Benchmarking Single-Image Dehazing and Beyond

Boyi Li, Wenqi Ren, *Member, IEEE*, Dengpan Fu, Dacheng Tao, *Fellow, IEEE*, Dan Feng, *Associate Member*, Wenjun Zeng, *Fellow, IEEE*, and Zhangyang Wang, *Member, IEEE*

***Abstract*— We present a comprehensive study and evaluation of existing single-image dehazing algorithms, using a new large-scale benchmark consisting of both synthetic and real-world hazy images, called REalistic Single-Image DEhazing (RESIDE). RESIDE highlights diverse data sources and image contents, and is divided into five subsets, each serving different training or evaluation purposes. We further provide a rich variety of criteria for dehazing algorithm evaluation, ranging from full-reference metrics to no-reference metrics and to subjective evaluation, and the novel task-driven evaluation. Experiments on RESIDE shed light on the comparisons and limitations of the state-of-the-art dehazing algorithms, and suggest promising future directions.**

***Index Terms*— Dehazing, detection, dataset, evaluations.**

I. Introduction

A. Problem Description: Single Image Dehazing

IMAGES captured in outdoor scenes often suffer from poor visibility, reduced contrasts, fainted surfaces and color shift, due to the presence of haze. Caused by aerosols such as dust, mist, and fumes, the existence of haze adds complicated, nonlinear and data-dependent noise to the images, making the haze removal (a.k.a. *dehazing*) a highly challenging image restoration and enhancement problem. Moreover, many computer vision algorithms can only work well with the scene radiance that is haze-free. However, a dependable vision system must reckon with the entire spectrum of degradations from unconstrained environments. Taking autonomous driving for example, hazy and foggy weather will obscure the vision of on-board cameras and create confusing reflections and glare, leaving state-of-the-art self-driving cars in struggle [1]. Dehazing is thus becoming an increasingly desirable technique for both computational photography and computer vision tasks, whose advance will immediately benefit many blooming application fields, such as video surveillance and autonomous/assisted driving [2].

While some earlier works consider multiple images from the same scene to be available for dehazing [3]–[6], the *single image dehazing* proves to be a more realistic setting in practice, and thus gained the dominant popularity. The *atmospheric scattering model* has been the classical description for the hazy image generation [7]–[9]:

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

where $I(x)$ is observed hazy image, $J(x)$ is the haze-free scene radiance to be recovered. There are two critical parameters: A denotes the global atmospheric light, and $t(x)$ is the transmission matrix defined as:

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

where β is the scattering coefficient of the atmosphere, and $d(x)$ is the distance between the object and the camera.

We can re-write the model (1) for the clean image as the output:

$$J(x) = \frac{1}{t(x)} I(x) - A\frac{1}{t(x)} + A. \tag{3}$$

Most state-of-the-art single image dehazing methods exploit the physical model (1), and estimate the key parameters A and $t(x)$ in either physically grounded or data-driven ways. The performance of top methods have continuously improved [10]–[17], especially after the latest models embracing deep learning [18]–[20].

B. Existing Methodology: An Overview

Given the atmospheric scattering model, most dehazing methods follow a similar three-step methodology: (1) estimating the transmission matrix $t(x)$ from the hazy image $I(x)$; (2) estimating A using some other (often empirical) methods; (3) estimating the clean image $J(x)$ via computing (3).

Manuscript received April 6, 2018; revised July 21, 2018 and August 19, 2018; accepted August 22, 2018. Date of publication August 30, 2018; date of current version September 25, 2018. The work of W. Ren was supported in part by the National Natural Science Foundation of China under Grant 61802403, in part by the Open Projects Program of the National Laboratory of Pattern Recognition, and in part by the CCF-Tencent Open Fund. The work of D. Tao was supported by the Australian Research Council Projects under Grant FL-170100117 and Grant DP-180103424. The work of D. Feng was supported in part by the National Natural Science Foundation of China under Grant U1705261 and Grant 61772222, and in part by the Engineering Research Center of Data Storage Systems and Technology, Ministry of Education, China. The work of Z. Wang was supported by the National Science Foundation under Grant 1755701. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jie Liang. *(Boyi Li, Wenqi Ren, and Dengpan Fu contributed equally to this work.) (Corresponding author: Zhangyang Wang.)*

B. Li is with the Computer Science Department, Cornell University, Ithaca, NY 14850 USA (e-mail: boyilics@gmail.com).

W. Ren is with the State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China (e-mail: rwq.renwenqi@gmail.com).

D. Fu is with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230000, China (e-mail: fdpan@mail.ustc.edu.cn).

D. Tao is with the UBTECH Sydney Artificial Intelligence Centre and the School of Information Technologies, Faculty of Engineering and Information Technologies, The University of Sydney, Darlington, NSW 2008, Australia (e-mail: dacheng.tao@sydney.edu.au).

D. Feng is with the Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: dfeng@hust.edu.cn).

W. Zeng is with Microsoft Research, Beijing 100080, China (e-mail: wezeng@microsoft.com).

Z. Wang is with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843-3126 USA (e-mail: atlaswang@tamu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2018.2867951

Usually, the majority of attention is paid to the first step, which can rely on either physically grounded priors or fully data-driven approaches.

A noteworthy portion of dehazing methods exploited natural image priors and depth statistics. Reference [21] imposed locally constant constraints of albedo values together with decorrelation of the transmission in local areas, and then estimated the depth value using the albedo estimates and the original image. It did not constrain the scene's depth structure, thus often leads to the inaccurate estimation of color or depth. References [22] and [23] discovered the dark channel prior (DCP) to more reliably calculate the transmission matrix, followed by many successors. However, the prior is found to be unreliable when the scene objects are similar to the atmospheric light [19]. Reference [12] enforced the boundary constraint and contextual regularization for sharper restorations. Reference [14] developed a color attenuation prior and created a linear model of scene depth for the hazy image, and then learned the model parameters in a supervised way. Reference [24] jointly estimated scene depth and recover the clear latent image from a foggy video sequence. Reference [15] proposed a non-local prior, based on the assumption that each color cluster in the clear image became a haze-line in RGB space.

In view of the prevailing success of Convolutional Neural Networks (CNNs) in computer vision tasks, several dehazing algorithms have relied on various CNNs to directly learn $t(x)$ fully from data, in order to avoid the often inaccurate estimation of physical parameters from a single image. DehazeNet [18] proposed a trainable model to estimate the transmission matrix from a hazy image. Reference [19] came up with a multi-scale CNN (MSCNN), that first generated a coarse-scale transmission matrix and gradually refined it. Despite their promising results, *the inherent limitation of training data is becoming a increasingly severe obstacle for this booming trend*: see Section II-1 for more discussions.

Besides, a few efforts have been made beyond the sub-optimal procedure of separately estimating parameters, which will cause accumulated or even amplified errors, when combining them together to calculate (3). They instead advocate simultaneous and unified parameter estimation. Earlier works [25], [26] modeled the hazy image with a factorial Markov random field, where $t(x)$ and A were two statistically independent latent layers. In addition, some researchers also examined the more challenging night-time dehazing problem [27], [28], which falls beyond the focus of this paper.

Another line of researches [29], [30] tries to make use of Retinex theory to approximate the spectral properties of object surfaces by the ratio of the reflected light. Very recently, [20] presented a re-formulation of (2) to integrate $t(x)$ and A into one new variable. As a result. their CNN dehazing model was fully end-to-end: $J(x)$ was directly generated from $I(x)$, without any intermediate parameter estimation step. The idea was later extended to video dehazing in [31].

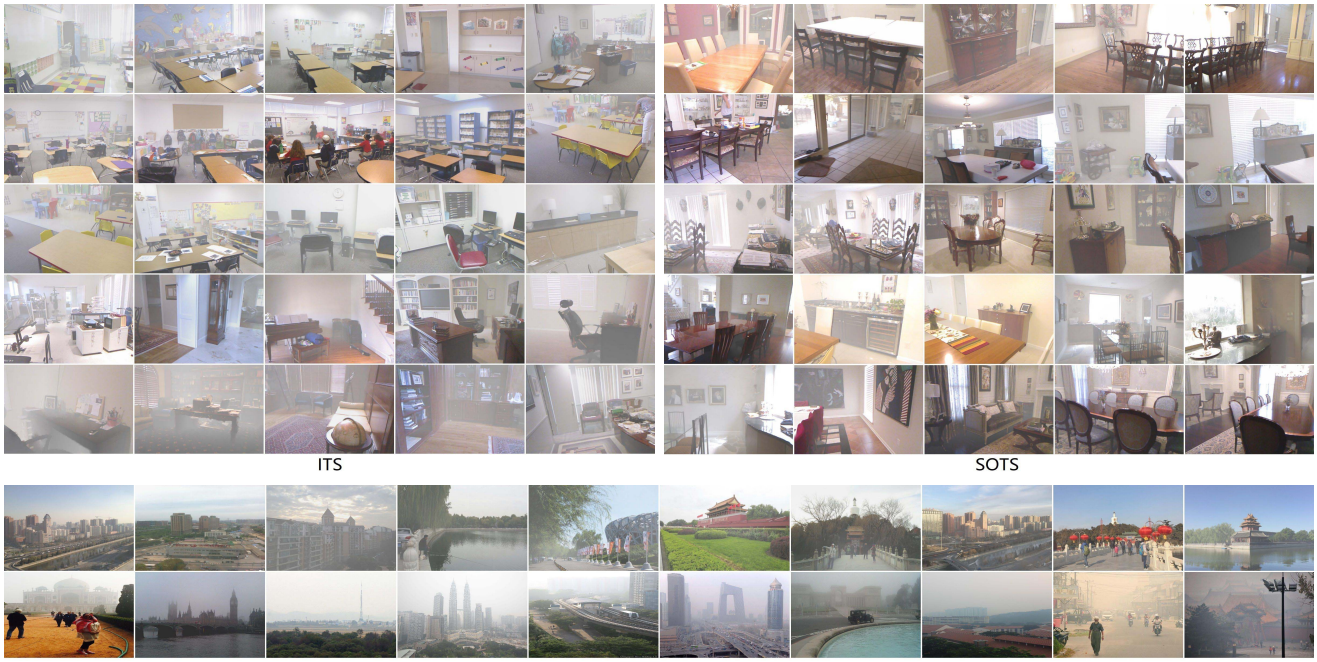
C. Our Contribution

Despite the prosperity of single image dehazing algorithms, there have been several hurdles to the further development of this field. There is a lack of benchmarking efforts on state-of-the-art algorithms on a large-scale public dataset. Moreover, current metrics for evaluating and comparing image dehazing algorithms are mostly just PSNR and SSIM, which turn out to be insufficient for characterizing either human perception quality or machine vision effectiveness.

