Single Image Super-resolution for Chest X-ray Images

Zeynab Saeedi, Department of Mathematics & Computer Science, Amir Kabir University of Technology, Tehran, Iran, mozhan@aut.ac.ir

Abstract:

Super-resolution has been a very attractive research topic over the last three decades as images with high resolution are desirable in many applications such as medical imaging, video surveillance, astronomy etc. In medical imaging, X-ray images are widely used to diagnose a variety of diseases, also X-rays commonly have low resolution and a significant amount of noise, because radiation levels are minimized to maintain patient safety. Thus, it is necessary to increase image resolution and reduce noise. Lots of methods have been presented for achieving high quality images in SR. This paper aims to combine architecture of two SR methods (SRGAN and VDSR) to obtain higher-resolution X-ray images.

1. Introduction

Super resolution is the process of upscaling and or improving the details within an image or obtaining one or more high-resolution images from one or more low-resolution observations. Often a low-resolution image is taken as an input and the same image is upscaled to a higher resolution, which is the output. The details in the high-resolution output are filled in where the details are essentially unknown.

According to the number of LR inputs images, SR can be classified into: Single Image Super Resolution (SISR) which is more Efficient and Popular, and Multiple Image Super Resolution (MISR)

The benefits are gaining a higher quality image from one where that never existed or has been lost, this could be beneficial in many areas or even life-saving on medical applications.

In medical imaging, images are obtained for medical investigative purposes and for providing information about the anatomy, the physiologic and metabolic activities of the volume below the skin. Medical imaging is an important diagnosis instrument to determine the presence of certain diseases. Therefore, increasing the image resolution should significantly improve the diagnosis ability for corrective treatment. Furthermore, a better resolution may substantially improve automatic detection and image segmentation results.

Despite the advances in acquisition technology and the performance of optimized reconstruction algorithms over the two last decades, it is not easy to obtain an image at a desired resolution due to imaging environments, the limitations of physical imaging systems as well as quality-limiting factors such as Noise and Blur.

In radiography, when x-rays impinge on a surface such as image receptor, there is no force that causes them to be uniformly distributed over the surface. Due to this uneven distribution, one area of the receptor surface might receive more photons than the other, even when both are exposed to the same average intensity. Within every small area of the image receptor, the amount of noise is primarily

determined by the variation in photon concentration from one point to another point. There is an amount of noise produced in all imaging procedures that uses photons because of the random manner in which photons are distributed.

A solution to this problem is the use of Super Resolution (SR) techniques which can be used for processing of such images.

In the following the two related methods are explained:

1.1. VDSR

VDSR [1] is the first very deep model used in SISR and one of the classical state-of-the-art SR approaches which was published in 2016 CVPR.

As shown in Fig. 1, VDSR is a 20-layer VGG-net. The VGG architecture sets all kernel sizes as 3×3 (the kernel size is usually odd and takes the increase in the receptive field into account, and 3×3 is the smallest kernel size).

To train this deep model, the authors used a relatively high initial learning rate to accelerate convergence and used gradient clipping to prevent the annoying gradient explosion problem. VDSR takes the bicubic of LR as input. During training, VDSR uses deep CNN to learn the mapping from the bicubic to the residual between the bicubic and HR. The authors argued that residual learning could improve performance and accelerate convergence. VDSR loss function is defined as:

$$\frac{1}{2}\|r - f(x)\|^2$$

where r = y - x (y = HR image, x = LR image, f(x) = the network prediction). Thus, the network is learning the residual errors between the output and input instead of learning the HR output directly.

Anyway, VDSR improves the image quality gradually and creates blurry outputs and unsharp images.

