

# Deep Learning based Super-Resolution

Fernando Zapata Barron, Jose Manuel Mejia Muñoz, Boris Jesus Mederos Madrazo, and Leticia Ortega Maynez

Universidad Autónoma de Ciudad Juárez  
Departamento de Ingeniería Eléctrica y Computación  
Avenida del Charro No. 450 Norte, Ciudad Juárez, Chihuahua, México  
a1164459@alumnos.uacj.mx, {jose.mejia,boris.mederos,lortega}@uacj.mx  
<http://www.uacj.mx>

**Abstract.** We propose a method for image super-resolution with basis on deep learning. This method makes use of a convolutional neural network to find the similarities between low-resolution and high-resolution patches of an image and learn a mapping between them. The network is capable of outputting a high-resolution image, taking a low-resolution image as an input, it can handle three color channels, and it's performant enough for use in real-time systems.

**Keywords:** Super-Resolution · Deep Learning · Convolutional Neural Network

## 1 Introduction

It's desirable to make use of high-resolution images on applications that require the use of digital imaging, given how these can contain details critical for various applications[15] in comparison to their low-resolution equivalents, as seen in Figure 1. A doctor can rely on a high-resolution image to make a correct diagnosis, objects can be easily distinguished on a high-resolution satellite image[11], or a pattern recognition algorithm can achieve a higher effectiveness when provided with high-resolution image samples.

The simplest methods to augment the resolution of a captured image are either diminishing the size of the pixel, or enlarging the image sensor itself. However, these approaches can prove to be prohibitively costly on large scale applications, or on those requiring a high degree of precision. Because of this, it's preferable to apply an algorithmic approach which doesn't depend on the development of new sensors and allows for the use of already existing image capture systems.

For this purpose, there exists a method known as super-resolution. Although there are many super-resolution techniques, most are based on the same basic idea: make use of the information contained within a multitude of distinct images of the same scene to reconstruct a new, higher resolution image [10]. This is unlike traditional image scaling methods, which synthesize details based on the existing information within a single image[14].

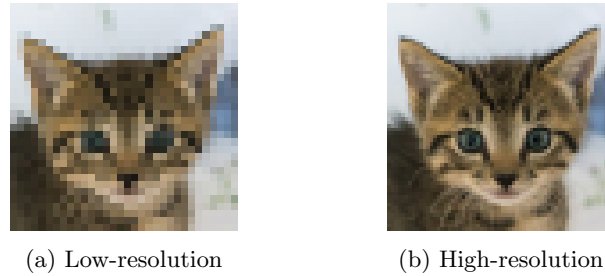


Fig. 1: Example of a low-resolution image compared to a high-resolution sample of the same scene.

In [6], one of the earliest spatial domain super-resolution methods is proposed, this is, a method that works directly on the pixels of an image. Expanding upon this, [7] presents a computationally inexpensive method that makes super-resolution viable in current computational systems. On the other hand, [8] a method is proposed, making use of a pre-established dictionary of images as additional information to perform the process of super-resolution, what is now known as example based super-resolution. Additionally, [9] proposes an evolution on this premise, now constructing a dictionary of patches extracted from a single image.

In this paper, we propose an architecture to increment the resolution of an input image. This is achieved with a neural network of convolutional layers capable of implicitly learning the dictionary of image patches typically used on example based super-resolution, in addition to performing the image reconstruction within the network itself.

The rest of the paper is organized as follows: section 2 reviews the theory of super-resolution and convolutional networks, section 3 describes the proposed architecture, section 4 show the obtained results, and finally, section 5 offers a conclusion.

## 2 Theory

In this section we review the super-resolution and convolutional network theory.

### 2.1 Image super-resolution

As mentioned previously, super-resolution is a process that generates one or more high-resolution images from a set of low-resolution images[10,7].

These images are subsampled and displaced with a subpixel precision. This allows using the information contained on each of these images to obtain a higher resolution one, as shown in Figure 2. If the images were displaced on integer units, they would contain the same information as one another and they wouldn't be useful for the reconstruction of high-resolution images[15].

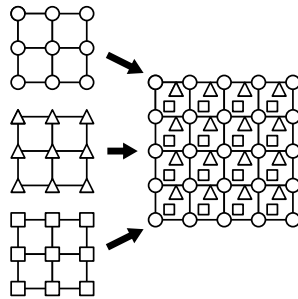


Fig. 2: The subpixel displacement provides the complementary information between multiple low-resolution images, making the reconstruction of a high-resolution image possible[3].

