The Supplementary Materials for "Learning Distortion Invariant Representation for Image Restoration from A Causality Perspective"

Xin Li[1], Bingchen Li[1], Xin Jin[2], Cuiling Lan[3]†, Zhibo Chen[1]†
[1]University of Science and Technology of China [2]Eastern Institute for Advanced Study
[3]Microsoft Research Asia
{lixin666, lbc31415926}@mail.ustc.edu.cn, jinxin@eias.ac.cn,
culan@microsoft.com, chenzhibo@ustc.edu.cn

Section 1 provides the systematic introduction for the related notations of the back-door criterion in causal learning.
Section 2 explains the counterfactual distortion augmentation from the causality perspective.
Section 3 theoretically derives the parallel sampling in Eq. 5 of our paper.
Section 4 clarifies the implementations of four variants of our DIL, which can help the readers to reproduce our methods more easily.
Section 5 describes the more detailed experimental settings and the construction of distortion/confounder set D in different image restoration tasks.
Section 6 visualizes more subjective comparisons on different image restoration tasks.

1. The Back-door Criterion in Causality.

In this section, we clarify the related notations and derivations for the back-door criterion in causality.

Structure causal Model. As described in [6, 14], we can describe the causal relationship between different vectors with a directed Structural causal Model (SCM) like Fig. 1. A directed arrow $X \rightarrow Y$ represent X is the cause of the Y. The difference between correlation and causation is as follows: 1) In causation, given $X \rightarrow Y$, changing the X will cause the effect on Y. But changing Y does not have an effect on X since Y is not the cause of X. 2) In correlation, we can compute the correlation between X and Y with conditional probability $P(Y|X)$ and $P(X|Y)$ no matter whether there is causation between X and Y. In general, model training in deep learning is a process to fit the correlation between inputs and their labels instead of the causation.

Confounder. The confounder is defined based on the SCM, which represents the variables (*e.g.*, C in Fig. 1) that are the common cause between two other variables (*e.g.,* X and Y in Fig. 1). The fork connection $X \leftarrow C \rightarrow Y$ causes the spurious correlation for X and Y, which has a confounding effect on the estimation of the causal relationship between X and Y. In other words, the correlation between X and Y learned by the model also is implicitly conditioned on the confounder C.

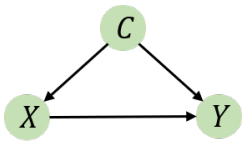

Figure 1. A structural causal model for back-door structure.

***do* operation.** A *do* operation means to cut off the connection from the $C \rightarrow X$, which is shown in Fig. 2. In this way, the correlation introduced from the path $X \leftarrow C \rightarrow Y$ is removed from *do* operation. Then the correlation learned by the model is only from the $X \rightarrow Y$, which are represented as $P(Y|do(X))$. And this causal correlation is independent of the confounder C and is what we expect the model to learn.

Back-door criterion. The back-door criterion is proposed in [6, 14], which aims to implement the *do* operation and eliminate the spurious correlation existed in $X \leftarrow C \rightarrow Y$. It removes the confounding effects of confounder C by computing the average causal effects between $X \rightarrow Y$ by traversing all values of C as:

$$P(Y|do(X)) = \sum_{c} P(Y|X, C = c)P(C = c) \quad (1)$$

Based on Eq. 1, we can achieve the *do* operation in Fig. 2 (b).

The back-door criterion in Image Restoration As shown in Fig. 2 in our paper, we model the image restoration process as a structural causal model, where $D = \{d_i|1 \leq i \leq n\}$ are the confounders between the distorted images I_d and the expected reconstruction images I_o, which satisfies the

†Corresponding Author

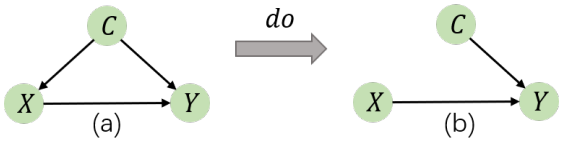

Figure 2. Back-door Criterion in Causality.

back-door criterion. Therefore, we can derive the back-door criterion in image restoration as:

$$P(I_o|do(I_d)) = \sum_{i=1}^{n} P(I_o|I_d, D = d_i)P(D = d_i) \quad (2)$$

2. A proof for counterfactual distortion augmentation.

The conterfactuals aims to answer the question "*"if X been x, in the situation U, what $Y_{X=x}(U)$ would be?"*. The three variables are in the same structural causal model (SCM), and X and U are the cause of Y. As described in [6, 14], the calculating of counterfactuals follows three steps: 1) Abduction: Use evidence e to determine the value of U. 2) Action: Remove the structural equations for the variables X to modify the model M (*i.e.,* the SCM). Then, set the X as $X = x$ to obtain the modified M_x. 3) Prediction: Use the M_x and $U = u$ to compute the value of Y (*i.e.,* the consequence of the counterfactual).

