A Comprehensive Benchmark Analysis of Single Image Deraining: Current Challenges and Future Perspectives

Siyuan Li[1] · Wenqi Ren[2] · Feng Wang[1] · Iago Breno Araujo[5] · Eric K. Tokuda[5] · Roberto Hirata Junior[5] · Roberto M. Cesar-Jr.[5] · Zhangyang Wang[6] · Xiaochun Cao[2,3,4]

Abstract

The capability of image deraining is a highly desirable component of intelligent decision-making in autonomous driving and outdoor surveillance systems. Image deraining aims to restore the clean scene from the degraded image captured in a rainy day. Although numerous single image deraining algorithms have been recently proposed, these algorithms are mainly evaluated using certain type of synthetic images, assuming a specific rain model, plus a few real images. It remains 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). We further provide a comprehensive suite of criteria for deraining algorithm evaluation, including full- and no-reference metrics, subjective evaluation, and the novel task-driven evaluation. The proposed benchmark is accompanied with extensive experimental results that facilitate the assessment of the state-of-the-arts on a quantitative basis. Our evaluation and analysis indicate the gap between the achievable performance on synthetic rainy images and the practical demand on real-world images. We show that, despite many advances, image deraining is still a largely open problem. The paper is concluded by summarizing our general observations, identifying open research challenges and pointing out future directions. Our code and dataset is publicly available at http://uee.me/ddQsw.

Keywords Image deraining · Image quality assessment · Deep convolution network · Benchmark analysis

---

Communicated by Torsten Sattler.

✉ Wenqi Ren
rwq.renwenqi@gmail.com

Siyuan Li
siyuan@tju.edu.cn

Iago Breno Araujo
ibacaraujo@gmail.com

Eric K. Tokuda
tokudaek@usp.br

Roberto Hirata Junior
rhiratajr@usp.br

Roberto M. Cesar-Jr.
rmcesar@usp.br

Zhangyang Wang
atlaswang@utexas.edu

Xiaochun Cao
caoxiaochun@iie.ac.cn

1 College of Intelligence and Computing, Tianjin University, Tianjin 300350, China

2 Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

3 School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

4 Cyberspace Security Research Center, Peng Cheng Laboratory, Shenzhen 518055, China

5 Institute of Mathematics and Statistics (IME), University of São Paulo, São Paulo, Brazil

6 Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA

1 Introduction

Images captured in rainy days suffer from noticeable degradation of scene visibility. For example, raindrops inevitably adhered to camera lenses or windscreens in a rainy day, which occlude and deform some image areas and make the performances of many algorithms in the vision systems (such as object detection, tracking, recognition, etc.) significantly degraded. The goal of single image deraining algorithms is to generate sharp images from a rainy image input, which can potentially benefit both the human visual perceptual quality, and many computer vision applications, such as intelligent vehicles and outdoor surveillance systems (Sheng et al. 2020; Tokuda et al. 2020).

The recent years have witnessed significant progress in single image deraining. The progress in this field can be attributed to various natural image priors (Sun et al. 2014; Kang et al. 2012; Chen and Hsu 2013; Bossu et al. 2011) and deep convolutional neural network (CNN)-based models (Fu et al. 2017b; Qian et al. 2018; Zhang and Patel 2018). 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. In this work, we put our focus on image deraining techniques and how they have been extended or applied to high-level vision systems based on our proposed new benchmark. To the best of our knowledge, this is the first comprehensive benchmark and the first review in the literature that focuses on image deraining and its corresponding applications.

This work is organized as follows. First, Sect. 2 review the rainy image models and explain important background concepts that will be necessary throughout the rest of the paper. Next, Sect. 3 surveys the model-based and learning-based single-image deraining approaches and the existing datasets used in the rain removal literature. Next, Sect. 4 provides a comprehensive description as well as an analysis of the proposed benchmark of multi-purpose image deraining (MPID). Section 5 analyzes typical metrics and evaluation protocols for the deraining methods and provides quantitative results for them on the proposed benchmark. Finally, Sect. 6 summarizes the paper by presenting a brief discussion on the presented benchmark and enumerates potential future research directions.

1.1 Our Contribution

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: (1) the modeling of rain is simplified, i.e., each method considers and is evaluated with one type of rain only (e.g., Kang et al. 2012; Chen and Hsu 2013; Li et al. 2016, 2017; Jiang et al. 2017; Lei et al. 2017; Wei et al. 2019; Ren et al. 2019) focus on rain streaks removal, and (Qian et al. 2018; You et al. 2016) concentrate on removing raindrops); (2) most quantitative results are reported on synthetic images, which often fail to capture the complexity and characteristics of real rain. Although there are some real deraining datasets are proposed, these databases lack sufficient real-world images and without any semantic annotation for diverse evaluations. (3) 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 (Lai et al. 2016; Li et al. 2019a) or high-level computer vision utility (Dai et al. 2016, 2020; Sakaridis et al. 2018; Hahner et al. 2019).

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 3, and image examples are displayed in Fig. 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 eight 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.

In this paper, we extend our preliminary work (Li et al. 2019c) in the following aspects.

– *Evaluations of more image deraining algorithms* In Li et al. (2019c), we evaluate six different deraining methods

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

on the proposed multi-purpose image deraining (MPID) dataset. In this manuscript, we also evaluate two very recent image deraining work of DAF-Net (Hu et al. 2019) and STL (Wei et al. 2019), which perform better than conventional deraining approaches on existing deraining datasets. In particular, STL (Wei et al. 2019) is the first semi-supervised learning network toward the image deraining task.

– *Extension of detection methods* In Li et al. (2019c), we use Faster R-CNN (FRCNN) (Ren et al. 2015), YOLO-V3 (Redmon and Farhadi 2018), SSD-512 (Liu et al. 2016), and RetinaNet (Lin et al. 2018) to detect objects after using a deraining algorithm. In this paper, we add a new state-of-the-art detection model of CenterNet (Zhou et al. 2019) to conduct the task-driven comparisons. As a result, our employed detection methods including

tow-stage, one-stage anchor, and one-stage anchor-free detection algorithms. In addition, we also found that the recent CenterNet performs better than the conventional deep detection models.

– *Detailed results of object detection* In addition to mAP results reported in Li et al. (2019c), we further show all the AP results in each object class for different deraining algorithms for more detailed comparative analysis.

– *Datasets survey* In this paper, we summarize the existing image draining datasets used to measure and compare the performance of image deraining algorithms. We found that existing datasets are either too small in scale or limited to one rain type, or lack sufficient real-world images for diverse evaluations. In addition, none of them has any semantic annotation nor consider any subsequent task performance.

– *More analysis* We add more analysis about different deraining algorithms in terms of various evaluation criteria (full- and no-reference objective, subjective, and task-specific metrics) to show the current challenges of the performance gap between synthetic and real-world images. Based on the comprehensive results, we further conclude some possible research directions for image deraining in the future.

2 Rainy Image Formulation Models

In this section, we review the commonly-used rain synthesis models in the literature. 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. 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 (Garg and Nayar 2005). Most existing deraining works assume one rain model (usually rain streak), which might have over-simplified 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 super-imposition 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 (You et al. 2016) 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. 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 a composition of the rain streak model and the atmospheric scattering haze model (McCartney 1976):

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

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.

There are two main drawbacks of existing evaluation approaches. First, synthetic rainy images usually fail to capture the characteristics of real degradation in rainy day. For example, the models of (1) and (2) only consider one factor. Halder et al. (2019) recently proposed a physically-based rendering method to improve the realism of these synthetic rainy images. They used a more complex pipeline to simulate and insert rain streaks taking into account its amount to generate a more convincing visual result. This method has achieved the goal of creating visually appealing rainy images. However, the method still only generates rain streaks without considering the mist effects in a rainy image. Second, existing deraining approaches use PSNR and SSIM to evaluate image restoration performance, which does not correlate well with human perception (Lai et al. 2016) and high-level visual algorithms (Li et al. 2019a). The lack of human and machine perceptual studies makes it difficult to compare the performance of deraining algorithms. While numerous full- and no-reference image quality metrics have been proposed, it is unclear whether these metrics can be applied to measure the quality of derained images.

3 Related Work

3.1 Overview of Deraining Algorithms

Early methods often require multiple frames for deraining (Ren et al. 2017; Santhaseelan and Asari 2015; Jiang et al. 2017; You et al. 2016). Garg and Nayar (2004) 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. Garg and Nayar (2005) further improved the performance by selecting camera parameters without appreciably altering the scene appearance. However, those methods are not applicable to single image deraining.

Compared to multi-frame based deraining approaches which have temporal redundant knowledge, deraining from a single image is more challenging since less information is available. To address this problem, the design of single image deraining algorithm has attracted more research attention. The existing single image deraining methods can be roughly divided into two categories: model-based (non-deep-learning) and data-driven (deep-learning) approaches.

There is a summary of single image rain removal methods in Table 1.

3.1.1 Model-Driven Algorithms

The model-driven methods especially focus on sufficiently utilizing and encoding physical properties of rain and prior knowledge of background scenes into an optimization problem and designing rational algorithms to solve it. These algorithms can be mainly divided into three main categories: filter based methods, low-rankand sparse-coding

based algorithms, and Gaussian Mixture Model (GMM) based approaches, etc.

Filter based algorithms Zheng et al. (2013) presented a multiple guided filter based method using low frequency part of a single image. Ding et al. (2016) designed a guided L0 smoothing filter based on L0 gradient minimization to remove rain streaks in a rainy image. Santhaseelan and Asari (2015) first detected rain streaks based on phase congruence features from input rainy videos, then the variation of features from frame to frame is capitalized to remove rain from videos.

