RVSL: Robust Vehicle Similarity Learning in Real Hazy Scenes Based on Semi-supervised Learning Supplementary Material

Wei-Ting Chen[1*], I-Hsiang Chen[2*], Chih-Yuan Yeh[2], Hao-Hsiang Yang[2], Hua-En Chang[2], Jian-Jiun Ding[2], and Sy-Yen Kuo[2]

[1] Graduate Institute of Electronics Engineering, National Taiwan University, Taiwan
[2] Department of Electrical Engineering, National Taiwan University, Taiwan
{f05943089,f09921058,f09921063}@ntu.edu.tw,
islike8399@gmail.com,{r10921a35,jjding,sykuo}@ntu.edu.tw

https://github.com/Cihsaing/rvsl-robust-vehicle-similarity-learning--ECCV22

1 More Experimental Results and Implementation Details

1.1 Comparison of Using Different Image Restoration Methods in Two-stage strategies

We apply ReID methods (i.e., HRCN [12] and CAL [8]) and combine them with different existing dehazing (i.e., CAP [13], DCP [5], DCPDN [11], DehazeNet [2], EPDN [7], GRID [6], MPR-Net [10], MSBDN [3], and NLI [1]) and image restoration (i.e., URIE [9]) methods for pre-processing. Note that, URIE is a comprehensive image restoration technique for the down-stream application. Followed by its original setting, we combine this method before the ReID models and train the combination in an end-to-end fashion. The evaluation is conducted on the real-world dataset and the results are reported in Table 1. We can see that the two-stage strategy may not always benefit the performance of the ReID model under haze scenarios and sometimes the performance may be even worse than the original model trained on clear images. By contrast, in the proposed method, the domain transformation and ReID models are trained jointly with three-stage optimization. It can outperform other two-stage strategies.

1.2 Evaluation on Existing Benchmarks

To verify the performance of our method on the existing dataset, we adopt Vehicle 1M [4] which contains both clear and hazy scenarios. We apply the full test set which consists of 5527 IDs and 5527 images in the query set and 5527 IDs and 85953 images in the gallery set to compare the performance with other methods. We adopt the training setting (iv) (i.e., the same training images used in our method including synthetic haze, real-world clear, and real-world haze datasets) in this experiment. The results are shown in Table 2. One can see that our method trained without ID labels of real-world data

*Indicates equal contribution.

Table 1: **Comparison of using different dehazing/image restoration methods for pre-processing in the two-stage strategy on the real-world dataset.**

Pre-processing Method	ReID	Metric			
		mAP	CMC@1	CMC@5	CMC@10
-	**HRCN**	71.77	85.30	95.40	97.50
	CAL	75.94	91.70	97.60	98.40
CAP	**HRCN-dehaze**	72.83	85.30	95.30	97.50
	CAL-dehaze	77.59	92.90	97.50	98.60
DCP	**HRCN-dehaze**	72.45	85.10	96.00	98.40
	CAL-dehaze	76.46	91.70	98.00	98.70
DCPDN	**HRCN-dehaze**	66.22	80.70	94.00	96.90
	CAL-dehaze	68.67	87.80	96.50	97.70
DHN	**HRCN-dehaze**	73.07	85.10	96.60	98.40
	CAL-dehaze	77.80	92.40	97.60	98.80
EPDN	**HRCN-dehaze**	71.57	82.00	95.10	97.90
	CAL-dehaze	75.56	92.10	97.60	98.70
GRID	**HRCN-dehaze**	72.78	84.60	95.60	97.80
	CAL-dehaze	76.81	92.30	97.90	98.70
MPR	**HRCN-dehaze**	72.78	84.60	96.10	97.80
	CAL-dehaze	77.49	94.00	98.00	98.80
MSBDN	**HRCN-dehaze**	68.55	79.40	92.90	95.40
	CAL-dehaze	74.12	89.20	95.10	96.60
NLI	**HRCN-dehaze**	68.41	81.00	94.60	97.00
	CAL-dehaze	72.81	89.70	97.60	98.80
URIE	**HRCN-dehaze**	78.13	89.20	96.20	98.10
	CAL-dehaze	80.44	95.00	98.50	99.10
-	**Ours**	**84.12**	**95.60**	**98.60**	**99.30**
-	**Ours-F**	**87.72**	**96.90**	**98.40**	**99.60**

