

Deep Learning Image Inpainting Techniques: An Overview

Shylaja B.¹ and Dr. R. Saravana Kumar²

¹Research Scholar, Computer Science and Engineering, Dayananda Sagar Academy of Technology & Management,
Bengaluru, Karnataka, India -560082
Visvesvaraya Technological University, Belagavi -590018
Email: shylaja2007.b@gmail.com

²Associate Professor, Computer Science and Engineering, Dayananda Sagar Academy of Technology & Management,
Bengaluru, Karnataka, India -560082
Visvesvaraya Technological University, Belagavi -590018
Email: saravanakumar.rsk28@gmail.com

Abstract—Images can have flaws when captured, may become impaired, or degrade over time. Paintings and other art works symbolize a significant part of one's legacy and modern culture. But, because of the type of supplies used in these art works, they are susceptible to impairment and they are gradually degraded. Traditionally, these were repaired manually to improve image quality. With the advent of digital age we can avoid manual restoration work by using digital image restoration techniques. The primary aim of inpainting process is recreating the missing portions so that the restored image resembles the original by predicting the missing part of the image by using the details from the other parts of the same image or by referring related images from the bulk data set. This paper presents an overview and comparison of image inpainting techniques using deep learning over the past decade. The steps involved and the pros and cons of each method is discussed. This will serve as a reference for the researchers in choosing the suitable method for their problem.

Index Terms— Inpainting, GAN, Restoration, Deep Learning, Neural Networks, Convolution Neural Networks, Preprocessing.

I. INTRODUCTION

Missing data is one of the most vital statistical and design problems in research. Missing data diminishes representativeness of the sample [2]. Images can also have missing regions. Image inpainting is the task of predicting and filling missing values in a specific part of the input image. The main goal of this task is recreating unavailable portions or impaired input image so that the restored image cannot be differentiated with the original image [1].

Image inpainting method originated from a prehistoric method done by art workers to repair impaired art works such as paintings or photographs having minor flaws like scratch, torn canvases, crack, smudges, dirt and patches to retain the quality as that of the original art work. Image degradation is a result of numerous causes such as rough handling, improper storage, due to light exposure, temperature variations, humidity and other environment factors. [3].

But the inpainting task requires more time and is a skilled task and it is susceptible to human mistakes and it requires high precision. Therefore the digital restoration of images and arts are very useful in this task. There are two types of image inpainting: Traditional methods and Data Driven Deep Learning methods.

A. Traditional Image Inpainting methods

Copy-paste is the simple way to fill in an image with missing regions. Primarily in this, most similar image spots will be searched from the residual pixels of the same image or from the related images from the large dataset containing thousands of images, and then patches on the missing areas will be pasted directly. There are five sub-types in Traditional Image Inpainting methods as shown in figure [1].

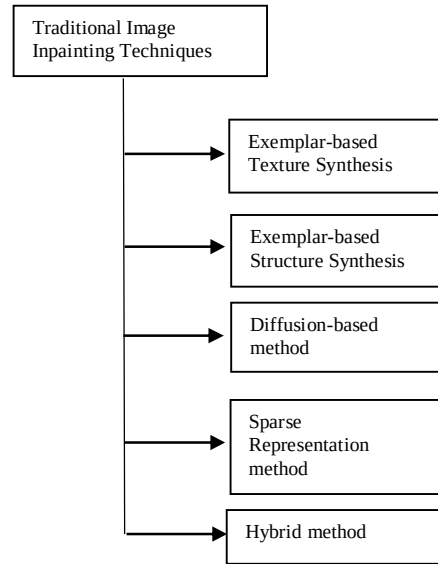


Figure 1. Traditional Image Inpainting methods

Traditional inpainting methods are sometimes adequate, but not always. It is computationally expensive. Search procedure takes more duration and it uses distance based metrics. It is not much efficient, requires improvement. Modern methods using deep Neural Networks are the most powerful techniques and are best suited for handling complex images and images with large missing area.

B. Deep Learning methods

Deep Inpainting methods are data driven techniques that predicts the unknown parts in the image with better global consistency and local fine texture details.

Deep Learning Methods are sub-divided into Convolution Neural Network (CNN) and Generative Adversarial Network (GAN) [1] as shown in figure 2.