Sparse coding based algorithms Many deraining methods capitalized on clean image or rain type priors to remove rain (Sun et al. 2014; Luo et al. 2015; Barnum et al. 2010). Kang et al. (2012) 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. (2017) introduced a rain removal method based on the prior that rain streaks typically span a narrow range of directions. Chen and Hsu (2013) decomposed the background and rain streak layers based on low-rank priors.

GMM based algorithms Li et al. (2016) used patch-based priors for both the clean background and rain layers in the form of Gaussian mixture models. Based on Li et al. (2016, 2017) further introduced a structure residue recovery step to further separate the background residues and improve the decomposition quality for image deraining.

However, all of the above approaches rely on handcrafted image priors, which cannot hold in some real-world scenes. As a result, these model-driven algorithms tend to have unsatisfactory performances and generate some artifacts on real-world images with complicated scenes and rain forms.

3.1.2 Data-Driven Algorithms

Recent methods often adopt the data-driven algorithms by designing specific network architectures to learn network parameters for attaining complex rain removal functions. Most of these methods aim at certain insightful aspects of rain removal and have their applicability and advantages on

some specific scenarios. We briefly discuss the popular deep neural networks employed for image deraining in this section.

CNN models A CNN architecture typically includes convolutional layers, pooling layers and fully connected layers. CNN is powerful in learning feature representation of different abstraction levels from large-scale data.

Recently, CNNs have achieved dominant success for image restoration (Ren et al. 2016; Zhang et al. 2017) including single image deraining (Fu et al. 2017a; Eigen et al. 2013). Fu et al. (2017b) proposed a deep detail network (DDN) for removing rain from single images with details preserving. Yang et al. (2017) 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 Zhang and Patel (2018) for joint rain density estimation and image deraining. Hu et al. (2019) formulated a depth-guided attention mechanism to learn depth-attentional features and regress a residual map, and prepared a new dataset RainCityscapes for rain removal. However, existing deep networks usually have an enormous number of parameters. To remedy this, Fu et al. (2020) proposed a lightweight deep network that is based on the classical Gaussian-Laplacian pyramid for single image deraining.

GAN models GAN is proposed to train generative models through a two-player game between a generator and a discriminator. Specially, the generator aims to generate synthesized data of the same distribution of real data, and tries to fool the discriminator. Discriminator is trained to distinguish synthesized data from real samples. During training, the generator and the discriminator compete with each other and improve themselves to help the two players to generate realistic derained images (Qian et al. 2018; Li et al. 2019b).

Qian et al. (2018) addressed a different problem of removing raindrops from single images by using visual attention with a generative adversarial network (GAN). Zhang et al. (2019) proposed a novel single image deraining method called Image Deraining conditional generative adversarial network (CGAN), which considers quantitative, visual and also discriminative performance into the objective function. Li et al. (2019b) proposed an integrated two-stage neural network and novel streak-aware decomposition to adaptively separate the image into a high-frequency component containing rain streaks and a low-frequency component containing rain accumulation. Yu et al. (2020) proposed a fully end-to-end image dehazing algorithm FD-GAN, which directly outputs haze-free images without the estimation of intermediate parameters.

Semi/Unsupervised models Semi-supervised learning is a learning paradigm concerned with the study of how comput-

ers and natural systems such as humans learn in the presence of both labeled and unlabeled data. While unsupervised learning means learning by only unlabeled data, i.e., using real captured rainy images with the corresponding ground truths. Wei et al. (2019) firstly proposed a semi-supervised transfer learning framework for single image rain removal. They rationally formulate the residual between the expected output clean images and their original noisy ones through a likelihood term imposed on a parameterized distribution designed based on the domain understanding for residuals. Jin et al. (2019) proposed an unsupervised generative adversarial network (UD-GAN) with self-supervised constraints for image deraining.

Despite the progress of deep-learning-based approaches compared with prior-based rain removal, their performance hinge on the synthetic training data, which may become problematic if real rainy images show a domain mismatch.

3.2 Datasets

In the computer vision field, widely accepted and commonly used databases have achieved objective comparisons and promoted scientific progress (Katrin et al. 2016; Szeliski et al. 2008; Schops et al. 2017). Several rainy image datasets were also used to measure and compare the performance of deraining algorithms. Li et al. (2016) introduced 12 images using photo-realistic rendering techniques. Zhang et al. (2019) synthesized a set of training and testing images with rain streak, using the same way in Li et al. (2016). The training set consists of 700 images and the testing set consists of 100 images. In addition, Zhang et al. (2019) also collected a dataset of 92 real-world rainy images downloaded from the web for qualitative visual comparison. Qian et al. (2018) released a set of clean and rain-drop corrupted image pairs, using a special lens equipment. To address heavy rain removal problem, Li et al. (2019b) created a new synthetic rain dataset named NYU-Rain and another outdoor rain dataset on a set of outdoor clean images, denoted as Outdoor-Rain. Specifically, they provided a new synthetic data generation pipeline by synthesizing the mist effect according to the scene depth. To make the synthesized images more realistic, they also added Gaussian blur on both the transmission map and the background to simulate the effect of scattering in heavy rain scenarios. Meanwhile, Wang et al. (2019) constructed a large-scale real-world paired rain and clean dataset by a semi-automatic method that incorporates temporal priors and human supervision.

We note that the recent work of Li et al. (2019b) and Wang et al. (2019) are two large-scale rain removal datasets with more realism than those from conventional deraining datasets. However, the data from Li et al. (2019b) only includes synthesized images, while the generated ground truths in Wang et al. (2019) may contain some noise, blur,

and shaking due to the misalignment between neighboring frames in the captured videos. We summarized the most used datasets for image deraining in Table 2. As shown, existing datasets are either too small in scale or limited to one rain type (rain streaks or raindrops), or lack sufficient real-world images for diverse evaluations. Although the recent proposed Weather Kitti dataset (Halder et al. 2019) includes a large-scale number of images, none of the existing databases has any semantic annotation or subsequent task performance. In contrast, our dataset contains synthetic, real-world, as well as annotated rainy images for a comprehensive evaluation of single image deraining algorithms. The images in our dataset cover various rain types and scenarios and include actual challenges and variations from the real world.

4 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" (Li et al. 2019a; Kupyn et al. 2018) 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 3.

4.1 Training Sets: Three Synthesis Models

Following the three rain models in Sect. 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 settings 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). Specifically, we synthesize rainy images according to the following two aspects. First, we follow the common protocol used in Li et al. (2016), Zhang et al. (2019) to generate rain streaks. We also noticed the wet ground/overcast sky issue during data synthesis, and man-

---

[1] Note that for Rain drop (T), the data generation used physical simulation (Qian et al. 2018), i.e., with/without lens, rather than algorithm simulation.

Table 1 An overview of single-image deraining methods

Method	Category	Advantages	Publication
	Model-driven		
Guided filter (Zheng et al. 2013)	Filter	The rain-removed image was acquired by taking the minimum value of rainy image and the coarse recovery image obtained by merging low frequency part (LFP) of rainy image with high frequency part (HFP) of rain-free image	Zheng et al. (2013)
Image decomposition (Kang et al. 2012)	Sparse coding	The first automatic MCA-based image decomposition framework is proposed. The learning of the dictionary for decomposing rain steaks is fully automatic and self-contained, where no extra training samples are required	Kang et al. (2012)
DSC (Luo et al. 2015)		This approach provided a variational model, together with a simple iterative numerical method, and there is no any other image processing module involved	Luo et al. (2015)
JCAS (Gu et al. 2017)		This work is to sparsely approximate the patches of two layers by very high discriminative codes over a learned dictionary with strong mutual exclusivity property	Gu et al. (2017)
GMM (Li et al. 2016, 2017)	GMM	This work uses simple patch-based priors for both the background and rain layers which are based on Gaussian mixture models and can accommodate multiple orientations and scales of the rain streaks.	Li et al. (2016, 2017)
	Data-driven		
CNN (Eigen et al. 2013)	Deep CNN	This work trained a convolutional neural network which learns how to map corrupted image patches to clean ones, implicitly capturing the characteristic appearance of dirt and water droplets in natural images	Eigen et al. (2013)
Lightweight CNN (Fu et al. 2020)		This work proposed a lightweight deep network based on the classical Gaussian–Laplacian pyramid for single image deraining	Fu et al. (2020)
JORDER (Yang et al. 2017)		The first method to model the rain-streak binary mask and jointly detect and remove rains from single images. A contextualized dilated network is used to obtain more context while preserving rich local details	Yang et al. (2017)
DAF-Net (Hu et al. 2019)		The work proposed a new dataset called Rain Cityscapes and designed an end-to-end deep neural network which can learn depth-attentional features via a depth-guided attention mechanism, and regress a residual map to produce the rain-free image output	Hu et al. (2019)
PReNet (Ren et al. 2019)		This work adopted intra-stage recursive computation of ResNet to notably reduce network parameters with unsubstantial degradation, and took both stage-wise result and original rainy image as input to each ResNet	Ren et al. (2019)
FFA-Net (Xu et al. 2020)		This work proposed an attention-based feature fusion (FFA) structure, this structure can retain shallow layers' information and pass it into deep layers	Xu et al. (2020)

