

Supplementary Material for Prior-based Domain Adaptive Object Detection for Hazy and Rainy Conditions

Vishwanath A. Sindagi*, Poojan Oza*, Rajeev Yasarla, and Vishal M. Patel

Department of Electrical and Computer Engineering,
Johns Hopkins University, 3400 N. Charles St, Baltimore, MD 21218, USA
{vishwanathsindagi,poza2,ryasar11,vpatel136}@jhu.edu

This contains the supplementary material for the paper *Prior-based Domain Adaptive Object Detection for Hazy and Rainy Conditions*. Due to the space limitations in the submitted paper, in this supplementary material we provide additional details like network configuration, details of the newly introduced Rainy-Cityscapes dataset, additional analysis and discussion about results.

1 Additional Results

1.1 Results with ResNet-152

Table 1 shows the additional results on the Cityscapes→Foggy-Cityscapes experiments, when ResNet-152 network architecture is used as backbone of detection network. From the results we can see that ResNet-152 performs better compared to the corresponding VGG16 baselines. For FRCNN+P₄₅+R₄₅ baseline with ResNet-152, residual feature recovery blocks and prior estimation networks are applied on fourth and fifth conv block of the network. The results in Table 1 show that the proposed approach generalizes well for different network architectures.

Table 1. Performance comparison for the Cityscapes → Foggy-Cityscapes experiment. Red and Blue color fonts show best and second best performance.

| Method | | prsn | rider | car | truc | bus | train | bike | bicycle | mAP |
|-------------------|--|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| DAFaster | | 25.0 | 31.0 | 40.5 | 22.1 | 35.3 | 20.2 | 20.0 | 27.1 | 27.6 |
| SCDA | | 33.5 | 38.0 | 48.5 | 26.5 | 39.0 | 23.3 | 28.0 | 33.6 | 33.8 |
| SWDA | | 29.9 | 42.3 | 43.5 | 24.5 | 36.2 | 32.6 | 30.0 | 35.3 | 34.3 |
| DM | | 30.8 | 40.5 | 44.3 | 27.2 | 38.4 | 34.5 | 28.4 | 32.2 | 34.6 |
| MTOR | | 30.6 | 41.4 | 44.0 | 21.9 | 38.6 | 40.6 | 28.3 | 35.6 | 35.1 |
| NL | | 35.1 | 42.1 | 49.2 | 30.1 | 45.3 | 26.9 | 26.8 | 36.0 | 36.5 |
| Ours (VGG16) | FRCNN | 25.8 | 33.7 | 35.2 | 13.0 | 28.2 | 9.1 | 18.7 | 31.4 | 24.4 |
| | FRCNN+P ₄₅ +R ₄₅ | 36.4 | 47.3 | 51.7 | 22.8 | 47.6 | 34.1 | 36.0 | 38.7 | 39.3 |
| Ours (ResNet-152) | FRCNN | 32.4 | 42.2 | 36.0 | 19.8 | 26.4 | 4.7 | 22.7 | 32.6 | 27.1 |
| | FRCNN+P ₄₅ +R ₄₅ | 34.9 | 46.4 | 51.4 | 29.2 | 46.3 | 43.2 | 31.7 | 37.0 | 40.0 |

* equal contribution

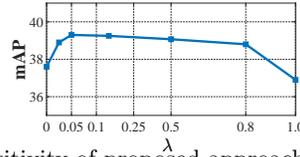


Fig. 1. Performance sensitivity of proposed approach with varying λ parameter.

Table 2. Results of the ablation experiments from Cityscapes \rightarrow Foggy-Cityscapes. Here, * indicates additional experiments that are not included in the paper.

| Method | mAP |
|--|-------------|
| FRCNN | 24.4 |
| FRCNN+D ₅ | 30.0 |
| FRCNN+D ₅ +R ₅ | 32.9 |
| FRCNN+D ₄₅ * | 33.2 |
| FRCNN+P ₅ +R ₅ | 36.5 |
| FRCNN+P ₄₅ * | 37.4 |
| FRCNN+P ₄₅ +R ₄₅ | 39.3 |

Table 3. Network configuration details for Prior Estimation Networks.

| Prior Estimation Network |
|--|
| Gradient Reversal Layer |
| Conv, 1 \times 1, 64, stride 1, BN, ReLU |
| Conv, 3 \times 3, 64, stride 1, BN, ReLU |
| Conv, 3 \times 3, 64, stride 1, BN, ReLU |
| Conv, 3 \times 3, 3, stride 1, Tanh |

Table 4. Network configuration details for Residual Feature Recovery Blocks.

| Residual Feature Recovery Block - Conv4 | Residual Feature Recovery Block - Conv5 |
|--|--|
| Maxpool, 2 \times 2, stride 2 | Maxpool, 2 \times 2, stride 2 |
| Conv, 3 \times 3, 256, stride 1, padding 1, ReLU | Conv, 3 \times 3, 512, stride 1, padding 1, ReLU |
| Conv, 3 \times 3, 512, stride 1, padding 1, ReLU | Conv, 3 \times 3, 512, stride 1, padding 1, ReLU |
| Conv, 3 \times 3, 512, stride 1, padding 1 | Conv, 3 \times 3, 512, stride 1, padding 1 |

1.2 Ablation Analysis

The Table 2 provides additional ablation experiments with different network configuration. The analysis is done with VGG-16 network architecture as backbone for detection network.

