10 Papers You Must Read for Deep Image Inpainting

Chu-Tak Li Nov 30, 2020 · 12 min read

Hello! This post can be regarded as a revision of deep image inpainting for my old friends and introductory deep image inpainting for newcomers. I have written more than 10 posts related to deep learning approaches for image inpainting. It's time to briefly review what we have learned and also provide a highway for newcomers to join us for fun!

What is Image Inpainting?

Figure 1. Examples of Image Inpainting Applications. Image by Jiahui Yu et al. from their paper, DeepFill v2 [13]

as unwanted object(s) removal and interactive image editing are shown in Figure 1. There are also many possible applications as long as you can imagine.

Terminology

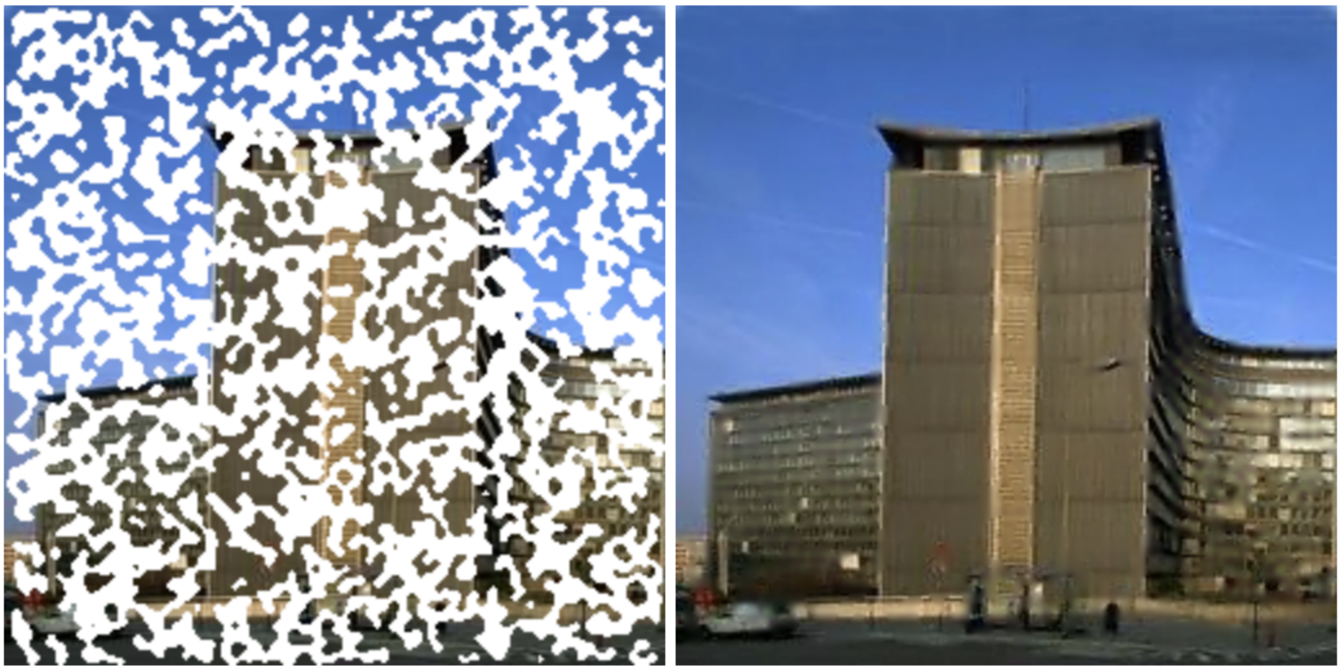


Figure 2. An example of a masked input image (left) and a completed image (right). Image by the author extracted from his github page

Given a corrupted/masked input image as shown in Figure 2 (left), we usually define **i) invalid/missing/hole pixels as the pixels located at the region(s) to be filled**; **ii) valid/remaining/ground truth pixels as the pixels we can use to help filling in the missing pixels**. Note that we can directly copy the valid pixels and paste them on the filled image at their corresponding locations.

Introduction

To fill in an image with some missing parts, the simplest way is to copy-and-paste. The core idea is to **first search** for the most similar image patches from the remaining pixels of the image itself or a large dataset with millions of images, **then directly paste** the patches on the missing parts. **However, the search algorithm could be time-consuming and it involves hand-crafted distance measure metrics**. Its generalization and efficiency still have plenty of room for improvement.

missing pixels in an image with good global consistency and local fine textures. We will focus on 10 famous deep learning-based inpainting approaches in this post. I am sure that you can understand other inpainting papers/works after you realizing these 10 approaches. Let's Go:)

Context Encoder (1st GAN-based inpainting, 2016)