Table 1 continued

Method	Category	Advantages	Publication
CGAN (Zhang et al. 2019)	GAN	A densely-connected generator sub-network that is specifically designed for the single image deraining task, without the use of any additional post-processing	Zhang et al. (2019)
Attentive GAN (Qian et al. 2018)		Attention map is injected into both the generative and discriminative networks for learning to attend raindrop regions and percept their surroundings	Qian et al. (2018)
Heavy Rain Restoration (Li et al. 2019b)		This work introduced an integrated two-stage neural network and an novel streak-aware decomposition to adaptively separate the image into components	Li et al. (2019b)
FD-GAN (Yu et al. 2020)		This model can generator more natural and realistic dehazed images with less color distortion and fewer artifacts using Fusion-discriminator which takes frequency information as additional priors	Yu et al. (2020)
Semi-supervised CNN (Wei et al. 2019)	Semi/Unsupervised	This work firstly proposed a semi-supervised transfer learning framework which rationally formulate the residual between the expected output clean images and their original noisy ones through a likelihood term imposed on a parameterized distribution	Wei et al. (2019)
UD-GAN (Jin et al. 2019)		UD-GAN is proposed to tackle the problems of insufficient labeled-supervised constraints by introducing self-supervised constraints from the intrinsic statistics of unpaired rainy and clean images	Jin et al. (2019)

Table 2 Summary of the most used datasets for image deraining

Name	Number (train/test)	Real/synthetic	Year	Resulotion (train)	Annotations
Rain12 (Li et al. 2016)	12	Synthetic	2016	481×321	N/A
Rain100L (Yang et al. 2017)	1800/200	Synthetic	2017	481×321	N/A
Rain100H (Yang et al. 2017)	1800/200	Synthetic	2017	481×321	N/A
Rain14000 (Fu et al. 2017b)	9100/4900	Synthetic	2017	512×384	N/A
RainCityscapes (Hu et al. 2019)	9432/1188	Synthetic	2017	2048×1024	N/A
GAN-Pired-Dataset (Zhang et al. 2019)	700/192	Synthetic/real	2017	1024×384	N/A
Rain12000 (Zhang and Patel 2018)	12000/4000	Synthetic	2018	1024×512	N/A
RainDrop-Dataset (Qian et al. 2018)	861/307	Synthetic	2018	720×480	N/A
NYU-Rain (Li et al. 2019b)	13500/2700	Synthetic	2019	640×480	N/A
Outdoor-Rain (Li et al. 2019b)	9000/1500	Synthetic	2019	720×480	N/A
Weather Kitti (Halder et al. 2019)	7481	Synthetic	2019	1216×352	N/A
Weather Cityscapes (Halder et al. 2019)	2995	Synthetic	2019	1024×512	N/A
STL-Syn-Dataset (Wei et al. 2019)	147	Real	2019	253×199 to 1600×1200	N/A
Real Rain Dataset (Wang et al. 2019)	28500/1000	Real	2019	512×512	N/A

ually inspected/selected clear overcast images, on which we synthesized rain. Second, we follow the widely-accepted routine in Li et al. (2019a), Sakaridis et al. (2018), Ren et al. (2016, 2018a, b, 2020) to generate mist. We first estimate

depth from clear overcast outdoor images, and then synthesizing mist images as like in Ren et al. (2016).

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-

Table 3 Overview of the proposed MPID dataset

Subset	Number	Real/synthetic	Resolution	Metrics	Annotation
Training set					
Rain streak (T)	2400 (pairs)	Synthetic	512×384	/	No
Raindrop (T)	861 (pairs)	Synthetic	720×480	/	No
Rain and mist (T)	700 (pairs)	Synthetic	1280×720	/	No
Testing set					
Rain streak (S)	200 (pairs)	Synthetic	640×480	PSNR, SSIM, NIQE, BLIINDS-II, SSEQ	No
Rain streak (R)	50	Real	1024×681 to 1920×1080	NIQE, BLIINDS-II, SSEQ	No
Raindrop (S)	149 (pairs)	Synthetic	720×480	PSNR, SSIM, NIQE, BLIINDS-II, SSEQ	No
Raindrop (R)	58	Real	720×480	NIQE, BLIINDS-II, SSEQ	No
Rain and mist (S)	70 (pairs)	Synthetic	720×1080 to 5312×2998	PSNR, SSIM, NIQE, BLIINDS-II, SSEQ	No
Rain and mist (R)	30	Real	1024×681 to 2976×3968	NIQE, BLIINDS-II, SSEQ	No
Task-driven evaluation set					
RID	2495	Real	1920×990 to 4023×3024	mAP	Yes
RIS	2048	Real	640×368 to 640×480	mAP	Yes

parameters to Li et al. (2016), Zhang et al. (2019). The Rain drop (T) set was borrowed from Qian et al. (2018)'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 Liu et al. (2016), Li et al. (2018) as recommended by Li et al. (2017), 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.

4.2 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) (Mittal et al. 2013), spatial-spectral entropy-based quality (SSEQ) (Liu et al. 2014), and blind image integrity notator using DCT statistics (BLIINDS-II) (Saad et al. 2012), 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.

4.3 Task-Driven Evaluation Sets

As pointed out by a plenty of recent works (Wang et al. 2016; Liu et al. 2018, 2019, 2020; Scheirer et al. 2020; Yang et al. 2020; Hahner et al. 2019), the performance of high-level computer vision tasks, such as object detection and recognition, will deteriorate in the presence of various sensory and environmental degradation. In particular, Sakaridis et al. (2018) studied the effect of image dehazing on semantic segmentation by a synthesized Foggy Cityscapes dataset with 20,550 images. This work carefully investigated the practicability of image dehazing for semantic foggy

---

[2] Note that in Liu et al. (2014) and Saad et al. (2012), 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.

scene understanding (SFSU) and found that image dehazing marginally advances SFSU in most cases. Dai et al. (2016) evaluated several image super-resolution methods on high-level vision tasks and concluded that super-resolution approaches are usually helpful for other vision tasks. In these cases, the low-level image processing methods improved the performances on the high-level tasks. While deraining could be used as pre-processing for many computer vision tasks executed in the rainy conditions, there has been no systematic study on deraining algorithms' impact on those target tasks. The recent work of Halder et al. (2019) evaluated the robustness of high-level tasks on rainy conditions. However, the evaluation did not include a study on the usefulness of deraining algorithms for the high-level 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 evaluating 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 4. 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 2495 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 2048 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). Specifically, we found very few bicycles in the RIS set, which consists of common sense that one will not go cycling when

Table 4 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

it rains. Therefore, we annotated trucks rather than bicycles in the RIS dataset. Finally, 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.

We carefully selected images containing these objects in the scene. We observed that rainy images tend to present a lesser number of objects in the scene, which is a natural disposition given that persons usually avoid getting out on the street when it is raining. These efforts result in a rich base of outdoor images in rainy and sunny weather conditions with the most common objects annotated.

Adverse weather condition like rain and haze, affects visual quality of images. Images captured in such conditions are intrinsically degraded. This leads to computer vision systems to have their performances decreased. Thus, tasks like deraining and dehazing are extremely challenging and important. Recently, many efforts have been made to remove rainy and hazy effects or, at least, attenuate their impairments. Despite the success of recent algorithms, a real rainy scenario continues to constitute a demanding problem to handle. We believe there is a gap between the synthetic rainy datasets used so far for training the current models and real rainy images. Hence, to improve on this task, we need to consider real information from rainy scenes. Motivated by this, we present a new dataset containing real rainy images from surveillance video cameras. We further provide a sunny set of images for evaluation and comparison in the object detection task. In this way, deraining strategies might benefit from promising results that image-to-image translation have shown for domain adaptation.

5 Experimental Comparison

We evaluate eight representative state-of-the-art algorithms on MPID: Gaussian mixture model prior (GMM) (Li et al. 2016), joint rain detection and removal (JORDER) (Yang et al. 2017), deep detail network (DDN) (Fu et al. 2017b), conditional generative adversarial network (CGAN) (Zhang et al. 2019), density-aware image de-raining method using a multistream dense network (DID-MDN) (Zhang and Patel 2018), depth-attentional features network (DAF-Net) (Hu

Table 5 Average full- and no-reference evaluations results on synthetic rainy images

	Degraded	GMM (Li et al. 2016)	JORDER (Yang et al. 2017)	DDN (Fu et al. 2017b)	CGAN (Zhang et al. 2019)	DID-MDN (Zhang and Patel 2018)	DeRaindrop (Qian et al. 2018)	DAF-Net (Hu et al. 2019)	STL (Wei et al. 2019)
Rain streak									
PSNR ↑	25.95	<u>26.88</u>	26.26	**29.39**	21.86	26.80	–	15.8370	26.7417
SSIM ↑	0.7565	0.7674	**0.8089**	0.7854	0.6277	<u>0.8028</u>	–	0.3120	0.7359
SSEQ ↑	70.24	67.46	<u>73.70</u>	**75.95**	70.02	60.05	–	66.0204	71.8536
NIQE ↓	5.4529	4.4248	<u>4.2337</u>	**3.9834**	4.6189	4.8122	–	4.8352	5.1800
BLINDS-II ↑	78.89	75.95	<u>84.21</u>	**91.71**	79.29	67.90	–	78.45	76.3571
Raindrops									
PSNR ↑	25.40	24.85	<u>27.52</u>	25.23	21.35	24.76	**31.57**	20.2375	21.4198
SSIM ↑	<u>0.8403</u>	0.7808	0.8239	0.8366	0.7306	0.7930	**0.9023**	0.7319	0.6503
SSEQ ↑	78.48	64.73	**84.32**	77.62	63.15	58.42	72.42	62.1607	<u>83.3950</u>
NIQE ↓	3.8126	5.1098	4.3278	4.1462	**3.3551**	4.1192	5.0047	5.1996	<u>3.6879</u>
BLINDS-II ↑	<u>92.50</u>	75.95	88.05	91.95	73.85	64.70	**96.45**	76.90	86.7778
Rain and mist									
PSNR ↑	26.84	29.37	<u>30.37</u>	**32.98**	22.44	28.77	–	17.0765	23.5496
SSIM ↑	0.8520	0.8960	<u>0.9262</u>	**0.9350**	0.7636	0.8430	–	0.6526	0.7148
SSEQ ↑	<u>72.37</u>	65.39	70.55	69.80	68.71	65.33	–	54.2606	**81.8984**
NIQE ↓	3.4548	3.2117	<u>2.8595</u>	2.9970	**2.8336**	3.0871	–	7.0598	3.4893
BLINDS-II ↑	82.95	74.90	83.75	<u>85.75</u>	80.20	76.35	–	53.45	**92.0556**

