From Sky to the Ground: A Large-scale Benchmark and Simple Baseline Towards Real Rain Removal

Yun Guo[1,†], Xueyao Xiao[1,†], Yi Chang[1,*], Shumin Deng[1], Luxin Yan[1]
[1]National Key Laboratory of Science and Technology on Multispectral Information Processing,
School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, China
{guoyun, xiaoxueyao, yichang, shumindeng, yanluxin}@hust.edu.cn
https://github.com/yunguo224/LHP-Rain

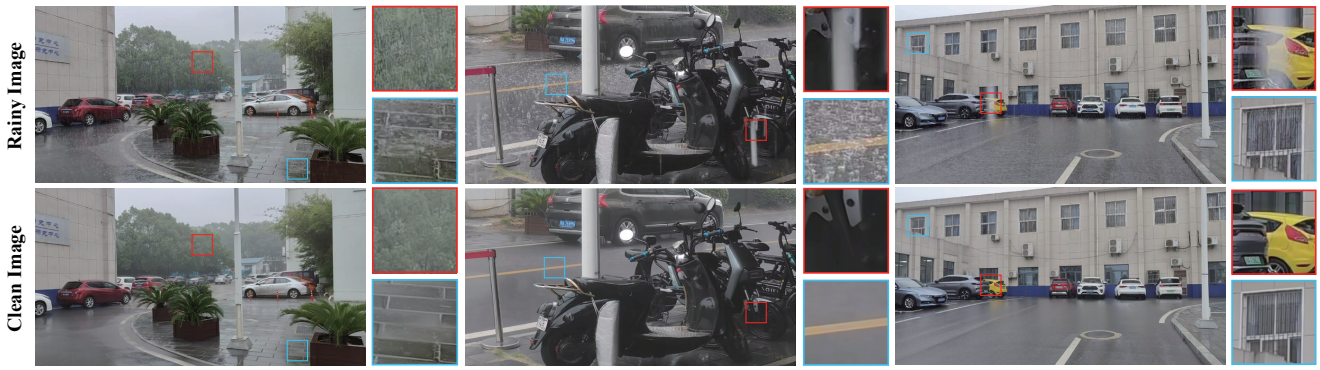

Figure 1: Illustration of the proposed paired rainy/clean images dataset. The proposed real rain dataset is to cope with various rain categories, such as rain streaks, veiling effect, occlusion, and **ground splashing**. We recommend **zooming in** the figure on PC for better visualization.

Abstract

*Learning-based image deraining methods have made great progress. However, the lack of large-scale high-quality paired training samples is the main bottleneck to hamper the real image deraining (RID). To address this dilemma and advance RID, we construct a Large-scale High-quality Paired real rain benchmark (LHP-Rain), including 3000 video sequences with 1 million high-resolution (1920*1080) frame pairs. The advantages of the proposed dataset over the existing ones are three-fold: rain with higher-diversity and larger-scale, image with higher-resolution and higher-quality ground-truth. Specifically, the real rains in LHP-Rain not only contain the classical rain streak/veiling/occlusion in the sky, but also the **splashing on the ground** overlooked by deraining community. Moreover, we propose a novel robust low-rank tensor recovery model to generate the GT with better separating the static background from the dynamic rain. In addition, we design a simple transformer-based single image deraining baseline, which simultaneously utilize the self-attention and cross-layer attention within the image and rain layer with discriminative feature representation. Extensive experiments verify the superiority of the proposed dataset and deraining method over state-of-the-art.*

雨絲（rain streaks）
霧化效果（veiling）
遮擋效果（occlusion）

1. Introduction

Single image deraining is to improve the imaging quality by separating rain from image background. In recent years, significant progress has been made in learning-based single image deraining by various sophisticated CNN architectures [14, 36, 53, 10] and powerful Transformer models [35, 44]. Although these state-of-the-art supervised methods have achieved impressive results on simulated datasets, a fact cannot be ignored that those competitive methods perform unsatisfactory on diverse real rainy scenes. The core reason is the domain shift issue between the simplified synthetic rain and complex real rain [50, 32, 41, 51].

To solve this problem, an intuitive idea is to try the best to make rain degradation model as real as possible [11]. The researchers formulate the rain imaging procedure into a comprehensive rain simulation model [9, 48, 12, 24, 18],

[†]These authors contributed equally to this work.
[*]Corresponding author.

Table 1: Summary of existing real rain datasets.

Datasets	Year	Source	Sequence	Frame	Resolution	Rain Categories	Annotation	Paired
RID/RIS[19]	2019	Cam/Internet	None	4.5K	640*368	streak, raindrop	Object detection	-
NR-IQA[43]	2020	Internet	None	0.2K	1000*680	streak, veiling	None	-
Real3000[25]	2021	Internet	None	3.0K	942*654	streak, veiling	None	-
FCRealRain[51]	2022	Camera	None	4.0K	4240*2400	streak, veiling	Object detection	-
SPA-Data[37]	2019	Cam/Internet	170	29.5K	256*256	streak	None	✓
RainDS[32]	2021	Cam	None	1.0K	1296*728	streak, raindrop	None	✓
GT-Rain[1]	2022	Internet	202	31.5K	666*339	streak, veiling	None	✓
RealRain-1K[20]	2022	Cam/Internet	1120	1.1K	1512*973	streak, veiling, occlusion	None	✓
LHP-Rain	2023	Camera	**3000**	**1.0M**	1920*1080	streak, veiling, occlusion, **splashing**	Object detection/**Lane**	✓

in which different visual appearance of rain streaks[9], accumulation veiling [48], haze [12, 18], and occlusion [24] factors are taken into consideration. Unfortunately, these linear simulation models still cannot well accommodate the distribution of realistic rains. For example, in Fig. 1, realistic rain streak is usually not exactly a regular line-pattern streak but possesses irregular non-uniform in terms of the intensity and width. Apart from the rain streaks, the existing rain simulation models could not handle the complicated rain splashing on the ground, which presents as dense point-shape texture, droplets or water waves, ruining visibility of traffic signs, such as lane lines, and also causes enormous negative effects for the high-level vision.

Another research line obtains the 'clean' counterpart from the realistic rainy videos [37, 20], which leverages the motion discrepancy between static image background and dynamic rain. Unfortunately, they simply employ naive filtering strategies such as percentile filter [37] and median filter [20], resulting in unsatisfactory GT with residual rain or over-smoothing phenomenon. Moreover, the number and diversity of the existing real paired rain datasets are still limited. Few datasets have considered the rain splash on the ground, which is commonly observed in the real world but still rarely mentioned in deraining community. And the number of existing video sequences and image frames are not sufficient to cover diverse rains in terms of the varied rain angle, intensity, density, length, width and so on. Last but not least, the existing realistic rainy images are mostly downloaded from the Internet with low-quality: compression, watermark, low-resolution, without annotation and so on. As such, constructing a large-scale high-quality paired realistic rain dataset is highly necessary.

