

Equipping Diffusion Models with Differentiable Spatial Entropy for Low-Light Image Enhancement

Wenyi Lian
Uppsala University
Sweden
shermanlian@163.com

Wenjing Lian
Northeastern University
China
abbylian04@gmail.com

Ziwei Luo
Uppsala University
Sweden
ziwei.luo@it.uu.se

Abstract

Image restoration, which aims to recover high-quality images from their corrupted counterparts, often faces the challenge of being an ill-posed problem that allows multiple solutions for a single input. However, most deep learning based works simply employ ℓ_1 loss to train their network in a deterministic way, resulting in over-smoothed predictions with inferior perceptual quality. In this work, we propose a novel method that shifts the focus from a deterministic pixel-by-pixel comparison to a statistical perspective, emphasizing the learning of distributions rather than individual pixel values. The core idea is to introduce spatial entropy into the loss function to measure the distribution difference between predictions and targets. To make this spatial entropy differentiable, we employ kernel density estimation (KDE) to approximate the probabilities for specific intensity values of each pixel with their neighbor areas. Specifically, we equip the entropy with diffusion models and aim for superior accuracy and enhanced perceptual quality over ℓ_1 based noise matching loss. In the experiments, we evaluate the proposed method for low light enhancement on two datasets and the NTIRE challenge 2024. All these results illustrate the effectiveness of our statistic-based entropy loss. Code is available at <https://github.com/shermanlian/spatial-entropy-loss>.

1. Introduction

The quest for high-quality image restoration has led to advancements in neural networks and objective functions. Even so, achieving perceptually convincing restorations remains a challenge. Traditional pixel-wise loss functions, such as ℓ_1 and ℓ_2 , fall short in capturing the perceptual qualities of images, often resulting in high fidelity (reflected in high PSNR scores) but over-smoothed outputs [22, 31, 62]. This highlights the necessity for methods that extend beyond pixel accuracy to embrace a broader spectrum of im-

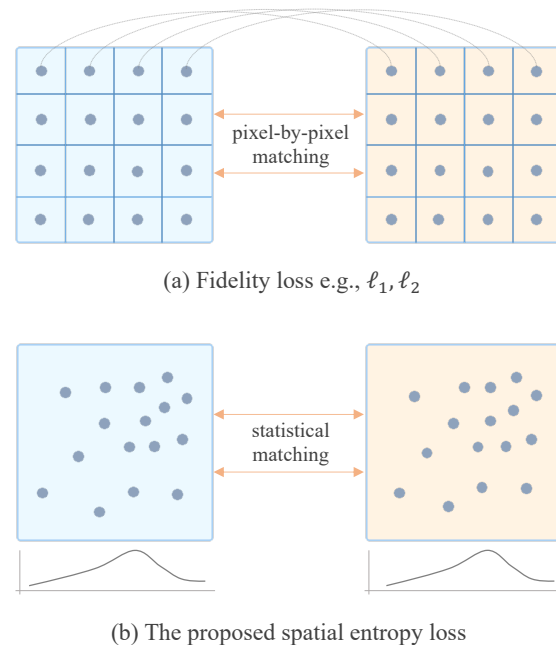


Figure 1. Illustration of the traditional fidelity loss (a) and the proposed spatial entropy loss (b). The fidelity loss uses pixel-by-pixel matching for two images while the proposed entropy loss adapts a statistical matching strategy for realistic image generation.

age attributes.

To address this issue, some studies utilize generative adversarial networks (GANs) [11] and perceptual metrics (e.g., VGG loss [19] and LPIPS loss [62]) to improve the visual quality. While these methods have shown promise in generating more realistic images, they still rely on the ℓ_1 loss to maintain the restoration accuracy. Moreover, both GAN and these perceptual metrics need additional (pre-trained) networks and are sensitive to different datasets and tasks. This not only increases the complexity of the model but also introduces a level of unpredictability, pointing to a need for more adaptable and self-sufficient solutions.

Given the limitations of GANs and perceptual metrics, the traditional SSIM [53] offers a promising alternative by evaluating image similarity through statistical attributes like means and variances of local patches, rather than relying on direct pixel-wise distances. While SSIM is often considered superior to PSNR (ℓ_2 -based metric) in image quality assessment, its reliance on basic statistical measures (mean and variance only) limits its effectiveness. When used as a loss function, SSIM may also produce over-smoothed images, akin to the limitations observed with ℓ_1 or ℓ_2 loss due to its restricted spatial information.

In this work, our objective is to design a novel statistic-based loss function which measures the distribution similarity between predicted images and ground truth images, as shown in Figure 1. The simplest way to represent an image distribution is through a 1D histogram, which calculates the probability of all pixel intensity values. This representation can then be utilized to compute meaningful statistic metrics such as the Shanno entropy [46] or relative entropy (KL-divergence) between two images. However, the limitation of a 1D histogram lies in its inability to comprehensively capture spatial information, thus rendering it inadequate for representing the entirety of an image. Additionally, the fact that the counting operation in histogram calculation is not differentiable prevents its use in training within a deep learning framework.

To address the above problems, this work proposes a differentiable spatial entropy. This method utilizes neighboring pixels to gather extra spatial information and employs kernel density estimation (KDE) to ensure the differentiability of the histogram/entropy. Moreover, to enhance the richness of pixel neighbor information, we suggest incorporating of randomly shuffled weights, which also improves the robustness of distribution matching. In experiments, we apply our entropy loss to diffusion models for low light image enhancement. The results shows that our new loss can preserve image restoration performance (e.g., PSNR/SSIM metrics) and further enhance the visual quality of predictions (in LPIPS [62], FID [14] scores).

