Toward Real-world Single Image Deraining: A New Benchmark and Beyond

Wei Li*, Qiming Zhang*, Student Member, IEEE, Jing Zhang, Member, IEEE, Zhen Huang, Xinmei Tian, Member, IEEE, and Dacheng Tao, Fellow, IEEE

Abstract—Single image deraining (SID) in real scenarios attracts increasing attention in recent years. Due to the difficulty in obtaining real-world rainy/clean image pairs, previous real datasets suffer from low-resolution images, homogeneous rain streaks, limited background variation, and even misalignment of image pairs, resulting in incomprehensive evaluation of SID methods. To address these issues, we establish a new high-quality dataset named RealRain-1k, consisting of 1,120 high-resolution paired clean and rainy images with low- and high-density rain streaks, respectively. Images in RealRain-1k are automatically generated from a large number of real-world rainy video clips through a simple yet effective rain density-controllable filtering method and have good properties of high image resolution, background diversity, rain streaks variety, and strict spatial alignment. RealRain-1k also provides abundant rain streak layers as a byproduct, enabling us to build a large-scale synthetic dataset named SynRain-13k by pasting the rain streak layers on abundant natural images. Based on them and existing datasets, we benchmark more than 10 representative SID methods on three tracks: (1) fully supervised learning on RealRain-1k, (2) domain generalization to real datasets, and (3) syn-to-real transfer learning. The experimental results (1) show the difference of representative methods in image restoration performance and model complexity, (2) validate the significance of the proposed datasets for model generalization, and (3) provide useful insights on the superiority of learning from diverse domains and shed lights on the future research on real-world SID. The datasets have been released at https://github.com/hiker-lw/RealRain-1k.

Index Terms—Single image deraining, real-world derain dataset, domain generalization.

I. INTRODUCTION

Images captured in rainy days suffer from noticeable degradation of visual quality, thereby affecting both human visual perception and practical applications of advanced computer vision systems such as autonomous driving [30], [45]. The aim of single image deraining (SID) is to restore the clean and rain-free background images from the rainy ones. Benefitting from the rapid development of deep convolutional neural networks (CNNs) [19], recent years have witnessed an explosive spread of using CNN models to perform SID [5], [6], [9], [21], [32], [38]–[40].

Recently, many research efforts have been made to solve SID from synthetic images to the real-world rainy images, where two real datasets are proposed, i.e., SPA-data [32] and RainDS [27]. While they contribute to advancing the research for real-world SID, the difficulty in collecting high-quality rainy and clean pairs in the real world leads to various issues, including low resolution (256 × 256), homogeneous rain streaks, limited background variation (Figure 1 (b)), and even spatial misalignment (Figure 1 (a)), limiting the generalization ability of SID models trained on them in dealing with various real-world rainy images and implying limited practical significance of them.

As a result, there are several important research questions that have not been well studied or even remain unexplored: 1) How the existing SID methods perform in complex real-world scenarios, where the rainy images are captured at high resolutions and have diverse rain streaks and backgrounds? 2) How about the quantitative performance of existing methods when generalizing to an unseen real scene? 3) How to well utilize existing large-scale synthetic datasets to obtain a better performance on real rainy images?

To answer these questions, it is necessary to establish a new high-quality real-world dataset and pave the way for future research in this field. In this paper, we make an attempt towards this goal by proposing two datasets, RealRain-1k and SynRain-13k. Specifically, we first collect 1,120 high-resolution rainy videos, covering different scenes with diverse rain streaks and backgrounds. Then, we use a simple yet effective method to control the rain density in our generated images and carefully check their spatial alignment. In this way, we obtain 1,120 high-quality image triplets (of clean, light rain, heavy rain), resulting in two versions of our RealRain-1k dataset, i.e., RealRain-1k-L and RealRain-1k-H, respectively.

Besides, RealRain-1k also provides abundant real rain streak layers as a byproduct by reversing the linear superimposition model on clean and rainy images. We thus construct a large-scale synthetic dataset named SynRain-13k by pasting the rain streak layers on abundant natural images. It has a smaller scale synthetic dataset named SynRain-13k by pasting the rain streak layers on abundant natural images. It has a smaller domain gap to real rainy images compared with previous synthetic datasets, thereby facilitating training better SID models.

Based on RealRain-1k and SynRain-13k, we benchmark more than ten SID models on three tracks: 1) SL track, where these models are trained and evaluated on RealRain-1k in a supervised learning manner. We compare their performance in terms of restoration quality, model complexity, inference throughput, and convergence speed; 2) DG track, where these models are trained on different datasets and tested on unseen real datasets to investigate their syn-to-real and real-to-real domain generalization ability as well as the value of synthetic datasets; 3) TL track, where SID models are pre-trained on synthetic datasets, further fine-tuned and tested on real datasets to investigate the benefit of transfer learning. Based on the experiment results, we gain some useful insights on the limitations of existing SID methods and datasets and discuss...
future research directions.

