Strong Baseline for Single Image Dehazing with Deep Features and Instance Normalization

Zheng Xu[1]
xuzh@cs.umd.edu

Xitong Yang[1]
xyang35@cs.umd.edu

Xue Li
echolixue@gmail.com

Xiaoshuai Sun[2]
xiaoshuaisun@hit.edu.cn

[1] Department of Computer Science,
University of Maryland,
College Park, MD

[2] Visual Intelligence Lab,
Harbin Institute of Technology,
Harbin, P. R. China

Abstract

We propose a novel deep neural network architecture for the challenging problem of single image dehazing, which aims to recover the clear image from a degraded hazy image. Instead of relying on hand-crafted image priors or explicitly estimating the components of the widely used atmospheric scattering model, our end-to-end system directly generates the clear image from an input hazy image. The proposed network has an encoder-decoder architecture with skip connections and instance normalization. We adopt the convolutional layers of the pre-trained VGG network as encoder to exploit the representation power of deep features, and demonstrate the effectiveness of instance normalization for image dehazing. Our simple yet effective network outperforms the state-of-the-art methods by a large margin on the benchmark datasets.

1 Introduction

Images captured in the wild are often degraded in visibility, colors, and contrasts caused by haze, fog and smoke. Recovering high-quality clear images from degraded images (a.k.a. image dehazing) is beneficial for both low-level image processing and high-level computer vision tasks. Dehazed images are more visually appealing to generate for image processing tasks. Dehazed images can improve the robustness of vision systems that often assume clear images as input. Typical applications that benefit from image dehazing include image super-resolution, visual surveillance, and autonomous driving. Image dehazing is highly desired because of the increasing demand of deploying visual system for real-world applications.

Image dehazing is a challenging problem. The effect of haze is caused by atmospheric absorption and scattering that depend on the distance of the scene points from the camera. In computer vision, the hazy image is often described by a simplified physical model, i.e., the atmospheric scattering model [15, 23, 26, 28],

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

where $I(x)$ is the observed hazy image, $J(x)$ is the scene radiance (clear image), $t(x)$ is the medium transmission map, and A is the global atmospheric light. When the atmosphere is homogeneous, $t(x)$ can be further expressed as a function of the scene depth $d(x)$ and the scattering coefficient β of the atmosphere as $t(x) = \exp(-\beta d(x))$. The goal of image dehazing is to recover clear image $J(x)$ from hazy image $I(x)$. Single image dehazing is particularly challenging. It is under-constrained because haze is dependent on many factors, including the unknown depth information that is difficult to recover from a single image.

The atmospheric scattering model (1) has been extensively used in previous methods for single image dehazing [4, 6, 10, 11, 15, 27, 38, 40]. These works either separately or jointly estimate the transmission map $t(x)$ and the atmospheric light A to generate the clear image from a hazy image. Due to the under-constrained nature of single image dehazing, the success of previous methods often relies on hand-crafted priors such as dark channel prior [15], contrast color-lines [11], color attenuation prior [50], and non-local pior [4]. However, it is difficult for these priors to be always satisfied in practice. For example, dark channel prior is known to be unreliable for areas that are similar to the atmospheric light.

More recent works learn convolutional neural networks (CNNs) to estimate components in the atmospheric scattering model for image dehazing [5, 22, 24, 29, 44, 45]. These methods are often trained with limited (synthetic) images, and use only a few layers of convolutional filters. The learned shallow networks have limited capacity to represent or process images, making them difficult to surpass the prior-based methods. In contrast, training deep neural networks with large-scale data has made significant progress and achieved state-of-the-art performance in many vision tasks [16, 21, 37]. Moreover, the deep features extracted by a pre-trained deep network are used as powerful image representation in many applications, such as domain invariant recognition [8], perceptual evaluation [49], and characterizing image statistics [12]. More recently, the architecture of CNNs itself has been recognized as a prior for image processing [42]. In this paper, we study how to release the power of *deep* network for single image dehazing.

