Implicit Learning of Rain Factor Disentangled Representation for Single Image De-raining

1. Response to General Concerns

1.1. More clear explanation of RFDR-Net

Motivation and Contributions: To learn an explainable, controllable, and generalizable de-raining model with convolutional neural network, imposing some meta-priors (*e.g.*, disentanglement, hierarchical organization of features) on the learned representation should be an effective way. In this paper, we turn to the disentangled representation learning mechanism, and design a model that captures the independent factors (*e.g.*, rainy pattern related feature and background related feature) of a given rainy image in such a way that if the rainy pattern related factor changes, the others remain unaffected. If done successfully, such a de-raining model will be able to (1) deal with different rainy conditions (*e.g.*, different density, shape, orientation *etc*.) (2) generalize well to unseen examples. Most importantly, models with those properties will undoubtedly work well on real rainy images. Specifically, without the availability of labels depicting the properties of rainy factors, the disentanglement is learned in a weakly supervised manner by using the clean images as implicit supervision. To achieve this, a multi-task learning framework is designed involving four tasks of clean-to-clean translation, rainy-to-clean translation, clean-to-rainy translation, and rainy-to-rainy translation (Ref Fig.3 of the paper), and an implicit and dynamic knowledge transfer procedure is implemented with such framework enabling the disentangled representation learning (Ref Eqns(4-9) of the paper). Accordingly, the overall contribution of our paper includes:

1. An explainable and generalizable deraining framework is built on the representation disentanglement mechanism, and we provide an elegant knowledge transfer strategy to achieve the disentanglement.
2. To guarantee a steady and smooth knowledge transfer so as to obtain high-quality disentangled representation, the multi-task learning model is trained with an elaborately designed adversarial loss formulation. We argue that such a loss formulation can be used generally to train all the existing GAN models and should provide competitive results compared with those obtained with existing adversarial loss formulation.
3. The proposed framework can be used as a) referenced model for learning disentangled representation in a weakly supervised manner, and b) general framework for other image restoration tasks and universal image-to-image translation task.

Limitation: The limitation of the proposed framework shows as inductive bias, which exists in nearly all existing disentangled representation learning framework [1]. For the proposed model, as illustrated in Fig.3 of the paper, the learning of two different factors are implemented in a mutual way. This means, the quality of the learned factor relies on the quality of the other one. Unluckily, due to the absence of direct rainy factor related supervision, the implicit supervision results in an impure learning of rainy related factor (maybe mixed with factor describing the background attribute), thus further inducing the unsatisfactory learning of incomplete background related factor. Due to the possible information loss in the background related factor, some of the de-rained results will be incomplete, showing as the over-deraining phenomenon in some results in Fig. 7 of the paper and Fig. R1 of this letter. In the future, inspired by [3], we plan to train a Gaussian Mixture Model with multiple components to model the distribution of complex rainy related patterns, and then define another loss function to estimate the KL divergence between the GMM prior and the rainy related factor. Benefited from the powerful GMM prior, the learned rainy related factors are regularized and improved, thus improving the completeness of the learned background factor and reducing the over-smoothing problems.

1.2. More results on real rainy images

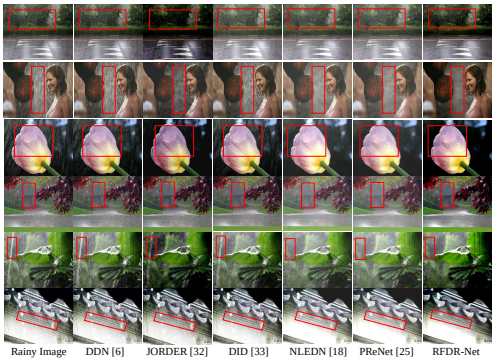

Figure R1: Visual quality comparison on some real rainy images (zoom in or click `here` to see higher resolution images).

As shown in Fig. R1, at first glance, very promising results are obtained by the proposed RFDR-Net, which can on the one hand remove the rain streaks thoroughly, and on the other hand preserve the detailed structure well with high contrast (*e.g.*, the results in the $2^{nd}/4^{th}/6^{th}$ rows). On the contrary, the other methods usually generate blurry results with important details missing. Among the comparative methods, JORDER generally performs the best. However, compared with our RFDR-Net, the results of JORDER are still blurry

(*e.g.*, the results in the 2nd/6th rows) and some important structural details of background cannot be recovered well (*e.g.*, the result in the 1st row is very dark with low contrast caused by the over-deraining effect, and the structure of the wall in the 4th row is not recovered well).

Unsatisfactory cases: Despite the superior results obtained by RFDR-Net, due to the inherent "inductive bias" issue (as analyzed in Sec1.1 of this letter), some over-deraining results are observed as in the 3rd/5th rows of Fig. R1.

1.3. Detailed description of experimental setup

For fair comparison, the experiments are carried out by following the setup that are adopted by most de-raining papers (we especially follow the setup in PReNet [2]). Specifically, the detailed setups are provided in Table R1 for explaining how the results from the deep learning-based methods are obtained (results in Table 3 of the paper). For the non-deep learning methods, we adopt the public codes directly to obtain both the quantitative and qualitative results.

