This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Towards a Universal Image Degradation Model via Content-Degradation Disentanglement

Wenbo Yang
University of Waterloo
w243yang@uwaterloo.ca

Zhongling Wang*
AI Center-Toronto, Samsung Electronics
z.wang2@samsung.com

Zhou Wang
University of Waterloo
zhou.wang@uwaterloo.ca

Abstract

*Image degradation synthesis is highly desirable in a wide variety of applications ranging from image restoration to simulating artistic effects. Existing models are designed to generate one specific or a narrow set of degradations, which often require user-provided degradation parameters. As a result, they lack the generalizability to synthesize degradations beyond their initial design or adapt to other applications. Here we propose the **first** universal degradation model that can synthesize a broad spectrum of complex and realistic degradations containing both homogeneous (global) and inhomogeneous (spatially varying) components. Our model automatically extracts and disentangles homogeneous and inhomogeneous degradation features, which are later used for degradation synthesis without user intervention. A disentangle-by-compression method is proposed to separate degradation information from images. Two novel modules for extracting and incorporating inhomogeneous degradations are created to model inhomogeneous components in complex degradations. We demonstrate the model's accuracy and adaptability in film-grain simulation and blind image restoration tasks. The demo video, code, and dataset of this project will be released at Yang's GitHub account .*

1. Introduction

Image degradation synthesis plays an important role in numerous applications across multiple domains. In image restoration, it is often used indirectly to generate pristine-distorted training pairs for supervised learning [22, 32, 46, 50], while some approaches incorporate it directly as an integral module within restoration frameworks [1, 8, 30, 41, 42, 48]. Beyond restoration, accurate degradation synthesis also plays a significant role in artistic applications. For example, cinematographers rely on precise film grain synthesis tor enhance viewer experience [29], and efficient haze and rain synthesis offers a computationally efficient alternative to complex physical simulations in computer graphics.

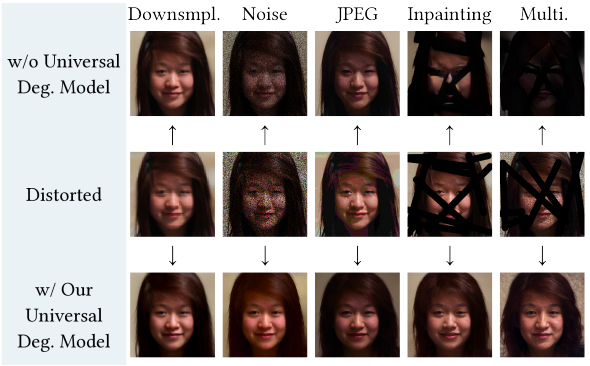


Figure 1. Demonstration of the applicability of our universal image degradation model in GAN inversion-based image restoration. When a fixed degradation model is used, only the modeled degradation (downsampling, first column) receives high-quality results. However, when our universal degradation model is used, all results are of pleasant quality. Best to be viewed zoomed in.

However, existing degradation models, whether handcrafted, statistical, or deep learning-based, are predominantly designed for specific degradation types or narrow degradation sets, as summarized in Table 1. These models often employ restrictive design choices to optimize performance for particular degradations, sacrificing generalizability across diverse scenarios. Consequently, they cannot be adapted to synthesize degradations beyond their original design. However, in many downstream applications, the ability to synthesize multiple types of degradation is essential. A set of visual examples in image restoration is given in Fig. 1. Furthermore, most existing models necessitate user-specified degradation parameters, limiting their applicability in contexts where frequent user input is impractical, such as film grain transfer or blind image restoration. While some degradation models can estimate the degradation parameters, they only apply to simple, well-defined degradations, failing to extend to complex real-world degra-

*Work done at the University of Waterloo.

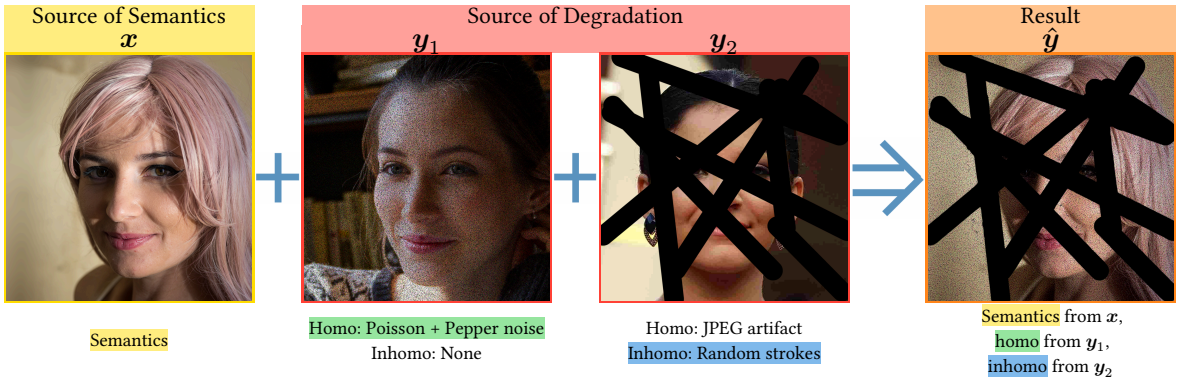


Figure 2. Visualization of our models' degradation disentanglement and transfer ability. Our model can extract homogeneous and inhomogeneous degradation information from the same or different distorted image(s) and reapply it to a pristine image.

dations that are usually hard to describe parametrically. Furthermore, due to the intrinsic design limitations, current methods cannot synthesize *inhomogeneous* degradations that vary spatially, a common characteristic of real-world distortions (examples are available in Supp. Sec. 7). Among existing works, only Chen et al. [12] attempted to develop a model without degradation-specific design. However, their approach still requires training separate models for each distortion due to the lack of degradation representation and cannot synthesize inhomogeneous degradations.