We propose an encoder-decoder architecture as an end-to-end system for single image dehazing. We exploit the representation power of deep features by adopting the convolutional layers of the deep VGG net [37] as our encoder, and pre-train the encoder on large-scale image classification task [33]. We add skip connections with instance normalization between the encoder and decoder, and then train decoder with both ℓ_2 reconstruction loss and VGG perceptual loss [49]. We show that the recently proposed instance normalization [41], which is designed for image style transfer, is also effective in image dehazing. The proposed method effectively learns the statistics of clear images based on the deep feature representation, which benefits the dehazing process on the input image. Our approach outperforms the state-of-the-art results by a large margin on a recently released benchmark dataset [23], and performs surprisingly well in several cross-domain experiments. Our method depends on neither the explicit atmospheric scattering model nor the hand-crafted image priors, and only exploits the deep network architecture and pre-trained models to tackle the under-constrained single image dehazing problem. Our simple yet effective network can serve as a strong baseline for future study in this topic.

2 Related work

Traditional methods focus on representing human knowledge as priors for image processing. Tan [38] assumes higher contrast of clear images and proposes a patch-based contrast-

maximization method. Fattal [10] assumes the transmission and surface shading are locally uncorrelated, and estimates the albedo of the scene. Dark channel prior (DCP) [15] assumes local patches contain low intensity pixels in at lease one color channel and hence estimates the transmission map. Fast visibility restoration (FVR) [40] is a filtering approach by atmospheric veil inference and corner preserving smoothing. Meng et al. [27] uses boundary constraint and contextual regularization (BCCR), and Chen et al. [6] uses gradient residual minimization (GRM) to surpress artifacts. Tang et al. [39] combines priors by learning with random forests model. Color attenuation prior (CAP) [50] assumes a linear model of brightness and the saturation and then learns the coefficients. Berman et al. [4] assumes each color cluster in the clear image becomes a line in RGB space, and proposes non-local image dehazing (NLD).

There is an increasing interest in applying convolutional neural networks (CNNs) for image dehazing. DehazeNet [5] and multi-scale convolutional neural networks (MSCNN) [29] are trained to estimate the transmission map. AOD-Net[22] estimates a new variable based on the transformation of the atmospheric scattering model. Zhang and Patel [45], Zhang et al. [48] and Li et al. [24] estimate transmission map and atmospheric light by separate CNNs. Yang et al. [44] adversarially train generators for components of the atmospheric scattering model. Ren et al. [30] train network to fuse three derived inputs from an original hazy image. These methods use relatively small CNNs and do not exploit the pre-trained deep networks for image representation. A few days before our submission, we notice a preprint [7] that also uses the pre-trained deep networks. The proposed method is quite different from [7]: we use encoder-decoder with skip connections, while Cheng et al. [7] only use feature maps extracted from one layer of the pre-trained network as input; we study instance normalization and demonstrate its effectiveness; we train an end-to-end system from hazy image to clear image, while Cheng et al. [7] estimate transmission map and atmospheric light; we can generate impressive results without explicitly applying the atmospheric scattering model.

Deep neural networks can be used as "priors" for image generation and image processing. The architecture of CNNs itself can be a constraint for image processing [42] and image generation [14, 20]. A pre-trained deep networks can be used as general purpose feature extractors [8] and perceptual metric [49]. The second-order information of the features extracted by a pre-trained network describes the style of images [12]. Instance normalization layers that effectively change the statistics of deep features are widely used for image style transfer [9, 13, 17, 41, 43]. Image translation tasks with adversarial networks are often trained with batch normalization and batch size one [19, 46, 47], which may suffer from the statistics mismatch between training and testing.

3 VGG-based U-Net with instance normalization

We propose an end-to-end encoder-decoder network architecture for single image dehazing, as shown in fig. 1. The input is a hazy image, and the output is the desired clear image. We introduce different components of the network in the following paragraphs of this section.

