Uncertainty-Driven Dehazing Network

Ming Hong[1, *], **Jianzhuang Liu**[2], **Cuihua Li**[1], **Yanyun Qu** [1, †]

[1] Xiamen University,
[2] Huawei Noah's Ark Lab
mingh@stu.xmu.edu.cn, liu.jianzhuang@huawei.com, chli@xmu.edu.cn, yyqu@xmu.edu.cn

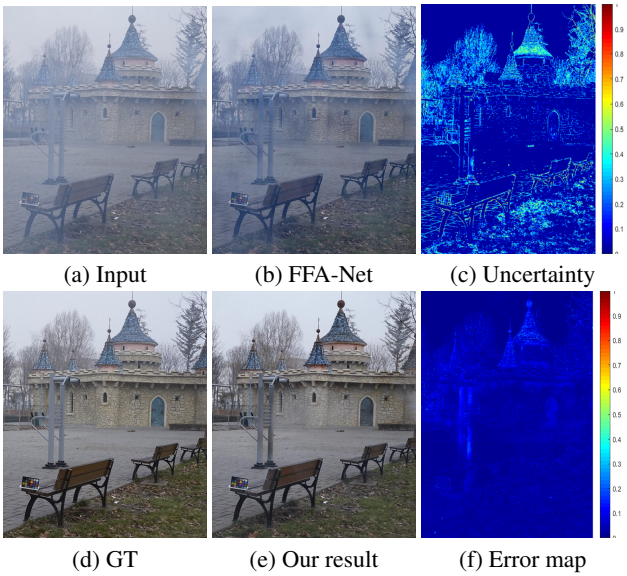
Abstract

Deep learning has made remarkable achievements for single image haze removal. However, existing deep dehazing models only give deterministic results without discussing their uncertainty. There exist two types of uncertainty in the dehazing models: aleatoric uncertainty that comes from noise inherent in the observations and epistemic uncertainty that accounts for uncertainty in the model. In this paper, we propose a novel uncertainty-driven dehazing network (UDN) that improves the dehazing results by exploiting the relationship between the uncertain and confident representations. We first introduce an Uncertainty Estimation Block (UEB) to predict the aleatoric and epistemic uncertainty together. Then, we propose an Uncertainty-aware Feature Modulation (UFM) block to adaptively enhance the learned features. UFM predicts a convolution kernel and channel-wise modulation coefficients conditioned on the uncertainty weighted representation. Moreover, we develop an uncertainty-driven self-distillation loss to improve the uncertain representation by transferring the knowledge from the confident one. Extensive experimental results on synthetic datasets and real-world images show that UDN achieves significant quantitative and qualitative improvements, outperforming state-of-the-arts.

Introduction

Hazy images often suffer from the degradation of visual quality such as limited visibility and low contrast [Tan 2008], leading to the failure of subsequent high-level visual tasks such as object detection [Liu et al. 2018], semantic segmentation [Ren et al. 2018] and so on. Hence image dehazing is highly demanded by vision-based applications.

Traditional dehazing methods [He, Sun, and Tang 2009], [Berman, Treibitz, and Avidan 2016], [Zhu, Mai, and Shao 2015] assume various image priors based on the observation and statistic analysis of natural hazy images, which are susceptible to violation in the wild world. Recently, deep learning has made great success in single image dehazing. The existing deep dehazing models can be roughly divided into physical-model dependent methods and physical-model free methods. The former firstly estimate the transmission map and global atmospheric light, and then restore the clear image by inversely transforming the physical scattering model [Ren et al. 2016], [Cai et al. 2016]. The latter directly explore an end-to-end mapping from hazy images to their haze-free counterparts [Qu et al. 2019], [Qin et al. 2020], [Liu et al. 2019]. Although these methods achieved significant improvement, they only give dehazing results without knowing how uncertain the results are.



(a) Input (b) FFA-Net (c) Uncertainty
(d) GT (e) Our result (f) Error map

Figure 1: Dehazing results on a dense haze photo from [Ancuti et al. 2018]. (a) Input hazy image. (b) Result of FFA-Net [Qin et al. 2020]. (d) Ground Truth. (e) Result of our model. (c) and (f) are the uncertainty map and absolute reconstruction error map obtained by our model. The more uncertain (less confident) are the dehazing results, the larger the reconstruction errors.

