HazDesNet: An End-to-End Network for Haze Density Prediction

Jiahe Zhang, Xiongkuo Min, *Member, IEEE*, Yucheng Zhu, Guangtao Zhai, *Senior Member, IEEE*, Jiantao Zhou, *Senior Member, IEEE*, Xiaokang Yang, *Fellow, IEEE*, and Wenjun Zhang, *Fellow, IEEE*

***Abstract*— Vision-based intelligent systems such as driver assistance systems and transportation systems should take into account weather conditions. The presence of haze in images can be a critical threat to driving scenarios. Haze density measures the visibility and usability of hazy images captured in real-world conditions. The prediction of haze density can be valuable in various vision-based intelligent systems, especially in those systems deployed in outdoor environments. Haze density prediction is a challenging task since the haze and many scene contents have a lot in common in appearance. Existing methods generally utilize different priors and design complex handcrafted features to predict the visibility or haze density of the image. In this article, we propose a novel end-to-end convolutional neural network (CNN) based method to predict haze density, named as HazDesNet. Our HazDesNet takes a hazy image as input and predicts a pixel-level haze density map. The density map is then refined and smoothed, and the average of the refined map is calculated as the global haze density of the image. To verify the performance of HazDesNet, a subjective human study is performed to build a Human Perceptual Haze Density (HPHD) database, which includes 500 real-world hazy images and 100 synthetic hazy images, and the corresponding human-rated perceptual haze density scores. Experimental results show that our method achieves the best haze density prediction performance on our built HPHD database and existing databases. Besides the global quantitative results, our HazDesNet is capable of predicting a continuous, stable, fine, and high-resolution haze density map. We will make the database and code publicly available at https://github.com/JiaheZhang/HazDesNet.**

***Index Terms*— Haze detection, haze density, haze, visibility, deep learning.**

I. Introduction

VISION based intelligent driver assistance systems and transportation systems may malfunction when they encounter adverse weather conditions such as haze, snow, hail, and rain [1]. This issue is mainly caused by the degraded visibility of captured images in bad weather conditions [2]. Since poor performance may be introduced by various adverse weather conditions, many specific methods have been proposed for severe weather conditions, such as nighttime visibility estimation [3], adverse weather forecasting system [4] and traffic surveillance [5]. Besides, visibility enhancement methods related to intelligent transportation systems have also been widely researched [6], [7].

Among these systems and applications, single image dehazing [2], [8]–[10], haze density prediction [11]–[13], and quality evaluation of dehazing algorithms [14], [15] have been widely studied by researchers due to the common appearance of hazy weather in driving images. For instance, fog detection methods based on on-board cameras for driving assistance systems have been proposed in [16], [17]. Haze density quantifies visibility of images captured in hazy conditions, and provides an important clue to understanding the machine perception of the environment. For example, the measurement of haze density can be considered as a warning signal for autonomous vehicles, so that the strategies of autonomous systems can be adjusted according to haze density prediction. Highly relevant to the visibility, the usability of images can be also measured according to the haze density. Besides, visual tasks such as image acquisition can adjust their parameters to the best according to the haze density. (We do not differentiate the haze and fog in this article, since the visibility degradation caused by haze and fog is similar).

It is a challenging task to predict haze density, since the haze density is highly related to the uncertain image depth. Besides, many image scene contents which have similar appearance with haze can be easily deemed as haze. With a corresponding haze-free image as a reference, haze density prediction becomes easy and precise. However, it is almost impossible to obtain the corresponding haze-free image of exactly the same scene in practice. Similar to no-reference (NR) image quality assessment (IQA) [18]–[20], we can possibly measure haze density from a single hazy image. However, haze perception is different from the perception of traditional digital image distortions.

Although, the haze density prediction is a difficult visual task for machines, humans can perceive haze at a short glance without much prior knowledge. With the success of deep learning-based models for image dehazing [21]–[24], it is reasonable to propose a deep learning based model to predict haze density. In these image dehazing methods, a critical

Manuscript received November 2, 2019; revised April 11, 2020 and August 16, 2020; accepted October 6, 2020. Date of publication October 22, 2020; date of current version March 29, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61901260, Grant 61831015, Grant 61521062, and Grant 61527804. The Associate Editor for this article was Z. Duric. *(Jiahe Zhang and Xiongkuo Min contributed equally to this work.) (Corresponding authors: Xiongkuo Min; Guangtao Zhai.)*

Jiahe Zhang, Xiongkuo Min, Yucheng Zhu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang are with the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhangjiahe@sjtu.edu.cn; minxiongkuo@sjtu.edu.cn; zyc420@sjtu.edu.cn; zhaiguangtao@sjtu.edu.cn; xkyang@sjtu.edu.cn; zhangwenjun@sjtu.edu.cn).

Jiantao Zhou is with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Taipa, Macau (e-mail: jtzhou@umac.mo).

Digital Object Identifier 10.1109/TITS.2020.3030673

step is to reconstruct the transmission and predict the global atmospheric light from a hazy image. The transmission map is determined by the scene depth, and it may indicate the haze density to some extent. However, it is not precise enough to use the transmission map to describe the haze density map. A detailed discussion about the differences between the transmission map and haze density map is given in Section VI-B. Thus, despite the good performance of these models in image dehazing, the outputs of these methods might not be directly used for the haze density prediction. The image dehazing and haze density prediction are two different tasks. To the best of our knowledge, the deep learning based haze density prediction method is still absent in the literature.

Data shortage is a critical problem that hinders the development and the application of data-driven based methods to some degree. It is the same for deep learning based haze density prediction. It is hard to build a large scale haze density database with human labeled haze density maps. Although some methods overcome the data shortage problem by augmentation [25], [26] and few-shot learning [27], and some other methods are proposed using a small number of training samples [28], [29], these tasks are different from the haze density prediction task.

In this article, we propose HazDesNet, a novel CNN based method to predict haze density, which is the first of its kind in the literature. This model predicts haze density of a hazy image in an end-to-end way. The system overview is shown in Fig. 1, which includes a training procedure and an inference procedure. To overcome the issue of data shortage, we propose to use synthetic hazy images for training. Specifically, hazy images are synthesized from haze-free images using the widely accepted haze model [30]. In Section IV-C, we apply full-reference (FR) image quality assessment (IQA) metrics to measure the visibility degradation from the haze-free image to the synthesized hazy image. We find that the structural similarity (SSIM) metric can describe the haze density well. Thus, we propose to use the scores computed by FR IQA metric as the training labels of the haze densities of the synthetic hazy images. HazDesNet is trained by these synthetic hazy patches and the corresponding haze density scores labeled by SSIM.

In the inference procedure, the HazDesNet is fixed. Different from the training procedure in which synthetic hazy images are fed, any real-world hazy image can be taken as input in the inference procedure. We find that the HazDesNet trained with synthetic hazy images generalizes very well to realistic hazy images. The input of the network can be an RGB image of any size, and the output is the haze density map whose size is about half of the input image. The global haze density score is the average of the refined density map. Compared to the baseline method FADE [11], our proposed approach makes great progress in both qualitative results and quantitative results. Specifically, the high-resolution pixel-level haze density map predicted by our model is continuous, and does not have blocking effect. Besides, our method does not rely on extra distance information, the corresponding haze-free images, or complex handcrafted features.

The interpretability and visualization for deep neural networks are as important as the theory [31]. Some researchers have proposed methods to enable deep neural networks to be interpretable to human [32], [33]. As for our proposed HazDesNet, it is also very interesting to explore the output feature of intermediate layers and establish its connection with traditional haze related features. In our experiments, we find that a certain layer output of our proposed model is similar to the dark channel [8]. This similarity can be explained mathematically.

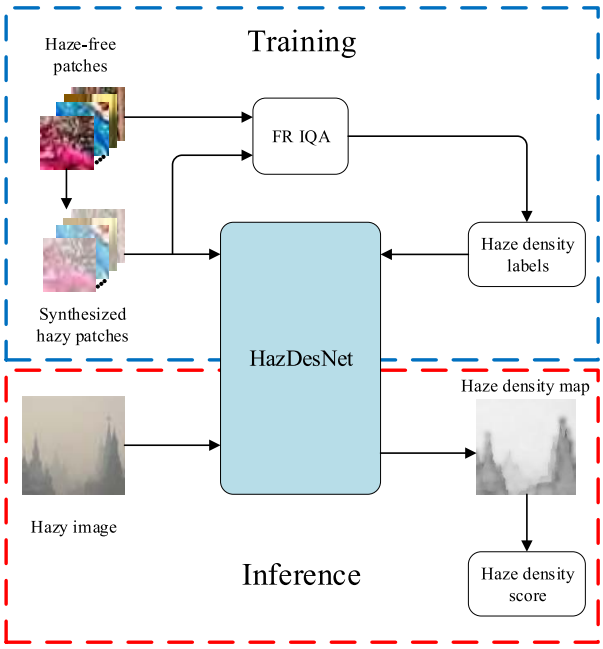


Fig. 1. The system overview. This system illustrates the training and inference modules of our proposed HazDesNet.

Our proposed model is first evaluated on the LIVE Defogging Image database [11] which contains 100 hazy images and the corresponding human perceptual judgments. However, as a haze density benchmark, this database is kind of small. To this end, we conduct a subjective haze perception experiment to build a Human Perceptual Haze Density (HPHD) database. The HPHD database contains two parts. One part includes 500 real-world hazy images, and another part includes 100 synthetic hazy images. All 600 hazy images are labeled with the human rated haze density scores. Quantitative results demonstrate that our HazDesNet outperforms existing methods on both databases in terms of Pearson's linear correlation coefficient (PLCC) and Spearman's rank ordered correlation coefficient (SROCC). Another advantage of the proposed HazDesNet lies in that it can also predict a pixel-level haze density map, which is absent from the existing methods. The HPHD database and the code of our HazDesNet will be publicly available to promote further study in this field.

