Super Resolution in Medical Imaging using SRGAN

Rewa Sood[1], Binit Topiwala[1], Karthik Choutagunta[1], Rohit Sood[2], Mirabela Rusu[3]

1. Department of Electrical Engineering, Stanford University
2. Patient Technology Solutions, Medical Imaging, Parexel Informatics
3. Department of Radiology, Stanford University

Motivation

Acquiring high-resolution (HR) magnetic resonance (MR) images usually requires the patient to remain still for long periods of time, which causes patient discomfort and increases the probability of motion induced image artifacts. A possible solution to this problem is allowing the radiologist to acquire a low-resolution (LR) image and process this image to create a HR version. One such processing method is the Super Resolution Generative Adversarial Network (SRGAN) [1]. GANs are deep learning frameworks that contain a generator network, which generates images from some input, and a discriminator, which discriminates between real and generated images. These two networks are trained adversarially, in that each network improves from the other's errors. The SRGAN was originally used with non-medical images to produce a factor of 4 upscaling. In this work, we apply the SRGAN framework to MR images of the prostate to improve the in-plane resolution for upscaling factors of both 4 and 8.

Methods

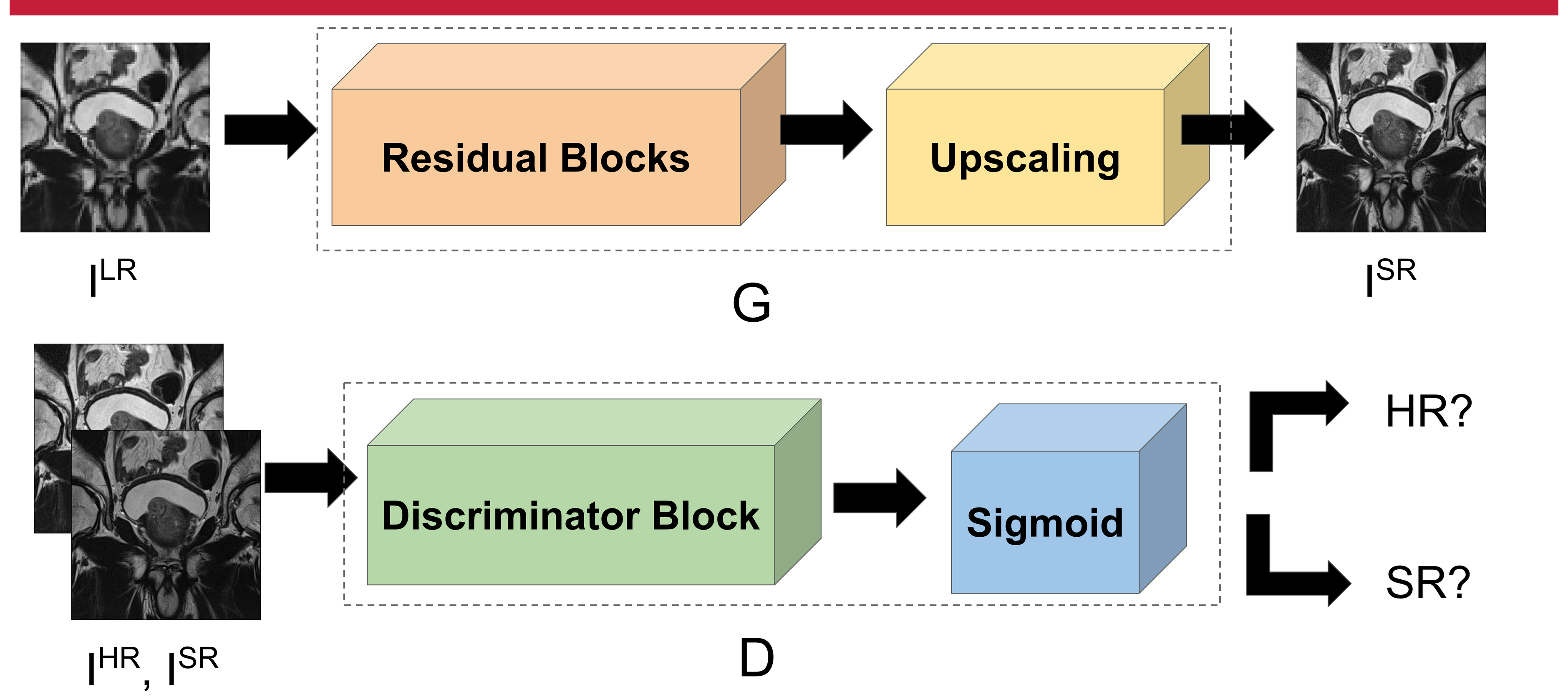

Figure 1: SRGAN Model. I^{LR}- Low Resolution image. I^{SR}- Super Resolved (SR) image. I^{HR}- High resolution Image. G- generator network. D- discriminator network.