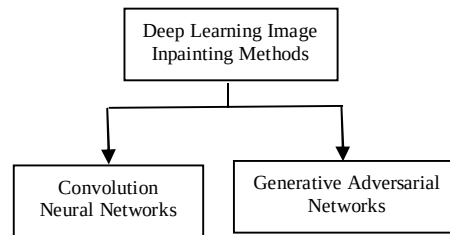


Figure 2. Deep Learning Inpainting methods

In section II, the description of the major inpainting techniques based on Deep Learning concept and the steps involved are discussed. In section III, the merits and demerits are of each method is discussed, followed by Conclusion in section IV.

II. DEEP LEARNING BASED IMAGE INPAINTING TECHNIQUES

A. Context Encoder

Context Encoder technique is based on the deep learning method Generative Adversarial Networks (GANs). The word “Context” refers to understanding the whole image. The main idea of Context Encoder is based on “Channel-wise Fully Connected Layer (middle layer in the network)” [4].

Feature locations at the current layer depend on feature locations at the previous layer just like standard fully connected layer. So that it helps to study the correlation among all feature locations. It helps to understand the in-depth semantic understanding of the complete image [4].

The steps involved in the Context Encoder technique is illustrated in figure 3.

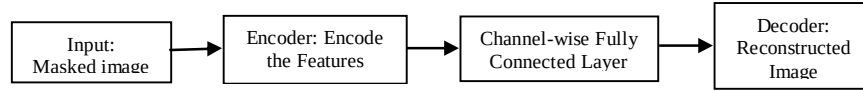


Figure 3. Steps involved in Context Encoder

Steps:

- Masked image or image with a center missing hole is used as input
- To obtain the encoded features, the input is fed into the encoder
- To get better semantic features, in between the encoded and the decoded features, Channel-wise Fully Connected Layer is placed
- By using the bottleneck features, decoder reconstructs the missing parts of the image.

B. “Multi-Scale Neural Patch Synthesis”

“Multi-Scale Neural Patch Synthesis” [5] is an improved version of Context Encoder. The missing portions in an image are predicted using a revised Context Encoder. To enhance the visual quality of the filled images, Texture Network is used. Texture Network is based on Style Transfer. “To improve the local texture details, the style of the most alike valid pixels to the generated pixels is transferred” [5].

The steps involved in the “Multi-Scale Neural Patch Synthesis” is illustrated in figure 4.



Figure 4. Multi-Scale Neural Patch Synthesis

Steps:

- Masked image or image with a center missing hole is used as input
- To obtain the encoded features, the input is fed into the Content Network
- Reconstructed image is obtained after filling in the missing parts.
- Texture Network is used to enhance the pictorial quality of the reconstructed image by refining texture details.
- The output of the is the Texture Network final enhanced image with better visual quality

C. “Globally and Locally Consistent Image Completion” [6]

“Globally and Locally Consistent Image Completion defines the Fully Convolution Network with Dilated Convolutions for deep image inpainting and there by sets the milestone in deep image inpainting” [6].

The network is can understand the context of an image by using dilated convolutions without using costly fully connected layers. Therefore it is able to handle different size images.

“It uses two discriminators at two scales are trained together with the generator network” [6]. The filled image will have good global and local consistency with both the global and local discriminators [6].

The steps involved in the “Globally and Locally Consistent Image Completion” is illustrated in figure 5.

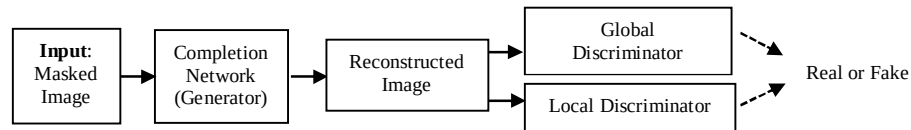


Figure 5. Globally and Locally Consistent Image Completion

Steps:

- Masked image or image with a center missing hole is used as input
- To obtain the encoded features, the input is fed into the Completion Network (Generator)
- Reconstructed image is obtained after filling in the missing parts.
- “A global discriminator handles the whole image” [6]
- A local discriminator handles the filled center hole i.e., sub-image around the filled region in a local sense.
- Uses simple post processing method to further enhance the visual quality of the image. Fast Marching method and Poisson image blending techniques are used.