In this work, we construct a new large-scale high-quality paired real rain benchmark. The strength of our benchmark is threefold. First, the LHP-Rain contains diverse rain categories with very large-scale, including 3000 video sequences with over 1 million frame pairs. Second, apart from the conventional streak and veiling, our benchmark is capable of removing the representative challenging ground splashing rain in the real world. Third, the LHP-Rain is collected by the smartphone with high-resolution (1920*1080 pixels) and abundant objects under self-driving and surveillance scenes are captured for comprehensive evaluation. Moreover, we propose a novel robust low-rank tensor recovery method (RLRTR) for video deraining, which can generate higher-quality GT with better rain removal from sky to the ground and image structure preserving. We summary the main contributions as follows:

- We construct a large-scale high-quality paired real rain benchmark for real single image deraining. To our best knowledge, LHP-Rain is the largest paired real rain dataset (3000 video sequences, 1 million frames) with high image resolution (1920*1080), and the first benchmark to claim and tackle the problem of ground splashing rain removal.
- We design a novel robust low-rank tensor recovery model for video deraining to better acquire paired GT. We provide detailed analysis to show RLRTR can better differ the rain from static background than previous datasets.
- We propose a new transformer-based single image deraining baseline, which exploits both self-attention and cross-layer attention between the rain and image layer for better representation. Extensive experiments on different real datasets verify the superiority of proposed method.

2. Related Work

Real rain datasets. At present, the researchers have mostly focused on the network architecture design, while relative fewer attention has been paid on the real rain dataset. The insufficiency of realistic rain dataset is the main bottleneck to hamper single image deraining. In Table 1, we provide a comprehensive summary of existing real rain datasets, which can be classified into two categories: rainy image only and paired rain-image. The former can be utilized via semi-supervised [40, 13] or unsupervised methods [7, 51]. The latter can be conveniently utilized by the supervised training.

The key to paired real dataset is how to acquire the pseudo-'clean' image from its rainy counterpart. There are two main ways to construct the pairs: video-based generation (SPA-data [37], RealRain-1K [20]), and time-interval acquisition (RainDS [32], GT-Rain [1]). All these datasets should ensure that the camera is strictly immobile during the acquisition process. SPA-data [37] was the first presented paired real dataset which utilized the human-supervised percentile video

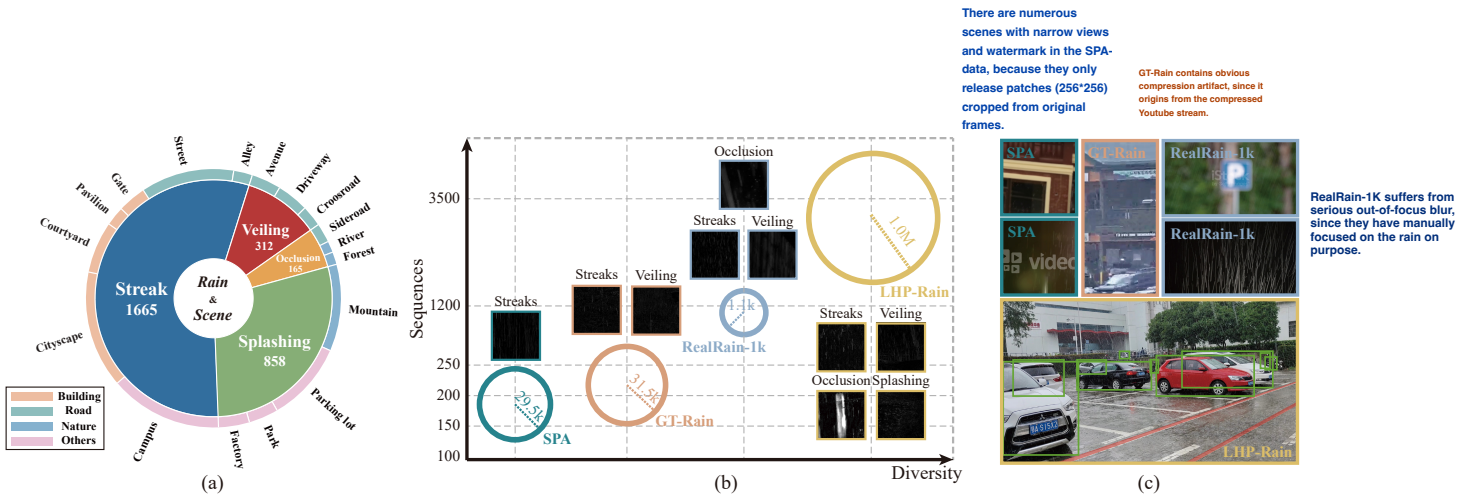

Figure 2: Features of the proposed benchmark LHP-Rain. (a) Distribution of rain and scene of the proposed benchmark. (b) Our proposed LHP-Rain outperforms others in terms of rain diversity and sequence amount. (c) LHP-Rain collects high-resolution and annotated rainy images without copyright, compression and blur.

filtering to obtain the GT. Instead of generation, GT-Rain [1] collected the pairs of same scene under rainy and good weather, respectively. Similar idea has been adopt in RainDS [32] by manually mimicking rainfall with a sprinkler.

Despite the rapid development promoted by those pioneer datasets, there remain some important issues to be solved: insufficient number and rain diversity (not consider ground splashing). In this work, we contribute a large-scale high-quality paired real rain dataset (Section 3.1) with diverse rain and abundant objects. Moreover, a novel GT generation method is proposed with higher-quality pairs (Section 3.3).

Single image deraining. The single image deraining methods have made great progress in last decade including the optimization models [26, 22, 4], deep convolutional network [12, 36, 10] and transformer [35, 44]. Fu *et al*. [9] first introduced the residual deep CNN for single image deraining. Latter, the researchers have further improved the network depth by stacking similar modules, such as the well-known recurrent [21, 49] and multi-stage progressive [33, 53] strategies. Meanwhile, the multi-scale has been widely explored to improve the representation such as the multi-scale fusion [14] and wavelet decomposition [46]. Further, the side information about the rain attribute: density [54], depth [12], directionality [38], location [47], non-local [17] have been extensively utilized to enhance the deraining performance.

Benefiting from self-attention mechanise for long-range relationships modelling, transformer-based methods have achieved significant performance for single image deraining [39, 35, 44, 6]. Very recently, Xiao *et al*.[44] proposed an image deraining transformer (IDT) with relative position enhanced and spatial-based multihead self-attention. Chen *et al*. [6] proposed a sparse Transformer architecture to solve the redundant feature issue. In this work, we propose a simple yet effective dual-branch transformer baseline which simultaneously utilizes the self-attention within rain/image layer and cross-layer attention between the rain and image layer, so as to jointly improve the discriminative disentanglement between the rain and image layer (Section 4).

3. Large-scale high-quality paired Benchmark

3.1. Benchmark Collection and Statistics

Due to the difficulty and inconvenience of collecting real rain videos, the video sequences and frames of existing paired real rain datasets are still limited as shown in Table 1. In this work, we collect the real rain sequences by smartphones with 24mm lens focal length, sampled in 30 fps. The data collection process is illustrated in Fig. 3(a). Firstly, to keep the camera immobile, we employ tripod to capture real rain videos with static background (no moving object except rain). For each sequence, we record approximate 15 seconds and extract the intermediate steady 10s into our dataset. Then, we manually pick out moving object to remove unexpected disturbance. Finally, we employ the proposed RLRTR (Section 3.3) to obtain high-quality GT.

