Implicit Euler ODE Networks for Single-Image Dehazing

Jiawei Shen*[1], Zhuoyan Li*[1], Lei Yu†[1], Gui-Song Xia[2], Wen Yang[1]

[1]School of Electronic and Information, Wuhan University
[2]School of Computer Science, Wuhan University
Wuhan 430072, China

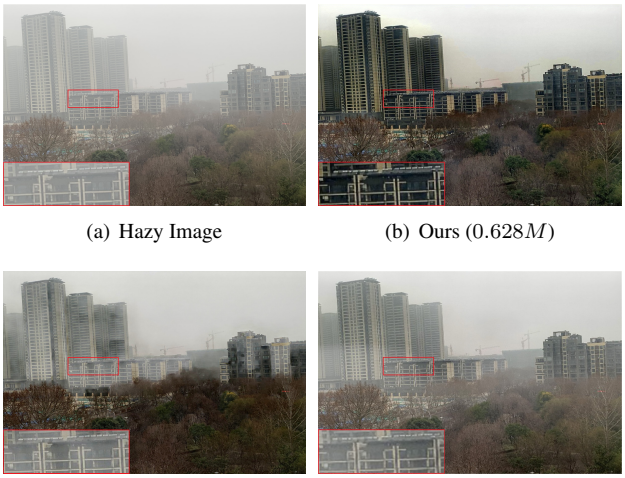
{shenjiawei0116, xzflzy}@gmail.com, {ly.wd, guisong.xia, yangwen}@whu.edu.cn

Abstract

Deep convolutional neural networks (CNN) have been applied for image dehazing tasks, where the residual network (ResNet) is often adopted as the basic component to avoid the vanishing gradient problem. Recently, many works indicate that the ResNet can be considered as the explicit Euler forward approximation of an ordinary differential equation (ODE). In this paper, we extend the explicit forward approximation to the implicit backward counterpart, which can be realized via a recursive neural network, named IM-block. Given that, we propose an efficient end-to-end multi-level implicit network (MI-Net) for the single image dehazing problem. Moreover, multi-level fusing (MLF) mechanism and residual channel attention block (RCA-block) are adopted to boost performance of our network. Experiments on several dehazing benchmark datasets demonstrate that our method outperforms existing methods and achieves the state-of-the-art performance.

1. Introduction

Images captured from the harsh environment are often hazy and exhibit a reduced visibility of scenery [15] with the loss of contrast, color fidelity and edge information. Such degradation of images may greatly decrease the accuracy and robustness for the subsequent high-level computer vision tasks [8,30], which makes the single-image dehazing an urgent but challenging task [7, 13]. Existing approaches, either exploiting the physical prior [1, 15, 25, 32] or learning the inverse mapping of degradation with large datasets [3, 11, 20, 28], have been extensively investigated.



(a) Hazy Image (b) Ours ($0.628M$)

(c) FFA-Net ($4.46M$) (d) GRID-Net ($0.958M$)

Figure 1. Visual comparison of our method and other state-of-the-art works, FFA [26] and GRID-Net [24]. The numbers in brackets indicate the size of the networks.

For CNN-based image dehazing methods, residual block (Resblock) [16] is widely applied [24, 26, 28, 42]. Zhang et al. [42] introduce a densely connected pyramid network with residual block, Liu et al. [24] has leveraged the residual dense block in their GRID-Net and Qin et al. [26] develop a residual attention block in their FFA-Net. Overall, residual block is not only widely applied in high-level tasks but also low-level tasks as image dehazing. On the other hand, we notice that many recent works relate ResNet with explicit Euler forward approximation of ODE.

Researching CNNs' theoretical properties and behavior has drawn considerable attention from the perspective of ODE [6, 29, 34]. Chen et al. [6] introduce the relation between CNN and ODE. Ruseckas et al. [29] prove that not only residual networks, but also feedforward neural networks with small nonlinearities can be related to the dif-

∗ The first two authors contributed equally and should be regarded as co-first authors.
† Corresponding author.

ferential equations. Thorpe et al. [34] propose that residual neural network model is a discretization of an explicit Euler ODE and the deep-layer limit coincides with a parameter estimation problem for a nonlinear ordinary differential equation. These works are all closely associated with ResNet, which is an explicit Euler scheme. However, focusing on parameters convergence and system stability, implicit Euler has been proved to be better than explicit one since implicit Euler is unconditionally stable [19]. Moreover, considering that the original objective of Resblock applied in deep neural network is to overcome vanishing gradient problem, it's suboptimal to directly apply Resblock in low-level tasks as image dehazing.

In this work, we propose a novel efficient network, multi-level implicit network (MI-Net), for single image dehazing. Specially, the structure of MI-Net is shown in Fig. 2. We introduce an implicit block (IM-block) combining the merits of implicit Euler scheme and CNN to realize implicit discretization of an implicit Euler ODE issue. In network's architecture, we cascade three IM blocks to learn the mapping relation between hazy image and clean image. Inspired from [23], we integrate features from three IM blocks with different weight coefficients generated from MLF block. Noting that the fused feature would inevitably incorporate artifact, the fused feature is fed into RCA-block to suppress artifact of channels. As we shall see in the experiments, our method achieves the best balance between the performance and size.

