Aerial Image Super Resolution via Wavelet Multiscale Convolutional Neural Networks

Tingwei Wang, Wenjian Sun, Hairong Qi, and Peng Ren

***Abstract*— We develop an aerial image super-resolution method by training convolutional neural networks (CNNs) with respect to wavelet analysis. To this end, we commence by performing wavelet decomposition to aerial images for multiscale representations. We then train multiple CNNs for approximating the wavelet multiscale representations, separately. The multiple CNNs thus trained characterize aerial images in multiple directions and multiscale frequency bands, and thus enable image restoration subject to sophisticated culture variability. For inference, the trained CNNs regress wavelet multiscale representations from a low-resolution aerial image, followed by wavelet synthesis that forms a restored high-resolution aerial image. Experimental results validate the effectiveness of our method for restoring complicated aerial images.**

***Index Terms*— Convolutional neural networks (CNNs), super resolution, wavelet analysis.**

I. Introduction

AERIAL images are widely used in both military and civil fields, such as military reconnaissance, environment monitoring, and weather forecast. However, aerial images captured by normal imaging devices usually have limited resolution that hardly satisfies requirements in operational tasks. We address this limitation and present a new super-resolution method that restores a high-resolution aerial image from one single low-resolution image.

One straightforward strategy for single image super resolution is interpolation such as bicubic interpolation. Though efficient, interpolation methods do not consider image characteristics and can hardly achieve accurate super resolution. Schultz and Stevenson [1] developed a Bayesian restoration model that utilized both the spatial and temporal information to obtain high-resolution images. Nguyen *et al*. [2] proposed an efficient approximation technique based on the Lanczos algorithm and the Gauss quadrature theory to estimate the point spread function parameters for super resolution. Recently, sophisticated machine learning methods have been comprehensively exploited for image super resolution. Dong *et al*. [3] described how to use a convolutional neural network (CNN) to perform an end-to-end mapping between low-resolution and high-resolution images. Lei *et al*. [4] made a preliminary study to bridge deep learning and multiscale analysis by employing an local-global-combined network (LGCNet) for capturing restoration features at both local and global levels in spatial domain. Wei *et al*. [5] formed a very deep residual neural network to improve the accuracy of multispectral image pan-sharpening. Li *et al*. [6] utilized a spectral mixture analysis and spatial–spectral group sparsity for the hyperspectral image super resolution. Though these methods exploited sophisticated machine learning schemes in the original spatial or spectral domain, they did not consider the image characteristics in the transformed domains such as frequency bands.

Aerial images exhibit a strong culture variability such that the image textures represent changes at different directions with various frequency characteristics. It is observed that image representations at different frequencies reflect different features for super resolution, where high-frequency details play an important role [7]. This motivates us to train multiple models at different frequency bands, which characterize culture variation at different scales, for super resolution. Wavelet analysis operates in terms of filter banks which can represent images in multiscale frequency bands. In these scenarios, two recent wavelet-based methods [8], [9] share the similar terminologies. However, our framework differs from [8] by exploiting multiple parallel CNNs and differs from [9] by employing wavelet synthesis for complete reconstruction. Therefore, there exist significant differences between our framework and the studies in [8] and [9] in terms of model structure, feature representation and learning spaces. In contrast to most deep learning-based super-resolution strategies [e.g., super-resolution CNN (SRCNN) and LGCNet that process images in the original spatial domain, we train multiple CNNs in multiscale frequency bands generated by the wavelet analysis, for the purpose of restoring frequency features in different directions. Compared with the edge preserving wavelet-based super resolution [10] that just considers the local details, we characterize super-resolution features at different frequency scales via multiple CNNs, which have greater representational power. Specifically, our method not only captures high-frequency local changes but also characterizes low-frequency global layouts.