Encoder. Our encoder uses the convolutional layers of the VGG net [37] pre-trained on Imagenet large-scale image classification task [33]. VGG net contains five blocks of convolutional layers, and we use the first three blocks and the first convolutional layer of the forth block. Each block contains several convolutional layers, and each convolutional layer is equipped with ReLU [21] as activation function. The width (number of channels) and size (height and width) of convolutional layers are shown in fig. 1. There is a maxpooling layer
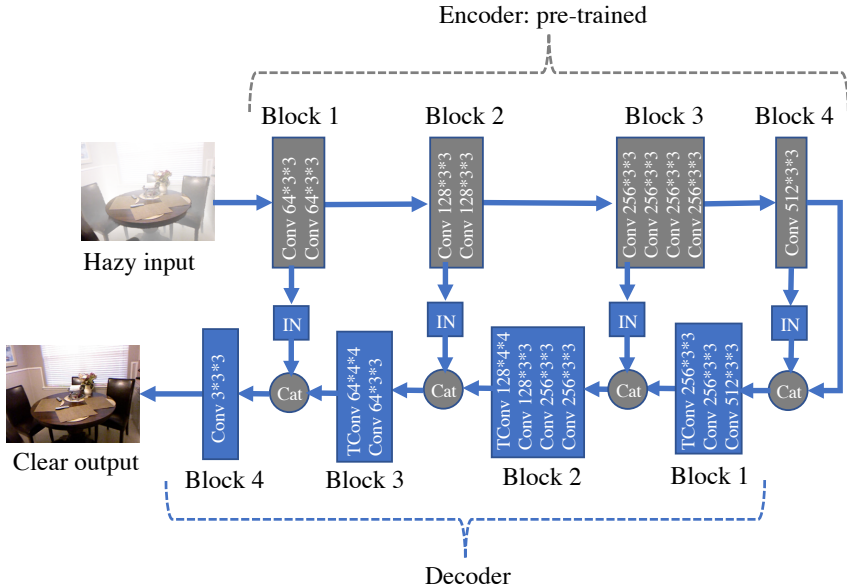
Figure 1: The proposed network: encoder-decoder with skip connections and instance normalization (IN); convolutional layers of pre-trained VGG [37] are used as encoder; ℓ_2 reconstruction loss and VGG perceptual loss are used for training decoder and IN layers.

of stride two between blocks, which enlarges the receptive field of higher layers. The width of convolutional layer is doubled after the subsampling of feature maps by maxpooling.

The pre-trained VGG net is a powerful feature extractor for perceptual metric [49] and image statistics [12]. Our encoder is deep and wide, and the extracted deep features are capable to capture the semantic information of the input image. We fix the encoder during training to exploit the power of pre-trained VGG net as "priors", and avoid overfitting from relatively small number of samples in image dehazing dataset.

Decoder and skip connection. Our decoder is designed to be roughly symmetric to the encoder. The decoder also contains four blocks, and each block contains several convolutional layers. The last layer of the first three blocks of the decoder uses transposed convolutional layer to upsample the feature maps. We use ReLU activation for convolutional and transposed convolutional layers except for the last layer, where we use Tanh as activation function.

We add skip connections from the output of the first convolutional layer of encoder block 1,2,3 to the input of decoder block 4,3,2 by concatenating (cat) the feature maps, respectively. Hence our deep encoder-decoder network has a U-Net [19, 51] structure except that our skip connections are based on blocks instead of layers . We use trainable instance normalization for skip connections, and have instance normalization before each convolutional layer in decoder except the first one. Our deep encoder-decoder network has large capacity, and skip connections make the information smoothly flow to easily train a large network [16, 25].

