ICCV 2019

SROBB: Targeted Perceptual Loss for Single Image Super-Resolution

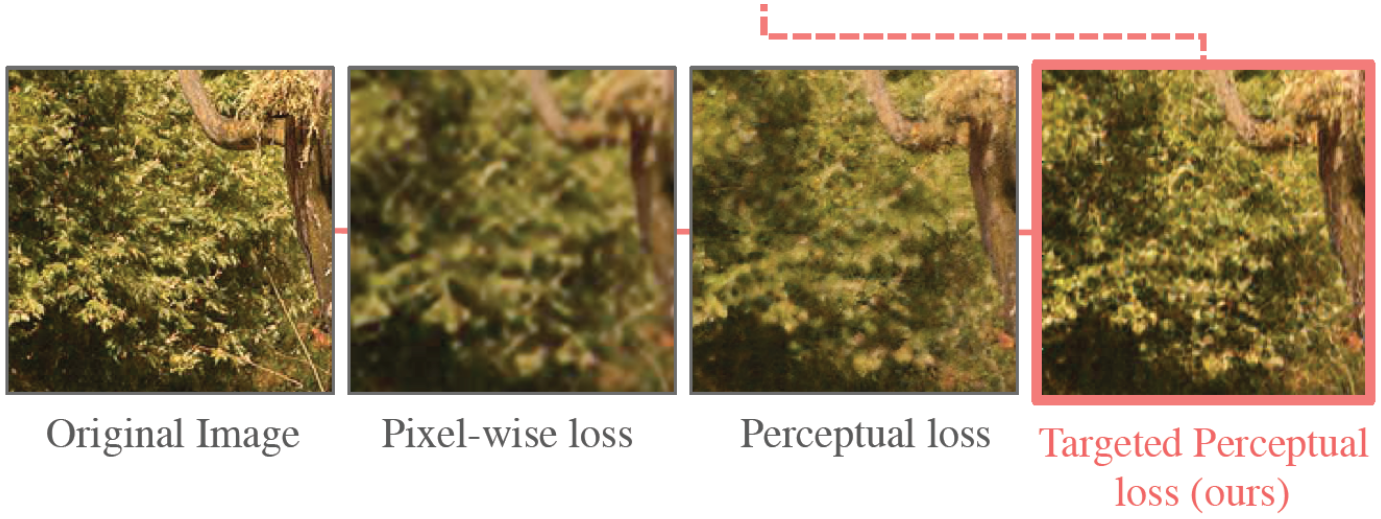
Mohammad Saeed Rad, Behzad Bozorgtabar, Urs-Viktor Marti,
Max Basler, Hazım Kemal Ekenel, Jean-Philippe Thiran

EPFL and Swisscom AG, Switzerland; ITU, Turkey

Slides compiled by Mengzhou Li

Introduction

- Perceptual loss helps a lot in generating photo-realistic images.
- The loss is usually calculated on the entire image without considering any semantic information.
- The authors notice that the features of object, background and boundary may need (1) **different weights**, and (2) different calculation strategies in perceptual loss to generate more photo-realistic images:

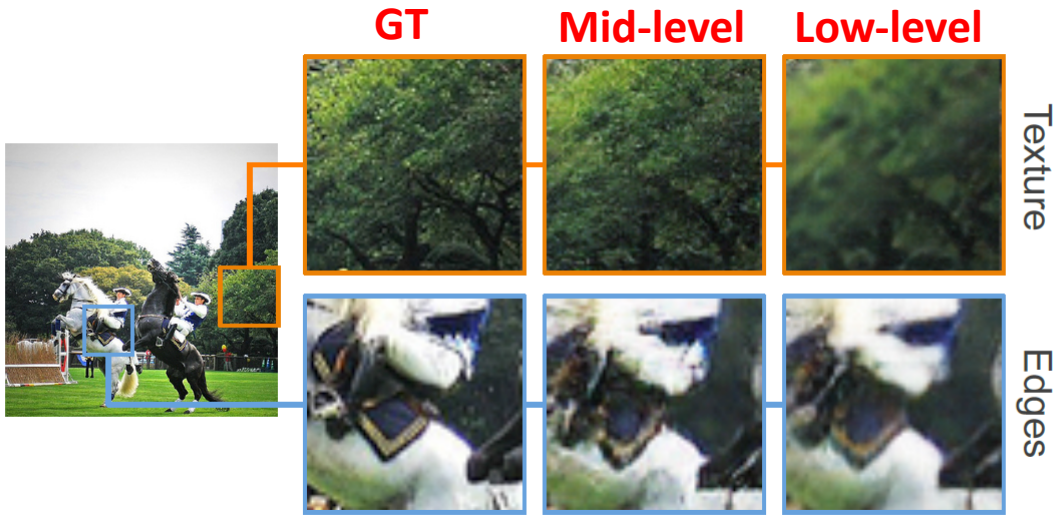


Minimizing the loss for details of the edges inside a **random texture** => unnecessary penalty and learn less informative features

The texture of a tree could still be realistic in the SR image without having close edges to the HR images.

