Analysis of image inpainting and object removal methodologies

Shiney Thomas¹*, Dr. Juby Mathew²*

¹Research Scholar, ²Associate Professor,
¹Department of Computer Science and Engineering, Amal Jyothi College of Engineering, Kanjirappally, Kottayam, Kerala, India

Email: shineythomas@amaljyothi.ac.in

Abstract: Image inpainting is a process that tries to fill in missing parts of a degraded image or remove objects from an image, but maintain the realistic content in the image. Various methods can be employed to perform inpainting. Here three benchmark methods of which two are conventional and other GAN based are evaluated to check the effectiveness of the inpainting process, for different class of images. The visual quality evaluation as well as quantification parameter namely the SSIM score was computed. Based on the test score it was seen that the GAN based method was capable of providing visually good restored images with an average SSIM score of 0.9516, which is an indicator on the quality of the image. The GAN based methods can be further enhanced to provide more realistic images by incorporating functions to extract both local and global contextual content.

1. Introduction

Image is one of the most common forms of information source that is used in everyday life with the help of many application/tools. The information content will be critically affected if these images are corrupted, missing or damaged. Image inpainting or restoration models include operations that try to alter an image by filling in or removing details from the targets regions in the image. The process tries to remove objects like text, date, or broken regions as well as add objects into the image but still provide a realistic image. Image restoration is a very active research content in the field of image processing [1].

A generic architecture of any image inpainting models can include various processes. The first process is the acquisition of the image that needs to be processed. This is followed by image pre-processing; these can incorporate operations enhancement operation like brightness or contrast, colour to greyscale conversion or noise removal. The next process would involve identification of the area of interest where the inpainting operation needs to be performed. Once this has been identified an appropriate mask has to be created so that image inpainting process can be initiated.

There are numerous papers in image inpainting focusing on different aspects. Image inpainting can be done by two methods- patch matching using low level image features and deep neural networks based on feed forward generative models. Patch based methods detects plausible fixed textures but are not successful in detecting complex scenes, human faces and some objects. Deep generative models learns from large scale datasets to capture semantics of the complex images and apply feature continuity of the generated contents around the mask region.[12]. When a key region missing (it is the only information in the image, which cannot be obtained from existing information, such as the nose loss of a single face picture), the target regions cannot be filled from existing information [1]. In [9] authors elaborate on an inpainting method based on a fast non-iterative nonlinear analysis of stationary first order transport equations using higher-order partial differential equations. Structure
tensors are employed to estimate the image coefficients in the coherence direction robustly using fast marching method which is run only once while transporting. Based on the coherence parameter the method switches between two operations namely diffusion and directional support. Another benchmark method is discussed in [10][11] that employ exemplar-based image inpainting, where the process is divided into have two phases namely, deciding the filling-in order and selecting good exemplars. Therefore, the traditional method does not have a semantic inpainting. In [3], a framework for coarse-to-fine generative image inpainting was introduced along with a novel contextual attention module. The addition of contextual attention module considerably improves image inpainting results by learning feature representations for explicitly matching and attending to relevant background patches. The input to the network is a 256 × 256 image with a rectangle missing region sampled randomly during training, and the trained model can take an image of different sizes with multiple holes in it. The authors in [4], proposes a gated convolution for free-form image inpainting where a dynamic feature gating mechanism is computed for both RGB channel and spatial location. Initially the features are computed based on the gating function defined by \( g = \sigma(w_g x) \) where \( \sigma \) and \( w_g \) represents the sigmoid function and learnable variable respectively. The final output operation is defined by the operation \( y = \Phi(wx)\Theta g \) where \( \Phi \) can be any activation function and \( \Theta \) represents the gating function. In other words the operation is simply a multiplication of learned feature and gating values. The main advantage of gated convolution is its ease of computation and that it can be effective used on arbitrary shaped masks. The gated convolution modules are stacked for designing a encoder-decoder network as discussed in [3]. For previous inpainting networks which try to fill a single rectangular hole, an additional local GAN is used on the masked rectangular region to improve results [3][5]. The task of free-form image inpainting where there may be multiple holes with any shape at any location employing SN-Patch GAN has been discussed in [4]. More recent work [13] focuses on encoder combined with a context loss function. In order to obtain clear repaired images and ensure that the semantic features of images are fully learned with the generator and discriminator operating on a fusion of squeeze-and-excitation networks deep residual networks.

