Deep Multi-Model Fusion for Single-Image Dehazing

Zijun Deng[1,*], Lei Zhu[3,*], Xiaowei Hu[2], Chi-Wing Fu[2], Xuemiao Xu[1,5,6,†],
Qing Zhang[7], Jing Qin[8], and Pheng-Ann Heng[2,4]
[1] South China University of Technology, [2] The Chinese University of Hong Kong,
[3] Guangdong Provincial Key Laboratory of Computer Vision and Virtual
Reality Technology, Shenzhen Institutes of Advanced Technology, CAS
[4] CAS Key Laboratory of Human-Machine Intelligence-Synergy
Systems, Shenzhen Institutes of Advanced Technology, CAS
[5] State Key Laboratory of Subtropical Building Science
[6] Guangdong Provincial Key Lab of Computational Intelligence and Cyberspace Information
[7] Sun Yat-sen University [8] The Hong Kong Polytechnic University

Abstract

This paper presents a deep multi-model fusion network to attentively integrate multiple models to separate layers and boost the performance in single-image dehazing. To do so, we first formulate the attentional feature integration module to maximize the integration of the convolutional neural network (CNN) features at different CNN layers and generate the attentional multi-level integrated features (AMLIF). Then, from the AMLIF, we further predict a haze-free result for an atmospheric scattering model, as well as for four haze-layer separation models, then fuse the results together to produce the final haze-free image. To evaluate the effectiveness of our method, we compare our network with several state-of-the-art methods on two widely-used dehazing benchmark datasets, as well as on two sets of real-world hazy images. Experimental results demonstrate clear quantitative and qualitative improvements of our method over the state-of-the-arts.

1. Introduction

In hazy conditions, floating particles in the atmosphere absorb and scatter the light, thereby distorting the photo contents and degrading the accuracy of subsequent visual analysis. To overcome the issues, many methods [8, 11, 23, 28, 36, 35] have been proposed to recover the underlying haze-free image from the single hazy input.

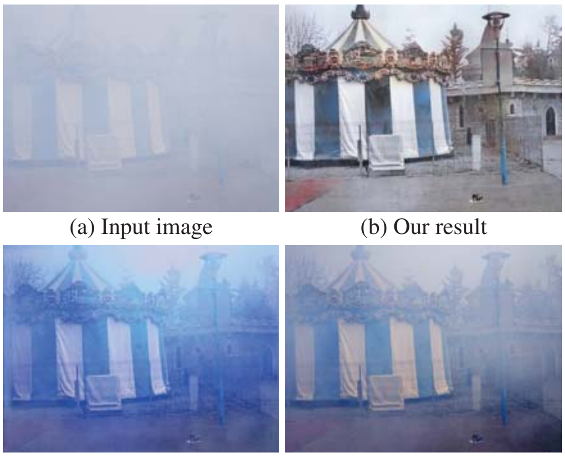

(a) Input image (b) Our result
(c) DCPDN [36] (CVPR'18) (d) PDNet [35] (ECCV'18)

Figure 1: Haze removal on a real-world photo under heavy haze. Results in (b)-(d) are obtained by training these networks using the training set of NTIRE-dehazing challenge.

The image degradation caused by the haze is usually formulated by an atmospheric scattering (AS) model [36, 35]:

$$I(p) = J(p) \times T(p) + A(p) \times (1 - T(p)) \ , \qquad (1)$$

where I is the observed hazy image; p is the pixel location; J is the underlying scene radiance image to be recovered; T is the transmission map, which represents the distance-dependent factor affecting the fraction of light that reaches the camera sensor; and A is the global atmospheric light, indicating the ambient light intensity.

Early dehazing methods employed hand-crafted priors based on the statistics of clean images to estimate the transmission map T [5, 9, 30, 2], such as local max contrast [31], dark channel prior [13], color-line prior [10], and color attenuation prior [44], then use the atmospheric scattering

[*] Zijun Deng and Lei Zhu are the joint first authors of this work.
[†] Corresponding author (xuemx@scut.edu.cn)

model to recover the underlying haze-free results. Although improving the overall scene visibility, using hand-crafted priors tend to introduce undesirable artifacts such as color distortions [19]. Recently, learning-based methods, such as convolutional neural network (CNN) based frameworks, have shown remarkable improvements by learning the transmission map from the labelled datasets [6, 27, 19], or by directly learning the mapping from the input hazy images to haze-free counterparts [23, 28, 36, 35]. However, most existing dehazing networks are based only on the haze related atmospheric scattering model (Eq. (1)) to learn the transmission maps or haze-free images, thus tend to over-dehaze or under-dehaze input images; see Figures 1 (c)-(d).

Similar to other image restoration tasks (e.g., image denoising [33, 12, 38, 41], image smoothing [42], and image deraining [43, 37, 16]), we can model the image dehazing as a layer separation problem by considering the input hazy image as a combination of multiple layers. The image dehazing separates the input hazy image (I) into a haze-free layer (J) and another layer (H), which contain haze information:

$$I = \Phi(J,\ H)\ , \tag{2}$$

where Φ denotes the layer separation function for the complex hazing process, and we explore four specific layer decompositions for the function Φ; see Section 3 for details.

In this work, we develop an end-to-end deep multi-model fusion network by integrating dehazed results recovered from the atmospheric scattering model and the hazing layer separation model into a single network architecture for improving the dehazing performance. To do so, we first utilize a CNN to generate feature maps with different scales, then produce an attentional multi-level integrated feature (AMLIF) map by integrating features from different CNN layers. Based on the AMLIF, we obtain a dehazed result from the atmospheric scattering model and four results from the layer separation models with different hazing layer decompositions. Lastly, we develop an attentional fusion module to integrate these results into our final result. Overall, we summarize the major contributions of this work as:

- First, we develop an end-to-end deep neural network by fusing the atmospheric scattering model and hazing layer separation model for improving dehazing performance.