II. Wavelet Multiscale Convolutional Neural Networks for Aerial Image Super Resolution

This section describes how to train wavelet multiscale CNNs (WMCNNs) for characterizing culture variations at

Manuscript received October 12, 2017; revised January 1, 2018 and February 2, 2018; accepted February 21, 2018. Date of publication March 15, 2018; date of current version April 20, 2018. This work was supported in part by the National Natural Science Foundation of China under Project 61671481, in part by the Qingdao Applied Fundamental Research under Project 16-5-1-11-jch, and in part by the Fundamental Research Funds for the Central Universities under Project 18CX05014A. *(Corresponding author: Peng Ren.)*

T. Wang, W. Sun, and P. Ren are with the College of Information and Control Engineering, China University of Petroleum (East China), Qingdao 266580, China (e-mail: wtw_upc@163.com; swj_upc@163.com; pengren@upc.edu.cn).

H. Qi is with the College of Engineering, The University of Tennessee, Knoxville, TN 37996 USA (e-mail: hqi@utk.edu).

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/LGRS.2018.2810893

different scales, and how to restore an aerial image via the trained WMCNNs.

A. Wavelet Representations of High-Resolution Aerial Images as Multiscale Regression Features

We perform a multiscale analysis of an aerial image via wavelet decomposition. We implement wavelet decomposition in terms of filter banks consisting of biorthogonal high-frequency pass filters and low-frequency pass filters. Let $\mathbf{L}$ denote a low-frequency pass filter matrix whose columns represent coefficients of a low-frequency pass filter, and $\mathbf{H}$ denote a high-frequency pass filter matrix whose columns represent coefficients of a high-frequency pass filter. Explanations of the matrices $\mathbf{H}$ and $\mathbf{L}$ for specific wavelet decompositions are referred to [11]. We use one original high-resolution image as the initial level representation $\mathbf{C}_0$ and perform the wavelet decomposition as follows:

$$\begin{aligned} \mathbf{C}_j &= (\mathbf{L}^\dagger \mathbf{C}_{j-1}\mathbf{L})_{s\downarrow}, \quad \mathbf{D}_j^h = (\mathbf{L}^\dagger \mathbf{C}_{j-1}\mathbf{H})_{s\downarrow} \\ \mathbf{D}_j^v &= (\mathbf{H}^\dagger \mathbf{C}_{j-1}\mathbf{L})_{s\downarrow}, \quad \mathbf{D}_j^d = (\mathbf{H}^\dagger \mathbf{C}_{j-1}\mathbf{H})_{s\downarrow} \end{aligned} \tag{1}$$

where j indicates the decomposition level and $s_\downarrow$ denotes the operation of downsampling to $\frac{1}{s}$ of the original resolution. The recursive downscaling decomposition (1) characterizes an aerial image with respect to multiple spatial ratios, which favors comprehensive distant observations based on aerial images. $\mathbf{C}_j$, $\mathbf{D}_j^h$, $\mathbf{D}_j^v$, and $\mathbf{D}_j^d$ represent the overall low frequency, horizontal high frequency along with vertical low frequency, vertical high frequency along with horizontal low frequency, and overall high-frequency characteristics of the last level representation $\mathbf{C}_{j-1}$, respectively. Therefore, the wavelet decomposition results in multiscale presentations for an aerial image in terms of spatial ratios, orientations, and frequency ranges. We use the wavelet multiscale representations as regression features for training multiple CNNs for super resolution in Section II-B.

B. Regressing Wavelet Multiscale Features From Low-Resolution Images Based on Convolutional Neural Networks

We train multiple CNNs for regressing wavelet multiscale features from low-resolution images. An example of the training framework based on one-level wavelet decomposition is illustrated in Fig. 1 (left dashed line box). The exemplary framework trains four CNNs. Each CNN processes the low-resolution image layer by layer and the final layer aims to regress one of the multiscale representations of the high-resolution image described in Section II-A.

For constructing each individual CNN, the structure of the state-of-the-art SRCNN [3] is adopted. Given a low-resolution aerial image I_L downsampled from high-resolution image as the input, the output of the nth convolutional layer is

$$f_n(I_L, W_n, b_n) = \sigma(W_n * f_{n-1}(I_L) + b_n) \tag{2}$$

where W_n and b_n are network weights and bias for training, respectively. σ is a rectified linear function [e.g., $\max(0, x)$], which enables the fast convergence for CNNs.

