CS584 Project: Report

Fall 2022 | December 2nd, 2022

Single-Image Facial Super-Resolution Using GANs

https://github.com/groundsada/Single-Image-Facial-Superresolution-GAN

Mohammad Firas Sada

Department of Computer Science

Illinois Institute of Technology

Chicago, Illinois USA

msada@hawk.iit.edu

Antoine Courbot

Department of Computer Science

Illinois Institute of Technology

Chicago, Illinois USA

acourbot@hawk.iit.edu

Aleksander Popovic

Department of Computer Science

Illinois Institute of Technology

Chicago, Illinois USA

apopovic@hawk.iit.edu

0. Abstract

Open-source code: https://github.com/groundsada/Single-Image-Facial-Superresolution-GAN

Super Resolution is the process of generating higher resolution images from lower resolution images. High resolution images offer a higher pixel density and therefore offer more details than the original images. Face super resolution, also known as face hallucination, is a domain-specific image super resolution problem. In the real world there are plenty of scenarios where face images are of low resolution or low quality due to multiple factors like photographic conditions and physical imaging. [1]

Current models use facial prior knowledge to super-resolve facial images, by utilizing Generative Adversarial Networks (GAN) to reach results. Well-trained GAN models can provide sufficient priors to generate high resolution face images. The problem with this approach is the limitations of the GAN models which is summarized in two points. The first is that a face GAN is exhausting on the system resources as a face GAN is not designed specifically for face super resolution and will need extra adaptations. The second is that GAN models are trained on limited datasets, and therefore they can only deal with a specific kind of face images. [1]

One more issue is that GAN models output one generated result from a specific input. Ideally, users will want to be able to control the generative strength to meet requirements. Prior work regarding the problem includes face super resolution with the use of face priors and face super resolution with the use of GAN priors. With face priors, we use facial priors to preserve the identity information and generate faithful face details. The facial priors are fed to the network to create facial landmarks, face parsing maps, and facial attributes, and we utilize a convolution neural network to obtain face component heatmaps to get to facial super resolution. With GAN priors, we make even more advancements in face super resolution in extracting feature maps and latent vectors and using them to adapt the face GAN prior to handle face super resolution tasks. [1]

The proposal is creating a controllable generative face super resolution framework that can construct faithful face images with realistic face details without the need for additional priors like facial priors and GAN priors, while also achieving state-of-the-art performance with the task of face super resolution. Additionally, the framework's generative strength is adjustable to meet requirements such as continuous upscaling factors. The proposed framework is composed of an encoder-generator architecture, a style modulator, and a facial modulator. The model is end-to-end trainable and converges fast. [2]

1. Problem statement

In the age of deep learning and neural networks, a key pressing issue is still largely unresolved: how do we recover the finer texture details when we super-resolve at large upscaling factors? Despite advances in accuracy and speed of single image super-resolution using faster and deeper convolutional neural networks, this issue is still at the forefront of the field. The choice of the objective function plays a major role in how optimization-based super-resolution approaches behave. [2]

Minimizing the mean squared reconstruction error has been the main focus of recent studies. The resulting estimates have high peak signal-to-noise ratios, but they are perceptually disappointing in that they frequently fall short of the fidelity anticipated at the greater resolution. [1]

2. Proposed solution

In our project, we investigate literature and build a SRGAN, a generative adversarial network (GAN) for superresolution of images (SR). We are basing our work on the literature that suggests the first framework capable of predicting natural images that are photorealistic for scaling factors of 4. We provide a perceptual loss function that combines an adversarial loss with a content loss to accomplish this. [3]

Super-resolution refers to the extremely difficult task of estimating a high-resolution (HR) image from its low-resolution (LR) equivalent (SR). SR has a variety of uses and has attracted a lot of interest from the computer vision research community. In our work, we propose a super-resolution generative adversarial network (SRGAN), for which we use a deep residual network (ResNet) with skip-connection as the sole optimization target and diverge from MSE. Unlike previous works, we define a novel perceptual loss by combining high-level feature maps from the VGG network with a discriminator that encourages solutions that are perceptually difficult to distinguish from HR reference images. [1]