We use bold and underline to indicate the best and suboptimal performance, respectively

Table 6 Average no-reference evaluations results of derained results on real rainy images

	Degraded	GMM(Li et al. 2016)	JORDER (Yang et al. 2017)	DDN (Fu et al. 2017b)	CGAN (Zhang et al. 2019)	DID-MDN (Zhang and Patel 2018)	DeRaindrop (Qian et al. 2018)	DAF-Net (Hu et al. 2019)	STL (Wei et al. 2019)
Rain streak									
SSEQ ↑	65.77	61.63	<u>64.00</u>	63.51	59.32	55.11	–	55.18	**72.3867**
NIQE ↓	3.5236	**3.2117**	<u>3.5371</u>	3.5811	3.5374	5.1255	–	6.0403	3.7404
BLINDS-II ↑	78.04	75.54	82.62	**85.81**	78.42	66.65	–	54.40	<u>85.6111</u>
Raindrops									
SSEQ ↑	78.23	64.77	69.26	67.62	62.18	60.65	**79.83**	67.86	<u>73.0466</u>
NIQE ↓	3.8229	4.3801	<u>3.6579</u>	3.8290	4.4692	4.5631	**3.5953**	4.6526	4.1245
BLINDS-II ↑	84.46	71.21	80.04	77.75	66.29	66.63	**87.13**	75.90	<u>80.7222</u>
Rain and mist									
SSEQ ↑	73.86	59.51	65.18	64.56	<u>70.04</u>	63.85	–	58.1319	**80.8053**
NIQE ↓	3.2602	4.4808	3.3238	3.7261	**2.9532**	<u>3.2260</u>	–	6.8706	3.4411
BLINDS-II ↑	84.00	62.70	78.62	81.67	<u>84.91</u>	76.08	–	53.3889	**89.6667**

We use bold and underline to indicate the best and suboptimal performance except the degraded inputs, respectively

Fig. 2 Visual comparisons of derained results on real images: rain streak (first image), raindrop (second image), and rain and mist (third image)

et al. 2019), semi-supervised transfer network (STL) (Wei et al. 2019), and DeRaindrop (Qian et al. 2018). All except GMM are state-of-the-art CNN-based deraining algorithms.

Evaluation Protocol The first seven 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 training dataset. In addition, we create a cascaded pipeline, by first running each of the five rain streak removal algorithms, followed by feeding into a dehazing model, as like in Yang et al. (2016), Li et al. (2019b). Based on the rain and mist model in (3), we can remove the

Table 7 Average subjective scores of derained results on 27 real images

	Rainy	GMM (Li et al. 2016)	JORDER (Yang et al. 2017)	DDN (Fu et al. 2017b)	CGAN (Zhang et al. 2019)	DID-MDN (Zhang and Patel 2018)	DeRaindrop (Qian et al. 2018)	DAF-Net (Hu et al. 2019)	STL (Wei et al. 2019)
Rain streak	0.64	0.80	0.91	<u>1.15</u>	**1.26**	0.97	–	0.61	0.77
Raindrops	0.80	**1.14**	0.75	0.83	0.85	<u>0.95</u>	0.80	0.72	0.94
Rain and mist	0.44	1.00	0.70	0.90	<u>1.22</u>	**1.40**	–	0.52	0.86

additive rain streak first without the damage of haze in theory. Therefore, we first remove rain for the rain and mist input then restore clean image by feeding the derained result to MSCNN (Ren et al. 2016), which is trained on the synthesized hazy images based on the Middlebury stereo database (Scharstein and Szeliski 2002). The very recent work (Li et al. 2019b) also demonstrates that using the combination strategy of deraining first and then dehazing shows better performance than dehazing first and then deraining. We choose the MSCNN dehazing algorithm since recent dehazing studies (Li et al. 2019a; Liu et al. 2018) 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.

5.1 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 Mittal et al. 2012, SSEQ Liu et al. 2014, and BLIINDS-II Saad et al. 2012).

As seen from Table 5, the results have high consensus levels on synthetic data. First, the method of DDN (Fu et al. 2017b) by Fu et al. is the obvious winner on the rain streak (S) set, followed by the approach of JORDER (Yang et al. 2017). Second, DerainDrop (Qian et al. 2018) performs the best on the rain drop (S) set, especially significantly surpassing the others in terms of full-reference of PSNR and SSIM, as well as no-reference metric of BLIINDS-II, showing that its specific structure indeed suits the raindrop removal problem. Other rain streak removal models seem to even hurt PSNR, SSIM and BLIINDS-II, compared to the input rainy images. For example, CGAN (Zhang et al. 2019) decreases both PSNR and SSIM on the rain drop (S) set. The main reason may be that GANs tend to generate some unrealistic details in the scenes. Finally, for the rain and mist images, DDN (Fu et al. 2017b) also perform consistently the best

according to PSNR and SSIM. Since STL (Wei et al. 2019) is trained to adapt real unsupervised diverse rain types through transferring from the supervised synthesized rain, this model achieves the highest SSEQ and BLINDS-II values although it performs worse in terms of full-reference metrics, which is aligned with the emerging trend of semi-supervised or unsupervised learning methods by using real-world training images .

The effectiveness of the winners can be ascribed to the two-step strategy of rain detection and removal, i.e., first estimate a mask of rain streaks or raindrops, then remove rain artifacts capitalized on the mask. We note that DDN (Fu et al. 2017b) focuses on high frequency details during training stage, while JORDER (Yang et al. 2017) also first detects the locations of rain streak, then removes rain based on the estimated rain streak regions. Coincidentally, DeRaindrop (Qian et al. 2018) also uses an attentive generative network to learn about raindrop regions and their surroundings first then derain images using the information of the learned masks. Therefore, removing background interference and attentively focusing on rain regions seem to be the main reason of the winners in Table 5. In addition, different from conventional deep learning methods which only use supervised image pairs, the recent work of STL (Wei et al. 2019) put real rainy images into the network training process and therefore obtain the best performance in terms of no-reference metrics.

We then show the derained results on the real-world images in Table 6, using three no-reference metrics (NIQE, SSEQ, and BLIINDS-II). Figure 2 shows three corresponding visual comparison examples. The raindrop (R) and rain and mist (R) sets show consistent results with their synthetic cases: DerainDrop (Qian et al. 2018) and STL (Wei et al. 2019) rank top-two on the raindrop dataset, while STL (Wei et al. 2019) still dominates on the rain and mist set. In particular, DerainDrop (Qian et al. 2018) ranks first in term of all the three no-reference metrics, thanks to the raindrop attention map learned by the attentive-recurrent network. However, some different tendency is observed on the rain streak (R) set: although DDN (Fu et al. 2017b) still obtain the highest BLIINDS-II value, it has a worse performance according to

