DRCNet: Dynamic Image Restoration Contrastive Network

Fei Li[1,3,4,*], Lingfeng Shen[2,*], Yang Mi[1,4], and Zhenbo Li[1,3,4,5,6]✉

[1] College of Info. and Electr. Eng., China Agricultural University, China
{leefly072,miy,lizb}@cau.edu.cn
[2] Department of Computer Science, Johns Hopkins University, USA
lshen30@jh.edu
[3] National Innovation Center for Digital Fishery, Ministry of Agriculture and Rural Affairs, China
[4] Key Lab. of Agricultural Info. Acquisition Tech., Ministry of Agriculture and Rural Affairs, China
[5] Key Lab. of Smart Breeding Tech., Ministry of Agriculture and Rural Affairs, China
[6] Precision Agriculture Techn. Integrated Scientific Experiment Base, Ministry of Agriculture and Rural Affairs, China

This supplementary material is organized as follows: Section 1 contains comprehensive proof of **Theorem 1**. Section 2, 3, and 4 describe the results under different thresholds t, the coefficient hyper-parameter δ, and computation complexity. Section 6 and 7 contain additional visualization results.

1 Proof of Theorem 1

Firstly, we provide a standard derivation of $\mathcal{R}_{\text{sen}}$ based on asymptotic attributes in statistics. Recall the definitions of $\ell(s,\theta)$ and $\mathcal{R}_{\text{sen}}(s^*)$, we give the assumption for $\ell_1(s,\theta)$.

Assumption 1 *$\ell_1(s,\theta)$ and $\ell_c(s^*,\theta)$ are twice differentiable and strongly convex in θ.*

Based on Assumption 1, we give the lemma below:

Lemma 1. $\mathcal{R}_{\text{sen}}(s^*) = \left|H_\theta^{-1} \sum_{s\in S} \nabla_\theta \ell_c(s^*,\theta)\right|$

The empirical risk $R(\theta)$ is formulated as follows:

$$R(\theta) = \sum_{s\in S} \left[\ell_1(s,\theta) + \alpha \cdot \ell_c(s^*,\theta)\right] \tag{1}$$

Then we make derivation on $R(\theta)$, the Hessian matrix of $R(\theta)$ can be obtained based on Assumption 1:

$$H_\theta = \nabla^2 R(\theta) = \sum_{s\in S} \left[\nabla_\theta^2 \ell_1(s,\theta) + \alpha \cdot \nabla_\theta^2 \ell_c(s^*,\theta)\right] \tag{2}$$

* Equal contribution. ✉ Corresponding author.

Assumption 1 guarantees that H_θ exists and is positive definite and guarantees the existence of H_θ^{-1}, which will be used in the following derivation.

Recall the definition of the minimizer θ_{α,s^*} of $\mathcal{R}_{\text{sen}}$:

$$\theta_{\alpha,s^*} = \text{argmin}_{\theta\in\Theta} \sum_{s\in S} [\ell_1(s,\theta) + \alpha \cdot \ell_c\left(s^*,\theta\right)] \tag{3}$$

Define the parameter change $\Delta_{\theta_{\alpha,s^*}} = \theta_{\alpha,s^*} - \theta_{0,s^*}$ when α is close to 0, then the change of θ towards α can be defined as follows:

$$\frac{d\theta_{\alpha,s^*}}{d\alpha} = \frac{\Delta_{\theta_{\alpha,s^*}}}{d\alpha} \tag{4}$$

Besides, let $L_1(s,\theta)$ and $L_c(s^*,\theta)$ be as follows for brevity:

$$L_1(s,\theta) = \sum_{s\in S} \ell_1(s,\theta)$$

$$L_c(s^*,\theta) = \sum_{s\in S} \ell_c\left(s^*,\theta\right)$$

Considering that θ_{α,s^*} is a minimizer of $R(\theta_{\alpha,s^*})$, then we can obtain the following property:

$$\nabla_{\theta_{\alpha,s^*}} L_1(s,\theta_{\alpha,s^*}) + \alpha \cdot \nabla_{\theta_{\alpha,s^*}} L_c\left(s^*,\theta_{\alpha,s^*}\right) \approx 0 \tag{5}$$

