DANet: Image Deraining via Dynamic Association Learning

Kui Jiang[1] , **Zhongyuan Wang**[1*] , **Zheng Wang**[1] , **Peng Yi**[1] , **Junjun Jiang**[2] ,
Jinsheng Xiao[3] , **Chia-Wen Lin**[4]

[1]NERCMS, School of Computer Science, Wuhan University
[2]School of Computer Science and Technology, Harbin Institute of Technology
[3]School of Electronic Information, Wuhan University
[4] National Tsing Hua University

Abstract

Rain streaks and background components in a rainy input are highly correlated, making the deraining task a composition of the rain streak removal and background restoration. However, the correlation of these two components is barely considered, leading to unsatisfied deraining results. To this end, we propose a dynamic associated network (DANet) to achieve the association learning between rain streak removal and background recovery. There are two key aspects to fulfill the association learning: 1) DANet unveils the latent association knowledge between rain distribution and background texture recovery, and leverages it as an extra prior via an associated learning module (ALM) to promote the texture recovery. 2) DANet introduces the parametric association constraint for enhancing the compatibility of deraining model with background reconstruction, enabling it to be automatically learned from the training data. Moreover, we observe that the sampled rainy image enjoys the similar distribution to the original one. We thus propose to learn the rain distribution at the sampling space, and exploit super-resolution to reconstruct background details for computation and memory reduction. Our proposed DANet achieves the approximate deraining performance to the state-of-the-art MPRNet but only accounts for 52.6% and 23% inference time and computational cost, respectively.

1 Introduction

Images taken outdoors are more susceptible to rain weather conditions which decrease the visual quality of the captured images. Both rain streaks and rain accumulation can cause severe occlusion on the background scene and significantly degrade the contrast and visibility. These negative visual effects result in the degradation of many high-level computer vision tasks like image understanding, object detection [Xu et al., 2021b] and identification [Xu et al., 2021a]. Therefore, there is a pressing need to develop algorithms that remove perturbation while recovering the clean background textures.
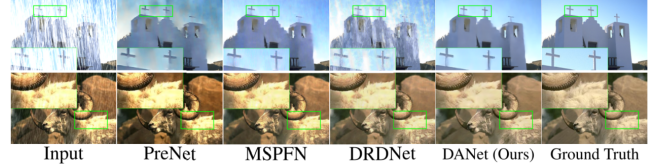

Figure 1: Visual comparison results. PreNet and MSPFN remove rain streaks in some cases, but fail to cope with additional degradation effects of **missing details and contrast bias**. Our DANet reconstructs credible textures with visually pleasing contrast.

Past decades have witnessed the significant progress of image deraining technologies. The early model-based methods [Garg and Nayar, 2005] rely more on statistical analyses of rain streaks and background scenes, and enforce handcrafted priors (*e.g.,* sparsity and nonlocal means filtering) on both rain and background. Since the restrictive assumptions are introduced for specific scenes or rain patterns, these methods [Kang et al., 2012] become still biased and less generalizable on complex or real scenarios.

More recently, data-driven methods have overtaken model-based methods on both deraining performance and popularity. These methods promote further progress by sophisticated architectures [Jiang et al., 2021] and training practices [Jiang et al., 2020a] for effective and efficient models. However, similar to image denoising, most of the prevalent models regard the image deraining as a simple rain streaks removal problem, *i.e.,* they produce deraining results by subtracting the predicted rain residual from rainy input based on the simple additive model. Consequently, the background contents suffer from incomplete and undesirable restoration, such as over-smoothing, halo artifacts, and color distortion, due to ignoring the additional degradation side effects. Recurrent- and cascaded-based methods [Ren et al., 2019; Deng et al., 2020] have made further steps on fidelity and textural plausibility by approaching both rain streaks removal and detail recovery. As shown in Figure 1, the cascaded (DRDNet) or recurrent learning (PreNet) method could bring considerable improvements, but is still far from generating desirable results under heavy rain conditions since the rain estimation and detail recovery are isolated. In addition, they perform inference in the whole resolution space using several cascaded stages or sub-networks, which limits their representation efficiency and interpretability since the rain streaks in
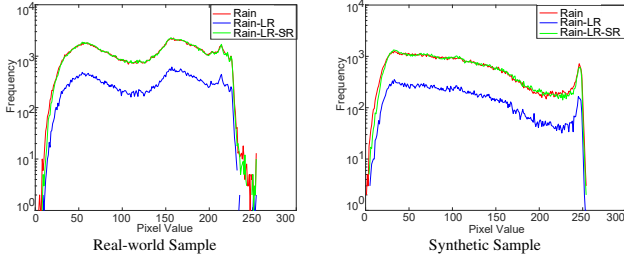
---
*Corresponding Author

Figure 2: Fitting results of "Y" channel histogram for **Real** and **Synthetic** samples. "Rain" and "Rain-LR" denote the original and corresponding low-dimension space distribution of rainy image. "Rain-LR-SR" is the distribution via Bilinear interpolation from "Rain-LR". The fitting results show that the reconstructed sample ("Rain-LR-SR") from "Rain-LR" can get the similar statistical distribution with that of the original input.

rainy images are commonly sparse and have the canonical samples.

