Supplementary Materials for Paper ID:1346

Xingyu Jiang[1], Hongkun Dou[1], Chengwei Fu[1], Bingquan Dai[1], Tianrun Xu[2], and Yue Deng[1⋆]

[1] School of Astronautics, Beihang University
ydeng@buaa.edu.cn
[2] North China University of Technology

1 Derivation of Equation 6

The derivation details of eq.6 (see manuscript) in the paper are as follows,

$$
\begin{aligned}
&\nabla_{\phi}\left(J_e\left(\mathcal{D}_v;\theta_t\right)+\alpha\Omega(\boldsymbol{w})\right) && \text{(6a)}\\
&\qquad \textit{(apply chain rule :)}\\
&=\nabla_{\theta}J_e\left(\mathcal{D}_v;\theta_t\right)^{\top}\cdot\nabla_{\phi}\theta_t(\phi)+\alpha\nabla_{\boldsymbol{w}}\Omega(\boldsymbol{w})^{\top}\cdot\nabla_{\phi}\boldsymbol{w} && \text{(6b)}\\
&\qquad \textit{(partitioned matrix multiplication by rows :)}\\
&=\nabla_{\theta}J_e\left(\mathcal{D}_v;\theta_t\right)^{\top}\nabla_{\phi}\theta_t(\phi)+\alpha\sum_i\frac{\partial\Omega(\boldsymbol{w})}{\partial w(x_i,y_i;\phi)}\nabla_{\phi}w(x_i,y_i;\phi) && \text{(6c)}\\
&\qquad \textit{(substitute } \theta_t \textit{ from eq.5 :)}\\
&=\nabla_{\theta}J_e\left(\mathcal{D}_v;\theta_t\right)^{\top}\cdot\nabla_{\phi}\left(\theta_{t-1}-\nabla_{\theta}J_w\left(\mathcal{D}_t;\theta_{t-1},\phi\right)\right)+\ \alpha\sum_i\frac{\partial\Omega(\boldsymbol{w})}{\partial w(x_i,y_i;\phi)}\nabla_{\phi}w(x_i,y_i;\phi) && \text{(6d)}\\
&\qquad \textit{(assume } \nabla_{\phi}\theta_{t-1}\approx 0)\\
&\approx-\nabla_{\theta}J_e\left(\mathcal{D}_v;\theta_t\right)^{\top}\cdot\nabla_{\phi}\left(\nabla_{\theta}J_w\left(\mathcal{D}_t;\theta_{t-1},\phi\right)\right)+\ \alpha\sum_i\frac{\partial\Omega(\boldsymbol{w})}{\partial w(x_i,y_i;\phi)}\nabla_{\phi}w(x_i,y_i;\phi) && \text{(6e)}\\
&\qquad \textit{(substitute } J_w(\cdot) \textit{ from eq.3 :)}\\
&=-\nabla_{\theta}J_e\left(\mathcal{D}_v;\theta_t\right)^{\top}\cdot\nabla_{\phi}\left(\sum_i w(x_i,y_i;\phi)\nabla_{\theta}\ell\left(x_i,y_i;\theta_{t-1}\right)\right)+\ \alpha\sum_i\frac{\partial\Omega(\boldsymbol{w})}{\partial w(x_i,y_i;\phi)}\nabla_{\phi}w(x_i,y_i;\phi) && \text{(6f)}\\
&\qquad \textit{(extract similar terms } \nabla_{\phi}w(x_i,y_i;\phi) \textit{ :)}\\
&=-\sum_i\nabla_{\phi}w(x_i,y_i;\phi)\cdot\left[\nabla_{\theta}J_e\left(\mathcal{D}_v;\theta_t\right)^{\top}\nabla_{\theta}\ell\left(x_i,y_i;\theta_{t-1}\right)-\alpha\frac{\partial\Omega(\boldsymbol{w})}{\partial w(x_i,y_i;\phi)}\right] && \text{(6g)}\\
&\qquad \textit{(utilize } \nabla f(x)=f(x)\nabla\log f(x) \textit{ :)}\\
&=-\sum_i w\left(x_i,y_i;\phi\right)\nabla_{\phi}\log w\left(x_i,y_i;\phi\right)\cdot\underbrace{\left[\nabla_{\theta}J_e\left(\mathcal{D}_v;\theta_t\right)^{\top}\nabla_{\theta}\ell\left(x_i,y_i;\theta_{t-1}\right)-\alpha\frac{\partial\Omega(\boldsymbol{w})}{\partial w(x_i,y_i;\phi)}\right]}_{R_i} && \text{(6h)}\\
&\qquad \textit{(substitute } \Omega(\cdot) \textit{ by } \mathcal{H}(\cdot) \textit{ :)}\\
&=-\sum_i w\left(x_i,y_i;\phi\right)\nabla_{\phi}\log w\left(x_i,y_i;\phi\right)\cdot\underbrace{\left[\nabla_{\theta}J_e\left(\mathcal{D}_v;\theta_t\right)^{\top}\nabla_{\theta}\ell\left(x_i,y_i;\theta_{t-1}\right)+\alpha\log w\left(x_i,y_i;\phi\right)+\alpha\right]}_{R_i} && \text{(6i)}
\end{aligned}
$$

