Project Report - ECE 285

Wei-Ru Lin
Electrical and Computer Engineering
A59003068

Zheng Li
Electrical and Computer Engineering
A59004851

Abstract

In this project, we will utilize machine learning models to recover corrupted, damaged traditional Chinese landscape painting. In order for image inpainting for irregular stains or damaged and filling image in holes. We use partial convolution network with an UNet-like architecture to recover painting from flawed and damaged image. Trying to recover an original, perfect traditional Chinese landscape paintings.

1 Introduction

The application of computer vision is very extensive on many fields especially in objection detection and segmentation, classification. Convolution Neural Networks can achieve high performance image classification. However, in recent years, research on inpainting, filling the holes with images, has become very popular in Computer Vision and Machine Learning. Inpainting is also utilized in many fields and application. This technique is able to reconstruct missing or damaged areas of images and videos.

Here we focus on image inpainting. Currently, there are many image inpaintings research done by many people in different dataset e.g. Animates, Celebrities. In this paper, we use image inpainting technique and traditional Chinese landscape paintings as a dataset to recover its damaged parts. The reason why we choose traditional Chinese landscapes paintings is that there are few researchers focus on the antique and antique recovering is very important for helping us be better understanding in ancient, culture, environment, etc.

Many traditional Chinese landscape paintings existing thousands of years ago are damaged or destroyed by some reasons in history. Although it is not possible to restore the original piece, the latest advances in image inpaiting technique may help us to get a good estimate of the original art piece.

In the work of Liu et al., they developed a U-Net like convolutional neural network with partial convolution layers to complete images with irregular holes[1]. Images generated by their model are not only semantically consistent with the context original image, but also shows great details, which other image inpainting models failed to achieve. We apply the idea from Liu et al. Build partial convolution layer and modify transform learning implement to generate a complete, flawless traditional Chinese landscape paintings.

In Section II, we will discuss related work. In Section III, it is devoted to some methods to build our model. In Section IV, we present the utilized dataset and its format and the experimental results. The conclusion is reported in Section V and Contribution in Section VI.

2 Related Work

Unsupervised Learning This is a patterns from untagged data. Through imitating, the model can build a internal representation of its and then generate outcome. Unsupervised Feature Learning with Auto-encoders is De-noising Auto-encoders.[2] This Auto-encoders take corrupted, damaged images as input, (the original, flawless image added with some masks), and reconstruct the original image. However, these features are not very useful for discriminative tasks.

Inpainting and hole-filling There are many different size of holes in paintings. Non-learning methods can deal with some narrow or small holes by their neighboring pixels utilizing some distance region mechanisms in [3, 4, 5] However, if the paintings have big holes, those methods cannot get a good result and cause ambiguous in the paintings. Patch-based methods [6, 7] However, it often comes a lot of computation cost. PatchMatch [8] speeds can accelerate the process for patch searching algorithm. However, those methods are not efficient for real-time.

3 Method

Our work combines the techniques such as partial convolution layer and mask updating from Liu et al. and the traditional Chinese landscape paintings dataset. We borrowed the idea of transfer learning and fined tuned the partial convolution model with the traditional Chinese landscape painting dataset.

3.1 Partial Convolutional Layer

The idea of partial convolution layer was introduced by Liu et al. The partial convolution layer consists of partial convolution operation and mask update. Denote the convolution filter weights as $\mathbf{W}$ and the corresponding bias term b. Denote the feature values for the current convolution as $\mathbf{X}$ and $\mathbf{M}$ as the corresponding binary mask. At each single pixel, the partial convolution operation is expressed as:

$$x' = \begin{cases} \mathbf{W}^T(\mathbf{X} \odot \mathbf{M})\frac{size(\mathbf{M})}{sum(\mathbf{M})} + b, & if sum(\mathbf{M}) > 0 \\ 0, & otherwise \end{cases}$$

Here, $\odot$ denotes element-wise multiplication. and $size(\mathbf{M})$ is the number of elements in tensor $\mathbf{M}$. It is clear that in this equation, output values depend only on the unmasked inputs, which is consistent with the our definition of masking. The $size(\mathbf{M})/sum(\mathbf{M})$ part is a scaling factor that adjusts according to the number of unmasked input elements/pixels.