Table 2: **Evaluation results on the Vehicle 1M dataset.**

Method	Metric			
	mAP	CMC@1	CMC@5	CMC@10
VRCF-all	62.20	87.30	94.90	96.80
VOC-all	75.70	81.50	91.90	94.80
DMT-all	87.80	94.10	98.00	98.70
VehicleX-all	72.75	79.55	92.18	95.01
PVEN-all	87.33	92.79	97.43	98.45
HRCN-all	74.27	80.21	91.39	95.06
CAL-all	86.48	91.75	97.40	98.39
Ours	87.27	92.89	97.52	98.55
Ours-F	**89.54**	**94.37**	**98.05**	**98.90**

can achieve competitive performance in Vehicle 1M compared with other methods. Moreover, we can achieve state-of-the-art performance if we adopt the ID labels of real-world data. Thus, the results prove the robustness of our proposed method.

Algorithm 1 Training Stage for Synthetic Dataset

Input:
Input synthetic pair image: K_H^S, $(K_H^S)_{GT}$
Encoder modules for clear and haze: $\mathbf{E_C}(\cdot)$, $\mathbf{E_H}(\cdot)$
Decoder modules for 'Clear to Haze' and 'Haze to Clear': $\mathbf{D_H}(\cdot)$, $\mathbf{D_C}(\cdot)$
Decoder modules for ReID: $\mathbf{D_{ReID}}(\cdot)$
iteration number: n_{max}
Initialization:
Initialize the all module parameters by Kaiming normalization.
while $n < n_{max}$ **do**
  extract feature: $F_H = \mathbf{E_H}(K_H^S)$, $F_C = \mathbf{E_C}((K_H^S)_{GT})$;
  rendered images: $K_{H \leftarrow C} = \mathbf{D_H}(F_C)$, $K_{C \leftarrow H} = \mathbf{D_C}(F_H)$;
  extract ReID features: F_{ReID} by $\mathbf{D_{ReID}}(F_H)$ and $\mathbf{D_{ReID}}(F_C)$, respectively;
  update $\mathbf{E_H}, \mathbf{E_C}, \mathbf{D_H}, \mathbf{D_C}$ using $\mathcal{L}_{DT_s} = \mathcal{L}_{DT_s}^{C \to H} + \mathcal{L}_{DT_s}^{H \to C}$;
  update $\mathbf{E_H}, \mathbf{E_C}, \mathbf{D_{ReID}}$ using $\mathcal{L}_{ReID_s} = \mathcal{L}_{Tri} + \mathcal{L}_{ID}$;
end
Output: Save the all parameters of the trained modules: θ_{stage1}.

1.3 Implementation Detail

Our proposed method combines domain transformation and the ReID technique with the three-stage optimization scheme. To illustrate the training procedure clearly, we present the three stages in detail: the training stage for synthetic, real clear, and real haze data, respectively.

Algorithm 2 Training Stage for Real Clear Dataset

Input:
Input real clear image: K_C^R
Encoder modules for clear and haze: $\mathbf{E_C}(\cdot)$, $\mathbf{E_H}(\cdot)$
Decoder modules for 'Clear to Haze' and 'Haze to Clear': $\mathbf{D_H}(\cdot)$, $\mathbf{D_C}(\cdot)$
Decoder modules for ReID: $\mathbf{D_{ReID}}(\cdot)$
Discriminator modules: $\psi(\cdot, Real/Fake)$
Iteration number: n_{max}
Balancing Weights: $\lambda_{rc_1} = 10^{-3}, \lambda_{rc_2} = 10$
Initialization:
Initialize all module parameters except $\psi()$ by Synthetic stage parameters θ_{stage1}.
while $n < n_{max}$ **do**
  render Hazy forward: $F_C = \mathbf{E_C}(K_C^R)$, $K_H^{R'} = \mathbf{D_H}(F_C)$;
  render Clear forward: $F_H = \mathbf{E_C}(K_H^{R'})$, $K_C^{R''} = \mathbf{D_C}(F_H)$;
  extract ReID features: F_{ReID} by $\mathbf{D_{ReID}}(F_H)$ and $\mathbf{D_{ReID}}(F_C)$, respectively;
  discriminator forward: $\mathcal{L}_{Dis} = \psi(K_H^{R'}, Real)$;
  update $\mathbf{E_C}, \mathbf{D_H}$ using $\mathcal{L}_{CR} + \lambda_{rc_1}\mathcal{L}_{MIDC} + \lambda_{rc_2}\mathcal{L}_{Dis}$;
  update $\mathbf{E_H}, \mathbf{E_C}, \mathbf{D_H}, \mathbf{D_C}$ using $\mathcal{L}_{RC}$;
  update $\mathbf{E_H}, \mathbf{E_C}, \mathbf{D_{ReID}}$ using $\mathcal{L}_{ReID_{rc}} = \mathcal{L}_{EC}$;

