Deep CNN-Based Super-Resolution Using External and Internal Examples

Jun Young Cheong, *Student Member, IEEE*, and In Kyu Park, *Senior Member, IEEE*

***Abstract*—The external example-driven single image super-resolution (SISR) method that uses a deep convolutional neural network (CNN) has exhibited superior performance as compared to previously developed SISR methods. However, the advantages of jointly using external and internal examples on a deep CNN framework have not been sufficiently investigated. In this letter, we present a novel method for single image super-resolution by exploiting a complementary relation between external and internal example-based SISR methods. The proposed deep CNN model consists of two subnetworks, a global residual network and a self-residual network, to utilize the advantages of both external and internal examples. In contrast with conventional joint SISR methods, the proposed method is the first deep CNN-based SISR method that does not require a retraining process, which tends to be inefficient. The proposed method outperformed existing methods in both quantitative and qualitative evaluations.**

***Index Terms*—Deep convolutional neural network (CNN), external example, internal example, single image super-resolution (SISR).**

I. Introduction

THE goal of single-image super-resolution (SISR) is to restore the fine details in high-resolution (HR) images using a single low-resolution (LR) image. However, due to the limited number of pixels in LR images, the SISR problem is highly ill-posed. Traditionally, interpolation-based (bilinear, bicubic, etc.) and image statistics-based SISR methods with natural image priors have been used to reduce the solution space [1]–[4]. Recent state-of-the-art SISR algorithms mostly use exemplar pairs to learn the nonlinear mapping functions from LR to HR images. Example-driven SISR methods have been developed along two directions, i.e., external example-driven and internal example-driven techniques.

Given LR and HR patch pairs, conventional external example-driven methods train mapping functions using dictionary learning [5], [6], regression [7], [8], and random forest [9]. Lately, a convolutional neural network (CNN) [10]–[12] has exhibited superior performance to conventional SISR methods. While passing through multiple convolutional layers in the CNN model, feature maps are gradually nonlinearized by following the activation function, and finally represent nonlinear features of the mapping functions correctly. These CNN-based SISR methods restore the details of HR images by utilizing a large receptive field, which is beneficial for preserving the structure of LR images.

On the contrary, internal example-driven methods restore HR images by increasing the image scale [13]–[16]. The restoration approach is justified by the self-similarity property, whereby small patches of natural images often occur repetitively at various scales in the same image.

Zhu *et al.* exploited deformable patches by optical flow to address the problem of the limited size of the self-exemplar HR patch space [17]. Huang *et al.* synthesized self-exemplar patches by considering multiple perspective distortions in an input image [18]. These internal example-driven methods can restore fine details recurring at various scales in the input image. However, since the size of self-exemplar patches is small, a tiny HR structure is challenging to restore.

In order to overcome the limitations of single example-driven SISR methods, Wang *et al.* proposed mapping LR to HR images using both internal and external example-driven dictionaries [19]. They subsequently tried to use self-similar patches to fine-tune of a pretrained CNN model [20]. However, since both methods require retraining, i.e., fine-tuning the internal dictionary and the CNN model for each test input, the methods have limited applicability.

In this letter, we propose a novel deep CNN that jointly uses external and internal examples to solve the SISR problem. The proposed method exploits a complementary relation between external and internal example-based SISR methods to recover the fine details of an HR structure. Compared with existing SISR methods [19], [20], the proposed method is the first deep CNN-based SISR method that does not require a retraining process and hence avoids unreliability of self-tuned data. Moreover, it is easily applicable to various internal example-driven SISR methods to improve performance.

