

# Super Resolution of DS-2 Satellite Imagery using Deep Convolutional Neural Network

Nour Aburaed<sup>a</sup>, Alavi Panthakkan<sup>a</sup>, Saeed Almansoori<sup>b</sup>, and Hussain Al-Ahmad<sup>a</sup>

<sup>a</sup>University of Dubai, Dubai, United Arab Emirates

<sup>b</sup>Mohammed bin Rashid Space Center, Dubai, United Arab Emirates

## ABSTRACT

Nowadays, Satellite images are used for various analysis, including building detection and road extraction, which are directly beneficial to governmental applications, such as urbanization and monitoring the environment. Spatial resolution is an element of crucial impact on the usage of remote sensing imagery. High spatial resolution means satellite images provide more detailed information. To improve the spatial resolution at the sensor level, many factors are ought to be taken into consideration, such as the manufacturing process. Moreover, once the satellite is launched, no further action can be taken from the perspective of hardware. Therefore, a more practical solution to improve the resolution of a satellite image is to use Single Image Super Resolution (SISR) techniques.

This research proposal deals with the re-design, implementation, and evaluation of SISR technique using Deep Convolutional Neural Network with Skip Connections and Network in Network (DCSCN) for enlarging multispectral remote sensing images captured by DubaiSat-2 (DS-2) and estimating the missing high frequency details. The goal is to achieve high performance in terms of quality, and to test whether training the network using luminance channel only, which is extracted from YCbCr domain, can achieve high quality results. For this purpose, DCSCN is trained, evaluated, and tested using a dataset collected from DS-2. A single low resolution DS-2 image is used to construct its high resolution version by training the model from scratch and fine-tuning its hyper-parameters to produce optimal results. The performance is evaluated using various quality indices, such as Structural Similarity Index Measurement (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Wavelet domain Signal-to-Noise Ratio (WSNR). The performance is compared to other state-of-the-art methods, such as Bi-linear, Bi-cubic, and Lanczos interpolation.

**Keywords:** Image Reconstruction, Super Resolution, Deep Learning, Multispectral, Peak Signal to Noise Ratio (PSNR), Wavelet domain Signal to Noise Ratio (WSNR), Structure Similarity Index Measurement (SSIM).

## 1. INTRODUCTION

Image Super Resolution deals with image processing algorithms that construct High Resolution (HR) images from Low Resolution (LR) ones, an ill-posed problem due to the fact that LR images lack high frequency components that are needed to construct HR images. Generally, there are three approaches known in the literature for SISR problems, which are interpolation, reconstruction, and machine learning techniques. Interpolation methods are considered as the most basic approaches to convert from one resolution to another. The basic principle is to approximate unknown values in between known values using certain kernel types. The process is performed in one dimension; row by row and column by column. The kernel types include Bi-linear, Bi-cubic, B-spline, and other hybrid kernels. For Bi-linear interpolation, the unknown value is estimated from the weighted sum of the four closest intensities. On the other hand, for Bi-cubic interpolation, the unknown value is estimated from the sixteen closest intensities. As for B-spline, the interpolation is a special piece-wise polynomial known as spline. A hybrid interpolation method was proposed by Sun et al.,<sup>1</sup> where the authors propose using covariance-based interpolation method for edges, and iterative curvature based method for smooth areas. According to the literature review presented by Dubey et al.,<sup>2</sup> this method has the best time performance, Peak Signal-to-Noise Ratio (PSNR), as well as visual quality compared to the previously mentioned interpolation methods.

---

Further author information: (Send correspondence to A.A.A.)

N.A.: E-mail: noaburaed@ud.ac.ae

A.P.: E-mail: apantakkan@ud.ac.ae

An interpolation-based SISR technique was proposed by M. Agarwal.<sup>3</sup> High frequency sub-bands are generated from an input image using Lifting Wavelet Transform (LWT), and interpolated with spatial information of input LR image that is generated in parallel. The image is then reconstructed using Inverse Lifting Wavelet Transformation (ILWT) to produce HR image. Currently, interpolation methods are still widely used due to their high computational speed. However, the high frequency details in the resulting images are usually blurry, which creates smoothness effect despite the fact that there is no anti-aliasing.

