

Research advanced in deep learning-based image restoration

Chengyue Xie

Qilu University of Technology, Ji'nan, 250000, China

202003090021@stu.qlu.edu.cn

Abstract. In the computer vision community, which tries to address issues with image data change, loss, and missing that may arise in the process of image collecting, transmission, and preprocessing, image restoration has always been a hot study topic. Learning the contextual information of the visible regions in an image is the basic idea of image restoration algorithms, which perform noise removal, missing region repair and transformation on the regions to be restored to reproduce the original image content. Most of the early image restoration algorithms rely on the design of manual features, whose accuracy and generalization ability can not meet the practical application requirements. Due to the rapid advancement of CNN, in recent years, image repair accuracy and effect have made breakthroughs, and have been widely used in old photos, movies, videos and other repair scenes. In this paper, three restoration algorithms (sequence-based restoration, CNN-based restoration and GAN-based restoration) are comprehensively introduced while the research progress of image restoration based on depth learning is reviewed. Secondly, we quantitatively compare the performance of different repair algorithms and analyze their advantages and disadvantages. Finally, we summarize the challenges facing the area of image restoration research and evaluate its potential future directions.

Keywords: deep learning, image restoration, CNN, GAN.

1. Introduction

In the field of computer vision, which seeks to address issues with image data change, loss, and missing that may arise during the collecting, transmission, and preprocessing of images, image restoration has long been a hot topic for research. For example, some image content is contaminated, hazy, etc. The art of image restoration The basic idea of an artisan's restoration of a damaged artwork still depends on the residual semantics of the restored image after the loss and uses the limited information in the image to reproduce the missing content of the image. CNN has grown rapidly in recent years, image restoration based on depth learning has made breakthroughs in accuracy and generalization, and has important applications in the storage of old photos, facial repair and defect artifact repair.

The early image restoration methods were mainly based on the methods actually used by artists in the restoration process (first restoring the outline and then the details), such as the consistency of the boundaries and the continuity of the restored area as the first consideration in designing the restoration method. The partial differential-based image restoration algorithm is the most representative traditional image restoration method, which uses the idea of diffusion to expand the data values near the lost area and then diffuse them to the area to be restored. It can achieve good restoration results in

removing small scratches, small objects, etc. (such as holes in images). However, for applications such as repairing damaged images over large areas and removing large objects from images, the local difference-based image restoration algorithm is a localized image restoration based on feature information, and it is difficult to show better restoration results than repairing small areas in these applications. In terms of the underlying principle of restoration, when restoring artwork, especially when the missing area is large, the high-level feature semantic information of the object being restored is the first thing the artist considers (what particular environment the image is in, what information is missing, etc.). Then, based on the vast amount of information collected in the past, a better fix is made.

In recent times, because of the advances in convolutional neural networks (CNN), deep learning based image restoration has gradually become a new research hot spot. Each artificial neuron of CNN only responds to the surrounding units within a part of the coverage, which makes CNN network perform well in large image processing. In addition, CNN-based deep learning networks are able to extract abstract information from images at a higher feature level. The features obtained from these studies can also be further combined with the objective function, make the resulting network image is semantically more similar to the aim image. According to the idea of image restoration, deep learning based image restoration methods can be organized as the following three, sequence based, CNN, and GAN. Sequence-based restoration methods include image block-based restoration and diffusion-based restoration methods, which are more effective in restoring images with small damaged areas, but less effective when encountering large distortion of global information. GAN is a generative model, which can generate clearer and more realistic samples than other models. Xue et al. proposed the method named low gradient regularization [1], this method makes a reduction in the penalty of small gradients while penalizing non-zero gradients and the statistical features on the depth image are well represented. And for better naturalness of the restored image, a prior kind of optimization of the loss mask part of the image is proposed to make the restored image with better naturalness of the restoration [2]. In literature [3], a layered depth image (LDI) method is proposed to improve the spatial and temporal coherence for better restoration of images with artifacts. It can be found that the results and progress presented can be considered very good at smaller masks of images. Among the CNN-based methods, feedforward generative networks with content-aware layers [4-5] are proposed to make better restoration predictions. And among the GAN network based methods, a new network architecture (PEPSI) relying on GAN methods [6] was proposed to obtain superior performance and significantly reduce the operation time.