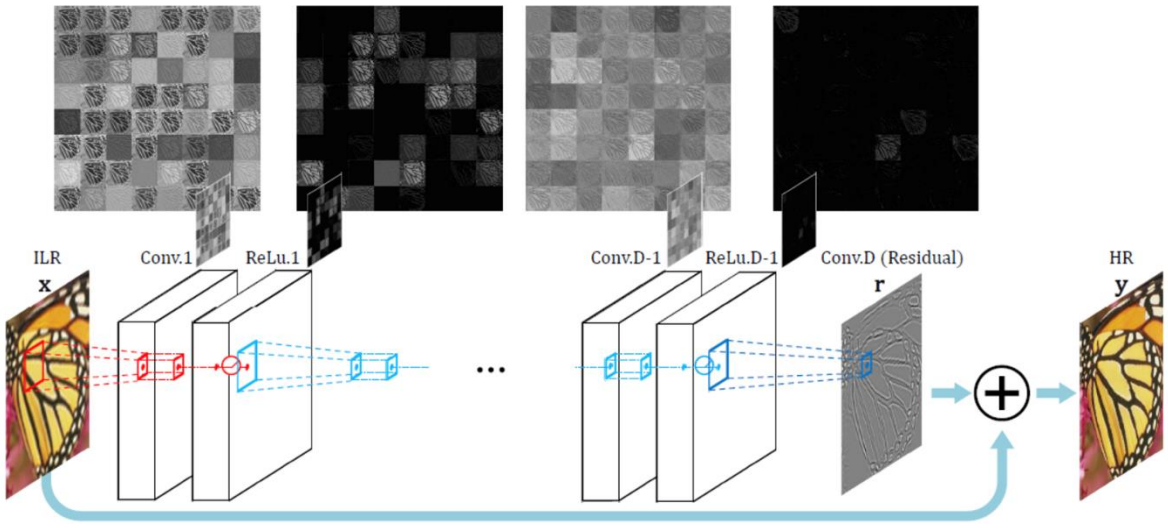

Figure 1: VDSR Network Structure

1.2. SRGAN

Super-Resolution Generative Adversarial Network [2], or SRGAN, is a generative adversarial network (GAN) that applies a deep network in combination with an adversary network to produce higher resolution images (state of art in 2017 in SISR). During the training, A high-resolution image (HR) is down sampled to a low-resolution image (LR). SRGAN generator up samples LR images to super-resolution images (SR). it uses a discriminator to distinguish the HR images and backpropagate the GAN loss to train the discriminator and the generator.

Fig. 2, shows the network design for the generator and the discriminator. It mostly composes of convolution layers, batch normalization and parameterized ReLU (PRelU). The generator also implements skip connections. The convolution layer with "k3n64s1" stands for 3x3 kernel filters outputting 64 channels with stride 1.

The loss function for the generator composes of the content loss (reconstruction loss) and the adversarial loss:

$$l^{SR} = \underbrace{\underbrace{l_{\mathrm{X}}^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}}}_{\text{perceptual loss (for VGG based content losses)}}$$

The adversarial loss is defined as:

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

SRGAN uses a perceptual loss measuring the MSE of features extracted by a VGG-19 network. Content loss extracts feature map of HR image and fake HR image from VGG-19 and compute the MSE between these two features.

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

$\phi_{i,j}$ **The feature map for the j-th convolution (after activation) before the i-th maxpooling layer.**

To train the discriminator, the loss function uses the typical GAN discriminator loss:

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{\text{train}}(I^{HR})}[\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))]$$