In summary, our contributions are as follows:

- We propose a simple yet effective implicit scheme framework for single image dehazing, which simplifies the net significantly but boosts the dehazing performance compared to the residual framework.
- We propose a residual channel attention block based on attention mechanism to alleviate the artifact but retain abundant texture features.
- Experimental results on several widely used benchmark datasets show superior performance of our method compared to state-of-the-art methods, which verifies the effectiveness of our method.

2. Related Work

2.1. Image Dehazing

Prior-driven Methods: The procedure of classical dehazing task is the reverse procedure of the atmospheric scattering model described as $I(p) = t(p)R(p) + A(1 - t(p))$, where $I(p)$ is the hazy image, $t(p)$ represents transmission map, $R(p)$ is the clean image, p represents the pixel location and A is the global atmospheric light constant value. The traditional methods use assumptions which derive from the statistical characteristics of hazy image to compensate for the loss information in the atmospheric model. Tan et al. [33] builds Markov random field cost function to obtain the clean image by maximizing the contrast of corrupted image ,assuming that the contrast of clean image is higher than that of hazy image and the smoothness of global atmospheric light. He et al. [15] proposes the dark channel prior, which is based on a statistical observation that at least one of color channels in the non-sky regions of a hazy image has a value close to zero, to estimate the transmission map. Fattal et al. [12] uses a color-line method line for dehazing based on observation that the color distribution of small patches of image in the RGB space is one-dimensional. Although these methods could handle the problem to some extent, the priors they based on would be invalid in some real scenes, which enables the methods struggle.

CNN-based Methods: Recently, many CNN-based methods have been applied in the dehazing by leveraging the advancement of powerful GPU and large-scale datasets. Early CNN-based methods still are based on global atmospheric scattering model and recover clean image by estimating the transmission map and global atmospheric light. Ren et al. [27] design a coarse-to-fine multiscale network to estimate a refined transmission map. Cai et al. [3] develop a Dehaze-Net embedded with feature layers for prediction of transmission map. However, the transmission map is susceptible to the noise and hence reduces the quality of dehazing performance. Therefore, end-to-end CNNs have been proposed to output a clean image from a hazy image directly without global atmospheric scattering model. Chen et al. [5] use an encoder-decoder net with smoothed dilated convolution in the net to alleviate the gridding artifacts. Zhang et al. [42] propose a densely-connected pyramid densely network (DCPDN) to jointly learn clean image, atmospheric light and transmission map. Although the DCPDN improves the performance to some extent, it enlarges the size of model greatly. Deng et al. [10] develop a net that fuses the atmospheric scattering model with extracted haze together to improve the dehazing results. We note that most recent works still rely on deepening the net to improve the quality whatever the principles they based on. Being a low-level task, image dehazing depends more on low-level features compared to high-level features. We consider the dehazing network could be simplified into a low-level network.

2.2. CNN from Ordinary Differential Equation

Along with the development of CNN, many researches study networks' theoretical properties from the perspective of ordinary differential equation (ODE) [2, 9, 17]. Weinan et al. [38] firstly view ResNet [16] as an approximation to ODE, which exploits the possibility of using computational theory from ResNet's dynamic system. Similarly, Chang et al. [4] connect ResNet with nonlinear ODE, and extend

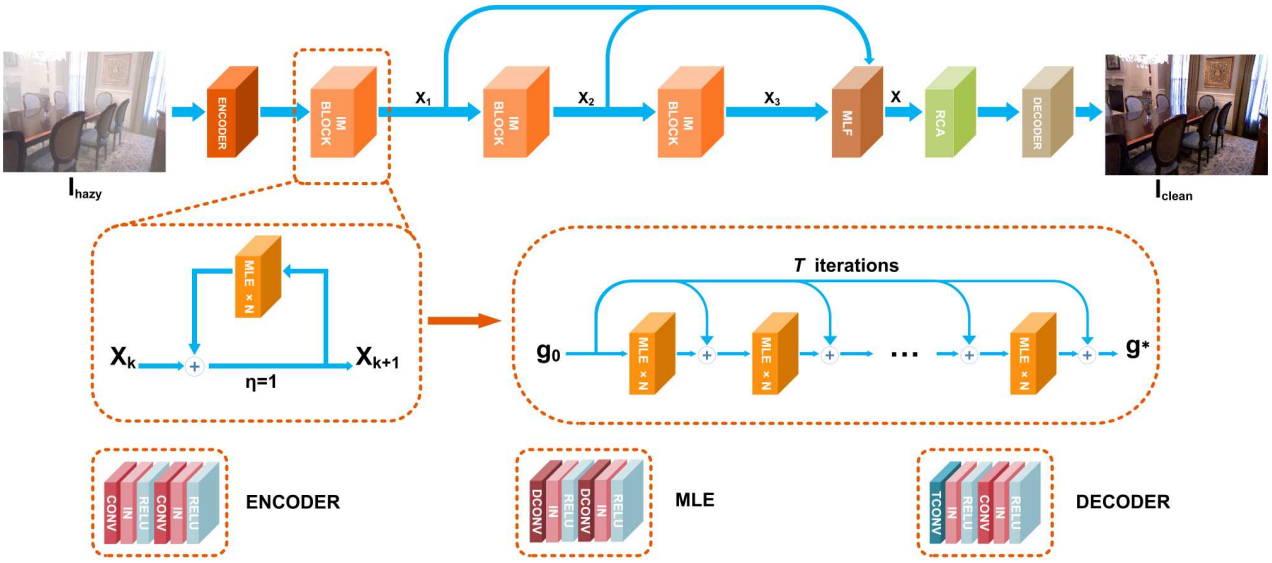


Figure 2. The architecture of multi-level implicit network (MI-Net). IM-block is illustrated in dashed box. IN and DCONV in MLE denotes instance normalization and dilated convolution respectively. $\times N$ in MLE represents the dilated rate and the dilated rates are different for each IM-block. TCONV in DECODER represents fractionally-strided convolution. MLF and RCA-block will be illustrated in proposed methods section.

