





# Inpainting with Sketch Reconstruction and Comprehensive Feature Selection

Siyuan Li<sup>1</sup> , Lu Lu<sup>1</sup>, Zhijing Li<sup>2</sup>, Kepeng Xu<sup>1</sup>, Matthieu Claisse<sup>3</sup>, Wenxin Yu<sup>1</sup> , Gang He<sup>1</sup>, Gang He<sup>4</sup>, Yibo Fan<sup>5</sup>, and Zhuo Yang<sup>6</sup>

<sup>1</sup> Southwest University of Science and Technology, Mianyang, China  
yuwenxin@swust.edu.cn

<sup>2</sup> Accenture Japan Ltd., Tokyo, Japan

<sup>3</sup> Graduate School in Computer Science and Mathematics Engineering France, Pau, France

<sup>4</sup> Xidian University, Xi'an, China

<sup>5</sup> State Key Laboratory of ASIC and System, Fudan University China, Shanghai, China

<sup>6</sup> Guangdong University of Technology, Guangzhou, China

**Abstract.** With the advent of the convolutional neural network, learning-based image inpainting approaches have received much attention, and most of these methods have been attracted by adversarial learning and various loss functions. This paper focuses on the enhancement of the generator model and guidance of structural information. Hence, a novel convolution block is proposed to comprehensively capture the context information among feature representations. The performance of the proposed method is evaluated on Place2 test dataset, which outperforms the current state-of-the-art inpainting approaches.

**Keywords:** Image inpainting · Deep learning · Feature selection · Edge guidance

## 1 Introduction

Image inpainting, also known as image completion, is the process of restoring the missing parts in a damaged image. Because the corrupted region only can be inferred through its neighborhood, it is still a challenging task to recover the details of the corrupted region to completely match the original image. From the tuition of human painting, the edge information in the filled region can guide the inpainting model to produce sharper results and away from the blurred edges, so as to improve inpainting quality and make filled region reasonable. Therefore, if we first paint the missing area with the fine structure or precise edges, the final results guided by the repaired sketch will be greatly improved.

In consideration of these pieces of knowledge, this paper proposes a two-stage, learning-based image inpainting approach with enhanced generator model and a new type of convolution block. Similar to one of the most advanced works [10],

the two stages of our proposed method are sketch reconstruction and image completion phases, and both of the two stages also introduce the generative adversarial network (GAN) [3].

The first stage, sketch reconstruction phase, aims to recover the gradient information in the missing area from corrupted sketch maps. In this paper, the Holistically-nested edge detection (HED) [12] is introduced to generate the sketches maps for training and testing. It is worth noting that Kamyar’s work [10] has conducted some experiments that also exploited the edge maps generated by HED, however, they assume the edge inpainting process is a relatively easy task and don’t put enough attention on the edge generator model. According to their experiments, their edge generator model fails to achieve better accuracy of HED prediction than applying Canny edge detector [1]. The sketch map generated by HED is considered as guiding information in this paper, and we enhance the sketch generator model by increasing the convolution layers to competent the sketch reconstruction task.

In the second phase, the goal of image completion networks is exploiting the repaired sketch maps and corrupted raw RGB image to color the sketch in the filled region. This paper applies a new convolution module named Comprehensive Feature Selection Block (CFS Block) in the second phase to comprehensively capture the saliency of context information among the feature maps. And the weight of the proposed module can be automatically updated during the training phase by the backpropagation.

Although our work is close to the combination of Kamyar’s [10] and Yu’s [13], our work proposed a novel convolution block to comprehensively select the features among the input features and convolved features in current module meanwhile enhancing the generator model through incorporating lots of popular technology to assure the accuracy of sketch and texture prediction.

