Supplementary for Syn2Real Transfer Learning for Image Deraining using Gaussian Processes

Rajeev Yasarla* Vishwanath A. Sindagi* Vishal M. Patel
Johns Hopkins University
Department of Electrical and Computer Engineering, Baltimore, MD 21218, USA
{ryasarl1, vishwanathsindagi, vpatel36}@jhu.edu

1. Introduction

Here we provide the supplementary material for the paper *Syn2Real Transfer Learning for Image Deraining using Gaussian Processes*. We provide more details such as network configuration, ablation experiments for different hyperparameters along with additional qualitative results.

2. Network configuration

The network consists of U-Net [5] style encoder decoder with dense connections [2], constructed using Denseblock. Denseblock contains a sequence of three 3×3 convolutional layers as shown in the Fig. 1. The proposed network consists of Encoder($h(., \theta_{enc})$) and Decoder($g(., \theta_{dec})$). Encoder($h(., \theta_{enc})$) with the following sequence:
Conv2d 3×3(3,16)-Denseblock(16,32)-Downsample-Denseblock(32,32)- Downsample-Denseblock(32,32)-Downsample-Denseblock(32,64)-Denseblock(64,64)-Denseblock(64,64)
Decoder($g(., \theta_{dec})$)) consists of the following sequence:
Denseblock(64,32)-Upsample-Denseblock(32,32)-Upsample-Denseblock(32,16)-Upsample-Conv2d 3×3(16,3),
where Conv2d $3 \times 3(m, n)$ is a 3×3 convolutional layer with m input channels and n output channels,Denseblock(m, n) meansDenseblock with m input channels and n output channels

3. Ablation study: Hyperparameters

Here, we provide the results of the ablation experiments conducted to analyze the effect of different hyperparameters: λ_{unsup} (see Eq. 18), number of nearest neighbours: N_n and N_f (see Eq. 16).

Table 1 illustrates the performance of the proposed method for different values of λ_{unsup}. For this experiment, we used 40% of data as labeled and rest was used as unlabeled. It can be observed that the proposed method is fairly robust to different values of λ_{unsup}.

Table 1: SSL experiments on Rain200H [8] dataset: The percentage of labeled data used for training is 40% and unlabeled is 60%.

Dataset	$\mathcal{D}_{\mathcal{L}}$	$\mathcal{D}_{\mathcal{U}}$	Metrics	λ_{unsup}			
				0.000015	0.00015	0.0015	0.015
Rain200H	40%	60%	PSNR	25.98	26.28	26.34	26.43
			SSIM	0.828	0.833	0.833	0.835

Further, we also analyze the performance of the network for different values of nearest neighbours used for obtaining the pseudo-GT via the Gaussian Processes. The results for this experiment are provided in Table 2. It can be observed that the results are approximately consistent for different values of N_n and N_f.

Table 2: SSL experiments on Rain200H [8] dataset: The percentage of labeled data used for training is 40% and unlabeled is 60%.

Dataset	$\mathcal{D}_{\mathcal{L}}$	$\mathcal{D}_{\mathcal{U}}$	Metrics	No. of nearest neighbors N_n No. of farthest neighbors N_f		
				N_n=16 N_f=16	N_n=32 N_f=32	N_n=64 N_f=64
Rain200H	40%	60%	PSNR	26.18	26.38	26.34
			SSIM	0.829	0.827	0.833

4. Additional Qualitative results

Figures 1-4 illustrate additional qualitative results on real-world and synthetic datasets for different experiments.

