Machine Learning Engineer Nanodegree

Capstone Project

Erick Costa de Farias
January 30th, 2021

I. Definition

Project Overview

Can you imagine identifying a tumor in a pixelated low resolution image? It's a hard task, even for the most experient radiologists. However, acquiring high-res computerized tomography images may be expensive and risky, as the level of contrast ingested by the patient needs to be higher. Thus the need to find ways improving artificially the resolution of a CT image.

In this project the DeepLesion dataset provided by the National Institutes of Health's Clinical Center will be used to train an implementation of a GAN super resolution algorithm named CIRCLE-GAN, whose objective is to super-resolved a given CT image.

Problem Statement

Medical images differ greatly from the usual images considered in pattern recognition and computer vision problems; these images convey an amount of per pixel information that surpasses the human capabilities in distinguishing among many levels of gray. The analysis of medical images involves many different techniques to measure spatial distributions of physical attributes of the human body, seeking a better understanding of diseases. In this regard, high-resolution images, which contain a high pixel density, are greatly desired and often required as these images can offer more detail and be critical for applications in medical

imaging. However, in order to obtain computerized tomography (CT) high-resolution images, it's necessary to expose the patient to long scan times and subject the inpatient to higher contrast doses. In either case, expenses and risks will increase dramatically, which is rarely clinically applicable. Thus, we need ways to super-resolve CT images, artificially, in order to improve the accuracy of subsequent diagnostic and prognostic steps.

Metrics

In order to measure the performance of our results, we'll look to:

- Qualitative similarity of super-resolved images to the original HR
- Peak signal-to-noise ratio[1] is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It denotes the ratio between the maximum possible intensity value of a signal and the distortion between the input and output images

$$\begin{aligned} PSNR &= 10 \cdot \log_{10}\left(\frac{(I_{in}^{(max)})^2}{MSE}\right) \\ &= 20 \cdot \log_{10}\left(\frac{I_{in}^{(max)}}{\sqrt{MSE}}\right), \end{aligned}$$

- The structural similarity index measure (SSIM) is a method for predicting the perceived quality of digital television and cinematic pictures, as well as other kinds of digital images and videos. SSIM is used for measuring the similarity between two images. The SSIM index is a full reference metric; in other words, the measurement or prediction of image quality is based on an initial uncompressed or distortion-free image as reference.

$$SSIM = \frac{(2\mu_{\mathbf{X}}\mu_{\mathbf{Y}} + \kappa_1)(2\sigma_{\mathbf{XY}} + \kappa_2)}{\left(\mu_{\mathbf{X}}^2 + \mu_{\mathbf{Y}}^2 + \kappa_1\right)\left(\sigma_{\mathbf{X}}^2 + \sigma_{\mathbf{Y}}^2 + \kappa_2\right)}.$$

