

# Exploiting Distortion Information for Multi-degraded Image Restoration

Wooksu Shin<sup>1</sup> Namhyuk Ahn<sup>2</sup> Jeong-Hyeon Moon<sup>1</sup> Kyung-Ah Sohn<sup>1</sup>

<sup>1</sup> Department of Artificial Intelligence, Ajou University    <sup>2</sup> NAVER WEBTOON AI

## Abstract

*In recent years, tremendous studies have been performed on the image distortion restoration task and deep learning-based methods have shown prominent performance improvement. However, assuming only a single distortion to an image may not be applicable in many real-world scenarios. To mitigate the issue, some studies have proposed multi-distortion datasets by applying the corruptions sequentially or spatially. In this work, we integrate the two perspectives on the multi-distortion nature and propose a new dataset that is a holistic multi-distortion dataset. To restore the multi-distortion effectively, we introduce a distortion information-guided restoration network, which exploits the conditional distortion information when reconstructing a given image. To do that, our framework first predicts the distortion type and their strength and delivers these to the restoration module. In our experiments, we show that the proposed model exceeds the others and we also demonstrate that any backbone network benefits from receiving the distortion information as prior knowledge.*

## 1. Introduction

Image restoration has been actively researched in computer vision field for several decades. It aims to reconstruct the degraded image into a cleaned one. There exist various sub-problems of the image restoration, such as image super-resolution [27,34], denoising [1], and deblurring [29]. Because the modeling of these tasks requires a one-to-many mapping function, it is challenging to develop the optimal solution. Nevertheless, many studies have been done on image restoration for various real-world applications.

Most of the works on image restoration tasks have been postulated that an input image is corrupted with a single and fixed-intensity distortion [8,14]. However, as various types of corruption with unknown strength can be applied in real-world applications, this assumption may not hold in many cases. Thus, some studies have introduced multiple distortion datasets [21,39] or methods [18,24,32,40] to close the gap between the simulated and the real environment.

Previous multi-distortion datasets are broadly catego-

rized into two types based on the dataset generation procedure. The first regime is the sequentially-applied (or mixed) distortion dataset [39] as shown in Figure 1b. This generates a corrupted image by overlapping the multiple distortions sequentially. However, because the same distortions are employed to the entire image (globally) this may not be suitable to the case where only a part of an image is corrupted. The second category is the spatially heterogeneous distortion [21], which corrupts an image with different distortions to each divided space (Figure 1c). This approach tackles the side that sequentially-applied distortion is overlooked, however, the underlying assumption of [21] is limited since this corrupts each region with a single distortion.

By integrating the idea of two aforementioned multi-distortion regimes, we introduce a holistic multi-distortion dataset (HMDD). To jointly implement the sequentially- and spatially-applied corruptions, we employ randomly selected distortions to the chunked region of a given image (Figure 1d). By doing so, our dataset can absorb the characteristic of two categories and we argue that HMDD is a general form of the previous multi-distortion datasets.

To effectively restore the multi-degraded image, we propose a distortion information-guided network (DIGNet). This framework first predicts the distortion information via the recognition module and injects this into the restoration network (Figure 2). We build the recognition component based on UNet [31] and train this in a supervised manner by utilizing the label information from a training dataset. With this, we can extract the conditional distortion information (CDI) that contains useful clues when reconstructing a corrupted image with spatially-variants multi-distortions. Then, we encode CDI to a form of restoration network to be received in a better way via mapping network. The restoration module reconstructs a given image by referring to the mapped CDI and we use the spatial feature transform (SFT) layer [38] to inject information. We show that our proposed DIGNet surpasses the other multi-distortion restoration methods. In addition, we also demonstrate the prominent performance gain of using CDI in various backbone networks. Our contribution can be summarized as follows:

- We introduce a new multi-distortion dataset by fusing the two previous assumptions on generating multi-

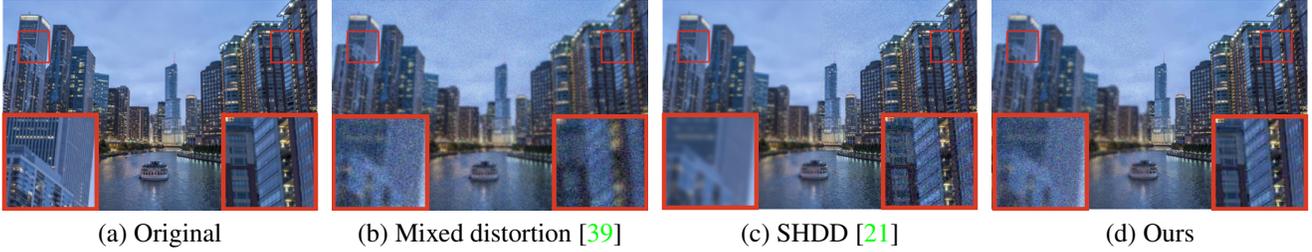


Figure 1. **Comparison of three different multi-distortion datasets.** (a) Original image. (b) Mixed distortions [39]. Multiple distortions are sequentially applied to an image. (Gaussian blur, Gaussian noise, JPEG in this case). (c) Spatially heterogeneous distortion dataset [21]. Multiple distortions are applied differently for each divided space of the image (left: Gaussian blur, right: Gaussian noise). (d) Our proposed HMDD. We integrate the idea of mixed-distortions and SHDD (left: Gaussian blur, Gaussian noise, JPEG, right: JPEG).

distortion. Our dataset can cover both sequentially and spatially corrupted distortion scenarios.