II. Example-Driven SISR

A. Internal Example-Driven SISR

1) Strength: Since HR patches are searched for in the self-exemplar HR patch space, the spatial search space is not limited to specific regions. Therefore, internal example-driven SISR methods are robust against the repetitive local HR structure

Manuscript received April 12, 2017; revised June 14, 2017; accepted June 20, 2017. Date of publication June 28, 2017; date of current version July 12, 2017. This work was supported by the National Research Foundation of Korea Grant funded by the Korea Government (MSIP) (No. NRF-2016R1A2B4014731). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Matteo Naccari. (*Corresponding author: In Kyu Park.*)

The authors are with the Department of Information and Communication Engineering, Inha University, Incheon 22212, South Korea (e-mail: cheongjunyoung@gmail.com; pik@inha.ac.kr).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2017.2721104

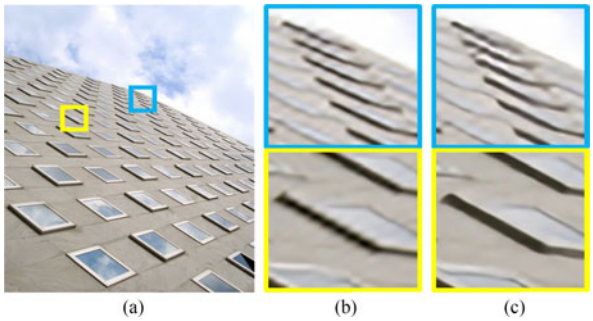

Fig. 1. Example of HR image and restored HR image regions. (a) Input image. (b) Result of internal example-driven SISR [18]. (c) Result of external example-driven SISR [12].

over the entire input image and across several image scales [see the blue box in Fig. 1(b)].

2) Weakness: The restoration approach based on a small self-exemplar HR patch often distorts fine HR structure. This stems from the difficulty in recognizing large HR structures by observing small overlapping HR patches only. This weakness is observed as a step artifact around the edges of large HR structures. Further, the oversharpening artifact is often observed due to iterative back-projection [21] [see the yellow box in Fig. 1(b)].

B. External Example-Driven SISR

1) Strength: Since external example-driven SISR methods provide a large size of receptive field when utilizing a deep CNN model, an LR image of large size is observed to restore a single pixel of an HR image. It is beneficial to model nonlinear mapping functions more precisely. Moreover, the residual training approach that restores residual images, subtracted from HR to interpolated LR images, shows powerful performance in restoring high-frequency components [12]. This strength yields clear edges around large HR structure [see the yellow box in Fig. 1(c)].

2) Weakness: Although example-driven SISR methods using deep CNN have a large receptive field, it is insufficient to fully exploit repetitive local HR information over the entire input image. As a result, high-frequency components in small HR structures are lost easily [see the blue box in Fig. 1(c)].

III. Proposed CNN Model

The proposed CNN model consists of two parallel subnetworks as shown in Fig. 3. The deeper subnetwork is the global residual network whereas the other is a self-residual network. As input to the proposed CNN model, we use an interpolated LR image as a guide image for the HR structure and an internal example-driven HR image (self-HR image) to extract the fine details of the local HR structure.

Given a training dataset $\{X^i, I^i_{\mathrm{HR}}\}_{i=1}^{K}$, the proposed CNN model M_p is optimized to predict the residual image using the input $\mathbf{X}$: $\{I_{\mathrm{LR}}, I_{\mathrm{self-HR}}\}$ by minimizing a cost function defined as

$$M_p = \min_M \frac{1}{2} \left\| I_{\mathrm{HR}} - (M(\mathbf{X}) + I_{\mathrm{LR}}) \right\|^2 . \tag{1}$$

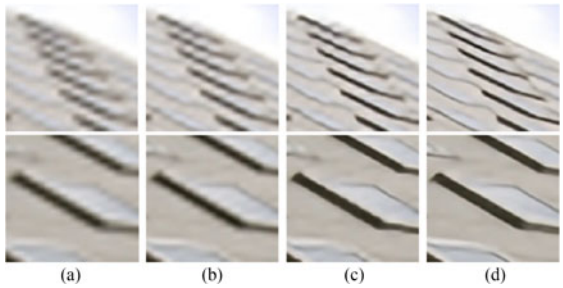

Fig. 2. Analysis of the proposed network. (a) Interpolated LR image. (b) Results using the self-residual network only. (c) Results of the dual input VDSR model. (d) Results of the proposed method.