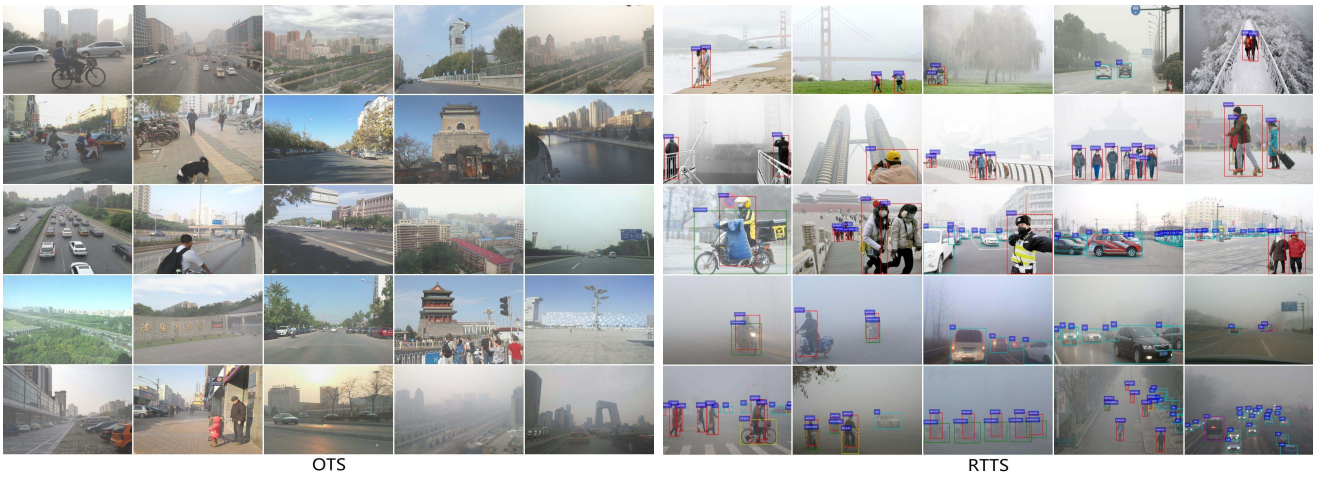
This paper is directly motivated to overcome the above hurdles, and makes three-fold technical contributions:

- We introduce a new single image dehazing benchmark, called the *Realistic Single Image Dehazing* (**RESIDE**) dataset. It features a large-scale synthetic training set, and two different sets designed for objective and subjective quality evaluations, respectively. We further introduce the RESIDE-β set, an exploratory and supplementary part of the RESIDE benchmark, including two innovative discussions on the current hurdles on training data content (indoor versus outdoor images) and evaluation criteria (from either human vision or machine vision perspective), respectively. Particularly in the latter part, we annotate a task-driven evaluation set of 4,322 real-world hazy images with object bounding boxes, which is first-of-its-kind contribution.
- We bring in an innovative set of evaluation strategies in accordance with the new RESIDE and RESIDE-β datasets. In RESIDE, besides the widely adopted PSNR and SSIM, we further employ both no-reference metrics and human subjective scores to evaluate the dehazing results, especially for real-world hazy images without clean ground truth. In RESIDE-β, we recognize that image dehazing in practice usually serves as the pre-processing step for mid-level and high-level vision tasks. We thus propose to exploit the perceptual loss [32] as a "full-reference" task-driven metric that captures more high-level semantics, and the object detection performance on the dehazed images as a "no-reference" task-specific evaluation criterion for dehazing realistic images [20].
- We conduct an extensive and systematic range of experiments to quantitatively compare nine state-of-the-art single image dehazing algorithms, using the new RESIDE and RESIDE-β datasets and the proposed variety of evaluation criteria. Our evaluation and analysis demonstrate the performance and limitations of state-of-the-art algorithms, and bring in rich insights. The findings from these experiments not only confirm what is commonly believed, but also suggest new research directions in single image dehazing.

An overview of RESIDE could be found in Table I. We note that some of the strategies used in this paper have been previously used in the literature to a greater or smaller extent, such as no-reference metrics in dehazing [33], subjective evaluation [34], and connecting dehazing to high-level tasks [20]. However, RESIDE is so far the first and only systematic evaluation, that includes a number of dehazing algorithms with multiple criteria on a common large-scale benchmark, which has long been missing from the literature.



(a)



(b)

Fig. 1. Example images from the five sets in RESIDE and RESIDE-β (see Table I. (a) RESIDE. (b) RESIDE-β.

TABLE I
STRUCTURE OF RESIDE(STANDARD) AND RESIDE-β

RESIDE(Standard)				
Subset	*Number of Images*	*real/synthetic*	*indoor/outdoor*	*annotations*
Indoor Training Set (ITS)	13,990	synthetic	indoor	No
Synthetic Objective Testing Set (SOTS)	500	synthetic	indoor	No
Hybrid Subjective Testing Set (HSTS)	20	real	outdoor	No
RESIDE-β				
Subset	*Number of Images*	*real/synthetic*	*indoor/outdoor*	*annotations*
Outdoor Training Set (OTS)	72,135	synthetic	outdoor	No
Real-world Task-driven Testing Set (RTTS)	4,322	real	outdoor	Yes

The RESIDE dataset is made publicly available for research purposes,[1] and we plan to periodically update our own benchmarking results for noticeable new dehazing algorithms. We also welcome authors to report new results on RESIDE, and to contact us to add their references on the website.

[1]Website: https://sites.google.com/site/boyilics/website-builder/reside

II. Dataset and Evaluation: Status Quo

1) Training Data: Many image restoration and enhancement tasks benefit from the continuous efforts for standardized benchmarks to allow for comparison of different proposed methods under the same conditions, such as [35] and [36]. In comparison, a common large-scale benchmark has been long missing for dehazing, owing to the significant challenge in collecting or creating realistic hazy images with clean ground truth references. It is generally impossible to capture the same visual scene with and without haze, while all other environment conditions stay identical. Therefore, recent dehazing models [34], [37] typically generate their training sets by creating synthetic hazy images from clean ones: they first obtain depth maps of the clean images, by either utilizing available depth maps for depth image datasets, or estimating the depth [38]; and then generate the hazy images by computing (1). Data-driven dehazing models could then be trained to regress clean images from hazy ones.

Fattal's dataset [37] provided 12 synthetic images. FRIDA [39] produced a set of 420 synthetic images, for evaluating the performance of automatic driving systems in various hazy environments. Both of them are too small to train effective dehazing models. To form large-scale training sets, [19], [20] used the ground-truth images with depth metadata from the indoor NYU2 Depth Database [40] and the Middlebury stereo database [41]. Recently, [34] generated Foggy Cityscapes dataset [42] with 25,000 images from the Cityscapes dataset, using incomplete depth information.

2) Testing Data and Evaluation Criteria: The testing sets in use are mostly synthetic hazy images with known ground truth too, although some algorithms were also visually evaluated on real hazy images [18]–[20].

With multiple dehazing algorithms available, it becomes pivotal to find appropriate evaluation criteria to compare their dehazing results. Most dehazing algorithms rely on the full-reference PSNR and SSIM metrics, with assuming a synthetic testing set with known clean ground truth too. As discussed above, their practical applicability may be in jeopardy even a promising testing performance is achieved, due to the large content divergence between synthetic and real hazy images. To objectively evaluate dehazing algorithms on real hazy images without reference, no-reference image quality assessment (IQA) models [43]–[45] are possible candidates. Reference [33] tested a few no-reference objective IQA models among several dehazing approaches on a self-collected set of 25 hazy images (with no clean ground truth), but did not compare any latest CNN-based dehazing models. A recent work [46] collected 14 haze-free images of real outdoor scene and corresponding depth maps, providing a small realistic testing set.

PSNR/SSIM, as well as other objective metrics, often align poorly with human perceived visual qualities [33]. Many papers visually display dehazing results, but the result differences between state-of-the-art dehazing algorithms are often too subtle for people to reliably judge. That suggests the necessity of conducting a subjective user study, towards which few efforts have been made so far [33], [47].