This approach we presented is evaluated on the test dataset of Place2 [15]. Compared with those state-of-art inpainting approaches, the produced results quantitatively achieve great improvement. To sum up, our paper makes the following contributions:

- *The introduction of the edge sketch produced by HED, which better represents the rough shape of objects in images.*
- *A reinforced edge generator that can repair or hallucinate the sketch map through rest of sketch.*
- *An integrated inpainting generator with a novel convolution block – CFS Block.*

## 2 Approach

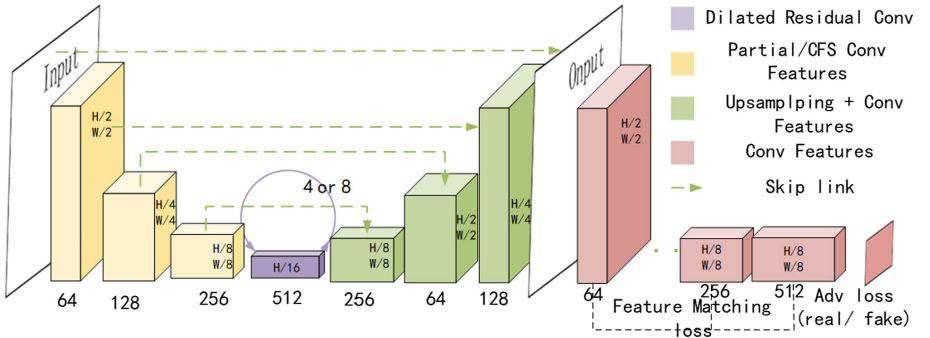
The proposed scheme divides the learning-based image inpainting process into two phases. At each stage, we create a generator model to produce target image meanwhile establishing a discriminator model that feedback to the generator to help produce high-quality results. In the first phase, the corrupted sketch and

grayscale image are concatenated as feature map, the features and the binary mask map (where 1 represent the non-damaged) region are fed into the generator with Partial Convolution, then the generator predicts a complete sketch as output. At the second stage, the restored sketch and damaged RGB image are considered as features together, they are inputted to a new generator built by CFS Blocks, aiming to get a complete RGB image.

In this section, we describe the detailed architecture of proposed networks in each phase and the detailed design of the Comprehensive Feature Selection Block (CFS Block) and briefly analyze what CFS Block actually doing.

## 2.1 Networks

The general architecture of the proposed model for each phase is illustrated as Fig. 1. As mentioned in [9], Spectral Normalization can further stabilize the training process through utilizing the maximum singular value of the weight matrix to reduce each weight matrix, which limits the Lipschitz constant of functions to 1. Thus we apply spectral normalization to the generators and discriminators in both phases.



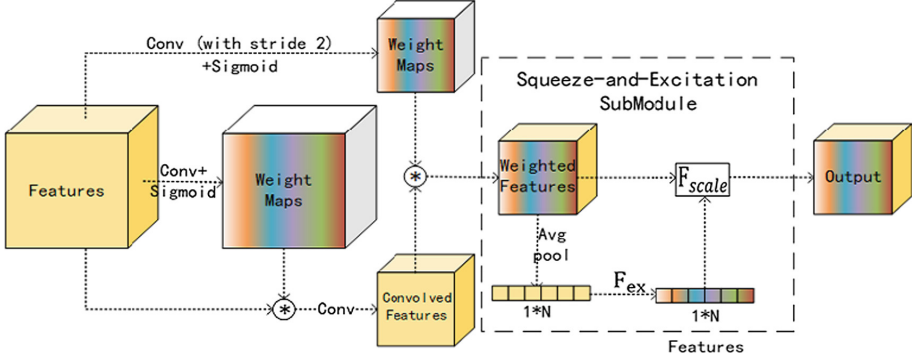
**Fig. 1.** The design of networks in one of stages.

The design of proposed discriminators in different phases are exactly the same. It's worth to note that the last two convolution layer with the same padding reduce the number of channels to 1 after increasing the number. In sketch reconstruction phase, the task of generator is relatively easy, thus we apply Partial Convolution [8] to each convolution layer in this generator instead of applying CFS Block, another difference is that we set 4 dilated residual block with Partial Convolution in the middle part of generator rather than 8 dilated residual blocks with CFS in image completion generator.

## 2.2 Comprehensive Feature Selection Block

The inspiration of the front part of CFS Block comes from classical Gated Recurrent Units (GRU) [2] and Gated Convolution [13], whileas the tail of the module

directly introduce the Squeeze-and-Excitation [4] block. The proposed integrated CFS Block aims to not only emphasize spatial relationships but also to further concern about channel correlation in feature selection process.



