Zoom-to-Inpaint: Image Inpainting with High-Frequency Details

Soo Ye Kim[1] Kfir Aberman[2] Nori Kanazawa[2] Rahul Garg[2] Neal Wadhwa[2]
Huiwen Chang[2] Nikhil Karnad[2] Munchurl Kim[1] Orly Liba[2]

[1]KAIST [2]Google Research

Contact: sooyekim@kaist.ac.kr

NTIRE Workshop

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

INTRODUCTION

★ Image inpainting

- Fill in missing regions in a masked image
- Output should be a natural completion of the scene with (i) plausible semantics, and (ii) realistic details and textures
- Difficult due to spectral bias: neural networks are biased towards learning low-frequency components

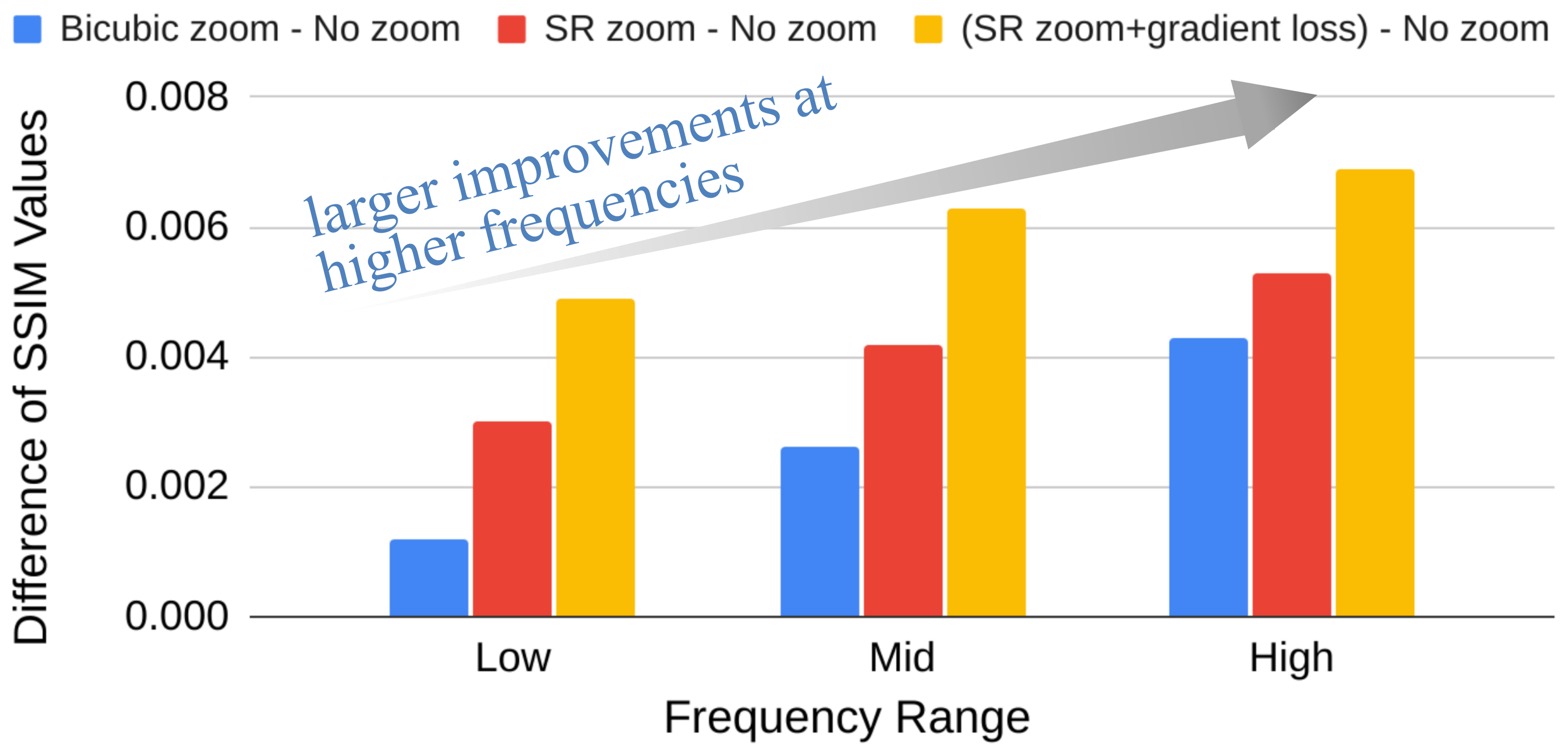
★ Super-resolution (SR)