Table 8 Detection results on the RID sets

		Rainy	JORDER (Yang et al. 2017)	DDN (Fu et al. 2017b)	CGAN (Zhang et al. 2019)	DID-MDN (Zhang and Patel 2018)	DeRaindrop (Qian et al. 2018)	DAF-Net (Hu et al. 2019)	STL (Wei et al. 2019)
mAP	FRCNN (Ren et al. 2015)	16.52	16.97	18.36	**23.42**	16.11	15.58	8.79	6.97
	YOLO-V3 (Redmon and Farhadi 2018)	**27.84**	26.72	26.20	23.75	24.62	24.96	26.48	24.15
	SSD-512 (Liu et al. 2016)	**17.71**	17.06	16.93	16.71	16.70	16.69	17.05	16.06
	RetinaNet (Lin et al. 2018)	**23.92**	21.71	21.60	19.28	20.08	19.73	21.71	17.76
	CenterNet (Zhou et al. 2019)	25.53	24.04	24.05	24.89	26.13	25.55	**26.91**	21.40
Person	FRCNN (Ren et al. 2015)	18.73	16.14	17.53	**21.01**	16.31	16.71	16.01	12.56
	YOLO-V3 (Redmon and Farhadi 2018)	**36.89**	34.96	34.94	29.91	30.38	32.13	33.95	30.78
	SSD-512 (Liu et al. 2016)	14.79	**14.85**	14.34	14.08	14.33	14.74	14.74	12.59
	RetinaNet (Lin et al. 2018)	**30.28**	27.65	28.10	23.74	23.74	25.14	26.49	23.68
	CenterNet (Zhou et al. 2019)	32.92	31.65	31.89	31.24	33.22	**34.53**	33.44	28.26
Bicycle	FRCNN (Ren et al. 2015)	14.49	12.54	13.87	**20.17**	10.42	10.13	1.92	1.21
	YOLO-V3 (Redmon and Farhadi 2018)	15.06	**16.73**	14.55	12.48	12.82	13.20	16.03	15.22
	SSD-512 (Liu et al. 2016)	**13.34**	12.63	12.85	12.16	11.11	12.47	11.36	11.74
	RetinaNet (Lin et al. 2018)	**17.83**	15.46	15.65	13.51	14.18	14.17	13.93	14.20
	CenterNet (Zhou et al. 2019)	**14.79**	13.26	13.88	13.93	13.24	13.92	14.75	14.37
Car	FRCNN (Ren et al. 2015)	38.56	34.73	36.15	**40.98**	35.21	34.79	30.23	24.77
	YOLO-V3 (Redmon and Farhadi 2018)	**65.98**	63.59	63.28	60.31	62.65	62.05	65.01	61.27
	SSD-512 (Liu et al. 2016)	**46.01**	44.13	44.60	45.00	43.23	41.55	44.60	42.33
	RetinaNet (Lin et al. 2018)	**54.99**	51.07	50.83	48.98	51.62	48.49	54.18	43.18
	CenterNet (Zhou et al. 2019)	60.20	58.18	56.64	62.13	66.25	60.78	**68.43**	54.87
Bus	FRCNN (Ren et al. 2015)	12.37	12.40	13.81	**17.82**	11.48	11.58	3.52	3.47
	YOLO-V3 (Redmon and Farhadi 2018)	**42.76**	39.47	39.01	34.90	37.11	38.03	38.18	33.97
	SSD-512 (Liu et al. 2016)	**28.95**	27.63	26.35	25.69	28.37	28.22	28.64	26.83
	RetinaNet (Lin et al. 2018)	**35.82**	32.03	34.43	26.38	27.13	27.94	31.09	22.73
	CenterNet (Zhou et al. 2019)	39.26	35.63	36.25	35.60	38.27	39.51	**39.42**	31.29
Motorbike	FRCNN (Ren et al. 2015)	8.74	7.33	9.36	**14.38**	6.61	5.60	1.17	2.92
	YOLO-V3 (Redmon and Farhadi 2018)	**6.33**	5.54	5.39	4.89	4.74	4.34	6.16	4.00
	SSD-512 (Liu et al. 2016)	3.17	3.15	3.44	3.31	3.16	3.13	**3.79**	3.07
	RetinaNet (Lin et al. 2018)	**4.59**	4.04	4.58	3.10	3.81	2.63	4.27	3.05
	CenterNet (Zhou et al. 2019)	5.59	5.53	5.82	6.33	**6.59**	4.87	5.61	3.92

the SSEQ and NIQE metrics. In contrast, STL (Wei et al. 2019) becomes the dominant winner on those real images again as like in the rain and mist case, outperforming DDN (Fu et al. 2017b) with a large margin in terms of SSEQ. As we observed, since CGAN (Zhang et al. 2019) 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 as shown in Fig. 2, especially for the rain streak and the rain and mist in

the first and the third images. Additionally, the result might also suggest a larger domain gap between synthetic and real rain and mist data.

From Tables 5 and 6, we can observe that despite certain discrepancy (e.g., when it comes to "bad performers"), the metrics agree reasonably well on ranking top performers. For example, the method of DeRaindrop is the clear winner, winning two full-reference metrics on synthetic raindrop images in Table 5 and three non-reference metrics on real-world

Table 9 Detection results on the RIS set

	Rainy	JORDER (Yang et al. 2017)	DDN (Fu et al. 2017b)	CGAN (Zhang et al. 2019)	DID-MDN (Zhang and Patel 2018)	DeRaindrop (Qian et al. 2018	DAF-Net (Hu et al. 2019)	STL (Wei et al. 2019)
mAP								
FRCNN (Ren et al. 2015)	**22.68**	21.41	20.76	18.02	18.93	19.97	13.02	6.35
YOLO-V3 (Redmon and Farhadi 2018)	**23.27**	20.45	21.80	18.71	21.50	20.43	14.52	11.22
SSD-512 (Liu et al. 2016)	8.19	7.94	8.29	7.10	8.21	8.13	6.20	2.35
RetinaNet (Lin et al. 2018)	12.81	10.71	10.39	9.36	10.33	10.85	**15.33**	8.60
CenterNet (Zhou et al. 2019)	23.35	24.01	22.77	20.21	**23.61**	20.94	16.95	9.90
Person								
FRCNN (Ren et al. 2015)	**14.65**	14.51	14.02	10.69	11.37	12.77	6.57	4.28
YOLO-V3 (Redmon and Farhadi 2018)	**21.25**	20.18	20.33	15.18	18.43	16.73	6.31	5.59
SSD-512 (Liu et al. 2016)	13.05	11.78	11.33	8.32	11.80	9.70	7.24	1.57
RetinaNet (Lin et al. 2018)	**10.59**	8.81	8.15	6.22	7.22	8.61	8.84	3.99
CenterNet (Zhou et al. 2019)	20.20	**22.04**	18.99	15.24	20.35	14.67	8.80	4.89
Truck								
FRCNN (Ren et al. 2015)	**20.29**	17.71	17.83	14.14	17.31	17.23	7.92	2.40
YOLO-V3 (Redmon and Farhadi 2018)	**21.25**	11.80	12.51	9.48	12.75	12.52	7.32	3.23
SSD-512 (Liu et al. 2016)	3.06	3.95	3.76	2.92	4.15	3.88	2.34	1.25
RetinaNet (Lin et al. 2018)	**8.14**	6.45	6.06	4.66	6.78	7.68	7.86	2.33
CenterNet (Zhou et al. 2019)	12.00	11.49	10.54	8.30	**12.09**	9.18	6.35	1.50
Car								
FRCNN (Ren et al. 2015)	**38.54**	38.00	37.54	35.80	36.37	35.80	33.04	14.57
YOLO-V3 (Redmon and Farhadi 2018)	**44.93**	40.21	43.23	42.07	42.61	42.39	37.17	27.64
SSD-512 (Liu et al. 2016)	20.87	19.60	21.42	21.36	20.25	20.39	19.73	7.34
RetinaNet (Lin et al. 2018)	28.80	26.19	26.36	26.26	25.31	25.20	**40.00**	23.77
CenterNet (Zhou et al. 2019)	43.88	43.54	44.39	43.04	**45.27**	40.72	41.29	23.47
Bus								
FRCNN (Ren et al. 2015)	**29.85**	26.77	25.41	22.08	23.28	27.04	7.49	6.44
YOLO-V3 (Redmon and Farhadi 2018)	**29.63**	25.40	25.55	21.85	27.79	25.87	16.36	15.97
SSD-512 (Liu et al. 2016)	1.96	2.29	2.24	2.01	1.87	3.79	1.81	1.39
RetinaNet (Lin et al. 2018)	11.29	8.67	7.88	6.30	9.04	9.10	**12.44**	11.26
CenterNet (Zhou et al. 2019)	31.31	30.54	29.68	27.09	**33.04**	29.34	21.00	16.66
Motorbike								
FRCNN (Ren et al. 2015)	10.06	10.07	9.02	7.41	6.29	7.00	**10.09**	4.05
YOLO-V3 (Redmon and Farhadi 2018)	6.92	4.64	**7.3**	4.97	5.90	4.64	5.46	3.30
SSD-512 (Liu et al. 2016)	2.02	2.08	2.69	0.91	2.96	2.89	0.84	0.45
RetinaNet (Lin et al. 2018)	5.25	3.44	3.49	3.37	3.29	3.67	**7.53**	1.66
CenterNet (Zhou et al. 2019)	9.38	**12.42**	10.24	7.37	7.29	10.79	7.29	2.98

raindrop images in Table 6. In addition, the semi-supervised method of STL (Wei et al. 2019) becomes the dominant winner on those real images, especially in the rain and mist case, which demonstrates that employing some real images as the training data is able to deal with rain in real-world cases.