Therefore, it is desirable to design a *universal* degradation model that can generalize to *complex* (combinations of) degradations, both homogeneous and *inhomogeneous*. However, achieving this is challenging for two main reasons. First, complex degradations are harder to separate from degraded images than simpler ones. Simple degradation models often assume linear (or affine) and content-independent degradation processes. However, real-world complex degradations are typically nonlinear and content-dependent, complicating the extraction of pure degradation information free of image content. Second, extracting and synthesizing inhomogeneous degradations presents additional challenges, as these spatially varying effects are more difficult to model than homogeneous ones.

In this paper, we present the first *universal image degradation model* capable of encoding and synthesizing complex combinations of homogeneous and inhomogeneous image degradations. As demonstrated in Figure 2, our model effectively separate degradation information from distorted images' contents, disentangle inhomogeneous and homogeneous degradations, and transfer degradations to other pristine images. To encourage the separation of degradation information from image content, we propose a novel *disentangle-by-compression* method to regularize the model behavior. To handle inhomogeneous degradations, we further propose two network modules, namely Inhomogeneous Degradation Embedding Network (IDEN) and Inhomogeneous Degradation Aware (IDA) network layer, which empowers the encoding and decoding of inhomogeneous degradations, respectively. Our model advances explainable degradation modeling and enables detailed analysis of degradation components. It also finds applications in a wide range of image restoration tasks and in the visual content generation and distribution industry.

We demonstrate our model's effectiveness through extensive ablation studies and in downstream applications such as film grain simulation and blind image restoration. For the latter, our model can be used as a drop-in replacement for the degradation synthesis module in generative model inversion-based image restoration methods for any type of image contents (as long as the inverted generative model supports), converting existing non-blind image restoration models into blind ones. Blind models do not require any degradation information from the user and have broader use cases in real-world applications, especially in cases where the degradation is complex and hard to describe parametrically. In summary, our key contributions are:

- We propose the first universal degradation model capable of handling complex combinations of homogeneous and inhomogeneous degradations.
- We introduce a novel disentangle-by-compression method that separates degradation information from distorted images' contents and encourages the components in degradation embedding to be independent.
- When used as a drop-in module for inversion-based image restoration, it enables, for the first time, generative model inversion in a blind setting to handle unknown and complex distortions.

2. Related Works

Table 1 summarizes representative degradation models. Most existing models focus on a single distortion type, while some handle basic fixed combinations of distortions. For instance, some super-resolution models [36, 44] con-

Degradation Type		Model	Mdl Type
Homogeneous	Noise	[50],[40],[46],[27],[18],[48]	Stat-based
		CA-NoiseGAN [11],Noise Flow [1],DANet [42]	Learned
	Downsample + Noise	RealSR [22]	Learned
	Rain	[32]	Learned
	Old Photo	[26]	Stat-based
	Grain	[41]	Stat-based
		[4],[2]	Learned
	Various single	[12]	Learned
Inhomo-	Haze	[21], [31], [49], [5]	Stat-based
		DehazeNet [8], AOD-Net [25]	Hybrid
Both	**Complex combination**	**Ours**	**Learned**

Table 1. A review of representative degradation models. Most of the models focus on single homogeneous distortion.

siders common degradations in the imaging pipeline such as blur, downsampling, noise and compression artifacts. However, each degradation component is modeled independently. Some components are handcrafted, requiring strong domain knowledge specific to the downstream task (super-resolution in this case). These degradation models cannot be easily transferred to other degradation types.

Attempting to model all possible degradations individually is impractical, highlighting the need for a universal degradation model that can generalize across degradations without requiring domain-specific knowledge from downstream applications. However, constructing such a model is challenging: complex combinations of distortions are typically intractable for explicit modeling and have not yet been effectively simulated through machine learning-based approaches. To our knowledge, only [12] attempted to construct a non-distortion-specific image degradation model. However, their approach still requires training separate models for each distortion type, as it does not accept degradation information as input. Furthermore, they did not consider the stochastic effects of image distortions, particularly for image noise, resulting in similar generated noise patterns across images.

Another major limitation of most existing degradation models is their assumption that degradations are homogeneous across the entire image. This simplification often fails to capture the complexity of real-world scenarios, where degradations can vary significantly across different regions of an image. Some haze simulation methods attempt to address this issue by using a depth map to model inhomogeneous haze. However, this approach is specific to haze and cannot be generalized to other types of degradations.

Another related area is Style Transfer (ST), which aims to impose the artistic style of a source image to the content of a target image. However, most ST methods require training a separate model for each specific artistic style. A few



Figure 3. Visual results for StyTr2 [16] retrained on the distortion dataset. It fails at separating distortion from content, hence, alters skin tone and texture while not transferring distortions (more examples in Supp. Sec. 3.5).

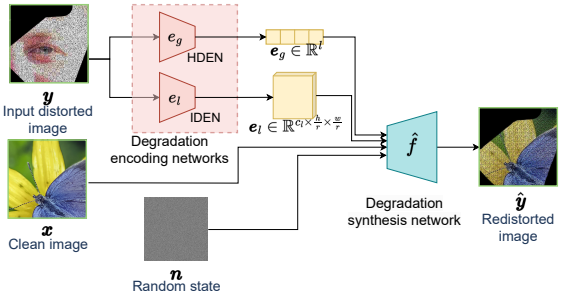


Figure 4. The overall architecture of our degradation encoding and decoding network (during the test phase).

SOTA methods handle multiple styles by extracting style information from a separate input image [9, 16, 23]. Nevertheless, image degradations exhibit distinct characteristics from artistic styles, necessitating specialized architecture for effective modeling. As a result, SOTA ST models (e.g., StyTr2 [16]) exhibit significant limitations (Fig. 3): they create undesirable changes to image content (e.g., skin tone), are poor at reproducing realistic noise patterns, are incapable of segregating homogeneous and inhomogeneous degradations, and fail at handling inhomogeneous degradations due to architectural limitations.