- We propose a restoration method for a multi-distortion environment. This method profits from utilizing conditional distortion information predicted by the distortion recognition module.
- Our experimental results show that provision of distortion clue as conditional information consistently improves the restoration methods.

## 2. Related work

**Image restoration.** The goal of this task is to reconstruct a corrupted image. Recently, deep learning-based approaches have shown remarkable performance in most of the sub-tasks such as image denoising [7, 8, 41, 42, 46], deblurring [23, 28, 33, 43, 45], and super-resolution [6, 14, 20, 25, 47]. The aforementioned studies improve their restoration ability by increasing the network capacity [20, 25], or by the novel architectural design [5, 41, 47].

**Multi-degraded image distortion.** In the real-world scenario, an image can be corrupted by the various distortions so that the model trained on the single-distortion dataset is possibly not suitable for this harsh environment. To make the assumption similar to the real-world application, recent studies have proposed datasets [4, 21, 26, 39] or methods [18, 21, 24, 32, 39, 40] for the multi-distortion restoration tasks. In these datasets, distortions are sequentially applied to the entire image [24, 39] or different corruptions distort the specific spatial region [4, 21]. In this work, we propose a new multiple distortion dataset by fusing the characteristic of both assumptions (sequential or spatial existences).

To recover a sequentially-degraded image, Yu et al., [39] exploited the “model bank” that stores the networks tailored to the designated corruptions. Subsequently, the restoration process of the given distorted image is determined through reinforcement learning. Suganuma et al., [32] utilized the attention mechanism, which gathers the features from the

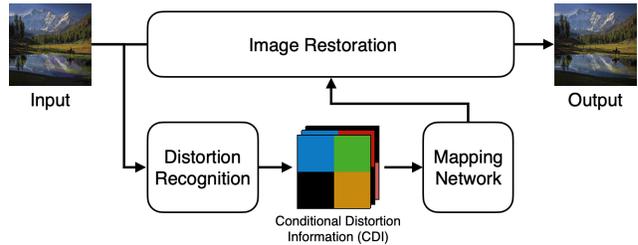


Figure 2. **Overview of our proposed DIGNet.** Our framework first recognizes the distortion and passes this to the restoration module as the conditional information.

multi-branch operations by the weighting scheme to guide the model to manage multiple distortions automatically. For the spatial-aware multi-distortions, Kim et al., [21] introduced a mixture of experts system in which each expert restores different distortions, and shares the parameters of all experts for the ease of learning shared representation.

**Image restoration in real world.** In recent studies, various real-world distortion datasets have been introduced [2, 9, 10, 30] to minimize the disparity between the synthetically generated distortions and the real ones. However, because of the high burden of collecting high- and low-quality image pairs, they confined the scope to the single distortion (*e.g.* noise [2, 30] or low-resolution [9, 10]). On the other hand, Real-ESRGAN [37] and BSRGAN [44] have advanced the generation process of the synthetic distortion to cover the majority of complex real-world distortions. The core spirit of our dataset is related to the latter approach, however, we focus on integrating the spatial axis of the distortion which may imply a complementary effect to these datasets.

In the blind image restoration, Cornillere et al., [13] tackled the spatially-varying degradation on super-resolution with kernel discriminator and degradation-aware network. Despite a similar motivation to ours, their approach is limited to the super-resolution task. From a viewpoint of the restoration network, controllable or flexible image restoration methods [15, 19, 36] lie in a similar line to ours. They

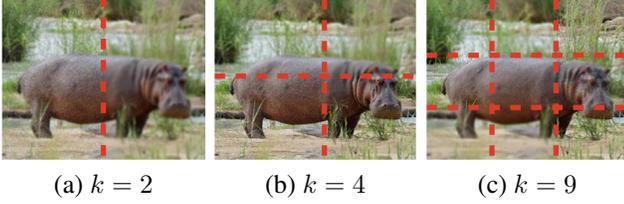


Figure 3. **Example of the dataset.** Here, we vary the dividing factor  $k$  from two to nine.

first receive (or predict in [19]) the control parameter from a user and restore an input image jointly referring to the conditional information. However, they apply the same information to the entire image, unlike our method that employs the information region-wise manner.

**Exploiting conditional information.** Recent studies have revealed that leveraging additional information improves the performance in many tasks [12, 16, 35, 38]. For example, SFT [38] employs a segmentation mask as a piece of conditional information for the super-resolution. Through this, the network can reconstruct the delicate but rich textures of a given image. In the image enhancement, HDRUNet [12] utilized the brightness and contrast information obtained by the external network according to the position of an input image. Motivated by the previous studies, in this work, we exploit the additional information to restore a image corrupted with multiple distortions.

### 3. Holistic multi-distortion dataset

In this section, we introduce a novel Holistic Multi-distortion Dataset (HMDD). Our proposed dataset integrates both sequential- [39] and spatial-aware [21] distortion natures so that the assumption of HMDD is a general form of the two regimes. To implement these to a single dataset, we first divide the image into random chunks as in below equation by following [21].

$$I_{gt} \longrightarrow \{I_{gt}^1, \dots, I_{gt}^k\}, \quad k \in \{2, 4, 9\}. \quad (1)$$

Here,  $I_{gt}$  is a clean image,  $k$  is the number of regions (randomly selected), and  $I_{gt}^1, \dots, I_{gt}^k$  are the split chunks.  $\longrightarrow$  denotes the image split process. As shown in Figure 3, a given image is divided through horizontal and vertical lines according to the number of split regions  $k$ , and these lines are randomly arranged so that the model does not memorize the position of each chunk [21].