5.2 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 (Bradley and Terry 1952) to estimate the subjective score for each method so that they can be ranked, with the exactly same rou-
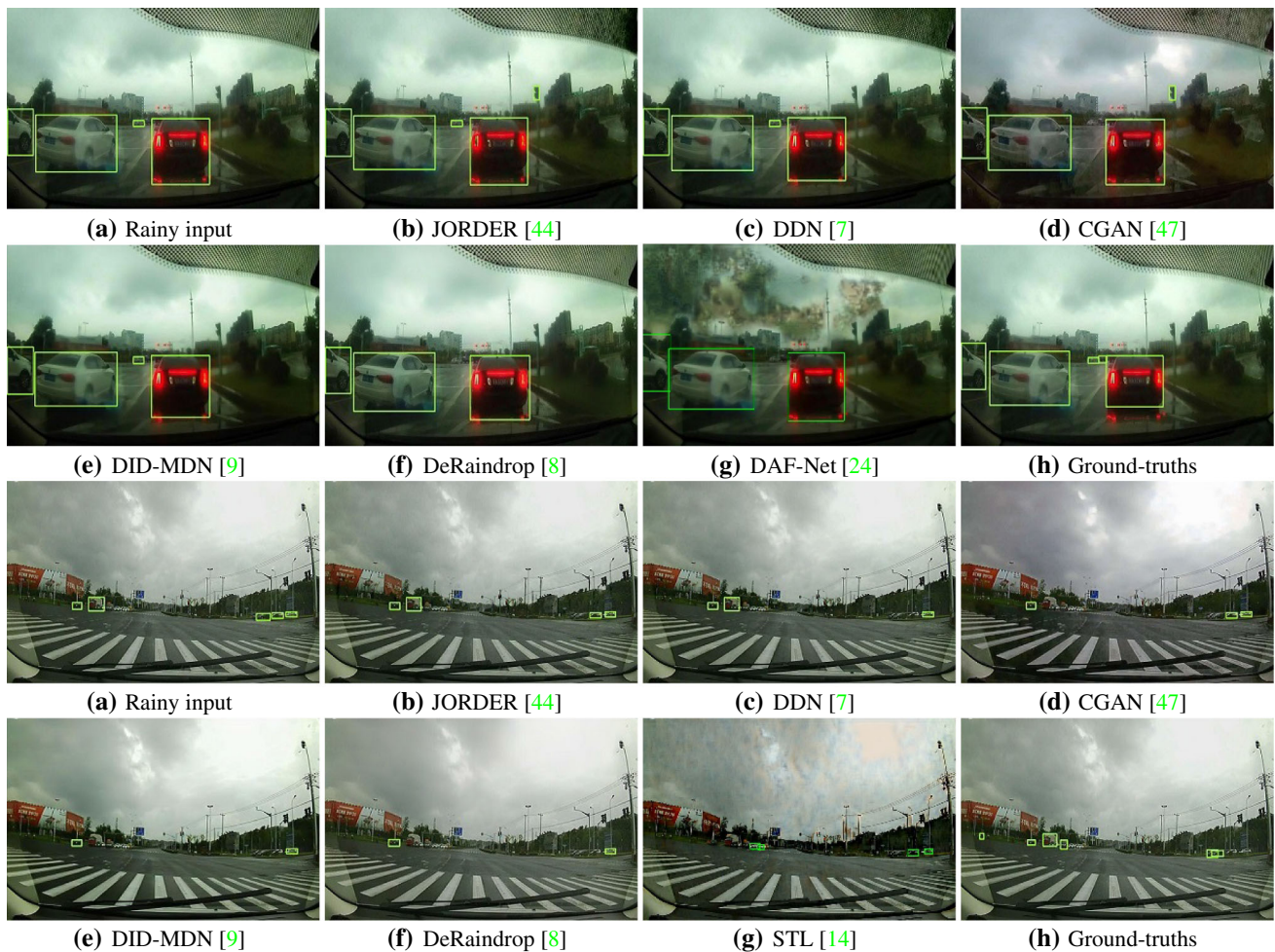
Fig. 3 Visualization of object detection results with YOLO-V3 after applying different deraining algorithms on two images from the **RID** dataset

tine as described in previous similar works (Li et al. 2019a). 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 and mist image is processed with each of the seven deraining algorithms (except DerainDrop Qian et al. 2018), and the seven 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 eight methods of GMM (Li et al. 2016), JORDER (Yang et al. 2017), DDN (Fu et al. 2017b), CGAN (Zhang et al. 2019), DID-MDN (Zhang and Patel 2018), DeRaindrop (Qian et al. 2018), DAF-Net (Hu et al. 2019), and STL (Wei et al. 2019). We collect the pair comparison results of human subject studies from 11 human raters, i.e., we use the paired comparison approach that requires each human subject to choose a preferred image from a pair of derained images. 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 7. 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 Fig. 2, the derained result generated by CGAN is more smooth than others. The main reason is that CGAN does not focus on designing a good prior or a good framework, but focus on ensuring the derained result should be indistinguishable from its corresponding clear image to a given discriminator. Therefore, CGAN generate derained results that is consistent with the human vision. On the raindrop images, it is somehow to our surprise that DerainDrop as well as other deep learning-based models is not favored by users; instead, the non-CNN-based GMM method, showed no advantage under previous objective metrics, was highly preferred by users.
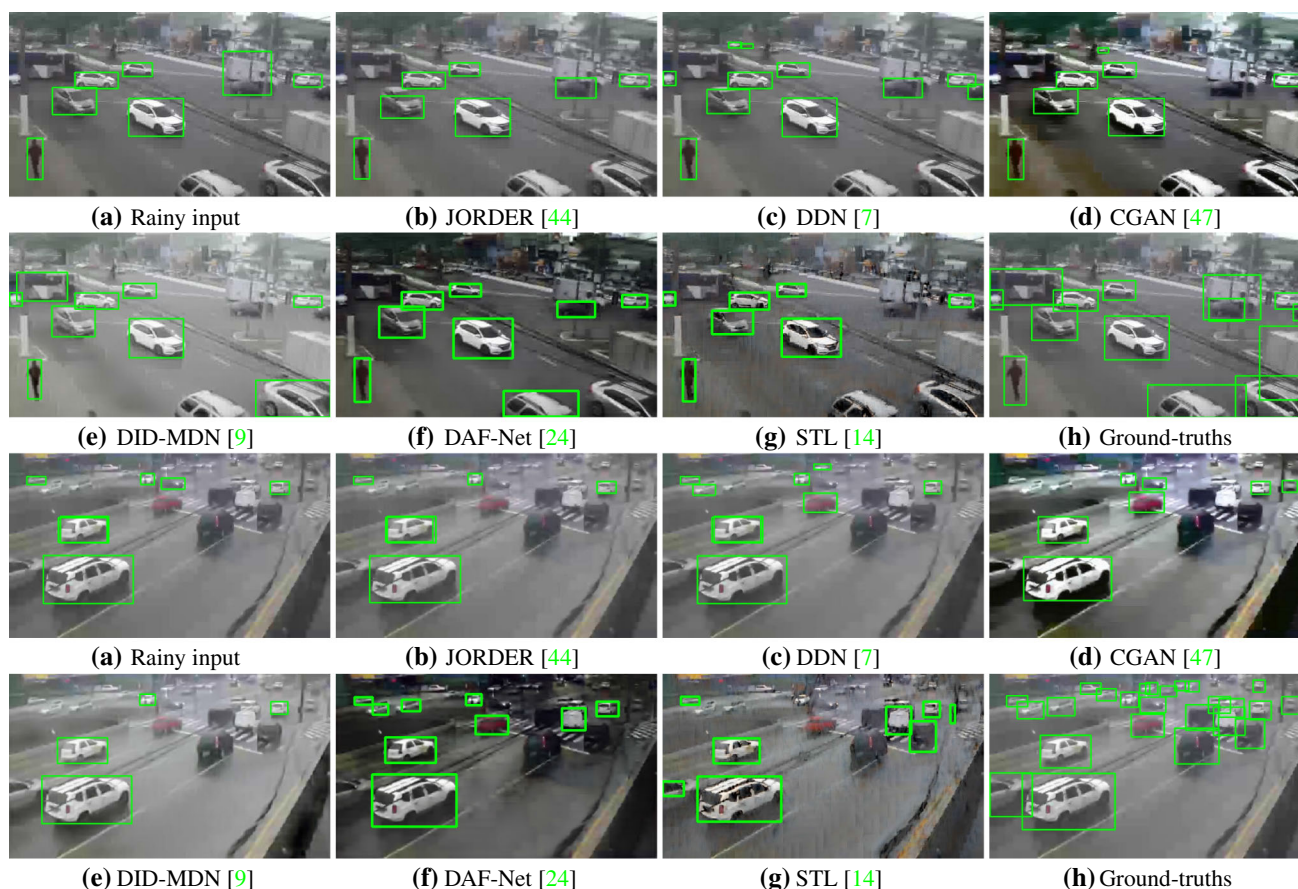
(a) Rainy input   **(b)** JORDER [44]   **(c)** DDN [7]   **(d)** CGAN [47]

(e) DID-MDN [9]   **(f)** DAF-Net [24]   **(g)** STL [14]   **(h)** Ground-truths

(a) Rainy input   **(b)** JORDER [44]   **(c)** DDN [7]   **(d)** CGAN [47]

(e) DID-MDN [9]   **(f)** DAF-Net [24]   **(g)** STL [14]   **(h)** Ground-truths

Fig. 4 Visualization of object detection results with YOLO-V3 after applying different deraining algorithms on two images from the **RIS** dataset

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 the rain-density subnetwork or GAN, that can provide more information of the scene context and hence improve generalization to complex rain conditions.

From Tables 5, 6 and 7, we can found that the 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 (Choi et al. 2015) 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.

5.3 Task-Driven Comparison

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

We then study object detection performance on the derained sets, using several state-of-the-art object detection models: Faster R-CNN (FRCNN) (Ren et al. 2015), YOLO-V3 (Redmon and Farhadi 2018), SSD-512 (Liu et al. 2016), RetinaNet (Lin et al. 2018), and CenterNet (Zhou et al. 2019). FRCNN is two-stage detection model, which recompute features for each potential box, then classify those features. YOLO-V3, SSD-512, RetinaNet are anchor-based one-stage detection model which slide a complex arrangement of candi-

---

[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.

Table 10 Average no-reference evaluations results of derained results on RID images

	JORDER (Yang et al. 2017)	DDN (Fu et al. 2017b)	CGAN (Zhang et al. 2019)	DID-MDN (Zhang and Patel 2018)	DeRaindrop (Qian et al. 2018)	DAF-Net (Hu et al. 2019)	STL (Wei et al. 2019)
SSEQ ↑	**76.9894**	74.1293	53.7457	49.0649	50.9312	41.1361	41.2826
NIQE ↓	**3.4373**	4.3186	6.2422	7.2907	7.6729	11.2619	8.6634
BLINDS-II ↑	86.1200	**86.5510**	51.0408	54.4388	51.0714	37.1500	50.1333

Table 11 Average no-reference evaluations results of derained results on RIS images

	JORDER (Yang et al. 2017)	DDN (Fu et al. 2017b)	CGAN (Zhang et al. 2019)	DID-MDN (Zhang and Patel 2018)	DeRaindrop (Qian et al. 2018)	DAF-Net (Hu et al. 2019)	STL (Wei et al. 2019)
SSEQ ↑	54.8768	53.4738	53.3636	54.6423	54.2766	54.3034	**59.7313**
NIQE ↓	6.11641	6.28624	5.02999	5.4222	5.80866	5.7002	**3.9031**
BLINDS-II ↑	58.772	57.5176	60.8572	57.9191	69.0596	61.7041	**77.1038**

date anchor boxes, over the image and classify them directly without specifying the box content. CenterNet is ancher-free one-stage detection model which detects a pair of corners of a bounding box and groups them to form the final detected bounding box. These are representative detection models in their respective fields. 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 the MS-COCO (Lin et al. 2014) dataset.

