International Journal of Engineering Research And Management (IJERM)
ISSN: 2349- 2058, Volume-08, Issue-10, October 2021

Generative Image Inpainting Through Edge Prediction Learning

Jialiang Yan

Abstract—**In the past few years, deep learning technology has significantly improved image restoration. However, many of these techniques cannot reconstruct reasonable structures because they are usually too smooth and/or blurry. This article develops a new image restoration method that can better reproduce the exquisite details of the filled area. We propose a two-stage adversarial GAN network that includes an edge generator, followed by an image completion network. The edge generator makes the edges of the missing areas (regular and irregular) of the image illusion, and the image completion network uses the illusion to fill the missing areas of the image Priori edge. We conducted an end-to-end evaluation of the publicly available datasets CelebA, Places2, and showed that it is superior in quantity and quality to the current state-of-the-art technology.**

Index Terms—**convolutional neural network,deep learning, generative adversarial network,image inpainting.**

I. Introduction

Image restoration, or image completion, includes filling in missing areas of the image. This is an important step in many image editing tasks. For example, it can be used to fill an image of a woman. Humans have an incredible ability to focus on visual inconsistencies. Therefore, the filled area must feel reasonable. In addition, the lack of fine structure in the filled area indicates something is wrong, especially when the rest of the image contains clear details. The work in this paper reminds us that many existing image inpainting techniques can produce excessively smooth and/or blurred areas that cannot reproduce fine details.

We divide image restoration into two stages: edge generation and image completion. Edge generation only focuses on creating hallucinated edges in missing areas. The image completion network uses the illusion edge and estimates the RGB pixel intensity of the missing area. Both stages follow an opposing framework [1] to ensure that the edges of the illusion and the RGB pixel intensity are visually consistent. Both networks contain loss based on deep features to achieve perceptually real results. The image structure is well represented in its edge mask, and we show that by adjusting the image inpainting network on the edges of the missing regions, better results can be produced. Obviously, we cannot access the edges of the missing area. Instead, we train an edge generator that produces hallucinogenic edges in these areas. Our “line first, color second” approach is partly inspired by our understanding of the way artists work [2]. We believe that edge recovery is easier than image completion. Our proposed model essentially eliminates the restoration of high-frequency and low-frequency information in the repair area.

We evaluated our proposed model on the standard data set CelebA [3] and Places2 [4]. We compared the performance of our model with current state-of-the-art solutions.

Manuscript received October 22, 2021

Jialiang Yan, School of Computer Science and Technology, Tiangong University,Tianjin,China

II. Procedure for Paper Submission

The repair technology proposed in this paper contains two completely different computer vision problems: image-to-edge and edge-to-image. There is a large amount of literature discussing the problem of "image to edge". For example, the Canny edge detector, an early scheme for constructing edge maps, has a history of about 30 years [5]. Dollár and Zitnick use structured learning [6] on random decision forests to predict local edge masks. Holistic Nested Edge Detection (HED) [7] is a fully convolutional network that learns edge information based on the importance of edge information as a feature of the entire image. In our work, we train on the edge map calculated using the Canny edge detector. The traditional "edge-to-image" approach usually follows a word-bag approach, where the image content is constructed through a set of predefined keywords. However, these methods cannot accurately construct fine-grained details, especially near object boundaries. Scribbler [8] is a learning-based model in which images are generated using line sketches as input. The results of their work are of artistic quality, where the color distribution of the generated results is guided by the use of colors in the input sketch. Isola et al. [9] proposed a conditional GAN framework [10], called pix2pix, for image-to-image translation. The solution can use the available edge information as a priori. CycleGAN [11] extended this framework and found a reverse mapping back to the original data distribution. This method produces better results because the goal is to learn the inverse of the forward mapping.