With the chunks, we then corrupt these by the random distortions listed up in Table 1 (upper part). Here, we select the widely used distortions in image restoration literature. In addition to this, we make another dataset based on the (subset of) distortions used in Hendrycks et al., [17] and we dub this dataset as HMDD-r (below part of Table 1).

Distortion	Values
Gaussian blur	$\sigma_b \in \{0.5, 1., 1.5, 2., 2.5, 3., 3.5, 4., 4.5\}$
Gaussian noise	$\sigma_n \in \{5, 10, 15, 20, 25, 30, 35, 40, 45\}$
JPEG quality	$q \in \{80, 60, 50, 40, 35, 30, 25, 20, 15\}$
Snow	$\mu_s \in \{0.85, 0.8, 0.75, 0.7, 0.65, 0.6, 0.55\}$ $\sigma_s \in \{4, 4, 5, 5, 5, 5, 6, 6, 6\}$
F-noise	$\alpha_f \in \{500, 250, 150, 100, 80, 60, 40, 25, 15\}$
Defocused blur	$\sigma_d \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 10\}$

Table 1. **A list of the distortions and their levels in our dataset.**

Since overall dataset generation processes are identical to both HMDD and HMDD-r, here we will explain the procedure based on HMDD. To simulate the sequentially-applied corruptions [39], we select the distortions and their strength values by the following policy.

$$\begin{aligned}
 D_b(\sigma_b) &= \begin{cases} \text{Gaussian blur}(\sigma_b) & \text{if } p_b \geq 0.5 \\ \text{Identity} & \text{if } p_b < 0.5 \end{cases}, \\
 D_n(\sigma_n) &= \begin{cases} \text{Gaussian noise}(\sigma_n) & \text{if } p_n \geq 0.5 \\ \text{Identity} & \text{if } p_n < 0.5 \end{cases}, \\
 D_j(q) &= \begin{cases} \text{JPEG compression}(q) & \text{if } p_j \geq 0.5 \\ \text{Identity} & \text{if } p_j < 0.5 \end{cases}.
 \end{aligned} \quad (2)$$

Here,  $p_b, p_n, p_j$  are the (random) probability values corresponding to Gaussian blur, Gaussian noise, and JPEG compression, respectively. In this policy, if  $p_b, p_n$ , or  $p_j$  are greater than 0.5, the corresponding distortions (with their degradation factors  $\sigma_b, \sigma_n$ , and  $q$ ) are chosen. In the case where all the probability values are less than 0.5, no distortion is employed (all identity operations). Finally, we now corrupt chunks to generate distortion image  $I_d$  by Eq. 3.

$$\begin{aligned}
 I_d^i &= D^i(I_{gt}^i); \quad D^i = D_j^i \circ D_n^i \circ D_b^i, \quad \text{for } i = 1, \dots, k \\
 I_d &\leftarrow \{I_d^1, \dots, I_d^k\}.
 \end{aligned} \quad (3)$$

Here,  $D^i$  is the composite corruptions corresponding to the  $i$ -th split region,  $I_{gt}^i$ . We first distort each chunk and produce  $I_d^i$  then combine these to make the final distorted image  $I_d$ . Note that we also generate the pixel-wise distortion label  $M$  simultaneously for the purpose of training the recognition network (Section 4.1). The distortion label  $M$  includes the multi-hot embedding for the presence (since multiple distortions can appear in the same position) of the distortion along with its strength value.

Following SHDD [21], we use DIV2K [3] as clean images; 750 images from the training dataset are used to create the training set of our dataset and 50 images for the validation set. The detailed distortion strength parameters used in our dataset are listed in Table 1. For snow corruption,  $\mu_s$  determines the amount of snow and  $\sigma_s$  indicates the standard deviation of motion blur to simulate the blown snowflake.