- Second, we develop the attention mechanism based module to integrate features from different convolutional layers of a CNN, and then predict dehazed results from the integrated features, based on the atmospheric scattering model and several specific layer separation formulations for fully exploiting the complementary information between different hazing models.

- Third, we evaluate the proposed method on two widely-used dehazing benchmark datasets and various real-world hazy images by comparing it with state-of-the-art dehazing methods. The experimental results show that the developed network outperforms other dehazing methods on all the benchmarks and real hazy images. Overall, the method in this work sets a new state-of-the-art performance on single image dehazing.

2. Related Work

Hand-crafted-prior-based methods investigated image priors from the hazy and clean images for estimating the transmission map for single-image dehazing, such as the dark channel prior (DCP) in He et al. [13], color-line priors in Fattal [10], and haze-line in Berman et al. [4]; please refer to Zhang et al. [36] for details. These methods tend to introduce undesirable artifacts (e.g., color distortions) in the results [28] since their hand-crafted priors from human observations do not always hold in diverse real-world images.

Deep learning-based methods have been developed for single-image dehazing by witnessing the success of convolutional neural networks (CNNs) in many computer vision tasks [26, 14, 7], Early attempts designed CNNs to only estimate the transmission map and then used the atmospheric scattering model (see Eq. (1)) for recovering the clean image. Ren et al. [27] first designed a coarse-scale network to predict a holistic transmission map and then a fine-scale network to refine the transmission map. Cai et al. [6] developed a DehazeNet equipped with BReLU based feature extraction layers for transmission map prediction. Hence, inaccuracies on the transmission map estimation tend to degrade the quality of the dehazed result.

Recently, end-to-end CNNs have been designed to directly learn the clean image from a hazy input for dehazing. Yang et al. [35] integrated the haze imaging model constraints and image prior learning into a single dehazing network for clean image prediction. Li et al. [23] introduced the VGG [29] features and an L_1-regularized gradient prior into conditional generative adversarial network (cGAN) [17] for clean image estimation. Ren et al. [28] designed an encoder-decoder network (GFN) to learn confidence maps from three derived inputs and fused them into the final dehazed result. However, these deep models formulated a disjoint optimization, so it failed to capture the relations among the transmission map, atmospheric light, and dehazed result, and hindered the overall dehazing performance. Unlike them, Zhang et al. [36] proposed a single dehazing network (DCPDN) to jointly learn the transmission map, atmospheric light and haze-free images for capturing their relations. Although improving the dehazing performance, the DCPDN [36] still under-dehaze or over-dehaze input hazy images, since only the atmospheric scattering model is considered when designing the CNN; see Figure 1 (c). To further boost clean image prediction, we consider the dehazing process as a layer separation model

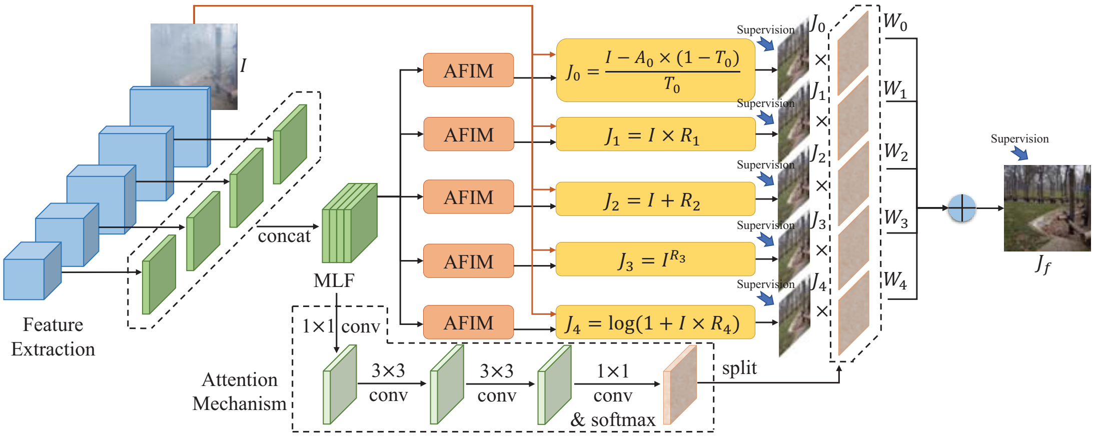


Figure 2: Overview of the developed DM^{2}F-Net: (i) it starts by generating multi-layer features (MLF) from different CNN layers; (2) we develop an attentional feature integration module (AFIM) (see Figure 3) to refine MLF, and then predict a dehazed result from the refined features by developing a scattering model based module (see Figure 5); (3) we formulate four specific hazing layer decompositions (see Figure 4) to predict their dehazed results (denoted as J_1, J_2, J_3 and J_4); (4) we fuse these dehazed results to produce our final result by learning weighting maps (W_0, W_1, W_2, W_3, and W_4). Note that convolutional parameters in the five AFIMs are not shared.

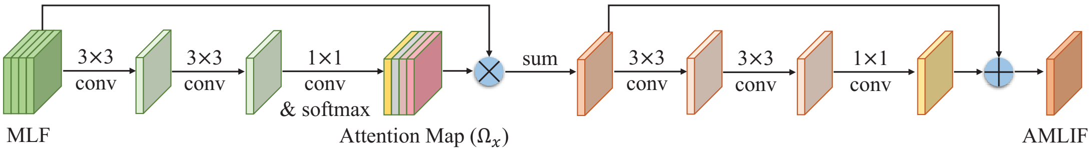


Figure 3: The schematic illustration about the attentional feature integration module (AFIM) of Figure 2.

(see Eq. (2)), and develop an efficient end-to-end dehazing network by fully fusing dehazed results from both the atmospheric scattering model and the layer separation.