Assumption 2 $\theta_{\alpha,s^*} \to \theta_{0,s^*}$ *when* $\alpha \to 0$

Then, we have a further derivation of Eq.(5) based on Assumption 2 and Taylor expansion, which is listed as follows:

$$\begin{aligned}&\nabla_{\theta_{\alpha,s^*}} L_1(s,\theta_{\alpha,s^*}) + \alpha \cdot \nabla_{\theta_{\alpha,s^*}} L_c\left(s^*,\theta_{\alpha,s^*}\right) + \\ &[\nabla^2_{\theta_{\alpha,s^*}} L_1(s,\theta_{\alpha,s^*}) + \alpha \cdot \nabla^2_{\theta_{\alpha,s^*}} L_c\left(s^*,\theta_{\alpha,s^*}\right)]\Delta_{\theta_{\alpha,s^*}} \approx 0\end{aligned}$$

where we have dropped high-order terms. Naturally, we can get the formulation of $\Delta_{\theta_{\alpha,s^*}}$:

$$\begin{aligned}\Delta_{\theta_{\alpha,s^*}} \approx &-[\nabla^2_{\theta_{\alpha,s^*}} L_1(s,\theta_{\alpha,s^*}) + \alpha \cdot \nabla^2_{\theta_{\alpha,s^*}} L_c\left(s^*,\theta_{\alpha,s^*}\right)]^{-1} \\ &[\nabla_{\theta_{\alpha,s^*}} L_1(s,\theta_{\alpha,s^*}) + \alpha \cdot \nabla_{\theta_{\alpha,s^*}} L_c\left(s^*,\theta_{\alpha,s^*}\right)]\end{aligned}$$

Then we give the assumption for θ_{α,s^*} as follows:

Assumption 3 $\nabla_{\theta_{\alpha,s^*}} L_1(s,\theta_{\alpha,s^*}) \approx 0$

Such an assumption is natural since empirical risk minimization (ERM) process still optimizes $L(\cdot)$ well when the weight α of regularizer is small [5].

Then, based on Assumption 3, we have $\nabla L(s,\theta_{\alpha,s^*}) \approx 0$. After abandoning the $o(\alpha)$ terms, we have

$$\Delta_{\theta_{\alpha,s^*}} \approx -\alpha[\nabla^2_{\theta_{\alpha,s^*}} L(s,\theta_{\alpha,s^*})]^{-1} \nabla_{\theta_{\alpha,s^*}} L_c\left(s^*,\theta_{\alpha,s^*}\right)$$

Combined with Eq.(2) and Eq.(4), we have

$$\frac{d\theta_{\alpha,s^*}}{d\alpha} \mid_{\alpha\to 0} = -H^{-1}_{\theta_{\alpha,s^*}} \nabla L_c(s^*, \theta_{\alpha,s^*}) \tag{6}$$

Since $\mathcal{R}_{sen}(s^*) = \left|\lim_{\alpha\to 0} \frac{d\theta_{\alpha,s^*}}{d\alpha}\right|$, we have that:

$$\mathcal{R}_{\text{sen}}(s^*) = \left|H_\theta^{-1}\nabla_\theta L_c(s^*, \theta)\right| \tag{7}$$

which completes the proof of Lemma 1.

Then we provide deeper derivation and present the Lemma 2. Note that we replace ∇_θ with ∇ for brevity, and the omitted θ is aligned to its following θ which is the same as the Proof for Lemma 1. For example, $\nabla L_c(s_1^*, \theta_{\alpha,s_1^*})$ means $\nabla_{\theta_{\alpha,s_1^*}} L_c(s_1^*, \theta_{\alpha,s_1^*})$.

Lemma 2. *Let s^*_{intra} (s_1^*) and s^*_{extra} (s_2^*) be the negative pairs constructed by Intra-CR and Extra-CR, we have*

$$|H^{-1}_{\theta_{\alpha,s_1^*}} \nabla L_c(s_1^*, \theta_{\alpha,s_1^*})| < |H^{-1}_{\theta_{\alpha,s_2^*}} \nabla L_c(s_2^*, \theta_{\alpha,s_2^*})|$$

Assumption 4 $H_{\theta_{\alpha,s_1^*}} \approx H_{\theta_{\alpha,s_2^*}}$

Assumption 5 $\theta_{\alpha,s_1^*} \approx \theta_{\alpha,s_2^*} = \theta_\alpha$

Since the objective can be decomposed into the dot product of two vectors:

$$\begin{aligned} &|H^{-1}_{\theta_{\alpha,s_1^*}} \nabla L_c(s_1^*, \theta_{\alpha,s_1^*})| \\ &= |H^{-1}_{\theta_{\alpha,s_1^*}}| \cdot |\nabla L_c(s_1^*, \theta_{\alpha,s_1^*})| \cdot |cos(\gamma)| \end{aligned} \tag{8}$$

where $H^{-1}_{\theta_{\alpha,s_1^*}} \in R^{\Theta\times 1}$, $\nabla L_c(s_1^*, \theta_{\alpha,s_1^*}) \in R^{1\times\Theta}$, and Θ is the total number of model's parameters. γ is the angle between such two vectors, which we assume the same.