The remainder of this article is organized as follows. In Section II, we review some existing methods. In Section III, we introduce some background of this work. The details and explanations of our proposed model are presented in Section IV. The subjective human study and our built HPHD database are introduced in Section V. Extensive experiments are performed to validate the proposed method in Section VI. Finally, we draw a conclusion in Section VII.

II. Related Works

There are some existing works that conducted in recent years on the haze density prediction. Huang *et al*. [34] propose a haze thickness estimation module to restore single-image visibility. But their proposed module is only used to refine the transmission map and it is limited to the sandstorm weather condition. Hautiére *et al*. [13] use an onboard camera to detect haze and estimate visibility distance. Nevertheless, their method depends on extra distance information obtained by the onboard camera, thus it works only in a certain condition and it is not a general method to predict haze density. Choi *et al*. [35] propose the first fog density prediction model, which is named as fog aware density evaluator (FADE) in [11]. Their model can predict haze density in general hazy weather conditions without a corresponding haze-free image, and also without other extra distance information. Unfortunately, their model involves many complex handcrafted features that may not generalize well. Moreover, their predicted haze density map has blocking effect and it is not smooth.

We overcome the above mentioned drawbacks existed in the existing methods. Specifically, our HazDesNet is a data-driven method so that it has better generalizability to predict haze density. We apply the end-to-end training method to avoid handcrafted features, which improves the accuracy of haze prediction to a large extend. Moreover, The input hazy images need not be divided into patches, thus the resolution of the predicted haze density map is much larger than that of the existing methods.

III. Background

In this section, we introduce some prior knowledge which includes atmospheric scattering model, structural similarity and dark channel prior.

A. Atmospheric Scattering Model

The atmospheric scattering model is proposed by McCartney [36] and simplified by Narasimhan and Nayar [30]. In our proposed method, the synthetic hazy image can be generated from the haze-free image by this model. The reflected light from particles will scatter in the atmosphere, and enter into the camera diffusely. The mathematical description [30] of the model is

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

where $I(x)$ is the hazy image, $J(x)$ is the haze-free image, $t(x) \in [0, 1]$ is the medium transmission at each pixel x, and A is the global atmospheric light.

The medium transmission $t(x)$ indicates the degree of unscattered light, which is defined as

$$t(x) = e^{-\beta d(x)}, \tag{2}$$

where $d(x)$ is the depth of the scene, and β is the medium attenuation coefficient of atmosphere.

B. Structural Similarity

The structural similarity (SSIM) index is proposed by Wang *et al*. [37] to measure the similarity between two images. The SSIM predicts the degradation of image based on the properties of the human visual system. Different from traditional image quality assessment, the perception-based SSIM considers more about the perceived change of image structure.

In this article, the SSIM index is applied to measure the degradation between the synthetic hazy image patch and the corresponding haze-free patch. As shown in Fig. 1, one of the most important steps of our framework is to measure the haze density of synthetic hazy images using FR IQA. We find that the SSIM scores can indicate the haze density well, thus the scores computed by SSIM metric are used as the training labels of HazDesNet, and the synthetic hazy image patches are used as the training input. The correlation of SSIM scores and haze densities perceived by human is shown in Section IV-C.

C. Dark Channel Prior

The dark channel prior is an image property which describes that at least one channel of the haze-free image patch has some low-density pixels [8]. The dark channel is an important haze feature used for haze removal [38], [39]. In our experiments, we find that intermediate result of our CNN based haze density prediction model is related to the dark channel.

IV. Methodology

To predict the haze density of a hazy image, we propose our end-to-end trainable network. In this section, we present the architecture of HazDesNet and discuss the reasons for the network design. In addition, the correlation between structural similarity and haze density labeled by human is explored. The training method is also presented. Finally, we visualize the extracted feature and establish the connection between the certain layer output of our model and the dark channel.

A. Model Architecture

Our proposed model consists of feature extraction, feature mapping, local maximum and average calculation, maximum and average mix, and sigmoid activation modules. These modules are implemented by convolutional layers and pooling layers. The design of our model is illustrated in Fig. 2. We explain each module in detail.

1) Feature Extraction: In many digital image processing algorithms, the first and important step is feature extraction. The CNNs are successfully used to extract features without human intervention. To avoid the duplicate layers of Maxout units [40], we propose our feature extraction module to efficiently extract haze-related features by cross-channel fusion. The effectiveness of this module is verified in Section IV-D. The module consists of two convolutional layers. The first convolutional layer includes 24 filters of size 5×5, and the second layer includes 24 filters of size 1×1. The 1×1 convolution operator is first introduced to enhance model discriminability in [41], and is used to increase more non-linearity in

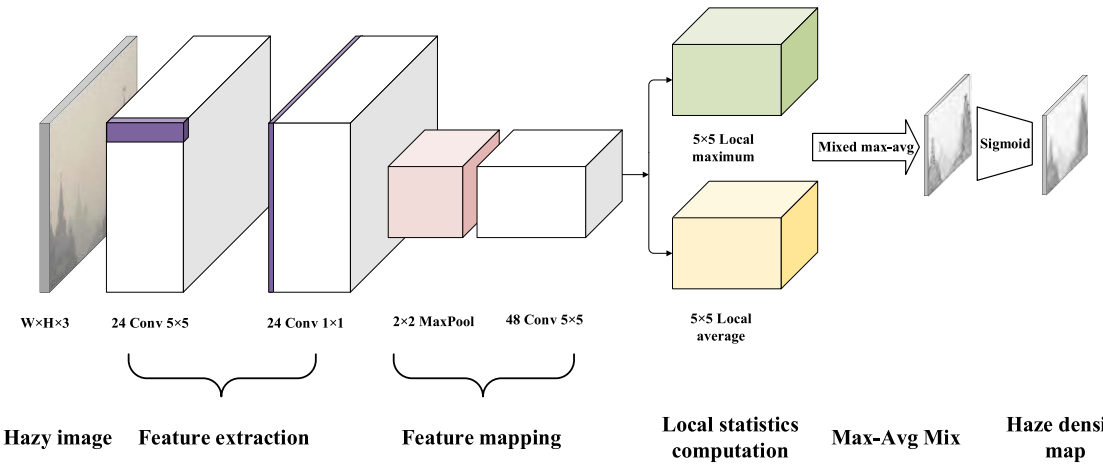


Fig. 2. The architecture of our proposed model. The input is a hazy image in arbitrary size, and the output is haze density map whose size is about half of the original image. The procedure includes feature extraction, feature mapping, local maximum and average calculation, maximum and average mix, and sigmoid activation.

GoogleNet [42] and ResNet [43]. We use 1×1 convolution to achieve cross-channel fusion in our feature extraction module. The output dimension of these two layers is both 24. Note that we do not set any zero paddings around the border, thus the output size of this module will be reduced by 4.

2) Feature Mapping: Max pooling layer is a downsampling operator performed along the spatial dimension. The max pooling is extensively used in CNNs for spatial size reduction, so that the parameters and computation can be reduced progressively [44]. We apply a max pooling layer of size 2×2 with a stride of 2 to downsample the feature. The output size is half of the previous layer size. A convolution layer is connected to the max pooling layer to map the features.

3) Local Statistics Computation: After the feature mapping, we calculate the local maximum and local average of the feature map. We assume that the medium transmission is locally constant. In other words, the transmission values within a small image patch (e.g. 16×16, 32×32) tend to be similar, since the pixels in a small patch have similar depths. This assumption is widely applied in haze removal methods, where the local maximum [21], [45] and local minimum [8], [9] are considered. Similarly, the haze density is also constant and continuous [35]. However, if the features are not very sparse, many local details are reduced by the local maximum computation [46]. To this end, we calculate the local average and local maximum together to maintain the details of features and density continuity.

4) Max-Avg Mix and Activation: The local maximum and local average have their own weakness [47]. The local maximum only considers the extremum and ignores the rapid change of the local region. On the other hand, the local average considers all the magnitudes of features, but reduces the contrast of the feature map. Therefore, many mix methods have been proposed [47]–[49]. These methods directly add the average and maximum together by a weight λ. The weight λ is always randomly initialized and learned by a new trainable layer. In our maximum and average (Max-Avg) mix module, two trainable convolutional layers are applied to mix the maximum and average, which is defined as

$$F = w_1 * F_{avg} + w_2 * F_{max} + b, \tag{3}$$

where F_{avg} and F_{max} are the local maximum and local average respectively, $w_1, w_2 \in \mathbb{R}^{6\times6\times48}$ and $b \in \mathbb{R}$ are the filter kernel and bias respectively.

The activation function is used for increasing the nonlinearity of deep neural networks. Common activation functions include Rectified Linear Unit (ReLU), sigmoid, TanH, etc. The ReLU is successfully used for image classification and overcomes the vanishing gradient issue. But there is no upper limit to the output of ReLU. The desired regression goal of our model is between 0 and 1. Thus, ReLU is not suitable for our regression task. Likewise, TanH is not suitable for our model. To sum up, the desired properties of activation function in our model include sufficient nonlinearity, range between 0 and 1, and continuity. Therefore, we choose sigmoid as our activation function. Although the vanishing gradient problem may occur with sigmoid, this issue can be alleviated through batch normalization [50]. The batch normalization layer is embedded into the feature mapping module. A summary of the configuration of our model is given in Table I.

The above modules construct our end-to-end trainable HazDesNet. The filter kernels and biases are parameters need to be learned. Based on the assumption that the medium transmission is locally constant, we can post-process the predicted haze density map. To get stable and continuous results, haze density map can be refined by performing a smoothing operator. Some simple and explicit filters such as average filter and Gaussian filter can smooth the haze density map, but can not preserve the edges. Therefore, we apply the guided filter [51] which is an edge-preserving smoothing technique to refine the predicted haze density map. Then, the global haze density score is the mean of the refined density map.

B. Training Method

It is not practical to collect plenty of hazy images and the corresponding haze-free images of exactly the same scene.