Based on the above three frameworks, in this paper, we first introduce the most representative image restoration algorithms, including their design ideas, basic pipelines, advantages and disadvantages. In Section 3, we quantify the performance of the different methods from different perspectives of the experiment. Finally, we summarize the existing research problems in the field of image restoration and discuss its future development direction.

2. Methods

2.1. Sequence-based restoration methods

2.1.1. Patch-based image restoration methods. A process known as image fusion, based on patch-based restoration, proposed by Soheil et al [7] for synthesizing the excess area between two input images with inconsistent colors, texture, and attribute structure between the two images gradually changes from one source image to the other. The approach consists of three modules: firstly adding a geometric luminosity transformation to enrich the search for image restoration patches, secondly the gradient integral of the image is added to the patch representation and the mask Poisson equation solver is used instead of the normal average color, and thirdly proposing new information about color and gradient based on L2/L0 mixing to perform smooth transitions at the constant expense of sharpness. This approach enables search patch-based image restoration solutions to address a wider range of image restoration problems, including inconsistencies in feature sources, fusion and repair of large images. Inconsistencies between the original image and the interior of the source image can be explained.

Some works repair images based on context-aware patch. Tijana et al. proposed a novel context-aware patch-based image repair method, which is a new top-down segmentation method that segments images into variable-sized blocks based on context, therefore, limit the search for candidate repair patches to regions of non-local images that have corresponding backgrounds [8]. This context-aware can be used in any image patch repair algorithm, and its main idea is to research patches that can be used to repair corrupted image regions according to contextual features., which speeds up the search for well-matched patches. Literature [9] proposed an efficient MRF based global restoration, which is a new optimization method that makes it appropriate for MRF-based image restore content with a large number of tags. This method is faster, consumes less and can process larger images compared to the content of the MRF. This method performs better in some small area loss and simple feature restoration, such as image scratches, text removal and other application scenarios.

2.1.1. Diffusion-based image restoration methods. The kernel of the divergence-based approach is to disperse all the semantics of the image edge pixels into the image for restoration and synthesize the new image texture content, which has greater limitations. Diffusion-based restoration algorithms are more difficult to obtain long-range image feature information and learn it correctly. With greater limitations in the reception of information around the lost region and the lack of understanding of higher depth features and semantic information of the entire image. As the diffusion range distance increases, the larger the lost region is, the less important structural and semantic features of the image are obtained. Therefore, this method is only suitable for restoration when the restoration area is small and fewer important features are lost. For those realistic scenes with complex texture features and larger areas to be repaired, hole repair is not effective for restoration.

2.2. Generative adversarial network (GAN) based approach

Yong-Goo Shin et al. proposed GAN-based parallel extended decoder path to the inpainting semantic network, where a contextual attention module called CAM is added to this network Structure [10]. As illustrated in Fig 1, this network composition has a complex stacked generative network that requires significant computational resources with convolutional operations and network parameters. To solve this problem, PEPSI module is constructed. The rough repair path produces the initial repair effect and trains the coding network to predict the features of the CAM. The restoration path is trained to obtain better image restoration quality through the reconstructed image features of the CAM.