3. Proposed Method

We note that all distorted process can be described by the following formulation:

$$\mathbf{y} = f(\mathbf{x}, \mathbf{d}, \mathbf{n}), \tag{1}$$

where $\mathbf{x}$ is a pristine image, $\mathbf{d}$ is the degradation information, $\mathbf{n}$ is a random state, and f is an operator synthesizing distortions based on the information. Due to the complexity of real-life degradations, it is intractable to explicitly construct a statistical model. Hence, end-to-end learned degradation encoding and decoding Convolutional Neural Networks (CNNs) are necessary. Figure 4 shows an overview of our framework. A network $\hat{f}$ is trained to simulate f, which degrades a pristine image $\mathbf{x}$ with the degradation extracted by a homogeneous and inhomogeneous degradation

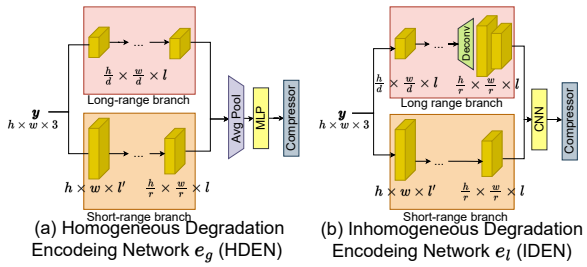


Figure 5. The architecture of the two Degradation Encoding Networks: HDEN and IDEN.

encoding networks e_g and e_l, respectively:

$$\hat{\mathbf{y}} := \hat{f}\left(\mathbf{x}, \mathbf{e}_g, \mathbf{e}_l, \mathbf{n}\right), \tag{2}$$

where $\mathbf{e}_g := e_g(\mathbf{y})$, and $\mathbf{e}_l := e_l(\mathbf{y})$. With our novel training strategy, our model learns to transfer degradation from one image $\mathbf{y}^{(1)}$ (or images $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$) to another image $\mathbf{x}^{(0)}$ without being supervised on the triplet $\left(\mathbf{y}^{(1)}, \mathbf{x}^{(0)}, \mathbf{y}^{(0)}\right)$ (or $\left(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \mathbf{x}^{(0)}, \mathbf{y}^{(0)}\right)$). Detailed explanations of $e_g, e_l, \hat{f}$ and training strategies are provided in the following sections.

3.1. Homo-/inhomogeneous degradation encoding

While most works ignore the spatial variance of image degradations, these inhomogeneous degradations occupy a large portion of real-world degradations. Hence, our model employs two degradation encoding networks: the *homogeneous degradation embedding network* (HDEN) e_g and the *inhomogeneous degradation embedding network* (IDEN) e_l. The homogeneous degradation embedding (HDE) $\mathbf{e}_g$ captures spatial-invariant degradation information, and the inhomogeneous degradation embedding (IDE) $\mathbf{e}_l$ retains spatial structure for local degradations.

As illustrated in Fig. 5, the HDEN has a dual-branch architecture. To accommodate degradations requiring different receptive fields (e.g., blur vs. noise), we incorporate (a) a short-range branch operating at the input image scale and (b) a long-range branch operating on a downsampled (by a factor d) resolution. This design, inspired by Zhou et al. [47]'s encoder, can effectively capture globally consistent degradation's characteristics. Additionally, to capture inhomogeneous degradations, we introduce IDEN, which has a modified long-range branch and a tail module that retains spatial structure. The post-processing Multi-layer perception (MLP) in IDEN is replaced with a small CNN to efficiently combine long-range and short-range information on the spatial domain.

3.2. Homo-/inhomogeneous degradation synthesis

Our degradation synthesis network has a U-Net structure [24] consisting of several IDA-SFT blocks, as shown in Figure 6. The global degradation embedding $\mathbf{e}_g$ and the local degradation embedding $\mathbf{e}_l$ are first combined into an intermediate degradation map $\mathbf{e}$, which then conditions each IDA-SFT block to degrade the input image $\mathbf{x}$.

We identify providing a smooth channel of incorporating local degradation as a practical requirement for disentangling local degradation information from global. Although performing convolution using a spatially-varying kernel is an ideal way of providing such channel, its space complexity is prohibitively expensive. The main cost is attributed to the inter-channel dependencies, while spatial dependencies are mainly handled by depthwise components. To reduce the computation cost, we propose an inhomogeneous degradation-aware (IDA) layer:

$$\mathrm{IDA}(\mathbf{F}_{\text{in}}, \mathbf{e}) = \mathrm{DConv}\left(\mathrm{DS}(\mathrm{DConv}(\mathbf{e})) \odot \mathrm{DS}(\mathbf{F}_{\text{in}})\right), \tag{3}$$

where DS is a downsampling operation achieved by convolution with stride and DConv is the deconvolution operation [43] that can upsample a latent. We have proved in Supp. Sec. 6.1 that a single IDA layer is more expressive and flexible than the combination of four depthwise convolution layers with spatially varying kernel.

The IDA layer excels at simulating inhomogeneous degradations for our application. We combine it in parallel with a Spatial Feature Transform (SFT) layer [35], which is good at introducing random states to latents and generating homogeneous degradations, to form the IDA-SFT layer:

$$\text{IDA-SFT}(\mathbf{F}_{\text{in}}, \mathbf{e}, \mathbf{n}) = \mathrm{IDA}(\mathbf{F}_{\text{in}}, \mathbf{e}) + \underbrace{\alpha(\mathbf{e}) \odot \mathbf{F}_{\text{in}} + \beta(\mathbf{e}, \mathbf{n})}_{\text{Spatial Feature Transform}}, \tag{4}$$

where α and β are two small MLPs.