A. Self-Residual Network

The purpose of a self-residual network is to transfer local details in the self-HR image. Inspired by [22], the proposed self-residual network has three pixel-encoding layers (a 1×1 convolutional layer and a ReLu activation function) to transfer information in the encoded pixels. Mathematically, after passing through the ith pixel-encoding layer, the output feature maps can be written as follows:

$$F_i = \max(0, W_i \otimes F_{i-1} + B_i) \tag{2}$$

where F_i denotes the output feature map, and W_i and B_i denote the convolutional filters and bias in the ith pixel-encoding layer, respectively. The output feature maps are encoded to preserve local HR details without spatial interference from neighboring pixels.

B. Global Residual Network

The global residual network is proposed here to restore fine HR structure and compensate for the distortion error in the self-HR image. Following discussion in Section II-B, we build the global residual network with a large receptive field by serializing residual blocks.

We apply skip-connection and batch normalization layers to stabilize the gradient values of the convolutional filters during training. For each residual block, the input feature maps pass through two 3×3 convolutional layers and the number of output feature maps is maintained at 64. Following N residual blocks and the last 3×3 convolutional layer, the receptive field of the global residual network increases from 3×3 to $(4N + 5) \times (4N + 5)$. The structure of the residual blocks is shown in Fig. 3.

C. Analysis of the Proposed CNN Model

To analyze the contribution from each subnetwork in restoring the HR image, we activate only the self-residual network to exclusively reconstruct it. For the regions of repetitive HR structure, a pixel-wise encoded HR structure of the self-HR image is transferred to the HR image to remove the ambiguity in the small interpolated LR structure [see Fig. 2(a) and (b)]. As a result, the global residual network focuses on training high-frequency components of less ambiguous structure. By exploiting the cooperation between the subnetworks, the proposed

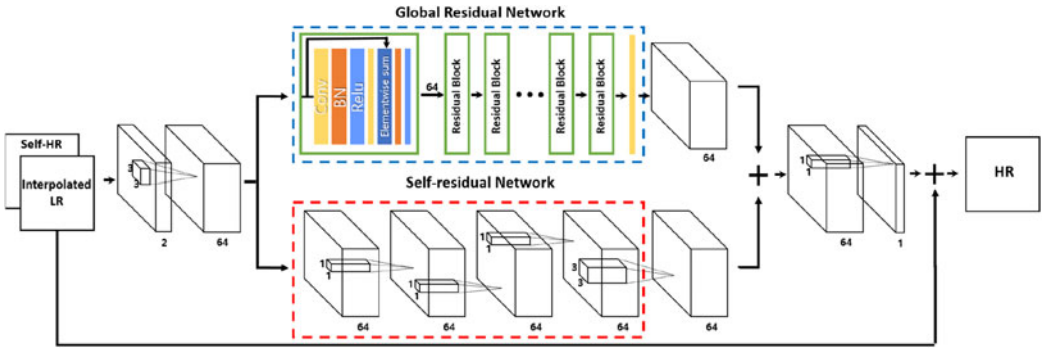

Fig. 3. Proposed deep CNN-based SISR model.