Introduction

- Perceptual loss helps a lot in generating photo-realistic images.
- The loss is usually calculated on the entire image without considering any semantic information.
- The authors notice that the features of object, background and boundary may need (1) different weights, and (2) **different calculation strategies** in perceptual loss to generate more photo-realistic images:

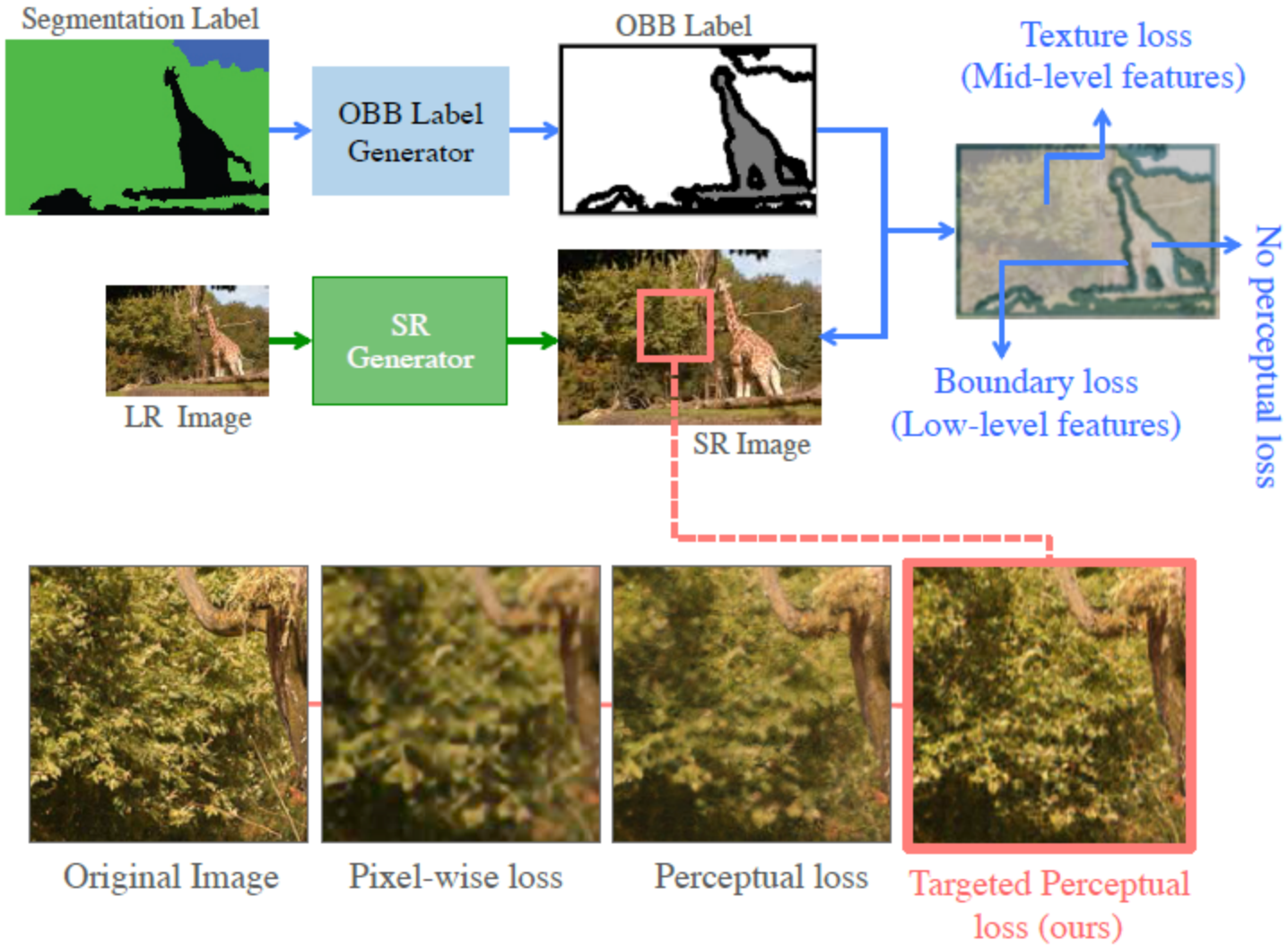


Mid-level => textures
Low-level => edges
High-level => semantic information

Low-level: ReLU 1-2, VGG-16
Mid-level: ReLU 4-1, VGG-16

Contributions

- Proposed a novel targeted perceptual loss function.
  - ➢Exploiting the segmentation labels during training;
  - ➢Edges at object boundaries;
  - ➢Texture on the background;