III. Edge Prediction Generator

We propose an image restoration network, including two stages: 1) edge generator, 2) image completion network (Figure 1). Both stages follow the opposite model [12], that is, each stage consists of a generator/discriminator pair. Let G_1 and D_1 be the generator and discriminator of the edge generator, and G_2 and D_2 be the generator and discriminator of the image completion network. In order to simplify the symbols, we will also use these symbols to represent the function mapping of their respective networks. Our generator follows a similar architecture to the method proposed by Johnson et al. [13]. Specifically, the generator consists of an encoder that downsamples twice, 6 remaining blocks [14], and a decoder that upsamples the image back to its original size. In the remaining layer, an expanded convolution with an expansion factor of 2 is used instead of regular convolution, so that a receptive field of 205 is generated at the final remaining block. For the discriminator, we use the 70×70 PatchGAN [15] architecture, which determines whether overlapping image blocks with a size of 70×70 are real. We use instance normalization in all layers of the network [16].

Let I_{gt} be ground truth images. Their edge map and grayscale counterpart will be denoted by C_{gt} and I_{gray}, We use the masked grayscale image $I_{gray} = I_{gray} \odot (1\text{-}M)$ as the input, its edge map $C_{gt} = C_{gt} \odot (1\text{-}M)$, and image mask M as a pre-condition.Here, denotes the Hadamard product. The generator predicts the edge map for the masked region. We use Cgtand Cpredconditioned on Igrayas inputs ofthe discriminator that predicts whether or not an edge map is real. The network is trained with an objective comprised of an adversarial loss.

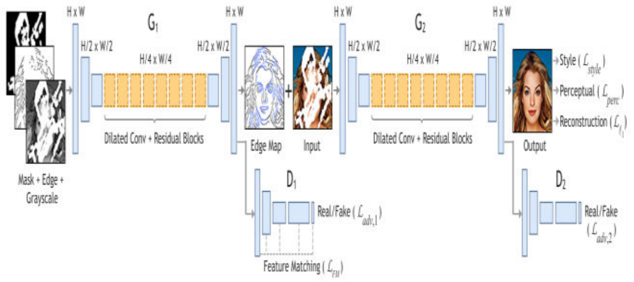

Figure 1: our network

IV. Image Completion Network

Our image completion network uses the incomplete color image $I_{gt} = I_{gt} \odot (1\text{-}M)$ as input, conditioned using a composite edge map Ccomp. The composite edge map is constructed by combining the background region of ground truth edges with generated edges in the corrupted region from the previous stage. $C_{comp} = C_{gt} \odot (1\text{-}M) + C_{pred} \odot M$. The network returns a color image Ipred, with missing regions filled in, that has the same resolution as the input image. We include the two losses proposed in commonlyknown as perceptual loss L_{perc} and style loss L_{style}. As the name suggests, L_{perc} penalizes results that are not perceptually similar to labels by defining a distance measure between activation maps of a pre-trained network.

V. Experiments

Our proposed model is evaluated on the datasets CelebA and Places2.Results are compared against the current state-of-the-art methods both qualitatively and quantitatively. Shown in Figure 2 and Figure 3. The image inpainting indicators are compared in Table 1 and Table 2.

