Super-Resolution of Degraded Low-Resolution X-Ray Images Using a Novel Patch-Based Generative Adversarial Network

Aria Vikram

under the direction of

Prof. Yueh Lee
Associate Professor of Radiology
University of North Carolina at Chapel Hill

Research Science Institute
August 2, 2021

Abstract

As X-rays become increasingly important, having low-dose, high resolution X-ray detectors could help reduce the cost of X-rays and minimize radiation exposure while maintaining image detail. In this work, we propose a patch-based generative adversarial network for degraded low-resolution images. The generated X-ray images showed a structural similarity of 0.85 to the source X-ray images. It was observed that our model outperforms interpolation techniques in reconstructing image detail and shows promise in recovering edges in degraded images. It also allows for memory efficient computation without sacrificing image fidelity. The network was trained and tested for X-ray images which were smoothed and downsampled by a factor of 4, corresponding to a 16x reduction in radiation dose. This could allow for the reconstruction of high resolution X-rays from lower dose X-ray images.

Summary

High resolution X-rays generally require greater radiation dosages compared to lower resolution X-rays. However, lower resolution low-dose X-rays often lack the pathological detail and clarity required for diagnosis and examination. Therefore, we need a way to convert these lower dose but degraded low resolution images to higher resolutions. As interpolation techniques cannot address this problem adequately, we propose a generative adversarial network to recover lost image detail from X-ray images of lower resolution and reconstruct high-fidelity super resolution X-ray images. This network outperforms existing techniques in recovering fine-grained high frequency features, while providing for memory efficient computation. It is also robust and can be used in a multitude of practical situations as it works for images which are not only downsampled but also further obfuscated at varying levels. This super resolution tool for X-rays could help us move towards a future where imaging is ubiquitous and cost-effective while significantly reducing the radiation dosage per X-ray.

1 Introduction

In the age of COVID-19, X-rays have become increasingly important as diagnostic tools. Traditionally, a radiation dose of 0.1 mSv is needed to generate a chest X-ray image with sufficient detail to examine and diagnose pathologies [1]. However, the radiation dose required to construct these high-resolution (HR) X-rays is much larger than that of their lower resolution counterparts.

X-ray dose falls as pixel size increases. When we decrease the resolution of each pixel by a factor of n, the number of pixels increases by a factor of n^2. Each of the new pixels needs the same number of X-ray photons in order to maintain the same signal to noise ratio, causing radiation dose to increase by a factor of n^2 as well. Therefore low-resolution (LR) X-rays require less radiation dosage and are more economically viable than their HR counterparts [2]. Lower resolution X-rays also take shorter times to read-out, improving the diagnostic workflow [3]. However, the lower the dose, the fewer the number of pixels and consequently lower the image detail captured.

Capturing LR X-ray images and then converting these to HR images could help reduce the cost of X-rays and minimize radiation exposure while retaining X-ray image detail.

Traditional upsampling interpolation methods such as bilinear and bicubic interpolation are unable to reconstruct high-resolution images from LR images due to insufficient detail in these LR images. Generative Adversarial Networks (GANs) have grown in prominence as they provide a way to recover texture detail which may not exist in the low-resolution images. The super-resolution GAN (SRGAN) proposed by Ledig et al.[4] is one such GAN which is targeted towards creating super-resolution photos. Although super-resolution (SR) images produced using this approach are similar to the original HR images, the results of the SRGAN paper show that the generated images suffer from the introduction of reconstruction artifacts. An approach which tries to address the problem of super-resolution in X-rays

is the SNSRGAN [5] which uses auxiliary label information - X-ray disease information - to constrain feature generation. However, label information is costly and often not known beforehand. Therefore, this approach cannot address SR tasks without prior disease label information, rendering it infeasible in practical settings, as highlighted by the authors of the paper.

Many current super-resolution GAN approaches resize both the HR and LR images. This is done as the original high-resolution images are of a large size and often exceed GPU memory limits during computation. However, the act of downsampling itself loses considerable information from the original HR image. Moreover, existing SR techniques for imaging only support downsampled images. They do not address the problem of LR x-ray images which may have varying levels of degradation (i.e where images are oversmoothed without sufficient texture detail).

Therefore, existing approaches focus only on super-resolution whereas we have the dual goal of

1. de-smoothing: recovering lost edges and reconstructing image detail and
2. increasing resolution.

The main contributions are summarized as follows:

1. A heavier emphasis is placed on reconstructing fine-grained features and preventing artifacts as the loss is based on individual patches instead of a full resized image. This approach has been used for segmentation of images with promising results [6].

2. The patch-based local approach sidesteps the need for resizing images and thus preserves image fidelity while allowing for memory efficient computation. It also allows for the usage of the model on any image regardless of the size.

3. The developed model recovers lost image detail from blurred images in addition to increasing the resolution.

4. The training of the model with different filter strengths to simulate varying levels of smoothing allows for a more robust model which can be translated to practical settings.

2 Methods

Assume that I^{LR} is a low-resolution X-ray image of size $a \times b$ and I^{HR} is a high-resolution X-ray image of size $c \times d$. Their relationship can be expressed as:

$$I^{LR} = f(I^{HR}) \tag{1}$$

where f denotes the moving average filter and downsampling process that creates an LR counterpart from an HR image.

The goal of the SR process is to implement an approximate inverse function g $\approx f^{-1}$. This problem is ill-posed as there is insufficient detail in the degraded LR image to recreate the original HR image. Therefore, we make use of priors to obtain an estimation for f^{-1} .