three reversible network architectures. Chen et al. [6] propose ODE-Net that combines ODE and CNN. Based on above, we speculate that network designed with certain theoretical basis has great potential for exploration. Thorpe et al. [34] considered that the deep layer limit coincides with a parameter estimation problem for nonlinear ODE [9]. In reality, convergent parameters means stable system which is closely related to model performance. Haber et al. [14] also associated gradient explosion and disappearance about neural networks with the stability of discrete ODE and suggested that stable networks generalize better. In fact, although implicit Euler scheme has larger computational cost compared to explicit Euler scheme, implicit one allows greater step size and is more stable since implicit scheme is unconditionally stable. Moreover, for low-level task as image dehazing, the increased computational cost could be ignored. Considering these all factors, we adopt the implicit Euler scheme in CNN to determine the dehazing model.

3. Proposed Methods

As shown in Fig. 2, we build an end-to-end network to establish mapping relation between hazy image and clean image. In this section, we illustrate the function of each component in MI-Net's architecture. Firstly, the hazy image input I_{hazy} will be encoded from RGB space to feature space. The encoded features will be enhanced by three cascaded IM-block. Additionally, the features x_1, x_2, x_3 output from each IM-block will be fused in MLF block. MLF block is an operation block, three features are input into

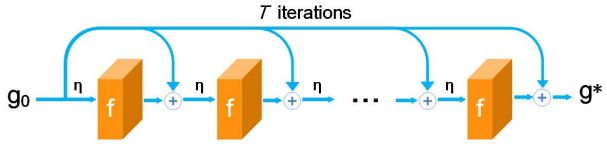


Figure 3. IM-block: Implicit scheme with specific function f.

MLF then MLF learns the weight coefficient of each feature x_i, and the output of MLF is a fusion of three features combined with different weights. The fused feature is input to RCA-block to refine the feature. Lastly, the refined feature will be decoded into RGB space to get the clean image I_{clean}.

3.1. From Resblock to IM-block

Convolutional neural network (CNN) has been applied to tackle the single image dehazing problem. By increasing the depth of CNN, it is possible to improve the perceptual fields and increase the expressive ability, thus leading to better dehazing performance [11, 26]. Residual neural network (ResNet) [16] has been widely used in both high-level and low-level tasks to overcome the vanishing gradient problem for deep neural networks. Recently, from the point view of ODE, Resblock can be considered as the explicit Euler forward discretization of the continuous-time ODE: $\dot{x}(t) = f(x(t))$ [6] and the relation between two consecutive layers can be expressed as

$$\textbf{Resblock:} \quad x_{k+1} = x_k + \eta f(x_k), \tag{1}$$

I_{x_1} I_{x_2} I_{x_3}

Figure 4. The visual comparison of $I_{x_1}, I_{x_2}, I_{x_3}$. I_{x_1} has clear detail information but still has hazy area. I_{x_3} has gridding effect and color distortion area, i.e. the edge of ceiling lamp. Please zoom in for a better illustration.

where η denotes the discretization step.

On the other hand, implicit Euler backward scheme [19] has been proven to be more stable and accurate compared to the explicit counterpart (1). From this point, the implicit scheme that bridges two consecutive layers can be written as

$$\textbf{IM-block:} \quad x_{k+1} = x_k + \eta f(x_{k+1}). \tag{2}$$

However, the above implicit algebraic equation cannot be solved analytically in general, and Newton iteration method is generally used to find x_{k+1}. Letting $g_0 \triangleq x_k$, the following iterations can be applied to approximate the solution:

$$g_{\kappa+1} = g_0 + \eta f(g_\kappa) \tag{3}$$

and $x_{k+1} = g^*$, with g^* the equilibrium point of (3), i.e.

$$g^* = g_0 + \eta f(g^*) \tag{4}$$

Then we have the following theorem,

Theorem 1 *Let $\partial_x f(x) \in \mathbb{R}^{n\times n}$ be the partial derivative of the vector field $f(x)$ at x, and note λ_i as its eigenvalue for $1 \le i \le n$. Then* (3) *is stable if*

$$|\lambda_i| < 1/\eta \tag{5}$$

for all $1 \le i \le n$.

The proof is straightforward. The above theorem implies that the vector field $f(x)$ should be well defined to guarantee the convergence of (3). Thus in our proposed MI-Net, the instance normalization [35] IN is exploited to ensure (5), as shown in Fig. 2. On the other hand, iterative approximation usually uses a large finite number as infinite iterations and we illustrate the unfolded form of (3), named IM-block in Fig. 3, which is exploited as the basic block instead of ResNets to construct our proposed MI-Net.