TABLE I
CONFIGURATION OF OUR MODEL

Layer	Filter	Num	Stride	Output Shape
Input	/	/	/	32×32×3
Conv2	5×5×3	24	1	28×28×24
Conv2	1×1×24	24	1	28×28×24
MaxPool2	2×2	24	2	14×14×24
Batch Normalization	/	/	/	14×14×24
Conv2	5×5×24	48	1	10×10×48
MaxPool/AvgPool	5×5	48×2	1	6×6×48×2
Max-Avg Mix	6×6×48×2	1	1	1×1×1
Activation	sigmoid			

Therefore, the lack of training images and their corresponding haze density scores is a challenging issue. Fortunately, hazy images can be synthesized with haze-free images using the atmospheric scattering model in Equation (1). Besides, we find that SSIM scores between synthetic hazy images and haze-free images can represent the haze densities well. Thus, we propose to use the synthesized images and the corresponding SSIM scores for training. It is difficult to train the model that takes a full-size synthetic hazy image as the input and its SSIM map as the target. Thus, HazDesNet is trained using hazy image patches and the corresponding SSIM labels, as shown in Fig. 1. An assumption can be introduced that the image content is independent of the transmission. This assumption is made based on the fact that the same image patch (image content) might have different scene depths. With this assumption, a haze-free image patch can be synthesized to hazy image patches with various transmissions. Therefore, we randomly crop haze-free image patches and synthesized hazy patches with various transmissions. The SSIM scores between the synthesized hazy patches and the corresponding original patches serve as the regression targets. The training dataset includes the synthetic hazy patches and the corresponding SSIM labels.

To sum up, the model can be denoted as $\mathcal{F}$ and training parameters are denoted as Θ. The loss function is defined as

$$L(\Theta) = \frac{1}{N}\sum_{i=1}^{N}\left\|\mathcal{F}\left(\Theta, I_i^P\right) - \mathrm{SSIM}_i^P\right\|^2, \tag{4}$$

where I^P is hazy image patch, SSIM^P is the SSIM index of I^P, $\|\cdot\|$ is L2 norm, and i is the index of image patches and the corresponding SSIM labels.

C. Relation Between SSIM and Haze Density

Hazy images are usually characterized with low contrast, shifted intensity, and faint color [35], thus it is possible to design objective algorithms to precisely measure the haze density of hazy images. But when designing and evaluating objective haze density prediction algorithms, the corresponding ground-truth of haze density is needed. Based on the fact that humans are good at identifying hazy areas [21], and humans are usually regarded as the ultimate arbiters of appearance of visual signals [11], [18], [35], [52], [53], we believe that human can perceive haze and judge haze density for a single hazy image accurately, thus the human labeled haze density scores are used as the ground-truth during the evaluation, which is the same as what has been done in [35]. The mean squared error (MSE), the peak signal-to-noise ratio (PSNR), and the structural similarity (SSIM) are three common metrics of FR IQA. The SSIM metric is more consistent with the human perception than MSE and PSNR [37]. To this end, we use the SSIM scores between the synthetic hazy images and the haze-free images as the training labels in the training procedure. SSIM scores are used because we observe that they are more suitable to describe the haze densities of synthetic hazy images than the MSE and PSNR scores. To justify this, we have included 100 synthetic hazy images in our subjective haze density study which is described in Section V. The Mean Opinion Scores (MOSs) in our subjective study represent the ground-truth haze density scores. The correlations between the three FR IQA scores and the ground-truth haze density scores are analyzed here.

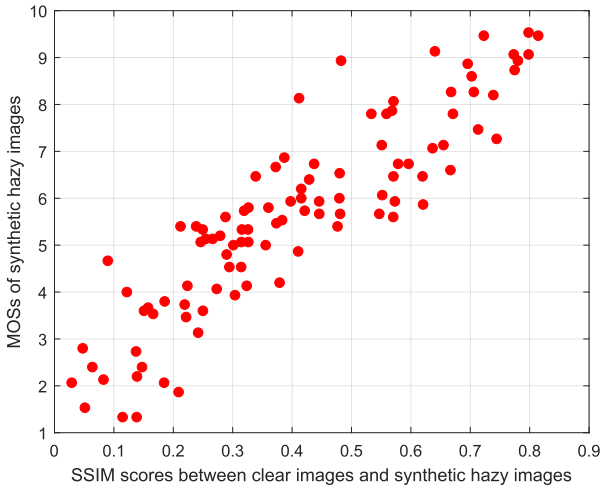


Fig. 3. SSIM scores and MOSs of synthetic hazy images.

TABLE II
PLCC AND SROCC BETWEEN FR IQA METRICS AND THE MOSS OF SYNTHETIC HAZY IMAGES

Metric	SSIM	PSNR	MSE
PLCC	**0.9095**	0.8060	0.6842
SROCC	**0.8947**	0.7817	0.7817

The Spearman and Pearson corelation coefficients are calculated, as shown in Table II. Quantitatively, the SROCC and the PLCC between the SSIM scores and the MOSs are 0.8947 and 0.9095, respectively, which are better than the correlations for the PSNR and MSE. It is clear that the SSIM is a much better metric to describe the haze density than PSNR and MSE. A scatter plot between the SSIM scores (mapped using a logistic non-linearity function introduced in Section VI-C) and the ground-truth human labeled haze density scores is shown in Fig. 3. It is observed that the SSIM and ground-truth haze density scores are highly correlated.

Fig. 4. Feature visualization. (a) is the original image, (b) is the dark channel with 2×2 slide window, and (c) is the output feature at the 14th slice of the feature extraction module.

D. Relation Between Inter-Network Feature and the Dark Channel

Deep neural networks need to be interpretable and explainable to humans [31]. The feature visualization plays an important role to understand the neural network and find out the function of a certain layer. We find that the output features of a certain layer in HazDesNet are related with the dark channel of the image. It can be explained in terms of the model architecture. The calculation process of dark channel includes two steps, i.e., a $r \times r$ slide windowing and an RGB cross-channel selection. This process is similar to our feature extraction module which contains a 5×5 convolutional operator and a cross-channel fusion layer.

We extract the dark channel of hazy images and show these images in Fig. 4 (a) and (b). The output features at the 14th slice of the feature extraction module are also illustrated in Fig. 4 (c). These output features are mapped into the range between 0 and 1 linearly for better visualization. It is clear that the dark channel features and our extracted features look very similar.

V. Subjective Assessment of Perceptual Haze Density

For evaluation of the image haze density prediction algorithms, ground-truth haze density scores of the hazy images are needed. Since it is not practicable to collect real-world hazy images and corresponding haze-free images, Choi *et al*. [11] perform a subjective human study using 100 images to construct the LIVE Image Defogging database to evaluate the FADE. In their human study, they ask the subjects to give a perceptual rating of the haze density for each hazy image. The statistical average of the ratings for each image is calculated as the ground-truth of the haze density. This database can be found in [54]. However, the database is kind of small for comprehensive evaluation. Therefore, we build another database called Human Perceptual Haze Density (HPHD) which contains a real-world hazy image (RHI) subset and a synthetic hazy image (SHI) subset. The former includes 500 real-world hazy images and the later includes 100 synthetic hazy images. A similar subjective human study is performed on this database. In total, the HPHD database includes 600 hazy images and their Mean Opinion Scores (MOSs) which represent the ground-truth haze densities. We evaluate the performance of HazDesNet using the HPHD and LIVE databases in Section VI.

Fig. 5. Some samples and their corresponding MOSs in our HPHD database.

A. Test Image

In the real-world hazy image (RHI) subset, we collect 500 hazy images from Flickr, which is an image hosting service. All images can be shared under the creative commons license. These images are searched by keywords like “haze”, “fog”, “mist”, etc. The content of these images is of diversity including square, mountain, forest, road, etc. These images have no weather restrictions. Since the database in [11] has included some well-known hazy test images, we do not repeat these images. The image resolution varies from 368×650 to 2048×1751. Fig. 5 shows some examples of the collected 500 hazy images. The MOS results of these samples are also listed for a better intuition of our subject study.

In the synthetic hazy image (SHI) subset, we collect synthetic haze images from various haze image datasets, which include I-HAZE [55], D-hazy [56], and HazeRD [57]. These three synthetic haze image datasets include various scenes with various haze densities. The resolutions of these images are different, and we resize the longest dimension of the image to 800 while maintaining its aspect ratio. Finally, we select 100 images from these synthetic haze image datasets to constitute the SHI subset. The number of images of the SHI subset is only 100, which is smaller than that of the RHI subset, because the main objective of the SHI subset is to verify the assumption that the SSIM metric can describe the haze density well. Besides, synthetic hazy images are not as realistic as real-world hazy images, thus we tend to include more images in the RHI subset.

B. Subjective Study

We invite 15 volunteers as subjects of our human study. All subjects have a corrected-to-normal vision. Before the

TABLE III
SUBJECTIVE EXPERIMENT SETTINGS

Categroy	Item	Detail
Monitor	Model	Dell U2417H
	Resolution	1920×1080
Methodology	Method	Single-stimulus
	Quality-scale	11-grade categorical
	Presentation order	Random
Test setting	Subjects number	15
	Viewing distance	3 times screen height
	Envrionment	Laboratory

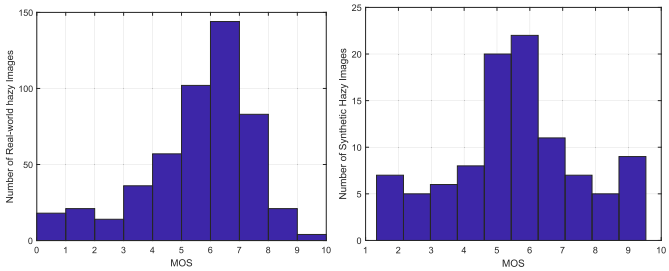
experiment starts, all subjects are instructed to evaluate the haze density of the shown image. A training session is added before the formal study. The subjects can have a sensation of the haze density range of the whole database and also learn how to rate the haze density in the training session.

A MATLAB user interface program is developed in a Windows PC for this study. This user interface displays a hazy image, and the subjects can give the corresponding haze density rating. The screen has a resolution of 1920×1080 pixels and its refresh rate is 60 Hz. The user interface is presented at the center of 24" monitor whose type is Dell U2417H.