Considering the generation process of the distorted images $I_d = g(I_c, d)$, where I_c and d are the clean images and distortion type/degree, respectively. g is the degradation process. The generation process can be modeled as a structural causal model $I_c \rightarrow I_d \leftarrow d$. To construct the datasets for the training of DIL, it is better to collect various distorted/clean image pairs with different distortions but the same content. However, in the real world, it is non-trivial to collect the datasets to satisfy this. Therefore, we can construct the ideal datasets by answering the counterfactual question *"if D is d_i, what the I_d would be with I_c invariant?"*. We call the construction counterfactual distortion augmentation.

Analogously, the computing of counterfactuals in distortion augmentation follows a three-step procedure. 1) Abduction: Use the distorted image I_d to determine the value of I_c, *i.e.,* $P(I_c|I_d)$. 2) Action: Modify the degradation model, g, so that D is adjusted to the counterfactual value d_i, that might rarely existed in real-world (*e.g.,* the synthesised distortions). 3) Prediction: Compute the consequence I_{d_i} of the counterfactual based on estimated I_c and modified degradation model g_{d_i}.

It is fortunate that there are amounts of high-quality images captured by professional devices, that are only degraded by some extremely mild distortions. We can regard these images as clean images I_c. Therefore, the first step in counterfactuals is unnecessary and can be ignored. We can implement the counterfactual distortion augmentation by adding different synthetic distortion types or degrees to the same image contents I_c.

3. The derivation of the parallel sampling.

In this section, we will give the derivation of our parallel sampling in Eq. 5 of our paper. From Eq. 3 and Eq. 4 in our paper as:

$$\begin{aligned} \theta^* &= \arg\min_{\theta} \mathbb{E}_{(I_d,I_c)\sim\mathcal{D}}[\frac{1}{n}\sum_{d_i\in D}\mathcal{L}(f_{\phi_{d_i}}(I_d), I_c)], \\ &where \quad \phi_{d_i} = \theta - \alpha\nabla_\theta\mathcal{L}(f_\theta(I_{d_i}), I_c) \end{aligned} \quad (3)$$

we can conduct the Talyor expansion for the above equation at position θ as:

$$\begin{aligned} \theta^* &= \arg\min_{\theta} \mathbb{E}_{(I_d,I_c)\sim\mathcal{D}}\{\frac{1}{n}\sum_{d_i\in D}[\mathcal{L}(f_\theta(I_d), I_c) \\ &\quad - \alpha\nabla_\theta\mathcal{L}(f_\theta(I_{d_i}), I_c)\nabla_\theta\mathcal{L}(f_\theta(I_d), I_c) \\ &\quad + o(\nabla_\theta\mathcal{L}(f_\theta(I_d), I_c))]\} \\ &= \arg\min_{\theta} \mathbb{E}_{(I_d,I_c)\sim\mathcal{D}}\{\mathcal{L}(f_\theta(I_d), I_c) \\ &\quad - \frac{1}{n}\sum_{d_i\in D}\alpha[\nabla_\theta\mathcal{L}(f_\theta(I_{d_i}), I_c)]\nabla_\theta\mathcal{L}(f_\theta(I_d), I_c) \\ &\quad + o(\nabla_\theta\mathcal{L}(f_\theta(I_d), I_c))]\}\} \\ &= \arg\min_{\theta} \mathbb{E}_{(I_d,I_c)\sim\mathcal{D}}\{\mathcal{L}(f_\theta(I_d), I_c) \\ &\quad - \alpha\nabla_\theta[\sum_{d_i\in D}\frac{1}{n}\mathcal{L}(f_\theta(I_{d_i}), I_c)]\nabla_\theta\mathcal{L}(f_\theta(I_d), I_c) \\ &\quad + o(\nabla_\theta\mathcal{L}(f_\theta(I_d), I_c))]\}\} \end{aligned} \quad (4)$$