We used the 'PROSTATEx' and 'Prostate-Diagnosis' datasets from the Cancer Imaging Archive [2], which contains over 20,000 individual MR image slices over about 450 patients. The train/test split was 80/20. The low resolution version of these images were used as input to the generator. The generator has the difficult task of creating new images while the discriminator has a simple classification task. Thus, to ensure that the discriminator does not dominate and prevent the generator from learning, we pretrain the generator component of the SRGAN for 20 epochs followed by training both the generator and discriminator networks for 50 epochs. The initial learning rate was 0.0001, which was automatically decreased to 0.00001 after 25 epochs. We performed two types of evaluation: visual and quantitative. For the quantitative evaluation, we used the PSNR (Peak Signal to Noise Ratio), SSIM (Structural SIMilarity), and MOS (Mean Opinion Score) metrics.

Results

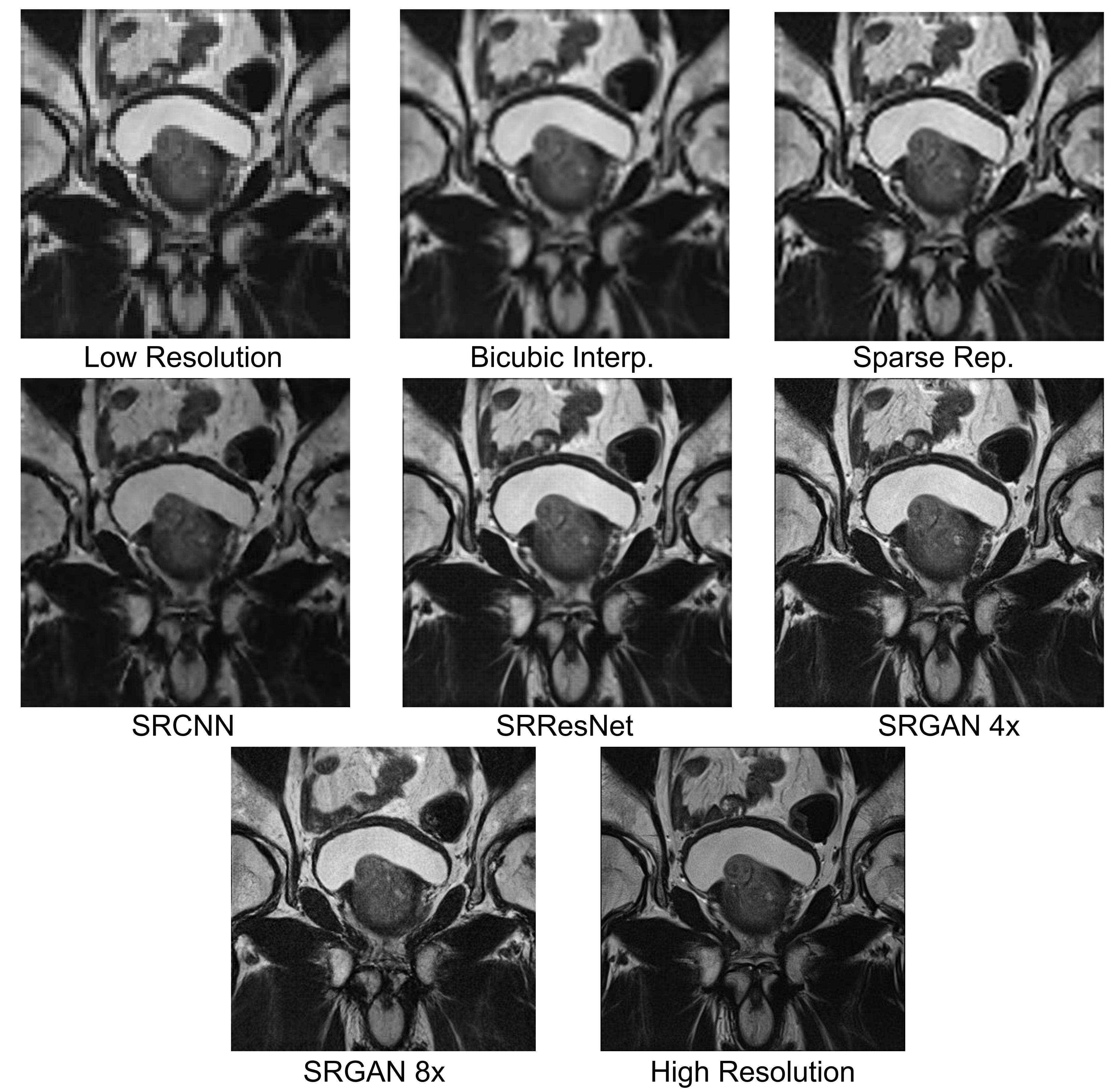

Figure 2: Comparison between SRGAN and other model results

In comparing the SRGAN to the other models, the outputs from the SRGAN model are the visually closer to the original HR ground truth images. Interestingly, this perceptual accuracy is not reflected in the PSNR (eq 2), where MAX_I is the maximum pixel value in the image, and SSIM (eq 3), where μ is the mean, σ is the standard deviation, and the constants c_1 and c_2 are used to stabilize the division.. The PSNR is proportional to the log of the MSE (e q 1) between the HR image *I* The MSE (eq1) involves the width W, height H, and scale r of the HR image and another image K. The MOS (eq 4), where R is the individual score and N is the number of samples, shows that the SRGAN has the best visual appeal overall. The MOS is the average of the scores of every instance of certain type of image in a sample set.