Our main contributions are summarized as follows:

1. We present a novel statistic-based objective that optimizes distribution similarity rather than pixel-wise distance. Compared to other commonly used loss functions such as ℓ_1 , ℓ_2 , or GAN loss, it provides a new perspective for generative modelling.
2. We leverage KDE to make this spatial entropy differentiable. By simply equipping it with diffusion models for noise matching, the diffusion performance for realistic image generation is significantly improved.
3. Extensive experiments on the low light enhancement task demonstrate that the proposed entropy is effective for diffusion-based image generation.

2. Preliminaries

2.1. Image 1D Entropy

Consider an image X of size $H \times W$, with pixel intensities as non-negative integers from 0 to L . The 1D Shannon entropy [46] $H(\cdot)$ is defined as:

$$H(X) = - \sum_{i=0}^L p_i \log(p_i), \quad (1)$$

where L is the maximum grey level and p_i is the probability of i -th intensity value given by:

$$p_i = \frac{\sum_{x \in X} \mathbb{I}\{x = i\}}{H \times W}, \quad (2)$$

where $\mathbb{I}\{x = i\}$ is the indicator function that counts the occurrence of pixels with intensity i , and the denominator $H \times W$ represents the total number of pixels in the image, with H being the height and W the width of the image.

Eq. (2) calculates the frequency of each intensity level i through a counting operation, effectively capturing the intensity distribution across an image. However, it overlooks the spatial arrangement and the relationships between pixels, which can be valuable for uncovering the deeper context and structural details of the image.

2.2. Image Spatial Entropy

To enhance the representation of an image, spatial entropy considers additional spatial characteristics of pixels by examining their immediate surroundings [37, 49]. Specifically, for any given pixel x , we define \mathbb{N} as the set of its 8 neighboring pixel values. And $\tilde{x} \in \Omega$ is the rounded value of the mean of \mathbb{N} . Spatial entropy is then calculated as follows:

$$H_s(X) = - \sum_{i=0}^L \sum_{j \in \Omega} p_{i,j} \log p_{i,j}, \quad (3)$$

where $p_{i,j}$ denotes the probability of occurrence for a tuple of values (i, j) , where i represents the intensity value of a current pixel and j represents the average intensity value of its surrounding eight neighbors \mathbb{N} . Similar to Eq. (2), $p_{i,j}$ is obtained by:

$$p_{i,j} = \frac{\sum_{x \in X} \mathbb{I}\{x = i \ \& \ \tilde{x} = j\}}{H \times W}. \quad (4)$$

Compared to 1D entropy, which only considers the frequency of individual pixel intensities, the calculation of spatial entropy (Eq. (4)) incorporates spatial relationships by statistically evaluating the mean intensity of pixels in a neighborhood. This approach enables spatial entropy to more accurately represent an image’s spatial distribution, offering a deeper insight into its textural and structural complexity.

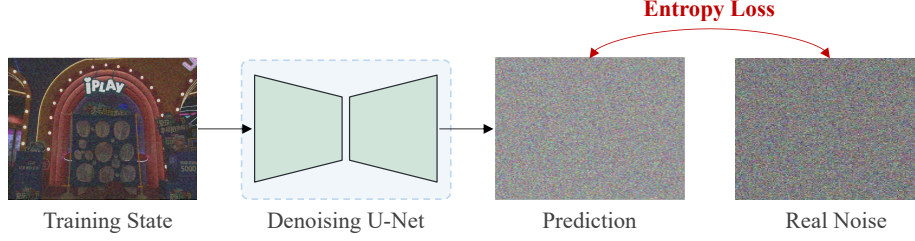


Figure 2. Overview of the training process for a diffusion model with the proposed spatial entropy loss for noise matching.

However, it is worth noting that both Eq. (2) and Eq. (4) contain a non-differentiable indicator function, which means we cannot use them as intermediate steps in the training of neural networks.

2.3. Diffusion Model

The general diffusion model consists of a forward process and a backward process. Specifically, given the input variable x_0 , the forward process iteratively adds noise to it with satisfying

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I), \quad (5)$$

where β_t is the predefined variance. With the nice property of Gaussian, we could have

$$q(x_t | x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)I), \quad (6)$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$. Note that as $t \rightarrow \infty$, the sample of $q(x_t | x_0)$ converges to a Gaussian noise $\mathcal{N}(0, I)$. Then the backward process reverses the noise to the initial variable x_0 , which naturally models the generative process and can reconstruct high-quality images.

Importantly, the diffusion forward process can be also modeled as a forward Stochastic Differential Equation (SDE) with continuous variables, defined by

$$dx = f(x, t) dt + g(t) dw, \quad x(0) \sim p_0(x), \quad (7)$$

where f and g are the drift and dispersion functions, respectively, w is a standard Wiener process, and $x(0) \in \mathbb{R}^d$ is an initial condition. And similarly, such an SDE can be reversed to model the generative process:

$$dx = \left[f(x, t) - g(t)^2 \nabla_x \log p_t(x) \right] dt + g(t) d\hat{w}, \quad (8)$$

where $x(T) \sim p_T(x)$ and \hat{w} is a reverse-time Wiener process and $p_t(x)$ stands for the marginal probability density function of $x(t)$ at time t . The score function $\nabla_x \log p_t(x)$ is in general intractable and thus SDE-based diffusion models approximate it by training a time-dependent neural network $s_\theta(x, t)$ under a so-called score matching objective.