There exist multiple methods for super-resolution: from the earliest ones, based on the displacement and aliasing properties of the Fourier transform[18], to more modern ones, like the previously mentioned example based super-resolution, which performs the process bases on a preset database of image patches generated via the degradation of high-resolution images[9,8], or single image super-resolution, a method based on the similarities between patches extracted from a single image[9].

## 2.2 Convolutional neural networks

A neural network is a system capable of learning to perform a task through the analysis of samples. For example, in the case of image classification, the samples could consist of images of different objects along with the label identifying the type of object they contain.

The principal components of a neural network are the nodes called neurons, which work based on the principles of the perceptron, and are inspired on the working of biological neurons[16]. These neurons are interconnected, allowing them to send signals between themselves, in the form of real numbers. These signals are calculated on function of the sum of the signals a neuron receives multiplied by its weights, values which modify the strength of the input signals and are adjusted during the training phase, making learning possible.

The most common way of modelling a neural network is in the shape of an acyclical graph[17], this is, the neurons are connected in such a way the outputs of certain neurons can act as the input of other ones without creating cycles between them. These connections are organized in layers: sets of neurons whose inputs are connected to the outputs of the previous layer, and whose outputs are connected to the inputs of the next layer[2].

A special type of network is the convolutional neural network, which assume the input signal are meant to represent an image, which allows for the codification of certain characteristics within its architecture, in order to make it more efficient for image processing purposes.

The principal component of the convolutional network is the convolutional layer. Unlike the layers of a traditional network, the convolutional layer's neurons are connected only to a small part of the previous layer. This allows these layers to process data of larger sizes, as are images, without needing to operate on large amounts of weights, reducing their computational requirements.

The way in which each neuron connects only to a small area of its input makes it so the operation of a layer is equivalent to the convolution of a filter conformed of the neuron's weights over the input[13]. As such, convolutional layers are defined by four parameters:  $K$ , the amount of filters or neurons on the layer;  $F$ , the size per side of said filters;  $S$ , the stride, or the amount of pixels the filter moves for each step of the convolution; and  $P$ , the thickness of the zero padding added around the input.

### 3 Methods

As previously mentioned, it's possible to perform the process of super-resolution on a single image, through a technique similar to sample based super-resolution, making use of patches extracted from the image itself[8,9]. We propose a method capable of learning these patches through the use of a convolutional neural network.

#### 3.1 Network architecture

As seen in Figure 3, the network architecture is based on four convolutional layers, in addition to a scaling layer which repeats its input data so its output is twice the size of the input received. The figure shows the parameters used for each layer according to the notation described at the end of section 2. Once trained, this network is capable of generating a high resolution image based on a low resolution one given as input.

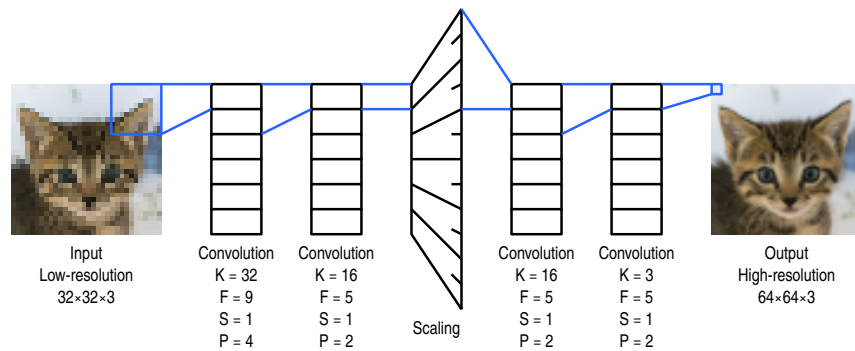


Fig. 3: The proposed architecture contains four convolutional layers, split by a scaling layer in the middle of the network.

This network performs the following tasks[5]:

**Patch extraction** This operation extracts patches from the low-resolution images and stores them as the weights of the neurons on the convolutional layer.

**High resolution-low resolution mapping** This operation makes use of the scaling layer to establish a relationship between the low-resolution patches extracted on the last operation with the high-resolution patches of the target image.

**High resolution image reconstruction** This operation merges the high-resolution patches to reconstruct the high-resolution target image, which is expected to be similar to the image of the original scene.

The neural network was implemented using Keras, a high-level neural network library for the Python language[4], running on top of TensorFlow, a machine learning framework[1].