For generating features that most regress the wavelet multiscale representation, each CNN is penalized by a loss function for measuring the distance between the representation generated by the CNN from the low-resolution image and the representation from the wavelet decomposition of the corresponding high-resolution image. The loss function for the top CNN in Fig. 1 is given as

$$\ell = \frac{1}{2K}\sum_{k=1}^{K} || \mathbf{C}(k) - \widetilde{\mathbf{C}}(k) ||_2^2 \tag{3}$$

where k is the pixel index of the wavelet multiscale representation, and $\mathbf{C}$ is the wavelet multiscale representation that preserves the two direction smoothing characteristics of a high-resolution aerial image, $\widetilde{\mathbf{C}}$ is the representation generated by the CNN from the low-resolution image I_L. In Fig. 1, the top patch in the solid line box is one CNN-generated representation for regressing $\mathbf{C}$. The loss functions for the rest three CNNs can be given by replacing $\mathbf{C}$ in (3) by $\mathbf{D}^h$, $\mathbf{D}^v$, and $\mathbf{D}^d$, separately.

The multiple CNNs are trained through back propagation for minimizing their loss functions separately such that each CNN is trained to learn features characterized by the associated wavelet representation. The multiple CNNs thus trained capture the multiscale image features in different directions and multiple frequency bands.

C. Super Resolution via Wavelet Multiscale Convolutional Neural Networks

We commence super resolution by feeding a low-resolution image into the multiple CNNs, separately. We then perform a wavelet synthesis on the CNN-generated representations from the jth level to the $(j-1)$th level as follows:

$$\widetilde{\mathbf{C}}_{j-1} = \mathbf{L}(\widetilde{\mathbf{C}}_j)_{s\uparrow}\mathbf{L}^\dagger + \mathbf{H}(\widetilde{\mathbf{D}}_j^h)_{s\uparrow}\mathbf{L}^\dagger + \mathbf{L}\,(\widetilde{\mathbf{D}}_j^v)_{s\uparrow}\mathbf{H}^\dagger + \mathbf{H}(\widetilde{\mathbf{D}}_j^d)_{s\uparrow}\mathbf{H}^\dagger. \tag{4}$$

The tidal symbol indicates that $\widetilde{\mathbf{C}}_j$, $\widetilde{\mathbf{D}}_j^h$, $\widetilde{\mathbf{D}}_j^v$, and $\widetilde{\mathbf{D}}_j^d$ are CNN-generated representations, which are different from those representations in (1) obtained from the wavelet decomposition. As one CNN-generated representation is endowed with the image features characterized in one frequency band with certain orientations specified the wavelet decomposition, the wavelet synthesis intrinsically ensembles the land cover and object diversity from multiscale frequency bands and directions and achieves an effective super resolution. In Fig. 1, the right dashed box describes a one-level wavelet synthesis example for super resolution, in which four CNN-generated representations are synthesized, restoring a high-resolution aerial image.

D. Observations

Two key factors enabling the effectiveness of CNNs are the local processing and the multiscale analysis, which highly resemble the human brain receptive fields. The local processing is effected by training network weights that operate local filtering over the one whole image. The multiscale analysis is implemented in terms of pooling that involves downsampling