Then we conduct the Taylor inverse expansion for the Eq. 4. The Eq. 4 can be derived as:

$$\begin{aligned} \theta^* &= \arg\min_{\theta} \mathbb{E}_{(I_d,I_c)\sim\mathcal{D}}[\\ &\quad \mathcal{L}(f_{(\theta-\alpha\nabla_\theta\sum_{d_i\in D}\frac{1}{n}\mathcal{L}(f_\theta(I_{d_i}),I_c))}(I_d), I_c)] \end{aligned} \quad (5)$$

Let $\phi_{\bar{d}} = \theta - \alpha\nabla_\theta\sum_{d_i\in D}\frac{1}{n}\mathcal{L}(f_\theta(I_{d_i}), I_c)$, we can obtain the final equation as Eq. 5 of our paper as:

$$\begin{aligned} \theta^* &= \arg\min_{\theta} \mathbb{E}_{(I_d,I_c)\sim\mathcal{D}}[\mathcal{L}(f_{\phi_{\bar{d}}}(I_d), I_c)], \\ where\ \phi_{\bar{d}} &= \theta - \alpha\nabla_\theta\sum_{d_i\in D}\frac{1}{n}\mathcal{L}(f_\theta(I_{d_i}), I_c), \end{aligned} \quad (6)$$

4. The detailed algorithms on four variants of DIL.

We further demonstrate the algorithm details of four variants of our proposed DIL in the Alg. 1 (DIL$_{ps}$), Alg. 2

Algorithm 1 DIL$_{ps}$ (The variant of DIL with parallel sampling and second-order optimization)

Input: Training dataset $\mathcal{D} = \{I_{d_i}, I_c | 1 \leq i \leq n\}$, where n is number of distortion types and degrees (*i.e,* confounders), and $D = \{d_i | 1 \leq i \leq n\}$ is the confounder set.
Init: Restoration network f with the parameters θ, learning rate α for virtually updating, β for the training.
while not converge **do**
  Sample training pairs (I_d, I_c) from D.
  Sample training pairs $\{I_{d_i}, I_c\}_{i=1}^{n}$ from $\mathcal{D}$.
  Virtual updating for the parameters θ as : $\phi_{\bar{d}} \leftarrow \theta - \alpha \nabla_\theta \sum_{d_i \in D} \frac{1}{n} \mathcal{L}(f_\theta(I_{d_i}), I_c)$.
  Updating the parameters θ with second-order gradient: $\theta \leftarrow \theta - \beta \mathcal{L}(f_{\phi_{\bar{d}}}(I_d), I_c)$
end while

Algorithm 2 DIL$_{ss}$ (The variant of DIL with serial sampling and second-order optimization)

Input: Training dataset $\mathcal{D} = \{I_{d_i}, I_c | 1 \leq i \leq n\}$, where n is number of distortion types and degrees (*i.e,* confounders), and $D = \{d_i | 1 \leq i \leq n\}$ is the confounder set.
Init: Restoration network f with the parameters θ, learning rate α for virtually updating, β for the training.
while not converge **do**
  Sample training pairs (I_d, I_c) from D.
  for $1 \leq i \leq n$ **do**
    Sample training pairs (I_{d_i}, I_c) from $\mathcal{D}$.
    Virtual updating for the parameters θ: $\phi_{d_i} \leftarrow \theta - \alpha \nabla_\theta \mathcal{L}(f_\theta(I_{d_i}), I_c)$
    Compute the loss for the second-order gradient: $\mathcal{L}(f_{\phi_{d_i}}(I_d), I_c)$
  end for
  Updating the parameters θ with second-order gradient: $\theta \leftarrow \theta - \beta \frac{1}{n} \sum_{d_i \in D} \nabla_\theta \mathcal{L}(f_{\phi_{d_i}}(I_d), I_c)$
end while

(DIL$_{ss}$), Alg. 3 (DIL$_{pf}$), and Alg. 4 (DIL$_{sf}$). As derived in Eq. 6, we can utilize the parallel data sampling for all distortions D to substitute the serial sampling based optimization. The implementation differences between the two sampling strategies can be observed by comparing the Line $5-6$ in the Alg. 1 and Line $5-9$ in the Alg. 2. We can find that parallel sampling can reduce the number of parameter updating by $1/n$. By comparing the Alg. 1 and Alg. 3, we can find that only first-order gradient existed in the DIL$_{pf}$, which is an approximation of the second-order optimization in Alg. 1. The related proof can be found in the [13].