Overall, we collect 3000 video sequences with approximate 1 million frames across 8 cities from 4 countries around the world, China (2091 sequences), England (51 sequences), Philippines and Indonesia (858 sequences). We visualize the per-country image counts and location distribution of LHP-Rain in Fig. 3(b). The rainfall levels are varied from light rain (10mm/day) to rainstorm (300mm/day) due to diversity of local climates. Over 17 typical scenes are captured, including the parking lot, street, alley, playground, courtyard, and forest, etc. Besides, 2490 sequences are captured at daytime and 510 sequences are at night. More Rainy/GT pairs from LHP-Rain are displayed in Fig. 3(c). With the changing of location, backgrounds are varied from nature to city with diverse rain patterns, such as rain streak, veiling effect, occlusion and splashing. Note that, although the background is the same for each video sequence, the rain in each frame is vastly different from the appearance, including streak, veiling, occlusion and splashing. Here we separate 2100 sequences as training set, 600 sequences as validation set and the other 300 sequences as test set. To further visualize the quantity distribution of rain and scene for each sequence, a sunburst chart is illustrated in Fig. 2(a).

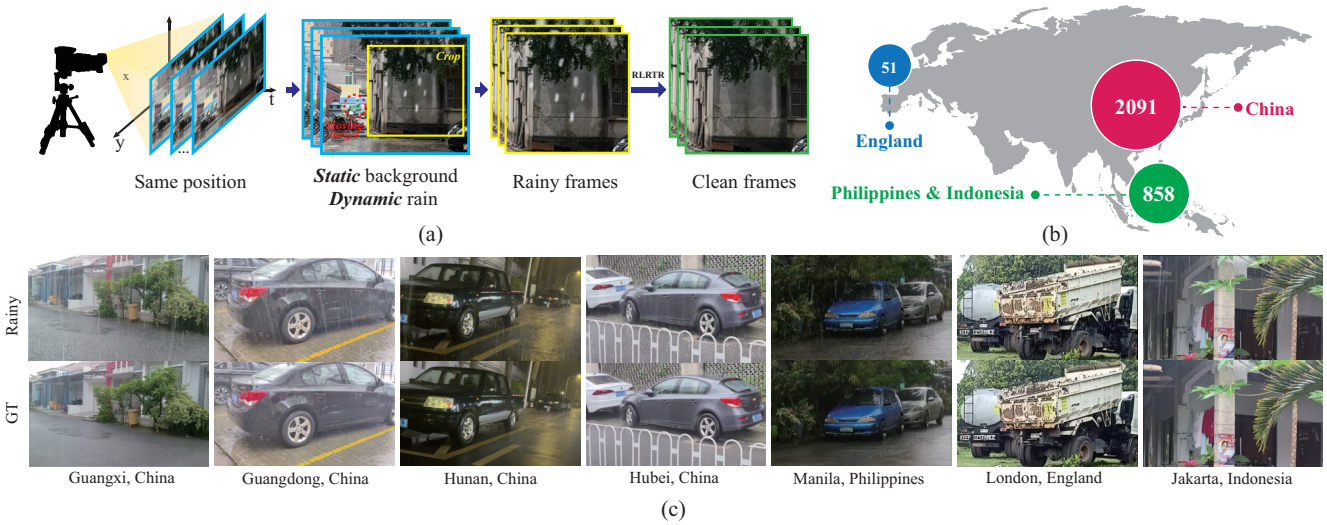

Figure 3: Illustration of the proposed benchmark LHP-Rain. (a) Overall procedure of obtaining rainy/clean image pair. (b) Quantity and location distribution of LHP-Rain. (c) Rainy/GT samples of LHP-Rain from different locations. Scenes are varied from nature to city, over day and night with diverse rain patterns from sky to the ground.

3.2. Benchmark Features

Rain with higher-diversity and larger-scale. We concern not only the sequence number and total frames but also the rain diversity of realistic rain, which both have great impact on the generalization for real-world rain. In Fig. 2(b), we show the statistic distribution of rain diversity and sequences/frames in typical real paired rain dataset. All datasets concern about the noticeable rain streak, especially SPA-data [37]. RainDS [32] additionally takes the raindrop into consideration with 1000 frames, while GT-Rain [1] and RealRain-1K [20] further capture the accumulation veiling artifact in heavy rain. LHP-Rain contains not only rain streak and veiling in the sky, but also challenging highlight occlusion and splashing water on the ground. To our best knowledge, the proposed LHP-Rain is the first benchmark to collect and tackle ground splashing rain in the real world, which is commonly ignored by previous existing datasets.

Image with higher-resolution and abundant objects. The existing datasets pay much attention to the rain, ignoring that the high-quality image is also what we really need. Unfortunately, existing realistic rainy images are generally downloaded from the Internet with various problems about the image: compression artifact, watermark, low-resolution, out-of-focus blur, without objects to name a few, which may cause challenges for high-level vision applications. In Fig. 2(c), we show the typical examples in each dataset. The image background of RealRain-1K [20] suffers from serious out-of-focus blur, since they have manually focused on the rain on purpose. GT-Rain [1] contains obvious compression artifact, since it origins from the compressed Youtube stream. There are numerous scenes with narrow views and watermark in the SPA-data [37], because they only release patches (256*256) cropped from original frames.

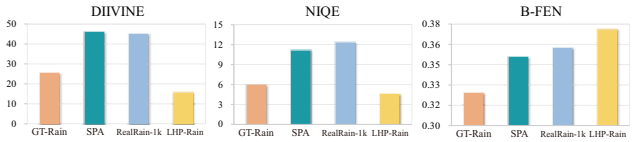

Figure 4: The GT quality of LHP-Rain is superior to others on DIIVINE (lowest), NIQE (lowest) and B-FEN (highest).

To improve the image quality, we personally capture high-resolution (1920*1080) realistic rain videos by smartphones. Moreover, LHP-Rain is not only designed for rain restoration, but also important for object detection and segmentation tasks under adverse weather, with abundant objects which are oriented for self-driving and video surveillance scenes. Thus, we provide annotations for object detection and lane segmentation. Five typical objects including person, car, bicycle, motorcycle and bus are annotated by bounding box with 326,961 instances totally. For lane segmentation, we annotate 24,464 lane masks to evaluate the effect of rain splashing removal. Note that the same object in different frames will be regarded as different instances because rain is inconstant and changing frame by frame.

Higher-quality ground-truth. The quality of GT is critical for paired real rain dataset. It is difficult to determine what is good or bad GT in an absolutely fair way. In this paper, we assume that the better the rain removal, the better the image quality is. Therefore, we employ several no reference image quality assessments: DIIVINE [29], NIQE [28] and B-FEN [43], to evaluate the image quality of the rain-free image. The former two are hand-crafted based general image quality indexes, and the last one B-FEN is the learning based index especially designed for de-raining quality assessment. We select all the video backgrounds in SPA-data [37], GT-Rain [1], RealRain-1K [20] and LHP-Rain for evaluation. In