Instance normalization. We briefly review instance normalization [41], and discuss our motivation in applying instance normalization for single image dehazing. Let $x \in \mathbb{R}^{N \times C \times H \times W}$ represent the feature map of a convolutional layer from a minibatch of sam-

ples, where N is the batch size, C is the width of the layer (number of channels), H and W are height and width of the feature map. x_{nchw} denotes the element at height h, width w of the cth channel from the nth sample, and instance normalization layer can be written as,

$$IN(x_{nchw}) = \gamma_{nc} \left(\frac{x_{nchw} - \mu_{nc}}{\sigma_{nc}} \right) + \beta_{nc}, \text{ where}$$

$$\mu_{nc} = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} x_{nchw}, \ \sigma_{nc} = \sqrt{\frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} (x_{nchw} - \mu_{nc})^2 + \varepsilon}, \quad (2)$$

γ_{nc}, β_{nc} are learnable affine parameters, ε is a very small constant, and μ_{nc}, σ_{nc}^2 represent the mean and variance for each feature map per channel per sample.

If we replace instance level variables $\gamma_{nc}, \beta_{nc}, \mu_{nc}, \sigma_{nc}^2$ with batch level variables $\gamma_c, \beta_c, \mu_c, \sigma_c^2$ that are estimated for all samples of a minibatch, we get the well-known batch normalization layer [18]. We show instance normalization is preferred than batch normalization for single image dehazing in our experimental ablation study.

The learnable affine parameters γ_{nc}, β_{nc} of instance normalization shift the first and second order statistics (mean and variance) of the feature maps. Instance normalization is effective for image style transfer, and the style of images can be represented by learned affine parameters [9]. Shifting the statistics of deep features extracted by pre-trained networks has achieved impressive results for arbitrary style transfer [17]. Shifting the statistics of images is intuitive for dehazing, however, it can be difficult to decide the exact amount to change because haze depends on the unknown depth. The deep features extracted by a pre-trained VGG net contain semantic information to effectively infer depth for haze, and hence the learned affine parameters effectively shift the statistics of images. We apply instance normalization on the deep features extracted by pre-trained VGG net for single image dehazing.

Training loss. Our network is trained with both reconstruction loss and VGG perceptual loss. Denoting the training pairs of hazy image and clear image as $(I_n, T_n), n = 1, \ldots, N$, we use the mean squared loss,

$$\min_F \frac{1}{N} \sum_{n=1}^{N} \|F(I_n) - T_n\|^2 + \lambda \|g(F(I_n)) - g(T_n)\|^2, \quad (3)$$

where F represents the trainable instance normalization and decoder layers in our network, g represents the perceptual function, and λ is a hyperparameter. We set $\lambda = 1$, and use the features extracted by the first convlutional layer of the third block from the pre-trained VGG net as perceptual function.

4 Experiments

In this section, we conduct various experiments on both synthetic and natural images to demonstrate the effectiveness of the proposed method. The atmospheric scattering model is widely used to synthesize images for both training and testing. The hazy images are synthesized from groundtruth clear images and grountruth depth images [3, 23], or estimated depth images [34].

We train our model on the recently released RESIDE-standard dataset [23]. RESIDE-standard contains 13,990 images for training, and 500 images for testing. These images are

Table 1: Quantitative results on RESIDE-standard dataset [23].

	DCP [15]	FVR [40]	BCCR [27]	GRM [6]	CAP [50]
PSNR	16.62	15.72	16.88	18.86	19.05
SSIM	0.8179	0.7483	0.7913	0.8553	0.8364
	NLD [4]	DehazeNet [5]	MSCNN [29]	AOD-Net [22]	**Ours**
PSNR	17.29	21.14	17.57	19.06	**27.79**
SSIM	0.7489	0.8472	0.8102	0.8504	**0.9556**

generated by existing indoor depth datasets, NYU2 [36] and Middlebury stereo [35]. The atmospheric scattering model is used, where atmospheric lights A is randomly chosen between (0.7, 1.0) for each channel, and scattering coefficient β is randomly selected between (0.6, 1.8).

We also apply our model trained on RESIDE-standard for cross-domain evaluation on D-Hazy [3], I-Haze [1] and O-Haze [2] dataset. D-Hazy dataset [3] is another synthetic dataset, which contains 23 images synthesized from Middlebury and 1449 images synthesized from NYU2, with atmospheric lights $A = (1,1,1)$ and scattering coefficient $\beta = 1$. Though D-Hazy dataset use the same clean images as RESIDE-standard, the generated hazy images are quite different. I-Haze [1] and O-Haze [2] are two recent released datasets on natural indoor and outdoor images, respectively. I-Haze contains 35 pairs of indoor images and O-Haze contains 45 pairs of outdoor images, where the hazy images are generated by using a physical haze machine.

