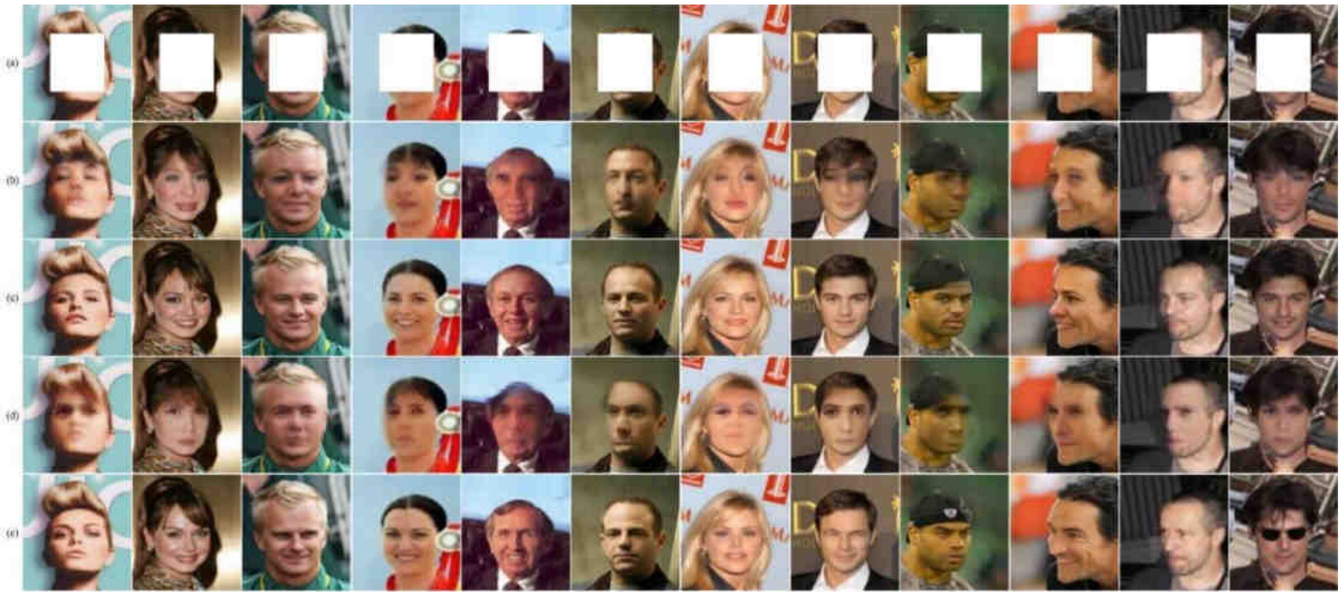
The resource consumption Table. 4 of U-Net class is shown. It can be seen that DF-Net has obvious advantages in the most important test time, and can complete image restoration in less computing time. But in terms of the use of video memory, it is obviously higher than the Shift-Net [13] method.

Table 4 Resource consumption

Resource（CelebA/Place2）	Train time(h)	Test time(s)
Shift-Net[13]	49.9/317.0	0.210/0.360
DFNet[12]	153.8/208.3	0.060/0.065
PEN-Net[14]	157.0/220.5	0.154/0.160