Principles

- Targeted loss function tries to favor more **realistic textures** around areas, where the type of the textures seems to be important, i.e., tree, while trying to resolve **sharper edges** around boundary area.
- Three types of regions:
  - Background (4 classes: sky, plant, ground and water)
    - Overall texture more important than local spatial relations and edges
    - Mid-level features to estimate the perceptual similarity (ReLU 4-3, AGG-16)
  - Boundaries
    - All edges separating objects and the background are considered as boundaries.
    - Low-level features to estimate the perceptual distance (ReLU 2-2, AGG-16)
  - Objects
    - Hard to tell which type of features is more important, textures or edges
    - Do not consider the perceptual loss, and only rely on the MSE and adversarial losses.
    - Intuitively, changes in “boundary” and “background” would result in more appealing objects.

Approaches

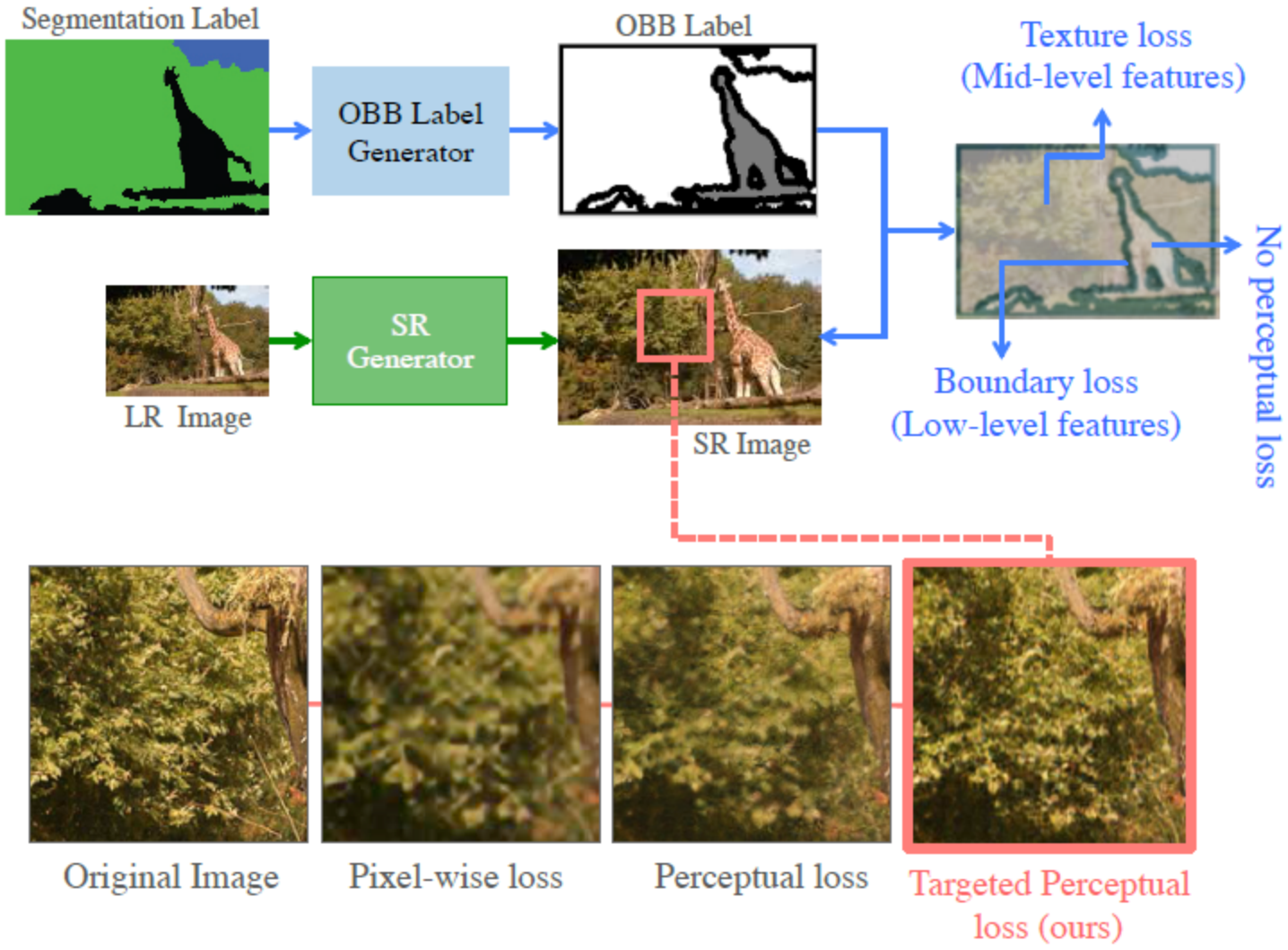
- Binary masks

$$M_{OBB}^{boundaries}$$

$$M_{OBB}^{background}$$

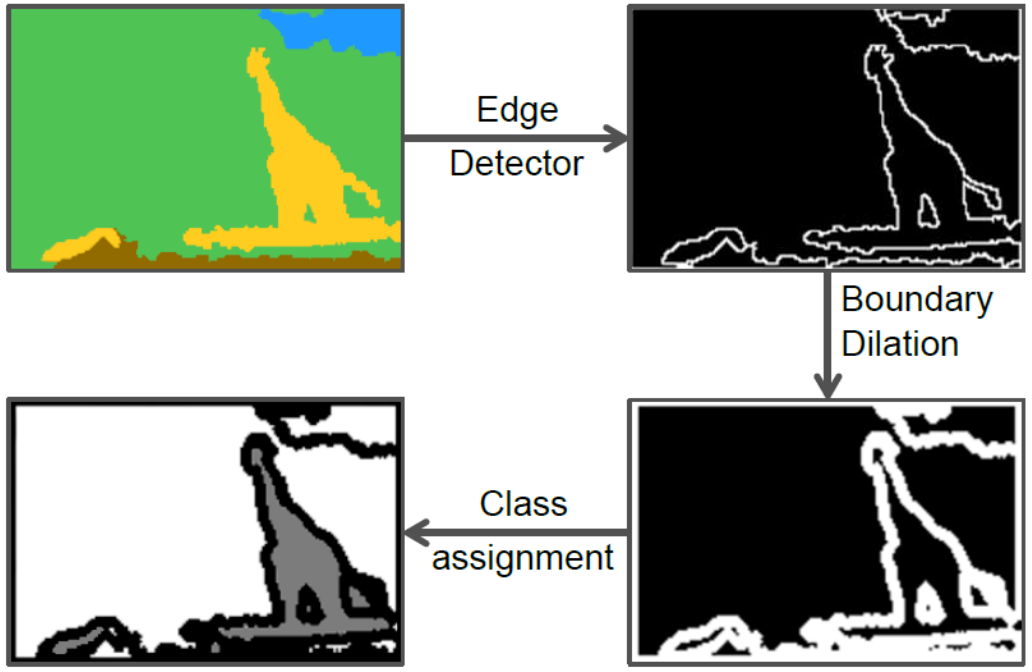
- Overall targeted perceptual loss

$$\mathcal{L}_{perc.} = \alpha \cdot \mathcal{G}_e(I^{SR} \circ M_{OBB}^{boundary}, I^{HR} \circ M_{OBB}^{boundary}) + \beta \cdot \mathcal{G}_b(I^{SR} \circ M_{OBB}^{background}, I^{HR} \circ M_{OBB}^{background})$$