We compare our results quantitatively and qualitatively with previous methods. We compare with prior-based methods, DCP [15], FVR [40], BCCR [27] , GRM [6], CAP [50] and NLD [4] . We also compare with learning-based methods DehazeNet [5], MSCNN [29] , and AOD-Net [22]. We have provided a brief review of these baseline methods in section 2. We use peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as metrics for quantitative evaluation. For the benchmark evaluation on RESIDE-side, all the learning-based methods are trained on the same dataset. For cross-domain evaluation on D-Hazy, O-Haze and I-Haze, we use the released best pre-trained model for the learning-based baseline methods.

We train our model by SGD with minibatch size 16 and learning rate 0.1 for 60 epochs, and linearly decrease the learning rate after 30 epochs. We use momentum 0.9 and weight decay 10^{-4} for all our experiments. We will release our Pytorch code and pre-trained models [1].

4.1   Quantitative evaluation on benchmark dataset

We present the performance of our network and baseline methods on the RESIDE-standard benchmark dataset [23] in table 1. Our network and the learning-based baselines [5, 22, 29] are trained on the provided synthetic data, and evaluated on the separate testing set. We evaluate our results by metrics provided by [23], and compare with the baseline results reported in [23]. The learning-based methods perform slightly better than the prior-based method. CAP [50] performs best in prior-based method, which has a learning phase for the coefficients of the linear model. DehazeNet [50] performs best in baseline methods, which uses a relatively small network to predict components.

---

[1]https://github.com/nightldj/dehaze_release

| Hazy | Clear | NA-NA | IN-IN | IN-IN-Percep |

Figure 2: An example of qualitative results in ablation study. We zoom in the bottom left corner of the images to show more details in the second row.

Table 2: Ablation study on RESIDE-standard dataset.

Skip	NA	BN	IN	NA	BN
Dec	NA	NA	NA	BN	BN
PSNR	18.24	25.67	26.00	25.99	26.38
SSIM	0.7945	0.9442	0.9414	0.9385	0.9519
Skip	IN	NA	BN	IN	Perceptual
Dec	BN	IN	IN	IN	loss
PSNR	26.89	26.57	27.67	27.75	**27.79**
SSIM	0.9535	0.9381	0.9543	0.9549	**0.9556**

Our approach outperforms all the baseline methods on both PSNR and SSIM by a large margin. The synthetic data for both training and testing are generated by the atmospheric scattering model, and the baseline methods explicitly use the atmospheric scattering model. In contrast, our approach only uses instance normalization to transform the statistics of deep features . The superior performance of our network on the benchmark dataset demonstrate the effectiveness of *deep* networks and instance normalization for single image dehazing.

4.2 Ablation study

We provide more discussion on the proposed network. We verify the effectiveness of instance normalization with ablation study on network structures, as shown in table 2. We use no normalization (NA), batch normalization (BN), or instance normalization (IN) for skip connections and decoders, respectively. The normalization layers are added before each convolutional layer of the decoder except for the first layer. All the results in table 2 are obtained by only using reconstruction loss ($\lambda = 0$ in loss function (3)) except for the last one, where IN and combined loss ($\lambda = 1$) are used. We train and evaluate our network on the RESIDE-standard dataset.

First, comparing the NA results in table 2 with previous best results in table 1, our encoder-decoder only achieves competitive results. Second, adding normalization to either skip connections or decoder significantly improves the performance of our network. The normalization layers for decoder are implicitly applied to the features from the skip connections, which makes the result of only normalizing decoder slightly better than only normalizing skip connections. Third, instance normalization works better than batch normalization,

Table 3: Quantitative results for cross-domain evaluation.