Uncertainty is important for an agent to make a decision for action. When given a real-world hazy image, how much can we trust the results obtained by these models? The probabilistic interpretation would be the desired tool to evaluate the uncertainty of dehazing models. In the MRI reconstruction [Zhang et al. 2019b] and image segmentation [Zheng et al. 2021], with the help of the uncertainty estima-

[*]Part of this work was done during an internship in Huawei Noah's Ark Lab.
[†]Corresponding author

tion, the models not only output the results but also provide the confidence values, which is favorable to an agent's inference. Besides, it is observed in [Yasarla and Patel 2019] and [Chen, Wen, and Chan 2021] that uncertainty measurement accompanied by pixel regression can lead to a more informed decision, and even improve the prediction quality. In these cases, the uncertainty estimation is treated as a regularization term or conditional input. However, the relationship between uncertain regions and confident ones is rarely studied, and how to exploit the knowledge in the confident representation to improve the uncertain one is still open.

The degradation degree of a hazy pixel varies with its color and depends on the distance of the scene to the camera. To differently treat each pixel, FFA-Net [Qin et al. 2020] proposes pixel attention and channel attention mechanisms to focus on regions with thick haze. However, these attention mechanisms are implicitly learned which lack interpretability and fail to build an explicit and direct relationship with the reconstruction result. To solve this problem, we propose to use uncertainty estimation to drive feature learning. The pixel-wise uncertainty is explicitly predicted and can measure the confidence values of the dehazing result. Generally, the larger the uncertain value, the larger the reconstruction error, and vice versa.

In this paper, we propose an Uncertainty-driven Dehazing Network (UDN) which focuses on enhancing the learned features and improving the dehaizng results conditioned on the predicted uncertainty. Specifically, we develop an Uncertainty Driven Module (UDM) to improve the uncertain representation guided by the confident one. In each UDM, an Uncertainty Estimation Block (UEB) is developed to estimate the combined pixel-wise aleatoric and epistemic uncertainty, and an Uncertainty-aware Feature Modulation (UFM) block is developed to adaptively strengthen the features based on the predicted uncertainty. UFM achieves feature adaption by predicting a convolution kernel and channel-wise modulation coefficients from the uncertainty weighted representation. Moreover, we mine the similarity relationship of pixels in the clear image and propose an uncertainty-driven self-distillation loss to improve an uncertain pixel's representation by transferring the confident knowledge from its similar pixels' representations. As shown in Figure 1, our model can predict the pixel-wise uncertainty of the dehazing result, and thereby identify uncertain regions that are likely to contain large reconstruction errors, *i.e.,* the edges of the objects. As a result, we achieve more convincing dehazing results. To summarize, this paper makes the following contributions: (1) We propose a novel Uncertainty-driven Dehazing Network (UDN) to effectively and adaptively improve the feature quality and generate confident hazy-free images. Compared with the state-of-the-arts, UDN achieves the best dehazing results on both synthetic datasets and real-world images. (2) We propose an Uncertainty Estimation Block (UEB) to capture the aleatoric and epistemic uncertainty of each pixel's dehazing result together.(3) We develop an Uncertainty-aware Feature Modulation (UFM) block to adaptively enhance the learned features, which predicts a convolution kernel and channel-wise modulation coefficients conditioned on the uncertainty weighted representation. (4) We present an uncertainty-driven self-distillation loss to improve the uncertain representation of a pixel via transferring confident knowledge from its similar pixels' representations.

Related Work

Single image dehazing. Image dehazing methods can be divided into traditional and deep learning-based methods. Most traditional dehazing methods depend on the physical scattering model which needs to estimate the transmission map and the global atmospheric light. The representative work is prior based dehazing, such as the dark channel prior (DCP) [He, Sun, and Tang 2009], non-local haze-line prior [Berman, Treibitz, and Avidan 2016] and color attenuation prior [Zhu, Mai, and Shao 2015], etc. Despite achieving promising results, these methods are not robust due to the strong assumptions.