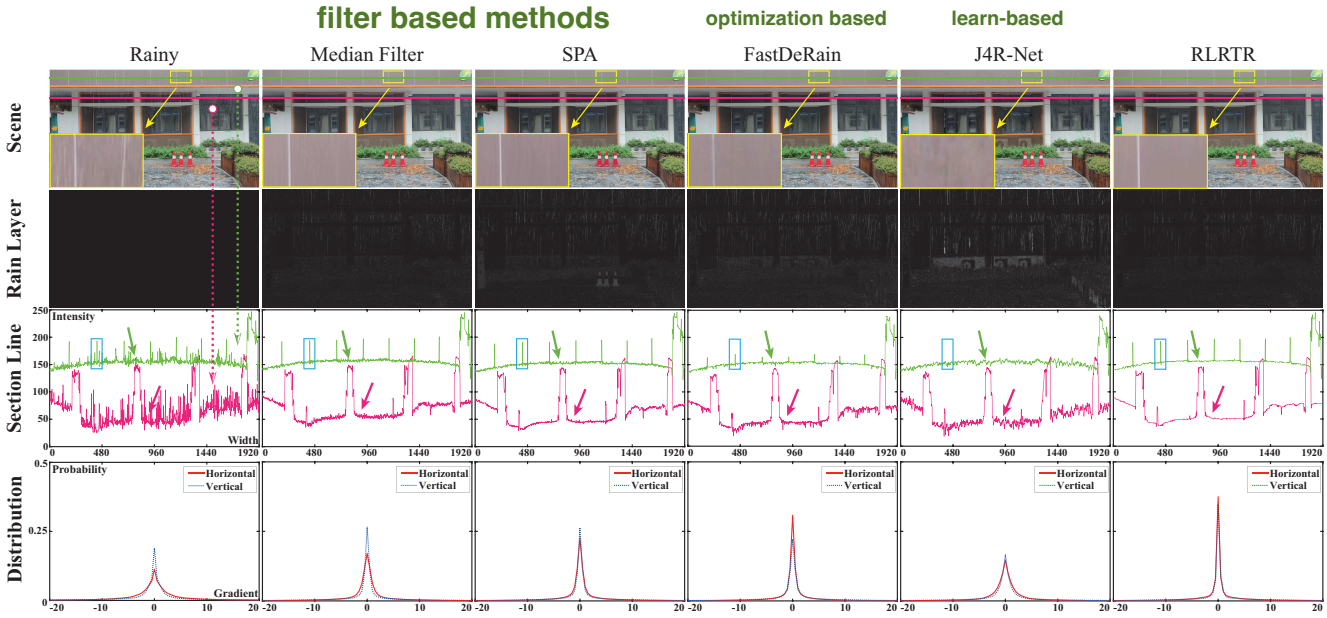

Figure 5: Analysis of different video deraining results on our dataset. From left to right, the first column is the original rainy frame, and the remaining five columns represent different methods, namely Median Filter, SPA, FastDeRain, J4R-Net and the proposed RLRTR. From top to bottom, the first row shows the deraining results, the second row is the rain layer of the deraining results, the third row denotes the section line of the deraining results and the last row represents the horizontal and vertical gradient distributions of the deraining results.

Fig. 4, we can observe that the proposed LHP-Rain consistently obtains the best results in terms of different evaluation indexes which strongly support the higher-quality of the GT.

3.3. Robust Low-rank Tensor Recovery Model

Given the rainy video $\mathcal{O} \in \mathbb{R}^{h\times w\times t}$, the key is how to properly obtain the paired GT. The existing methods simply employ the naive filtering technique benefiting from the temporal consistency of the static background. Due to the slight camera vibration caused by the wild wind, we further leverage an affine transformation operator τ [52] to achieve the pixel-level alignment of each frame. Thus, a multi-frame rainy video can be described as the following formula:

$$\mathcal{O}\circ\tau = \mathcal{B} + \mathcal{R} + \mathcal{N}, \tag{1}$$

where $\mathcal{B} \in \mathbb{R}^{h\times w\times t}$ is the rain-free video, $\mathcal{R} \in \mathbb{R}^{h\times w\times t}$ represents the rains, $\mathcal{N} \in \mathbb{R}^{h\times w\times t}$ denotes the random noise, and τ denotes the affine transformation to ensure the rainy video of each frame is pixel-level aligned. In this work, we formulate the video deraining into inverse problem via the *maximum-a-posterior* as follow:

$$\min_{\mathcal{B},\mathcal{R},\tau} \frac{1}{2}||\mathcal{B}+\mathcal{R}-\mathcal{O}\circ\tau||_F^2 + \omega P_b(\mathcal{B}) + \mu P_r(\mathcal{R}), \tag{2}$$

where P_b and P_r are the prior knowledge for the image and rain, respectively, ω and μ are the corresponding hyper-parameters. As for the aligned rainy video, when there are no moving objects except the rain, the rain-free background image is the same for all rainy frames. That is to say, clean video $\mathcal{B}$ has extreme *global* low-rank property along the temporal dimension, ideally its rank is equal to one for each scene. On the other hand, the clean video $\mathcal{B}$ also has very *non-local* low-rank property along the spatial dimension, due to the self-similarity widely employed in image restoration [8]. Moreover, we further take the *local* smoothness of the video $\mathcal{B}$ into consideration via the total variation (TV) regularization [4]. Thus, the joint global-nonlocal-local prior along both the spatial and temporal dimension has been fully exploited for better representation of the static video $\mathcal{B}$:

$$P_b(\mathcal{B}) = \omega\sum_i \left(\tfrac{1}{\lambda_i^2}||\mathcal{S}_i\mathcal{B}\times_3 Q_i - \mathcal{J}_i||_F^2 + ||\mathcal{J}_i||_{tnn}\right) + \gamma||\nabla_t\mathcal{B}||_1, \tag{3}$$

where $\mathcal{S}_i\mathcal{B} \in \mathbb{R}^{p^2\times k\times t}$ is the constructed 3-D tensor via the non-local clustering of a sub-cubic $u_i \in \mathbb{R}^{p\times p\times t}$ [3], p and k are the spatial size and number of the sub-cubic respectively, $Q_i \in \mathbb{R}^{d\times t}(d \ll t)$ is an orthogonal subspace projection matrix used to capture the temporal low-rank property, $\times_3$ is the tensor product along the temporal dimension [16], $\mathcal{J}_i$ represents the low-rank approximation variable, $||\bullet||_{tnn}$ means the tensor nuclear norm for simplicity [3], ∇_t is the difference operator, γ and λ_i is the regularization parameters. As for the rain $\mathcal{R}$, we formulate it as the sparse error [42] via the L_1 sparsity. Thus, the Eq. (2) can be expressed as:

$$\begin{aligned}\left\{\hat{\mathcal{B}},\hat{\mathcal{R}},\hat{\mathcal{J}}_i,\hat{\tau},\hat{Q}_i\right\} = \arg\min_{\mathcal{B},\mathcal{R},\mathcal{J}_i,\tau,Q_i} \tfrac{1}{2}||\mathcal{B}+\mathcal{R}-\mathcal{O}\circ\tau||_F^2 \\ +\mu||\mathcal{R}||_1 + \omega\sum_i\left(\tfrac{1}{\lambda_i^2}||\mathcal{S}_i\mathcal{B}\times_3 Q_i - \mathcal{J}_i||_F^2 + ||\mathcal{J}_i||_{tnn}\right) + \gamma||\nabla_t\mathcal{B}||_1.\end{aligned} \tag{4}$$

Optimization. Due to the difficulty of estimating multiple variables directly, we adopt the alternating minimization

Table 2: Quantitative comparisons with SOTA supervised methods on paired real datasets SPA-data (A), GT-Rain (B) and proposed LHP-Rain (C) under 9 different task settings. X→Y means training on the dataset X and testing on the dataset Y. The degraded results of the three datasets are also provided. Top 1_{st} and 2_{nd} results are marked in **red** and **blue** respectively.