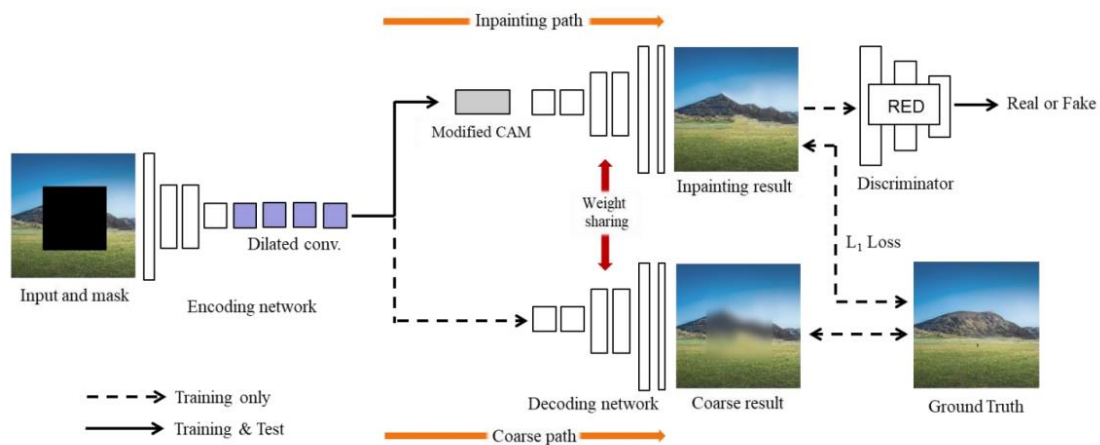


Figure 1. Network framework of PEPSI.

In [11], a CNN-based approach is utilized to simultaneously predict foreground separation masks and depth maps for conditional generative adversarial networks (GANs). In traditional depth image estimation the geometric scene of an image can be inferred from a single RGB image, but the masked scene region in the source image cannot be estimated, which restricts the application of depth

prediction in practice and in VR. This work extends the image restoration task to RGB-D data, using CNN to predict the depth data and mask data of an image, generated based on GAN on incomplete RGB images and on the condition of depth samples. Generative adversarial networks is widely used for various image restoration tasks, the most representative of which are algorithms based on deep generative models trained with undamaged and restored images for comparison, although these methods rely excessively on the training process, which leads to problems such as blurring of the generated images. So Chen et al, presented an improved approach based on GAN [12]. Starting with a low-resolution image and using the previous results to gradually complete the high-resolution image results in a clearer and more complete restoration with finer scale detail.

2.3. Convolutional neural network-based approach

Deepak et al. [13] utilize deep learning to address the challenge of image restoration, and proposed a contextual encoder. As shown in Fig.2, the contextual encoder learns not only the appearance but also the structural semantic of an image when restoring it using a contextual encoder. One of the adversarial losses is based on GAN [14].

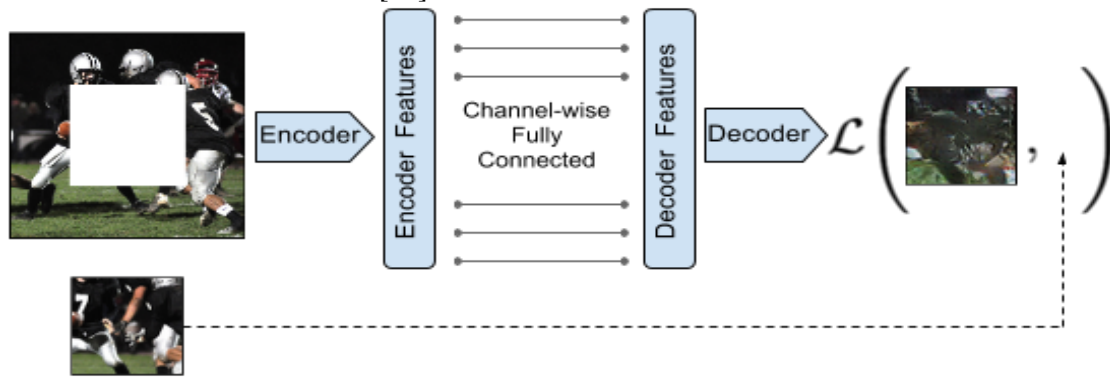


Figure 2. The structure of Context Encoder.