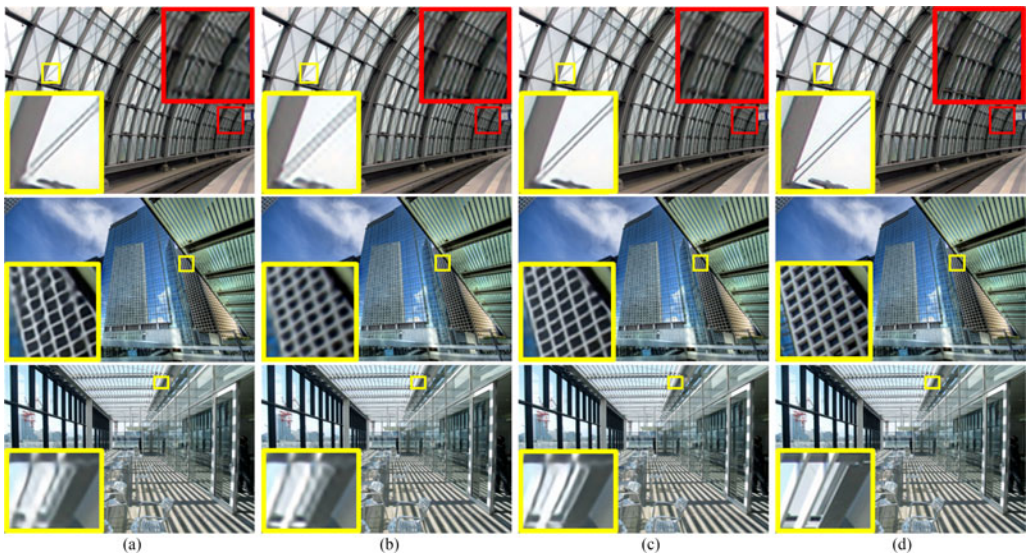

Fig. 4. Super-resolution results on *Urban100* dataset with a scale factor of $\times 4$. We recommend a computer display. (a) Results of VDSR [12]. (b) Results of SelfExSR [18]. (c) Results of the proposed method. (d) Ground truth.

method maintains HR structure of both external and internal examples as shown in Fig. 2(d).

As an alternative to the proposed network, one might consider a dual input very deep super-resolution (VDSR) model which is similar to the original VDSR model [12] except that it has the same first input layer as the proposed model. In Fig. 2(c), the dual input VDSR model exploits self-HR images to refine the repetitive local patterns. However, it is observed that it converges to nearly identity mapping. In addition, it produces artifacts in the regions with strong edges which is commonly observed in self-HR images.

D. Training

The proposed CNN model is trained using subimages cropped from 291 natural images [9] with data augmentation. For the self-HR image, the results of SelfExSR are used with default parameters in [18].

The training data are resized to multiple scales ($\times 2$, $\times 3$, and $\times 4$) to train the proposed CNN model, which can restore images at multiple scales. Based on back-propagation [23], the minibatch gradient descent method is used with a batch size of 64, a momentum of 0.9, and an L2-weighted decay parameter of 0.0001. We adopt gradient clipping [12] for faster training. The learning rate decreased from 0.01 by a factor of 10 every 10 epochs. It is kept constant after 40 epochs and training finishes after 60 epochs.

IV. Experimental Results

We evaluate the performance of the proposed algorithm and compare it against several state-of-the-art SISR methods. The HR results of the proposed method are reconstructed on an Intel Core i5 CPU (3.2 GHz) with 12 GB RAM and an NVIDIA Titan GPU. The evaluation is performed on datasets of various natural [24]–[26] and urban [18] scenes. Moreover, the *Sun-Hays80* [27] uses for qualitative evaluation with scale factor of $\times 8$.

We set nine residual blocks in order for the global residual network to have a receptive field of the same size as VDSR [12] (41×41) and predict only single-luminance channels. We use the results of [15], [16], and [18] from Huang *et al.*'s webpage [28] as benchmark. Unlike the original papers [10], [12], we perform the quantization (floating point to integer conversion) of pixel values in LR image generation for more realistic representation of real-world LR images as well as for fair comparison with other methods. The proposed method is implemented using the MatConvNet deep learning toolbox [29], and takes approximately 10 hours to train. We also show that the proposed method can be used with other internal example-driven SISR methods to improve performance.

Fig. 5. Super-resolution results on *Sun-Hays80* dataset with a scale factor of ×8. (a) Result of VDSR [12]. (b) Result of SelfExSR [18]. (c) Result of the proposed method. (d) Ground truth.