One can easily find that the ResNet (1) actually equals to IM-block when it contains only one recursion (3) with $\kappa = 0$. While the implementation of IM-block shown in Fig. 3 exhibits many recursions with shared weights for different layers. Theoretically, IM-block is equivalent to one layer implicit Euler approximation (2). In some sense, the IM-block could possess the perceptual field as large as the whole image but with a very shallow depth, and thus could be a more efficient structure than ResNet to capture low-level features. This property is favorable for low-level tasks [31] and motivates us to turn to IM-block for the image dehazing problem.

Consequently, we stack three IM-blocks as shown in Fig. 2 to capture features with different level of scales. To achieve this target, the multi-level feature fusion strategy is exploited in next subsection.

3.2. Multi-Level Feature Fusion

Dilated convolutions [41] support exponential expansion of the receptive field without loss of resolution or coverage, which can be utilized in IM-block to further enlarge the receptive field in MI-Net. Thus we adopt a dilated convolutional block to serve as f in (3). To overcome the gridding effect of dilated convolution, Wang et al. [36] proposed hybrid dilated convolution (HDC). Therefore, we adopt HDC structure that with coprime dilated rates, e.g. 1,2,5, to gain the best result for each IM-block. In order to help the convergence of (3), we adopt instance norm layers instead of setting the discretization step η a small value [22]. And the layers with dilated convolutions and instance normalizations are grouped as the multi-level extraction (MLE) block, as shown in Fig. 2, where the DCONV represents dilated convolutions and the IN layer denotes instance normalization [35].

However, the HDC structure can only suppress the gridding effects, color distortion still exists in the third IM-block. We visualize the difference among features of each IM-block, we extract the output features of IM-blocks x_1, x_2, x_3 with $x_i \in \mathbb{R}^{c\times h\times w}$, where c represents channel number, h and w represents image's height and width of input images. Without being fed into MLF, the features are directly input into DECODER block to transform features into images, i.e $I_{x_1}, I_{x_2}, I_{x_3}$. The corresponding reconstructions are plotted in Fig. 4. Note that feature from MI-Net at first IM block contains detail information but the non-haze details of image reconstructed from this feature would still be corrupted with haze severely. Third IM block of MI-Net has a larger receptive field but ignore background details. Hence, we adopt the attention based multi-level fusion block (MLF), and this block leverages convolutional layer to obtain a weight coefficient matrix $W \in \mathbb{R}^{3\times h\times w}$ of each pixel of feature $x \in \mathbb{R}^{c\times h\times w}$. In this case, the gridding effect area, the hazy area and the color distortion area will have a small weight coefficient but the weight coefficient of clean area will be larger.

Specially, as shown in Fig. 2, MLF block obtains output features of three IM-blocks in three different levels x_1, x_2, x_3, the concatenation of three features $X = [x_1, x_2, x_3]$ where $X \in \mathbb{R}^{3c\times h\times w}$ is fed into MLF layer to output weight coefficient $W = [W_1, W_2, W_3]$ where

Table 1. Quantitative comparisons of image dehazing on SOTS dataset from RESIDE, TestA and MiddleBury.

PSNR (dB)	DCP [15]	AOD-Net [20]	DCPDN [42]	GFN [28]	GRID-Net [24]	FFA-Net [26]	FD-GAN [11]	OURS
SOTS	16.62	20.86	28.13	21.14	32.16	**36.12**	23.15	35.51
TestA	13.91	20.46	23.27	20.02	22.33	19.96	18.82	**29.45**
MiddleBury	11.94	13.94	14.31	14.01	12.83	13.76	14.63	**17.44**

SSIM	DCP [15]	AOD-Net [20])	DCPDN [42]	GFN [28]	GRID-Net [24]	FFA-Net [26]	FD-GAN [11]	OURS
SOTS	0.8179	0.8788	0.9592	0.8500	0.9836	**0.9886**	0.9207	0.9841
TestA	0.8642	0.8379	0.8398	0.8160	0.9123	0.7715	0.8614	**0.9394**
MiddleBury	0.7620	0.7426	0.7643	0.7545	0.6755	0.6983	0.7812	**0.8465**

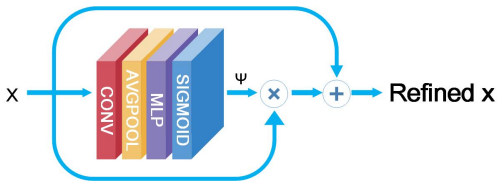


Figure 5. The schematic structure of Residual Channel Attention block.

$W \in \mathbb{R}^{3\times h\times w}$:

$$W = \text{MLF}(X) \tag{6}$$

where MLF is a group convolutional layer.

Finally, we generate fused feature x by element-wise multiplying $\circ$, Hadamard product, x_1, x_2, x_3 with weight coefficient W_1, W_2, W_3 linearly in each c channel:

$$x = W_1 \circ x_1 + W_2 \circ x_2 + W_3 \circ x_3 \tag{7}$$

Then the fused feature x will be fed into RCA-block.

3.3. Residual Channel Attention block

The fused feature x is inevitably mixed up with artifacts generated during the process of CNN, which would degrade the final performance significantly. We note that the existed disturbance would reduce the information of each channel in feature map has and the distribution of artifacts in each channel is uneven. The channel attention mechanism provides insight into this problem. Inspired by [18], we propose a modified residual attention block (RCA-block) to mitigate this. RCA-block generates an attention map to reweigh feature map of each channel, which treats channel unequally thus provides extra flexibility in suppressing background disturbance.