TABLE II
Comparison Between Existing Hazy Datasets and RESIDE

	Synthetic		Real	
	indoor	outdoor	outdoor	annotated
Fattal [37]	4	8	31	-
FIRDA [39]	-	480	-	-
Ma [33]	3	22	-	-
HazeRD [46]	-	14	-	-
Sakaridis [34]	-	25,000	101	101
RESIDE	14,490	72,135	9,129	4,322

All the aforementioned hazy image datasets, as well as RESIDE, are compared in Table II. As shown, most of the existing datasets are either too small in scale, or lack sufficient real-world images (or annotations) for diverse evaluations.

III. A New Large-Scale Dataset: RESIDE

We propose the *REalistic Single Image DEhazing* (**RESIDE**) dataset, a new large-scale dataset for fairly evaluating and comparing single image dehazing algorithms. A distinguishing feature of RESIDE lies in the diversity of its evaluation criterion, ranging from traditional full-reference metrics, to more practical no-reference metrics, and to the desired human subjective ratings. A novel set of task-driven evaluation options will be discussed later in this paper.

A. Dataset Overview

The REISDE training set contains 13, 990 synthetic hazy images, generated using 1, 399 clear images from existing indoor depth datasets NYU2 [40] and Middlebury stereo [41]. We synthesize 10 hazy images for each clear image. An optional split of 13, 000 for training and 990 for validation is provided. We set different atmospheric lights A, by choosing each channel uniformly randomly between $[0.7, 1.0]$, and select β uniformly at random between $[0.6, 1.8]$. It thus contains paired clean and hazy images, where a clean ground truth image can lead to multiple pairs whose hazy images are generated under different parameters A and β.

The REISDE testing set is composed of *Synthetic Objective Testing Set* (SOTS) and the *Hybrid Subjective Testing Set* (HSTS), designed to manifest a diversity of evaluation viewpoints. SOTS selects 500 indoor images from NYU2 [40] (non-overlapping with training images), and follow the same process as training data to synthesize hazy images. We specially create challenging dehazing cases for testing, e.g., white scenes added with heavy haze. HSTS picks 10 synthetic outdoor hazy images generated in the same way as SOTS, together with 10 real-world hazy images collected real world outdoor scenes [48],[2] combined for human subjective review.

B. Evaluation Strategies

1) From Full-Reference to No-Reference: Despite the popularity of the full-reference PSNR/SSIM metrics for evaluating dehazing algorithms, they are inherently limited due to the unavailability of clean ground truth images

[2]Image Source: http://www.tour-beijing.com/real_time_weather_photo/

TABLE III
AVERAGE FULL- AND NO-REFERENCE EVALUATIONS RESULTS OF DEHAZED RESULTS ON SOTS

	DCP [10]	FVR [11]	BCCR [12]	GRM [13]	CAP [14]	NLD [15]	DehazeNet [18]	MSCNN [19]	AOD-Net [20]
PSNR	16.62	15.72	16.88	18.86	19.05	17.29	21.14	17.57	19.06
SSIM	0.8179	0.7483	0.7913	0.8553	0.8364	0.7489	0.8472	0.8102	0.8504
SSEQ	64.94	67.75	65.83	63.30	64.69	67.46	65.46	65.31	67.65
BLIINDS-II	74.41	75.63	74.45	73.46	73.41	74.85	71.71	74.34	79.02

in practice, as well as their often poor alignment with human perception quality [33]. We thus refer to two no-reference IQA models: spatial-spectral entropy-based quality (SSEQ) [45], and blind image integrity notator using DCT statistics (BLIINDS-II) [44], to complement the shortness of PSNR/SSIM. Note that the score of SSEQ and BLIINDS2 used in [45] and [44] are range from 0 (best) to 100 (worst), and we reverse the score to make the correlation consistent to full-reference metrics.

We will apply PSNR, SSIM, SSEQ, and BLIINDS-II, to the dehazed results on SOTS, and examine how consistent their resulting ranking of dehazing algorithms will be. We will also apply the four metrics on HSTS (PSNR and SSIM are only computed on the 10 synthetic images), and further compare those objective measures with subjective ratings.

2) From Objective to Subjective: Reference [33] investigated various choices of full-reference and no-reference IQA models, and found them to be limited in predicting the quality of dehazed images. We then conduct a subjective user study on the quality of dehazing results produced by different algorithms, from which we gain more useful observations. Ground-truth images are also included when they are available as references.

In the previous survey [33], [49] a participant scored each dehazing result image with an integer from 1 to 10 that best reflects its perceptual quality. We adopt a different pipeline: (1) asking participants to give pairwise comparisons rather than individual ratings, the former often believed to be more robust and consistent in subjective surveys, which has also be adopted by [34] and [50]; (2) decomposing the perceptual quality into two dimensions: the dehazing *Clearness* and *Authenticity*, the former defined as how thoroughly the haze has been removed, and the latter defined as how realistic the dehazed image looks like. Up to our best knowledge, such two disentangled dimensions have not been explored before in similar literature. They are motivated by our observations that some algorithms produce naturally-looking results but are unable to fully remove haze, while some others remove the haze at the price of unrealistic visual artifacts.

During the survey, each participant is shown a set of dehazed result pairs obtained using two different algorithms for the same hazy image. For each pair, a participant needs to independently decide which one is better than the other in terms of Clearness, and then which one is better for Authenticity. The image pairs are drawn from all the competitive methods randomly, and the images winning the pairwise comparison will be compared again in the next round [51], until the best one is selected. We fit a Bradley-Terry [52] model to estimate the subjective scores for each dehazing algorithm so that they can be ranked.

As the same for peer benchmarks [53], [54], the subjective survey is not "automatically" scalable to new results. However, it is extremely important to study the correlation between human perception and objective metrics, which helps analyze the effectiveness of the latter. We are preparing to launch a leaderboard, where we will accept selective result submissions, and periodically run new subjective reviews.

IV. ALGORITHM BENCHMARKING

Based on the rich resources provided by RESIDE, we evaluate 9 representative state-of-the-art algorithms: Dark-Channel Prior (DCP) [10], Fast Visibility Restoration (FVR) [11], Boundary Constrained Context Regularization (BCCR) [12], Artifact Suppression via Gradient Residual Minimization (GRM) [13], Color Attenuation Prior (CAP) [14], Non-local Image Dehazing (NLD) [15], DehazeNet [18], Multi-scale CNN (MSCNN) [19], and All-in-One Dehazing Network (AOD-Net) [20]. The last three belong to the latest CNN-based dehazing algorithms. For all data-driven algorithms, they are trained on the same RESIDE training set.

A. Objective Comparison on SOTS

We first compare the dehazed results on SOTS using two full-reference (PSNR, SSIM) and two no-reference metrics (SSEQ, BLIINDS-II). Table III displays the detailed scores of each algorithm in terms of each metric.[3]

In general, since learning-based methods [14], [18]–[20] are optimized by directly minimizing the mean-square-error (MSE) loss between output and ground truth pairs or maximizing the likelihood on large-scale data, they clearly outperform earlier algorithms based on natural or statistical priors [10]–[13], [15] in most cases, in terms of PSNR and SSIM. Especially, DehazeNet [18] achieves the highest PSNR value, AOD-Net [20] and CAP [14] obtain the suboptimal and third PSNR score. Although GRM [13] achieves the highest SSIM score, AOD-Net [20] and DehazeNet [18] still obtain the similar SSIM values.

However, when it comes to no-reference metrics, the results become less consistent. AOD-Net [20] still maintains competitive performance by obtaining the best BLIINDS-II result on indoor images, thanks to end-to-end pixel correction. On the other hand, several prior-based methods, such as FVR [11] and NLD [15] also show competitiveness: FVR [11] ranks first in term of SSEQ, and NLD [15] achieves the suboptimal SSEQ and BLIINDS-II. We visually observe the results, and find that

[3]We highlight the top-3 performances using red, cyan and blue, respectively.

TABLE IV
AVERAGE FULL-EVALUATIONS RESULTS OF DEHAZED RESULTS ON SOTS WITH DIFFERENT HAZE LEVEL

	DCP [10]	FVR [11]	BCCR [12]	GRM [13]	CAP [14]	NLD [15]	DehazeNet [18]	MSCNN [19]	AOD-Net [20]
$\beta \in [0.6, 0.9]$									
PSNR	16.10	17.18	16.91	18.64	20.88	17.52	24.24	19.72	22.40
SSIM	0.8158	0.7682	0.7978	0.8528	0.8597	0.7558	0.9044	0.8489	0.8980
$\beta \in [1.0, 1.4]$									
PSNR	16.58	16.00	17.07	18.74	19.68	17.37	22.02	17.25	19.61
SSIM	0.8210	0.7538	0.7942	0.8576	0.8450	0.7487	0.8870	0.8110	0.8616
$\beta \in [1.5, 1.8]$									
PSNR	17.15	14.42	17.14	19.11	17.21	17.06	18.67	15.10	16.16
SSIM	0.8259	0.7289	0.7906	0.8555	0.8120	0.7438	0.8454	0.7723	0.8064

