Single Image Deraining: A Comprehensive Benchmark Analysis

Siyuan Li[1]*, Iago Breno Araujo[2]*, Wenqi Ren[3]†, Zhangyang Wang[4]†, Eric K. Tokuda[2],
Roberto Hirata Junior[2], Roberto Cesar-Junior[2], Jiawan Zhang[1], Xiaojie Guo[1], Xiaochun Cao[3]
[1]*Tianjin University* [2]*University of Sao Paulo* [3]*SKLOIS, IIE, CAS* [4]*Texas A&M University*

https://github.com/lsy17096535/Single-Image-Deraining

Abstract

Numerous single image deraining algorithms have been recently proposed. However, these algorithms are mainly evaluated using certain type of synthetic images, assuming a specific rain model, plus a few real images. It is thus unclear how these algorithms would perform on rainy images acquired "in the wild" and how we could gauge the progress in the field. This paper aims to bridge this gap. We present a comprehensive study and evaluation of existing single image deraining algorithms, using a new large-scale benchmark consisting of both synthetic and real-world rainy images of various rain types. This dataset highlights diverse rain models (rain streak, rain drop, rain and mist), as well as a rich variety of evaluation criteria (full- and no-reference objective, subjective, and task-specific) Our evaluation and analysis indicate the performance gap between synthetic rainy images and real-world images and allow us to better identify the strengths and limitations of each method as well as future research directions.

1. Introduction

Images captured in rainy days suffer from noticeable degradation of scene visibility. The goal of single image deraining algorithms is to generate sharp images from a rainy image input. Image deraining can potentially benefit both the human visual perception quality of images, and many computer vision applications, such as outdoor surveillance systems and intelligent vehicles.

The recent years have witnessed significant progress in single image deraining. The progress in this field can be attributed to various natural image priors [1, 2, 3, 4, 5] and deep convolutional neural network (CNN)-based models [6, 7, 8]. However, a fair comprehensive study of the problem, the existing algorithms, and the performance metrics have been absent so far, which is the goal of this paper.

*The first two authors contributed equally.
†indicates corresponding author.

1.1. Rainy Image Formulation Models

As a complicated atmospheric process, rain could cause several different types of visibility degradations, due to a magnitude of environmental factors including raindrop size, rain density, and wind velocity [9]. When a rainy image is taken, the visual effects of rain on that digital image further hinges on many camera parameters, such as exposure time, depth of field, and resolution [10]. Most existing deraining works assume one rain model (usually rain streak), which might have oversimplified the problem. We group existing rain models in literature into three major categories: **rain streak**, **raindrop**, as well as **rain and mist**.

A rain streak image $\mathbf{R}_s$ can be modeled as a linear superimposition of the clean background scene $\mathbf{B}$ and the sparse, line-shape rain streak component $\mathbf{S}$:

$$\mathbf{R}_s = \mathbf{B} + \mathbf{S}. \tag{1}$$

Rain streaks $\mathbf{S}$ accumulated throughout the scene reduce the visibility of the background $\mathbf{B}$. This is the most common model assumed by the majority of deraining algorithms.

Adherent raindrops [11] that fall and flow on camera lenses or a window glasses can obstruct and/or blur the background scenes. The raindrop degraded image $\mathbf{R}_d$ can be modeled as the combination of the clean background $\mathbf{B}$, and the blurry or obstruction effect of the raindrops $\mathbf{D}$ in scattered, small-sized local coherent regions:

$$\mathbf{R}_d = (1 - \mathbf{M}) \odot \mathbf{B} + \mathbf{D}. \tag{2}$$

$\mathbf{M}$ is a binary mask and $\odot$ means element-wise multiplication. In the mask, a pixel x is part of a raindrop region if $\mathbf{M}(x) = 1$, and otherwise belongs to the background.

Further, rainy images often contain both rain and mist in real cases [12]. In addition, distant rain streaks accumulated throughout the scene reduce the visibility in a manner more similarly to fog, creating a mist-like phenomenon in the image background. Concerning this, we can define the rain and mist model for the captured image $\mathbf{R}_m$, based on