Indeed, rain streaks removal and texture restoration are highly correlated, where the predicted rain distribution maps essentially reveal the degradation location and intensity, which are helpful for the accurate background restoration. This motivates us to study the following key question. ***Question One:*** Whether the predicted rain distribution can provide degradation priors to help background texture recovery?

Inspired by the dynamic network, it can adapt structures or parameters to their input during inference, leading to notable advantages in terms of accuracy, computational efficiency and adaptiveness [Han *et al.*, 2021]. Because of the sparse distribution of rain streaks in rainy scenes, the feature-selection dynamic strategy provides the alternative solution for accurate image deraining and background restoration. Specifically, it allows the network to focus on the informative region through the pixel-wise dynamic weights (rain degradation mask), thus enjoying favorable properties those are absent in aforementioned deraining models. This way, it is reasonable to employ the dynamic learning to predict the degradation mask from the rain input to guide the accurate texture restoration. We thus construct a novel dynamic associated network (DANet) for single image deraining. DANet is a two-stage framework, whose key idea is to cast the image deraining task as a composition of rain streaks removal, texture (detail and contrast) recovery, and their association learning. More specifically, the dynamic degradation prior is learned directly from the rainy input, and then produces the degradation mask to help extract the complementary information for accurate texture restoration. In detail, a deraining sub-network (DSN) and a reconstruction sub-network (RSN) are respectively designed to approach these two sub-problems, respectively. Meanwhile, an associated learning module (ALM) is used to associate these two sub-tasks. It provides the dynamic prediction in terms of the location and degradation intensity from rainy inputs, which is more helpful for the accurate texture restoration. To achieve the optimal approximation, we introduce joint constraints for enhancing the compatibility of deraining model with background reconstruction, automatically learned from the training data.

In addition, we observe that the reconstructed rainy image

via bilinear interpolation from the sampled rainy image has the similar statistical distribution to the original one, shown in Figure 2. This raises another key question. ***Question Two:*** Whether the rain distribution can be inferred from the sampling space? Simultaneously, the background details follow this observation.

This inspires us to learn the prediction of rain streak distribution and degradation mask at the sampling space, and then exploit the super-resolution to reconstruct the spatial details. There are two substantial superiorities: 1) predicting and characterizing rain distribution in the sub-space from the sampled low-resolution rainy image can greatly simplify the first learning task; 2) all the feature representations, including DSN and RSN are processed at the low-dimension space, greatly saving the computational cost. In particular, besides introducing the association learning and super-resolution technologies to alleviate the information loss due to spatial sampling, compared to the efficiency proceeds, we have verified that the performance decline is acceptable. Overall, the main contributions are summarized as follows.

1. We propose a novel dynamic association network (DANet) for image deraining. It consists of the rain streak removal, background restoration and their association learning. To the best of our knowledge, we are the first to approach image deraining from a new perspective of joint rain streaks removal and texture reconstruction via dynamic association learning and optimization.

2. We design a novel associated learning module (ALM) to associate rain streaks removal and texture recovery tasks via the dynamic degradation prior learning. It significantly alleviates the learning burden while promoting recovery accuracy. We propose a selective fusion block (SFB) for effective multi-scale residue fusion.

3. Comprehensive experiments on image deraining and detection task have verified that our proposed DANet method is more suitable for preserving the textural details while removing rain streaks.

2 Proposed Method

This section details the architecture of our proposed dynamic association network (DANet) and its essential components.

2.1 Architecture and Model Optimization

Figure 3 outlines the framework of our proposed DANet, which contains a deraining sub-network (DSN), an associated learning module (ALM), and a reconstruction sub-network (RSN). For convenience, DSN and RSN share the same dual-branch fusion network, elaborated in Section 2.2: Dual-branch Fusion Network.