A multi-column generative network for picture restoration was suggested in the literature [15], which parallelizes the synthesis of various image components. This restoration network, which consists of a restoration generator, a discriminator for the generated results, and a VGG network for computing the losses, can be trained from beginning to end. The development of texture consistency benefits greatly from one of the ID-MRF regularizations. Deep learning has yielded positive results in the application of image restoration, but in [10] such methods the restored images often produce distorted and blurred textures that are inconsistent with the surrounding area, mainly due to the inadequate extraction of features by convolutional neural networks at a distance from the restored holes. As a result, in [16], the authors suggest a new depth-based generative model approach that uses references from nearby picture features to produce superior predictions. A contextual attention layer is proposed to focus on features at more distant spatial locations. Higher quality image restoration is accomplished in dataset tests for numerous scenarios.

3. Experiment and performance analysis

3.1. Common datasets

Currently, it is quite difficult to find enough professional datasets, and experimenters often choose a dataset that matches the current experimental direction when conducting image restoration experiments, and then use the input image to simulate the restored area by adding a randomly added mask. Some of the more common masks are rectangular masks and random masks with irregular shapes. When selecting a dataset, most use a public existing dataset, and different image categories determine the effectiveness of each method, the most commonly used datasets are Paris Street View, Places, foreground-aware dataset, segmentation dataset ImageNet etc. Some methods try to refer to RGB images or RGB-D images, for instance, in [1], RGB-D image patches are clustered and grouped

using a low-level mutual complementation based to enhance the depth image using the corresponding color image.

3.2. *Evaluation metric*

In objective quantitative analysis, the metric used for image restoration are mostly the same, and to evaluate the restoration effectiveness of the methods used, researchers have used MSE, PSNR, SSIM [17] etc., for example, in [8] PSNR and MSE quantitative analysis were used in The PSNR and SSIM are used to evaluate the robustness of the image after restoration of, for example, scratches, text, random mask blocks, etc., which can also reflect the intuition of the human eye. In addition, the quantitative analysis of the image after restoration is also related to the size of the area to be restored and to the diversity of the dataset.

3.3. *Performance comparison*

With the invocation of deep learning techniques, most of the tasks based on computer vision can be learned automatically with unsupervised features. The learning is done using CNN i.e. Convolutional Neural Network, which enables some computer vision tasks to be somewhat more robust in some aspects, and in the method [13]. However, with the special network structure design and ID-MRF regularization, the improvement is obvious, for example, in [15]. The objective quantitative data are significantly better than the method [13]. The method [10] The parallel decoding network is based on the GAN network structure with the addition of adaptive expansion convolutional layer, and the performance is better than that of the method from the table [12]. It is easy to see from the table that its performance is better than that of the method. Collectively, the two best-performing methods in the table are [10] and [15] in the table, the [15] The performance of the multi-column convolutional network is better than that of the method, especially on the data set CelebA. [15] The multi-column convolutional network does a better job of feature extraction and semantic consistency as well as texture repair, which also reflects the future research direction. It can be found from the table that different methods use different datasets, and most of them use CelebA, Places and Paris Street, which shows that the researchers' requirements for datasets are broad and diverse, like some new datasets such as Paris Street, Places, ImageNet has complex street scenes, and the accuracy of the restoration is unsurprisingly lower than other simple datasets, such as in [15] the same way as in the other simple datasets. The quantitative data reveals that the restoration results of different masks with different restoration scenes also vary greatly, for example [10] In the case of CelebA, different masks and different datasets are used, one can find that this dataset has a better performance in CelebA face images, The effect decreases in the case of random masks. It is found by the icons [10] and [15] the restoration effect is more significant compared to other methods, which shows that CNN and GAN-based restoration methods are still the mainstream and more effective image restoration solutions.