Specially, as illustrated in Fig. 5, RCA-block firstly generates attention map Ψ from the fused features x:

$$\Psi = \text{sigmoid}(\text{MLP}(\text{GAP}(\text{conv}(x)), \Psi \in \mathbb{R}^{c\times 1\times 1} \tag{8}$$

where conv is 1×1 convolutional layers. Then Global Average Pooling (GAP) ($\mathbb{R}^{c\times h\times w} \rightarrow \mathbb{R}^{c\times 1\times 1}$) is applied to convert the global spatial information into a channel descriptor. To obtain the attention map, channel descriptor passes through multi-layer perceptron (MLP), which consists of two fully connected layers, and sigmoid activation function.

Then the input x element-wise multiplies $\circ$ with the attention map Ψ then adds input x to get the refined feature x_r:

$$x_r = \Psi \circ x + x \tag{9}$$

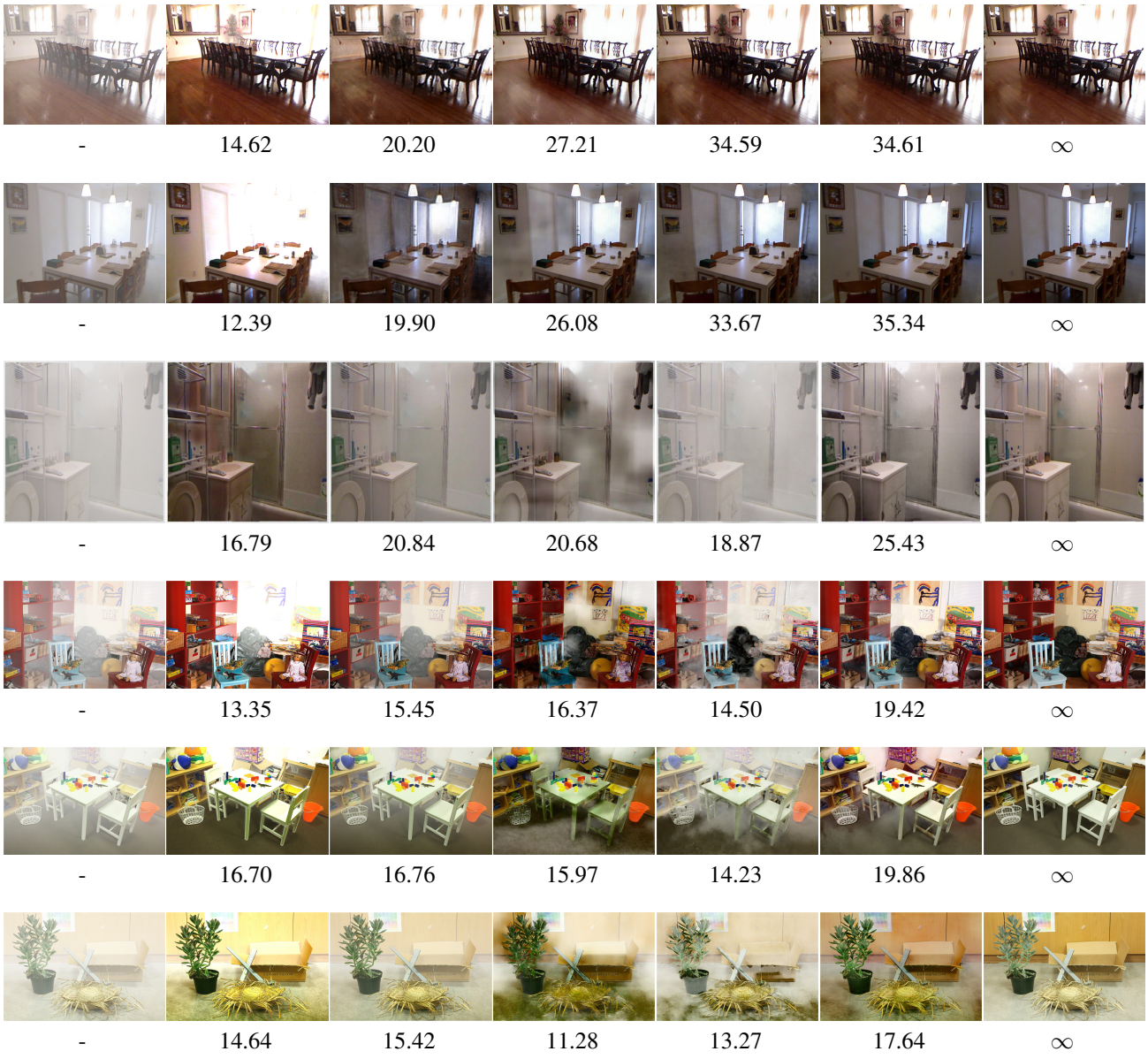
where each weight in attention map Ψ where $\Psi \in \mathbb{R}^{c\times 1\times 1}$ reflects the information in the corresponding channel. A higher value of weight indicates more information channel has. Therefore, multiplying feature with attention map Ψ can suppress background noise effectively.

Considering the back propagation, operation of adding allows network firstly rely on the cue of the local feature then gradually learn parameters of attention map Ψ in a global view.

