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:

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

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

$$l^{SR} = \alpha l^{SR}_{MSE} + \beta l^{SR}_{VGG16_19/i.j} + \gamma l^{SR}_D$$

More precisely, the content loss components are defined as follows:

$$l^{SR}_{MSE} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I^{HR}_{x,y} - G_{\theta_G}(I^{LR})_{x,y})^2$$

$$l^{SR}_{VGG1619/i,j} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\Phi_{i,j}(I^{HR})_{x,y} - \Phi_{i,j}(G_{\Theta_G}(I^{LR}))^2_{x,y}$$

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

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

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 investigated on the proposed loss by the authors of [7], by training multiple networks with only parts of the loss. The reults can be seen in figure 1.

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

We observed that those widely used metrics are not sufficient for our purposes. While they did not change anymore during training, the visual appearence still improved. This observation was also made the authors of [7].

(a) VGG1619 perceptual, adversarial, image loss

(b) VGG1619 perceptual, adversarial loss

(c) VGG1619 perceptual, image loss

(d) VGG1619 perceptual loss

(e) VGG1619 adversarial, image loss

(f) VGG1619 image loss

(g) VGG19 perception, adversarial, image loss

(h) VGG16 perception, adversarial, image loss

Figure 1: Comparison of several loss configurations

5. Conclusion

In our work we achieved very good results using the proposed loss and the residual network. During our analysis we found the perceptual loss to be crucial for achieving high performance, but when used on its own it led to weird artifacts in the output images.

The removal of the discriminator had no impact on the resulting image quality in our work, as the generator performs very well with the other parts of the loss. This leads to huge difficultuies in training the GAN.

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.

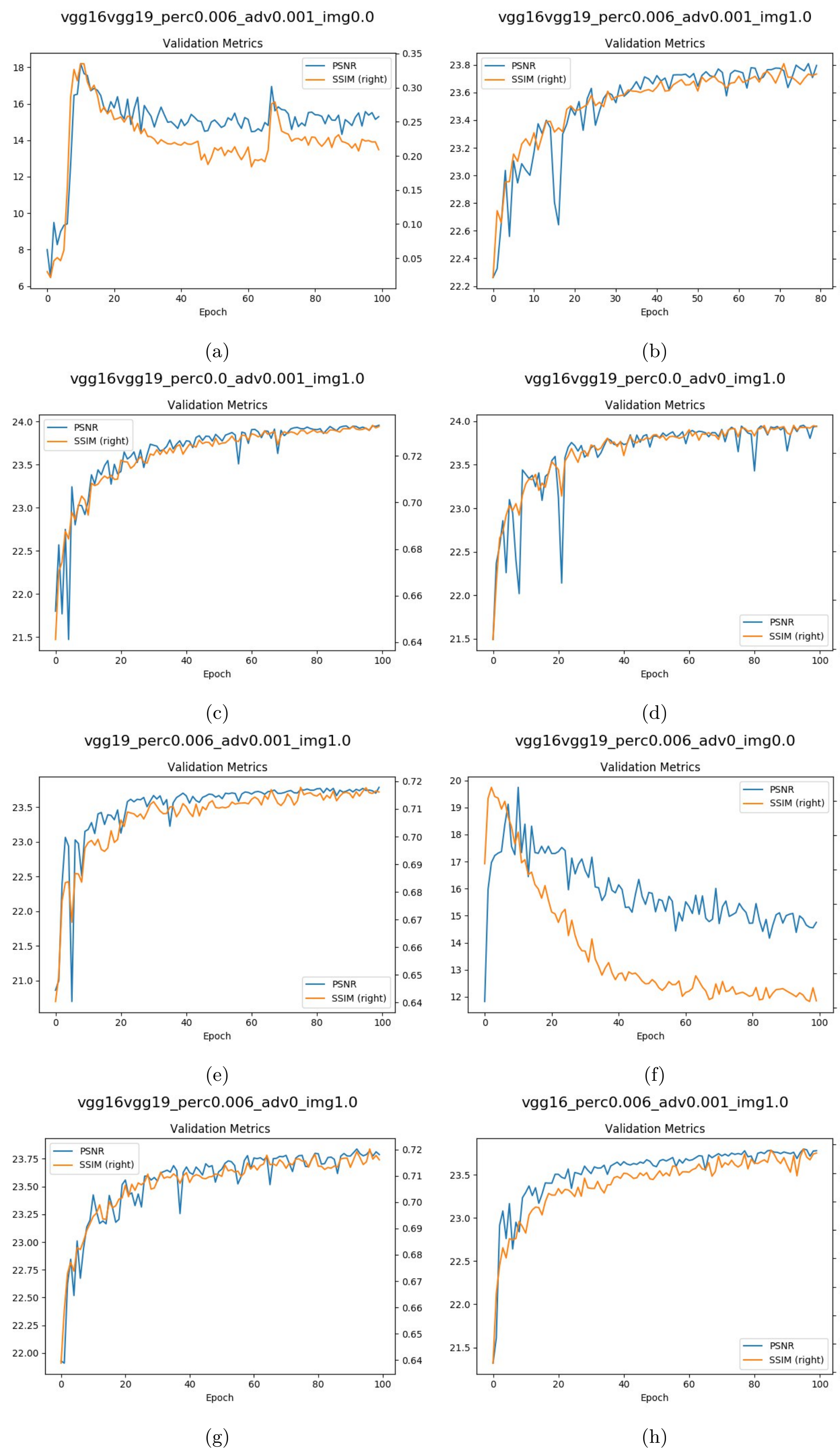

Figure 2: Curves of metrics during training