Single-Image Facial Super-Resolution Using GANs

Mohammad Firas Sada
Antoine Courbot
Aleksander Popovic

Image Super-resolution

Image super-resolution is the process of restoring and upscaling low-resolution images to a higher resolution.

Face super-resolution (FSR), also known as face hallucination, which is aimed at enhancing the resolution of low-resolution (LR) face images to generate high-resolution face images.

There have been many approaches that utilize GAN priors to accomplish Face super-resolution before.

In our project, we investigate single-image facial super-resolution with no priors.

Our Approach

There are various works of literature investigating generic image super-resolution.

For our project, we have selected a Generative Adversarial Network that accomplished single-image super-resolution.

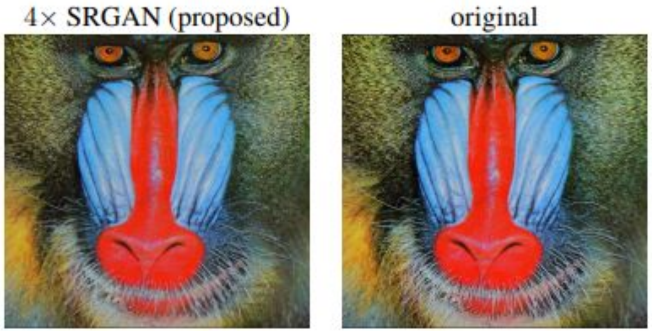


Figure 1: Super-resolved image (left) is almost indistinguishable from original (right). [4× upscaling]

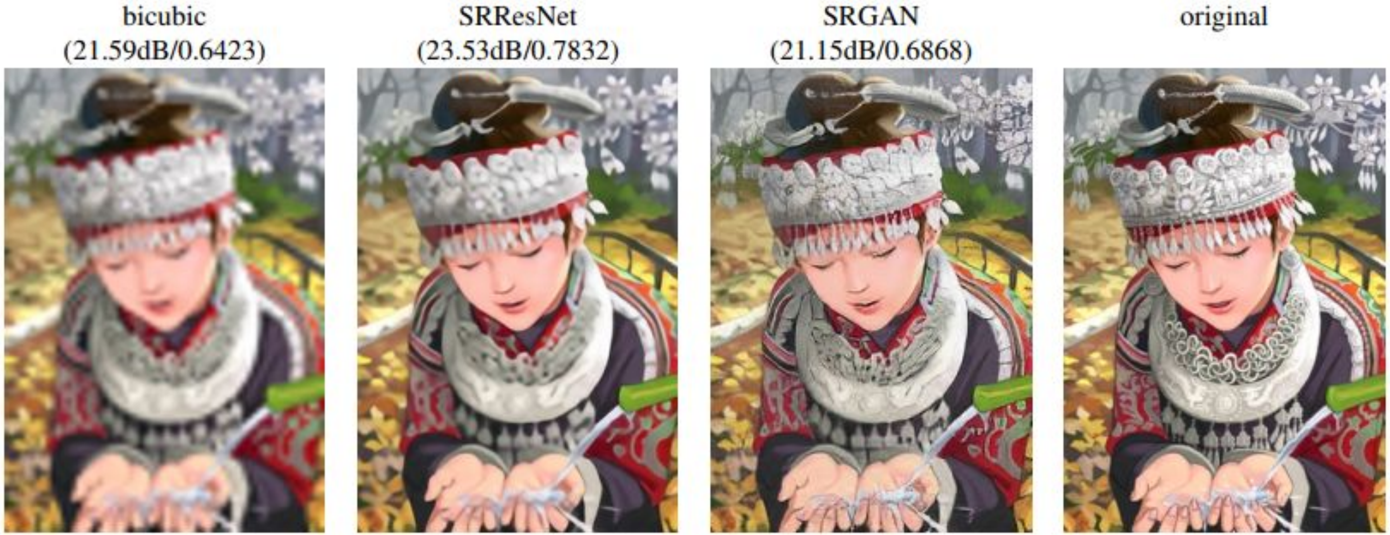
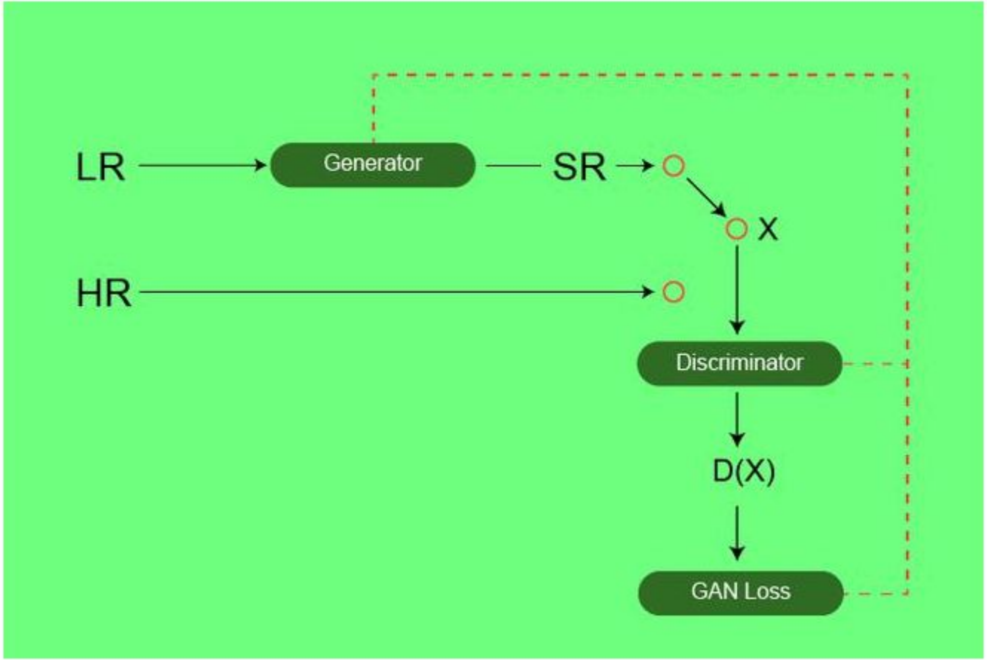


Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

Generative Adversarial Networks

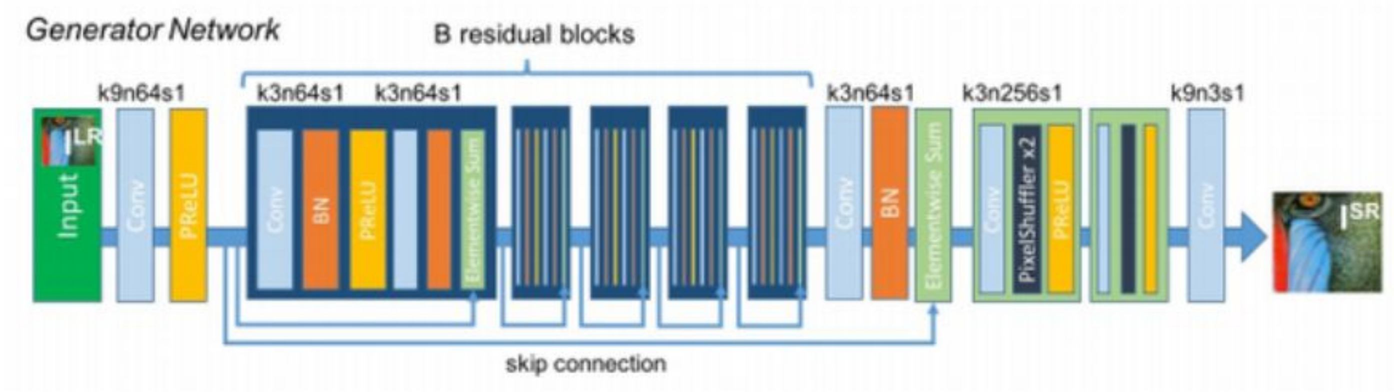
- A generative adversarial network is a class of machine learning frameworks where two neural networks contest with each other in the form of a zero-sum game, where one agent's gain is another agent's loss.
- GANs provide a powerful framework for generating plausible-looking natural images with high perceptual quality.
- The two neural networks for SR are the **generator** and **discriminator**.