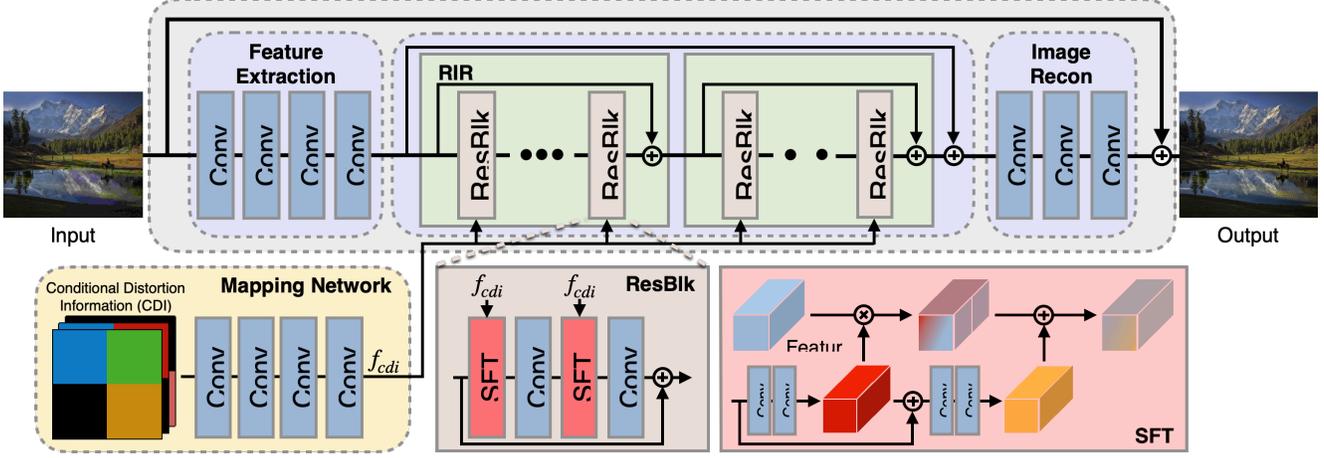


Figure 4. **Restoration module.** This network is composed of the mapping network and the reconstruction network (ReconNet). Unlike the plain reconstruction component (such as RCAN [47]), ours contains the spatial feature transform (SFT) layers [38] inside of the residual blocks to effectively combine the conditional distortion information embedding ( $f_{cdi}$ ), mapped feature of the predicted distortion by the recognition component.  $\otimes$  and  $\oplus$  symbols are element-wise multiplication and addition operations, respectively.

## 4. Method

As shown in Figure 2, our proposed restoration framework, dubbed as Distortion Information-guided Network (DIGNet), is composed of the distortion recognition module (Section 4.1) and the restoration component (Section 4.2), which reconstruct the corrupted image by following the distortion guidance from the recognition module.

### 4.1. Recognition Module

The goal of the recognition module is to recognize the distortion types and intensity from a given corrupted image. Since we assumed different distortions can be applied for each region, we formulate this as the segmentation problem. The reason for using the segmentation framework is two-fold: 1) distortions may lay pixel-wise and 2) multiple distortions can be overlapped, which become multi-label recognition. We will explain how to handle the multi-label information later in this section. The recognition network  $f_{recog}$  takes the distorted image  $x$  and predicts  $\widehat{M}$ .

$$\widehat{M} = f_{recog}(x) \quad (4)$$

We train the recognition module in a supervised manner by calculating the mean-squared error between the distortion label map  $M$  and the segmentation prediction  $\widehat{M}$ .

We design the recognition module based on UNet [31], which is the representative semantic segmentation network widely used in many vision tasks. To manage the multi-label recognition problem, we utilize distortion-wise decoders (upsampling blocks) as shown in Figure 5. We observed that naïvely using multiple segmentation networks cannot properly deal with the sequentially-applied distortions since overlapped corruptions influence each other.

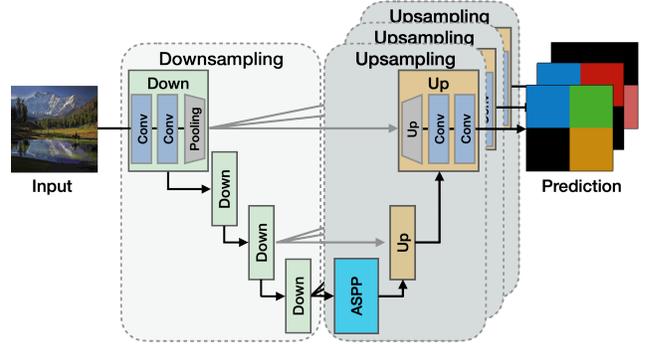


Figure 5. **Recognition module.** This component is designed based on UNet [31] with ASPP [11]. Unlike the vanilla segmentation network, we place multiple decoders (one decoder per distortion) to deal with the multi-label problem appropriately. The predictions of each upsampling block are concatenated for final product.

On the other hand, shared encoder (downsampling blocks) architecture enables the model to learn the interactions between the distortions, while the decoder outputs the distortion-specific representations. Consequently, the final output,  $\widehat{M}$  is obtained by concatenating the predicted distortion information as the following equation.

$$\begin{aligned} F_0 &= f_{enc}(X), \quad \widehat{M}^k = f_{dec}^t(F_0) \\ \widehat{M} &= [\widehat{M}^t], \quad \text{for } t = 1, \dots, T \end{aligned} \quad (5)$$

where  $T$  is the number of the distortion.  $f_{enc}/f_{dec}$  are the encoder/decoder of the recognition module and  $[\cdot]$  denotes the channel-wise concatenation operation. The predicted output of  $t$ -th distortion  $\widehat{M}^t$  is  $\mathbb{R}^{N \times H \times W \times S}$  dimension tensor with  $N$  batch size,  $H, W$  image resolution and  $S$

strength intensity, respectively. With this output design, we can manage the distortion strength apart from the distortion type. Note that we place the ASPP block [11] at the first layer of the decoder (upsampling block). ASPP is proposed to capture the multi-scale contextual information by overlapping Atrous pooling layers with different dilation rates. Because distortion can have different distributions (*i.e.*, disparate scales and shapes), passing the multi-scale features from ASPP improves the recognition performance.

**Discussion.** It is known that the deep network easily detects the distortion type [4] and we also observed a similar performance (Section 5). As a consequence, we believe that our framework can effectively recognize the distortion presented in recent synthetic dataset process [37, 44] as well. Although our method is limited to identifying the discrete distortion strength because we formulate it as a distortion segmentation, the recognition module is flexible to increasing the number of the distortion intensity since we separate the distortion type and intensity in the model output.