Other algorithms use learning-based approaches, such as Markov Network,<sup>4-6</sup> Neighborhood Embedding,<sup>7-9</sup> Sparse Representation,<sup>10-13</sup> and Deep Convolutional Neural Networks (DCNN). A comprehensive review for Deep Learning in SISR is presented by Yang et al.<sup>14</sup> In this paper, a DCNN with Skip Connections and Network in Network (DCSCN)<sup>15</sup> is trained from scratch using DS-2 images. The hypothesis is that training the network using Y channel taken from YCbCr color space, as opposed to training it using RGB channels, is more efficient and produces results that are of no less quality to that of using RGB. For RGB color space, the images are directly used to train the network. For YCbCr color space, RGB images are converted to YCbCr domain, and the network is trained using Y channel only in order to generate an upsampled Y channel. As for Cb and Cr channels, their HR versions are generated using Bi-cubic interpolation. Afterward, all Y, Cb, and Cr are merged together to produce a full HR image. The results from training Y channel as opposed to training full RGB images are compared. The results are also compared against commonly used methods, such as Bi-cubic, Bi-linear, and Lanczos interpolation. The rest of the paper is organized as follows: Section 2 explores the dataset from DS-2, Section 3 illustrates DCSCN model and the experiment that will be carried out, Section 4 demonstrates and analyzes the results, and finally, Section 5 concludes the paper and briefly discusses the future work.

## 2. DATASET

DS-2 is the Emirati remote sensing earth observation satellite owned and operated by Mohammed Bin Rashid Space Centre (MBRSC) located in Dubai - UAE. DS-2 provides electro-optical data with a spatial resolution of 1-m for panchromatic imagery and 4-m for multispectral imagery that includes 4 bands; red, green, blue, and Near-infrared (NIR) with a quantization rate of 10 bits.

1730 RGB images of sizes 512×512 were collected from DS-2. The images were selected in such a way that they do not fully consist of water bodies or sand only. Selecting images that contain a variety of structures such as several buildings and roads represent meaningful input in order to train the network effectively.

For all images, the range of values falls between 0-255, therefore, normalization does not need to be applied as a pre-processing step. Calculating the mean value of all the images in the dataset gives a single gray image with a minimum value of 132 and maximum value of 143, and the standard deviation gives a gray image as well with a minimum value of 49 and maximum of 55. This indicates that the dataset has a good variety of images, which is beneficial for the training process.

## 3. DCSCN MODEL AND EXPERIMENTAL SETUP

DS-2 images are used to train and evaluate DCSCN model, which is shown in Figure 2. This model has only 11 layers and, according to the developers of the model,<sup>15</sup> it achieves 10 to 100 times better computational time than other DCNN methods, which makes it efficient to use in devices that cannot handle computationally extensive tasks. The original image itself is used as direct input to the network. The image first goes through feature extraction layers, and then reconstruction layers. Additionally, the image is sent to the reconstruction layers via skip connection in order to learn the features directly from it. A full explanation and analysis of the model can be viewed in the paper authored by Yamanaka et al.<sup>15</sup>

In this paper, this model is trained and evaluated using DS-2 images through tuning its various hyper-parameters, such as number of layers, filters, minimum filters, and scale factor. The experiments and hyper-parameters tuning is tested on two different color spaces, RGB and YCbCr. Human eye is more sensitive to luminance than chrominance. Therefore, this research hypothesizes that training the network using only Y channel, the luminance component, by itself produces results of at least the same quality as using RGB images. This is important because training the network using one channel only saves computational time and resources, especially if there is no degradation in the result's quality. As for the chrominance components, Cb and Cr, they