  discriminator forward: $\mathcal{L}_{Dis}$ = $\psi(K_C^R, Real)$ + $\psi(K_C^{R''}, Fake)$+$\psi(K_H^{R'}, Fake)$;
  update $\psi(\cdot)$ using $\mathcal{L}_{Dis}$;
end
Output: Save the parameters of the trained model except discriminator modules: θ_{stage2}.

Algorithm 3 Training Stage for Real Haze Dataset

Input:
Input real hazy image: K_H^R
Encoder modules for clear and haze: $\mathbf{E_C}(\cdot)$, $\mathbf{E_H}(\cdot)$
Decoder modules for 'Clear to Haze' and 'Haze to Clear': $\mathbf{D_H}(\cdot)$, $\mathbf{D_C}(\cdot)$
Decoder modules for ReID: $\mathbf{D_{ReID}}(\cdot)$
Discriminator modules: $\psi(\cdot, Real/Fake)$
iteration number: n_{max}
Balancing Weight: $\lambda_{rh_1} = 10$
Initialization:
Initialize all module parameters except $\psi()$ by real clear stage parameter θ_{stage2}.
while $n < n_{max}$ **do**
  render Clear forward: $F_H = \mathbf{E_C}(K_H^R)$, $K_C^{R'} = \mathbf{D_C}(F_H)$;
  render Hazy forward: $F_C = \mathbf{E_C}(K_C^{R'})$, $K_H^{R''} = \mathbf{D_H}(F_C)$;
  extract ReID features: F_{ReID} by $\mathbf{D_{ReID}}(F_H)$ and $\mathbf{D_{ReID}}(F_C)$, respectively;
  discriminator forward: $\mathcal{L}_{Dis} = \psi(K_C^{R'}, Real)$;
  update $\mathbf{E_H}, \mathbf{D_C}$ using $\mathcal{L}_{CR} + \mathcal{L}_{DC} + \mathcal{L}_{TV} + \lambda_{rh_1}\mathcal{L}_{Dis}$;
  update $\mathbf{E_H}, \mathbf{E_C}, \mathbf{D_H}, \mathbf{D_C}$ using $\mathcal{L}_{RC}$;
  update $\mathbf{E_H}, \mathbf{E_C}, \mathbf{D_{ReID}}$ using $\mathcal{L}_{ReID_{rh}} = \mathcal{L}_{EC}$;

  discriminator forward: $\mathcal{L}_{Dis} = \psi(K_H^R, Real) + \psi(K_H^{R''}, Fake) + \psi(K_C^{R'}, Fake)$;
  update $\psi(\cdot)$ using $\mathcal{L}_{Dis}$;
end
Output: Save the final parameters of the trained model except discriminator modules: θ_{stage3}.

References

1. Berman, D., Avidan, S., et al.: Non-local image dehazing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1674–1682 (2016) 1
2. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing **25**(11), 5187–5198 (2016) 1
3. 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: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2157–2167 (2020) 1
4. Guo, H., Zhao, C., Liu, Z., Wang, J., Lu, H.: Learning coarse-to-fine structured feature embedding for vehicle re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018) 1
5. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence **33**(12), 2341–2353 (2010) 1
6. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention-based multi-scale network for image dehazing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7314–7323 (2019) 1
7. Qu, Y., Chen, Y., Huang, J., Xie, Y.: Enhanced pix2pix dehazing network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019) 1
8. Rao, Y., Chen, G., Lu, J., Zhou, J.: Counterfactual attention learning for fine-grained visual categorization and re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1025–1034 (2021) 1
9. Son, T., Kang, J., Kim, N., Cho, S., Kwak, S.: Urie: Universal image enhancement for visual recognition in the wild. In: European Conference on Computer Vision. pp. 749–765. Springer (2020) 1
10. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14821–14831 (2021) 1
11. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3194–3203 (2018) 1
12. Zhao, J., Zhao, Y., Li, J., Yan, K., Tian, Y.: Heterogeneous relational complement for vehicle re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 205–214 (2021) 1
13. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE transactions on image processing **24**(11), 3522–3533 (2015) 1