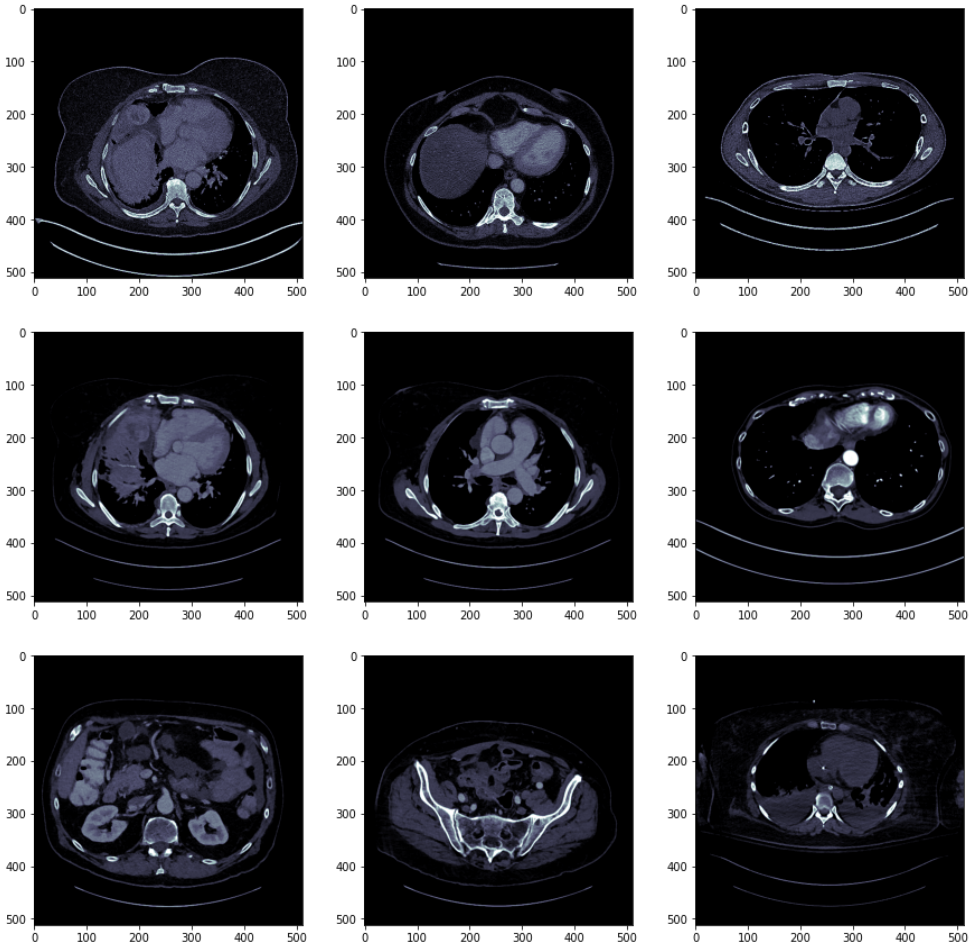
[1] Implementation details of these metrics can be found in *L. Rundo et al., "MedGA: A novel evolutionary method for image enhancement in medical imaging systems," Expert Syst. Appl., vol. 119, pp. 387–399, 2019.*

II. Analysis

Data Exploration & Visualization

The images provided in the deep lesion dataset will be used for this project.

"The National Institutes of Health's Clinical Center has made a large-scale dataset of CT images publicly available to help the scientific community improve detection accuracy of lesions. While most publicly available medical image datasets have less than a thousand lesions, this dataset, named DeepLesion, has over 32,000 annotated lesions identified on CT images, representing 4,400 unique patients"[2]



[2] Available on https://www.nih.gov/news-events/news-releases/nih-clinical-center-releases-dataset-32000-ct-images

All images are provided in a .png format, with the identification of patient and slice in the file name. Each image represents a CT slice of size 512x512px. The pixel values have been transformed to a positive integer, ranging roughly from -31000 to 35000, so these images could be distributed in a simple format.

Some slices are purely air or water, depending on the body region being represented.

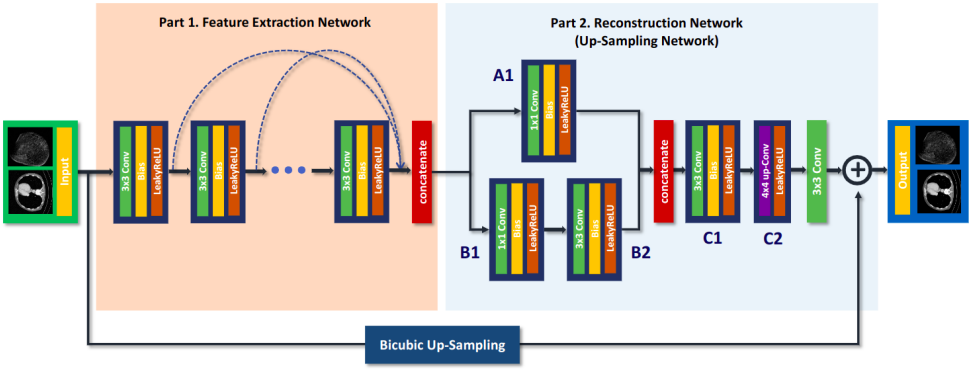
Algorithms and Techniques

The algorithm implemented is a CIRCLE GAN, as proposed by You et al, in "CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE)." arXiv preprint2018, arXiv:1808.04256.".

