

# FastGAN: Faster and Stabilized GAN

Authors:

**Mauro Ficorella 1941639**

**Martina Turbessi 1944497**

**Valentina Sisti 1952657**



Sapienza  
May 2021

## ABSTRACT

*The main aim of FastGAN is to allow users with limited computing budget and resources to train a GAN. Moreover it eliminates the requirement of a big dataset for training. These are big advantages since traditional GANs required a lot of GPU computational power (i.e. one or more server-level GPUs with at least 16 GB of vRAM in StyleGAN2) and a large number of images for training. This implementation allowed to train from scratch on a NVIDIA GeForce RTX 2070 SUPER and a NVIDIA GeForce GTX 1050-Ti, obtaining good results also on a small dataset.*

## 1 Introduction

In this project our objective is to transform blurry, LR images in order to obtain sharp, realistic and high-resolution images. This is a very important task in a lot of disciplines (such as medicine, astronomy, satellite imagery, and so on), since the cost and hardware restrictions for obtaining high-resolution images are very high, and so it often occurs that low-resolution, blurry images are captured. In other situations there could be used old, blurry and out of focus images, so our algorithm is a good choice also in these cases. Using those kind of images, in addition to be visually unattractive, worsen the use of the "downstream analysis methods", such as disease diagnosis, since they depend on having high-resolution images. Another field of interest for our project is the recently increased popular demand for sharp images related to the grown of screen resolution, in particular regarding laptops, televisions and mobile phones. All these applications motivated the recent interest in the task of *image super-resolution*, in other words the creation of realistic high-resolution images to which a given low-resolution input image could correspond. We have just seen all the benefits of methods for image super-resolution. Conversely the difference in information content between HR and LR images (especially at high scale factors) obstruct efforts to develop such techniques. In particular, since LR images have less high-variance informations, the details can be blurred and so being visually indistinct. Is important to point out that the problem of recovering the true HR image from an LR input is inherently ill-posed, since many high-resolution images can correspond to the same low-resolution image, so we can only generate a set of potential such HR images.

Training a model (like a convolutional neural network) in order to minimize the mean-squared error (pixel-wise) between the obtained super-resolved images and the corresponding real HR images, as done by traditional supervised super-resolution algorithms, led to ignore perceptually important details related to the photorealism of HR images. But regarding mean squared error (MSE) the ideal approach is to use the weighted average (pixel-wise) of the set of the realistic images that downscale correctly to the LR images taken in input. However the result of this solution is smoothing in areas of high variance (for example parts of the image with complex patterns or textures), and so we can conclude that MSE should not be used alone as a metric of the image quality for super-resolution.

An attempt to solve this issue was proposed by some researchers, that tried to mitigate the smoothing effect of the MSE optimizing metrics related to realism; this brought to a better perceptual image quality, but also produced the unattended result that there were no more guarantees that the generated images were realistic; moreover there were still signs of blurring in high variance parts of the images, as seen before.

So, in order to avoid these problems, the goal should be to obtain realistic images within the set of feasible solutions, and so, to find points which actually lie on the natural image manifold and also downscale correctly.

The method that we used outputs images through a pretrained generative model that approximates the distribution of natural images that we are taking into account. Given an input LR

image, we go through the manifold of natural images, that is parametrized by the latent space of the generative model; we do this in order to find regions that downscale correctly. Since this approach is entirely self-supervised, with no training needed, it avoids the need for supervised training; this is not true regarding the unsupervised generative model.

This algorithm has the following benefits: first, it allows the same network to be used on images with different degradation operators even if there is not a database of corresponding LR-HR pairs of images; secondly, unlike previous methods, it obtains state-of-the-art level results in terms of generative modeling, without requiring retraining and super-resolution network architectures, which are very difficult to develop, without providing a real insight into the problem. This project works with GANs and any other type of generative models; but, since in the last years GANs have improved on the field of resolution and sharpness, StyleGAN is the chosen one for this work. Moreover, in this work the main focus is on face images, but with a different training, it can be extended also to other contexts.

So, concluding, we can say that this work yields realistic and unique images that downscale correctly to the LR image taken as input, meaning that any of them could be the original image for the LR input.

## 2 Method

We denote the low-resolution input image by  $I_{LR}$ . We aim to learn a conditional generating function  $G$  that brings to a higher-resolution super-resolved image  $I_{SR}$ , when applied to  $I_{LR}$ . Let  $I_{LR} \in \mathbb{R}^{m \times n}$  and denote the function  $SR$  as a map  $\mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{M \times N}$ , where  $M > m, N > n$ . Now we can define the super-resolved image  $I_{SR} \in \mathbb{R}^{M \times N}$  as:

$$I_{SR} := SR(I_{LR}).$$

We use a framework for single image super resolution. Let  $\mathcal{M}$  be the natural image manifold in  $\mathbb{R}^{M \times N}$  and, given  $I \in \mathbb{R}^{M \times N}$  let be  $DS(I) = I_{LR}$ .

NOME

**NOME**

**NOME**

**NOME**

NOME



**NOME**

NOME