3. Method

3.1. Differentiable Spatial Entropy

The idea is to introduce the kernel density estimation (KDE) [41] to estimate the smooth probability density function (PDF) instead of directly counting the number of the same intensity value (Eq. (2)) or tuple (Eq. (4)). Formally, we define the one-dimensional KDE for the probability of a specific intensity value i as:

$$p_i \approx \hat{f}_h(i) = \frac{1}{2Nh} \sum_{x \in X} \mathbb{I} \left\{ \frac{|x - i|}{h} < 1 \right\} \quad (9)$$

$$= \frac{1}{Nh} \sum_{x \in X} \mathcal{K} \left(\frac{x - i}{h} \right), \quad (10)$$

where $N = H \times W$ is the total number of pixels, $\mathcal{K}(\cdot)$ represents the generalized kernel substituting the indicator function $\frac{1}{2}\mathbb{I}\{\cdot\}$ and h is a smoothing parameter called the bandwidth. The above 1D KDE is derived from Eq. (2) and it can be made differentiable by choosing an appropriate kernel function such as Gaussian distribution. Similarly, to estimate the spatial probability density for the tuple (i, j) , we rewrite Eq. (4) to a spatial KDE for $p_{i,j}$ as follows:

$$\hat{f}_h(i, j) = \frac{1}{2Nh} \sum_{x \in X} \mathbb{I} \left\{ \frac{|x - i|}{h} < 1 \ \& \ \frac{|\tilde{x} - j|}{h} < 1 \right\} \quad (11)$$

$$= \frac{1}{2Nh} \sum_{x \in X} \mathbb{I} \left\{ \frac{|x - i|}{h} < 1 \right\} \cdot \mathbb{I} \left\{ \frac{|\tilde{x} - j|}{h} < 1 \right\}. \quad (12)$$

By replacing the indicator function $\frac{1}{2}\mathbb{I}\{\cdot\}$ with $\mathcal{K}(\cdot)$, we have the following:

$$\hat{f}_h(i, j) = \frac{2}{Nh} \sum_{x \in X} \frac{1}{2} \mathbb{I} \left\{ \frac{|x - i|}{h} < 1 \right\} \cdot \frac{1}{2} \mathbb{I} \left\{ \frac{|\tilde{x} - j|}{h} < 1 \right\} \quad (13)$$

$$= \frac{2}{Nh} \sum_{x \in X} \mathcal{K}_1 \left(\frac{x - i}{h} \right) \cdot \mathcal{K}_2 \left(\frac{\tilde{x} - j}{h} \right) \quad (14)$$

Here \mathcal{K}_1 and \mathcal{K}_2 could be the same kernel function. In particular, valid kernels should be real-valued, non-negative,

symmetric, and satisfy the normalization: $\int \mathcal{K}(t)dt = 1$, which guarantees the spatial KDE is a probability density function. We usually choose Gaussian or the derivative of sigmoid [2] as the kernel function.

3.2. Spatial Entropy Loss for Image Restoration

To make use of the differentiable spatial KDE in image generation, we introduce statistical losses to measure the distribution distance between the predicted image and ground truth. A typical example is the relative entropy (also called KL-divergence). Mathematically, let P and Q represent the ground truth’s and prediction’s probability distributions computed from Eq. (14), the spatial relative entropy is defined as follows:

$$H_s(P, Q) = D_{KL}(P||Q) = - \sum_{i=0}^L \sum_{j \in \Omega} p_{i,j} \log \frac{q_{i,j}}{p_{i,j}}. \quad (15)$$

Note that the choice of the loss function is flexible, other statistic measurements like Cross-entropy and Hellinger distance [50] can also be used in our framework. In addition, although we directly employ spatial relative entropy as our loss function for network training, the statistics of the whole image are too global and spatially non-stationary, which cannot guide the reconstruction of local details. To address this issue, we further propose to compute the relative entropy locally to provide a spatially varying quality map of the image, which is more stable and efficient for different image distortions. Similar to SSIM [53], we slide a 11×11 square window to compute the local probabilities, which can be regarded as a convolution operation by setting all weights to one and then replace the N with the total number of pixels i.e. 11×11 in Eq. (14).

Shuffle Weights for Neighborhoods Augmentation In previous works, the mean value of neighborhood pixels often serves as the spatial information. However, only using one pattern of averaging neighborhoods is still not sufficient to represent the whole image. To better improve the spatial information during training, we propose adding random weights to neighboring pixels for each entropy calculation, as a strategy for weighted averaging augmentation. In theory, this approach of differentiable spatial entropy, enhanced by the random weights strategy, can be more broadly and effectively applied across various image generation frameworks and applications.

Equipping Entropy to Diffusion Models Most diffusion models adopt noise-matching loss (NML) for generative modelling. In particular, NML uses ℓ_2 to match predicted noise and a real Gaussian noise at a random sampled time t . To equip the proposed entropy to common diffusion models, we just simply replace the ℓ_2 term in noise matching. In this

paper, we use Refusion [35] as the base diffusion framework which learns a mean-reverting SDE [34] (a special case in Eq. (7)) for image restoration. Figure 2 illustrated the combination of diffusion models with entropy loss. To further improve the distortion performance (in terms of PSNR and SSIM metrics), we follow Refusion to change NML to the maximum likelihood loss (MLL) [34] for optimal path reversing.

4. Experiments

In this section, we provide evaluations on the low light enhancement task with two datasets from LoL-v1 and LoL-v2 in Sec. 4.2 and analyze the effectiveness of the entropy in Sec. 4.3. Moreover, in Sec. 4.4, we further report the final results of the corresponding NTIRE 2024 low light enhancement challenge. Finally, the limitations and future work are presented in Sec. 4.5.