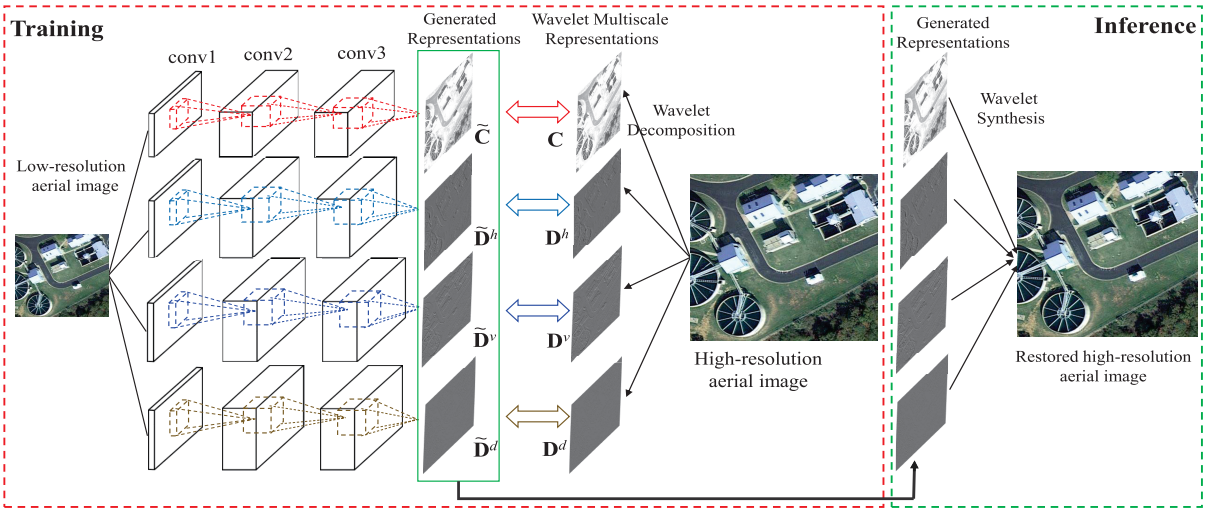

Fig. 1. Proposed wavelet multiscale CNNs.

as the main manipulation. However, the state-of-the-art CNN-based super-resolution methods such as SRCNN tend to just involve local processing but discard multiscale analysis. This is because, in contrast to the feature condensation required for accurate pattern recognition, super-resolution tasks aim to upscale image feature representations. Therefore, straightforward downsampling operations evidently contradict the upscaling goal of super resolution. In order to fully exploit the representational power of CNNs in terms of both the local processing and multiscale analysis, we propose to train multiple CNNs without pooling based on image wavelet representations.

The CNN network weights are obtained through training and the CNNs conduct supervised filtering on aerial image representations. On the other hand, the wavelet decomposition and synthesis employ off-the-shelf wavelet filters and perform unsupervised filtering on the multiscale representations. Thus, they enrich the local processing effects in addition to filtering performed by the CNN network weights. Such enhanced local processing tends to capture more comprehensive land cover and object characteristics. The downsampling operations in the wavelet decomposition form an effective substitute of the pooling operations discarded from the existing CNN-based super-resolution methods. Furthermore, the wavelet analysis involves multiple frequency band filtering with different orientations, which forms a more general multiscale analysis. Additionally, the multiscale representations not only improve the detailed visual characterization but also favor holistic descriptions for aerial images. Finally, the wavelet synthesis upscales the multiscale representations with filtering revisions conjugate to the wavelet decomposition, which results in the super resolution.

For multilevel analysis, the four CNNs trained at one level can be reused for regressing wavelet mulitscale representations at multiple levels in a recursive manner. This recursive implementation can upscale an aerial image to large resolutions. However, this also gives rise to one restriction that the resolution of one aerial image can only be upscaled by the multiples of two. This is because the wavelet analysis in terms of filter banks strictly requires that the downsampling and upsampling should be scaled by two.

Furthermore, we compare the computational complexities of different methods. Compared with the existing deep learning-based super-resolution methods, our model is efficient in both the training and testing procedures. It is obvious that the complexity of our model is four times than that of SRCNN, because it adopts four CNNs with each having the same architecture with SRCNN. Though our model complexity is larger than SRCNN, it is acceptable because its complexity follows a linear growth with respect to the number of CNNs and does not incur exponentially increased overheads. On the other hand, in contrast to the deep super-resolution models such as very deep super-resolution (VDSR) [15] which have 20 weight layers, our three-layered model is not very deep. Different from the very deep structure, we improve the super-resolution ability of the CNN model by not making it deeper but expanding it broader, i.e., training multiple three-layered CNNs in parallel in decomposed frequency subbands.

III. Experiments

We make an empirical evaluation of our framework, with empirical comparison with other super-resolution methods on the RSSCN7 data set [12]. RSSCN7 is an aerial image data set collected by Google Earth. We use this data set, because it exhibits great variability over land covers and objects and thus extends great challenges for super resolution. We used 70% of the images of RSSCN7 data set, i.e., 1960 images, to train the proposed model and the rest 30%, i.e., 840 images, to validate the performance of the trained model. For constructing the individual CNNs, we adopt the structures and settings of SRCNN in [3]. Specifically, we test the convergence of comparison models, evaluate their performance in terms of peak signal-to-noise ratio (PSNR, dB) and structural similarity (SSIM).