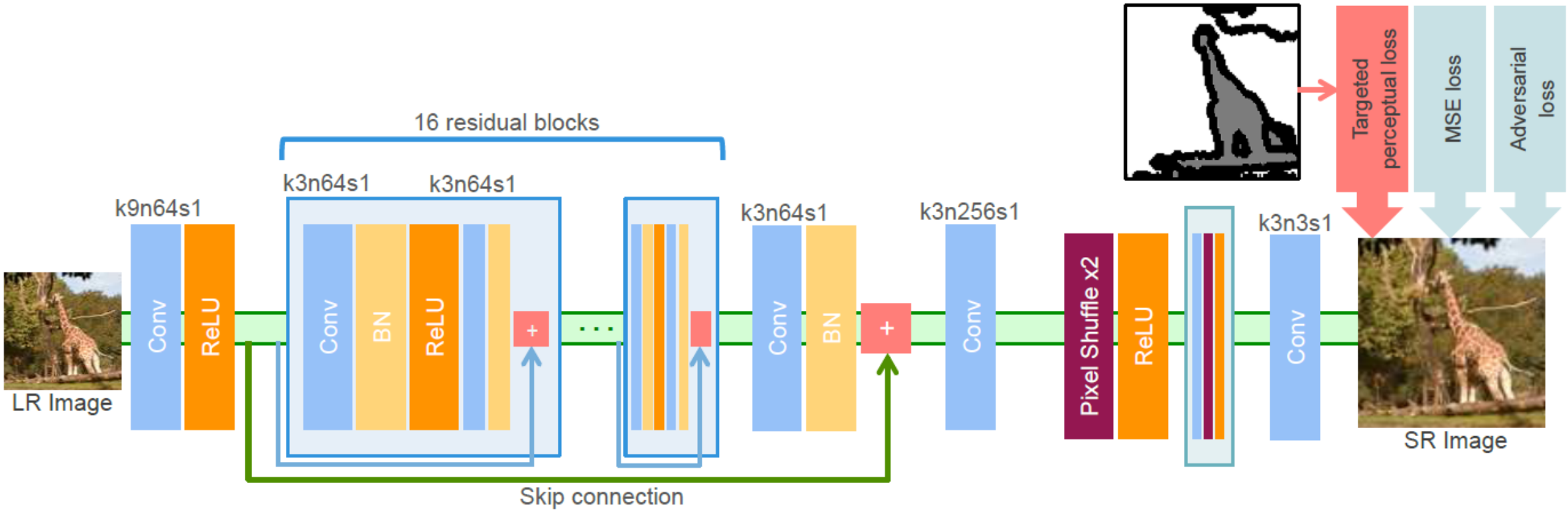
Approaches

- Binary masks
  - Calculate the derivative of the **segmentation label**
  - Dilation with a disk of d_1
  - Type assignments
    - *Boundary* class
    - The 4 classes from the segmentation labels as *background*
    - All remaining is *object*



Experiments

- Architecture (SRGAN, scale factor of 4)



Experiments

- Datasets
  - A random set of 50K images from the COCO-Stuff dataset
  - Only considered landscapes with one or more of the “sky”, “plant”, “ground”, and “water” classes. These classes are grouped into the “background” class.
  - MATLAB *imresize* (bicubic kernel) => Low resolution images
- Training
  - First 25 epochs with only MSE
  - Proposed targeted perceptual loss + adversarial loss for 55 more epochs

Testing (Set5 and Set14)