4.1. Datasets and Implementation Details

Datasets. Our method is evaluated on the Low Light Paired (LoL) datasets: LOL-v1 [55] and LOL-v2-real [58]. The LOL-v1 dataset is the first of its kind, comprising image pairs captured from real-world scenarios. It consists of 500 pairs of low-light and normal-light images, with 485 pairs for training and 15 for testing. These images maintain a consistent resolution of 400×600 pixels and mostly depict indoor scenes with typical low-light noise. In contrast, the LOL-v2-real dataset contains 689 pairs of low-/normal-light images for training and additional 100 pairs for testing. This dataset captures a diverse range of scenes, both indoors and outdoors, under varying lighting conditions. The diversity of the LOL-v2-real dataset allows us to learn from a broader spectrum of low-light scenarios, enhancing its evaluation of image enhancement techniques.

Implementation Details. We develop the image restoration SDE (IR-SDE) [34] model targeting low-light image enhancement with a Conditional NAFNet [5] architecture. This model employs the AdamW optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.99$ over 200,000 iterations. The initial learning rate is 5×10^{-5} , reduced to 1×10^{-6} through a cosine annealing schedule. Training samples are processed in batches of 8, with 128×128 patches randomly cropped from low-/normal-light image pairs. Data augmentation includes random rotation and flipping. In addition, we set the SDE timestep to 100 in both training and testing. All our experiments are implemented using the PyTorch framework and performed on an NVIDIA A100 GPU with 2 days of training for each dataset.

4.2. Comparison with State-of-the-Arts

Comparison Methods In this section, we compare the proposed method with the following state-of-the-

Table 1. Quantitative comparison between the proposed method with other state-of-the-art low-light enhancement approaches on the LOL-v1 [55] test set. The best results are marked in **bold**.

Method	Distortion		Perceptual	
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FID \downarrow
NPE [52]	16.97	0.484	0.400	104.1
SRIE [10]	11.86	0.495	0.353	88.73
LIME [13]	17.55	0.531	0.387	117.9
RetinexNet [55]	16.77	0.462	0.417	126.3
DSLRL [27]	14.82	0.572	0.375	104.4
DRBN [58]	16.68	0.730	0.345	98.73
Zero-DCE [12]	14.86	0.562	0.372	87.24
MIRNet [59]	24.14	0.830	0.250	69.18
EnlightenGAN [17]	17.61	0.653	0.372	94.70
ReLLIE [63]	11.44	0.482	0.375	95.51
RUAS [28]	16.41	0.503	0.364	102.0
DDIM [48]	16.52	0.776	0.376	84.07
CDEF [23]	16.34	0.585	0.407	90.62
SCI [36]	14.78	0.525	0.366	78.60
URetinex-Net [56]	19.84	0.824	0.237	52.38
SNRNet [57]	24.61	0.842	0.233	55.12
Uformer [54]	19.00	0.741	0.354	109.4
Restormer [60]	20.61	0.797	0.288	73.00
Palette* [43]	11.77	0.561	0.498	108.3
UHDFour _{2x} [24]	23.09	0.821	0.259	56.91
WeatherDiff [40]	17.91	0.811	0.272	73.90
GDP [9]	15.90	0.542	0.421	117.5
Retinexformer [4]	25.16	0.845	0.130	71.21
DiffLL [16]	26.34	0.845	0.217	48.11
Entropy-SDE (Ours)	24.05	0.848	0.081	37.20

art methods: NPE [52], SRIE [10], LIME [13], RetinexNet [55], DSLR [27], DRBN [58], Zero-DCE [12], MIRNet [59], EnlightenGAN [17], ReLLIE [63], RUAS [28], DDIM [48], CDEF [23], SCI [36], URetinex-Net [56], SNRNet [57], Uformer [54], Restormer [60], Palette [43], UHDFour_{2x} [24], WeatherDiff* [40], GDP [9], Retinexformer [4], DiffLL [16]. These methods can be roughly split into four categories: optimization-based approaches, learning-based approaches, transformer-based approaches, and diffusion-based approaches. Our method belongs to the last category. For most methods, we report their results from the DiffLL [16] paper. And we compare the visual results with EnlightenGAN [17] and URetinex-Net [56] by re-testing their official pretrained models.

Evaluation Metrics As the purpose of this work is to improve the visual quality of diffusion-based results in low-light enhancement, we thus focus more on the perceptual metrics: Learned Perceptual Image Patch Similarity (LPIPS) [62] and Fréchet inception distance (FID) [14]. In

Table 2. Quantitative comparison between the proposed method with other state-of-the-art low-light enhancement approaches on the LOL-v2-real [58] test set. The best results are marked in **bold**.

Method	Distortion		Perceptual	
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FID \downarrow
NPE [52]	17.33	0.464	0.396	100.0
SRIE [10]	14.45	0.524	0.332	78.83
LIME [13]	17.48	0.505	0.428	118.2
RetinexNet [55]	17.72	0.652	0.436	133.9
DSLRL [27]	17.00	0.596	0.408	114.3
DRBN [58]	18.47	0.768	0.352	89.09
Zero-DCE [12]	18.06	0.580	0.352	80.45
MIRNet [59]	20.02	0.820	0.233	49.10
EnlightenGAN [17]	18.68	0.678	0.364	84.04
ReLLIE [63]	14.40	0.536	0.334	79.84
RUAS [28]	15.35	0.495	0.395	94.16
DDIM [48]	15.28	0.788	0.387	76.39
CDEF [23]	19.76	0.630	0.349	74.06
SCI [36]	17.30	0.540	0.345	67.62
URetinex-Net [56]	21.09	0.858	0.208	49.84
SNRNet [57]	21.48	0.849	0.237	54.53
Uformer [54]	18.44	0.759	0.347	98.14
Restormer [60]	24.91	0.851	0.264	58.65
Palette [43]	14.70	0.692	0.333	83.94
UHDFour _{2x} [24]	21.79	0.854	0.292	60.84
WeatherDiff* [40]	20.00	0.829	0.253	59.67
GDP [9]	14.29	0.493	0.435	102.4
Retinexformer [4]	22.80	0.840	0.169	62.46
DiffLL [16]	28.86	0.876	0.207	45.36
Entropy-SDE (Ours)	21.31	0.832	0.120	49.61

addition, distortion metrics like PSNR and SSIM [53] are also reported to measure the image fidelity for reference.

