Deep Network-Enabled Haze Visibility Enhancement for Visual IoT-Driven Intelligent Transportation Systems

Ryan Wen Liu, *Member, IEEE*, Yu Guo, Yuxu Lu, Kwok Tai Chui, *Member, IEEE*, Brij B. Gupta, *Senior Member, IEEE*

***Abstract*—The Internet of Things (IoT) has recently emerged as a revolutionary communication paradigm where a large number of objects and devices are closely interconnected to enable smart industrial environments. The tremendous growth of visual sensors can significantly promote the traffic situational awareness, traffic safety management, and intelligent vehicle navigation in intelligent transportation systems (ITS). However, due to the absorption and scattering of light by the turbid medium in atmosphere, the visual IoT inevitably suffers from imaging quality degradation, e.g., contrast reduction, color distortion, etc. This negative impact can not only reduce the imaging quality, but also bring challenges for the deployment of several high-level vision tasks (e.g., object detection, tracking and recognition, etc.) in ITS. To improve imaging quality under the hazy environment, we propose a deep network-enabled three-stage dehazing network (termed TSDNet) for promoting the visual IoT-driven ITS. In particular, the proposed TSDNet mainly contains three parts, i.e., multi-scale attention module for estimating the hazy distribution in the RGB image domain, two-branch extraction module for learning the hazy features, and multi-feature fusion module for integrating all characteristic information and reconstructing the haze-free image. Numerous experiments have been implemented on synthetic and real-world imaging scenarios. Dehazing results illustrated that our TSDNet remarkably outperformed several state-of-the-art methods in terms of both qualitative and quantitative evaluations. The high-accuracy object detection results have also demonstrated the superior dehazing performance of TSDNet under hazy atmosphere conditions. The source code is available at https://github.com/gy65896/TSDNet.**

***Index Terms*—Visual Internet of Things, smart traffic services, image dehazing, deep network, intelligent transportation systems**

I. Introduction

With the rapid developments of communications, networks, sensor devices, and data science technologies, the Internet of Things (IoT) has attracted significant attention from both industry and academia [1]–[5]. The tremendous growth of visual sensors can specialize in sensing and gathering imaging data [6], directly contributing to the extension of industrial IoT in vision-empowered video surveillance. With the advanced IoT, visual perception and computation have become an integral part of smart traffic services, ranging from traffic situational awareness, traffic safety management to intelligent vehicle navigation. As shown in Fig. 1, the distributed computing and high extensibility properties of IoT make the deployment of intelligent transportation systems (ITS) practicable. In the visual IoT-enabled ITS, intelligent computing technology is essential to promote intelligent monitoring, intelligent navigation, collision avoidance, and transport efficiency, etc. However, due to the scattering and absorption of ambient light by turbid medium in atmosphere, the video/image captured by the observers is inevitably attenuated under the hazy imaging condition. This degenerate phenomenon will seriously reduce the contrast and color fidelity across the image. Meanwhile, the foreground and background will become blurred and deformed, causing the critical target information to be hidden. In particular, the visibility degradation will be more serious as the distance between the camera and the object or scene becomes larger. Therefore, the haze commonly presents non-homogeneous distribution in realistic imaging scenarios. It is necessary to design an advanced dehazer to effectively enhance visibility for promoting the visual IoT-enabled ITS.

To eliminate the haze, many handcrafted prior-based methods have been proposed. For example, He *et al.* [7] designed a dark channel prior (DCP) method to suppress the haze from the captured RGB image. However, DCP easily fails in the large sky regions and white scenes, resulting in unsatisfactory dehazing performance. Similarly, other prior-based dehazing methods [8]–[13] easily suffer from the low generalization ability, since these prior assumptions fail to accurately model the image degradation process in complex imaging situations. Inspired by deep learning, convolutional neural network (CNN)-based learning methods have been applied in various low-level visual tasks, e.g., low-visibility enhancement, denoising, and deraining, etc. Although the learning-based dehazing strategies [14]–[24] have received continuous attention, few methods perform haze visibility enhancement in different IoT-enabled ITS scenarios. Unlike other tasks, the dehazing network employed in IoT-enabled ITS firstly should have more stable and gratifying effects in outdoor traffic scenarios. Secondly, since the real-world haze is irregularly distributed in the spatial position, it is

This work was supported by the National Natural Science Foundation of China (No.: 51609195). (Corresponding authors: Kwok Tai Chui; Brij B. Gupta.)

R. W. Liu, Y. Guo, and Y. Lu are with the School of Navigation, Wuhan University of Technology, Wuhan 430063, China (e-mail: {wenliu, yuguo, yuxulu}@whut.edu.cn).

K. T. Chui is with the School of Science and Technology, Hong Kong Metropolitan University, Hong Kong (jktchui@hkmu.edu.hk).

B. B. Gupta is with the Department of Computer Science and Information Engineering, Asia University, Taichung 413, Taiwan (China), & Research and Innovation Department, Skyline University College, Sharjah P.O. Box 1797, United Arab Emirates, & Staffordshire University, Stoke-on-Trent ST4 2DE, UK (gupta.brij@gmail.com).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3170594, IEEE Transactions on Industrial Informatics

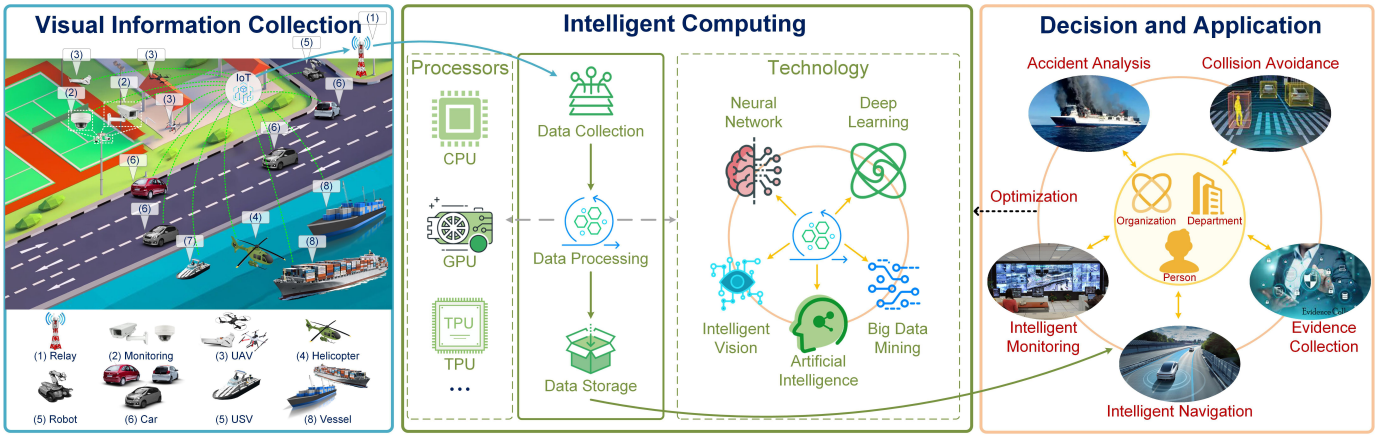


Fig. 1. The flowchart of the visual IoT-enabled intelligent transportation system, which includes visual information collection, intelligent computing, and decision and application. Note that UAV and USV, respectively, denote the unmanned aerial vehicle and unmanned surface vehicle.

more challenging to fully remove the haze and highlight the potential information for improving visibility and detection accuracy. Furthermore, it is necessary to real-timely restore the haze-degraded images in practical application.

To concern the above issues comprehensively, we propose to design a three-stage dehazing network (termed TSDNet). Given the previous studies, the main contributions of this work are as follows

- We propose a deep network-enabled haze visibility enhancement method for promoting the visual IoT-based ITS under hazy conditions. The proposed TSDNet is mainly composed of multi-scale attention, two-branch extraction, and multi-feature fusion modules.
- The proposed multi-scale attention module is capable of estimating the hazy distribution. The two-branch extraction network performs well in refining the hazy features from both spatial image and frequency domains. The outputs of attention and extraction modules are finally fused by a residual network to reconstruct the haze-free image with more structures and high-frequency details.
- We conduct comprehensive dehazing experiments on synthetic and real-world images to demonstrate the superiority of our TSDNet over other competitive methods. In addition, object detection experiments and running time analysis are also implemented on different devices to verify the practicability and efficiency of our method.

II. Related Works

In current literature, numerous dehazing methods have been proposed to improve imaging quality. We will briefly review the recent progress of image dehazing in this section.