The main contribution of the paper is two-fold. (1) We establish a high-quality real dataset RealRain-1k for SID study, which consists of 1,120 high-resolution clean and rainy image pairs with diverse rain streaks, abundant backgrounds, and strict spatial alignment. We also generate a large-scale synthetic dataset of 13k image pairs, i.e., SynRain-13k, by leveraging the diverse real rain layers from RealRain-1k. (2) Based on them, we comprehensively benchmark more than ten representative SID methods on three tracks including SL, DG and TL. The experiment results validate the value of our datasets and provide useful insights on the limitations of existing SID methods.

II. RELATED WORK

A. Single image deraining

Traditional methods Generally, SID methods can be categorized as prior-based methods and data-driven deep learning ones. Traditional methods mainly use hand-crafted features and priors to describe the features of rain streaks. For example, Kang et al. [17] propose to first decompose an image into low and high frequency parts and remove rain streaks in the high frequency layer via dictionary learning. Luo et al. [26] present a discriminative sparse coding method for separating rain streaks from the background. Li et al. [22] use Gaussian mixture models (GMM) as the prior to model rain streaks and background separately for rain removal. Zhu et al. [47] utilize layer-specific priors to discriminate rainy regions. Albeit achieving good performance on certain scenarios, the prior-based methods rely on specific priors, which may not applicable to many real scenes where the rain streaks are complex and diverse. As a result, they have a limited generalization ability.

CNN-based methods Recently, many deep-learning based approaches [1], [2], [6], [9], [15], [16], [21], [28], [31], [32], [36]–[38], [46], have been proposed for rain streak removal and made significant progress in this area. For example, Fu et al. first adopt a shallow CNN [8] and then a deeper ResNet [9] to remove rain streaks. Zhang et al. [42] take the rain density into account and present a multi-task CNN for joint rain density estimation and deraining. Ren et al. [28] provided a simpler baseline deraining network by considering network architecture, input and output, and loss functions. Jiang et al. [16] proposed a multi-scale coarse-to-fine progressive fusion network to remove rain streaks.

Transformer-based methods Vision transformer [7], [29] has recently been introduced in image restoration tasks and obtains significant performance due to its powerful modeling ability [3], [14], [18], [24], [35]. Chen et al. [4] first applies a standard transformer block equipped with multiple heads and tails for multiple low-level vision tasks including deraining. Liang et al. [23] propose a Swin Transformer-based model by combining CNN and transformer and achieves superior performance while maintaining computational efficiency. Wang et al. [34] constructs a U-shape transformer based on Swin Transformer for image restoration. Apart from the U-shape architecture, Restormer [39] utilizes the channel-wise attention and convolution and obtains better SID results.

B. Deraining datasets

Multiple datasets for SID have been proposed in previous works [9], [13], [20], [22], [16], [43], [44], most of which are generated by first creating rain streaks using the photo-realistic rendering technique [11], [12] and then blending them with clean images. For example, Jiang et al. [16] propose a large and diverse synthetic dataset containing 13k paired images gathered from multiple aforementioned datasets. As for real data, there are two paired datasets in this field as far as we know, named SPA-data [32] and RainDS [27], respectively. When revisiting the rainy images in RainDS, we
find that they are generated by using sprinklers in front of the camera to simulate the rainy scenes, leading to depth information loss and homogeneous rain streaks due to the less-varied rain-to-camera distance and the less-varied physical characteristics of the water from sprinklers. We also identify the spatial misalignment issue in RainDS as indicated by the green rectangles in Figure 1 (a), which will cause erroneous loss that misleads the training process and result in inaccurate evaluation performance. The rainy images in SPA-data are the frames from 170 videos with static backgrounds, among which drizzling videos are dominating. Consequently, the small number of videos and sparse rain streaks in each image lead to the limited diversity of rain and background appearances as shown in Figure 1 (b), which cannot comprehensively reflect the complex and diverse scenes in real world with rain at various densities. Besides, the images in SPA-data are small patches with low resolution as shown in Table 1. Although these two datasets have advanced the research in this field, there are still some problems needed to be solved. On the one hand, due to the domain gap between synthetic and real rainy images, SID models trained on existing synthetic datasets always do not generalize well on real-world rainy images. On the other hand, existing real-world datasets have limited diversity in rain streaks, backgrounds, and scenes, making it difficult to benchmark SID methods comprehensively. Therefore, it is necessary to establish a high-quality real-world dataset containing abundant scenes, diverse rain streaks, and various rain densities, and benchmark existing methods to show the overall progress in this area.

