SYDE 672 Report

mkmmcleod

December 2018

Contents

Abstract

Two methods for super resolution detailed in the papers "Zero-Shot Super Resolution using Deep Internal Learning" (ZSSR) and "Deep Image Prior" (DIP) are summarized and then compared against one another. This comparison raises the question of what is the driving force behind the efficacy of the ZSSR method. The ZSSR method uses a convolutional neural network (CNN) to learn the internal image statistics to perform super resolution. The DIP paper, however, shows that the inherent structure of a CNN alone is enough to obtain good super resolution performance. To answer the question of whether the performance gains of the ZSSR method are due to the learning of internal image statistics or the use of a CNN, a set of images chosen to tease apart this question are analyzed. It is shown that the ZSSR method is learning internal image statistics which result in distinct high resolution images as compared to the DIP method.

1 Introduction

Super resolution is a set of techniques that attempt to accurately increase the resolution of an image. Enhancing the resolution of an image is a challenging problem with many commercial and scientific applications. For example, the resolution of photos taken on mobile phones are limited by the hardware yet many displays such as TVs are of much higher resolution. Another example is the pinch zooming on the smart phone itself. Google recently published a paper outlining their method of using super resolution to enhance the zooming on their Pixel 2 smartphone [1].

2 Background

The following explains relevant information for the report. It is assumed the reader has some basic knowledge about image processing and deep learning. More complex ideas directly related to super resolution are explained in the following subsections.

2.1 The challenge of super resolution

Single image super resolution produces a suitable high resolution image from a low resolution image. For a given low resolution image, there are many possible high resolution image counterparts. This fails the uniqueness property described in the textbook "Statistical Image Processing and Multidimensional Modelling" written by Dr. Fieguth [2]. This means that the problem of super resolution is ill-posed and some type of prior is required to complete the task.

A simple and common approach is to assert a smoothing prior as found in interpolation methods or smoothing methods [3]. Interpolation methods, such as bicubic interpolation, work by taking a weighted average of the local pixels in the lower resolution image and assigning it to the higher resolution image. However, as this prior encourages smooth gradients between pixels in the high resolution image, it leaves the image blurry. This is an undesirable effect of super resolution as clear precise edges are desired for the higher resolution image.

2.2 What is zero-shot single image super resolution

Single image super resolution means that only a single image is being upscaled. The alternative to this is video where the images in a video can be used together to create an effective higher resolution video.

"Zero-shot" is a term borrowed from the pattern recognition/classification domain which means there are no other instances of the same class in the training data [4]. In the context of super resolution, it means that the entire process to create the higher resolution image from the low resolution target image only requires the image itself. Examples of zero-shot super resolution methods are the interpolation methods such as bicubic interpolation described previously in section "The challenge of super resolution".

2.3 State of the Art Deep Learning Methods

The current state of the art methods for super resolution use deep learning based primarily on CNN [5],[6]. The CNN is trained on a large corpus of data that consists of low resolution images and their corresponding high resolution counterpart. The CNN is trained over the dataset on how to transform the low resolution image into the high resolution image. The success of deep learning, and specifically CNNs in super resolution is commonly attributed to the network learning complex priors that best create the high resolution images [5] [6]. The learned priors are stored implicitly in the weights and connections of the neural network during the training process of the CNN. The state of the art methods, such as [5], are not "zero-shot" methods since they require a dataset of training images to create the model capable of performing super resolution.

2.4 PSNR

PSNR is a common metric used to evaluate the performance of super resolution techniques [7]. PSNR is a function of the pixel-wise mean squared error between two images. Equation 1 shows the equation for calculating the PSNR between two images,

$$\text{PSNR} = 20 \log_{10} \frac{\text{MAX}}{\sqrt{\text{RMSE}}}, \tag{1}$$

where MAX is the largest possible value in the images (in 8bit images, it is 256) and RMSE is the root-mean square difference between each pixel of the two images. While this metric is useful for evaluating performance, it is not perfect since the squared error does not distinguish between important differences between images and irrelevant differences between images [8][9]. Thus, PSNR should be taken as a signal of the quality, and not the objective sole measurement [5]. Visual inspection can be another useful method to determine image quality.

3 Paper Summaries and Review

The following section will summarize two methods for performing zero-shot single image super resolution using convolutional neural networks (CNNs).

3.1 Zero-Shot Super Resolution Using Deep Internal Learning (ZSSR)

3.1.1 Theory behind ZSSR

Research has shown that there is a recurrence of small image patches in natural images across the same scale as well as varying scales [10]. This recurrence of information has been used to create effective image denoising algorithms [10]. This recurrence of information at different scales has also been used to create a super-resolution algorithm by using a nearest neighbour approach [10]. This method works relatively well, but performance suffers when the image patches do not have any close neighbours. The hypothesis behind this paper is that using a CNN as opposed to a nearest neighbour approach will allow better capturing of the prior that can be learned from the recurrence of information at varying spatial scales as well as enable better generalization for the patches of an image that do not have any "close neighbours" [11].

3.1.2 Methods

The test image I is downscaled at varying scales to generate a series of "high resolution" fathers (HR fathers). The HR fathers are downscaled by the desired scaling coefficient `s` and are the "lower resolution" sons (LR sons). Image augmentation on the HR fathers and LR sons is performed to increase the training dataset. Each pair is rotated at $90, 180$ and 270 degrees and mirrored horizontally and vertically. This adds 8x images to the training dataset. To allow for a large SR scale factor, s, the scaling is broken up into intermediate steps $(s_1, s_2, s_3 .... s_m = s)$. At each of these intermediate steps

during training, the HR fathers and LR sons from the previous step are included in the training data set. This enlarged training dataset captures as much information as possible about the image.

The neural network is an 8 layer hidden layer convolutional network with 64 channels in each layer. The ReLU activation function is used in the netwok. A ReLU activation is the max of the preactivation value and 0. The CNN is trained with L_1 loss using the Adam optimizer which is a more sophisticated optimization method than gradient descent. The initial learning rate is 0.001 and is reduced to one tenth the value if the standard deviation of the reconstruction error falls below a dynamic threshold. The training stops when the learning rate reaches a value of 10^{-6}.

3.2 Deep Image Prior (DIP)

3.2.1 Theory behind DIP

The "Deep Image Prior" paper argues that the inherent structure of a CNN is a strong enough prior to deliver strong performance in super resolution tasks [12]. The inverse problem of super resolution can be modelled as an energy minimization problem with the form:

$$x^* = \min_x E(x; x_0) + R(x) \tag{2}$$

where x is the desired high resolution image, x_0 is the low resolution image, $E(x; x_0)$ is the data dependent term, and $R(x)$ is the regularizer. CNNs work by applying network weights θ to an input z to generate the output x. Therefore, the output image x can be thought of as a function of θ for a fixed z. Hence,

$$\theta^* = \min_\theta E(f(\theta); x_0) + R(f(\theta)) \tag{3}$$

Since the regularizer imposes a prior on the model, and the desire is for the only prior to be the CNN structure itself, the regularizer term is dropped. This leaves the following minimization problem.

$$\theta^* = \min_\theta E(f(\theta); x_0) \tag{4}$$

For the task of super resolution, the data term $E(f(\theta); x_0)$ can be set to $\left\|d(x) - x_0\right\|^2$ which is equal to $\left\|d(f(\theta)) - x_0\right\|^2$ where $d(x)$ is the high resolution image downsampled to match the low resolution image x_0.

Therefore, by modifying the weights of a CNN to minimize the error between the low resolution image and the downscaled high resolution image, a natural-looking high resolution image can be generated.

3.2.2 Methods

A randomly initialized vector z is initialized and is used as input into an hour glass shaped convolutional neural network. The loss function, $\left\|d(f(\theta)) - x_0\right\|^2$, is the squared error between the downscaled high resolution image generated by the CNN and the original low resolution image. Adam optimization is used with default parameters to train the CNN. Since the CNN will overfit the image (especially since there is no other prior acting as a regularizer), early stopping is used and the training process ends after 2000 iterations.

