Single Image Super-Resolution using Back-Propagation Neural Networks

Salman Taseen Haque
Electrical, Electronic and Computer Science Department
North South University, Bangladesh

Mohammad S. Hasan
School of Computing and Digital Technologies
Staffordshire University, UK

***Abstract*—There are several existing mathematical algorithms for colour image upscaling like Nearest Neighbour, Bicubic and Bilinear. This paper firstly investigates the performances of these three and it has been found that Bicubic performs the best in terms of structural similarity and execution time. A Bicubic with backpropagation based ANN method has been proposed to improve the results. Bicubic with ANN shows 6.5% higher SSIM, 6.9% higher PSNR, 8.7% higher SNR and 30.23% lower MSE values than pure Bicubic. The results of Bicubic with ANN are also compared with state of the art super-resolution techniques like SRCNN. Bicubic with ANN produces 1.48% higher SSIM and 3.44% higher PSNR than SRCNN.**

Keywords—Artificial Neural Network (ANN), Image Upscaling, Backpropagation, Super-resolution, Bicubic.

I. INTRODUCTION

Image super-resolution or upscaling is a method of constructing a higher resolution image from one or more lower resolution source images. There are many applications of this method. For example, it can be used when a low-resolution video or photo is being displayed on a higher resolution screen and increasing the apparent quality of the picture or video being shown. Medical imaging can also benefit from image upscaling to identify small fractures in bones or other similarly small injuries. It can also be used for surveillance purposes e.g. if a face of a suspect is caught on CCTV camera and is not of high enough quality for proper identification, intelligent image upscaling can approximate the details of the face for more accurate identification.

One of the two ways to achieve high-resolution images from lower resolution source images is called single image super-resolution where the output high-resolution image is achieved from a single low-resolution image. The other method is called multi-image super-resolution where the high-resolution image is created from multiple source images. Both of these methods are used depending on the use case. Generally, multi-image super-resolution provides better results as there are more data to begin with. However, in certain circumstances, obtaining multiple minutely different images of the same subject is not practical. For instance, if one were to upscale an image from a CCTV video footage, then there will not be multiple images from the same angle at the same instant- hence making multi-image super-resolution techniques futile. This paper investigates upscaling of facial images from a single source image.

This paper is organised as follows. Existing works regarding single image super-resolution is discussed in section II. Section III presents the design of the experiments. The results of the experiments and the analysis of the results are done in section IV. Some conclusion and future work directions are given in section V.

II. EXISTING WORKS

There are various existing research works e.g. [1], [2], [3], [4] etc. that discuss numerous methods for image super-resolution. Both multi and single frame image super-resolution are commonly discussed topics in computer vision. Normally, single image upscaling can be done via a number of ways without the assistance of neural networks. The most commonly used algorithms for upscaling images include Bilinear, Nearest Neighbour and Bicubic interpolation methods, etc. The simplest of these algorithms is Nearest Neighbour where it selects the pixel value of its nearest pixel and does not take into account the values of other neighbouring pixels. The Bilinear interpolation method considers the closest 2 by 2 neighbourhood pixel block surrounding the pixel to be created and then takes a weighted average of these four pixels to calculate the new pixel value. On the other hand, the Bicubic algorithm considers the closest 4 by 4-pixel neighbourhood and calculates the weighted average of these 16 pixels to come up with the new pixel value to give more weight to the pixels closest to the new pixel.

In theory, the higher the number of pixels used to generate a new pixel, the more accurate the result is. There are higher-order methods available which are more accurate since they take into account more surrounding pixels to come up with the new pixel value. But these algorithms are computationally far more demanding, and also the visible improvement by these higher-order interpolation algorithms like Spline and Sinc over lower order interpolation methods e.g. Bicubic is not high enough to justify the added computational cost, especially for a single step enlargement [5]. The adaptive methods, on the other hand, consider image texture when interpolating- resulting in generally more natural looking high-resolution images. However, these algorithms are generally tailored to minimise artefacts and maximise details in enlarged photos. So in most cases, these cannot be used to distort or rotate an image [6] which limits their flexibility when viewing.

