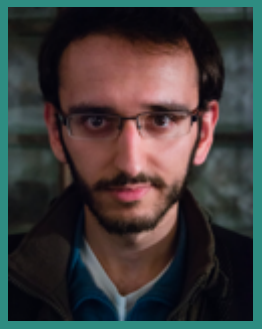
EnhanceNet

Single Image Super-Resolution Through Automated Texture Synthesis



Mehdi S. M. Sajjadi

Bernhard Schölkopf

Michael Hirsch

Max Planck Institute for Intelligent Systems

ICCV17

Single Image Super-Resolution

?

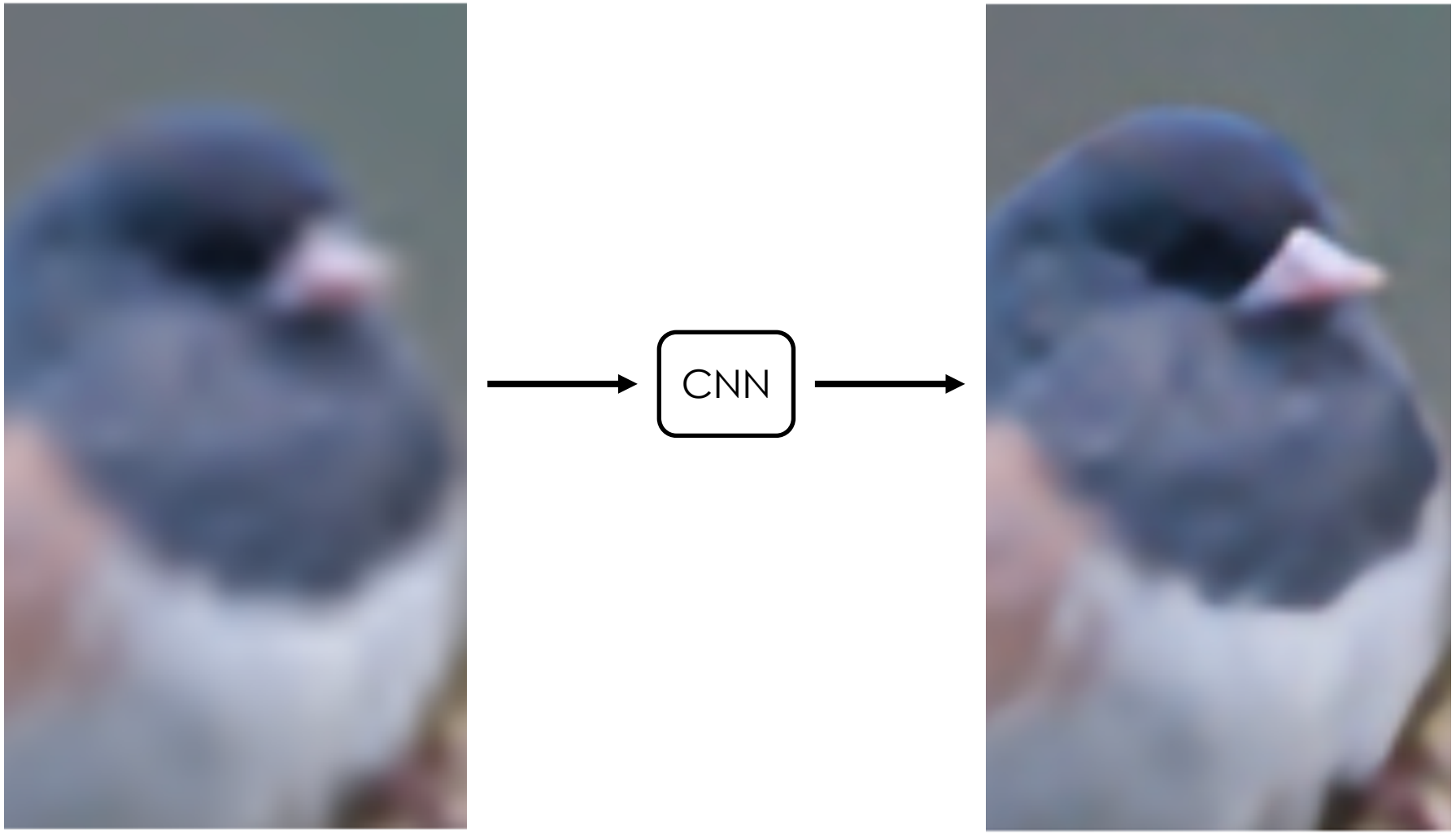
Low-resolution input

Original high-res image

Method: Classical approach

- Minimize mean-squared error
  - $\text{MSE} = \frac{1}{\text{nm}} \sum_i^n \sum_j^m \left(\tilde{I}_{ij} - I_{ij}\right)^2$
- Evaluation by Peak signal-to-noise ratio (PSNR)
  - $\text{PSNR} = -10 \log_{10}(MSE)$
- The field is dominated by convolutional neural nets
  - SRCNN (Dong et al., ECCV 2014)
  - DRCN, VDSR (Kim et al., CVPR 2016)
  - NTIRE challenge on image super-resolution (CVPR 2017)

State of the art by PSNR



Low-resolution input

High-resolution output

Is PSNR the right metric?

