SAPNet: Segmentation-Aware Progressive Network for Perceptual Contrastive Deraining

Shen Zheng, Changjie Lu, Yuxiong Wu, Gaurav Gupta

College of Science and Technology, Wenzhou-Kean University

Outline

- Introduction
- Related Works
- Methodology
- Experiment
- Conclusion

Introduction

Figure 1: Deraining comparison at a synthetic rainy image from Rain100H (top) and a real rainy image (bottom). From left to right: Rainy, PreNet (CVPR 2019), MSPFN (CVPR 2020), MPRNet (CVPR 2021), SAPNet (ours).

Challenges

- Trained on Synthesized Images -> Poor Generalization to Real Rains
- Semi-/Unsupervised Learning -> Inflexible with Diverse Rain Patterns
- Muli-Scale Methods -> Huge Parameter and Long Inference Time
- Low-Level Restoration only -> Sub-Optimal High-Level Semantics

HOW TO ADDRESS THESE CHALLENGES?

Related Works

Supervised Methods

- DetailNet [Fu et al., 2017]
- Jorder [Yang et al., 2017]
- DID-MDN [Zhang and Patel, 2018]
- MSPFN [Jiang et al., 2020]
- RESCAN [Li et al., 2018]
- PreNet [Ren et al., 2019]

Semi-/Unsupervised Methods

- SIRR [Wei et al., 2019]
- Syn2Real [Yasarla et al., 2020]
- MOSS [Huang et al., 2021]

Strategies for Deraining

- Recurrent Networks [Ren et al., 2019]
- Multi-Scale Fusion [Jiang et al., 2020]
- Dilated Convolution [Yang et al., 2017]
- Semantic Segmentation [Wang et al., 2019]
- Contrastive learning [Wu et al., 2021]

Model Architecture

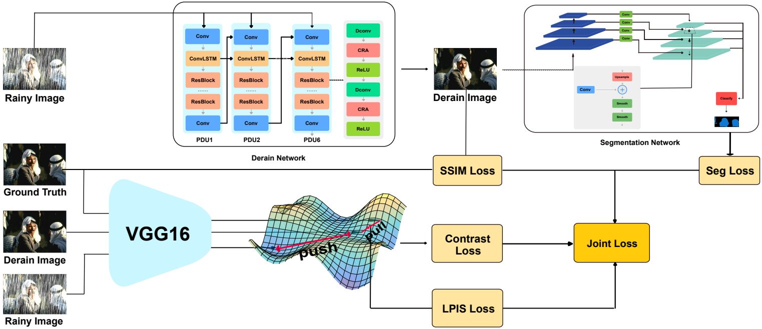


Figure 2: SAPNet joins a derain network for supervised rain removal, a segmentation network for unsupervised background segmentation, and a VGG-16 network for perceptual contrast.

Progressive Dilation

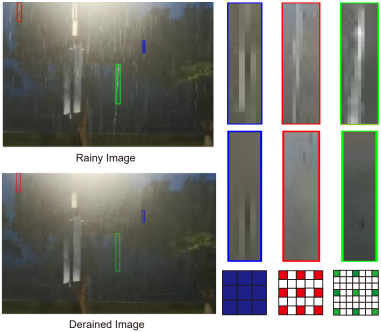


Figure 3: Visual Illustration of Progressive Dilation. Each progressive dilated unit utilize different dilation rate to address multi-scale rain streaks.

Loss Function

- SSIM loss

$$\mathcal{L}_{\text{ssim}} = -\,\mathsf{SSIM}\left(\mathbf{x}^D, \mathbf{x}^G\right) \tag{1}$$

- Segmentation Loss

$$\mathcal{L}_{\text{seg}} = \frac{1}{HW} \sum_{1 \le i \le H, 1 \le j \le W} -\alpha \left(1 - p_{i,j}\right)^{\gamma} \log p_{i,j} \tag{2}$$

- Perceptual Contrastive Loss

$$\mathcal{L}_{\text{pcl}} = \sum_{i=1}^{n} \omega_i \cdot \frac{L1\left(V_i(\mathbf{x}^D), V_i(\mathbf{x}^G)\right)}{L1\left(V_i(\mathbf{x}^D), V_i(\mathbf{x}^R)\right)} \tag{3}$$

Loss Function

- Learned Perceptual Image Similarity Loss

$$\mathcal{L}_{\text{lpisl}} = \sum_{i=1}^{n} \frac{1}{H_i W_i} \sum_{h,w} \left\| \theta_i \odot \left(V_i(\mathbf{x}^D) - V_i(\mathbf{x}^G) \right) \right\|_2^2 \tag{4}$$

- Total Loss

$$\mathcal{L} = \lambda_1 \times \mathcal{L}_{\text{ssim}} + \lambda_2 \times \mathcal{L}_{\text{seg}} + \lambda_3 \times \mathcal{L}_{\text{pcl}} + \lambda_4 \times \mathcal{L}_{\text{lpisl}} \tag{5}$$

