Single Image Dehazing via Deep Learning-based Image Restoration

Chia-Hung Yeh[1,2], Chih-Hsiang Huang[2], Li-Wei Kang[3,*], and Min-Hui Lin[2]
[1]Department of Electrical Engineering, National Taiwan Normal University, Taipei, Taiwan
E-mail: chyeh@ntnu.edu.tw
[2]Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan
E-mail: tommy830725@gmail.com
[3]Graduate School of Engineering Science and Technology-Doctoral Program, and Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan
[*]E-mail: lwkang@yuntech.edu.tw

***Abstract*—Images/videos captured from outdoor visual devices are usually degraded by turbid media, such as haze, smoke, fog, rain, and snow. Haze is the most common one in outdoor scenes due to the atmosphere conditions. This paper presents a deep learning-based architecture for single image dehazing via image restoration. Instead of learning an end-to-end mapping between each pair of hazy image and its corresponding haze-free one adopted by most existing approaches, we propose to transform the problem into the restoration of the image base component. By first decomposing the hazy image into the base and the detail components, haze removal can be achieved by learning a CNN (convolutional neural network) only for mapping between hazy and haze-free base components, while the detail component can be further enhanced. As a result, the final dehazed image is obtained by integrating the haze-removed base and the enhanced detail image components. Experimental results have demonstrated good efficacy of the proposed method, compared with state-of-the-art results.**

I. INTRODUCTION

Different weather conditions, such as haze, fog, smoke, rain, or snow will cause unpleasing visual effects in images [1]–[3]. Such artifacts may significantly degrade the performances of several outdoor vision systems, such as event detection and understanding, object detection, tracking, and recognition, and scene analysis and classification, for outdoor surveillances or ADAS (advanced driver assistance systems) applications [4]. In addition, hazing artifacts have been also shown to significantly influence the performance of image/video compression [5], which will degrade the transmission and storage efficiency for digital visual data. Therefore, removal of weather effects in images/videos has been important and received much attention. In this paper, we focus on haze removal (or dehazing) from a single image.

Based on the fact that haze would depend on unknown depth, dehazing is therefore very challenging. Moreover, if the available input is only single hazy image, the problem is under-constrained and more challenging. Therefore, most traditional dehazing approaches have been presented relying on using multiple hazy images as input or additional prior knowledge [6], which are usually impractical in several real applications.

To achieve single image dehazing [7]–[9], the atmospheric scattering model [1]–[3] has been commonly used to describe the image formation in the presence of haze, and great success has been obtained relying on stronger priors or assumptions. In this model, a hazy image is formulated as

$$I(x) = J(x)t(x) + A\big(1 - t(x)\big), \tag{1}$$

where I is the observed hazy image, J is the scene radiance (the original haze-free image to be recovered), t is the medium transmission (or transmission map) indicating the portion of the light that is not scattered and reaches the camera, and A is the global atmospheric light. Based on this image degradation model, it was proposed in [7] to estimate the albedo of the scene and the medium transmission under the assumption that the transmission and the surface shading are locally uncorrelated for single image dehazing. Moreover, an effective image prior, called dark channel prior [8] was proposed for single image dehazing, where the key observation is that most local patches in outdoor haze-free images include some pixels whose intensity is very low in at least one color channel. This approach [8] was then further extended to pixel-based dark channel prior in [9] with performance improved. Nevertheless, these traditional approaches may not well restore heavily hazy images and may fail in the cases where the constraints/priors/ assumptions are invalid.

To achieve better dehazing performance, based on the deep investigation in haze-relevant image features, a learning framework for single image dehazing was presented in [10]. Furthermore, relying on the rapid development of deep learning techniques with great success in numerous perceptual tasks (e.g., object detection, image understanding, and speech recognition) [11]–[12], some deep learning-based single image dehazing frameworks [13]–[15] have been presented. They have shown that haze-relevant features would be automatically learned through a deep CNN (convolutional neural network) to get better dehazing results.