4 Analysis of ZSSR and DIP

4.1 Similarities and differences between ZSSR and DIP

"Zero-Shot Super Resolution Using Deep Internal Learning" (ZSSR) and "Deep Image Prior" (DIP) were uploaded to Arxiv on Dec 17th 2017 and Nov 29th 2017, respectively. The obvious similarity between the two methods is that they are both perform zero-shot super resolution methods as opposed

to state of the art methods using a training dataset [5]. This means that the prior used by the model for ZSSR and DIP is derived solely from the image and the structure of the model. ZSSR and DIP both use a CNN as the structure of the model.

The core difference between ZSSR and DIP is the implicit prior of the model. The prior in the ZSSR model is a combination of the CNN structure as well as the weights of the CNN learning internal image statistics, particularly the recurrence of information at varying scales found in the image. The ZSSR method attributes the success of the method to the prior being learned from analyzing the internal image statistics. The ZSSR process is designed to capture as much information as possible about the image across spatial scales by generating training images of varying scales as well as performing data augmentation on this data. The DIP method does not attempt any of this, and instead shows that the inherent structure of a CNN is a strong enough prior to generate good super resolution images. While the paper says that training on a large corpus does improve super resolution performance, it questions whether the largest contributor to the performance of CNN based super resolution techniques is due to the prior being learned by the CNN, or is it due to the nature of the CNN itself. It may be of interest to the reader to see what the produced super resolution image looks like during the intermediate iterations of the training process. Appendix B contains an example of this.

4.2 Defining the Question

The performance reported by ZSSR and DIP relative to the state of the art benchmarks were quite similar. This raises the question of whether the performance of the ZSSR is due to it learning a prior due to the internal image statistics as the authors suggest, or is it due to the structure of the CNN and the additional information of the internal image statistics are not being captured effectively. The authors of the ZSSR paper do not present any information on the effects of the number of HR father and LR son pairs have on the efficacy of the model which would help show how the learned prior is affected by this technique.

The report will attempt to answer the question of is the ZSSR method learning a prior beyond what is provided by the CNN structure itself by comparing ZSSR's and DIP's performance on the same images. If the images generated by the two super resolution methods are significantly different, it indicates that the prior learned by the ZSSR method is different than the prior imposed by the CNN structure as represented by the performance of the DIP method. However, if the two super resolution images are similar, it is inconclusive as to whether the prior of the ZSSR method is substantively different as different priors can create a similar looking super resolution images. To tease apart this attribution of performance, a scenario needs to be constructed such that the two philosophies governing the two approaches would yield different answers. ZSSR is based on the approach using the recurrence of information at varying scales to a create the prior. Therefore, as the recurrence of information in the image increases, the performance of the super resolution should increase as well. However, if the quality of the super resolution done by ZSSR remains on par with DIP as the recurrence of information increases, then it is not using the information any more effectively than just the inherent structure of the CNN. The mandrill image, a common natural image used in image processing, will be used to evaluate the methods on a standard natural image. Next, an image of the Taj Mahal will be used to test if increasing the recurrence of information at varying scales in the image improves the relative performance of the ZSSR method as compared to the DIP method. Finally, a fractal image will be used for the case of a large amount of repetition in the image. If DIP performs as well as ZSSR on upscaling a fractal image, it is likely that the ZSSR method is not learning any additional prior beyond the prior imposed by the structure of the CNN.

It is important to note some caveats to this analysis. The actual structure of the CNN differs between the two papers, so differences in performance may also be due to one CNN structure imposing a better prior than the other. In addition, only three images per method are being tested. This is due to limitations on computational resources. It takes approximately 3 hours and 5 hours for the methods ZSSR and DIP to upscale an image. This is unfortunate since with so few images, it is difficult to make any definitive statements. Finally, the magnitude of the recurrence of information at varying scales in the tested images are estimated visually rather than empirically derived. This is

due to time constraints of finding an appropriate algorithmic method for determining this property in images. Since only the relative magnitude is required for this analysis, it was thought that a visual inspection would be good enough.

4.3 Testing Framework

The code for the ZSSR method and DIP method are both available on Github [13][14]. The only modification to the code was the changing of it to use CPU processing instead of GPU processing as there was no access to a suitable GPU. The results of the super resolution methods are plotted on a multipanel figure where the first image is the ground truth, the second image is the bicubic interpolation, the third image is the ZSSR method and the fourth image is the DIP method. The input fed into each of the three super resolution methods is the ground truth image downscaled by a factor of 4 using bicubic downscaling. In each test image, the super resolution methods scale the image by a factor of 4.

4.4 Test Case 1: Natural Image

The mandrill image is a common image used in image processing and was retrieved from the University of Southern California Signal and Image Processing Institute [15]. It exhibits fine details with edges and texture as well as flat edge-free regions. This is a sample natural image to test if there is a significant difference in image quality between the DIP and ZSSR method. This can be established by either distinctive visual differences in the image or a large difference in PSNR value. If the high resolution images generated by the DIP and ZSSR methods are significantly different, then this is an indication that the priors being used for the super resolution task are different. Figure 1 shows the ground truth mandrill image at the top left as well as super resolution versions generated by bicubic interpolation, ZSSR and DIP.

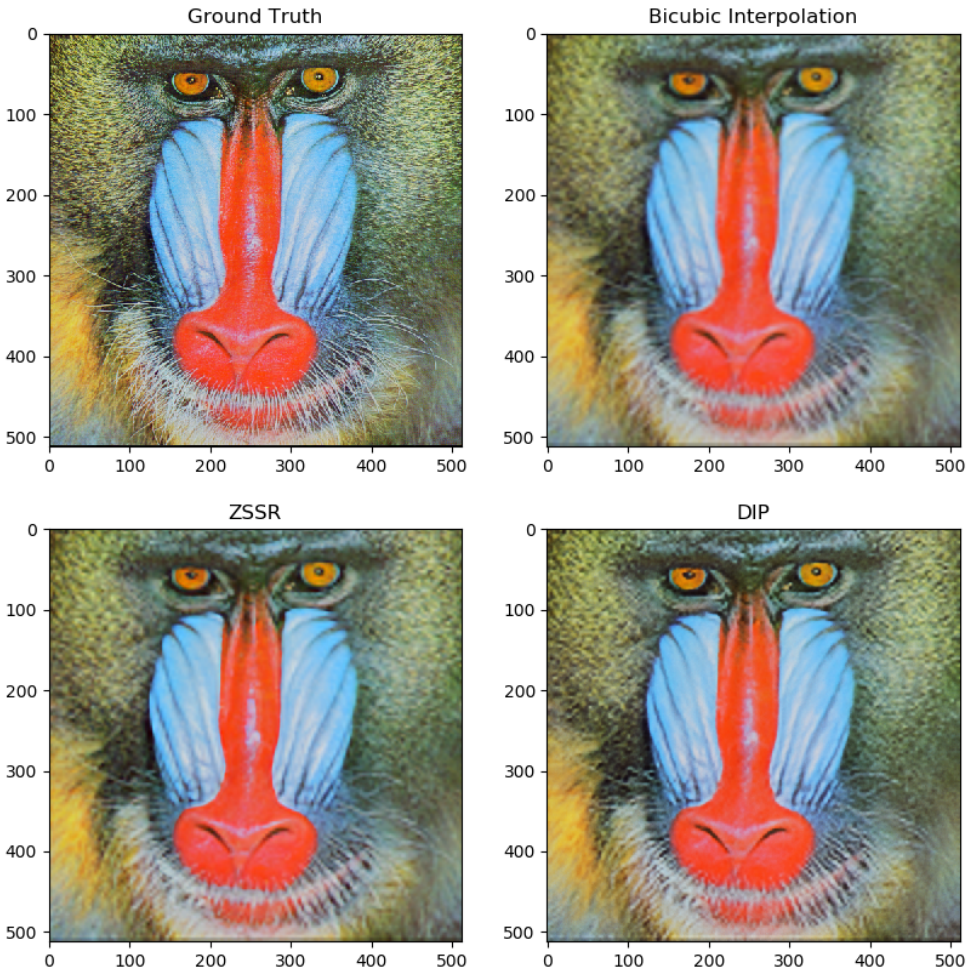

Figure 1: Ground truth mandrill image is shown in the top left of the figure. The three other images are the respective super resolution techniques which took in as input the ground truth image downscaled by a factor of four.