**Fig. 2.** The internal structure of Comprehensive Feature Selection Block.

As shown in Fig. 2, we adopt two convolution layer followed by sigmoid activation from the input features to calculate the weights of input features and the weights of convoluted features. Before the convolution, the input features are multiplied by their corresponding weights, and the convolved values also are multiplied by their weights. The slider-wise process is described as follows.

$$g = \sigma(W_g^T x + b_g) \quad (1)$$

$$G = \sigma(W_G^T x + b_G) \quad (2)$$

$$f = G \cdot \phi(W_f^T (g \cdot x) + b_f) \quad (3)$$

In which the  $g$  represents the gating value of the original feature in one of sliding windows and  $G$  is the gating values (weight) of the convolved feature map, the  $\sigma$  denotes sigmoid function and  $\phi$  represents the Leaky ReLU activation with the slope of 0.2. The  $f$  corresponds to the weighted feature computed by those foregoing units.

**Discussion About CFS Block.** All of the parameters in CFS Block (except  $r$ ) are learnable in the training process, this means that all the gating values (weight maps) in CFS Block can automatically be updated from data, it enables the generator to learn weight maps dynamically thus can select features both in input feature maps and the features in the next level. Because the importance of the damaged image, masks and sketches are obviously different, it is reasonable to set an additional gated convolution to select the input features rather than directly applying cross-level learning pattern [13]. Furthermore, the Squeeze-and-Excitation submodule is set to reinforce the ability of the model to notice some

essential channels in the computed features. However, the number of parameters in CFS Block is a bit large, this is the reason why the relatively easy task, the sketch inpainting task, adopts Partial Conv instead of CFS Block.

### 2.3 Loss Function

For convenience of formulizing, this paper denotes the generator and discriminator in sketch reconstruction as  $G_s$  and  $D_s$ , denotes the generator and discriminator in image completion as  $G_i$  and  $D_i$ , and the binary mask that labels the valid pixels as 1 (invalid pixels as 0) is expressed as  $M$ , the ground truth images in dataset is represented as  $\mathbf{I}_{gt}$ , damaged image can be represented as  $\dot{\mathbf{I}} = \mathbf{I}_{gt} \odot M$ , the complete sketch generated by HED is  $S_{gt}$ , the incomplete sketch is  $\dot{\mathbf{S}} = \mathbf{S}_{gt} \odot M$ , the composite sketch is described as  $\mathbf{S}_{comp} = \mathbf{S}_{pred} \odot (1 - M) + \mathbf{S}_{gt} \odot M$ .

The sketch generator predicts the complete sketch by considering the concatenation of damaged grayscale image and damaged sketch as the features, meanwhile inputting the mask  $M$  for Partial Conv. While in the image completion generator, it concatenates the inpainted sketch, damaged image and mask as the input feature.

$$\mathbf{S}_{pred} = G_s(\left[\dot{\mathbf{I}}_{gray}, \dot{\mathbf{S}}\right], M) \quad (4)$$

$$\mathbf{I}_{pred} = G_i\left(\left[\dot{\mathbf{I}}, \mathbf{S}_{comp}, M\right]\right) \quad (5)$$

On account of the feature-matching loss [11] is introduced as one of loss terms for further optimizing the generator, the total loss for sketch generator is interpreted as Eq. (6).

$$\min_{G_s} \max_{D_s} \mathcal{L}_{G_s} = \min_{G_s} \left( 0.1 \max_{D_s} (\mathcal{L}_{D_s}) + 10\mathcal{L}_{FM} \right) \quad (6)$$

The task of the discriminator is to distinguish whether the input sketch in discriminator belongs to the grayscale image of the corresponding ground truth image, this paper adopts the hinge loss as the target function of discriminator, which train the discriminator more strictly.

$$\begin{aligned} \mathcal{L}_{D_s} = & \mathbb{E}_{\mathbf{S}_{gt}} [\psi(1 - D_s(\mathbf{S}_{gt}, \mathbf{I}_{gray}))] \\ & + \mathbb{E}_{\mathbf{S}_{pred}} [\psi(1 + D_s(\mathbf{S}_{pred}, \mathbf{I}_{gray}))] \end{aligned} \quad (7)$$

Image completion generator adopts L1 distance and the perceptual loss ( $\mathcal{L}_{style}$  and  $\mathcal{L}_{perc}$ ) [6]. It enables the model to learn the high-level representation and remove the checkboard artifacts from the predicted image, the total loss of completion generator is express as

$$\mathcal{L}_{G_i} = \mathcal{L}_{\ell_1} + 0.1\mathcal{L}_{D_i} + 300\mathcal{L}_{perc} + 300\mathcal{L}_{style} + 10\mathcal{L}_{FM} \quad (8)$$

where  $\mathcal{L}_{D_i}$  is similar to  $\mathcal{L}_{D_s}$  however the input of  $\mathcal{L}_{D_i}$  has a slight difference, it just input the  $\mathbf{I}_{gt}$  and  $\mathbf{I}_{pred}$  without any grayscale images.

