

# Review of image inpainting in practical challenge

**Yuhang Shen**

University of Electronic Science and Technology of China, No. 4, Section 2, Jianshe North Road, Chenghua District, Chengdu City, Sichuan Province, 611731, China

964550721@qq.com

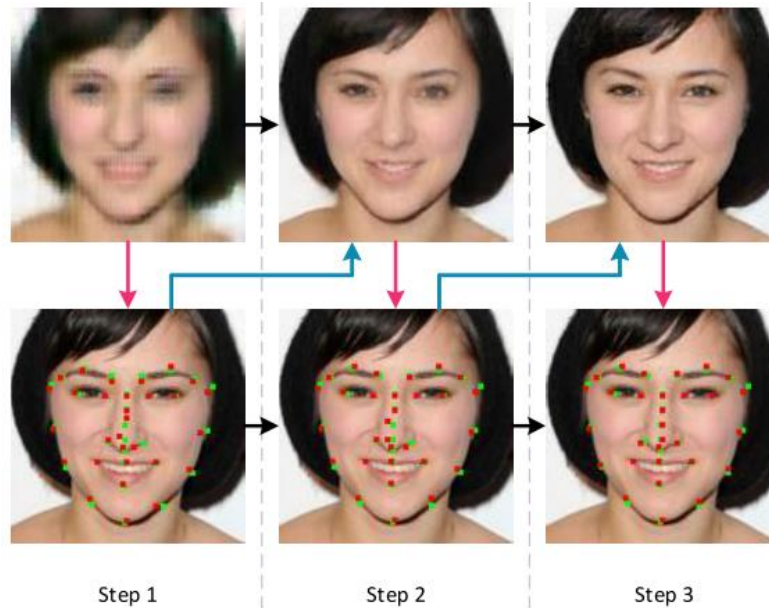
**Abstract.** As a very popular computer vision direction in recent years, image inpainting has developed very rapidly and produced many valuable results, but there are still many problems that need to be solved. This paper focuses on classic model research, related data sets, and evaluation standards. Through the analysis of specific examples, it is found that the effect of the existing model on the inpainting of some unique pictures is still unsatisfactory, and a solution is proposed to optimize the specificity of the training set by modifying the training set so that the model and the corrupt image are more correlated, which will result in more relative results.

**Keywords:** image inpainting, deep learning, training dataset.

## 1. Introduction

In today's society, pictures play an irreplaceable role in our daily lives. People take photos to share their lives with friends, use photos to convey information, and communicate with other people. Everyone, to some extent, has used photographs. However, due to the impact of the environment or various emergencies when taking photos, the content of the photos taken in many cases will not meet people's wishes to a certain extent, such as blurred photos due to water damage or an occluded target when taking photos. So photo processing has naturally become a very popular research direction in computer vision. From image quality to image content, image inpainting is mainly divided into two important branches: quality restoration and content restoration.

Quality restoration is to convert low-resolution images to high-resolution images or to eliminate the impact of the environment, such as heavy fog, on the quality of the photo. Through image noise reduction or neural networks, low-resolution photos can be repaired into high-resolution photos so that old photos can be revived, such as by using iterative collaboration in two recurrent networks to create high-resolution images [1]. Restoring images with a reversible encoder using a two-stage neural network [2]. There is even a well-trained network to improve the quality of comics [3].