1.3 Parameter Sensitivity

In Fig.1, we provide sensitivity of the proposed approach with respect to λ parameter. The parameter λ controls the effect of regularization applied on residual feature norm coming from residual feature recovery blocks. The parameter sensitivity experiment was performed for adaptation from Cityscapes \rightarrow Foggy-Cityscapes with VGG16 network architecture as backbone of detection network.

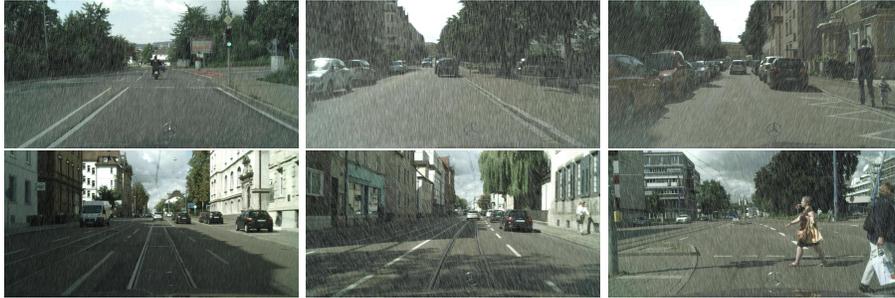


Fig. 2. Sample images from the Rainy-Cityscapes dataset.

2 Network configurations

The network configuration details of different modules such as Residual Feature Recovery Blocks (RFRB) and Prior Estimation Network (PEN) are shown in Table 3 and 4.

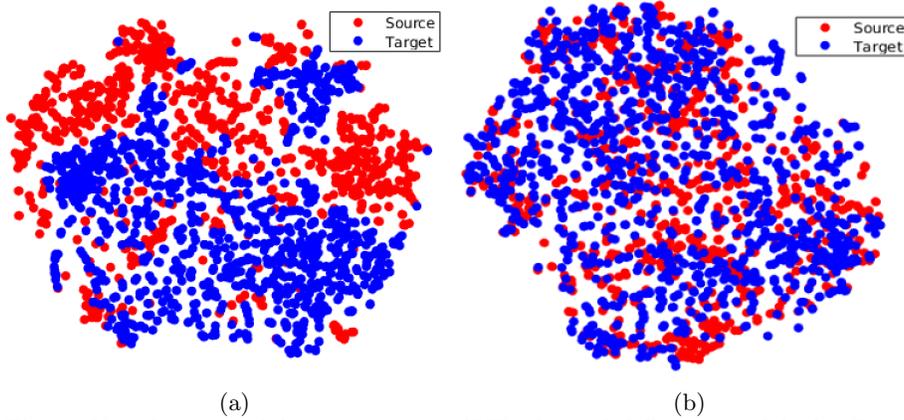


Fig. 3. Visualization of features using t-SNE plots of different models for Foggy-Cityscapes. (a) Model trained using only the domain adaptive loss. (b) Model trained using the prior adversarial loss. With the domain adaptive loss, the features are not perfectly aligned. Introducing the prior adversarial loss results in better alignment.

3 Rainy-Cityscapes

Fig. 2 shows few sample image examples from Rainy-Cityscapes dataset introduced in the paper.

4 Qualitative Results

4.1 t-SNE Feature Visualization

Fig. 3 shows the distribution of the source and target features for Cityscapes \rightarrow Foggy-Cityscapes experiment. Fig. 3(a) shows the distribution of features from the model which is trained using only domain adversarial loss. Fig. 3(b) shows the distribution of features from the model which is trained using the proposed prior adversarial loss.

4.2 Cityscapes \rightarrow Foggy-Cityscapes

Fig. 4 shows comparison of detection results from the proposed method with DA-Faster RCNN [2] on the Cityscapes \rightarrow Foggy-Cityscapes adaptation.

4.3 Cityscapes \rightarrow Rainy-Cityscapes

Fig. 5 shows comparison of detection results from the proposed method with DA-Faster RCNN [2] on the Cityscapes \rightarrow Rainy-Cityscapes adaptation.

4.4 Cityscapes \rightarrow RTTS

Fig. 6 shows comparison of detection results from the proposed method with DA-Faster RCNN [2] on the Cityscapes \rightarrow RTTS adaptation.