For a given rainy image I_{Rain} and its clean version I_B, we first sample them with Bilinear operation to generate the corresponding sub-samples ($I_{Rain,S}$ and $I_{B,S}$). $I_{Rain,S}$ is then input to DSN to predict the distribution of rain streaks and produce the corresponding deraining result via $\mathcal{G}_{DSN}(\cdot)$. The above procedures can be expressed as

$$I_{R,S}^* = \mathcal{G}_{DSN}(I_{Rain,S}), \quad I_{B,S}^* = I_{Rain,S} - I_{R,S}^*. \quad (1)$$
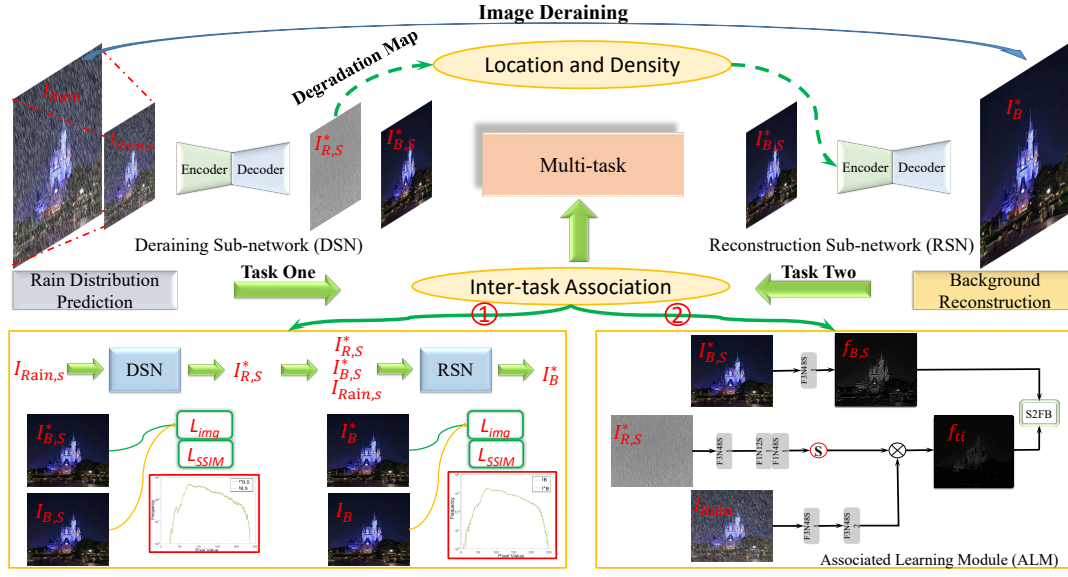
Figure 3: Framework of our proposed dynamic association network (DANet). It consists of a deraining sub-network (DSN), a reconstruction sub-network (RSN), and their association learning. DSN learns the corresponding rain distribution $I^*_{R,S}$ from the sub-sample $I_{Rain,S}$, and produces the coarse deraining result $I^*_{B,S}$ by subtracting $I^*_{R,S}$. There are two strategies to achieve the association learning: ① We introduce the joint optimization of rain streak removal and background reconstruction tasks to promote the compatibility. ② ALM takes $I^*_{R,S}$ and I_{Rain} as inputs to learn the dynamic degradation map, which guides the network to focus on the valuable components f_{ti} (background texture information) to help accurate texture recovery.

To promote texture recovery, we propose the associated learning module (ALM) based on dynamic learning, shown in Figure 3. ALM is designed to associate rain streaks removal and texture recovery tasks. Specifically, it takes the predicted derained image $I^*_{B,S}$, rain residue $I^*_{R,S}$ at sub-space and the original rainy image I_{Rain} as inputs, and first learns a global rain-distribution mask from $I^*_{R,S}$. The mask provides global degradation priors (location and density) and allows the network to borrow background texture information f_{ti} from I_{Rain}. A novel selection fusion block (SFB), described in Sec. 2.3 and **Supplementary**, is then used to distill the features for enhancement. These procedures of ALM are expressed as

$$f_{ti} = F_M(I^*_{R,S}) \circ F_R(I_{Rain}),$$
$$f_{ALM} = F_{SFB}(f_{ti}, F_B(I^*_{B,S})). \tag{2}$$