All three super resolution images look quite similar to each other. The deep learning methods have slightly more distinctive edges. It is most visible on the edges found on the blue parts of the mandrill's face as well the eyes of the mandrill. The PSNR for the methods DIP, ZSSR and bicubic are 18.33, 18.37 and 18.38 when compared to the ground truth image. This does not prove or disprove whether the priors being used by the ZSSR method and the DIP method are the same. Interestingly, the PSNR between the high resolution DIP image and the high resolution ZSSR image is 20.93. That is over 2dB more than the PSNR calculated against the ground truth. This indicates that there is at least some commonality of the prior between the two methods. Images with high degrees of informational recurrence at varying spatial scales are needed to better tease apart the question outlined in the section above.

4.5 Test Case 2: High Recurrence of Information

The Taj Mahal image was retrieved from the Wikipedia page for the Taj Mahal and is shown in Figure 13 at the top left [16]. The size of the image is 1024x840 pixels. The Taj Mahal exhibits a relatively large amount of reoccuring components of the image at varying scales. The architecture itself of the Taj Mahal has many repeating structures at varying scales due to both the archecture and the angle of photography. It will be important to examine the difference in super resolution quality around the pillars, spires and domes. In addition, the trees also have varying scale, so quality of super resolution can also be examined by looking at those components.

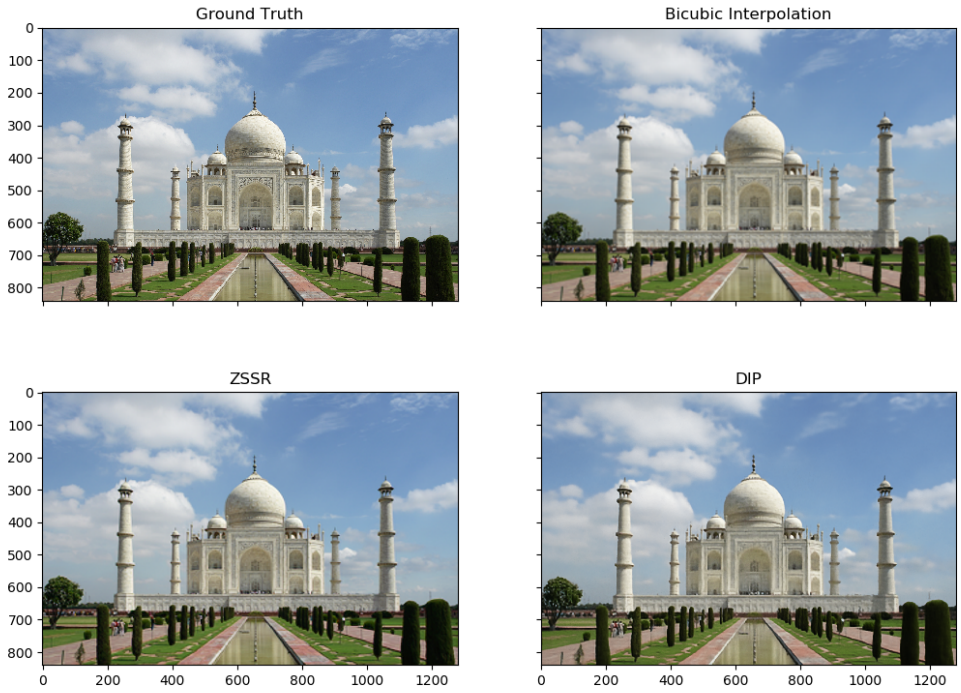

Figure 2: Ground truth Taj Mahal image is shown in the top left of the figure. The three other images are the respective super resolution techniques which took in as input the ground truth image downscaled by a factor of four.

Both deep learning methods have more distinct edges than the simple bicubic upscaling. This is most visible around the trees as well as the Taj Mahal building itself. Figure 3 shows the images zoomed in on the trees in front of the Taj Mahal and Figure 4 shows the image zoomed in on the building itself.