### 3 Experiments

All of the experiments in this paper are conducted in the dataset of Place2 [15], the sketches are inferred from RGB images through HED [12] approach. The irregular mask dataset used in this paper comes from the work of Liu [8]. With these data groups (image sketch mask), We train on single NVIDIA 1080TI with a batch size of 6 until the generators converge using Adam optimizer [7].

**Table 1.** Comparison of quantitative results, these data are taken from this paper [10].

Mask		CA [14]	GLCIC [5]	PConv [8]	EdgeCnt [10]	Ours
10–20%	PSNR	24.36	23.49	28.02	27.95	<b>30.85</b>
	SSIM	0.893	0.862	0.869	0.920	<b>0.951</b>
20–30%	PSNR	21.19	20.45	24.90	24.92	<b>26.87</b>
	SSIM	0.815	0.771	0.777	0.861	<b>0.900</b>
30–40%	PSNR	19.13	18.50	22.45	22.84	<b>24.17</b>
	SSIM	0.739	0.686	0.685	0.799	<b>0.841</b>
>=40	PSNR	17.75	17.17	20.86	21.16	<b>21.74</b>
	SSIM	0.662	0.603	0.589	0.731	<b>0.7695</b>

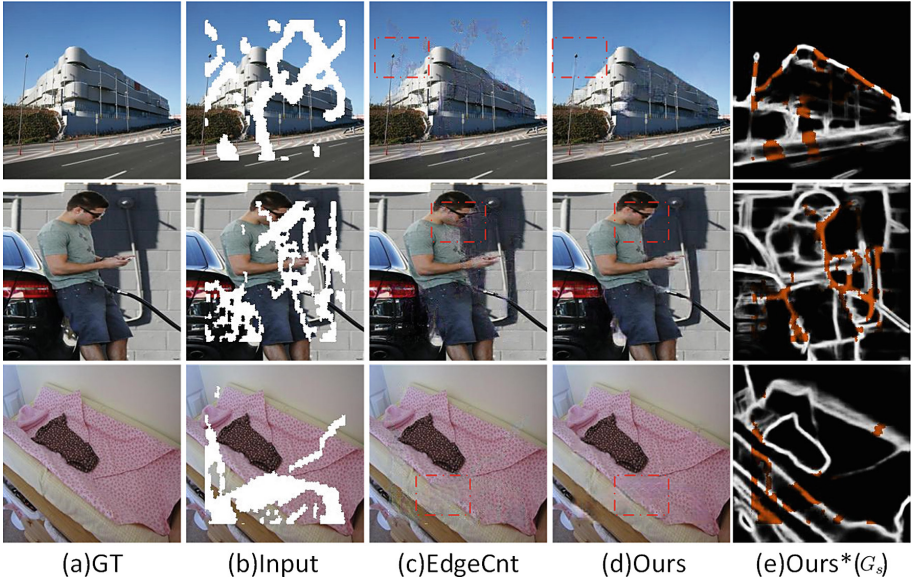
#### 3.1 Quantitative Results

The quantitative results in the test dataset of Place2 are shown in Table 1, this table also shows some results that produced popular inpainting methods in comparison. In the case of all different ratios of the damaged region, the table indicates that our results outperform the others in PSNR (Peak Signal-to-Noise) and SSIM (Structural Similarity) metric, especially in the case of small masks. In addition, the unique sketch prediction task achieved 77% accuracy (better than [10]), it indicates the quantitative improvement is benefited from the enhanced sketch generator.