Image Processing

For training and inference on ultra-high-resolution images with limited GPU memory, existing techniques employ resizing of original images to a smaller size. To avoid this, we take patches of fixed sizes from the original image and pass each patch through the model individually. Finally, we stitch the generated SR patches back together to create a composite image. Any pixelation at the seams (where the patches are joined together) can be reduced by taking overlapping windows such as those proposed by Pielawski et al. for image segmentation [7].

Network Architecture

In our approach, which we have called XPGAN (X-ray patch GAN), we train two feed-forward convolutional neural networks(CNN) : the first network, the generator, has the goal of fooling the other network, the discriminator. A GAN is used as it helps in the generation of more realistic images and encourages perceptual similarity to actual high-resolution images[4]. The network flow is depicted in Figure 1.

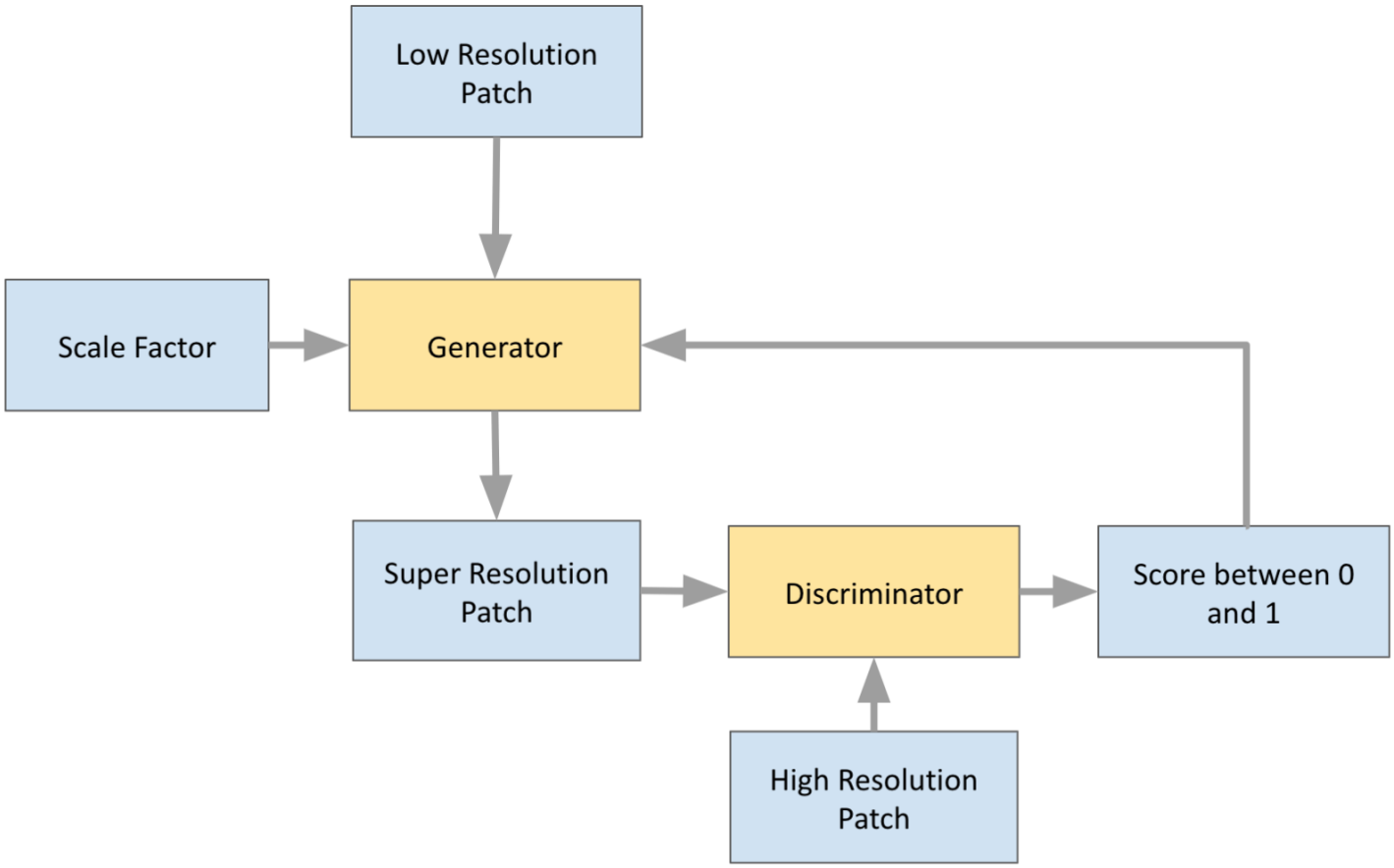


Figure 1: Network flow: The generator takes an LR patch and its upsampling factor as the inputs and outputs an SR patch. The discriminator is called with an HR patch as its input and it tries to predict whether this HR patch is a real HR patch or a generated SR patch. Similarly the discriminator is also called with an SR patch as its input and it tries to predict whether this SR patch is a real HR patch or a generated SR patch.

Our generator network G takes in two inputs, the LR image and the upsampling factor. It utilizes residual blocks to concatenate feature maps in the previous blocks to the feature maps in the later blocks. Additionally, the high level features extracted by the residual blocks are combined with the information from the original image through a skip connection crossing over the residual blocks. These skip connections improve training and keep the generated images structurally similar to the input LR images [8]. Each residual block is composed of two convolutional layers with 3×3 kernels and 64 feature maps followed by batch-normalization layers. ParametricReLU(PReLU) is used as the activation function for each residual block as it speeds up training and prevents the vanishing gradient problem[9]. The resolution of the input image is increased using an upsampling block with trained sub-pixel convolution layers[10]. The generator structure is depicted in Figure 2.

