DS5500 Phase-1 Project Report
Super-Resolution of Images using Deep Learning

Nitin Kumar Mittal

1 Motivation

Computer vision algorithms have become a critical component of applications we use every day. Major use-cases include object detection, object classification, semantic segmentation, and pattern recognition. However, the performance of these algorithms is dependent on the quality of input data which is mostly pictorial. In practical applications, often the available pictorial data is of low quality, affected by factors such as low resolution, presence of noise, hallucination, and pixelation of objects within an image, etc.

Therefore, the project explores state-of-the-art techniques in deep learning to generate high-resolution images from otherwise low-resolution counterparts to resolve these challenges. In real-world applications, improving quality of available pictorial data will help computer vision techniques achieve better results.

2 Dataset

There are 2 phases of this project. In phase-1, the project specifically targeted to provide deep learning models to perform super-resolution on images of humans faces. For this, the project used the fake human faces dataset from Kaggle [1] which contains nearly 5000 images of fake human faces generated by GANs. In phase-2, the project will target to provide deep learning models to perform super-resolution of breast cancer histology images. For this, the project will use the BACH dataset from ICIAR 2018 grand challenge which contains nearly 400 H&E stained breast histology microscopy images. [2].

3 Implementation

3.1 Methodology

In phase-1, the project trained and tested End-2-End Pixel-2-Pixel SR-UNet architecture to generate high-resolution images of human faces from low-resolution counterparts. This approach is self-supervised because during training the low-resolution input images are produced from ground-truth high-resolution images using various interpolation methods. Figure 1. represents the workflow used in phase-1.

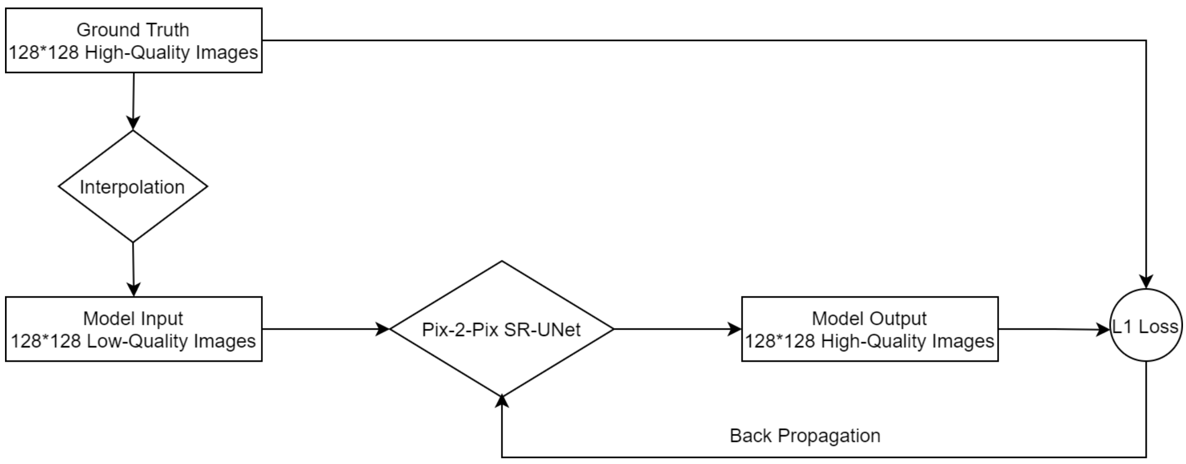

Figure 1: Workflow to generate high-resolution images of human faces

Bicubic interpolation was used to generate low-quality images by reducing resolution by nearly 3 times.

3.2 Model Architecture

The architecture of SR-UNet is illustrated in Figure 2. The illustrated architecture is inspired by original U-Net architecture.

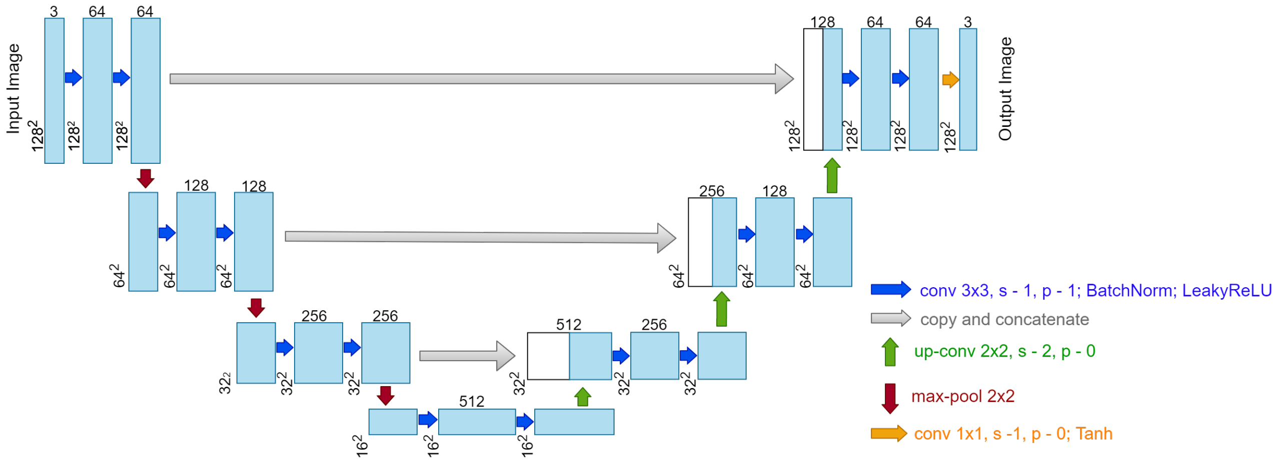

Figure 2: SR-UNet Architecture

A general UNet architecture consists of a contracting path/ analysis path (left side) and an expansive path/ synthesis path (right side). In the analysis path, deep features are learned, and in the synthesis path, segmentation is performed based on the learned features. Additionally, UNet uses connections between contracting and expansive paths to propagate dense feature maps from the analysis path to the corresponding layers in the synthesis part. In this way, the spatial information is applied to the deeper layer, which significantly produces a more accurate output segmentation map.