D. “Patch-Based Image Inpainting with Generative Adversarial Networks” [7]

In the generator network, Dilated and interpolated convolutions are employed. By using the dilated residual blocks the inpainting results are enhanced. To form a dilated residual block, residual connection and dilated convolution are combined. “The traditional GAN discriminator is replaced by the PatchGAN discriminator to provide better local texture details and global structure consistency. The main difference between traditional GAN discriminator and PatchGAN discriminator is that traditional GAN discriminator generates a single predicted label (0 to 1) to show if the input is real or fake” [7]. PatchGAN discriminator generates a matrix of predicted labels (0 to 1) to show the authenticity of each local region of the input image. Every matrix element indicates a local region of the input [7].

The steps involved in the “Patch-Based Image Inpainting” is illustrated in figure 6.

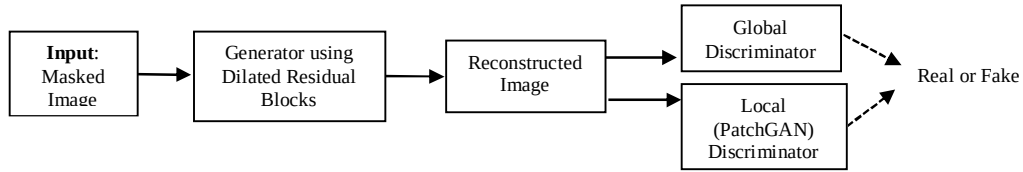


Figure 6. “Patch-Based Image Inpainting with Generative Adversarial Networks” [7]

Steps:

- Masked image or image is used as input
- To obtain the encoded features, the input is fed into the Generator using Dilated Residual Blocks
- Reconstructed image is obtained after filling in the missing parts.
- A global discriminator handles the complete image
- A local discriminator handles local region using patchGAN.
- The output is in matrix form and every element in the matrix indicates local region in the input image. It generates 1 if the local region is real otherwise 0.
- The discriminator checks multiple local image patches and decides each patch is real or not.

E. “Shift-Net: Image Inpainting via Deep Feature Rearrangement” [8]

“Shift-Net Deep Feature Rearrangement by using both modern data-driven CNNs and the conventional “Copy-and-Paste” method by using the shift-connection layer” [8]. To embed copy-and-paste method in modern CNNs, a shift-connection layer is used to provide inpainting results with both accurate detailed textures and global semantic structure [8].

The steps involved in the “Shift-Net Image Inpainting” is illustrated in figure 7.

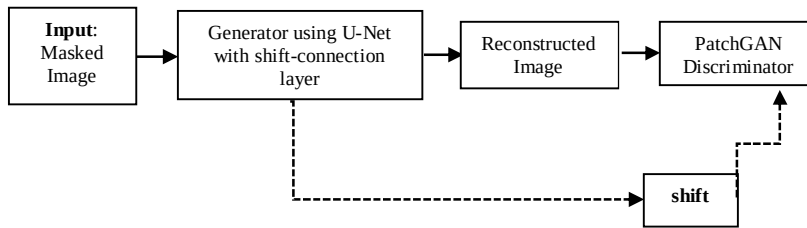


Figure 7. Shift-Net Image Inpainting

Steps:

- Masked image or image is used as input
- To obtain the encoded features, the input is fed into the Generator using U-Net with shift connection layer.
- During the decoding process, a guidance loss is used to allow the network (Shift Net) to learn to fill in the missing parts.
- To match the decoded feature inside the missing region to the encoded feature outside the missing region, a shift-connection layer is used.
- Then each matched location of the encoded feature outside the missing region is moved to the corresponding location inside the missing region.
- This captures the information about the most similar local image patches found outside the missing region and this information is concatenated to the decoded feature for further reconstruction.
- The patchGAN discriminator is used.

F. “Deep Image Inpainting for Irregular Holes Using Partial Convolutions (PartialConv or PConv)” [10]

“Partial Convolutions provides a method to handle masked images with multiple irregular holes. By using PartialConv, the results of convolution will only depend on the valid pixels, henceforth we can control the information to be passed inside the network” [10]. This method aims at handling irregular holes. Whereas the previous inpainting methods were can only handle regular (rectangular) masked images.

“It is called partial convolution because during convolutions the missing pixels are separated from the valid pixels so that the results of convolutions are influenced only by the valid pixels. The convolution is partially done on the input based on a binary mask image that is automatically updated” [10].