How It All Works



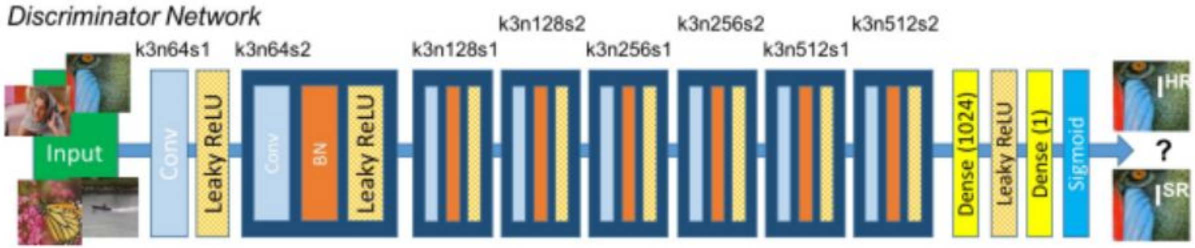
How It All Works: The Generator

- The generator is the neural network we are seeking after, that generates high-resolution images from low-resolution ones.
- The generator uses CNN with residual blocks.
- This generator architecture uses ReLU as an activation function



How It All Works: The Discriminator

- The task of the discriminator is to discriminate between real HR images and generated SR images.
- The discriminator architecture used in the research paper we are investigating is similar to DC- GAN architecture with LeakyReLU as activation.



How It All Works: The Loss

- The SRGAN uses **image content loss** as a loss function for the generator, which is basically the MSE of pixel data.
- It uses **adversarial loss** for the discriminator.
- The model takes the average loss, for the GAN backpropagation.

How It All Works: The Dataset

- The task of image super-resolution can be looked at as a supervised learning for both neural networks.
- Training is done by feeding LR images into the proposed pipeline, which generates HR images from them.
- For the generator, the LR images are the features and the HR images are the labels.
- The discriminator is fed generator SR images, and HR images as labelled training data.

How It All Works: The Dataset

- The model is trained using the COCO dataset (~118k images).

Our Investigation

In our project, we propose:

- Reproducing the original results
- Pivoting the model to generate face hallucinations (face SR images)
- Creating a more robust “face-specific” SRGAN.
- Tuning the hyperparameters and testing different optimizers and loss functions

Reproducing the Results

- We have obtained the source code referenced in the paper.
- We recreated the models using keras and Google Colab.
- We trained the model in the same conditions using the same Coco dataset.

Reproduced Results

50x50 input | SRGAN result | SRGAN result | 150x150 input

Face Super-Resolution Trained Model

- We retrained the model using the FFHQ (~70,000) dataset, which is just face images.
- Our hopes were to get better face super-resolution.
- Model uses Adam as an optimizer, we use AdaBelief which also literature suggests offers better results that Adam or RMSprop. AdaBelief aims to offer optimal convergence time and generalization.

Face Super-Resolution Trained Model

- We retrained the model using the FFHQ (~70,000) dataset, which is just face images.
- Our hopes were to get better face super-resolution.

50x50 input | result | 100x100 input | result | 250x250 input | result

What's next?

- Training with different datasets, and comparing results.
- Tuning hyperparameters, and testing different optimizers and loss functions.
- Model uses MSE, our proposal suggests a combination of MSE and DSSIM, which some works of literature indicate that it offers better performance for facial images.