Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network

Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi {first name initial}{surname}@twitter.com

Twitter

#KeyIdeas

SRGAN: Employ a generative adversarial network (GAN) [1] for image super-resolution (SR)

Perceptual Loss: Optimize a perceptual loss function based on a content loss calculated in VGG feature space [2,4] and a adversarial loss [1,3] that pushes the solutions to the natural image manifold.

MOS-Testing: Perform an extensive mean-opinion-score (MOS) test to confirm hugely significant gains in perceptual quality and limitations of mean-squared-error (MSE) based quality measures.

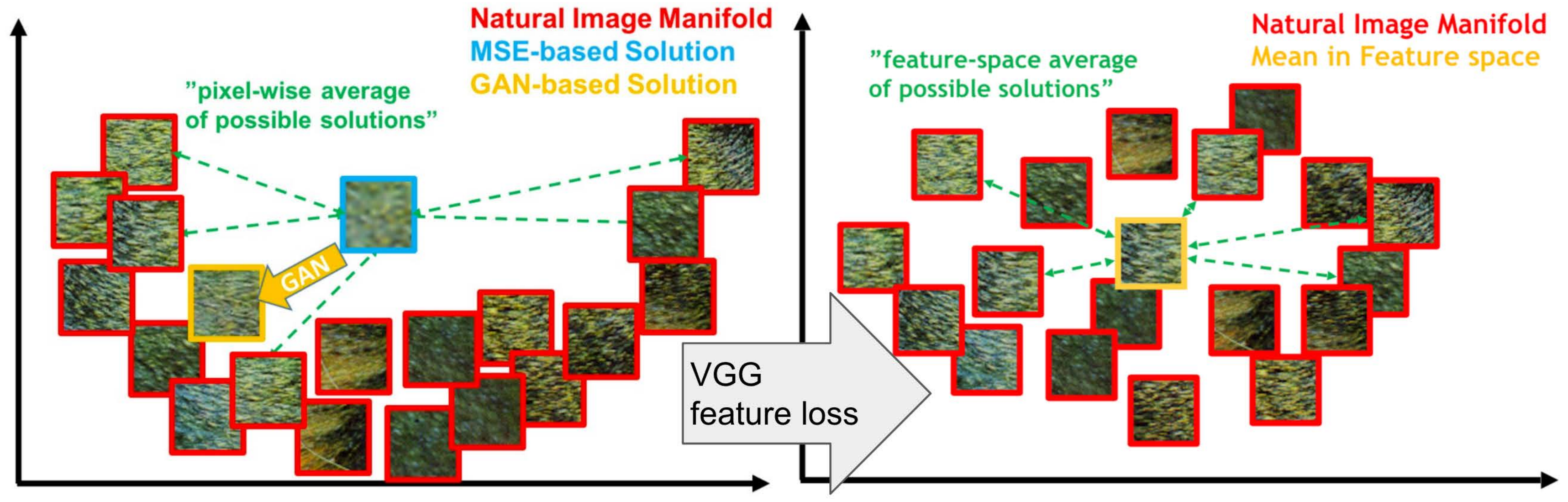
#TheProblem

Super-resolve a low-resolution input image $\mathbf{I}^{LR}$ that was obtained by 4x downscaling a high-resolution image $\mathbf{I}^{HR}$. We train a convolutional neural network (CNN) with optimal parameters such that:

$$\hat{\theta}_G = \arg\min_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} l^{SR}(G_{\theta_G}(I_n^{LR}), I_n^{HR})$$

#Motivation

The Limitation of MSE based optimization is that it encourages average-like solutions that are overly smooth and generally not reside on the manifold of natural images.