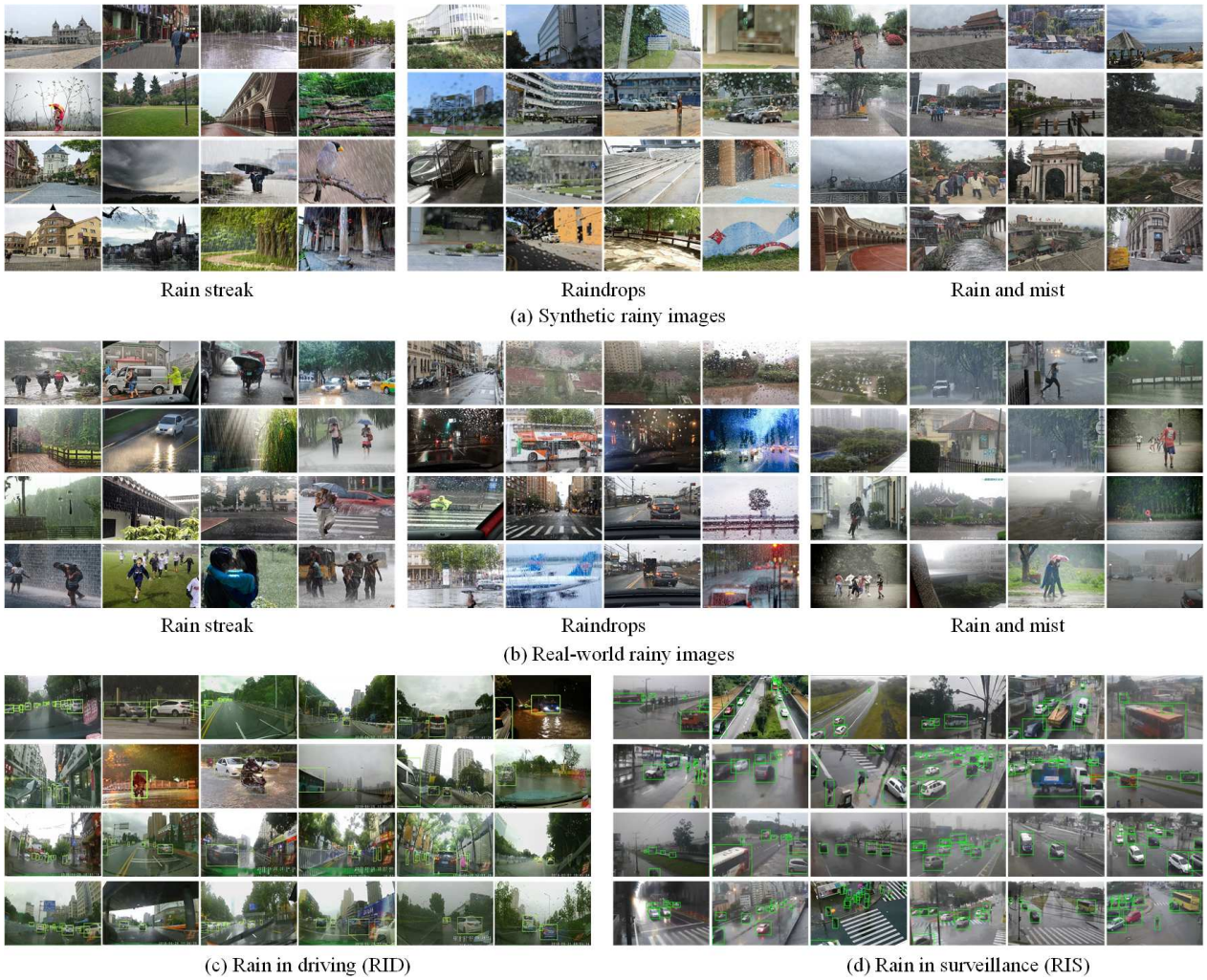


Figure 1. Example images from the MPID dataset. The proposed dataset contains both synthetic and real-wold rainy images of rain streak, raindrops, and rain & mist. In addition, we also annotate two sets of real-world images with object bounding boxes from autonomous driving and video surveillance scenarios.

a composition of the rain streak model and the atmospheric scattering haze model [13]:

$$\mathbf{R}_m = \mathbf{B} \odot t + A\,(1-t) + \mathbf{S}, \tag{3}$$

where $\mathbf{S}$ is the rain streak component; t and A are the transmission map and atmospheric light that determines the fog/mist component, respectively.

1.2. Our Contribution

Regardless of what rain models to follow, image deraining is a heavily ill-posed problem. Despite many impressive methods published in recent few years, the lack of a large dataset and algorithm benchmarking makes it difficult to evaluate the progress made, and how practically useful those algorithms are. There are several unclear and unsatisfactory aspects of current deraining algorithm development, including but not limited to: i) the modeling of rain is oversimplified, i.e., each method considers and is evaluated with one type of rain only, e.g., rain streak; ii) most quantitative results are reported on synthetic images, which often fail to capture the complexity and characteristics of real rain; iii) as a result of the last point, the evaluation metrics have been mostly limited to (the full-reference) PSNR and SSIM for image restoration purposes. They may become poorly related when it comes to other task purposes, such as human perception quality [14] or computer vision utility [15].

In this paper, we aim to systematically evaluate state-of-the-art single image deraining methods, in a comprehensive and fair setting. To this end, we construct a large-scale benchmark, called Multi-Purpose Image Deraining (MPID). An overview of MPID could be found in Table 1, and image examples are displayed in Figure 1. Compared with existing synthetic sets, the MPID dataset covers a much larger diversity of rain models (rain streak, raindrop, and rain and mist), including both synthetic and real-world

images for evaluation, and featuring diverse contents and sources (for real rainy images). In addition, as the first-of-its-kind efforts in image deraining, we have annotated two sets of real-world rainy images with object bounding boxes from autonomous driving and video surveillance scenarios, respectively, for task-specific evaluation.

Using the MPID benchmark, we evaluate six state-of-the-art single image deraining algorithms. We adopt a wide range of full-reference metrics (PSNR and SSIM), no-reference metrics (NIQE, BLIINDS-II, and SSEQ), as well as human subjective scores to thoroughly examine the performance of image deraining methods. A human subjective study is also conducted. Furthermore, as image deraining might be expected as a preprocessing step for mid- and high-level computer vision tasks, we also evaluate current algorithms in terms of their impact on subsequent object detection tasks, as a "task-specific" evaluation criterion. We reveal the performance gap in various aspects, when these algorithms are applied on synthetic and real images. By extensively comparing the state-of-the-art single image deraining algorithms on the MPID dataset, we gain insights into new research directions for image deraining.

2. Related Work

2.1. Overview of Deraining Algorithms

Multi-frame based approaches: Early methods often require multiple frames to deal with the deraining problem [4, 16, 17, 18, 19, 5, 20, 11]. Garg and Nayar [21] proposed a rain streak detection and removal method from a video by taking the average intensity of the detected rain streaks from the previous and subsequent frames. [10] further improved the performance by selecting camera parameters without appreciably altering the scene appearance. However, those methods are not applicable to single image deraining.

Prior based algorithms: Many deraining methods capitalize on clean image or rain type priors to remove rain [22, 1, 23, 24, 25]. Kang et al. [2] decomposed an input image into its low and high frequency components. Then they separated the rain streak frequencies from the high frequency layer via sparse coding. Zhu et al. [26] introduced a rain removal method based on the prior that rain streaks typically span a narrow range of directions. Chen and Hsu [3] decomposed the background and rain streak layers based on low-rank priors. Li et al. [27] use patch-based priors for both the clean background and rain layers in the form of Gaussian mixture models. All of the above approaches rely on good (and relatively simple) crafted priors. As a result, they tend to have unsatisfactory performances on real images with complicated scenes and rain forms.