One way to improve the results from these algorithms even further is to modify the existing algorithms for specific purposes. For instance, a modification of the Bicubic interpolation method called "Edge Directed Bicubic Interpolation" produces higher peak signal to noise ratio (PSNR) values than normal Bicubic algorithm [7]. Another way to improve the results is to use Machine Learning (ML) algorithms [8]. A number of researches is available on this topic e.g. [9], [10] etc. Most of these papers discuss various ML approaches when interpolating images instead of relying on the mathematical methods discussed above. These methods, while providing very good results, require long training phases and complicated filter designs [11]. For quick results, such networks are mostly impractical. In most of these cases, Convolutional Neural Networks (CNN) is used. In some cases, modified versions of neural networks such as Super-Resolution Convolutional Neural Network (SRCNN) [12] is used to provide very good results. However, all of these methods rely on creating completely new gradient mapping filters using machine learning. This approach is very complex and time-consuming to design and also requires a lot of training time. There has also been some work [13] that uses frequency domain approach to tackle this problem.

Another approach to solve this problem is to develop a neural network that modifies the output of already upscaled images created by computationally cheap algorithms like Bicubic. For example, Google has recently developed an algorithm called "Rapid and Accurate Image Super Resolution (RAISR)" [14]. RAISR uses computationally non-demanding algorithms to upscale an image and then applies hash based ML on that to develop filters. Since the input images are already upscaled to begin with, the training phase takes much less time to come up with a filter that accurately approximates the actual image. The results are much faster than other state-of-the-art algorithms like SRCNN and Anchored Neighbourhood Regression (ANR), while providing broadly similar results.

Another paper [15] discusses using artificial neural networks to achieve the same goal. It also reveals that using feed-forward neural networks improve the results achieved by Bicubic interpolation on binary images.

III. Experiment Design

The approach of this paper in solving this problem can be broken down into several separate phases which can be seen in Figure 1. This paper takes the approach of first using a cheap upscaling method and then developing a back propagation based ANN to improve its results, rather than upscaling the low-resolution image from scratch. The phases are explained below.

A. *Selecting the cheap upscaling method*

The initial method is chosen from Nearest Neighbour, Bilinear and Bicubic based on the algorithm's performance with regards to time and structural similarity index. The test is performed with a set of 10 images from MATLAB's image processing toolbox.

B. *Selecting the input and target images*

A set [16] of 450 images of faces (350 pixel x 350 pixel) in different lighting conditions is used as target images to aid learning. Lower resolution versions of the same images are used as input images in the ANN.

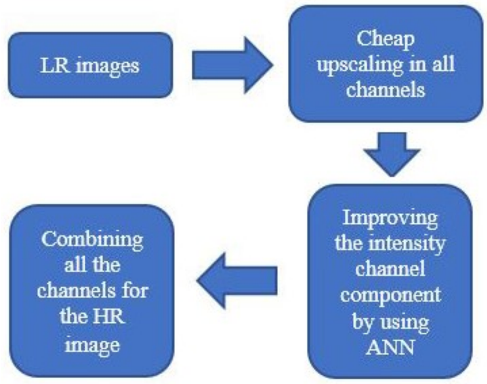

Figure 1: Flowchart of the proposed method.

C. *Preprocessing the images*

All of the input images are then split into their hue, saturation and intensity channels. Since the hue and saturation components of a colour image are not as susceptible to artefacts when interpolated, using the computationally cheaper method when upscaling only the intensity channel is sufficient. Hence, the intensity channel components of the low-resolution images are then used as inputs to the neural network and the intensity components of the original high-resolution versions of the same images are used as target images.

The original higher resolution pictures are first downscaled by a factor of 0.25 or one-fourth of their original resolution. These lower resolution versions are then used as input images and the higher resolution original versions are used as target images to train the neural network.