## 4.2. Restoration Module

The restoration module consists of two parts: the mapping and the reconstruction network (Figure 4). By using the conditional distortion information (CDI),  $\widehat{M}$ , which is predicted by the recognition module, the mapping network  $F_{map}$  embeds to produce feature  $f_{cdi}$  as follows:

$$f_{cdi} = F_{map}(\widehat{M}). \quad (6)$$

Here, we build  $F_{map}$  to have four convolution layers.

To convey the distortion information effectively, we utilize spatial feature transform (SFT) [38]. In detail, we place this layer before every convolution located inside of the residual blocks. With SFT, we modify the intermediate feature  $f_h \in R^{(C \times W \times H)}$  as in the equation below.

$$\begin{aligned} (\alpha, \beta) &= t(f_{cdi}), \\ SFT(f_h; \alpha, \beta) &= \alpha \otimes f_h + \beta, \end{aligned} \quad (7)$$

where  $t$  is the layer outputs a modulation parameter pair  $(\alpha, \beta)$  and  $\otimes$  denotes element-wise multiplication. The mapped distortion information is transmitted by scaling and shifting  $f_h$  through the modulation parameter pair  $\alpha \in R^{(C \times W \times H)}$ ,  $\beta \in R^{(C \times W \times H)}$ , respectively.

The main reconstruction network takes a corrupted image  $x$  and generates  $\hat{y}$ , the final clean counterpart. We construct the feature extraction module  $F_e(\cdot)$  to have four convolutional layers unlike the previous distortion restoration methods [7, 25, 47] that use very shallow ones. Moreover, we gradually increase the number of the channels from 3 to 256 for the first three convolution layers and distill this as a 64-dimension at the last layer of the extraction module. The reason for such a design choice is because the multi-distortion circumstance requires more representation power

Method	HMDD		HMDD-r	
	PSNR	SSIM	PSNR	SSIM
OWAN [32]	23.52	0.5948	22.25	0.5694
+ CDI	25.96	0.7323	27.13	0.7885
MEPSNet [21]	25.77	0.7257	26.08	0.7757
+ CDI	26.60	0.7606	28.43	0.8270
EDSR [25]	26.25	0.7461	26.70	0.7795
+ CDI	26.63	0.7622	28.56	0.8401
Ours w/o CDI	26.52	0.7528	27.91	0.8177
+ CDI (DIGNet)	<b>26.74</b>	<b>0.7634</b>	<b>28.70</b>	<b>0.8560</b>

Table 2. Quantitative comparison to other methods.

compared to the single distortion. Noe we extract the initial features  $f_0$  with the extraction module  $F_e$  as:  $f_0 = F_e(x)$ .

Then,  $f_0$  is passed to the feature learning module  $F_f(\cdot)$  that has two residual-in-residual (RIR) blocks [47] each of which consists of ten residual blocks. However, unlike [47], we employ the SFT layer to model the conditional distortion information. We now produce intermediate feature  $f_o$  as:

$$f_f = F_f(f_0; \alpha, \beta), \quad f_o = f_f + f_0. \quad (8)$$

Finally, we generate the cleaned image  $\hat{y}$  by using the image reconstruction module  $F_r$  as  $\hat{y} = F_r(f_o) + x$ . We train the entire restoration module with L1 loss between the ground-truth image  $y$  and the reconstructed result  $\hat{y}$ .

## 5. Experiments

**Implementation details.** In the training process, we set (batch size, patch size) = (16,  $224^2$ ) for the recognition and (16,  $48^2$ ) for the restoration module. As an optimizer, Adam [22] with a setting of  $(\beta_1, \beta_2, \epsilon) = (0.9, 0.999, 10^8)$  and weight decay as  $10^{-4}$  are used. We train the recognition and restoration components for 200K and 300K steps respectively. We use the initial learning rate  $10^{-4}$  and reduce it half at 40K and 80K steps for recognition, 50K and 100K for the restoration network.

### 5.1. Comparison to the other methods

**Baselines.** We compare our method to the following multi-degraded image restoration methods: OWAN [32], MEPSNet [21] and EDSR [25]. EDSR is the method designated to the super-resolution task, however, we observed that it also shows prominent performance in other distortion restoration tasks. We use the official code for OWAN and MEPSNet with default settings and we decrease the number of the blocks for EDSR to match the capacity to the others.

**Quantitative comparison.** Our DIGNet surpasses the other competitors on both HMDD and HMDD-r (Table 2). Compared to EDSR, which is the second-best method, our model outperforms with a 0.49dB margin in PSNR on HMDD