3.3. Degradation disentanglement

Our model achieves three key disentangling effects: separating (1) degradation from distorted images contents, (2) inhomogeneous from homogeneous distortions, and (3) individual degradation components. These effects enable degradation transfer between images and direct manipulation of degradation characteristics. We attribute this to our novel disentangle-by-compression approach, implemented through **entropy regularization losses**:

$$\mathcal{L}_{\text{rate_g}} = \sum_i H\left(e_g^{(i)}\right) = -\sum_i \mathbb{E} \log_2 p\left(e_g^{(i)}\right),$$

$$\mathcal{L}_{\text{rate_l}} = \sum_{i,j} H\left(e_l^{(i,j)}\right) = -\sum_{i,j} \mathbb{E} \log_2 p\left(e_l^{(i,j)}\right),$$

where p denotes the probability density function. The density $p\left(e_g^{(i)}\right)$ and $p\left(e_l^{(i,j)}\right)$ are estimated using [7] and [6], respectively. Due to the large sample size used in the estimation of p, its estimation is accurate enough (while $p\left(\mathbf{e}_g\right)$

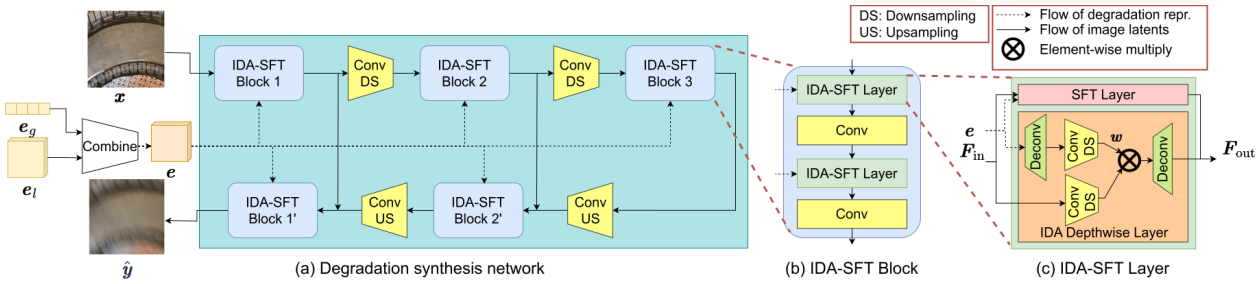


Figure 6. The degradation synthesis network.

cannot be reliably estimated, due to its high dimensionality).

We shall justify the entropy regularization loss's effects in remainder of this section. Since $\mathbf{e}_l$ and $\mathbf{e}_g$ are similar in nature, we shall drop the subscript until the discussion of inhomogeneous and homogeneous disentanglement. The sum of entropy $\mathbf{e}$ can be expanded as [39]

$$\sum_i H\left(e^{(i)}\right) = \underbrace{H\left(\mathbf{e}\right)}_{\text{separate degrad. from img}} + \underbrace{D_{\mathrm{KL}}\left(p\left(\mathbf{e}\right) \parallel q\left(\mathbf{e}\right)\right)}_{\text{enforce entries' independence}}, \tag{5}$$

where $q\left(\mathbf{e}\right) := \prod_i p\left(e^{(i)}\right)$. By optimizing the left-hand-side of Eq. (5), we are simultaneously minimizing the entropy of $\mathbf{e}$ and the KL-divergence of its distribution towards q, which has the same marginal distribution as $p\left(\mathbf{e}\right)$ but with independent entries. The latter makes the entries in $\mathbf{e}$ independent.

The following mild assumptions are made: (**A-1**) With information about the degradation process and the clean image, the degraded image can be exactly reconstructed, albeit the random state difference. (**A-2**) The distribution of distortion processes is independent of the clean images'. (**A-3**) The distortion can be inferred from the degraded image. (**A-4**) When an appropriate **perceptual similarity loss** $\mathcal{L}_{\text{sim}} = d\left(\mathbf{y}, \hat{\mathbf{y}}\right)$ is used, the reconstructed distorted image $\hat{\mathbf{y}}$ will be be similar enough to $\mathbf{y}$. (See Supp. Sec. 2.2 for detailed justifications.) With these assumptions, we are able to rigorously prove (in Supp. Sec. 6.2) that

$$H\left(\mathbf{e}\right) = I\left(\mathbf{e}; \mathbf{x}\right) + H\left(\mathbf{d}\right). \tag{6}$$

Since $H\left(\mathbf{d}\right)$ is a fixed (but unknown) constant, by minimizing $H\left(\mathbf{e}\right)$, the mutual information $I\left(\mathbf{e}; \mathbf{x}\right)$ is minimized, which encourages $\mathbf{e}$ to contain no information in $\mathbf{x}$ (which contains all content information), disentangling the degradation from image content.

We further observe that there exists two kinds of degradations: the homogeneous one $\mathbf{d}_g$ is the same for all geometry location, and the inhomogeneous one $\mathbf{d}_l^{(i)}$ is different for different geometry location (i). Based on our observations, we assume (**A-5**) they are jointly independent. Then, we show in the supplementary materials that, with a flexible enough network (and λ_l and λ_g satisfying mild conditions), the optimal solution for $\lambda_g \mathcal{L}_{\text{rate_g}} + \lambda_l \mathcal{L}_{\text{rate_l}}$ guarantees that $\mathbf{e}_l$ only contains spatially dependent information, and $\mathbf{e}_g$ only contains homogeneous distortion information.

3.4. Loss functions

We choose DISTS [17] as our perceptual distance measure d since it has the least unnecessary sensitivity to noise's random state (which cannot be perceived by human observers, details are in Supp. Sec. 6.2). To ensure diverse outputs, we use SSIM [37] as the diversity loss due to its sensitivity to random states:

$$\mathcal{L}_{\text{diver}} = -\text{SSIM}(\hat{\mathbf{y}}, \hat{\mathbf{y}}')$$

where $\hat{\mathbf{y}}$ and $\hat{\mathbf{y}}'$ are generated using the same inputs but different random states. To address the limitations of general-purpose IQAs in capturing human-perceived fidelity, we incorporate an adversarial loss $\mathcal{L}_{\text{gan}}$ for more realistic degradation generation. Following conventions, our total loss function also incorporates a contrastive loss $\mathcal{L}_{\text{contra}}$ [33] and color loss $\mathcal{L}_{\text{color}}$ [19]. Our total loss function is

$$\begin{aligned}\mathcal{L} := \quad & \mathcal{L}_{\text{sim}} + \lambda_g \mathcal{L}_{\text{rate_g}} + \lambda_l \mathcal{L}_{\text{rate_l}} + \lambda_c \mathcal{L}_{\text{contra}} \\ & + \lambda_r \mathcal{L}_{\text{color}} + \lambda_d \mathcal{L}_{\text{diver}} + \lambda_g \mathcal{L}_{\text{gan}},\end{aligned} \tag{7}$$

where all $\lambda_\cdot$'s are trade-off weights.