D. *Designing a Backpropagation based artificial neural network*

In this case, a Feedforward Backpropagation network is used. The Backpropagation training method subtracts the training output from the target images and generates the errors. It then back propagates and adjusts the weights in the hidden layers to reduce the error - only terminating when it is below a sufficient threshold.

The final network uses Traingdx as its training function and Learngdm as its adapting learning function. Log-based or tan-based sigmoid transfer functions provide better results in cases like this as the mapping of the high-resolution images to the low-resolution images is not linear. Hence, linear transfer functions e.g. Purelin is ignored. According to [17], Tansig transfer functions provide lower Mean Squared Error (MSE) values than Logsig transfer functions. Hence, Tansig is used in this case.

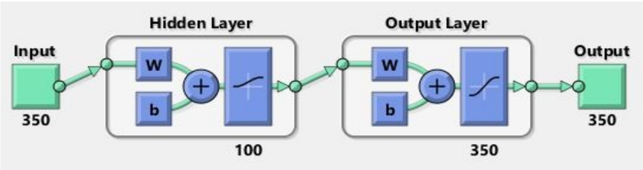

Figure 2: Structure of the Feedforward Backpropagation network.

The performance of the network is evaluated by calculating MSE. According to the results shown in [18], the depth of the network is kept relatively shallow for performance benefits. The network has 20 hidden layers and 100 neurons. The exact number of neurons and layers are selected by a trial and error basis. The final structure of the network can be seen in Figure 2.

When designing the network, only facial images are used as inputs to construct it. According to [2], neural networks tuned for a specific type of image input e.g. human faces tend to produce better results than general purpose networks.

E. *Applying the network on the sample images*

After developing a network that provides satisfactory results, 10 new low-resolution sample images of faces are then used as inputs. The high-resolution outputs of those images are compared with results obtained by the selected computationally cheap method on the basis of Structural Similarity (SSIM) index, Peak Signal to Noise Ratio (PSNR), Signal to Noise Ratio (SNR) and MSE.

All the tests are conducted on an Intel® Core™ i5-7200U CPU based computer running at 2.50GHz with 8GB of RAM. The software used is MATLAB R2016a.

IV. Results and analysis

Testing of the three cheap methods of interpolation, namely Nearest Neighbour, Bilinear and Bicubic reveals the tradeoffs associated with all of them. To calculate the time, for each image, 100 iterations of each algorithm is performed and then averaged out for a single iteration's execution time. The results from the 10 images are then averaged to come up with the single execution times per image.

It is clearly evident from Figure 3 that Nearest Neighbour, being the simplest of the three is faster than the other two methods and Bicubic is the slowest by around 15%.

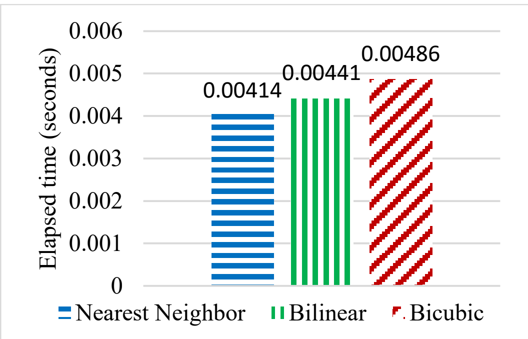

Figure 3: Execution time comparison between Nearest Neighbour, Bilinear and Bicubic interpolation methods.

On the other hand, when SSIM is measured, Bicubic provides the best results of the three as shown in Figure 4 where higher numbers corresponding to better results.

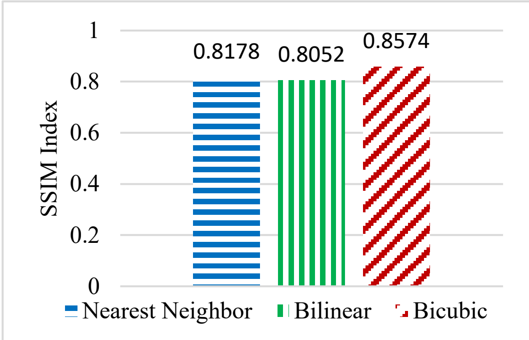

Figure 4: Structural similarity comparison between Nearest Neighbour, Bilinear and Bicubic interpolation methods.