**Figure 1.** The process of using two recurrent networks for quality restoration [1].

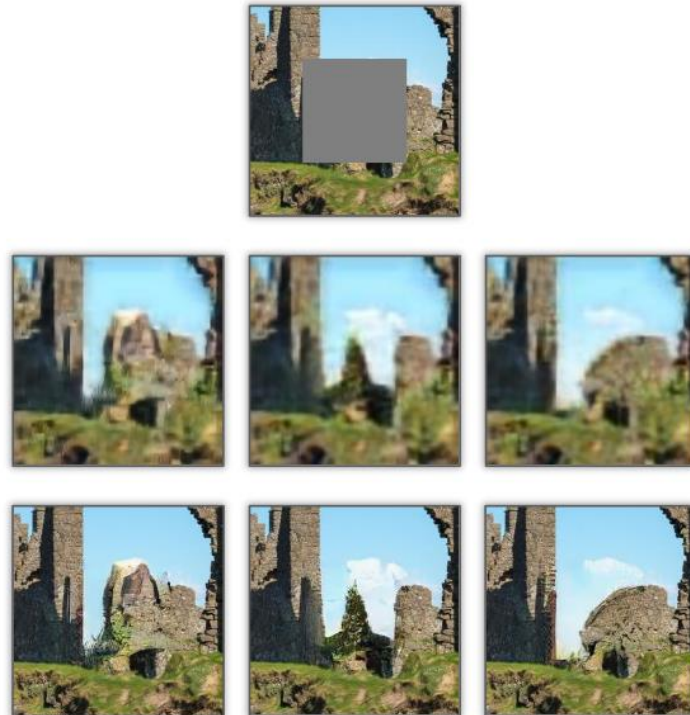
Content restoration is to restore the content of the picture, and there are two popular directions at present. The first is to help you remove unwanted objects in the photo, which can well eliminate the impact of the lack of image integrity caused by the absence of the object and reconstruct the integrity of the image, such as through a two-stage model that accomplishes the goal of human body de-occlusion through prior knowledge and attention-focused modules [4]. The second is that it completes the missing part of the picture based on the information in the remaining part, which is of great significance for precious photos that are accidentally damaged.

At present, there are mainly two mainstream image restoration techniques. One is to generate smooth, consistent patterns and colors. The patch-based repair method mainly infers the information of the missing part from the information surrounding the missing part and repairs the picture hole in the form of diffusion, but the generated picture does not have semantic content and may generate some incomprehensible images. Besides, because of the generality of the learning process, the repaired area may have an overly smooth boundary and weird lines in the picture structure. To resolve these problems, some networks adopt a two-stage generation mode [5], by first generating the structure of the picture and then filling the texture, and some networks go further, using a two-stage confrontation generation network, using prior knowledge to better outline the texture boundary [6]. Many subsequent patch-based repair networks are improvements and enhancements made to this two-stage structure.

The second is to use the method of deep learning to train the model through a large-scale training set, understand the semantics of the picture, and then repair the picture. At this stage, some methods have been used to improve the ability of the network. Such as using prior knowledge to improve the model's understanding ability and sharing it with the target image [7], or using relevant pictures of the background to repair the missing part [3]. Some networks adopt a two-stage generation mode [8], by first generating the structure of the picture and then filling in the texture, and some networks go further, using a two-stage confrontation generation network, using prior knowledge to better outline the texture boundary [9]. Contextual encoders from context-based unsupervised learning networks [10] and image inpainting learned through networks such as CNN, GAN, etc. Many subsequent patch-based repair networks are improvements and enhancements made to this two-stage structure.

The network based on deep learning has a great advantage over patches in inpainting pictures. It can learn the intrinsic semantics of pictures instead of inferring the missing parts of pictures only based on color blocks and structures. With the support of a large amount of data, the neural network

based on deep learning can generate images that are more in line with the original meaning of the original picture in terms of semantics instead of only considering the external structure of the picture.



**Figure 2.** Example of content restoration [5].

Despite the fact that this field has been developed for some time, there is still no good summary of some achievements and future development directions. This paper will provide a brief overview of some key milestones in image inpainting. Moreover, because each photo has its own uniqueness, such as the uniqueness of the face or the uniqueness of time and space, this paper will focus on the accuracy of image restoration, talking about how the model trained by big data can repair unique images. whether the restored image can show the original appearance of the missing part well.

The following paper will introduce the current state of existing work, introduce several commonly used data sets and evaluation criteria, analyze the known situation in conjunction with specific experiments, and propose some potential problems and solutions.

## **2. Related work**

In recent years, the majority of image inpainting has been developed using machine learning. From the initial restoration of presumed missing areas relying on color blocks [6], they gradually moved closer to image restoration associated with semantics and some extended adversarial networks for neural networks like GANs.

There are some typical models below.

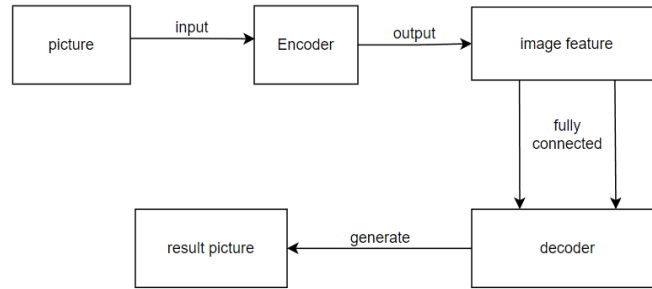
### *2.1. Context encoders*

This is an earlier model of applying deep learning to image repair. In this paper, the author proposes a trained convolutional neural network to repair defective images [10].