A single-stimulus continuous quality evaluation (SSCQE) [58] strategy is adopted. The subjective study is performed in a quiet laboratory environment, and no external events will interfere the subjects during the study. The subject is required to rate the haze density of the displayed image. There are 11 integer score labels ranging from 0 to 10 in the rating bar. These score labels indicate the degrees of haze density. Score 0 represents that there is hardly any haze in the image. Score 10 indicates that the image is excessively hazy. After finishing rating an image, the user interface automatically displays the next image right away. The display order of testing images is randomly set for each subject. There is no time limit for rating each image, and the whole test procedure of one subject lasts about 30 to 45 minutes. A summary of the experiment settings is given in Table III.

C. Data Processing

We follow the method in [58] to exclude outliers and reject subjects. The raw haze density rating for a hazy image is detected as outlier, if it is far from the average (2 or $\sqrt{20}$ standard deviations for the Gaussian or non-Gaussian case). Besides, a subject with more than 5% outlier ratings is rejected as an outlier subject. Both outlier ratings and outlier subjects are excluded from following processes. Since the rating score of our subjective human study ranges from 0 to 10, we do not apply score mapping to preprocess the raw ratings. Mean Opinion Score (MOS) is calculated to represent the ground-truth of the haze density of each image. The histograms of MOS of two sets are shown in Fig. 6. It is clear that the MOS distributes widely from 0 to 10.



(a) (b)

Fig. 6. Statistics of MOS for two sets of HPHD database. (a) is histogram of MOS of the RHI subset, (b) is histogram of MOS of the SHI subset.

VI. EXPERIMENTAL VALIDATION

In this section, various experiments and studies are performed to validate HazDesNet. We first introduce the experimental settings including training data, training parameters, etc. Then, the performance of our model architecture is illustrated. The quantitative results of our model and the FADE [11] are compared on different databases. Finally, the qualitative results on real-world images are shown, and they demonstrate the improvements of our model.

A. Network Training Details

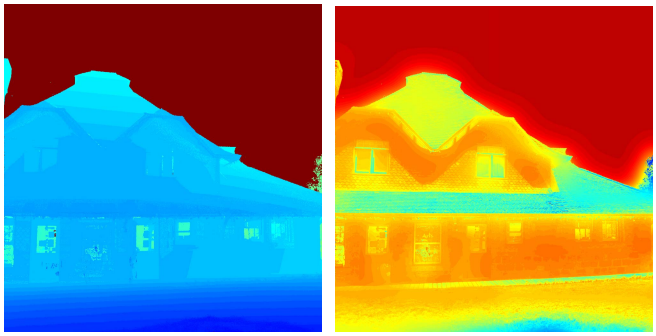
To train our model, we collect some haze-free images from the Internet to generate our training set. A total of 4,000 haze-free patches of size 32×32 are randomly cropped from these images. For each patch, we uniformly sample $t \in (0, 1)$ to generate 10 synthetic hazy patches and calculate the SSIM scores between the hazy and haze-free patches. Thus, a total of 40,000 hazy image patches and the corresponding SSIM labels are used to train our model. These patches are split into two parts randomly: a training set with 75 percent of patches and a validation set with 25 percent of patches.

We implement our model using the $Keras$ package. The configuration of the model is shown in Fig. 2 and summarized in Table I. RMSprop is used to optimize our model with a learning rate of 0.001 and a default rho of 0.9. The decay rate of learning rate is 10^{-3} for every epoch. We set the batch size to 512 and train the model with 1,000 rounds. We do not apply any data augmentation techniques to enlarge our training set, since it is easy to generate more training data and the current data is enough for the network training. Based on these configurations, HazDesNet is trained on a server with Nvidia GeForce GTX 1080 GPU.

B. Comparison Between the Transmission Map and the Haze Density Map

There is no doubt that both transmission map and haze density map can reflect the degree of hazing in the image, however they still have significant differences. In this section, the differences between the transmission map and the haze density map are discussed. According to the atmospheric scattering model, $t(x) = e^{-\beta d(x)}$ is the so called transmission, where $-\beta d(x)$ is the optical thickness. Therefore, if β is a constant, the transmission map is determined by the depth

(a) Haze-free image. (b) Synthetic hazy image.



(c) Reversed transmission map. (d) Haze density map.

Fig. 7. Comparison between the transmission map and the haze density map.

of a scene. The pixels at the same depth have the same transmission. However, these pixels of the same transmission might have different visual features, such as edges, textures, colors, etc. Thus, even if the pixels have the same transmission, they still have different haze perceptions, and transmission map for these pixels will also be different from their haze density map. Fig. 7 is an illustration, showing a haze-free image, a synthetic haze image, the corresponding transmission map and a predicted haze density map. The transmission map is reversed to be positively related to the haze density map. It is clear that the texture variation of the wall is not reflected in the transmission map, however it is reflected in the haze density map. Another apparent example is the region of the large white window, which has a similar transmission but a relatively lower haze density than its surroundings. The difference between the large white window and its surroundings is easily observed in the haze density map, while the transmission map can not reflect this difference.

C. Performance Evaluation Criteria

The Pearson's linear correlation coefficient (PLCC) and Spearman's rank ordered correlation coefficient (SROCC) between the haze density scores D and MOSs are calculated to evaluate the performance. Before PLCC and SROCC computation, the predicted haze density scores D are mapped by a logistic non-linearity function [18], [59], [60].

D. Ablation Study: Performances of HazDesNet With Different Configurations

In the proposed HazDesNet, two components are designed and chosen especially for certain purposes. The maximum

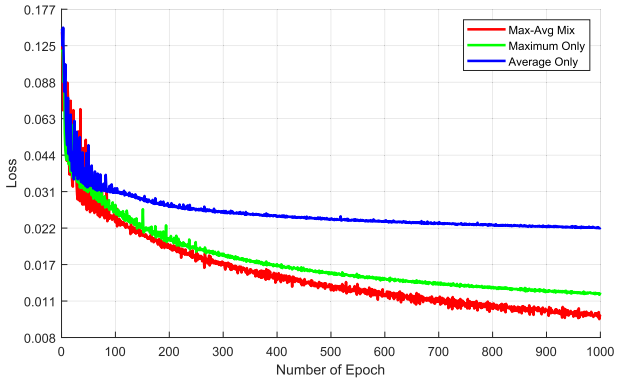


Fig. 8. Training process with different maximum and average layers.

TABLE IV
HAZE ESTIMATION ACCURACY OF HAZDESNET WITH DIFFERENT FUSION LAYERS

Fusion Layer	Max-only	Avg-only	Max-Avg Mix
LIVE Defogging Database			
PLCC	0.8819	0.8451	**0.9156**
SROCC	0.8708	0.8326	**0.9056**
RHI subset of HPHD Database			
PLCC	0.7784	0.7210	**0.8184**
SROCC	0.7938	0.7381	**0.8392**
SHI subset of HPHD Database			
PLCC	0.8891	0.8634	**0.9082**
SROCC	0.8712	0.8418	**0.8822**

and average (Max-Avg) mix module is designed for local maximum and local average fusion in an automatically learning way. The sigmoid activation function is chosen because of the reasonable nonlinearity, range, and continuity. In this section, the effectiveness of Max-Avg mix module and sigmoid function is illustrated. In addition, the number of filters in the feature mapping module is an important hyperparameter, and we explore the trade-off between the number of parameters and performance.

1) Effectiveness of Max-Avg Mix: To evaluate the effectiveness of Max-Avg fusion, we remove the Max-Avg mix module and replace it with either a local maximum layer or a local average layer. These modified models are trained under the same settings.

The performance of Max-Avg mix module and the effectiveness of Max-Avg fusion are illustrated in Fig. 8, which shows the training process of HazDesNet with a Max-Avg mix module, or with the maximum-only or average-only layer. The convergence speed of Max-Avg mix module is the fastest, and the convergence result is also the best. Besides, the haze estimation accuracy is also compared to verify the effectiveness of Max-Avg fusion, as shown in Table IV.

2) Comparison of Activation Functions: The performance of sigmoid activation is compared with the performance of Tanh, linear, and BReLU activation functions. The BReLU [21] is the special case of adjustable bounded rectifier [61]. BReLU is useful for image restoration whose definition is $f(x) = \min(0, \max(1, x))$.

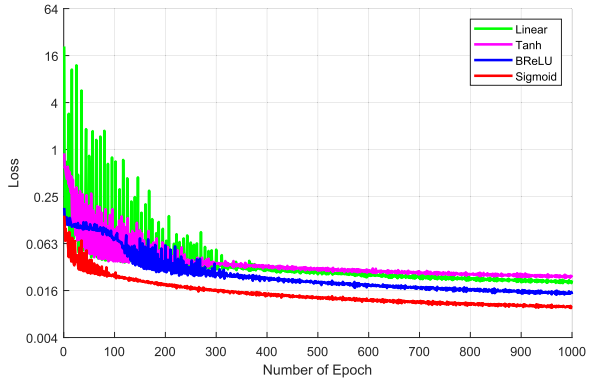


Fig. 9. Training process with different activation functions.

TABLE V
HAZE ESTIMATION ACCURACY OF HAZDESNET WITH DIFFERENT ACTIVATION METHODS

Activation	Linear	Tanh	BRelu	Sigmoid
LIVE Defogging Database				
PLCC	0.8564	0.8560	0.8767	**0.9156**
SROCC	0.8455	0.8437	0.8621	**0.9056**
RHI subset of HPHD Database				
PLCC	0.7409	0.7342	0.7589	**0.8184**
SROCC	0.7510	0.7493	0.7797	**0.8392**
SHI subset of HPHD Database				
PLCC	0.8708	0.8698	0.8944	**0.9082**
SROCC	0.8491	0.8428	0.8751	**0.8822**

Fig. 9 illustrates the comparison of different activations during the training process. The sigmoid has the smallest convergence loss. The loss of linear and Tanh activations vibrates intensely during the first 300 rounds. The final stable convergence of linear and Tanh is larger than the BReLU and sigmoid, since the range of BReLU and sigmoid is between 0 and 1 which is consistent with SSIM index. The BReLU has a good performance in [21] because their model applies more nonlinear mapping layers than HazDesNet. Clearly, the nonlinearity of sigmoid improves convergence precision. In addition, the sigmoid also improves haze estimation accuracy. In Table V, the PLCC and SROCC performances of sigmoid are the best comparing with other activation functions.