Results The quantitative comparison results on two datasets (LoL-v1 [55] and LoL-v2-real [58]) are illustrated in Table 1 and Table 2. It is observed that most advanced image restoration approaches (such as the MIRNet, SNR-Net, Restormer, and Retinexformer) only perform well on the distortion performance (high PSNR and SSIM), meaning that they are easy to fit pixels but tend to produce relatively smooth or blurry results. DiffLL [16] uses a specific wavelet-based diffusion model for low light enhancement and achieves the overall best results for both datasets. By leveraging the entropy loss in training, our method (Entropy-SDE) performs comparably with DiffLL on perceptual metrics in the LoL-v2-real dataset and even achieves the best performance over most metrics on the LoL-v1 dataset. These results illustrate that combining the proposed entropy loss with diffusion models works well on various datasets. In addition, it is reasonable that our fidelity results are inferior to other methods since we only lever-

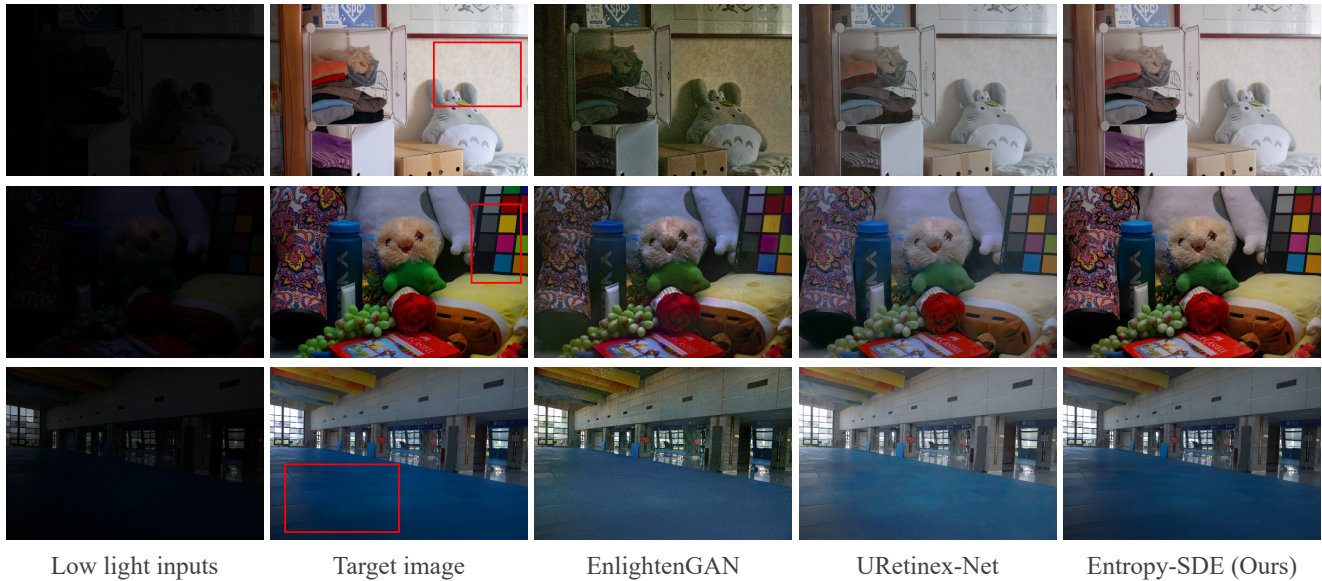


Figure 3. Visual comparison of the proposed model with other low light enhancement approaches on the LoL-v1 [55] dataset.