3. Our Approach

Figure 2 shows the architecture of our network (denoted as DM^{2}F-Net), which fuses the atmospheric scattering (AS) and layer separation models for dehazing. Given an input hazy image, we develop attentional feature integration modules (AFIMs; see Section 3.1) to produce feature maps (denoted as AMLIF) by learning attention maps to leverage the complementary information among different CNN features. Then, we predict the AS model based result (denoted as J_0) from AMLIF by joint learning. Moreover, we compute four dehazed results (denoted as J_1, J_2, J_3 and J_4) for four-layer separation formulations from another four AMLIF. Finally, we learn attention maps to weight all these dehazed results for generating the final result; see Section 3.2.

3.1. Attentional Feature Integration Module

Note that the features at shallow layers in a convolutional neural network (CNN) are responsible for discovering the fine detail information but lack of semantic information of input hazy image. Hence, the dehazing prediction from these features can capture most of the background details, but many non-haze details are also corrupted with haze. On the other hand, features at deep CNN layers are responsible for capturing the semantic information to remove most of the haze in the input image but somehow lack of non-haze background details due to their relatively larger receptive fields than shallow layers. Hence, we design an attentional feature integration module (AFIM) to leverage complementary among different CNN layers for the clean image prediction by automatically learning attention maps for weighting concatenated features from different CNN layers; see Figure 3 for the AFIM architecture.

To do so, taking concatenated features (denoted as MLF) from different CNN layers as the input, the AFIM first utilizes three convolutional layers and a softmax function to produce attention weights Ω_x (see Figure 3):

$$\Omega_x = \mathrm{Softmax}(\sigma(\Theta * \mathrm{MLF} + b)) \ , \tag{3}$$

where Θ and b are the weights and bias of three convolutional layers on the MLF; The three convolution kernel sizes are 3×3, 3×3, and 1×1; and σ is the ReLU activation function [18]. Then, the attention map Ω_x is multiplied to the concatenated features (MLF) in a layer-by-layer manner, and then the multiplied features are added together across the channel direction. After that, we employ a residual block [14] to produce the output attentional concatenated

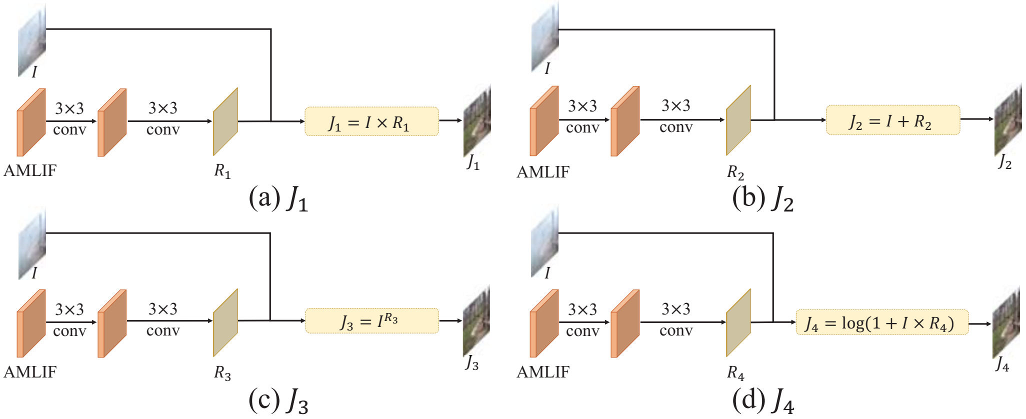


Figure 4: The schematic illustration about how to produce the dehazed results (denoted as J_1, J_2 J_3 and J_4) using four layer decompositions formulations (see Figures 4 (a)-(d)). Note that the channels of R_1, R_2, R_3, and R_4 are 3.

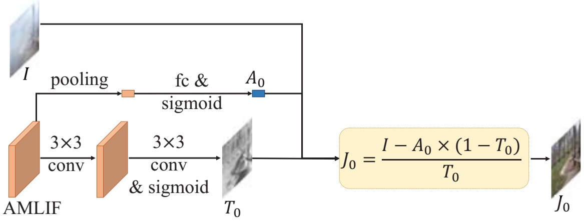


Figure 5: The schematic illustration about how to produce the dehazed result (denoted as J_0) using the AS model.

multi-level features (AMLIF) of our AFIM. In the residual block, we use two 3×3 and a 1×1 convolutional layers to produce the residual component; see Figure 3.

3.2. Dehazing Prediction

This section shows how to predict dehazed results for the atmospheric scattering model (Section 3.2.1), and the layer separation models (Section 3.2.2), as well as merge them for our final result (Section 3.2.3).

3.2.1 Prediction from Atmospheric Scattering Model

To predict the dehazed result for the atmospheric scattering (AS) model, we develop a AFIM to generate AMLIF (see Section 3.1), and then jointly estimate the transmission map, atmospheric light and the dehazed result from AMLIF by embedding the AS model to the network. Figure 5 shows the detailed architecture. Specifically, we employ two 3×3 convolutional layers and a sigmoid function on the AMLIF for computing the transmission map. Then, we use a global average pooling [15] on the AMLIF, followed by two fully connected layers and a sigmoid function to estimate the atmospheric light. After that, we compute the dehazed result (denoted as J_0) by re-formulating AS model in Eq. (1) as:

$$J_0(p) = \frac{I(p) - A_0 \times (1 - T_0(p))}{T_0(p)} , \tag{4}$$

where p denotes the pixel location; I is the input hazy image; A_0 is the computed atmospheric light; and T_0 is the estimated transmission map.