	D-Hazy-NYU [3]		D-Hazy-MB [3]		I-Haze [1]		O-Haze [2]	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCP [15]	11.56	0.6695	12.13	0.6752	13.41	0.4930	17.01	0.4875
CAP [50]	13.29	0.7266	_14.36_	**0.7526**	15.27	0.5603	16.68	0.4810
DehazeNet [5]	13.02	0.7256	13.78	0.7342	**16.73**	_0.6263_	**17.90**	**0.5514**
MSCNN [29]	_13.67_	_0.7413_	13.97	_0.7488_	15.93	0.5896	16.27	0.4947
AOD-Net [22]	12.44	0.7147	13.48	0.7470	15.00	0.5828	16.22	0.4142
Ours	**18.11**	**0.8268**	**15.63**	0.7338	_16.04_	**0.6332**	_17.46_	_0.5337_

which demonstrates the effectiveness of shifting the mean and variance of deep features at instance level.

Finally, the perceptual loss only helps a little for quantitative evaluation, but it can help generate more visually appealing output images. We show an qualitative example in fig. 2, where the hazy input, the groundtruth clear image, outputs of our network without normalization layers and no perceptual loss (NA-NA), our network with instance normalization and no perceptual loss (IN-IN), and our network with instance normalization and perceptual loss (IN-IN-Percep). We enlarge the bottom left corner of the results to show more details. The results of IN-IN look much better than NA-NA. The enlarged area of the result with perceptual loss (IN-IN-Percep) looks sharper and clearer.

4.3 Cross-domain evaluation

In this section, we focus on the cross-domain performance by evaluating our network trained on RESIDE-standard [23] on the cross domain datasets, D-Hazy [3], I-Haze [1] and O-Haze [2]. We compare with baseline methods that have publicly available code, and these are strong baselines according to benchmark evaluation in table 1. For learning-based methods DehazeNet [5], MSCNN [29], and AOD-Net [22], we use the best model the authors have released. MSCNN [29] and AOD-Net [22] are trained with synthetic images similar to RESIDE-standard, while DehazeNet [5] is trained with patches of web images.

We present the quantitative results in table 3, where we use bold to label the best results and underline to label the second best results. Our approach achieves best results, or close to the best results for all the cross-domain evaluations. Our first observation is that the learning-based methods [5, 22, 29], including ours, generalize reasonably well and perform equally or better than the prior-based methods [15, 50].

Our network performs well on the cross-domain D-Hazy dataset [3]. Particularly, our approach outperforms all baseline methods by a large margin on the images synthesized from NYU depth dataset. D-Hazy dataset is synthesized by the same clear images as our training data RESIDE-standard, but uses different parameters of the atmospheric scattering model. Our trained network has effectively captured the statistics of the deep features of the desired clear images.

I-Haze [1] and O-Haze [2] images look quite different from our training images, and our network may have difficulty to infer the exact statistics of deep features for these images. DehazeNet [5] may have gained some advantage on these two datasets because it is trained on patches of web images. Our approach still produces competitive results compared with DehazeNet [5], and outperforms all the other baselines. Notice again that our network does not use the powerful atmospheric scattering model, and is only trained on a limited number

of indoor synthetic images. The cross-domain evaluation further demonstrates the power of *deep* features and instance normalization in our approach.
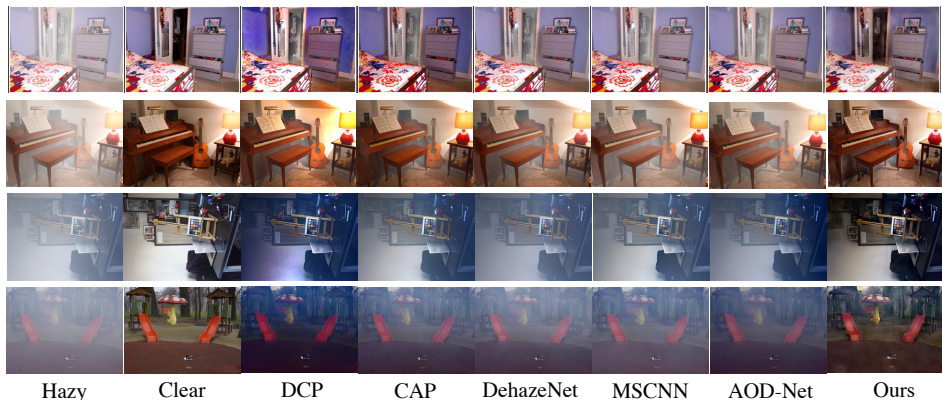
4.4 Qualitative evaluation