Table 1. Performance comparison of different restoration methods.

Method	Datasets	Type distortion	PSNR	SSIM
[10]	CelebA	Blocks	25.6	0.901
		Free-form mask	28.6	0.929
	Places	Blocks	21.5	0.839
		Free-form mask	25.2	0.889
[12]	CelebA	Center block	21.45	0.851
	Stanford		15.02	0.725
[13]	Cars	Random region	17.59	/
	Paris Street			
	Paris Street			
[15]	ImageNet	Blocks	22.43	0.894
	Places2		20.16	0.863
	CelebA		25.70	0.955
[16]	Places2	Blocks	18.91	/

4. Discussion

With the widespread use of CNN and GAN, a variety of deep learning-based network structures have been proposed, allowing the field of image restoration in computer vision to flourish, and after comparing some of the more typical approaches mentioned above, we can find that:

(1) For the network structure, the convolutional neural network-based image restoration method is still one of the most mainstream methods, and some researchers add other network structures such as RNN to try to improve the restoration effect, and the combination of different network structures is expected to be the mainstream research direction in the future. Finer and more realistic textures.

(2) In adversarial generative networks, GAN can effectively train existing datasets and apply the trained data models to image restoration, which has higher and clearer generation quality compared to other generative networks. However, the training of GAN models is relatively difficult and consumes relatively large computational resources. Researchers of GAN-based methods also tend to combine with other networks, such as GAN and VAE for image restoration, which is a better solution to the problem of training difficulties.

(3) It can be seen that the present increase in the depth and breadth of the network can be better adapted to the characteristics of the data, and when the number of layers and individual ku degrees of the network are expanded without limit, the size of the model and the training difficulty will then increase. Therefore, employing a thin and deep network structure is the more common method to minimize the number of parameters. In the method of solving chromatic aberrations and artifacts, the use of partial convolution [18] or gated convolution [19] is a better approach, which is based on the improvement of the convolution level, as opposed to increasing the depth and width.

(4) In the progressive restoration of images, the complexity of the network model is much higher than before, and the amount of computer computation and resources called is very large, so it is necessary to study how to better use the encoder to narrow down the number of settings, so as to better improve the training efficiency of the model, for example, in [10] For example, the PEPSI network structure incorporates a parallel decoding network, which can improve the training efficiency of the model.

It can be seen that the main problems of current image restoration methods are insufficient semantic details and structure after restoration, the quality of the restored image texture is more blurred, and the restored image will have more obvious artifacts and semantic inconsistencies. Within the network structure, the depth and width of the network are growing, the computational scale of the model will become larger and larger, and the use of computational resources will become higher and higher. The mainstream solution and research idea are to add more diverse network structures for

better extraction of semantic features, such as the combination of RNN and GAN, or to add a better attention mechanism to make the semantic more consistent and to make the overall image look more realistic. The depth and width of the network structure are adopted to reduce the number of settings needed for training and to increase the computer efficiency of training. It is especially important to reduce the training difficulty and increase the computational speed more effectively. Possible future research directions may include more novel ways of model training, including contrast learning, etc., to learn deeper features by comparing with high-quality source images.

5. Conclusion

This paper introduces the research progress of image restoration based on deep learning, and systematically introduces three main repair frameworks and corresponding representative repair algorithms: sequence based restoration, CNN based restoration and GAN based restoration. In addition, this paper further quantitatively compares the performance of different repair algorithms, and analyzes their advantages and disadvantages. Finally, the existing research problems and further development of image restoration are discussed.