From a visual perspective, Bicubic also provides the best balance between these three algorithms as it is neither too sharp nor jagged like the Nearest Neighbour or too smooth and soft like Bilinear interpolation. A real-world example of this can be seen in Figure 5.

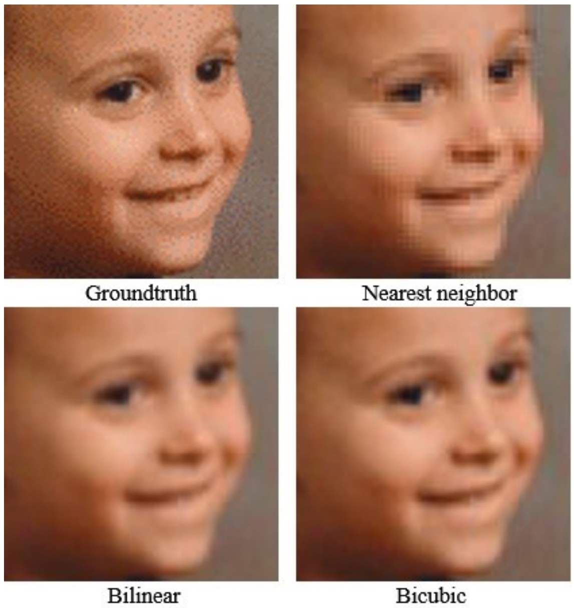

Figure 5: Comparison between computationally cheap interpolation algorithms. Among these, Nearest Neighbour (top right) produces the most jagged artefacts, Bilinear (bottom left) produces the softest results and Bicubic (bottom right) produces the best balance overall. The ground truth image (top left) can be observed for reference.

From these tests, it is evident while Bicubic is the slowest of the three methods; however, is also the best in terms of structural similarity. The relative time differences between Nearest Neighbour and Bicubic is small, but Bicubic provides higher SSIM values and much fewer artefacts as shown in Figure 5. When comparing Bicubic with Bilinear, the time difference is negligible (by about 9%). However, Bicubic provides better SSIM values and sharper results as shown in Figure 5. As a result, Bicubic is chosen to be the preferred computationally cheap method and is also used to downscale all of the images in the dataset.

After using the set of 450 high-resolution and low-resolution pairs of the same image for training the neural network and using the trained network on the set of 10 images, the acquired results in terms of SSIM, PSNR, SNR and MSE, averaged from the 10 samples are discussed below.

In the case of Figure 6 and Figure 7, the higher the number is, the better the result is. However, since Figure 8 is a measure of MSE, in this case, the better result is given by the lower number. From Figure 6, Figure 7 and Figure 8, it is clear that using the neural network on the upscaled images results in much better SSIM index values, which is also a good indicator of visual similarity. Moreover, using the neural network also reduces the noise in the image resulting in higher PSNR and SNR values and fewer artefacts. This also results in lower MSE values.

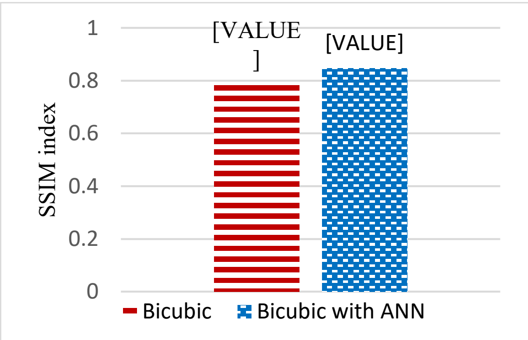

Figure 6: SSIM index comparison between Bicubic interpolation and Bicubic interpolation with ANN.