Dataset	DDN	JORDER	DID	NLEDN	PReNet	RFDR-Net
DDN-Data	①	①	①	④	②	④
DID-Data	①	①	②	④	①	④
Rain100H	①	③	①	④	②	④

Table R1: Detailed experimental setups: ① represents using the codes provided by the authors, and re-train the model on the specific dataset to obtain the results; ② represents using the pre-trained weights provided by the authors to obtain the results; ③ represents calculating the quantitative results with the de-rained images provided by the authors; ④ represents implementing the codes by ourselves and train the model on the specific dataset to obtain the results. All the codes will be released after paper acceptance.

Besides, for Rain100H data, it is found that 546 images in the training set have the same background with the testing set. We have noticed this during preparing the paper, and conducted all the re-training procedures on the remaining 1254 training images, thus the results are convincing. Also, all the results in the ablation study of the paper are obtained by training different models on the 1254 training images.

1.4. Feature maps of task relevant/irrelevant factorss

As can be seen from Fig. R2, as expected, the rainy pattern are more highlighted from the task-irrelevant layer while the background pattern are more highlighted from the task-relevant layer. However, as analyzed in Sec1.1 of this letter, the existing implicit learning formulation cannot guarantee a well-constrained separation of these two factors, resulting in the feature maps in the task-irrelevant layer also encode some background patterns while inducing some desired information loss in the task-relevant layer, and such drawbacks cause the over-deraining results, as shown and analyzed in Sec2 of this letter.

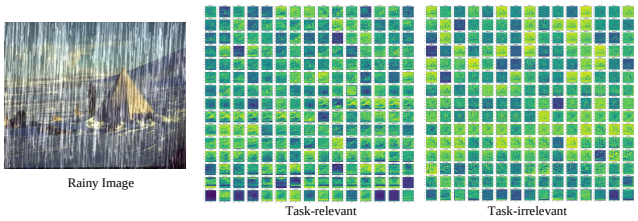

Figure R2: Activations from the task-relevant/irrelevant layers before ReLU (zoom in or click here to see higher resolution images).

2. Response to Specific Comments

Reviewer #1: • During training, four tasks including a) clean-to-clean image translation, b) clean-to-rainy image translation, c) rainy-to-rainy image translation, and d) rainy-to-clean image translation, are involved for learning disentangled representation. During testing, only the rainy-to-clean image translation branch are used for performing de-raining task. For Eqns (4-5) in the paper, the latent code of clean image is only involved in the task of clean-to-rainy image synthesis task, serving as an auxiliary branch for learning the task-irrelevant representation. Thus, the latent code from clean image doesnt involve the de-raining branch directly. • For the Rain100H dataset, when training the model, we have excluded the 546 images in the training set that have the same background with the testing set, refer to Sec1.3 of this letter for detailed explanation. • More results on real rainy images and the corresponding analysis are provided in Sec 2 of this letter. Note that the deraining results (see the 2nd row of Fig. R1) on real image with heavy rain are very satisfactory and much better than all the other SOTA models.

Reviewer #2: • We claim that our experimental setups are fair and consistent with that from most existing SOTA deraining papers. Refer to Sec3 of this letter for detailed explanation. • Limitation of the proposed RFDR-Net is analyzed theoretically (Sec1.1 of this letter) and experimentally (Sec1.2 of this letter). • More results on real rainy images and corresponding analysis are provided on Sec2 of this letter. • We argue that the removal of the red things is resulted from decoding only the task-relevant factor (the background statistics) for reconstructing the de-raining results. Differently, the existing SOTA refers to a direct rainy-to-clean mapping formulation, without a very explicit design of removing the task-irrelevant factors (*e.g.*, the raindrops and red things in the third example of Fig.5 of the paper).

Reviewer #3: • In this paper, we focus more on providing a much better learning formulation that are on the one hand can describe and solve deraining problem better, and on the other hand are generalizable to be applied to any image-to-image translation network. Experimental results turn out our formulation can achieve much better results when compared with those SOTA models with complex hand-engineered

network architecture (*e.g.*, NLEDN or DID). Moreover, the design of AutoML and NAS technology make it unnecessary to design network architecture in a hand-engineered manner. We look forward to seeing the combination of our proposed formulation and auto-designed/searched network architecture to obtain a flexible, lightweight, and powerful de-raining model. • Analysis on the learned feature maps are provided in Sec1.4 of this letter. Also refer to Sec1.1 of this letter to get a better understanding on the pros and cons of the proposed model. Q3: In the future, for these missing references, we will (1) add the analysis to the Section2 (Related Work) of the paper, and (2) add both the quantitative and qualitative results to the Section4.3 of the paper.

References

[1] F. Locatello, S. Bauer, M. Lucic, S. Gelly, B. Schölkopf, and O. Bacheme. Challenging common assumptions in the unsupervised learning of disentangled representations. In *ICML*, 2019.

[2] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng. Progressive image deraining networks: A better and simpler baseline. In *CVPR*, 2019.

[3] W. Wei, D. Meng, Q. Zhao, Z. Xu, and Y. Wu. Semi-supervised transfer learning for image rain removal. In *CVPR*, 2019.