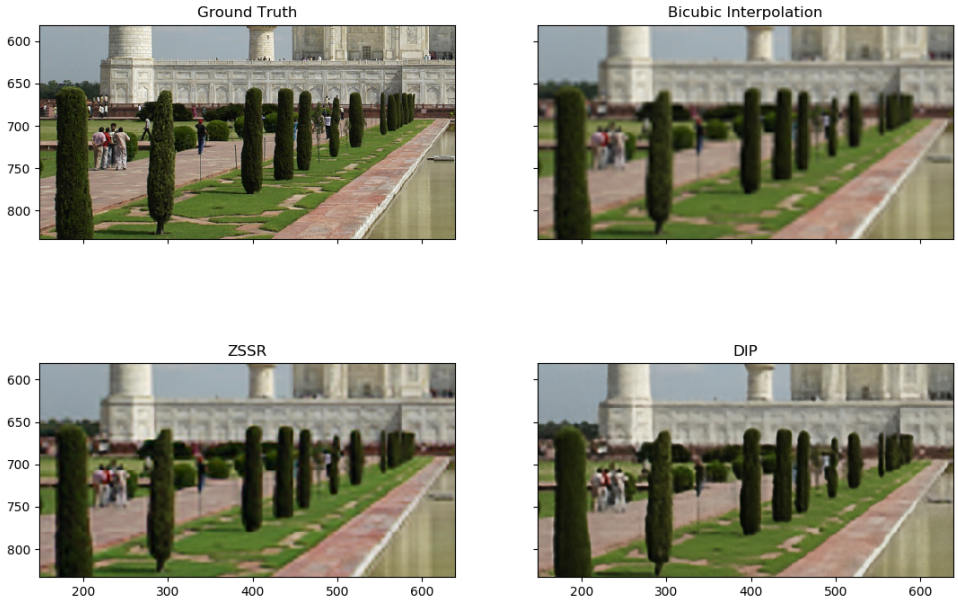

Figure 3: Taj Mahal image zoomed in on the trees in front of the building

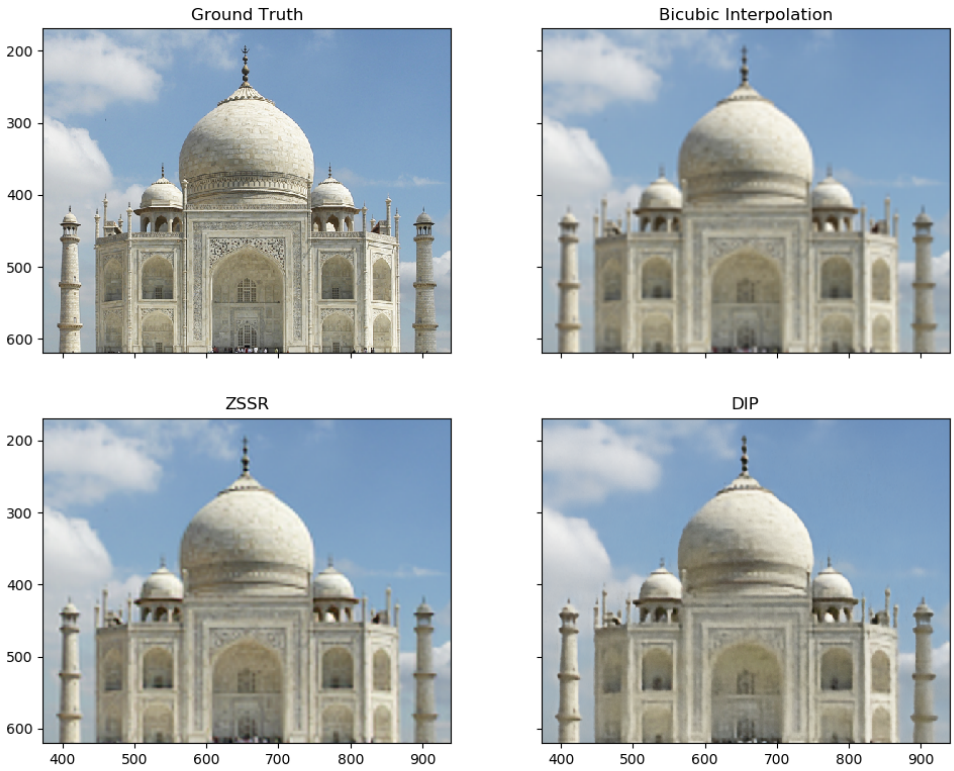

Figure 4: Taj Mahal image zoomed in on the Taj Mahal building

The most noticeable difference between ZSSR image and the DIP image is shown in Figure 4. The ZSSR image has more clear and accurate spires atop the domes of the Taj Mahal than the DIP method. The DIP method lacks this detail and for some spires, looks lumpy. This indicates that the ZSSR method has been able to incorporate additional information about the spires into the prior for more accurate super resolution. In addition, the carving pattern on the face of the Taj Mahal is more clear in the ZSSR method. However, ZSSR method is not better in every regard. As shown in Figure 5, which is a zoomed image of one of the pillars of the Taj Mahal, the DIP method has more clear edges between pillar and sky than the ZSSR method.

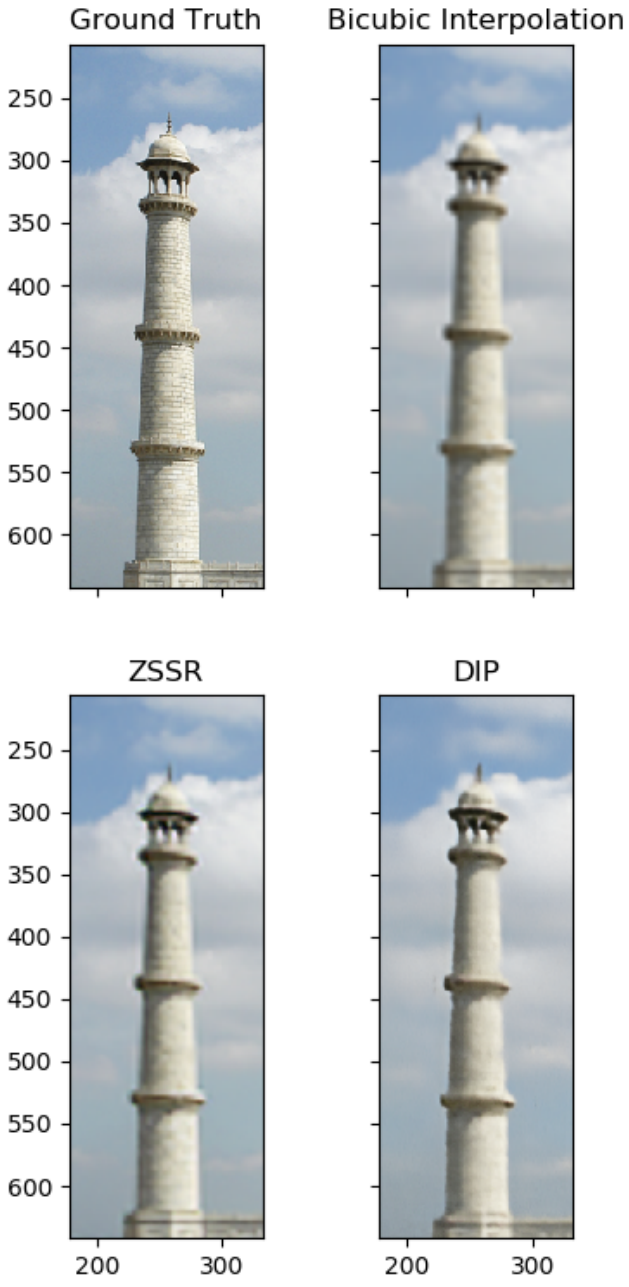

Figure 5: Zoomed on a pillar of the Taj Mahal

Pillars are another object that appear in the image at varying spatial scales, yet the DIP image is visually better. The PSNR values for the DIP, ZSSR and bicubic methods are 15.8, 16.53 and 16.62, respectively. The PSNR values are quite close to each other, indicating that the super resolution images are of similar quality. From this image, it is difficult to say confidently if the prior used by the ZSSR method is substantively different than the prior imposed by the CNN structure. The overall image quality is similar and while the spires were of higher detail in the ZSSR method, the pillars of the Taj Mahal were clearer in the DIP method.

4.6 Test Case 3: Very High Recurrence of Information

The Sierpinski triangle is created by iteratively subtracting smaller and smaller equilateral triangles from the larger equilateral triangle. The top left image of Figure 6 shows the ground truth Sierpinski triangle. The image was retrieved from the Wikipedia page on Sierpinski triangles [17] with a resolution of 1024x887 pixels. The Sierpinski triangle exhibits repeating information across spatial scales of the image. Looking at Figure 6, one can see the repeating subunits of the fractal image. This should be the ideal case for the ZSSR method since the image is copies of itself at varying scales. One expects the ZSSR method should outperform the DIP method by a relatively large margin if the ZSSR method is successfully learning a prior derived from the internal statistics of the image.

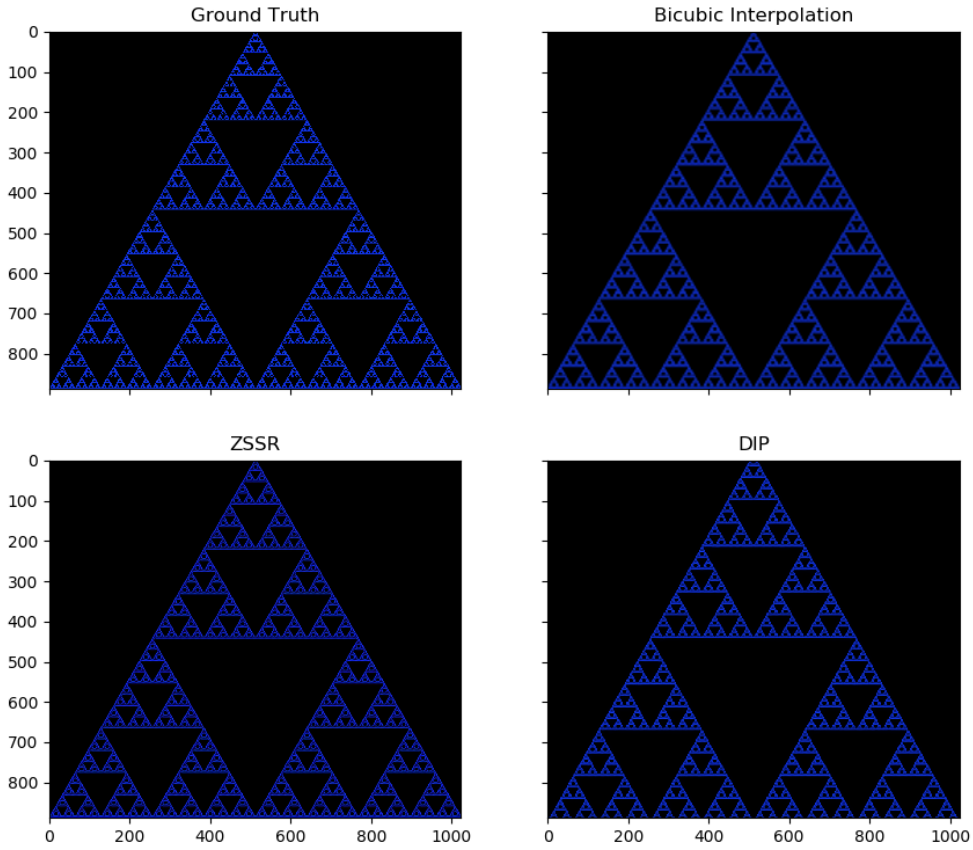

Figure 6: Ground truth Sierpinski triangle image is shown in the top left of the figure. The three other images are the respective super resolution techniques which took in as input the ground truth image downscaled by a factor of four.