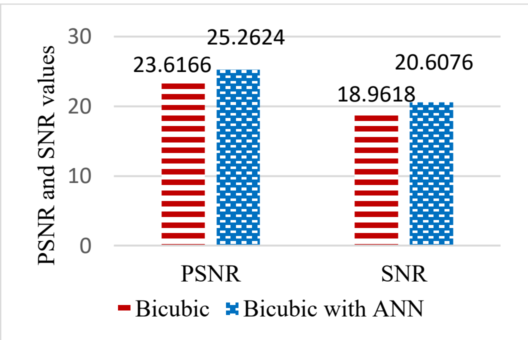

Figure 7: PSNR and SNR value comparison between Bicubic interpolation and Bicubic interpolation with ANN.

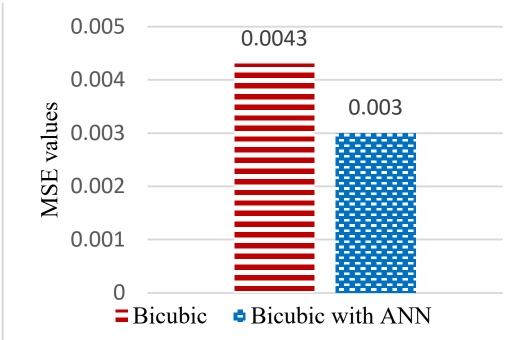

Figure 8: MSE comparison between Bicubic interpolation and Bicubic interpolation with ANN.

Visually, the differences between Bicubic only and Bicubic with ANN is even more noticeable than the numbers suggest as shown in Figure 9. Here, the outputs from the proposed method (Bicubic with ANN) can be visually compared with the results obtained from only Bicubic interpolation. In both cases, 4 times upscaling is done. In Figure 9, the ground truth images and the original low-resolution images without any upscaling can be seen for reference.

The areas of difference in structural similarity between only Bicubic and Bicubic with ANN can be seen in the local SSIM image in Figure 10. Here, the darker areas of the image represent more difference and lighter areas of the image represent more similarity- with pure white meaning that the images are structurally identical in that particular area.

Figure 9: Visual comparison between original high-resolution images (row 1), low-resolution versions of those images with 1/4th resolution with no upscaling (row 2), upscaled high-resolution versions obtained using only Bicubic upscaling (row 3) and high-resolution versions obtained using Bicubic upscaling with Backpropagation artificial neural network (row 4).

Figure 10: Local SSIM of the images used in Figure 9. The darker areas represent the least structural similarity between Bicubic with ANN and Bicubic only, while the white areas represent the most structural similarity. Areas of smoother details are similar in both images, as can be seen in the smoother skin around the cheeks. However, areas with finer details like eyes, teeth, wrinkles and hair are structurally dissimilar- implying a greater improvement in these areas of the image obtained by the Bicubic with ANN.

When the percentage of improvement is concerned, this result is comparable to that of highly regarded algorithms like SRCNN e.g. [19]. The improvement over Bicubic in percentages can be seen in Table 1.

Table 1: Comparison of improvement percentages between SRCNN and the proposed method.

Improvement Category	**SRCNN**	**Bicubic with ANN**
SSIM	5.49%	6.97%
PSNR	4.44%	7.88%

However, it should be noted that although SRCNN lags behind the proposed method in this specific scenario, the former is a general purpose upscaling procedure. In other cases when not upscaling facial images, SRCNN gives very good results for a wide variety of inputs [20], [21], [22].

V. Conclusion

From the testing, it can clearly be concluded that using a Backpropagation based ANN to improve the results of computationally cheaper methods like Bicubic interpolation provides consistently better image quality overall than using just Bicubic interpolation. Moreover, since the input images are already interpolated, the network takes much less time to train than if it had to do all the upscaling itself. In essence, the method provides higher SSIM and PSNR values and lower MSE values than conventional upscaling methods like Bicubic on their own. It offers the advantages of being much simpler and less time-consuming than designing a neural network that can do all the processing by itself. For future improvements, different and larger datasets can be used to make it more adaptable for a wider variety of uses.