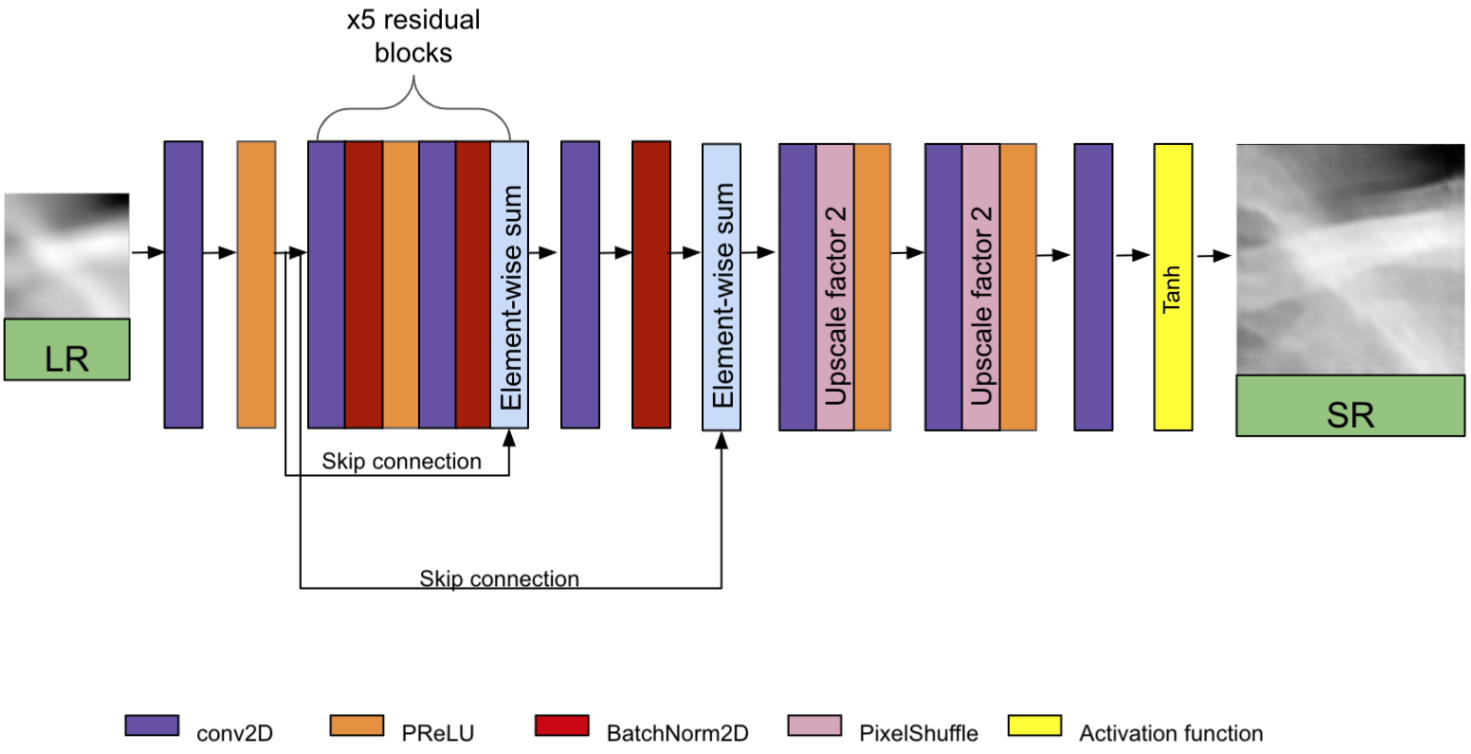


Figure 2: Generator architecture

Our discriminator network D contains convolutional layers with an increasing number of filter kernels which in turn increases the number of feature maps. Strided convolutions are used to decrease image resolution as we go deeper down the network. An average pooling layer transforms the last feature map to a 1 x 1 feature map. This helps in dimensionality reduction: reducing the number of feature maps while retaining the important features. The final activation function is a sigmoid to obtain a probability for whether the input image is a generated SR image or a real HR image. The discriminator structure is depicted in Figure 3.