TABLE V
AVERAGE SUBJECTIVE SCORES, AS WELL AS FULL- AND NO-REFERENCE EVALUATIONS RESULTS, OF DEHAZING RESULTS ON HSTS

	DCP [10]	FVR [11]	BCCR [12]	GRM [13]	CAP [14]	NLD [15]	DehazeNet [18]	MSCNN [19]	AOD-Net [20]
	Synthetic images								
Clearness	1.26	0.18	0.62	0.75	0.50	1	0.29	1.22	0.86
Authenticity	0.78	0.14	0.50	0.95	0.86	1	1.94	0.54	1.41
PSNR	14.84	14.48	15.08	18.54	21.53	18.92	24.48	18.64	20.55
SSIM	0.7609	0.7624	0.7382	0.8184	0.8726	0.7411	0.9153	0.8168	0.8973
SSEQ	86.15	85.68	85.60	78.43	85.32	86.28	86.01	85.56	86.75
BLIINDS-II	90.70	87.65	91.05	82.30	85.75	85.30	87.15	88.70	87.50
	Real-world images								
Clearness	0.39	0.46	0.45	0.75	1	0.54	1.16	1.29	1.05
Authenticity	0.17	0.20	0.18	0.62	1	0.15	1.03	1.27	1.07
SSEQ	68.65	67.75	66.63	70.19	67.67	67.96	68.34	68.44	70.05
BLIINDS-II	69.35	72.10	68.55	79.60	63.55	70.80	60.35	62.65	74.75

DCP [10], BCCR [12] and NLD [15] tend to produce sharp edges and highly contrasting colors, which explains why they are preferred by BLIINDS-II and SSEQ. Such an inconsistency between full- and no-reference evaluations aligns with the previous argument [33] that existing objective IQA models are very limited in providing proper quality predictions of dehazed images.

We have further conducted an experiment using standard evaluation metrics, with different haze concentration levels (i.e.,β values), to detail the suitability of each method for each distinct haze density. As show in Table IV, we split the SOTS dataset into three groups according to the ranges of β. It makes clear that DehazeNet is consistently the best for light and medium haze, and GRM achieves the highest PSNR and SSIM for thick haze.

B. Subjective Comparison on HSTS

We recruit 100 participants from different educational backgrounds for the subjective, using HSTS which contains 10 synthetic outdoor and 10 real-world hazy images. We fit a Bradley-Terry [52] model to estimate the subjective score for each method so that they can be ranked. In the Bradley-Terry model, the probability that an object X is favored over Y is assumed to be

$$p(X \succ Y) = \frac{e^{s_X}}{e^{s_X} + e^{s_Y}} = \frac{1}{1 + e^{s_Y - s_X}}, \tag{4}$$

where s_X and s_Y are the subjective scores for X and Y. The scores $\mathbf{s}$ for all the objects can be jointly estimated by maximizing the log likelihood of the pairwise comparison observations:

$$\max_{\mathbf{s}} \sum_{i,j} w_{ij} \log\left(\frac{1}{1 + e^{s_j - s_i}}\right), \tag{5}$$

where w_{ij} is the (i, j)-th element in the winning matrix $\mathbf{W}$, representing the number of times when method i is favored over method j. We use the Newton-Raphson method to solve Eq. (5). Note that for a synthetic image, we have a 10×10 winning matrix $\mathbf{W}$, including the ground truth and nine dehazing methods' results. For a real-world image, its winning matrix $\mathbf{W}$ is 9×9 due to the absence of ground truth. For synthetic images, we set the score for ground truth method as 1 to normalization scores.

Figures 3 and 4 show qualitative examples of dehazed results on a synthetic and a real-world image, respectively. Quantitative results can be found in Table V and the trends are visualized in Figure 2. We also compute the full- and no-reference metrics on synthetic images to examine their consistency with the subjective scores.

A few interesting observations could be drawn:

- The subjective qualities of various algorithms' results show different trends on synthetic and real hazy images. On the 10 synthetic images of HSTS, DCP [10] receives the best clearness score and DehazeNet is the best in authenticity score. On the 10 real images, CNN-based methods [18]–[20] rank top-3 in terms of both clearness and authenticity, in which MSCNN [19] achieves the best according to both scores.
- The clearness and authenticity scores of the same image are often not aligned. As can be seen from Figure 2, the two subjective scores are hardly correlated on

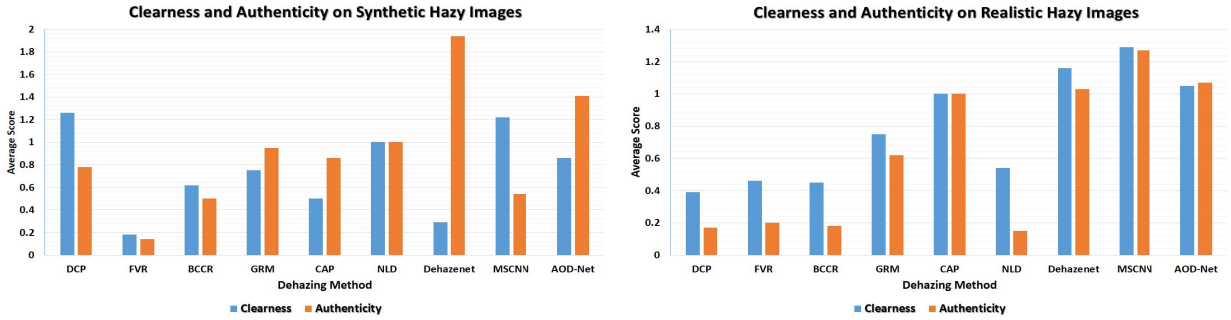


Fig. 2. Averaged clearness and authenticity scores: (a) on 10 synthetic images in HSTS; and (b) on real-world images in HSTS.

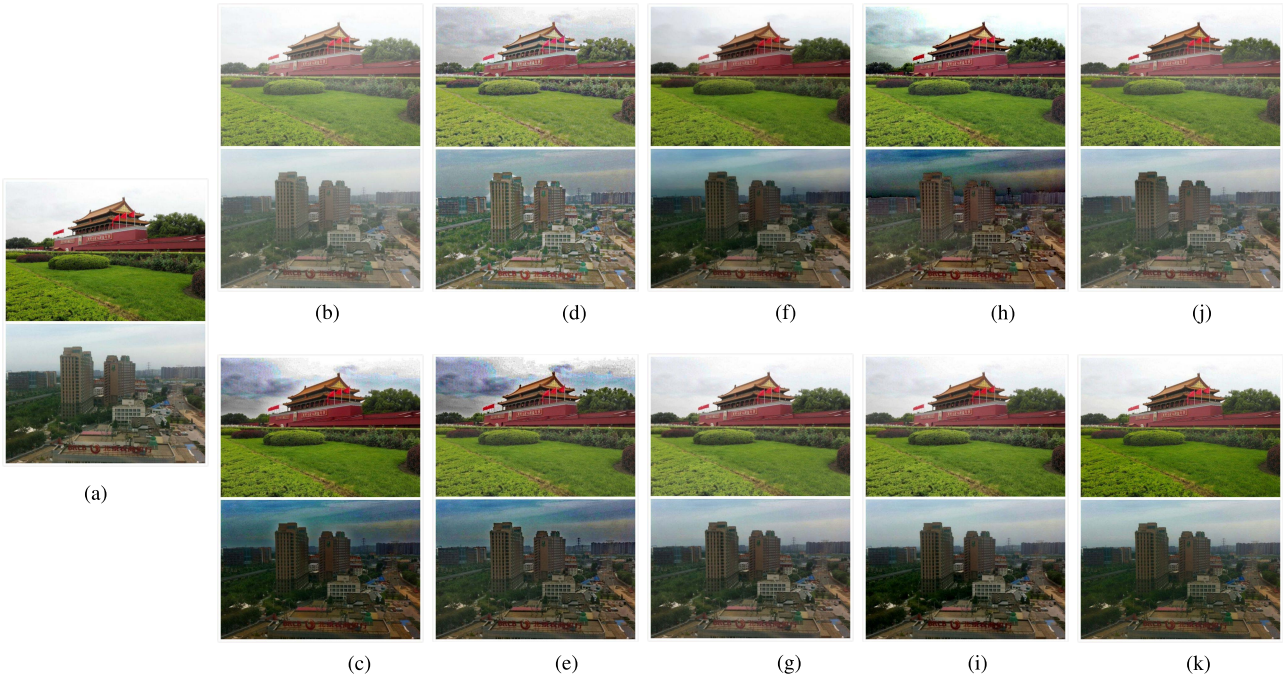


Fig. 3. Examples of dehazed results on a synthetic hazy image from HSTS. (a) Clean Image. (b) Hazy Image. (c) DCP. (d) FVR. (e) BCCR. (f) GRM. (g) CAP. (h) NLD. (i) DehazeNet. (j) MSCNN. (k) AOD-Net.

synthetic images; their correlation shows better on real images. That reflects the complexity and multi-facet nature of subjective perceptual evaluation.

- From Table V, we observe the divergence between subjective and objective (both full- and no-reference) evaluation results. For the best performer in subjective evaluation, MSCNN [19], its PSNR/SSIM results on synthetic indoor images are quite low, while SSEQ/BLIINDS-II on both synthetic and outdoor images are moderate. As another example, GRM [13] receives the highest SSEQ/BLIINDS-II scores on real HSTS images. However, both of its subjective scores rank only fifth among nine algorithms on the same set.

C. Running Time

Table VI reports the per-image running time of each algorithm, averaged over the synthetic indoor images (620×460) in SOTS, using a machine with 3.6 GHz CPU and 16G RAM. All methods are implemented in MATLAB, except AOD-Net by Pycaffe. However, it is fair to compare AOD-Net with other methods since MATLAB implementation has superior efficiency than Pycaffe as shown in [20]. AOD-Net shows a clear advantage over others in efficiency, thanks to its light-weight feed-forward structure.