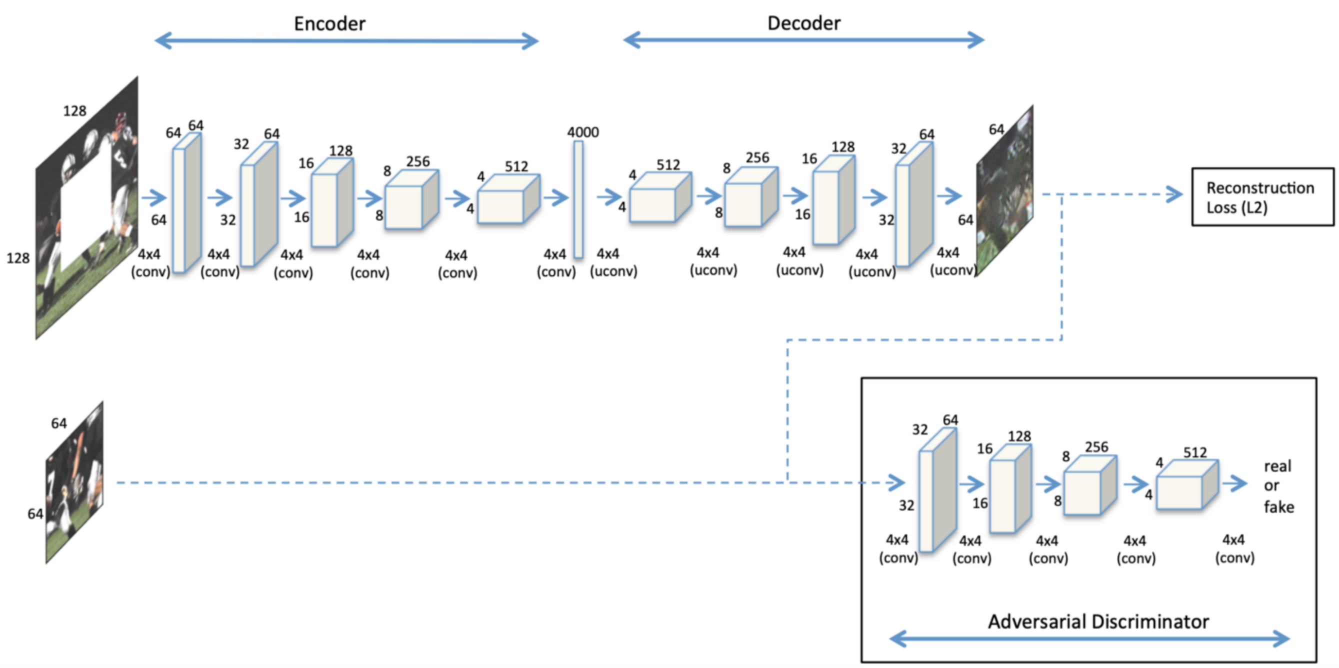


Figure 3. Network architecture of the proposed Contextu Encoder (CE). Image by Deepak Pathak et al. from their paper [1]

Context Encoder (CE, 2016) [1] is the first Generative Adversarial Networks (GANs) [2] based inpainting algorithm. This paper spots some useful basic concepts for the task of image inpainting. The term **"Context" relates to the understanding of the entire image itself** and the core idea of CE is **Channel-wise Fully Connected Layer** (the middle layer at the network as shown in Figure 3). Similar to the standard Fully Connected Layer, the main point is that all the feature locations at the previous layer would contribute to each feature location at the current layer. By doing so, the network can learn the relationship between all the feature locations and a deeper semantic understanding of the entire image can be obtained. CE has been regarded as a baseline and you can learn more about it from my previous post [here].

MSNPS (Enhanced Context Encoder, 2016)