The mask is updated after each convolution operation by this equation:

$$m' = \begin{cases} 1, & if sum(\mathbf{M}) > 0 \\ 0, & otherwise \end{cases}$$

If the convolution was able to condition its output on at least one valid input value ($sum(\mathbf{M}) > 0$), the pixel will be labeled as unmasked. In this way, if the number of partial convolution layer is large enough, any mask will eventually be all removed and thus the image is recovered.

3.2 Loss Functions

We use the same loss function as defined by Liu et al. in their original paper of partial convolution. The loss functions leverages both per-pixel reconstruction accuracy and how well the pixel fits in the surrounding context. In the original paper the loss function has multiple components and each has very complex definition. But the overall loss funtion is the following.

$$\mathcal{L}_{total} = \mathcal{L}_{valid} + 6\mathcal{L}_{hole} + 0.05\mathcal{L}_{perceptual} + 120\left(\mathcal{L}_{style_{out}} + \mathcal{L}_{style_{comp}}\right) + 0.1\mathcal{L}_{tv}$$

3.3 Transfer learning Implementation

As described in the original paper, the model's network has a UNet-like architecture. Compared with general U-net models, the only difference is the convolution layers is replaced by the partial convolution layers we mentioned above. The skip connection in the U-nets connects both the feature and the masks. In the original paper, the authors suggested the training process took ten days on a single Nvidia V100 GPU. However, we do not have as much resource as they had at Nvidia. So in our project, we use the implementation of the Partial convolution layer from Gruber[9] and the pre-trained model weights as a starting point of our project. After proper croping and resizing of images, we then fine-tune the model on the traditional Chinese landscape dataset to make adapt the original model to our specific task.

4 Experiments

4.1 Dataset

We want to find a high-quality traditional Chinese landscape paintings. At the beginning, we use Google search engines to find dataset of Chinese landscape paintings. However, we encounter many problems. Firstly, we get many different size of paintings. We want to target at a lot of same size of paintings. Secondly, the quality and quantity of paintings are uneven, some are good and some are bad. Then, we find a collection of traditional Chinese landscape painting. [10]The dataset has totally 2,192 high-quality traditional Chinese landscape paintings collected from Princeton University Art Museum(362 paintings), Harvard University Art Museum(101 paintings), Metropolitan Museum of Art(428 paintings), Smithsonian's Freer Gallery of Art(1,301 paintings). All paintings are sized 512×512. Those paintings are high-quality and uniform size in preparation for training.

4.2 Irregular Mask Dataset

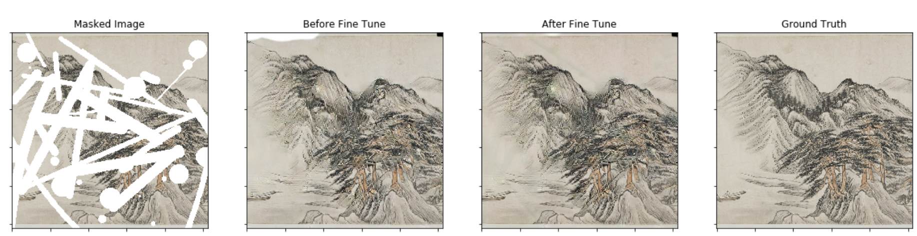
In order to simulate the holes, stains, or damage on traditional Chinese landscape paintings, we create mask generation with OpenCV. We start by generating masks of random holes and stain of arbitrary shapes in the same way as described in the original paper. In total, 100 masks for training and 50 masks for testing was generated. All the masks for training and testing are with the size of 512×512. We create a test set by starting with the 50 raw masks and adding random dilation, rotation and cropping. We split the test set into two: masks with and without holes close to border.