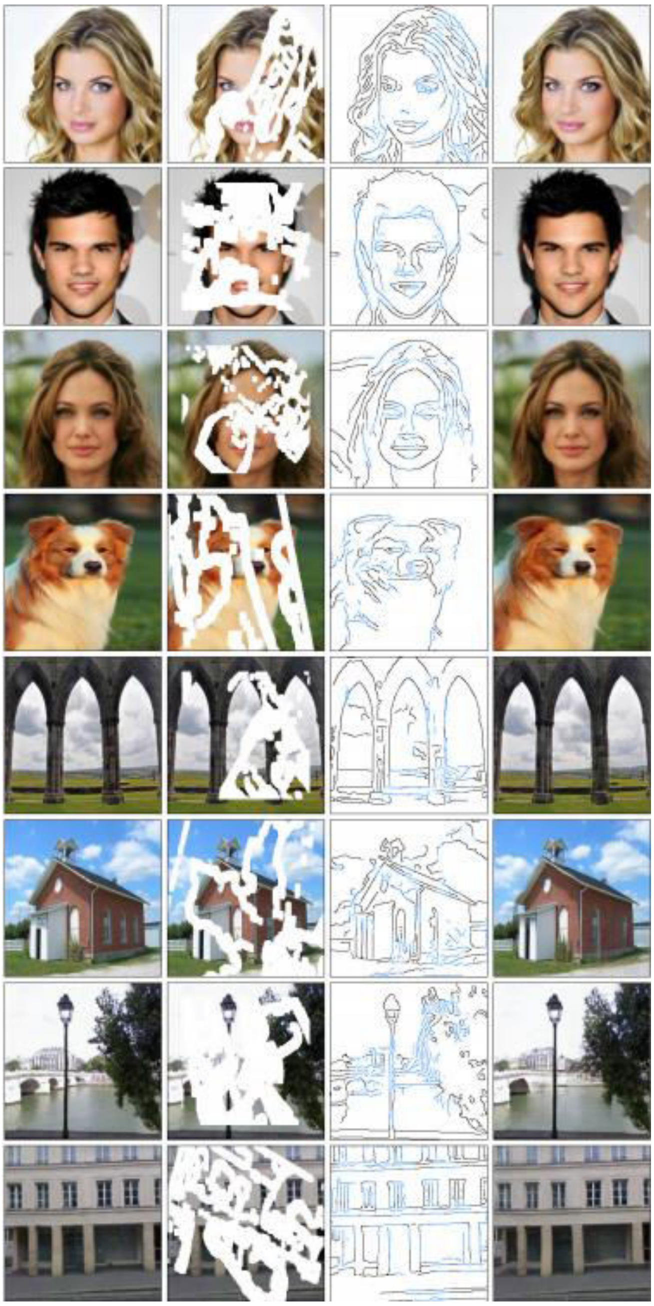

Figure 2 Our results show in celebA and Place2

Method	PSNR	SSIM
CA	20.65	0.818
GLCIC	21.34	0.814
PConv	21.42	0.819
Our	**21.75**	**0.823**

Table1 Evaluations in Places2

Method	PSNR	SSIM
CA	25.34	0.882
GLCIC	22.13	0.847
PConv	23.18	0.891
Our	**25.49**	**0.891**

Table2 Evaluations in celebA

VI. Conclusion

A new deep learning model is proposed for image restoration tasks. The network consists of an edge generator and an image completion network, both of which follow a confrontation model. We proved the important role of edge information in image restoration. Our method achieves state-of-the-art results on standard benchmarks and can handle images with multiple irregularly shaped missing regions.

References

[1] Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu,D. Warde-Farley, S. Ozair, A. Courville, and Y . Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014. 1, 3

[2] M. Eitz, J. Hays, and M. Alexa. How do humans sketch objects? ACM Transactions on graphics (TOG), 31(4):44 – 1,2012. 2

[3] Z. Liu, P . Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In Proceedings of International Conference on Computer Vision (ICCV), 2015. 2, 5

[4] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba.Places: A 10 million image database for scene recognition.IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017. 2, 5

[5] J. Canny. A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence,(6):679 – 698, 1986. 3

[6] S. Nowozin, C. H. Lampert, et al. Structured learning and prediction in computer vision. F oundations and Trends R ? inComputer Graphics and Vision, 6(3 – 4):185 – 365, 2011. 3

[7] S. Xie and Z. Tu. Holistically-nested edge detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1395 – 1403, 2015. 3, 7,12

[8] P . Sangkloy, J. Lu, C. Fang, F. Y u, and J.Hays.Scribbler:Controlling deep image synthesis with sketch and color. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), volume 2, 2017. 3

[9] P . Isola, J.-Y . Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 3, 11

[10] M. Mirza and S. Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014. 3

[11] J.-Y . Zhu, T. Park, P . Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In The IEEE International Conference on Computer Vision (ICCV), 2017. 3, 11

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu,D. Warde-Farley, S. Ozair, A. Courville, and Y . Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672 – 2680, 2014. 1, 3

[13] J. Johnson, A. Alahi, and L. Fei-Fei.Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision (ECCV), pages 694 – 711.Springer, 2016. 3, 4, 11

[14] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),pages 770 – 778, 2016. 3

[15] P . Isola, J.-Y . Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 3, 11

[16] D. Ulyanov, A. V edaldi, and V . Lempitsky. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 3