V. What Are Beyond: From RESIDE to RESIDE-β

RESIDE serves as a sufficient benchmark for evaluating single image dehazing as a traditional image restoration problem: either to ensure signal fidelity or to please *human vision*. However, dehazing is increasingly demanded in *machine vision* systems in outdoor environments, whose requirement is not naturally met by taking an image restoration viewpoint. To identify and eliminate the gaps between *current dehazing research* and the *practical application need*, we introducing the RESIDE-β part, as an exploratory and supplementary part of the RESIDE benchmark, including two innovative explorations on solving two hurdles, on training data content and evaluation criteria, respectively. Being our novel try,

Fig. 4. Examples of dehazed results on a real-world hazy image from HSTS. (a) Hazy Image. (b) DCP. (c) FVR. (d) BCCR. (e) GRM. (f) CAP. (g) NLD. (h) DehazeNet. (i) MSCNN. (j) AOD-Net.

TABLE VI
COMPARISON OF AVERAGE PER-IMAGE RUNNING TIME (SECOND) ON SYNTHETIC INDOOR IMAGES IN SOTS

	DCP [10]	FVR [11]	BCCR [12]	GRM [13]	CAP [14]	NLD [15]	DehazeNet [18]	MSCNN [19]	AOD-Net [20]
Time	1.62	6.79	3.85	83.96	0.95	9.89	2.51	2.60	0.65

RESIDE-β has a "beta stage" nature and is meant to inspire more followers.

A. Indoor Versus Outdoor Training Data

Up to our best knowledge, almost all data-driven dehazing models have been utilizing synthetic training data, because of the prohibitive difficulty of simultaneously collecting real-world hazy RGB images and their "hazy-free" ground truth. Most outdoor scenes contain object movements from time to time, e.g. traffic surveillance and autonomous driving. Even in a static outdoor scene, the change of illumination conditions etc. along time is inevitable. Despite their positive driving effects in the development of dehazing algorithms, those synthetic images are collected from indoor scenes [40], [41], while dehazing is applied to outdoor environments.

The content of training data thus significantly diverges from the target subjects in real dehazing applications. Such a mismatch might undermine the practical effectiveness of the trained dehazing models. Reference [46] collected 14 outdoor clean images with accurate depth information, and proposed to generate hazy images from them with parameters that are chosen to be physically realistic. Their meaningful and delicate efforts are however not straightforward to scale up and generate large-scale training sets.

Aiming for automatic generation of large-scale realistic outdoor hazy images, we first examine the possibility of utilizing existing outdoor depth datasets. While several such datasets, e.g., Make3D [55] and KITTI [56], have been proposed, their depth information is less precise and incomplete compared to indoor datasets. For example, due to the limitations of RGB-based depth cameras, the Make3D dataset suffer from at least 4 meters of average root mean squared error in the predicted depths, and the KITTI dataset has at least 7 meters of average error [57]. In comparison, the average depth errors in indoor datasets, e.g., NYU-Depth-v2 [40], are usually as small as 0.5 meter. For the outdoor depth maps can also contain a large amount of artifacts and large holes, which renders it inappropriate for direct use in haze simulation. We choose Make3D to synthesize hazy images in the same way as we did for RESIDE training set, a number of examples being displayed at the first row of Figure 5. It can be easily seen that they suffer from unrealistic artifacts (e.g., notice the "blue" regions around the tree), caused by inaccurate depth map. A possible remedy is to adopt recent approaches of depth map denoising and in-painting [34], [58], which we leave for future.

TABLE VII
PERCEPTUAL LOSS ON SOTSāłł INDOOR IMAGES

	Haze	DCP [10]	FVR [11]	BCCR [12]	GRM [13]	CAP [14]	NLD [15]	DehazeNet [18]	MSCNN [19]	AOD-Net [20]
Relu2_2	0.0558	0.0473	0.0601	0.0593	0.0395	0.0380	0.0523	0.0314	0.0417	0.0394
Relu3_3	0.0814	0.0731	0.0988	0.0885	0.0626	0.0617	0.0805	0.0520	0.0651	0.0634
Relu4_3	0.0205	0.0190	0.0256	0.0217	0.0168	0.0165	0.0206	0.0143	0.0172	0.0186
Relu5_3	0.0264	0.0158	0.0280	0.0188	0.0161	0.0195	0.0186	0.0151	0.0204	0.0173

Fig. 5. Visual comparison between the synthetic hazy images directly generated from Make3D (first row) and from OTS (second row).

Another option is to estimate depth from outdoor images and then synthesizing hazy images. After comparing different depth estimation methods, we find the algorithm in [38] to produce fewest visible depth errors and to cause much less visual artifacts on natural outdoor images, same as [20] observed. We display a few synthetic hazy examples generated by using [38] for depth estimation, in the second row of Figure 5. By comparing them with the first row (Make3D), one can see that using depth estimation [38] leads to much more visually plausible results.

We thus extend to a large scale effort, collecting 2, 061 real world outdoor images from [48], among which we carefully excluded those originally with haze and ensure their scenes to be as diverse as possible. We use [38] to estimate the depth map for each image, with which we finally synthesize 72, 135 outdoor hazy images with β in [0.04, 0.06, 0.08, 0.1, 0.12, 0.16, 0.2] and A in [0.8, 0.85, 0.9, 0.95, 1]. This new set, called Outdoor Training Set (OTS), consists of paired clean outdoor images and generated hazy ones. It is included as a part of RESIDE-β, and could be used for training. Despited that depth estimation could potentially be noisy, we visually inspect the new set and find most generated hazy images to be free of noticeable artifacts (and much better than generating using Make3D). As we observed from preliminary experiments, including this outdoor set for training performed in general similarly on SOTS in the sense of PSNR/SSIM, but improved the generalization performance on real-world images, in terms of visual quality.

B. Restoration Versus High-Level Vision

It has been recognized that the performance of high-level computer vision tasks, such as object detection and recognition, will deteriorate in the presence of various degradations, and is thus largely affected by the quality of image restoration and enhancement. Dehazing could be used as pre-processing for many computer vision tasks executed in the wild, and the resulting task performance could in turn be treated as an indirect indicator of the dehazing quality. Such a "task-driven" evaluation way has received little attention so far, despite its great implications for outdoor applications.

A relevant preliminary effort was presented in [20], where the authors compared a few CNN-based dehazing models by placing them in an object detection pipeline, but their tests were on synthetic hazy data with bounding boxes. Reference [34] created a relatively small dataset of 101 real-world images depicting foggy driving scenes, which came with ground truth annotations for evaluating semantic segmentation and object detection. We notice that [34] investigated detection and segmentation problems in hazy images as well, evaluated on a small image set with only three dehazing methods.

1) Full-Reference Perceptual Loss Comparison on SOTS: Since dehazed images are often subsequently fed for automatic semantic analysis tasks such as recognition and detection, we argue that the optimization target of dehazing in these tasks is neither pixel-level or perceptual-level quality, but the utility of the dehazed images in the given semantic analysis task [59]. The perceptual loss [32] was proposed to measure the semantic-level similarity of images, using the VGG recognition model[4] pre-trained on ImageNet dataset [60]. Here, we compared the Euclidean distance between clean images and dehazed images with different level features including relu2_2, relu3_3, relu4_3 and relu5_3. Since it is a full-reference metric, we compute the perceptual loss on the SOTS dataset, as listed in Table VII. We also compute the perceptual loss on the 10 synthetic images in HSTS, to examine how well it agrees with the perceptual quality, as seen from Table VIII. DehazeNet and CAP consistently lead to the lowest perceptual loss differences on both sets, which seem to be in general aligned with PSNR results, but not SSIM or other two no-reference metrics.

On HSTS synthetic images, we observe the perceptual loss to be correlated to the authenticity score to some extent (e.g., DehazeNet and AOD-Net perform well under both), but hardly correlated to the clearness. It might imply that for preserving significant semantical similarities for recognition, it is preferable to keep a realistic visual look than to thoroughly remove haze. In other words, "under-dehazed" images might be preferred over "over-dehazed" images, the latter potentially losing details and suffering from method artifacts.

[4]Public available at http://www.robots.ox.ac.uk/ṽgg/software/very_deep/caffe/VGG_ILSVRC_16_layers.caffemodel

TABLE VIII
PERCEPTUAL LOSS ON HSTS 10 SYNTHETIC OUTDOOR IMAGES

	Haze	DCP [10]	FVR [11]	BCCR [12]	GRM [13]	CAP [14]	NLD [15]	DehazeNet [18]	MSCNN [19]	AOD-Net [20]
Relu2_2	0.0595	0.0544	0.0593	0.0635	0.0443	0.0334	0.0541	0.0233	0.0452	0.0356
Relu3_3	0.0918	0.0838	0.0973	0.0944	0.0659	0.0538	0.0829	0.0392	0.0728	0.0596
Relu4_3	0.0234	0.0213	0.0274	0.0240	0.0183	0.0145	0.0217	0.0108	0.0264	0.0165
Relu5_3	0.0347	0.0184	0.0320	0.0207	0.0196	0.0181	0.0213	0.0122	0.0192	0.0178

TABLE IX
DETAILED CLASSES INFORMATION OF RTTS

Category	*person*	*bicycle*	*car*	*bus*	*motorbike*	Total
Normal	7,950	534	18,413	1,838	862	29,597
Difficult	3,416	164	6,904	752	370	11,606
Total	11,366	698	25,317	2,590	1,232	41,203

2) No-Reference Task-Driven Comparison on RTTS: For real-world images without ground-truth, following [20], we adopt a task-driven evaluation scheme for dehazing algorithms, by studying the object detection performance on their dehazed results. Specially, we used several state-of-the-art pre-trained object detection models, including Faster R-CNN (FRCNN) [61], YOLO-V2 [62], SSD-300 and SSD-512 [63],[5] to detect objects of interests from the dehazed images, and rank all algorithms via the mean Average Precision (mAP) results achieved.

