CS230

Through Thick and Thin: MRI Super-Resolution Using a Generative Adversarial Network

Andrew Narcomey
aon1@stanford.edu

Angela Gu
angelajg@stanford.edu

Abstract

Image super-resolution, or generating a high-resolution image from a low-resolution one, is a popular problem within deep learning, but has not yet been applied widely to medical images. We seek to develop a technique for MRIs that could produce high-resolution scans from low-resolution ones, which would shorten MRI scanning time for vulnerable patients. Basing our model off of SR-GAN, a generative adversarial network that prioritizes features over pixel-to-pixel comparisons, we find that our preliminary results preserve some interesting features and show promise for future work.

1 Introduction

Because MRI (Magnetic Resonance Imaging) does not emit X-rays or other radiation, it is the preferred scanner for many medical evaluations, particularly when frequent imaging is necessary or when examining soft tissue differentiation or brain trauma. However, the typical scan requires patients to lie still inside the scanner for half an hour or more, which may be difficult, particularly for pediatric, claustrophobic, or very ill patients [1]. In order to allow a physician to diagnose cancer and determine its stage, the MRI scan must be of a sufficiently high resolution. Therefore, there may be a trade-off between quality of the scan and the patient's comfort during the process.

Working with Dr. Heike Daldrup-Link and Dr. Anuj Pareek, we hope to develop a deep learning algorithm that can reconstruct high-resolution images from rapidly acquired low-resolution data. This would shorten overall scanning time, thus increasing hospital workflow and decreasing patient discomfort. Using high- and low-resolution MRI scans from sixteen patients, we used a generative adversarial network (GAN) based off of the SRGAN network developed by Ledig *et. al* [2]. We modified this model to work with the single color channel of our data; to incorporate actual low-resolution data, instead of artificially deteriorating high-resolution images; and to incorporate the sequential nature of an MRI.

2 Related work

Our project falls under two larger problems: the application of deep learning to medical image analysis, and algorithms for image super-resolution. Although the former is still a new field, Litjens *et. al.* provide a fairly comprehensive survey on deep learning techniques in medical image analysis [3]. Patient privacy makes large datasets difficult; therefore, most of the literature focuses on convolutional neural networks (CNNs) and uses transfer learning due to the smaller datasets available. For instance, Bien *et. al.* uses AlexNet to diagnose different knee injuries through about 1,400 MRIs [4], while Fauw *et. al.* applies a 3D-CNN to screen head CT images for neurological damage [5]. As

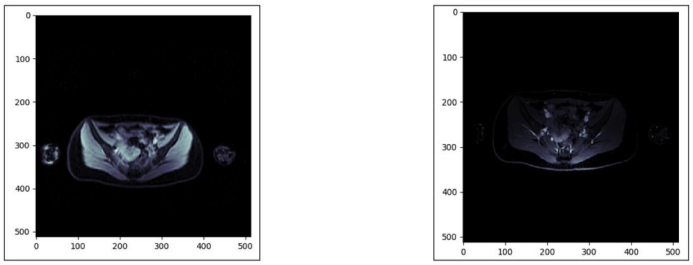
CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

of the survey's most recent publication (June 2017), there were no peer-reviewed articles applying GANS to medical images [3]. However, GANs offer a promising opportunity of research due to their unsupervised nature; due to patient privacy laws, there is much more readily available unlabeled medical data than labeled, and unsupervised methods could take advantage of this fact.

On the other hand, image super-resolution–increasing the size of a small image while minimizing its drop in quality–is a popular field within computer vision and has applications to a wide variety of fields. Traditional super-resolution techniques use interpolation, but this fails to preserve the sharp edges that are necessary for medical imaging [6]. SRCNN, the first deep learning method to outperform traditional ones, is a simple CNN with three convolutional layers, but is very sensitive to hyperparameter changes [7]. VDSR is a deeper network that utilizes VGG-Net [8]. Finally, SRGAN uses a residual network (SRResNet) as the generator and bases its discriminator off of VGG-Net [2]. We ultimately chose SRGAN due to its unique loss function, discussed further in section 4.

3 Dataset and Features

This dataset was provided by Dr. Heike Daldrup-Link and Dr. Anuj Pareek, and contains paired high- and low-resolution scans from 16 patients. Each scan is composed of several hundred "slices" of varying thickness. Images were dicom (.dcm) files, which each consisted of an array of gray-scale pixels (512 x 512) with 1 channel, as well as meta-data about the scan and the patient.