Algorithm 3 DIL$_{pf}$ (The variant of DIL with parallel sampling and first-order optimization)

Input: Training dataset $\mathcal{D} = \{I_{d_i}, I_c | 1 \leq i \leq n\}$, where n is number of distortion types and degrees (*i.e,* confounders), and $D = \{d_i | 1 \leq i \leq n\}$ is the confounder set.
Init: Restoration network f with the parameters θ, learning rate α for virtually updating, β for the training.
while not converge **do**
  $\tilde{\theta} \leftarrow \theta$
  for step=1 to 2 **do**
    Sample training pairs $\{I_{d_i}, I_c\}_{i=1}^{n}$ from $\mathcal{D}$.
    Virtual updating: $\tilde{\theta} \leftarrow \tilde{\theta} - \alpha \nabla_{\tilde{\theta}} \sum_{d_i \in D} \frac{1}{n} \mathcal{L}(f_{\tilde{\theta}}(I_{d_i}), I_c$.
  end for
  Updating the parameters θ: $\theta \leftarrow \theta - \beta(\tilde{\theta} - \theta)$
end while

Algorithm 4 DIL$_{sf}$ (The variant of DIL with serial sampling and first-order optimization)

Input: Training dataset $\mathcal{D} = \{I_{d_i}, I_c | 1 \leq i \leq n\}$, where n is number of distortion types and degrees (*i.e,* confounders), and $D = \{d_i | 1 \leq i \leq n\}$ is the confounder set.
Init: Restoration network f with the parameters θ, learning rate α for virtually updating, β for the training.
while not converge **do**
  $\tilde{\theta} \leftarrow \theta$
  for $1 \leq i \leq n$ **do**
    Sample training pairs (I_{d_i}, I_c) from $\mathcal{D}$.
    Virtual Updating: $\tilde{\theta} \leftarrow \tilde{\theta} - \alpha \nabla_{\tilde{\theta}} \mathcal{L}(f_{\tilde{\theta}}(I_{d_i}), I_c)$
  end for
  Updating the parameters θ: $\theta \leftarrow \theta - \beta(\tilde{\theta} - \theta)$
end while

5. Implementation Details.

5.1. Overall Settings.

For all image restoration tasks (except for the image deraining task) in this paper, we use 800 images from DIV2K [2] and 2650 images from Flickr2K [16] as the clean images to construct the datasets for training. Following the common setting [12, 21], In the training process, we randomly crop the distorted/clean image pairs with the size of 64×64 from the training images, and feed them to the restoration network to optimize the parameters. In the process of the counterfactual distortion augmentation, the distorted patches I_d are generated online according to dis-

tortion set D of different image restoration tasks. For ERM, we use Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. For DIL$_{sf}$ and DIL$_{pf}$ training paradigms, the same Adam optimizer with ERM is used for the training optimization for the above two variants. For the virtual updating process, we adopt the Adam optimizer with $\beta_1 = 0$ and $\beta_2 = 0.999$ following [13]. For DIL$_{ss}$ and DIL$_{ps}$, we utilize the same two Adam optimizers as that used in ERM for the virtual updating step and training optimization step. We set the batch size to 8 on each GPU. The total training iterations and initial learning rate are set to 400K and 1e-4, respectively. The learning rate will reduce by half at [200K, 300K]. All the tasks are optimized with the L1 loss if not mentioned. In the image deraining task, we utilize Charbonnier [4] Loss as:

$$\mathcal{L}_{char} = \sqrt{\|I_o - I_c\|^2 + \epsilon^2} \tag{7}$$

where I_o and I_c denotes the reconstructed images and clean images, respectively. Following previous works [8, 21], we set ϵ to 1e-3.

5.2. The distortion/confounders set D for different tasks.

In this section, we describe the specific construction of the distortion/confounder set D in the counterfactual distortion augmentation strategy.