III. THE PROPOSED DATASETS

A. The real rainy dataset: RealRain-1k

Data collection We first collect a large number of high-resolution rainy videos from diverse real-world scenes. Specifically, 972 high-resolution rainy videos with static background are captured using a tripod to avoid the jittering artifacts and achieve strict spatial alignment. Moreover, we also collect 148 online available videos and have 1,120 videos in total, which cover various daylight and night outdoor scenes such as buildings, avenues, and parks. In addition, they have diverse rain streaks, different rain densities, and various backgrounds.

Triplet generation We notice that the rain in the real world significantly varies from light to heavy. To this end, we use a simple yet effective rain density-controllable filtering method to generate rainy and clean image pairs with different rain densities to mimic different rainy scenarios in the real world. The process is illustrated in Figure 2. Based on the observation that the pixel in a rainy image always has higher intensity if covered by rain streaks, we propose a simple maximum filtering method over the video clips to select and retain the bright pixels, i.e., pixels with rain. Specifically, given the pixel intensity \( P(x,y) \) at the position \((x,y)\) of the \(i_{th}\) frame from a rainy video, the rainy images \( R \) are generated as follows:

\[
R(x,y) = \max\{P(x,y) | l \leq i \leq r\},
\]

where \(l\) and \(r\) denote the selected start and end frame index in the video. Obviously, if a longer video clip is selected in Eq. 1 (i.e., a larger \(|r - l|\)) for rainy image generation, more pixels covered by rain streaks will be kept and thus the rain density increases, and vice versa. In this way, we can mimic the real-world rainy images with various rain densities by simply adjusting the length of the selected video clip to control the rain density in the generated rainy images. Specifically, we generate rainy images with two levels of rain densities, i.e., light and heavy, respectively, where the clip length is determined manually for each video since our captured videos have different rain densities. Figure 4 shows some examples with different rain densities.

It is assumed that the pixel is infrequently covered by rain streaks in a video clip lasting for a long period, especially in the mild rainy conditions. Thereby, most pixels at the same spatial position of the consecutive frames may be clean pixel candidates. Based on this assumption, we propose a simple median filtering method to obtain a reference clean (background) image for each video. Specifically, let \( B(x,y) \) denotes the pixel value at the position \((x,y)\) in the generated clean image, it can be calculated as follows:

\[
B(x,y) = \text{median}\{P(x,y) | l \leq i \leq r\},
\]

where \( P(x,y) \), \( l \), and \( r \) have the same meanings as Eq. 1.

To obtain high-quality rainy and clean image pairs, we carefully check them during the generation process. Three rounds of cross-checking and re-generation (e.g., select the start frame index and adjust the length of the video clip) are carried out to ensure the quality of the generated image pairs, i.e., in terms of strict spatial alignment, clean images without rain streaks, various rain streak appearances, and diverse backgrounds. In total, we generate 1,120 image triplets (clean, light rain, heavy rain) respectively for two versions of our RealRain-1k dataset, i.e., RealRain-1k-L and RealRain-1k-H respectively.

How to avoid the misalignment issue? It’s difficult to capture high-quality rainy and clean pairs with exact the same scene at the same time. We thus conduct the following strategies to guarantee the alignment quality in RealRain-1k: 1) Many images in our dataset consist of static objects (such as buildings, cars, etc.) that will not be affected when raindrops hit. 2) Thanks to the proposed rain density-controllable
generation method, we don’t have strict weather limitations to obtain the rainy videos with diverse rain densities. We instead choose the relatively normal rainy days to reduce the impact of the raindrops hitting the easy-to-move objects (flowers, leaves, etc.). Similarly, breezy days are selected to avoid wind disturbance. 3) We also carefully filter out some video clips with large disturbances (caused by wind, rain, and human disturbances) with many human efforts to ensure the quality of the dataset regarding alignment. Then each pair in our dataset undergoes repeated (at least three times) comparisons between rainy images and groundtruth. If disturbances are still unavoidable after filtering out the undesirable video clips, we choose to throw away the pairs from our dataset.

We acknowledge that it is impossible to obtain completely aligned real rainy datasets, and thus we made many efforts to include strict human supervision in the loop to make the misalignment almost invisible to the naked eyes. Compared to current valuable real rainy datasets SPA-data and RainDS, the misalignment issue is relieved in RealRain-1k. It is noted that we don’t claim the alignment issue is completely resolved and don’t mean to criticize existing real datasets. They have played an important role in promoting the development of real-world SID. We just hope RealRain-1k can make a positive supplement to the existing real rain datasets and our work can help to shed some light on the community of real-world SID.