- Upscale a low-resolution (LR) image to high-resolution (HR)

ANALYSIS

★ Frequency domain comparison

- 2-level Laplacian pyramid with 5-tap Gaussian kernel
- SSIM differences plotted vs "No zoom" model (baseline)
- Bicubic zoom: bicubic upsampling instead of SR network



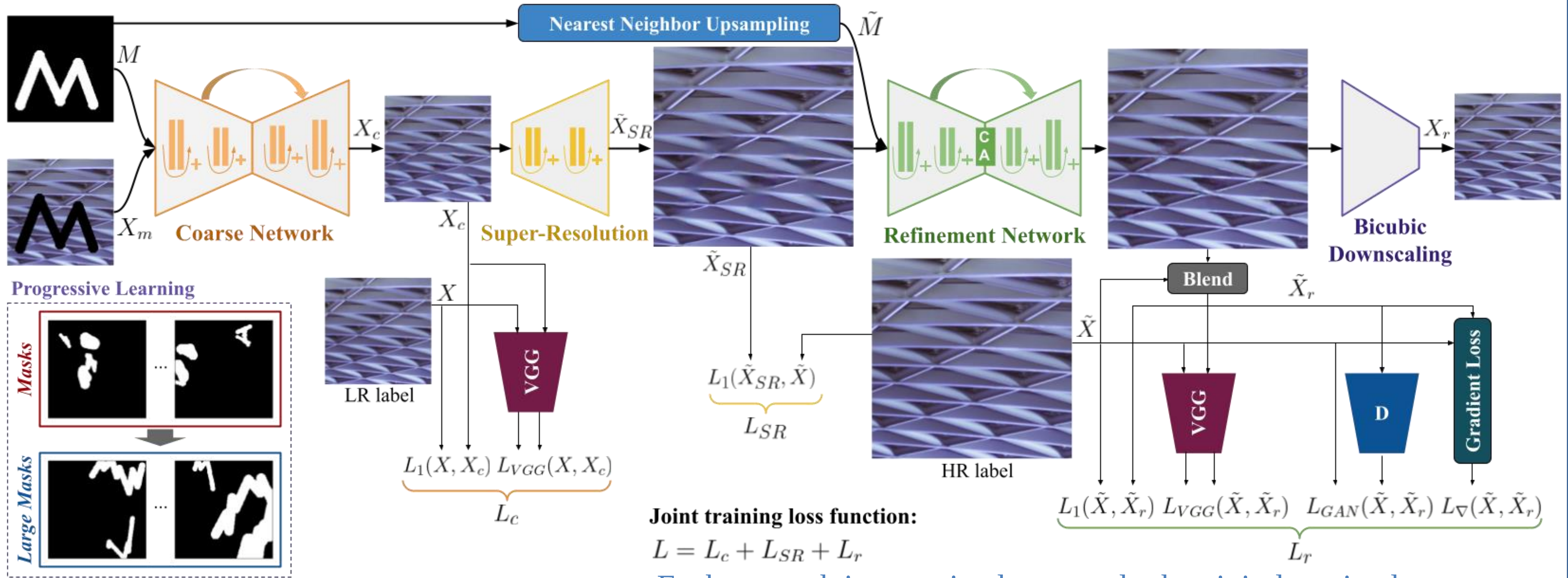
★ User study

Method	HiFill	Pluralistic	DeepFill-v2	EdgeConnect
Preference of Ours over compared	75.49%	89.13%	69.23%	64.21%

PROPOSED METHOD

★ Zoom-to-Inpaint

- Coarse network → Zoom in → Refinement network → Zoom out
- SR network is inserted between the coarse network and the refinement network for zooming in
- Refine after zooming in to correct local irregularities at a finer level
- Inject high-frequency details into the resulting image by learning from HR labels



Each network is pretrained separately then jointly trained

★ Gradient loss for high-frequency details

- $L_\nabla = \frac{1}{2}\left(\left\|(\tilde{X}_r - \tilde{X})_{\nabla_x}\right\|_2^2 + \left\|(\tilde{X}_r - \tilde{X})_{\nabla_y}\right\|_2^2\right)$
- $\tilde{X}_r$: output of refinement network, $\tilde{X}$: ground truth HR label, ∇_x: horizontal image gradient, ∇_y: vertical image gradient

★ Progressive learning for faster convergence

- Train the network in N steps by increasing the size of masks
- If trained directly on amsks at n=2, 2M iterations needed
- With progressive learning, 1.2M iterations are sufficient

<Increasing masks for progressive learning (N=2)>

QUALITATIVE COMPARISON

★ Places2, DIV2K dataset