Robust Image Outpainting With Learnable Image Margins

Cheng-Yo Tan*, Chiao-An Yang*, Shang-Fu Chen, Meng-Lin Wu, Yu-Chiang Frank Wang

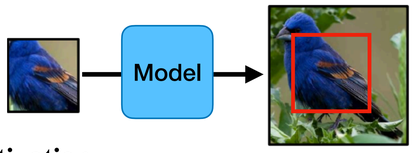
國立臺灣大學 National Taiwan University

Qualcomm

ICIP 2021
Paper ID: 1520

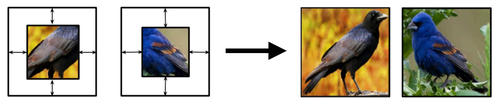
Introduction

- **Image Outpainting**
  - Definition: Given an input image, synthesize image pixels with limited guidance of its surrounding regions.

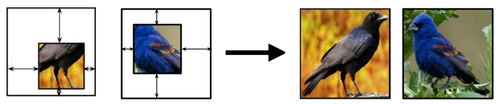


- **Motivation**
  - While existing image outpainting methods achieve impressive results, they either lack the ability to extend image regions in arbitrary directions or require the filling image margins to be given in advance.

(1) fixed margin



(2) arbitrary margin (SRN)



(3) no margin (Ours)

- **Contributions**

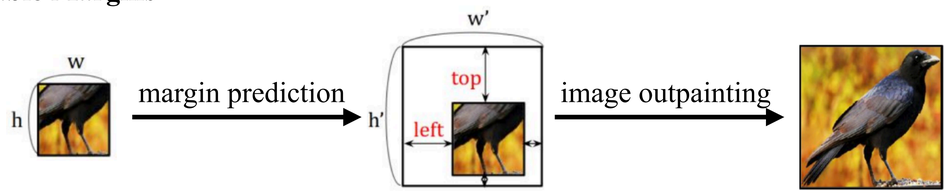
(1) We present a deep learning framework for image outpainting, which outpaints input images in all four possible directions **without the filling margins to be known** in advance.

(2) The **margin prediction network** introduced in our framework takes the partial image input and predicts desirable image margins for outpainting purposes.

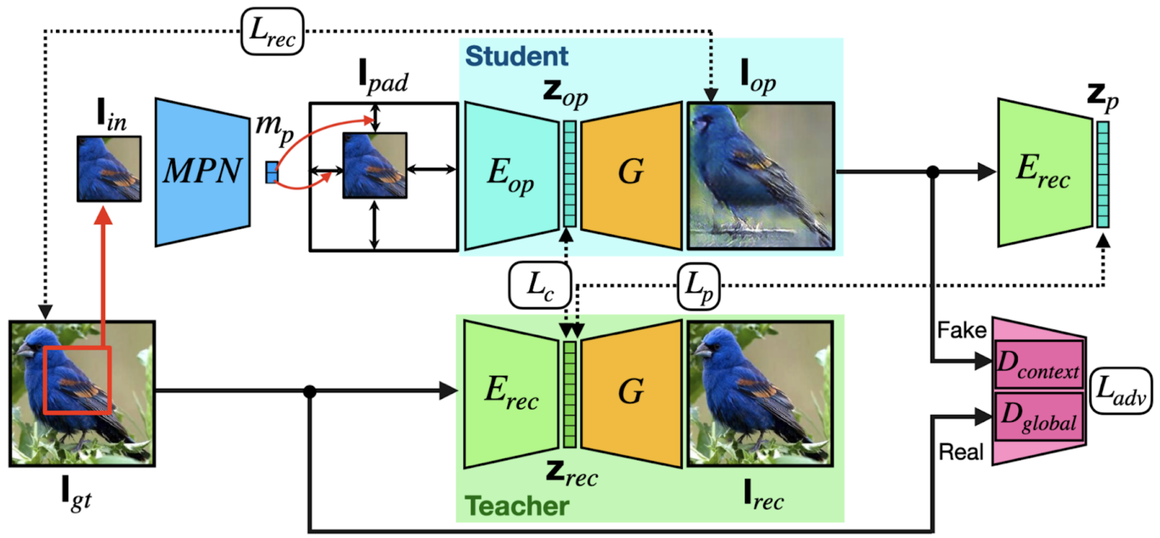
(3) Regularized by a **teacher auto-encoder network**, our encoder-decoder outpainting network serves as a student module for image outpainting with both image appearance recovery and perceptual feature consistency jointly observed

Approach

- **Learnable Margins**



- **Framework Overview**



- **Margin Prediction Network (MPN)**
  - We propose a Margin Prediction Network (MPN) to leverage visual cues from input images to predict corresponding filling margins.
  - During inference, MPN predicts the filling margins of the input images, and then an encoder-decoder network outpaints the input images according to its predicted margins.

- **Teacher-Student Network for Image Outpainting**
  - To improve the quality of the outpainted images, we introduce an auto-encoder teacher network in our training framework to help guide the encoder-decoder student network with both image-level appearance recovery and feature-level consistency guarantees.
  - The teacher network ensures the student network extracts features with sufficient information for fair reconstruction and satisfactory context for image outpainting.

- **Learning Constraints**

(1) margin prediction
$$L_m = \|m_{gt} - m_p\|_1$$

(2) image reconstruction
$$L_{rec} = \|I_{gt} - I_{op}\|_1$$

(3) contextual adversarial
$$L^n_{adv} = E_{I_{gt}}[D_n(I_{gt})] - E_{I_{op}}[D_n(I_{op})] + \lambda_{gp} E_{\bar{I}}[(\|\nabla_{\bar{I}} D_n(\bar{I})\|_2 - 1)^2]$$
where $n \in \{context, global\}$

(4) feature consistency
$$L_c = \|z_{rec} - z_p\|_2$$

(5) perceptual loss
$$L_p = \|z_{rec} - z_{op}\|_2$$

Experiments

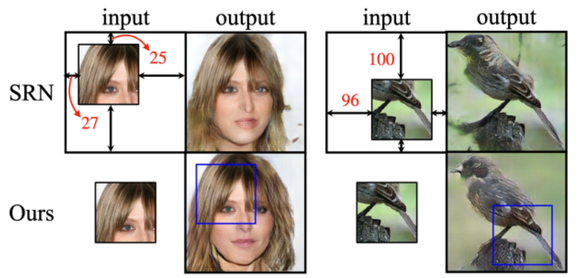
- **Datasets**
  - CelebA-HQ
    28000/2000 training/testing face images
  - CUB200
    10000/1788 training/testing bird images

- **Implementation Details**
  - Given input images with a resolution of 128 × 128 pixels, our models outpaint images to 256 × 256 pixels.
  - Our MPN consists of a ResNet-50 as the feature ex-tractor followed by a fully connected layer with two output units.
  - During training, we use Adam optimizers with a learning rate of 0.0001.

- **Quantitative Comparisons**

	CelebA-HQ-2K			CUB200-1.7K		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
ED	14.25	0.5961	0.3237	14.85	0.5923	0.4035
SRN [10]	14.48	0.6016	0.3203	15.13	0.6207	**0.3856**
Ours	14.10	0.5040	0.3313	13.82	0.4672	0.4983
Ours*	**14.54**	**0.6112**	**0.3110**	**15.62**	**0.6319**	0.3943

- **Visual Comparisons**



- **Visualization of outpainting with arbitrarily cropped partial inputs**