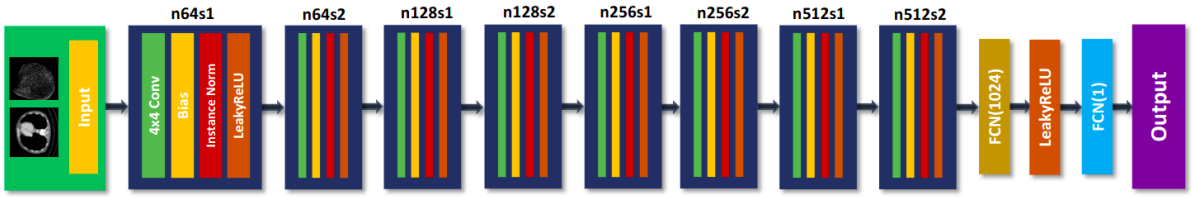
Adversarial learning has become a popular strategy to enable the learning of feature representation from complex data distributions. It's based on a generative adversarial network, which is defined as a mini-max game, where a generator G(x) competes with a critic D(x). The task of the generator is to learn how to map an image from a source domain X to a domain Y, whereas the critic D(x) must learn to distinguish the images generated by G(x) and the real ones.

The CIRCLE-GAN is a network in network architecture that performs adversarial learning in a cyclic fashion. It enforces the mappings between the source and target domains by combining four types of loss functions: adversarial loss; cycle-consistency loss; identity loss and joint sparsifying transform loss.

Generators architecture:



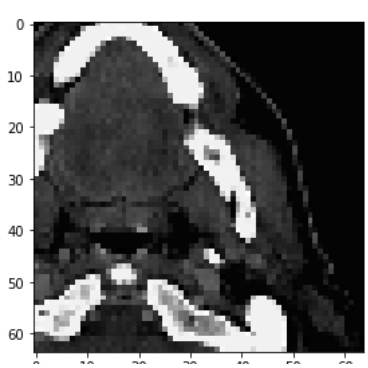
Discriminators architecture:

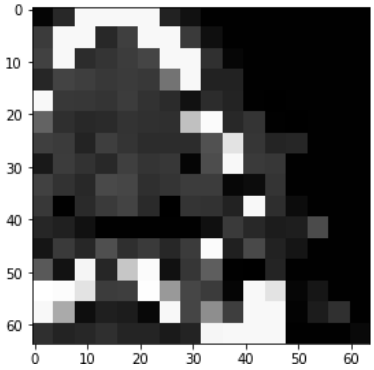


Benchmark

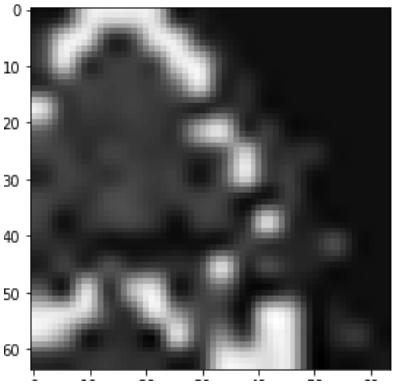
As a benchmark, a bicubic interpolation algorithm will be considered, as it's a classic interpolation method used in the filed.

