

# Shift-Net: Image Inpainting via Deep Feature Rearrangement

Zhaoyi Yan<sup>1</sup>[0000–0001–5155–5741], Xiaoming Li<sup>1</sup>[0000–0003–3844–9308], Mu Li<sup>2</sup>[0000–0002–7327–3304], Wangmeng Zuo<sup>1</sup>(✉)[0000–0002–3330–783X], and Shiguang Shan<sup>3</sup>[0000–0002–8348–392X]

- <sup>1</sup> School of Computer Science and Technology, Harbin Institute of Technology, China  
yanzhaoyi@outlook.com, csxmli@hit.edu.cn, wzmzuo@hit.edu.cn
- <sup>2</sup> Department of Computing, The Hong Kong Polytechnic University, Hong Kong  
csmuli@comp.polyu.edu.hk
- <sup>3</sup> Institute of Computing Technology, CAS, Beijing, 100049, China  
sgshan@ict.ac.cn

**Abstract.** Deep convolutional networks (CNNs) have exhibited their potential in image inpainting for producing plausible results. However, in most existing methods, e.g., context encoder, the missing parts are predicted by propagating the surrounding convolutional features through a fully connected layer, which intends to produce semantically plausible but blurry result. In this paper, we introduce a special shift-connection layer to the U-Net architecture, namely Shift-Net, for filling in missing regions of any shape with sharp structures and fine-detailed textures. To this end, the encoder feature of the known region is shifted to serve as an estimation of the missing parts. A guidance loss is introduced on decoder feature to minimize the distance between the decoder feature after fully connected layer and the ground-truth encoder feature of the missing parts. With such constraint, the decoder feature in missing region can be used to guide the shift of encoder feature in known region. An end-to-end learning algorithm is further developed to train the Shift-Net. Experiments on the Paris StreetView and Places datasets demonstrate the efficiency and effectiveness of our Shift-Net in producing sharper, fine-detailed, and visually plausible results. The codes and pre-trained models are available at <https://github.com/Zhaoyi-Yan/Shift-Net>.

**Keywords:** Inpainting · feature rearrangement · deep learning

## 1 Introduction

Image inpainting is the process of filling in missing regions with plausible hypothesis, and can be used in many real world applications such as removing distracting objects, repairing corrupted or damaged parts, and completing occluded regions. For example, when taking a photo, rare is the case that you are satisfied with what you get directly. Distracting scene elements, such as irrelevant people or disturbing objects, generally are inevitable but unwanted by the



