A. Model Convergence

Our WMCNNs improve SRCNN by training multiple CNNs to characterize wavelet multiscale representations.

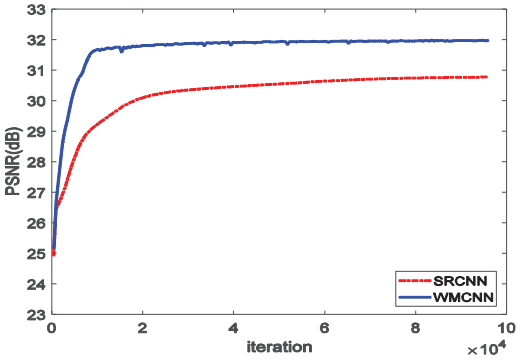

Fig. 2. Convergence curves.

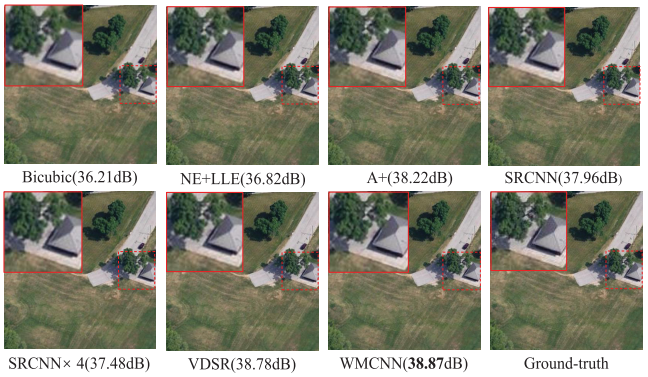

Fig. 3. Restoring grass images with the upscaling factor two.

We compare their convergence speeds for training. The experiments are conducted with the same network configuration and computation environment. The convergence curves with respect to PSNR are shown in Fig. 2. The SRCNN achieves convergence within 8×10^4 and our WMCNN converges to optima within 4×10^4 iterations with the better accuracies.

SRCNN straightforwardly restores the one whole aerial image. In contrast, our WMCNN regresses the wavelet multiscale representations of the whole aerial image. The wavelet-based divide (in terms of multiple orientations and frequency bands) and conquer (via training CNNs) strategy renders a more powerful representation than the sole holistic representations. The efficient training convergence of our WMCNN reflects that CNNs learn aerial image intrinsic faster from wavelet multiscale representations than from the whole images.

B. Quantitative and Qualitative Evaluations

We make quantitative and qualitative empirical comparisons between the bicubic interpolation, NE + LLE [13], A+ [14], SRCNN [3], SRCNN × 4, VDSR [15], and the proposed WMCNNs. Experiments with upscaling factors two and four are conducted. Table I shows the quantitative super-resolution results in terms of PSNR and SSIM.

Our WMCNN notably outperforms SRCNN in terms of both PSNR and SSIM for the upscaling two. Such improvement over SRCNN benefits from the wavelet multiscale analysis that not only provides a remedy for the missing pooling operations in SRCNN but also enriches its local filtering effects by wavelet filters. Furthermore, we empirically compare WMCNN and SRCNN × 4, which is constructed in terms of concatenating four copies of the original SRCNN architectures. Experimental results in Table I show that the proposed model outperforms SRCNN × 4 in terms of PSNR and SSIM over upscaling factor two and four. This empirical comparison validates that the wavelet analysis is the key to the effectiveness of WMCNN rather than the larger model size. We observe that our method performs better than VDSR in terms of PSNR and SSIM. It should be noted that VDSR has 20 weight layers, while our framework consists of four three-layered CNNs resulting in an overall 12-layered concise structure. This observation reflects that our method has much lower structure complexity than VDSR but still achieves comparable performance with them. One reason for the contrastive results is that VDSR takes the whole image as input for training one overall model, and our method trains the CNNs to restore different frequency subbands, which guarantees each frequency representations are restored properly.

The overall performance with the upscaling factor four is inferior to that with two, because larger upscaling restoration induces more super-solution uncertainty. In this situation, some comparison methods and our WMCNN show comparable performance for certain classes. However, our WMCNN still performs better than other methods.