3) Filter Number of Feature Mapping: In the feature mapping module of HazDesNet, the number of filters highly affects the performance of the model. In common sense, the larger the number of filters, the more precise convergence. However, with a large number of filters, the model parameters increase. Therefore, the trade-off between the filters number and performance needs to be investigated. In the feature mapping module, we use 24, 48, and 96 as the number of filters to fine-tune HazDesNet.

In Table VI, we list the training results using different numbers of filters. This table includes the information including training/validation mean square error (MSE) and number of parameters. From Table VI, it is shown that the 24-filter layer has the smallest number of parameters, but the training performance is the worst. In addition, the MSE of the 96-filter layer is nearly equal to the MSE of the 48-filter layer. However, the smaller network is preferred if the desired goal is accomplished. That is why we apply a 48-filter layer in the feature mapping module in our model.

TABLE VI
TRAINING RESULTS USING DIFFERENT FILTER NUMBERS

Filter No.	Training MSE	Validation MSE	Parameter No.
24	0.0121	0.0196	14424
48	0.0098	0.0180	28848
96	0.0093	0.0178	57696

E. Quantitative Evaluation

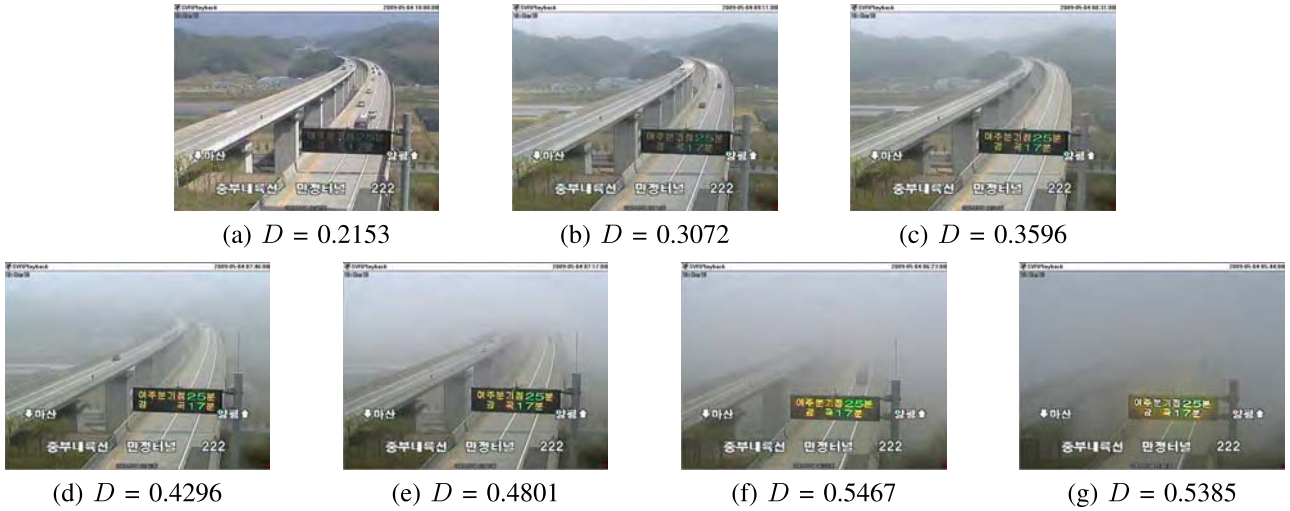
In this subsection, we evaluate the proposed HazDesNet quantitatively, and compare it with the state-of-the-art. HazDesNet takes a hazy image as input and predicts its haze density map. The haze density map is refined by the guided filter. The mean of the refined density map is calculated as the overall haze density D. The performance of the HazDesNet is evaluated on 7 representative surveillance images, LIVE Defogging database, and our built Human Perceptual Haze Density (HPHD) database.

1) Compared Methods: The previous studies of haze density prediction are limited to certain conditions, such as the onboard camera, geographical information, sandstorm environment, etc. Our HazDesNet is not limited by these restrictions, thus it is not compared with these methods with limitations. Instead, our HazDesNet is compared with the following methods: a haze density prediction model and three baseline models created on the basis of several widely recognized network architectures.

- **FADE**: It can predict the haze density using only one single hazy image. It is the most widely recognized haze density prediction model in the literature.
- **ResNet50**: We remove the fully connected layers at the end of the pre-trained models and add a 1×1 convolutional filter, and the mean of the final feature maps is used to represent the haze density.
- **GoogleNet**: The first seven inception modules of GoogleNet are applied to extract the feature maps. We also add a 1×1 convolutional filter, and calculate the average of the final feature maps as the haze density.
- **NASNet-Mobile**: We set the input size to $32 \times 32 \times 3$, and the output size is exactly $1 \times 1 \times 1$, which represents the haze density of the input image patch.

The training method of these models is the same with our HazDesNet. The activation functions of the final layers of these three models are all sigmoid. For all three networks except the traditional method FADE, we apply transfer learning, and fine-tune the models pre-trained on ImageNet.

2) Quantitative Results on Representative Surveillance Images: It is difficult to evaluate haze prediction algorithms and haze removal algorithms, because the corresponding haze-free image of a hazy image is unavailable. Fortunately, there are still a small number of surveillance images for haze prediction evaluation. First, seven representative surveillance



(a) $D = 0.2153$ (b) $D = 0.3072$ (c) $D = 0.3596$

(d) $D = 0.4296$ (e) $D = 0.4801$ (f) $D = 0.5467$ (g) $D = 0.5385$

Fig. 10. Seven representative hazy surveillance images of the same scene with different haze densities and their predicted haze density scores D.

TABLE VII

PLCC AND SROCC BETWEEN PREDICTED DENSITY SCORES D AND THE MOSS OF 7 REPRESENTATIVE SURVEILLANCE IMAGES

Image index	a	b	c	d	e	f	g
Density D	0.2153	0.3072	0.3596	0.4296	0.4801	0.5467	0.5385
MOS	8.10	16.45	30.85	43.90	59.95	77.05	81.90
PLCC	0.9960						
SROCC	0.9643						

images, as shown in Fig. 10, are used to evaluate our method. These seven surveillance images are captured in the same scene, at the same location, but at different times. With the weather changing, hazy images of different haze densities are captured by the surveillance camera. These time-varying real-world images are suitable to measure the performance of haze prediction method. Fig. 10 shows these seven surveillance hazy images from LIVE Defogging Image database. We calculate haze density score D for each hazy image using HazDesNet, and observe the consistency between these predicted scores and the MOSs labeled by human. The PLCC and SROCC are listed in Table VII.

From Table VII, we can observe that the PLCC and SROCC are 0.9960 and 0.9643, respectively. It is concluded that the predicted density scores of HazDesNet have a high correlation with the MOS of human judgment on these surveillance images. In addition, it is also proves that our method can be applied to monitor the haze condition in surveillance systems and autonomous vehicles.

The predicted haze densities of Fig. 10 (f) and (g) are not consistent with MOSs, i.e., the predicted haze density of Fig. 10 (f) is larger but its MOS is smaller. This inconsistency may be introduced by the following reasons. The brightness of Fig. 10 (g) is darker than Fig. 10 (f), thus the haze density of Fig. 10 (g) perceived by human is larger. However, in our training procedure, we set the global atmospheric light in Equation (1) as $A = 1$ for simplification. If the original image $J(x) \neq 1$ and $t(x) \neq 1$, this simplification results in an increase of brightness. Therefore, in this surveillance situation, our HazDesNet gives Fig. 10 (f) a higher haze density comparing with Fig. 10 (g). This also suggests that some future work is needed to improve the proposed method's robustness against global luminance variations.

3) Quantitative Results on LIVE Defogging Database: Besides a small number of surveillance images, our proposed model HazDesNet is also evaluated on LIVE Defogging database [54]. The quantitative results of HazDesNet are compared with FADE [11] and other three deep learning based methods in this part. The LIVE Defogging database contains 100 real-world hazy images and their corresponding MOSs. We also utilize PLCC and SROCC between the predicted haze density scores and MOSs to verify the performance of our method.

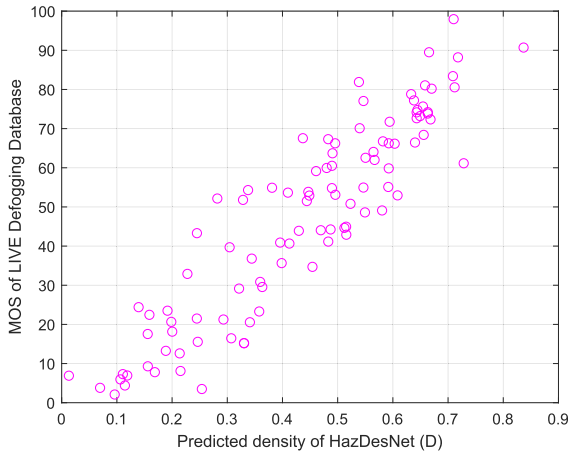
The performances of these methods are shown in Table VIII. FADE needs to divide the hazy image into small patches first and then estimates the haze density of the entire image. Therefore, FADE needs to select a best patch size during practice. Except for NASNet whose input is $32 \times 32 \times 3$, other deep learning based methods can predict haze density maps for hazy images of any size. In Table VIII, the performances of FADE with different patch sizes which range from 4×4 to 32×32 are tabulated. The best PLCC and SROCC of FADE are 0.8934 and 0.8756, respectively, but they are calculated with different patch sizes. The PLCC and SROCC of our HazDesNet are 0.9156 and 0.9056, respectively, which are better than FADE. What's more, HazDesNet has an advantage that it does not need any extra parameters such as patch sizes. Besides, our HazDesNet outperforms other deep learning based methods. That is mainly because these models are originally designed for other purposes, and the model architectures are not proper for this task.