Data-driven CNN models: Recently, CNNs have achieved dominant success for image restoration [28, 29] including single image deraining [30, 31]. Fu et al. [6] proposed a deep detail network (DDN) for removing rain from single images with detailed preserved. Yang et al. [32] presented a CNN based method to jointly detect and remove rain streaks, using a multi-stream network to capture the rain streak component. A density-aware multi-stream densely connected convolutional neural network was introduced in [8] for joint rain density estimation and image deraining. Qian et al. [7] addressed a different problem of removing raindrops from single images, using visual attention with a generative adversarial network (GAN). Despite the progress of deep-learning-based approaches compared with prior-based rain removal methods, their performance hinge on the synthetic training data, which may become problematic if real rainy images show a domain mismatch.

2.2. Datasets

Several datasets were used to measure and compare the performance of image deraining algorithms. Li et al. [27] introduced a set of 12 images using photo-realistic rendering techniques. Zhang et al. [33] synthesized a set of training and testing images with rain streak, using the same way in [27]. The training set consists of 700 images and the testing set consists of 100 images. In addition, [33] also collects a dataset of 50 real-world rainy images downloaded from the web for qualitative visual comparison. [7] released a set of clean and rain-drop corrupted image pairs, using a special lens equipment. However, existing datasets are either too small in scale and limited to one rain type (streak or drop), or lack sufficient real-world images for diverse evaluations. Besides, none of them has any semantic annotation nor consider any subsequent task performance.

3. New Benchmark: Multi-Purpose Image Deraining (MPID)

We present a new benchmark as a comprehensive platform, for evaluating single image deraining algorithms from a variety of perspectives. Our evaluation angles range from traditional PSNR/SSIM, to no-reference perception-driven metrics and human subjective quality, to "task-driven metrics" [15, 34] indicating how well a target computer vision task can be performed on the derained images. Fitting those purposes, we generate/collect images in large scale, from both synthesis and real world sources, covering diverse real-life scenes, and annotate them when needed. The new benchmark, dubbed *Multi-Purpose Image Deraining* (**MPID**), is introduced below in details. An overview of MPID can be found in Table 1.

3.1. Training Sets: Three Synthesis Models

Following the three rain models in Section 1.1, we create three training sets, named *Rain streak (T), Rain drop (T)* and *Rain and mist (T)* sets (*T* short for "training"), respectively. All three sets are synthesized in controlled set-

Table 1. Overview of the proposed MPID dataset.

Training Set				
Subset	*Number of Images*	*Real/synthetic*	*Annotations*	*Metrics*
Rain streak (T)	2400 (pairs)	synthetic	No	/
Raindrop (T)	861 (pairs)	synthetic	No	/
Rain and mist (T)	700 (pairs)	synthetic	No	/
Testing Set				
Subset	*Number of Images*	*Real/synthetic*	*Annotations*	*Metrics*
Rain streak (S)	200 (pairs)	synthetic	No	PSNR, SSIM, NIQE, BLIINDS-II, SSEQ
Rain streak (R)	50	real	No	NIQE, BLIINDS-II, SSEQ
Raindrop (S)	149 (pairs)	synthetic	No	PSNR, SSIM, NIQE, BLIINDS-II, SSEQ
Raindrop (R)	58	real	No	NIQE, BLIINDS-II, SSEQ
Rain and mist (S)	70 (pairs)	synthetic	No	PSNR, SSIM, NIQE, BLIINDS-II, SSEQ
Rain and mist (R)	30	real	No	NIQE, BLIINDS-II, SSEQ
Task-Driven Evaluation Set				
Subset	*Number of Images*	*Real/synthetic*	*Annotations*	*Metrics*
RID	2496	real	Yes (bounding boxes)	mAP
RIS	2048	real	Yes (bounding boxes)	mAP

tings from clean images.[1]. All clean images used are collected from the web, and we specifically pick those outdoor rain-free, haze-free photos taken in cloudy daylight, so that the synthesized rainy images look more realistic in terms of lighting condition (for example, there will be no rainy photo in a sunny daylight background).

The Rain streak (T) set contains 2,400 pairs of clean and rainy images, where the rainy images are generated from the clean ones using (1), with the identical protocol and hyper-parameters to [27, 33]. The Rain drop (T) set was borrowed from [7]'s released training set consisting of 861 pairs of clean and rain-drop corrupted images, upon their authors' consent. The Rain and mist (T) set is synthesized by first adding haze using the atmospheric scattering model: for each clean image, we estimate depth using the algorithm in [35, 36] as recommended by [37], set different atmospheric lights A by choosing each channel uniformly randomly between $[0.7, 1.0]$, and select β uniformly at random between $[0.6, 1.8]$. Then from the synthesized hazy version, we further add rain streaks in the same way as Rain streak (T). We end up with 700 pairs for the Rain and mist (T) set.

3.2. Testing Sets: From Synthetic To Real

Corresponding to three training sets, we generate three synthetic testing set in the same way: denoted as *Rain streak (S), Rain drop (S)*, and *Rain and mist (S)* (*S* short for "synthetic testing"), consisting of 200, 149, and 70 pairs, respectively. On each testing set, we evaluate the restoration performance of deraining algorithms, using classical PSNR and SSIM metrics. Further, to predict the derained image's perceptual quality to human viewers, we introduce the usage of three no-reference IQA models: Naturalness Image Quality Evaluator (NIQE) [38], spatial-spectral entropy-based quality (SSEQ) [39], and blind image integrity notator using DCT statistics (BLIINDS-II) [40], to complement the shortness of PSNR/SSIM. NIQE is a well-known no-reference image quality score to indicate the perceived "naturalness" of an image: a smaller score indicates better perceptual quality. The score of SSEQ and BLIINDS-II that we used range from 0 (worst) to 100 (best).[2]

Besides the three above synthetic test sets, we collect three sets of real-world images, that fall into each of three defined rain categories, to evaluate the deraining algorithms' real-world generalization. The three sets, denoted as *Rain streak (R), Raindrop (R)*, and *Rain and mist (R)* (*R* short for "real-world testing"), are collected from the Internet and are carefully inspected to ensure that images in each set fit the pre-defined rain type well. Due to the unavailability of ground truth clean images in real world, we evaluate NIQE, SSEQ, and BLIINDS-II on the three real-world sets. In addition, we also pick a small set of real-world images for human subjective rating of derained results.