In the derivation above, we make Markov assumption for the parameters θ, which indicates that current ϕ_t is only relate to θ_t, having nothing to do with earlier value of θ_{t-1}. Therefore, we assume that $\nabla_{\phi}\theta_{t-1}\approx 0$ in the above derivation process, which simplifies the complexity of optimization. The assumption is also proved to be reasonable and effective in our experiment results.

⋆ Corresponding author

In our task, we choose to maximize entropy $\mathcal{H}(\boldsymbol{w})$ as $\Omega(\boldsymbol{w})$ to prevent over concentration of sample weight. Therefore, our R_i be expressed as,

$$R_i = \nabla_\theta J\left(\mathcal{D}_v; \theta_t\right)^\top \nabla_\theta \ell\left(x_i, y_i; \theta_{t-1}\right) + \alpha \log w\left(x_i, y_i; \phi\right) + \alpha \tag{6j}$$

2 Datasets

We adopt RESIDE dataset[6] for synthetic dehazing training. The RESIDE dataset consists of five parts: Indoor Training Set(ITS), Outdoor Training Set(OTS), Synthetic Objective Testing Set(SOTS), Real World Task-driven Testing Set(RTTS) and Hybrid Subjective Testing Set(HSTS). ITS contains 13990 indoor hazy images for training and OTS contains 72135 outdoor hazy images for training. SOTS contains 500 indoor hazy images and 500 outdoor hazy images for testing. In our work, we introduce the validation set which can not overlap with testing set to optimize the external reweighting loop. Specially considering the time cost and partition ratio of training, validation and testing sets, we hold out 10% of training data for validation and ensure the number of samples in validation is the same as that in testing set. To this end, we select subsets of 5000 hazy images from ITS and OTS as our training set, respectively.

We adopt O-HAZE[2] and NH-HAZE[1] datasets for real-world dehazing training. O-HAZE is the NTIRE2018 challenge dataset for real-world dehazing task, in which hazy images are degraded by real haze, generated by professional haze machines. And O-HAZE is divided into three parts: 35 training images, 5 validation images and 5 test images, which is in line with our requirements for dehazing training. The image sizes of O-HAZE range from 1286×947 to 4056×3412. NH-HAZE is the NTIRE2020 challenge dataset for non-homogeneous dehazing task, which is also divided into three parts: 45 training images, 5 validation images and 5 test images. The image size of NH-HAZE is 1600×1200.

3 Implementation Details

The bi-level dehazing (BILD) framework is composed of two main modules (see Fig.1): the patch reweighting network for external loop and weighed dehazing network for internal loop. The patch reweighting network consists of three simple downsampling + residual blocks for extracting high-level image features at different scales and one fully connected + softmax block for generating different learning weights of different patches. As for weighed dehazing network, we adopt AODNet[5], DehazeNet[3], GridDehazeNet[7], MSBDN[4] and FFANet[8] to substitute its dehazing module and use DDU module[9] to alleviate GPU memory overhead due to the image size in O-HAZE and NH-HAZE.