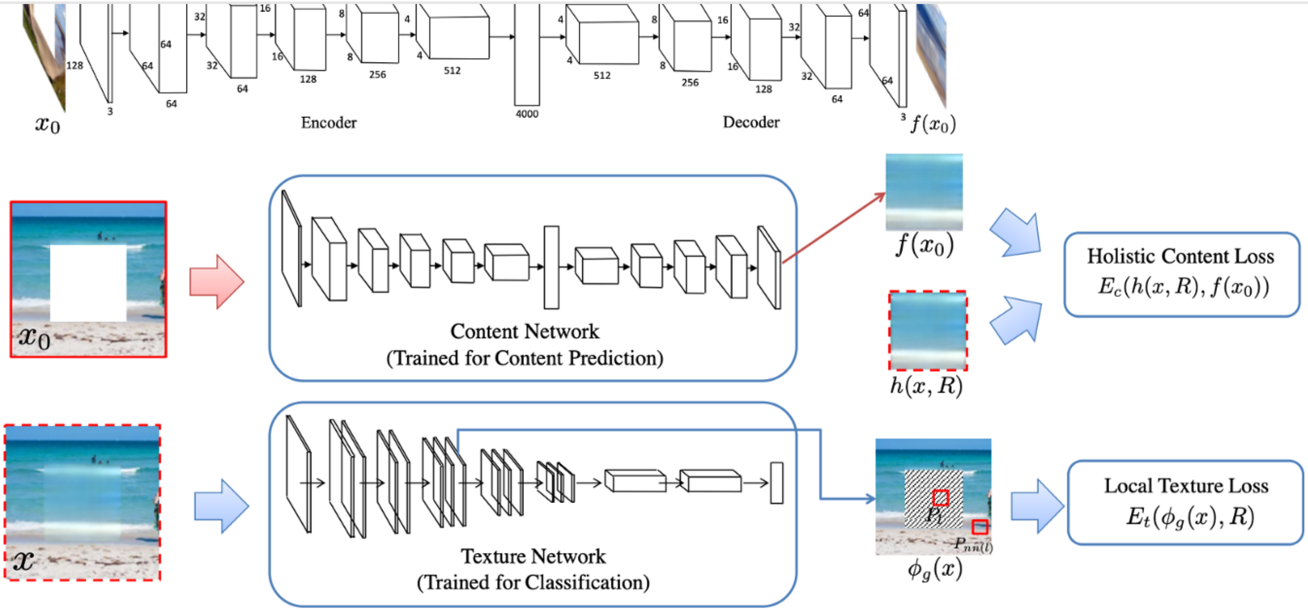


Figure 4. Overview of the content network (modified CE [1]) and the texture network (VGG-19). Image by Chao Yang et al. from their paper [3]

Multi-Scale Neural Patch Synthesis (MSNPS, 2016) [3] can be regarded as an enhanced version of CE [1]. The authors of this paper employed a modified CE to predict the missing parts in an image and a texture network to decorate the prediction about the missing parts to improve the visual quality of the filled images. **The idea of the texture network is from the task of style transfer. We would like to transfer the style of the most similar valid pixels to the generated pixels to enhance the local texture details.** I would say that this work is an early version of the two-stage coarse-to-fine network structure. The first content network (i.e. CE here) is responsible for the reconstruction/prediction of the missing parts while the second network (i.e. texture network here) is responsible for the refinement of the filled parts.

Apart from the typical pixel-wise reconstruction loss (i.e. L1 loss) and the standard Adversarial loss, **the concept of texture loss proposed in this paper plays an important role in later inpainting papers.** Actually, texture loss is related to perceptual loss and style loss that are widely used in many image generation tasks such as neural style transfer. To know more about this paper, you may refer to my previous post [here].

GLCIC (A Milestone in Deep Image Inpainting, 2017)

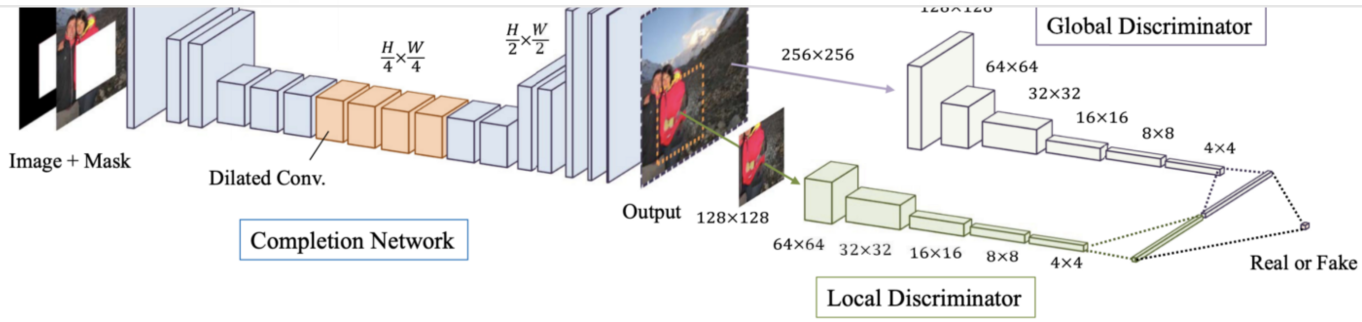
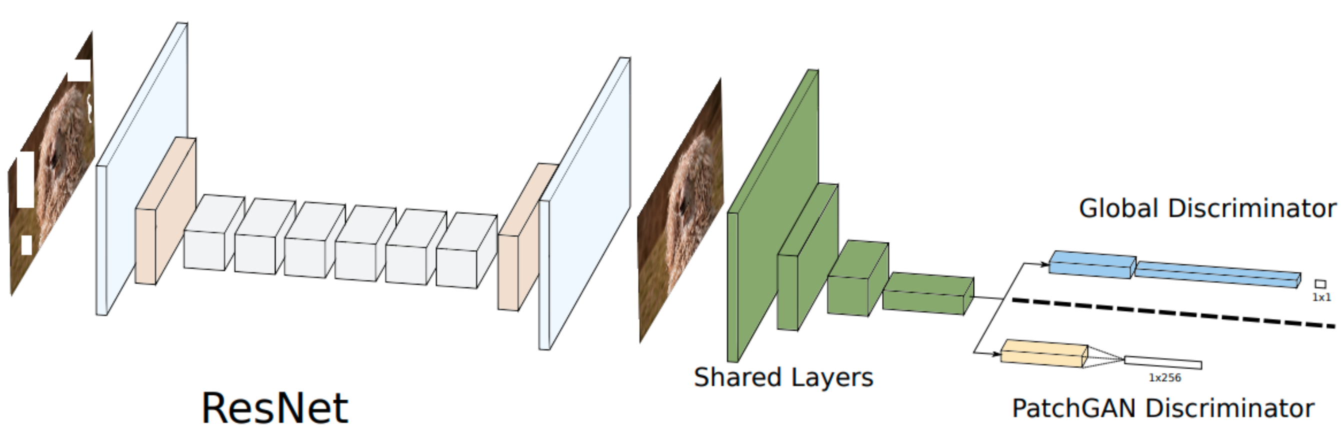


Figure 5. Overview of the proposed model which consists of a completion network (Generator network), a global discriminator, and a local discriminator. Image by Satoshi Iizuka et al. from their paper [4]

Globally and Locally Consistent Image Completion (GLCIC, 2017) [4] is a milestone in deep image inpainting as it defines the **Fully Convolution Network with Dilated Convolutions** for deep image inpainting and actually this is a typical network structure for deep image inpainting. By using Dilated convolutions, the network is able to understand the context of an image without employing expensive fully connected layers and hence it can handle images of different sizes.