Here we set λ_1 to 1, λ_2, λ_3 and λ_4 to 0.1

Implementation Details

Training details

- Training Datasets: 1800 paired images from Rain100H (Train)
- Optimizer: Adam [Kingma and Ba, 2014]
- Framework: Pytorch [Paszke et al., 2019]
- Batch Size 12
- Epoch: 100
- Initial Learning Rate: 0.0001
- Learning Rate Decay: 80% $\downarrow$ on epoch 30, 50, and 80

Evaluation Dataset

Name	Category	Test Samples
Rain12	Synthetic	12
Rain100L	Synthetic	100
Rain100H	Synthetic	100
Rain800	Real	50
SIRR	Real	147
MOSS	Real	48
COCO150	Synthetic	150
CityScape150	Synthetic	150

Table 1: Dataset Description

Ablation Study

	M1	M2	M3	M4	M5	Ours
CRA	✓	✓	✓	✓	✓	✓
UBS		✓	✓	✓	✓	✓
PCL			✓	✓	✓	✓
Dilation				✓	✓	✓
Decay					✓	✓
LPISL						✓
PSNR	27.94	28.34	28.56	28.93	29.36	**29.46**
SSIM	0.882	0.886	0.887	0.891	0.896	**0.897**

Table 2: Ablation result for SAPNet with different model (M) components.

Quantitative Comparison: PSNR↑ / SSIM ↑

Methods	Rain12	Rain100L	Rain100H
Rainy	28.82/0.836	25.52/0.825	12.13/0.349
DDN	28.89/0.897	26.25/0.856	12.65/0.420
RESCAN	33.60/0.953	31.76/0.946	27.43/0.841
PreNet	34.79/<u>0.964</u>	**36.09**/<u>0.972</u>	28.06/<u>0.884</u>
Syn2Real	28.06/0.893	24.24/0.871	15.18/0.397
MSPFN	34.17/0.945	30.55/0.915	26.29/0.798
MOSS	28.82/0.835	27.27/0.885	16.82/0.487
EffDerain	28.11/0.836	25.72/0.800	14.82/0.439
MPRNet	**36.53**/0.963	34.73/0.959	<u>28.52</u>/0.872
Ours	<u>35.50</u>/**0.968**	<u>34.77</u>/**0.973**	**29.46/0.897**

Table 3

Quantitative Comparison: UNIQUE↑ / BRISQUE↓

Methods	Rain800	SIRR	MOSS
Rainy	0.755/26.63	0.672/29.13	0.786/26.47
DDN	0.741/**18.12**	0.670/25.46	0.790/19.92
RESCAN	0.761/21.54	0.671/25.67	0.794/19.02
PreNet	<u>0.762</u>/20.08	0.674/24.17	<u>0.797</u>/18.26
Syn2Real	0.750/<u>20.04</u>	0.689/24.11	0.783/<u>17.96</u>
MSPFN	0.749/22.17	0.657/<u>20.71</u>	0.732/22.64
MOSS	0.743/22.05	0.691/29.06	0.788/24.45
EffDerain	0.737/31.86	0.679/39.33	0.773/38.10
MPRNet	0.754/21.57	**0.697**/28.48	<u>0.797</u>/24.22
Ours	**0.767**/22.21	<u>0.696</u>/**20.68**	**0.798**/**17.88**

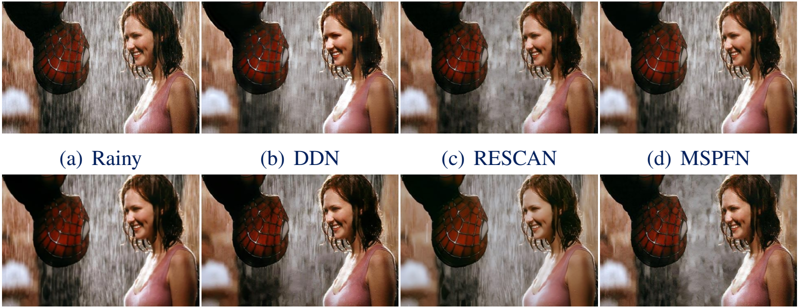
Table 4

Qualitative Comparison: mAP/mPA/mIOU

Metrics	Rainy	DDN	RESCAN	PreNet	EffDerain	Syn2Real	MOSS	**Ours**	GT
mAP (%)	52.1	65.1	78.5	81.0	68.2	55.4	73.2	**82.2**	85.4
mPA (%)	65.3	66.4	70.3	73.8	67.3	59.9	76.6	**77.2**	78.8
mIOU (%)	50.7	53.6	57.3	56.3	56.7	49.9	60.1	**62.2**	66.7