5. Perceptual loss

Inspired by the importance of the perceptual loss in many image restoration tasks [3, 10], we use it to further improve the visual quality of the de-rained images. The perceptual loss is feature based loss, and in our case, extracted features from layer $relu1\_2$ of pretrained network VGG-16[6], and computed perceptual loss similar to method proposed

*equal contribution

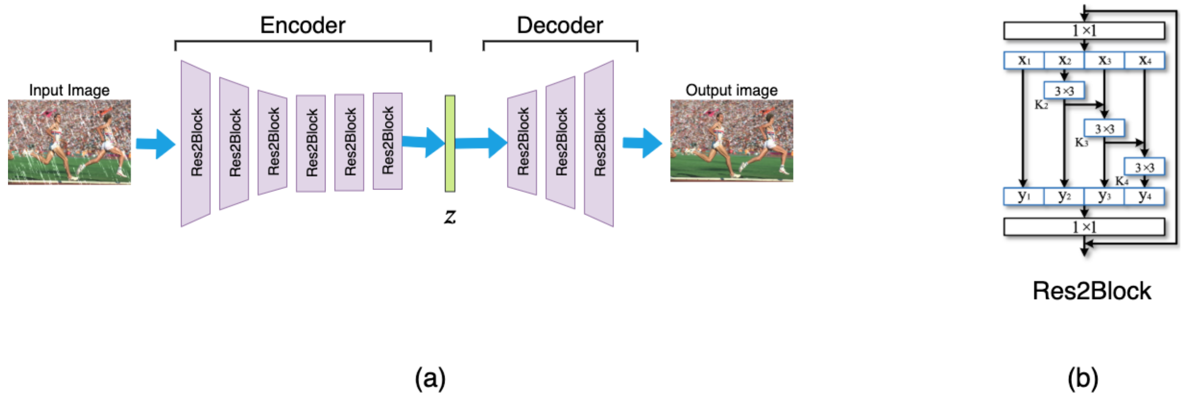


Figure 1: (a) U-Net based encoder and decoder with dense blocks (b) Res2Block structure. Note z is latent vector explained in main paper

in [4, 9]. Let $\Phi_{VGG}(.)$ denote the features obtained using the VGG16 model [6], then the perceptual loss is defined as follows

$$\mathcal{L}_p = \|\Phi_{VGG}(y_l^{pred}) - \Phi_{VGG}(y_l)\|_2^2, \tag{1}$$

where $y_l^{pred} = g(z, \theta_{dec})$ is the predicted output, y_l is the ground-truth, $z = h(x, \theta_{enc})$, and x is rainy image.

References

[1] X Fu, J Huang, D Zeng, X Ding, Y Liao, and J Paisley. Removing rain from single images via a deep detail network. *In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1715–1723, 2017. 3, 4

[2] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 1

[3] J Johnson, A Alahi, and L Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. *In European Conference on Computer Vision(ECCV)*, pages 694–711, 2016. 1