Consider the following HR image, With a low resolution counterpart, downscaled to ¼ of its original resolution:





Applying the bicubic interpolation to the low resolution version of the image results in this:



III. Methodology

Data Preprocessing

The raw images were distributed as .png files, of size 512x512 pixels.
In order to feed the training of the proposed algorithm, these images were subject to the following preprocessing pipeline:

1. Transformation to an array of shape (512, 512, 1). There is only 1 channel in the third axis, because the images are in grayscale;

2. The pixel values were transformed to CT Hounsfield Value range (HU) and then scaled to the unit interval [0,1].

3. Performed a random crop, extracting a patch of 128x128 pixels;

4. Cropped patches were downscaled using bicubic interpolation to ½ of the initial resolution, resulting in 64x64 images. These were treated as the ground-truth HRCT images;

5. Random noise was added to the image with 30% of chance;

6. Gaussian blur was applied to the image with 30% of chance;

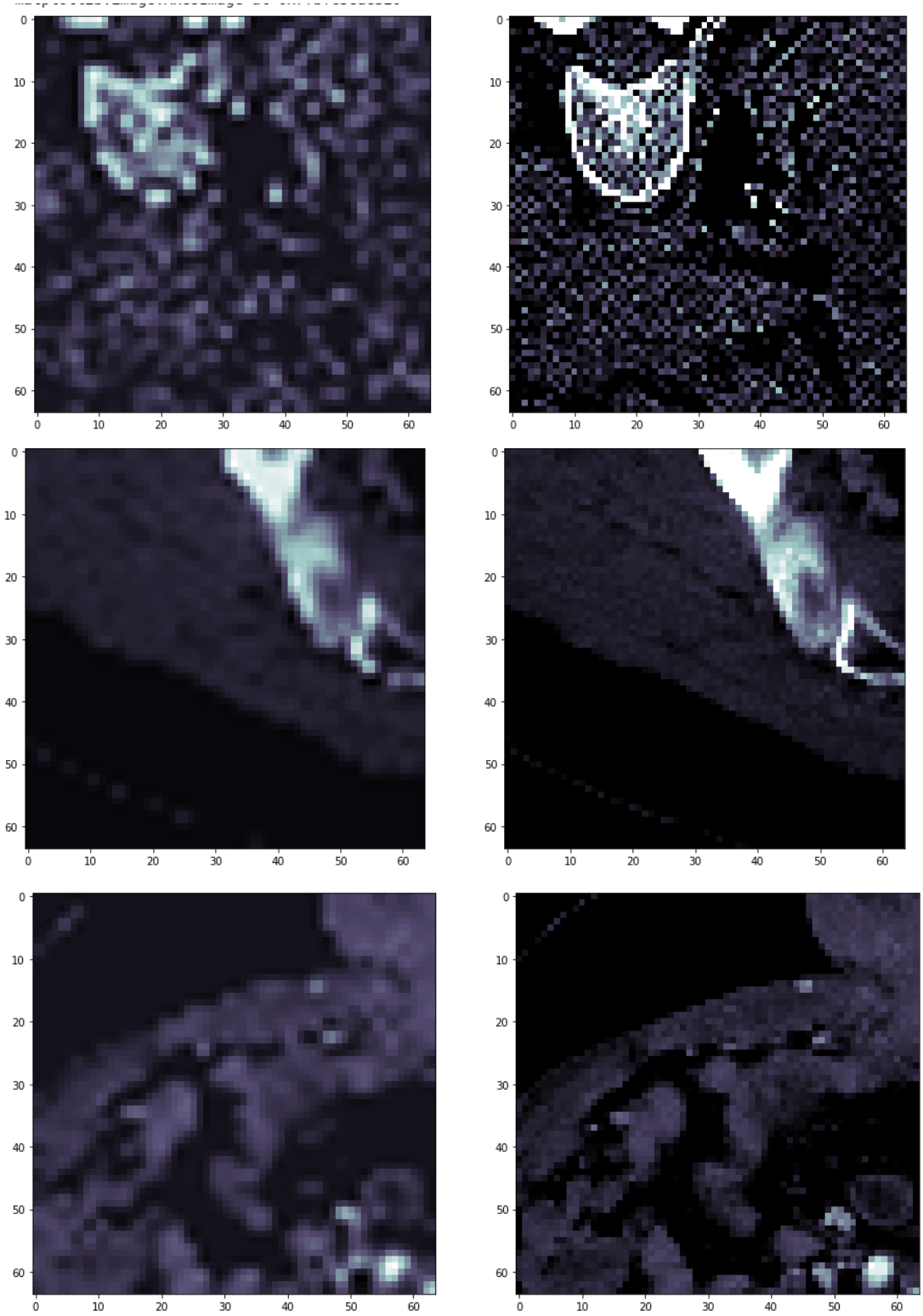
7. A random flip was applied to the image with 50% of chance;