A. Prior-Based Dehazing Methods

Generally, prior-based methods assume that the formation of haze is defined as follows

$$I(x) = e^{-\beta d(x)}(J(x) - 1) + A, \tag{1}$$

where x is the pixel index of the 2D image, I, J, and A denote the haze-degraded image, haze-free image, and atmospheric light value, respectively. $e^{-\beta d(x)}$ represents the transmission map t related to the scattering coefficient β and the scene depth d. In prior-based methods, many methods are proposed to directly estimate A and t. According to the statistical analysis of massive outdoor haze-free images, He *et al.* [7] found the haze-free image contains some pixels with a low value in at least one color channel and proposed dark channel prior (DCP). Based on the DCP theory and atmospheric scattering model, the restored image can be generated by estimating the thickness of the haze. Intuitively, some pixels in the white scenes and sky regions do not follow the DCP, which will generate the results with low-illumination and unnatural artifacts. For the sake of better dehazing effect, many DCP-based methods are proposed [8]–[10]. In particular, zhu *et al.* [9] designed a fusion of luminance and dark channel prior (F-LDCP) method, which exploits a sigmoid function to fuse the transmission maps generated by the luminance and DCP models for guaranteeing the sky naturalness. However, F-LDCP easily produces serious distortion in some real-world hazy images.

Furthermore, several physically-grounded prior strategies have been constructed [11]–[13]. For example, Fattal [11] proposed a local feature model (named CL), which explained the color-lines in the hazy scene. CL can generate the haze-free image based on the offset of the line from the origin. Zhu *et al.* [12] used color attenuation prior (CAP) to model the scene depth of hazy images and estimated the depth information through the supervised learning method. Berman [13] proposed a non-local prior (NL) model to characterize clean images. Although massive prior-based researches have been implemented, it is generally challenging to accurately model the haze in practical scenarios. The effectiveness of the prior-based dehazing methods strongly depends on the accuracy of assumptions. When the prior is insufficient to characterize the hazy degradation, algorithms will provide an unsatisfactory restored result.

B. Learning-Based Dehazing Methods

The data-driven convolutional neural network (CNN) has been applied in the image recovery task. Early CNN-enabled

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3170594, IEEE Transactions on Industrial Informatics

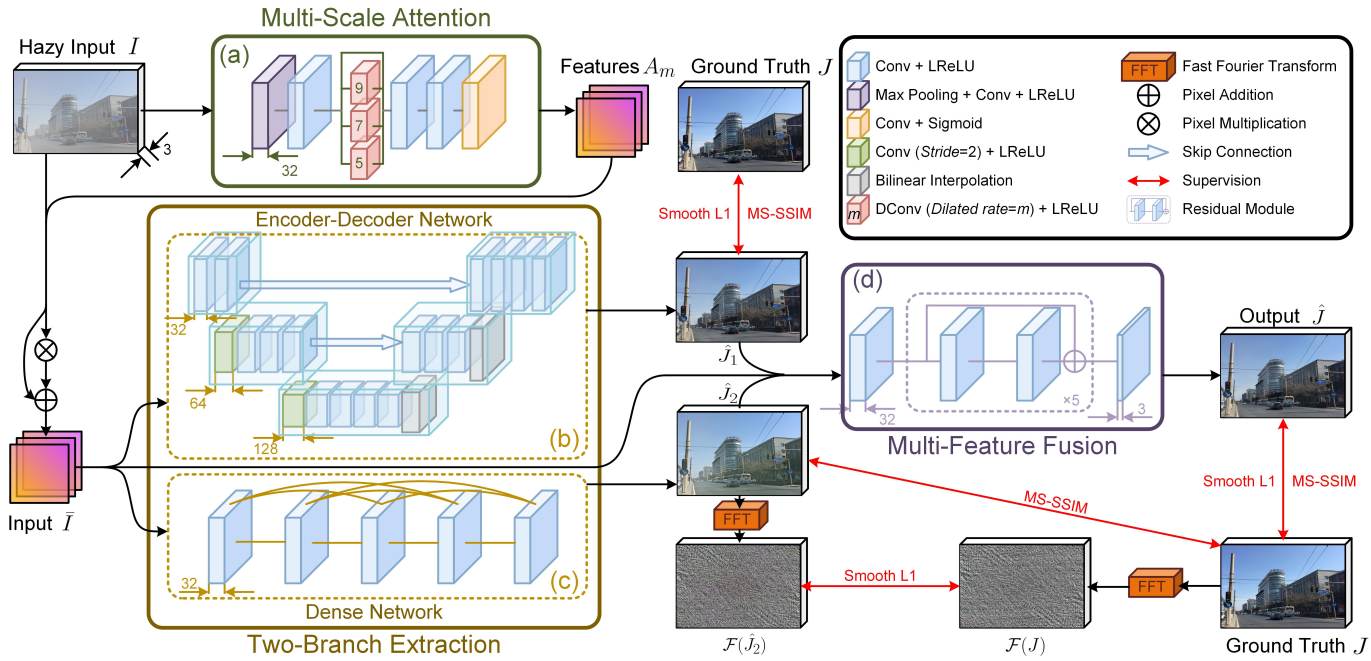


Fig. 2. The architecture of our three-stage dehazing network (termed TSDNet). TSDNet consists of three components, i.e., multi-scale attention, two-branch extraction, and multi-feature fusion. The Conv, DConv, and LReLU represent the convolution, dilated convolution, and leaky rectified linear unit activation function, respectively.

dehazing attempts reconstructed the haze-free images by estimating the transmission of the atmospheric scattering model, e.g., DehazeNet [14] and MSCNN [15]. Meanwhile, Zhang *et al.* [16] proposed a densely connected pyramid dehazing network (DCPDN) to jointly learn the transmission map and atmospheric light. However, it is a tricky issue to produce the labels of model parameters accurately. Once the estimation of parameters is inaccurate, a significant cumulative error will appear between the restored and clear image.

Recently, various end-to-end neural networks have been constructed to learn the translation of the haze-to-clear image. For instance, Li *et al.* [17] designed an all-in-one dehazing network (AODNet) based on the re-formulated atmospheric scattering model. Since AODNet collects hazy information using fewer parameters, it tends to provide restored images with low-illumination effects. In [18], a gated context aggregation network (GCANet) was constructed to learn the end-to-end mapping of hazy and clear images. GCANet introduces a smooth dilated convolution to solve the passive impact of gridding artifacts caused by the dilation technique. However, the non-homogeneous haze still remains in the result recovered by GCANet. To solve this issue, Liu *et al.* [19] designed an attention-based multi-scale end-to-end network (termed GridDehazeNet) to focus on the hazy distribution in the image domain. In particular, GridDehazeNet designs a channel-wise attention mechanism to flexibly adjust the weights of different scales for feature fusion. To further improve the capability of image detail extraction, a feature fusion attention network (FFANet) [20] was proposed to obtain the hazy distribution in channels and pixels. Meanwhile, many attempts based on generative adversarial network (GAN) have also been applied in image dehazing [21]–[24]. Engin *et al.* [21] designed an end-to-end cycle-consistent adversarial dehazing network (Cycle-Dehaze) without paired datasets. To more accurately estimate the transmission map in the real-world hazy scene, Dudhane *et al.* [22] designed an encoder-decoder-based generator network and an optical model to generate the haze-free image. Meanwhile, a residual initial (RI) module [23] composed of dense connections within the multi-scale convolutional layer is designed for learning the potential rich features in hazy images more efficiently. Mehta *et al.* [24] improved the visual quality of hazy images by generating an adversarial network and analyzing hyperspectral information.

Although extensive CNN- and GAN-based dehazing methods are oriented to improve imaging quality by increasing network parameters and depth, the expensive computation and redundant feature information will hinder the application of the hazy elimination technology. In addition, insufficient detail recovery capabilities will seriously affect the performance of visual IoT-enabled ITS.

III. TSDNet: Three-Stage Dehazing Network

Fig. 2 provides the schematic configuration of our TSDNet, which mainly contains three parts, i.e., multi-scale attention module for estimating the hazy distribution in the RGB image domain, two-branch extraction module for learning the hazy features, and multi-feature fusion module for reconstructing the latent haze-free image. In addition, a mixed loss function is reconstructed to improve the robustness of our network.