V. PROBLEMS AND CHALLENGES FACED

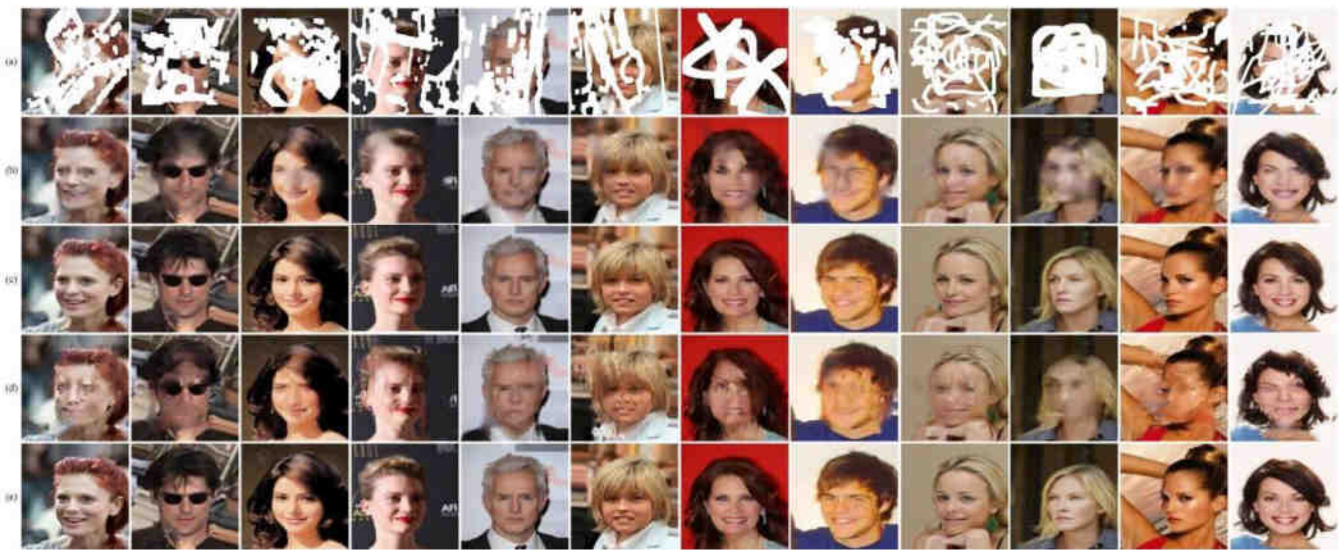
There is no doubt that defect image restoration has a clear application background and outstanding practical value, which is an indispensable and important research topic in the fields of computer vision, image processing, machine learning and so on. The emergence and development of deep neural network technology effectively promotes the overall technical level of this field. However, the emergence of deep network is not long, and large-scale application of image inpainting is only a few years. The overall technical level is not mature, and there is still room for improvement. In addition to the accuracy issues already mentioned, other problems and challenges currently facing may include:



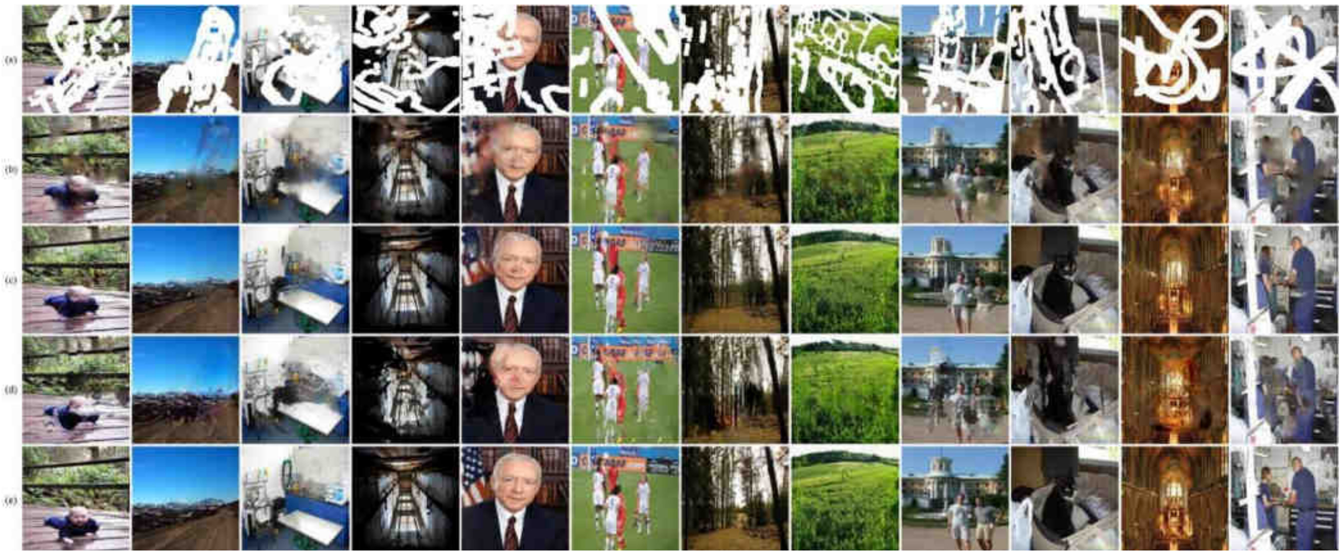
(A) CelebA datasets

(B) Place2 datasets

Fig. 4 The repair effect of U-Net methods when the central area is defective. (a) Input image, (b) Shift-Net method restoration result, (c) DFNet method restoration result, (d) PEN-Net method restoration result, (e) Ground Truth



(A) CelebA datasets



(B) Place2 datasets

Fig. 5 The repair effect of U-Net methods when the random area is defective. (a) Input image, (b) Shift-Net method restoration result, (c) DFNet method restoration result, (d) PEN-Net method restoration result, (e) Ground Truth

(1) Defect inpainting involves not only high-level semantic knowledge, but also low-level pixel information. Only by integrating the two parts of information in high-level structure can we approach the level of image restoration of human visual system.

(2) Due to the huge computational cost caused by the huge network structure, the current deep neural network method can only deal with images with low resolution.

(3) In the field of image restoration, a large number of methods need to know the defect area in advance, so it is not a complete end-to-end solution.

(4) There is still much room for improvement in the generalization ability of the model. In the current related research, the method still has to be trained on the high correlation training set in order to achieve better results on the corresponding test set.

(5) Specific metrics are still lacking.

VI. CONCLUSION

In this paper, the image defect repair technology based on U-Net network is studied systematically. Its core ideas and technical characteristics are summarized and analyzed. For the three methods of public code, systematic evaluation experiments and performance comparison are carried out. Finally, the existing problems and challenges in the current related research are introduced and elaborated.

REFERENCES

[1] Huang Y, Sun S, Duan X, et al A study on deep neural networks framework//Proceedings of IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, Xi'an, China, 2016:1519-1522.

[2] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation//Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention. Munich, Germany, 2015:234-241.

[3] Johnson D H. Signal-to-noise ratio. Scholarpedia, 2006, 1(12): 2088.

[4] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[5] Liu Z, Luo P, Wang X, et al. Deep learning face attributes in the wild//Proceedings of the IEEE International Conference on Computer vision. Santiago, Chile, 2015:3730-3738..

[6] Zhou B, Lapedriza A, Khosla A, et al. Places: A 10 million image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(6): 1452-1464..

[7] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 3431-3440.

[8] LiuY, QiN, Zhu Q, et al.CR-U-Net: Cascaded U-Net with Residual Mapping for Liver Segmentation in CT Images//Proceedings of theIEEE Visual Communications and Image Processing.Sydney, Australia, 2019:1-4.

[9] Han Y,YeJ C, Framing u-net via deep convolutional framelets: application to sparse-view ct. IEEE Transactions on Medical Imaging, 2018,37(6):1418-1429..

[10] Zhang Z, Liu Q and Wang Y. Road Extraction by Deep Residual U-Net. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5):749-753.

[11] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556,2014.

[12] Hong X, Xiong P, Ji R, et al. Deep fusion network for image completion. arXiv preprint arXiv:1904.08060, 2019.

[13] Yan Z, Li X, Li M, et al. Shift-net: Image inpainting via deep feature rearrangement//Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018:1-17.

[14] Zeng Y, Fu J, Chao H, et al. Learning pyramid-context encoder network for high-quality image inpainting//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. California, USA, 2019: 1486-1494.

Kai Xue received the B.Sc. degree in Computer Science and Technology from Xi'an University of Posts & Telecommunications in 2018. Currently, he is a postgraduate at Tiangong University. His research interests include image inpainting and generation, image occlusion reconstruction, face reconstruction, etc.