Visual Recognition-Driven Image Restoration for Multiple Degradation with Intrinsic Semantics Recovery

Zizheng Yang* Jie Huang* Jiahao Chang Man Zhou
Hu Yu Jinghao Zhang Feng Zhao†
University of Science and Technology of China
{yzz6000, hj0117, changjh, manman, yuhu520, jhaozhang}@mail.ustc.edu.cn
fzhao956@ustc.edu.cn

Appendix

A.1. Discussion about the SAD Pre-Training

In Sec.3.3 of the main body, we introduce the prior-ascribing optimization strategy and experimentally prove its effectiveness in Sec.4.5. We also summarize the SAD pre-training scheme and its two possible alternatives. Here, we analyze each of them in detail.

The first one is that we only utilize clean image semantic features to train the SAD, as shown in Fig. A1(a). However, the SAD trained by clean images only is hard to ensure the sensitivity to degraded semantics without having seen them. This will cause a disaster—the SAD may reconstruct an image with high visual quality but semantic less when a degraded semantic feature is input. Unlike [5], invertible network is not applicable in our setting since no task-specific annotations is available. We hope that f_{SAD} is monotonic in the domain $\mathcal{X} = \{cle, deg_1, ..., deg_N\}$, which means the more degradation the input feature semantics suffer, the worse the quality of the image reconstructed by f_{SAD}.

One possible solution is to train f_{SAD} with both clean and degraded images, with the goal to reconstruct the corresponding images from the input semantic representations, as shown in Fig. A1(b). Such training endow f_{SAD} with injective property, which guarantees one-to-one mapping from features to images in domain $\mathcal{X}$. However, this requires a larger number of parameters and a complex structure for f_{SAD}, since it needs to reconstruct images accurately for diverse input features. As a part of VRD-IR, the f_{SAD} should be simple and lightweight.

In fact, we hope the SAD can perceive the different semantic representations, and reconstruct the recognition-friendly images from the semantic features without degradation. The strict one-to-one mapping is not essential to f_{SAD}. Based on that, the similarity ranking loss $\mathcal{L}_{sr}$ is constrained, which encourages the f_{SAD} to map the clean semantic features to high-quality images, while mapping the diverse degraded semantic features to a common manifold space that do not follow the distribution of natural images, as illustrated in Fig. A1(c). From another point of view, we simplify the domain $\mathcal{X}$ to $\mathcal{X}' = \{cle, deg\}$, and encourage f_{SAD} to be monotonic in domain $\mathcal{X}'$. Our training schedule reduces the complexity and parameter amount required for SAD.

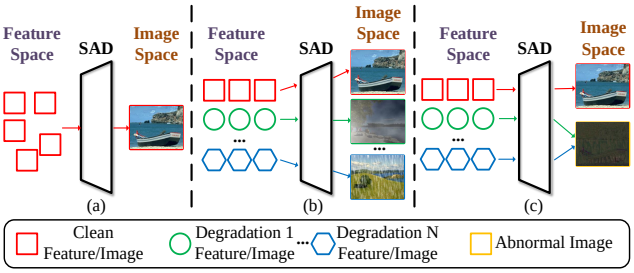

Figure A1. Illustration of different training strategy for SAD. (a) training on clean images only, (b) training on both clean and degraded images with objective to reconstruct both of them accurately, (c) training on both clean and degraded images with our proposed similarity ranking loss.

A.2. More Details about Datasets

A.2.1. Training Datasets

Dehazing Datasets. We utilize the RESIDE [8] as the training datasets for image dehazing. It has an Outdoor Training Set (OTS) consisting of 72,135 outdoor hazy-clean image pairs, and a Synthetic Objective Testing Set (SOTS) consisting of 500 outdoor hazy-clean image pairs.

Denoising Datasets. We use the combination of BSD400 [11] and WED [10] as training set for image de-

*Equal contribution
†Corresponding Author

Table A1. Complete comparison with state-of-the-art IRSD (*i.e.*, image restoration for single degradation) and IRMD (*i.e.*, image restoration for multiple degradation) approaches for image classification on CUB dataset among two different degradation. The best results are marked as **bold** and the second ones are masked by underline.