Fig. 11 illustrates the scatter plots of HazDesNet and FADE on the LIVE Defogging database. These plots show the relation between the predicted haze density scores and MOSs judged by human subjects in the LIVE Defogging database. Higher density score represents heavier perceptual haze in the

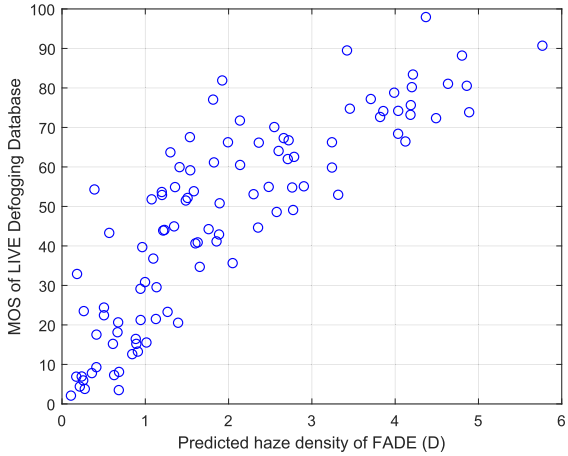
TABLE VIII

PLCC AND SROCC BETWEEN THE PREDICTED HAZE DENSITY SCORES AND THE MOSS OF HAZY IMAGES ON THE LIVE DEFOGGING DATABASE

Method	HazDesNet	FADE					ResNet50	GoogleNet	NASNet-Mobile
Patch size	/	4×4	8×8	10×10	16×16	32×32	/	/	32×32
PLCC	**0.9156**	0.8896	0.8899	0.8922	0.8934	0.8835	0.8752	0.8643	0.8514
SROCC	**0.9056**	0.8720	0.8756	0.8742	0.8723	0.8647	0.8641	0.8589	0.8464



(a) Scatter plot of HazDesNet on the LIVE Defogging database.



(b) Scatter plot of FADE on the LIVE Defogging database.

Fig. 11. (a) and (b) are the scatter plots of HazDesNet and FADE on the Live Defogging database, respectively. Both of these plots show the predicted haze density scores and MOSs of LIVE Defogging database.

TABLE IX

PLCC AND SROCC BETWEEN THE PREDICTED HAZE DENSITY SCORES AND THE MOSS OF HAZY IMAGES ON THE HPHD DATABASE

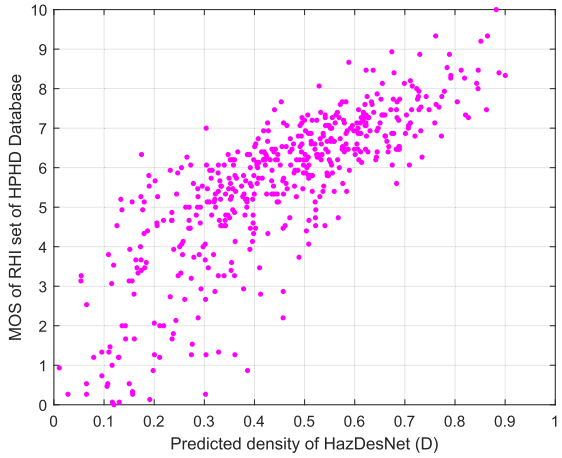
Method	HazDesNet	FADE					ResNet50	GoogleNet	NASNet-Mobile
Patch size	/	4×4	8×8	10×10	16×16	32×32	/	/	32×32
RHI subset									
PLCC	**0.8184**	0.6973	0.7066	0.7127	0.7154	0.7156	0.7744	0.7588	0.7489
SROCC	**0.8392**	0.7454	0.7550	0.7593	0.7608	0.7592	0.7613	0.7381	0.7334
SHI subset									
PLCC	**0.9082**	0.8754	0.8949	0.8993	0.9064	0.9062	0.8996	0.8967	0.8755
SROCC	**0.8822**	0.7914	0.8335	0.8419	0.8600	0.8733	0.8704	0.8581	0.8421

image. Fig. 11 (a) indicates that the predicted density scores of HazDesNet are highly correlated with human perception. Comparing the two plots in Fig. 11, it is observed that the scatter points of the HazDesNet are more convergent, which means that the HazDesNet has better predictions. What's more, it is also observed that the predictions of HazDesNet is also more linear than FADE.

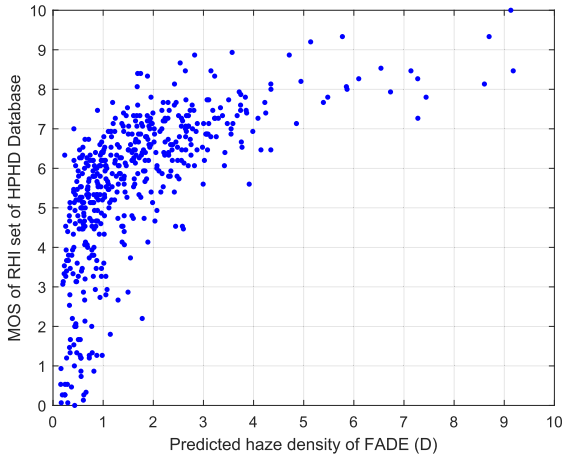
4) Quantitative Results on HPHD Database: Although the LIVE Defogging database includes hazy images of different contents and various haze densities, the number of images of this database is only 100 which is kind of small. Therefore, we build the HPHD database which includes a RHI subset and a SHI subset, and conduct a subjective human study in Section V. HPHD database also contains hazy images of different contents and their corresponding human judged haze densities which are described by MOSs.

In this part, HazDesNet is evaluated on the RHI subset and the SHI subset of the HPHD database, and it is also compared with FADE and other three fine-tuned neural network based methods. Similar to what have been done in Section VI-E, PLCC and SROCC are calculated to verify the performance of HazDesNet. The predicted haze density scores are passed through the non-linearity function. The performances of HazDesNet and FADE other methods are summarized in Table IX.

From Table IX, it is clearly observed that the performance of our HazDesNet is also the best. After a non-linearity mapping, the PLCC and SROCC between predicted density scores D of HazDesNet and MOSs of the RHI subset are 0.8184 and 0.8392, respectively. In terms of FADE, the best PLCC and SROCC on the RHI subset are 0.7156 and 0.7608, using the patch size 32×32 and 16×16, respectively. On the SHI subset, the PLCC and SROCC of HazDesNet are also better than FADE. It proves that our HazDesNet is better than FADE. The performance of HazDesNet is also better than other fine-tuned models on both the RHI subset and the SHI subset. On the other hand, compared with the performances of these two methods on the LIVE Defogging database, the performance

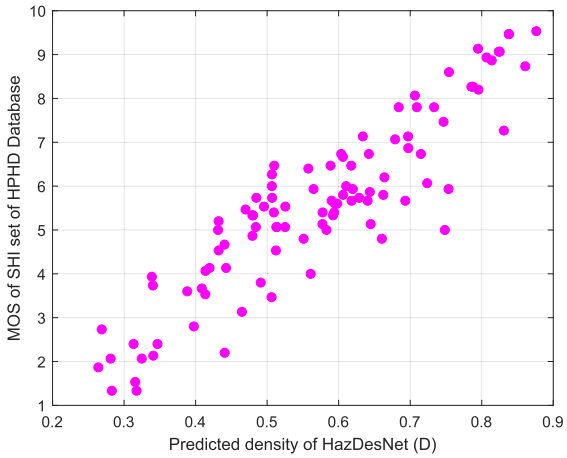


(a) Scatter plot of HazDesNet on the RHI subset.

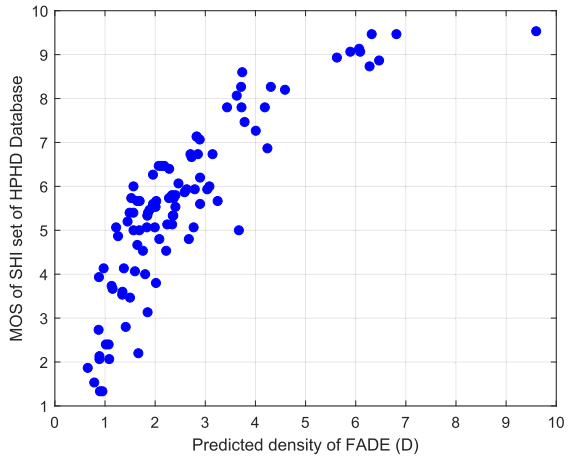


(b) Scatter plot of FADE on the RHI subset.

Fig. 12. (a) and (b) are the scatter plots of HazDesNet and FADE on the RHI subset, respectively. Both of these plots show the predicted haze density scores and MOSs of the RHI subset.



(a) Scatter plot of HazDesNet on the SHI subset.



(b) Scatter plot of FADE on the SHI subset.

Fig. 13. (a) and (b) are the scatter plots of HazDesNet and FADE on the SHI subset, respectively. Both of these plots show the predicted haze density scores and MOSs of the SHI subset.

on the HPHD database decreases a lot. That is because our HPHD database includes more hazy images and the included hazy images are also more challenging.

Fig. 12 (a) illustrates the scatter plot between the predicted haze density scores of HazDesNet and MOSs on the RHI subset of the HPHD database. The distribution of points in Fig. 12 (a) demonstrates that the predicted scores of HazDesNet are correlated well with the haze densities to some extent. However, the points in the left-bottom corner of Fig. 12 (a) are scattered, which reveals that HazDesNet could be further improved in this region. This imperfect distribution is probably because some white or gray objects in images are deemed as haze by the network and thus results in low predicted haze density scores. Meanwhile, the scatter plot between the predicted haze density scores of FADE and the MOSs on the RHI subset of the HPHD database is shown in Fig. 12 (b). From the scatter plot of FADE, we can observe that 86.8% of predicted haze density scores are smaller than 3 and only 8 density scores are larger than 7, which means that the predictions of FADE are uneven and highly nonlinear. Fig. 13 (a) illustrates the scatter plot between the predicted haze density scores of HazDesNet and MOSs on the SHI subset of the HPHD database, and Fig. 13 (b) illustrates the scatter plot between the predicted haze density scores of FADE and MOSs on the SHI subset of the HPHD database. In conclusion, our HazDesNet has a stronger ability to predict haze densities than FADE.