In Equation (2), $F_R(\cdot)$ denotes the feature embedding function, including an initial layer and a strided convolution layer to learn the embedding representation of I_{Rain}. $F_M(\cdot)$ refers to the mask function, which is composed of an initial layer and a bottleneck unit to learn the rain-distribution matrices. A Sigmoid operator is followed to normalize them into $[0, 1.0]$. Through the positional guidance and weights, we can borrow the rain-free texture information f_{ti} from the embedding representation of I_{Rain} through the point-wise multiplication $\circ$. The visualization of the mask and f_{ti} are provide in Figure 4. $F_B(\cdot)$ is the embedding function to generate the initial representation of $I^*_{B,S}$. $F_{SFB}(\cdot)$ denotes the fusion function in SFB. Following that, RSN takes f_{ALM} as input to achieve the texture reconstruction as follows:

$$I^*_B = \mathcal{G}_{RSN}(f_{ALM}) + F_{UP}(I^*_{B,S}), \tag{3}$$

Figure 4: Visualization of the mask and extracted background texture information (f_{ti}) in ALM. The mask provides the distortion-localization information while assigning dynamically predicted weights to the input features of undergo little or no distortion locations, helping textures restoration via the extracted complementary texture f_{ti}. For a better visual effect, we respectively select one of the channels (48) from the mask and f_{ti}, and then rescale their pixel values into $[0, 255]$.

where $\mathcal{G}_{RSN}(\cdot)$ denotes the super-resolving function of RSN, and $F_{UP}(\cdot)$ is the Bilinear interpolation.

To train these two sub-networks jointly, we introduce the association optimization, including the image loss (Charbonnier penalty loss [Lai *et al.*, 2017]) and a structural similarity (SSIM) loss, which supervise networks to achieve image and structural fidelity restoration simultaneously. The loss functions are given by

$$\mathcal{L}_{DSN} = \sqrt{(I^*_{B,S} - I_{B,S})^2 + \varepsilon^2} + \alpha \times SSIM(I^*_{B,S}, I_{B,S}),$$
$$\mathcal{L}_{RSN} = \sqrt{(I^*_B - I_B)^2 + \varepsilon^2} + \alpha \times SSIM(I^*_B, I_B),$$
$$\mathcal{L} = \mathcal{L}_{DSN} + \lambda \times \mathcal{L}_{RSN},$$
$$\tag{4}$$

where α and λ are used to balance the loss components, and experimentally set as -0.15 and 1, respectively. The penalty coefficient ε is set to 10^{-3}.

2.2 Dual-branch Fusion Network

The backbone of DSN and RSN is a deep dual-branch fusion network, involving an original resolution branch (ORB) and an encoder-decoder branch. The former is used to preserve spatial structure from the input to the output image. The latter is designed to infer locally-enriched textures.

2.3 Selective Fusion Block

Considering the feature redundancy and knowledge discrepancy among residual blocks and encoding stages, we introduce a novel selective fusion block (SFB) where the low-level contextualized features of earlier stages help consolidate high-level features of the later stages (or scales). Specifically, we incorporate depth-wise separable convolutions and the channel attention layer into SFB to discriminatively aggregate multi-scale features in spatial and channel dimensions. More analyses and discussions to verify its effectiveness are presented in Section 3.2: Ablation Study.

3 Experiments

To validate our proposed DANet, we conduct extensive experiments on synthetic and real-world rainy datasets, and compare DANet with eleven image deraining methods. These methods include MPRNet [Zamir *et al.*, 2021], SW-AL [Huang *et al.*, 2021], DRDNet [Deng *et al.*, 2020], M-SPFN [Jiang *et al.*, 2020b], IADN [Jiang *et al.*, 2020a], PreNet [Ren *et al.*, 2019], UMRL [Yasarla and Patel, 2019], DIDMDN [Zhang and Patel, 2018], RESCAN [Li *et al.*, 2018], LPNet [Fu *et al.*, 2020] and DDC [Li *et al.*, 2019b]. Five evaluation metrics, such as Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM), Feature Similarity (FSIM), Naturalness Image Quality Evaluator (NIQE) and Spatial-Spectral Entropy-based Quality (SSEQ) [Liu *et al.*, 2014], are employed for comparison.

3.1 Implementation Details

Data Collection. Since there exits the discrepancy in training samples for all comparison methods, following [Jiang *et al.*, 2020b], we use $13,700$ clean/rain image pairs from [Zhang *et al.*, 2020; Fu *et al.*, 2017] for training all comparison methods with their publicly released codes by tuning the optimal settings for a fair comparison. For testing, four synthetic benchmarks (Test100 [Zhang *et al.*, 2020], Test1200 [Zhang and Patel, 2018], R100H and R100L [Yang *et al.*, 2017]) and three real-world datasets (Rain in Driving (RID), Rain in Surveillance (RIS) [Li *et al.*, 2019a] and Real127 [Zhang and Patel, 2018]) are considered for evaluation.