8. Validated if the generated crop had more than 50% of tissue, in order to ignore images composed mainly by air or water, as this could impair the network learning ability;

9. Cropped patches were downscaled using proximal interpolation (nearest neighbor) to ½ of the initial resolution, resulting in 32x32 images. These were treated as the input data, i.e. LRCT images;

10. For convenience in training our proposed network, we up-sampled the LR image via proximal interpolation to ensure that x and y are of the same size

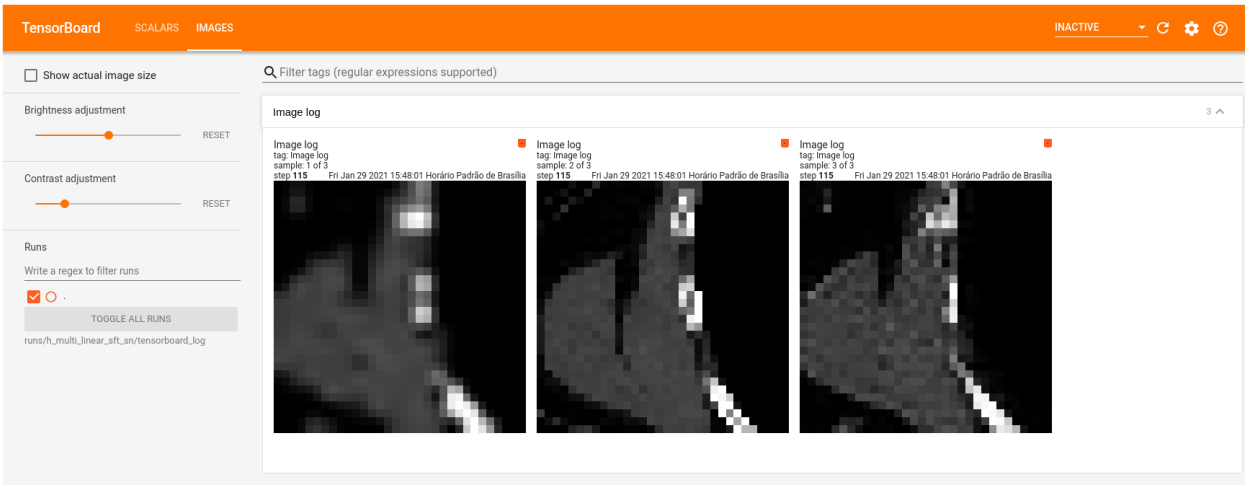
Examples of LRCT | HRCT input patches:



Implementation

After generating the training data, the resulting images (~15000) were split in batches of 32 and fed into the CIRCLE GAN, which was trained for 150 epochs.
The discriminator was trained with a learning rate of 0.00005, and the generator with rate of 1/3 of it, following the Two-Times Update Rule, which is known to support GAN convergence.
Moreover, due to the varied sizes of real lesions in CT images, we utilize the Spatial Pyramid Pooling (SPP) layer in the output of discriminators, to avoid the effect of resizing, similar to what is reported in Kim, Daeun Dana, et al. *"Generating Pedestrian Training Dataset Using DCGAN"*. Proceedings of the 2019 3rd International Conference on Advances in Image Processing, ACM, 2019, p. 1–4. DOI.org (Crossref), doi:10.1145/3373419.3373458.

After each epoch, we tracked the "eyes" of the GAN, logging into tensorboard the LR, HR and SR extracted from the test data, in order to verify the learning.



The CIRCLE-GAN proposed loss function, as described below, is hard to inform if the learning is converging or not, as common in general adversarial training problems.