4 Additional Experimental Results

In this section, we show additional experimental results on both synthetic and real-world hazy images compared against SOTA methods. As shown in Fig.2,

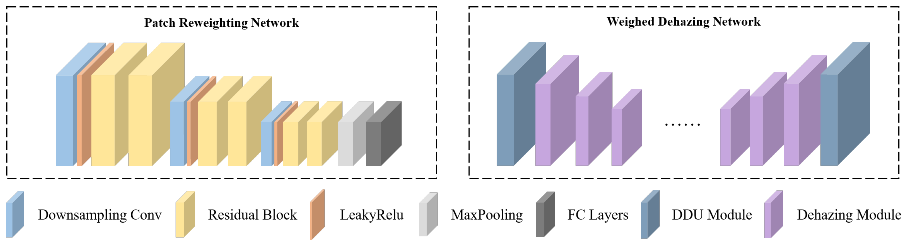

Fig. 1: The architecture of the proposed patch reweighting network for external loop and weighed dehazing network for internal loop. The main body of the patch reweighting network is composed of simple residual blocks, downsampling blocks and fully connected blocks, in which LeakyRelu is adopted as activation function. For weighted dehazing network, it consists of two modules: DDU module[9] for downsampling&upsampling(only used in O-HAZE and NH-HAZE dehazing training) and dehazing module for image dehazing, which can be replaced by general supervised dehazing methods.

our proposed framework (BILD) restores more image details and performs better patch-level haze removal on SOTS dataset[6]. Also, we can observe that our proposed framework performs better color restoration and generates cleaner images on O-HAZE[2] and NH-HAZE[1] datasets compared with SOTA methods in Figs.3 and 4. Besides, we demonstrate more experimental results to show the effectiveness of our proposed framework (BILD) in Figs.5, 6 and 7.

References

1. Ancuti, C.O., Ancuti, C., Timofte, R.: NH-HAZE: an image dehazing benchmark with non-homogeneous hazy and haze-free images. In: CVPRW (2020)
2. Ancuti, C.O., Ancuti, C., Timofte, R., Vleeschouwer, C.D.: O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In: CVPRW (2018)
3. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE TIP **25**(11), 5187–5198 (2016)
4. Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., Yang, M.H.: Multi-scale boosted dehazing network with dense feature fusion. In: CVPR. pp. 2157–2167 (2020)
5. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: ICCV. pp. 4770–4778 (2017)
6. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. IEEE TIP **28**(1), 492–505 (2019)
7. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention-based multi-scale network for image dehazing. In: ICCV. pp. 7314–7323 (2019)
8. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: Ffa-net: Feature fusion attention network for single image dehazing. In: AAAI. pp. 11908–11915 (2020)
9. Zheng, Z., Ren, W., Cao, X., Hu, X., Wang, T., Song, F., Jia, X.: Ultra-high-definition image dehazing via multi-guided bilateral learning. In: CVPR. pp. 16180–16189 (2021)