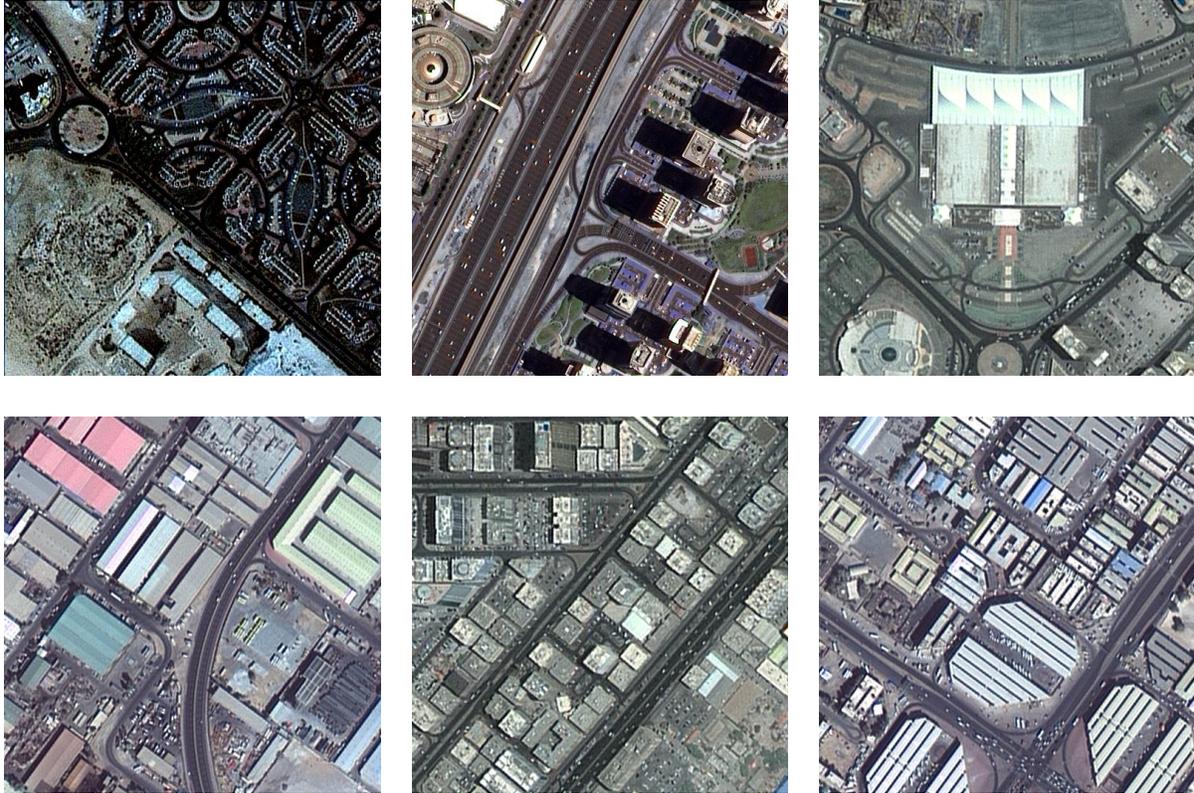


Figure 1. Samples of DS-2 RGB images of size  $512 \times 512$ .

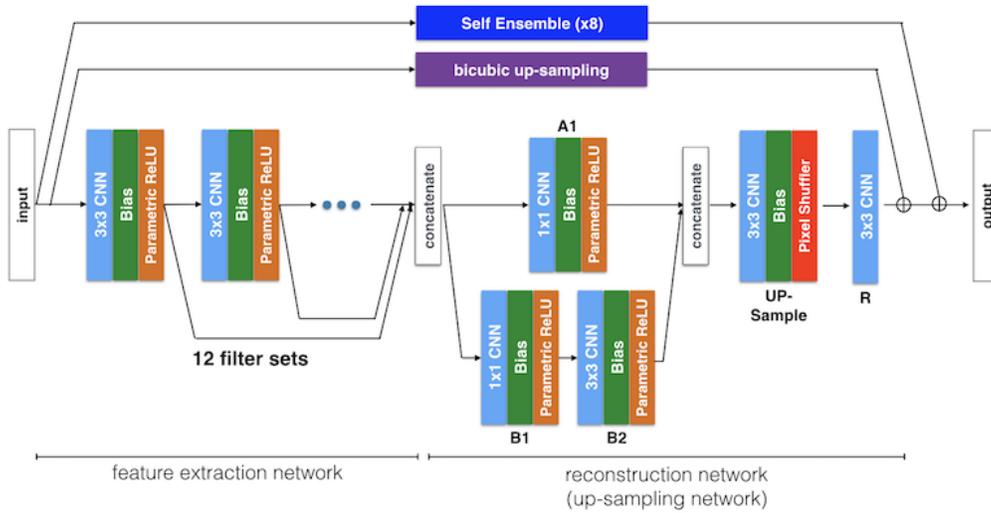


Figure 2. DCSCN Model.<sup>15</sup>

are upsampled using Bi-cubic interpolation, and then combined again with Y channel that was upsampled using DCSCN. There are three cases deployed to test the network and the hypothesis. In Case 1, the number of layers is varied from 2 to 12. In Case 2, the numbers of filters and minimum filters are varied from 4 to 196 and from 4 to 48, respectively. Finally, in Case 3, the scale factor is varied between 2, 3, and 4. These 3 cases are tested for RGB as well as YCbCr. A summary of the experiments is shown in Figure 3.