### 3.2 Training

In order to learn the mapping between low-resolution and high-resolution, the training phase seeks to reduce the loss between the reconstructed high-resolution images and the original samples. The loss function utilized in this network is the Mean Squared Error, defined by:

$$MSE = \frac{1}{n} \sum_{i=1}^n (F(Y; \Theta) - X_i)^2 \quad (1)$$

where  $n$  is the number of training samples,  $F(Y, \Theta)$  is the set of reconstructed high-resolution images,  $X$  is the set of the original high-resolution images and  $Y$  is the set of low resolution ones. The loss between the images is minimized using Adam, a method for stochastic optimization with low computational and memory requirements[12].

The training data consists of a set of 2577 image patches extracted from a set of larger images, mostly captured from architectural or outdoors scenes. The set of low-resolution images is generated from these patches by decimating them, discarding 75% of their content. The high-resolution patches have dimensions of  $256 \times 256$ , while the generated low-resolution ones have half the size per side, at  $128 \times 128$ . Additionally, the images have 3 color channels, corresponding to red, green and blue.

## 4 Results

### 4.1 Learned filters

The weights learned by a convolutional layer can be interpreted as a set of filters that, when convolved with the input data, activate due to the presence of specific

visual characteristics, for example, borders, patterns or colors. An example of this can be seen on Figure 4, a visualization of a set of filters extracted from the first layer of the network.

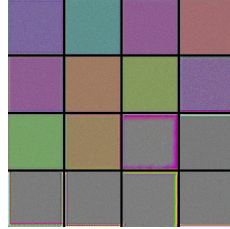


Fig. 4: Filters obtained in the first convolutional layer after performing the training process.

## 4.2 Generated images

As stated, the purpose of this neural network is generating a high-resolution image from a low resolution input. Specifically, the network is capable of doubling the resolution of the input image, and while it was trained using pairs of  $128 \times 128$  and  $256 \times 256$ , in practice, the network can apply the super-resolution process over an image of any size, limited only by the computational and memory capacity of the system it executes on.

Examples of the generated images can be seen on Figure 5 and Figure 6, each showing the original image, the low-resolution sample and the high-resolution image generated by the neural network.

In comparison to the generated images, the low-resolution samples have a more pixelated appearance when scaled at double their size.

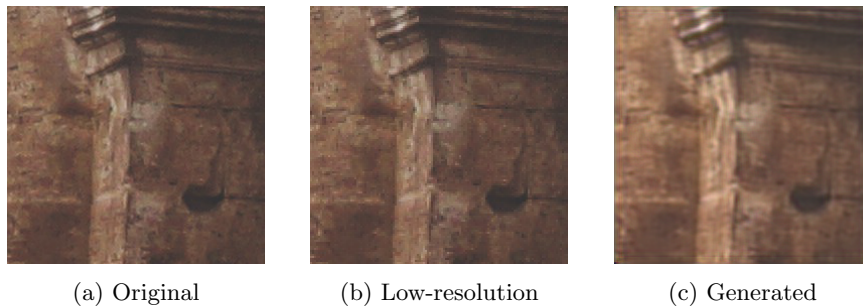


Fig. 5: Original image, low-resolution image and high-resolution generated image from an architectural scene.

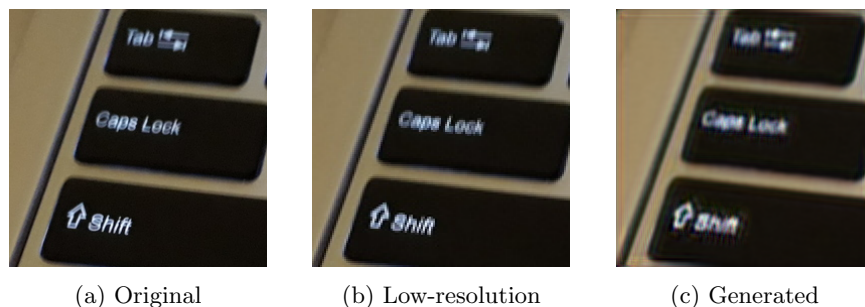


Fig. 6: Original image, low-resolution image and high-resolution generated image from a picture of a keyboard containing text.

Furthermore, sample images were compared using PSNR and SSIM to measure the difference between them. These are methods utilized to approximate the human perceived quality of an image in comparison to another one [19], and are traditionally used for the purpose of measuring the quality of compressed images for data transmission.

PSNR is defined on a scale of decibels, with identical images having a measurement of 0, indicating the absence of noise, and 25 being considered an adequate quality for wireless transmission. On the other hand, SSIM is defined as a real number, with a maximum of 1, which indicates the compared images are identical.

The comparison can be seen in Figure 7 and Table 1, which show a sample of the compared image along with the measurements obtained from the comparison with the generated high-resolution image.