3.3. Task-Driven Evaluation Sets

As pointed out by several recent works [41, 15, 42, 43], the performance of high-level computer vision tasks, such as object detection and recognition, will deteriorate in the presence of various sensory and environmental degradations. While deraining could be used as pre-processing for many computer vision tasks executed in the rainy conditions, there has been no systematical study on deraining algorithms' impact on those target tasks. We consider the resulting task performance after deraining as an indirect indicator of the deraining quality. Such a "task-driven" evaluation way has received little attention and can have great implications for outdoor applications.

To conduct such task-driven evaluations, realistic annotated datasets are necessary. To our best knowledge, there has been no dataset available serving the purpose of evaluat-

[1]Note that for Rain drop (T), the data generation used physical simulation [7] , i.e., with/without lens, rather than algorithm simulation.

[2]Note that in [39] and [40], a smaller SSEQ/BLIINDS-II score indicates better perceptual quality. We reverse the two scores (100 minus) to make their trends look consistent to full-reference metrics: in our tables the bigger the two values, the better the perceptual quality. We did not do the same to NIQE, because NIQE has no bounded maximum value.

Table 2. Object Statistics in RID and RIS sets.

Categories	*Car*	*Person*	*Bus*	*Bicycle*	*Motorcycle*
RID Set	7332	1135	613	268	968
Categories	*Car*	*Person*	*Bus*	*Truck*	*Motorcycle*
RIS Set	11415	2687	488	673	275

ing deraining algorithms in task-driven ways. We therefore collect two sets by our own: a *Rain in Driving* (**RID**) set collected from car-mounted cameras when driving in rainy weathers, and a *Rain in surveillance* (**RIS**) set collected from networked traffic surveillance cameras in rainy days.

For either set, we annotate object bounding boxes, and evaluate object detection performance after applying deraining. A summary with object statistics on both RID and RIS sets can be found in Table 2. The two sets differ in many ways: rain type, image quality, object size and angle, and so on. They are representative of real application scenarios where deraining may be desired.

Rain in Driving (RID) Set This set contains 2,495 real rainy images from high-resolution driving videos. As we observe, its rain effect is closest to "raindrops" on camera lens. They were captured in diverse real traffic locations and scenes during multiple drives. We label bounding boxes for selected traffic objects: car, person, bus, bicycle, and motorcycle, that commonly appear on the roads of all images. Most images are of 1920×990 resolution, with a few exceptions of 4023×3024 resolution.

Rain in Surveillance (RIS) Set This set contains 2,048 real rainy images from relatively lower-resolution surveillance video cameras. They were extracted from a total of 154 surveillance cameras in daytime, ensuring diversity in content (for example, we do not consider frames too close in time). As we observe, its rain effect is closest to "rain and mist" (many cameras have mist condensation during rain, and the low resolution will also cause more foggy effects). We selected and annotated the most common objects in the traffic surveillance scenes: car, person, bus, truck, and motorcycle. The vast majority of cameras have the resolution of 640×368, with a few exceptions of 640×480.

4. Experimental Comparison

We evaluate six representative state-of-the-art algorithms on MPID: Gaussian mixture model prior (GMM) [27], JOint Rain DEtection and Removal (JORDER) [32], Deep Detail Network (DDN) [6], Conditional Generative Adversarial Network (CGAN) [33], Density-aware Image Deraining method using a Multistream Dense Network (DID-MDN) [8], and DeRaindrop [7]. All except GMM are state-of-the-art CNN-based deraining algorithms.

Evaluation Protocol. The first five models are specifically developed for removing rain streaks, while the last one targets at removing rain drops. Therefore, we compare them for rain streak sets. Since DeRaindrop is the only recent published method for raindrop removal, to provide more baselines for its performance, we also re-train and evaluate the other five models on the raindrop sets. Finally, since no published method was targeted for removing rain and mist together, we create a cascaded pipeline, by first running each of the five rain streak removal algorithms, followed by feeding into a pre-trained MSCNN dehazing network [28]. MSCNN was chosen because recent dehazing studies [15, 48] endorsed it both to produce the best human-favorable, artifact-free dehazing results, and to benefit subsequent high-level task in haze most. Such cascaded pipeline can be tuned from end to end, and we freeze the MSCNN part during tuning in order to focus on comparing deraining components. All models will be re-trained on the corresponding MPID training set, when evaluated on a certain rain type.

4.1. Objective Comparison

We first compare the derained results on the synthetic images using two full-reference (PSNR and SSIM) and three no-reference metrics (NIQE, SSEQ, and BLIINDS-II). As seen from Table 3, the results have high consensus levels on synthetic data. First, DDN is the obvious winner on the rain streak (S) set, followed by JORDER; the same two methods also perform consistently the best on the rain and mist (S) set. Second, DerainDrop performs the best on the rain drop (S) set, especially significantly surpassing the others in terms of PSNR and SSIM, showing that its specific structure indeeds suits this problem. Other rain streak removal models seem to even hurt PSNR, SSIM and BLINDS-II, compared to the degraded images.

The effectiveness of the winners can be ascribed to the two-step strategy of rain detection and removal. We note that DDN focuses on high frequency details during training stage, while JORDER also first detects the locations of rain streak, then removes rain based on the estimated rain streak regions. Coincidentally, DeRaindrop also uses an attentive generative network to generate raindrops mask first then derain images capitalizing on the masks. Therefore, removing background interference and attentively focusing on rain regions seem to be the main reason of the winners.