5.2.1 Cross distortion degrees

Image Denoising. For image denoising, the distortion/confounder set is composed of Additive White Gaussian Noise (AWGN) with the noise intensity of [5, 10, 15, 20], which is added to the clean images from DF2K [2, 16] to construct the training data. For testing, we utilize several unseen noise intensities, including [30, 40, 50] to estimate the generalization capability of different schemes.

Image Deblurring. For image deblurring, we obtain I_d by applying the distortion/confounding set D to I_c, which contains the 2D gaussian filter with different blurring sigma of [1.0, 2.0, 3.0, 4.0]. For testing, we validate the generalization capability of different schemes on the sigma [4.2, 4.4, 4.6, 4.8, 5.0].

Hybrid distortion restoration. Following the [10, 20], the hybrid distortions are degraded with the superposition of blur, noise, and Jpeg compression artifacts in a sequence manner. The distortion/confounder set D for training is composed of multiple levels of severe hybrid distortions. The test datasets are composed of unseen distortion levels, including mild and moderate hybrid distortions.

5.2.2 Cross distortion types

Real Image Super-resolution. For real image super-resolution, we utilize the degradation model introduced by [17] for training. To simplify the training process, we adopt the one-order distortion synthesis mode in [17] to construct the distortion/confounder set D, where different $d_i \in D$ are divided with different noise types and blur types in the degradation model of [17]. To validate the generalization capability of different schemes for the "cross distortion types", we exploit the RealSR [3] and DRealSR [18] for the real image super-resolution as our test data.

Real Image Denoising. For real image denoising, we obtain I_d based on the ISP process introduced by [7]. We divide this degradation model into four different distortion types based on the different color filter arrays (CFA) to construct D. We follow previous works [1, 15] and utilize real denoising datasets DND [15] and SIDD [1] as the benchmarks to validate the generalization capability of different schemes.

Image Deraining. For image deraining, we utilize three datasets with three different raining types, including Rain14000 [5], DID-MDN [22], and Heavy Rain Dataset [9], to construct the training data $\mathcal{D}$, and test the generalization capability of the restoration network on other three unseen raining types, including 100 rainy images from Rain100L [19], 100 rainy images from Rain800 [23], and 12 rainy images from Rain12 [11]. It is noteworthy that the synthesis strategies of the above raining types are rarely released. Therefore, in this task, we relax the content consistency between different raining types for training. And our DIL is still effective for improving the generalization capability of the restoration network.

6. More Subjective Visualization

We provide more visual comparisons for different image restoration tasks in this section. As shown in Fig. 3, the commonly-use ERM and our proposed DIL all achieve similar reconstructed results on the seen noise level ($\sigma = 15$). But ERM fails to restore high-quality images on unseen noise levels well, (*e.g.,* $\sigma = 40$ and $\sigma = 50$), which indicates that ERM lacks of enough generalization ability for the unseen distortion degrees. In contrast, our DIL$_{sf}$ can recover noise-free and structure-preserved images despite the distortion degrees do not exist in the training data. This further proves the correctness and effectiveness of our proposed DIL.

We show more visual comparisons of image deblurring in Fig. 4. When dealing with unseen blur degrees, our proposed DIL can restore the clear structures, while ERM produces overshooting artifacts on the edges. More visualizations for real image denoising can be found in Fig. 5. And more visualizations for image deraining can be found

