

Super-Resolution with GANs

Nathanael Bosch
nathanael.bosch@tum.de

Thomas Grassinger
thomas.grassinger@tum.de

Jonas Kipfstuhl
jonas.kipfstuhl@tum.de

Pierre Springer
pierre.springer@tum.de

1. Introduction

We call super-resolution (SR) the task of estimating a high-resolution (HR) image from its low-resolution (LR) counterpart. Recent work with optimization-based methods largely focuses on minimizing the mean squared reconstruction error, which results in high peak signal-to-noise ratios (PSNR) but very smooth pictures.

The authors of the paper[7] propose using a generative adversarial network (GAN) with an improved loss function. We reimplemented the paper and questioned the purpose of the discriminator and the suggested loss function.

2. GANs

GANs consist of two different networks, a Generator Network and a Discriminator Network. The concept behind this is that the generative network estimates a super-resolved image from its LR version with the goal to become highly similar to real images that the discriminator network fails to distinguish.

Therefore we optimize the discriminator network D_{Θ_D} in an alternating manner along with the generative network G_{Θ_G} to solve the adversarial min-max problem:

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

The perceptual loss l^{SR} we defined as weighted sum of a content loss and an discriminative loss component:

$$l^{SR} = \alpha l_{MSE}^{SR} + \beta l_{VGG16_9/i,j}^{SR} + \gamma l_D^{SR}$$

More precisely, the content loss components are defined

as follows:

$$l_{MSE}^{SR} = \frac{1}{r^2 W H} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2 \\ l_{VGG16_9/i,j}^{SR} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\Phi_{i,j}(I^{HR})_{x,y} - \Phi_{i,j}(G_{\Theta_G}(I^{LR}))_{x,y})^2$$

with downsampling factor r and W, H defining the tensor size. Finally the discriminative loss is defined as follows

$$l_D^{SR} = \sum_{n=1}^N -\log D_{\Theta_D}(G_{\Theta_G}(I^{LR}))$$

3. Setup

For training we used the PASCAL VOC Dataset[3] with more than 10,000 images as well as the NITRE Dataset[1] with 800 images.

We used pretrained VGG networks in different configurations. The best results were obtained when we used a VGG1619 network. We also considered a VGG16 network. Although faster at training it did not yield results of equal quality.

4. Results

We observed that the widely used metrics, SSIM and PSNE, i. e., structural similarity and peak signal to noise ratio respectively, are not sufficient for our purposes. This observation was also made by other authors dealing with image super resolution. Therefore, in addition to the numerical values of the metrics, we also had to trust our perception of the images. This is similar to the mean opinion score (MOS) used in [7].

For comparison of the generated images with several loss configurations, i.e. setting the coefficients to zero or using the values of the paper, see figure ??.

The curves of the PSNR and SSIM values during training may be seen in figure 2.



Figure 1: Comparison of several loss configurations

5. Conclusion

One main point is the training. It was very hard to keep the discriminator and the generator in line. Often we killed our discriminator such that it could not contribute anymore.

We also observed that super resolution works even without any discriminator activated, as was suggested by the paper we started with.

Another key learning point is the quality of the produced images. The metrics don't change much from some point on but the visual impression of the generated images still improves.

References

- [1] E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, July 2017. 1
- [2] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath. Generative adversarial networks: An overview. October 2017.
- [3] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>. 1
- [4] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. June 2014.
- [5] R. Jia Deng, R. Wei Dong, R. Socher, R. Li-Jia Li, R. Kai Li, and R. Li Fei-Fei. Imagenet: A large-scale hierarchical image database. pages 248–255. IEEE, June 2009.
- [6] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim. Learning to discover cross-domain relations with generative adversarial networks. March 2017.
- [7] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image

super-resolution using a generative adversarial network. September 2016. [1](#)

- [8] Y. Li, L. Song, X. Wu, R. He, and T. Tan. Anti-makeup: Learning a bi-level adversarial network for makeup-invariant face verification. September 2017.
- [9] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. September 2014.
- [10] J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros. Generative visual manipulation on the natural image manifold. September 2016.