This model consists of an encoder and a decoder. The encoder will extract the features of its pictures according to the image, and the decoder will reconstruct the missing area pictures based on the features. Compared with other models, inference and repair are carried out by extracting information about missing parts. The model can handle more complex scenes where a large number of

areas are missing and some of the missing information cannot be inferred from the surrounding information, which means that it needs to have a better understanding of the picture scene.

Through the joint training mode, the model uses the method of minimizing reconstruction and confrontation generation so that it can better understand the overall content and predict the loss area.

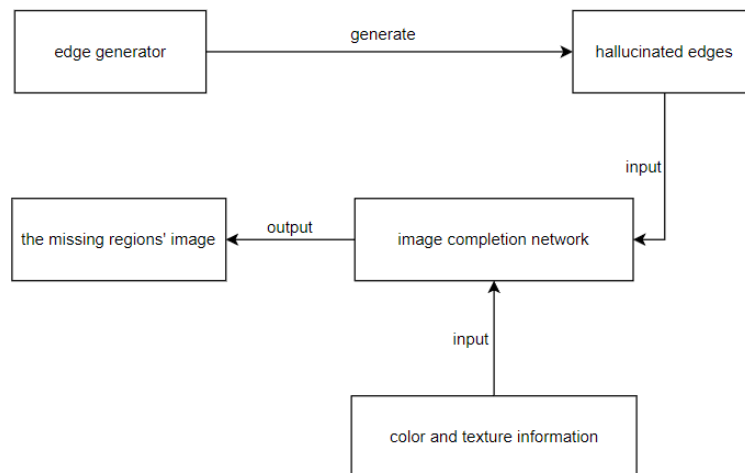


**Figure 3.** Working flowchart of context encoders.

## 2.2. EdgeConnect

This paper proposes a two-stage confrontation model called EdgeConnect, which is used to solve the problem of unreasonable structure in some other models' image inpainting and to improve the quality of image repair [9].

This model consists of an edge generator and an image completion network. The edge generator only pays attention to the edge structure of the missing part. By generating possible phantom edges and then generating corresponding textures through the image completion network to complete the missing areas, the two aspects of structure and texture can be well repaired.



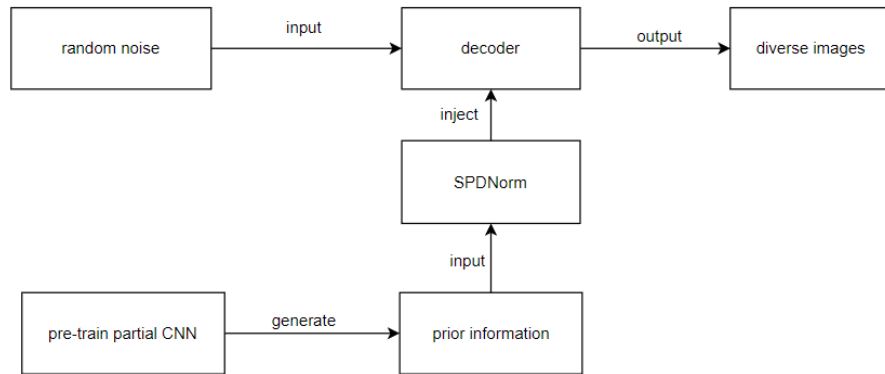
**Figure 4.** Working flowchart of EdgeConnect.

## 2.3. PD-GAN

This paper proposes a GAN for image inpainting that introduces random noise as a parameter to generate resulting images with different contents [11].

In order to solve the problem of the diversity reduction caused by the reconstruction loss, the model allows the GAN network to generate different images by inputting random noise, injecting prior information when decoding, and following when filling. The principle states that the closer the missing part is, the higher the degree of freedom there is to generate the corresponding image. To this

end, a spatially probabilistic diversity normalization is proposed to balance realism and diversity and to generate better-quality images.



**Figure 5.** Working flowchart of PD-GAN.

From the above models, it can be seen that, with the support of a large amount of data, the neural network based on deep learning can generate images that are more in line with the original meaning of the original picture in terms of semantics instead of only considering the external structure of the picture.