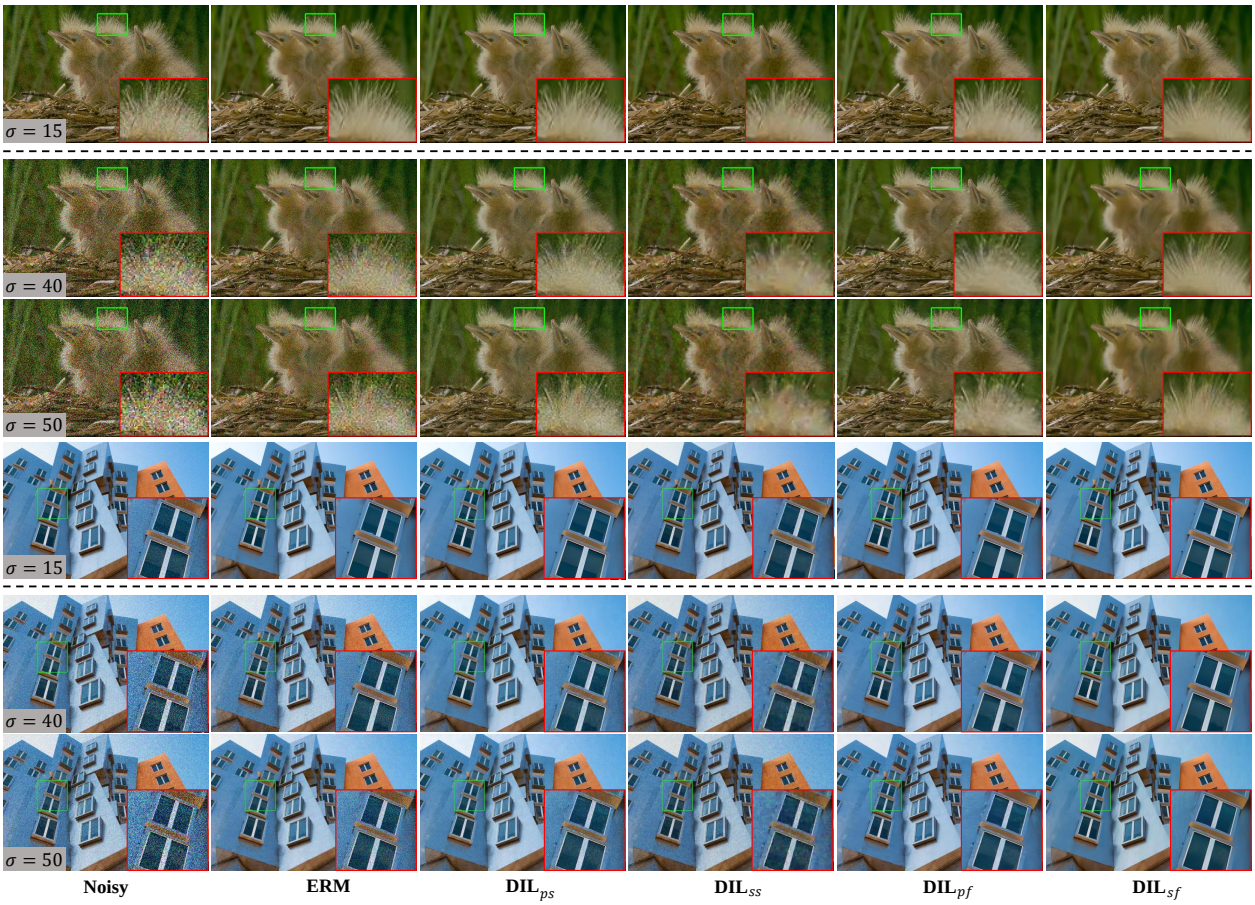

Figure 3. Visual comparison of the commonly-used ERM and our proposed four variants of DIL on image denoising task. The noise level above the dash line is seen, while noise levels below are unseen.