A. Multi-Scale Attention

Theoretically, the human visual system (HVS) selectively focuses on a certain specific signal while ignoring other

visible information. Inspired by this phenomenon, the attention mechanism is widely used in natural language processing (NLP), text recognition, image classification, etc. Since the haze is generally non-homogeneous, it is feasible and critical to employ an attention module for estimating the hazy distribution of the image domain before dehazing.

Fig. 2 (a) shows the proposed multi-scale attention subnetwork, which consists of a dilated convolution-based atrous spatial pyramid pooling (ASPP) module and multiple convolutional layers. In particular, we first adjust the size of the feature map as $1/2$ of the original resolution through a max-pooling operator to reduce the computational cost. A convolutional layer with the leaky rectified linear unit (LReLU) function is then arranged after max-pooling. Let ω, C, and $\mathcal{A}_l$, respectively, denote the max-pooling, convolution, and LReLU, this computing process can be mathematically expressed as follows

$$F_1^a = \mathcal{A}_l(C(\omega(I))), \tag{2}$$

with I and F_1^a, respectively, being the input haze-degraded image and output feature map. Three dilated convolutions C_m^d with the dilated rate $m \in \{5, 7, 9\}$ are simultaneously employed to extract the multi-scale hazy information, i.e.,

$$G_m = \mathcal{A}_l(C_m^d(F_1^a)), \tag{3}$$

with G_m being the feature map generated by the dilated convolutional layer with the dilated rate of m. Meanwhile, the three outputs and F_1^a are merged into one feature map, which is given by

$$F_2^a = [F_1^a, G_5, G_7, G_9]. \tag{4}$$

Finally, we adopt two convolutional layers and a bilinear interpolation upsampling operator to obtain the attention map A_m from F_2^a. The final output of multi-scale attention module $\bar{I}$ will be written as

$$\bar{I}(x) = I(x) * A_m(x) + I(x). \tag{5}$$

B. Two-Branch Extraction

The two-branch extraction module employs an encoder-decoder network and a dense network to obtain hazy features. Enormous studies have proved that the encoder-decoder network can get significant achievement in image restoration tasks. In particular, this network can fully extract the structure and texture features (i.e., multi-scale information) from input image [25]. Therefore, we use a three-scale encoder-decoder network to achieve hazy feature extraction, shown in Fig. 2 (b). In particular, two convolutions with step size $stride = 2$ are exploited to reduce image size in the encoder, and the decoder uses two bilinear interpolation operators to restore image size. Meanwhile, two skip connection operations are adopted between the encoder and decoder. These operations are able to transfer the hierarchical features learned by the encoder to the decoder, resulting in avoiding the loss of spatial information caused by down-sampling.

In addition, Fig. 2 (c) shows the proposed dense network, which only contains five convolutional layers. To avoid information loss, this network adopts many concatenation strategies to guarantee feature reuse and prevent gradient disappearance. Specifically, the output of all previous convolutional operations will be cascaded and fed to the current layer. For instance, let F_i^c be the output of i-th layer in the dense network, the operation of the 5-th convolutional layer can be written as

$$F_5^c = \mathcal{A}_l(C([F_4^c, F_3^c, F_2^c, F_1^c])). \tag{6}$$

Since the features yielded by each convolutional layer are reused, this dense network can extract rich feature information through fewer convolutional kernels. Meanwhile, our dense network adopts frequency domain supervision based on the fast Fourier transform to extract hazy features different from the encoder-decoder network. The specific details about frequency domain supervision in the dense network will be introduced in Section III-D1.

C. Multi-Feature Fusion

For the sake of better visual performance, the multi-feature fusion module employs a simple but effective residual network to integrate all hazy information. As shown in Fig. 2 (d), the outputs of the multi-scale attention module and the two-branch extraction module are fed into the multi-feature fusion module to reconstruct haze-free images. In particular, the proposed fusion module includes two convolutional layers and five residual modules. Each residual module is composed of two convolutional layers and a pixel addition operation, which can suppress the feature loss issue caused by convolution.

D. Reconstruction of Loss Function

This section introduces the mixed loss function of the proposed network. To achieve satisfactory imaging results, network parameters can be optimized by minimizing this loss function. In particular, this loss function $\mathcal{L}$ mainly consists of smooth L1 loss $\mathcal{L}_1^s$ and MS-SSIM loss $\mathcal{L}_{\text{MS-SSIM}}$. Let γ_1 and γ_2 be the trade-off parameters, the mixed loss function can then be given by

$$\mathcal{L} = \gamma_1 \mathcal{L}_1^s + \gamma_2 \mathcal{L}_{\text{MS-SSIM}}. \tag{7}$$

1) Smooth L1 Loss: Hang *et al.* [26] have verified L1 loss function has better effects than L2 in image restoration task. We thus use a robust L1 loss (termed smooth L1 [27]) to supervise the final output $\hat{J}$ and the outputs of the two-branch extraction module $(\hat{J}_1, \hat{J}_2)$, i.e.,

$$\mathcal{L}_1^s = L_1^s(J - \hat{J}) + \gamma_3 L_1^s(J - \hat{J}_1) + \gamma_4 L_1^s(\mathcal{F}(J) - \mathcal{F}(\hat{J}_2)), \tag{8}$$

with J being the ground truth, i.e., haze-free image. In Eq. (8), γ_3 and γ_4 denote the penalty coefficients. Since the encoder-decoder network and dense network are designed for respectively extracting feature information from the spatial and frequency domains, the fast Fourier transform operator $\mathcal{F}(\cdot)$ is used to convert J and $\hat{J}_2$ into the frequency domain. Let P denote the input, the operation of L_1^s is defined as follows

$$L_1^s(P) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{S}_{l1}(P(i)), \tag{9}$$

TABLE I
THE DETAILS OF THREE DATASETS USED IN OUR EXPERIMENTS. "REAL" AND "DEPTH" DENOTE WHETHER THIS DATASET IS REAL-WORLD HAZY IMAGES AND WHETHER IT CONTAINS DEPTH INFORMATION, RESPECTIVELY.

Dataset		Train	Test-S	Test-R	Test-D	Real	Depth
RESIDE [29]	OTS	2000	35	0	0		✔
	RTTS	0	0	45	4320	✔	
Seaships [30]		2000	35	0	0		
VOC2007 [31]		0	0	0	4952		

where i denotes the pixel index, N represents the sum of the image pixel. The smooth L1 operator $\mathcal{S}_{l1}$ is given by

$$\mathcal{S}_{l1}(P(i)) = \begin{cases} 0.5P^2(i), & \text{if } |P(i)| < 1, \\ |P(i)| - 0.5, & \text{otherwise.} \end{cases} \tag{10}$$

By comparison, the smooth L1 loss is less sensitive to outliers than L2 and has a faster convergence than raw L1.

2) MS-SSIM Loss: Although the smooth L1 loss preserves the color and luminance of each region with equal weight, the contrast of high-frequency regions in the image still needs to be improved. Therefore, we adopt MS-SSIM loss for capturing the superior hazy characteristics. Specifically, the MS-SSIM loss used in our TSDNet can be written as follows

$$\begin{aligned} \mathcal{L}_{\text{MS-SSIM}} = L_{\text{MS-SSIM}}(J, \hat{J}) + \gamma_5 L_{\text{MS-SSIM}}(J, \hat{J}_1) \\ + \gamma_6 L_{\text{MS-SSIM}}(J, \hat{J}_2), \end{aligned} \tag{11}$$

where $L_{\text{MS-SSIM}}$ represents the multi-scale structure similarity operator, γ_5 and γ_6 denote the penalty weights. Let $\hat{Y}$ and Y be the restored image and ground truth, SSIM value for pixel x is defined as follows

$$\begin{aligned} \text{SSIM}(x) &= \frac{2\mu_Y \mu_{\hat{Y}} + c}{\mu_Y^2 + \mu_{\hat{Y}}^2 + c} \cdot \frac{2\sigma_{Y\hat{Y}} + c_*}{\sigma_Y^2 + \sigma_{\hat{Y}}^2 + c_*} \\ &= l(x) \cdot cs(x), \end{aligned} \tag{12}$$

where c and c_* are two constants for avoiding a zero denominator. The means μ_Y, $\mu_{\hat{Y}}$, standard deviations σ_Y, $\sigma_{\hat{Y}}$, and covariance $\sigma_{Y\hat{Y}}$ are computed by a Gaussian filter. Finally, the operation of MS-SSIM is given by

$$L_{\text{MS-SSIM}} = 1 - l_M^{\alpha} \cdot \prod_{j=1}^{M} [cs_j]^{\beta_j}, \tag{13}$$

where l_M and cs_j represent the terms we defined in Eq. (12), M denotes the default parameter of scales. Please refer to [28] for detail parameter settings on MS-SSIM.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

This section mainly introduces the implementation details of the proposed method. Meanwhile, extensive experiments on both synthetic and real-world hazy images are conducted to verify the effectiveness of our TSDNet. Furthermore, the application of the dehazing method in object detection and running time analysis will be further discussed.