We then show the derained results on the real-world images in Table 4, using three no-reference metrics (NIQE, SSEQ, and BLIINDS-II). The rain streak (R) and raindrop (R) sets show consistent results with their synthetic cases: JORDER and DDN rank top-two on the former, while DerainDrop still dominates on the raindrop set. However, some different tendency is observed on the rain and mist (R) set: CGAN becomes the dominant winner on those real images, outperforming both DDN and JORDER with large margins. As we observed, since CGAN is most free of physical priors or rain type assumptions, it has the largest flexibility for re-training to fit different data. Its results is also most photo-realistic due to the adversarial loss. Additionally, the result might also suggest a larger domain gap

Table 3. Average full- and no-reference evaluations results on synthetic rainy images. We use **bold** and underline to indicate the best and suboptimal performance, respectively.

	Degraded	GMM [27]	JORDER [32]	DDN [6]	CGAN [33]	DID-MDN [8]	DeRaindrop [7]
	rain streak						
PSNR	25.95	26.88	26.26	**29.39**	21.86	26.80	/
SSIM	0.7565	0.7674	**0.8089**	0.7854	0.6277	0.8028	/
SSEQ	70.24	67.46	73.70	**75.95**	70.02	60.05	/
NIQE	5.4529	4.4248	4.2337	**3.9834**	4.6189	4.8122	/
BLINDS-II	78.89	75.95	84.21	**91.71**	79.29	67.90	/
	raindrops						
PSNR	25.40	24.85	27.52	25.23	21.35	24.76	**31.57**
SSIM	0.8403	0.7808	0.8239	0.8366	0.7306	0.7930	**0.9023**
SSEQ	78.48	64.73	**84.32**	77.62	63.15	58.42	72.42
NIQE	3.8126	5.1098	4.3278	4.1462	**3.3551**	4.1192	5.0047
BLINDS-II	92.50	75.95	88.05	91.95	73.85	64.70	**96.45**
	rain and mist						
PSNR	26.84	29.37	30.37	**32.98**	22.44	28.77	/
SSIM	0.8520	0.8960	0.9262	**0.9350**	0.7636	0.8430	/
SSEQ	**72.37**	65.39	70.55	69.80	68.71	65.33	/
NIQE	3.4548	3.2117	2.8595	2.9970	**2.8336**	3.0871	/
BLINDS-II	82.95	74.90	83.75	**85.75**	80.20	76.35	/

Table 4. Average no-reference evaluations results of derained results on real rainy images. We use **bold** and underline to indicate the best and suboptimal performance except the degraded inputs, respectively.

	Degraded	GMM [27]	JORDER [32]	DDN [6]	CGAN [33]	DID-MDN [8]	DeRaindrop [7]
	rain streak						
SSEQ	65.77	61.63	**64.00**	63.51	59.32	55.11	/
NIQE	3.5236	**3.2117**	3.5371	3.5811	3.5374	5.1255	/
BLINDS-II	78.04	75.54	82.62	**85.81**	78.42	66.65	/
	raindrops						
SSEQ	78.23	64.77	69.26	67.62	62.18	60.65	**79.83**
NIQE	3.8229	4.3801	3.6579	3.8290	4.4692	4.5631	**3.5953**
BLINDS-II	84.46	71.21	80.04	77.75	66.29	66.63	**87.13**
	rain and mist						
SSEQ	73.86	59.51	65.18	64.56	**70.04**	63.85	/
NIQE	3.2602	4.4808	3.3238	3.7261	**2.9532**	3.2260	/
BLINDS-II	84.00	62.70	78.62	81.67	**84.91**	76.08	/

between synthetic and real rain and mist data.

4.2. Subjective Comparison

We next conduct a human subjective survey to evaluate the performance of image deraining algorithms. We follow a standard setting that fits a Bradley-Terry model [49] to estimate the subjective score for each method so that they can be ranked, with the exactly same routine as described in previous similar works [15]. We select 10 images from Rain streak (R), 6 images from Rain drop (R), and 11 images from Rain and mist (R), taking all possible care to ensure that they have very diverse contents and quality. Each rain streak or rain & mist image is processed with each of the five deraining algorithms (except DerainDrop), and the five deraining results, together with the original rainy image, are sent for pairwise comparison to construct the winning matrix. For a rain drop image, the procedure is the same except that it will be processed by all six methods. We collect the pair comparison results of human subject studies from 11 human raters. Despite the relatively small numbers of raters, we observed good consensus and small inter-person variances among raters, on same pairs' comparison results, which make scores trustworthy.

The subjective scores are reported in Table 5. Note that we did not normalize the scores: so it is the score rank rather than the absolute score values that makes sense here. On the rain streak images, it seems that most human viewers prefer CGAN first, and then DDN. As shown in the first row of Figure 2, the derained result generated by CGAN is more smooth than others. On the raindrop images, it is somehow to our surprise that DerainDrop is not favored by users; instead, the non-CNN-based GMM method, which showed no advantage under previous objective metrics, was highly preferred by users. We conjecture that the patch-based Gaussian mixture prior can treat and remove both rain streaks and raindrops as "outliers", and is less sensitive to training/testing data domain difference. Finally on the rain and mist images, DID-MDN receives the highest scores, while CGAN is next to it. This is mainly thanks to incorporating th rain-density subnetwork or GAN, that can provide more information of the scene context and hence improve generalization to complex rain conditions.

Table 5. Average subjective scores of derained results on 10 real images.

	rainy	GMM [27]	JORDER [32]	DDN [6]	CGAN [33]	DID-MDN [8]	DeRaindrop [7]
rain streak	0.64	0.80	0.91	1.15	**1.26**	0.97	–
raindrops	0.80	**1.14**	0.75	0.83	0.85	0.95	0.80
rain and mist	0.44	1.00	0.70	0.90	1.22	**1.40**	–

Table 6. Detection results (mAP) on the RID and RIS sets. Detailed results for each class can be found in the supplementary material.