For example, in [13], a multi-scale deep neural network for single image dehazing was proposed by learning the mapping between hazy images and their corresponding transmission

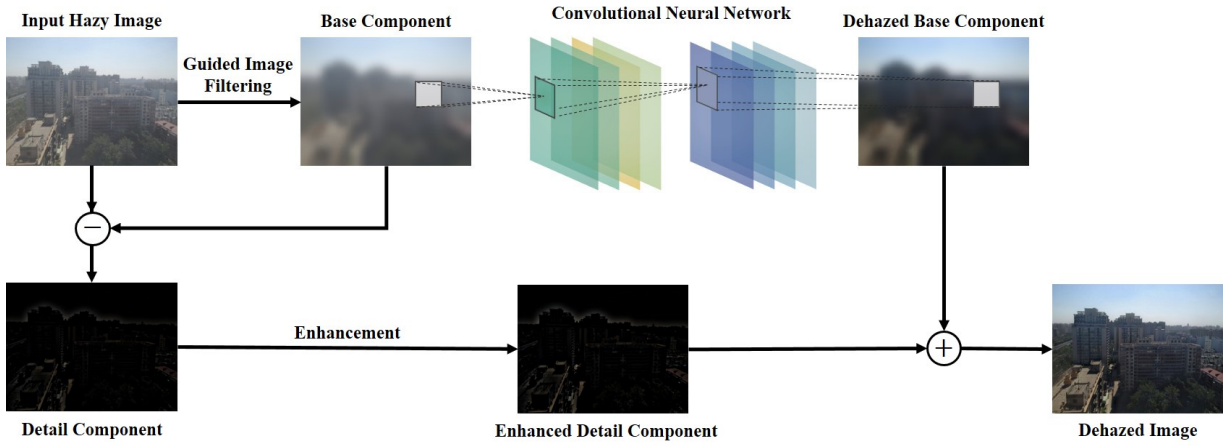

Fig. 1. The diagram of the proposed deep learning-based single image dehazing framework.

maps. The dehazed image can then be produced based on the atmospheric scattering model (denoted by MSCNN). Similarly, a trainable end-to-end deep CNN (denoted by DehazeNet) for medium transmission estimation was also proposed in [14], demonstrating that most haze-relevant features can be learned through the network. Furthermore, a CNN-based single image dehazing network is designed in [15] based on a reformulated atmospheric scattering model for directly generating the dehazed image.

In this paper, to get better dehazing results with lower run-time computational complexity, a novel deep CNN architecture is proposed for single image dehazing. The major contribution of this paper is three-fold: (1) by first decomposing the hazy image into the base and the detail components, haze removal can be achieved by learning a CNN only for mapping between hazy and haze-free base components, while the detail component can be further enhanced separately to avoid possible blurring effects (loss of details) in most existing methods; (2) the architecture of our CNN is relatively simple, is with lower run-time complexity, and is suitable for real applications; and (3) our dehazing process can be iterated several times to progressively refine the visual quality of output images without resulting in over-enhancement.

The rest of this paper is organized as follows. In Sec. II, we present the proposed single image dehazing framework. In Sec. III, experimental results are demonstrated. Finally, Sec. IV concludes this paper.

II. Proposed Single Image Dehazing Framework

As shown in Fig. 1, the proposed deep learning-based single image dehazing framework consists of the three stages: (1) the preprocessing stage via image decomposition; (2) the dehazing stage relying on CNN to produce the haze-removed base image component; and (3) the generation stage of final dehazed image by integrating the clean base image component and the enhanced detail image component. The details of the three stages shall be elaborated in the following three subsections, respectively.

A. Preprocessing and Problem Formulation

To avoid possible detail loss in the image dehazing process, in our method, we first decompose an input hazy image into the base component and the detail component via image filtering. The main principle is to keep the haze and low-frequency information in the base component, while keeping the image details (edge/texture information) in the detail component. To achieve this goal, the well-known guided image filter [16] is employed based on that the filter performs well as an edge-preserving smoothing operator in linear time.

For an input hazy image I, we perform guided image filtering to I to obtain the base component I_{base} and the detail component I_{detail}, such that

$$I = I_{base} + I_{detail}, \tag{2}$$

where $I_{base} = G(I)$, and G denotes the guided image filtering operator. This signal decomposition-based idea has been also shown to be efficient in several other applications (e.g., [17]–[23])