(a) Low-resolution scan. (b) Corresponding high-resolution scan.

Figure 1: Low- and high-resolution images, paired through our preprocessing. The high-resolution scan appears thinner because each scan contains more images, and thus each image features a thinner slice.

For each patient, these scans were obtained by sequentially low-resolution MRIs scan less thoroughly and, thus, have thicker and fewer slices. We used the meta-data to pair each low-resolution scan with its best-fitting high-resolution scan by minimizing the Euclidean distance between the position of the patient at each scan. Altogether, we created 8479 image pairings.

4 Methods

Image super-resolution techniques typically focus on minimizing MSE (mean-squared error) between the pixels of the two images:

$$MSE(I, \hat{I}) = \frac{1}{n} \sum_{i,j} (I_{i,j} - \hat{I}_{i,j})^2$$

However, like any element-wise comparison, MSE can be very sensitive to small shifts in the image. This is a large concern for our problem for two reasons. Our paired images showed noticeable differences in the pixel values and placement (see Figure 1) due to the slight shifts in patient positioning between the two MRIs. Additionally, the edges of an MRI, which detect different organs, bones, tissues, etc., are more important for radiologists than the actual pixel values.

As a result, we chose to base our work after SRGAN, which focuses more on feature representation than pixel value in its loss function. Instead of MSE, SRGAN defines their content loss as the Euclidean distance between the feature representations of a reconstructed image $G_{\theta_G}(I^{LR})$ and its

reference high-resolution image I^{HR}:

$$l_{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2,$$

where $W_{i,j}$ and $H_{i,j}$ are the dimensions of the respective feature maps within the VGG network and $\phi_{i,j}$ refers to the feature map after the j-th convolution before the i-th max-pooling layer within the VGG19 network.

The adversarial loss is

$$l_{Gen} = \sum_{n=1}^{N} -\log D_{theta_D}(G_{theta_G}(I^{LR})),$$

and the perceptual loss is a weighted sum of adversarial and content loss:

$$l^{SR} = l_{VGG} + 10^{-3} l_{Gen}$$

For its generator, our GAN utilized SRResNet, a network with 16 residual blocks, each consisting of 2 convolutional layers with same padding and random normal initialization with a standard deviation of 0.02, followed by a batch normalization layer each. To improve the generator before adding the discriminator and to look at a baseline model, we trained the MSE-based SRResNet with our data prior to training the actual GAN. For our discriminator, we used a model similar to VGG-19, with alterations for our specific dataset (Model 1).

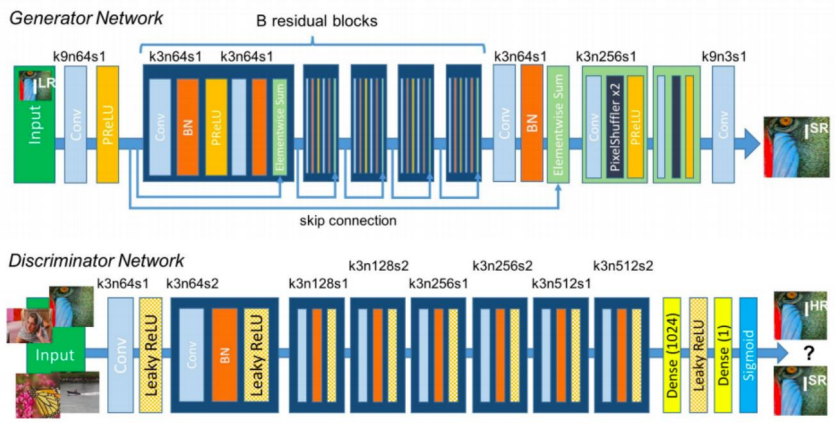
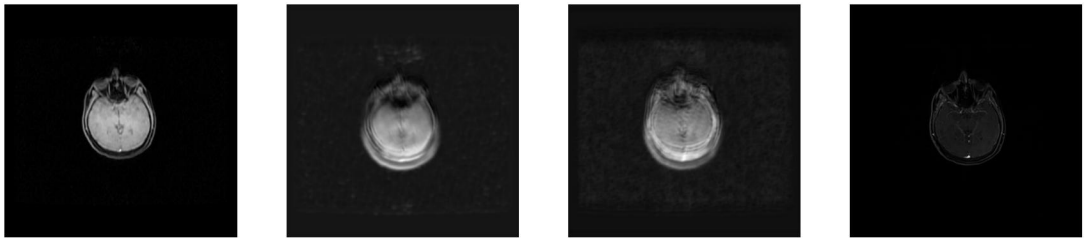


Figure 2: Original architecture of SRGAN [2], modified for our dataset for number of channels, dimensions, and low-resolution data (not constructed through bicubic downsampling, as the original model does). Note that, due to the padding, the dimensions of the generated output were the same as the input. This enabled us to use the generator as a baseline model for image super-resolution.