	Model	Top-1 V (%)	Top-1 R (%)	PSNR (dB)	SSIM
IRSD	DehazeNet [1]	6.67	17.24	14.29	0.5225
	AODNet [7]	24.01	42.06	13.28	0.6415
	EPRN [13]	12.05	25.93	14.39	0.6864
	FDGAN [3]	62.04	74.23	16.76	0.7545
	FFANet [12]	62.32	74.56	16.82	0.7658
	DDP [17]	48.26	63.02	15.32	0.7002
	DL [4]	64.93	75.24	16.23	0.7567
	MPRNet [20]	70.38	78.62	19.23	0.8134
	AirNet [6]	68.67	77.14	17.42	0.7981
	Restormer [19]	**72.47**	**82.85**	**19.68**	**0.8186**
	VRD-IR (Ours)	72.26	80.67	17.72	0.7853
IRMD	DL [4]	65.62	75.95	16.18	0.7316
	MPRNet [20]	69.59	78.15	**18.83**	**0.8000**
	AirNet [6]	68.19	76.81	16.97	0.7692
	VRD-IR (Ours)	**72.11**	**80.55**	17.64	0.7790

(a) dehazing

	Model	Top-1 V (%)	Top-1 R (%)	PSNR (dB)	SSIM
IRSD	CBM3D [1]	20.16	25.41	22.67	0.5237
	DnCNN [7]	24.48	38.92	23.01	0.5474
	IRCNN [13]	26.98	43.37	25.52	0.7121
	FFDNet [3]	22.69	37.51	25.16	0.6982
	BRDNet [12]	25.05	41.96	26.28	0.7552
	DL [4]	24.84	40.48	26.47	0.7654
	MPRNet [20]	25.23	42.35	27.05	0.7894
	AirNet [6]	27.85	45.98	26.57	0.7696
	Restormer [19]	29.18	47.54	**27.46**	**0.7972**
	VRD-IR (Ours)	**31.95**	**49.97**	26.35	0.7648
IRMD	DL [4]	25.12	42.07	26.33	0.7635
	MPRNet [20]	25.18	42.15	**26.74**	**0.7764**
	AirNet [6]	27.77	45.72	26.42	0.7653
	VRD-IR (Ours)	**32.00**	**50.26**	26.41	0.7669

(b) denoising

noising. BSD400 contains 400 clean natural images and WED includes 4,744 natural images.

Deraining Datasets. Rain100L [18] is adopted as the training set for image deraining, which consists of 200 rainy-clean training pairs.

A.2.2. Testing Datasets

Dataset for Classification. We choose CUB [16] as our test dataset for classification evaluation, which has 11,788 images of 200 bird species. It consists of 5,994 images for training and 5,794 images for testing. Note that the recognition models (*i.e.*, VGG16, ResNet50 in Sec.4.2 of the main body) are first pre-trained on clean CUB training set, and then utilized to evaluate the test images restored by different restoraion methods from the degraded test set.

Dataset for Detection. To verify the effectiveness of VRD-IR on detection, we use CrowdHuman [14] as the evaluation dataset. It consists of 15,000, 4,370, and 5,000 images for training, validation, and testing, respectively.

Dataset for Person ReID. We utilize Market1501 [25] for ReID evaluation. It contains 32,668 person images of 1,501 identities captured by 6 cameras. The training set consists of 12,936 images of 751 identities, the query set consists of 3,368 images, and the gallery set consists of 19,732 images of 750 identities.

Table A2. Complete performance comparisons with state-of-the-art IRSD and IRMD approaches for object detection on CrowdHuman dataset among two different degradation. ↑ means higher the better. The best results are marked as **bold** and the second ones are masked by underline.

	Model	AP ↑	JI ↑	MR ↓	PSNR (dB)
IRSD	DehazeNet [1]	46.77	41.39	85.53	13.13
	AODNet [7]	61.08	53.18	75.79	12.13
	EPRN [13]	55.45	47.69	81.04	13.26
	FDGAN [3]	74.75	63.43	65.32	17.27
	FFANet [12]	74.71	63.11	65.56	17.35
	DL [4]	76.85	65.13	64.92	18.44
	MPRNet [20]	78.49	66.20	63.85	19.16
	AirNet [6]	78.21	65.73	64.25	19.07
	Restormer [19]	79.11	67.20	63.41	**19.53**
	VRD-IR (Ours)	**79.20**	**67.24**	**63.35**	18.19
IRMD	DL [4]	76.65	64.82	65.08	18.31
	MPRNet [20]	78.64	66.52	63.60	**19.27**
	AirNet [6]	78.27	65.87	64.03	19.00
	VRD-IR (Ours)	**79.33**	**67.58**	**63.21**	18.25