Apart from the fully convolution network with dilated convolutions, two discriminators at two scales were also trained together with the generator network. **A global discriminator looks at the whole image while a local discriminator looks at the filled centre hole**. With both the global and local discriminators, the filled image would have better global and local consistency. Note that many later inpainting papers follow this multi-scale discriminator design. If you are interested in this paper, please visit my previous post [here] for more details.

Patch-based Image Inpainting with GANs (A Variant of GLCIC, 2018)



Open in app

Figure 6. The proposed Generative ResNet architecture and PGGAN discriminator. Image by Ugur Demir and Gozde Unal from their paper [5]

Patch-based Image Inpainting with GANs [5] can be regarded as a variant of GLCIC [4]. Simply speaking, two advanced concepts namely **residual learning** [6] and **PatchGAN** [7] were embedded in GLCIC to further boost its inpainting performance. **The authors of this paper combined residual connection and dilated convolution to form a dilated residual block.** The traditional GAN discriminator was also replaced by the PatchGAN discriminator to encourage better local texture details and global structure consistency.

The core difference between traditional GAN discriminator and PatchGAN discriminator is that traditional GAN discriminator only gives a single predicted label (from 0 to 1) to indicate the realness of the input while PatchGAN discriminator gives a matrix of predicted labels (also from 0 to 1) to indicate the realness of each local region of the input. Note that each element of the matrix represents a local region of the input. You can also have a review of the residual learning and PatchGAN by visiting my previous post [here].

Shift-Net (Deep Learning-based "Copy-and-Paste", 2018)

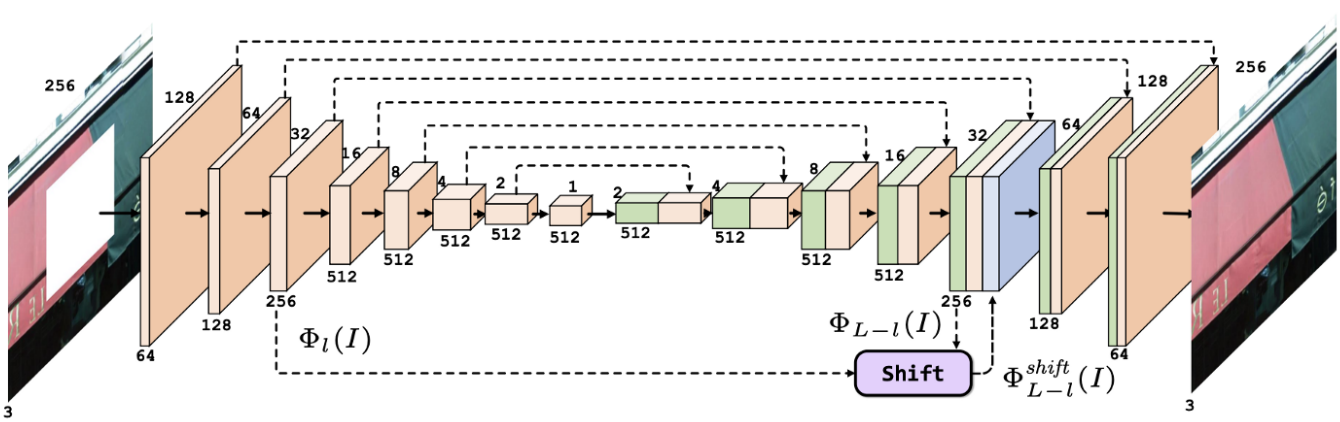


Figure 7. Network architecture of Shift-Net. The shift-connection layer is added at resolution of 32×32. Image by Zhaoyi Yan et al. from their paper [8]

Shift-Net [8] takes both the advantages of modern data-driven CNNs and the conventional "Copy-and-Paste" method in the form of Deep Feature Rearrangement by

First, the authors proposed guidance loss that encourages the decoded features of the missing parts (given a masked image) to be close to the encoded features of the missing parts (given a good-conditioned image). As a result, the decoding process is able to fill in the missing parts with a reasonable estimation of the missing parts in the good-conditioned image (i.e. the ground truth of the missing parts).

Second, the proposed shift-connected layer with shift operation allows the network to effectively borrow the information given by the nearest neighbours outside the missing parts to refine both the global semantic structure and local texture details of the generated parts. Simply speaking, we are providing suitable references to refine our estimation. I think it's good for readers who are interested in image inpainting to consolidate the ideas proposed in this paper. I highly recommend you to read my previous post [here] for details.