With the rising of deep learning, deep dehazing models are developed, which are grouped into two classes: physical-model dependent methods and physical-model free ones. In the first class, the intermediate variables of the physical scatter model may not be estimated accurately, which results in the degradation of the dehazing performance. Recently, more and more attention is paid to physical-model free dehazing. Qu *et al.* [Qu et al. 2019] built the Pix2Pix dehazing model via adversarial training. Liu *et al.* [Liu et al. 2019] introduced an attention-based multi-scale estimation network named GridDehaze for image dehazing. However, the methods mentioned above are not robust to the uneven distribution of haze. FFA-Net [Qin et al. 2020] was proposed to implement a joint attention mechanism that combines channel attention and pixel attention for uneven haze removal. However, the attention maps are learned in an implicit and unexplainable way that fails to build an explicit and direct relationship with the restoration of each pixel. To further boost clean image prediction, we embed uncertainty estimation into our dehazing network and pay more attention to improving the uncertain representation.

Uncertainty Estimation. Bayesian deep learning can be used to model two types of uncertainty [Kendall and Gal 2017]: 1) aleatoric uncertainty that accounts for the noisy measurement and 2) epistemic uncertainty that accounts for uncertainty in the model parameters. Recently, in [Yasarla and Patel 2019], uncertainty is used to guide a deep deraining model in blocking the flow of incorrect estimation in rain streaks. In [Zhang et al. 2019a], the uncertainty of image reconstruction is investigated which is caused by the partially k-space observation for MRI reconstruction. In our approach, we focus on reducing the uncertain representation by making full use of the confident one to obtain an accurate and confident dehaizng result.

Method

Our goal is to restore a clear image $\hat{J}$ with better confidence (or less uncertainty) from its hazy observation I. To this end, we present an Uncertainty-driven Dehazing Network (UDN), focusing on reducing the uncertain feature repre-

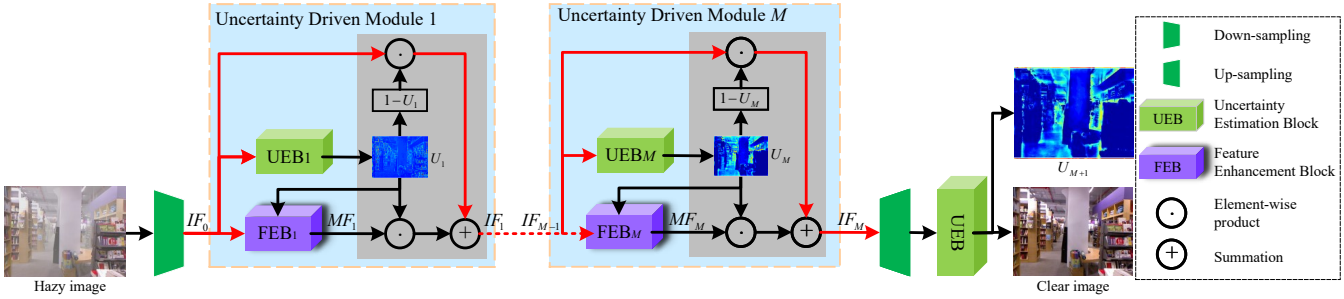


Figure 2: Overview of the proposed UDN. The red lines represent the flow of improved feature IF_m $\{m = 1, ..., M\}$. UDN consists of M Uncertainty Driven Modules (UDMs), in which UEB$_m$ is used to estimate the uncertainty map U_m together with an intermediate dehazing result $\hat{J}_m$. FEB$_m$ is used to modulate IF_{m-1} to generate the modulated feature MF_m, where FEB$_m$ enhances the uncertain feature from IF_{m-1}, and a gate unit (the gray part) is used to aggregate IF_{m-1} and MF_m to obtain a more confident improved representation IF_m.

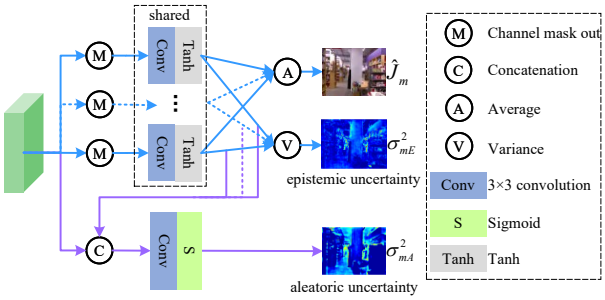


Figure 3: The Uncertainty Estimation Block (UEB).