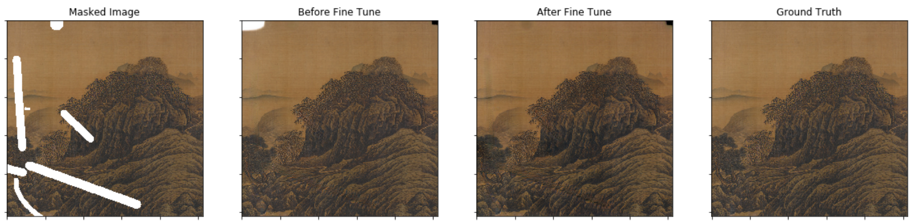
4.3 Training Process

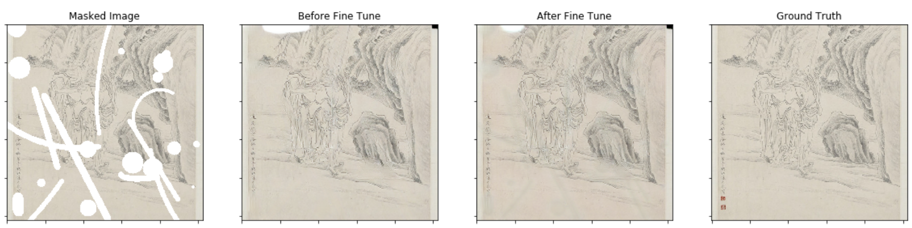
As we have discussed, the original partial convolution network was trained on Imagenet for 10 days using a single GPU. Since it's not feasible for our project, we used the pre-trained weights as a starting point, and fine-tuned our model on the traditional Chinese landscape dataset. Due to time and resource constraint, we were only able to use 10% of the whole dataset as training set to tune the model and another 5% of the dataset as the validation set. We then sampled another 5% of data as the test set. Because of unstable GPU connection, we could only finish 15 epoches of training on our local CPU.

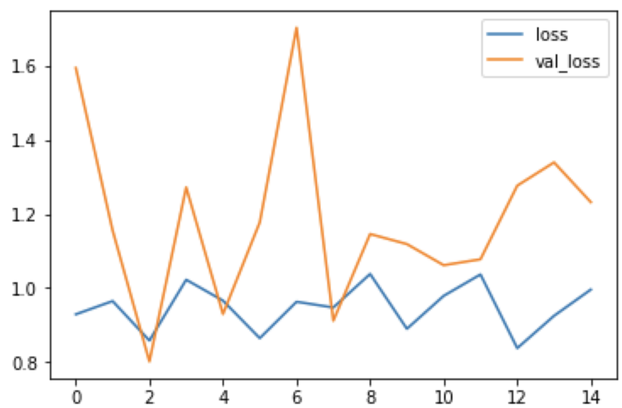
4.4 Results

We have successfully generated images from the traditional Chinese landscape paintings with irregular masks after fine-tuning. We can observe some improvements from figures below over the pre-trained model by comparing with the images generated by the fine-tuned model. However, due to our limited training time and amount of data, we do not see any meaningful change in the loss curve.









5 Summary

Based on the result, we can perfectly recover to a complete, flawless traditional Chinese landscapes paintings. If we have more resource, we can also get pretty low loss value. In the future, we can apply our methods to western paintings with stains or damaged or even to destroyed architect, pottery and so on for recovering mission.

6 Contribution

Zheng Li: collecting dataset and existing code/ library, fine-tuning task, writing report (Method, Training Process, Results, Contribution)

Wei-Ru Lin: literature survey, collecting dataset and existing code/ library, writing report (Abstract, Introduction, Related Work, Dataset, Irregular Mask Dataset, Summary, Contribution)

References

[1] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 85–100.

[2] X. Wang and A. Gupta, "Unsupervised learning of visual representations using videos," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2794–2802.

[3] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 417–424.

[4] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera, "Filling-in by joint interpolation of vector fields and gray levels," *IEEE transactions on image processing*, vol. 10, no. 8, pp. 1200–1211, 2001.

[5] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of graphics tools*, vol. 9, no. 1, pp. 23–34, 2004.

[6] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001, pp. 341–346.

[7] V. Kwatra, I. Essa, A. Bobick, and N. Kwatra, "Texture optimization for example-based synthesis," in *ACM SIGGRAPH 2005 Papers*, 2005, pp. 795–802.

[8] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, p. 24, 2009.

[9] M. Gruber, "Pconv-keras," https://github.com/MathiasGruber/PConv-Keras, 2019.

[10] A. Xue, "End-to-end chinese landscape painting creation using generative adversarial networks," 2020.