4. Experiments

4.1. Datasets and Metrics

We use the benchmark synthetic dataset RESIDE [21], TestA [42] and Middlebury [39] for the evaluation of our method. Indoor Training Set (ITS) and Synthetic Objective Testing set (SOTS) from RESIDE are adopted in our experiment. ITS which contains 13,500 synthetic indoor hazy images is applied as the training set. SOTS consists of 500 indoor images and 200 outdoor images with light, medium and high level of haze. TestA is a synthesis dataset introduced by DCPDN. Due to the low resolution and similar scene of images in the SOTS testing set and TestA, we adopt Middlebury synthetic dataset, a high-resolution stereo datasets with subpixel-accurate ground truth, as an assistant testing set. The high resolution of image in Middlebury can provide abundant processing details. Besides, we have done evaluation on real-world images to validate the performance of our proposed network. The dehazing performance



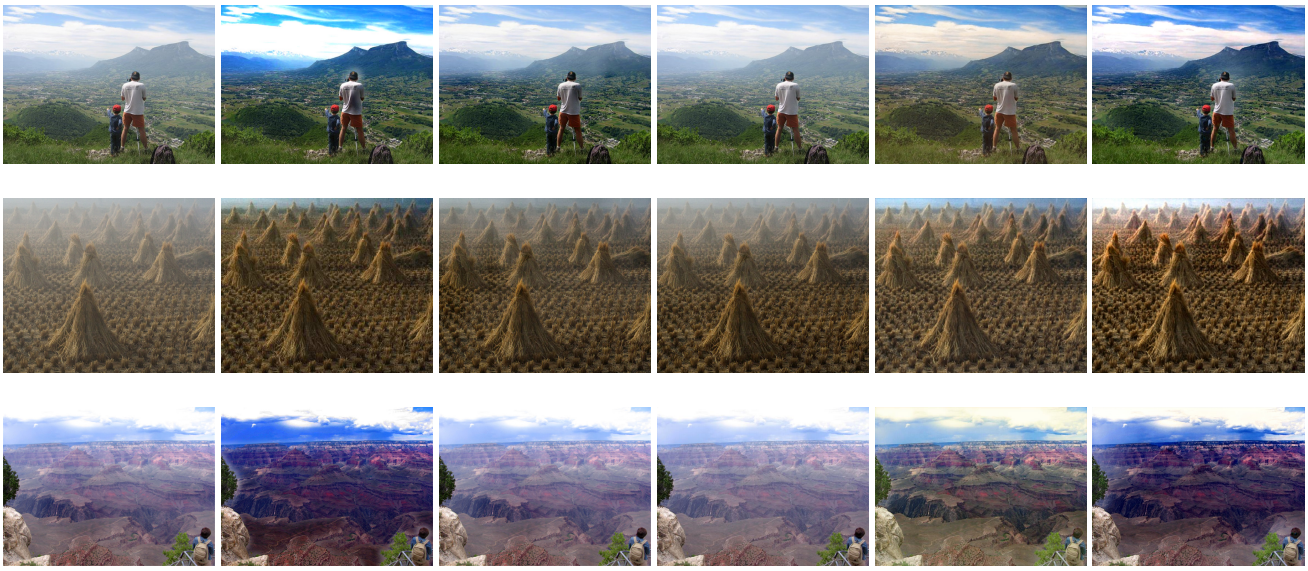
(a) HAZY (b) DCP [15] (c) AOD-Net [20] (d) GRID-Net [24] (e) FFA-Net [26] (f) OURS (g) GT

Figure 6. Qualitative comparisons with different state-of-the-art dehazing methods for indoor synthesis hazy images. The top two rows are from SOTS, the third row is from TestA dataset and the bottom three rows are from MiddleBury dehazing dataset. The numbers below image are PSNR (dB) value of each image.

on synthetic dataset is evaluated with Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) [37]. Due to the lack of groundtruth of the real-world images, real-world images are evaluated by visualization.

4.2. Implementation Details

For training, we directly use RGB images as input instead of using image patches since image patches will lose structure information of original image. The parameter settings for our proposed MI-Net are as follows. For the encoding and decoding blocks shown in Fig. 2, the convolutional layers have 3 filters of size 3×3. While for the other blocks, each convolution layer has 64 filters of size 3×3. Besides, The dilated rate of MLE blocks are $\times 1$, $\times 2$, and $\times 5$ respectively. The zero padding is adopted to fix the size of feature maps. We use Adam optimizer with $\beta_1 = 0.99$, $\beta_2 = 0.999$ to train our MI-Net for $350,000$ iterations. The learning rate is initially set to $1e^{-3}$, and then



(a) INPUT (b) DCP [15] (c) GRID-Net [24] (d) FFA-Net [26] (e) FD-GAN [11] (f) OURS

Figure 7. Qualitative comparisons with different dehazing state-of-the-art methods for real hazy images.

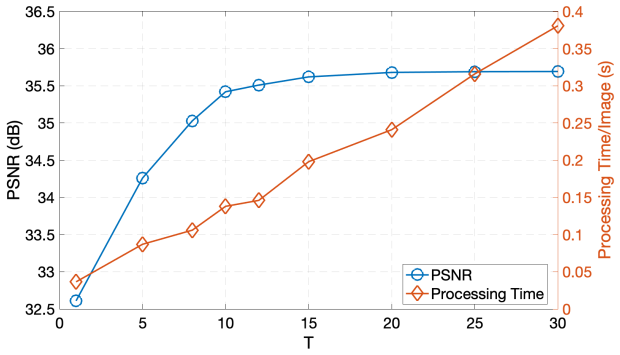


Figure 8. Performance and processing time curve for MI-Net with different recursion number T for IM-blocks. The processing time is tested with NVIDIA GTX 1060 GPU.