3.2.2 Prediction from Layer Separation Model

Apart from the atmospheric scattering model (see Eq. 1), we integrate the dehazed results from layer separation models together for improving the dehazing performance, since these models can learn the complementary dehazing information of the scattering model. Note that the image hazing process is pretty complicated and accurate layer decomposition in the single image dehazing task is non-trivial. In this regard, we empirically explore four specific layer formulations (with common mathematical operations on the layer composition) as the decomposition basis and use the attention mechanism to linearly combine these four bases to obtain the dehazed results respectively; Figure 4 shows how to predict dehazed results (denoted as J_1, J_2, J_3 and J_4) using the four-layer decomposition basis. For a specific layer decomposition, we apply the developed AFIM to generate AMLIF and then use the decomposition formulation to obtain the dehazed result from the AMLIF.

Specifically, we first consider the layer multiplication mechanism for the hazing layer decomposition model:

$$J_1(p) = I(p) \times R_1(p) , \tag{5}$$

where p is the pixel location; I is the hazy input; and J_1 and R_1 denote the two layers, which are decomposed from the I using the Eq. (5). Figure 4 (a) shows the architecture of predicting dehazed result (denoted as J_1) based on the Eq. (5) by taking AMLIF and I as the input. Specifically, we apply two 3×3 convolutional layers on the AMLIF for predicting R_1, and then compute the dehazed result J_1 by using Eq. (5) with the estimated R_1 and input I.

Secondly, we model the dehazing layer separation as a

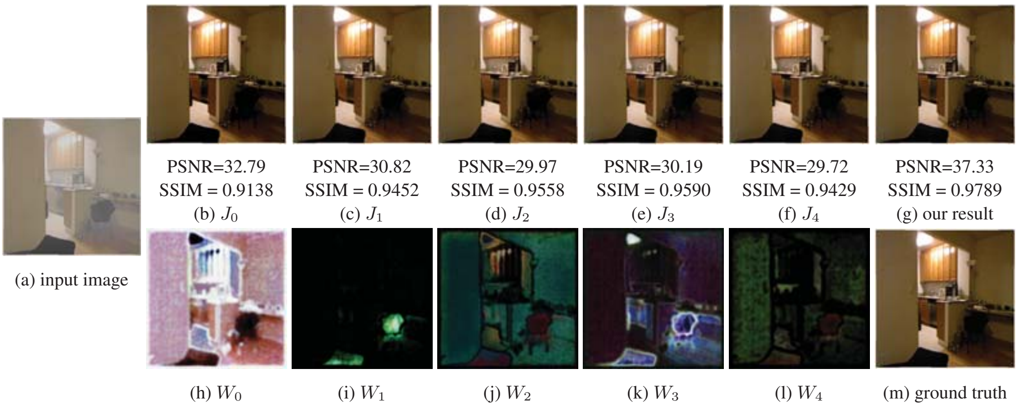


Figure 6: Visualization of dehazed results predicted by the the atmospheric scattering (AS) model (J_0) and the four layer separation models (J_1-J_4), as well as the corresponding attention weights learned in five dehazing models: W_0, W_1, W_2, W_3 and W_4. (a) Input hazy image; (b)-(f): J_0 to J_4; (g) our result; (h)-(l): W_0 to W_4; and (m) the haze-free ground truth.

classical linear combination with an addition operation:

$$J_2(p) = I(p) + R_2(p) \ , \tag{6}$$

where J_2 and R_2 denote two layers after performing the layer decomposition in the Eq. (6). With the linear combination formulation in Eq. (6), We estimate the dehazed result (denoted as J_2) by first utilizing two 3×3 convolutional layers on the AMLIF to compute R_2, and then adding R_2 into the input I, as shown in Figure 4 (b).

The third formulation is to explore the exponentiation operation for separating the hazy input I into J_3 and R_3:

$$J_3(p) = (I(p))^{R_3(p)} \ , \tag{7}$$

Figure 4 (c) shows how to obtain the dehazed result J_3 for Eq. (7). Specifically, we use two 3×3 convolutional layers on the AMLIF for predicting R_3, and then compute J_3 according to Eq. (7).

Our last layer separation for image dehazing is given by:

$$J_4(p) = log(1 + I(p) \times R_4(p)) \ , \tag{8}$$

where J_4 and R_4 are two decomposed layers for Eq. (8). We use two 3×3 convolutional layers on AMLIF to estimate R_4 and then Eq. (8) to compute J_4; see Figure 4 (d).

3.2.3 Final Result

After obtaining results of different hazing models, we leverage the attention mechanism [21, 40] to integrate these predictions for final result of our network. To do so, we learn five attention maps from the multi-layer integration features (AMIF) for different predictions by performing a 1× 1 convolutional layer, two 3× 3 convolutional layers, a 1× 1 convolutional layer, and a softmax layer; see Figure 2. Then, the final result (denoted as J_f) is computed as:

$$\begin{aligned} J_f = W_0 \times J_0 + W_1 \times J_1 + W_2 \times J_2 \\ + W_3 \times J_3 + W_4 \times J_4 + W_5 \times J_5 \ , \end{aligned} \tag{9}$$

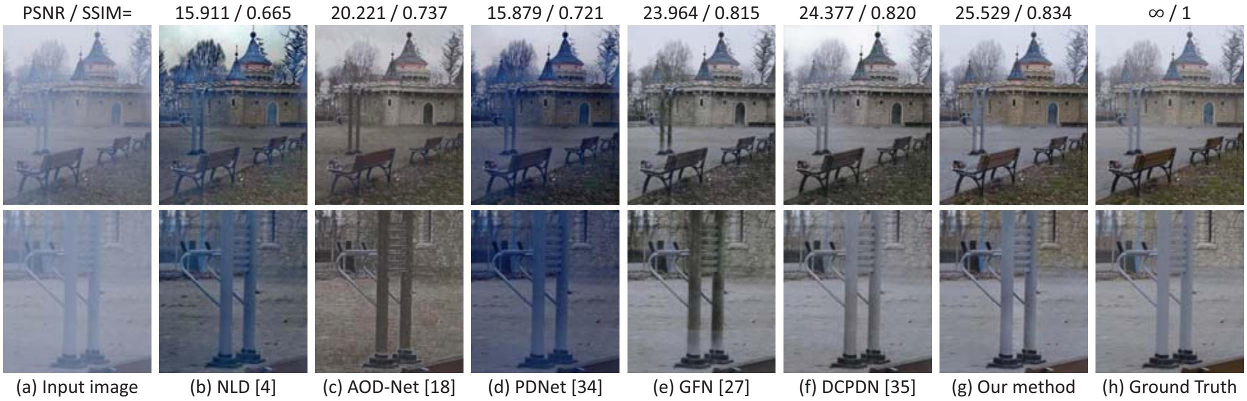
where W_0, W_1, W_2, W_3 and W_4 are the learned attention maps for dehazed results J_0, J_1, J_2, J_3 and J_4, respectively.