Then, based on Assumption 4 and Assumption 5, we only need to investigate the relation between $|\nabla L_c(s_1^*, \theta_{\alpha,s_1^*})|$ and $|\nabla L_c(s_2^*, \theta_{\alpha,s_2^*})|$.

Then we directly compare $|\nabla L_c(s_1^*, \theta_{\alpha,s_1^*})|$ and $|\nabla L_c(s_2^*, \theta_{\alpha,s_2^*})|$. Firstly, we rewrite the formulation of $\nabla L_c(s_1^*, \theta_{\alpha,s_1^*})$ as follows:

$$\nabla L_c(s_1^*, \theta_\alpha) = \frac{L_c(s_1^*, \theta_\alpha + \epsilon) - L_c(s_1^*, \theta_\alpha)}{d\epsilon} \tag{9}$$

Similarily, we also have:

$$\nabla L_c(s_2^*, \theta_\alpha) = \frac{L_c(s_2^*, \theta_\alpha + \epsilon) - L_c(s_2^*, \theta_\alpha)}{d\epsilon} \tag{10}$$

where $\epsilon \to 0$ is a perturbation. Since θ_α is the minimizer for $L_c(s_1^*, \theta_\alpha)$ and $L_c(s_2^*, \theta_\alpha)$, respectively, which indicates that any small perturbation ϵ on θ will increase the empirical risk. Therefore, $L_c(s_2^*, \theta_\alpha + \epsilon) - L_c(s_2^*, \theta_\alpha) > 0$ for any small perturbation ϵ.

Based on this, we present our subsequent derivation for $\nabla L_c(s_1^*, \theta_\alpha)$ and $\nabla L_c(s_2^*, \theta_\alpha)$. First of all, $\nabla L_c(s_2^*, \theta_\alpha)$ and $\nabla L_c(s_2^*, \theta_\alpha)$ are positive due to the property of θ_{α,s_1^*} and θ_{α,s_2^*}. Therefore, we have:

$$\begin{aligned}
&|\nabla L_c(s_2^*, \theta_\alpha)| - |\nabla L_c(s_1^*, \theta_\alpha)| \\
&= |\frac{L_c(s_2^*, \theta_\alpha + \epsilon) - L_c(s_2^*, \theta_\alpha)}{d\epsilon}| \\
&- |\frac{L_c(s_1^*, \theta_\alpha + \epsilon) - L_c(s_1^*, \theta_\alpha)}{d\epsilon}| \\
&= (\frac{L_c(s_2^*, \theta_\alpha) - L_c(s_1^*, \theta_\alpha)}{d\epsilon}) \\
&- (\frac{L_c(s_2^*, \theta_\alpha + \epsilon) - L_c(s_1^*, \theta_\alpha + \epsilon)}{d\epsilon})
\end{aligned} \tag{11}$$

Recall the Assumption 1 and the fact that θ is the local minimizer of empirical risk of L_c, such conditions indicate that the norm of gradient on θ is smaller than the one on $\theta + \epsilon$. Based on the theoretical justification of empirical risk [10] and convexity of loss function [2], the change of input s^* causes higher risk change at $\theta + \epsilon$ than θ, which means:

$$\begin{aligned}
&|\frac{L_c(s_1^*, \theta_\alpha) - L_c(s_2^*, \theta_\alpha)}{d\epsilon}| \\
&< |\frac{L_c(s_1^*, \theta_\alpha + \epsilon) - L_c(s_2^*, \theta_\alpha + \epsilon)}{d\epsilon}|
\end{aligned} \tag{12}$$