$$\mathrm{MSE}= \frac{1}{r^2WH}\sum_{i=0}^{rW-1}\sum_{j=0}^{rH-1}(I^{HR}(i,j)-K(i,j))^2 \quad (1)$$

$$PSNR = 10\cdot\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (2)$$

$$\mathrm{SSIM(I,K)}=\frac{(2\mu_I\mu_K+c_1)(2\sigma_{IK}+c_2)}{(\mu_I^2+\mu_K^2+c_1)(\sigma_I^2+\sigma_K^2+c_2)} \quad (3)$$

$$MOS = \frac{\sum_{n=1}^{N} R_n}{N} \quad (4)$$

Discussion

The equations below delineate the losses involved in SRGAN. The adversarial loss (eq 6) is particular to GANs and is the component that forces the generator and discriminator to learn from each other. By applying features obtained from the VGG19 network trained on Imagenet to both the HR and generated image and optimizing on the difference, the VGG loss (eq 5) emphasizes perceptual similarity. The sum of the MSE and VGG losses form the content loss, while the sum of the content loss and the scaled adversarial loss form the overall perceptual loss. The MSE loss is calculated with r set to either 4 or 8 and with K as the SR output image. In the VGG loss, Φ represents the features of a particular image. In the adversarial loss, Θ represents the weights and biases of a particular network.

$$l_{VGG}^{SR}=\frac{1}{WH}\sum_{i=0}^{W-1}\sum_{j=0}^{H-1}(\phi(I^{HR})-\phi(G(I^{LR})))^2 \quad (5)$$

$$l_G^{SR} = \sum_{n=1}^{N} -log D_{\theta_D}(G_{\theta_G}(I^{LR})) \quad (6)$$

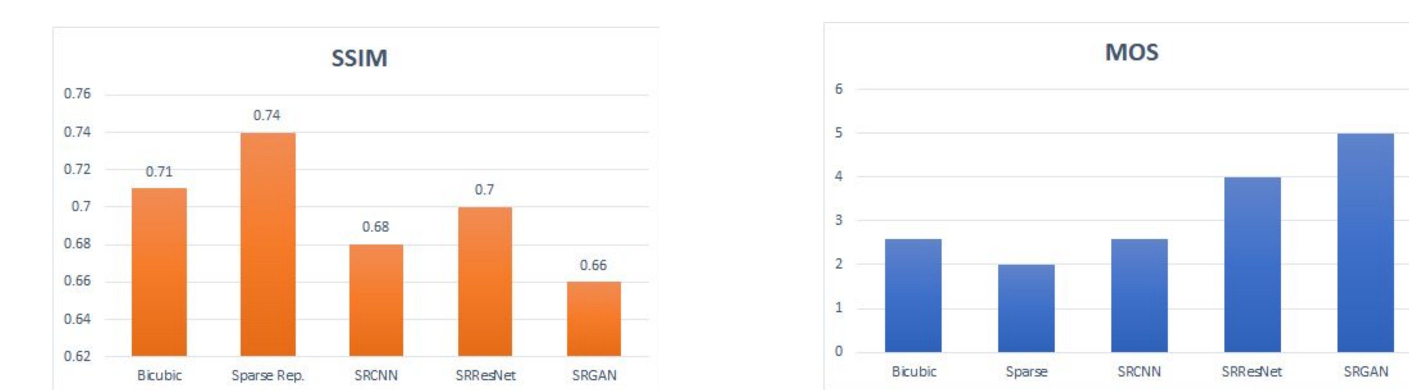

Figure 3: SSIM and MOS metrics Results

The SRGAN output images contain high frequency information that is similar to the HR ground truth images while the Sparse Representation and SRCNN models tend to smooth the images out to achieve a high PSNR. The SRCNN is especially biased toward smoothing the image because the network only has MSE loss. The SRResNet has both MSE and perceptual loss yet fails to outperform the SRGAN. Clearly, the discriminator network seeks out the high frequency information that differentiates HR and LR images, thus forcing the SRGAN output to have far more high frequency details than the output of the SRResNet.

Conclusion

The SRGAN network produces images that appear visually closer to the HR ground truth images than the images produced by the other models. The SRGAN output images contain far more high-resolution details than any of the other baseline images. This fact is reflected in the higher average MOS that the SRGAN images receive. While the other images remove the pixelation found in the LR image, they tend to blur the image, due to their emphasis on the MSE loss. By using the perceptual loss and the GAN format of training, the SRGAN does not smooth the image while still removing pixelation.

References

[1] Ledig et al. Photo-realistic single image super-resolution using a generative adversarial network. CoRR, abs/1609.04802, 2016.

[2] Clark K et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging*. 2013; 26(6): 1045-1057