sentation and making full use of the confident one pixel-wise and channel-wise. As illustrated in Figure 2, UDN is built with M Uncertainty Driven Modules (UDMs) that are connected by a path to improve the feature, denoted as Improved Feature (IF). In the mth UDM $(m = 1, ..., M)$, an Uncertainty Estimation Block (UEB) is designed to estimate the uncertainty map U_m together with an intermediate dehazing result $\hat{J}_m$, and a Feature Enhancement Block (FEB) is employed to modulate IF_{m-1} via exploiting the confident feature and obtain the Modulated Feature MF_m. We linearly combine MF_m and IF_{m-1} with the uncertainty map U_m as the gate. Thus, UDM updates the representation in IF_{m-1} with MF_m and outputs a more confident improved representation IF_m as follows:

$$IF_m = (1 - U_m) \odot IF_{m-1} + U_m \odot MF_m, \qquad (1)$$

where $\odot$ represents the element-wise product.

Concretely, we first down-sample the input to 1/16 of its original size with two strided convolutions for fast computing and obtain the initial feature IF_0. After that, IF_0 is updated gradually by M UDMs to obtain $IF_1, IF_2, \cdots, IF_M$. Finally, IF_M is upsampled to the original size by two up-sampling layers. And the final haze-free output along with its uncertainty map U_{M+1} is generated by another UEB.

Aleatoric and Epistemic Uncertainty Estimation

Bayesian deep learning can be used to model two types of uncertainty: 1) aleatoric uncertainty that comes from noise inherent in the observations; 2) epistemic uncertainty that accounts for uncertainty in the model [Kendall and Gal 2017]. These two kinds of uncertainty also exist in the dehazing models, but previous methods can only obtain a deterministic result without knowing its confidence. In this paper, we introduce an Uncertainty Estimation Block (UEB) to model each pixel's aleatoric uncertainty σ_A^2 and epistemic uncertainty σ_E^2 together. As shown in Figure 3, UEB contains an aleatoric uncertainty estimation branch and an epistemic uncertainty estimation branch.

Inspired by [Kendall and Gal 2017], we assume the dehazing pixel-wise output being $p(J|\hat{J}, \theta)$ is a Gaussian distribution, with the mean being the ground-truth image J and variance being σ^2, where θ, σ, and $\hat{J}$ denote the network parameters, the observation noise of all pixels, and the restored result, respectively. To find the optimal $\hat{\theta}$, we perform Maximum A Posteriori (MAP) inference:

$$\begin{aligned} \hat{\theta} &= \arg\max_{\theta} log(p(J|\hat{J}, \theta)) \\ &= \arg\max_{\theta} \{-\frac{1}{2\sigma^2}\|J - \hat{J}\|_2^2 - \frac{1}{2} log\sigma^2\}. \end{aligned} \qquad (2)$$

We treat $\sigma_A^2 = \sigma^2$, and predict σ_A^2 and $\hat{J}$ using two branches in UEBs. To capture σ_A^2, the mth UEB uses a Conv-Sigmoid layer to predict it conditioned on the dehazing results and constrained by the following minimization objective,

$$L_r^m = \frac{1}{D_m} \sum_{i=1}^{D_m} [\frac{1}{2(\sigma_{mA}^i)^2}(J^i - \hat{J}_m^i)^2 + \frac{1}{2} \log(\sigma_{mA}^i)^2], \qquad (3)$$

where D_m is the number of the output pixels, and the superscript i denotes the pixel index. Different from aleatoric uncertainty, epistemic uncertainty is a property of the model which is captured by replacing the deterministic network's weight parameters with distributions over these parameters and averaging over all possible weights. However, it is time-consuming to perform inference. To build an efficient

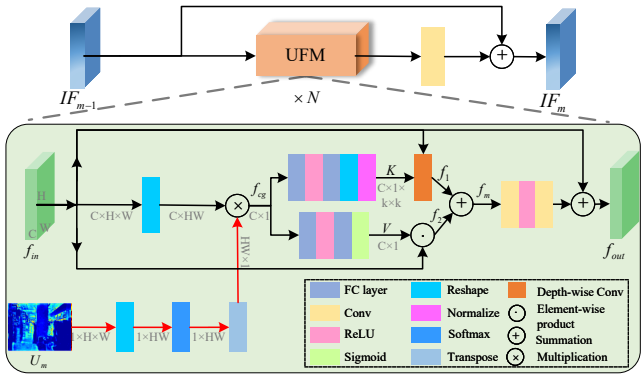


Figure 4: Feature Enhancement Block (FEB) that includes N Uncertainty-aware Feature Modulation (UFM) blocks.