**Data organization and statistics** The 1,120 image pairs in RealRain-1k-L (and RealRain-1k-H) are further split into three disjoint subsets, i.e., training set, validation set, and test set, at the ratio of 7:1:2 respectively. Compared with SPA-data and RainDS, our RealRain-1k contains higher resolution images, i.e., an average resolution of 1,512 × 973. Some examples are shown in Figure 3 (a) and Figure 6. The statistics of the proposed RealRain-1k and previous representative datasets are summarized in Table I.

**B. The synthetic rainy dataset: SynRain-13k**

As a byproduct, we can extract the real rain layers from the rainy images in the training set of RealRain-1k-L after obtaining the rainy and clean image pairs. As shown in Figure 5, it can be achieved by reversing the linear superimposition model as follows:

\[
L = R - B, \quad (3)
\]

where \( R \) and \( B \) represent the rainy and clean images, respectively. After double-checking the appearances and density of the extracted rain streaks, we manually select 440 clean rain streak layers finally. Some examples are shown in the bottom row of Figure 3 (a). Then, we randomly paste them...
TABLE I

The comparison of the proposed RealRain-1k and SynRain-13k datasets and existing ones. The number separated by the slashes in the second column denotes the number of image pairs in the training set, validation set, and test set in each dataset, respectively.

| Dataset          | Image pairs | Data source                                                                 | Real | Average resolution | Features                                                | Rain of night scenes |
|------------------|-------------|------------------------------------------------------------------------------|------|--------------------|---------------------------------------------------------|----------------------|
| RainDS [27]      | 150/0/98    | using sprinkler                                                             | ✓    | 1,296×728          | not strictly aligned, homogenous rain streaks           | 0%                   |
| SPA-data [32]    | 638,492/0/1,000 | 49% videos are captured using mobile phone                                     | ✓    | 256×256          | mostly light rain, homogenous rain streaks              | 7.53%                |
| Rain13k [16]     | 13,711/0/4,298 | synthetic rain streaks and background images                                  | ✓    | 482×420          | large domain gap between real and synthetic              | -                    |
| RealRain-1k      | 784/112/224 | 1,120 high-resolution videos                                                | ✓    | 1,512×973         | strictly aligned, diverse rain streaks, controllable rain density | 47.95%                |
| SynRain-13k      | 13,711/198/1,200 | real rain layer + background                                                  | ✓    | 482×420          | small domain gap between real and synthetic              | -                    |

on the clean images from Rain13k [16] following the linear superimposition model and the process is illustrated in Figure 5. In this way, we generate a new large-scale synthetic rainy dataset containing 13k rainy images, i.e., SynRain-13k, by keeping the same dataset volume as Rain13k. Compared with previous synthetic datasets that are generated based on simulated rain streaks [25], [36], we directly use the real rain layers, thereby reducing the domain gap between the synthetic and real images, which is never explored before. Some examples are shown in Figure 3(b). As demonstrated in the experiment parts, SID models trained on our SynRain-13k show better generalization performance on real rainy images.

IV. Experiments

Based on our proposed real and synthetic datasets RealRain-1k and SynRain-13k, we benchmark thirteen representative SID methods including two traditional ones [22], [26] and eleven state-of-the-art deep learning-based ones [5], [6], [8]–[10], [16], [21], [32], [38]–[40].

Specifically, we conduct the experiments on three tracks, i.e., (1) supervised learning (SL) track, (2) domain generalization (DG) track, and (3) cross-domain transfer learning (TL) track. The detailed settings of each track will be presented later. We use the common PSNR and SSIM [33] metrics to measure the performance of different methods. In addition, we present some visual results for subjective evaluation.

A. Supervised learning track

In this track, we compare different methods by training and testing them on our RealRain-1k-L and RealRain-1k-H, respectively. For each method, we use the original implementation code provided by the authors and adopt the same training settings including augmentation, batch size, learning rate, optimizer, etc. We follow the common practice for all experiments, i.e., all the methods are trained on the training set, validated on the validation set, and evaluated on the test set.
### Main results
The results as well as model size and FLOPs are summarized in Table II. As can be seen, the deep learning-based method of both architectures, i.e., CNNs and transformers, significantly outperform the traditional prior-based methods on both RealRain-1k-H and RealRain-1k-L. The results validate the superiority of the deep neural networks, owing to their strong representation ability and the data-driven learning paradigm. When comparing CNN-based and the advanced transformer-based methods, Restormer outperforms MSPFNN by a large margin, although they have the similar model size. Comparing the results of each method on RealRain-1k-H and RealRain-1k-L, we find that the performance of all the methods drops significantly when dealing with heavy rainy images. It shows that the proposed RealRain-1k-H and RealRain-1k-L have different levels of difficulties as implied by their names. It is noteworthy that our datasets can be easily extended to cover different levels of rain densities and it is interesting to investigate how different methods perform on rainy images with different rain densities, which we leave as our future work. Among all the methods, HINet achieves the best performance of 40.82 dB and 41.98 dB on RealRain-1k-H and RealRain-1k-L, respectively. However, it is also the largest model that has 88M parameters and high computational cost, i.e., about 170G FLOPs. Therefore, it also matters to develop efficient SID methods that can achieve a better Pareto front between the deraining performance and model complexity.