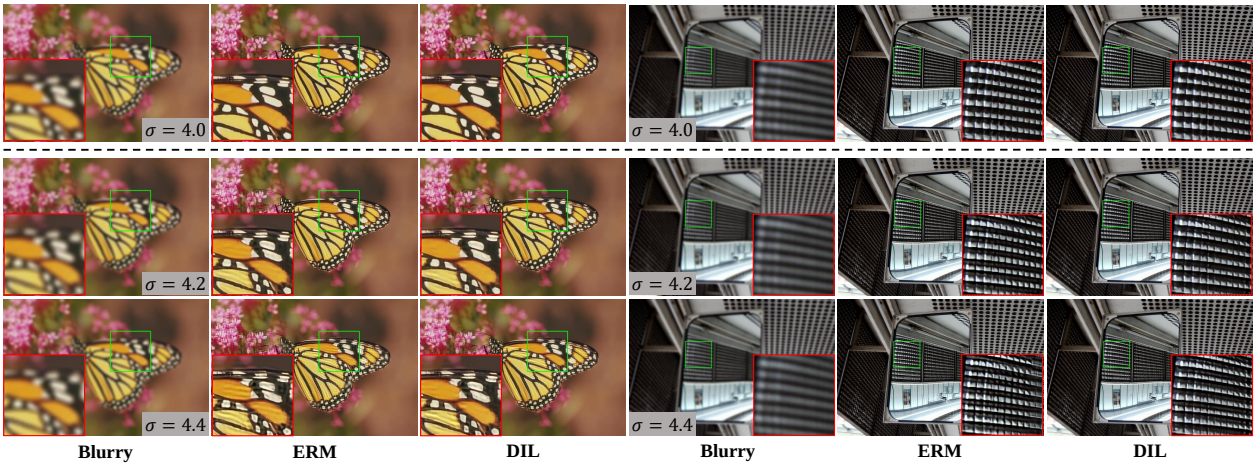

Figure 4. Visual comparison of the commonly-used ERM and DIL on image deblurring task. The blur level above the dash line is seen, while blur levels below are unseen.

in Fig. 6. We also visualize the subjective comparisons on hybrid-distorted image restoration in Fig. 7.