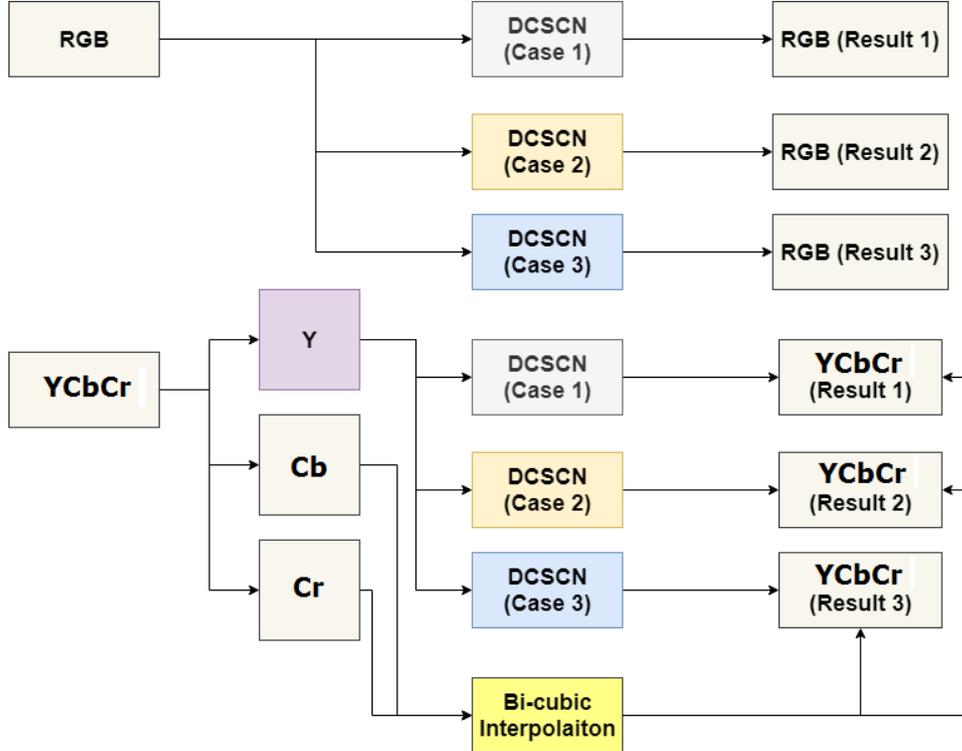


Figure 3. Experimental setup.

#### 4. RESULTS

From the results seen in Tables 1, 2, and 3, it can be concluded that the hypothesis is correct. Training the network over Y channel only, and using Bi-cubic interpolation to enlarge Cb and Cr, gives overall similar results to those of RGB in terms of SSIM and WSNR. In some cases, SSIM for YCbCr prevails over RGB. What is interesting to note is that in Table 3, as the scaling factor becomes larger, training with YCbCr shows more resilience. That is, the gap between RGB and YCbCr results becomes smaller in all metrics, and this is especially clear in SSIM. Moreover, the results from YCbCr also show superiority against commonly used interpolation methods, such as Bi-linear, Bi-cubic, and Lanczos, as seen in Table 4. From Figure 4 It can be seen that, visually, RGB and YCbCr produce the same quality.

Table 1. Results summary of Case 1.

Layers	PSNR		SSIM		WSNR	
	RGB	Y	RGB	Y	RGB	Y
2	30.2707	28.4613	0.8736	0.8835	47.9977	46.6666
4	30.3785	28.5488	0.8756	0.8852	48.1001	46.7867
8	30.4487	28.5885	0.8771	0.8859	48.1293	46.7986
12	30.4713	28.6406	0.8775	0.8873	48.2225	46.8289

#### 5. CONCLUSION AND FUTURE WORK

In this paper, DCSCN has been trained and evaluated using DS-2 images. The proposed approach hypothesizes that by converting to YCbCr, Y channel by itself can be used to train the network, while Cb and Cr can be upsampled using conventional Bi-cubic interpolation. This hypothesis is verified by training the model using

Table 2. Results summary of Case 2.