4. Experiments

4.1. Degradation reproduction and transfer

Our model is the first to encode, synthesize, and transfer arbitrary degradation combinations from existing degraded images. It is designed to train on paired clean-distorted datasets with any type of degradation. However, the diversity of degradations and content in current datasets remains limited. Additionally, real-world datasets often lack ground-truth references, making it challenging to benchmark degradation transfer performance. To address this, we supplement real-world datasets [28, 34] with synthetic datasets inspired by real-world scenarios for quantitative

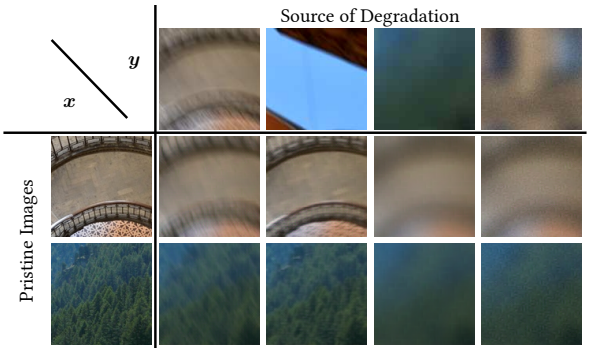


Figure 7. Synthetic distortion transfer. Each grid is a synthesized degraded image, generated by distorting a pristine image $\mathbf{x}$ with degradation from $\mathbf{y}$. Best to be viewed zoomed in.

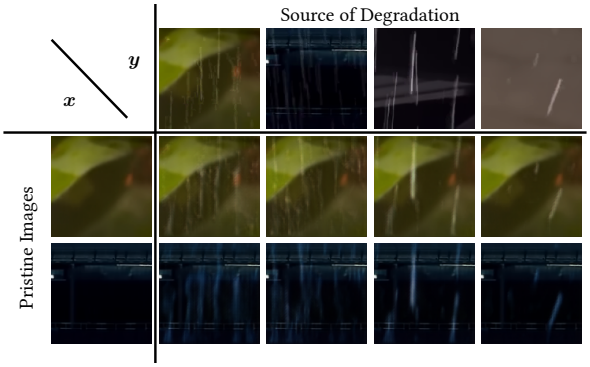


Figure 8. Realistic distortion transfer on the raindrop dataset [34].

evaluations and more challenging tasks. We believe our model can serve as a baseline for degradation synthesis and transfer, encouraging the development of more comprehensive paired real-world degradation datasets.

To test our model's generalizability to image semantics, we curated a dataset of 300K Wikimedia Quality Images (WQIs), each of which is inspected by Wikipedia or its sister projects' editors for high technical and aesthetic quality [20]. From this, we generated 40K training pairs using about 13K images in the dataset. Each pair consists of a pristine image $\mathbf{x}$ and a distorted image $\mathbf{y}$, generated by distorting $\mathbf{x}$ with distortions randomly selected from a pool of degradation combinations designed according to typical image processing pipelines. Details of the dataset and training settings are in Supp. Sec. 4. Figure 7 showcases our model's ability to reproduce and transfer various degradations (e.g., Gaussian noise, blur, motion blur, JPEG artifacts) and their combinations across different images. Unlike [12], our model generates diverse noise patterns, demonstrating both effective representation learning and output diversity.

By analyzing variances, we identified five active dimensions in $\mathbf{e}_g$ for the model trained on the distorted WQI dataset. The model learns to disentangle degradation components and enables human to find interpretable meaning of each active latent dimension.

	MS-SSIM ↑	SSIM↑	LPIPS ↓	DISTS ↓
Reproduce	0.879	0.860	0.295	0.141
Transfer	0.875	0.856	0.306	0.147

Table 2. Quantitative evaluation results of our model in reproducing image degradations.

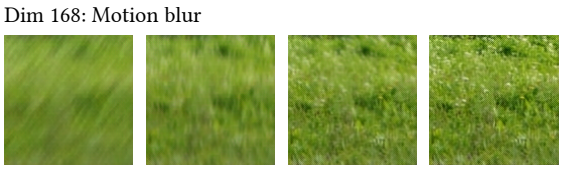


Figure 9. Demonstration of degradation latent disentangling by varying single entry in $\mathbf{e}_g$ on a selected dimension.

Authentic image degradations often include inhomogeneous elements that cannot be easily modeled by simple statistical models; for example, real-world raining effects [34]. As shown in Figure 8, our model successfully reproduces and transfers these complex, authentic distortions. Extra tests on the GoPro dataset [28] is in Supp. Sec. 3.2.

Since this is the first universal image degradation model, it is difficult to find comparable models. Comprehensive ablation studies are available in Section 4.5. To quantitatively evaluate our model's degradation reproduction and transfer capabilities, we created a test set of 2,000 image triplets $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)}, \mathbf{y}^{(1)})$ from 4,000 pristine WQIs that were not used for training. Here, $\mathbf{x}^{(0)}$ is the pristine version of $\mathbf{y}^{(0)}$, and $\mathbf{y}^{(1)}$ is generated from a different pristine image using the same degradation as in $\mathbf{y}^{(0)}$. The reproduction score is calculated as $d\left(\mathbf{y}^{(0)}, \hat{f}\left(\mathbf{x}^{(0)}, \mathbf{e}_g^{(0)}, \mathbf{e}_l^{(0)}\right)\right)$, and the transfer score is calculated as $d\left(\mathbf{y}^{(0)}, \hat{f}\left(\mathbf{x}^{(0)}, \mathbf{e}_g^{(1)}, \mathbf{e}_l^{(1)}\right)\right)$, where $\mathbf{e}_\cdot^{(i)} = e_\cdot\left(\mathbf{y}^{(i)}\right)$. Results in Table 2 demonstrate the model's high accuracy in degradation reproduction and transfer, indicating effective separation of degradation information from distorted images.