Method	A→A		B→A		C→A		A→B		B→B		C→B		A→C		B→C		C→C	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Rainy Image	32.60 / 0.9173								19.48 / 0.5849						29.97 / 0.8497			
SPANet	38.53	0.9875	22.93	0.8207	31.46	0.9612	20.01	0.6148	21.51	0.7145	19.20	0.5706	28.00	0.8905	20.10	0.8061	31.19	**0.9346**
PReNet	37.05	0.9696	22.44	0.7713	**32.46**	0.9387	20.29	0.5860	20.65	0.6005	19.34	0.5530	27.57	0.8595	20.91	0.7222	32.13	0.9177
RCDNet	39.74	0.9661	22.51	0.8392	32.30	0.9378	20.09	0.5785	21.04	0.6106	19.09	0.5264	25.46	0.7959	21.38	0.8047	32.34	0.9152
JORDER-E	40.63	0.9794	23.47	0.7426	31.23	0.9234	19.98	0.5799	21.24	0.6854	18.76	0.4861	27.13	0.8531	22.14	0.8433	31.24	0.8847
MPRNet	46.06	0.9894	24.27	0.8428	32.37	0.9379	19.87	0.6286	22.00	0.6515	19.47	0.5889	28.41	0.8807	**23.82**	0.8052	33.34	0.9309
GT-Rain	37.21	0.9827	**25.30**	**0.9243**	26.46	0.9145	20.07	**0.6941**	**22.51**	**0.7300**	**21.14**	0.5698	28.62	0.8675	23.19	0.8098	32.18	0.9132
Uformer-B	**46.42**	**0.9917**	24.08	0.8979	32.21	**0.9667**	19.70	0.6875	21.60	0.7124	19.10	**0.6622**	**28.74**	**0.9262**	22.91	**0.8734**	**33.56**	0.9317
IDT	45.74	0.9889	23.80	0.8334	32.38	0.9422	**20.34**	0.6306	21.98	0.6536	19.44	0.5977	26.90	0.8742	23.34	0.7897	33.02	0.9310
SCD-Former	**46.89**	**0.9941**	**26.13**	**0.9122**	**34.38**	**0.9798**	**20.98**	**0.6985**	**22.79**	**0.7684**	**21.71**	**0.6893**	**29.41**	**0.9127**	**23.56**	**0.8626**	**34.33**	**0.9468**

scheme to solve the Eq. (4) with respect to each variable.

1) *Affine Transformation τ*: Since $\mathcal{O} \circ \tau$ is a nonlinear geometric transform, it's difficult to directly optimize τ. A common technique is to linearize around the current estimate and iterate as follows: $\mathcal{O} \circ \tau + \nabla\mathcal{O}\triangle\tau = \mathcal{B} + \mathcal{R} + \mathcal{N}$ [31], where $\nabla\mathcal{O}$ is the Jacobian of the image $\mathcal{O}$ with respect to τ. This method iteratively approximates the original nonlinear transformation with a locally linear approximation [31].

2) *Rain Estimation $\mathcal{R}$*: By ignoring variables independent of $\mathcal{R}$, we can obtain following subproblem:

$$\hat{\mathcal{R}} = \arg\min_{\mathcal{R}} \tfrac{1}{2}||\mathcal{B} + \mathcal{R} - \mathcal{O} \circ \tau||_F^2 + \mu||\mathcal{R}||_1. \quad (5)$$

Eq. (5) is a L_1 minimization problem which can be easily solved by soft thresholding with closed-form solution [23].

3) *Subspace Projection Q_i*: We enforce the orthogonal constraint on $Q_i^T Q_i = I$ with the following subproblem:

$$\hat{Q}_i = \arg\min_{Q_i^T Q_i = I} \frac{1}{\lambda_i^2}||\mathcal{S}_i\mathcal{B}\times_3 Q_i - \mathcal{J}_i||_F^2. \quad (6)$$

According to [45], Eq. (6) has the closed-form solution, which can be obtained by the *rank-d* singular value decomposition of the folding matrix of $\mathcal{S}_i\mathcal{B}$, where d is the measurement of the intrinsic subspace of the temporal dimension. In this work, we empirically set $d \leq 3$.

4) *Low-rank Approximation $\mathcal{J}_i$*: Dropping the irrelevant variables, we can get following subproblem:

$$\hat{\mathcal{J}}_i = \arg\min_{\mathcal{J}_i} \tfrac{1}{\lambda_i^2} \parallel \mathcal{S}_i\mathcal{B}\times_3 Q_i - \mathcal{J}_i \parallel_F^2 + ||\mathcal{J}_i||_{tnn}. \quad (7)$$

This is a typical tensor nuclear norm minimization problem, can be solved by singular value thresholding algorithm [2, 3].

5) *Clean Video Estimation $\mathcal{B}$*: We fix the other variables and optimize $\mathcal{B}$ with the following subproblem:

$$\min_{\mathcal{B}} \tfrac{1}{2}||\mathcal{B} + \mathcal{R} - \mathcal{O} \circ \tau||_F^2 + \omega\textstyle\sum_i \frac{1}{\lambda_i^2}||\mathcal{S}_i\mathcal{B}\times_3 Q_i - \mathcal{J}_i||_F^2 + \gamma||\nabla_t\mathcal{B}||_1. \quad (8)$$

Due to the non-differentiability of the L_1 norm in Eq. (8), we apply the ADMM [23] to decouple this problem into several sub-problems with closed-form solutions. Please refer to the supplementary material for the whole algorithm details.

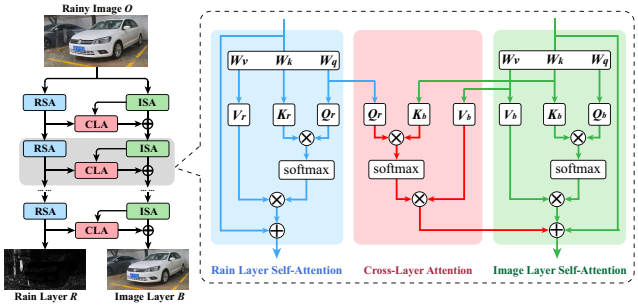

Figure 6: Overall framework of the SCD-Former. It utilizes self and cross-layer attention from rain layer to image layer which serves as side information to recover image layer.

Discussion. Figure 5 illustrates the comparison results of representative video deraining methods: filter-based methods (median filter[20], SPA[37]), optimization-based (FastDeRain[15]) and learning-based methods (J4R-Net[24]). The first and second rows show the video deraining of the image and rain layer, respectively. RLRTR removes almost all the rain from sky to the ground, including rain streaks and splashing, and preserve the image details well, while other methods more or less have rain streaks residual but also cause noticeable damage to the image. In third row, we randomly choose two 1D section lines of the deraining results. The section line of RLRTR is smoother with less burr than other methods. Moreover, the SPA, FastDeRain and the J4R-Net unexpectedly attenuate the spike signal of the share edge, while the proposed RLRTR has well preserve the spike signal. It is well-known the natural image is isotropic and its gradient distribution along different directions should be close to each other [34]. Compared with other results, in forth row the gradient distributions along the vertical and horizontal directions of RLRTR are most similar to each other, which further indirectly verify the naturalness of the deraining result and produce better paired clean-rainy GT.