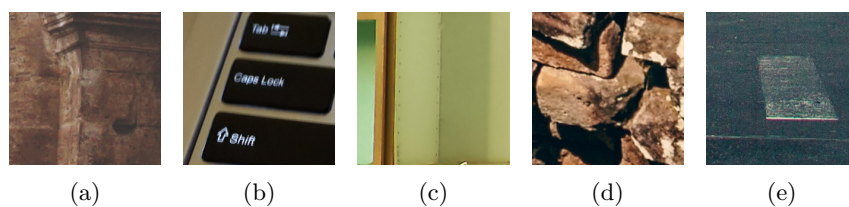


Fig. 7: Images used for the comparison using PSNR and SSIM.

## 5 Conclusion

We have proposed a method for image super-resolution based on deep convolutional networks. The proposed method is capable of learning a mapping between low resolution and high resolution images, while maintaining a simple architecture and achieving an adequate performance for use in real-time applications.

Image	MSE	PSNR	SSIM
(a)	226.39168294	24.58219892	0.86677955
(b)	218.77011108	24.73092373	0.83722339
(c)	301.35599772	23.34000521	0.87755187
(d)	383.97914632	22.28772722	0.80628970
(e)	259.54080200	23.98874718	0.62289394

Table 1: Measurements obtained from the comparison of a set of original samples against the images generated through the neural network.

This deep learning based method could be further refined through experimentation with different filter sizes, an expanded amount of layers, or the use of a different set of training data.

In addition, the incorporation of other types of layers in the proposed network architecture could enable the network to perform other operations along with super-resolution, for example, image denoising, segmentation, feature recognition, etc; or even further augment the resolution of the input image, by stacking the layers already present in this network.

## References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), <https://www.tensorflow.org/>, software available from tensorflow.org
2. Burger, J.T.: A basic introduction to neural networks. <http://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html> (1996)
3. Choi, E., Choi, J., Kang, M.G.: Super-resolution approach to overcome physical limitations of imaging sensors: An overview. *International Journal of Imaging Systems and Technology* **14**(2), 36–46 (2004). <https://doi.org/10.1002/ima.20006>
4. Chollet, F., et al.: Keras. <https://keras.io> (2015)
5. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: *European conference on computer vision*. pp. 184–199. Springer (2014)
6. Elad, M., Hel-Or, Y.: A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. In: *21st IEEE Convention of the Electrical and Electronic Engineers in Israel. Proceedings (Cat. No.00EX377)*. pp. 402–405 (2000). <https://doi.org/10.1109/EEEL.2000.924450>
7. Farsiu, S., Robinson, M.D., Elad, M., Milanfar, P.: Fast and robust multiframe super resolution. *IEEE transactions on image processing* **13**(10), 1327–1344 (2004)
8. Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super-resolution. *IEEE Computer graphics and Applications* **22**(2), 56–65 (2002)



9. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: 2009 IEEE 12th International Conference on Computer Vision. pp. 349–356 (9 2009). <https://doi.org/10.1109/ICCV.2009.5459271>
10. Infognition Co. Ltd.: What is super-resolution? [http://www.infognition.com/articles/what\\_is\\_super\\_resolution.html](http://www.infognition.com/articles/what_is_super_resolution.html) (2009), accessed: 2017-01-31
11. Kim, S.P., Bose, N.K., Valenzuela, H.M.: Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **38**(6), 1013–1027 (6 1990). <https://doi.org/10.1109/29.56062>
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *CoRR* **abs/1412.6980** (2014), <http://arxiv.org/abs/1412.6980>
13. LeCun, Y., Jackel, L., Bottou, L., Cortes, C., Denker, J.S., Drucker, H., Guyon, I., Muller, U., Sackinger, E., Simard, P., et al.: Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural networks: the statistical mechanics perspective* **261**, 276 (1995)
14. Nasrollahi, K., Moeslund, T.B.: Super-resolution: A comprehensive survey. *Machine Vision and Applications* **25**(6), 1423–1468 (2014)
15. Park, S.C., Park, M.K., Kang, M.G.: Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine* **20**(3), 21–36 (2003)
16. Rosenblatt, F.: The perceptron, a perceiving and recognizing automaton. Cornell Aeronautical Laboratory (1957)
17. Schmidhuber, J.: Deep learning in neural networks: An overview. *Neural Networks* **61**, 85 – 117 (2015). <https://doi.org/https://doi.org/10.1016/j.neunet.2014.09.003>
18. Tsai, R.Y., Huang, T.S.: Multi-frame image restoration and registration. *Advances in computer vision and Image Processing* **1**, 317–339 (1984)
19. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**(4), 600–612 (April 2004). <https://doi.org/10.1109/TIP.2003.819861>