### Convergence comparison
We select five representative methods to investigate their convergence performance by training them on RealRain-1k-L and RealRain-1k-H with different epochs, including 200, 500, 1000, and 2000 epochs, respectively. For methods like Restormer that splits the training pipeline into several stages, we determine the training epochs in each stage proportionally.

The results are shown in Table III. As can be seen, except the smallest model, i.e., DDN whose performance fluctuates in different settings, all other methods benefit from a longer training schedule. When the model size increases, the model tends to have faster convergence speed owing to the stronger representation ability. For example, HINet trained with 200 and 500 epochs obtains comparable performance to MPRNet trained with 500 and 1000 epochs, respectively. An interesting finding is that the superiority of HINet over MPRNet diminishes with the increase of training epochs, where HINet is 20× larger than MPRNet. It implies that big models may have better data efficiency and it also matters to investigate the performance of different methods under the few-shot learning paradigm.

### B. Domain generalization track
In this track, we quantitatively show the generalization performance of different methods, which are trained on a synthetic or real dataset and tested on an unseen real dataset. It is a common scenario when applying the developed SID methods in real-world applications, but has not been well studied. The experiments are conducted at two settings, i.e., synthetic-to-real DG and real-to-real DG.

**Synthetic-to-real DG** We benchmark 11 deep learning-based methods based on the synthetic datasets, i.e., Rain13k [16] and our SynRain-13k, and real datasets, i.e., SPA-data [32], RainDS [27], and our RealRain-1k-L and RealRain-1k-H. All the methods are trained on the training set of these synthetic datasets and evaluated on the test sets.
of these real datasets, following the original settings in their papers. We use the models trained on Rain13k from their official repositories if provided. The results are summarized in Table IV. Figure 7 shows some visual examples. As can be seen, when trained on SynRain-13k and tested on both SPA-data and RainDS, all the methods have better generalization performance than those trained on Rain13k. The results demonstrate the value of the proposed SynRain-13k which is generated using the real rain layers and has a smaller synthetic to real domain gap. Besides, Restormer trained on SynRain-13k achieves the best generalization performance in all scenarios, validating the good generalization ability of transformer as studied in [41].

We also show the generalization performance of different methods on the test sets of our RealRain-1k-H and RealRain-1k-L. As shown in Table V, the models trained on our SynRain-13k generalize better on RealRain-1k, which is consistent with the results in Table IV validating the value of the proposed SynRain-13k.

**Real-to-real DG** We benchmark three most representative SID methods in this track based on RealRain-1k-H, RealRain-1k-L, and SPA-data, whose rain density decreases in order. The models trained on RealRain-1k are collected from the experiments in Table I while the models trained on SPA-data are collected from the official repositories or retrained following the same setting.

The results are summarized in Table VI. Comparing Table VI and Table IV we can see that our RealRain-1k datasets can help methods generalize well to an unseen real domain. For example, both Restormer and HINet in the RealRain-1k-L→SPA-data setting perform better than the counterparts in the Rain13k→SPA-data setting. It is also interesting to find that the methods trained on our SynRain-13k outperform those trained on our RealRain-1k, owing to the real rain layers and large data volume of SynRain-13k. Besides, the experiment results also show that the generalization performance decreases when the difference of rain density in training and test sets becomes larger. For example, the methods in RealRain-1k-L→SPA-data generalize better than those in RealRain-1k-H→SPA-data. We also show some visual examples in Figure 8. The promising deraining results further validate the value and effectiveness of RealRain-1k.