4.2. Degradation latent disentanglement

For demonstration, we modified each active dimension of the degradation embedding for a pristine test image. Figure 9 shows visual results for one of the active dimensions, with additional examples available in the Supp. Sec. 3.3. Most degradation embedding dimensions control a group of degradations, aligning with our understanding of natural biases and masking effects [38] in the dataset.

4.3. Film grain encoding and transfer

It is a common practice in the film-making industry to simulate film grains for digital movies, as directors and produc-

	MS-SSIM ↑	SSIM↑	LPIPS ↓	DISTS ↓
Reproduce	0.869	0.726	0.209	0.051
Transfer	0.872	0.728	0.209	0.053

Table 3. Quantitative evaluation results of our model in reproducing film grain.

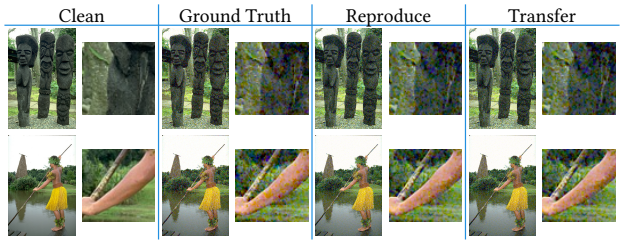


Figure 10. The result of film grain synthesis.

	Accur. ↓		Fidelity ↓		Realism ↓	
	w/o Ours	w/ Ours	w/o Ours	w/ Ours	w/o Ours	w/ Ours
2 degradations						
NA	.680	**.513**	.424	**.334**	228.8	**28.1**
AP	.506	**.463**	.232	**.219**	45.1	**16.9**
UA	.519	**.512**	**.274**	.275	47.8	**30.0**
NP	.713	**.485**	.187	**.081**	221.9	**20.6**
UN	.700	**.564**	.273	**.177**	211.2	**41.4**
UP	.541	**.478**	.130	**.097**	57.7	**38.8**
3 degradations						
UNP	.715	**.590**	.156	**.082**	215.6	**60.9**
UAP	.588	**.526**	.169	**.115**	80.5	**34.8**
UNA	.669	**.569**	.407	**.298**	175.5	**34.5**
NAP	.686	**.508**	.282	**.182**	213.7	**25.3**
4 degradations						
UNAP	.666	**.560**	.251	**.141**	147.4	**31.7**

Table 4. Accuracy (LPIPS [45]), fidelity (LPIPS [45]) and realism (pFID [10]) comparison of the same image restoration method [30] under blind settings w/o and w/ our universal degradation model on the restoration of images degraded by complex combinations of distortions. The method that has the best score in each category is shown in **bold fonts**.

ers consider it essential for creating cinematic atmosphere and immersive viewing experiences [29]. We fine-tune our model on the FilmGrainStyle [3] dataset and evaluate it using their test set, both containing professionally inspected grainy images. The film industry has two primary requirements for film grain synthesis (a) encoding and reproducing film grains from grainy images to grain-free versions, and (b) to transfer realistic film grain from existing images to new ones. Table 3 shows quantitative results of both tasks with visuals in Figure 10, from which realistic film grain patterns can be observed. To evaluate transfer performance, we simulate real application scenarios by transferring grains from a training image (with similar grain style as indicated by the test set's label) to a grain-free test image, and compare it with the ground-truth grainy image. The comparable performance between transfer and reproduction scores demonstrates our model's effectiveness in disentangling degradation information. Notably, despite being designed as a universal model, our approach outperforms the specialized baseline model from [3] in reproduction scores.

4.4. Inversion-based image restoration

A major limitation of existing Inversion-Based Image Restoration (IBIR) methods is their reliance on an impractical *non-blind* assumption: degradation parameters must be known during inference [13, 14, 30], an assumption rarely aligns with real-world scenarios. To address this, our model can be plugged into existing non-blind IBIR methods, converting them into blind restoration approaches. We upgraded two state-of-the-art non-blind restoration methods with our model: Robust StyleGAN Inversion (RSG) [30] and Diffusion Posterior Sampling (DPS) [13]. In this section, we focus on RSG due to its simplicity and lower computational cost. While our computational resources limit the scale of experiments for DPS, qualitative results are provided in the Supp. Sec. 5.4.

A generative model $G : \mathbf{w} \rightarrow \mathbf{x}$ maps a randomly sampled input $\mathbf{w}$ from a known distribution $\mathcal{W}$ to a generated image $\mathbf{x}$. For IBIR tasks, the output high-quality image $G(\mathbf{w})$ must be re-distorted for comparison. In particular, Poirier-Ginter and Lalonde [30] propose an optimization method to empirically solve

$$\min_{\mathbf{w}} \text{LPIPS}\left(\mathbf{y}, D_{\theta_d}\left(G\left(\mathbf{w}\right)\right)\right), \tag{8}$$

where D_{θ_d} is the continuous approximation of the degradation operation parametered by the already recorded θ_d. To make it a *blind* method, we replace D_{θ_d} with our model:

$$\min_{\mathbf{w},\mathbf{e}_g,\mathbf{e}_l} \mathbb{E}_{\mathbf{n}}\text{LPIPS}\left(\mathbf{y}, \hat{f}\left(G\left(\mathbf{w}\right), \mathbf{e}_g, \mathbf{e}_l, \mathbf{n}\right)\right), \tag{9}$$

where $\hat{f}$ is our degradation synthesis module, and $\mathbf{e}_g$ and $\mathbf{e}_l$ are predicted by our model from $\mathbf{y}$.