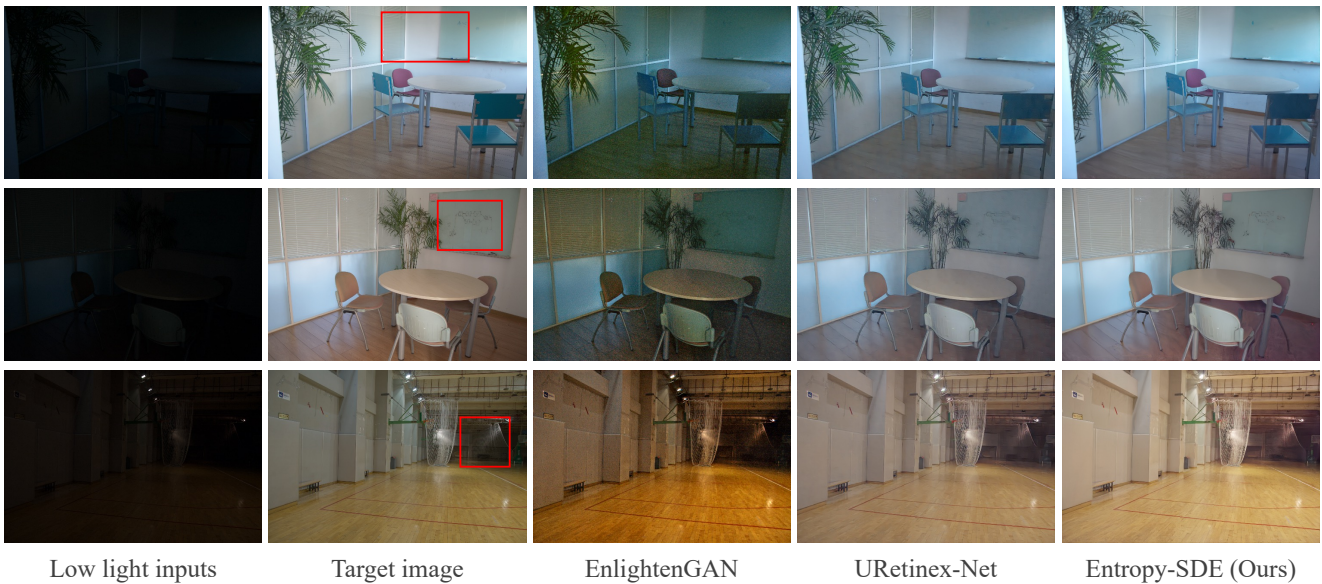


Figure 4. Visual comparison of the proposed model with other low light enhancement approaches on the LoL-v2-real [58] dataset.

age statistical information but PSNR is computed based on pixel-by-pixel distance.

The visual comparison of our method with EnlightenGAN [17] and URetinex-Net [56] on two datasets are shown in Figure 3 and Figure 4, respectively. Note that EnlightenGAN always encounters the color shift problem due to the unstable GAN training process. The results produced by URetinex-Net look better in colors but lose some details in reconstruction (such as the blue point in the first-row of Figure 4). Our results are consistent with the target images

while also look more realistic.

4.3. Effectiveness of the Entropy Loss

In this section, we include ablation experiments further to analyze the effectiveness of the proposed spatial entropy Loss. In particular, the Refusion model is used as the baseline and we add individual ℓ_1 loss and entropy to it. Here, we use a comprehensive set of metrics for evaluation, including PSNR, SSIM [53], LPIPS [62], DISTS [7], FID [14], and NIQE [38]. Note the NIQE is a non-reference

Table 3. Ablation experiment for the proposed entropy loss on LOL-v1 [55] test set. The Refusion [35] is used as the baseline.

Method	Distortion		Perceptual			
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	DISTS \downarrow	FID \downarrow	NIQE \downarrow
Baseline /w ℓ_1 loss	23.21	0.826	0.114	0.109	51.99	4.67
Baseline /w Entropy	24.05	0.848	0.081	0.071	37.20	4.40

Table 4. Ablation experiment for the proposed entropy loss on LOL-v2-real [58]. The Refusion [35] is used as the baseline.

Method	Distortion		Perceptual			
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	DISTS \downarrow	FID \downarrow	NIQE \downarrow
Baseline /w ℓ_1	18.54	0.812	0.151	0.136	51.99	3.98
Baseline /w Entropy	21.31	0.832	0.120	0.119	49.61	4.09

metric that only measures the visual quality of outputs. In contrast, all other metrics measure both visual quality and the consistency between inputs and outputs.

The results on two datasets are shown in Table 3 and Table 4. Obviously, baseline models with the proposed entropy loss can improve the results significantly. It is worth noting that the entropy loss only employs statistical information from images but still outperforms the ℓ_1 loss version, further demonstrating the effectiveness of the proposed spatial entropy loss and the importance of differentiable solutions to the entropy.

4.4. NTIRE Low Light Enhancement Challenge

Our model was also evaluated in the NTIRE 2024 Low Light Enhancement Challenge [29], which provides a high-quality dataset containing 230 training scenes, along with 35 validation and 35 testing ones. We use the same model and settings as described in Sec. 4.1. In this dataset, most images are 4K resolution that are hard to process with our diffusion model. Thus we downsample these large images with a factor of 0.5 for memory efficiency and resize them back to the original resolution after enhancement. The final results and rank are shown in Table 5, in which we choose the Top 10 teams and report our results in the last row. **Notably, our method outperforms all top teams in terms of LPIPS.**

In Figure 5, we visualize two examples from the challenge validation dataset and illustrate the results of two diffusion models: learn noise matching with ℓ_1 and noise matching with the proposed spatial entropy. Although there are no ground truth images, it is easy to observe that spatial entropy improves the visual quality of diffusion models.

4.5. Limitation

Although the proposed spatial entropy can achieve good perceptual performance, the training is computationally

Table 5. Evaluation and Rankings in the NTIRE 2024 Low Light Enhancement Challenge.

Team	PSNR	SSIM	LPIPS	Final Rank
SYSU-FVL-T2	25.52	0.8637	0.1221	1
Retinexformer	25.30	0.8525	0.1424	2
DH-AISP	24.97	0.8528	0.1235	3
NWPU-DiffLight	24.78	0.8556	0.1673	4
GiantPandaCV	24.83	0.8474	0.1353	5
LVGroup_HFUT	24.88	0.8395	0.1371	6
Tryltry8	24.49	0.8483	0.1359	7
Pixel_warrior	24.74	0.8416	0.1514	8
HuiT	24.13	0.8484	0.1436	9
X-LIME	24.28	0.8446	0.1298	10
221B (Ours)	22.04	0.8141	0.1084	17

costly since we need to count the pixel numbers for all bandwidths repeatedly. Therefore, we have to use a small patch size in training, which might affect the final performance. In addition, it is worth noting that the proposed entropy loss can be also used for other tasks and frameworks. However, since the loss is purely based on statistical feature matching, it reduces the weight of fidelity parts thus leading an inferior results on other metrics like PSNR and SSIM. That means directly applying it to some scenes would produce undesired artifacts. An example of applying it to image deblurring is shown in Figure 6. Although this problem can be alleviated using diffusion models in their iterative denoising process, it is still challenging to apply it to common image restoration frameworks. We regard this as our future work.