#PerceptualLossFunction

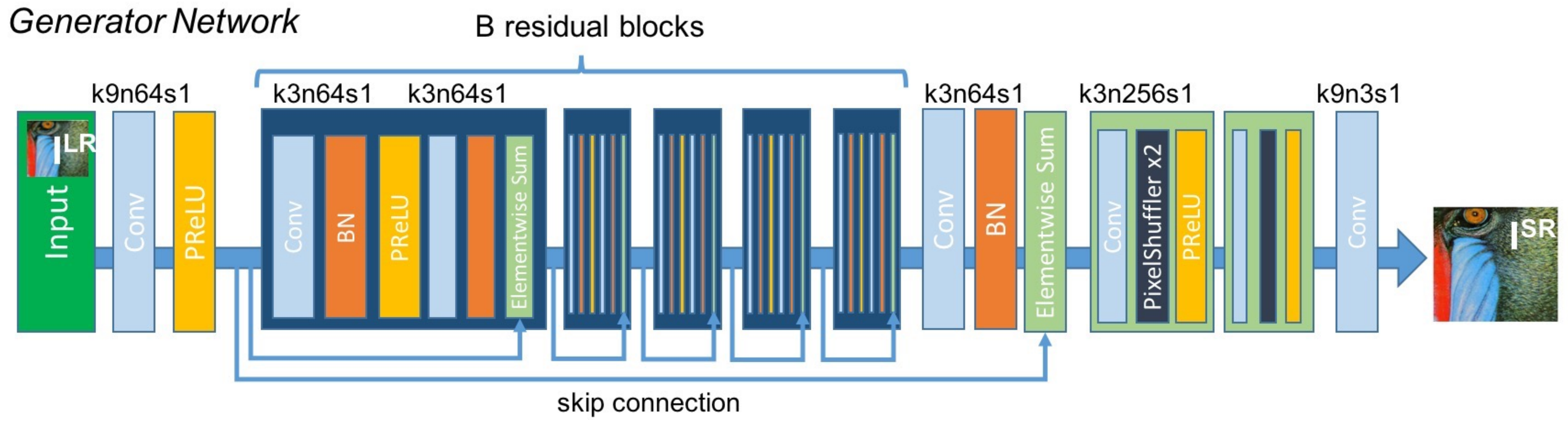
Content loss that ensures high level content is preserved without penalizing low level details [2,3,4,6].

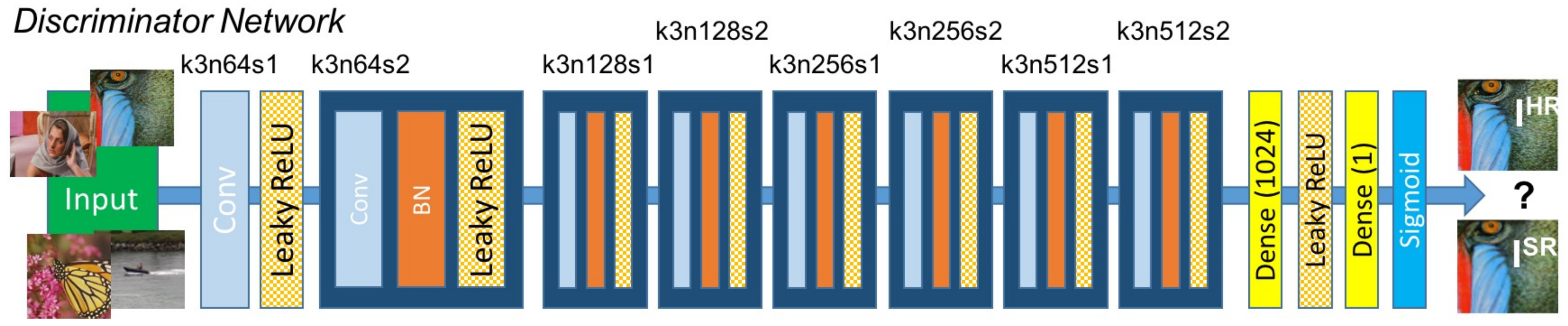
$$l^{SR} = \underbrace{\underbrace{\overset{\text{MSE or VGG}}{l_{X}^{SR}}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}}}_{\text{perceptual loss (for VGG based content losses)}}$$

Adversarial loss that encourages solutions from the natural image manifold [1,3].

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{\text{train}}(I^{HR})}[\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))]$$

#Architectures





- Discriminator architecture based on VGG [2], DCGAN [5].