Table 5: mAP↑, mPA↑ and mIOU↑ comparison

Qualitative Comparison: SIRR

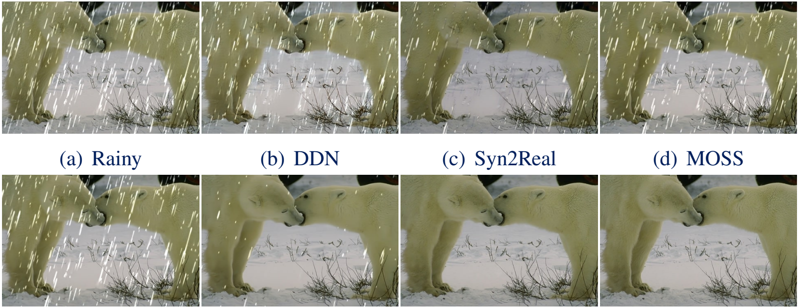


(a) Rainy (b) DDN (c) RESCAN (d) MSPFN

(e) EffDerain (f) MOSS (g) MPRNet (h) Ours

Figure 4: Visual comparison at SIRR

Qualitative Comparison: Rain100L

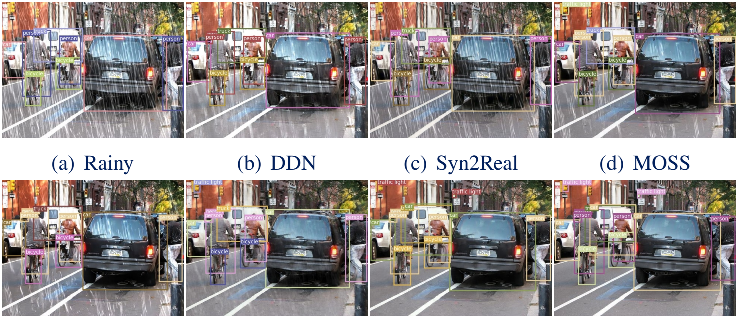


(a) Rainy (b) DDN (c) Syn2Real (d) MOSS

(e) MSPFN (f) MPRNet (g) Ours (h) GT

Figure 5: Visual comparison at Rain100L

Object Detection at Rainy Images



(a) Rainy (b) DDN (c) Syn2Real (d) MOSS

(e) EffDerain (f) RESCAN (g) Ours (h) GT

Figure 6: Object detection result at COCO150

Semantic Segmentation at Rainy Images

(a) Rainy (b) DDN (c) Syn2Real (d) MOSS (e) EffDerain

(f) RESCAN (g) PreNet (h) MSPFN (i) Ours (j) GT

Figure 7: Semantic segmentation result at CityScape150

Conclusion and Future Works

- Segmentation-Aware Progressive Network (SAPNet)
- Progressive Dilated Unit (PDU)
- Unsupervised Background Segmentation (UBS)
- Perceptual Contrastive Loss (PCL)
- Learned Perceptual Image Similarity Loss (LPISL)
- Future: Detection-Driven Deraining, Low-Light Deraining

References I

Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., and Paisley, J. (2017).
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.

Huang, H., Yu, A., and He, R. (2021).
Memory oriented transfer learning for semi-supervised image deraining.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7732–7741.

Jiang, K., Wang, Z., Yi, P., Chen, C., Huang, B., Luo, Y., Ma, J., and Jiang, J. (2020).
Multi-scale progressive fusion network for single image deraining.
In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8346–8355.

Kingma, D. P. and Ba, J. (2014).
Adam: A method for stochastic optimization.
arXiv preprint arXiv:1412.6980.

Li, X., Wu, J., Lin, Z., Liu, H., and Zha, H. (2018).
Recurrent squeeze-and-excitation context aggregation net for single image deraining.
In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 254–269.

References II

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019).
Pytorch: An imperative style, high-performance deep learning library.
Advances in neural information processing systems, 32:8026–8037.

Ren, D., Zuo, W., Hu, Q., Zhu, P., and Meng, D. (2019).
Progressive image deraining networks: A better and simpler baseline.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3937–3946.

Wang, S., Wen, B., Wu, J., Tao, D., and Wang, Z. (2019).
Segmentation-aware image denoising without knowing true segmentation.
arXiv preprint arXiv:1905.08965.

Wei, W., Meng, D., Zhao, Q., Xu, Z., and Wu, Y. (2019).
Semi-supervised transfer learning for image rain removal.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3877–3886.

References III

Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., and Ma, L. (2021).
Contrastive learning for compact single image dehazing.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10551–10560.

Yang, W., Tan, R. T., Feng, J., Liu, J., Guo, Z., and Yan, S. (2017).
Deep joint rain detection and removal from a single image.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1357–1366.

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

Zhang, H. and Patel, V. M. (2018).
Density-aware single image de-raining using a multi-stream dense network.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 695–704.