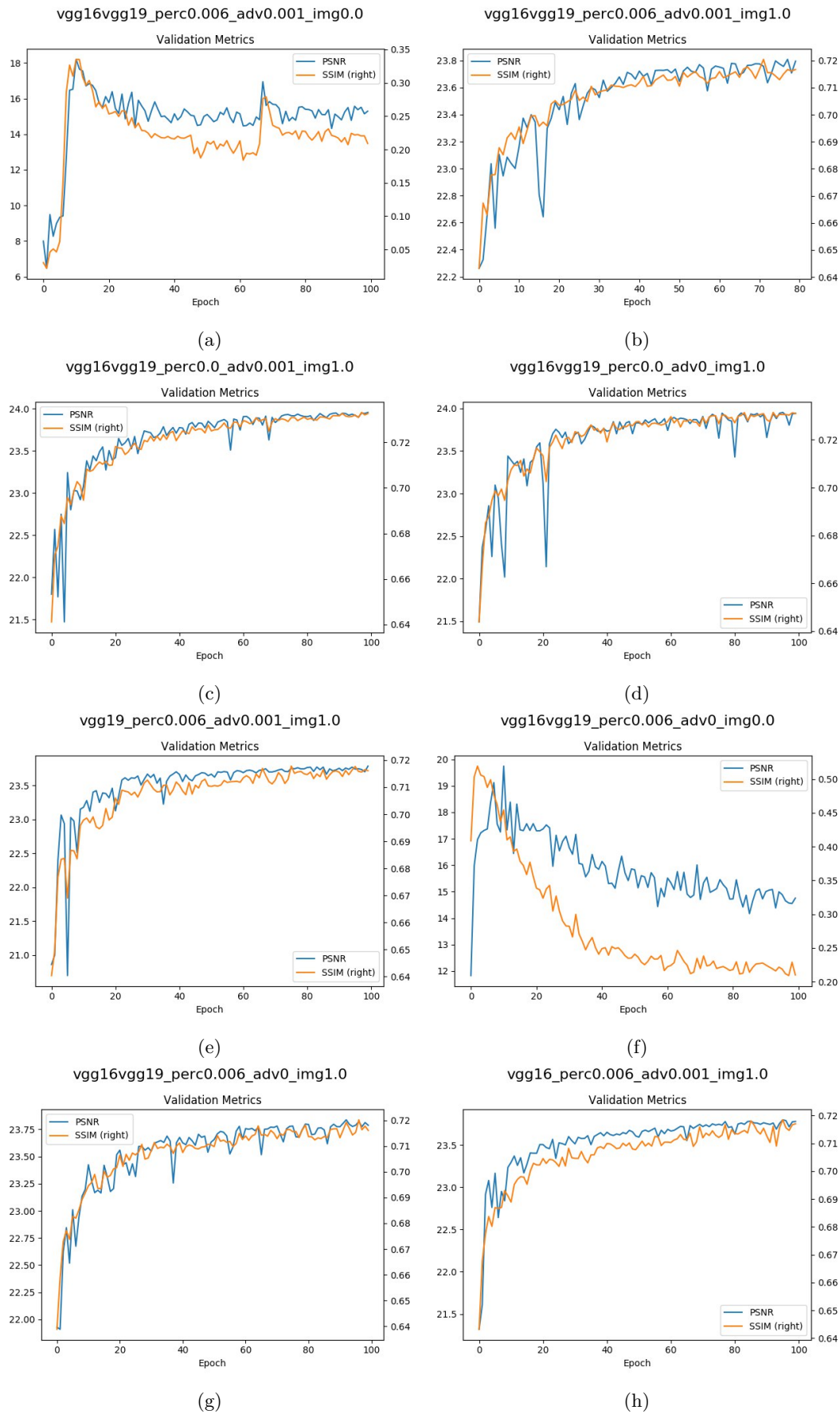


Figure 2: Curves of metrics during training