To this end, we formulate the single image dehazing problem as the mapping between a hazy base image component and its corresponding hazy-free component. To solve the problem, we intend to learn a mapping function F subject to the loss function ℓ as:

$$\ell(\Theta) = \frac{1}{N}\sum_{j=1}^{N}\left\|F(Y_j;\Theta) - X_j\right\|_2^2, \tag{3}$$

where Θ denotes the set of parameters trained for our CNN (described in Sec. II.B), N is the number of training patch pairs, and F is the mapping function between each pair of training patches. Here, a pair of training patches includes a patch Y_j extracted from a hazy base component and its corresponding patch X_j extracted from the corresponding hazy-free base

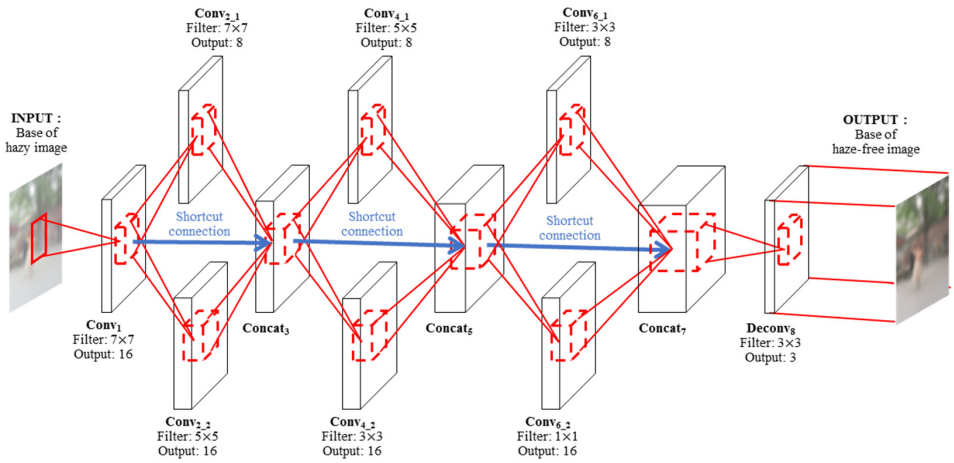

Fig. 2. The proposed deep convolutional neural network architecture for single image dehazing.

component. The loss function ℓ is minimized using the Adam optimization algorithm designed for first-order gradient-based optimization of stochastic objective function [24].

B. *Learning of CNN for Haze Removal of Base Component*

As shown in Fig. 2, the proposed CNN aims at transforming an input hazy base image component to its corresponding hazy-removed base component. Our CNN consists of eight layers, where the first is a convolutional layer (denoted by Conv_i for the i-th layer in Fig. 2) with 16 filters of kernel size 7×7, and the last is a deconvolutional layer with 3 filters of kernel size 3×3. In addition, the 2nd, 4th, and 6th layers are the multi-scale convolutional layers with filter sizes of 7×7 and 5×5, 5×5 and 3×3, and 3×3 and 1×1, respectively. For each sub-layer (denoted by Conv_{i_j} for the j-th sub-layer of the i-th layer in Fig. 2) with larger filter size of the i-th layer (i = 2, 4, 6), 8 feature maps will be produced (the upper half of Fig. 2), whereas for the sub-layer with smaller filter size of the i-th layer, 16 feature maps will be produced (the lower half of Fig. 2). All of the convolutional operations (except the last layer) in our CNN are all with the rectified linear unit (ReLU) function [25] as the activation function. ReLU is mainly utilized for nonlinearity, which has been shown to allow for faster training than other non-linear separators.

Moreover, to avoid the problem of possible color distortion in a dehazed image, we propose to fully preserve the feature maps produced by a previous layer, which will be fed into the next layer. To achieve this, we include the three shortcut connections and three concatenation layers (denoted by Concat_i for the i-th layer in Fig. 2), inspired by the deep residual learning framework proposed in [26], which utilized the shortcut connections to skip one or more layers. More specifically, let's consider the i-th layer (i = 2, 4, 6, in Fig. 2, as an example), including two convolutional sub-layers (with ReLU functions) and a shortcut connection, and assume x and y are the input and output vectors, respectively, of this layer. Formally, in this paper, we consider the building block of our CNN, defined as:

$$y = \sigma(x * W_{i\_1}) + \sigma(x * W_{i\_2}) + x, \tag{4}$$

where "$*$" denotes the convolution operation, σ denotes the ReLU function, W_{i_1} and W_{i_2} are the filter weights to be learned of the sub-layer Conv_{i_1} and Conv_{i_2}, of the i-th layer, respectively, and the "+" operation integrates the three items realized by a concatenation layer. Since the depths of the feature maps produced by the three items are different, instead of utilizing the element-wise addition used in [26], the concatenation layer is used to stack all of the feature maps, which is then fed into the next layer.