Steps:

- An input with its equivalent binary mask image (0 for hole pixels, 1 for valid pixels) with fixed weights are used.
- U-Net network is used with skip connections. Standard convolution layers are swapped by partial convolution layers.
- After each partial convolution, update the binary mask.
- The binary mask will be gradually updated to have all 1’s when we perform further partial convolutions
- Other than the standard L1 loss and total variation loss, two high-level feature losses are used for completing the fine textured masked images
- In this method discriminator is not used.

G. Image Inpainting with Adversarial Edge Learning (EdgeConnect)

EdgeConnect uses two steps: “edge prediction and image completion based on the predicted edge map. First edges in the missing regions are predicted by the Generator to obtain the overall image structure” [11]. Then according to the predicted edge information which is a binary map to show the outline of the image, the image is completed by the second Generator by filling the missing details and improved texture details.

The steps involved in the “EdgeConnect” method is illustrated in figure 8.

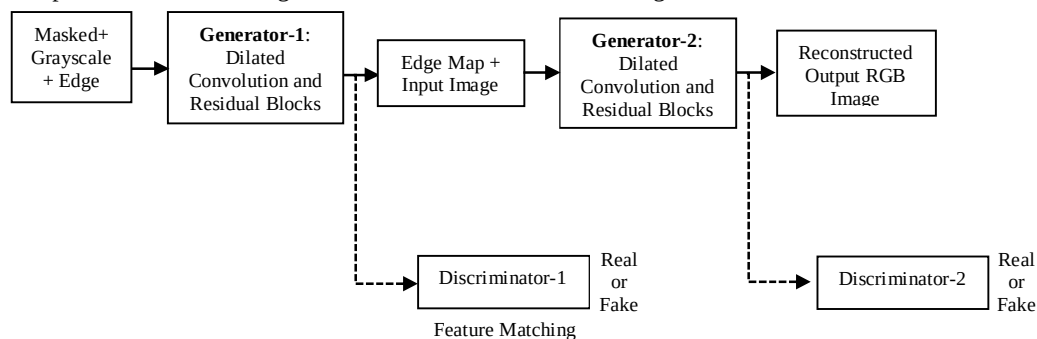


Figure 8. Image Inpainting with Adversarial Edge Learning (EdgeConnect)

Steps:

- Mask image, masked edge image and the masked grayscale image is sent to Generator-1 as input and it predicts the Edge Map

- Generator-1 is trained using the standard adversarial loss and the feature matching loss.
- Generator-2 is trained using the style loss, perceptual loss, L1 reconstruction loss and the standard adversarial loss.
- It accepts the predicted edge map and the masked RGB image as input and outputs a completed RGB image.

H. “DeepFill v2: Free-Form Image Inpainting with Gated Convolution” [12]

“DeepFill-v2 uses Gated Convolution, a learnable version of the Partial Convolution by adding an extra standard convolutional layer followed by a sigmoid function, can learn if each pixel is valid and therefore user sketch input is also accepted which is optional. Along with Gated Convolution, SN-PatchGAN is also adopted to improve the training of the GAN model. DeepFill-v2 achieves high quality, free-form inpainting compared to other previous techniques” [12].

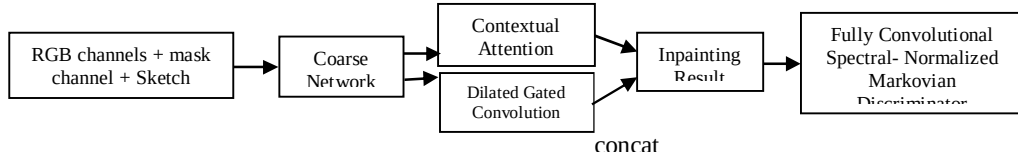


Figure 9. Free form image inpainting

Steps:

- “The coarse generator takes the RGB channels, mask channel, and an optional sketch image as input for a coarse reconstruction of the missing areas.
- Then for refinement, the coarse filled image will be passed to the second refinement generator network. Contextual Attention layer is used in the refinement network [9].
- PatchGAN is used in the discriminator,
- Spectral Normalization is applied to each standard convolutional layer of the discriminator to stabilize the training” [12].

III. MERITS AND DEMERITS: COMPARISON

Table I summarizes the advantages and disadvantages of the inpainting methods discussed in this paper in Section II.