The underlying hypothesis behind this evaluation protocol is: (1) an object detector trained on clean natural images will perform the best when the input is also from the clean image domain or close; (2) 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 Kupyn et al. (2018), Li et al. (2019a).

Tables 8 and 9 report the mAP results and AP results in each object class comparison for different deraining algorithms, achieved using five 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, for YOLO-V3, SSD-512, and RetinaNet. Our observation concurs the conclusion of another recent study (on dehazing) (Pei et al. 2018): since those deraining algo-rithms 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.

The two exceptions are FRCNN and CenterNet, where deraining algorithms could help detection a bit particularly on the RID dataset. However, the overall mAP results by FRCNN are often the worst or second worst. That implies a strong domain mismatch, suggesting that FRCNN results might not be as reliable an indicator for deraining performance as the others. In contrast, when combined with a deraining algorithm, CenterNet is almost the best detection methods on both RID and RIS datasets. Particularly, the cascaded manner of DAF-Net following the CenterNet achieves better detection results on the RID set than others, and the manner of DID-MDN following the CenterNet obtains the optimal detection results on the RIS set. This demonstrates that densely connected convolutional neural network based on rain density could apply to surveillance images in rain.

Both results on RID and RIS sets in Tables 8 and 9 show that YOLO-V3 achieves best detection performance, independently of deraining algorithms applied. Figures 3 and 4 show 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 (Redmon and Farhadi 2018).

We further notice a weak correlation in comparing the mAP results with the full- and no-reference evaluation results

on RID (Tables 8, 10) and RIS (Tables 9, 11) images. We can observe this taking STL (Wei et al. 2019) as an example. Despite having obtained the highest SSEQ, NIQE, and BLINDS-II scores on the RIS dataset in Table 11, STL (Wei et al. 2019) has almost the lowest mAP values between all deraining approaches for all detection models performances in Table 9. The main reason may be that the unsupervised training strategy in Wei et al. (2019) is able to output images with sharp edge and contrast information close to real-world images, which consists of the features (e.g., high-frequency details and image edges) in no-reference metrics. However, the derained results by STL have some blocking artifacts as shown in Fig. 4g, which result in the lower detection result by detection models. Besides that, the two best deraining competitors (DAF-Net and DID-MDN) in terms of detection metric did not achieve best results in any no-reference evaluation metric.

All the results on real-world data experiments in terms of non-reference and the proposed task-specific metrics demonstrate that deraining could be further complicated when entangled with other practical degradation. There is no single metric in perfect tune with the human subjective score. Therefore, when designing a deraining algorithm, one needs to be clear about the end purpose. No-reference metrics are more appropriate when measuring the visual effect of real-world images, while the proposed task-specific metric is more reliable for high-level machine task performance.

6 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. While the state-of-the-art image deraining methods can recover satisfactory sharp images on those standard benchmark datasets, they tend to fail on real-world rainy images. The main reason is that real-world images are often degraded by several factors other than a single rainy type, such as low-resolution, low-light, noise, and blur (Kupyn et al. 2019). To deal with the real complicated, varying rains, one might need to consider a mixture model of experts. Another practically useful direction is to develop scene-specific deraining, e.g., for traffic surveillance views.

- There is no single best deraining algorithm under all metrics. The most popular evaluation metrics for image deraining are still PSNR and SSIM. They directly compare the pixel differences between derained images and the ground-truths when available. However, PSNR and SSIM cannot measure the perceptual quality precisely. Therefore, when designing a deraining algorithm, one needs to be clear about its end purpose. In addition, since the 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. Semi-supervised learning (Wei et al. 2019), domain generalization (Chen et al. 2020), or unpaired training (Zhu et al. 2017; Jiang et al. 2019) can take advantage of real data even without clean ground truth. They can potentially boost no-reference metrics and could be interesting to explore. A recent work (Yasarla et al. 2020) seems to make meaningful progress along this direction.

- Existing deraining algorithms are ineffective in dealing with different rain types due to domain gaps between these synthetic images and real-world rainy images, as the rain models (e.g., rain streak, rain drop, and rain & mist) are oversimplified. Therefore, we advocate for more research attention on a better model design to handle rains in a complex and mixed scene.

- 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 pre-processing; there are other domain adaptation type options, e.g., Chen et al. (2018), which we will discuss in future work.

.

References

Barnum, P. C., Narasimhan, S., & Kanade, T. (2010). Analysis of rain and snow in frequency space. *International Journal of Computer Vision*, *86*(2–3), 256.

Bossu, J., Hautière, N., & Tarel, J.-P. (2011). 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.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, *39*(3/4), 324–345.

Chen, W., Yu, Z., Wang, Z. & Anandkumar, A. (2020). Automated synthetic-to-real generalization. *arXiv preprint*arXiv:2007.06965.

Chen, Y., Li, W., Sakaridis, C., Dai, D., & Van Gool, L. (2018). Domain adaptive faster r-cnn for object detection in the wild. In *IEEE conference on computer vision and pattern recognition* (pp. 3339–3348).

Chen, Y.-L. & Hsu, C.-T. (2013). A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *IEEE international conference on computer vision* (pp. 1968–1975).

Choi, L. K., You, J., & Bovik, A. C. (2015). Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing*, *24*(11), 3888–3901.

Dai, D., Sakaridis, C., Hecker, S., Gool, V., & Luc,. (2020). Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. *International Journal of Computer Vision*, *128*(5), 1182–1204.

Dai, D., Wang, Y., Chen, Y., & Van Gool, L. (2016). Is image super-resolution helpful for other vision tasks? In *Winter conference on applications of computer vision*. IEEE.

Ding, X., Chen, L., Zheng, X., Huang, Y., & Zeng, D. (2016). Single image rain and snow removal via guided l0 smoothing filter. *Multimedia Tools and Applications*, *75*(5), 2697–2712.

Eigen, D., Krishnan, D., & Krishnan, R. (2013). Restoring an image taken through a window covered with dirt or rain. In *IEEE international conference on computer vision*.

Fu, X., Huang, J., Ding, X., Liao, Y., & Paisley, J. (2017a). Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, *26*(6), 2944–2956.

Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., & Paisley, J. (2017b). Removing rain from single images via a deep detail network. In *IEEE conference on computer vision and pattern recognition*.

Fu, X., Liang, B., Huang, Y., Ding, X., & Paisley, J. W. (2020). Lightweight pyramid networks for image deraining. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(6), 1794–1807.

Garg, K. & Nayar, S. K. (2004). Detection and removal of rain from videos. In *IEEE conference on computer vision and pattern recognition*.

Garg, K., & Nayar, S. K. (2005). When does a camera see rain? In *IEEE International conference on computer vision*.

Gu, S., Meng, D., Zuo, W., & Zhang, L. (2017) Joint convolutional analysis and synthesis sparse representation for single image layer separation. In *IEEE international conference on computer vision* (pp. 1717–1725).

Hahner, M., Dai, D., Sakaridis, C., Zaech, J.-N., & Van Gool, L. (2019). Semantic understanding of foggy scenes with purely synthetic data. In *Intelligent transportation systems conference* (pp. 3675–3681).

Halder, S. S., Lalonde, J.-F., & de Charette, R. (2019). Physics-based rendering for improving robustness to rain. In *IEEE international conference on computer vision* (pp. 10203–10212).

Hu, X., Fu, C.-W., Zhu, L. & Heng, P.-A. (2019). Depth-attentional features for single-image rain removal. In *IEEE conference on computer vision and pattern recognition* (pp. 8022–8031).

Jiang, T.-X., Huang, T.-Z., Zhao, X.-L., Deng, L.-J., & Wang, Y. (2017). A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors. In *IEEE conference on computer vision and pattern recognition*.

Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., & Wang, Z. (2019). Enlightengan:

Deep light enhancement without paired supervision. *arXiv preprint*arXiv:1906.06972.

Jin, X., Chen, Z., Lin, J., Chen, Z., & Zhou, W. (2019). Unsupervised single image deraining with self-supervised constraints. In *IEEE international conference on image processing* (pp. 2761–2765).

Kang, L.-W., Lin, C.-W., & Fu, Y.-H. (2012). Automatic single-image-based rain streaks removal via image decomposition. *IEEE Transactions on Image Processing*, *21*(4), 1742.

Katrin, H., Ole, J., Daniel, K., & Bastian, G. (2016). A dataset and evaluation methodology for depth estimation on 4d light fields. In *Asian conference on computer vision* (pp. 19–34).

Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., & Matas, J. (2018). Deblurgan: Blind motion deblurring using conditional adversarial networks. In *IEEE conference on computer vision and pattern recognition* (pp. 8183–8192).

Kupyn, O., Martyniuk, T., Wu, J., & Wang, Z. (2019). Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE international conference on computer vision* (pp. 8878–8887).

Lai, W.-S., Huang, J.-B., Hu, Z., Ahuja, N., & Yang, M.-H. (2016). A comparative study for single image blind deblurring. In *IEEE conference on computer vision and pattern recognition* (pp. 1701–1709).

