Dunhuang Image Restoration with Partial Convolution and Structural Similarity Loss

B05705014 鄭揚 B05705022 李信穎 B05705043 唐瑋廷

Abstract

In this project, we employed NVIDIA's Partial Convolutions with an U-Net architecture and structural similarity loss function to restore incomplete Dunhuang images. The model surpassed baseline with quantitative evaluation and produced high-quality restoration result.

Partial Convolutions

In general implementation of Convolutional Neural Network (CNN), all kernels perform convolution with all pixels of images. However, in the scenario of image restoration and inpainting, hold regions become noise during encoding steps. To reduce such noise, Liu et al. [1] proposed Partial Convolution, where the convolution is masked and renormalized to be conditioned on only unmasked pixels. Updated mask for the next layer are also automatically generated and passed forward.

$$x' = \begin{cases} W^T(x \odot M)\frac{1}{sum(M)} + b, & if\ sum(M) > 0 \\ 0, & otherwise \end{cases}$$

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

Structural Similarity Loss Function

To achieve higher SSIM score and directly force model to learn from it, we made SSIM metric differentiable through PyTorch implementation of SSIM using Gaussian filter convolutions. The SSIM loss is then subtracted from the loss so that a higher SSIM will result in lower loss.

Training

- We randomly cropped 400*400 pixel patch from each image for mini-batch training.
- We used Adam optimizer with learning rate 5e-4 and batch size 8.