Low-resolution input

Generated image

Original image

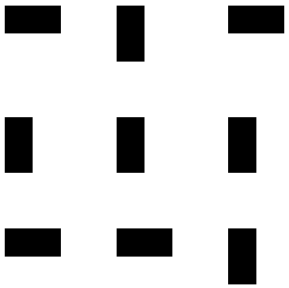
EnhanceNet's output

Low-resolution input

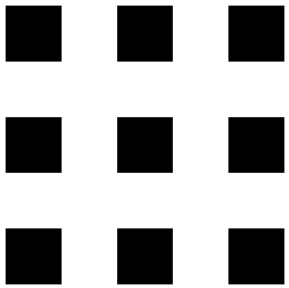
Our result

Original image

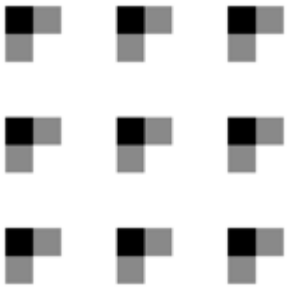
PSNR vs. visual similarity



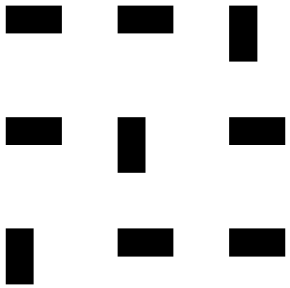
High-resolution image



Low-resolution image



Optimal PSNR



Realistic image
Low PSNR

Method: residual CNN with loss

- **E**uclidean distance / mean squared error (MSE)
  - $\left|\left|\tilde{I} - I\right|\right|_2^2 = \frac{1}{NMC}\sum_{ij}\left(\tilde{I}_{ij} - \tilde{I}_{ij}\right)^2$
- **P**erceptual loss (Dosovitskiy and Brox 2016, Johnson *et al*. 2016)
  - $\left|\left|\phi(\tilde{I}) - \phi(I)\right|\right|_2^2$ MSE in VGG feature space
- **T**exture loss / style transfer (Gatys et al. 2015)
  - $\left|\left|G\left(\phi(\tilde{I})\right) - G\left(\phi(I)\right)\right|\right|_2^2$ MSE of correlation in VGG feature space
- **A**dversarial loss / GAN (Goodfellow et al. 2014)
  - $D(\tilde{I}),\ D(I) \in [0,1]$ discriminator rates realism of image patches