It is difficult to see the details in the image, so Figure 7 presents a zoomed image of the top triangle of the original images.

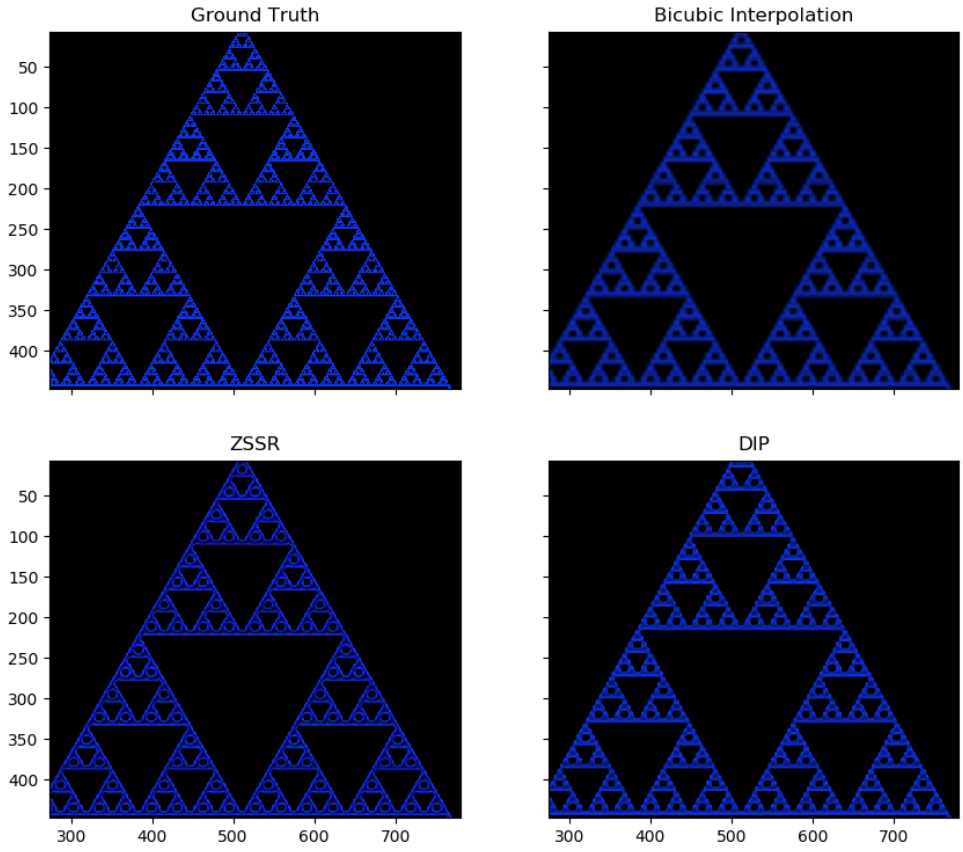

Figure 7: Ground truth and super resolution images zoomed in on the top triangle large triangle of Figure 6

Similar to the Taj Mahal image, the deep learning methods have significantly more defined edges than the bicubic method. A common pattern for all the super resolution images are the disappearance of the smallest holes that are barely visible in the ground truth image of Figure 7 and the reconstruction of the second smallest triangular holes as circles instead. This is because in the scaled down image shown in Figure 8 these holes are absent and the second smallest holes are genuinely circular.

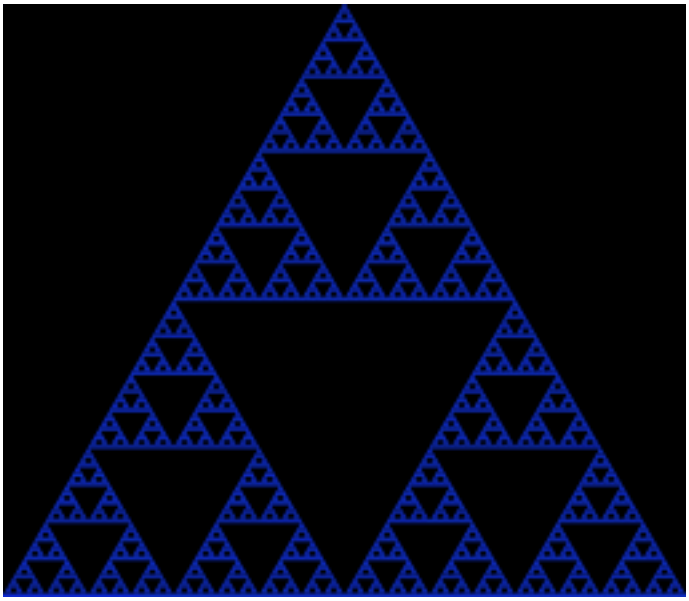

Figure 8: Downscaled Sierpinski triangle which is passed into the super resolution methods

There is a significant difference between the ZSSR image and the DIP image shown in Figures 6 and 7. The vertices found in the larger triangles in the DIP image all form sharp 60 degree angles. The ZSSR method has distinctive rounded vertices for each of the triangular holes across the spatial scales in the image. It is clear that a prior across scales has been learned by the ZSSR method which is not present in the DIP method. While the rounding of the triangle vertices is "incorrect", it is understandable as to why that prior was so strongly learned. Due to the fractal nature of the image, the smallest triangles found in the Sierpinski triangle make up around 50% of the total triangles in the image. Since the smallest holes of the downscaled image are circular (a byproduct of the downsampling method), this means that 50% of the holes/black triangles are circular. The ZSSR method emphasizes learning a prior that applies across scales in the image which results in the larger triangles having curved vertices.

The important thing is that this image demonstrably shows that the ZSSR method is learning a prior at one scale and applying it to other scales. This type of prior is not found in the DIP method where the larger triangles maintain sharp 60 degree edges and only the smallest holes form a circle. This shows that a substantively different prior can be learned by the two methods where the ZSSR learns a prior derived from internal statistical information about the image across scales.

As expected, the PSNR for the ZSSR method is significantly larger than the DIP method or bicubic interpolation with a PSNR value of 20.7 as opposed to 19.7 and 19.6, respectively. The PSNR value of the ZSSR method would most likely be higher if the prior learned by the model had not been corrupted by the circular holes in the downscaled input image.

5 Further Work

There are a number of ideas that could be done to better understand how the recurrence of information at varying scales in an image gets incorporated into the prior of the ZSSR method. A way to improve the efficacy of the analysis is to ensure that the ZSSR method and the DIP method have the same CNN structure. A confounding variable to the analysis conducted by this report is that the structure of the CNNs are different, hence it is likely that this imposes different priors on the method.

A way to extend the analysis is by choosing/defining a metric to quantify the recurrence of spatial information across scales in the image. Then, one could compare the two methods against a spectrum of images and see how increasing the recurrence of information impacts the relative performance of the

ZSSR method and the DIP method. This would give a more detailed understanding of the relationship between these variables. The analysis in this report visually estimates the relative magnitude of the recurrence of information in the images.

Finally, the ZSSR method itself could be analyzed further. By varying the number of HR fathers and LR sons generated in the training dataset, the strength of the learned prior that captures the information at varying scales could be altered. Analyzing this effect may yield interesting results.