As a result, by performing deep residual learning to the CNN, the output dehazed base image component can be produced almost without color distortion. On the other hand, our CNN also benefits the inherited advantages from the deep residual learning, such as preventing gradient explosions while avoiding the vanishing gradient problem, and fast convergence [26]. In addition, by utilizing the multi-scale convolutional layers of different filter sizes enables our CNN to extract multi-scale image features, which would be also beneficial to recover hazy images.

C. *Reconstruction of Haze-free Image*

As a result, the final haze-removed image J (for the input hazy image I) is obtained by integrating the dehazed base and enhanced detail components as:

$$J = D(I_{base}) + E(I_{detail}), \tag{5}$$

where $D(I_{base})$ denotes the dehazed base component (I_{base}) via the proposed deep network, and $E(I_{detail})$ is the enhanced detail component (I_{detail}) of the input hazy image. In our method, $E(I_{detail}) = M \times I_{detail}$, where M is a factor related to the entropy of the input hazy image I, which is obtained by exploring the relationship between each pair of detail image

(a) (b) (c) (d) (e)

Fig. 3. Dehazing results on synthetic images from SOTS dataset: **(a)** haze-free image; **(b)** synthetic hazy version of (a); the dehazed results obtained by **(c)** MSCNN [13]; **(d)** DehazeNet [14]; and **(e)** Proposed method.

(a) (b) (c) (d) (e)

Fig. 4. Dehazing results on synthetic images from HSTS dataset: **(a)** haze-free image; **(b)** synthetic hazy version of (a); the dehazed results obtained by **(c)** MSCNN [13]; **(d)** DehazeNet [14]; and **(e)** Proposed method.

TABLE I
AVERAGE QUANTITATIVE RESULTS ON SYNTHETIC HAZY IMAGES IN SOTS DADASET, IN TERMS OF PSNR (IN DB) AND SSIM METRICS.

	MSCNN [13]	DehazeNet [14]	Proposed
PSNR	21.084	24.179	**25.3849**
SSIM	0.780	0.820	**0.872**

TABLE II
AVERAGE QUANTITATIVE RESULTS ON SYNTHETIC HAZY IMAGES IN HSTS DADASET, IN TERMS OF PSNR (IN DB) AND SSIM METRICS.

	MSCNN [13]	DehazeNet [14]	Proposed
PSNR	20.273	**25.528**	25.121
SSIM	0.722	0.836	**0.860**

and its groundtruth in 30,000 training image pairs via nonlinear regression.

On the other hand, a unique property of our method is that iteratively performing the proposed dehazing process would progressively enhance the dehazing performance. That is, by feeding the output haze-removed image into the system as an input hazy image again would obtain better dehazing result. This advantage mainly comes from that for each dehazing process, we first decompose the input hazy image into the base and the detail components. As a result, the image details would be well preserved while the color distortion would be avoided based on our CNN even if the dehazing process is iteratively performed several times.

III. EXPERIMENTAL RESULTS

A. Network Training and Parameter Settings

To train the proposed deep CNN, the Outdoor Training Set (OTS) from the RESIDE (REalistic Single Image DEhazing) dataset with a large diversity of scenes [27] was used. OTS includes 313,950 images synthesized from collected real world outdoor scenes, where each image pair contains one hazy image and its haze-free groundtruth. To create our training dataset, we randomly extracted several pairs of patches of size 16×16 from each hazy image and its corresponding haze-free groundtruth in OTS. As a result, we totally obtained 1,000,000 pairs of patches. For each image patch, we applied our preprocessing process via guided image filtering [16] to get its base component to form the training patch. During the training

(a) (b) (c) (d)

Fig. 5. Dehazing results on challenging natural hazy images: **(a)** hazy image; the dehazed results obtained by **(b)** MSCNN [13]; **(c)** DehazeNet [14]; and **(d)** Proposed method.

process, the filter weights are initialized by drawing randomly from a Gaussian distribution. The learning rate is set to 1×10^{-5}, and our network is trained with a batch-size of 20 in 100 epochs.