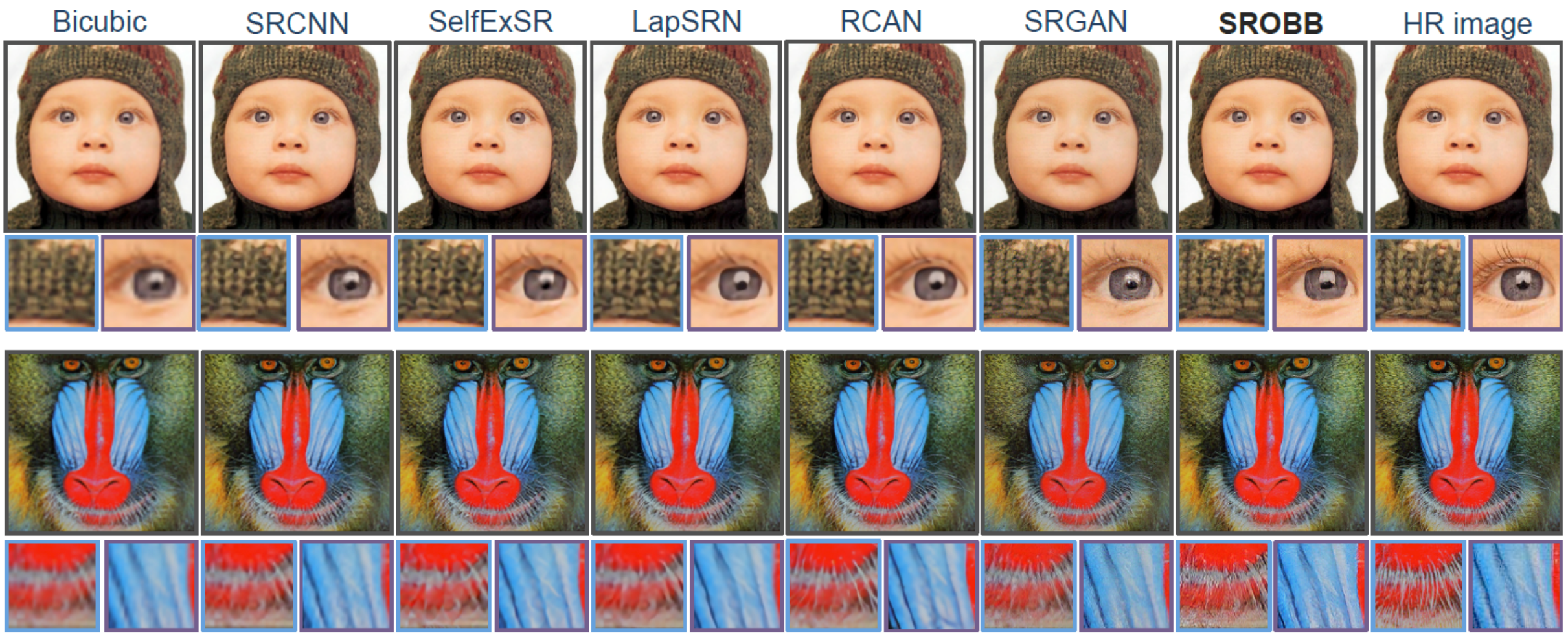
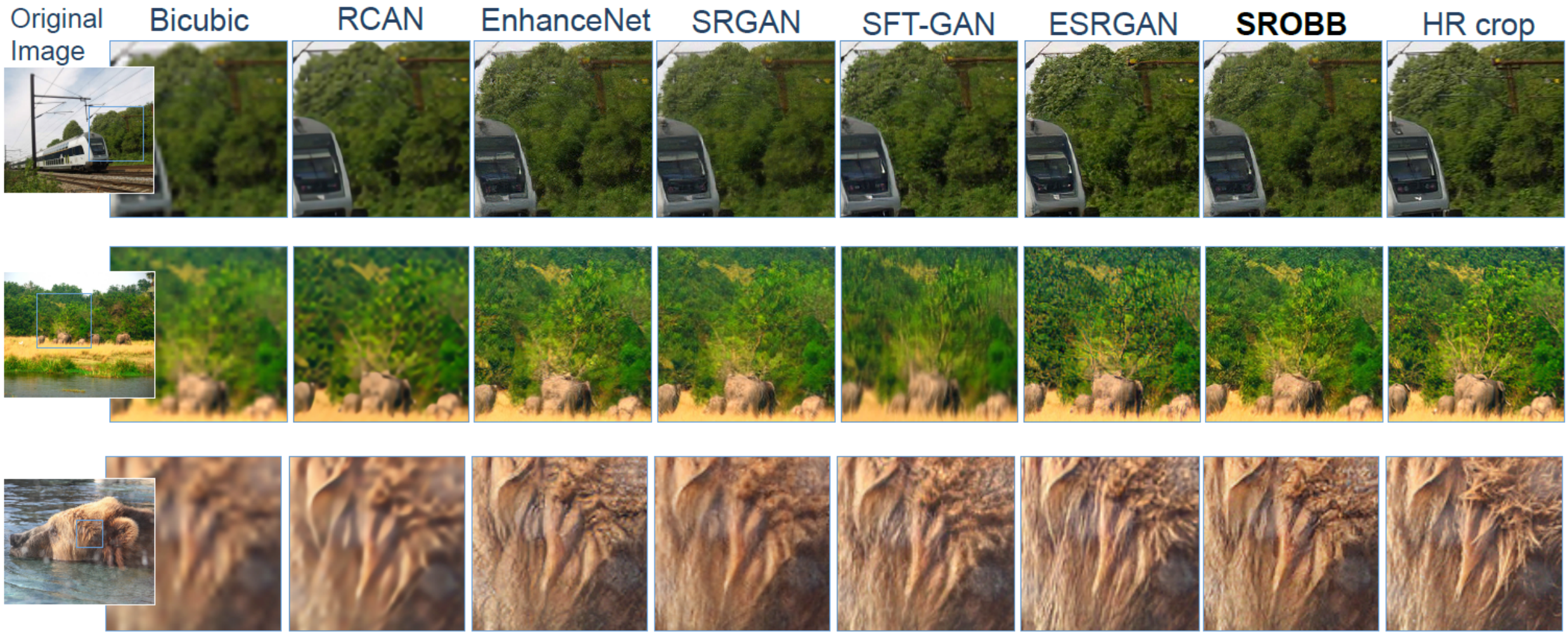


Figure 5. Sample results on the "baby" (top) and "baboon" (bottom) images from Set5 [1] and Set14 5 datasets, respectively. From left to right: bicubic, SRCNN [5], SelfExSR [14], LapSRN [19], RCAN [44], SRGAN [20] and SROBB (ours), HR image, respectively.