(a) dehazing

	Model	AP ↑	JI ↑	MR ↓	PSNR (dB)
IRSD	CBM3D [1]	48.51	41.93	83.69	20.15
	DnCNN [7]	56.48	48.89	78.62	22.59
	IRCNN [13]	59.08	50.76	78.12	24.10
	FFDNet [3]	57.88	50.63	78.27	23.94
	BRDNet [12]	58.05	50.98	78.20	24.52
	DL [4]	58.36	51.17	77.48	24.72
	MPRNet [20]	58.77	51.49	77.15	24.86
	AirNet [6]	59.14	51.78	77.16	24.69
	Restormer [19]	59.52	51.96	76.43	**25.09**
	VRD-IR (Ours)	**59.60**	**52.33**	**76.14**	24.60
IRMD	DL [4]	58.21	51.06	77.57	24.61
	MPRNet [20]	58.98	51.61	76.89	**24.98**
	AirNet [6]	59.36	51.95	77.08	24.66
	VRD-IR (Ours)	**59.80**	**52.57**	**75.89**	24.69

(b) denoising

A.3. Additional Results

A.3.1. More Comparisons on Image Classification

In Sec.4.2 of the main body, we demonstrate that our VRD-IR can benefit the image classification and show results in Tab.1, Fig.6, and Fig.7. Here, we extend the results in Tab.1 and show the complete results of the comparison between VRD-IR and state-of-the-art methods in Tab A1 on dehazing and denoising. We also re-train IRMD methods with two settings, *i.e.*, one-by-one and all-in-one [6]. As we can see, the VRD-IR is superior to all the compared methods in most cases. On IRSD dehazing, Restormer [19] achieves better performance. However, Restormer adopts transformer-based network. Note that although DL can handle multiple degradation simultaneously, it requires the corruption types and levels.

A.3.2. More Comparisons on Object Detection

In Sec.4.3 of the main body, we have shown some comparison results about object detection. Here, we extend the results in Tab.2 and show the complete results in Tab. A2. As we can see, the VRD-IR outperforms all compared baseline in detection. Note that the evaluation protocols for visual recognition (*e.g.*, Top-1, AP, JI) are more important than those for visual quality (*e.g.*, PSNR) in our setting. The experiment results further demonstrate that higher visual quality does not mean higher recognition quality. The reason for this phenomenon is that signal fidelity metrics (*e.g.*, PSNR, and SSIM) have dominated in the research of image restoration method, and the optimization for signal fidelity cannot be optimal for semantic quality due to the trade-off between them [9].

A.3.3. More Comparison on Person ReID

In Sec.4.4 of the main body, we evaluate the effectiveness of the VRD-IR in person ReID and show some comparison results in Tab.3. Here, we show the complete comparison results in Tab A3.

Table A3. Complete performance (%) comparisons of different methods for person ReID on Market1501 dataset in dehazing and denoising.

Category	Method	mAP (%)
Dehazing	DehazeNet [1]	36.85
	AODNet [7]	40.74
	EPRN [13]	50.56
	FDGAN [3]	72.98
	FFANet [12]	73.74
	MPRNet [20]	75.12
	AirNet [6]	74.21
	VRD-IR (Ours)	**75.83**
Denoising	CBM3D [2]	20.41
	DnCNN [21]	23.45
	IRCNN [22]	54.28
	FFDNet [23]	53.33
	BRDNet [15]	52.89
	MPRNet [20]	54.56
	AirNet [6]	54.45
	VRD-IR (Ours)	**55.64**

A.4. Generalization Ability

We further conduct experiments to evaluate the generalization ability of the VRD-IR. Note that we train different methods with hazy, noisy, and rainy images. Then, we test them on low contrast images, which is an unseen corruption. Tab A4 shows that our proposed VRD-IR can achieve better generalization ability to unseen corruptions compared with other IRMD methods. Fig A2 describes the qualitative results of different methods along with their feature maps.

Table A4. Performance (%) comparison of different IRMD methods on CUB VGG16 classification when facing the unseen corruption low contrast.

Method	Top-1 V (%)
MPRNet [20]	36.92
AirNet [6]	35.57
VRD-IR (Ours)	**37.40**

A.5. Experiments about SAD Pre-Training

In Sec. A.1, we discuss the different SAD pre-training schemes in details. Here, we show the SAD architecture and performance of different training schemes. The structure of SAD is illustrated in Fig A3. As we can see, it only consists of a couple of RCAB [24] blocks, which is lightweight and simple. We compare the three SAD pre-training schemes in image classification on dehazing, as shown in Tab. A5. Among them, ***C-PT*** means pre-training SAD with clean images only, as shown in Fig. A1(a). ***CD-PT*** means pre-training SAD using both clean and various degraded images with the objective to reconstruct all of them accurately, as shown in Fig. A1(b). ***PA-PT*** denotes pre-training SAD using our proposed prior-ascribing optimization strategy with the similarity ranking loss, as shown in Fig. A1(c).