[4] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. 2016. 2

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1

[6] K Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*, 2014. 1, 2

[7] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3877–3886, 2019. 3, 4

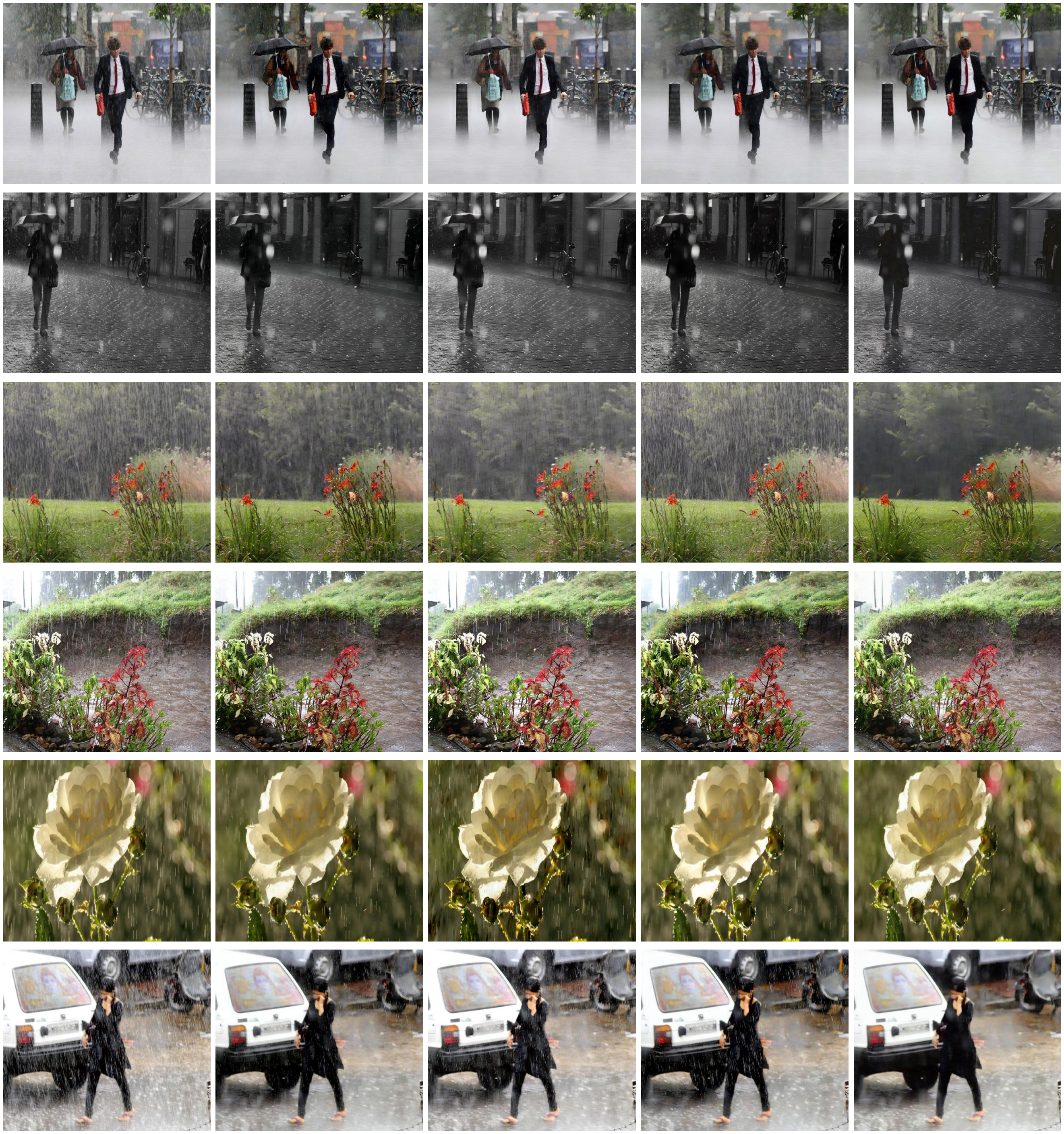
[8] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1357–1366, 2017. 1

[9] Hang Zhang and Kristin Dana. Multi-style generative network for real-time transfer. *arXiv preprint arXiv:1703.06953*, 2017. 2

[10] H. Zhang and Vishal M Patel. Convolutional sparse and lowrank coding-based rain streak removal. *7 IEEE Winter Conference In Applications of Computer Vision(WACV)*, pages 1259–1267, 2017. 1

[11] H. Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, abs/1802.07412, 2018. 3, 4

Figure 2: Qualitative results on DDN-SIRR synthetic test set.



Rainy Image | DDN [1](CVPR'17) | DID-MDN [11](CVPR'18) | SIRR [7](CVPR'19) | Ours