Testing (COCO-Stuff)



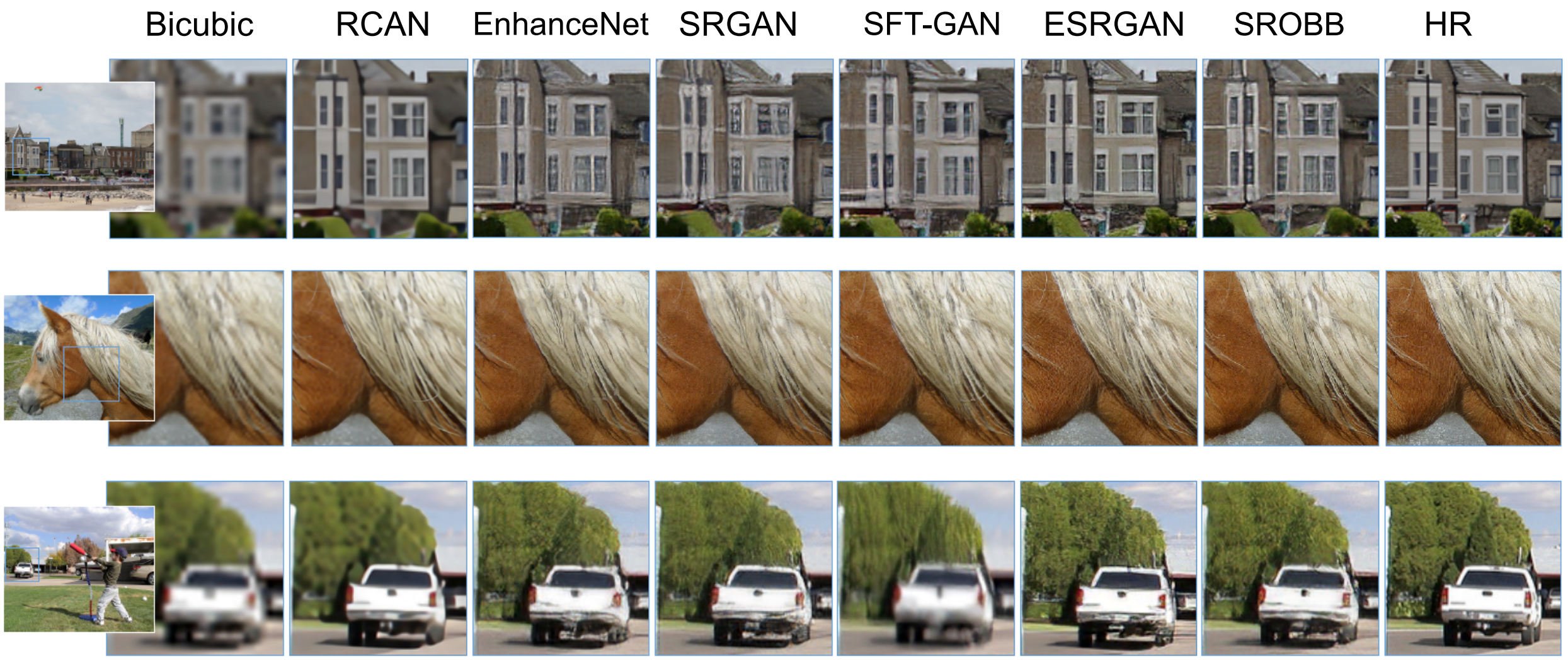


Figure 6. Qualitative results on a subset of the COCO-Stuff dataset [4] images. Cropped regions are zoomed in with a factor of 2 to 5 to have a better comparison. Results from left to right: bicubic, RCAN [44], EnhanceNet [27], SRGAN [20], SFT-GAN [35], ESRGAN [36], SROBB (ours) and a high resolution image. Zoom in for the best view.

ESRGAN **SROBB** HR crop

ESRGAN produces over-sharpened edges which sometimes leads to an unrealistic reconstruction and dissimilar to ground-truth.