MRIs have an additional feature in that they are sequential; therefore, a slice's neighbors may contain useful information. Therefore, we ran a similar model, but with our "ground truth" input as a 3-channeled image containing a slice and its neighbors (Model 2). 6008 image pairings (one low-resolution slice–its best-matching high-resolution slice) were used to train Model 1 and 2741 were used for evaluation. Due to memory constraints, 907 image pairings (one low-resolution slice–its best-matching high-resolution slice, along with its neighbors) were used to train Model 2 and 749 were used for evaluation.

Ledig *et. al.* ran SRResNet over 100 iterations and SRGAN over 2000 iterations, which took several weeks–a process that would have cost thousands of dollars if we used an external service. We did not have the computational resources to do this, so all results are in a preliminary stage.



(a) LR Slice (b) SRResNet Slice (c) SRGAN Slice (d) HR Slice

Figure 3: Starting from the left: a low resolution slice, the corresponding generated image from SRResNet, the generated image from SRGAN, and the closest matching high resolution slice using Model 1.

5 Experiments, Results, & Discussion

5.1 Training

All networks were trained with the dataset described in Section 3. Due to GPU access constraints, experiments were affected in a number of ways. To avoid memory overflow constraints, a batch size of 1 was used. For training of Model 1, 11 patients' scans were used (6008 paired scans, or 12,016 512 x 512 x 1 images in total). The SRResNet for Model 1 was trained with 1 epoch over the dataset, and SRGAN was trained with 5 epochs. 5 patients were used for evaluation of these networks. For training of Model 2, due to storage constraints from having double the number of images per training example as a result of using the neighbors for the target image, 2 patients' scans were used. The SRResNet for Model 2 was trained with 5 epochs over the dataset and SRGAN was trained with 20 epochs. 1 patient was used for evaluation of these new networks.

5.2 Parameters

Adam optimization was used for training of all networks, with $\beta_1 = .9$ and $\beta_2 = .999$. A fixed learning rate of 10^{-4} was used for SRResNet and SRGAN used two learning rates, 10^{-4} and 10^{-5}, which were switched halfway through training. These hyperparameters were chosen to match those tested by Ledig *et. al.* [2]. Given that SRGAN is such a deep network, it is less sensitive to hyperparameter changes.

5.3 Evaluation

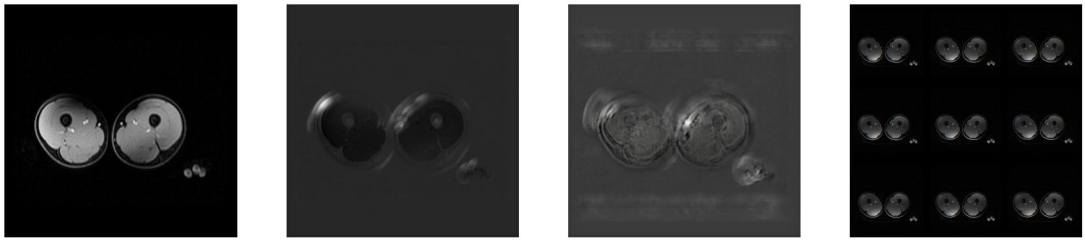
As the goal of this paper is to provide physicians with interpretable and clear images, our evaluation metrics include mean opinion score (MOS) and, as a rough approximation for feature preservation, MSE. Due to time constraints, MOS has not yet been evaluated, and is to be done by radiologists in the Stanford School of Medicine. MSE values are shown below in Table 1.

	Model 1	Model 2
SRResNet	$5.39 * 10^7$	$3.90 * 10^5$
SRGAN	$4.73 * 10^7$	$3.46 * 10^5$

Table 1: MSE for each model, evaluated on both networks. We use MSE as a rough approximation for feature preservation; MOS is a more important metric, since ultimately radiologists will be the users of this model, and since we designed our model to minimize VGG loss (feature loss).