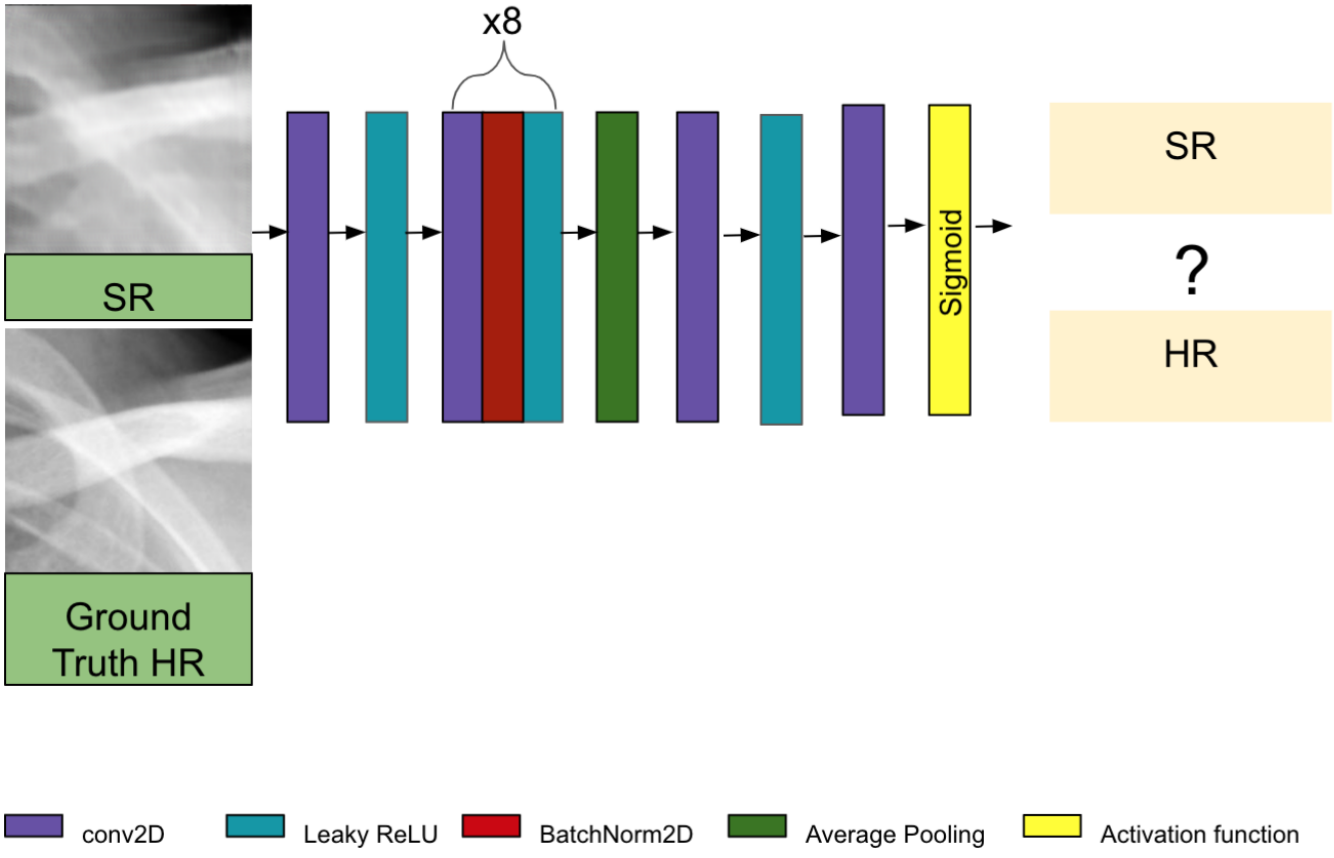


Figure 3: Discriminator architecture

Network Losses

In this section, we take I^{HR} as the high-resolution patch and I^{LR} as the corresponding low-resolution patch. I^{SR} denotes the super-resolution generated patch. D is the discriminator with the parameter θ_d and G is the generator with the parameters denoted by θ_g.
For our generator loss function, we combine four different losses.

Adversarial Loss: We define the generator's adversarial loss function as

$$L_{adv} = \sum_{n=1}^{N} -logD_{\theta_d}(G_{\theta_g}(I^{LR})) \tag{2}$$

where N is the total number of HR-LR training pairs. This encourages the generator network to favor solutions that fool the discriminator.

Mean Squared Error Loss: We also use pixel-wise mean squared error (MSE) loss which can be represented as

$$L_{mse} = \frac{1}{r^2WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_g}(I_{x,y}^{LR}))^2 \tag{3}$$

in which we represent I^{LR} by a tensor of size W x H x C and I^{HR} by a tensor of size rW x rH x C where r is the upscaling factor [12].

Perceptual Loss: In order to capture the high level features, we use the VGG loss derived from the pretrained VGG19 model originally introduced in the SRGAN paper. This can be represented by

$$L_{vgg/i,j} = \frac{1}{r^2W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}I^{HR} - \phi_{i,j}(G_{\theta_g}(I^{LR})))^2 \tag{4}$$

where ϕ indicates the feature map obtained by the jth convolution before the ith maxpooling in our pretrained VGG19 network. $W_{i,j}$ and $H_{i,j}$ represent the dimensions of the respective feature maps within the fine-tuned VGG network.

Gradient Loss: We use the gradient loss as the regularization term to penalize blurring

and preserve edges [11]. It can be defined as

$$L_{gra} = \sum_{i=1}^{W-1}\sum_{j=1}^{H-1}(G_{\theta_g}(I^{LR})^{i+1,j} - G_{\theta_g}(I^{LR})^{i,j})^2 + \sum_{i=1}^{W-1}\sum_{j=1}^{H-1}(G_{\theta_g}(I^{LR})^{i,j+1} - G_{\theta_g}(I^{LR})^{i,j})^2 \quad (5)$$

Generator Loss: Combining the above terms, we can formulate the loss function of the generator as

$$L_{gen} = \alpha(L_{adv}) + \beta(L_{mse}) + \gamma(L_{vgg/i,j}) + \delta(L_{gra}) \quad (6)$$

where α, β, γ and δ are the weights for the loss terms.

Discriminator Loss

The discriminator attempts to assign a score of 1 for the real HR images and a score of 0 for the generated SR images. Therefore, the loss function of the discriminator can be defined as:

$$L_{disc} = 1 - \frac{1}{N}\sum_{n=1}^{N} D_{\theta_d}(I^{HR}) + \frac{1}{N}\sum_{n=1}^{N} D_{\theta_d}(I^{SR}) \quad (7)$$

where N is the total number of HR-LR training pairs.