We benchmark the augmented method on the categories and the test set (FFHQ-X) proposed in [30]. The test set consists of face images, due to the underlying GAN model's limitations. However, our approach can be applied to any generative model and image content (results using DPS [13] on ImageNet [15] are available in the Supp. Sec. 5.4). For fair comparison, we only show performance against the only blind GAN inversion-based IR method (*i.e.*, without re-degradation). The benchmarked categories include single degradations (**U**psampling, de**N**oising, de**A**rtifacting, and in**P**ainting) and their combinations. Figure 11 shows examples of restoration results from our model and the naive blind restoration method. Without utilizing our model for redegradation, the result produced by blind GAN-inversion method is visually unacceptable. The quantitative

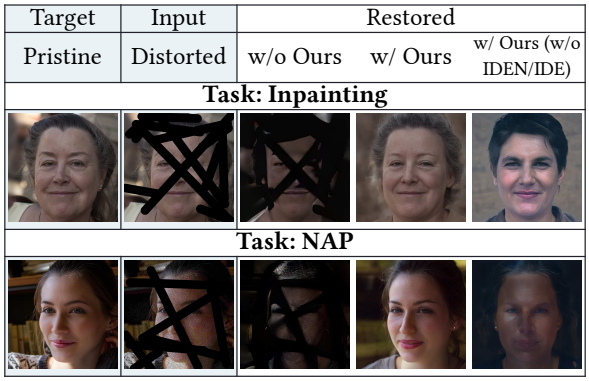


Figure 11. Image restoration results for a single/complex combination of homo- and inhomogeneous degradations. The naive blind GAN inversion method, without universal degradation modeling, produces unsatisfactory results. In contrast, with our universal degradation model, the otherwise identical model achieves significantly improved results, while remaining fully blind.

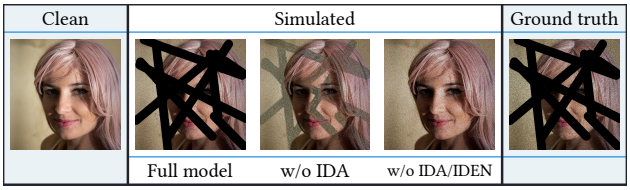


Figure 12. Examples of degradation transfer from mixed sources. Without the IDA layer or IDEN, the model is unable to properly reproduce inhomogeneous degradation. Inputs to the models $(\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2)$ are shown in Fig. 2.

results are shown in Tab. 4. Due to space limitations, we only present results for multi-degradation restoration; the full table is available in the Supp. Sec. 5.3. When compared under the same conditions, our model outperforms the vanilla blind restoration method in accuracy, fidelity, and in realism for most tasks.

4.5. Ablation Studies

We construct a test dataset from FFHQ-X [30] using their proposed degradations to verify our newly introduced modules. For each pristine image $\mathbf{x}^{(0)}$, we compute the following transfer scores:

- Direct: Transferring both homogeneous and inhomogeneous degradation from a distorted image $\mathbf{y}^{(1)}$ with different contents.
- Global-only: Transferring both homogeneous degradation from a distorted image $\mathbf{y}^{(2)}$ with different content that only contain homogeneous degradation
- Mixed: Transferring homogeneous degradation from one distorted image $\mathbf{y}^{(1)}$ and inhomogeneous degradation from another distorted image $\mathbf{y}^{(3)}$. All three involved images have different contents.

Model	Direct	Mixed
Full	.271	.286
No disentanglement (w/o entropy loss)	.290 (+.018)	.334 (+.048)

Table 5. Ablation study results on the disentangle effects of the disentangle-by-compression loss on degradation transfer. LPIPS scores (↓) for each modified model are shown.

Model	Global-only	Direct	Mixed
Full	.388	.271	.286
No IDA	.404 (+.016)	.298 (+.027)	.356 (+.070)
No IDA, No IDEN	.394 (+.006)	.577 (+.306)	.588 (+.302)

Table 6. Ablation study results on the effects of inhomogeneous degradation-related structures on degradation transfer. LPIPS scores (↓) for each modified model are shown.

Since all three cases use distortions synthesized from Python scripts, we are able to generate ground-truth distorted images to evaluate our performance. The distorted FFHQ dataset constructed in the previous subsection is used to train the models analyzed in the ablation study.

Disentangling-by-compression To benchmark the disentangle-by-compression method, we removed the entropy regularization loss functions. The results, shown in Table 5, indicate that this change led to decreased performance in degradation transfer, with a more pronounced effect on mixed transfer. This demonstrates the effectiveness of the disentangle-by-compression method in separating degradation information from distorted images and disentangling local from global degradations.

IDA layer and IDEN We introduced the IDE network and IDA layer for modeling inhomogeneous degradations. Our ablation study incrementally removed these components from the full model. The results, presented in Table 6 with a visual example in Figure 12, show that the IDA layer and IDE network not only enhance the reproduction and transfer of inhomogeneous degradations but also improve homogeneous degradation reproduction. Additional visual examples are available in Supp. Sec. 3.4.

5. Conclusion

We present the first universal image degradation model. The proposed disentangling-by-compression strategy enables the model to effectively separate distortion information from the distorted image's content while encouraging independence among the components of distortion embedding. The introduction of the IDA-SFT layer and the IDEN extends our model's capability to handle both homogeneous and inhomogeneous degradations, significantly broadening its applicability. Notably, our model transforms the non-blind restoration methods [13, 30] into blind ones, achieving competitive performance even without any distortion information.

Acknowledgements

This work is supported in part by National Sciences and Engineering Research Council of Canada under the Canada Research Chair and Discovery Grant programs.

References

[1] Abdelrahman Abdelhamed, Marcus A. Brubaker, and Michael S. Brown. Noise flow: Noise modeling with conditional normalizing flows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 1, 3

[2] Zoubida Ameur, Claire-Hélène Demarty, Olivier Le Meur, Daniel Ménard, and Edouard François. Style-based film grain analysis and synthesis. In *Proceedings of the 14th Conference on ACM Multimedia Systems*, page 229–238, New York, NY, USA, 2023. Association for Computing Machinery. 3

[3] Zoubida Ameur, Claire-Hélène Demarty, Olivier Le Meur, Daniel Ménard, and Edouard François. Style-based film grain analysis and synthesis. In *Proceedings of the 14th ACM Multimedia Systems Conference*, page 229–238, New York, NY, USA, 2023. Association for Computing Machinery. 7

[4] Zoubida Ameur, Wassim Hamidouche, Edouard François, Miloš Radosavljević, Daniel Menard, and Claire-Hélène Demarty. Deep-based film grain removal and synthesis. *IEEE Transactions on Image Processing*, 32:5046–5059, 2023. 3