### 3. Dataset and evaluation

Whether you use CNN or GNN, the model involved in machine learning is inseparable from a useful data set to train the model. Most of the training sets at the present stage are trained by artificially destroying the existing dataset images, comparing the images generated by the model with the real images, and comparing the difference with the expected value. Later in this paper, we will introduce some frequently used datasets, analyze the composition of the data set, and train different models that are better for trade-offs.

#### 3.1. ImageNet

ImageNet was produced by Li Feifei's team at Stanford University. It is a network in the form of nodes. Under each noun node, there are hundreds of corresponding pictures. It contains 21,841 different categories and accommodates 14,197,122 pictures. As a general image recognition model, it contains the appearance of various objects in different states in various scenarios.

#### 3.2. Paris street view

Paris Street View is a data set containing 14,900 images, mainly composed of urban street views, surface buildings, and other images of various cities. It is useful for training and testing models for the restoration of buildings' images.

#### 3.3. CelebA dataset

CelebA Dataset, whose full name is CelebFaces Attributes Dataset, is a dataset containing 202,599 face images of 10,177 celebrities, which contains five marker points and 40 related attributes. It is very helpful when training and testing the model for face restoration.

#### 3.4. Places2 dataset

The Places2 Dataset is a dataset containing more than 100,000 images of 400 different scenes, ranging from natural scenes to rooms and bathrooms. At the same time, it will generate different labels and rank them according to their confidence level for the same scene. It is very helpful when training and testing the model to repair the scene image.

In order to evaluate the performance of an image inpainting model, researchers usually use results such as the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) for



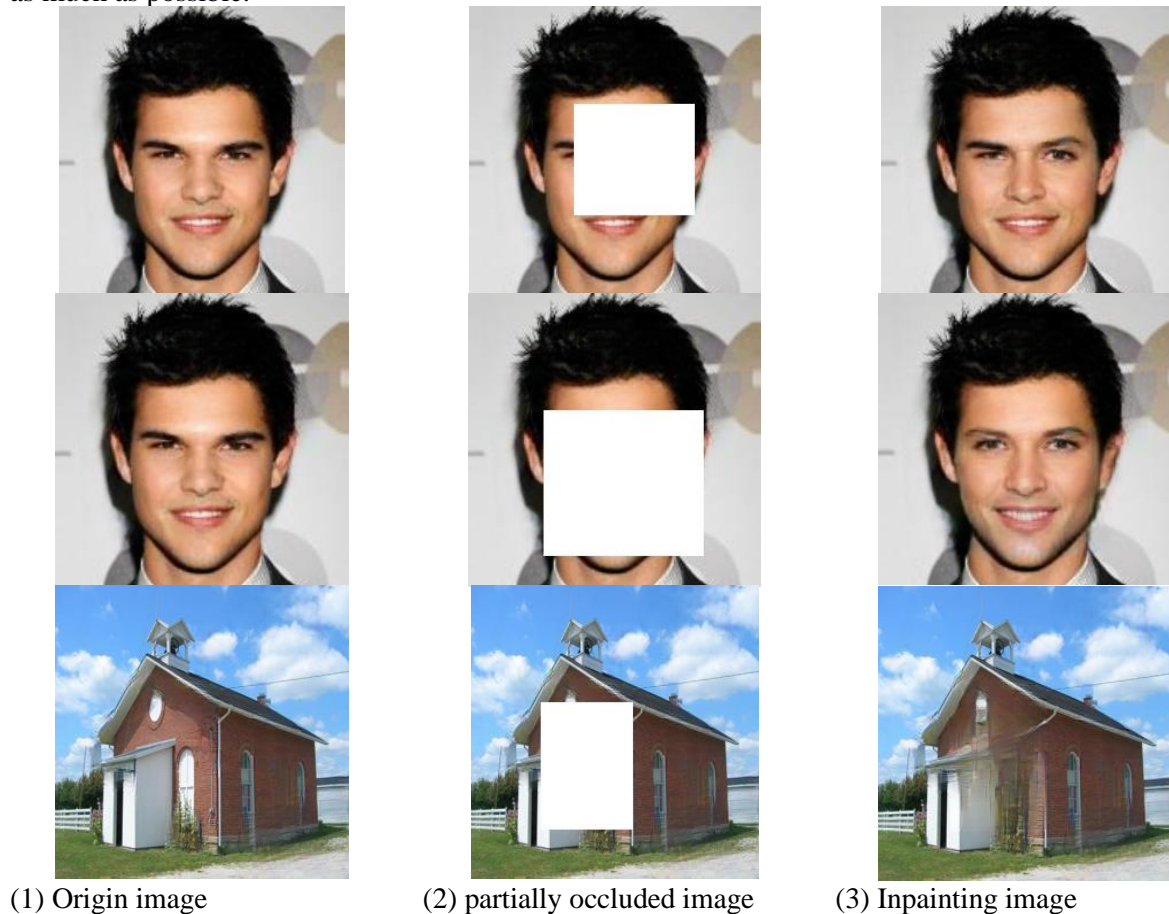
comparison in order to quantitatively analyze the results of the model [10][11], use PSNR to measure image quality and SSIM to measure image structure.