6 Conclusion

Two approaches for zero shot super resolution were analyzed and compared. The "Zero-Shot Super Resolution using Deep Internal Learning" paper creates a CNN that learns a prior derived from the recurrence of spatial relationships at varying scales in an image to perform super resolution [11]. The "Deep Image Prior" paper shows that the CNN structure is enough to assert an accurate enough prior for super resolution [12]. Given that the two methods perform similarly well against state of the art methods, the question is raised as to whether the efficacy found in the ZSSR method is due to the learning of the prior derived from internal image statistics, or is it due to the prior inherent to a CNN structure described in the DIP paper. Three images consisting of varying degrees of recurrence of spatial information at different scales were analyzed to answer this question. The first two images were inconclusive but the last image, a fractal, showed that the ZSSR method does learn the spatial relationships occurring at varying scales found in the image.

References

[1] P. Milanfar, "Enhance! raisr sharp images with machine learning." {https://ai.googleblog.com/2016/11/enhance-raisr-sharp-images-with-machine.html}, 2013.

[2] P. Fieguth, *Statistical Image Processing and Multidimensional Modelling*. Springer, 2011.

[3] Y. X. Hong Chang, Dit-Yan Yeung, "Super-resolution through neighbor embedding," *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[4] Y. Xian, B. Schiele, and Z. Akata, "Zero-shot learning – the good, the bad and the ugly," *CVPR*, 2017.

[5] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *CVPR*, 2017.

[6] M. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Processing Magazine, volume 34 , Issue: 6*, 2017.

[7] C.-Y. Yang, C. Ma, and M.-H. Yang, "Single-image super-resolution: A benchmark," *ECCV*, 2014.

[8] H. R. S. Z. Wang, A. C. Bovik and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, 2004.

[9] E. P. S. Z. Wang and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *Conference on Signals, Systems and Computers, volume 2, pages 9–13*, 2003.

[10] M. I. Maria Zontak, "Internal statistics of a single natural image," *CVPR*, 2011.

[11] M. I. Assaf Shocher, Nadav Cohen, ""zero-shot" super-resolution using deep internal learning," *CVPR*, 2018.

[12] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," *CVPR*, 2018.

[13] M. I. Assaf Shocher, Nadav Cohen, ""zero-shot" super-resolution using deep internal learning." {https://github.com/assafshocher/ZSSR}, 2018.

[14] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior." {https://github.com/DmitryUlyanov/deep-image-prior}, 2017.

[15] U. S. California, "Mandrill." {http://sipi.usc.edu/database/database.php?volume=misc&image=10#top}, 2018.

[16] Yann and J. Carter, "Taj mahal." {https://commons.wikimedia.org/wiki/File:Taj_Mahal_(Edited).jpeg}, 2010.

[17] B. Stanislaus, "Sierpinski triangle." {https://en.wikipedia.org/wiki/Sierpinski_triangle#/media/File:Sierpinski_triangle.svg}, 2013.

Appendix A: Higher Resolution Images

Mandrill Images

The ground truth high resolution image of the mandrill is shown in Figure 9

Figure 9: Ground truth image of the mandrill with resolution 512x512

Figures 10, 11 and 12 show the 4x super resolution of the mandrill image using DIP, ZSSR and bicubic interpolation, respectively.

Figure 10: 4x super resolution of the mandrill using the deep image prior (DIP) method

Figure 11: 4x super resolution of the mandrill using the ZSSR method

Figure 12: 4x super resolution of the mandrill using bicubic interpolation

Taj Mahal Images

Figure 13: Ground truth image of the Taj Mahal with resolution 1024x840

Figures 14, 15 and 16 show the 4x super resolution of the Taj Mahal image using DIP, ZSSR and bicubic interpolation, respectively.

Figure 14: 4x super resolution of the Taj Mahal using the deep image prior (DIP) method

Figure 15: 4x super resolution of the Taj Mahal using the ZSSR method

Figure 16: 4x super resolution of the Taj Mahal using bicubic interpolation

Sierpinski Triangle Images

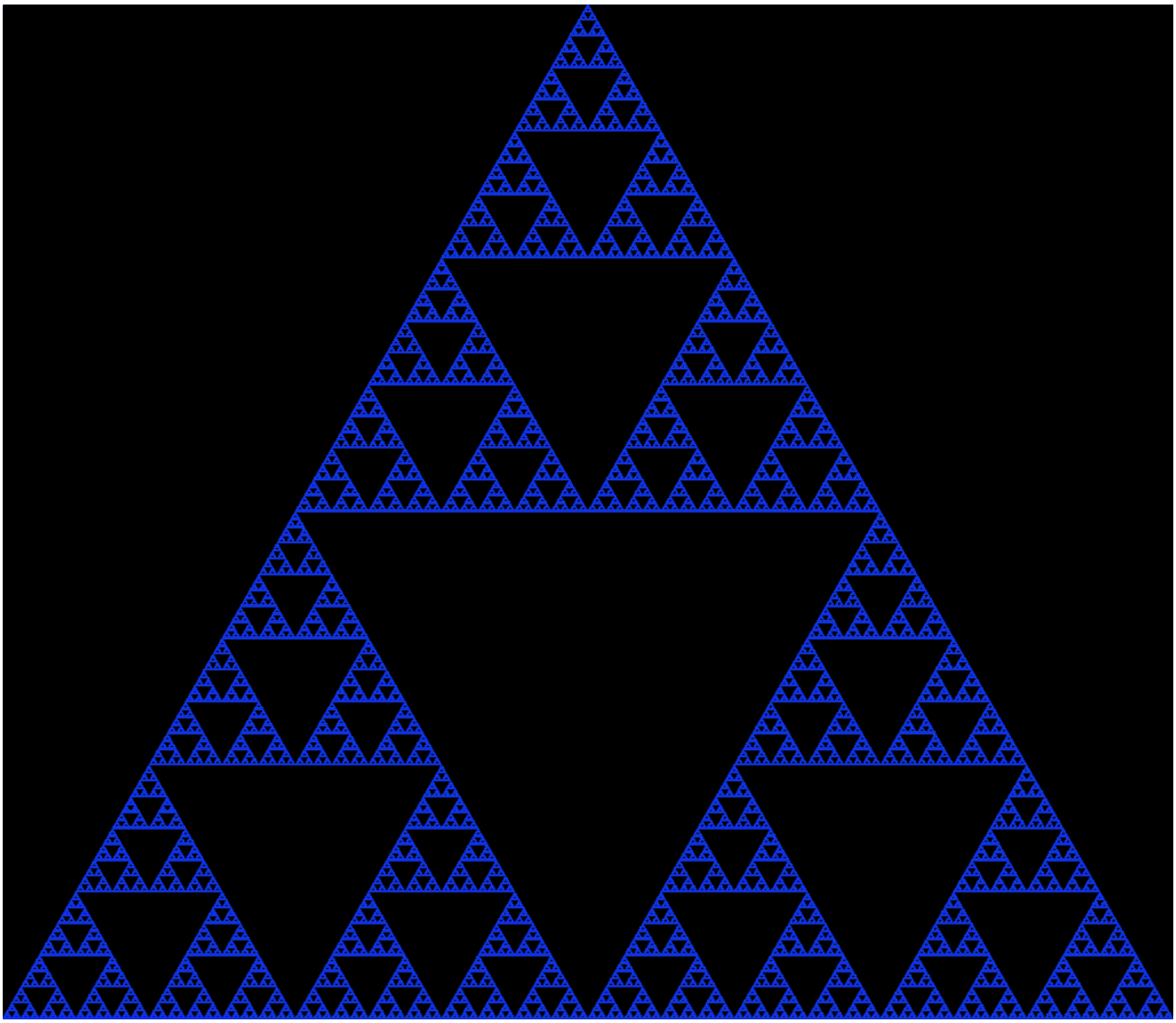

Figure 17: Ground truth image of the Sierpinksi Triangle with resolution 1024x887

Figures 18, 19 and 20 show the 4x super resolution of the Sierpinski triangle image using DIP, ZSSR and bicubic interpolation, respectively.