For that purpose, we collect a *Real-world Task-driven Testing Set* (RTTS), consisting of 4, 322 real-world hazy images crawled from the web, covering mostly traffic and driving scenarios. Each image is annotated with object categories and bounding boxes, and RTTS is organized in the same form as VOC2007 [64]. We currently focus on five traffic-related categories: car, bicycle, motorbike, person, bus. We obtain 41, 203 annotated bounding boxes, 11, 606 of which are marked as "difficult" and not used in this paper's experiments. The class details of RTTS are shown in Table IX. Additionally, we also collect 4,807 unannotated real-world hazy images, which are not exploited in this paper, but may potentially be used for domain adaption in future, etc. The RTTS set is the largest annotated set of its kind.

Table X compares all mAP results.[6] The results are not perfectly consistent among four different detection models, the overall tendency clearly shows that MSCNN, BCCR, and DCP are the top-3 choices that are most favored by detection tasks on RTTS. If comparing the ranking of detection mAP with the no-reference results on the same set (see Table XI), we can again only observe a weak correlation. For example, BCCR [12] achieves highest BLIINDS-II value, but MSCNN has lower SSEQ and BLIINDS-II scores than most competitors. We further notice that MSCNN also achieved the best clearness and authenticity on HSTS real-world images (see Table V). Figure 6 display the object detection results using FRCNN on an RTTS hazy image and after applying nine different dehazing algorithms.

3) Discussion: Optimizing Detection Performance in Haze?: Reference [20] for the first time reported the promising performance on detecting objects in the haze, by concatenating and jointly tuning AOD-Net with FRCNN as one unified pipeline, similar to other relevant works [65], [66], [69]. The authors trained their detection pipeline using an annotated dataset of synthetic hazy images, generated from VOC2007 [64]. Due to the absence of annotated realistic hazy images, they only reported quantitative performance on a separate set of synthetic annotated images. While their goal is different from the scope of RTTS (where a fixed FRCNN is applied on dehazing results for fair comparison), we are interested to explore whether we could further boost the detection mAP on RTTS realistic hazy images using such a joint pipeline. We also point to other recent works utilizing domain adaptation [70].

In order for further enhancing the performance of such a dehazing + detection joint pipeline in realistic hazy photos or videos, there are at least two other noteworthy potential options as we can see for future efforts:

- Developing *photo-realistic* simulation approaches of generating hazy images from clean ones [71], [72]. That would resolve the bottleneck of handle-labeling and supply large-scale annotated training data with little mismatch. The technique of haze severity estimation [73] may also help the synthesis, by first estimating the haze level from (unannotated) testing images and then generating training images accordingly.
- If we view the synthetic hazy images as the source domain (with abundant labels) and the realistic ones as the target domain (with scarce labels), then the unsupervised domain adaption can be performed to reduce the domain gap in low-level features, by exploiting unannotated realistic hazy images. For example, [74] provided an example of pre-training the robust low-level CNN filters using unannotated data from both source and target domains, leading to much improved robustness when applied to testing on the target domain data. For this purpose, we have included 4,322 unannotated realistic hazy images in RESIDE that might help build such models.

Apparently, the above discussions can be straightforwardly applied to other high-level vision tasks in uncontrolled outdoor environments (e.g., bad weathers and poor illumination), such as tracking, recognition, semantic segmentation, etc.

[5]Here we use py-faster-rcnn and its model is trained on VOC2007_trainval, while official implementations are used for YOLO-V2 and SSDs and their models are trained on both VOC2007_trainval and VOC2012_trainval

[6]For FVR, only 3,966 images are counted, since for the remaining 356 FVR fails to provide any reasonable result.

TABLE X

ALL DETECTION RESULTS ON RTTS(in %), PLEASE NOTE THAT, THE MODEL USED IN FRCNN IS TRAINED ON VOC2007_TRAINVAL DATASET, WHILE THE MODELS USED IN YOLO-V2 AND SSDS ARE TRAINED ON VOC2007_TRAINVAL + VOC2012_TRAINVAL

		Haze	DCP [10]	FVR [11]	BCCR [12]	GRM [13]	CAP [14]	NLD [15]	DehazeNet [18]	MSCNN [19]	AOD [20]
mAP	FRCNN [61]	37.58	40.58	35.01	41.56	28.90	39.63	40.03	40.54	41.34	37.47
	YOLO-V2 [62]	40.37	39.81	38.06	40.65	29.41	39.80	39.93	40.10	40.76	40.53
	SSD-300 [63]	50.26	49.40	47.04	51.57	35.59	50.31	49.84	50.14	51.82	49.77
	SSD-512 [63]	55.55	55.71	52.29	57.17	39.18	55.70	54.99	55.40	56.88	55.29
Person	FRCNN [61]	60.84	61.54	57.72	64.51	50.22	61.29	60.53	61.40	61.43	61.22
	YOLO-V2 [62]	61.24	61.14	60.00	61.16	50.13	61.24	60.49	61.16	61.30	61.20
	SSD-300 [63]	68.60	68.18	66.36	69.12	53.91	68.78	66.96	68.18	69.20	68.28
	SSD-512 [63]	72.58	72.72	69.45	73.34	56.74	72.50	71.20	72.34	73.13	72.62
Bicycle	FRCNN [61]	40.72	40.77	38.76	44.57	30.71	40.48	40.21	40.68	41.69	40.33
	YOLO-V2 [62]	44.63	43.39	40.08	43.66	28.81	42.65	43.56	42.34	43.53	44.55
	SSD-300 [63]	54.92	51.36	49.35	53.33	34.48	53.38	53.42	53.08	55.73	54.18
	SSD-512 [63]	58.45	56.70	54.57	58.57	36.70	57.49	56.38	57.50	58.76	57.91
Car	FRCNN [61]	35.18	42.15	34.74	42.69	26.30	41.52	42.30	41.74	42.61	35.13
	YOLO-V2 [62]	39.39	38.93	37.22	39.88	29.91	39.03	38.96	39.35	40.00	39.49
	SSD-300 [63]	54.14	54.98	50.81	56.32	40.21	55.08	54.98	55.27	56.32	54.62
	SSD-512 [63]	63.05	64.95	61.54	65.80	47.79	64.15	65.04	64.21	65.22	64.05
Bus	FRCNN [61]	20.90	24.18	19.06	24.66	14.81	24.74	23.74	25.20	25.25	20.56
	YOLO-V2 [62]	20.57	19.34	19.42	20.01	12.86	18.90	18.22	19.07	19.63	19.09
	SSD-300 [63]	30.13	30.87	30.98	33.70	19.72	30.90	30.43	30.86	32.26	29.42
	SSD-512 [63]	34.60	36.51	33.47	37.69	22.81	35.47	34.31	35.18	37.42	34.13
Motorbike	FRCNN [61]	30.24	34.25	24.78	34.34	22.44	30.10	33.36	33.70	35.72	30.09
	YOLO-V2 [62]	37.84	36.23	33.59	38.54	25.33	37.10	38.40	38.59	39.33	38.31
	SSD-300 [63]	43.48	41.61	37.72	45.38	29.63	43.41	43.40	43.30	45.60	42.35
	SSD-512 [63]	49.08	47.69	42.40	50.46	31.85	48.89	48.04	47.79	49.87	47.76

TABLE XI

AVERAGE NO-REFERENCE METRICS OF DEHAZED RESULTS ON RTTS

	DCP [10]	FVR [11]	BCCR [12]	GRM [13]	CAP [14]	NLD [15]	DehazeNet [18]	MSCNN [19]	AOD-Net [20]
SSEQ	62.87	63.59	63.31	58.64	60.66	59.37	60.01	62.31	65.35
BLIINDS-II	68.34	67.68	74.07	54.54	65.15	68.32	52.54	56.59	71.05