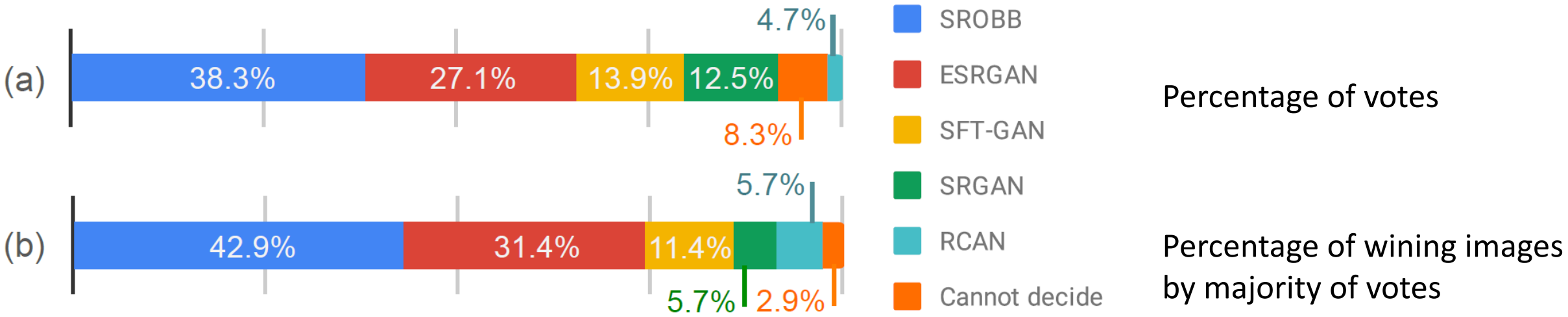
Quantitative Metrics (Set5 and Set14)

Image	Metric	Bicubic	LapSRN	SRGAN	SROBB
baby	SSIM	0.936	**0.951**	0.899	0.905
	PSNR	30.419	**32.019**	28.413	28.869
	LPIPS	0.305	0.237	0.112	**0.104**
baboon	SSIM	0.645	**0.677**	0.615	0.607
	PSNR	20.277	**20.622**	19.147	18.660
	LPIPS	0.632	0.537	**0.220**	0.245

Table 1. Comparison of bicubic interpolation, LapSRN [19], SRGAN [20] and SROBB (ours) for the "baby" and "baboon" images from Set5 and Set14 test sets. Best measures (SSIM, PSNR [dB], LPIPS) are highlighted in bold. The visual comparison is shown in Figure 5.

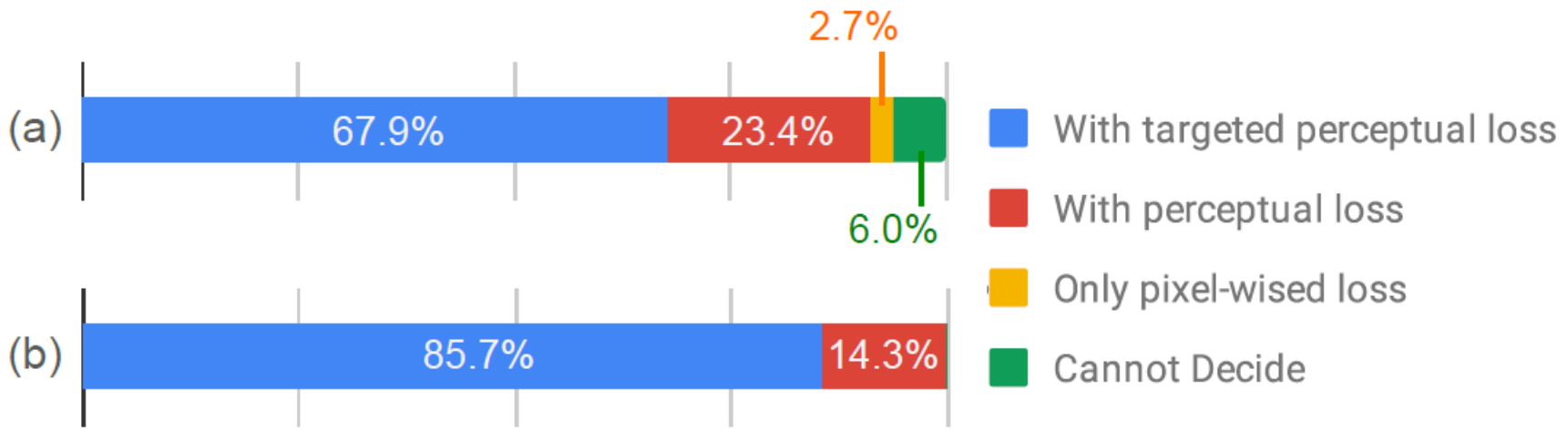
These metrics would not reflect superior reconstruction quality, and user study will be a better choice.

User Study



Results of five methods together with the high resolution references. Users were requested to vote for more appealing images with respect to the ground-truth image.
35 images from COCO-Stuff, outdoor scenes. 46 persons participated in the survey.

Ablation study



Percentage of votes

Percentage of wining images by majority of votes

35 images from COCO-Stuff, outdoor scenes. 51 persons participated in the survey.

Take-homes

- Incorporating the semantic information into the perceptual loss
  - Using low-level features to limit the boundary mismatch (sharpen edges)
  - Using mid-level features to raise the background texture fidelity

to generate overall more photo-realistic results.

Thanks for your attention!