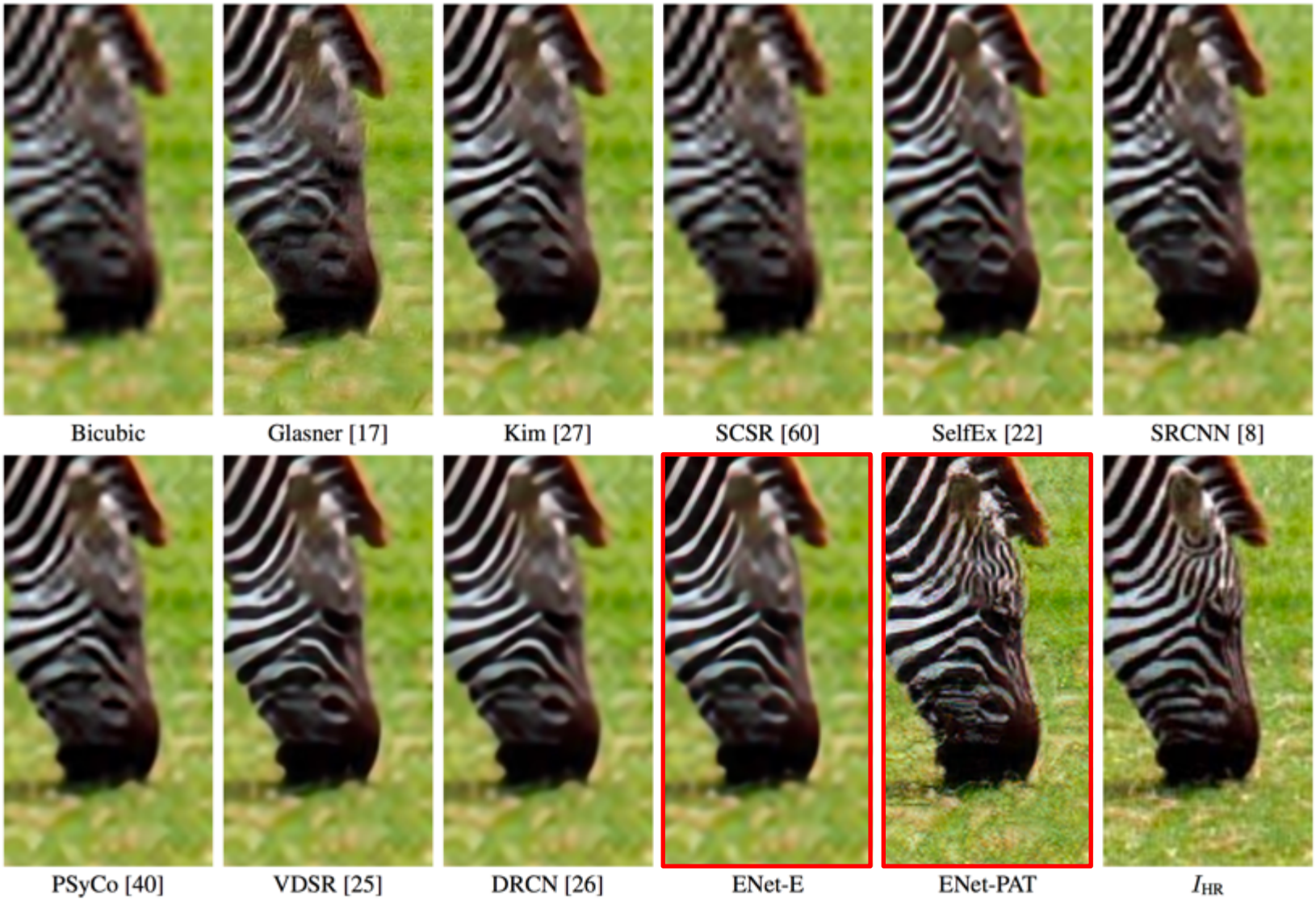
Best result by PSNR vs. our best

Evaluation

- PSNR/SSIM/IFC: **ENet-E SOTA**; ENet-PAT low scores
- Survey: **ENet-PAT** is preferred over ENet-E in **91.0%** images
- Object recognition image quality benchmark
  - Feed ImageNet through super-resolution models
  - Run pre-trained object recognition network on results
  - **ENet-PAT leads to lowest error**

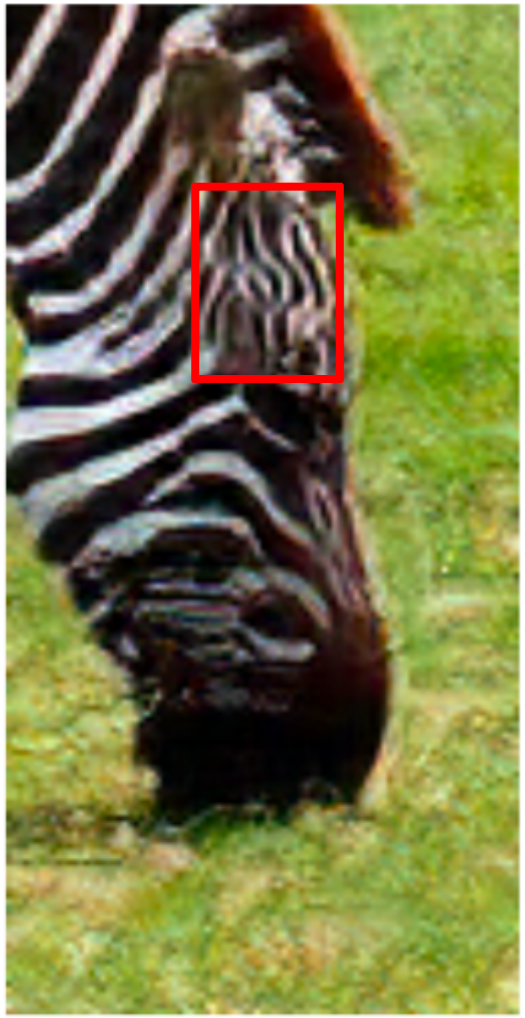
Evaluation	Bicubic	DRCN [26]	PSyCo [40]	ENet-E	ENet-EA	ENet-PA	ENet-PAT	Baseline
Top-1 error	0.506	0.477	0.454	0.449	0.407	0.429	**0.399**	0.260
Top-5 error	0.266	0.242	0.224	0.214	0.185	0.199	**0.171**	0.072
Confidence	0.754	0.727	0.728	0.754	0.760	0.783	**0.797**	0.882

Comparison w/ other methods



A closer look at ENet-PAT

Bicubic



ENet-PAT

I_{HR}

Conclusion

- Introduce a novel combination of loss functions for single image super-resolution
- State of the art in quantitative + qualitative benchmarks
- Propose object detection image quality benchmark
  - Let's see if it works in other domains as well
- Outlook: lots of room for improvements
  - Perceptual evaluation still an unsolved problem