Filters	Minimum Filters	PSNR		SSIM		WSNR	
		RGB	Y	RGB	Y	RGB	Y
4	4	30.0813	27.9607	0.8696	0.8699	47.7541	46.0091
32	8	30.3785	28.4885	0.8756	0.8835	48.1001	46.6576
96	48	30.5655	28.6412	0.8808	0.8884	48.6004	46.9500
196	48	30.6091	28.6824	0.8816	0.8892	48.6333	46.9778

Table 3. Results summary of Case 3.

Scale	PSNR		SSIM		WSNR	
	RGB	Y	RGB	Y	RGB	Y
2	30.3785	28.4885	0.8756	0.8835	48.1001	46.6576
3	26.5707	25.1010	0.7004	0.7654	33.2006	32.8768
4	24.9340	23.5251	0.5855	0.6437	39.3986	39.0840

Table 4. Comparison between YCbCr results and interpolation methods.

Average	Case 1 (YCbCr)	Case 2 (YCbCr)	Bi-linear	Bi-cubic	Lanczos
PSNR	28.5598	28.4432	25.5146	26.6785	26.7350
SSIM	0.88548	0.8828	0.7792	0.8284	0.8302
WSNR	46.7702	46.6486	38.1541	43.6338	43.4691

RGB, and then using Y channel, and comparing both results. The results from YCbCr are similar qualitatively and quantitatively to those of RGB. Additionally, training with YCbCr approach shows superiority against conventional interpolation methods. In the future, the performance can be boosted further by providing a larger dataset, as 1530 is relatively a small number to train a network from scratch. Furthermore, this approach can be further inspected against other color domains.