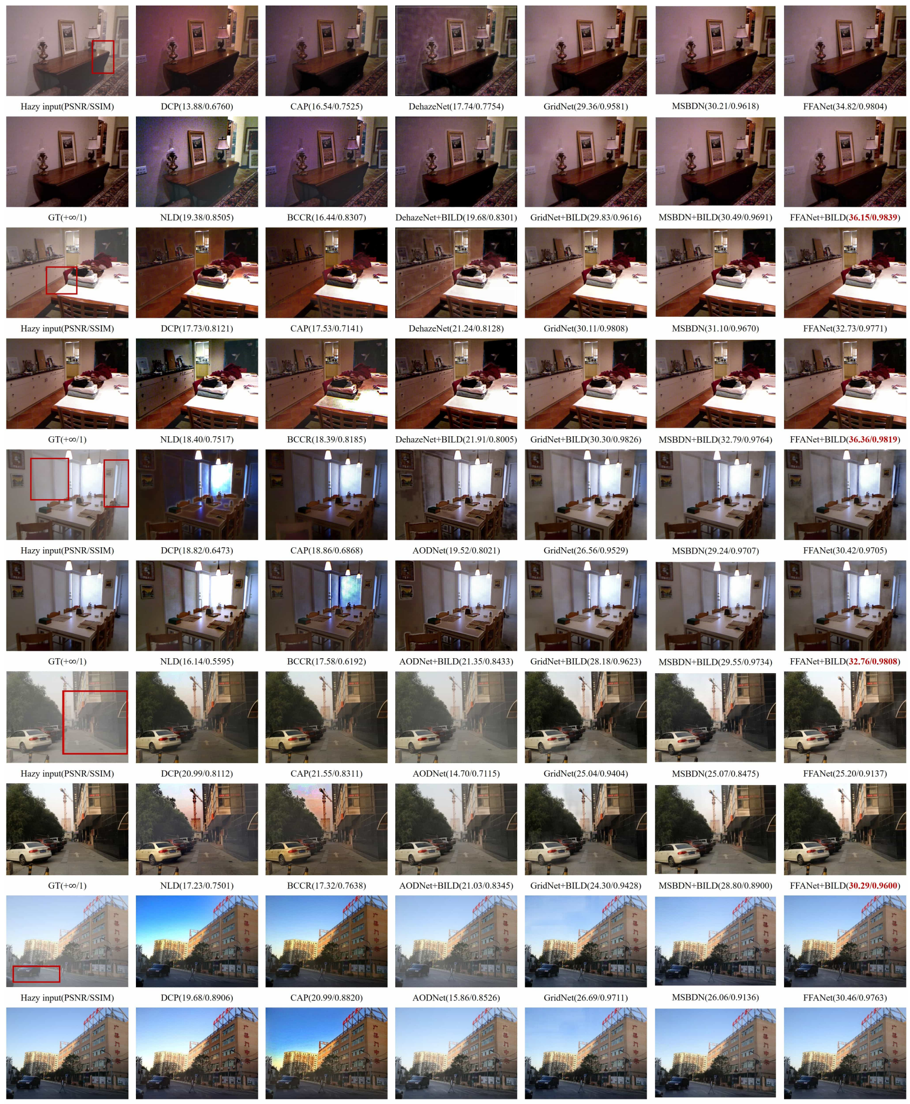

Fig. 2: Visual comparisons on the SOTS dataset. Our framework performs better color restoration and patch-level haze removal against the state-of-the-art methods.

Fig. 3: Visual comparisons on the O-HAZE dataset. Our framework performs better color restoration and patch-level haze removal against the state-of-the-art methods.

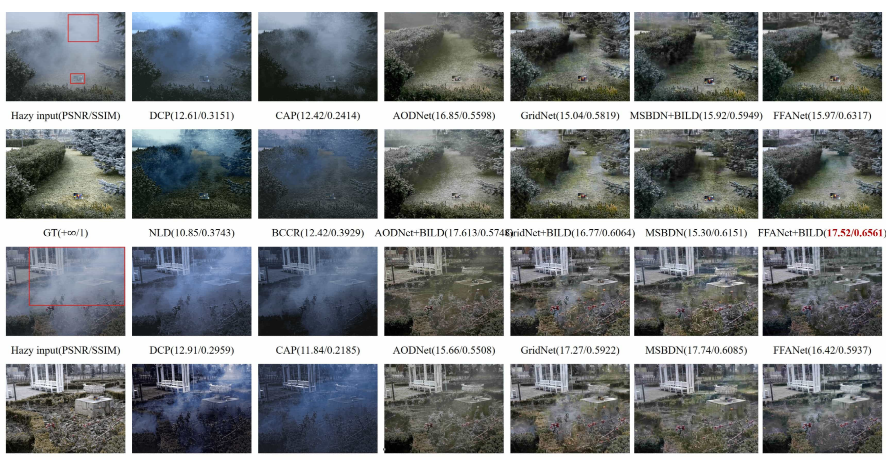

Fig. 4: Visual comparisons on the NH-HAZE dataset. Our framework performs better color restoration and patch-level haze removal against the state-of-the-art methods.