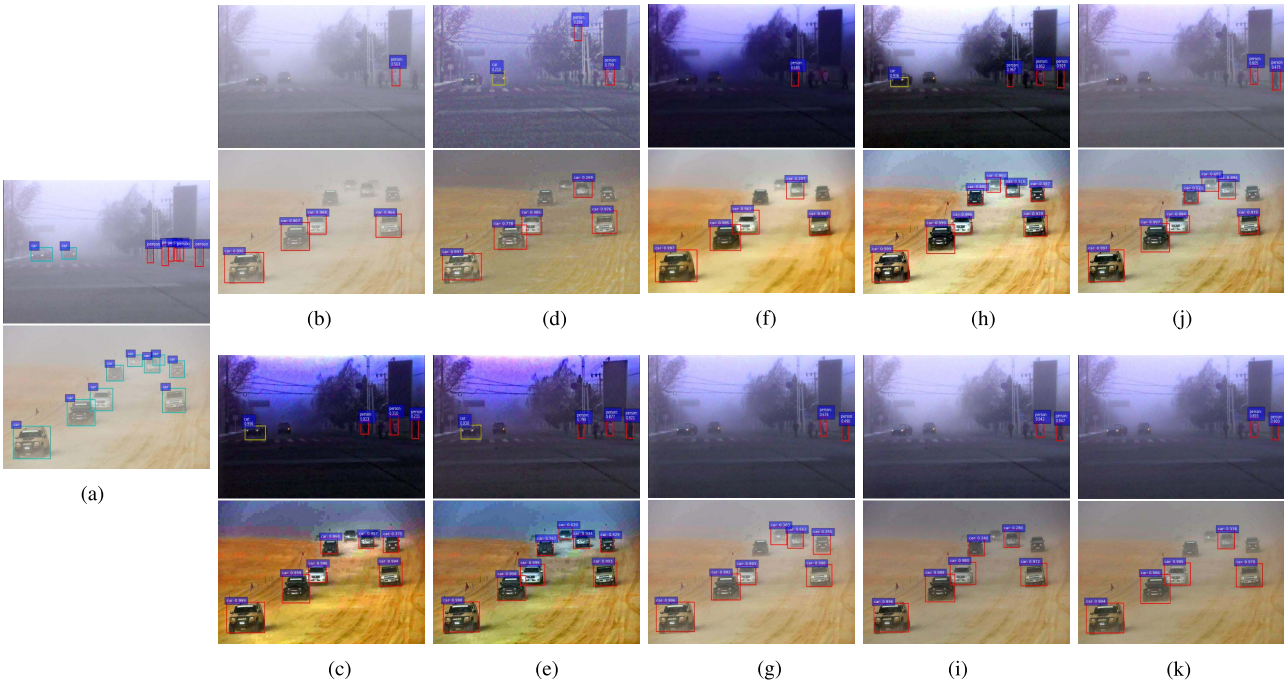


Fig. 6. Visualization of two RTTS images' object detection results after applying different dehazing algorithms. (a) Ground Truth. (b) RawHaze. (c) DCP. (d) FVR. (e) BCCR. (f) GRM. (g) CAP. (h) NLD. (i) DehazeNet. (j) MSCNN. (k) AOD-Net.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we systematically evaluate the state-of-the-arts in single image dehazing. From the results presented, there seems to be no single-best dehazing model for all criteria: AOD-Net and DehazeNet are favored by PSNR and SSIM; DCP, FVR and BCCR are more competitive in terms of no-reference metrics; DehazeNet performs best in terms of perceptual loss; MSCNN shows to have the most appreciated subjective quality and superior detection performance on real hazy images; and AOD-Net is the most efficient among all. The reason why each dehazing method might succeed or fail in each evaluation case is certainly complicated, e.g., depending on the prior it uses or the model's design choices. Some overall remarks and empirical hypotheses made by the authors are:

- Deep learning methods [18]–[20], especially with the end-to-end optimization towards reconstruction loss [20],

are advantageous under traditional PSNR and SSIM metrics. However, the two metrics do not necessarily reflect human perceptual quality, and those models may not always generalize well on real-world hazy images.

- Classical prior-based methods [11], [12], [22] seem to generate results favored more by human perception. That is probably because their priors explicitly emphasized illumination, contrasts, or edge sharpness, to which human eyes are particularly sensitive. On the other hand, the typical MSE loss used in deep learning methods tend to over-smooth visual details in results, which are thus less preferred by human viewers. We refer the readers to a later manuscript [70] for more related discussions.
- The detection results on RTTS endorse MSCNN [19] in particular, which is aligned with the current trend in object detection to use multi-scale features [75].

Based on the RESIDE study and its extensions, we see the highly complicated nature of the dehazing problem, in both real-world generalization and evaluation criteria. For future research, we advocate to be evaluate and optimize dehazing algorithms towards more dedicated criteria (e.g., subjective visual quality, or high-level target task performance), rather than solely PSNR/SSIM, which are found to be poorly aligned with other metrics we used. In particular, correlating dehazing with high-level computer vision problems will likely lead to innovative robust computer vision pipelines that will find many immediate applications. Another blank to fill is developing no-reference metrics that are better correlated with human perception, for evaluating dehazing results. That progress will accelerate the needed shift from current full-reference evaluation on only synthetic images, to the more realistic evaluation schemes with no ground truth.

ACKNOWLEDGMENT

The authors appreciate the support from the authors of [15] and [21]. They also acknowledge Dr. Changxing Ding, South China University of Technology, for his indispensable support to our data collection and cleaning.

REFERENCES

[1] *How Autonomous Vehicles Will Navigate Bad Weather Remains Foggy*. [Online]. Available: https://www.forbes.com/sites/centurylink/2016/11/29/how-autonomous-vehicles-will-navigate-bad-weather-remains-foggy/#1aff07088662

[2] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[3] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Trans. Pattern Anal. Mach. Learn.*, vol. 25, no. 6, pp. 713–724, Jun. 2003.

[4] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Dec. 2001, p. 1.

[5] T. Treibitz and Y. Y. Schechner, "Polarization: Beneficial for visibility enhancement?" in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 525–532.

[6] J. Kopf *et al.*, "Deep photo: Model-based photograph enhancement and viewing," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 116-1–116-10, Dec. 2008.

[7] E. J. McCartney, *Optics of the Atmosphere: Scattering by Molecules and Particles*. New York, NY, USA: Wiley, 1976.

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

[9] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.*, vol. 48, no. 3, pp. 233–254, 2002.

[10] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1956–1963.

[11] J.-P. Tarel and N. Hautière, "Fast visibility restoration from a single color or gray level image," in *Proc. 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 2201–2208.

[12] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 617–624.

[13] C. Chen, M. N. Do, and J. Wang, "Robust image and video dehazing with visual artifact suppression via gradient residual minimization," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 576–591.

[14] 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.

[15] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1674–1682.

[16] Y. Jiang, C. Sun, Y. Zhao, and L. Yang, "Image dehazing using adaptive bi-channel priors on superpixels," *Comput. Vis. Image Understand.*, vol. 165, pp. 17–32, Dec. 2017.

[17] M. Ju, Z. Gu, and D. Zhang, "Single image haze removal based on the improved atmospheric scattering model," *Neurocomputing*, vol. 260, pp. 180–191, Oct. 2017.

[18] 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.

[19] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 154–169.

[20] 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.*, Oct. 2017, pp. 4780–4788.

[21] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, p. 72, Aug. 2008.

[22] 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.

[23] 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.

[24] Z. Li, P. Tan, R. T. Tan, D. Zou, S. Zhiying Zhou, and L.-F. Cheong, "Simultaneous video defogging and stereo reconstruction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 4988–4997.

[25] L. Kratz and K. Nishino, "Factorizing scene albedo and depth from a single foggy image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Oct. 2009, pp. 1701–1708.

[26] K. Nishino, L. Kratz, and S. Lombardi, "Bayesian defogging," *Int. J. Comput. Vis.*, vol. 98, no. 3, pp. 263–278, 2012.

[27] Y. Li, R. T. Tan, and M. S. Brown, "Nighttime haze removal with glow and multiple light colors," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 226–234.

[28] J. Zhang, Y. Cao, S. Fang, Y. Kang, and C. W. Chen, "Fast haze removal for nighttime image using maximum reflectance prior," in *Proc. IEEE CVPR*, Jul. 2017, pp. 7016–7024.

[29] D. Nair, P. A. Kumar, and P. Sankaran, "An effective surround filter for image dehazing," in *Proc. ACM Int. Conf. Interdiscipl. Adv. Appl. Comput.*, 2014, p. 20.

[30] J. Zhou and F. Zhou, "Single image dehazing motivated by retinex theory," in *Proc. 2nd Int. Symp. Instrum. Meas., Sensor Netw. Automat. (IMSNA)*, Dec. 2013, pp. 243–247.

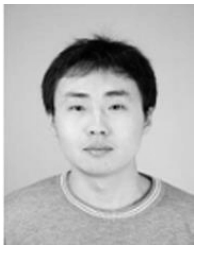
[31] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. (2017). "End-to-end united video dehazing and detection." [Online]. Available: https://arxiv.org/abs/1709.03919

[32] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 694–711.

[33] K. Ma, W. Liu, and Z. Wang, "Perceptual evaluation of single image dehazing algorithms," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3600–3604.