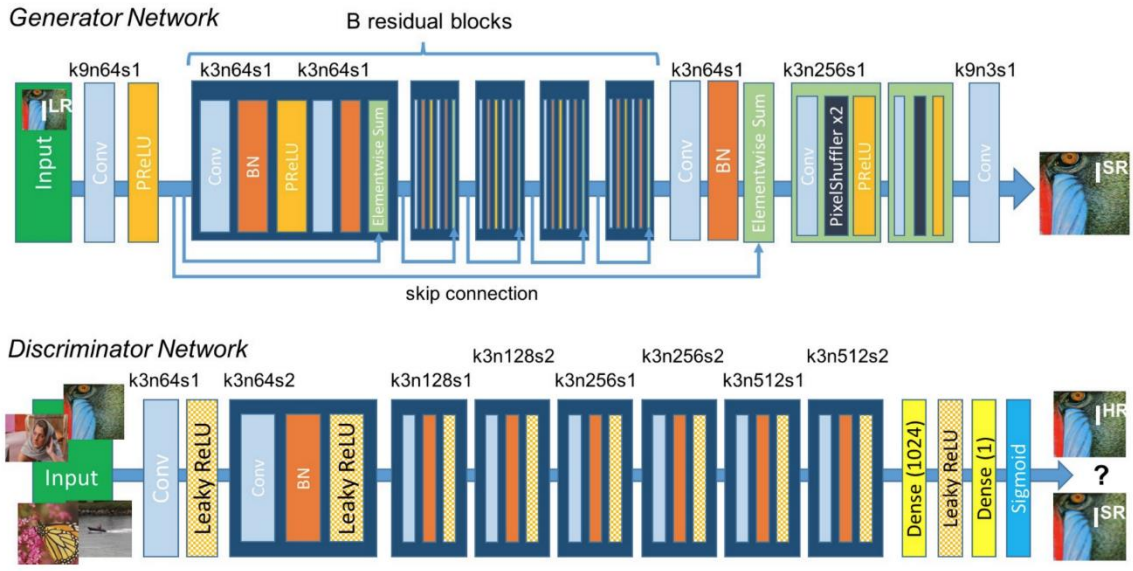

Figure 2: Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

Images generated from SRGAN have sharper and clear details than VDSR outputs, but some texture will be distorted and deformed. It's because GAN fills something else to the texture part which does not exist in the original picture (SRGAN generates sharp results but with more artifacts). Also residual blocks in SRGAN network extract low-level features in each block while VDSR network extracts high-level features from input. In this situation, maybe combining SRGAN and VDSR architecture generates more satisfiable images.

2. New Method

This new method is a Generative Adversarial Network which the discriminator network and loss function are the same as SRGAN's and its generator network is the combination of SRGAN's generator and VDSR network.

New generator network architecture follows as below:

- 1 convolution layer with 64 filters 9x9 and PRelu activation function
- 8 residual blocks including 2 convolution layers with 64 filters 3x3, PRelu activation function and Batch normalization layer
- 20 convolution layers with 64 filters 3x3 and PRelu activation function (VDSR Network)
- 1 Batch normalization layer
- 1 residual layer (Elementwise sum)
- 1 convolution layer with 3 filter 3x3 and Tanh activation function (non-trainable sharpening layer)

- 1 Up-sampling block
- 1 convolution layer with 16 filters 3x3 and PRelu activation function
- 1 convolution layer with 3 filters 5x5 and Tanh activation function

3. Dataset

National Institutes of Health Chest X-Ray Dataset [3] is comprised of 112,120 X-ray images with disease labels from 30,805 unique patients. These images are grayscale and in PNG format with size 1024 x 1024. In this experiment, a random sample of 500 images of this Dataset has been used (300 images for training and 200 for testing).

3.1 Data Preparation

Before using images for training and testing, they'll be resized to 256 x 256. Low-resolution images for each network will be obtained as follows:

For VDSR, images will be down-sampled using bicubic interpolation with down-sampling factor r=4, and up-sampled using bicubic interpolation with up-sampling factor r=4.

For SRGAN, images will be down-sampled using bicubic interpolation with down-sampling factor r=4.

For new method, images will be down-sampled using bicubic interpolation with down-sampling factor r=4, and up-sampled using bicubic interpolation with up-sampling factor r=2.

4. Results

all three methods have been trained in 6 epochs and batch size = 4 and results are shown in the table below:

Table 1: Comparing PSNR (Peak Signal-to-Noise Ratio), UQI (Universal Quality Image Index), SSIM (The Structural Similarity Index) and MS-SSIM (Multiscale Structural Similarity Index) of HR and SR outputs from VDSR, SRGAN and new method.

Model/Evaluation	PSNR	UQI	SSIM	MS-SSIM
VDSR	28.80188	0.986009	0.912259	0.972789
SRGAN	27.24132	0.956506	0.801246	0.953419
New Method	28.43138	0.957970	0.828199	0.961132

Visual results:

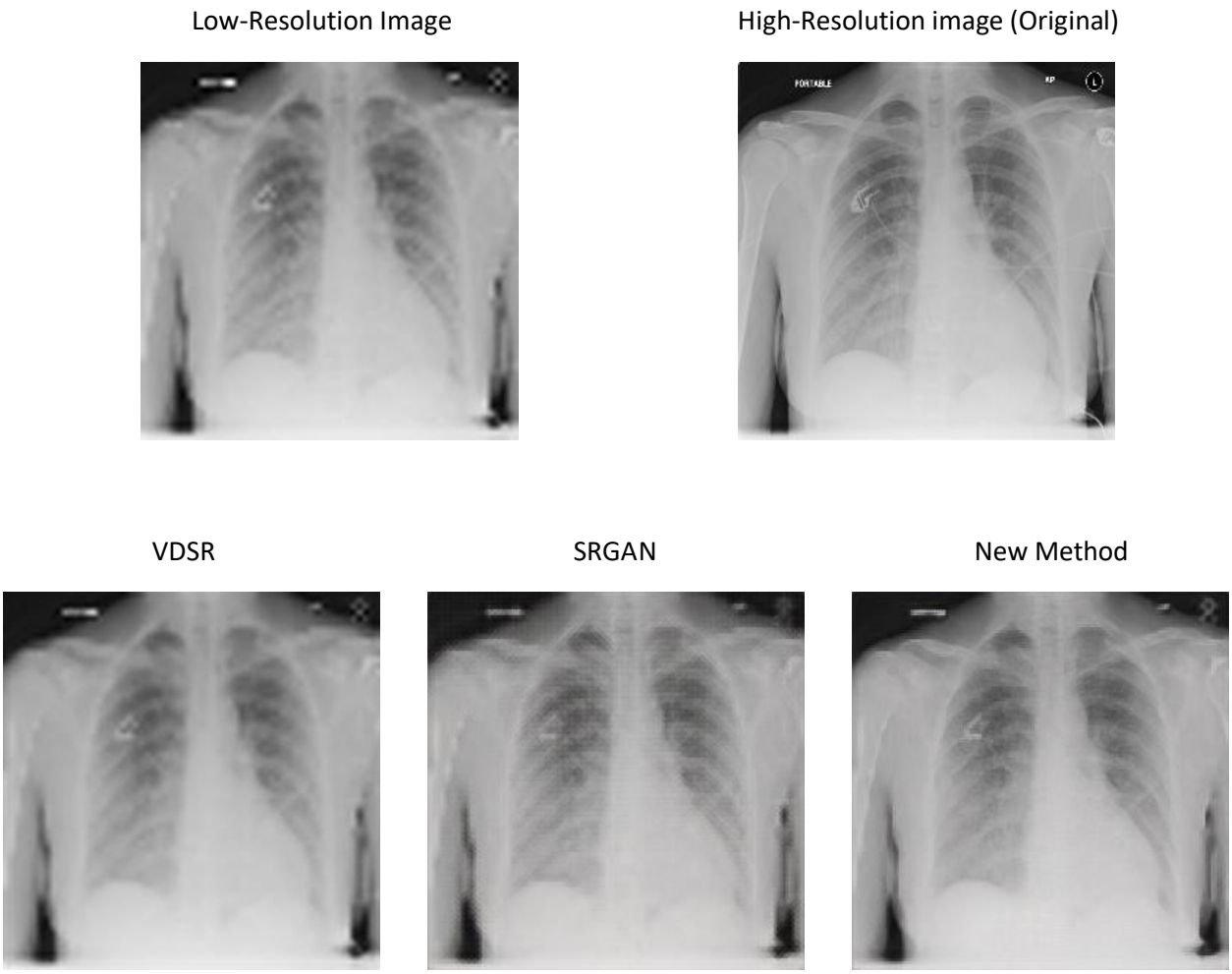

Figure 3: From left to right: VDSR, SRGAN and New Method outputs of corresponding LR input.

References

[1] J. Kim, J. K. Lee and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 1646-1654, doi: 10.1109/CVPR.2016.182.

[2] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 105-114, doi: 10.1109/CVPR.2017.19.

[3] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri and R. M. Summers, "ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 3462-3471, doi: 10.1109/CVPR.2017.369. (Original source files and documents: https://nihcc.app.box.com/v/ChestXray-NIHCC/folder/36938765345)

[4] W. Yang, X. Zhang, Y. Tian, W. Wang, J. Xue and Q. Liao, "Deep Learning for Single Image Super-Resolution: A Brief Review," in IEEE Transactions on Multimedia, vol. 21, no. 12, pp. 3106-3121, Dec. 2019, doi: 10.1109/TMM.2019.2919431.

[5] M. Shimizu, H. Kariya, T. Goto, S. Hirano and M. Sakurai, "Super-resolution for X-ray images," 2015 IEEE 4th Global Conference on Consumer Electronics (GCCE), Osaka, 2015, pp. 246-247, doi: 10.1109/GCCE.2015.7398662.

[6] J. S. Isaac and R. Kulkarni, "Super resolution techniques for medical image processing," 2015 International Conference on Technologies for Sustainable Development (ICTSD), Mumbai, 2015, pp. 1-6, doi: 10.1109/ICTSD.2015.7095900.

[7] Fitzgerald R (2000) Phase-sensitive x-ray imaging. Phys Today 53(7): 23-26