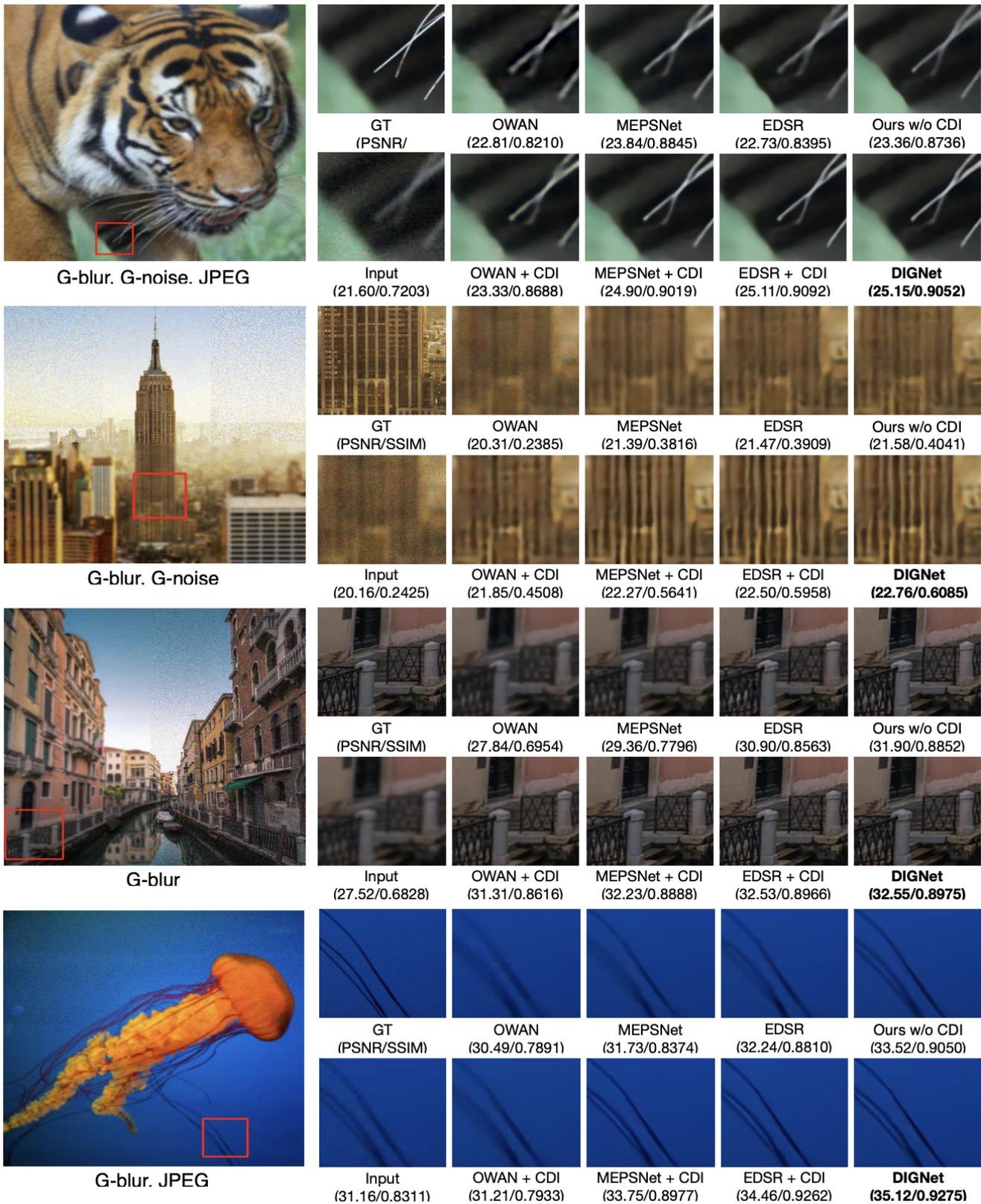


Figure 6. Qualitative comparison to the other methods on HMDD.