Experiments

In order to train and test our network, XPGAN, we use images from the Chest X-ray 2016 dataset [12] which consists of 5,863 X-ray images characterized as pneumonia and normal. Instead of the traditional resizing approach, we break up the training dataset into patches of 256×256. This size was chosen according to GPU memory constraints and ease of computation. Images are broken up into patches using random offsets for the stride length. We randomly divided the dataset into 30,000 patches for training, 600 for validation and 3000 patches for testing. Before taking patches, the original images are zero-padded so that equal sized patches can be taken. Each HR 256×256 patch is downsampled using a bicubic kernel with downsampling factor $r = 4$ and further smoothed by a simple moving average (SMA) filter with a randomly chosen kernel size of k where $k \in [1, 40]$ to obtain the 64×64

LR patches. A 40×40 filter was chosen as the maximum value of degradation as we found it to be a high level of smoothing which does not completely obfuscate the image.

The generator is called with the scale factor (chosen as 4 as this is a sufficiently large downsampling factor) and the low-resolution input patch of 64×64. The Adam optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.5$) is used to update the generator and discriminator. Weights for the generator loss terms were set according to the results of cross-validation. Training continued for 200 epochs with a mini-batch size of 16 and an initial learning rate of 2×10^{-4}. The image processing pipeline during training is shown in Figure 4. The link to the implementation code is given in Appendix A.

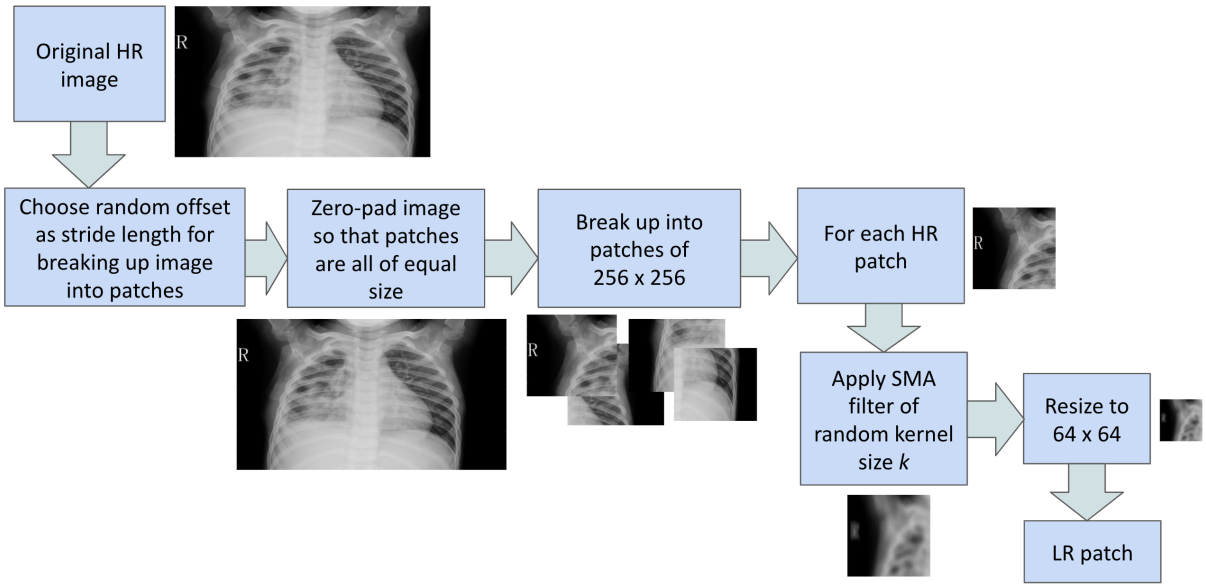


Figure 4: Image Preprocessing Pipeline Before Training

For testing the model, we used the previously mentioned dataset of 3000 patches. The images used for testing are downsampled by 4 to give 64×64 sized patches which are also smoothed using SMA filters. The testing dataset is converted into 4 datasets, each with the same 3000 patches, with 10×10, 20×20, 30×30 and 40×40 filters respectively applied on the patches of each dataset. This is done to observe the progression of the structural similarity and blurriness in SR and interpolated images as the level of degradation in the LR images

is increased. We also test with an additional dataset derived from the same testing dataset where a random filter k is applied on each image with $k \in [1, 40]$. The generator is called with an upsampling factor of 4. To evaluate the images, we use the structural similarity index measure (SSIM). The SSIM measure quantifies the perceptual similarity of the generated SR images to the real HR images[13]. We also use the variance of the Laplacian to detect the amount of blur in the images. The higher this measure, the more well-defined the edges in the image. This is a useful measure for us as texture structure and sharpness is important in medical image evaluation [14]. We compare the SSIM and variance of Laplacian measures of XPGAN-generated SR images with upsampled images produced by bicubic interpolation. The quantitative image analysis was done using OpenCV and scikit-learn. An illustrated pipeline of the processing done on the LR image during inference is given in Appendix B.

3 Results

We compare the performance of XPGAN with that of bicubic interpolation on the basis of the SSIM metric. Table 1 shows this comparison on the aforementioned testing datasets.