5. Related Work

Low Light Image Enhancement Low light enhancement aims to recover the correct illumination of images that are captured in the dark [21, 39]. To make the transformed images visually satisfactory, numerous works have made efforts to improve the contrast and details in restoration [10, 13, 18, 47]. Traditional methods mainly adopt histogram equalization (HE) [3, 39, 47] and Retinex theory [10, 21]. Built upon them, deep learning approaches are developed and have achieved impressive progress. LLNet [30] is the first work applying deep neural networks to low light enhancement. Following this, RetinexNet [55] proposes a real-world captured low light dataset with a Retinex theory based network. To improve the details, EnlightenGAN [17] combines GAN loss with unpaired images for practical model training. And some other concurrent works mostly adopt similar architectures or traditional theories for data-driven image enhancement [4, 12, 28, 51, 56, 57, 64, 65]. Most recently, DiffLL [16] further utilizes a wavelet-based diffusion model to produce illuminated results with satisfactory perceptual fidelity.

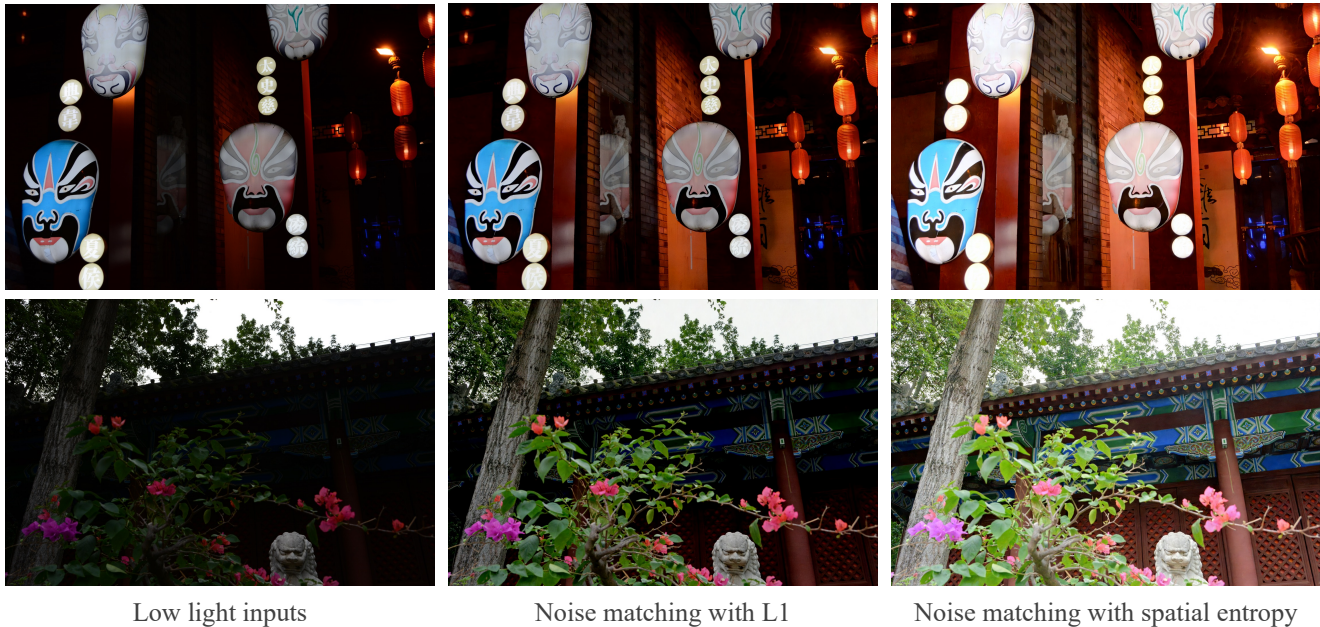


Figure 5. Visual results of the proposed method on the NTIRE 2024 Low Light Enhancement Challenge dataset.



Figure 6. A case of applying the proposed spatial entropy for image deblurring. Here we use a simple U-Net as the base network and directly train it with our differentiable spatial entropy loss.

Diffusion Models for Image Restoration Image restoration aims to reconstruct a high-quality image from its corrupted counterpart [1, 25, 26, 32]. It contains a wide range of practical applications such as image deblurring [45], denoising [61], super-resolution [8], etc. Existing deep learning based methods directly train neural networks with ℓ_1/ℓ_2 loss based on collected data pairs, which is effective for image reconstruction but would produce over-smooth results [22, 66]. Recently, the generative diffusion model has drawn increasing attention due to its stable training process and remarkable performance in producing realistic images and videos [15, 42, 44]. Inspired by it, recent researchers started converting various image restoration tasks into diffusion processes to obtain high-perceptual results [6, 20, 33, 34, 67]. Notably, all these methods still use ℓ_1/ℓ_2 for noise matching to learn the diffusion process. In this paper, our method is the first that proposes to use a purely statistical matching approach for noise matching.

6. Conclusion

This paper presents a statistic-based matching approach (spatial entropy loss) for image restoration. Specifically, we introduce the kernel density estimation (KDE) to make the spatial entropy differentiable. Then the spatial entropy can be used for different learning-based frameworks for image reconstruction. By equipping it into the diffusion models (to substitute the ℓ_1 or ℓ_2), we obtain a novel statistical noise matching loss for realistic image restoration. We then apply this model to the low light enhancement task to illustrate its effectiveness. Our model achieves the best LPIPS performance in the NTIRE Low Light Enhancement challenge. All these results demonstrate that the spatial entropy loss is effective for the high perceptual diffusion learning process.

Acknowledgements The computations were enabled by the *Berzelius* resource provided by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre.