DeepFill v1 (A Breakthrough in Image Inpainting, 2018)

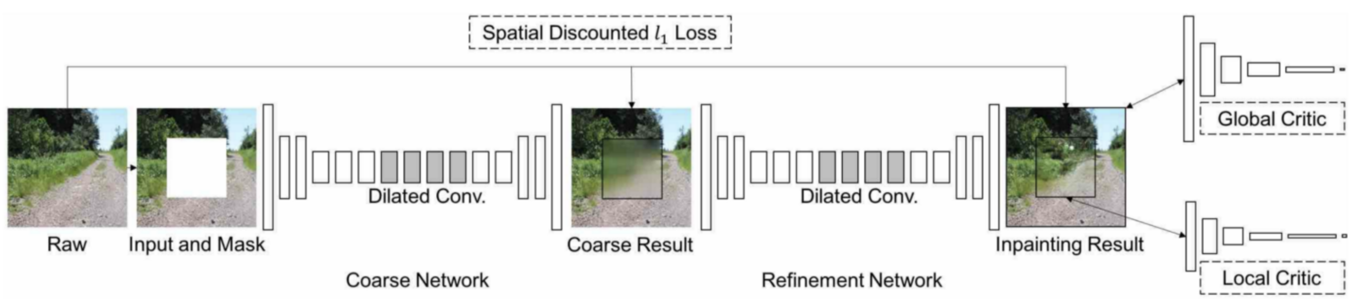


Figure 8. Network architecture of the proposed inpainting framework. Image by Jiahui Yu et al. from their paper [9]

Generative Image Inpainting with Contextual Attention (CA, 2018) (also called DeepFill v1 or CA) [9] can be regarded as an enhanced version or a variant of Shift-Net [8]. The authors further develop the idea of copy-and-paste and propose a contextual attention layer which is differentiable and fully convolutional.

Similar to the shift-connection layer in [8], **by matching the generated features inside the missing hole and the features outside the missing hole, we can know the contributions of all the features outside the missing hole to each location inside the missing hole. Hence, the combination of all the features outside can be used to**

Open in app

not differentiable), the CA layer in this paper employs a soft assignment (is differentiable) in which all the features have their own weights to indicate their contributions to each location inside the missing hole. To know more about the Contextual Attention, please visit my previous post [here] and look for more concrete examples.

GMCNN (Multi-branch CNNs for Image Inpainting, 2018)

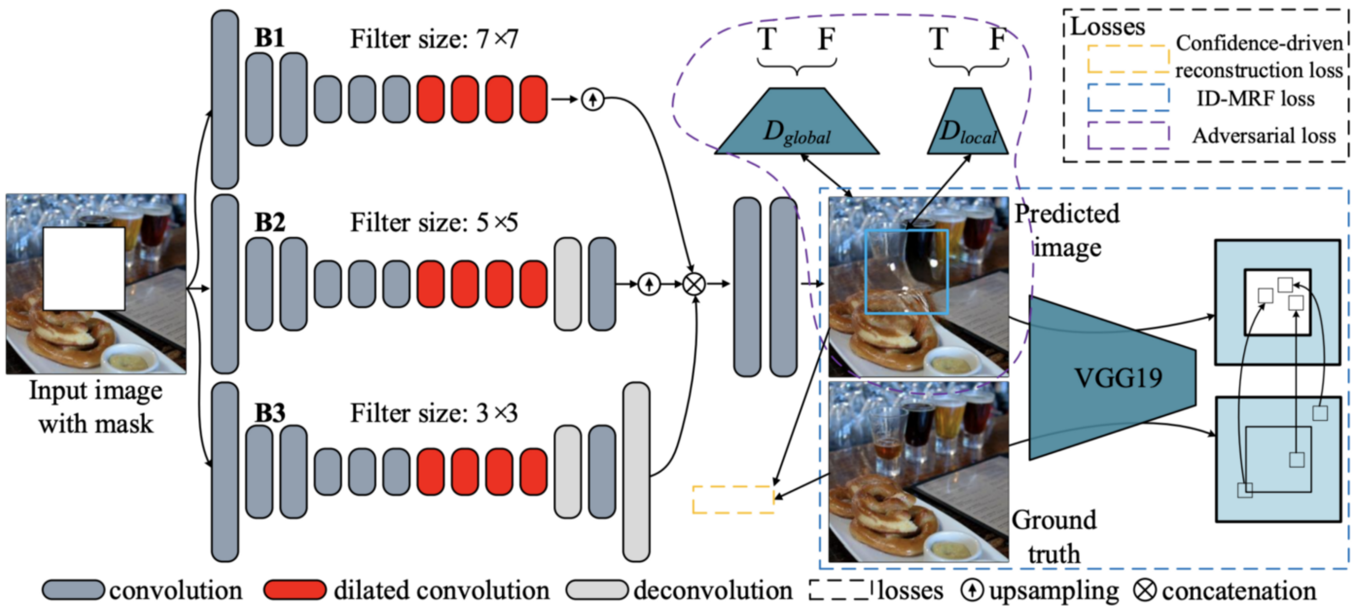


Figure 9. The proposed network architecture. Image by Yi Wang et al. from their paper [10]

Generative Multi-column Convolutional Neural Networks (GMCNN, 2018) [10] **expands the importance of sufficient receptive fields for image inpainting and proposes new loss functions to further enhance local texture details of the generated content**. As shown in Figure 9, there are three branches/columns and three different filter sizes are used at each branch. The use of the multiple receptive fields (filter sizes) is due to the fact that the size of the receptive field is important to the task of image inpainting. As the local neighbouring pixels are missing, we have to borrow information given by distant spatial locations to fill in the local missing pixels.