F. Qualitative Results

Fig. 14 shows the haze density maps predicted by FADE and our HazDesNet. These real-world hazy images are samples of the RHI subset of the HPHD database, as shown in Fig. 14 (a). Fig. 14 (b), (c), (d) show the results of FADE using patch sizes of 4×4, 16×16, and 32×32, respectively. Fig. 14 (e) shows the results of our HazDesNet.

It is challenging to predict the haze density map for the entire hazy image. FADE uses the image patch to predict local

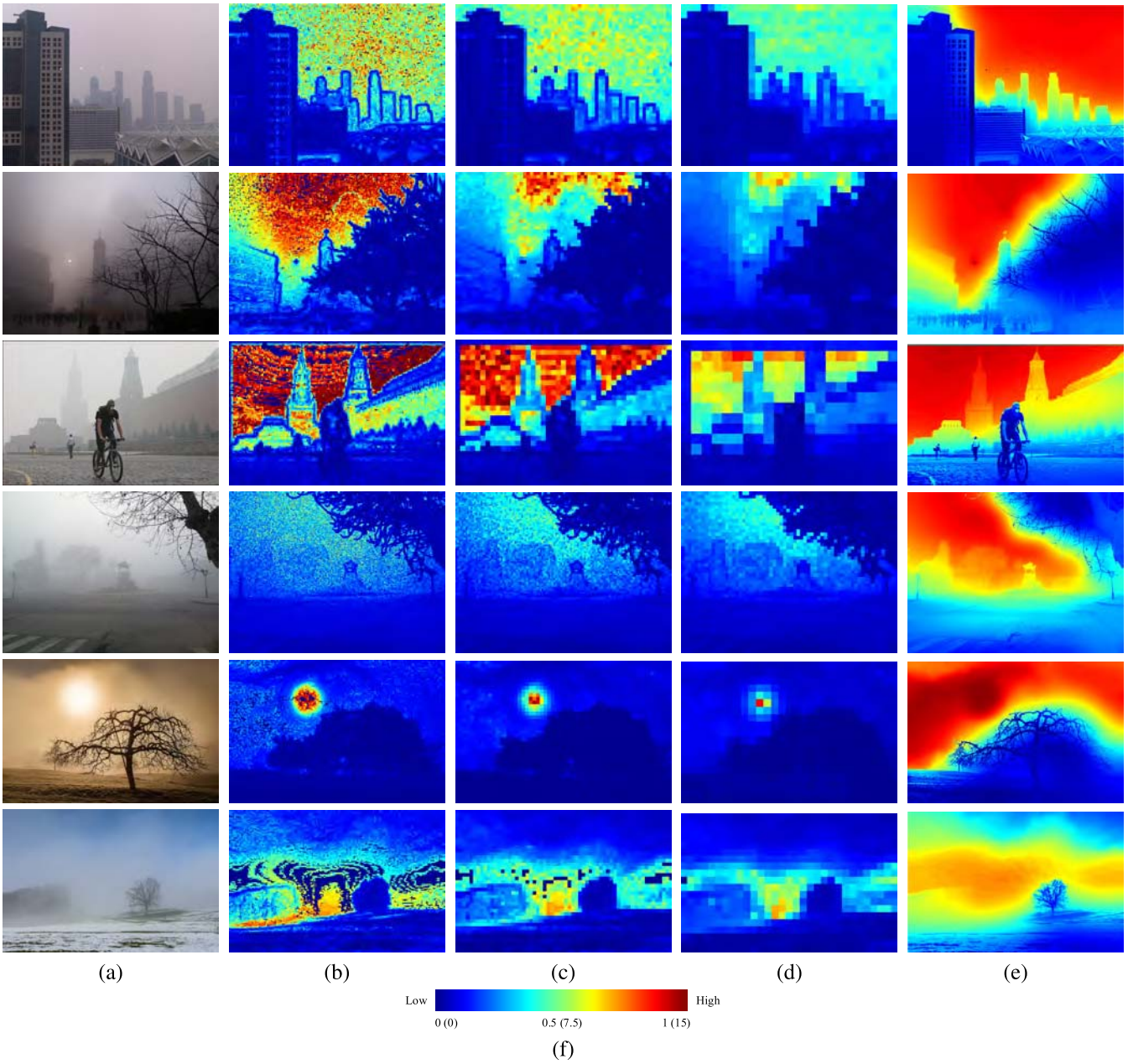


Fig. 14. Original hazy images and predicted haze density maps. (a) is original hazy images. (b), (c), and (d) are haze density maps predicted by FADE using 4×4, 16×16, and 32×32 patch size, respectively. (e) is haze density maps predicted by HazDesNet. (f) is colorbar showing color scale, where 0-15 is the indication of (b)-(d) and 0-1 is the indication of (e).

haze density, and then the haze density map is constructed by these local densities. To get a high-resolution density map, the patch size that FADE used needs to be as small as possible. However, the predicted density map of FADE is not continuous even when using a small patch size such as 4×4, as shown in Fig. 14 (b). To get a stable density map, FADE needs to apply a large patch size such as 16×16 and 32×32. However, these large patch sizes cause the blocking effect, as shown in Fig. 14 (c) and Fig. 14 (d). On the contrary, our HazDesNet is capable of predicting a continuous, stable, and high-resolution density map, as shown in Fig. 14 (e). The resolution of the predicted haze density map is about half of the resolution of the input haze image.

In terms of prediction precision, the predicted density maps of HazDesNet are of high quality and robust. The fourth and fifth rows of Fig. 14 demonstrate that our HazDesNet is more reliable than FADE in different scenes.

VII. Conclusion

In this article, we have proposed a novel end-to-end CNN-based method to predict haze density. Since it is hard to build a large scale haze density database, we first apply structural similarity (SSIM) index to measure the haze densities of synthetic hazy image patches. It is found that these SSIM scores indicate the haze densities of hazy images well. Thus, these hazy image patches and the corresponding SSIM

labeled haze densities are fed to train the HazDesNet. Although our HazDesNet is trained by synthetic hazy images, it can be generalized well for real-world hazy images. As another main contribution of this article, we construct a Human Perceptual Haze (HPHD) database which is the largest of its kind and includes 500 real-world hazy images and 100 synthetic hazy images, and the corresponding haze density labels judged by human. This database is used to evaluate the generalization ability of HazDesNet. The quantitative results of our proposed system on the HPHD database and existing databases demonstrate that HazDesNet has a good ability of haze density prediction. Another advantage of the proposed HazDesNet lies in that it can predict a high-resolution pixel-level haze density map, which describes the haze densities of single pixels in the image. Such a high-quality haze density map which is absent from the current methods can be of great value in many haze-relevant applications.

REFERENCES

[1] K. C. Dey, A. Mishra, and M. Chowdhury, "Potential of intelligent transportation systems in mitigating adverse weather impacts on road mobility: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1107–1119, Jun. 2015.

[2] M. Negru, S. Nedevschi, and R. I. Peter, "Exponential contrast restoration in fog conditions for driving assistance," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2257–2268, Aug. 2015.

[3] R. Gallen, A. Cord, N. Hautiere, E. Dumont, and D. Aubert, "Nighttime visibility analysis and estimation method in the presence of dense fog," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 310–320, Feb. 2015.

[4] V. R. Tomas, M. Pla-Castells, J. J. Martinez, and J. Martinez, "Forecasting adverse weather situations in the road network," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2334–2343, Aug. 2016.

[5] C. H. Bahnsen and T. B. Moeslund, "Rain removal in traffic surveillance: Does it matter?" *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 2802–2819, Aug. 2019.

[6] H. Kuang, X. Zhang, Y.-J. Li, L. L. H. Chan, and H. Yan, "Nighttime vehicle detection based on bio-inspired image enhancement and weighted score-level feature fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 927–936, Apr. 2017.

[7] M. Rezaei, M. Terauchi, and R. Klette, "Robust vehicle detection and distance estimation under challenging lighting conditions," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2723–2743, Oct. 2015.

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

[9] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015.

[10] S.-C. Huang, B.-H. Chen, and Y.-J. Cheng, "An efficient visibility enhancement algorithm for road scenes captured by intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2321–2332, Oct. 2014.

[11] L. Kwon Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3888–3901, Nov. 2015.

[12] M. Pavlic, H. Belzner, G. Rigoll, and S. Ilic, "Image based fog detection in vehicles," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2012, pp. 1132–1137.

[13] N. Hautiére, J.-P. Tarel, J. Lavenant, and D. Aubert, "Automatic fog detection and estimation of visibility distance through use of an onboard camera," *Mach. Vis. Appl.*, vol. 17, no. 1, pp. 8–20, Apr. 2006.

[14] X. Min, G. Zhai, K. Gu, X. Yang, and X. Guan, "Objective quality evaluation of dehazed images," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 2879–2892, Aug. 2019.

[15] X. Min *et al.*, "Quality evaluation of image dehazing methods using synthetic hazy images," *IEEE Trans. Multimedia*, vol. 21, no. 9, pp. 2319–2333, Sep. 2019.

[16] R. Spinneker, C. Koch, S.-B. Park, and J. J. Yoon, "Fast fog detection for camera based advanced driver assistance systems," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 1369–1374.

[17] M. Negru and S. Nedevschi, "Image based fog detection and visibility estimation for driving assistance systems," in *Proc. IEEE 9th Int. Conf. Intell. Comput. Commun. Process. (ICCP)*, Sep. 2013, pp. 163–168.

[18] G. Zhai and X. Min, "Perceptual image quality assessment: A survey," *Sci. China Inf. Sci.*, vol. 63, no. 11, Nov. 2020, Art. no. 211301.

