Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions: Supplementary Material

Wenyu Liu[1,2], **Gaofeng Ren** [3], **Runsheng Yu** [4], **Shi Guo** [5], **Jianke Zhu** [1,2 *], **Lei Zhang** [3,5]

[1] Colleage of Computer Science and Technology, Zhejiang University
[2] Alibaba-Zhejiang University Joint Institute of Frontier Technologies
[3] DAMO Academy, Alibaba Group
[4] The Hong Kong University of Science and Technology
[5] The HongKong Polytechnic University
{liuwenyu.lwy, jkzhu}@zju.edu.cn, {gaof.ren, runshengyu}@gmail.com, {csshiguo, cslzhang}@comp.polyu.edu.hk

Appendix

Defog Filter Design

Motivated by the conventional dark channel prior method (He, Sun, and Tang 2009), we design a defog filter with a learnable parameter. In the atmospheric scattering model (McCartney 1976; Narasimhan and Nayar 2002), the formation of a hazy image can be formulated as follows:

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{1}$$

where $I(x)$ is the foggy image, and $J(x)$ represents the scene radiance (clear image). A is the global atmospheric light, and $t(x)$ is the medium transmission map.

In order to recover the clear image $J(x)$, the key is to obtain the global atmospheric light A and the medium transmission map $t(x)$. To this end, we first compute the dark channel map and pick the top 1000 brightest pixels. Then, A is estimated by averaging these 1000 pixels in the haze image $I(x)$. From Eq. (1), we can derive that

$$\frac{I^C(x)}{A^C} = t(x)\frac{J^C(x)}{A^C} + (1 - t(x)) \tag{2}$$

where C donates the RGB color channel. By taking two min operations, one on the channels and one on a local patch, in the above equation, we can obtain:

$$\min_C(\min_{y\in\Omega(x)} \frac{I^C(y)}{A^C}) = t(x)\min_C(\min_{y\in\Omega(x)} \frac{J^C(y)}{A^C}) + (1 - t(x)) \tag{3}$$

Based on the dark channel prior, we can get that

$$J^{dark}(x) = \min_C(\min_{y\in\Omega(x)} J^C(y)) = 0 \tag{4}$$

Since A^C is always positive, Eq. (4) can be written as:

$$\min_C(\min_{y\in\Omega(x)} \frac{J^C(y)}{A^C}) = 0 \tag{5}$$

By substituting Eq. (5) into Eq. (2), we can obtain:

$$t(x) = 1 - \min_C(\min_{y\in\Omega(x)} \frac{I^C(y)}{A^C}) \tag{6}$$

We further introduce a parameter ω to control the degree of defogging. There is:

$$t(x, \omega) = 1 - \omega \min_C(\min_{y\in\Omega(x)} \frac{I^C(y)}{A^C}) \tag{7}$$

Since the above operation is differentiable, we can optimize ω through back propagation to make *defog filter* more conducive to foggy image detection.

Experiments

Experiments on Foggy Images We compare our method with the baseline YOLOv3 (Redmon and Farhadi 2018), $Defog + Detect$ (Hang et al. 2020; Liu et al. 2019), domain adaptation (Hnewa and Radha 2021), and multi-task learning (Huang, Le, and Jaw 2020). For domain adaptation approach, we employ the one-stage multi-scale domain-adaptive detector DAYOLO (Hnewa and Radha 2021) with multiple domain adaptation paths and the corresponding domain classifiers at different scales of YOLOv3. We set the loss weight $\lambda = 0.1$ for training, and each batch has 2 images, one from the source domain and the other from the target domain. Other hyperparameters are set the same as in the original paper.

Fig. 1 shows several visual examples of our IA-YOLO method, the baseline YOLOv3 II and the $Defog + Detect$ methods. Both GridDehaze (Liu et al. 2019) and MS-BDN (Hang et al. 2020) can reduce the haze effect, which is generally beneficial to detection. Our IA-YOLO method not only reduces the haze, but also enhances the local image gradients, which lead to better detection performance.

Fig. 2 shows two examples on how the CNN-PP module predicts DIP's parameters , including detailed parameter values and the images processed by each sub-filter. The CNN-PP is able to learn a set of DIP parameters for each image according to its brightness, color, tone and weather-specific information. After the input image is processed by the learned DIP module, more image details are revealed, which are conducive to the subsequent detection task.

Experiments on Low-light Images The total number of images in VOC_norm_trainval, VOC_norm_test and ExDark_test are 12334, 3760 and 2563, respectively. The numbers of instances are listed in Table 1.

*corresponding author

Figure 1: Detection results by different methods on real-world RTTS foggy images. From left to right: YOLOv3 II, GirdDehaze + YOLOv3 I, MSBDN + YOLOv3 I and our IA-YOLO. The proposed method learns to reduce the haze and enhance the image contrast, which leads to better detection performance with fewer missed and wrong detections.