		Rainy	JORDER [32]	DDN [6]	CGAN [33]	DID-MDN [8]	DeRaindrop [7]
RID	FRCNN [44]	16.52	16.97	18.36	23.42	16.11	15.58
	YOLO-V3 [45]	27.84	26.72	26.20	23.75	24.62	24.96
	SSD-512 [46]	17.71	17.06	16.93	16.71	16.70	16.69
	RetinaNet [47]	23.92	21.71	21.60	19.28	20.08	19.73
RIS	FRCNN [44]	22.68	21.41	20.76	18.02	18.93	19.97
	YOLO-V3 [45]	23.27	20.45	21.80	18.71	21.50	20.43
	SSD-512 [46]	8.19	7.94	8.29	7.10	8.21	8.13
	RetinaNet [47]	12.81	10.71	10.39	9.36	10.33	10.85

While we are in the process of recruiting more human raters to solidify our subject score results more, our results seem to be consistent so far, and might in turn imply that off-the-shelf no-reference perceptual metrics (SSEQ, NIQE, BLINDS-II) do not align well with the real human perception quality of deraining results. In fact, recent works [50] already discovered similar misalignments, when applying standard no-reference metrics to estimating defogging perceptual quality, and proposed fog-specific metrics. Similar efforts have not been found for deraining yet, and we expect this worthy effort to take place in near future.

4.3. Task-driven Comparison

We first apply all deraining algorithms except GMM[3], to pre-processing the two task-driven testing sets. Due to their different rain characteristics, for the RID set, we use deraining algorithms trained on the *rain and mist* case; for the RIS set, we use deraining algorithms trained on the *raindrop* case. We visually inspected the derained results and found the rains to be visually attenuated after applying the selected deraining algorithms. We show some derained results on the RID and RIS sets in the supplementary material.

We then study object detection performance on the derained sets, using several state-of-the-art object detection models: Faster R-CNN (FRCNN) [44], YOLO-V3 [45], SSD-512 [46], and RetinaNet [47]. Finally, we compare all deraining algorithms via the mean Average Precision (mAP) results achieved. It is important to note that our primary goal is not to optimize detection performance in rainy days, but to use a strong detection model as a fixed, fair metric on comparing deraining performance from a complementary perspective. In this way, the object detectors should not be adapted for rainy or derained images, and we use all authors' pre-trained models on MS COCO. The underlying hypothesis is: i) an object detector trained on clean natural images will perform the best, when the input is also from the clean image domain or close; ii) for detection in rain, the better the rain is removed, the better an object detection model (trained on clean images) will then perform. Such task-specific evaluation philosophy follows [34, 15].

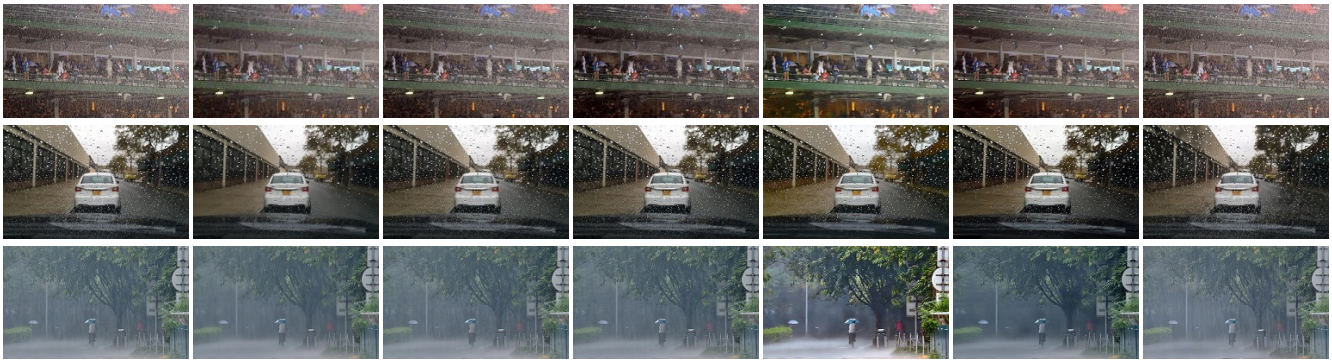
Table 6 reports the mAP results comparison for different deraining algorithms, achieved using four different detection models, on both RID and RIS sets. We find that quite aligned conclusions could be drawn from the two sets.

Perhaps surprisingly at the first glance, we find that almost **all existing deraining algorithms will deteriorate the detection performance compared to directly using the rainy images**[4], for YOLO-V3, SSD-512, and RetinaNet. Our observation concurs the conclusion of another recent study (on dehazing) [51]: since those deraining algorithms were not trained/optimized towards the end goal of object detection, they are unnecessary to help this goal, and the deraining process itself might have lost discriminative, semantically meaningful true information.

Both results on RID and RIS sets in Table 6 show that YOLO-V3 achieves best detection performance, independently of deraining algorithms applied. Figure 3 shows detections using YOLO-V3 on the respectives rainy images and their derained results for all deraining algorithms considered in this comparison. Since both RID and RIS have many small objects due to their relative long distance from the camera, we believe that here YOLO-V3 benefits from its new multi-scale prediction structure, that is known to improve small object detection dramatically [45]. We further notice a fairly weak correlation between the mAP results with the no-reference evaluation results of the derained images: see supplementary for more details.