To better reveal how and why the models behave differently when encountering the different rain density gap. We conduct several experiments to further support our claim. We split each rainy image into two non-overlapped regions according to the difference threshold between the rainy image and groundtruth, *i.e.*, rainy region and rain-free region, where the rainy region denotes the area whose difference is above the threshold. We show the restoration performance on the two kinds of regions

---

**TABLE IV**

| Method | Rain13k → SPA-data | SynRain-13k → SPA-data | Rain13k → RainDS | SynRain-13k → RainDS |
|--------|-------------------|------------------------|-----------------|---------------------|
| Rain13k | PSNR (dB) | SSIM | PSNR (dB) | SSIM | PSNR (dB) | SSIM |
| DDN [9] | 32.63 | 0.9282 | 32.63 | 0.9282 | 21.80 | 0.6084 |
| DerainNet [8] | 30.13 | 0.9132 | 31.22 | 0.9288 | 22.31 | 0.5973 |
| RESCAN [11] | 32.04 | 0.9385 | 35.50 | 0.9510 | 23.18 | 0.6499 |
| LPNet [10] | 29.66 | 0.9201 | 31.10 | 0.9372 | 23.10 | 0.6429 |
| SPANet [12] | 32.82 | 0.9337 | 33.81 | 0.9470 | 22.83 | 0.6289 |
| DRDNet [6] | 28.35 | 0.8943 | 31.56 | 0.9403 | 22.78 | 0.6164 |
| MSPFN [16] | 22.46 | 0.8952 | 34.89 | 0.9548 | 17.33 | 0.5865 |
| SPDNNet [32] | 32.44 | 0.9343 | 33.88 | 0.9531 | 23.41 | 0.6642 |
| MPRNet [40] | 32.77 | 0.9396 | 35.08 | 0.9555 | 23.25 | 0.6563 |
| HINet [39] | 32.35 | 0.9361 | 34.62 | 0.9521 | 23.50 | 0.6639 |
| Restormer [39] | 32.65 | 0.9358 | 35.48 | 0.9586 | 22.96 | 0.6480 |

**TABLE V**

| Method | Rain13k → RealRain-1k-H | SynRain-13k → RealRain-1k-H | Rain13k → RealRain-1k-L | SynRain-13k → RealRain-1k-L |
|--------|-------------------------|-----------------------------|-----------------------|-----------------------------|
| Rain13k | PSNR (dB) | SSIM | PSNR (dB) | SSIM | PSNR (dB) | SSIM |
| DDN [9] | 22.27 | 0.7657 | 22.27 | 0.7657 | 25.93 | 0.8651 |
| DerainNet [8] | 27.00 | 0.8468 | 26.06 | 0.8270 | 29.94 | 0.9047 |
| RESCAN [11] | 21.91 | 0.7420 | 23.50 | 0.8483 | 25.40 | 0.8474 |
| LPNet [10] | 25.32 | 0.8591 | 28.66 | 0.9057 | 28.98 | 0.9178 |
| SPANet [12] | 26.66 | 0.8824 | 27.54 | 0.8934 | 28.97 | 0.9207 |
| DRDNet [6] | 24.17 | 0.8304 | 24.96 | 0.8440 | 27.40 | 0.8959 |
| MSPFN [16] | 25.05 | 0.8131 | 27.42 | 0.8801 | 27.03 | 0.8634 |
| SPDNNet [32] | 17.92 | 0.7728 | 30.05 | 0.9399 | 19.50 | 0.8552 |
| MPRNet [40] | 24.63 | 0.8585 | 29.78 | 0.9272 | 27.94 | 0.9111 |
| HINet [39] | 23.39 | 0.8346 | 30.32 | 0.9402 | 26.51 | 0.8949 |
| Restormer [39] | 23.82 | 0.8382 | 30.66 | 0.9390 | 26.56 | 0.8920 |

Q is a placeholder for quantization, U is for uncertainty, and sid is for subjective image quality.
TABLE VI
QUANTITATIVE RESULTS OF REPRESENTATIVE SID METHODS ON THE REAL-TO-REAL DG TRACK.

|                  | RealRain-1k-L → SPA-data | RealRain-1k-H → SPA-data | SPA-data → RealRain-1k-L | SPA-data → RealRain-1k-H |
|------------------|--------------------------|--------------------------|--------------------------|--------------------------|
|                  | PSNR (dB)                | SSIM                     | PSNR (dB)                | SSIM                     |
| Rainy images     | 32.63                    | 0.9282                   | 32.63                    | 0.9282                   |
| MPRNet           | 32.50                    | 0.9482                   | 30.70                    | 0.9460                   |
| HINet            | 35.37                    | 0.9634                   | 34.08                    | 0.9575                   |
| Restormer        | 33.84                    | 0.9590                   | 32.93                    | 0.9545                   |