[34] C. Sakaridis, D. Dai, and L. Van Gool. (2017). "Semantic foggy scene understanding with synthetic data." [Online]. Available: https://arxiv.org/abs/1708.07819
[35] 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.
[36] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1122–1131.
[37] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graph.*, vol. 34, no. 1, pp. 13-1–13-14, Dec. 2014.
[38] F. Liu, C. Shen, G. Lin, and I. Reid, "Learning depth from single monocular images using deep convolutional neural fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 2024–2039, Oct. 2016.
[39] J. P. Tarel, N. Hautiere, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer, "Vision enhancement in homogeneous and heterogeneous fog," *IEEE Intell. Transp. Syst. Mag.*, vol. 4, no. 2, pp. 6–20, Apr. 2012.
[40] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 746–760.
[41] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2003, p. 1.
[42] M. Cordts *et al.*, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3213–3223.
[43] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[44] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
[45] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process., Image Commun.*, vol. 29, no. 8, pp. 856–863, 2014.
[46] 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.
[47] Z. Chen, T. Jiang, and Y. Tian, "Quality assessment for comparing image enhancement algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3003–3010.
[48] *Beijing Realtime Weather Photos*. Accessed: 2000. [Online]. Available: http://goo.gl/svzxLm
[49] Y. Li, S. You, M. S. Brown, and R. T. Tan, "Haze visibility enhancement: A survey and quantitative benchmarking," *Comput. Vis. Image Understand.*, vol. 165, pp. 1–16, Dec. 2017.
[50] Z. Wang, Y. Yang, Z. Wang, S. Chang, J. Yang, and T. S. Huang, "Learning super-resolution jointly from external and internal examples," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4359–4371, Nov. 2015.
[51] U. Pulkkinen, "Bayesian analysis of consistent paired comparisons," *Rel. Eng. Syst. Saf.*, vol. 43, no. 1, pp. 1–16, 1994.
[52] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. The method of paired comparisons," *Biometrika*, vol. 39, nos. 3–4, pp. 324–345, 1952.
[53] C.-Y. Yang, C. Ma, and M.-H. Yang, "Single-image super-resolution: A benchmark," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2014, pp. 372–386.
[54] W.-S. Lai, J.-B. Huang, Z. Hu, N. Ahuja, and M.-H. Yang, "A comparative study for single image blind deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1701–1709.
[55] A. Saxena, M. Sun, and A. Y. Ng, "Make3D: Learning 3D scene structure from a single still image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 824–840, May 2009.
[56] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.
[57] F. Ma and S. Karaman. (2017). "Sparse-to-dense: Depth prediction from sparse depth samples and a single image." [Online]. Available: https://arxiv.org/abs/1709.07492
[58] L. Wang, H. Jin, R. Yang, and M. Gong, "Stereoscopic inpainting: Joint color and depth completion from stereo images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.
[59] D. Liu, D. Wang, and H. Li, "Recognizable or not: Towards image semantic quality assessment for compression," *Sens. Imag.*, vol. 18, p. 1, Dec. 2017.
[60] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[61] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[62] J. Redmon and A. Farhadi. (2017). "Yolo9000: Better, faster, stronger." [Online]. Available: https://arxiv.org/abs/1612.08242
[63] W. Liu *et al.*, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 21–37.
[64] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. *The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results*. Accessed: Nov. 6, 2007. [Online]. Available: http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html
[65] Z. Wang, S. Chang, Y. Yang, D. Liu, and T. S. Huang, "Studying very low resolution recognition using deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 4792–4800.
[66] B. Cheng *et al.*, "Robust emotion recognition from low quality and low bit rate video: A deep learning approach," in *Proc. 7th Conf. Affect. Comput. Intell. Interact.*, Oct. 2017, pp. 65–70.
[67] D. Liu, B. Wen, X. Liu, Z. Wang, and T. S. Huang. (2017). "When image denoising meets high-level vision tasks: A deep learning approach." [Online]. Available: https://arxiv.org/abs/1706.04284
[68] D. Liu, Z. Wang, Y. Fan, X. Liu, Z. Wang, S. Chang, and T. Huang, "Robust video super-resolution with learned temporal dynamics," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2526–2534.
[69] D. Liu *et al.*, "Learning temporal dynamics for video super-resolution: A deep learning approach," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3432–3445, Jul. 2018.
[70] Y. Liu *et al.* (2018). "Improved techniques for learning to dehaze and beyond: A collective study." [Online]. Available: https://arxiv.org/abs/1807.00202
[71] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb. (2016). "Learning from simulated and unsupervised images through adversarial training." [Online]. Available: https://arxiv.org/abs/1612.07828
[72] K. Li, Y. Li, S. You, and N. Barnes, "Photo-realistic simulation of road scene for data-driven methods in bad weather," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Worksho*, Oct. 2017, pp. 491–500.
[73] Y. Li, J. Huang, and J. Luo, "Using user generated online photos to estimate and monitor air pollution in major cities," in *Proc. ACM 7th Int. Conf. Internet Multimedia Comput. Service*, 2015, p. 79.
[74] Z. Wang *et al.*, "Deepfont: Identify your font from an image," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 451–459.
[75] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. CVPR*, vol. 1, no. 2, 2017, p. 4.

Boyi Li is currently pursuing the Ph.D. degree with the Computer Science Department, Cornell University. Her research focuses on computer vision, multimedia art, and machine learning. Her previous research on image and video dehazing was published by the ICCV, CVPR Workshop, and AAAI. She was a recipient of the MSRA STAR of Tomorrow Internship Program Award and the National Scholarship.



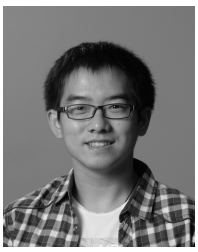
Wenqi Ren received the Ph.D. degree from Tianjin University in 2017. From 2015 to 2016, he was a joint-training Ph.D. student in electrical engineering and computer science with the University of California at Merced, Merced, CA, USA. He is currently an Assistant Professor with the Institute of Information Engineering, Chinese Academy of Sciences, China. His research interests include image/video analysis and enhancement, and related vision problems.

Dengpan Fu received the B.S. degree in electronic engineering from the University of Science and Technology of China in 2015. He is currently pursuing the joint Ph.D. degrees with the University of Science and Technology of China and Microsoft Research Asia. His research interests include computer vision and reinforcement learning.

Dacheng Tao (F'15) is currently a Professor of computer science and an ARC Laureate Fellow with the School of Information Technologies and the Faculty of Engineering and Information Technologies, and the Inaugural Director of the UBTECH Sydney Artificial Intelligence Centre, The University of Sydney. He mainly applies statistics and mathematics to artificial intelligence and data science. His research results have expounded in one monograph and over 200 publications at prestigious journals and prominent conferences, such as the IEEE T-PAMI, T-IP, T-NNLS, IJCV, JMLR, NIPS, ICML, CVPR, ICCV, ECCV, and ICDM; and ACM SIGKDD, with several best paper awards, such as the Best Theory/Algorithm Paper Runner Up Award in IEEE ICDM'07, the Best Student Paper Award in IEEE ICDM'13, the Distinguished Paper Award in the 2018 IJCAI, the 2014 ICDM 10-Year Highest-Impact Paper Award, and the 2017 IEEE Signal Processing Society Best Paper Award. He is a Fellow of the Australian Academy of Science, AAAS, IAPR, OSA and SPIE.

Dan Feng (A'17) received the B.E., M.E., and Ph.D. degrees in computer science and technology from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 1991, 1994, and 1997, respectively. She is currently a Professor and the Dean of the School of Computer Science and Technology, HUST. Her research interests include computer architecture, non-volatile memory technology, distributed and parallel file systems, and massive storage systems. She has over 100 publications in major journals and international conferences, including IEEE TC, IEEE TPDS, ACM-TOS, FAST, USENIX ATC, EuroSys, ICDCS, HPDC, SC, ICS, IPDPS, DAC, and DATE. She is a member of the Association for Computing Machinery, and the Chair of the Information Storage Technology Committee, Chinese Computer academy. She has served on the program committees of multiple international conferences, including SC 2011 and 2013, and MSST 2012 and 2015.

Wenjun (Kevin) Zeng (M'97–SM'03–F'12) received the B.E. degree from Tsinghua University, the M.S. degree from the University of Notre Dame, and the Ph.D. degree from Princeton University. He was with PacketVideo Corp., Sharp Labs of America, Bell Labs, and Panasonic Technology. He was with the University of Missouri from 2003 to 2016, where he is a Full Professor. He is currently a Principal Research Manager and a member of the Senior Leadership Team, Microsoft Research Asia. He has been leading the video analytics research empowering the Microsoft Cognitive Services and Azure Media Analytics Services since 2014. He has contributed significantly to the development of international standards (ISO MPEG, JPEG2000, and OMA). His current research interests include mobile-cloud media computing, computer vision, social media analysis, and multimedia communications and security. He was a recipient of several best paper awards. He served as the Steering Committee Chair for the IEEE ICME in 2010 and 2011, and has served as the General Chair or TPC Chair for several IEEE conferences (e.g., ICME'2018 and ICIP'2017). He was an Associate Editor-in-Chief of the *IEEE Multimedia Magazine*, and an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT), the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, and the IEEE TRANSACTIONS ON MULTIMEDIA (TMM). He was a Special Issue Guest Editor for the Proceedings of the IEEE, TMM, ACM TOMCCAP, TCSVT, and IEEE *Communications Magazine*. He was on the Steering Committee of the IEEE TRANSACTIONS ON MOBILE COMPUTING and the IEEE TMM.



Zhangyang Wang received the B.E. degree from the University of Science and Technology of China in 2012 and the Ph.D. degree from the Electrical and Computer Engineering Department, University of Illinois at Urbana–Champaign, working with Prof. Thomas S. Huang. He was a former Research Intern with Microsoft Research (summer 2015), Adobe Research (summer 2014), and the United States Army Research Laboratory (summer 2013). He is currently an Assistant Professor of computer science and engineering with Texas A&M University. He has co-authored over 50 papers, and published several books and chapters. He holds three patents. His research has been addressing machine learning, computer vision, and multimedia signal processing problems, as well as their interdisciplinary applications, using advanced feature learning and optimization techniques. He was a recipient of 20 research awards and scholarships.