Different GAN models vary in their implantation strategies in terms of number of layers, types of layers, parameters used in the layers and so forth. In this papers a GAN model is explored that can be used for both filling in missing regions as well as removing objects from the scene. The efficacy of the model is compared with existing conventional model. Section 2 provides a crisp discussion on the methodology employed and the results and findings are elaborated in sections 3. The manuscript is concluded in section 4.

2. Methodology

The GAN has basically two modules namely the generator and discriminator. Both the networks are trained simultaneously to achieve its required level namely: Generator is trained to generate data that is capable of fooling the discriminator and the discriminator is trained such that is capable to distinguish between real and generated data. A GAN generator and discriminator that can be employed for inpainting application is illustrated in figure 1a and 1b. The choice of the layers and filter kernels determine the effectiveness of the model as well as the novelty.

From the figure, it can be seen that the common functional layers in the generator and discriminator are Convolution, Batch Normalization and Leaky ReLU. Some other design constraints in the generator include the embedding dimension of 50; each convolution layer consist of 5x5 filter kernel and the number of filters are reduced at subsequent layers, with stride of 2. The last convolution layer has 3 filters of size 5x5 corresponding to the RGB color space. This is followed by a tanh activation function. In the discriminator the each convolution layer consist of 5x5 filter kernel and the number of
filters is increased at subsequent layers, with stride of 2. Here the scale factor of the Leaky ReLU layer is taken as 0.2.

![GAN model for Inpainting](image)

**Figure 1:** GAN model for Inpainting a. Generator b. Discriminator

3. Results and Finding

To get a clearer understanding on efficacy of the proposed GAN based inpainting process, it is compared with two conventional benchmark models based on [9] and [11]. For the purpose of training and testing of the image inpainting models images from various publically available datasets were taken these include CelebA [6], CelebA-HQ [7] and Places2 [8]. For the purpose of evaluation illustration few critical outputs are illustrated in Figure 2. From the figure following observations were made:

1. All the methods perform equally well in removing an object from a scene and is able to maintain good realistic content.
2. In the case of text removal from a scene all the methods perform equally well, but with slight reconstruction error in method using coherence transport.
3. When multiple patch removal is involved the coherence transport method provides a poor quality scene reproduction when with the other two methods performing equally well.
4. Considering the task of filling the missing region in a scene, celebrity faces data is used as an illustration. The main difference between the two input images is the size of missing patch. It can be seen from first celebrity face that both the conventional methods fail to create a realistic image. The GAN model gave a far superior image, but still sharper features are still missing. In the case of the second celebrity image all the methods gave an equally good result. Reason why these methods gave goods result can be attributed to two parameters namely the size of patch and area of the patch. The patch is located in the region of least texture transition.
5. Finally scene filling scenario is considered next. In the first scenario the window pane was perfectly reconstructed by the GAN model and the exemplar method provided a marginally good reconstruction followed by coherence transport method. The same result is observed in the last scenario, but realistic or the contextual content can still be improved in the GAN model.

To obtain a quantitative analysis on the performance of the algorithm structural similarity index measure (SSIM) has been employed which is given equation (1). SSIM takes on values between 0 and 1, with the values greater than 0.95 representing good quality images.

\[
SSIM(x, y) = \frac{2\mu_x\mu_y + C_1 + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}
\]  

(1)

SSIM has been evaluated on all the reconstructed images and average scores of the methods are quantified in Table 1.