Fig. 7. Visual comparison of three representative methods MSPFN [16], MPRNet [40] and Restormer [39] trained on the Rain13k and our SynRain-13k respectively.
Our DG experiments reveal that the models behave differently when generalizing from heavy to light rain, i.e., models trained on light rain images can obtain larger performance gain when generalizing to heavy rain images, while the model generalizing from heavy to light tends to change the details of the rain-free region, resulting in relatively lower improvement. These findings have never been revealed by previous works, owing to the comparisons on the proposed rain density-controllable dataset. It confirms the value of our proposed dataset. Besides, it is also noteworthy that the proposed RealRain-1k-L and RealRain-1k-H are more challenging than the SPA-data since their rainy images have smaller PSNR values due to the larger rain density. As shown in Table VII in the paper, the performance is far from saturation on the RealRain-1k-H dataset, calling for more research efforts on image deraining, especially for heavy rain images.

**Real-to-real DG with different training schedules**

The results on the setting of real-to-real domain generalization with different training schedules are shown in Table IX and Table X. As can be seen, the generalization performance on the SPA-data is getting better with the increase of training epochs on our RealRain-1k dataset. Again, the results confirm the value of our dataset for training a better model that generalizes well to real rainy images. It is noteworthy that the results of different methods in Table IX and Table X are not consistent with those in Table VII because we train them according to their default settings (i.e., with different training epochs for different methods) for the experiment in Table VII.

**C. Cross-domain transfer learning track**

In this track, we compare two representative methods, i.e., MPRNet [40] and Restormer [39] on different settings. Compared to the synthetic-to-real DG setting, we collect the models trained on the synthetic datasets including Rain13k and SynRain-13k, and further fine-tune them on the training sets of SPA-data, RainDS, RealRain-1k-H, and RealRain-1k-L. The results are given in Table XI. Again, both methods pre-trained on our SynRain-13k outperform those pre-trained on Rain13k or without pre-training, showing the benefit of the real rain layers. And the results after fine-tuning are much better than those in the synthetic-to-real and real-to-real DG settings, which is reasonable since the training data from the downstream datasets are used. Note that the performances of Restormer [39] pre-trained on the synthetic datasets are

---

**TABLE VII**

| RealRain-1k-H → SPA-data | Rainy region PSNR (dB) | Rain-free region PSNR (dB) | RealRain-1k-L → SPA-data | Rainy region PSNR (dB) | Rain-free region PSNR (dB) |
|--------------------------|------------------------|---------------------------|--------------------------|------------------------|---------------------------|
| Rainy images             | 22.96                  | 37.34                     | Rainy images             | 22.96                  | 37.34                     |
| MPRNet [40]              | 26.95                  | 30.65                     | MPRNet [40]              | 26.83                  | 33.31                     |
| Gain                     | 3.99                   | -6.68                     | Gain                     | 3.87                   | -4.02                     |
| Restormer [39]           | 27.59                  | 31.60                     | Restormer [39]           | 27.83                  | 32.71                     |
| Gain                     | 4.63                   | -5.74                     | Gain                     | 4.87                   | -4.62                     |

**TABLE VIII**

| SPA-data → RealRain-1k-H | Rainy region PSNR (dB) | Rain-free region PSNR (dB) | SPA-data → RealRain-1k-L | Rainy region PSNR (dB) | Rain-free region PSNR (dB) |
|--------------------------|------------------------|---------------------------|--------------------------|------------------------|---------------------------|
| Rainy images             | 22.72                  | 29.38                     | Rainy images             | 23.72                  | 31.15                     |
| MPRNet [40]              | 25.73                  | 31.78                     | MPRNet [40]              | 23.79                  | 33.47                     |
| Gain                     | 3.01                   | 2.40                      | Gain                     | 3.67                   | 2.32                      |
| Restormer [39]           | 25.21                  | 31.23                     | Restormer [39]           | 27.88                  | 33.33                     |
| Gain                     | 2.49                   | 1.85                      | Gain                     | 4.16                   | 2.18                      |
Fig. 8. Some deraining results of HINet [5] which is trained on our RealRain-1k-L and evaluated on SPA-data [32]. We can see that the deraining results are fairly promising demonstrating that our datasets can generalize to various rainy scenarios.

| Training epochs | RealRain-1k-L → SPA-data |
|-----------------|--------------------------|
|                 | 200          | 500          | 1000         | 2000         |
| Rainy images    | PSNR (dB)    | SSIM         | PSNR (dB)    | SSIM         | PSNR (dB)    | SSIM         |
| MPRNet [40]     | 30.94        | 0.9391       | 33.10        | 0.9512       | 33.63        | 0.9529       | 34.73        | 0.9601       |
| HINet [5]       | 32.91        | 0.9489       | 34.26        | 0.9556       | 34.74        | 0.9593       | 35.14        | 0.9623       |
| Restormer [39]  | 32.75        | 0.9451       | 33.49        | 0.9545       | 33.65        | 0.9575       | 34.52        | 0.9623       |
inferior to the one without pre-training. Maybe it is because the backgrounds in the training set of SPA-data overlap with the ones in the test set. Therefore, Restormer could achieve a very good performance even without pre-training.