| Hazy | Clear | DCP | CAP | DehazeNet | MSCNN | AOD-Net | Ours |

Figure 3: Qualitative evaluation on cross-domain dataset. The four examples are from D-Hazy-NYU [4], D-Hazy-MB [4], I-Haze [1] and O-Haze [2], respectively. Best viewed in color and zoomed in.

We present qualitative results from cross-domain evaluation in fig. 3. The images are from D-Hazy-NYU [4], D-Hazy-MB [4], I-Haze [1] and O-Haze [2], respectively. We show the hazy image and groundtruth clear image, and compare our results with DCP [15], CAP [50], DehazeNet [5], MSCNN [29], and AOD-Net [22]. We use the best released model for the learning-based baselines [5, 22, 29], and train our network on RESIDE-standard [23].

Our network makes the best efforts to remove haze and recover the real color of images, as shown in fig. 3. The results of baselines still have a large amount of undesired haze and look blurry (row 2,3,4). Particularly, the baselines have difficulty in dark areas of the image, and DCP also has difficulty in area of white and blue walls (row 1,3). For the outdoor image (row 4), our network produces a little artifact due to the significant domain difference between the desired images and the training indoor images. Use regularizers such as total variation [32] may help reduce these artifacts, and we plan to investigate it in the future. Our simple yet effective network has generated visually appealing results, without depending on extra constraints like the atmospheric scattering model.

5 Discussion

We proposed a simple yet effective end-to-end system for single image dehazing. Our network has an encoder-decoder architecutre with skip connections. We manipulated the statistics of deep features extracted by pre-trained VGG net and demonstrated the effectiveness of instance normalization for image dehazing. Moreover, without explicitly using the atmospheric scattering model, our approach outperforms previous methods by a large margin on the benchmark datasets. Notice that both the training and testing data are generated by the

atmospheric scattering model, and the baseline methods all explicitly use the model. Our network effectively learns the transformation from hazy image to clear image with limited synthetic data, and generalizes reasonably well.

The atmospheric scattering model is powerful and has been successfully deployed for image dehazing in the past decade. However, the atmospheric scattering model, as a simplified model, also constrained the learnable components to be "linearly" combined by element-wise multiplication and summation, which may not be ideal for training deep models. Our study sheds light on the power of *deep* neural networks and the *deep* features extracted by pre-trained network for single image dehazing, and encourages the rethinking on how to effectively exploit the physical model for haze. How will physical model help when training powerful deep networks? It is still an open question, and our approach serves as a strong baseline for future study.

Our network outperforms state-of-the-art methods by a large margin on the benchmark dataset, and achieves competitive results on cross-domain evaluation. The key idea of our approach is to apply instance normalization to shift the statistics of deep features for image dehazing. For cross-domain evaluation, it may be difficult to effectively infer the desired statistics of deep features of clear images that is quite different from the training data. Our generalization ability can be significantly improved by training from large-scale natural images. In the future, we will explore adversarial training to use unpaired hazy and clear images that are easier to collect from the web.

Acknowledgment

References

[1] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: a dehazing benchmark with real hazy and haze-free indoor images. *arXiv preprint arXiv:1804.05091*, 2018.

[2] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. *arXiv preprint arXiv:1804.05101*, 2018.

[3] Cosmin Ancuti, Codruta O Ancuti, and Christophe De Vleeschouwer. D-hazy: A dataset to evaluate quantitatively dehazing algorithms. In *ICIP*, pages 2226–2230. IEEE, 2016.