Acknowledgment

The authors would like to acknowledge an Erasmus+ International Credit Mobility (ICM) fund for Bangladesh awarded to Dr. Mohammad Hasan at Staffordshire University, UK in 2016.

References

[1] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, and L. Zhang, “Image super-resolution: The techniques, applications, and future,” *Signal Processing*, vol. 128, pp. 389–408, 2016.

[2] D. Capel and A. Zisserman, “Computer vision applied to super resolution,” *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 75–86, 2003.

[3] X. Kang, S. Li, and J. Hu, “Fusing Soft-Decision-Adaptive and Bicubic methods for Image Interpolation,” *Int. Conf. Pattern Recognit.*, no. ICPR, pp. 1043–1046, 2012.

[4] S. Baker and T. Kanade, “Limits on super-resolution and how to break them,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, 2002.

[5] W. Virginia, “Comparison of Different Image Interpolation Abstract Comparison of Different Image,” 2008.

[6] “Understanding Digital Image Interpolation.” [Online]. Available: http://www.cambridgeincolour.com/tutorials/image-interpolation.htm. [Accessed: 25-Aug-2017].

[7] Z. Dengwen, “An edge-directed bicubic interpolation algorithm,” *Proc. - 2010 3rd Int. Congr. Image Signal Process. CISP 2010*, vol. 3, pp. 1186–1189, 2010.

[8] “Google sets RAISR to work: Hi-res images now hog less Google+ mobile bandwidth | ZDNet.” [Online]. Available:

http://www.zdnet.com/article/google-sets-raisr-to-work-hi-res-images-now-hog-less-google-mobile-bandwidth/. [Accessed: 26-Aug-2017].

[9] A. Van Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel Recurrent Neural Networks," *Proc. 33rd Int. Conf. Mach. Learn.*, vol. 48, pp. 1747–1756, 2016.

[10] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses Supplementary," *Arxiv*, pp. 1–5, 2016.

[11] N. Kumar, J. Mehta, A. Sethi, and P. K. Deswal, "Neural network based single image super resolution," *11th Symp. Neural Netw. Appl. Electr. Eng.*, pp. 213–218, 2012.

[12] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2016.

[13] P. Vandewalle, S. Süsstrunk, and M. Vetterll, "A frequency domain approach to registration of aliased images with application to super-resolution," *EURASIP J. Appl. Signal Processing*, vol. 2006, pp. 1–14, 2006.

[14] Y. Romano, J. Isidoro, and P. Milanfar, "RAISR: Rapid and Accurate Image Super Resolution," *arXiv*, vol. 1606.01299, pp. 1–27, 2016.

[15] C. Gershenson, "Artificial Neural Networks for Beginners," *Networks*, vol. cs.NE/0308, no. 16, p. 8, 2003.

[16] "Computational Vision: Archive." [Online]. Available: http://www.vision.caltech.edu/Image_Datasets/faces/faces.tar. [Accessed: 25-Aug-2017].

[17] M. M. S, "Performance Analysis of Neural Network Architecture Combined with DWT for Image Compression," vol. 58, no. 11, pp. 13–20, 2012.

[18] K. He, "Convolutional Neural Networks at Constrained Time Cost," pp. 5353–5360, 2015.

[19] C. Ledig *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *arXiv*, pp. 1–14, 2016.

[20] Y. Liang, J. Wang, S. Zhou, Y. Gong, and N. Zheng, "Incorporating image priors with deep convolutional neural networks for image super-resolution," *Neurocomputing*, vol. 194, pp. 340–347, 2016.

[21] H. Mao, Y. Wu, J. Li, and Y. Fu, "Super Resolution of the Partial Pixelated Images With Deep Convolutional Neural Network," *Proc. 2016 ACM Multimed. Conf. - MM '16*, pp. 322–326, 2016.

[22] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," *Cvpr*, pp. 1646–1654, 2016.