	XPGAN	Bicubic Interpolation
Randomly chosen $k \in [1, 40]$	0.85	0.81
$k = 40$	0.80	0.72
$k = 30$	0.83	0.74
$k = 20$	0.85	0.79
$k = 10$	0.89	0.86

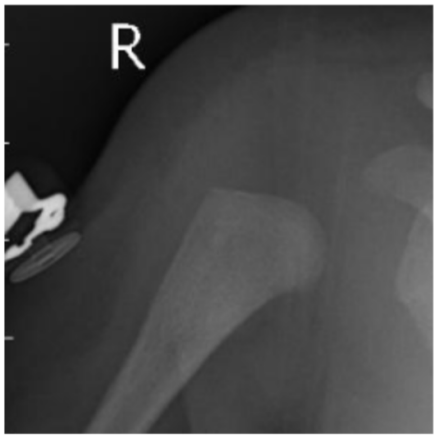
Table 1: SSIM Comparison

Table 2 compares the variance of the Laplacian between XPGAN-generated images and interpolated images. The average variance of the Laplacian of the original HR images is 144.21.

	XPGAN	Bicubic Interpolation
Randomly chosen $k \in [1, 40]$	30.58	1.72
$k = 40$	14.65	0.94
$k = 30$	23.85	0.99
$k = 20$	28.38	1.10
$k = 10$	31.08	1.69

Table 2: Variance of the Laplacian Comparison

The image patches shown below serve as a visual comparison of interpolated images with XPGAN-generated images. Figure 5 represents the original HR image with which we compare the interpolated and GAN-generated images in Figure 6 and Figure 7.



Real HR

Figure 5: Original HR image

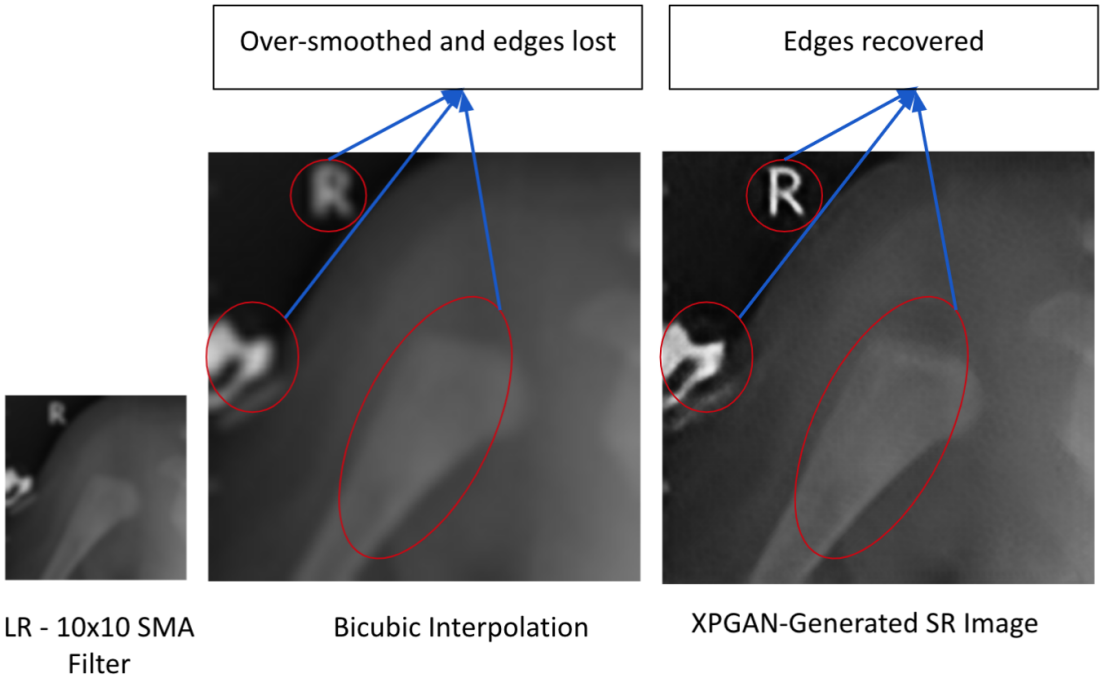


Figure 6: Results on sample test image with 10×10 SMA filter applied on LR images