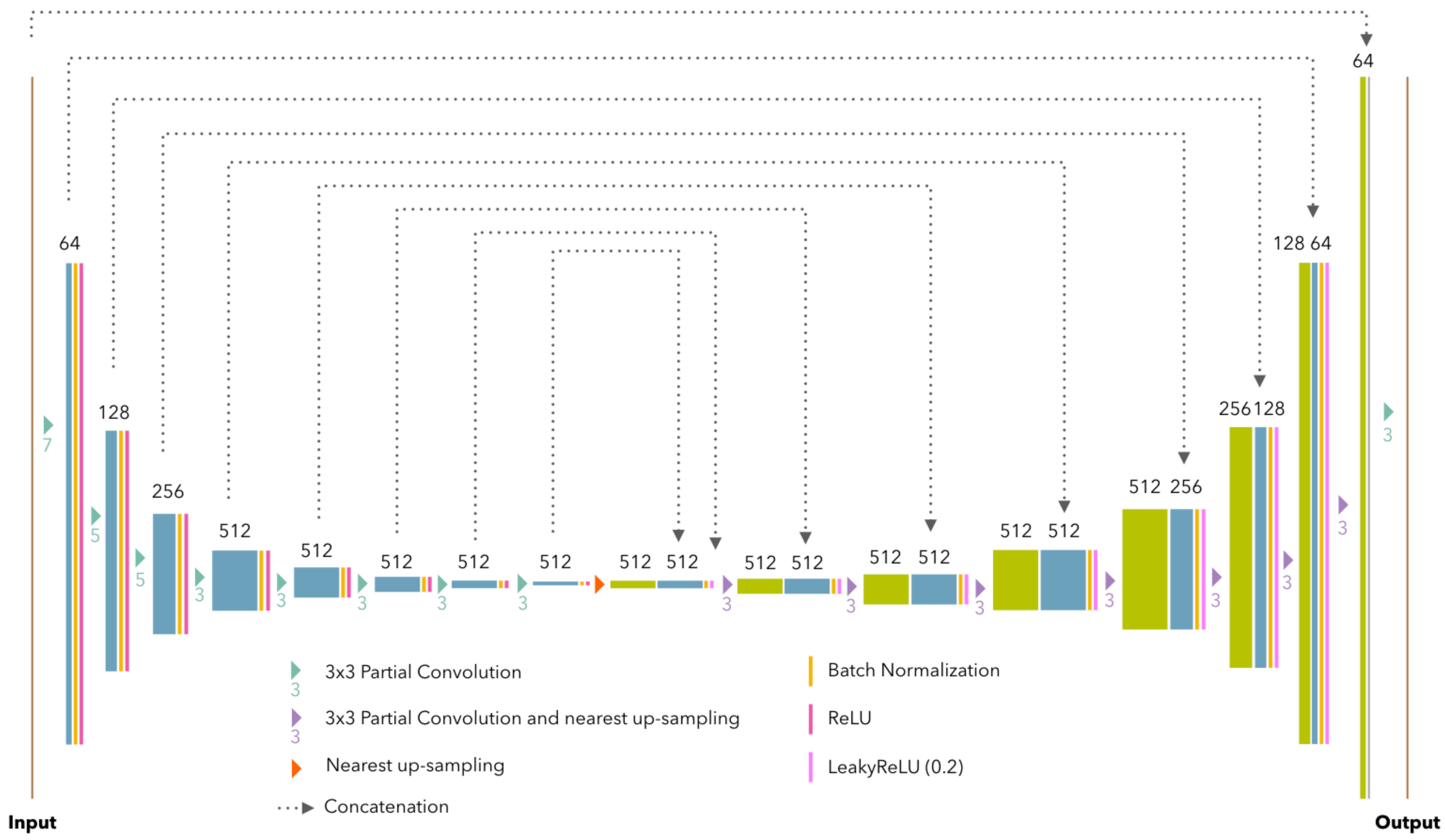
Ablation Study

	MSE	SSIM
PConv	39.373	0.786
PConv + SS	36.359	0.804
PConv + SS + post-process	35.153	0.805

Conclusion

We employed Partial Convolution and Structural Similarity loss function to perform image restoration, and proved both modules' effectiveness with experiments. Image inpainting is a challenging task attracting lots of works that we haven't explored in this project, such as Gated Convolution or coarse-to-fine network. We are exciting to keep diving deeper into the topic and can't wait to find more inspiring improvements.

Model Architecture



Loss

input of model: I_{in}

output of model: I_{out}

ground truth: I_{gt}

mask (1=valid, 0=hole): M

feature of I extracted by VGG16: ψ_I

$I_{comp} = M \odot I_{in} + (1 - M) \odot I_{out}$

Pixel Loss

$$\mathscr{L}_{hole} = \|(1 - M) \odot (I_{out} - I_{gt})\|_1$$

$$\mathscr{L}_{valid} = \|M \odot (I_{out} - I_{gt})\|_1$$

Style Reconstruction Loss

$$\mathscr{L}_{style_{out}} = \Sigma_n \|\psi_{I_{out}^T} \psi_{I_{out}} - \psi_{I_{gt}^T} \psi_{I_{gt}}\|_1$$

$$\mathscr{L}_{style_{comp}} = \Sigma_n \|\psi_{I_{comp}^T} \psi_{I_{comp}} - \psi_{I_{gt}^T} \psi_{I_{gt}}\|_1$$

Perceptual Loss

$$\mathscr{L}_{perceptual} = \|\psi_{I_{out}} - \psi_{I_{gt}}\|_1 + \|\psi_{I_{comp}} - \psi_{I_{gt}}\|_1$$

Total Variance Loss

$$\mathscr{L}_{tv} = \Sigma_{i,j} \|I_{comp}^{i,j+1} - I_{comp}^{i,j}\|_1 + \Sigma_{i,j} \|I_{comp}^{i+1,j} - I_{comp}^{i,j}\|_1$$