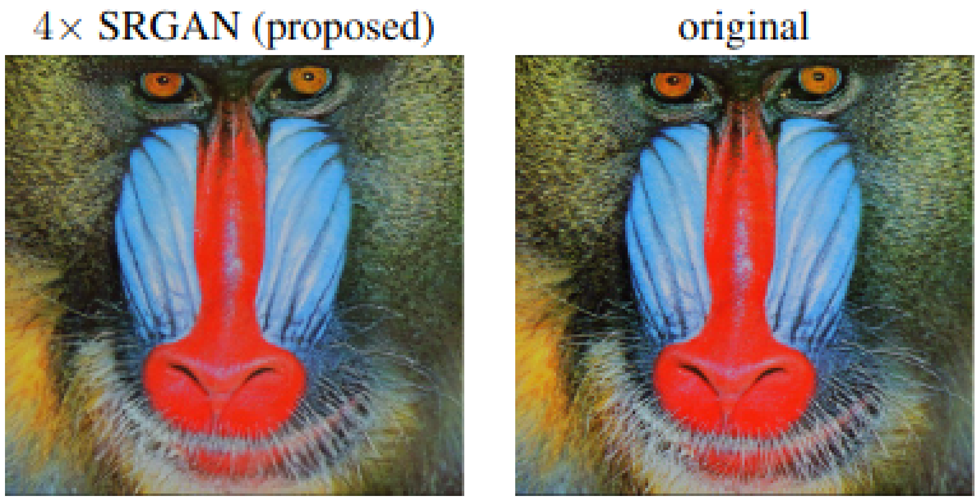

Figure 1 shows a photorealistic image that was superresolved with a 4 upscaling factor.

3. Implementation details

GANs, or Generative Adversarial Networks, are predictive models used for unsupervised learning. They were created by Ian Goodfellow in 2014. GANs are essentially made up of a system of two competing neural network models that can analyze, capture, and copy the variations within a dataset. A generator and a discriminator are both present in GANs. The Generator attempts to trick the Discriminator by creating phony samples of data (such as an image, audio, etc.). On the other hand, the Discriminator tries to tell the difference between the genuine and fraudulent samples. Both the Generator and the Discriminator are neural networks, and throughout the training phase, they compete with one another. The procedures are repeated multiple times, and each time, the Generator and Discriminator become better at what they are doing [4].

The generative model, which is trained to try to increase the likelihood that the Discriminator would make a mistake, captures the distribution of data. On the other hand, the Discriminator is based on a model that calculates the likelihood that the sample it received came from the training data and not the generator. The Discriminator is seeking to reduce its reward in the minimax game that the GANs are designed as, while the Generator is aiming to maximize its loss by minimizing the Discriminator's reward [5].

As the name implies, SRGAN is a method of creating a GAN that uses both an adversarial network and a deep neural network to create pictures with greater resolution. This specific sort of GAN is very helpful in enhancing features in native low-resolution photos while reducing mistakes [6].

One of the first methods that enables the model to reach an upscaling factor of approximately 4x for the majority of picture visualizations is the notion of SRGANs. It is a very difficult process to estimate and create a high-resolution image from a low-resolution image. In the past, CNNs were employed to create high-resolution pictures that trained more quickly and accurately. However, occasionally they are unable to retrieve more minute features and frequently produce fuzzy photos. Most of these problems are resolved by the proposed SRGAN architecture, which produces high-quality, cutting-edge pictures.