A. Datasets

In this paper, we employ three datasets to train and test our TSDNet, including realistic single image dehazing (RESIDE) [29], Seaships [30], and VOC2007 [31].

1) RESIDE: As a large-scale dataset specially applied in single image dehazing, RESIDE is divided into five subsets for different tasks. In particular, the outdoor training set (OTS) and real-world task-driven testing set (RTTS) from RESIDE are exploited because our dehazer is designed for special tasks. OTS is a synthetic hazy set, which includes 8970 haze-free images and the corresponding depth information. In contrast, RTTS is composed of 4322 real-world hazy images. Meanwhile, each image in RTTS has a corresponding annotation, which contains five object types, i.e., "Bicycle", "Bus", "Car", "Motorbike", and "Person".

2) Seaships: In 2018, Shao *et al.* [30] published a large-scale ship dataset (named Seaships), which contains 31455 images from 10800 real-world video segments. Massive high-visibility images can be extracted from the Seaships. Therefore, we will manually select extensive high-quality images from the Seaships as the training and testing datasets.

3) VOC2007: This dataset mainly contains 5011 training/validation images and 4952 test images. Meanwhile, VOC2007 has 20 types of object. Since our TSDNet is designed for outdoor traffic scenarios, we only consider eight types of object, i.e., "Aeroplane", "Bicycle", "Boat", "Bus", "Car", "Motorbike", "Person", and "Train".

To train the proposed network, we select 2000 haze-free images from RESIDE-OTS and 2000 high-visibility images from Seaships as the training dataset (termed Train). In particular, each image is cropped as several 200×200 patches. Each image patch is randomly rotated 90, 180, 270, and 360 degrees. Meanwhile, we randomly select 35 RESIDE-OTS images and 35 Seaships images as the synthetic test dataset (Test-S), 45 RESIDE-RTTS images as the real-world test dataset (Test-R), and 4952 VOC2007 images and 4320 RESIDE-RTTS images as the detection test dataset (Test-D), respectively. It is mentioned that the training and test datasets do not overlap. Since the Seaships and VOC2007 datasets lack the scene depth, we employ MegaDepth [33] to calculate the depth information of each pixel. Finally, the hazy images I can be generated by clear images J and depth information d through Eq. (1) with $\beta \in [0.08, 0.3]$ and $A \in [0.7, 1.0]$. The usage details of all datasets are organized in Table I.

B. Experimental Settings

Our TSDNet is trained on a PC with Intel (R) Core (TM) i5-10600KF CPU @ 4.10GHz and Nvidia GeForce RTX 2080 Ti GPU. In this work, the TSDNet is implemented based on the Pytorch 1.9.0 package in the Python 3.7 platform. Meanwhile, the Adam optimizer is exploited to train our network. The training epoch, initial learning rate, and batch size are set to 120, 10^{-4}, and 4, respectively. The learning rate is adjusted to $1/2$ by every 15 epoch. The six weight values are empirically set as $\gamma_1 = 1$ and $\gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = \gamma_6 = 0.5$, respectively. For a fair comparison, the parameters of all competing methods are optimally provided by the authors.

C. Visibility Enhancement Results on Synthetic Images

In this section, we evaluate our TSDNet and other state-of-the-art methods on the Test-S. In particular, four metrics,

TABLE II
PSNR, SSIM, FSIM, AND LPIPS RESULTS OF VARIOUS METHODS ON THE 35 RESIDE-OTS AND 35 SEASHIPS IMAGES FROM TEST-S. THE BEST RESULTS ARE IN **BOLD**.

Method	RESIDE-OTS [29]				Seaships [30]			
	PSNR↑	SSIM↑	FSIM↑	LPIPS↓	PSNR↑	SSIM↑	FSIM↑	LPIPS↓
Hazy	13.59±1.89	0.726±0.091	0.852±0.065	0.185±0.058	17.36±3.07	0.864±0.079	0.917±0.051	0.152±0.061
DCP [7]	15.34±2.67	0.826±0.081	0.938±0.026	0.136±0.060	12.30±1.52	0.677±0.065	0.904±0.018	0.354±0.083
CAP [12]	21.18±2.73	0.911±0.049	0.950±0.029	0.057±0.024	19.97±2.46	0.917±0.028	0.965±0.013	0.141±0.023
HL [32]	20.37±3.30	0.906±0.071	0.962±0.021	0.083±0.052	18.60±1.92	0.818±0.058	0.904±0.035	0.266±0.063
F-LDCP [9]	21.08±3.47	0.917±0.062	0.967±0.012	0.080±0.043	17.46±1.96	0.857±0.045	0.943±0.014	0.248±0.053
DehazeNet [14]	15.57±2.40	0.783±0.090	0.887±0.058	0.140±0.056	20.06±4.15	0.884±0.068	0.934±0.042	0.161±0.051
MSCNN [15]	18.67±2.80	0.869±0.081	0.947±0.031	0.084±0.046	23.87±3.35	0.940±0.026	0.971±0.016	0.111±0.019
AODNet [17]	17.99±1.78	0.854±0.062	0.888±0.040	0.100±0.032	17.45±2.57	0.879±0.032	0.916±0.023	0.174±0.034
GCANet [18]	19.93±3.28	0.887±0.055	0.949±0.025	0.190±0.047	19.86±3.36	0.893±0.060	0.960±0.012	0.154±0.057
GridDehazeNet [19]	24.04±3.43	0.950±0.026	0.980±0.012	0.030±0.014	24.26±3.13	0.943±0.027	0.965±0.014	0.065±0.024
FFANet [20]	22.63±3.69	0.927±0.052	0.968±0.019	0.048±0.036	21.31±3.69	0.919±0.047	0.952±0.029	0.093±0.040
TSDNet	**24.24±3.25**	**0.959±0.022**	**0.981±0.010**	**0.029±0.016**	**25.49±3.56**	**0.965±0.012**	**0.984±0.005**	**0.062±0.015**

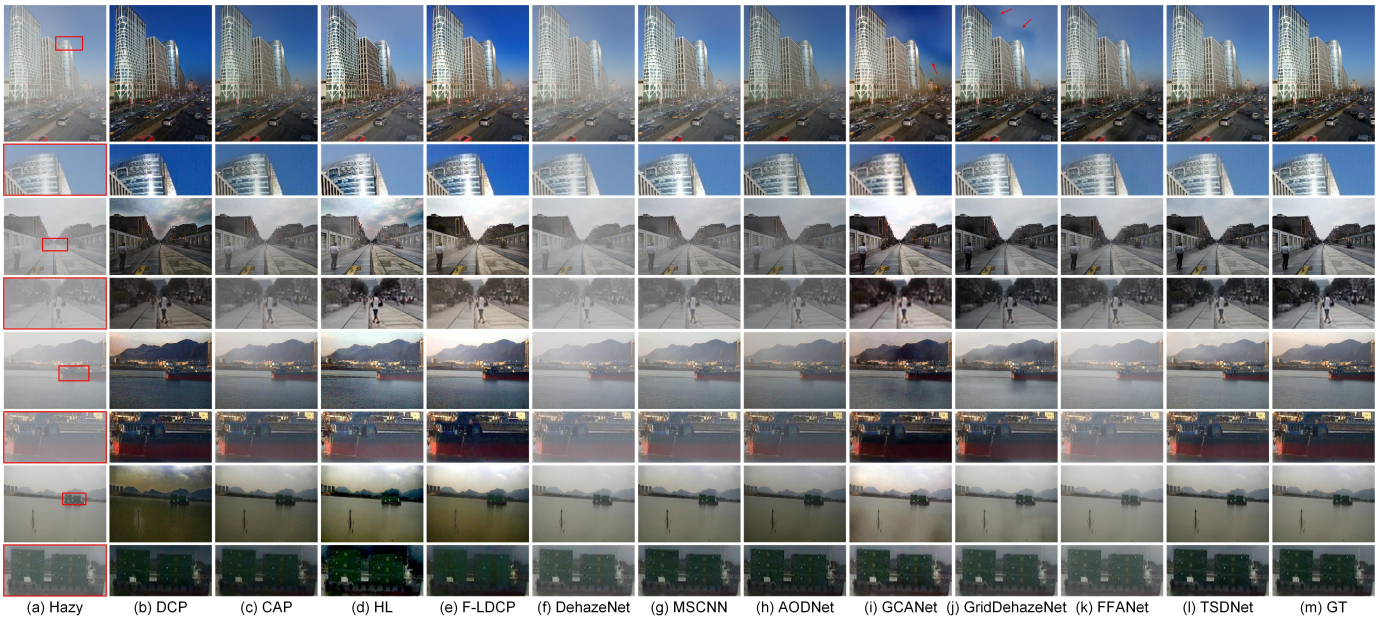


Fig. 3. Visual comparisons of various methods on two RESIDE-OTS and two Seaships images from Test-S. Our TSDNet can obtain clearer and more colorful results than other state-of-the-art dehazing methods.

i.e., peak signal-to-noise ratio (PSNR), structural similarity (SSIM), feature similarity (FSIM), and learned perceptual image patch similarity (LPIPS), are simultaneously employed to quantitatively compare the performance of various methods.