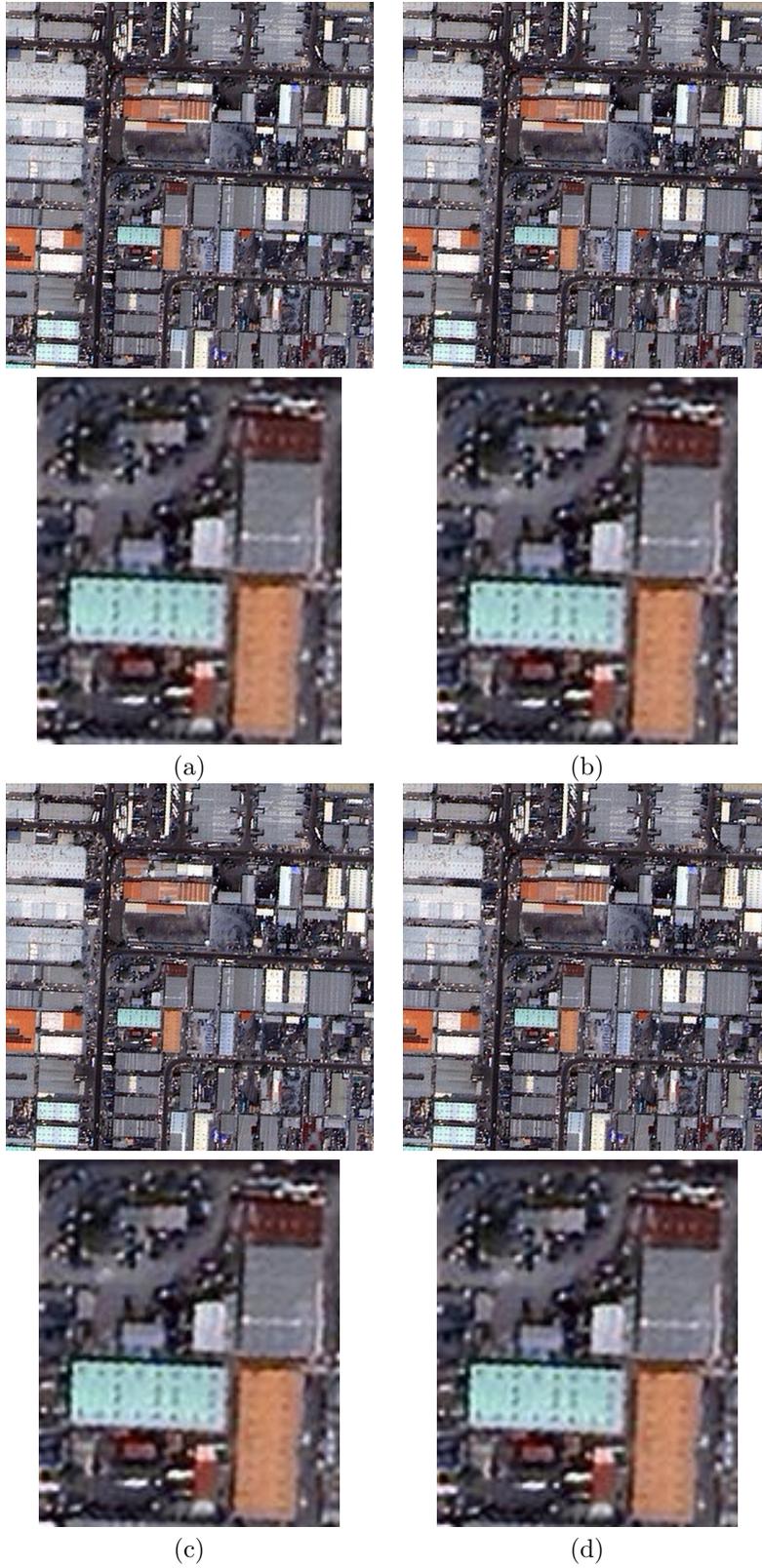


Figure 4. Sample results of RGB and YCbCr. (a) and (b) are the results of Case 1 for RGB and YCbCr, respectively. (c) and (d) are the results of Case 2 for RGB and YCbCr, respectively. Visually, there is no difference between RGB and YCbCr results.

## REFERENCES

- [1] Sun, Y., Sun, Y., Gu, G., Sui, X., and Liu, Y., “Compressive super-resolution imaging based on local and nonlocal regularizations,” *IEEE Photonics Journal* **8** (2016).
- [2] Dubey, S., Sahu, P., and Bazal, S., “Literature review on single image super resolution,” *International Journal of Trend in Scientific Research and Development (IJTSRD)* **2** (2018).
- [3] Agrawal, M. and Dash, R., “Image super-resolution by interpolating high frequency sub-bands of image using surface fitting,” in [*INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING AND COMMUNICATION (ICSC)*], (2014).
- [4] Freeman, W. T., Pasztor, E. C., and Carmichael, O. T., “Learning low-level vision,” *International Journal of Computer Vision* **40** (2000).
- [5] Bishop, C. M., Blake, A., and Marthi, B., “Super resolution enhancement of video,” in [*6th International Conference on Artificial Intelligence and Statistics*], 410–414 (2003).
- [6] Sun, J., Zheng, N.-N., Tao, H., and Shum, H.-Y., “Image hallucination with primal sketch priors,” in [*IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*], (2003).
- [7] Chang, H., Yeung, D.-Y., and Xiong, Y., “Super-resolution through neighborhood embedding,” in [*IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*], (2004).
- [8] Chan, T.-M., Zhang, J., Pu, J., and Huang, H., “Neighbor embedding based super-resolution algorithm through edge detection and feature selection,” *Pattern Recognition Letters* **30**(5), 494 – 502 (2009).
- [9] Zhang, K., Gao, X., Li, X., and Tao, D., “Partially supervised neighbor embedding for example-based image super-resolution,” *IEEE Journal of Selected Topics in Signal Processing* **5**, 230–239 (April 2011).
- [10] Yang, J., Wright, J., Huang, T., and Ma, Y., “Image super-resolution as sparse representation of raw image patches,” in [*IEEE Conference on Computer Vision and Pattern Recognition*], (2008).
- [11] Dong, W., Zhang, L., and Shi, G., “Centralized sparse representation for image restoration,” in [*2011 International Conference on Computer Vision*], 1259–1266 (2011).
- [12] Yang, J., Wang, Z., Lin, Z., Cohen, S., and Huang, T., “Coupled dictionary training for image super-resolution,” *IEEE Transactions on Image Processing* **21**(8), 3467–3478 (2012).
- [13] Yang, S., Wang, M., Chen, Y., and Sun, Y., “Single-image super-resolution reconstruction via learned geometric dictionaries and clustered sparse coding,” *IEEE Transactions on Image Processing* **21**(9), 4016–4028 (2012).
- [14] Yang, W., Zhang, X., Tian, Y., Wang, W., and Xue, J., “Deep learning for single image super-resolution: A brief review,” *CoRR* **abs/1808.03344** (2018).
- [15] Yamanaka, J., Kuwashima, S., and Kurita, T., “Fast and accurate image super resolution by deep CNN with skip connection and network in network,” *CoRR* **abs/1707.05425** (2017).