Experimental Setup. In our baseline, the number of RCAB is empirically set to 2 for each stage in the encoder-decoder branch and 5 for the original resolution branch with filter numbers of 48. The training images are coarsely cropped into small patches with a fixed size of 128×128 pixels to obtain the training samples. We use Adam optimizer with the learning rate (4×10^{-4} with the decay rate of 0.8 at every 80 epochs till 500 epochs) and batch size (16) to train our DANet on a single NVIDIA Titan Xp GPU.

Model	SR	DSC	SFB	ALM	SSIM	PSNR↑	SSIM↑	Par.↓	Time.↓	GFlops↓
Rain Image	–	–	–	–	–	22.16	0.732	–	–	–
w/o DSC	✓	✗	✓	✓	✓	32.06	0.912	1.596	0.077	43.56
w/o SFB	✓	✓	✗	✓	✓	31.98	0.904	1.520	0.056	54.66
w/o ALM	✓	✓	✓	✗	✓	30.12	0.874	1.544	0.078	48.94
w/o SSIM	✓	✓	✓	✓	✗	32.35	0.914	1.547	0.079	49.11
w/o all	✓	✗	✗	✗	✗	29.48	0.862	1.582	0.053	45.57
DANet*	✗	✓	✓	✓	✓	32.37	0.916	1.535	0.107	166.83
DANet	✓	✓	✓	✓	✓	32.57	0.917	1.547	0.079	49.11

Model	ORB	EDB	PSNR	SSIM	Par.	Time.	GFlops
w/o ORB	✗	✓	31.42	0.896	1.557	0.081	27.59
w/o EDB	✓	✗	32.22	0.910	1.583	0.045	101.48
DANet	✓	✓	32.57	0.917	1.547	0.079	49.11

Table 1: Ablation study on the depth-wise separable convolutions (DSC), associated learning module (ALM), selective fusion block (SFB), SSIM loss, super-resolution (SR), original resolution branch (ORB) and encoder-decoder branch (EDB) on *Test1200* dataset. We obtain the model parameters (Million (M)), average inference time (Second (S)), and calculation complexity (GFlops (G)) of deraining on images with the resolution size of 512×512.

3.2 Ablation Study

Validation on Basic Components. We conduct ablation studies to validate the contributions of individual components, including the depth-wise separable convolutions (DSC), selective fusion block (SFB) and associated learning module (ALM) to the final deraining performance. For simplicity, we denote our final model as DANet and devise the baseline model *w/o* all by removing all these components above. Quantitative results in terms of deraining performance and inference efficiency on the Test1200 dataset are presented in Table 1, revealing that the complete deraining model DANet achieves significant improvements over its incomplete versions. Compared to *w/o ALM* model (removing ALM from DANet), DANet achieves 2.45dB performance gain since the association learning in ALM can help the network to fully exploit the background texture from the rainy image under the guidance of predicted rain distribution. In addition, disentangling the image deraining task into rain streaks removal and texture reconstruction at the low-dimension space exhibits considerable superiority in terms of efficiency (26.2% and 70.6% more efficient in inference time and computational cost, respectively) and restoration quality (referring to the results of DANet and DANet* models (completing the deraining and texture recovery at the original resolution space [Deng *et al.*, 2020])). Moreover, using the depth-wise separable convolutions allows increasing the channel depth with approximately the same parameters, thus enhancing the representation capacity.

We also conduct ablation studies to validate the dual-branch fusion, involving an original resolution branch (ORB) and a U-shaped encoder-decoder branch (EDB). Based on DANet, we devise two comparison models (*w/o* ORB and *w/o* EDB) by removing these two branches in turn. Quantitative results are presented in Table 1. Removing ORB may greatly weaken the representation capability on the spatial structure, leading to obvious performance decline (1.15dB in PSNR) (referring to the results of DANet and *w/o* ORB models). Moreover, EDB allows the network to aggregate multi-scale textures, which are crucial for fine detail recovery.