References

- [1] H. Xue, S. Zhang and D. Cai, "Depth Image Inpainting: Improving Low Rank Matrix Completion With Low Gradient Regularization," in *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4311-4320, Sept. 2017, doi: 10.1109/TIP.2017.2718183.
- [2] M. Isogawa, D. Mikami, D. Iwai, H. Kimata and K. Sato, "Mask Optimization for Image Inpainting," in *IEEE Access*, vol. 6, pp. 69728-69741, 2018, doi: 10.1109/ACCESS.2018.2877401.
- [3] Muddala, S.M., Olsson, R. & Sjöström, M. Spatio-temporal consistent depth-image-based rendering using layered depth image and inpainting. *J Image Video Proc.* 2016, 9 (2016).
- [4] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu and T. S. Huang, "Generative Image Inpainting with Contextual Attention," 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5505-5514, doi: 10.1109/CVPR.2018.00577.
- [5] Dhama, H., Tateno, K., Laina, I., Navab, N., & Tombari, F. (2019). Peeking Behind Objects: Layered Depth Prediction from a Single Image. *pattern Recognit. lett.*, 125, 333-340.
- [6] Shin YG, Sagong MC, Yeo YJ, Kim SW, Ko SJ. PEPSI++: Fast and Lightweight Network for Image Inpainting. *IEEE Trans Neural Netw Learn Syst.* 2021 Jan;32(1): 252-265. doi: 10.1109/TNNLS.2020.2978501. epub 2021 Jan 4. PMID: 32203033.
- [7] S. Darabi, E. Shechtman, C. Barnes, D.B. Goldman, P. Sen, Image melding: Combining inconsistent images using patch-based synthesis, *ACM Trans. graphics (TOG)* 31(4) (2012) 82-81. 1, 2.
- [8] Ružić T, Pižurica A. Context-aware patch-based image inpainting using Markov random field modeling. *IEEE Trans Image Process.* 2015 Jan;24(1):444-56. doi: 10.1109/TIP.2014.2372479. epub 2014 Nov 20. pmid: 25420260.
- [9] T. Ružić, A. Pižurica, and W. Philips, "Neighborhood-consensus message passing as a framework for generalized iterated conditional expectations," *Pattern Recognit. Lett.* vol. 33, no. 3, pp. 309-318, Feb. 2012.
- [10] Y.-G. Shin, M.-C. Sagong, Y.-J. Yeo, S.-W. Kim and S.-J. Ko, "PEPSI++: Fast and Lightweight Network for Image Inpainting," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 252-265, Jan. 2021, doi: 10.1109/TNNLS.2020.2978501.
- [11] Dhama, Helisa & Tateno, Keisuke & Laina, Iro & Navab, Nassir & Tombari, Federico.(2018). Peeking Behind Objects: Layered Depth Prediction from a Single Image. *pattern Recognition Letters*. 125. 10.1016/j.patrec.2019.05.007.
- [12] Chen, Y., Hu, H. An Improved Method for Semantic Image Inpainting with GANs: Progressive Inpainting. *Neural Process Lett* 49, 1355-1367 (2019).
- [13] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell and A. A. Efros, "Context Encoders: Feature

- Learning by Inpainting," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2536-2544, doi: 10.1109/CVPR.2016.278.
- [14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. in NIPS, 2014.
- [15] Yi Wang et al. "Image Inpainting via Generative Multi-column Convolutional Neural Networks" neural information processing systems (2018): n. pag.
- [16] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu and T. S. Huang, "Generative Image Inpainting with Contextual Attention," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 5505-5514, doi: 10.1109/CVPR.2018.00577.
- [17] O. Elharrouss, A. Abbad, D. Moujahid, H. Tairi, Moving object detection zone using a block-based background model, IET Computer Vision 12 (1) (2017) 86- 94.
- [18] Guilin Liu, Fitsum A. Reda, Kevin J. Shih, Ting-Chun Wang, Andrew Tao, Bryan Catanzaro, Image inpainting for irregular holes using partial convolutions , in: Proc. ECCV, 2018. 3, 4, 6, 7, 85-100.
- [19] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, Thomas S. Huang, Free-form image inpainting with gated convolution, in: Proc. ICCV, 2019. 1, 3, 6, 7, 4471-4480.