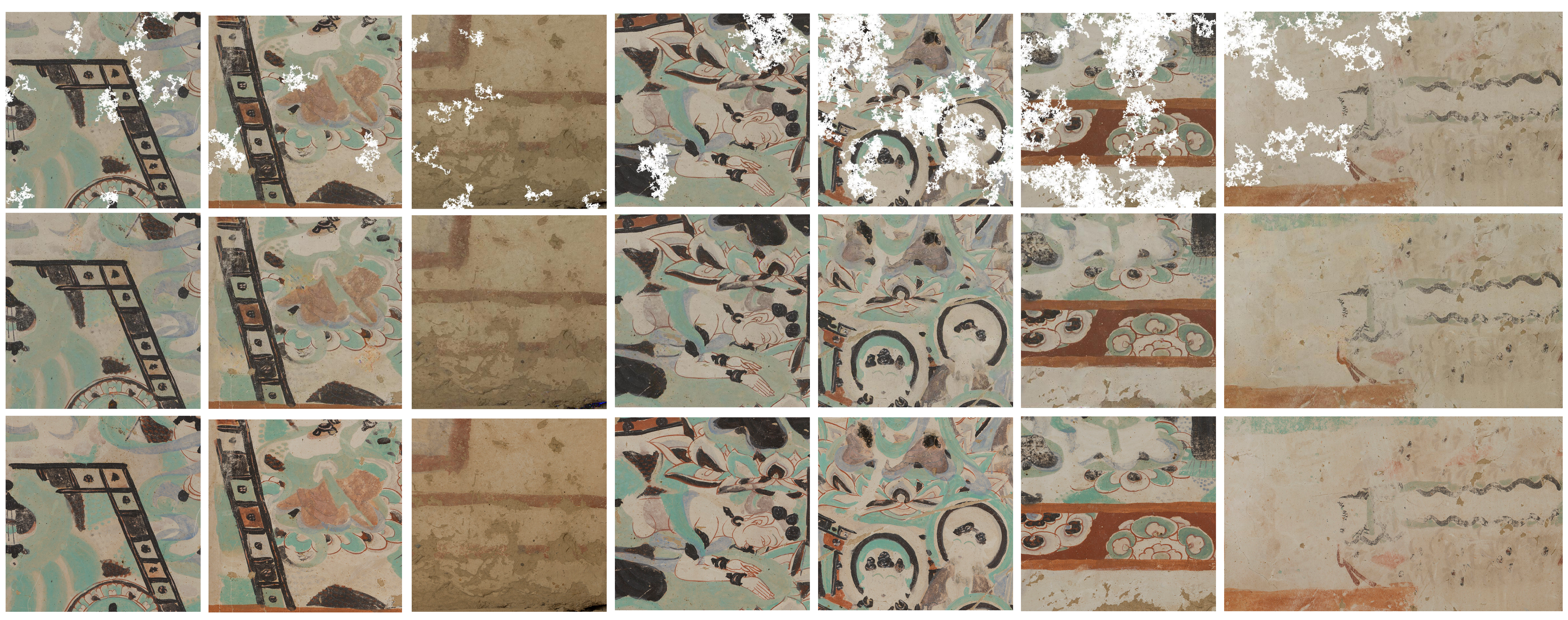
Structural Similarity Loss

$$\mathscr{L}_{struct} = -SSIM(I_{out}, I_{gt})$$

Total Loss

$$\mathscr{L} = \mathscr{L}_{valid} + 6\mathscr{L}_{hole} + 0.05\mathscr{L}_{perceptual} + 120(\mathscr{L}_{style_{out}} + \mathscr{L}_{style_{comp}}) + 0.1\mathscr{L}_{tv} + 5\mathscr{L}_{struct}$$

Result



Rows from up to down : input images, restoration result, ground truth

Reference

1. G. Liu, F. A. Reda, K. J. Shih, T. Wang, A. Tao and B. Catanzaro. Image Inpainting for Irregular Holes Using Partial Convolutions. In The European Conference on Computer Vision (ECCV) 2018
   Official Github : https://github.com/NVIDIA/partialconv
   Initial model weight from : https://github.com/naoto0804/pytorch-inpainting-with-partial-conv
2. Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. In IEEE Transactions on Image Processing, vol. 13, no. 4, Apr. 2004