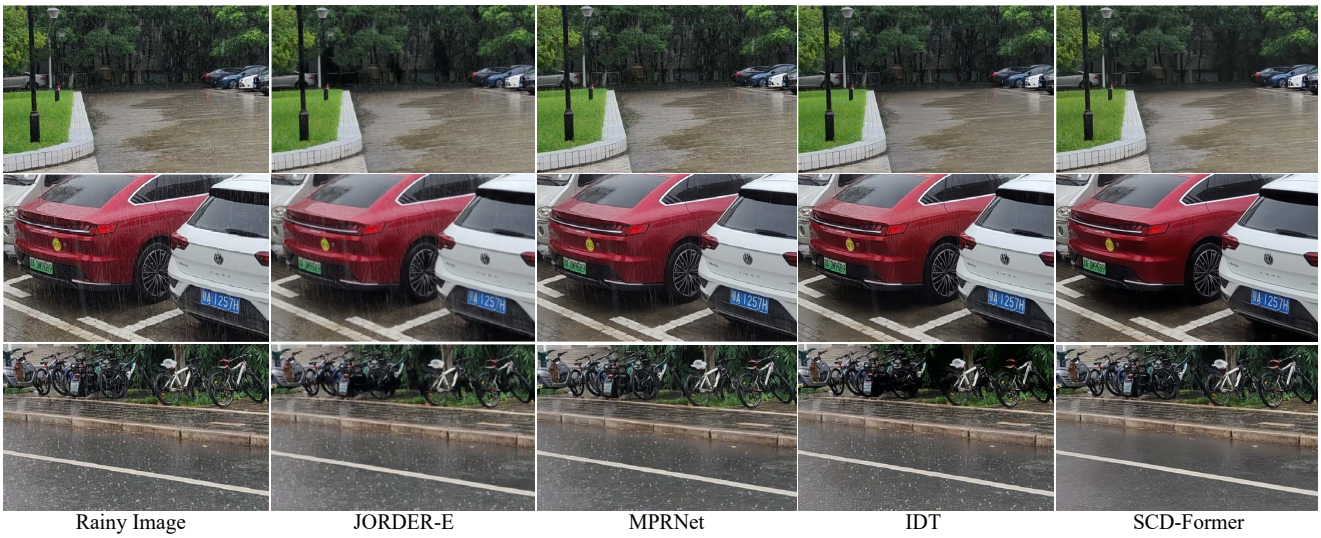

Figure 7: Visual comparisons on LHP-Rain. Comparing with state-of-the-arts, SCD-Former achieves more visual pleasing deraining results and it is capable of removing the highlight occlusion on the car and the splashing water on the ground.

4. SCD-Former: Image Deraining Baseline

The degradation procedure can be formulated as:

$$O = B + R. \tag{9}$$

It has been proved that the rain such as location [47] serving as an attention would be informative in CNN-based image restoration. In this work, we show the rain attention would also be beneficial in transformer. In Fig. 6, we design a simple two-stream Self- and Cross-attention Deraining Transformer (SCD-Former), in which the two-stream network is designed to restore rain and image layer respectively. On one hand, we utilize the self-attention in each rain/image stream independently; on the other hand, we further exploit the cross-layer attention between the rain and image streams. Thus, the rain collaboratively interactive with the image layer to further improve the discriminative representation.

Self-attention and cross-layer attention. In this work, we exploit self-attention on rain and image layer as Rain layer Self-Attention (RSA) and Image layer Self-Attention (ISA). Given the input feature X, it will be projected into query (Q), key (K) and value (V) by three learnable weight matrices W_q, W_k and W_v. Then dot-product, scaling and softmax among Q, K and V will be conducted. The self-attention function is defined as:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V. \tag{10}$$

We further design a Cross-Layer Attention (CLA) module which bridges the attention relationship between rain and image layer. The CLA conducts attention operation among Q_r from rain layer and K_b, V_b from image layer as follow:

$$CLA(Q_r, K_b, V_b) = softmax(\frac{Q_r K_b^T}{\sqrt{d_k}})V_b. \tag{11}$$

The token of rain layer serves as a query token Q_r to interact with the patch tokens K_b and V_b from the image layer through attention mechanism. By calculating the correlation degree between both layer, the highly attentive location of rain residual can be acquired, which provides an extra prior for enhanced feature representation. Note that, the CLA module has been stacked over the whole network. Compared with previous work, SCD-Former exploits not only the self-attention but also cross-layer attention of the rain and image layer for better restoration.

Implementation details. We train the network using the Charbonnier loss[5] supervised by the ground truth of rain and image layer:

$$\mathcal{L} = ||O - B - R||_F^2 + \lambda_r||R - \hat{R}||_1 + \lambda_b||B - \hat{B}||_1. \tag{12}$$

The framework is implemented with two RTX 3090 GPUs. We set the hyperparameter λ_b and λ_r as 1. The images are randomly cropped into 256 * 256 for training. The learning rate of network is set as 0.0002. The Adam optimizer is adopted for optimization with a batch size of 32.

5. Experiments

Datasets. We conduct the experiments on paired datasets SPA-data[37], GT-Rain[1] and LHP-Rain. For the SPA-data, training set is cropped from 29,500 images and 1000 rainy images are used for testing. For the GT-Rain, 89 sequences are used for training and 7 sequences are used for testing. For the LHP-Rain, 2100 sequences are used for training and 300 sequences are used for testing. To evaluate the deraining performance on real scenes, we choose typical real rainy images from Internet-data. For single image deraining methods, we select the representative supervised deraining methods, including the CNN-based SPANet[37],

實驗不公平，因為在利用 GT-Rain 等等其他資料集 train SCD-Former 時，根本沒有 distinct rain-layer for supervised learning（真實的雨圖），只有 clean image 做 supervised

合成的表現很差因為 huge domain gap

could remove real rain in the sky partially but they cannot handle the splashing water on the ground

Rainy Image | Rain100L | SPA-data | GT-Rain | LHP-Rain

Figure 8: Evaluation of the diversity of the LHP-Rain. We train SCD-Former on different datasets: Rain100L, SPA-data, GT-Rain and LHP-Rain, and test on other datasets. The model trained on LHP-Rain has achieved better deraining results.

PReNet[33], RCDNet[36], JORDER-E[47], MPRNet[27], GT-Rain[1], transformer-based Uformer[39] and IDT[44].

Evaluation metrics. We employ the full-reference PSNR and SSIM to evaluate the single image deraining results. Moreover, mean Average Precision (mAP) and Accuracy (Acc) are employed to evaluate object detection and lane segmentation after restoration by deraining methods.

5.1. Quantitative Evaluation

Deraining results on benchmarks. We make comparisons with state-of-the-art deraining methods on three datasets SPA-data (A), GT-Rain (B) and LHP-Rain (C). The quantitative results are reported in Table 2 under the following columns: A→A, B→B and C→C. It is observed that transformer-based methods perform better than most CNN-based methods except for MPRNet in terms of PSNR and SSIM because of the superior representation of self-attention. Note that SCD-Former outperforms the existing state-of-the-art methods on all benchmarks, which confirms the effectiveness of our method with both self and cross-layer attention.

5.2. Qualitative Evaluation

Evaluation on LHP-Rain. To further validate the deraining performance, we compare with the qualitative results of typical methods on LHP-Rain. As shown in Fig. 7, SCD-Former achieves more visual pleasing results without rain residual and artifacts comparing with other methods, which cleans the rain streaks and veiling effect on the trees, highlight occlusion on the red car and the ground splashing water.

Evaluation on real rainy images. To evaluate the performance on real rainy images, we train SCD-Former on synthetic rain Rain100L[48], real rain SPA-data, GT-Rain and LHP-Rain respectively and test on real rainy images. As shown in Fig. 8, the model trained on Rain100L performs poorly due to the huge domain gap. SPA-data and GT-Rain could remove real rain in the sky partially but they cannot handle the splashing water on the ground. The model trained on LHP-Rain has the best deraining performance which simultaneously removes rain streaks, veiling and ground splashing water without destroying image details.

