Super Resolution Using Cycle-Consistent Generative Adversarial Networks

Jonathan Griffin
Stanford University
jgriffi2@stanford.edu

1. Motivation

Generative Adversarial Networks (GANs) have only recently been used for the super resolution task [6]. This method lead to perceptually pleasing photos with higher levels of details not seen in photos generated using other methods [4]. Even then, Ledig et al. used a generic implementation of the GAN architecture, called SRGAN. They only changed the cost function to optimize the perceptual-loss as opposed to the mean-square-error.

In the same year, [10] proposed CycleGAN, a GAN network with two generators and two discriminators where one generator-discriminator pair tries to outperform the other. This method surpassed the baseline results in many areas, such as photo enhancement.

I think that combining the two, Ledig et al.'s metric of evaluation and Zhu et al.'s CycleGAN architecture, will lead to an overall better performance on the super resolution task.

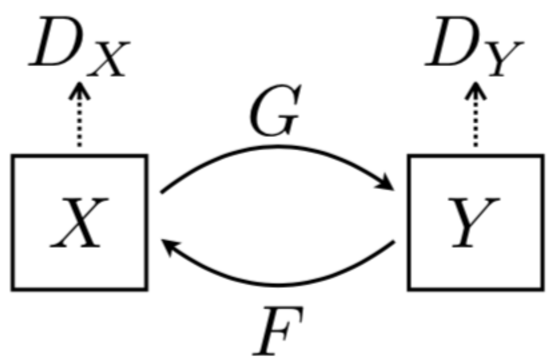

Figure 1. CycleGAN architecture, taken from [10]. D_X and D_Y are the two discriminators. G and F are the two generators. X and Y are images generated by G and F.

2. Related Work

Recent advancements related to the super resolution problem have come from the use of various deep learning architectures. Some apply conventional CNN frameworks with unconventional non-linear mapping strategies, such as sparse-coding-based SR methods [3]. This method was succeeded by [5] with the use of a recursive network that combines skip-connections and recursive supervision. [4] also outperforms Dong et al.'s algorithm by using a much larger network (up to 20 weighted layers) that only learns residuals and has high learning rates in order to speed up convergence.

Recently, [6] has outperformed all of the aforementioned methods by using a generic GAN architecture and a different evaluation metric. Ledig et al. optimizies the perceptual loss as opposed to the MSE metric. This leads to prioritizing perceptually pleasing, high detailed photos over similar PSNR values.

Dahl et al. focuses on a slightly different problem than the previous methods. Instead of solving how to increase the resolution of already high resolution images, they enhance small 8x8 resolution images to 32x32 resolution images. They tackle this problem with a recursive network simliar to PixelCNN [2].

3. Project Overview

The concept of this project is to replicate the CycleGAN architecture with a variant of the perceptual loss optimization function used in SRGAN. Optimizing hte perceptual loss will result in the high detail images that Ledig et al. obtained using SRGAN and using the CycleGAN architecture will boost the performance of the algorithm like it did with many other GAN-related tasks outlined in [10].

For reference, the CycleGAN architecture is shown in Figure 1.

The perceptual loss function I will be optimizing is shown in Equation 1. In Equation 1, l_X^{SR} measures the content loss and l_{Gen}^{SR} measures the adversarial loss. The content loss and the adversarial loss are shown in Equations 2 and 3, respectively. In these equations, $G_{\theta_G}(I^{LR})$ represents the reconstructed image, I^{HR} is the reference image, and $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ is the probability that $G_{\theta_G}(I^{LR})$ is a natural image.

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR} \tag{1}$$

$$l_{X/i,j}^{SR} = \frac{1}{H_{i,j}W_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2 \tag{2}$$

$$l_{Gen}^{SR} = \sum_{n=1}^{N} -log(D_{\theta_D}(G_{\theta_G}(I^{LR}))) \tag{3}$$

For evaluation, I will be using **SET5** [1], **SET14** [9], and **BSD100** [7] datasets, the same datasets that Ledig et al. used. For training, I will be using images from **ImageNet** [8], similar to Ledig et al.

Unfortunately, I do not think I will have time to perform the same evaluation as Ledig et al., since they used the mean-opinion-score (MOS) metric which involves having user-feedback. Therefore, I will base my evaluation on PSNR values unless I have adequate time to gather enough participants to perform the MOS evaluation.

4. Milestones, Timeline, and Goals

The timeline I would like to keep is outlined below:

15 February: Gather necessary data

22 February: Have a basic implementation of CycleGAN

1 March: Adjust cost functions to prioritize perceptual loss

8 March: Have testing and evaluation completed

13 March: Paper submitted and poster ready to present

In terms of my goals for the project, I would like to be able to at least replicate the results of the Ledig et al. paper. However, I think I will be able to improve upon it since CycleGAN has been shown to improve image enhancement [10].

References

[1] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. *BMVC*, 2012.

[2] R. Dahl, M. Norouzi, and J. Shlens. Pixel recursive super resolution. *ICCV*, 2017.

[3] C. Dong, C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE*, 2016.

[4] J. Kim, J. Lee, and K. Lee. Accurate image super-resolution using very deep convolutional networks. *CVPR*, 2016.

[5] J. Kim, J. Lee, and K. Lee. Deeply-recursive convolutional network for image super-resolution. *CVPR*, 2016.

[6] 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. *CVPR*, 2017.

[7] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating seg- mentation algorithms and measuring ecological statistics. *IEEE*, 2001.

[8] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. B. et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 2014.

[9] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. *Curves and Surfaces*, 2012.

[10] J. Zhu, T. Park, P. Isola, and A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *CVPR*, 2017.