[4] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *CVPR*, 2016.

[5] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE TIP*, 25(11):5187–5198, 2016.

[6] Chen Chen, Minh N Do, and Jue Wang. Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In *ECCV*, pages 576–591. Springer, 2016.

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

[8] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *ICML*, pages 647–655, 2014.

[9] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. A learned representation for artistic style. *ICLR*, 2017.

[10] Raanan Fattal. Single image dehazing. *ACM TOG*, 27(3):72, 2008.

[11] Raanan Fattal. Dehazing using color-lines. *ACM TOG*, 34(1):13, 2014.

[12] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *CVPR*, pages 2414–2423. IEEE, 2016.

[13] Golnaz Ghiasi, Honglak Lee, Manjunath Kudlur, Vincent Dumoulin, and Jonathon Shlens. Exploring the structure of a real-time, arbitrary neural artistic stylization network. *BMVC*, 2017.

[14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*, pages 2672–2680, 2014.

[15] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE TPAMI*, 33(12):2341–2353, 2011.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[17] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *CVPR*, pages 1501–1510, 2017.

[18] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, pages 448–456, 2015.

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

[20] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.

[22] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *ICCV*, pages 4770–4778, 2017.

[23] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Reside: A benchmark for single image dehazing. *arXiv preprint arXiv:1712.04143*, 2017.

[24] Chongyi Li, Jichang Guo, Fatih Porikli, Huazhu Fu, and Yanwei Pang. A cascaded convolutional neural network for single image dehazing. *arXiv preprint arXiv:1803.07955*, 2018.

[25] Hao Li, Zheng Xu, Gavin Taylor, and Tom Goldstein. Visualizing the loss landscape of neural nets. *arXiv preprint arXiv:1712.09913*, 2017.

[26] Earl J McCartney. Optics of the atmosphere: scattering by molecules and particles. *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976.

[27] 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. IEEE, 2013.

[28] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3):233–254, 2002.

[29] 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. Springer, 2016.

[30] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. *CVPR*, 2018.

[31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015.

[32] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[33] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *IJCV*, 2015.

[34] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *arXiv preprint arXiv:1708.07819*, 2017.

[35] Daniel Scharstein, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nešić, Xi Wang, and Porter Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition*, pages 31–42. Springer, 2014.

[36] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. *ECCV*, pages 746–760, 2012.

[37] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint*, 2014.

[38] Robby T Tan. Visibility in bad weather from a single image. In *CVPR*, 2008.

[39] Ketan Tang, Jianchao Yang, and Jue Wang. Investigating haze-relevant features in a learning framework for image dehazing. In *CVPR*, pages 2995–3000, 2014.

[40] Jean-Philippe Tarel and Nicolas Hautiere. Fast visibility restoration from a single color or gray level image. In *ICCV*, pages 2201–2208. IEEE, 2009.

[41] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Instance normalization: The missing ingredient for fast stylization. *CoRR*, abs/1607.08022, 2016.

[42] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Deep image prior. *CoRR*, abs/1711.10925, 2017.

[43] Zheng Xu, Michael Wilber, Chen Fang, Aaron Hertzmann, and Hailin Jin. Beyond textures: Learning from multi-domain artistic images for arbitrary style transfer. *arXiv preprint arXiv:1805.09987*, 2018.

[44] Xitong Yang, Zheng Xu, and Jiebo Luo. Towards perceptual image dehazing by physics-based disentanglement and adversarial training. *AAAI*, 2018.

[45] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. *CVPR*, 2018.

[46] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. *arXiv preprint arXiv:1802.07412*, 2018.

[47] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *arXiv preprint arXiv:1701.05957*, 2017.

[48] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Joint transmission map estimation and dehazing using deep networks. *arXiv preprint arXiv:1708.00581*, 2017.

[49] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. *CVPR*, 2018.

[50] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE TIP*, 24(11):3522–3533, 2015.