[5] Codruta O. Ancuti, Cosmin Ancuti, Chris Hermans, and Philippe Bekaert. A fast semi-inverse approach to detect and remove the haze from a single image. In *Computer Vision – ACCV 2010*, pages 501–514, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. 3

[6] Johannes Ballé, Valero Laparra, and Eero P Simoncelli. End-to-end optimized image compression. *arXiv preprint arXiv:1611.01704*, 2016. 4

[7] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. Variational image compression with a scale hyperprior. In *International Conference on Learning Representations*, 2018. 4

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

[9] Qiang Cai, Mengxu Ma, Chen Wang, and Haisheng Li. Image neural style transfer: A review. *Computers and Electrical Engineering*, 108:108723, 2023. 3

[10] Lucy Chai, Michaël Gharbi, Eli Shechtman, Phillip Isola, and Richard Zhang. Any-resolution training for high-resolution image synthesis. In *Computer Vision – ECCV 2022*, pages 170–188, Cham, 2022. Springer Nature Switzerland. 7

[11] Ke-Chi Chang, Ren Wang, Hung-Jin Lin, Yu-Lun Liu, Chia-Ping Chen, Yu-Lin Chang, and Hwann-Tzong Chen. Learning camera-aware noise models. In *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV*, page 343–358, Berlin, Heidelberg, 2020. Springer-Verlag. 3

[12] Li-Heng Chen, Christos G. Bampis, Zhi Li, and Alan C. Bovik. Learning to distort images using generative adversarial networks. *IEEE Signal Processing Letters*, 27:2144–2148, 2020. 2, 3, 6

[13] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. 7, 8

[14] Arthur Conmy, Subhadip Mukherjee, and Carola-Bibiane Schönlieb. Stylegan-induced data-driven regularization for inverse problems. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3788–3792. IEEE, 2022. 7

[15] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 7

[16] Yingying Deng, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. Stytr2: Image style transfer with transformers. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 3

[17] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2567–2581, 2022. 5

[18] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008. 3

[19] Manuel Fritsche, Shuhang Gu, and Radu Timofte. Frequency separation for real-world super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3599–3608, 2019. 5

[20] Gnangarra, et al. Commons:quality images - Wikimedia Commons. https://commons.wikimedia.org/wiki/Commons:Quality_images, n.d. 6

[21] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1956–1963, 2009. 3

[22] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2020. 1, 3

[23] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*, 26(11):3365–3385, 2019. 3

[24] Seunghwan Lee and Tae Hyun Kim. Noisetransfer: Image noise generation with contrastive embeddings. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 3569–3585, 2022. 4

[25] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. AOD-Net: All-in-one dehazing network. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4780–4788, 2017. 3

[26] Xuan Luo, Xuaner Cecilia Zhang, Paul Yoo, Ricardo Martin-Brualla, Jason Lawrence, and Steven M. Seitz. Time-travel rephotography. *ACM Trans. Graph.*, 40(6), 2021. 3

[27] Markku Mäkitalo and Alessandro Foi. Poisson-gaussian denoising using the exact unbiased inverse of the generalized anscombe transformation. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1081–1084, 2012. 3

[28] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 257–265, 2017. 5, 6

[29] Andrey Norkin and Neil Birkbeck. Film grain synthesis for av1 video codec. In *2018 Data Compression Conference*, pages 3–12, 2018. 1, 7

[30] Yohan Poirier-Ginter and Jean-François Lalonde. Robust unsupervised stylegan image restoration. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22292–22301, 2023. 1, 7, 8

[31] Robby T. Tan. Visibility in bad weather from a single image. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008. 3

[32] Hong Wang, Zongsheng Yue, Qi Xie, Qian Zhao, Yefeng Zheng, and Deyu Meng. From rain generation to rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14791–14801, 2021. 1, 3

[33] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10576–10585, 2021. 5

[34] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson W.H. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12262–12271, 2019. 5, 6

[35] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 4

[36] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1905–1914, 2021. 2

[37] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. 5

[38] Zhongling Wang, Shahrukh Athar, and Zhou Wang. Blind quality assessment of multiply distorted images using deep neural networks. In *Image Analysis and Recognition*, pages 89–101, Cham, 2019. Springer International Publishing. 6

[39] Satosi Watanabe. Information theoretical analysis of multivariate correlation. *IBM J. Res. Dev.*, 4(1):66–82, 1960. 5

[40] Kaixuan Wei, Ying Fu, Yinqiang Zheng, and Jiaolong Yang. Physics-based noise modeling for extreme low-light photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8520–8537, 2022. 3

[41] J.C.K. Yan, P. Campisi, and D. Hatzinakos. Film grain noise removal and generation for color images. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, pages 2957–2960 vol.5, 1998. 1, 3

[42] Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng. Dual adversarial network: Toward real-world noise removal and noise generation. In *Computer Vision – ECCV 2020*, pages 41–58, Cham, 2020. Springer International Publishing. 1, 3

[43] Matthew D. Zeiler, Dilip Krishnan, Graham W. Taylor, and Rob Fergus. Deconvolutional networks. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2528–2535, 2010. 4

[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 (ICCV)*, pages 4791–4800, 2021. 2

[45] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 7

[46] Yi Zhang, Hongwei Qin, Xiaogang Wang, and Hongsheng Li. Rethinking noise synthesis and modeling in raw denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4593–4601, 2021. 1, 3

[47] Yifeng Zhou, Chuming Lin, Donghao Luo, Yong Liu, Ying Tai, Chengjie Wang, and Mingang Chen. Joint learning content and degradation aware feature for blind super-resolution. In *Proceedings of the 30th ACM International Conference on Multimedia*, page 2606–2616, New York, NY, USA, 2022. Association for Computing Machinery. 4

[48] Fengyuan Zhu, Guangyong Chen, and Pheng Ann Heng. From noise modeling to blind image denoising. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 420–429, 2016. 1, 3

[49] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, 2015. 3

[50] Yunhao Zou and Ying Fu. Estimating fine-grained noise model via contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12682–12691, 2022. 1, 3