Image Inpainting using Partial Convolution

Minting Chen
Electrical and Computer Engineering
A59004799

Xinyue Wei
Electrical and Computer Engineering
A59002575

Abstract

Image inpainting is the task of reconstructing the hole-content of a corrupted image, which has a wide application in daily life. In recent years, deep learning is applied to this area and achieved good performance, while previous works only focus on rectangle holes in image center, and irregular holes are rarely explored. The method we implement in this project focuses on recovering irregular holes and proposes a novel *partial convolution*. It only utilizes valid elements within a feature map and removes the effect of initial hole values. We design experiments on CIFAR-10 and Paris dataset with different types of irregular hole masks to show the model performance.

1 Problem Definition

Image inpainting is a useful image editing technique for removing unwanted parts. It also can be used for image repairing. Solving image inpainting problem with deep learning-based methods has been a popular choice for many researchers. Deep learning-based methods are more robust and efficient than traditional methods, like the one implemented in PatchMatch [1]. Context Encoder [5] describes some basic concepts of image inpainting. It utilizes channel-wise fully connected layers in the context encoder; so the current layer can capture all features from the previous layer. However, Context Encoder [5] only works on images with rectangular holes and it makes predictions based on both valid and invalid pixels, which could cause color discrepancy and blurriness. In contracts, ParticalConv [4], the first paper that handles irregular holes, makes predictions only on existing valid pixels. Some interesting highlights from [4] make it a paper that we want to re-implement: (1) the mask (holes) can be automatically updated after the convolution layers; (2) a comprehensive loss function that takes not only the L1 loss and perceptual loss but also style loss and TV loss into account. Based on our understanding, the key and challenges of the image inpainting task are making the prediction (filling pixel) fits in the input image. For example, unsatisfying output could be produced due to huge corrupted areas or complicated scenes and structures in the input image.

2 Tentative Method

In this section, we describe the details of the method and the reason why we choose it. The method we implement for this project is the Partial Conv proposed by Liu *et al.* [4]. The key idea is to replace the normal convolution with a novel *partial convolution*, in order to avoid the usage of invalid mask pixels. A mask updating mechanism is proposed to automatically adjust mask size for the next layers.

Partial Convolution. Compared to normal convolution, partial convolution takes the mask into consideration. It selects the valid pixels to do the convolution operation according to mask, which can be written as:

$$x' = W^T(X \circ M)\frac{N}{N_{valid}} + b, \text{if } N_{valid} > 0 \tag{1}$$

where W denotes weights of convolution filters, X denotes input features, M denotes the corresponding binary mask map and $\circ$ denotes element-wise multiplication; let N be the number of total elements within M and N_{valid} be the number of valid pixels. The partial convolution only works when $N_{valid} > 0$, thus for an input features with all elements invalid, the output is zero.

The mask updating mechanism works on the mask M after each partial convolution operation. For each M, if $N_{valid} > 0$, then the whole region is update to be valid. As long as the partial convolution was able to acquire valid information from input features, all the locations in the corresponding mask will be valid in the next operation. Note the mask is updated by a convolution operation together with features, thus keeping the corresponding size all the time.

Network Architecture. The whole inpainting network is based on a U-Net architecture similar to [2], with normal convolution replaced by partial convolution. The encoder stage contains a couple of partial convolution layers with a stride of two to downsample the features of each operation; the decoder stage uses the nearest neighbor upsampling layer to recover features into the original resolution. The last convolution layer takes features, the original image and original mask as input, making it possible for the model to copy valid pixels directly.

Loss Functions. In this project, we utilize four types of loss functions in all to achieve the best reconstruction quality of corrupted images, including L_1 content loss, perceptual loss and style loss, as well as total variation (TV) loss. Content loss is widely used in image generation topic, which restricts the consistency between the predicted image and ground truth at pixel level. Perceptual loss and style loss are introduced for constraining higher-level features, to better recover textures. And total variation loss is used for removing *checkerboard artifacts* generated by perceptual loss, as suggested by [3]. All these four losses are combined during optimization.

Partial convolution network is good at dealing with irregular holes, which is rarely explored in previous works. People used to focus on the center square hole, which is not applicable enough in various situations. Besides, unlike other neural network approaches, partial convolution avoids using invalid pixels, thus has no dependency on the initial hole values, which leads to better texture reconstruction.

3 Experiments

PartialConv [4] uses three datasets for training and testing: ImageNet, Places2, and CelebA-HQ. So we would like to implement the model on different datasets like Paris [6] and CIFAR-10. The Paris dataset contains 6412 images of some famous landmarks in Paris. CIFAR-10 is another dataset that is commonly used for deep learning. There are 60000 images in 10 different classes. For the mask dataset, we are planning to use QD-IMD which contains masks drew by human hand. The first experiment we are planning to perform is to re-implement the PartialConv network architecture; then we use the network to test the performance. The second experiment we will conduct is to test the trained network on masks with different hole ratios. From this experiment, we can see the relationship between the mask size versus the performance of the model.

References

[1] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 28(3), August 2009.

[2] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[3] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, pages 694–711. Springer, 2016.

[4] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 85–100, 2018.

[5] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. Context encoders: Feature learning by inpainting, 2016.

[6] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.