5.4 Discussion

As seen in the results of Table 1, SRResNet had greater MSE than SRGAN for both models. This was also shown in Ludig *et al.* and is expected in this context as well as it is a simpler model [2]. Between Model 1 and Model 2, Model 2 had a much lower MSE, possibly a result of training with a smaller portion of the data for a greater number of iterations, and the smaller evaluation size.



(a) LR Slice (b) SRResNet Slice (c) SRGAN Slice (d) HR Slice

Figure 4: Starting from the left: a low resolution slice, the corresponding generated image from SRResNet, the generated image from SRGAN, and the closest matching high resolution slice with its neighbors using Model 2.

Figure 3 and Figure 4 are examples of the resulting images we obtained by processing slices through SRResNet and SRGAN, for both models tested. Using Model 1, the results from SRResNet and SRGAN largely mirror the low resolution, but become slighter blurrier as they have more noise around the edges. SRGAN improves upon SRResNet visually by reducing the noise or blur in the image. The images from SRGAN and SRResNet also still maintain most of the coloring of the original low resolution slice. We also observe a small detail at the bottom of the skull in the example in Figure 4 where the region starts to brighten and stand out more in the image, as appears in the HR slice. Using Model 2, the results are noticeably different, where both SRResNet and SRGAN results start to learn a different coloring to match some of the darker features of the HR images, as well as start to pick up finer details.

6 Conclusion & Future Work

We were heavily restricted by computational resources, particularly access to GPUs, storage, and training time. To determine how effective our model is for image super-resolution would require much longer training. Nonetheless, our preliminary results for SRResNet and SRGAN show an opportunity for improvement, and we are excited to see results after more extensive testing.

Future work would focus first and foremost on increased computational power, storage, and time training. Due to patient privacy issues, as well as monetary and time restrictions, we were heavily restricted in our access to GPUs. We would also research other ways to incorporate the sequential nature of MRIs: recurrent neural networks, 3D CNNs, or research into video interpolation techniques. Finally, in order to reduce our computational workload, we would begin our model with an image detection portion in order to locate the relevant scan area. This would allow us to reduce the dimensions of the image before feeding it into the GAN.

7 Contributions

Andrew and Angela contributed equally to the project, including the data preprocessing; the generator, discriminator, and overall incorporation into a model; the model evaluation; and the project and poster.

A GitHub for this project with our implementations and trained model weights can be found at `https://github.com/narcomey/mri-superresolution`.

References

We would like to thank Dr. Heike Daldrup-Link and Dr. Anuj Pareek for providing the data and offering guidance throughout the project.

[1] The Pulse on Health, Science & Tech (2018) Overcoming one of the biggest MR imaging challenges...scan time. In GE Healthcare, `http://newsroom.gehealthcare.com/overcoming-one-of-the-biggest-mr-imaging-challenges-scan-time-ecr18/`.

[2] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A.P., Tejani, A., Totz, J., Wang, Z. and Shi, W. (2017) Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *CVPR*, 2.3, p. 4. GitHub at `https://github.com/tensorlayer/srgan`.

[3] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B. & Sánchez, C.I. (2017) A survey on deep learning in medical image analysis. In *Medical image analysis*, 42, pp.60-88.

[4] Bien, N., Rajpurkar, P., Ball, R.L., Irvin, J., Park, A., Jones, E., Bereket, M., Patel, B.N., Yeom, K.W., Shpanskaya, K. and Halabi, S. (2018) Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. In *PLoS medicine*, 15(11), p.e1002699.

[5] De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D. and van den Driessche, G. (2018) Clinically applicable deep learning for diagnosis and referral in retinal disease. In *Nature medicine*, 24(9), p.1342.

[6] K. Katarzyna & Paweł Goliński (2017) Using deep learning for Single Image User Resolution. In deepsense.ai, `https://deepsense.ai/using-deep-learning-for-single-image-super-resolution/`.

[7] Dong, C., Loy, C.C., He, K. and Tang, X. (2016) Image super-resolution using deep convolutional networks. In *IEEE transactions on pattern analysis and machine intelligence*, 38(2), pp.295-307.

[8] Kim, J., Kwon Lee, J. and Mu Lee, K. (2016) Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1646-1654.