In order to verify the impact of certain functional components on the overall performance or the optimization effect of existing models, researchers usually use ablation studies for analysis. Each module is evaluated and analyzed individually in the form of control variables to confirm whether it is meaningful for improving the results. For example, soft and hard SPDNorm are used in PD-GAN and SPADE, a degraded version of SPNNorm, is used to prove that combining soft and hard SPDNorm can obtain the best repair effect [11].

#### 4. Challenge analysis

The information presented above provides an overview of current image inpainting research methods. In the forefront of image restoration, machine learning is an unavoidable topic. However, a large number of data sets used in machine learning training will naturally lead to a problem. The learning method of the model is to extract a large number of features from the data set and optimize each parameter of the model through the features so that it can obtain a wider range. Feature results are used to repair images, but in real life, images such as human faces are always unique. Can a model generated from big data be used to repair unique images well?

By artificially covering some areas of the face with inpainting by the EdgeConnect, it can be seen that although the model can restore the approximate image very well, the restoration of the right eye can clearly find the difference when compared with the original image, while the restoration of the left eye with a small amount of information can be better restored. If a larger part of the area is occluded, then the model can only use the average value of the training set to restore the appearance of the face as much as possible.



**Figure 6.** The test results inpainting by EdgeConnect.

If the corresponding picture is a building, the effect is similar. When a certain area is completely covered, the model can only use the data in the training set to make assumptions and repairs. Such effects often do not meet people's expectations.

For face restoration, in the absence of anchor points, researchers can optimize the data set by supplementing previous photos, relatives, and even pictures of people with similar characteristics. Training set: by training the model in a fixed direction, a better repair effect can be obtained. The same is true for buildings and so on. By changing the content of the training set, there is a better repair effect when dealing with large areas of continuously missing pictures.

## 5. Conclusion

This paper starts with several classic models and introduces several data sets and evaluation indicators commonly used in image repair. By using existing models for testing, it is confirmed that the existing models are insufficient for repairing unique pictures, and a solution that utilizes a special training set is proposed.

## References

- [1] Ma,C. , Jiang,Z. , Rao,Y. , Lu,J. ,& Zhou , J.(2020).Deep Face Super-Resolution With Iterative Collaboration Between Attentive Recovery and Landmark Estimation.2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- [2] Kim, Insoo, et al. "Quality-agnostic image recognition via invertible decoder." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.
- [3] Zhou, Yuqian, et al. "Transfill: Reference-guided image inpainting by merging multiple color and spatial transformations." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.
- [4] Ma, Cheng, et al. "Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
- [5] Peng, Jialun, et al. "Generating diverse structure for image inpainting with hierarchical VQ-VAE." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.
- [6] Yan, Zhaoyi, et al. "Shift-net: Image inpainting via deep feature rearrangement." *Proceedings of the European conference on computer vision (ECCV)*. 2018.
- [7] Liao, Liang, et al. "Image inpainting guided by coherence priors of semantics and textures." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.
- [8] Ren, Yurui, et al. "Structureflow: Image inpainting via structure-aware appearance flow." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.
- [9] Nazeri, Kamyar, et al. "Edgeconnect: Generative image inpainting with adversarial edge learning." *arXiv preprint arXiv:1901.00212* (2019).
- [10] Pathak, D. , et al. "Context Encoders: Feature Learning by Inpainting." *IEEE, 10.1109/CVPR.2016.278*. 2016.
- [11] Liu, H. , et al. "PD-GAN: Probabilistic Diverse GAN for Image Inpainting." (2021).