B. Quantitative Results on Synthetic Hazy Images

To quantitatively evaluate the performance of the proposed deep learning-based single image dehazing method, two state-of-the-art methods, MSCNN [13] and DehazeNet [14], were used for comparisons. The two well-known metrics, PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index) [28] were used for quality assessment of haze-removed images.

In the experiments, two testing image sets, SOTS (Synthetic Objective Testing Set) and HSTS (Hybrid Subjective Testing Set), from the RESIDE dataset [27] were used. We picked the 500 outdoor scene images and their corresponding synthetic hazy images from SOTS, and 10 synthetic outdoor hazy images (with groundtruths) from HSTS. The subjective and objective results for the SOTS testing images obtained by MSCNN [13], DehazeNet [14], and the proposed method are shown in Fig. 3 and Table I, respectively. Moreover, Fig. 4 and Table II, respectively, display the subjective and objective results for the HSTS testing images.

As revealed by Figs. 3-4 and Tables I-II, the proposed method significantly outperforms MSCNN [13], and outperforms or is comparable with DehazeNet [14]. For example, based on Fig. 4, MSCNN may suffer from color distortion (e.g., sky color) while the proposed method better keeps the color information beneficial from our deep residual learning network structure. In addition, based on Fig. 3, DehazeNet may loss some image details in darker regions (e.g., grass regions) while our method better keeps the texture information beneficial from preservation of details in our image decomposition architecture.

C. *Subjective Results on Real-World Hazy images*

To evaluate our method on realistic hazy images, we selected some challenging natural images, which were found to be highly challenging for dehazing [7]–[8]. As revealed by Fig. 5, our method is found to generalize well on realistic hazy images in visual quality, and preservations of details and color information in dehazed images. The two compared state-of-the-art methods still suffer from the problems of possible color distortion or lack of details in dehazed images.

D. Iterative Improvements of Proposal Method

To verify the unique property for iteratively performing the proposed dehazing process would progressively enhance the dehazing performance, we show the dehazing results on a real-world hazy image by performing our method with one and two iterations, respectively, in Fig. 6. It can be observed that as the number of iterations increases, the dehazing result is progressively improved while preserving image details and color information.

E. Run-Time Analysis

The proposed method was implemented in Python programming language with TensorFlow on a personal computer equipped with Intel® Core™ Core i7-4790 processor, 16 GB memory, and NVIDIA GeForce GTX 1080 GPU. We estimated the run-time (in seconds with only CPU utilized) for dehazing 500 synthetic hazy images from SOTS dataset, and report the average processing time per image (of size 620×460) in Table III (with those of MSCNN and DehazeNet). It should be noted that the online available public open sources of MSCNN [13] and DehazeNet [14] were implemented by

(a) (b) (c) (d) (e)

Fig. 6. Dehazing results on challenging natural hazy images: **(a)** hazy image; the dehazed results obtained by **(b)** MSCNN [13]; **(c)** DehazeNet [14]; **(d)** Proposed (one iteration); and **(e)** Proposed (two iterations) methods.

Matlab. In addition, with the utilization of GPU, our method only takes 0.48 seconds to process one hazy image of size 620×460. As a result, our method reveals good efficiency, thanks to its lightweight deep architecture.

TABLE III
COMPARISON OF AVERAGE RUNNING TIME (IN SECONDS) FOR DEHAZING PER IMAGE ON 500 SYNTHETIC HAZY IMAGES IN SOTS.

	MSCNN [13]	DehazeNet [14]	Proposed
Time	1.92	1.85	**1.21**

IV. CONCLUSIONS

In this paper, we have proposed a novel deep CNN architecture for single image haze removal. Our method first decomposes an input hazy image into the base and detail components. The base component is fed into our lightweight CNN with deep residual learning to obtain the haze-removed base component while preserving the color information for the image. The detail component is enhanced by nonlinear regression-based image enhancement while preserving the original image details. Our experimental results show that the proposed method achieves better or comparable performance with state-of-the-art single image dehazing algorithms.

ACKNOWLEDGEMENT

This work was supported in part by Ministry of Science and Technology, Taiwan, under the Grants NSC 102-2221-E-110-032-MY3, MOST 103-2221-E-110-045-MY3, MOST 105-2221-E-110-094-MY3, MOST 106-2221-E-110-083-MY2, MOST 107-2218-E-003-003-, MOST 107-2218-E-110-004-, and MOST 105-2628-E-224-001-MY3. This work was also financially supported by the "Artificial Intelligence Recognition Industry Service Research Center (Project No.107-N04-2)" from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Educa-tion (MOE) in Taiwan.