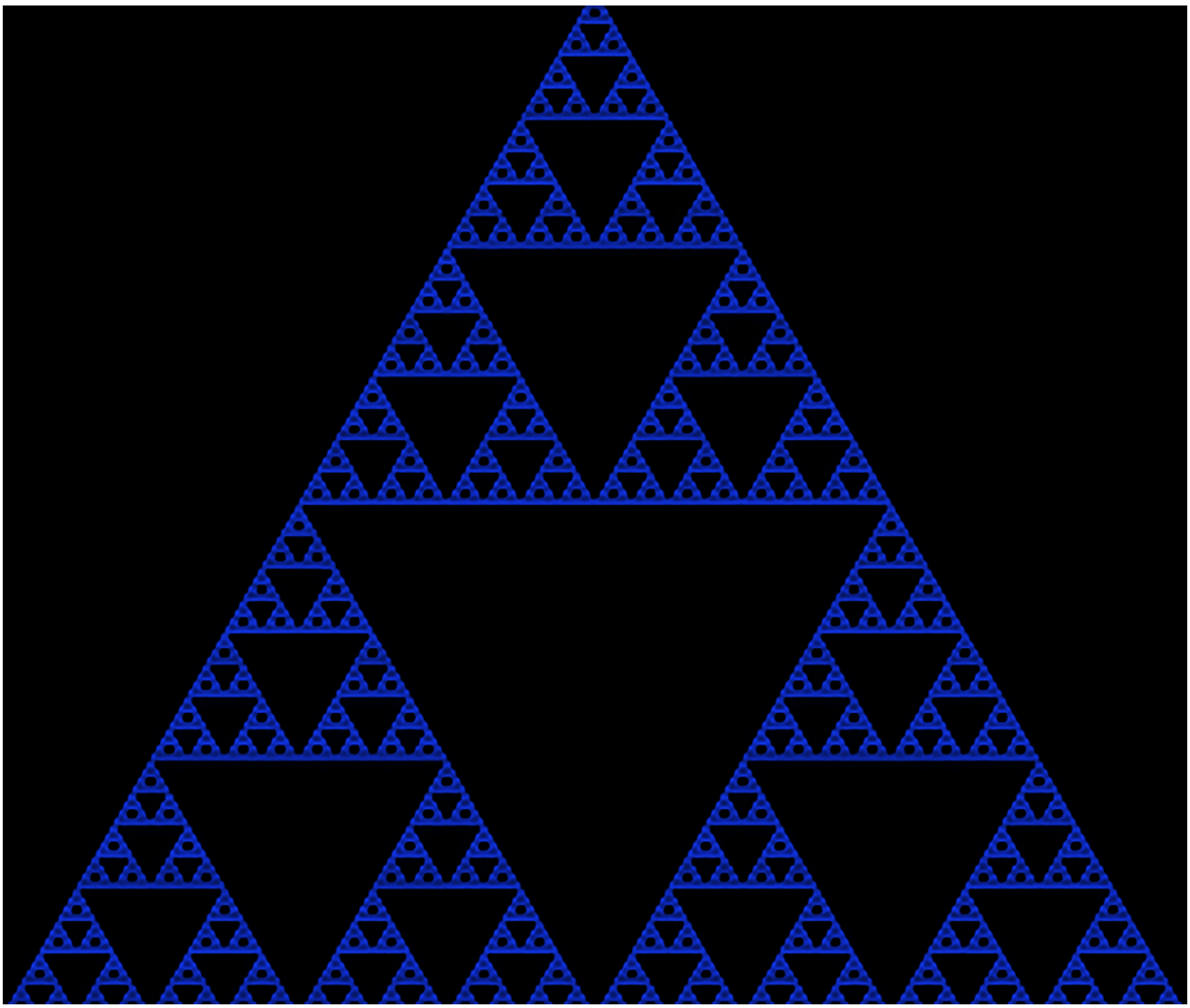

Figure 18: 4x super resolution of the Sierpinski triangle using the deep image prior (DIP) method