Methods	RESCAN⋆	DIDMDN	UMRL	PreNet⋆	DDC	IADN	MSPFN	DRDNet	MPRNet	SWAL	DANet (Ours)	DANet♠ (Ours)
Datasets							**Test100/Test1200**					
PSNR↑	25.00/30.51	22.56/29.65	24.41/30.55	24.81/31.36	23.47/28.65	26.71/32.29	27.50/32.39	28.06/26.73	**30.27**/32.91	28.47/30.40	29.23/32.57	29.90/**33.10**
SSIM↑	0.835/0.882	0.818/0.901	0.829/0.910	0.851/0.911	0.806/0.854	0.865/0.916	0.876/0.916	0.874/0.824	**0.897**/0.916	0.889/0.892	0.892/0.917	0.893/**0.919**
FSIM↑	0.909/0.944	0.899/0.950	0.910/0.955	0.916/0.955	0.898/0.936	0.924/0.958	0.928/0.960	0.925/0.920	**0.939**/0.959	0.936/0.950	0.935/0.961	0.938/**0.962**
Datasets							**R100H/R100L**					
PSNR↑	26.36/29.80	17.35/25.23	26.01/29.18	26.77/32.44	15.53/27.60	27.86/32.53	28.66/32.40	21.21/29.24	**30.41/36.40**	29.30/34.60	28.83/34.07	29.96/35.85
SSIM↑	0.786/0.881	0.524/0.741	0.832/0.923	0.858/0.950	0.450/0.877	0.835/0.934	0.860/0.933	0.668/0.883	0.889/**0.965**	0.887/0.958	0.870/0.950	**0.889**/0.962
FSIM↑	0.864/0.919	0.726/0.861	0.876/0.940	0.890/0.956	0.664/0.902	0.875/0.942	0.890/0.943	0.797/0.903	0.910/**0.969**	0.908/0.963	0.897/0.956	**0.911**/0.967
Par.(M)↓	0.150	0.372	0.984	0.169	31.46	0.980	13.35	5.230	3.637	156.54	1.547	2.943
Time (S)↓	0.546	0.315	0.112	0.163	0.125	0.132	0.507	1.426	0.207	0.116	0.079	0.109
GFlops (G)↓	129.28	–	65.74	265.76	42.49	80.99	708.39	–	565.81	614.35	49.11	130.91

Table 2: Comparison results of average PSNR, SSIM, and FSIM on Test100/Test1200/R100H/R100L datasets. We obtain the model parameters (Million) and average inference time (Second) of deraining on images with the size of 512×512. ⋆ denotes the recursive network using the parameter sharing strategy. ♠ denotes the high-accuracy version of our DANet.

Input RESCAN UMRL PreNet MSPFN IADN DRDNet DANet (Ours) Ground Truth

Figure 5: Visual comparison of derained images obtained by seven methods on synthetic rainy datasets.

3.3 Comparison with State-of-the-arts

Synthesized Data. Quantitative results on Test1200 and Test100 are provided in Table 2. As expected, our high-accuracy DANet♠ model (by setting the number of RCAB in ORB to 10 with the channel depth of 64) achieves the best and second-best evaluated scores on Test1200 and Test100. Visual comparisons in Figure 5 show that high-accuracy methods, such as PreNet, UMRL and MSPFN, can effectively eliminate the rain layer and thus bring an improvement in visibility. But they fail to generate visual appealing results due to **the inaccurate color correction**. Likewise, DRDNet focuses on the detail recovery, but shows undesired deraining performance. Besides recovering cleaner and more credible image textures, our DANet produces results with better contrast and less color distortion. Please refer to the "skiing" and "airplane" scenarios. These results further demonstrate the effectiveness of our proposed dynamic association learning.

The results on R100H/R100L datasets are provided in Table 2, indicating that most of the deraining models obtain impressive performance on light rain cases with high consistency. However, only our DANet and MPRNet still perform favorably on heavy rain conditions, exhibiting great superiority over other competing methods in terms of PSNR. **Note that MPRNet achieves limited superiority (average 0.29dB on 4 datasets) over our DANet♠, but requires 89.9%, 23.6% and 332.2% more inference time, parameters and computational cost.** For more convincing evidence, we also provide additional visual comparisons in Figure 5. Our DANet model can remove more rain streaks and achieve better image fideli-

ty than the competing methods. Especially for the heavy rain condition, other deraining methods fail to eliminate the rain streaks, and **introduce considerable artifacts and unnatural color appearance**. (please refer to the first two scenarios in Figure 5). We speculate that these visible improvements on restoration quality may benefit from the elaborate design of the framework, including the novel deraining scheme – rain streaks removal and texture reconstruction as well as the dynamic association learning in ALM. These strategies are integrated into a unified framework, allowing the network to fully exploit the contextual information for feature representation and complete the deraining task more easily.