Figure 3: Qualitative results on DDN-SIRR real-world test set

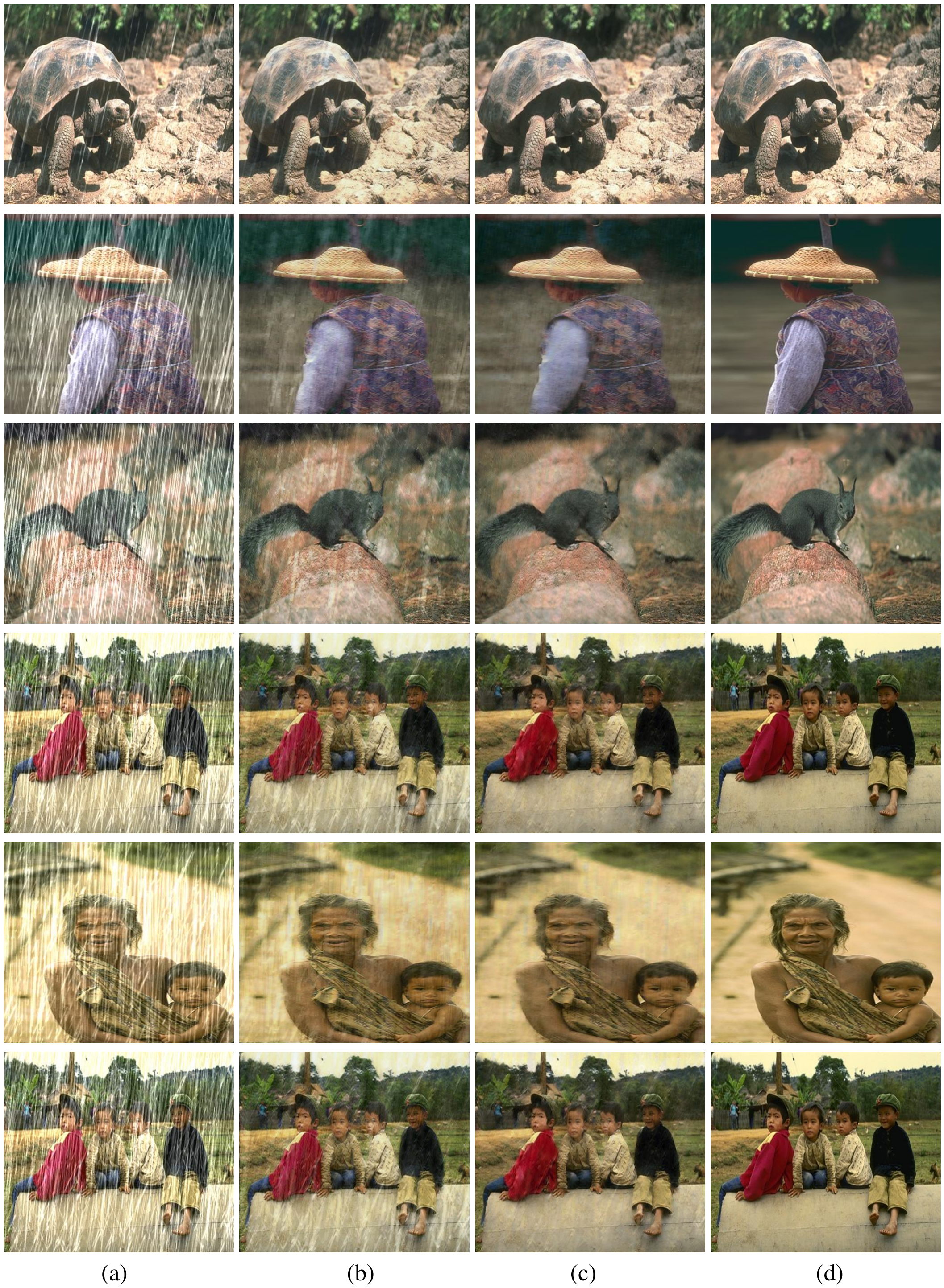
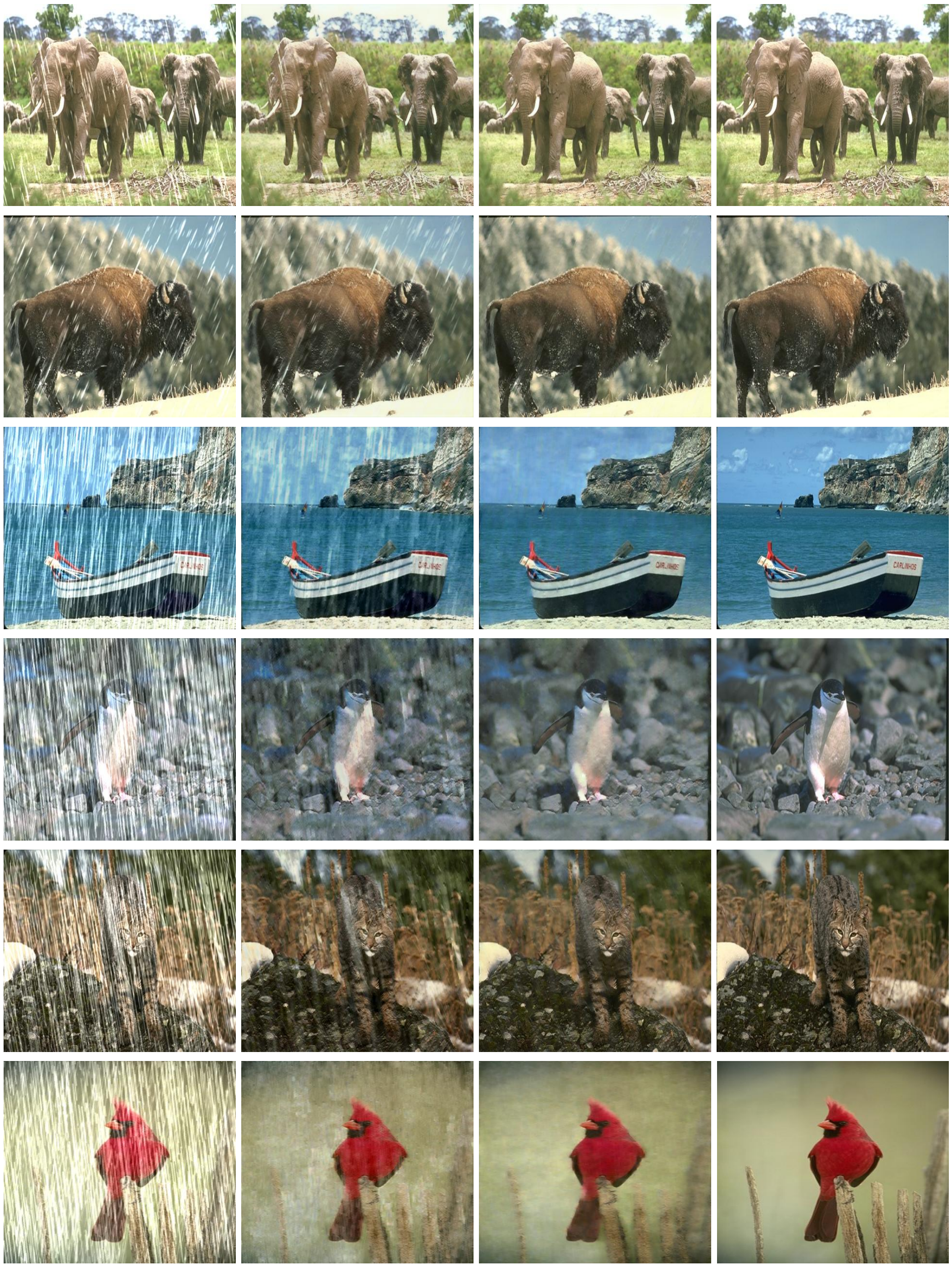


Figure 4: Results of experiments with 10% labeled data on Rain200H (a) Input rainy image (a) Using only labeled data (c) Using labeled and unlabeled data.(d) Ground-Truth image



(a) (b) (c) (d)

Figure 5: Results of experiments with 40% labeled data on Rain200H (a) Input rainy image (a) Using only labeled data (c) Using labeled and unlabeled data.(d) Ground-Truth image