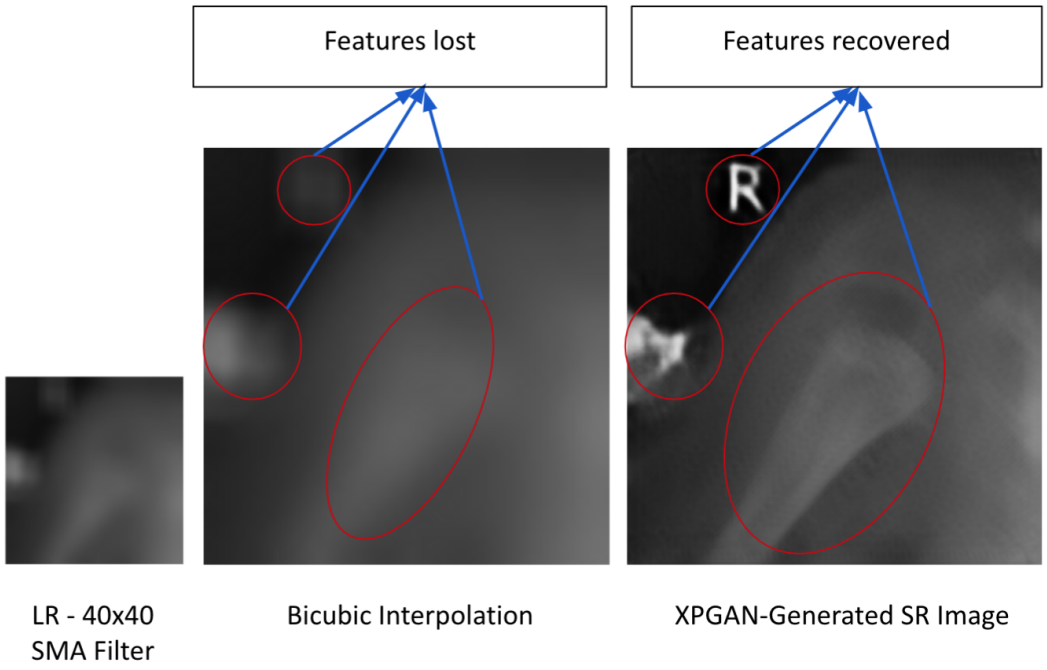


Figure 7: Results on sample test image with 40×40 SMA filter applied on LR images

4 Discussion

The quantitative results demonstrate that XPGAN generates more competitive SR images compared with existing interpolation methods.

The higher SSIM values of XPGAN-generated images suggests that these images are more structurally similar to the actual HR images than interpolated upsampled images.

The variance of the Laplacian measure shows that XPGAN consistently produces images with sharper edges as compared to existing interpolation methods.

We can also see from the progression of SSIM values that as the kernel size of the filter increases, the SSIM values of interpolated images decrease more sharply. This is because interpolation fails as images become more obfuscated. Further, when LR images are less degraded (smaller kernel size of filter), interpolated images retain the overall structures but lose edges and have over-smooth textures. When LR images are more degraded (larger kernel size of filter), interpolated images lose the overall structures in addition to becoming more blurry. XPGAN images on the other hand, try to reconstruct detail based on the prior learned distribution and are hence able to recover image detail and edges even when the LR images passed to the XPGAN model are more degraded. This is further elucidated by the example images shown in Figure 6 and 7.

Thus, we can infer from the results that XPGAN outperforms interpolation methods and is particularly useful in recovering fine-grained details and edges even for heavily degraded images. The examination of individual patches during training and inference instead of a resized full image preserves the fidelity of the original images and prevents the introduction of artifacts as indicated by the SSIM values. As the patches are smaller than the original images, this model allows for memory efficient computation of SR images.

The high SSIM and variance of Laplacian values for the dataset with random filters indicates that the model is robust and can be used for images with varying levels of degradation.

All experiments are performed with a scale factor of 4× between LR and HR images. This corresponds to a 16× reduction in image pixels and consequently a 16× reduction in radiation dose.

5 Future Work

One limitation of this study is the lack of a subjective assessment of produced SR results to ascertain whether the images produced are acceptable to radiologists. In order to analyze the SR images qualitatively, we plan to conduct a Mean Opinion Score (MOS) test with radiologists to check for acceptability of the SR images.

We additionally envision a place for this model in the testing of novel X-ray sources. In order to test these X-ray sources, we use CT data to simulate the phantoms. This CT data is traditionally $1 \times 1 \times 1$ mm. X-ray detectors on the other hand capture a resolution of 100×100 microns. When we try to test novel X-rays sources on phantoms constructed from lower resolution CT data, the projections obtained from these simulated phantoms are of lower resolution and degraded (i.e projected images are over-smoothed without sharp edges and lack texture detail). Our method can be applied to generate realistic HR images from these degraded LR images.

We can also apply this model for conversion of CT scans to X-rays. This is because the pixel resolution of a CT is higher than pixel resolution in an X-ray [16]. So translation of a CT to an X-ray results in blurry low-resolution X-rays which lack detail.

6 Conclusion

This paper has presented a super-resolution scheme for recovery of degraded low-resolution X-ray images. To attack the potential risk of losing information by resizing images due to

GPU constraints, we introduce a patch-based GAN method. This also helps place a heavier emphasis on recovering the fine-grained details while penalizing artifacts. This method allows for memory efficient computation during training and inference as it takes in smaller patches as inputs. Training and testing the model with different filtration strengths helps make it more robust and feasible for application in practical situations. Quantitative measures also support the fact that images produced using XPGAN are sharper and more perceptually similar to the real HR X-ray images than those produced by existing state of the art interpolation techniques. The proposed approach - XPGAN- has produced promising results in recovering edges and reconstructing image detail for degraded low-resolution X-ray images.

7 Acknowledgments

I would like to express my gratitude to my mentor, Prof. Yueh Lee, whose advice and mentorship throughout the last two months have played a pivotal role in the development of this project. I would also like to thank Dr. Pew-Thian Yap and Dr. Marc Niethammer for their guidance on the deep learning components of the project. I would like to thank Dr. Jenny Sendova, Anne Liang and Gabrielle Kaili-May Liu for their invaluable feedback on the project. I am also grateful to my family for their support and encouragement throughout this process. I also thank my sponsors, Mr. Nicholas Nash and Ms. Phalgun Raju, for their generosity in making this research possible. Finally, I would like to thank CEE, MIT and the entire RSI community for giving me this opportunity.

References

[1] R. a. ACR. Radiation dose in x-ray and ct exams, Apr 2020.

[2] 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)*, 2015.

[3] T. Dreier, U. Lundström, and M. Bech. Super-resolution x-ray imaging with hybrid pixel detectors using electromagnetic source stepping. *Journal of Instrumentation*, 15(03), 2020.