| a. Input Image | b. Using Coherence Transport | c. Exemplar Based | d. GAN Based |
|----------------|-----------------------------|-----------------|--------------|
| ![Input Image](image1.png) | ![Using Coherence Transport](image2.png) | ![Exemplar Based](image3.png) | ![GAN Based](image4.png) |
| ![Using Coherence Transport](image5.png) | ![Exemplar Based](image6.png) | ![GAN Based](image7.png) |
| ![Exemplar Based](image8.png) | ![GAN Based](image9.png) |

**Figure 2: Generic Inpainting Model**
Table 1: SSIM Score Average Across Different Images

| Method           | Average SSIM Score |
|------------------|--------------------|
| Coherence Transport | 0.8215             |
| Exemplar Based    | 0.9111             |
| GAN              | 0.9516             |

4. Conclusion

Image inpainting is a process used to fill in missing part or remove objects from an image, while maintaining the image integrity (realistic content). Benchmark conventional and GAN based methods has been evaluated on different class of images and their performances were evaluated on both visually and quantitatively using the SSIM score. In all the evaluations across different class of images the GAN based methods performed better than the conventional methods with higher average SSIM score (0.9516). Furthermore GAN based methods fail to address the contextual content perfectly which has lead to smaller SSIM scores for certain images. Hence use of contextual methods on both local and global level can improve the SSIM score as well as realistic content.

5. References

1. H. Liu, G. Lu, X. Bi, J. Yan and W. Wang 2018 Image Inpainting Based on Generative Adversarial Networks 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Huangshan, China, pp. 373-378, doi: 10.1109/FSKD.2018.8686914.
2. Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester 2000 Image inpainting Proceedings of the 27th annual conference on Computer graphics and interactive techniques (SIGGRAPH ’00) ACM Press/Addison-Wesley Publishing Co., USA, pp. 417–424, doi:10.1145/344779.344972.
3. J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu and T. S. Huang 2018 Generative Image Inpainting with Contextual Attention IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, pp. 5505-5514, doi: 10.1109/CVPR.2018.00577.
4. J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu and T. Huang 2019 Free-Form Image Inpainting With Gated Convolution IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), pp. 4470-4479, doi: 10.1109/ICCV.2019.00457.
5. Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa 2017 Globally and locally consistent image completion ACM Trans.on Grapic. Vol.36 Issue 4 Article 107 (July 2017), 14 pages. doi: 10.1145/3072959.3073659.
6. Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaou Tang December 2015 Deep Learning Face Attributes in the Wild Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV ’15) IEEE Computer Society, USA pp.3730–3738. doi:10.1109/ICCV.2015.425.
7. C. -H. Lee, Z. Liu, L. Wu and P. Luo 2020 MaskGAN: Towards Diverse and Interactive Facial Image Manipulation IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA pp.5548-5557, doi: 10.1109/CVPR42600.2020.00559.
8. Zhou B, Lapedriza A, Khosla A, Oliva A, Torralba A. June 2018 Places: A 10 Million Image Database for Scene Recognition IEEE Trans Pattern Anal Mach Intell. 40(6) pp.1452-1464 doi: 10.1109/TPAMI.2017.2723009. Epub 2017 Jul 4. PMID: 28692961.
9. Bornemann, F., März, T. 2007 Fast Image Inpainting Based on Coherence Transport. J Math Imaging Vis 28, pp. 259–278 .doi: 10.1007/s10851-007-0017-6.
10. Criminisi, A., P. Perez, and K. Toyama 2004 Region Filling and Object Removal by Exemplar-Based Image Inpainting *IEEE Transactions on Image Processing* **Vol. 13, No. 9**, pp. 1200–1212.

11. Le Meur, O., M. Ebdelli, and C. Guillemot 2013 Hierarchical Super-Resolution-Based-Inpainting *IEEE Transactions on Image Processing* **Vol. 22, No. 10**, pp. 3779–3790.

12. R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson and M. N. Do 2017 Semantic Image Inpainting with Deep Generative Models *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI*, pp. 6882-6890, doi: 10.1109/CVPR.2017.728.

13. Chen, Y., Liu, L., Tao, J. et al. 2020 The improved image inpainting algorithm via encoder and similarity constraint. *Vis Comput*. doi:10.1007/s00371-020-01932-3.