(a) Origin (b) IM-Net (OTS) (c) IM-Net (ITS)

Figure 9. The evaluation of our model trained on indoor training set (ITS) and outdoor training dataset (OTS) from RESIDE. For IM-Net (OTS), the haze in foreground is removed but not for the background.

decays by 0.1 every $20,000$ iterations. Two NVIDIA TITAN RTX are used for the training phase and one NVIDIA GTX 1060 is used for testing. The recursion number T of IM-blocks shown in Fig. 3 is determined by compromising between the dehazing performance and the processing time. As shown in Fig. 8, increasing T will boost the performance but increase the processing time, and vice versa. Finally, we choose $T_1 = T_2 = T_3 = 12$ as recursion number for each of three IM-blocks.

4.3. Quantitative and Qualitative Evaluation

We compare the proposed method with previous state-of-the-art methods, including DCP [15], AOD-Net [20],GFN [28], DCPDN [42], GRID-Net [24], FD-GAN [11], and FFA-Net [26]. We leverage the pre-trained models trained on RESIDE to reproduce the image comparison. In order to present our model's generality, all the aforementioned test datasets are evaluated using model trained on RESIDE training dataset. As shown in Tab. 1, our method outperforms previous methods except FFA on the SOTS synthetic dataset. Fig. 6 shows the visual comparisons on the RESIDE, Fig. 7 shows the comparison on real hazy image.

The most of synthesis datasets [21] are based on atmospheric scattering model [15], where depth information plays a crucial role. However, depth of transmission map is hard to measure from outside scene using depth telemeter. Without accurate depth information, outdoor synthesis hazy images are produced by setting depth information as a constant value. Compared to the outdoor synthesis hazing image, the indoor synthesis hazy image contains structure information due to accurate depth information measurement using depth telemeter. As shown in Fig. 9, the IM-Net (ITS) performs much better on background area. Based on this observation, different from [24, 26, 42], the quantitative

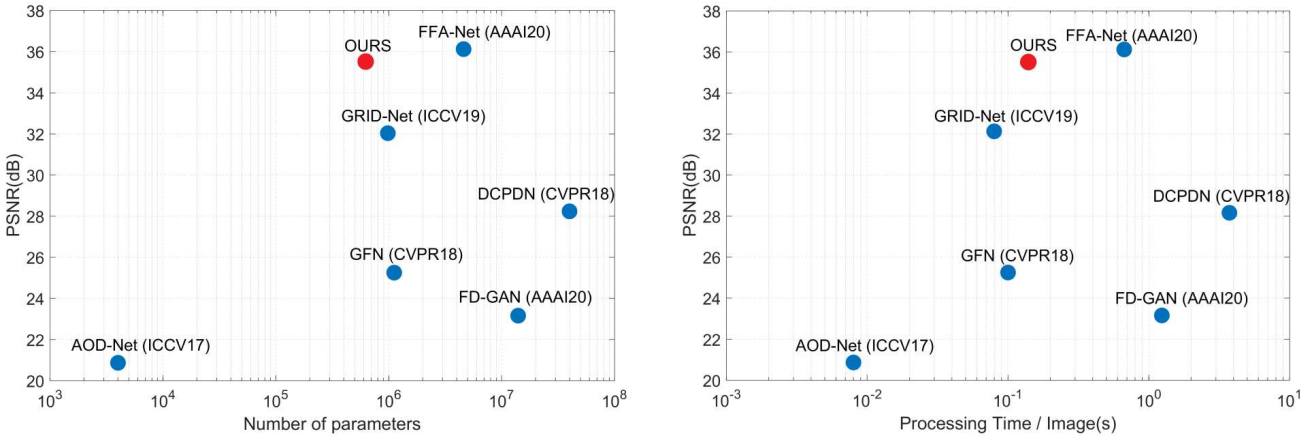


Figure 10. Performance of state-of-the-art methods versus the number of parameters on SOTS. The aforementioned methods are tested with NVIDIA GTX 1060 GPU. The results show that our work gets the best balance between performance and model size.

Table 2. Ablation study on SOTS dataset.

IM-block			✓	✓	✓	✓
Resblock-1	✓					
Resblock-T		✓				
RCA	✓	✓	✓			✓
OA					✓	
MLF	✓	✓		✓	✓	✓
PSNR (dB)	32.61	34.79	34.63	34.92	35.31	35.51

comparisons of outdoor synthesis dataset are not presented in Table. 1.

Furthermore, in order to evaluate the computational efficiency of our proposed method to the aforementioned state-of-arts, we visualize the model size and running time per image versus performances with respect to PSNR on SOTS dataset, as shown in Fig. 10. Obviously, the results show that our proposed method gets the best balance between performance and model size, and thus can process very efficiently.

4.4. Ablation Study

To demonstrate the effectiveness of three mechanism referred in proposed methods section, we conduct ablation experiments to test the performance of model with and without specific component in MI-Net on SOTS dataset.

According to different scheme structure, we compare the implicit scheme structure, i.e. IM-block and the explicit scheme structure, ResNet (1). To develop an explicit structure, we directly replace the IM-block with Resblock, denoting Resblock-1 with -1 denoting only 1 Resblock. Moreover, in order to conduct a fair comparison, we also replicate the Resblock for T times (not shared) to achieve a comparable computational complexity as IM-block, denoting ResNet-T.