[4] L. et al. Photo-realistic single image super-resolution using a generative adversarial network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[5] L. Xu, X. Zeng, Z. Huang, W. Li, and H. Zhang. Low-dose chest x-ray image super-resolution using generative adversarial nets with spectral normalization. *Biomedical Signal Processing and Control*, 55:101600, 2020.

[6] W. Chen, Z. Jiang, Z. Wang, K. Cui, and X. Qian. Collaborative global-local networks for memory-efficient segmentation of ultra-high resolution images. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[7] N. Pielawski and C. Wählby. Introducing hann windows for reducing edge-effects in patch-based image segmentation. *Plos One*, 15(3), 2020.

[8] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. *Computer Vision – ECCV 2016 Lecture Notes in Computer Science*, page 630–645, 2016.

[9] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015.

[10] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[11] Q. Lyu, H. Shan, and G. Wang. Mri super-resolution with ensemble learning and complementary priors. *IEEE Transactions on Computational Imaging*, 6:615–624, 2020.

[12] D. Kermany. Labeled optical coherence tomography (oct) and chest x-ray images for classification, Jan 2018.

[13] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[14] A. Ahmadvand and M. R. Daliri. A review on texture analysis methods in biomedical image processing. *OMICS Journal of Radiology*, 05(02), 2016.

Appendix A Link to Code

Link to code: https://github.com/Caffetaria/XPGAN

Appendix B Inference pipeline for LR image

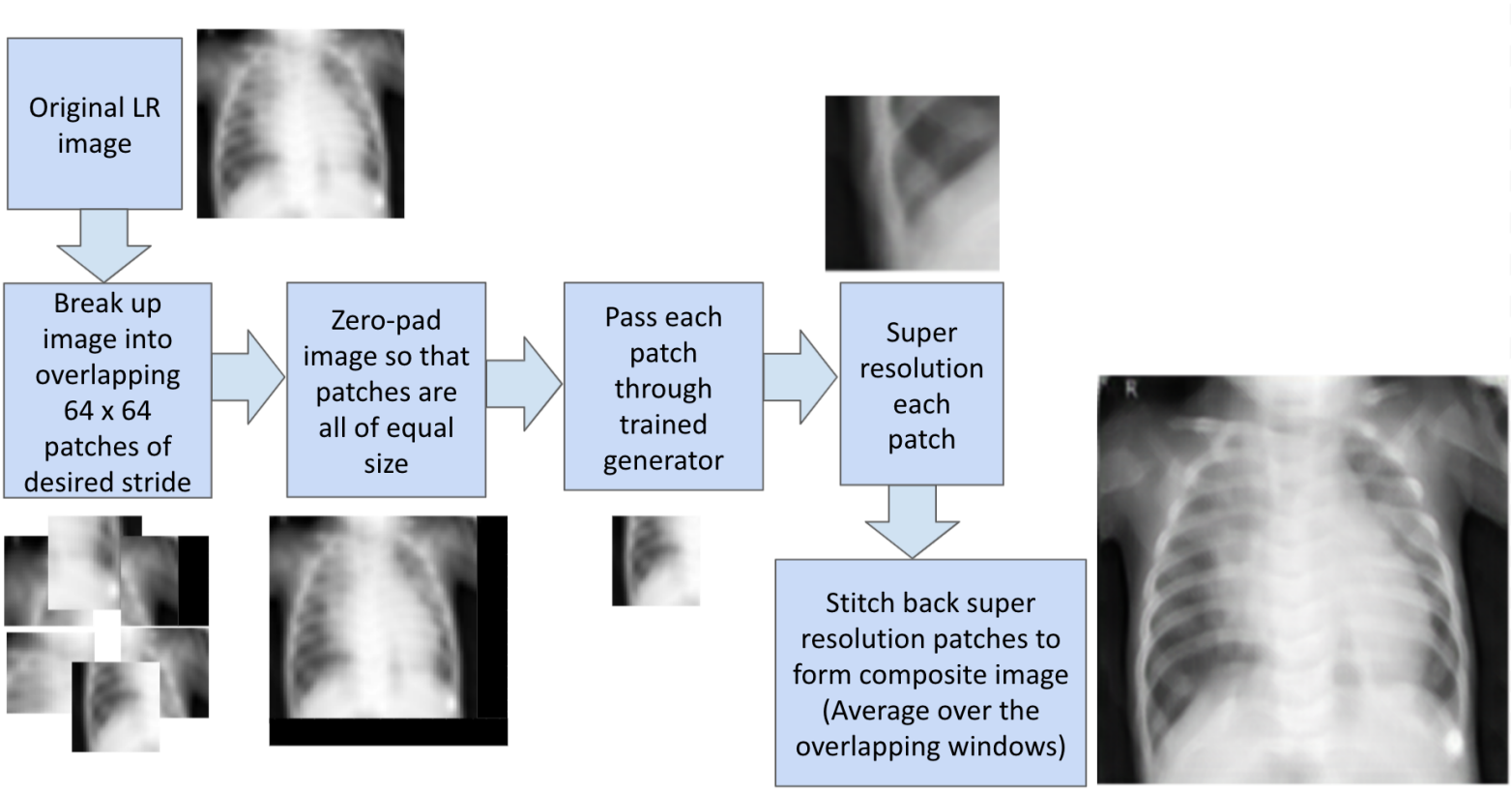


Figure 8: Inference pipeline for LR image