CIRCLE-GAN Loss function:

$$\begin{aligned}\mathcal{L}_{\text{GAN}-\text{CIRCLE}} = \mathcal{L}_{\text{WGAN}}(D_Y, G) + \mathcal{L}_{\text{WGAN}}(D_X, F) \\ + \lambda_1 \mathcal{L}_{\text{CYC}}(G, F) + \lambda_2 \mathcal{L}_{\text{IDT}}(G, F) \\ + \lambda_3 \mathcal{L}_{\text{JST}}(G),\end{aligned}$$

Refinement

During the training process, I realized the results were overly smooth. Researching about this, I found the implementation of a network conditioning called Spatial Feature Transform. The authors argued that this conditioning could be helpful to the recovery of texture granularity in super resolution tasks. I implemented the SFT, generating 10 priors to the network, binning the pixel intensity in 10 buckets. This improved the overall quality of the result.

IV. Results

Model Evaluation and Validation

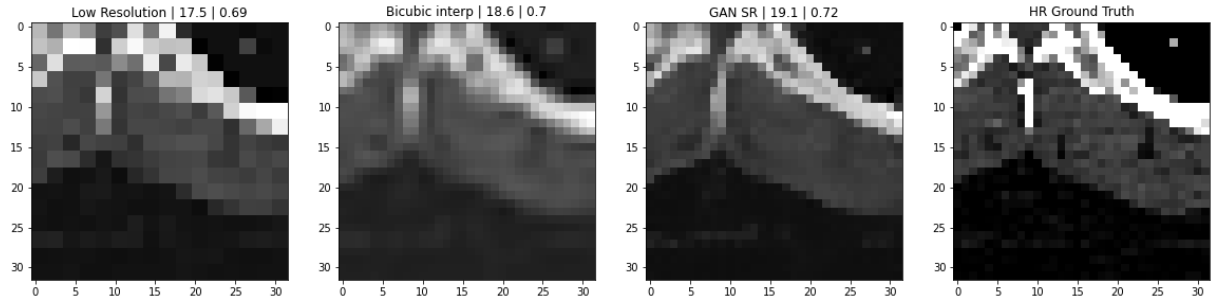
PSNR

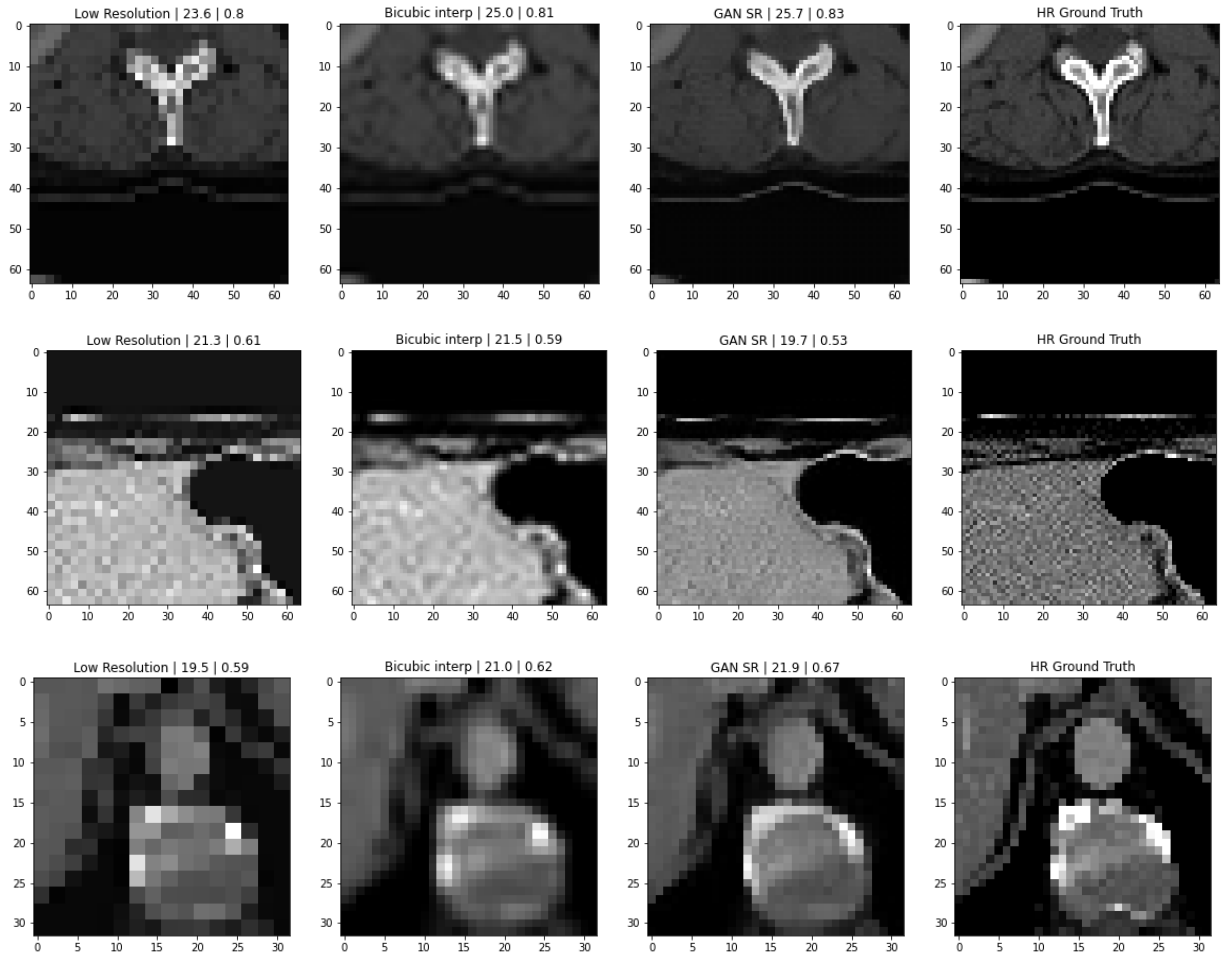
The average PSNR for the bicubic interpolation was 20.6 (3.9 std)
The average PSNR for the GAN model was 21.97 (4.8 std)

SSIM

The average PSNR for the bicubic interpolation was 0.665 (0.15 std).
The average PSNR for the GAN model was 0.674 (0.19 std).

Qualitative evaluation:





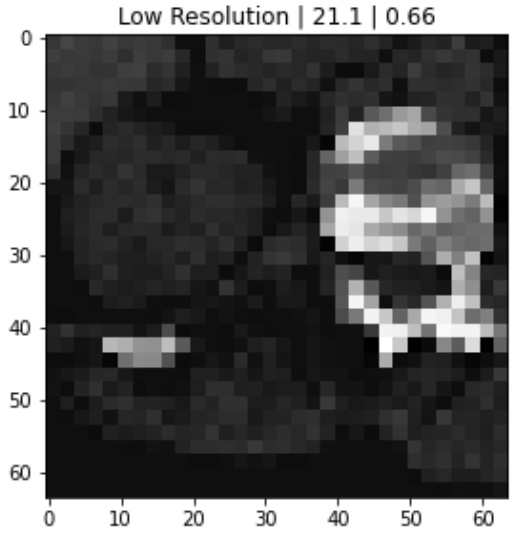
Justification

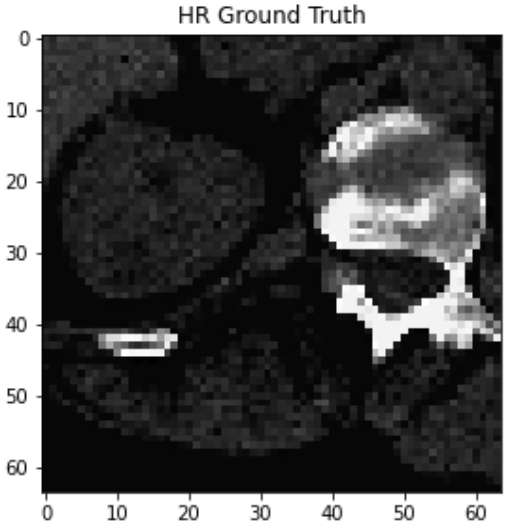
The proposed Super Resolution Generator performs better than the established baseline, quantitatively and clearly qualitatively as well, generating sharper images with a superior quality of reconstruction.