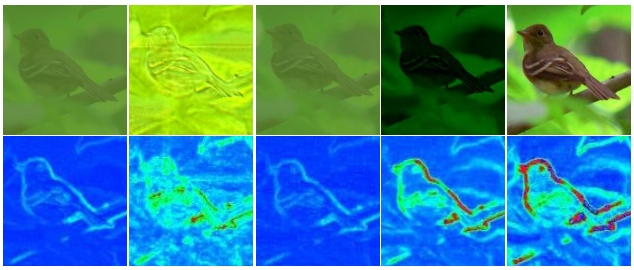

Figure A2. Qualitative results of different IRMD methods on unseen corruption: low contrast.

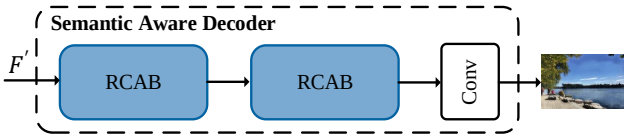

Figure A3. Architecture of the SAD.

As we can see, "C-PT" suffer a significant performance drop compared with "PA-PT" when using the same structure of SAD. On the other hand, "CD-PT" achieve comparable performance with "PA-PT", but "CD-PT" requires a larger number of parameters for SAD.

Table A5. Performance comparison of different SAD pre-training schemes on VGG16 classification in dehazing. "#Params." means the number of parameters of SAD.

Scheme	#Params.	Top-1 V (%)
C-PT	0.90M	64.36
CD-PT	1.43M	72.18
PA-PT	0.90M	72.11

A.6. Broader Impacts

As for the positive impact, the visual recognition-driven image restoration technology has broad impacts and practical values in real world scenarios, *e.g.*, autonomous vehicles, since image degradation is a common phenomenon in imaging systems. Our visual recognition-driven image restoration for multiple degradation can restore diverse degraded images from the perspective of machine vision, which can benefit a lot of downstream high-level tasks. Since the recognition-friendly restoration has long been

overlooked, we hope more methods can be motivated.

As for the negative impact, the image restoration may cause an invasion of privacy. In some case, the identity of certain persons will be masked by image degradation for privacy protection, while the image restoration methods will restore the degradation and reveal the certain identity. Thus, the usage of image restoration should be regularized.

References

[1] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. 2, 3, 4

[2] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. In *2007 IEEE International Conference on Image Processing*, volume 1, pages I–313. IEEE, 2007. 4

[3] Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10729–10736, 2020. 2, 3, 4

[4] Qingnan Fan, Dongdong Chen, Lu Yuan, Gang Hua, Nenghai Yu, and Baoquan Chen. A general decoupled learning framework for parameterized image operators. *IEEE transactions on pattern analysis and machine intelligence*, 43(1):33–47, 2019. 2, 3

[5] Insoo Kim, Seungju Han, Ji-won Baek, Seong-Jin Park, Jae-Joon Han, and Jinwoo Shin. Quality-agnostic image recognition via invertible decoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12257–12266, 2021. 1

[6] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17452–17462, 2022. 2, 3, 4

[7] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision*, pages 4770–4778, 2017. 2, 3, 4

[8] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. 1

[9] Dong Liu, Haochen Zhang, and Zhiwei Xiong. On the classification-distortion-perception tradeoff. *Advances in Neural Information Processing Systems*, 32, 2019. 3

[10] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, 2016. 1

[11] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001. 1

[12] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11908–11915, 2020. 2, 3, 4

[13] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8160–8168, 2019. 2, 3, 4

[14] Shuai Shao, Zijian Zhao, Boxun Li, Tete Xiao, Gang Yu, Xiangyu Zhang, and Jian Sun. Crowdhuman: A benchmark for detecting human in a crowd. *arXiv preprint arXiv:1805.00123*, 2018. 2

[15] Chunwei Tian, Yong Xu, and Wangmeng Zuo. Image denoising using deep cnn with batch renormalization. *Neural Networks*, 121:461–473, 2020. 4

[16] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 2

[17] Yang Wang, Yang Cao, Zheng-Jun Zha, Jing Zhang, and Zhiwei Xiong. Deep degradation prior for low-quality image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11049–11058, 2020. 2

[18] Wenhan Yang, Robby T Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE transactions on pattern analysis and machine intelligence*, 42(6):1377–1393, 2019. 2

[19] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022. 2, 3

[20] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021. 2, 3, 4

[21] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. 4

[22] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3929–3938, 2017. 4

[23] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018. 4

[24] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very

deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 4

[25] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, pages 1116–1124, 2015. 2