TABLE I
QUANTITATIVE EVALUATION (IN TERMS OF PSNR/SSIM) CONDUCTED ON *Set5*, *Set14*, *BSD100*, AND *Urban100*

	Set5			*Set14*			*BSD100*			*Urban100*	
	×2	×3	×4	×2	×3	×4	×2	×3	×4	×2	×4
SRCNN [10]	36.28 / 0.951	32.37 / 0.903	30.08 / 0.853	32.00 / 0.901	28.90 / 0.812	27.13 / 0.740	31.11 / 0.884	28.20 / 0.779	26.70 / 0.702	28.65 / 0.891	24.14 / 0.705
VDSR [12]	37.42 / 0.957	33.62 / 0.920	31.33 / 0.883	32.95 / 0.911	29.73 / 0.831	27.97 / 0.766	31.84 / 0.894	28.80 / 0.796	27.26 / 0.724	30.21 / 0.915	25.16 / 0.751
Glasner [15]	35.43 / 0.945	31.10 / 0.881	28.84 / 0.821	31.41 / 0.888	28.21 / 0.793	26.43 / 0.716	30.28 / 0.862	27.06 / 0.737	26.17 / 0.675	27.85 / 0.871	23.58 / 0.674
Sub-band [16]	N/A	N/A	N/A	N/A	N/A	N/A	30.73 / 0.877	27.88 / 0.771	26.60 / 0.702	28.34 / 0.882	24.19 / 0.712
SelfExSR [18]	36.50 / 0.954	32.62 / 0.909	30.33 / 0.862	32.23 / 0.904	29.16 / 0.820	27.40 / 0.752	31.18 / 0.886	28.30 / 0.784	26.85 / 0.711	29.38 / 0.903	24.82 / 0.739
Proposed + Glasner	36.93 / 0.954	32.68 / 0.908	30.23 / 0.861	32.57 / 0.905	29.20 / 0.817	27.42 / 0.747	31.39 / 0.885	28.05 / 0.771	26.73 / 0.701	29.75 / 0.908	24.43 / 0.722
Proposed + Sub-band	N/A	N/A	N/A	N/A	N/A	N/A	31.82 / 0.894	28.66 / 0.791	27.19 / 0.722	30.37 / 0.917	25.01 / 0.747
Proposed + SelfExSR	37.48 / 0.958	33.78 / 0.922	31.42 / 0.884	32.92 / 0.910	29.78 / 0.830	28.02 / 0.766	31.91 / 0.895	28.85 / 0.797	27.31 / 0.725	30.58 / 0.919	25.38 / 0.759

Red and blue color indicate the best and the second-best performance, respectively.

A. Qualitative Comparison

In Fig. 4, the test images shown contain various scales of repetitive HR structure, which is helpful in evaluating the strengths and weaknesses of both external and internal examples-driven SISR methods. Because the VDSR is designed to restore high-frequency components of HR structure, the strength is visually noticeable at edges around the large HR structure. However, due to ambiguities in small HR structures, the weakness is often observed as shown in Fig. 4(a).

In case of SelfExSR, the ambiguities in the small HR structure are clarified by searching for HR information in the self-exemplar HR patch space. However, restoring a large HR structure using small self-exemplar HR patches causes step artifacts around the edges, as shown in Fig. 4(b). On the contrary, in addition to simply compensating the weaknesses of these SISR methods, the proposed method improves performance by utilizing the collaborative strengths of SISR methods. As a result, the proposed method generate finer details of the HR structure.

Furthermore, we evaluate the performance at a scale factor (×8). Although the proposed method is not trained for a scale factor greater than ×4, we apply it during the restoration process of SelfExSR, which restores an HR image by increasing image scales gradually by a scale factor of 1.25. During the gradual HR restoration, the proposed method refines the HR image until the image scale factor reaches ×2, ×4, and ×8. Similar to the above approach, VDSR is applied to increase the HR image scale twice, and this process is repeated until it reaches an image scale factor of ×8. As shown in Fig. 5, the proposed method restores the shape of the fence more clearly than the other methods [12], [18].

B. Quantitative Comparison

The performance is evaluated quantitatively and compared with several state-of-the-art external-example based [10], [12] and internal-example based SISR methods [15], [16], [18]. Table I shows the results on *Set5* [24], *Set14* [25], *BSD100* [26], and *Urban 100* [18] dataset.