dehazing model, we only place the distributions over the last reconstruction layer (Conv-Tanh) of each UEB and introduce a mask-out operation to approach approximate inference. Specifically, we firstly randomly mask out $q\%$ of the input feature channels (*i.e.,* set the values of these channels to 0) and then pass them to a shared Conv-Tanh layer to reconstruct a clear image. We repeat this process for T times and output T different dehazing results $\{\hat{J}_{m,t}\}_{t=1}^{T}$. After that, we calculate the mean and variance (epistemic uncertainty) of $\{\hat{J}_{m,t}\}_{t=1}^{T}$, *i.e.,* the predicted mean $\hat{J}_m = \frac{1}{T}\sum_{t=1}^{T}\hat{J}_{m,t}$ and the epistemic uncertainty $\sigma_{mE}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{J}_{m,t}^2 - (\frac{1}{T}\sum_{t=1}^{T}\hat{J}_{m,t})^2$ that measures how much the model is uncertain about the prediction. This process is similar to the dropout that can be interpreted as a Bayesian Approximation [Gal and Ghahramani 2016].

To summarize, the predicted uncertainty of each pixel in the mth UEB can be approximated using:

$$\begin{aligned} U_m &\approx \sigma_{mE}^2 + \sigma_{mA}^2 \\ &= \frac{1}{T}\sum_{t=1}^{T}\hat{J}_{m,t}^2 - (\frac{1}{T}\sum_{t=1}^{T}\hat{J}_{m,t})^2 + \sigma_{mA}^2. \end{aligned} \tag{4}$$

With the help of the uncertainty map U_m from UEB$_m$, we can know the uncertainty of each pixel. In the following, we will describe how to exploit the predicted uncertainty map and how to transfer the knowledge learned by confident pixels to uncertain ones to improve the dehazing performance.

Feature Enhancement Block (FEB)

The goal of FEB is to improve the uncertain feature. Due to the difference between image features and uncertainty map, directly using U_m as a condition and concatenating it with the features as FEB's inputs introduce bad interference. It is observed that the estimated uncertainty map is directly related to the reconstruction error, and the hard-to-dehaze regions, *i.e.,* the edges of the objects, always have larger uncertainty values. According to this, we develop an Uncertainty-aware Feature Modulation (UFM) block to adapt the features based on U_m, as illustrated in Figure 4. Suppose the input and output features of UFM are $f_{in} \in \mathbb{R}^{C\times H\times W}$ and $f_{out} \in \mathbb{R}^{C\times H\times W}$, respectively. Firstly, we use U_m to distinguish each pixel and obtain an confidence-aware global statistic $f_{cg} = \sum_{i=1}^{D_m} w^i f_{in}^i \in \mathbb{R}^C$, where $w^i = 1 - \frac{e^{U_m^i}}{\sum_{j=1}^{D_m} e^{U_m^j}}$ is the normalized weight of pixel i, and the more uncertain pixel has a small weight.

Then, we adapt f_{in} using two seperate paths conditioned on f_{cg}. In the first path, we feed f_{cg} to two full-connected (FC) layers followed by a reshape layer to predict the kernel $K \in \mathbb{R}^{C\times 1\times k\times k}$ of a depth-wise convolution, where $k \times k$ is the kernel size. Then f_{in} is processed with a depth-wise convolution (using K) to produce f_1. Note that the generated kernel values can be very large or small for some features, and directly using K for convolution will make the network unstable. Thus, we propose a kernel normalize layer here, where $K_{lnkk} \leftarrow K_{lnkk}/\sqrt{\sum_{kk} K_{lnkk}^2 + \varepsilon}$, and ε is a small constant to avoid numerically unstable. This operation restores f_1 back to the unit standard deviation. Moreover, we feed f_{cg} to another two FC layers, which is followed by a sigmoid layer to predict each channel's weight $v \in \mathbb{R}^C$. Then we perform channel-wise feature modulation using v, and obtain $f_2 = f_{in} \odot v$, where $\odot$ is element-wise product. After that, we obtain the intermediate feature $f_m = f_1 + f_2$.