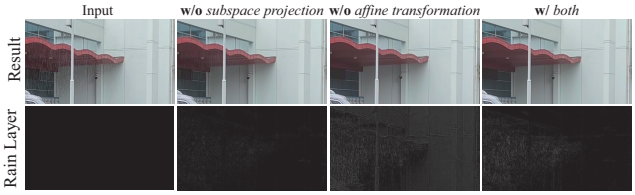

Figure 9: Ablation study of the robust low-rank tensor recovery model. The first row represents the deraining results, and the second row is corresponding rain. The first column is the original rainy frame and the remaining three columns represent the model without subspace projection, without affine transformation and full RLRTR.

Table 3: Ablation study of cross-layer attention.

Cross-layer attention	PSNR / SSIM
-	33.92 0.9384
✓	**34.33 0.9403**

5.3. Ablation Study

Q

Effectiveness of subspace projection. The subspace projection is used to characterize the extreme global low-rank property along the temporal dimension. In Fig. 9, there are obvious rain residual without subspace projection, implying that it is insufficient to characterize the property of the temporal dimension relying on the local prior of the temporal smoothness. Since the temporal low-rank property is neglected, leading to the rain residual in the results.

Effectiveness of affine transformation. The affine transformation is exerted to guarantee the pixel-level alignment of rainy video. As shown in Fig. 9, it can be observed that background residual and distortion obviously exist in the rain layer without affine transformation, because the extreme low-rank property along the temporal dimension is affected by data alignment among each frame in the video.

Effectiveness of cross-layer attention. The design aims to find the correlations between rain layer and image layer. After iterations, the image layer sub-network obtains more attention on the rain features brought by cross-layer attention. In Table 3, We show that the complementary guidance from rain layer promotes the restoration of image layer.

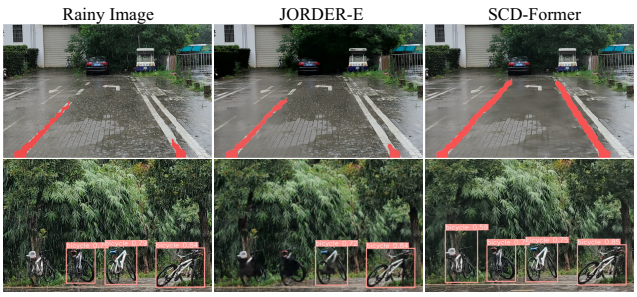

Figure 10: Evaluation of high-level tasks on lane segmentation and object detection. The performance improves significantly after removing rain streak and ground splashing from lanes and bicycles by our proposed SCD-Former.

Table 4: Evaluation of high-level tasks on deraining results.

Method	Det. (mAP)	Gain (Det.)	Seg. (Acc)	Gain (Seg.)
Rainy	0.543	-	0.237	-
SPANet	0.563	+0.020	0.268	+0.031
PReNet	0.560	+0.022	0.255	+0.018
RCDNet	0.556	+0.018	0.361	+0.124
JORDER-E	0.568	+0.025	0.385	+0.148
MPRNet	0.560	+0.017	0.350	+0.113
Uformer-B	0.568	+0.025	0.306	+0.069
IDT	0.570	+0.027	0.365	+0.128
SCD-Former	**0.575**	**+0.031**	**0.449**	**+0.212**

5.4. Discussion

Rain diversity of different datasets. The rain diversity of dataset can be validated by the experiment of training models on one dataset and testing on the other unseen datasets among SPA-data (A), GT-Rain (B) and LHP-Rain (C). As shown in Table 2, on one hand, the best results of C→A methods positively improve the performance of A, while all A→C results performing worse than degraded result of C. On the other hand, the best results of C→B have promotion while B→C drop severely. Therefore, the outcome proves that our LHP-Rain contains more diverse rain categories than others, because they could not handle extreme challenging cases such as occlusion and ground splashing.

Evaluation on downstream tasks. We further evaluate the image deraining results on high-level tasks. For object detection, we apply the official YOLOv5 model on deraning results and report the mean average precision (mAP) of different classes in Table 4, where SCD-Former reaches the best average mAP among typical objects. For lane segmentation, we choose the LaneNet[30] to predict the lane on LHP-Rain and SCD-Former has larger promotion on the segmentation accuracy. It is reasonable because SCD-Former performs well on removing ground splashing water and recovering the lane lines. The visualization results in Fig. 10 shows that the lane on the surface of ground and the bicycles could be properly predicted after deraining by SCD-Former.

User study on benchmarks quality. We look for 126 volunteers to anonymously vote for the benchmark with best quality. Among existing benchmarks, we randomly select

(a) Rainy Image (b) RLRTR Result

Figure 11: The limitation of RLRTR. The challenging occlusion effect could be removed by RLRTR from the background while the static haze is preserved.

Table 5: User study on benchmarks quality.

	LHP-Rain	SPA	RealRain1K	GT-Rain
Rain diversity	**63%**	8%	13%	16%
Image quality	**51%**	14%	20%	15%
GT quality	**55%**	15%	16%	14%

100 samples and conduct user study including: rain diversity, image quality (resolution, JEPG, blur) and GT quality. The result is listed in Table 5, where LHP-Rain consistently outperforms other benchmarks more than 50%.

6. Limitation

Our proposed video deraining method is limited to remove the haze in the heavy rain scenes. RLRTR is adept at separating the static background from the dynamic rain. However, due to the steadiness of mist in the short interval, which is almost motionless in the background, RLRTR cannot decompose the haze from image layer well. Fig. 11 shows examples of rain and mist in our benchmark. Although the challenging rain patterns such as rain streaks are clearly removed, the result still contains haze. We look forward to handling the static haze in the future.

7. Conclusion

In this paper, we propose a large-scale and high-quality paired real rain benchmark. Our proposed LHP-Rain provides diverse rain categories, especially the ground splashing rain issue which is first claimed in deraining community. The model trained on LHP-Rain could generalize well on various real rainy scenes with great rain removal performance. Moreover, the proposed low-rank tensor recovery model could generate high-quality GT and detailed analysis confirms better results than others. In addition, we propose a single deraining baseline which performs well on removing rain from sky to the ground. Extensive experiments verify the superiority of the proposed benchmark and significantly improves segmentation task after removing splashing.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China under Grant 61971460 and Grant 62101294, in part by JCJQ Program under Grant 2021-JCJQ-JJ-0060 and in part by the Fundamental Research Funds for the Central Universities, HUST: 2022JYCXJJ001.