The proposed method yields the best quantitative results, slightly better than those of VDSR, which has been known as the best algorithm to-date. Note that, the *Urban 100* dataset contains many repetitive HR structures at various scales, which provide the best condition for the proposed method.

Furthermore, we employ different internal example-driven SISR methods [15], [16] to generate the self-HR image. Note that we do not train the proposed CNN model again using corresponding internal example-driven SISR methods. Interestingly, although the proposed method is trained using only self-HR images of SelfExSR, the performance of both Glasner [15], and sub-band [16] is improved by the proposed method. On the *BSD100* and the *Urban100* datasets, the performance of both SISR methods is significantly enhanced.

While external example-driven SISR methods run fast to produce test images, the proposed method depends on the internal example-driven HR image which requires additional processing time. However, previous joint example-driven SISR methods [19] [20] require retraining of the internal dictionary or self-tuning, which requires a large amount of computation.

V. CONCLUSION

In this letter, we showed the strengths and weaknesses of external and internal SISR methods for large or repetitive small HR structures. Based on the complementary relation between these SISR methods, we proposed a novel SISR method using deep CNN that exhibited the best performance in both qualitative and quantitative comparison with state-of-the-art SISR methods. Unlike in conventional joint SISR methods [19] [20], additional retraining process is not required for the proposed method, and it is applicable to other internal SISR methods to improve performance.

REFERENCES

[1] J. Huang and D. Mumford, "Statistics of natural images and models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1999, pp. 541–547.

[2] K. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1127–1133, Jun. 2010.

[3] J. Sun, Z. Xu, and H. Shum, "Image super-resolution using gradient profile prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[4] R. Fablet and F. Rousseau, "Missing data super-resolution using non-local and statistical priors," in *Proc. IEEE Int. Conf. Image Process.*, 2005, pp. 676–680.

[5] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[6] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jan. 2011.

[7] C.-Y. Yang and M.-H. Yang, "Fast direct super-resolution by simple functions," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 561–568.

[8] R. Timofte, V. D. Smet, and L. V. Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. Asian Conf. Comput. Vis.*, 2014, pp. 111–126.

[9] S. Schulter, C. Leistner, and H. Bischof, "Fast and accurate image upscaling with super-resolution forests," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3791–3799.

[10] C. Dong, C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.

[11] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1874–1883.

[12] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.

[13] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, vol. 30, no. 2, p. 12, 2011.

[14] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2004, pp. 275–282.

[15] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 349–356.

[16] A. Singh and N. Ahuja, "Super-resolution using sub-band self-similarity," in *Proc. Asian Conf. Comput. Vis.*, 2014, pp. 552–568.

[17] Y. Zhu, Y. Zhang, and A. L. Yuille, "Single image superresolution using deformable patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2917–2924.

[18] J.-B. Huang, A. Singh, and N.Ahuja,"Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5197–5206.

[19] Z. Wang, Y. Yang, and Z. Wang, "Learning super-resolution jointly from external and internal examples," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4359–4371, Nov. 2015.

[20] Z. Wang *et al.*, "Self-tuned deep super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 1–8.

[21] M. Irani and S. Peleg, "Improving resolution by image registration," *Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, 1991.

[22] D. Cho, Y. W. Tai, and I. Kweon, "Natural image matting using deep convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 626–643.

[23] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[24] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 1–10.

[25] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparserepresentations," in *Proc. Int. Conf. Curves Surf.*, 2012, pp. 711–730.

[26] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2001, pp. 416–423.

[27] L. Sun and J. Hays, "Super-resolution from internet-scale scene matching," in *Proc. IEEE Int. Conf. Comput. Photography*, 2012, pp. 1–12.

[28] 2017. [Online]. Available: https://github.com/jbhuang0604/SelfExSR

[29] A. Vedaldi and K. Lenc, "MatConvnet—Convolutional neural networks for MATLAB," in *Proc. ACM Int. Conf. Multimedia*, 2015, pp. 689–692.