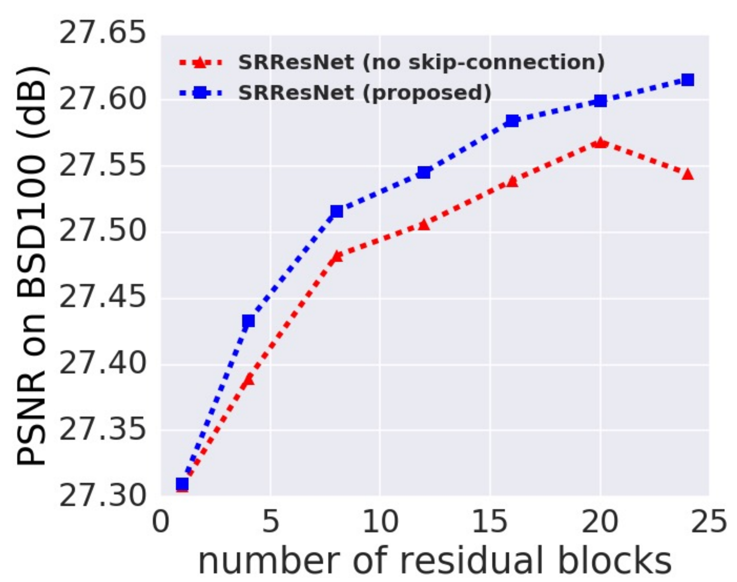
#ExperimentsAndResults

Evaluation Measures

MOS averaged over scores from 26 human raters. Scores range from 1 (worst, nearest neighbor) to 5 (best, original HR image)

Influence of Network Depth

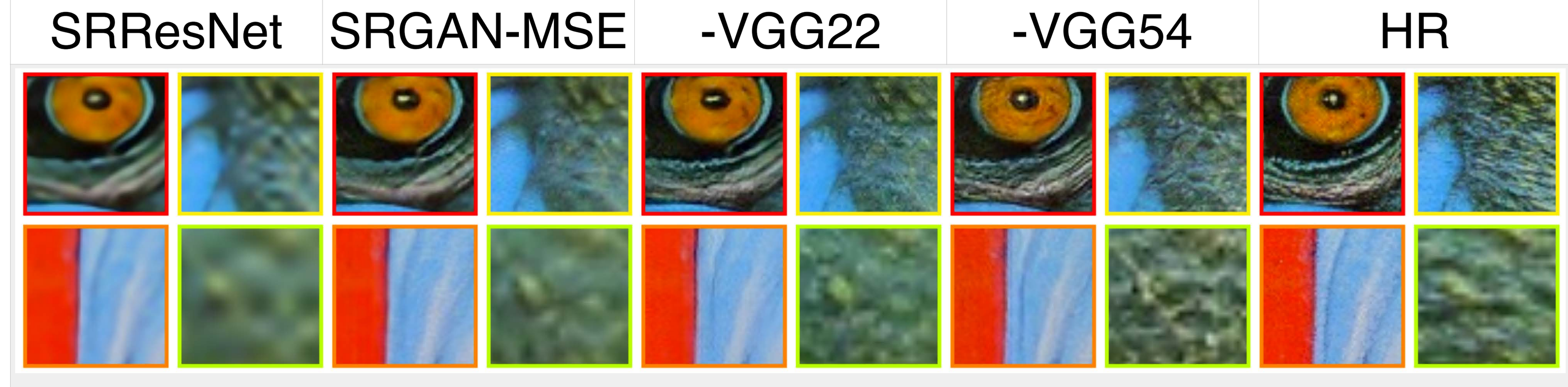
- Higher performance using skip-connections
- Depth is beneficial for PSNR
- SRGAN gets more difficult to train for deeper networks