5 Preliminary Study : Domain Adaptive Detection in Snow

Similar to rain, a snow image can be considered as superposition of clean image and snow-residues. If we denote snowy image as I_{snow} , clean image as I_{clean} and snow-residues as $I_{snow-res}$, then snow image can be mathematically written as,

$$I_{snow} = I_{clean} + I_{snow-res},$$

Here, $I_{snow-res}$ can be used as a prior and can be extracted from the snowy image with the help of GMM, similar to the case of rain. We use this model to perform a preliminary experiment on snowy conditions in Sec. 5.

Following the above mention model, we present a preliminary study of domain adaptive detection in snowy conditions. For the experiment, we consider adaptation scenario WIDER-Face \rightarrow UFDD-Snow, which adapts the detection network from labeled clean image dataset, WIDER-Face [8], to unlabeled snow affected image dataset, UFDD-Snow [4]. We compare with the state of art method SWDA [6]. For proposed approach we use the GMM prior as snow weather prior to extract snow residues as explained in the Sec. 6. For both methods we use VGG16 [7] as the backbone of the detection network. As we can see from the Table 5 that proposed method is able to improve $\sim 10\%$ over the Faster-RCNN baseline and $\sim 3\%$ the method SWDA.



Fig. 4. Detection results on Foggy-Cityscapes. (a) DA-Faster RCNN [2] (b) Proposed method. The bounding boxes are colored based on the detector confidence using the color map as shown. As we can see from the figures, the proposed method is able to produce high confidence predictions and is able to detect more objects in the image.

6 Extending to Other Weather Conditions

Many other weather conditions have been researched in the literature and have a mathematical model based on the physics of image formation [3], [9], [9]. Our



Fig. 5. Detection results on Rainy-Cityscapes. (a) DA-Faster RCNN [2] (b) Proposed method. The bounding boxes are colored based on the detector confidence using the color map as shown. As we can see from the figures, the proposed method is able to produce high confidence predictions and is able to detect more objects in the image.

method can be easily extensible to other conditions by utilizing these mathematical models. Here we provide some examples of the prior that can be used based



Fig. 6. Detection results on RTTS Dataset. (a) DA-Faster RCNN [2] (b) Proposed method. The bounding boxes are colored based on the detector confidence using the color map as shown. As we can see from the figures, the proposed method is able to produce high confidence predictions and is able to detect more objects in the image.

on their corresponding mathematical models:

Table 5. Results of the adaptation experiments from WIDER-Face \rightarrow UFDD-Snow.

| Method | mAP |
|-----------|-------------|
| FRCNN [5] | 52.1 |
| SWDA [6] | 58.7 |
| Proposed | 61.9 |

1. Low-light/Sunshine: Any image can be modeled with the help of Luminance (L) - Reflectance (R) model [1]. This model follows an additive formula in the logarithm and can be written as,

$$\log(I) = \log(L) + \log(R),$$

Here, I is the Image, L is luminance map and R is the reflectance map. Considering this model, for low-light/sunshine conditions, luminance map can be used as a prior information and homomorphic filtering [1] can be used to extract luminance map from the image.

2. Water puddle: The water puddle model has been extensively discussed in [3]. In the mathematical formulation provided by [3], we can use reflection attention as a prior. The details regarding how to extract the reflectance attention from the image is provided in [3].

3. Adherent water drops: A detailed discussion adherent water drops mathematical model based on physics of water-droplets is provided in [9]. In the model formulation, the term I_r (Eq. 7 [9]) can be used as prior.

For the paper we focused mainly on rainy and hazy conditions. However, in future we plan to study the above mentioned weather conditions with the help of priors extracted from the corresponding mathematical model as discussed above.

References

1. blogs.mathworks.com/steve/2013/06/25/homomorphic-filtering-part-1/
2. Chen, Y., Li, W., Sakaridis, C., Dai, D., Gool, L.V.: Domain adaptive faster r-cnn for object detection in the wild. 2018 IEEE Conference on Computer Vision and Pattern Recognition pp. 3339–3348 (2018)
3. Han, X., Nguyen, C., You, S., Lu, J.: Single image water hazard detection using fcn with reflection attention units. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 105–120 (2018)
4. Nada, H., Sindagi, V.A., Zhang, H., Patel, V.M.: Pushing the limits of unconstrained face detection: a challenge dataset and baseline results. arXiv preprint arXiv:1804.10275 (2018)
5. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp. 91–99 (2015)

6. Saito, K., Ushiku, Y., Harada, T., Saenko, K.: Strong-weak distribution alignment for adaptive object detection. CoRR **abs/1812.04798** (2018)
7. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
8. Yang, S., Luo, P., Loy, C.C., Tang, X.: Wider face: A face detection benchmark. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5525–5533 (2016)
9. You, S., Tan, R.T., Kawakami, R., Mukaigawa, Y., Ikeuchi, K.: Adherent raindrop modeling, detection and removal in video. IEEE transactions on pattern analysis and machine intelligence **38**(9), 1721–1733 (2015)