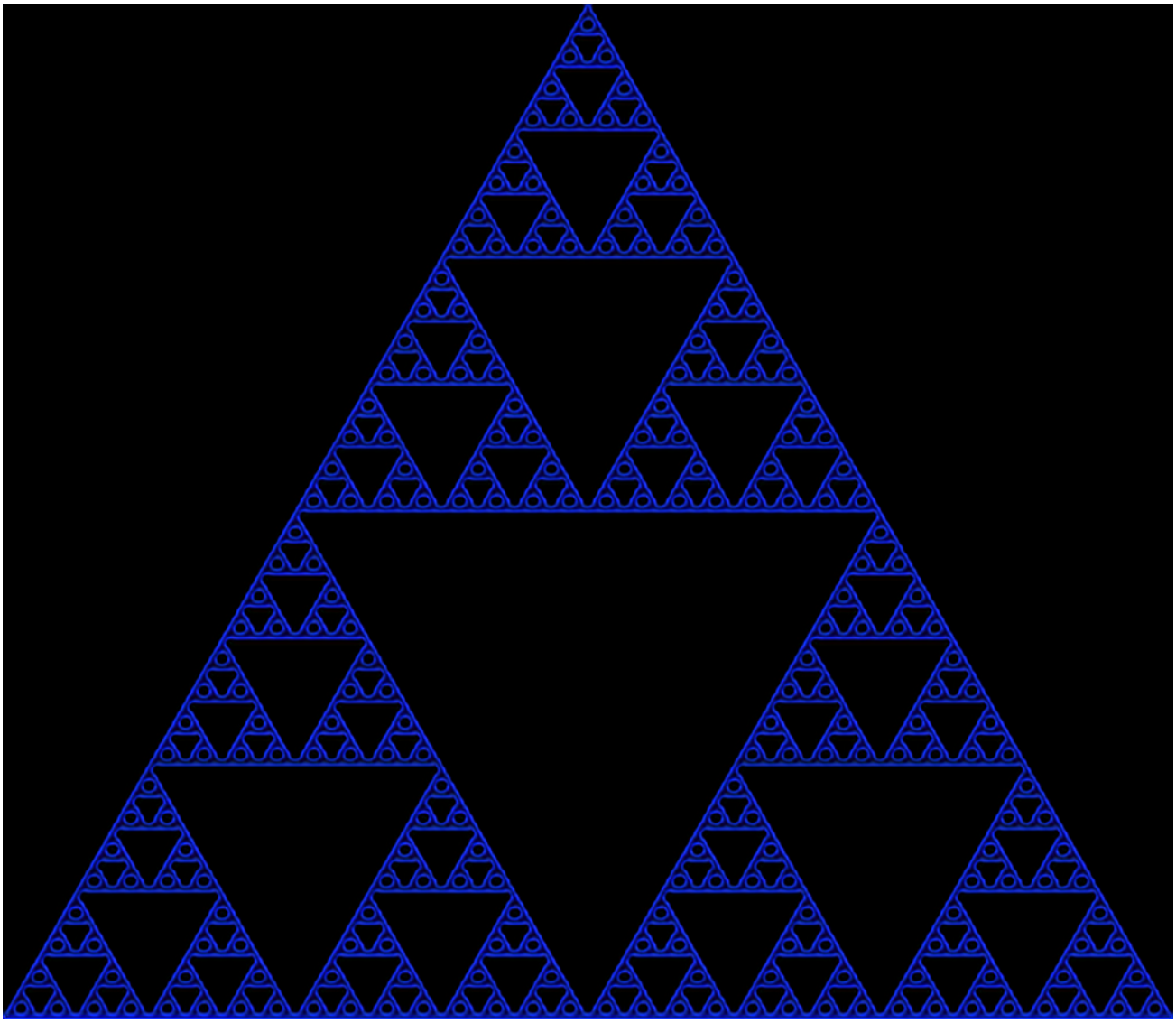

Figure 19: 4x super resolution of the Sierpinski triangle using the ZSSR method

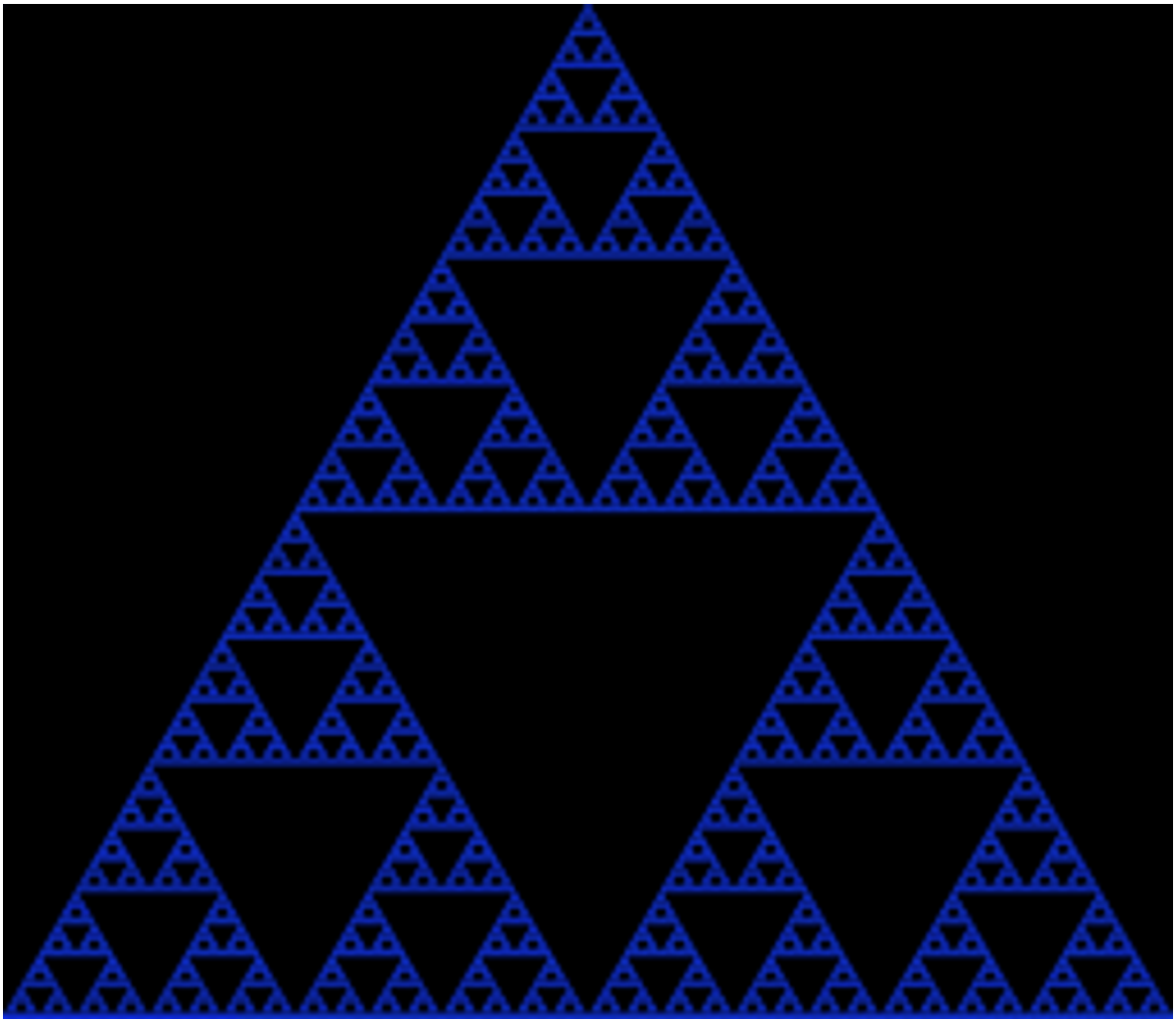

Figure 20: 4x super resolution of the Sierpinski triangle using bicubic interpolation

Appendix B: Higher Resolution Images

Figure 21 shows the evolution of the Taj Mahal image generated by the DIP method. The first row is the initial estimate, the second row is after 100 iterations of training, and the third row is after 200 iterations of training. The training continues up to 2000 iterations.

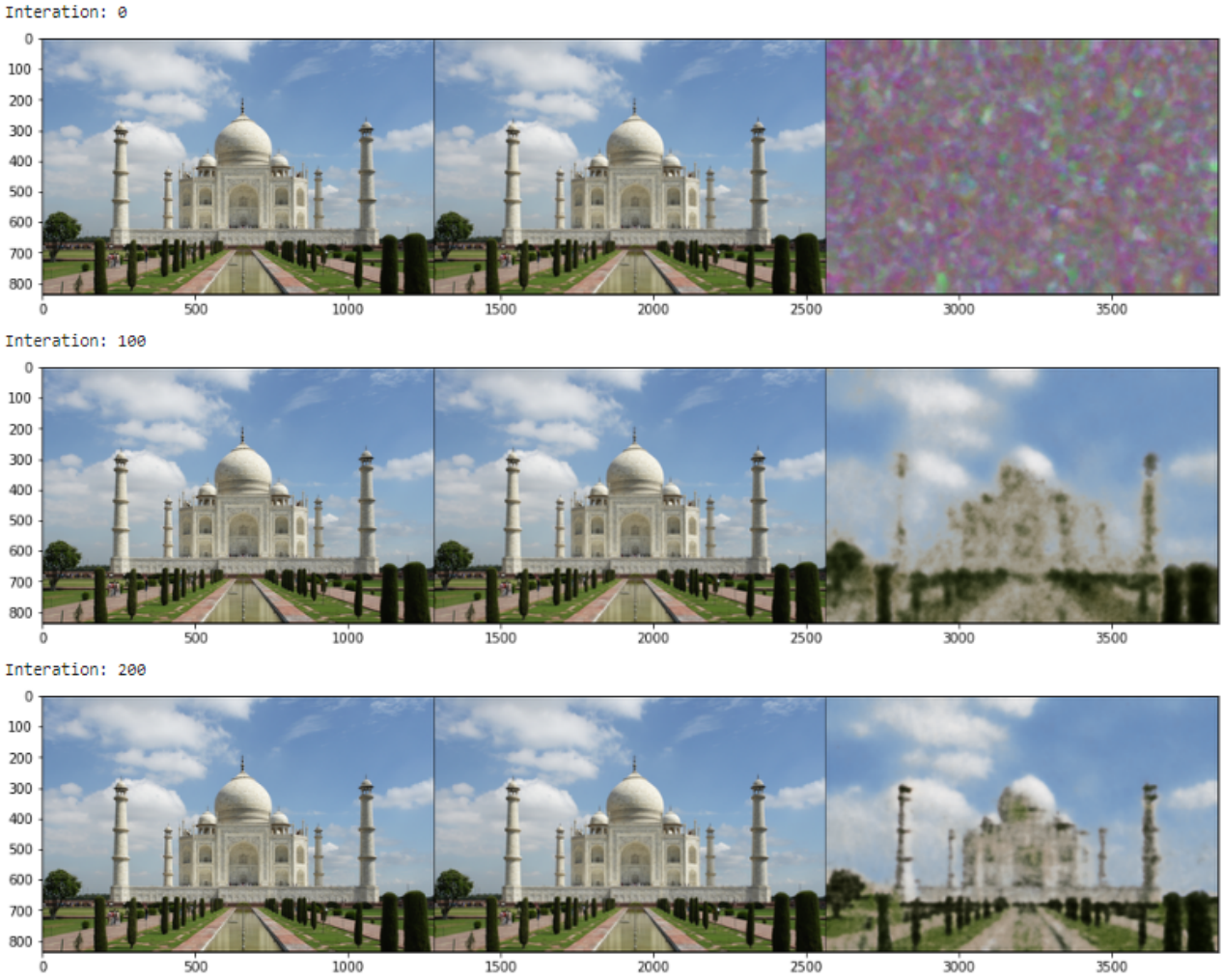

Figure 21: Left column is ground truth, middle column is bicubic interpolation and the right column is the generated super resolution image. Each row represents the image after 100 iterations of training.