Investigation of content loss

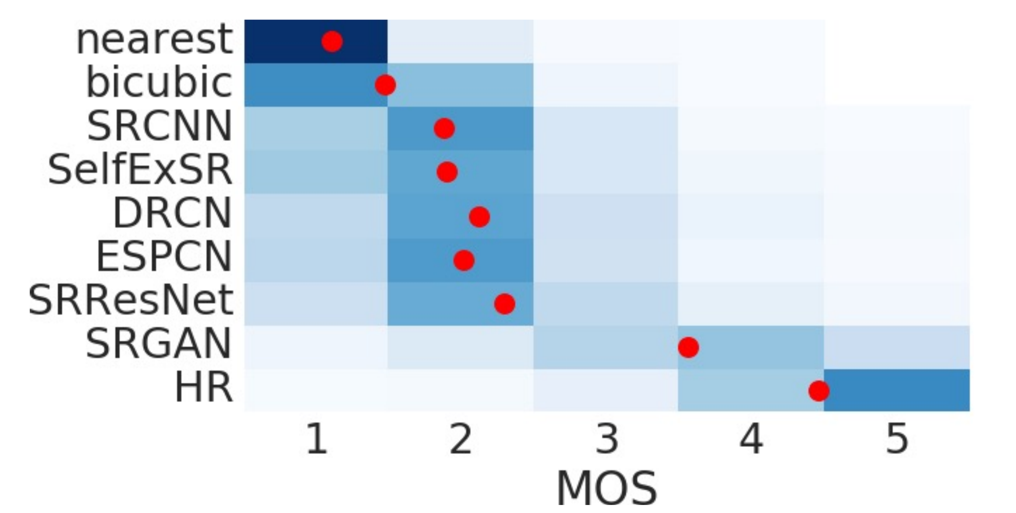
- Loss on higher level VGG features yields better texture detail

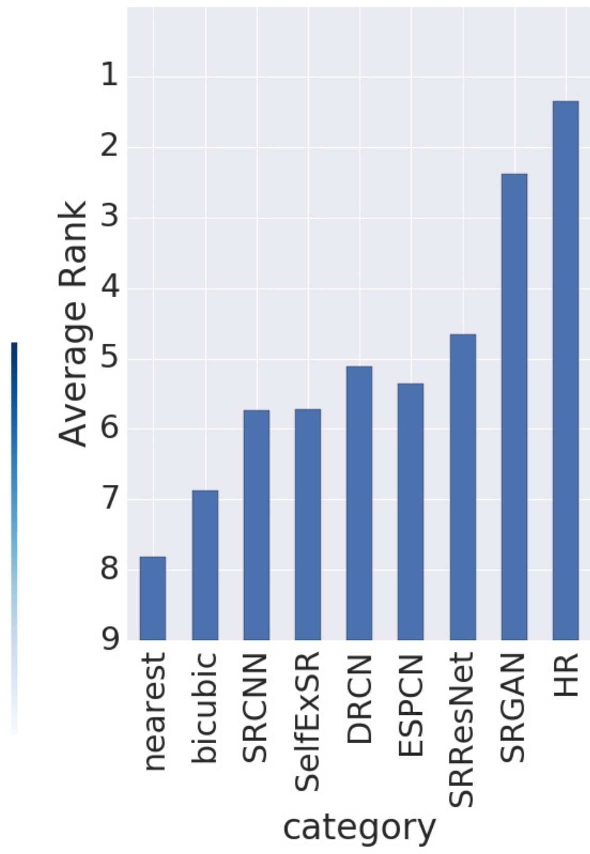
Set14	SRResNet-		SRGAN-		
	MSE	VGG22	MSE	VGG22	VGG54
PSNR	28.49	27.19	26.92	26.44	26.02
MOS	2.98	3.15*	3.43	3.57	3.72*



Comparison to state of the art (BSD100)

- SRGAN closes the MOS gap between the state of the art and original high resolution images by more than 50%





BSD100	NN	SRCNN	SelfEx	DRCN	ESPCN	SRResNet	SRGAN	HR
PSNR	25.02	26.68	26.83	27.21	27.02	**27.58**	25.16	∞
MOS	1.11	1.87	1.89	2.12	2.01	2.29	**3.56**	4.46

#Visuals

#References

[1] I. Goodfellow, et al., "Generative adversarial nets", NIPS, 2014.
[2] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition", ICLR, 2015.
[3] A. Dosovitskiy and T. Brox. "Generating images with perceptual similarity metrics based on deep networks", NIPS, 2016.
[4] J. Bruna, et al., "Super-resolution with deep convolutional sufficient statistics", ICLR, 2016
[5] A. Radford, et al., "Unsupervised representation learning with deep convolutional generative adversarial networks", ICLR, 2016
[6] J. Johnson, et al., "Perceptual losses for real-time style transfer and super-resolution", ECCV, 2016