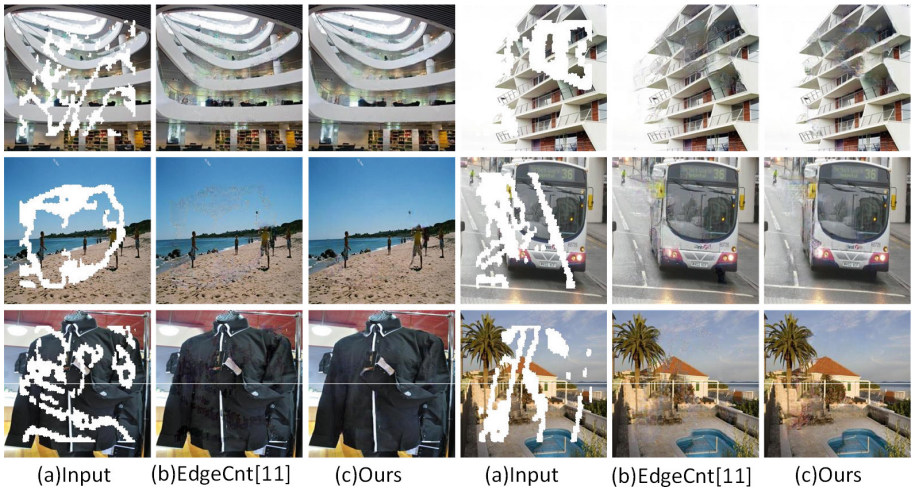
#### 3.2 Qualitative Results

The EdgeConnect approach [10] has recently shown surprising advancement in image inpainting, therefore we compare our result with this state-of-art work (Figs. 3 and 4). The red box in the figure indicates that the proposed approach with CFS Block can produce a clearer color image with the more obvious edges than EdgeConnect, the 5th column show the generated sketch in the proposed approach, the orange lines in these images represent the inpainted line by generator, it demonstrates the sketch generator with Partial Conv already can handle the sketch reconstruction task.





**Fig. 3.** The qualitative comparison of results. (a) Ground Truth image. (b) Corrupted image. (c) EdgeConnect [10]. (d) Ours. (e) Restored sketches of Ours. (Color figure online)



**Fig. 4.** The additional qualitative comparison.

## 4 Conclusions

This paper proposed a two-stage image inpainting approach with a novel feature selection mechanism – CFS Block, and proves that the enhanced sketch generator and the proposed comprehensive feature selection mechanism can significantly improve the inpainting results. The qualitative comparisons show that the proposed approach produces visually more pleasing results, and the objective evaluations in various sizes of masks demonstrate the superiority of the proposed approach.

**Acknowledgements.** This research was supported by Sichuan Provincial Science and Technology Department (No. 2018GZ0517, 2019YFS0146, 2019YFS0155), State Key Laboratory of ASIC & System (No. 2018KF003), Science and Technology Planning Project of Guangdong Province (No. 2017B010110007), National Natural Science Foundation of China (No. 61907009), Natural Science Foundation of Guangdong Province (No. 2018A030313802), Science and Technology Planning Project of Guangdong Province (No. 2017B010110007 and 2017B010110015). Supported by Postgraduate Innovation Fund Project by Southwest University of Science and Technology (19ycx0050).

## References

1. Canny, J.: A computational approach to edge detection. In: Readings in computer vision, pp. 184–203. Elsevier (1987)
2. Cho, K., van Merriënboer, B., Gülçehre, Ç., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. CoRR (2014). <http://arxiv.org/abs/1406.1078>
3. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
4. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. CoRR (2017). <http://arxiv.org/abs/1709.01507>
5. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Trans. Graph. (ToG) **36**(4), 107 (2017)
6. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 694–711. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46475-6\\_43](https://doi.org/10.1007/978-3-319-46475-6_43)
7. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
8. Liu, G., Reda, F.A., Shih, K.J., Wang, T.C., Tao, A., Catanzaro, B.: Image inpainting for irregular holes using partial convolutions. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 85–100 (2018)
9. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral normalization for generative adversarial networks. arXiv preprint [arXiv:1802.05957](https://arxiv.org/abs/1802.05957) (2018)
10. Nazeri, K., Ng, E., Joseph, T., Qureshi, F., Ebrahimi, M.: Edgeconnect: generative image inpainting with adversarial edge learning. arXiv preprint (2019)
11. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional GANs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8798–8807 (2018)



12. Xie, S., Tu, Z.: Holistically-nested edge detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1395–1403 (2015)
13. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Free-form image inpainting with gated convolution. arXiv preprint [arXiv:1806.03589](https://arxiv.org/abs/1806.03589) (2018)
14. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018
15. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: A 10 million image database for scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(6), 1452–1464 (2017)