Finally, we input f_m to two 3×3 convolution layers and add the result to f_{in} to produce the output feature f_{out}. We observe that simply stacking N UFMs to build a deep FEB can hardly achieve performance gain, so we design FEB by stacking N UFMs followed by a 3×3 convolutional layer and a skip connection, as illustrated in Figure 4.

UFM adapts the feature by learning a convolution kernel K and channel-wise coefficients v based on the uncertainty map. Therefore UFM can well exploit the uncertainty information. The adaptive K and v make every UFM different and dynamic. It is demonstrated in the ablation study that our UDN benefits from UFM to achieve better performance.

Uncertainty-Driven Self-Distillation Loss

It is observed that the colors of a hazy-free image are well approximated by a few hundred distinct colors [Berman, Avidan et al. 2016] and we can reconstruct a clear image by capturing the long-range dependency between pixels. The characteristic of non-local similarity implies that two distant pixels with the same color in the clear image may suffer from different degradations in their hazy image. A dehazing network needs to map these two different hazy pixels to the same color, forming a difficult many-to-one mapping. To ease the problem, we assume pixels with similar hazy-free colors are reconstructed from a similar representation. That is, the pixels which are similar in the clear image should also be similar in the feature domain. Based on this assumption, we propose a self-distillation loss $\mathcal{L}_{SD}$. Given a pair of hazy and clear images, we first search for the *i*-th pixel's S most similar pixels in the haze-free image and denote their indexes by a set $\Omega_m^i, (m = 1, \cdots, M)$. We treat the average representation of the S pixels as the prototype and force the feature of the *i*-th pixel IF_m^i close to it. In this way, the knowledge of non-local similar pixels is transferred each other. The whole process is illustrated in Figure 6 and the

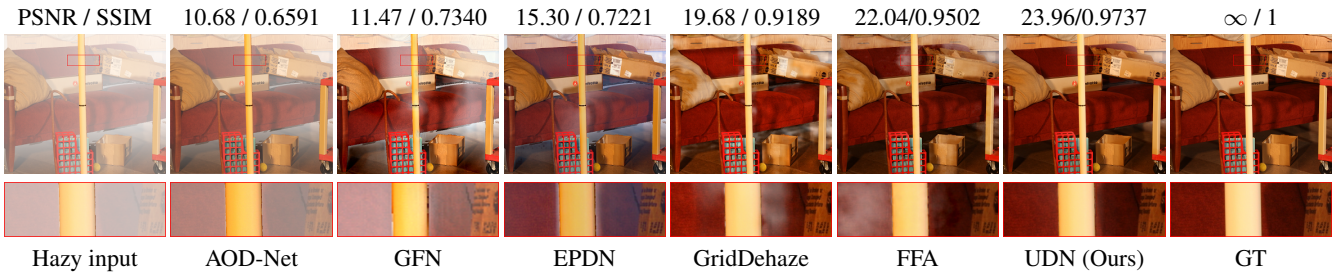


Figure 5: An example of dehazed images (better view by zooming in).

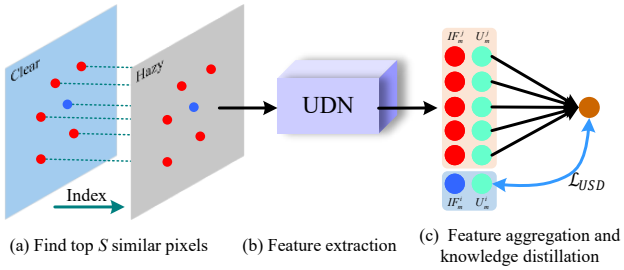


Figure 6: Illustration of the uncertainty-driven self-distillation loss.

self-distillation loss $\mathcal{L}_{SD}$ is formulated as:

$$\mathcal{L}_{SD}^m = \frac{1}{D_m}\sum_{i=1}^{D_m}(IF_m^i - \frac{1}{S}\sum_{j\in\Omega_m^i} IF_m^j)^2. \quad (5)$$