Then recall the definition of ℓ_c and construction of s_1^* and s_2^*:

$$\ell_c(s^*) = \frac{\ell_1(G(X), G(J))}{\ell_1(G(X), G(J^*))}$$

Since s_1^* and s_2^* are intra-class and extra-class image pairs, the distance between $G(X), G(J_{intra}^*)$ is lower than $G(X), G(J_{extra}^*)$, thus $\ell_c(s_{intra}^*) > \ell_c(s_{extra}^*)$. Therefore, we have follows:

$$L_c(s_2^*, \theta_\alpha) - L_c(s_1^*, \theta_\alpha) < 0 \tag{13}$$

Then we give further derivation on Eq.(11):

$$\begin{aligned}
&(\frac{L_c(s_2^*, \theta_\alpha) - L_c(s_1^*, \theta_\alpha)}{d\epsilon}) \\
&- (\frac{L_c(s_2^*, \theta_\alpha + \epsilon) - L_c(s_1^*, \theta_\alpha + \epsilon)}{d\epsilon}) \\
&= -|\frac{L_c(s_2^*, \theta_\alpha) - L_c(s_1^*, \theta_\alpha)}{d\epsilon}| \\
&+ |\frac{L_c(s_2^*, \theta_\alpha + \epsilon) - L_c(s_1^*, \theta_\alpha + \epsilon)}{d\epsilon}| > 0
\end{aligned} \tag{14}$$

The equation in Eq.(14) holds due to Eq.(13).

The result completes the Lemma 2. Combining Lemma 1 and Lemma 2, we complete the proof of the theorem.

Table 1. Comparison of the different parameter values of threshold t in three main restoration datasets in terms of PSNR (dB).

	t	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
	Derain [7]	26.48	26.73	26.82	27.10	27.22	**27.39**	27.31	27.23	27.21	27.13
PSNR	SIDD [1]	36.92	37.09	37.11	37.37	37.72	**37.76**	37.74	37.61	36.99	36.92
	GoPro [3]	26.05	29.12	29.17	29.71	30.12	**30.21**	30.19	28.17	28.15	27.10

Table 2. Comparison of the different parameter values of threshold δ in three main restoration dataset in terms of PSNR (dB).

	δ	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
	Derain [7]	25.08	25.13	25.42	25.56	25.57	**25.59**	25.54
PSNR	SIDD[1]	36.97	37.00	37.01	37.07	37.09	**37.12**	37.04
	GoPro [3]	31.51	31.58	31.68	31.69	31.69	**31.71**	31.60

2 Threshold of t

Threshold t is set in the spatial filter branch to detect the degraded pixels with a soft distinction. It could affect the masking progress of the degraded pixels and further affect the quality of restoration results. We conduct a set of experiments to explore the most suitable value of three t restoration tasks. The PSNR results on Derain [7], SIDD [1] and GoPro [3] show that the proposed method only using the spatial filter-branch achieves the best results when threshold t equals to 0.75, as shown in Table 1. A range of t varies from 0.5 to 0.95 with an interval of 0.05, and we test the model three times for each value of t, and select the mean as the result. For each time, the training progress takes less than 400,000 epochs.

3 Analysis of δ

In order to explore the best value of the coefficient hyper-parameter δ in the energy-based attention branch selected as it provides a good trade-off between the re-calibration feature effect with texture detail and PSNR. The value of δ varies from $10^{-1} \sim 10^{-7}$. The performance on the Derain [7], SIDD [1] and GoPro [3] show that the energy-based attention branch achieves the best results when δ equals to 10^{-6}, which provides good performance of three datasets, as shown in Table 2.

4 Computation Complexity

We show the comparisons in Table 3,4, and 5 on the three tasks test sets. Flops are obtained on a single NVIDIA V100 GPU with 256×256 image.

Table 3. Comparison of the Flops(G) on the derain test.

Methods	PReNet	MSPFN	SPAIR	HINet	MPRNet	DRCNet(Ours)
FLOPs (G)	66.44	620	-	171	545	124.2

Table 4. Comparison of the Flops(G) on the SIDD [1] dataset.

Methods	DnCNN	CBDNet	RIDNet	AINDNet	VDN	SADNet	DANet+	CycleISP	InvDN	MPRNet	DRCNet(Ours)
FLOPs (G)	24.38	40.33	196.5	187.9	99.0	43.3	30.6	219.2	47.8	1176	124.2

Table 5. Comparison of the Flops(G) on the GoPro [3]test.

Methods	Nah et al.	DeblurGAN-v2	SRN	MT-RNN	DMPHN	MPRNet	MIMO-UNet	HINet	DRCNet(Ours)
FLOPs (G)	336	43	173	164	235	760	154	341	124.2

5 More visual results of Energy-based attention module

We added the some visual results of each task in Figure 1, which indicated with (w/.) EA can retain more details of edge info than without (w/o) it.