From Tab. 2, the results are improved by introducing RCA-block into framework, which verifies the effectiveness of RCA-block. Additionally, we have done an ablation experiment between ordinary attention block (OA-block) [40] and our RCA-block, RCA-block's final PSNR increases 0.2dB compared to ordinary attention block.

Lastly, to demonstrate the function of multi-level fusion (MLF), we design a structure that directly fed x_3 into RCA-block without fusing multi-level features. The result indicates that combining these three mechanism can improve performance by a large margin.

5. Conclusion

We propose an end-to-end multi-level implicit network (MI-Net) for single image dehazing. The crucial idea of this work is to introduce a novel efficient implicit block that substitutes Resblock in image dehazing tasks. Moreover, MLF mechanism and RCA-block that modify the ordinary attention block are adopted to boost performance. Extensive experimental results demonstrate our method's superior performance over state-of-the-art methods.

Acknowledgement

This work is supported by National Natural Science Foundation of China under Grant 61871297. We thank Jingwei He and Binyi Su for their insightful discussion about this work.

References

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

[2] Carla Rezende Barbosa Bonin, Guilherme Cortes Fernandes, Rodrigo Weber dos Santos, and Marcelo Lobosco. Mathematical modeling based on ordinary differential equations: a promising approach to vaccinology. *Human vaccines & immunotherapeutics*, 13(2):484–489, 2017.

[3] 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.

[4] Bo Chang, Lili Meng, Eldad Haber, Lars Ruthotto, David Begert, and Elliot Holtham. Reversible architectures for arbitrarily deep residual neural networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[5] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, pages 1375–1383, 2019.

[6] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in neural information processing systems*, pages 6571–6583, 2018.

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

[8] Dengxin Dai, Christos Sakaridis, Simon Hecker, and Luc Van Gool. Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. *International Journal of Computer Vision*, pages 1–23, 2019.

[9] G David, CAIN SCHAEFFER, and W JOHN. *Ordinary differential equations: basics and beyond*. Springer, 2018.

[10] Zijun Deng, Lei Zhu, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Qing Zhang, Jing Qin, and Pheng-Ann Heng. Deep multi-model fusion for single-image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2453–2462, 2019.

[11] Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing. In *Thirty-Fourth AAAI Conference on Artificial Intelligence*, 2020.

[12] Raanan Fattal. Single image dehazing. *ACM Transactions on Graphics (TOG)*, 27(3):1–9, 2008.

[13] Adrian Galdran, Aitor Alvarez-Gila, Alessandro Bria, Javier Vazquez-Corral, and Marcelo Bertalmío. On the duality between retinex and image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8212–8221, 2018.

[14] Eldad Haber and Lars Ruthotto. Stable architectures for deep neural networks. *Inverse Problems*, 34(1):014004, 2017.

[15] 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, 2010.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[17] Stefan Hoops, Raquel Hontecillas, Vida Abedi, Andrew Leber, Casandra Philipson, Adria Carbo, and Josep Bassaganya-Riera. Ordinary differential equations (ODEs) based modeling. In *Computational Immunology*, pages 63–78. Elsevier, 2016.

[18] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[19] Mustafa İnç, Necdet Bildik, and Hasan Bulut. A comparison of numerical ODE solvers based on euler methods. *Mathematical and Computational Applications*, 3(3):153–159, 1998.

[20] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. AOD-Net: All-in-one dehazing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4770–4778, 2017.

[21] 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, 2018.

[22] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition workshops*, pages 136–144, 2017.

[23] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2117–2125, 2017.

[24] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. GridDehazeNet: Attention-based multi-Scale network for image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7314–7323, 2019.

[25] Soo-Chang Pei and Tzu-Yen Lee. Nighttime haze removal using color transfer pre-processing and dark channel prior. In *Proceedings of the IEEE International Conference on Image Processing*, pages 957–960, 2012.

[26] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. FFA-Net: Feature fusion attention network for single image dehazing. In *Thirty-Fourth AAAI Conference on Artificial Intelligence*, 2020.

[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 *European Conference on Computer Vision*, pages 154–169, 2016.

[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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3253–3261, 2018.

[29] Julius Ruseckas. Differential equations as models of deep neural networks. *arXiv:1909.03767*, 2019.

[30] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126(9):973–992, 2018.

[31] Anders Søgaard and Yoav Goldberg. Deep multi-task learning with low level tasks supervised at lower layers. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 231–235, 2016.

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

[33] Robby T Tan. Visibility in bad weather from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[34] Matthew Thorpe and Yves van Gennip. Deep limits of residual neural networks. *arXiv:1810.11741*, 2018.

[35] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv:1607.08022*, 2016.

[36] Panqu Wang, Pengfei Chen, Ye Yuan, Ding Liu, Zehua Huang, Xiaodi Hou, and Garrison Cottrell. Understanding convolution for semantic segmentation. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1451–1460, 2018.

[37] 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.

[38] E Weinan. A proposal on machine learning via dynamical systems. *Communications in Mathematics and Statistics*, 5(1):1–11, 2017.

[39] Porter Westling and Heiko Hirschmller. High-resolution stereo datasets with subpixel-accurate ground truth. In *36th German Conference on Pattern Recognition*, 2014.

[40] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–19, 2018.

[41] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv:1511.07122*, 2015.

[42] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3194–3203, 2018.