3.3. More analysis

Different models' Result Visualization. Figures 6 (b)-(f) demonstrate dehazed results of the atmospheric scattering (AS) model (J_0) and four layer separation models (J_1 to J_4). As can be seen, the AS model (J_0) can better recover the input hazy image than other layer separation models (J_1 to J_4), which are also verified by its higher PSNR/SSIM values. More importantly, when removing the haze, the AS model tends to over-smooth parts of non-haze background details, and those details are preserved in the dehazed results of layer separation models respectively, which demonstrates that our layer separation models can learn the complementary dehazing information of the AS model.

Attention map visualization. Figures 6 (h)-(i) visualize the learned attention weights (W_0, W_1, W_2, W_3 and W_4) of five dehazing models. Obviously, for each dehazing model, the learned attention map has smaller weights on their blurred regions, while automatically highlighting these regions, which are better recovered by this image dehazing model. Furthermore, since there are complementary information among the dehazed results of the five dehazing modes, the attention maps (W_0, W_1, W_2, W_3 and W_4) can automatically select the best one among all the five dehazed results to predict the final result of our method by highlighting different regions of the input image, as shown in W_0, W_1, W_2, W_3 and W_4. Hence, our method integrating these five dehazing models by using these learned attention maps in our method incurs a better performance of image dehazing, as shown in Figure 6 (g) (compared to the haze-free ground truth in Figure 6 (m)).

Why only four models. The main goal of our layer separation models is to separate the input hazy image into two layers (see Eq. 2): one is with haze-free background detail-



(a) Input image (b) NLD [4] (c) AOD-Net [18] (d) PDNet [34] (e) GFN [27] (f) DCPDN [35] (g) Our method (h) Ground Truth

Figure 7: Haze removal results by various methods on a real-world photo in O-HAZE [1]. Please zoom in for a better view.

s while another layer contains only haze information. Our four layer separation models (see Figure 2) contain common mathematical operations for two-layer combinations, and they are "+(-)" in J_1, "$\times(\div)$" in J_2, exponentiation in J_3, and logarithm in J_4. Furthermore, for better approximating the mathematical formulation in the presence of haze, we use the attention mechanism to produce weighting maps for linearly combining all these four models in the final haze-free predictions, and these weights are optimized when minimizing the training loss of our network, which is computed from many hazy and haze-free image pairs of the training set. Our superior performance on real-world and synthetic benchmarks have demonstrated the effectiveness of our four-layer separation models for image dehazing.

3.4. Training Strategy

Loss function. As shown in Figure 2, our network adds haze-free supervision on the dehazed results (J_0, J_1, J_2, J_3 and J_4) from atmospheric scattering model and layer separation models, as well as our final result (J_f). When predicting the dehazed result based on the scattering model, we also add a transmission map supervision on the estimated transmission map and an atmospheric light supervision on the computed atmospheric light. The total loss Θ is:

$$\begin{aligned}\Theta = \alpha_0 ||J_0 - G_H||_1 + \alpha_1 ||J_1 - G_H||_1 + \alpha_2 ||J_2 - G_H||_1 \\ + \alpha_3 ||J_3 - G_H||_1 + \alpha_4 ||J_4 - G_H||_1 + \alpha_4 ||J_f - G_H||_1 \\ + \alpha_6 ||T_0 - G_T||_1 + \alpha_7 ||A_0 - G_A||_1 \ ,\end{aligned} \tag{10}$$

where G_A, G_T and G_H denote ground truth of the atmospheric light, transmission map, and single-image dehazing; $||.||_1$ denotes the L_1 norm based loss for computing difference between the prediction and the corresponding ground truth. α_0, α_1, α_2, α_3, α_4, α_5, α_6 and α_7 are the weight of each L_1 loss. We empirically set α_6 as 10, while other weights are fixed as 1 in both training and testing stages.