For the proposed loss functions, **the main idea of the Implicit Diversified Markov Random Field (ID-MRF) loss is to guide the generated feature patches to find their nearest neighbours outside the missing areas as references and these nearest**

MSNPS [3]. I highly recommend you to read my previous post [here] for a detailed explanation of the loss.

PartialConv (Pushing the Limits of Deep Image Inpainting for Irregular Holes, 2018)

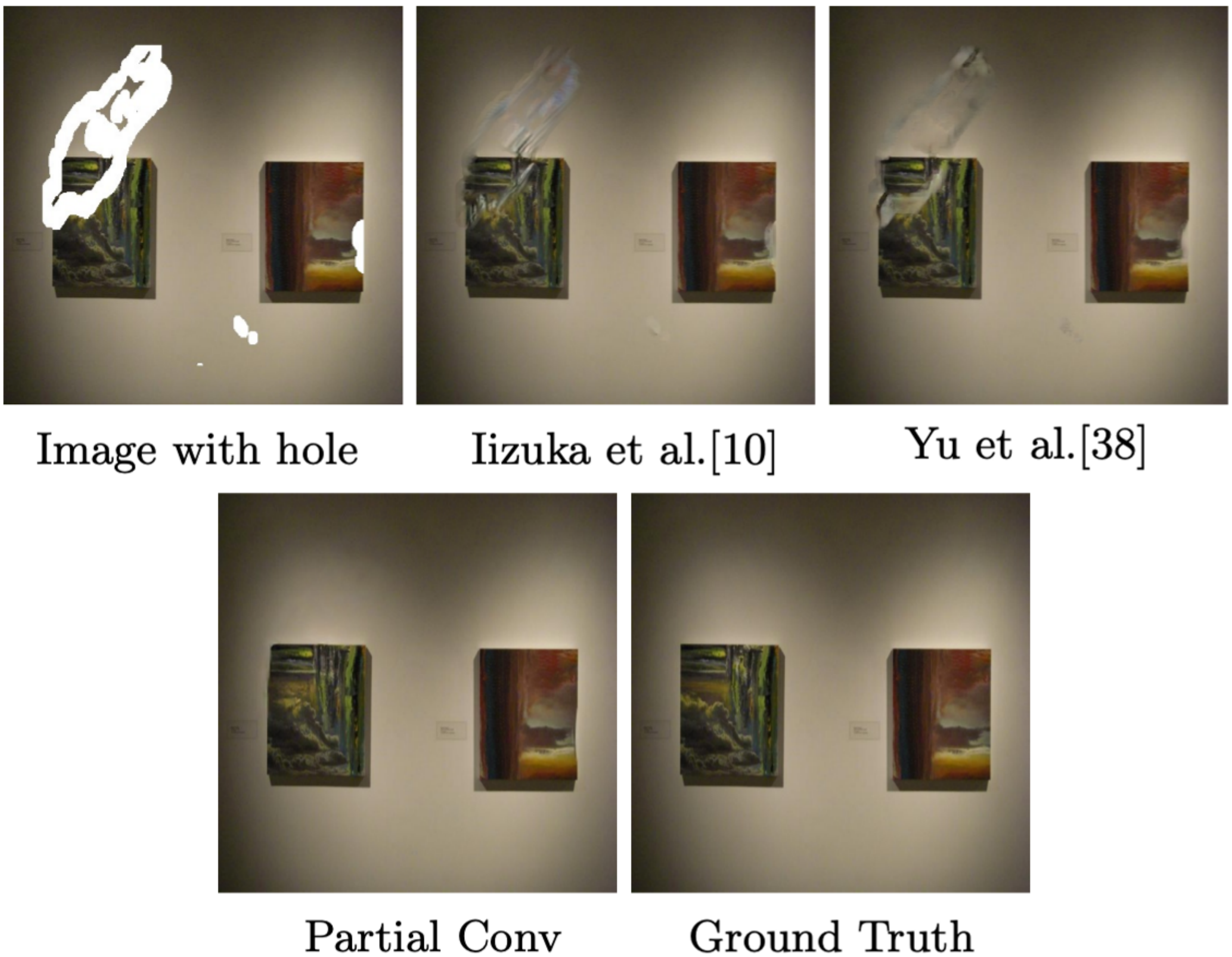


Figure 10. Visual comparisons of previous deep inpainting approaches trained by using regular masked images and the proposed Partial Conv. Image by Guilin Liu et al. from their paper [11]

Image Inpainting for Irregular Holes using Partial Convolutions (PartialConv or PConv) [11] pushes the limits of deep image inpainting by proposing a way to handle masked images with multiple irregular holes. Obviously, the core idea of this paper is the Partial Convolution. **By using PConv, the results of convolution would only depend on the valid pixels, hence we can have the control of the information to be passed inside** models are not suitable to complete irregular masked images.