[19] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, and C. W. Chen, "Blind quality assessment based on pseudo-reference image," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2049–2062, Aug. 2018.

[20] X. Min, G. Zhai, K. Gu, Y. Liu, and X. Yang, "Blind image quality estimation via distortion aggravation," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 508–517, Jun. 2018.

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

[22] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3194–3203.

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

[24] W. Ren *et al.*, "Gated fusion network for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3253–3261.

[25] A. Carlson, K. A. Skinner, R. Vasudevan, and M. Johnson-Roberson, "Modeling camera effects to improve visual learning from synthetic data," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 1–16.

[26] M. Afifi and M. Brown, "What else can fool deep learning? Addressing color constancy errors on deep neural network performance," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 243–252.

[27] M. Bucher, T.-H. Vu, M. Cord, and P. Pérez, "Zero-shot semantic segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 468–479.

[28] K. F. Hussain, M. Afifi, and G. Moussa, "A comprehensive study of the effect of spatial resolution and color of digital images on vehicle classification," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 1181–1190, Mar. 2019.

[29] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2247–2256, Aug. 2015.

[30] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2000, pp. 598–605.

[31] Q.-S. Zhang and S.-C. Zhu, "Visual interpretability for deep learning: A survey," *Frontiers Inf. Technol. Electron. Eng.*, vol. 19, no. 1, pp. 27–39, Jan. 2018.

[32] Q. Zhang, Y. N. Wu, and S.-C. Zhu, "Interpretable convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8827–8836.

[33] C.-C.-J. Kuo, M. Zhang, S. Li, J. Duan, and Y. Chen, "Interpretable convolutional neural networks via feedforward design," *J. Vis. Commun. Image Represent.*, vol. 60, pp. 346–359, Apr. 2019.

[34] S.-C. Huang, J.-H. Ye, and B.-H. Chen, "An advanced single-image visibility restoration algorithm for real-world hazy scenes," *IEEE Trans. Ind. Electron.*, vol. 62, no. 5, pp. 2962–2972, May 2015.

[35] L. K. Choi, J. You, and A. C. Bovik, "Referenceless perceptual fog density prediction model," in *Human Vis. Electron. Imag.*, vol. 9014, Feb. 2014, Art. no. 90140H.

[36] E. J. McCartney, "Optics atmosphere: Scattering by molecules particles," *Phys. Today*, vol. 30, no. 5, 1976, doi: 10.1063/1.3037551.

[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[38] J.-B. Wang, N. He, L.-L. Zhang, and K. Lu, "Single image dehazing with a physical model and dark channel prior," *Neurocomputing*, vol. 149, pp. 718–728, Feb. 2015.

[39] S.-C. Pei and T.-Y. Lee, "Nighttime haze removal using color transfer pre-processing and dark channel prior," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 957–960.

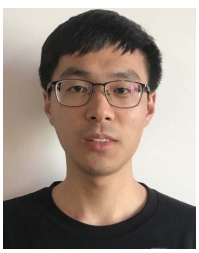
[40] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, "Maxout networks," 2013, *arXiv:1302.4389*. [Online]. Available: http://arxiv.org/abs/1302.4389

[41] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*. [Online]. Available: http://arxiv.org/abs/1312.4400

[42] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[45] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2995–3002.
[46] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 111–118.
[47] D. Yu, H. Wang, P. Chen, and Z. Wei, "Mixed pooling for convolutional neural networks," in *Rough Sets and Knowledge Technology*. Cham, Switzerland: Springer, 2014, pp. 364–375.
[48] C.-Y. Lee, P. W. Gallagher, and Z. Tu, "Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree," in *Proc. 19th Int. Conf. Artif. Intell. Stat.*, vol. 51, 2016, pp. 464–472.
[49] D. Fourure, R. Emonet, E. Fromont, D. Muselet, A. Tremeau, and C. Wolf, "Mixed pooling neural networks for color constancy," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3997–4001.
[50] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, vol. 37, Jul. 2015, pp. 448–456.
[51] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[52] X. Min, J. Zhou, G. Zhai, P. Le Callet, X. Yang, and X. Guan, "A metric for light field reconstruction, compression, and display quality evaluation," *IEEE Trans. Image Process.*, vol. 29, pp. 3790–3804, 2020.
[53] X. Min, K. Ma, K. Gu, G. Zhai, Z. Wang, and W. Lin, "Unified blind quality assessment of compressed natural, graphic, and screen content images," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5462–5474, Nov. 2017.
[54] L. K. Choi, J. You, and A. C. Bovik. (2015). *Live Image Defogging Database*. [Online]. Available: http://live.ece.utexas.edu/research/fog/fade_defade.html
[55] C. Ancuti, C. O. Ancuti, R. Timofte, and C. De Vleeschouwer, "I-HAZE: A dehazing benchmark with real hazy and haze-free indoor images," in *Advanced Concepts for Intelligent Vision Systems*. Cham, Switzerland: Springer, 2018, pp. 620–631.
[56] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, "D-HAZY: A dataset to evaluate quantitatively dehazing algorithms," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2226–2230.
[57] Y. Zhang, L. Ding, and G. Sharma, "HazeRD: An outdoor scene dataset and benchmark for single image dehazing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3205–3209.
[58] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document Recommendation ITU-R BT, 2002, pp. 500–513.
[59] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[60] X. Min, G. Zhai, J. Zhou, M. C. Q. Farias, and A. C. Bovik, "Study of subjective and objective quality assessment of audio-visual signals," *IEEE Trans. Image Process.*, vol. 29, pp. 6054–6068, Apr. 2020.
[61] Z. Wu, D. Lin, and X. Tang, "Adjustable bounded rectifiers: Towards deep binary representations," 2015, *arXiv:1511.06201*. [Online]. Available: http://arxiv.org/abs/1511.06201



Jiahe Zhang received the B.E. degree from Hangzhou Dianzi University, Hangzhou, China, in 2018. He is currently pursuing the master's degree with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include computer vision and multimedia signal processing.

Xiongkuo Min (Member, IEEE) received the B.E. degree from Wuhan University, Wuhan, China, in 2013, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2018. From January 2016 to January 2017, he was a Visiting Student with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He is currently a Post-Doctoral Fellow with Shanghai Jiao Tong University. His research interests include visual quality assessment, visual attention modeling, and perceptual signal processing. He received the Best Student Paper Award at IEEE ICME 2016.

Yucheng Zhu received the B.E. degree from Shanghai Jiao Tong University, Shanghai, China, in 2015, where he is currently pursuing the Ph.D. degree with the Institute of Image Communication and Network Engineering. His research interests include image quality assessment, visual attention modeling, and perceptual signal processing. He was a recipient of the Grand Challenge Best Performance Award in ICME 2017 and 2018.

Guangtao Zhai (Senior Member, IEEE) received the B.E. and M.E. degrees from Shandong University, Shandong, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009. From 2008 to 2009, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Post-Doctoral Fellow from 2010 to 2012. From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich Alexander University of Erlangen–Nuremberg, Germany. He is currently a Research Professor with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University. His research interests include multimedia signal processing and perceptual signal processing. He received the Award of National Excellent Ph.D. Thesis from the Ministry of Education of China in 2012.

Jiantao Zhou (Senior Member, IEEE) received the B.Eng. degree from the Department of Electronic Engineering, Dalian University of Technology, in 2002, the M.Phil. degree from the Department of Radio Engineering, Southeast University, in 2005, and the Ph.D. degree from the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, in 2009. He held various research positions with the University of Illinois at Urbana–Champaign, The Hong Kong University of Science and Technology, and McMaster University. He is currently an Associate Professor with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau. He holds four granted U.S. patents and two granted Chinese patents. His research interests include multimedia security and forensics, multimedia signal processing, artificial intelligence, and big data. He has coauthored two papers that received the Best Paper Award at the IEEE Pacific-Rim Conference on Multimedia in 2007 and the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo in 2016. He is also an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING.

Xiaokang Yang (Fellow, IEEE) received the B.S. degree from Xiamen University, Xiamen, China, in 1994, the M.S. degree from the Chinese Academy of Sciences, Shanghai, China, in 1997, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, in 2000.

From 2000 to 2002, he was a Research Fellow with the Centre for Signal Processing, Nanyang Technological University, Singapore. From 2002 to 2004, he was a Research Scientist with the Institute for Infocomm Research, Singapore. From 2007 to 2008, he visited the Institute for Computer Science, University of Freiburg, Freiburg im Breisgau, Germany, as an Alexander von Humboldt Research Fellow. He is currently a Distinguished Professor with the School of Electronic Information and Electrical Engineering and the Deputy Director of the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University. He has published over 200 refereed articles and has filed 60 patents. His current research interests include image processing and communication, computer vision, and machine learning.

Dr. Yang is a member of the Asia-Pacific Signal and Information Processing Association, the VSPC Technical Committee of the IEEE Circuits and Systems Society, and the MMSP Technical Committee of the IEEE Signal Processing Society. He is also the Chair of the Multimedia Big Data Interest Group of MMTC Technical Committee, and the IEEE Communication Society. He was a Series Editor of CCIS and an Editorial Board Member of *Digital Signal Processing*. He is also an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA and a Senior Associate Editor of the IEEE SIGNAL PROCESSING LETTERS.

Wenjun Zhang (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 1984, 1987, and 1989, respectively. After three years working as an Engineer with Philips, Nuremberg, Germany, he went back to his Alma Mater, in 1993, and became a Full Professor in electronic engineering, in 1995. As the Project Leader, he successfully developed the first Chinese HDTV prototype system, in 1998. He was one of the main contributors of the Chinese DTTB Standard (DTMB) issued, in 2006. He holds more than 76 patents and authored/coauthored more than 90 papers in international journals and conferences. He is the Chief Scientist of the Chinese Digital TV Engineering Research Centre, an industry/government consortium in DTV technology research and standardization, and the Director of Cooperative MediaNet Innovation Center (CMIC) an excellence research cluster affirmed by the Chinese Government. His main research interests include digital video coding and transmission, multimedia semantic analysis, and broadcast/broadband network convergence.