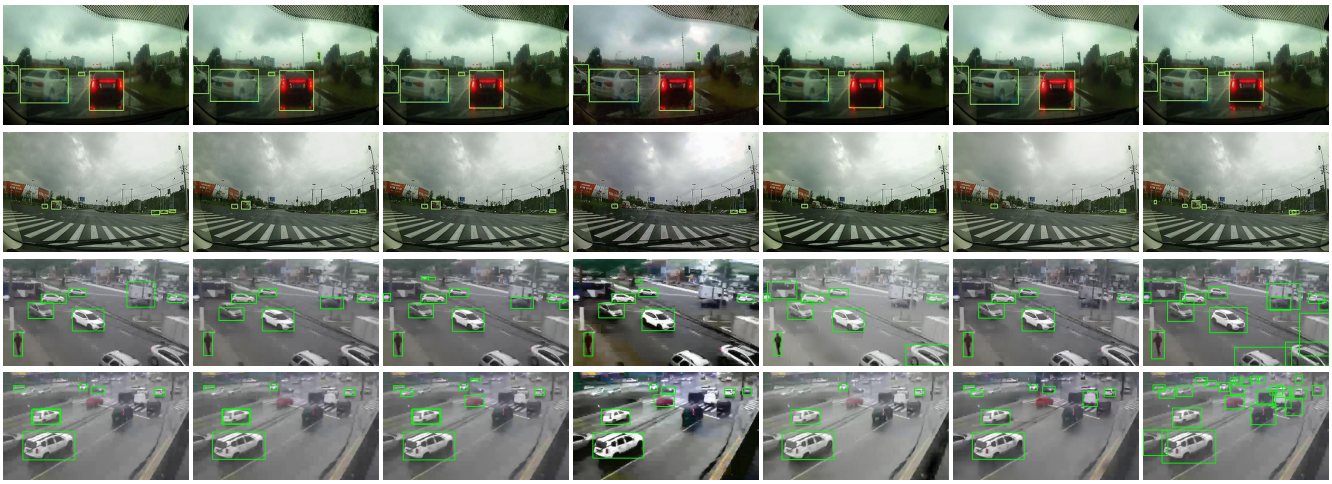
[3]We did not include GMM for the two sets, because (1) it did not yield promising results when we tried to apply it to (part of) the two sets; (2) it runs very slow, given we have two large sets.

[4]The only exception is FRCNN on the RID set. However, its overall mAP result is the worst compared to the other three. That implies a strong domain mismatch, suggesting that FRCNN results might not be as reliable an indicator for RID deraining performance as the other three.



(a) Rainy input (b) GMM [27] (c)JORDER [32] (d) DDN [6] (e) CGAN [33] (f) DID-MDN [8] (g) DeRaindrop [7]

Figure 2. Examples of derained results on real images: rain streak (first row), raindrop (second row), and rain and mist (third row).



(a) Rainy input (b) JORDER [32] (c) DDN [6] (d) CGAN [33] (e) DID-MDN [8] (f) DeRaindrop [7] (g) Ground-truths

Figure 3. Visualization of object detection results after applying different deraining algorithms on two images (first two rows) from the RID dataset and two examples (last two rows) from the RIS dataset.

5. Conclusions and Future Work

This paper proposes a new large-scale benchmark and presents a thorough survey of state-of-the-art single image deraining methods. Based on our evaluation and analysis, we present overall remarks and hypotheses below, which we hope can shed some light on future deraining research:

- Rain types are diverse and call for specialized models. Certain models or components are revealed to be promising for specific rain types, e.g., rain detection /attention, GANs, and priors like patch-level GMM. We also advocate a combination of appropriate priors and data-driven methods.
- There is no single best deraining algorithm for all rain types. To deal with the real complicated, varying rains, one might need consider a mixture model of experts. Another practically useful direction is to develop scene-specific deraining, e.g., for traffic views.
- There is also no single best deraining algorithm under all metrics. When designing a deraining algorithm, one needs be clear about its end purpose. Moreover, classical perceptual metrics themselves might be problematic to evaluate deraining. Developing new metrics could be as important as new algorithms.
- Algorithms trained on synthetic paired data may generalize poorly to real data, especially on complicated rain types such as rain and mist. Unpaired training [52] on all real data could be interesting to explore.
- No existing deraining method seems to directly help detection. That may encourage the community to develop new robust algorithms to account for high-level vision problems on real-world rainy images. On the other hand, to realize the goal of robust detection in rain does not have to adopt a de-raining preprocessing; there are other domain adaptation type options, e.g., [53], which we will discuss in future work.

Acknowledgments. This work is supported in part by the National Natural Science Foundation of China (No. 61802403, U1605252U1736219), Beijing Natural Science Foundation (No. L182057), and CCF-DiDi GAIA (YF20180101). The work of Z. Wang is supported by the US National Science Foundation under Grant 1755701.

References

[1] Shao-Hua Sun, Shang-Pu Fan, and Yu-Chiang Frank Wang. Exploiting image structural similarity for single image rain removal. In *IEEE International Conference on Image Processing*, pages 4482–4486, 2014. 1, 3

[2] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Transactions on Image Processing*, 21(4):1742, 2012. 1, 3

[3] Yi-Lei Chen and Chiou-Ting Hsu. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *IEEE International Conference on Computer Vision*, pages 1968–1975, 2013. 1, 3

[4] Xiaopeng Zhang, Hao Li, Yingyi Qi, Wee Kheng Leow, and Teck Khim Ng. Rain removal in video by combining temporal and chromatic properties. In *IEEE International Conference on Multimedia and Expo*, 2006. 1, 3

[5] Jérémie Bossu, Nicolas Hautière, and Jean-Philippe Tarel. Rain or snow detection in image sequences through use of a histogram of orientation of streaks. *International journal of computer vision*, 93(3):348–367, 2011. 1, 3

[6] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 1, 3, 5, 6, 7, 8

[7] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *CVPR*, 2018. 1, 3, 4, 5, 6, 7, 8

[8] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1, 3, 5, 6, 7, 8

[9] Sudipta Mukhopadhyay and Abhishek Kumar Tripathi. Combating bad weather part i: Rain removal from video. *Synthesis Lectures on Image, Video, and Multimedia Processing*, 7(2):1–93, 2014. 1

[10] Kshitiz Garg and Shree K Nayar. When does a camera see rain? In *IEEE International Conference on Computer Vision*, 2005. 1, 3

[11] Shaodi You, Robby T Tan, Rei Kawakami, Yasuhiro Mukaigawa, and Katsushi Ikeuchi. Adherent raindrop modeling, detectionand removal in video. *IEEE transactions on pattern analysis and machine intelligence*, 38(9):1721–1733, 2016. 1, 3