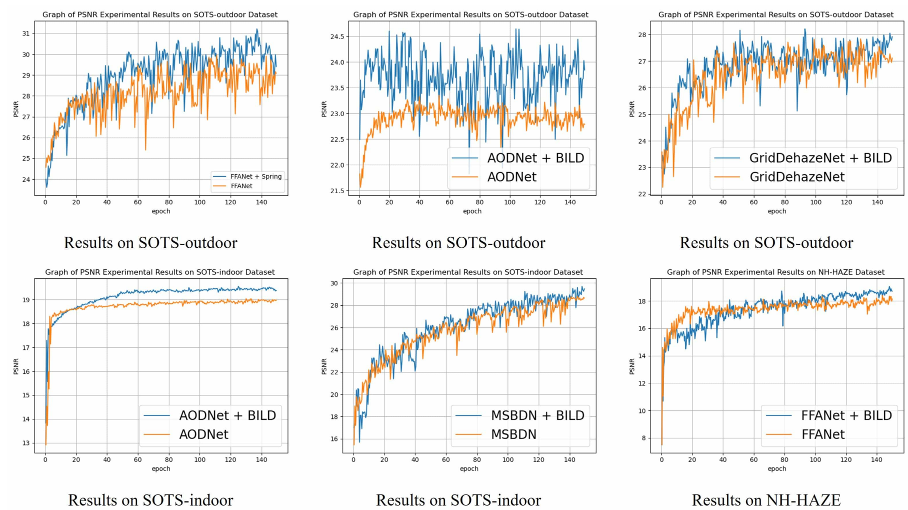

Fig. 5: Graph of PSNR with/without BILD framework during training process.

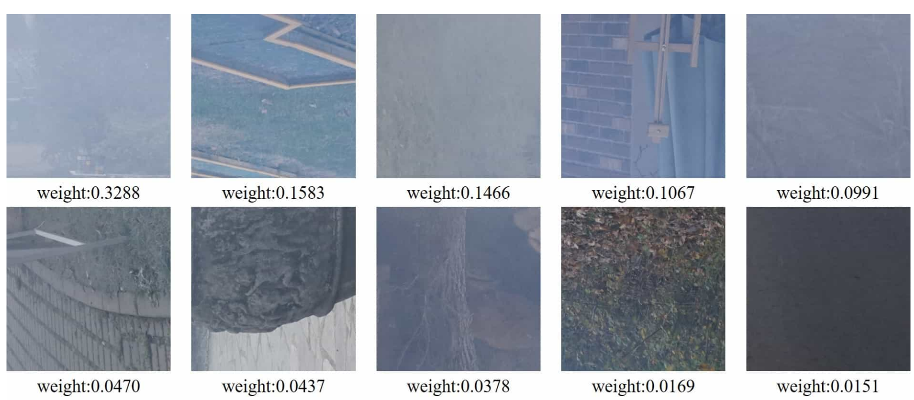

Fig. 6: The experimental results of patch reweighting in one batch. In the manuscript, as shown in Fig.6, we select figure 6(h) as the baseline for different patches, of which the learning weight is 0.1000

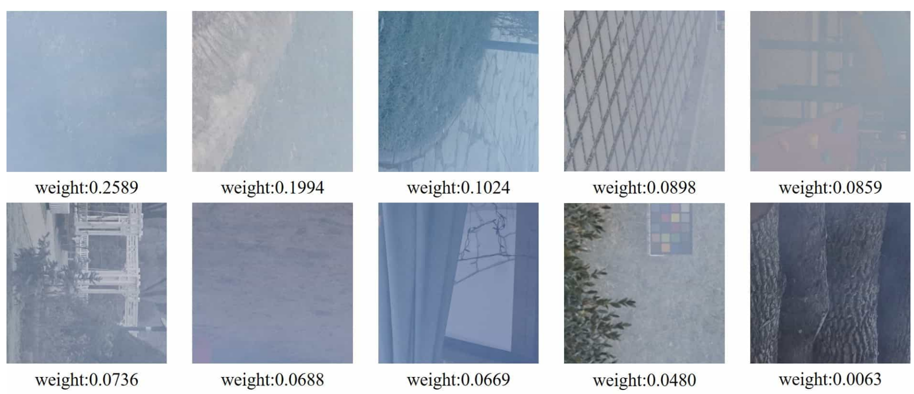

Fig. 7: The experimental results of patch reweighting in one batch.