I have provided a simple example to explain clearly how the partial convolution is performed in my previous post [here]. Please visit and have a look for details. Hope you enjoy it:)

EdgeConnect ("Lines First, Color Next" Deep Image Inpainting Approach, 2019)

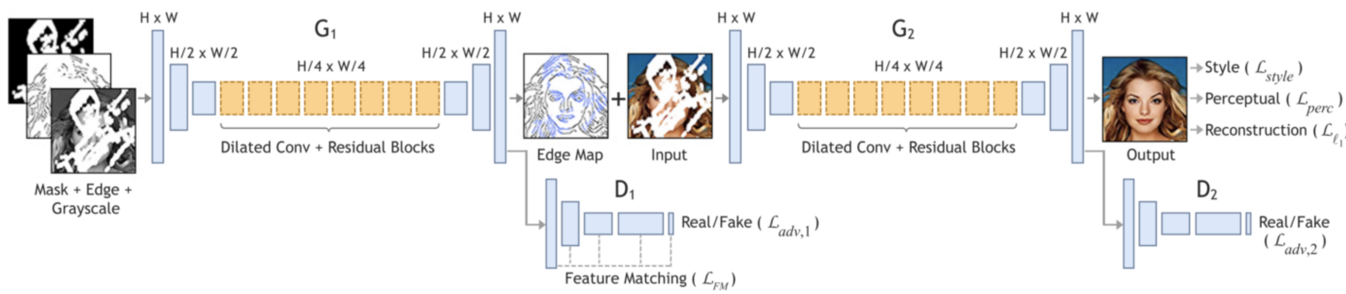
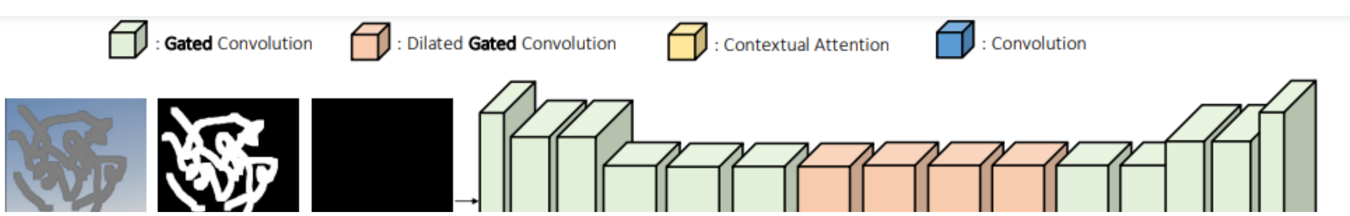


Figure 11. The network architecture of the proposed EdgeConnect. As you can see there are two generators and two discriminators. Image by Kamyar Nazeri et al. from their paper [12]

EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning (EdgeConnect) [12] provides an interesting way to the task of image inpainting. **The main idea of this paper is to separate the task into two simpler steps, namely edge prediction and image completion based on the predicted edge map**. They first predict the edges in the missing regions, then complete the image according to the predicted edge information. I would say that most of the techniques used in this paper have been covered in my previous posts. It's good for you to have a look at how various techniques can be used together to form a novel deep image inpainting approach. Perhaps, you may be able to develop your own inpainting model! Please see my previous post [here] for more details about this paper.

DeepFill v2 (A Practical Generative Image Inpainting Approach, 2019)