[12] Zheng Zhang, Wu Liu, Huadong Ma, and Xinchen Liu. Going clear from misty rain in dark channel guided network. 1

[13] Earl J McCartney. Optics of the atmosphere: scattering by molecules and particles. *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976. 2

[14] Wei-Sheng Lai, Jia-Bin Huang, Zhe Hu, Narendra Ahuja, and Ming-Hsuan Yang. A comparative study for single image blind deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1709. 2

[15] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019. 2, 3, 4, 5, 6, 7

[16] Abhishek Kumar Tripathi and Sudipta Mukhopadhyay. Removal of rain from videos: a review. *Signal, Image and Video Processing*, 8(8):1421–1430, 2014. 3

[17] Weihong Ren, Jiandong Tian, Zhi Han, Antoni Chan, and Yandong Tang. Video desnowing and deraining based on matrix decomposition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 3

[18] Varun Santhaseelan and Vijayan K Asari. Utilizing local phase information to remove rain from video. *International Journal of Computer Vision*, 112(1):71–89, 2015. 3

[19] Tai-Xiang Jiang, Ting-Zhu Huang, Xi-Le Zhao, Liang-Jian Deng, and Yao Wang. A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 3

[20] Jin-Hwan Kim, Jae-Young Sim, and Chang-Su Kim. Video deraining and desnowing using temporal correlation and low-rank matrix completion. *IEEE Transactions on Image Processing*, 24(9):2658–2670, 2015. 3

[21] Kshitiz Garg and Shree K Nayar. Detection and removal of rain from videos. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2004. 3

[22] De-An Huang, Li-Wei Kang, Yu-Chiang Frank Wang, and Chia-Wen Lin. Self-learning based image decomposition with applications to single image denoising. *IEEE Transactions on multimedia*, 16(1):83–93, 2014. 3

[23] Yu Luo, Yong Xu, and Hui Ji. Removing rain from a single image via discriminative sparse coding. In *IEEE International Conference on Computer Vision*, 2015. 3

[24] Peter C Barnum, Srinivasa Narasimhan, and Takeo Kanade. Analysis of rain and snow in frequency space. *International journal of computer vision*, 86(2-3):256, 2010. 3

[25] Xianhui Zheng, Yinghao Liao, Wei Guo, Xueyang Fu, and Xinghao Ding. Single-image-based rain and snow removal using multi-guided filter. In *International Conference on Neural Information Processing*, 2013. 3

[26] Lei Zhu, Chi-Wing Fu, Dani Lischinski, and Pheng-Ann Heng. Joint bilayer optimization for single-image rain streak removal. In *IEEE International Conference on Computer Vision*, 2017. 3

[27] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2736–2744, 2016. 3, 4, 5, 6, 7, 8

[28] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, 2016. 3, 5

[29] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 3

[30] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6):2944–2956, 2017. 3

[31] David Eigen, Dilip Krishnan, and Rob Fergus. Restoring an image taken through a window covered with dirt or rain. In *IEEE International Conference on Computer Vision*, 2013. 3

[32] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 3, 5, 6, 7, 8

[33] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *arXiv preprint arXiv:1701.05957*, 2017. 3, 4, 5, 6, 7, 8

[34] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. *arXiv preprint arXiv:1711.07064*, 2017. 3, 7

[35] Fayao Liu, Chunhua Shen, Guosheng Lin, and Ian Reid. Learning depth from single monocular images using deep convolutional neural fields. *IEEE transactions on pattern analysis and machine intelligence*, 38(10):2024–2039, 2016. 4

[36] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. End-to-end united video dehazing and detection. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018. 4

[37] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4770–4778, 2017. 4

[38] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a" completely blind" image quality analyzer. *IEEE Signal Process. Lett.*, 20(3):209–212, 2013. 4

[39] Lixiong Liu, Bao Liu, Hua Huang, and Alan Conrad Bovik. No-reference image quality assessment based on spatial and spectral entropies. *Signal Processing: Image Communication*, 29(8):856–863, 2014. 4

[40] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE transactions on Image Processing*, 21(8):3339–3352, 2012. 4

[41] Rosaura G VidalMata, Sreya Banerjee, Brandon Richard-Webster, Michael Albright, Pedro Davalos, Scott McCloskey, Ben Miller, Asong Tambo, Sushobhan Ghosh, Sudarshan Nagesh, et al. Bridging the gap between computational photography and visual recognition. *arXiv preprint arXiv:1901.09482*, 2019. 4

[42] Ding Liu, Bihan Wen, Xianming Liu, Zhangyang Wang, and Thomas S Huang. When image denoising meets high-level vision tasks: a deep learning approach. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 842–848. AAAI Press, 2018. 4

[43] Zhangyang Wang, Shiyu Chang, Yingzhen Yang, Ding Liu, and Thomas S Huang. Studying very low resolution recognition using deep networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4792–4800, 2016. 4

[44] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, pages 91–99. 2015. 7

[45] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 7

[46] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37, 2016. 7

[47] Tsung-Yi Lin, Priyal Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *IEEE transactions on pattern analysis and machine intelligence*, 2018. 7

[48] Yu Liu, Guanlong Zhao, Boyuan Gong, Yang Li, Ritu Raj, Niraj Goel, Satya Kesav, Sandeep Gottimukkala, Zhangyang Wang, Wenqi Ren, et al. Improved techniques for learning to dehaze and beyond: A collective study. *arXiv preprint arXiv:1807.00202*, 2018. 5

[49] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952. 6

[50] Lark Kwon Choi, Jaehee You, and Alan Conrad Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing*, 24(11):3888–3901, 2015. 7

[51] Yanting Pei, Yaping Huang, Qi Zou, Yuhang Lu, and Song Wang. Does haze removal help cnn-based image classification? *arXiv preprint arXiv:1810.05716*, 2018. 7

[52] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint*, 2017. 8

[53] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3339–3348, 2018. 8