### D. Model complexity

The model complexity of different methods is shown in Table XII, including model parameters, FLOPs, GPU memory footprint, and inference latency. From the table, we can see that the model size, the latency and FLOPs increase in recent years, probably due to the advance of GPUs and the trend of pursuing better performance. For example, MPRNet needs about 3s to process an image, while HINet has 88M parameters, limiting their usage in practical applications. Therefore, it deserves more research efforts to achieve a better balance between deraining performance and model complexity.

### V. Conclusion and discussion

In this paper, we establish a high-quality real rainy dataset RealRain-1k using a simple but effective rain density-controllable filtering method and a large-scale synthetic dataset SynRain-13k based on real rain layers. Based on them, we benchmark more than 10 deraining methods comprehensively on three tracks. The results demonstrate the value of the proposed datasets, the different representation abilities of CNNs- or transformers-based methods, and the generalization performance under different learning paradigms with different datasets. We hope RealRain-1k and SynRain-13k can serve as test beds to facilitate future research in this area.

**Limitation and discussion** Our real dataset only has 1k images in the current version, which can be scaled up by collecting more videos. Meanwhile, more rainy images of different rain densities can be generated to form a series of datasets, each of which has a specific level of rain density.

**Social impact.** RealRain-1k and SynRain-13k can benefit the study of image deraining that is of great practical significance. Attention should be drawn to the generalization ability of models trained on our datasets, which may not be good in unseen scenes with extremely different rain streaks.

### References

[1] N. Ahn, S. Y. Jo, and S.-J. Kang. Eagnet: Elementwise attentive gating network-based single image de-raining with rain simplification. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(2):608–620, 2021.

[2] L. Cai, Y. Fu, W. Huo, Y. Xiang, T. Zhu, Y. Zhang, H. Zeng, and D. Zeng. Multi-scale attentive image de-raining networks via neural architecture search. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.
[3] Q. Cai, Y. Qian, J. Li, J. Lv, Y.-H. Yang, F. Wu, and D. Zhang. Hippu: Hierarchical patch transformer for single image super resolution. arXiv preprint arXiv:2203.10247, 2022.

[4] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao. Pre-trained image processing transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12299–12310, 2021.

[5] L. Chen, X. Lu, J. Zhang, X. Chu, and C. Chen. Hinet: Half instance normalization network for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 1822–1922, 2021.

[6] S. Deng, M. Wei, J. Wang, Y. Feng, L. Liang, H. Xie, F. L. Wang, and M. Wang. Detail-recovery image deraining via context aggregation networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 14560–14569, 2020.

[7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image classification at scale. In International Conference on Learning Representations, 2020.

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

[9] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley. Removing rain from single images via a deep detail network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3855–3863, 2017.

[10] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley. Lightweight pyramid networks for image deraining. IEEE transactions on neural networks and learning systems, 31(6):1794–1807, 2019.

[11] K. Garg and S. K. Nayar. Photorealistic rendering of rain streaks. ACM Transactions on Graphics (TOG), 25(3):996–1002, 2006.

[12] K. Garg and S. K. Nayar. Vision and rain. International Journal of Computer Vision, 75(1):13–27, 2007.

[13] X. Hu, C.-W. Fu, L. Zhu, and P.-A. Heng. Depth-attentional features for single-image rain removal. IEEE Transactions on Image Processing, 3855–3863, 2017.

[14] X. Huang and V. M. Patel. Representation learning for deep image restoration: A better and simpler baseline. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9147–9156, 2021.

[15] R. Quan, X. Yu, Y. Liang, and Y. Yang. Removing raindrops and rain streaks in one go. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9147–9156, 2021.

[16] A. Vasilyev, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. Advances in neural information processing systems, 30, 2017.

[17] X. Hu, C.-W. Fu, L. Zhu, and P.-A. Heng. Depth-attentional features for single-image rain removal. International Journal of Computer Vision, 129(5):1650–1674, 2021.

[18] Y. Liu, Z. Yue, J. Pan, and Z. Su. Unpaired learning for deep image deraining with rain direction regularizer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 4753–4761, 2021.

[19] Y. Luo, Y. Xu, and H. Ji. Removing rain from a single image via discriminative sparse coding. In Proceedings of the IEEE international conference on computer vision, pages 3397–3405, 2015.