References

[1] Yunhao Ba, Howard Zhang, Ethan Yang, Akira Suzuki, Arnold Pfahnl, Chethan Chinder Chandrappa, Celso de Melo, Suya You, Stefano Soatto, Alex Wong, and Achuta Kadambi. Not just streaks: Towards ground truth for single image deraining. In *ECCV*, 2022. 2, 3, 4, 9

[2] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on optimization*, 20(4):1956–1982, 2010. 6

[3] Yi Chang, Luxin Yan, and Sheng Zhong. Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising. In *CVPR*, pages 4260–4268, 2017. 5, 6

[4] Yi Chang, Luxin Yan, and Sheng Zhong. Transformed low-rank model for line pattern noise removal. In *ICCV*, pages 1726–1734, 2017. 3, 5

[5] Pierre Charbonnier, Laure Blanc-Feraud, Gilles Aubert, and Michel Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *ICIP*, volume 2, pages 168–172. IEEE, 1994. 9

[6] Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. Learning a sparse transformer network for effective image deraining. In *CVPR*, pages 5896–5905, June 2023. 3

[7] Xiang Chen, Jinshan Pan, Kui Jiang, Yufeng Li, Yufeng Huang, Caihua Kong, Longgang Dai, and Zhentao Fan. Unpaired deep image deraining using dual contrastive learning. In *CVPR*, pages 2017–2026, 2022. 2

[8] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE TIP*, 16(8):2080–2095, 2007. 5

[9] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *CVPR*, pages 3855–3863, 2017. 1, 2, 3

[10] Xueyang Fu, Qi Qi, Zheng-Jun Zha, Yurui Zhu, and Xinghao Ding. Rain streak removal via dual graph convolutional network. In *AAAI*, pages 1352–1360, 2021. 1, 3

[11] Kshitiz Garg and Shree K Nayar. When does a camera see rain? In *ICCV*, pages 1067–1074, 2005. 1

[12] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. Depth-attentional features for single-image rain removal. In *CVPR*, pages 8022–8031, 2019. 1, 2, 3

[13] Huaibo Huang, Mandi Luo, and Ran He. Memory uncertainty learning for real-world single image deraining. *IEEE TPAMI*, 2022. 2

[14] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *CVPR*, pages 8346–8355, 2020. 1, 3

[15] Tai-Xiang Jiang, Ting-Zhu Huang, Xi-Le Zhao, Liang-Jian Deng, and Yao Wang. Fastderain: A novel video rain streak removal method using directional gradient priors. *IEEE TIP*, 28(4):2089–2102, 2018. 6

[16] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009. 5

[17] Guanbin Li, Xiang He, Wei Zhang, Huiyou Chang, Le Dong, and Liang Lin. Non-locally enhanced encoder-decoder network for single image de-raining. In *ACM MM*, pages 1056–1064, 2018. 3

[18] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *CVPR*, 2019. 1, 2

[19] Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaojie Guo, and Xiaochun Cao. Single image deraining: A comprehensive benchmark analysis. In *CVPR*, pages 3838–3847, 2019. 2

[20] Wei Li, Qiming Zhang, Jing Zhang, Zhen Huang, Xinmei Tian, and Dacheng Tao. Toward real-world single image deraining: A new benchmark and beyond. *arXiv preprint arXiv:2206.05514*, 2022. 2, 4, 6

[21] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *ECCV*, pages 254–269, 2018. 3

[22] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In *CVPR*, pages 2736–2744, 2016. 3

[23] Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. *NeurIPS*, 24, 2011. 6

[24] Jiaying Liu, Wenhan Yang, Shuai Yang, and Zongming Guo. Erase or fill? deep joint recurrent rain removal and reconstruction in videos. In *CVPR*, pages 3233–3242, 2018. 1, 2, 6

[25] Yang Liu, Ziyu Yue, Jinshan Pan, and Zhixun Su. Unpaired learning for deep image deraining with rain direction regularizer. In *ICCV*, pages 4753–4761, 2021. 2

[26] Yu Luo, Yong Xu, and Hui Ji. Removing rain from a single image via discriminative sparse coding. In *ICCV*, pages 3397–3405, 2015. 3

[27] Armin Mehri, Parichehr B Ardakani, and Angel D Sappa. Mprnet: Multi-path residual network for lightweight image super resolution. In *CVPR*, pages 2704–2713, 2021. 9

[28] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Sign. Process. Letters*, 20(3):209–212, 2012. 4

[29] Anush Krishna Moorthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE TIP*, 20(12):3350–3364, 2011. 4

[30] Davy Neven, Bert De Brabandere, Stamatios Georgoulis, Marc Proesmans, and Luc Van Gool. Towards end-to-end lane detection: an instance segmentation approach. In *IEEE intelligent vehicles symposium (IV)*, pages 286–291. IEEE, 2018. 11

[31] Yigang Peng, Arvind Ganesh, John Wright, Wenli Xu, and Yi Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE TPAMI*, 34(11):2233–2246, 2012. 6

[32] Ruijie Quan, Xin Yu, Yuanzhi Liang, and Yi Yang. Removing raindrops and rain streaks in one go. In *CVPR*, pages 9147–9156, 2021. 1, 2, 3, 4

[33] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *CVPR*, pages 3937–3946, 2019. 3, 9

[34] Antonio Torralba and Aude Oliva. Statistics of natural image categories. *Network: computation in neural systems*, 14(3):391, 2003. 6

[35] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *CVPR*, pages 2353–2363, 2022. 1, 3

[36] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In *CVPR*, pages 3103–3112, 2020. 1, 3, 9

[37] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson WH Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *CVPR*, pages 12270–12279, 2019. 2, 4, 6, 9

[38] Ye-Tao Wang, Xi-Le Zhao, Tai-Xiang Jiang, Liang-Jian Deng, Yi Chang, and Ting-Zhu Huang. Rain streaks removal for single image via kernel-guided convolutional neural network. *IEEE Trans. Neural Networks and Learning Systems.*, 32(8):3664–3676, 2020. 3

[39] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, pages 17683–17693, 2022. 3, 9

[40] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *CVPR*, pages 3877–3886, 2019. 2

[41] Yanyan Wei, Zhao Zhang, Yang Wang, Mingliang Xu, Yi Yang, Shuicheng Yan, and Meng Wang. Deraincyclegan: Rain attentive cyclegan for single image deraining and rainmaking. *IEEE TIP*, 30:4788–4801, 2021. 1

[42] John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. *NeurIPS*, 22, 2009. 5

[43] Qingbo Wu, Lei Wang, King Ngi Ngan, Hongliang Li, Fanman Meng, and Linfeng Xu. Subjective and objective deraining quality assessment towards authentic rain image. *IEEE TCSVT*, 30(11):3883–3897, 2020. 2, 4

[44] Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Image de-raining transformer. *IEEE TPAMI*, 2022. 1, 3, 9

[45] Qi Xie, Qian Zhao, Deyu Meng, and Zongben Xu. Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. *IEEE TPAMI*, 40(8):1888–1902, 2017. 6

[46] Wenhan Yang, Jiaying Liu, Shuai Yang, and Zongming Guo. Scale-free single image deraining via visibility-enhanced recurrent wavelet learning. *IEEE TIP*, 28(6):2948–2961, 2019. 3

[47] Wenhan Yang, Robby T Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE TPAMI*, 42(6):1377–1393, 2019. 3, 9

[48] 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 *CVPR*, pages 1357–1366, 2017. 1, 2, 9

[49] Youzhao Yang and Hong Lu. Single image deraining via recurrent hierarchy enhancement network. In *ACM MM*, pages 1814–1822, 2019. 3

[50] Yuntong Ye, Yi Chang, Hanyu Zhou, and Luxin Yan. Closing the loop: Joint rain generation and removal via disentangled image translation. In *CVPR*, pages 2053–2062, 2021. 1

[51] Yuntong Ye, Changfeng Yu, Yi Chang, Lin Zhu, Xi-Le Zhao, Luxin Yan, and Yonghong Tian. Unsupervised deraining: Where contrastive learning meets self-similarity. In *CVPR*, pages 5821–5830, 2022. 1, 2

[52] Hongwei Yong, Deyu Meng, Wangmeng Zuo, and Lei Zhang. Robust online matrix factorization for dynamic background subtraction. *IEEE TPAMI*, 40(7):1726–1740, 2017. 5

[53] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, pages 14821–14831, 2021. 1, 3

[54] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*, pages 695–704, 2018. 3