Lei, Z., Fu, C.-W., Dani, L., & Heng, P.-A. (2017). Joint bilayer optimization for single-image rain streak removal. In *IEEE international conference on computer vision*.

Li, B., Peng, X., Wang, Z., Xu, J. & Feng, D. (2017). Aod-net: All-in-one dehazing network. In *IEEE international conference on computer vision* (pp. 4770–4778).

Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2018). End-to-end united video dehazing and detection. In *AAAI conference on artificial intelligence*.

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., et al. (2019a). Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, *28*(1), 492–505.

Li, R., Cheong, L.-F. & Tan, R. T. (2019b). Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *IEEE conference on computer vision and pattern recognition* (pp. 1633–1642).

Li, S., Iago Araujo, B., Ren, W., Wang, Z., Tokuda, E. K., Junior, R. H., Cesar-Junior, R., Zhang, J., Guo, X., & Cao, X. (2019c). Single image deraining: A comprehensive benchmark analysis. In *IEEE conference on computer vision and pattern recognition* (pp. 3838–3847).

Li, Y., Tan, R. T., Guo, X., Lu, J., & Brown, M. S. (2016). Rain streak removal using layer priors. In *IEEE conference on computer vision and pattern recognition* (pp. 2736–2744).

Li, Y., Tan, R. T., Guo, X., Lu, J., & Brown, M. S. (2017). Single image rain streak decomposition using layer priors. *IEEE Transactions on Image Processing*, *26*(8), 3874–3885.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2018). Focal loss for dense object detection. *IEEE transactions on pattern analysis and machine intelligence*.

Lin, T.-Y., Maire, M., Belongie, S. J., Hays, J., Perona, P. Ramanan, D., Piotr. Dollár, C., & Zitnick, L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740–755).

Liu, D., Cheng, B., Wang, Z., Zhang, H., & Huang, T. S. (2019). Enhance visual recognition under adverse conditions via deep networks. *IEEE Transactions on Image Processing*, *28*(9), 4401–4412.

Liu, D., Wen, B., Jiao, J., Liu, X., Wang, Z., & Huang, T. S. (2020). Connecting image denoising and high-level vision tasks via deep learning. *IEEE Transactions on Image Processing*, *29*, 3695–3706.

Liu, D., Wen, B., Liu, X., Wang, Z., & Huang, T. S. (2018). When image denoising meets high-level vision tasks: A deep learning approach.

In *International joint conference on artificial intelligence* (pp. 842–848).

Liu, F., Shen, C., Lin, G., & Reid, I. (2016). Learning depth from single monocular images using deep convolutional neural fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(10), 2024–2039.

Liu, L., Liu, B., Huang, H., & Bovik, A. C. (2014). No-reference image quality assessment based on spatial and spectral entropies. *Signal Processing: Image Communication*, *29*(8), 856–863.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A. C. (2016). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21–37).

Liu, Y., Zhao, G., Gong, B., Li, Y., Raj, R., Goel, N., Kesav, S., Gottimukkala, S., Wang, Z., Ren, W., et al. (2018). Improved techniques for learning to dehaze and beyond: A collective study. *arXiv preprint*arXiv:1807.00202.

Luo, Y., Xu, Y., & Ji, H. (2015). Removing rain from a single image via discriminative sparse coding. In *IEEE international conference on computer vision*.

McCartney, E. J. (1976). *Optics of the atmosphere: Scattering by molecules and particles* (p. 421). New York: Wiley.

Mittal, A., Soundararajan, R., & Bovik, A. C. (2012). Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, *20*(3), 209–212.

Mittal, A., Soundararajan, R., & Bovik, A. C. (2013). Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, *20*(3), 209–212.

Pei, Y., Huang, Y., Zou, Q., Lu, Y., & Wang, S. (2018). Does haze removal help cnn-based image classification? arXiv:1810.05716.

Qian, R., Tan, R. T., Yang, W., Su, J., & Liu, J. (2018). Attentive generative adversarial network for raindrop removal from a single image. In *IEEE conference on computer vision and pattern recognition*.

Redmon, J. & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv:1804.02767.

Ren, D., Zuo, W., Hu, Q., Zhu, P., & Meng, D. (2019). Progressive image deraining networks: A better and simpler baseline. In *IEEE conference on computer vision and pattern recognition* (pp. 3937–3946).

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91–99).

Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., & Yang, M.-H. (2016). Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*.

Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., & Yang, M.-H. (2018b). Gated fusion network for single image dehazing. In *IEEE conference on computer vision and pattern recognition* (pp. 3253–3261).

Ren, W., Pan, J., Zhang, H., Cao, X., & Yang, M.-H. (2020). Single image dehazing via multi-scale convolutional neural networks with holistic edges. *International Journal of Computer Vision*, *128*(1), 240–259.

Ren, W., Tian, J., Han, Z., Chan, A., & Tang, Y. (2017). Video desnowing and deraining based on matrix decomposition. In *IEEE conference on computer vision and pattern recognition*.

Ren, W., Zhang, J., Xiangyu, X., Ma, L., Cao, X., Meng, G., et al. (2018a). Deep video dehazing with semantic segmentation. *IEEE Transactions on Image Processing*, *28*(4), 1895–1908.

Saad, M. A., Bovik, A. C., & Charrier, C. (2012). Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE Transactions on Image Processing*, *21*(8), 3339–3352.

Sakaridis, C., Dai, D., Gool, V., & Luc,. (2018). Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, *126*(9), 973–992.

Santhaseelan, V., & Asari, V. K. (2015). Utilizing local phase information to remove rain from video. *International Journal of Computer Vision*, *112*(1), 71–89.

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, *47*(1–3), 7–42.

Scheirer, W., VidalMata, R., Banerjee, S., RichardWebster, B., Albright, M, Davalos, P., McCloskey, S., Miller, B., Tambo, A., Ghosh, S., et al. (2020). Bridging the gap between computational photography and visual recognition. In *IEEE transactions on pattern analysis and machine intelligence*.

Schops, T., Schonberger, J. L., Galliani, S., Sattler, T., Schindler, K., Pollefeys, M. & Geiger, A. (2017). A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *IEEE conference on computer vision and pattern recognition* (pp. 3260–3269).

Sheng, H., Zheng, Y., Ke, W., Dongxiao, Yu., Cheng, X., Lv, W., et al. (2020). Mining hard samples globally and efficiently for person re-identification. *IEEE Internet of Things Journal*, *7*(10), 9611–9622.

Sun, S.-H., Fan, S.-P., & Wang, Y.-C. F. (2014). Exploiting image structural similarity for single image rain removal. In *IEEE international conference on image processing* (pp. 4482–4486).

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., et al. (2008). A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(6), 1068–1080.

Tokuda, E. K., Lockerman, Y., Ferreira, G. B. A., Sorrelgreen, E., Boyle, D., Cesar-Jr, R. M., et al. (2020). A new approach for pedestrian density estimation using moving sensors and computer vision. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, *6*(4), 1–20.

Wang, T., Yang, X., Xu, K., Chen, S., Zhang, Q., & Lau, R. W. H. (2019). Spatial attentive single-image deraining with a high quality real rain dataset. In *IEEE conference on computer vision and pattern recognition*.

Wang, Z., Chang, S., Yang, Y., Liu, D., & Huang, T. S. (2016). Studying very low resolution recognition using deep networks. In *IEEE conference on computer vision and pattern recognition* (pp. 4792–4800).

Wei, W., Meng, D., Zhao, Q., Xu, Z., Wu, Y. (2019). Semi-supervised transfer learning for image rain removal. In *IEEE conference on computer vision and pattern recognition*.

Xu Q., Wang, Z., Bai, Y., Xie, X., & Jia,H. (2020). Ffa-net: Feature fusion attention network for single image dehazing. In *Conference on artificial intelligence*.

Yang, W., Tan, R. T., Feng, J., Liu, J., Guo, Z., & Yan, S. (2016). Joint rain detection and removal via iterative region dependent multi-task learning. *CoRR, abs/1609.07769*, 2(3).

Yang, W., Tan, R. T., Feng, J., Liu, J., Guo, Z., & Yan, S.(2017). Deep joint rain detection and removal from a single image. In *IEEE conference on computer vision and pattern recognition*.

Yang, W., Yuan, Y., Ren, W., Liu, J., Scheirer, W. J., Wang, Z., et al. (2020). Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing*, *29*, 5737–5752.

Yasarla, R., Sindagi, V. A., & Patel, V. M. (2020). Syn2real transfer learning for image deraining using gaussian processes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2726–2736).

You, S., Tan, R. T., Kawakami, R., Mukaigawa, Y., & Ikeuchi, K. (2016). Adherent raindrop modeling, detection and removal in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(9), 1721–1733.

Yu, Y., Liu, Y., Zhang, H., Chen, S., & Qiao, Y. (2020). FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing. In *Conference on artificial intelligence*.

Zhang, H. & Patel, V. M. (2018). Density-aware single image de-raining using a multi-stream dense network. In *IEEE conference on computer vision and pattern recognition*.

Zhang, H., Sindagi, V., & Patel, V. M. (2019). Image de-raining using a conditional generative adversarial network. In *IEEE transactions on circuits and systems for video technology*.

Zhang, K., Zuo, W., Gu, S., & Zhang, L. (2017) Learning deep CNN denoiser prior for image restoration. In *IEEE conference on computer vision and pattern recognition*.

Zheng, X., Liao, Y., Guo, W., Fu, X., & Ding, X. (2013). Single-image-based rain and snow removal using multi-guided filter. In *International conference on neural information processing*.

Zhou, X., Wang, D., Krähenbühl, P. (2019). Objects as points. arXiv:1904.07850.

Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint*.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.