IV. CONCLUSION

Deep Learning methods using CNNs and GANs provide a great performance with respect to pixel-wise reconstruction accuracy, visual quality and speed. Overcoming the limitations of Traditional methods, Deep Learning based inpainting is becoming state-of-the-art techniques witnessing great progress in image inpainting field and currently it is one of the key research domain. The main challenge in inpainting is when the input image has a large, irregular missing area and when the image is complex with rich texture. Also high resolution image inpainting is another challenge. Therefore, Extreme Image Inpainting is going to be a promising direction. More effort can be made in developing simple and efficient inpainting methods to handle complex images.

TABLE I. ADVANTAGES AND DISADVANTAGES OF DEEP LEARNING METHODS

Method	Advantages	Disadvantages
Context Encoder	▪ It provides very deeper semantic understanding of the whole image. ▪ This method has been widely used in the later techniques. ▪ Used to restore tarnished images (can be extended to repairing of videos) ▪ The extracted feature representations are useful to additional operations like classification, detection and semantic segmentation.	▪ Additional improvement of the texture details of the generated pixels needed. ▪ Not possible to handle images with high resolution ▪ It can only predict the missing portions up to some degree of semantic correctness ▪ But textures and details of the predicted pixels should be improved further.
Multi-Scale Neural Patch Synthesis	▪ It provides a frame work to combine the techniques from Context Encoders and Neural Style Transfer. ▪ To handle high resolution images, it suggest a Multi-scale way. ▪ Enhances the texture details of the generated pixels using style transfer techniques	▪ For complex image, it is still difficult to fill the missing parts. ▪ Low speed. So real-time performance not possible.

TABLE I (CONTINUED)

Globally and Locally Consistent Image Completion	- The network can take input images at different sizes. - Without using fully-connected layers, Fully Convolutional Network with Dilated Convolution makes us to understand the context of an image - Multiple discriminators are helpful to improve the texture details of the completed images at various scales. - Can be used for object removal	For complex images, it is very difficult to fill in the missing parts. Not able to restore images with large missing area.
Patch-Based Image Inpainting with Generative Adversarial Networks	- Residual learning increases the inpainting performance - Good local texture details and global structure consistency - Provides results with good visual quality - Enhances the local texture details of the generated pixels.	Provides low performance on 256×256 testing images. Provides better local details, but poor global consistency
Shift-Net	- Improved global semantic structure and local texture details. - Benefits from both “Copy-and-Paste” and modern CNN methods. - Provides Good local visual details and reconstruction accuracy - Increased speed with sufficient details through deep feature rearrangement.	Shift operation can be complex Cannot handle irregular masked images
Partial Convolutions (PartialConv or PConv)	- Provides better performance - This method aims at handling irregular holes.	When the missing area is larger, it is difficult to reconstruct. Cannot handle complex images
EdgeConnect	- Instead of blur results, sharper inpainting results are achieved - Can be used in unwanted object removal, merging of two images etc. - Both regular and irregular masks can be restored.	Difficult to predict complex images Not able to restore images with large missing area.
DeepFill-v2	- Achieves high quality, free-form inpainting compared to other previous techniques. - Very practical deep image inpainting method in the literature. - Both regular and irregular masks can be restored.	Difficult to predict complex images Not able to restore images with large missing area.

REFERENCES

- [1] Jireh Jam, Connah Kendrick, Kevin Walker, Vincent Drouard, Jison Gee-Sern Hsu, Moi Hoon Yap, “A comprehensive review of past and present image inpainting methods”, Computer Vision and Image Understanding 203 (2021) 103147, 1077-3142, Elsevier Inc, 2021.
- [2] Shylaja B, Dr. R. Saravana Kumar, “Traditional Versus Modern Missing Data Handling Techniques: An Overview”, International Journal of Pure and Applied Mathematics, Volume 118 No. 14 2018, 77-84 ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version), 2018.
- [3] Varun Gupta, Nitigya Sambyal, Akhil Sharma, Praveen Kumar, “Restoration of artwork using deep neural networks”, Evolving Systems, <https://doi.org/10.1007/s12530-019-09303-7>, Springer-Verlag GmbH Germany, part of Springer Nature, 2019.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Liuchun Yuan, Congcong Ruan, Haifeng Hu, Dihui Chen, “Image Inpainting Based on Patch-GANs”, IEEE Access, VOLUME 7, 2019.
- [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] 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.
- [11] 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.
- [12] 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.