In this letter, we have just used SRCNN as the basic model for multiscale learning. As the wavelet multiscale representations establish comprehensive feature subspaces, it is anticipated that the wavelet multiscale learning based on alternative super-resolution models (e.g., VDSR) would also achieve promising performance.

We visualize the super-resolution results of different methods for qualitative evaluations. Fig. 3 shows the restored high-resolution aerial image with the upscaling factor two. The top-left boxes show the enlarged views of contents in the dashed line boxes. We can see that our WMCNN restores the textures more precisely and clearly than the other comparison methods.

C. Robustness of Proposed Model

We have conducted experiments to validate the robustness of the proposed method by adding varying Gaussian noise to the input images for generating low-high resolution image pairs. We normalize the pixel value to [0, 1] and then add Gaussian noise whose mean value is 0.05, 0.1, 0.15, and 0.2 to the normalized images. Table II shows the performance of SRCNN, SRCNN×4, VDSR, and WMCNN under different noise weights. Compared with the conventional methods and SRCNN×4, WMCNN achieves the best performance in terms of PSNR and SSIM with upscaling factor two. As for upscaling factor four, our method achieves comparable performance with VDSR, which is more complex and deeper than our method.

Compared with the conventional spatial-domain-based super-resolution methods, our WMCNN exploits the wavelet multiscale analysis for capturing scale-invariant statistics of aerial images. The properties make contributions to noise elimination and image restoration. The effective performance of our WMCNN indicates that CNNs eliminate the effect

TABLE I

COMPARISON OF PSNR AND SSIM BETWEEN BICUBIC INTERPOLATION, NE + LLE [13], A+ [14], SRCNN [3], SRCNN × 4, VDSR [15], AND WMCNNS

Methods	Upscaling factor	Grass PSNR/SSIM	Field PSNR/SSIM	Industry PSNR/SSIM	River Lake PSNR/SSIM	Forest PSNR/SSIM	Resident PSNR/SSIM	Parking PSNR/SSIM	Average PSNR/SSIM
Bicubic	2	36.59/0.84	36.42/0.66	26.29/0.69	29.92/0.77	28.70/0.72	26.57/0.71	23.64/0.74	29.73/0.73
NE+LLE		36.45/0.89	35.47/0.84	25.62/0.76	29.58/0.85	27.51/0.75	25.86/0.77	24.61/0.82	29.30/0.81
A+		38.23/0.89	36.82/0.84	27.36/0.77	31.25/0.86	28.91/0.75	27.48/0.77	27.32/0.84	31.05/0.82
SRCNN		38.21/0.90	37.08/0.73	27.13/0.76	30.85/0.82	29.54/0.80	27.36/0.77	25.27/0.81	30.78/0.80
SRCNN×4		38.68/0.90	37.11/0.73	27.64/0.78	31.55/0.83	29.63/0.80	27.96/0.78	26.87/0.84	31.35/0.81
VDSR		38.72/0.96	37.16/0.91	28.10/0.87	32.44/0.94	29.71/0.85	28.21/0.87	29.26/0.94	31.94/0.91
WMCNN		38.82/0.96	37.30/0.90	28.35/0.86	32.41/0.96	29.68/0.87	28.49/0.88	29.10/0.93	32.02/0.91
Bicubic	4	31.58/0.47	33.68/0.32	23.08/0.34	26.13/0.41	25.46/0.35	23.12/0.39	19.78/0.33	26.12/0.37
NE+LLE		31.28/0.58	32.59/0.55	22.16/0.32	25.24/0.53	24.34/0.30	22.29/0.38	18.90/0.33	25.26/0.43
A+		31.60/0.59	33.58/0.55	23.66/0.33	26.75/0.55	25.74/0.31	23.27/0.39	20.46/0.35	26.44/0.44
SRCNN		32.50/0.83	34.10/0.80	23.44/0.57	26.45/0.78	25.78/0.54	23.48/0.60	20.30/0.58	26.58/0.67
SRCNN×4		32.81/0.83	33.62/0.81	23.62/0.57	26.68/0.78	25.79/0.43	23.80/0.56	20.51/0.62	26.69/0.66
VDSR		32.92/0.84	33.94/0.80	23.78/0.60	27.25/0.81	25.85/0.56	24.12/0.65	20.53/0.63	26.91/0.70
WMCNN		33.13/0.85	34.12/0.81	23.85/0.65	27.08/0.80	25.81/0.57	24.02/0.65	20.58/0.64	26.94/0.71