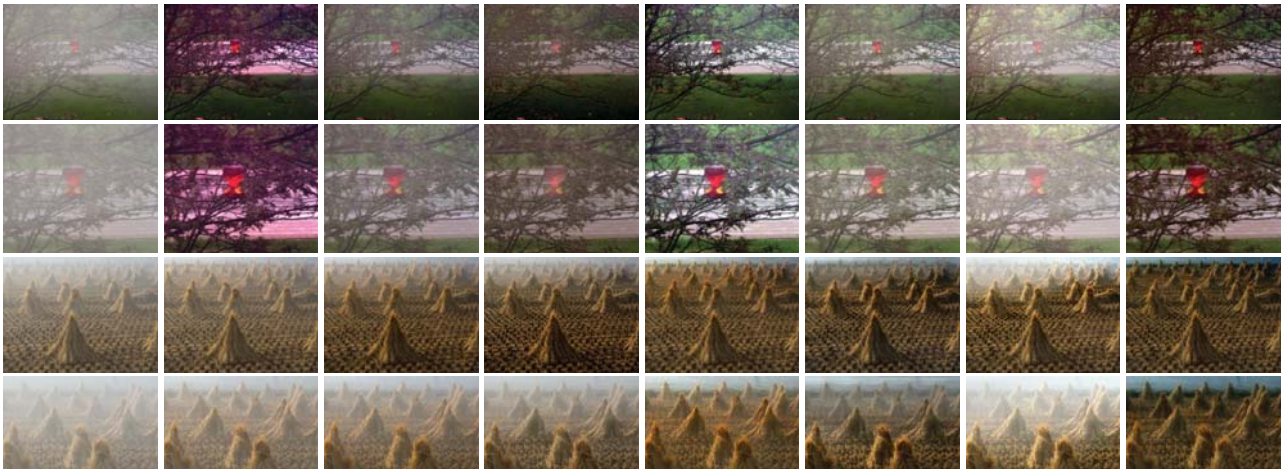
Training parameters. We initialize the parameters of the basic CNN by a pre-trained ResNeXt [34] on the ImageNet, and other parameters by Gaussian random noise. We randomly cropped 256× 256 image patches from the entire training images and adopt the Adam optimizer with iteration number of 20, 000 for training. The learning rate is adjusted by the poly policy [24] with the initial learning rate of 0.0002 and power of 0.9. We use a mini-batch size of 16 and 4 hours to train our model using a single NVIDIA GTX 1080Ti GPU based on the PyTorch library. Processing a 640 × 480 image takes around 0.032 sec.

4. Experimental Results

We compare our dehazing network against state-of-the-art methods, including DCP [13], NLD [4], MSCNN [27],, DehazeNet [6], AOD-Net [19], GFN [28], DCPDN [36], and PDNet [35]. Furthermore, we employ three widely-used metrics for quantitative comparisons, and they are peak signal to noise ratio (PSNR) [41], structural similarity index (SSIM) [32], and CIEDE2000 [39]. Our code, trained models, and dehazed results on the benchmark datasets are publicly available at https://github.com/zijundeng/DM2F-Net.

4.1. Results on Real-world Images

NTIRE 2018 outdoor dehazing challenge (O-HAZE). According to the final ranking of O-HAZE challenge [1], top 5 PSNR/SSIM results are 24.598/0.777 (Team: BJTU), 24.232/0.687 (Team: KAIST-VICLAB), 24.029/0.775 (Team: Scarlet Knights), 23.877/0.775 (Team: FKS), and 23.207/0.770 (Team: Ranjanisi). We use the training data of O-HAZE dataset [3] to train our network and test on its testing data, Table 1 reports the PSNR and SSIM results of our network and state-of-the-arts. Obviously, our method (PNSR/SSIM: 25.188/0.777) outperforms the top 5 teams and compared dehazing methods in terms of the PSNR and SSIM on a large margin. It demonstrates that our method can better restore the outdoor real-world hazy scenes, which is also verified by the visual comparisons in Figure 7.



(a) Input haze image (b) NLD [4] CVPR' 16 (c) DehazeNet [6] TIP' 16 (d) AOD-Net [19] ICCV' 17 (e) PDNet [35] ECCV' 18 (f) GFN [28] CVPR' 18 (g) DCPDN [36] CVPR' 18 (h) Our method

Figure 8: Dehazing real-world hazy photos using various methods (b)-(h). Please zoom in for a better illustration.

Table 1: Comparisons on real-world & synthetic dehazing datasets.

	O-HAZE [3]		HAZERD [39]		TestA-DCPDN [36]		SOTS [28]	
method	PSNR	SSIM	CIEDE2000	SSIM	PSNR	SSIM	PSNR	SSIM
DM^{2}F-Net (ours)	**25.188**	**0.777**	**12.9285**	**0.656**	**35.61**	**0.9829**	**34.29**	**0.9844**
DCPDN [36]	22.777	0.742	14.6251	0.546	29.27	0.9533	28.13	0.9592
GFN [28]	22.578	0.737	16.3619	0.511	25.59	0.9398	22.30	0.8800
PDNet [35]	17.403	0.658	16.9360	0.495	21.98	0.9083	22.83	0.9210
AOD-Net [19]	19.586	0.679	16.6743	0.500	20.46	0.8379	20.86	0.8788
DehazeNet [6]	16.207	0.666	17.1261	0.479	19.92	0.8575	21.14	0.8500
MSCNN [27]	19.068	0.765	13.7952	0.624	17.98	0.8203	17.57	0.8100
NLD [4]	16.610	0.750	16.4010	0.577	16.95	0.7959	17.27	0.7500
Li et al.[22]	14.43	0.583	15.91	0.623	15.34	0.781	17.05	0.794
Meng et al.[25]	23.92	0.725	16.85	0.578	24.33	0.904	23.49	0.936
DCP [13]	16.586	0.735	17.9014	0.534	13.91	0.8642	16.62	0.8179

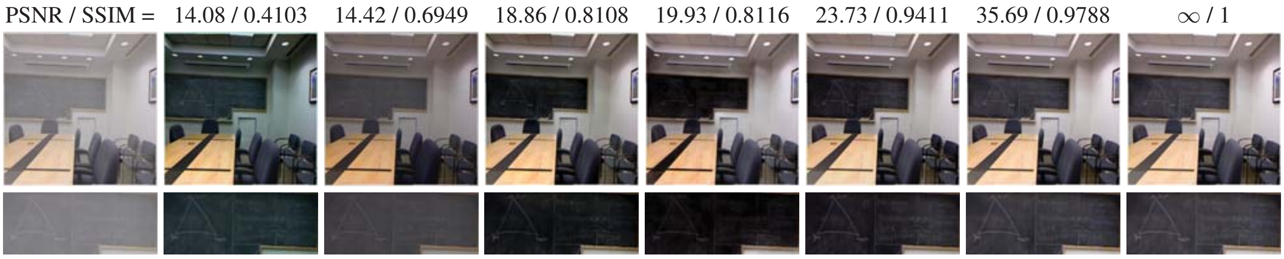
HAZERD. The HAZERD dataset [39] only has 15 hazy outdoor images with more realistic haze for testing. Hence, we train our network and competitors on the synthetic RESIDE dataset [20, 28] and test on the HAZERD dataset. Table 1 reports the quantitative results, and our network has larger SSIM and smaller CIEDE2000 than other competitors, demonstrating that our method has superior dehazing performance on realistic images of HAZERD.