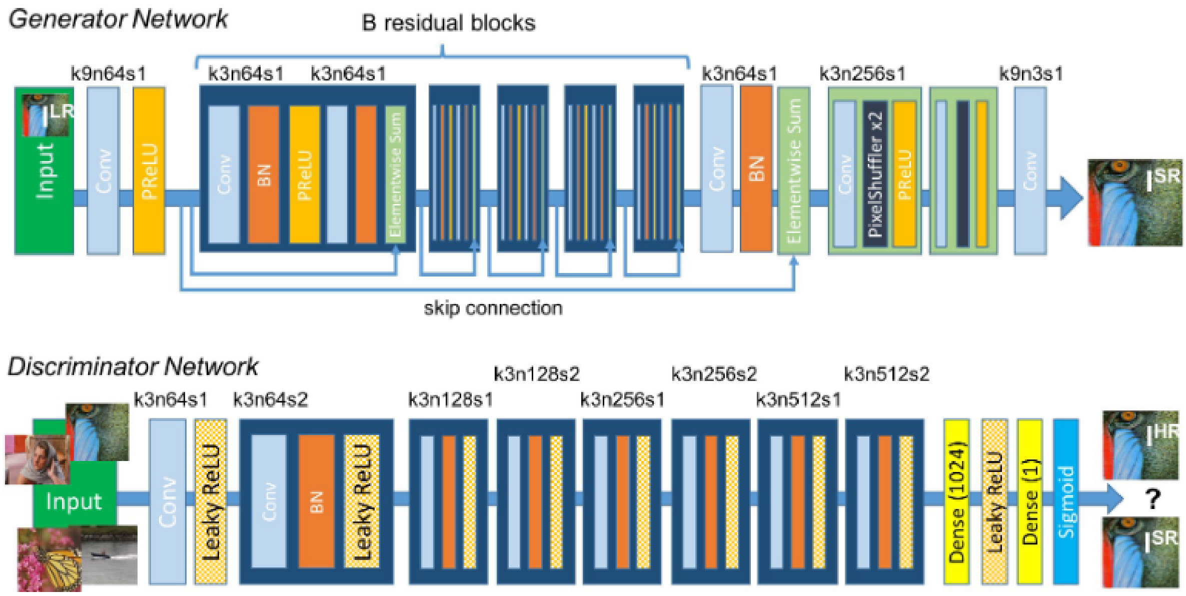

Figure 2: Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

In our SRGAN implementation, we reconstruct the GAN that is based upong the VGG19 model. Then, we attempt to fine-tune the GAN model with a dataset called FFHQ [7]. The dataset consists of ~70,000 images of facial data. This facial data is downloaded directly from source. Then, we wrote a python script that preprocesses the dataset into normalized High-Resolution images, and Low-Resolution images.

After that, we built our model in TensorFlow and Keras. The model is resource-intrusive, and our best local machine has an Nvidia GTX 1660 Ti which proved to be insufficient to reasonable time for training. Therefore, we had to resort splitting the training between the Google Colab platform and Paperspace platform to be able to train the model on the Coco dataset, and fine tune it on the FFHQ dataset.

Figure 3: The results of the project on low resolution images of Alan Cramb, a former president of the Illinois Institute of Technology. From left to right: input image with dimensions under, and next to it the result of the SRGAN.

4. Results and Discussion

In our project, we have investigated multiple approaches to facial image super-resolution. We picked the literature that promised the most state-of-the-art single-image superresolution and we based our work on it. We reconstructed the architecture suggested in the literature, and fine-tuned it with a face specific dataset called FFHQ (which is basically face images pulled from Flickr). We attempted to do the same with another dataset called Flickr8k, but fine-tuning with FFHQ yielded better results.

In our investigation and attempts to optimize, we have found that the Adam optimizer yielded the least loss, and the best quality superresolution. Some literature suggests that AdaBelief might be better suited for SRGANs. Further investigation is required on the subject matter. Future steps include utilizing more advanced models such as VGG56 and ResNet-56 to achieve higher-fidelity.

5. Research Team

Mohammad Firas Sada, Antoine Courbot, Aleksander Popovic

Research effort distribution:

- Preliminary work and first draft: Mohammad Firas Sada, Antoine Courbot, Aleksander Popovic
- SR GAN code reproduction, customization, and testing: Mohammad Firas Sada
- Dataset Generator: Mohammad Firas Sada, Antoine Courbot, Aleksander Popovic
- Extended literature review: Antoine Courbot, Mohammad Firas Sada
- Hyperparameter experiments: Aleksander Popovic, Antoine Courbot
- Fine-tuning and new functionalities: Mohammad Firas Sada, Aleksander Popovic
- Results analysis: Aleksander Popovic, Antoine Courbot
- Report background content: Antoine Courbot, Mohammad Firas Sada
- Report implementation content: Mohammad Firas Sada, Antoine Courbot, Aleksander Popovic
- Final report writing and editing: Mohammad Firas Sada, Antoine Courbot, Aleksander Popovic

6. References

[1] He, J., Shi, W., Chen, K., Fu, L., & Dong, C. (2022). GCFSR: A generative and controllable face Super Resolution Method without facial and gan priors. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr52688.2022.00193

[2] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr.2017.19

[3] Singla, K., Pandey, R., &; Ghanekar, U. (2022). A review on single image super resolution techniques using generative Adversarial Network. Optik, 266, 169607. https://doi.org/10.1016/j.ijleo.2022.169607

[4] Chudasama, V., & Upla, K. (2021). Computationally efficient progressive approach for single-image super-resolution using generative Adversarial Network. Journal of Electronic Imaging, 30(02). https://doi.org/10.1117/1.jei.30.2.021003

[5] Tang, Y., Liu, C., & Zhang, X. (2022). Single image super-resolution using Wasserstein generative adversarial network with Gradient Penalty. Pattern Recognition Letters, 163, 32–39. https://doi.org/10.1016/j.patrec.2022.09.012

[6] Naik, D. A., Sangeetha, V., & Sandhya, G. (2021). Generative adversarial networks based method for generating photo-realistic super resolution images. 2021 Emerging Trends in Industry 4.0 (ETI 4.0). https://doi.org/10.1109/eti4.051663.2021.9619393

[7] Liu, P., Hong, Y., & Liu, Y. (2018). Dual discriminator generative adversarial network for Single Image Super-Resolution. 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS). https://doi.org/10.1109/icsess.2018.8663715