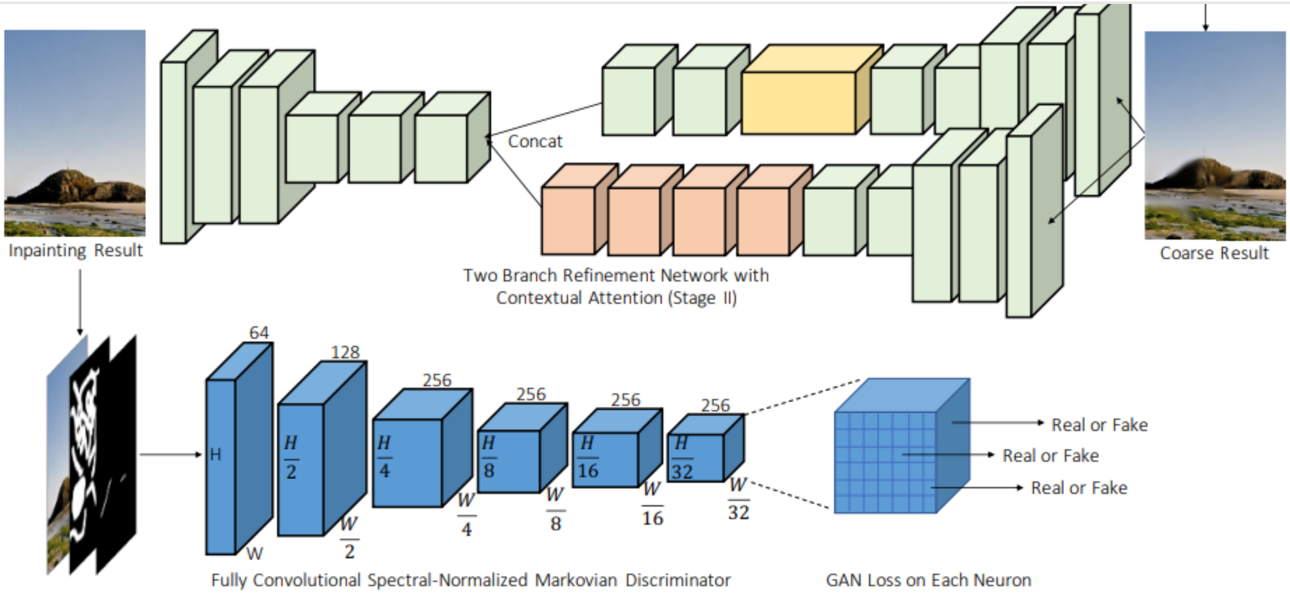


Figure 12. Overview of the network architecture of the proposed model for free-form image inpainting. Image by Jiahui Yu et al. from their paper [13]

Free-Form Image Inpainting with Gated Convolution (DeepFill v2 or GConv, 2019) [13] perhaps is the most practical image inpainting algorithm that can be directly used in your applications. This can be regarded as an enhanced version of DeepFill v1 [9], Partial Convolution [11], and EdgeConnect [12]. **The main idea of this paper is Gated Convolution which is a learnable version of the Partial Convolution. By adding an extra standard convolutional layer followed by a sigmoid function, the validness of each pixel/feature location can be learned and hence optional user sketch input is also allowed.** Apart from the Gated Convolution, SN-PatchGAN is also adopted to further stabilize the training of the GAN model. To know more about the difference between the Partial Convolution and the Gated Convolution, and how optional user sketch input can contribute to the inpainting results, please visit my latest post [here].

Final Thoughts

I hope that all of you can have a basic understanding of image inpainting now. I believe that most of the common techniques used in deep image inpainting have been covered in my previous posts. If you are my old friend, I think now you are able to understand other inpainting papers in the literature. If you are a newcomer, I would like to welcome you. I hope that you find this post useful for you. Actually, this post provides a highway for you to join us and learn together.

course, high-resolution image inpainting is also another challenging task. All these challenges can be categorized into Extreme Image Inpainting. I think that the coming state-of-the-art inpainting approach should be able to tackle some of these challenges.

Apart from paper reviews, I would like to try more new things and learn more. I have tried my best to show the development of deep image inpainting. I hope that you can see the trends in image inpainting and how a model is designed based on previous studies. I am really happy to share my knowledge with all of you here:)

Thanks for reading. If you have any questions, please feel free to leave comments here or send me an email. I am happy to hear from you and any suggestions are welcome. Hope to see you in the future:)

References

[1] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros, "Context Encoders: Feature Learning by Inpainting," *Proc. International Conference on Computer Vision and Pattern Recognition* (*CVPR*), 2016.

[2] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative Adversarial Nets," *in Advances in Neural Information Processing Systems* (*NeurIPS*), 2014.

[3] Chao Yang, Xin Lu, Zhe Lin, Eli Shechtman, Oliver Wang, and Hao Li, "High-Resolution Image Inpainting using Multi-Scale Neural Patch Synthesis," *Proc. International Conference on Computer Vision and Pattern Recognition* (*CVPR*), 2017.

[4] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa, "Globally and Locally Consistent Image Completion," *ACM Trans. on Graphics*, Vol. 36, №4, Article 107, Publication date: July 2017.

[5] Ugur Demir, and Gozde Unal, "Patch-Based Image Inpainting with Generative Adversarial Networks," https://arxiv.org/pdf/1803.07422.pdf.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," *Proc. Computer Vision and Pattern Recognition* (*CVPR*), 27–30

[7] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," *Proc. Computer Vision and Pattern Recognition* (*CVPR*), 21–26 Jul. 2017.

[8] Zhaoyi Yan, Xiaoming Li, Mu Li, Wangmeng Zuo, and Shiguang Shan, "Shift-Net: Image Inpainting via Deep Feature Rearrangement," *Proc. European Conference on Computer Vision* (*ECCV*), 2018.

[9] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S. Huang, "Generative Image Inpainting with Contextual Attention," *Proc. Computer Vision and Pattern Recognition* (*CVPR*), 2018.

[10] Yi Wang, Xin Tao, Xiaojuan Qi, Xiaoyong Shen, and Jiaya Jia, "Image Inpainting via Generative Multi-column Convolutional Neural Networks," *Proc. Neural Information Processing Systems*, 2018.

[11] Guilin Liu, Fitsum A. Reda, Kevin J. Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro, "Image Inpainting for Irregular Holes Using Partial Convolution," *Proc. European Conference on Computer Vision* (*ECCV*), 2018.

[12] Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Z. Qureshi, Mehran Ebrahimi, "EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning," *Proc. International Conference on Computer Vision* (*ICCV*), 2019.

[13] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas Huang, "Free-Form Image Inpainting with Gated Convolution," *Proc. International Conference on Computer Vision* (*ICCV*), 2019.

Deep Learning   Image Inpainting   Convolutional Network   Image Processing

Generative Adversarial