For the sake of better visual comparison, we show four dehazing cases to verify the superiority of our method, shown in Fig. 3. It can be observed that the enhanced results yielded by DCP [7] have serious distortion, which causes the image brightness to be lower. In contrast, the unremoved haze still remains in the restored versions of CAP [12] and HL [32]. Although F-LDCP [9] has the strong dehazing ability, the color distortion in the local region (e.g., water and sky) will affect the visual effect. The failure of these prior-based methods is that the outdoor condition is complex and changeable, leading to the inapplicability of prior knowledge. In the learning-based methods, DehazeNet [14], MSCNN [15], AODNet [17], and FFANet [20] fail to remove haze in the road and maritime scenes adequately. On the contrary, the excessive removal of haze by GCANet [18] and GridDehazeNet [19] results in abnormalities in some homogeneous regions, such as artifacts in the sky area shown in Fig. 3 (i) and (j). By visual comparison, our TSDNet can fully avoid the problems existing in other methods and has the best scores on PSNR, SSIM, FSIM, and LPIPS, shown in Table II.

D. Visibility Enhancement Results on Real-World Images

To verify that the proposed method can remove the haze in the real-world scene, we further conduct a dehazing experiment on 45 real-world hazy images from Test-R. Meanwhile, three no-reference indexes, i.e., naturalness image quality evaluator (NIQE), perception-based image quality evaluator (PIQE), and blind/referenceless image spatial quality evaluator (BRISQUE), are used for comparing the imaging performance of all methods.

Fig. 4 displays three real-world dehazing cases. Obviously, traditional prior-based methods [7], [9], [12], [32] unable to produce clear dehazing results and may further cause image degradation. Specifically, F-LDCP [9] not only fails to remove

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3170594, IEEE Transactions on Industrial Informatics

Fig. 4. Visual comparisons of various methods on three RESIDE-RTTS real-world images from Test-R. The proposed method can remove haze more fully on the premise of ensuring the natural scene.

TABLE III
NIQE, PIQE, AND BRISQUE RESULTS (MEAN±STD) OF VARIOUS METHODS ON THE 45 RESIDE-RTTS IMAGES FROM TEST-R. THE BEST RESULTS ARE IN **BOLD**.

Method	NIQE↓	PIQE↓	BRISQUE↓
Hazy	4.208±1.148	22.01±9.55	0.513±0.127
DCP [7]	3.633±0.658	19.23±8.22	0.542±0.165
CAP [12]	4.026±0.849	21.30±8.47	0.504±0.046
HL [32]	3.786±0.741	19.22±7.39	0.499±0.065
F-LDCP [9]	3.773±0.767	20.10±8.03	0.499±0.100
DehazeNet [14]	4.090±1.069	21.02±8.80	0.529±0.100
MSCNN [15]	3.847±0.867	19.93±8.24	0.504±0.080
AODNet [17]	3.973±0.798	19.68±8.07	0.498±0.055
GCANet [18]	4.070±0.751	22.05±8.74	0.502±0.049
GridDehazeNet [19]	3.852±0.816	18.79±8.01	0.509±0.192
FFANet [20]	3.992±0.892	19.86±8.86	0.538±0.147
TSDNet	**3.607±0.668**	**16.42±7.58**	**0.468±0.188**

TABLE IV
METRIC RESULTS (MAP±STD) OF YOLOV4-TINY/FASTER R-CNN/YOLOV4 ON THE 4952 VOC2007 SYNTHETIC HAZY IMAGES FROM TEST-D AND THE ENHANCED IMAGES YIELDED BY ALL DEHAZING METHODS (UNIT: %). THE BEST RESULTS ARE IN **BOLD**, AND THE SECOND-BEST ARE WITH UNDERLINE.

Method	YOLOv4-Tiny [34]	Faster R-CNN [35]	YOLOv4 [36]
Hazy	51.55±10.6	79.01±8.32	89.13±4.87
DCP [7]	53.92±8.70	80.92±7.74	91.16±3.41
CAP [12]	53.64±8.33	80.38±7.82	90.69±3.99
HL [32]	51.23±10.3	80.07±8.17	90.28±3.77
F-LDCP [9]	54.14±8.75	81.16±7.46	90.61±3.80
DehazeNet [14]	51.99±10.1	79.71±8.79	89.65±4.63
MSCNN [15]	53.37±9.48	80.96±8.04	90.67±4.04
AODNet [17]	52.41±8.92	80.42±7.70	90.71±4.26
GCANet [18]	52.97±9.19	80.04±7.97	90.68±3.93
GridDehazeNet [19]	54.09±9.10	80.96±8.04	91.33±3.92
FFANet [20]	54.28±8.91	81.51±8.37	91.01±4.21
TSDNet	54.83±8.61	81.57±7.98	91.40±3.45
GT	**55.78±8.31**	**82.03±7.58**	**91.74±3.65**

the haze but also produces serious color abnormalities. Since the transmission is overestimated, the illumination of scene recovered by DCP [7], CAP [12] and HL [32] becomes darker. For the learning-based methods, AODNet [17] tends to produce low-illumination results. DehazeNet [14], MSCNN [15], GridDehazeNet [19], and FFANet [20] fail to get the clean images, and retained a certain degree of haze. GCANet [18] can obtain natural visual effects, but the haze often residues in the image domain. By comparison, our TSDNet is more suitable for the dehazing task of outdoor traffic scenarios and tends to generate robust and high-contrast results. Furthermore, Table III provides the comparison of three no-reference evaluators on the Test-R. Our method generates the optimal values in all three indexes, better than the sub-optimal methods by -0.026 NIQE, -2.37 PIQE, and -0.030 BRISQUE. The superior dehazing performance of TSDNet benefits from the attention of hazy distribution and the dual feature extraction from image and frequency domains.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3170594, IEEE Transactions on Industrial Informatics

TABLE V
DETECTION AVERAGE PRECISION (%) OF YOLOV4-TINY/FASTER R-CNN/YOLOV4 ON THE 4320 REAL-WORLD HAZY IMAGES OF RESIDE-RTTS FROM TEST-D AND THE ENHANCED IMAGES BY ALL DEHAZING METHODS. THE BEST RESULTS ARE IN **BOLD**.