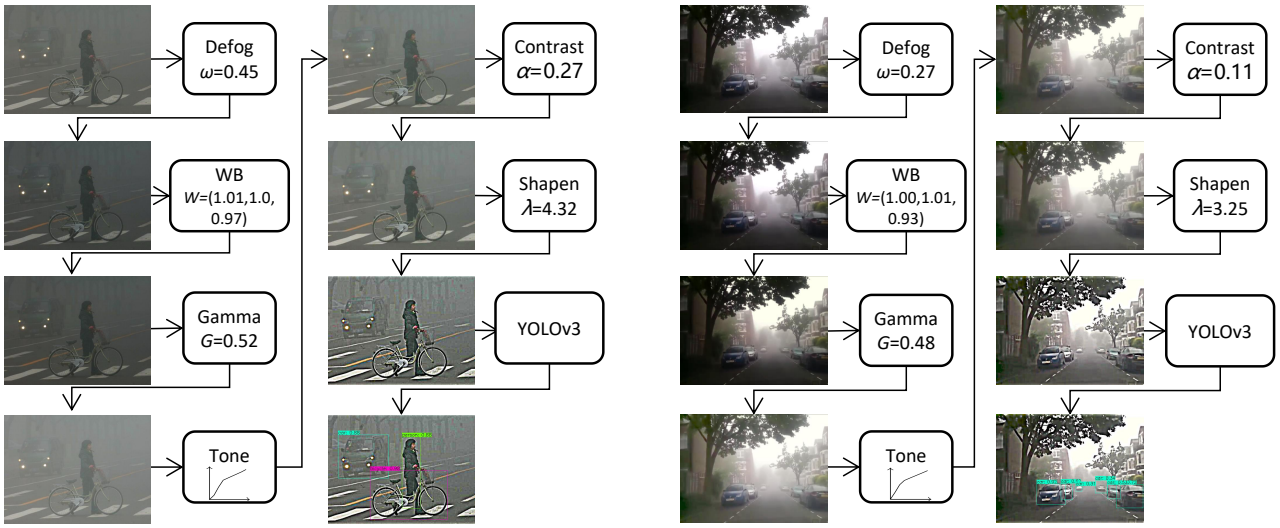

Figure 2: Examples of learned DIP module and their filtering outputs. The image-adaptive processing module can output the corresponding filter parameters according to the brightness, color, tone and weather-information of each input image, so as to get better detection performance.

Dataset	person	bicycle	car	bus	motorbike	boat	bottle	cat	chair	dog	Total
Voc_norm_trainval	13256	1064	3267	822	1052	1140	1764	1593	3152	2025	29135
Voc_norm_test	4528	337	1201	213	325	263	469	358	756	489	8939
ExDark_test	2235	418	919	164	242	515	433	425	609	490	6450

Table 1: Statistics of the used datasets.

Method	Additional Params	Speed(ms)
YOLOv3	/	31
YOLOv3_deep II	412K	35
ZeroDCE	79K	34
MSBDN	31M	94
GridDehaze	958K	51
IA-YOLO(Ours)	165K	44

Table 2: Efficiency analysis on the compared methods.

We compare our presented method with the baseline YOLOv3, $Enhance + Detect$ (Guo et al. 2020), DAYOLO, and DSNet on the three testing datasets. Fig. 3 shows several visual examples of our IA-YOLO method, the baseline YOLOv3 II and the $Enhance + Detect$ methods. It can be observed that both Zero-DCE (Guo et al. 2020) and IA-YOLO can brighten the image and reveal the image details. The proposed IA-YOLO can further increases the contrast of the input image, which is essential to object detection.

Efficiency Analysis

In our proposed IA-YOLO framework, we introduce a learning module of CNN-PP into YOLOv3, which is a small network containing five convolutional layers and two fully connected layers. Table 2 shows the efficiency analysis of some methods used in our experiments. The methods not listed are validated using the YOLOv3 architecture. The second column lists the number of additional parameters over the YOLOv3 model. The third column lists the running time on a $544 \times 544 \times 3$ resolution image with a single Tesla V100 GPU. It can be seen that IA-YOLO only adds 165K trainable parameters over YOLOv3 while achieving the best performance on all testing with comparable running time. Note that IA-YOLO has fewer trainable parameters than YOLO_deep II but its running time is longer. This is because that the filtering process in the DIP module incurs additional computation.

References

Guo, C. G.; Li, C.; Guo, J.; Loy, C. C.; Hou, J.; Kwong, S.; and Cong, R. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of IEEE/CVF Conference Computer Vision Pattern Recognition (CVPR)*, 1780–1789.

Hang, D.; Jinshan, P.; Zhe, H.; Xiang, L.; Xinyi, Z.; Fei, W.; and Ming-Hsuan, Y. 2020. Multi-Scale Boosted Dehazing Network with Dense Feature Fusion. In *Proceedings of IEEE/CVF Conference Computer Vision Pattern Recognition (CVPR)*.

He, K.; Sun, J.; and Tang, X. 2009. Single image haze removal using dark channel prior. In *Proceedings of IEEE/CVF Conference Computer Vision Pattern Recognition (CVPR)*.

Hnewa, M.; and Radha, H. 2021. Multiscale Domain Adaptive YOLO for Cross-Domain Object Detection. arXiv:2106.01483.

Huang, S.-C.; Le, T.-H.; and Jaw, D.-W. 2020. DSNet: Joint semantic learning for object detection in inclement weather conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Liu, X.; Ma, Y.; Shi, Z.; and Chen, J. 2019. GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

McCartney, E. J. 1976. Optics of the atmosphere: scattering by molecules and particles. *New York*.

Narasimhan, S. G.; and Nayar, S. K. 2002. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3): 233–254.

Redmon, J.; and Farhadi, A. 2018. Yolov3: An incremental improvement. arXiv:1804.02767.

Figure 3: Detection results of different methods on synthetic VOC_Dark_test images (top row), real-world ExDark_test low-light images (bottom two rows). From left to right: YOLOv3 II, ZeroDCE + YOLOv3 I and our IA-YOLO. The proposed method learns to make the image brighter with more details, which results in better detection performance with fewer missed and wrong detections.