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

The goal of this project is to gain a deeper understanding of Generative Adversarial Networks. We will then apply them to the image super resolution problem.

1.1. Related Works

- Generative Adversarial Networks [4]
- Generative Adversarial Networks: An Overview [2]
- Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network [5]

As we did not get responses for all datasets, for now, we decided to work on the Photo-Realistic Single Image Super-Resolution using a Generative Adversarial Network [5] paper. The main goal is to learn how to increase the resolution of images up to 4 times, while still creating *good* images that look realistic. This is achieved by training Generative Adversarial Networks on downscaled versions of high resolution images. The *generator* part of the GAN will do the actual task of super-resolution, while the *discriminator* is trained to distinguish between real images and super-resolved images. We will compare the results of this approach to regular methods, such as bicubic extrapolation. The work will be mostly based on [5].

2. Dataset

Our initial work towards *understanding GANs* will use the MNIST and CIFAR-10 datasets, which we already worked with in the lecture and exercises. We will not give detailed explanation to these datasets.

For the Image Super-Resolution [5] we will work with the DIV2k dataset [1]. The dataset consists of 1000 2K resolution images divided into: 800 images for training, 100 images for validation, 100 images for testing. As we do not have access to the 100 high resolution test images we will only use the 800+100 images provided, and split into train-validation-test accordingly. The images are provided in both high and low resolution images for 2, 3, and 4 downscaling factors. The structure of the given images is [1]:

> For the high resolution images: 0001.png, 0002.png, ..., 1000.png
>
> For the downscaled images:
> YYYYx2.png for downscaling factor x2; where YYYY is the image ID
> YYYYx3.png for downscaling factor x3; where YYYY is the image ID
> YYYYx4.png for downscaling factor x4; where YYYY is the image ID

We also have access to the ImageNet data but will most likely use a pretrained network to make use of its features such as the VGG network [6], which is already available within torchvision. Detailed information about ImageNet can be found in [3].

3. Methodology

Our initial work will setup on the cs241 course, which treated the basics of GANs and followed the work of [4]. In this stage, we want to understand what GANs are and how they work, as well as why people use them and which problems they were meant to solve. The cs241 course provides practical guidelines for initial implementations, but also allows us to modify the results and explore the many possibilities. We will start working with the MNIST and CIFAR-10 dataset by applying and comparing different GAN models such as fully connected GAN networks, convolution GAN networks, conditional CNN, and Adversarial Autoencoders to get a understanding of Generative Adversarial Networks.

In a second step, after having understood the principles of GANs, we will reproduce the results of the super-resolution paper [5] and try different modifications. These modifications include for example, changing the input features or choosing other pretrained models, such as different VGG layer models or even an entirely different model e.g. , AlexNet. Using our knowledge from the initial work on the

MNIST and CIFAR data, we might also find other ways to improve our models performance compared to that of the paper. These tasks (working on layers, data engineering, work on pretrained models, further research) can be easily split up, which provides an efficient workflow.

We would like to use the Google cloud service as presented in the lecture, but the local possibilities of the computer vision group might also be sufficient.

4. Outcome

Our target still remains: We want to *understand* GANs! Additionally we want to create a GAN which is able to transform low-resolution images into higher resolutions with similar or even better results as [5].

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, 2017. 1

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009. 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, 2014. 1

[5] 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, 2016. 1, 2

[6] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2014. 1