6 Image Deraining Results

Fig. 2, and 3 show deraining results of our DRCNet and those of the state-of-the-art methods on several challenging images from different datasets [6,8,9].

7 Image Deblurring Results

For the case of deblurring datasets, the visual results are shown in Fig. 4 on the RealBlur-J [4] datasets.

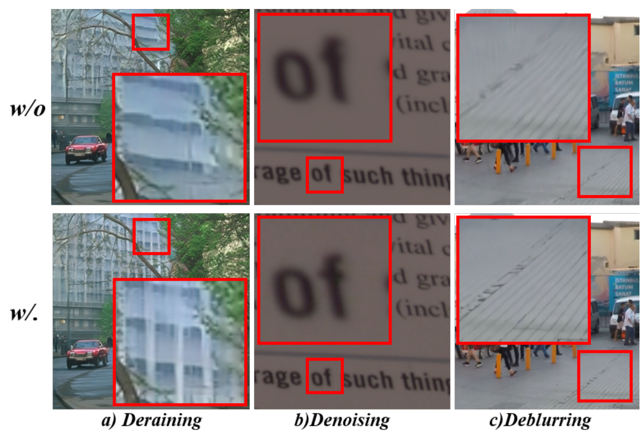

Fig. 1. Visual results of w. and w/o EA.

Fig. 2. Image deraining on the Test1200 dataset [8].

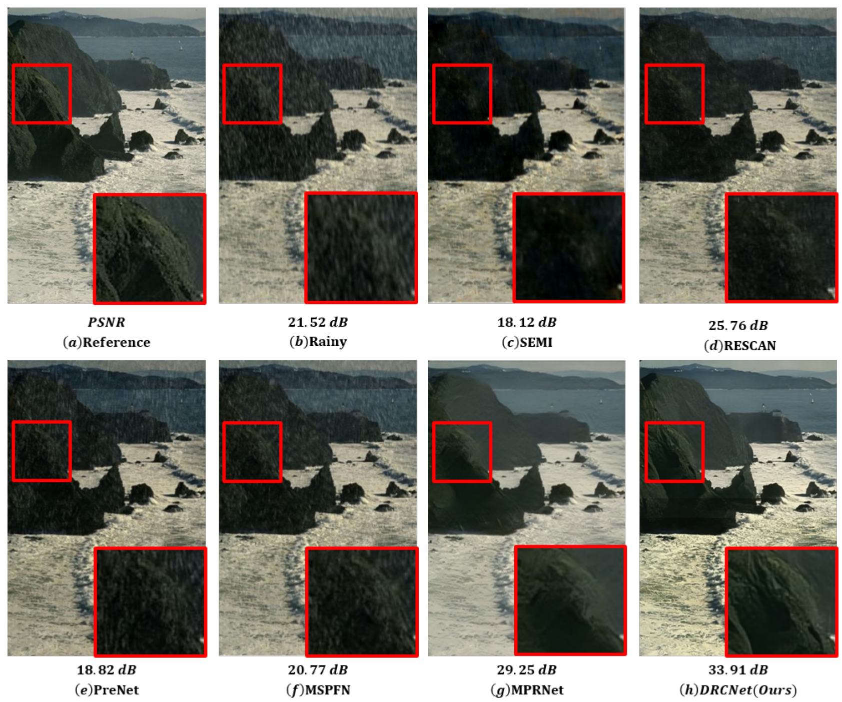

Fig. 3. Image deraining on the Test100 dataset [9].

Fig. 4. Image deblurring on the RealBlur-J dataset [4]

References

1. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1692–1700 (2018)
2. Boyd, S., Boyd, S.P., Vandenberghe, L.: Convex optimization. Cambridge university press (2004)
3. Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3883–3891 (2017)
4. Rim, J., Lee, H., Won, J., Cho, S.: Real-world blur dataset for learning and benchmarking deblurring algorithms. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 184–201 (2020)
5. Shalev-Shwartz, S., Ben-David, S.: Understanding machine learning: From theory to algorithms. Cambridge university press (2014)
6. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1357–1366 (2017)
7. 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 (CVPR). pp. 14821–14831 (2021)
8. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 695–704 (2018)
9. Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. IEEE Transactions on Circuits and Systems for Video Technology **30**(11), 3943–3956 (2019)
10. Zheng, Y., Zhang, R., Mao, Y.: Regularizing neural networks via adversarial model perturbation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8156–8165 (2021)