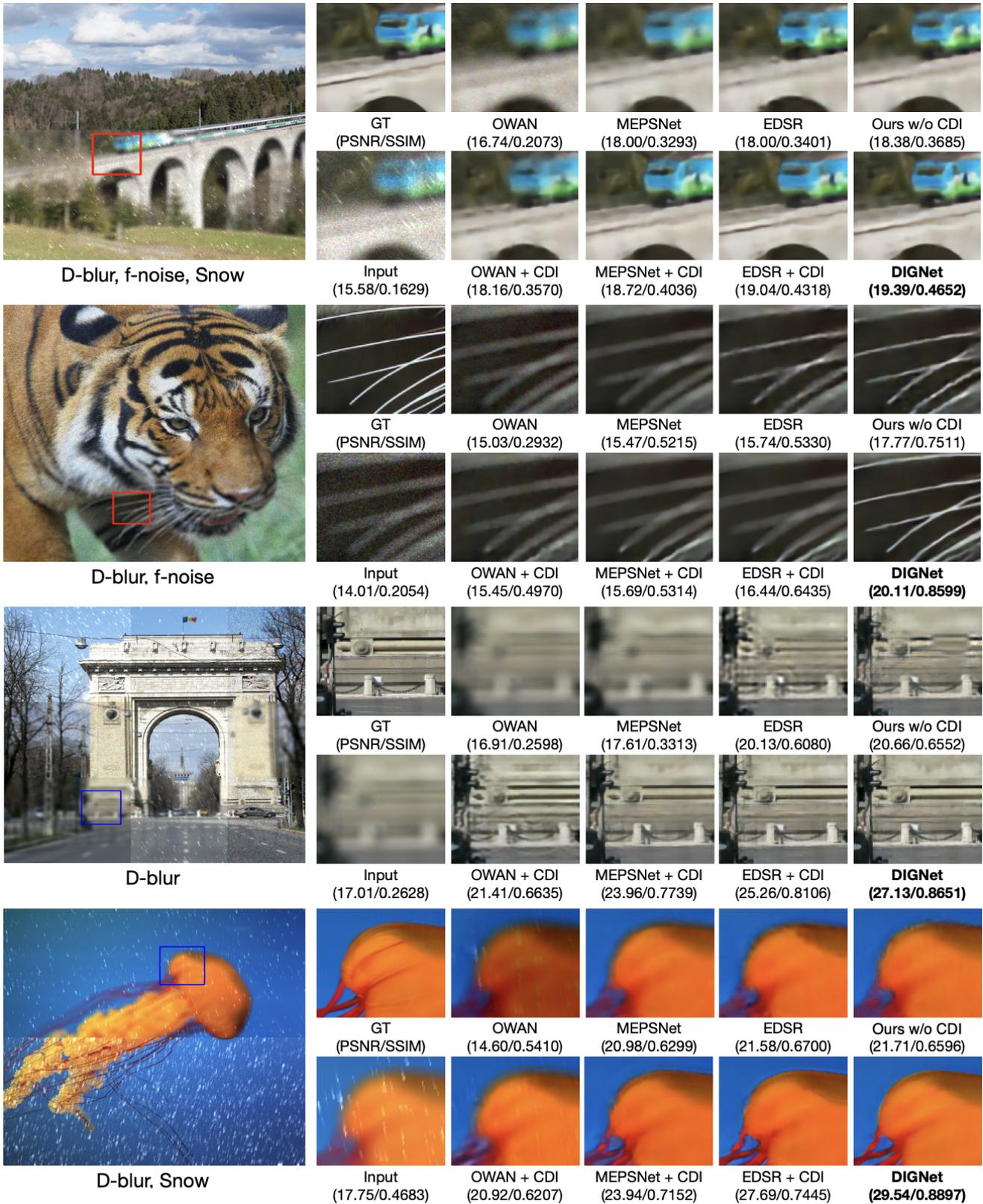


Figure 7. Qualitative comparison to the other methods on HMDD-r.

# Down.	# Up.	ASPP	Accuracy (HMDD)				→ DIGNet		HMDD-r	Accuracy
			G-blur	G-noise	JPEG	Pixel	PSNR	SSIM		
2	2		60.76	79.22	87.39	44.59	26.60	0.7559	Snow	88.36
4	2		71.72	84.26	<b>94.24</b>	60.29	26.68	0.7594	f-noise	97.40
4	2	✓	<b>79.12</b>	<b>87.93</b>	94.01	<b>64.49</b>	<b>26.74</b>	<b>0.7634</b>	D-blur	81.89
4	4	✓	78.77	85.38	93.98	63.01	26.71	0.7628	Pixel	67.49

Table 3. **Model analysis.** (Left) Ablation study on the recognition network. We compare the segmentation accuracy and the restoration performance (via DIGNet) by varying the number of the up- and down-sample layers as well as the existence of the ASPP layer. Here, G-Blur, G-Noise, and JPEG indicate pixel-wise accuracy for each distortion, and Pixel denotes the percentage of pixels that match the labels of three distortions exactly. (Right) Prediction accuracy of the best recognition network on HMDD-r.

(26.25 vs. 26.52). The performance gap is more clearly shown in the HMDD-r results; our method dominates the others with the gap of 2.00 dB in PSNR (EDSR: 26.70 vs. 28.70) which is the magnificent difference on the restoration tasks. We would like to emphasize that CDI utilization is flexible so that any backbone restoration method can adopt this. Our experiment demonstrates that with CDI adaptation, all the methods earn additional performance leap in all the datasets and metrics. In these results, one interesting observation is the small capacity networks (*e.g.* OWAN) enjoy more performance gain. We hypothesize that this is because the distortion information embedded in CDI enables the model to concentrate on the restoration task alone.

**Qualitative comparison.** Figures 6 and 7 show the qualitative results on HMDD and HMDD-r scenarios, respectively. As in the quantitative analysis, our method produces the strongest restoration capability regardless of the number of distortions applied to the image. The performance improvement of using CDI is valid for the other backbone networks as well (below rows in each method).

## 5.2. Model analysis

**Recognition network.** Here, we compare the performance of the recognition network by adjusting the number of the down/up blocks and switching the usage of ASPP [11] layer (Table 3, left). In most cases, four downsample and two up-sample blocks with ASPP work best. Interestingly, using four upsample blocks shows a lower performance. We observed that this is related to the assumption of the distortion application. Since distortions are applied region-wise, not pixel-wise, the corruptions are located unitively such as the cluster form. Therefore, when the output of the recognition module is finer (large resolution), the hole or ring shape predictions are also upscaled so as the hole to be bigger, and as a result, the performance could be degraded. In all our experiments, we use the settings of the third row in Table 3 for the recognition module. We also report the accuracy on HMDD-r (Table 3, right) using the best model on HMDD.

**Effect of CDI.** As shown in Table 3 (left), we perform the experiment on how the recognition performance affects the

restoration quality. As expected, the accuracy on HMDD is highly related to the performance of DIGNet. For example, conditional information with a 44.59 pixel accuracy results in 26.60 dB in PSNR, while DIGNet with a 64.49 pixel accuracy information achieves 26.74 dB.

## 6. Conclusion

In this study, we propose the holistic multi-distortion dataset (HMDD) and the distortion information-guided network (DIGNet) for effective multi-degraded image restoration. Our dataset integrates the two multi-distortion scenarios [21, 39] and with this, we argue that HMDD is the general form of the previous multi-degraded datasets. The core spirit of our DIGNet is providing distortion information to the reconstruction network and we expect that the restoration module is able to focus on “restoring” a given image. Several experiments show that the DIGNet outperforms the other multi-distortion image restoration methods.

One limitation is the supervised learning manner when training the distortion recognition network. Although we know which distortion may appear in advance many of the cases, there might exist potential scenarios where out-of-distribution distortions appear. To increase the generalization ability, in the future, we hope to extend our method by bridging it to the unsupervised training approach. Another possible future work is to fuse with the recent practical image restoration approach [37, 44], which builds a complex but pure synthetic distortion generation procedure. Ideally, extending our work to corrupt an image with irregular spatial patterns would encourage the restoration method to learn more robust representation in practical scenarios.