Collected real hazy photos. Additionally, Figure 8 shows the visual comparisons on real-world hazy photos we collected. As revealed in Figure 8, NLD suffers from the color distortions, while DehazeNet, AOD-Net, PDNet, GFN, and DCPDN again tend to leave haze or darken some regions. Contrarily, our method predicts better dehazed results in terms of effectively removing the haze while producing realistic colors, as shown in these blown-up views of Figure 8.

4.2. Results on Synthetic Images

We evaluate our network on two synthetic benchmarks: "TestA-DCPDN" [36] and "SOTS" [20, 28] and report our results using the same training strategy of [36, 28]. To do the fair comparisons, we obtained the results of compared methods by obtaining their released code and re-training deep networks by using training sets of two dehazing benchmarks. Table 1 also reports average PSNR and SSIM values of different dehazing methods on "TestA-DCPDN" and "SOTS". Deep learning-based dehazing competitors have larger PNSR and SSIM values than the hand-crafted prior based methods (DCP & NLD). Furthermore, our method has the largest PSNR and SSIM values on TestA-DCPDN [36] and SOTS [20, 28] among all the dehazing networks, which demonstrate that our method has a superior performance of recovering the clean images for the two dehazing datasets.

Figure 9 presents visual comparisons on a synthetic image of two benchmarks. NLD overestimates the haze thickness and thus causes color distortion. Although improving the dehazing performance than NLD, these dehazing networks (e.g., AOD-Net, GFN, PDNet, and DCPDN) tend to leave there are still some remaining haze or darken several areas in the results; see Figures 9 (c)-(f). In contrast, our dehazed result (Figure 9 (g)) is closest to the haze-free ground truth image (see Figure 9 (h)). Overall, the dehazed result of our network have higher visual quality and fewer color

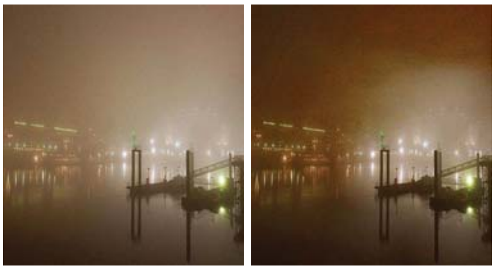


(a) Input image (b) NLD [4] (c) AOD-Net [19] (d) PDNet [35] (e) GFN [28] (f) DCPDN [36] (g) Our method (h) Ground truth

Figure 9: Haze removal on a synthetic hazy photo. Please zoom in for a better illustration.

Table 2: Average PSNR and SSIM values in ablation study.

	TestA-DCPDN [36]		SOTS [28]	
method	PSNR	SSIM	PSNR	SSIM
basic+AS	34.36	0.9679	32.42	0.9717
basic+J_1	30.57	0.9558	28.93	0.9486
basic+J_2	31.70	0.9656	30.92	0.9654
basic+J_3	32.30	0.9714	32.64	0.9758
basic+J_4	29.98	0.9510	28.85	0.9446
ours_w/o_AFIM	34.71	0.9810	33.93	0.9823
DM^{2}F-Net (ours)	**35.61**	**0.9829**	**34.29**	**0.9844**



Input haze image Our method

Figure 10: An example of a failure case.

distortions, which are also verified by the largest PSNR and SSIM value of our method in Figure 9.

4.3. Ablation Study

We perform an ablation study experiment to verify the major components of our network. Here, we consider six baseline networks, and report their results on TestA-DCPDN [36] and SOTS [20, 28]. The first baseline (denoted as "basic+AS") is constructed by only using the atmospheric scattering model of our network (see Figure 2) for dehazing; Then, we construct another four baselines by only taking J_1 ("basic+J_1"), J_2 ("basic+J_2"), J_3 ("basic+J_3") and J_4 ("basic+J_4") as the results of our network, respectively. The last baseline (denoted as "ours_w/o_AFIM") is built by removing the attentional feature integration module (AFIM) from our network (Figure 2) to verify the AFIM.

Table 2 compares our method against six baselines. Apparently, our method has better dehazed results than "basic+AS", which indicates that the layer separation model in our method can help to improve the dehazed results. Similarly, our method has a superior PSNR and SSIM performance than all four specific layer decompositions ("basic+J_1", "basic+J_2", "basic+J_3" and "basic+J_4"), demonstrating that the atmospheric scattering model in our method also contributes better results to our dehazing network. Lastly, our method has larger PSNR and SSIM values than "ours_w/o_AFIM", which shows that leveraging AFIM to integrate features at different CNN layers for the clean image prediction can also help our method to obtain superior dehazing results.

Failure cases. Like other works (e.g., [23]), our method might not work well for night hazy images; see an example input and result shown in Figure 10. It is because existing training datasets do not contain similar hazy conditions. This can be alleviated by collecting more data samples.

5. Conclusion

This work presents a multi-model fusing network for boosting the single-image dehazing. Our key idea is to design a new deep multi-modal fusion framework that allows us to simultaneously explore multiple dehazing models (including an atmospheric scattering (AS) model and four dehazing models) to combine their strengths and maximize the methods dehazing capability. On the contrary, existing dehazing methods mainly examine the AS model and tend to fail in various real-world complex hazing situations. Experimental results demonstrate the superior performance of our method over the state-of-the-arts.