Real-world Data. We further conduct experiments on three real-world datasets: Real127 [Zhang and Patel, 2018], Rain in Driving (RID), and Rain in Surveillance (RIS) [Li *et al.*, 2019a]. Quantitative results of NIQE and SSEQ [Liu *et al.*, 2014] are listed in Table 3, where smaller NIQE and SSEQ scores indicate better perceptual quality and clearer contents. Again, our proposed DANet is highly competitive, achieving the lowest average values on the RID dataset and the best average score of SSEQ on the Real127 dataset. We visualize the deraining results in Figure 6, showing that DANet produces rain-free images with cleaner and more credible contents, whereas the competing methods fail to remove rain streaks. These evidences indicate that our DANet model performs well in separating rain information while preserving textural details and image naturalness, even under real rain conditions.

Datasets	DIDMDN	RESCAN	DDC	LPNet	UMRL	PreNet	IADN	MSPFN	DRDNet	DANet (Ours)
Real127 (127)	3.929/32.42	3.852/30.09	4.022/29.33	3.989/29.62	3.984/29.48	3.835/29.61	**3.769**/29.12	3.816/29.05	4.208/30.34	3.817/**28.73**
RID (2495)	5.693/41.71	6.641/40.62	6.247/40.25	6.783/42.06	6.757/41.04	7.007/43.04	6.035/40.72	6.518/40.47	5.715/39.98	**4.565/38.20**
RIS (2348)	5.751/46.63	6.485/50.89	5.826/47.80	6.396/53.09	**5.615**/43.45	6.722/48.22	5.909/**42.95**	6.135/43.47	6.269/45.34	5.896/43.27

Table 3: Comparison of average NIQE/SSEQ scores with ten deraining methods on three real-world datasets.
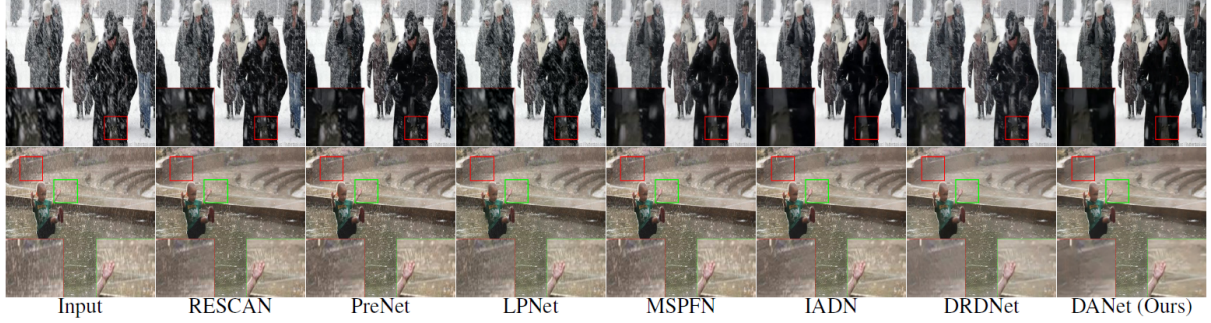

Figure 6: Visual comparison of derained images obtained by seven methods on two real-world scenarios. Please refer to the region highlighted in the boxes for a close up comparison.

Figure 7: Visual comparison of joint image deraining and object detection.

Methods	Rain input	PreNet	IADN	MSPFN	DANet (Ours)
Deraining; Dataset: **COCO350/BDD350**; Image Size: **640 × 480/1280 × 720**					
PSNR	14.79/14.13	17.53/16.90	18.18/17.91	18.23/17.85	**18.51/17.97**
SSIM	0.648/0.470	0.765/0.652	0.790/0.719	0.782/**0.761**	**0.804**/0.725
Ave.inf.time (s)	–/–	0.227/0.764	0.135/0.412	0.584/1.246	0.082/0.118
Object Detection; Algorithm: **YOLOv3**; Dataset: **COCO350/BDD350**; Threshold: 0.6					
Precision (%)	23.03/36.86	31.31/38.66	32.92/40.28	32.56/41.04	**33.51/41.35**
Recall (%)	29.60/42.80	37.92/48.59	39.83/50.25	39.31/50.40	**40.63/51.75**
IoU (%)	55.50/59.85	60.75/61.08	61.96/62.27	61.69/62.42	**62.40/62.53**

Table 4: Comparison results of joint image deraining and object detection on COCO350/BDD350 datasets.