TABLE II

COMPARISON OF PSNR AND SSIM BETWEEN SRCNN [3], SRCNN × 4, VDSR [15], AND WMCNNS UNDER DIFFERENT NOISE WEIGHTS

Methods	Upscaling factor	0.05 PSNR/SSIM	0.1 PSNR/SSIM	0.15 PSNR/SSIM	0.2 PSNR/SSIM
SRCNN	2	25.22/0.85	20.68/0.84	17.55/0.82	15.23/0.80
SRCNN×4		25.20/0.85	21.32/0.83	17.54/0.82	15.22/0.80
VDSR		25.66/0.87	20.85/0.86	17.63/0.83	15.26/0.81
WMCNN		26.08/0.89	22.35/0.88	18.67/0.88	16.82/0.87
SRCNN	4	23.24/0.67	19.79/0.67	17.15/0.64	14.88/0.63
SRCNN×4		23.49/0.67	19.77/0.66	17.09/0.65	14.96/0.63
VDSR		23.31/0.68	19.92/0.67	17.12/0.66	14.95/0.63
WMCNN		23.38/0.69	19.88/0.66	17.25/0.64	14.83/0.58

of noise more effectively by learning essence from wavelet multiscale representations than from the whole images.

IV. CONCLUSION

We have presented a single aerial image super-resolution framework based on WMCNNs. The local processing ability of a CNN is enhanced by wavelet filters. The lack of pooling operations in CNN-based super resolution is compensated by the downsampling in the wavelet decomposition. The proposed framework integrates the representational power of CNNs for learning specific features and the multiscale capability of the wavelet analysis for obtaining multiple orientation and frequency representations. Experimental evaluations have validated the effectiveness of our super-resolution framework.

The conventional assessment methods for single image super resolution are based on the differences between restored high-resolution images and original high-resolution images, and it is unfeasible in practice to obtain the original high-resolution images. In our future work, we will investigate how to develop a visual analysis and no reference quality indexes as assessment indicators. Furthermore, we will investigate how we can incorporate the visual analysis and no reference quality indexes into the objective function for the purpose of improving perceptual performance.

REFERENCES

[1] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 996–1011, Jun. 1996.

[2] N. Nguyen, P. Milanfar, and G. Golub, "Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1299–1308, Sep. 2001.

[3] 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, Feb. 2015.

[4] S. Lei, Z. Shi, and Z. Zou, "Super-resolution for remote sensing images via local–global combined network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1243–1247, Aug. 2017.

[5] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017.

[6] J. Li, Q. Yuan, H. Shen, X. Meng, and L. Zhang, "Hyperspectral image super-resolution by spectral mixture analysis and spatial–spectral group sparsity," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1250–1254, Sep. 2016.

[7] L. Yue, H. Shen, J. Li, Q. Yuanc, H. Zhang, and L. Zhang, "Image super-resolution: The techniques, applications, and future," *Signal Process.*, vol. 128, pp. 389–408, Nov. 2016.

[8] N. Kumar, R. Verma, and A. Sethi, "Convolutional neural networks for wavelet domain super resolution," *Pattern Recognit. Lett.*, vol. 90, pp. 65–71, Apr. 2017.

[9] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet-SRNet: A wavelet-based CNN for multi-scale face super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2017, pp. 1689–1697.

[10] H. Chavez-Roman and V. Ponomaryov, "Super resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1777–1781, Oct. 2014.

[11] A. Cohen and I. Daubechies, "A stability criterion for biorthogonal wavelet bases and their related subband coding scheme," *Duke Math. J.*, vol. 68, no. 2, pp. 313–335, 1992.

[12] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2321–2325, Nov. 2015.

[13] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2004, p. 1.

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

[15] 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.*, Jun. 2016, pp. 1646–1654.