Method	Bicycle	Bus	Car	Motorbike	Person	mAP ↑
Hazy	52.01/43.86/53.58	28.97/21.53/34.83	66.53/44.91/72.12	43.04/33.49/55.54	72.14/70.49/78.70	52.54/42.86/58.96
DCP [7]	49.42/44.80/53.58	28.85/24.43/38.38	69.32/47.63/72.34	40.83/37.39/55.09	74.12/72.59/78.84	52.51/45.37/59.65
CAP [12]	39.62/38.75/47.16	27.00/21.83/34.28	65.77/43.47/69.62	37.95/31.80/48.83	70.21/68.84/75.27	48.11/40.94/55.03
HL [32]	47.32/45.89/50.51	24.15/22.97/35.44	67.64/**49.47**/71.71	39.49/34.71/52.51	70.27/69.54/76.98	49.77/44.52/57.43
F-LDCP [9]	52.01/43.86/53.58	28.98/21.53/34.83	66.53/44.90/72.12	43.04/33.49/54.54	72.14/70.49/78.70	52.54/42.86/58.96
DehazeNet [14]	52.27/45.75/54.40	29.51/22.31/35.17	67.65/46.29/72.48	42.61/35.07/54.70	72.36/70.66/78.71	52.88/44.02/59.09
MSCNN [15]	53.95/46.81/55.55	31.77/24.51/38.30	69.42/48.66/73.44	46.34/37.86/56.05	73.96/72.43/79.10	53.95/46.05/60.49
AODNet [17]	53.63/46.99/56.17	27.65/22.61/35.61	67.17/45.72/72.28	42.22/34.76/53.55	72.93/72.08/78.82	52.72/44.43/59.29
GCANet [18]	52.83/47.33/53.83	31.54/24.60/36.72	69.08/47.59/72.15	46.52/37.50/54.69	73.07/71.53/78.13	54.61/45.71/59.10
GridDehazeNet [19]	53.92/47.85/55.64	31.14/24.33/36.33	68.56/47.78/73.20	47.02/38.52/**57.70**	74.07/72.41/79.30	54.94/46.18/60.43
FFANet [20]	52.59/45.86/54.87	29.66/22.62/35.23	67.59/46.36/72.79	44.74/35.86/56.98	72.85/71.43/78.93	53.49/44.42/59.76
TSDNet	**55.22/49.29/56.24**	**31.94/25.53/38.55**	**69.75**/49.12/**73.46**	**47.42/39.28**/56.65	**74.33/72.75/79.58**	**55.73/47.20/60.90**

TABLE VI
DETAILED CONFIGURATION OF ALL DEVICES USED IN RUNNING TIME ANALYSIS.

Number	Type	CPU	RAM	GPU	GDDR
Dev.1	Laptop	Intel Core i7-9750H	8GB	GTX 1050	3GB
Dev.2	Laptop	AMD Ryzen 7 5800H	16GB	GTX 3060	6GB
Dev.3	Desktop	Intel Core i5-8600	8GB	GTX 1060	6GB
Dev.4	Desktop	Intel Core i5-10600KF	16GB	RTX 2080 Ti	11GB
Dev.5	Desktop	Intel Core i9-10850K	32GB	RTX 2080 Ti	11GB

E. Object Detection after Visibility Enhancement

High-level vision tasks, e.g., object detection and recognition, have received continuous attention. However, these methods trained on haze-free images may have the poor precision under hazy conditions, which affect practical applications. Therefore, we study the reliability of object detection in the presence of haze and verify that our TSDNet can promote the detection performance. In particular, we conduct an experiment on 4952 VOC2007 synthetic hazy images and 4320 RESIDE-RTTS real-world hazy images. YOLOv4-Tiny [34], Faster R-CNN [35], and YOLOv4 [36] are simultaneously used to detect the object in the hazy images and the images restored by all dehazing methods.

Table IV displays the mean average precision (mAP) and standard deviation of all methods on the VOC2007 synthetic hazy images from Test-D. By comparing the detection results of hazy images and ground truth (GT), the synthetic fog will reduce the mAP of three object detection methods with different extents. The detection failure on the hazy image lies in that the fog buries the critical information, making the object invisible. Theoretically, the detection accuracy can be improved by enhancing the visibility of hazy images. Therefore, object detection after dehazing can achieve better precision than hazy images detection in most cases. In particular, our TSDNet can obtain the detection accuracy closest to the GT in each target detection method.

To verify that dehazing methods are also effective in the real-world hazy environment, Table V compares the average detection precision of all methods on the RESIDE-RTTS images from Test-D. Our TSDNet achieves the best performance with 55.73% mAP in YOLOv4-Tiny, 47.20% mAP in Faster R-CNN, and 60.90% mAP in YOLOv4. Meanwhile, our method can generate the best detection average precision in most cases. Although the proposed TSDNet is non-optimal for "Car" detection using Faster R-CNN and "Mobtorbike" detection using YOLOv4, our method gets the second and third places, which falls behind the optimal solution by a slight difference. Furthermore, Fig. 5 provides six real-world hazy scenes to observe the impact of the proposed dehazing method on YOLOv4. Obviously, it is challenging to accurately detect the target in the hazy images by YOLOv4, which often produces the issue of detection failure or misclassification. With the distance between the observer and the target increasing, the degradation phenomenon becomes more serious, which increases the risk of detection failure. Fortunately, our method can significantly improve detection accuracy by highlighting the critical information of the object in the hazy image.

F. Running Time Analysis

It is well known that many devices applied in the distributed intelligence-enabled Internet of Things (IoT) have different computing power. Therefore, we select five types of equipment with different performances for calculating the running time of all methods. The detailed configuration of all devices is displayed in Table VI. As shown in Table VII, we calculate the average running time on three different scale datasets, i.e., 480×640, 640×800, and 1080×720. Each dataset contains 20 images. Undoubtedly, four comparable algorithms implemented by Matlab [7], [9], [12], [32] have slower calculation speeds, which fail to meet the requirement of real-time processing in practical applications. On the contrary, the learning-based methods accelerated by GPU have a great advantage. Unfortunately, many competitive approaches require expensive computational costs due to the huge complexity of network. In deep learning methods, only AODNet [17], GridDehazeNet [19], and our TSDNet can satisfy the processing speed over than 30 frames per second in all datasets with different scales and devices with different performance. In contrast, our method has faster inference speed and better enhancement effects under the hazy environment. The robustness and efficiency of our method benefits from the novel strategy, which considers the challenging dehazing task as three simple works, i.e., haze distribution attention, feature extraction, and information fusion. Each part thus only needs fewer parameters and calculations to generate the efficient dehazing performance.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3170594, IEEE Transactions on Industrial Informatics

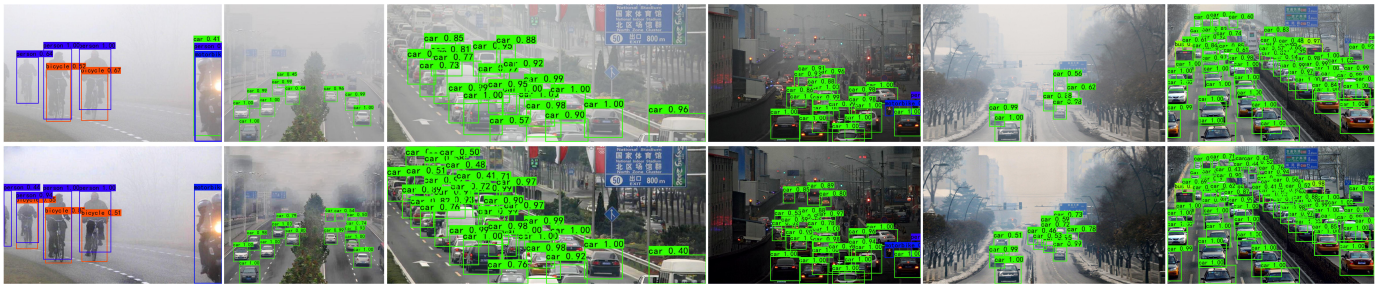


Fig. 5. The detection results on (Top) RESIDE-RTTS real-world hazy images from Test-D and (Bottom) restored images yielded by our TSDNet. It is obvious that our TSDNet can not only reduce haze but also improve the accuracy and robustness of YOLOv4-based object detection.

TABLE VII
AVERAGE RUNNING TIME COMPARISON OF THREE DATASETS ($480 \times 640/640 \times 800/1080 \times 720$) ON FIVE DEVICES (UNIT: SECOND). THE FASTEST TIMES ARE IN **BOLD**.