3.4 Impact on Downstream Vision Task

Eliminating the degraded effects of rain streaks under rainy conditions while preserving credible textural details is crucial for object detection. This motivates us to investigate the effect of deraining performance on object detection accuracy based on popular object detection algorithms (*e.g.*, Y-OLOv3 [Redmon and Farhadi, 2018]). Based on our DANet and several representative deraining methods, the restoration procedures are directly applied to the rainy images to generate corresponding rain-free outputs. We then apply the publicly available pre-trained models of YOLOv3 for the detection task. Table 4 shows that DANet achieves the highest P-SNR scores on COCO350 and BDD350 datasets [Jiang *et al.*, 2020b]. Meanwhile, the rain-free results generated by DANet facilitate better object detection performance than other methods. Visual comparisons on two instances in Figure 7 in-

dicate that DANet exhibit a notable superiority in terms of image quality and detection accuracy. We attribute the considerable performance improvements of both deraining and down-stream detection tasks to our association learning between rain streaks removal and detail recovery tasks.

4 Conclusion

We rethink the image deraining as a composite task of rain streak removal, textures recovery and their association learning, and propose a dynamic associated network (DANet) for image deraining. Accordingly, a two-stage architecture and an associated learning module (ALM) are adopted in DANet to account for twin goals of rain streak removal and texture reconstruction while facilitating the learning capability. Meanwhile, the joint optimization promotes the compatibility while maintaining the model compactness. Extensive results on image deraining and joint detection task demonstrate the superiority of our DANet model over the state-of-the-arts.

Acknowledgments

References

[Deng *et al.*, 2020] 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 *CVPR*, pages 14548–14557, 2020.

[Fu *et al.*, 2017] 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.

[Fu *et al.*, 2020] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley. Lightweight pyramid networks for image deraining. *IEEE TNNLS*, 31(6):1794–1807, 2020.

[Garg and Nayar, 2005] Kshitiz Garg and Shree K Nayar. When does a camera see rain? In *ICCV*, volume 2, pages 1067–1074, 2005.

[Han *et al.*, 2021] Yizeng Han, Gao Huang, Shiji Song, Le Yang, Honghui Wang, and Yulin Wang. Dynamic neural networks: A survey. *arXiv preprint arXiv:2102.04906*, 2021.

[Huang *et al.*, 2021] Huaibo Huang, Aijing Yu, Zhenhua Chai, Ran He, and Tieniu Tan. Selective wavelet attention learning for single image deraining. *IJCV*, 129(4):1282–1300, 2021.

[Jiang *et al.*, 2020a] K. Jiang, Z. Wang, P. Yi, C. Chen, Z. Han, T. Lu, B. Huang, and J. Jiang. Decomposition makes better rain removal: An improved attention-guided deraining network. *IEEE TCSVT*, pages 1–1, 2020.

[Jiang *et al.*, 2020b] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, and J. Jiang. Multi-scale progressive fusion network for single image deraining. In *CVPR*, pages 8343–8352, 2020.

[Jiang *et al.*, 2021] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Zheng Wang, Xiao Wang, Junjun Jiang, and Chia-Wen Lin. Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining. *IEEE TIP*, 30:7404–7418, 2021.

[Kang *et al.*, 2012] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. Automatic single-frame-based rain streaks removal via image decomposition. *IEEE TIP*, 21(4):3888–3901, 2012.

[Lai *et al.*, 2017] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *CVPR*, pages 624–632, 2017.

[Li *et al.*, 2018] 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.

[Li *et al.*, 2019a] 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.

[Li *et al.*, 2019b] Siyuan Li, Wenqi Ren, Jiawan Zhang, Jinke Yu, and Xiaojie Guo. Single image rain removal via a deep decomposition–composition network. *CVIU*, 186:48–57, 2019.

[Liu *et al.*, 2014] Lixiong Liu, Bao Liu, Hua Huang, and Alan Conrad Bovik. No-reference image quality assessment based on spatial and spectral entropies. *SPIC*, 29(8):856–863, 2014.

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

[Ren *et al.*, 2019] 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.

[Xu *et al.*, 2021a] Xin Xu, Lei Liu, Xiaolong Zhang, Weili Guan, and Ruimin Hu. Rethinking data collection for person re-identification: active redundancy reduction. *Pattern Recognition*, 113:107827, 2021.

[Xu *et al.*, 2021b] Xin Xu, Shiqin Wang, Zheng Wang, Xiaolong Zhang, and Ruimin Hu. Exploring image enhancement for salient object detection in low light images. *TOMM*, 17(1s):1–19, 2021.

[Yang *et al.*, 2017] 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.

[Yasarla and Patel, 2019] Rajeev Yasarla and Vishal M Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *CVPR*, pages 8405–8414, 2019.

[Zamir *et al.*, 2021] 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.

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

[Zhang *et al.*, 2020] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *IEEE TCSVT*, 30(11):3943–3956, 2020.