Acknowledgments. The work is supported by CUHK Research Committee Funding (Direct Grants) under project code - 4055103, Research Grants Council of the Hong Kong Special Administrative Region (No. CUHK 14201717), Science and Technology Plan Project of Guangzhou (No.201704020141), Shenzhen Science and Technology Program (Project no. JCYJ20170413162617606), NSFC (Grant No. 61772206, U1611461, 61472145), Guangdong R&D key project of China (Grant No. 2018B010107003), Guangdong High-level personnel program (Grant No. 2016TQ03X319), Guangdong NSF (Grant No. 2017A030311027), and Guangzhou key project in industrial technology (Grant No. 201802010027).

References

[1] Cosmin Ancuti, Codruta O. Ancuti, and Radu Timofte. N-TIRE 2018 challenge on image dehazing: Methods and results. In *CVPR Workshops*, pages 891–901, 2018. 6

[2] Codruta Orniana Ancuti and Cosmin Ancuti. Single image dehazing by multi-scale fusion. *IEEE Transactions on Image Processing*, 22(8):3271–3282, 2013. 1

[3] Codruta O. Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images. In *CVPR Workshops*, pages 754–762, 2018. 6, 7

[4] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *CVPR*, pages 1674–1682, 2016. 2, 6, 7, 8

[5] Dana Berman, Tali Treibitz, and Shai Avidan. Air-light estimation using haze-lines. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–9. IEEE, 2017. 1

[6] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. DehazeNet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. 2, 6, 7

[7] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018. 2

[8] Ziang Cheng, Shaodi You, Viorela Ila, and Hongdong Li. Semantic single-image dehazing. *arXiv preprint arXiv:1804.05624*, 2018. 1

[9] Raanan Fattal. Single image dehazing. *ACM Trans. on Graphics (SIGGRAPH)*, 27(3):72:1–10, 2008. 1

[10] Raanan Fattal. Dehazing using color-lines. *ACM Trans. on Graphics (SIGGRAPH)*, 34(1):13:1–14, 2014. 1, 2

[11] Adrian Galdran, Aitor Alvarez-Gila, Alessandro Bria, Javier Vazquez-Corral, and Marcelo Bertalmıo. On the duality between Retinex and image dehazing. In *CVPR*, pages 8212–8221, 2018. 1

[12] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *CVPR*, pages 2862–2869, 2014. 2

[13] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2011. 1, 2, 6, 7

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 2, 3

[15] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, June 2018. 4

[16] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. Depth-attentional features for single-image rain removal. In *CVPR*, pages 8022–8031, 2019. 2

[17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, pages 1125–1134, 2017. 2

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems (NIPS)*, pages 1097–1105, 2012. 3

[19] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. AOD-Net: An all-in-one network for dehazing and beyond. In *ICCV*, pages 4770–4778, 2017. 2, 6, 7, 8

[20] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019. 7, 8

[21] Guanbin Li, Yuan Xie, Liang Lin, and Yizhou Yu. Instance-level salient object segmentation. In *CVPR*, pages 247–256, 2017. 5

[22] Kunming Li, Yu Li, Shaodi You, and Nick Barnes. Photo-realistic simulation of road scene for data-driven methods in bad weather. In *ICCV Workshop*, pages 491–500, 2017. 7

[23] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In *CVPR*, pages 8202–8211, June 2018. 1, 2, 8

[24] Wei Liu, Andrew Rabinovich, and Alexander C Berg. ParseNet: Looking wider to see better. In *ICLR*, 2016. 6

[25] Gaofeng Meng, Ying Wang, Jiangyong Duan, Shiming Xiang, and Chunhong Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *ICCV*, pages 617–624, 2013. 7

[26] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems (NIPS)*, pages 91–99, 2015. 2

[27] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, pages 154–169, 2016. 2, 6, 7

[28] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *CVPR*, pages 3253–3261, 2018. 1, 2, 6, 7, 8

[29] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *ICLR*, 2015. 2

[30] Matan Sulami, Itamar Glatzer, Raanan Fattal, and Mike Werman. Automatic recovery of the atmospheric light in hazy images. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–11. IEEE, 2014. 1

[31] Robby T Tan. Visibility in bad weather from a single image. In *CVPR*, pages 1–8. IEEE, 2008. 1

[32] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6

[33] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In *Advances in neural information processing systems (NIPS)*, pages 341–349, 2012. 2

[34] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *CVPR*, pages 5987–5995, 2017. 6

[35] Dong Yang and Jian Sun. Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In *ECCV*, pages 702–717, 2018. 1, 2, 6, 7, 8

[36] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *CVPR*, pages 3194–3203, 2018. 1, 2, 6, 7, 8

[37] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*, pages 695–704, 2018. 2

[38] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 2

[39] Yanfu Zhang, Li Ding, and Gaurav Sharma. HAZERD: an outdoor scene dataset and benchmark for single image dehazing. In *ICIP*, pages 3205–3209. IEEE, 2017. 6, 7

[40] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *ECCV*, pages 121–136, 2018. 5

[41] Lei Zhu, Chi-Wing Fu, Michael S Brown, and Pheng-Ann Heng. A non-local low-rank framework for ultrasound speckle reduction. In *CVPR*, pages 5650–5658, 2017. 2, 6

[42] Lei Zhu, Chi-Wing Fu, Yueming Jin, Mingqiang Wei, Jing Qin, and Pheng-Ann Heng. Non-local sparse and low-rank regularization for structure-preserving image smoothing. In *Computer Graphics Forum*, volume 35, pages 217–226, 2016. 2

[43] Lei Zhu, Chi-Wing Fu, Dani Lischinski, and Pheng-Ann Heng. Joint bi-layer optimization for single-image rain streak removal. In *ICCV*, pages 2526–2534, 2017. 2

[44] Qingsong Zhu, Jiaming Mai, Ling Shao, et al. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, 2015. 1