**Acknowledgement.** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2022R1A2C1007434), and also by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2018-0-01431).

## References

- [1] Abdelrahman Abdelhamed, Mahmoud Afifi, Radu Timofte, and Michael S Brown. Ntire 2020 challenge on real image denoising: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 496–497, 2020. [1](#)
- [2] 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. [2](#)
- [3] 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] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Image distortion detection using convolutional neural network. In *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 220–225. IEEE, 2017. [2](#), [5](#)
- [5] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 252–268, 2018. [2](#)
- [6] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Efficient deep neural network for photo-realistic image super-resolution. *arXiv preprint arXiv:1903.02240*, 2019. [2](#)
- [7] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3155–3164, 2019. [2](#), [5](#)
- [8] Harold C Burger, Christian J Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with bm3d? In *2012 IEEE conference on computer vision and pattern recognition*, pages 2392–2399. IEEE, 2012. [1](#), [2](#)
- [9] 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. [2](#)
- [10] Chang Chen, Zhiwei Xiong, Xinmei Tian, Zheng-Jun Zha, and Feng Wu. Camera lens super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1652–1660, 2019. [2](#)
- [11] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. [4](#), [5](#), [8](#)
- [12] Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. Hdrnet: Single image hdr reconstruction with denoising and dequantization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 354–363, 2021. [3](#)
- [13] Victor Cornillere, Abdelaziz Djelouah, Wang Yifan, Olga Sorkine-Hornung, and Christopher Schroers. Blind image super-resolution with spatially variant degradations. *ACM Transactions on Graphics (TOG)*, 38(6):1–13, 2019. [2](#)
- [14] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014. [1](#), [2](#)
- [15] Jingwen He, Chao Dong, and Yu Qiao. Interactive multi-dimension modulation with dynamic controllable residual learning for image restoration. In *European Conference on Computer Vision*, pages 53–68. Springer, 2020. [2](#)
- [16] Jingwen He, Chao Dong, and Yu Qiao. Interactive multi-dimension modulation with dynamic controllable residual learning for image restoration. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16*, pages 53–68. Springer, 2020. [3](#)
- [17] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*, 2019. [3](#)
- [18] Zihao Huang, Chao Li, Feng Duan, and Qibin Zhao. H-owan: Multi-distorted image restoration with tensor 1x1 convolution. *arXiv preprint arXiv:2001.10853*, 2020. [1](#), [2](#)
- [19] Jiayi Jiang, Kai Zhang, and Radu Timofte. Towards flexible blind jpeg artifacts removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4997–5006, 2021. [2](#), [3](#)
- [20] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. [2](#)
- [21] Sijin Kim, Namhyuk Ahn, and Kyung-Ah Sohn. Restoring spatially-heterogeneous distortions using mixture of experts network. In *Proceedings of the Asian Conference on Computer Vision*, 2020. [1](#), [2](#), [3](#), [5](#), [8](#)
- [22] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. [5](#)
- [23] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018. [2](#)
- [24] 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. [1](#), [2](#)
- [25] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. [2](#), [5](#)
- [26] Xing Liu, Masanori Suganuma, Xiyang Luo, and Takayuki Okatani. Restoring images with unknown degradation factors by recurrent use of a multi-branch network. *arXiv preprint arXiv:1907.04508*, 2019. [2](#)
- [27] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Ntire 2020 challenge on real-world image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 494–495, 2020. [1](#)

- [28] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017. [2](#)
- [29] Seungjun Nah, Sanghyun Son, Suyoung Lee, Radu Timofte, and Kyoung Mu Lee. Ntire 2021 challenge on image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 149–165, 2021. [1](#)
- [30] 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. [2](#)
- [31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. [1](#), [4](#)
- [32] Masanori Suganuma, Xing Liu, and Takayuki Okatani. Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9039–9048, 2019. [1](#), [2](#), [5](#)
- [33] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8174–8182, 2018. [2](#)
- [34] 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. [1](#)
- [35] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10581–10590, 2021. [3](#)
- [36] Wei Wang, Ruiming Guo, Yapeng Tian, and Wenming Yang. Cfsnet: Toward a controllable feature space for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4140–4149, 2019. [2](#)
- [37] 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. [2](#), [5](#), [8](#)
- [38] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 606–615, 2018. [1](#), [3](#), [4](#), [5](#)
- [39] 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. [1](#), [2](#), [3](#), [8](#)
- [40] Ke Yu, Xintao Wang, Chao Dong, Xiaoou Tang, and Chen Change Loy. Path-restore: Learning network path selection for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. [1](#), [2](#)
- [41] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. *arXiv preprint arXiv:2111.09881*, 2021. [2](#)
- [42] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 492–511. Springer, 2020. [2](#)
- [43] Jiawei Zhang, Jinshan Pan, Jimmy Ren, Yibing Song, Linchao Bao, Rynson WH Lau, and Ming-Hsuan Yang. Dynamic scene deblurring using spatially variant recurrent neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2521–2529, 2018. [2](#)
- [44] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4791–4800, 2021. [2](#), [5](#), [8](#)
- [45] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2737–2746, 2020. [2](#)
- [46] 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. [2](#)
- [47] Yulun Zhang, Kungpeng 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. [2](#), [4](#), [5](#)