REFERENCES

[1] S. K. Nayar and S.G. Narasimhan, "Vision in bad weather," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, pp. 820-827, 1999.

[2] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.*, vol. 48, pp. 233-254, 2002.

[3] R. Tan, "Visibility in bad weather from a single image," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2008.

[4] M. S. Shehata, J. Cai, W. M. Badawy, T. W. Burr, M. S. Pervez, R. J. Johannesson, and A. Radmanesh, "Video-based automatic incident detection for smart roads: the outdoor environmental challenges regarding false alarms," *IEEE Trans. Intell. Transportation Syst.*, vol. 9, no. 2, pp. 349–360, June 2008.

[5] K. B. Gibson, D. T. Võ, and T. Q. Nguyen, "An investigation of dehazing effects on image and video coding," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 662-673, Feb. 2012.

[6] Y. Li, S. You, M. S. Brown, and R. T. Tan, "Haze visibility enhancement: A survey and quantitative benchmarking," *Comput. Vis. Image Understanding*, vol. 165, pp. 1–16, Dec. 2017.

[7] R. Fattal, "Single image dehazing," *ACM Trans. Graphics*, vol. 27, no. 3, Aug. 2008.

[8] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[9] C.-H. Yeh, L.-W. Kang, M.-S. Lee, and C.-Y. Lin, "Haze effect removal from image via haze density estimation in optical model," *Optics Express*, vol. 21, no. 22, pp. 27127–27141, Nov. 2013.

[10] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2995–3002.

[11] Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, 2012.

[13] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," *Proc. European Conf. Comput. Vis.*, 2016, pp. 154–169.

[14] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal, *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[15] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network, *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 4780–4788.

[16] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, June 2013.

[17] L. W. Kang, C. W. Lin, and Y. H. Fu, "Automatic single-image-based rain streaks removal via image decomposition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1742–1755, Apr. 2012.
[18] D.-A. Huang, L.-W. Kang, Y.-C. F. Wang, and C.-W. Lin, "Self-learning based image decomposition with applications to single image denoising," *IEEE Trans. Multimedia*, vol. 16, no. 1, pp. 83–93, Jan. 2014.
[19] D.-Y. Chen, C.-C. Chen, and L.-W. Kang, "Visual depth guided color image rain streaks removal using sparse coding," *IEEE Trans. Circuits Syst. Video Techn*., vol. 24, no. 8, pp. 1430–1455, Aug. 2014.
[20] C.-H. Yeh, L.-W. Kang, Y.-W. Chiou, C.-W. Lin, and S.-J. Fan Jiang, "Self-learning-based post-processing for image/video deblocking via sparse representation," *Journal of Visual Communication and Image Representation*, vol. 25, no. 5, pp. 891–903, July 2014.
[21] L.-W. Kang, C.-C. Hsu, B. Zhuang, C.-W. Lin, and C.-H. Yeh, "Learning-based joint super-resolution and deblocking for a highly compressed image," *IEEE Trans. on Multimedia*, vol. 17, no. 7, pp. 921-934, July 2015.
[22] L.-W. Kang, C.-H. Yeh, D.-Y. Chen, and C.-T. Lin, "Self-learning-based signal decomposition for multimedia applications: a review and comparative study," *Proc. APSIPA Annual Summit and Conf.*, Siem Reap, city of Angkor Wat, Cambodia, Dec. 2014.
[23] L.-W. Kang, C.-M. Yu, C.-Y. Lin, and C.-H. Yeh, "Image and video restoration and enhancement via sparse representation," *Biometrics: Concepts, Methodologies, Tools, and Applications*, IGI Global, pp. 501–528, 2017.
[24] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *Proc. Int. Conf. Learning Representations*, May 2015.
[25] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," *Proc. Int. Conf. Machine Learning*, Haifa, Israel, June 2010, pp. 807–814.
[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*., 2016, pp. 770–778.
[27] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "RESIDE: A benchmark for single image dehazing," *arXiv preprint arXiv*:1712.04143, Dec. 2017.
[28] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process*., vol. 13, no. 4, pp. 600–612, Apr. 2004.