Method	Language	Dev. 1	Dev. 2	Dev. 3	Dev. 4	Dev. 5
DCP [7]	Matlab (CPU)	1.2245/2.0567/3.0913	1.0354/1.7270/2.8429	1.0387/1.8224/2.7664	1.0641/1.7615/2.6398	0.9564/1.8347/2.7821
CAP [12]		1.0913/1.7804/2.7294	1.4375/1.8495/2.6752	1.0658/1.5789/2.2944	1.0543/1.4984/2.3438	0.8980/1.4901/2.2344
F-LDCP [9]		1.4564/2.3750/3.4490	1.4844/2.0617/3.3553	1.3766/2.1168/3.6571	1.2738/2.0313/2.9424	1.3224/2.0148/3.2681
HL [32]	Matlab (GPU)	5.3479/6.0275/7.0730	5.1868/5.8526/6.6321	5.2668/5.9034/6.6914	5.1599/5.8031/6.6118	5.1034/5.4900/6.6217
DehazeNet [14]		0.6934/1.1839/1.6592	0.6765/1.0387/1.5023	0.6868/1.0423/1.5568	0.6728/1.0299/1.4935	0.6680/0.9936/1.4289
MSCNN [15]		0.2911/0.5082/0.9572	0.2875/0.4569/0.8975	0.2875/0.4601/0.9012	0.2795/0.4459/0.8918	0.2391/0.4109/0.8701
AODNet [17]	Python (GPU)	0.0345/0.0517/0.0779	0.0140/0.0209/0.0186	0.0287/0.0466/0.0574	0.0126/0.0196/0.0287	0.0122/0.0195/0.0280
GCANet [18]		0.2950/0.4872/0.7540	0.1055/0.1395/0.1776	0.2937/0.4569/1.7016	0.0866/0.1355/0.1994	0.0798/0.1255/0.1858
GridDehazeNet [19]		0.0269/0.0286/0.0316	0.0171/0.0195/0.0162	0.0183/0.0194/0.0327	0.0213/0.0268/0.0270	0.0208/0.0252/0.0266
FFANet [20]		3.7178/6.2597/9.4397	0.7951/1.3194/1.8980	3.1832/4.8600/6.4772	0.4080/0.6731/1.1272	0.3981/0.6655/1.1228
TSDNet		**0.0143/0.0193/0.0225**	**0.0099/0.0112/0.0132**	**0.0124/0.0149/0.0171**	**0.0111/0.0128/0.0148**	**0.0102/0.0121/0.0166**

G. Ablation Study

To verify the effectiveness of the proposed strategies, we implement an ablation study in this work. In particular, we select 200 RESIDE-OTS and 200 Seaships images from Train mentioned in Table. I to train all dehazing models. Other training parameters are consistent with Section IV-B. As shown in Table VIII, we compare the PSNR and SSIM results of seven different models on 70 images from Test-S. The reconstructed images generally have lower PSNR and SSIM values when the single sub-network (e.g., TBEb, TBEc, and MFF) is directly adopted for dehazing. Compared with Model.1, Model.4 and Model.5 combine TBEb and TBEc with MFF, respectively, which can improve the dehazing performance with different degrees. Obviously, Model.6 achieves better results. The superior performance of Model.6 benefits from the dual feature extraction and fusion strategies in image domain and frequency domain. Therefore, Model.6 can fully extract the hazy features and reconstruct sharp images. Finally, Model.7 (i.e., TSDNet) further introduces a multi-scale attention module to achieve optimal dehazing results.

TABLE VIII
PSNR AND SSIM RESULTS OF SEVEN DIFFERENT MODELS ON TEST-S. NOTE THAT MSA, TBEB, TBEC, AND MFF REPRESENT THE MULTI-SCALE ATTENTION MODULE, ENCODER-DECODER NETWORK, DENSE NETWORK, AND MULTI-FEATURE FUSION MODULE SHOWN IN FIG. 2, RESPECTIVELY.

Number	MSA	TBEb	TBEc	MFF	PSNR	SSIM
Model.1		✔			20.94±2.98	0.932±0.029
Model.2			✔		17.40±3.19	0.859±0.072
Model.3				✔	19.33±2.91	0.878±0.060
Model.4		✔		✔	21.92±3.36	0.939±0.027
Model.5			✔	✔	19.89±3.06	0.887±0.061
Model.6		✔	✔	✔	23.03±3.07	0.942±0.031
Model.7	✔	✔	✔	✔	23.07±3.22	0.944±0.030

V. CONCLUSIONS AND FUTURE PERSPECTIVES

In this paper, we designed a three-stage dehazing network (named TSDNet) for promoting visual IoT-enabled ITS. To be specific, the proposed TSDNet mainly contains three parts, i.e., multi-scale attention, two-branch extraction, and multi-feature fusion modules. The multi-scale attention module consists of a dilated convolution-based ASPP module and several convolutional layers to obtain hazy distribution. The two-branch extraction module exploits an encoder-decoder network and a dense network to extract features separately. In particular, the dense network adopts the frequency domain supervision to obtain hazy characteristics different from the encoder-decoder network. Finally, the outputs of the attention and extraction modules are simultaneously fed into the multi-feature fusion module to reconstruct the latent haze-free image. Comprehensive experiments on synthetic and real-world scenarios have demonstrated the superior performance of TSDNet in terms of imaging reliability and robustness under different hazy imaging conditions. In addition, our TSDNet is able to significantly improve the robustness and accuracy of object detection for YOLOv4-Tiny, Faster R-CNN, and YOLOv4 under hazy weather. The imaging results meet the real-time requirement of detection applications in visual IoT-driven ITS. To make our work more reliable and applicable, the study shown in this work can be extended with the following directions.

- To further enhance the efficacy and robustness of our TSDNet, the weights $\gamma_{1\leq k\leq 6}$ of different loss functions in Section III-D should be adaptively adjusted for different imaging scenarios. Traditional parameter-selection methods, commonly performed using sufficient experimental experiences, can generate satisfactory imaging results, but are rather complex and time-consuming. To replace the constant weights with the adaptive ones, the homoscedastic uncertainty weighting strategy [37] could be considered to automatically select the optimal weights.
- In practical scenarios, the state-of-the-art visibility enhancement methods still inevitably suffer from the loss of fine details in restored images, leading to the limited improvement in object detection under hazy weather [38]. To further improve the detection accuracy and robustness, there is strong potential for simultaneously restoring the latent haze-free images and detecting the objects of interest. To achieve this challenging goal, we will further consider the influence of other factors (e.g., air pollution and season) on haze generation to produce more realistic hazy dataset. Meanwhile, the multi-task learning paradigm [39], where multiple learning tasks are solved jointly, could be exploited in future work.
- In addition, almost no previous dehazing studies consider the influences of ultraviolet (UV) light on visibility enhancement. The UV light is inevitably absorbed and scattered by fog and haze in long-range outdoor scenes, easily leading to color distortion in blue sky regions. The UV light filter, an additional optical device, is able to reduce the atmospheric haze through absorbing UV light [40]. There is thus a potential solution to combine our TSDNet (i.e., software dehazing) and UV light filter (i.e., hardware dehazing) to further enhance the imaging quality in visual IoT-enabled ITS.

Owing to the advantages of the introduced multi-scale attention, two-branch extraction, and multi-feature fusion modules, our TSDNet could effectively and robustly improve the haze visibility enhancement, leading to the high-quality images and the promoted object detection for visual IoT-driven ITS.

REFERENCES

[1] M. Serror, S. Hack, M. Henze, M. Schuba, and K. Wehrle, "Challenges and opportunities in securing the industrial internet of things," *IEEE Trans. Ind. Inf.*, vol. 17, no. 5, pp. 2985–2996, May 2020.

[2] C. Dai, X. Liu, H. Xu, L. T. Yang, and J. Deen, "Hybrid deep model for human behavior understanding on industrial internet of video things," *IEEE Trans. Ind. Inf.*, to be published, doi:10.1109/TII.2021.3058276.

[3] C. Wang, J. Shen, P. Vijayakumar, and B. B. Gupta, "Attribute-based secure data aggregation for isolated iot-enabled maritime transportation systems," *IEEE Trans. Intell. Transp. Syst.*, to be published, doi:10.1109/TITS.2021.3127436.

[4] Z. Zhou, A. Gaurav, B. B. Gupta, M. D. Lytras, and I. Razzak, "A fine-grained access control and security approach for intelligent vehicular transport in 6g communication system," *IEEE Trans. Intell. Transp. Syst.*, to be published, doi:10.1109/TITS.2021.3106825.

[5] H. Fatemidokht, M. K. Rafsanjani, B. B. Gupta, and C.-H. Hsu, "Efficient and secure routing protocol based on artificial intelligence algorithms with uav-assisted for vehicular ad hoc networks in intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4757–4769, Jul. 2021.

[6] P. Li, Z. Chen, L. T. Yang, Q. Zhang, and M. J. Deen, "Deep convolutional computation model for feature learning on big data in internet of things," *IEEE Trans. Ind. Inf.*, vol. 14, no. 2, pp. 790–798, Feb. 2017.

[7] 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.

[8] 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.

[9] Y. Zhu, G. Tang, X. Zhang, J. Jiang, and Q. Tian, "Haze removal method for natural restoration of images with sky," *Neurocomputing*, vol. 275, pp. 499–510, Jan. 2018.