V. Conclusion

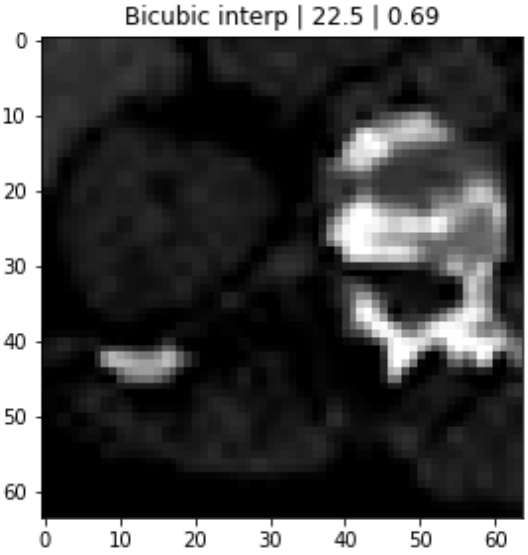
Free-Form Visualization

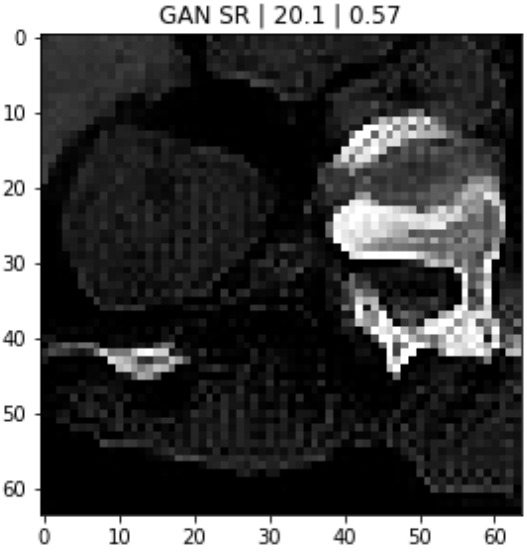
Considering the following LRCT and HRCT pair of images:





We have the following bicubic interpolated and GAN super-resolved versions:





Although the quantitative quality of the image can vary in terms of PSNR | SSIM (shown on the labels of each image), the GAN version results in a much sharper image, which can support better diagnostic and prognostic tasks when performed by specialists.

Reflection

Medical Imaging super resolution remains a challenging task, from the difficulty to acquire larger standardized datasets to the high computational needs to process them. However, I believe that the high potential in aiding the medical community to perform more accurate and fast diagnostic and prognostic tasks outweigh the considered costs.

The present project implemented a state-of-the-art architecture to solve a super resolution problem. The results have beaten the baseline performance and shown to be better than a very commonly used interpolation method, but still seem to be far from what was reported in the original paper. Perhaps the network could be trained for longer, but specific computation limitation was found by using the GPU available in google colaboratory.

Improvement

Lyu et al[3] proposed a novel method for SR using an ensemble of GANs.
In their proposed scheme five different model-based SR algorithms are used to process the LR images in order to obtain five different datasets, which are fed into separate GANs for training and generating five super-resolution counterparts. As a final step, a CNN is used as the ensemble learning to obtain the final super-resolution result. In this work, it was reported that the application of an ensemble resulted in a great reduction in artifacts and noise, with richer textual details in the super-resolution images, compared to the GAN-only versions. Quantitatively, it also reached the highest PSNR and SSIM values among all the results. This is a certain improvement to be considered.

[3] *Q. Lyu, H. Shan and G. Wang, "MRI Super-Resolution With Ensemble Learning and Complementary Priors," in IEEE Transactions on Computational Imaging, vol. 6, pp. 615-624, 2020, doi: 10.1109/TCI.2020.2964201.*