3.3 Training and Validation

Out of 5000 images of fake human faces, 4250 images were used for training, 500 for validation, and the remainder 250 for testing SR-UNet. Train and validation batch-size of 64 and 128 respectively were used. The model was trained for 370 epochs. L1/ MAE loss function was used for SR-UNet. The initial learning rate was set to 5e-4 and a learning-rate scheduler was employed to reduce the learning rate by a factor of .999 after every 100 training batch updates/steps. Adam with $b_1 = .5$ and $b_2 = .9$ was used as an optimizer for SR-UNet. The training and validation losses across training and validation steps are provided in Figure 3.

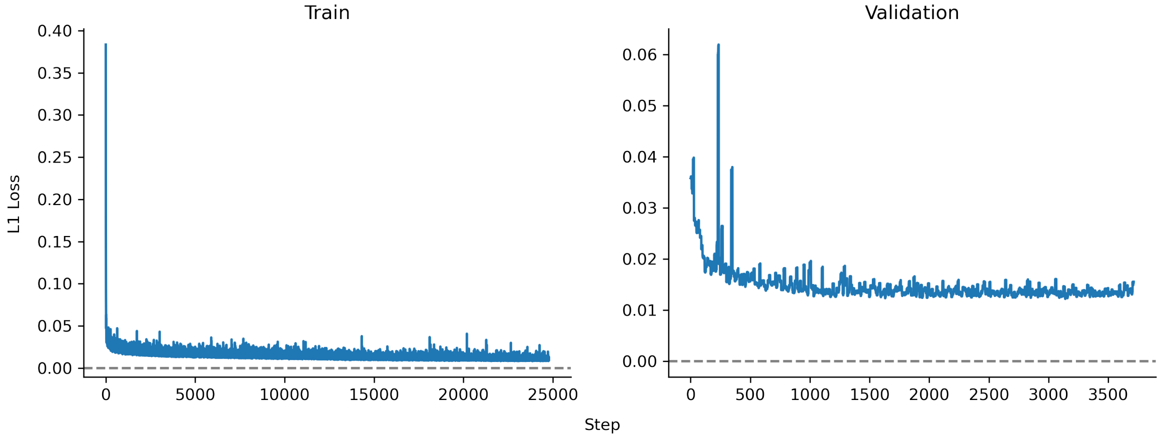

Figure 3: Traning and validation loss across training and validation steps respectively. Best model was picked on basis of minimum validation loss.

3.4 Testing

Peak-Signal-to-Noise-Ratio (PSNR) [3] and Structural similarity index measure (SSIM) [4] are used to evaluate the performance of trained SR-UNet. Figure 4. represents input and output SSIM and PSNR values for test images before and after super-resolution respectively. The larger SSIM and PSNR output values for test images indicate that SR-UNet can improve the quality of test images.

The sample test inputs and corresponding ground truth and model outputs can be seen be Figure 5.

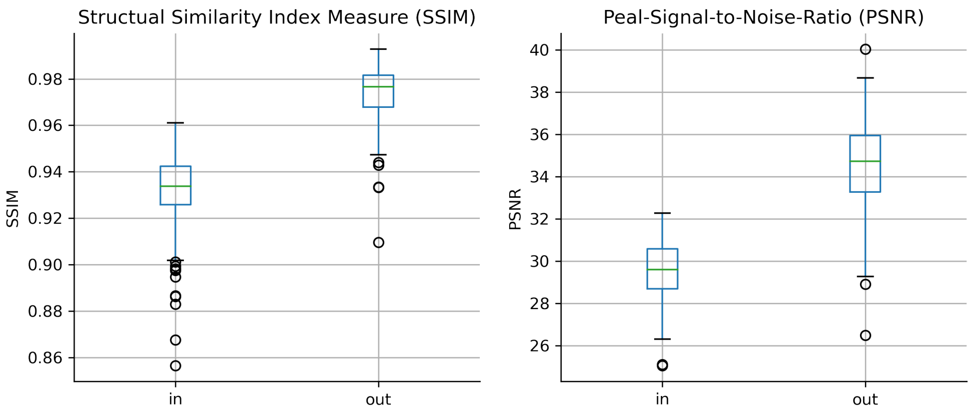

Figure 4: SSIM and PSNR value for test images

Figure 5: SR-UNet is able to improve resolution of human faces

4 Conclusion

Overall, SR-UNet can be used to improve the resolution of human faces, although thorough research is further required to identify the extent to which SR-UNet can achieve super-resolution. Also, this approach is limited to the classes and variations in samples on which the model is trained, i.e an SR-UNet trained to increase the resolution of human faces will not be able to improve the resolution of cars or other objects.

References

[1] Glasses or No Glasses - Kaggle

[2] ICIAR 2018 Grand Challenge on Breast Cancer Histology Images

[3] Peak-Signal-to-Noise-Ratio (PSNR)

[4] Structural similarity index measure (SSIM)