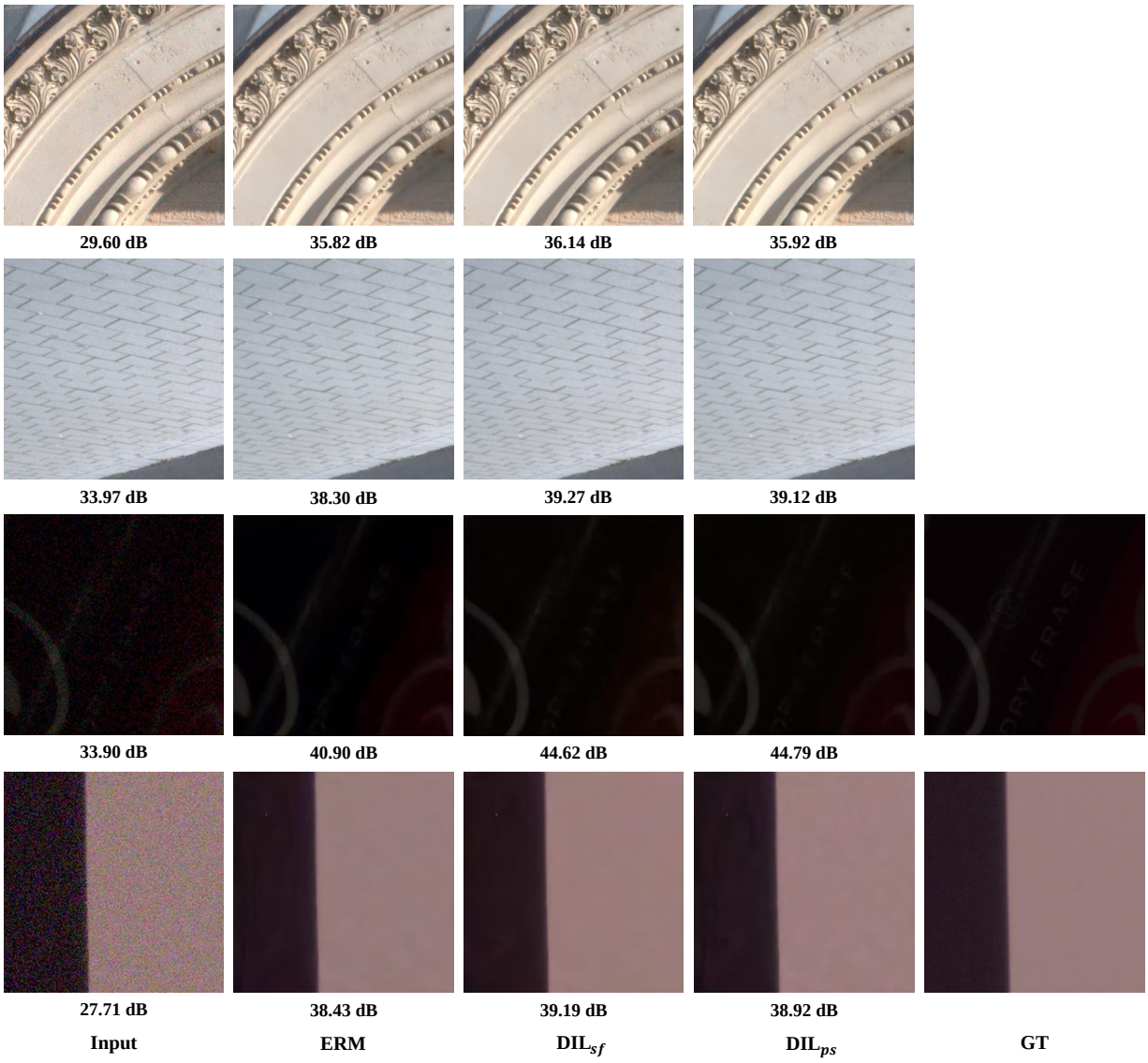

Figure 5. Visual comparison of the commonly-used ERM and DIL on real image denoising task. The top samples are from DND [15] while the bottom samples are from SIDD [1]. Brightening the third line for a better view.

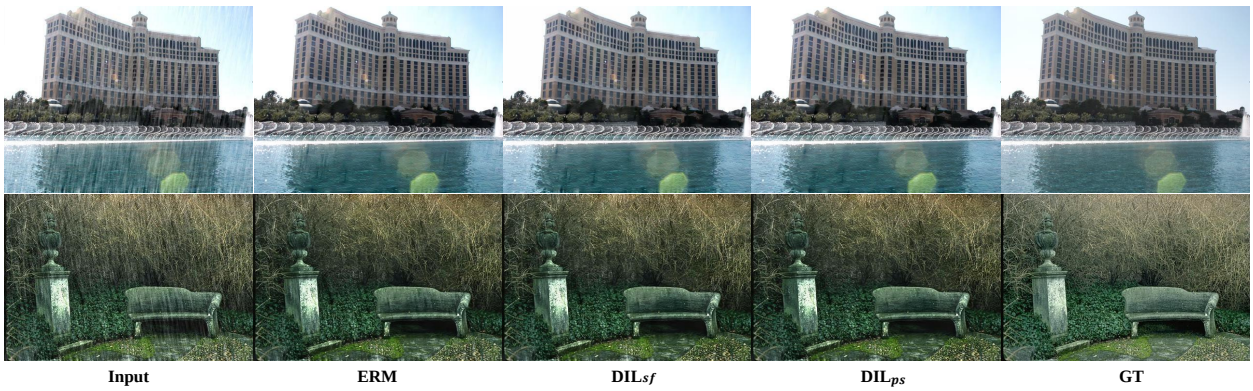

Figure 6. Visual comparison of the commonly-used ERM, our DIL$_{sf}$ and DIL$_{ps}$ on image deraining task.

Figure 7. Visual comparison of the commonly-used ERM and DIL on hybrid distortion removal task. Here, we show restoration results on the mild distortion level, which is the unseen distortion level for the restoration network.

References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1692–1700, 2018. 4, 6

[2] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017. 3, 4

[3] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3086–3095, 2019. 4

[4] Pierre Charbonnier, Laure Blanc-Feraud, Gilles Aubert, and Michel Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of 1st International Conference on Image Processing*, volume 2, pages 168–172. IEEE, 1994. 4

[5] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3855–3863, 2017. 4

[6] Madelyn Glymour, Judea Pearl, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016. 1, 2

[7] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1712–1722, 2019. 4

[8] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8346–8355, 2020. 4

[9] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1633–1642, 2019. 4

[10] Xin Li, Xin Jin, Jianxin Lin, Sen Liu, Yaojun Wu, Tao Yu, Wei Zhou, and Zhibo Chen. Learning disentangled feature representation for hybrid-distorted image restoration. In *European Conference on Computer Vision*, pages 313–329. Springer, 2020. 4

[11] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2736–2744, 2016. 4

[12] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 3

[13] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018. 3, 4

[14] Judea Pearl. Causal inference in statistics: An overview. *Statistics surveys*, 3:96–146, 2009. 1, 2

[15] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1586–1595, 2017. 4, 6

[16] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 114–125, 2017. 3, 4

[17] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1905–1914, 2021. 4

[18] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *European Conference on Computer Vision*, pages 101–117. Springer, 2020. 4

[19] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1357–1366, 2017. 4

[20] Ke Yu, Chao Dong, Liang Lin, and Chen Change Loy. Crafting a toolchain for image restoration by deep reinforcement learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2443–2452, 2018. 4

[21] 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. 3, 4

[22] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 695–704, 2018. 4

[23] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology*, 30(11):3943–3956, 2019. 4