[10] Q. Shu, C. Wu, Z. Xiao, and R. W. Liu, "Variational regularized transmission refinement for image dehazing," in *Proc. IEEE ICIP*, 2019, pp. 2781–2785.

[11] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graphics*, vol. 34, no. 1, pp. 1–14, Nov. 2014.

[12] 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.

[13] D. Berman, S. Avidan *et al.*, "Non-local image dehazing," in *Proc. IEEE CVPR*, 2016, pp. 1674–1682.

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

[15] W. Ren, J. Pan, H. Zhang, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks with holistic edges," *Int. J. Comput. Vision*, vol. 128, no. 1, pp. 240–259, Sep. 2020.

[16] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE CVPR*, 2018, pp. 3194–3203.

[17] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proc. IEEE ICCV*, 2017, pp. 4770–4778.

[18] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," in *Proc. IEEE WACV*, 2019, pp. 1375–1383.

[19] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," in *Proc. IEEE ICCV*, 2019, pp. 7314–7323.

[20] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 07, 2020, pp. 11 908–11 915.

[21] D. Engin, A. Genç, and H. Kemal Ekenel, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," in *Proc. IEEE CVPRW*, 2018, pp. 825–833.

[22] A. Dudhane and S. Murala, "Cdnet: Single image de-hazing using unpaired adversarial training," in *Proc. IEEE WACV*, 2019, pp. 1147–1155.

[23] A. Dudhane, H. Singh Aulakh, and S. Murala, "Ri-gan: An end-to-end network for single image haze removal," in *Proc. IEEE CVPRW*, 2019, pp. 1–10.

[24] A. Mehta, H. Sinha, P. Narang, and M. Mandal, "Hidegan: A hyperspectral-guided image dehazing gan," in *Proc. IEEE CVPRW*, 2020, pp. 212–213.

[25] H. Liu, B. Jiang, Y. Song, W. Huang, and C. Yang, "Rethinking image inpainting via a mutual encoder-decoder with feature equalizations," in *Proc. ECCV*, 2020, pp. 725–741.

[26] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imaging*, vol. 3, no. 1, pp. 47–57, Mar. 2016.

[27] R. Girshick, "Fast r-cnn," in *Proc. IEEE ICCV*, 2015, pp. 1440–1448.

[28] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE ACSSC*, vol. 2, 2003, pp. 1398–1402.

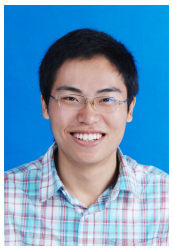
[29] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2018.

[30] Z. Shao, W. Wu, Z. Wang, W. Du, and C. Li, "Seaships: A large-scale precisely annotated dataset for ship detection," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2593–2604, Oct. 2018.

[31] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *Int. J. Comput. Vision*, vol. 88, no. 2, pp. 303–338, Sep. 2010.

[32] D. Berman, T. Treibitz, and S. Avidan, "Single image dehazing using haze-lines," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 3, pp. 720–734, Mar. 2018.

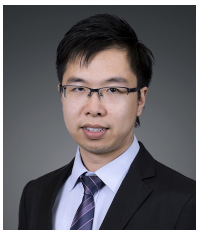
[33] Z. Li and N. Snavely, "Megadepth: Learning single-view depth prediction from internet photos," in *Proc. IEEE CVPR*, 2018, pp. 2041–2050.
[34] Z. Jiang, L. Zhao, S. Li, and Y. Jia, "Real-time object detection method based on improved yolov4-tiny," *arXiv preprint arXiv:2011.04244*, 2020.
[35] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Proc. NIPS*, 2015, pp. 91–99.
[36] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.
[37] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proc. IEEE CVPR*, 2018, pp. 7482–7491.
[38] R. W. Liu, W. Yuan, X. Chen, and Y. Lu, "An enhanced cnn-enabled learning method for promoting ship detection in maritime surveillance system," *Ocean Eng.*, vol. 235, p. 109435, Sep. 2021.
[39] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Trans. Knowl. Data Eng.*, to be published, doi:10.1109/TKDE.2021.3070203.
[40] C. Xu, H. Zhang, and J. Cheng, "Effects of haze particles and fog droplets on nlos ultraviolet communication channels," *Opt. Express*, vol. 23, no. 18, pp. 23 259–23 269, Aug. 2015.



Ryan Wen Liu (M'15) received the B.Sc. degree (Hons.) in information and computing science from the Department of Mathematics, Wuhan University of Technology, Wuhan, China, in 2009, and the Ph.D. degree in imaging informatics from The Chinese University of Hong Kong, Hong Kong, in 2015. He is currently an Associate Professor with the School of Navigation, Wuhan University of Technology. He was a Visiting Scholar with the Agency for Science, Technology and Research, Singapore. His research interests include computer vision, data mining, and intelligent transportation system.

Yu Guo received the B.Sc. degree in naval architecture and marine engineering from the School of Transportation, Wuhan University of Technology, Wuhan, China, in 2021. He is currently pursuing the Ph.D. degree in traffic information engineering and control with the School of Navigation, Wuhan University of Technology, Wuhan, China. He has received the Best Paper Award from AAAI-22 Workshop: AI for Transportation. His research interests include computer vision, machine learning, and intelligent navigation systems.

Yuxu Lu received the B.Sc. degree in navigation technology from the School of Navigation, Wuhan University of Technology, Wuhan, China, in 2020, where he will pursue the M.S. degree in traffic information engineering and control with the School of Navigation, Wuhan University of Technology, Wuhan, China. He has received the Excellent Oral Presentation from The 11th International Conference on Machine Learning and Computing (ICMLC 2019). His research interests include computer vision, machine learning, and intelligent navigation systems.



Kwok Tai Chui (S'14-M'17) received the B.Eng. degree in electronic and communication engineering - Business Intelligence Minor and Ph.D. degree in electronic engineering from City University of Hong Kong. He had industry experience as Senior Data Scientist in Internet of Things (IoT) company. He is with the Department of Technology, School of Science and Technology, at Hong Kong Metropolitan University as Assistant Professor.

He has more than 90 research publications including edited books, book chapters, journal papers, and conference papers. He has served as various editorial position in ESCI/SCIE-listed journals including Managing Editor of International Journal on Semantic Web and Information Systems, Topic Editor of Sensors, Associate Editor of International Journal of Energy Optimization and Engi-neering. His research interests include computational intelligence, data science, energy monitoring and management, intelligent trans-portation, smart metering, healthcare, machine learning algorithms and optimization.

Brij B. Gupta (SM'17) received the PhD degree in information and cyber security from Indian Institute of Technology (IIT) Roorkee, India. In more than 16 years of his professional experience, he published over 400 papers in journals/conferences including 30 books and 08 Patents with over 14000 citations. He has received numerous national and international awards including Canadian Commonwealth Scholarship (2009), Faculty Research Fellowship Award (2017), from the Govt. of Canada, MeitY, GoI, IEEE GCCE outstanding and WIE paper awards and Best Faculty Award (2018 & 2019), NIT Kurukshetra, respectively. Prof Gupta is also serving as Distinguished Research Scientist with LoginRadius Inc., USA which is one of leading cybersecurity companies in the world, especially in the field of customer identity and access management (CIAM). He is also selected in the 2021 and 2020 Stanford University's ranking of the world's top 2% scientists. He is also a visiting/adjunct professor with several universities worldwide. He is also an IEEE Senior Member (2017) and also selected as 2021 Distinguished Lecturer in IEEE CTSoc. Dr Gupta is also serving as Member-in-Large, Board of Governors, IEEE Consumer Technology Society (2022-204). Prof. Gupta is also leading IJSWIS, IJSSCI and IJCAC, IGI Global, as Editor-in-Chief. Moreover, he is also serving as lead-editor of a Book Series with CRC, World Scientific and IET press. He also served as TPC members and organized/special session chairs in ICCE-2021, GCCE 2014-2021 and TPC Chair in 2018 INFOCOM: CCSNA Workshop and Publicity Co-chair in 2020 ICCCN. Dr Gupta is also serving/served as Associate/Guest Editor of IEEE TII, IEEE TITS, IoT, IEEE Big Data, ASOC, FGCS, etc. At present, Prof. Gupta is working as Director, International Center for AI and Cyber Security Research and Innovations, and Full Professor with the Department of Computer Science and Information Engineering (CSIE), Asia University, Taiwan. His research interests include information security, Cyber physical systems, cloud computing, blockchain technologies, intruion detection, AI, social media and networking.