References

- [1] Harry C Andrews and Bobby Ray Hunt. *Digital image restoration*. Prentice Hall Professional Technical Reference, 1977. 8
- [2] Mor Avi-Aharon, Assaf Arbelle, and Tammy Riklin Raviv. Differentiable histogram loss functions for intensity-based image-to-image translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 4
- [3] Alan C Bovik. *Handbook of image and video processing*. Academic press, 2010. 7
- [4] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12504–12513, 2023. 5, 7
- [5] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022. 4
- [6] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 8
- [7] 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, 2020. 6
- [8] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*, pages 391–407. Springer, 2016. 8
- [9] Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9935–9946, 2023. 5
- [10] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2782–2790, 2016. 5, 7
- [11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 1
- [12] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1780–1789, 2020. 5, 7
- [13] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 5, 7
- [14] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 2, 5, 6
- [15] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 8
- [16] Hai Jiang, Ao Luo, Haoqiang Fan, Songchen Han, and Shuaicheng Liu. Low-light image enhancement with wavelet-based diffusion models. *ACM Transactions on Graphics (TOG)*, 42(6):1–14, 2023. 5, 7
- [17] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30:2340–2349, 2021. 5, 6, 7
- [18] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997. 7
- [19] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016. 1
- [20] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022. 8
- [21] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 7
- [22] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 1, 8
- [23] Xiaozhou Lei, Zixiang Fei, Wenju Zhou, Huiyu Zhou, and Minrui Fei. Low-light image enhancement using the cell vibration model. *IEEE Transactions on Multimedia*, 2022. 5
- [24] Chongyi Li, Chun-Le Guo, Man Zhou, Zhixin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. *arXiv preprint arXiv:2302.11831*, 2023. 5
- [25] Wenyi Lian and Wenjing Lian. Sliding window recurrent network for efficient video super-resolution. In *European Conference on Computer Vision*, pages 591–601. Springer, 2022. 8
- [26] Wenyi Lian and Shanglian Peng. Kernel-aware burst blind super-resolution. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 4892–4902, 2023. 8
- [27] Seokjae Lim and Wonjun Kim. Dslr: Deep stacked laplacian restorer for low-light image enhancement. *IEEE Transactions on Multimedia*, 23:4272–4284, 2020. 5
- [28] Risheng Liu, Long Ma, Jiao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior

- architecture search for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10561–10570, 2021. 5, 7
- [29] Xiaoning Liu, Zongwei Wu, Ao Li, Florin-Alexandru Vasluianu, Yulun Zhang, Shuhang Gu, Le Zhang, Ce Zhu, Radu Timofte, et al. NTIRE 2024 challenge on low light enhancement: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2024. 7
- [30] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 7
- [31] Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte. Srflow: Learning the super-resolution space with normalizing flow. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 715–732. Springer, 2020. 1
- [32] Ziwei Luo, Haibin Huang, Lei Yu, Youwei Li, Haoqiang Fan, and Shuaicheng Liu. Deep constrained least squares for blind image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17642–17652, 2022. 8
- [33] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for universal image restoration. In *The Twelfth International Conference on Learning Representations*, 2023. 8
- [34] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Image restoration with mean-reverting stochastic differential equations. *arXiv preprint arXiv:2301.11699*, 2023. 4, 8
- [35] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1680–1691, 2023. 4, 7
- [36] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5637–5646, 2022. 5
- [37] H. Maitre, I. Bloch, and M. Sigelle. Spatial entropy: a tool for controlling contextual classification convergence. In *Proceedings of 1st International Conference on Image Processing*, pages 212–216 vol.2, 1994. 2
- [38] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 6
- [39] Chen Hee Ooi and Nor Ashidi Mat Isa. Quadrants dynamic histogram equalization for contrast enhancement. *IEEE Transactions on Consumer Electronics*, 56(4):2552–2559, 2010. 7
- [40] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 5
- [41] Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962. 3
- [42] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 8
- [43] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–10, 2022. 5
- [44] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, 35:36479–36494, 2022. 8
- [45] Qi Shan, Jiaya Jia, and Aseem Agarwala. High-quality motion deblurring from a single image. *Acm transactions on graphics (tog)*, 27(3):1–10, 2008. 8
- [46] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948. 2
- [47] Kuldeep Singh, Rajiv Kapoor, and Sanjeev Kr Sinha. Enhancement of low exposure images via recursive histogram equalization algorithms. *Optik*, 126(20):2619–2625, 2015. 7
- [48] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 5
- [49] F. Tupin, M. Sigelle, and H. Maitre. Definition of a spatial entropy and its use for texture discrimination. In *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, pages 725–728 vol.1, 2000. 2
- [50] Aad W Van der Vaart. *Asymptotic statistics*. Cambridge university press, 2000. 4
- [51] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6849–6857, 2019. 7
- [52] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing*, 22(9):3538–3548, 2013. 5
- [53] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 2, 4, 5, 6
- [54] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 5
- [55] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 4, 5, 6, 7

- [56] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhao Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5901–5910, 2022. [5](#), [6](#), [7](#)
- [57] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17714–17724, 2022. [5](#), [7](#)
- [58] Wenhao Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3063–3072, 2020. [4](#), [5](#), [6](#), [7](#)
- [59] 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. [5](#)
- [60] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. [5](#)
- [61] 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. [8](#)
- [62] 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*, pages 586–595, 2018. [1](#), [2](#), [5](#), [6](#)
- [63] Rongkai Zhang, Lanqing Guo, Siyu Huang, and Bihan Wen. Rellie: Deep reinforcement learning for customized low-light image enhancement. In *Proceedings of the 29th ACM international conference on multimedia*, pages 2429–2437, 2021. [5](#)
- [64] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. [7](#)
- [65] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129:1013–1037, 2021. [7](#)
- [66] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016. [8](#)
- [67] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhong Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Pro-*

ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 1219–1229, 2023. [8](#)