The contextual information may be different between non-local similar pixels, leading to the features extracted from the hazy pixels with different degrees of uncertainty. Hence, treating all similar pixels in Ω_m^i equally may over-emphasize the uncertain features and ignore the confident ones. Besides, if a pixel can obtain a confident representation, there is no need to distill knowledge from its similar pixels. Therefore, we introduce the uncertainty estimate of every pixel into $\mathcal{L}_{SD}$ and propose the uncertainty-driven self-distillation loss $\mathcal{L}_{USD}$ as:

$$\mathcal{L}_{USD}^m = \frac{1}{D_m}\sum_{i=1}^{D_m} U_m^i(IF_m^i - \sum_{j\in\Omega_m^i}(1-z_m^j)IF_m^j)^2, \quad (6)$$

where every pixel has a loss weight U_m^i and more knowledge is distilled from the features of the more confident similar pixels with a normalized weight $z_m^j = \frac{e^{U_m^j}}{\sum_{j\in\Omega_m^i} e^{U_m^j}}$.

Training Strategy

Due to M UDMs in UDN, there are $M+1$ UEBs to estimate the dehazing results and uncertainty maps. We formulate the overall objective as follows:

$$\mathcal{L} = \sum_{m=1}^{M+1}\theta_m(\mathcal{L}_r^m + \lambda_p\mathcal{L}_p^m) + \lambda_u\sum_{m=1}^{M}\theta_m\mathcal{L}_{USD}^m, \quad (7)$$

where $\mathcal{L}_r^m$ is the reconstruction loss described in Eq. 3, $\mathcal{L}_p^m$ is the perceptual loss [Johnson, Alahi, and Fei-Fei 2016], $\mathcal{L}_{USD}^m$ is the uncertainty-driven self-distillation loss described in Eq. 6, and λ_p and λ_u are two weight factors. The perceptual loss $\mathcal{L}_p$ is defined as:

$$\mathcal{L}_p^m = \left\|\Phi(J) - \Phi(\hat{J}_m)\right\|_2, \quad (8)$$

where Φ is a feature extractor of VGG16 [Sim et al. 2018] pre-trained on ImageNet [Deng et al. 2009].

In UDN, the former UEBs can hardly reconstruct a good result, and thus their large losses make the training excessively focus on them and adversely affect the training of the latter UEBs. Hence, we gradually increase the loss weights of UEBs, and set $\theta_m = 0.8^{M+1-m}$ empirically.

Experiments

Implementation Details. We implement UDN in the PyTorch 1.2.0 framework with an NVIDIA RTX 2080 GPU. We use batch-size of 2 and a patch-size of 256 × 256 pixels for training. Samples are augmentated by random rotation and horizontally flipping. Adam optimizer is used with an initial learning rate of 0.0001 and is scheduled by cosine decay [Athiwaratkun et al. 2019]. The model is trained for 300 epoches. The parameters M and N are both set to 6, which means we use 6 UDMs in UDN and each contains 6 UFMs. All convolutional layers have $C = 64$ channels. Besides, we empirically set $\lambda_p = 1$, $\lambda_u = 0.1$, $S = 10$, $T = 5$, and $q = 10$. Source codes (implemented in MindSpore and Pytorch) will be released later.

Datasets. We evaluate the proposed method on the RESIDE dataset [Li et al. 2018] which contains an Indoor Training Set (ITS), an Outdoor Training Set (OTS) and a Synthetic Objective Testing Set (SOTS). Specifically, ITS contains 13,990 synthetic hazy images generated from NYU Depth V2 [Nathan Silberman and Fergus 2012], OTS contains 313,950 synthetic hazy images generated from 8,970 outdoor scenes, and SOTS consists of an indoor test set and an outdoor test set, which includes 1,000 indoor/outdoor hazy images generated from 100 different indoor/outdoor scenes. Besides, we evaluate our model trained with ITS on the Middlebury dataset [Cosmin Ancuti 2016] which contains 23 hazy images generated from high-quality real scenes. We also give a quantitative evaluation on a real-world dataset O-HAZE [Ancuti et al. 2018], which contains 45 pairs of outdoor scenes recorded in haze-free and hazy conditions.

(a) Input image (b) AOD-Net (c) GFN (d) EPDN (e) GridDehaze (f) FFA-Net (g) UDN

Figure 7: Dehazing real-world hazy photos using various methods. Please zoom in for a better view.

Method	SOTS indoor		SOTS outdoor		Middlebury		O-HAZE		#param
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
AOD-Net (ICCV'17)	19.06	0.8504	20.29	0.8765	13.40	0.7979	19.586	0.679	1833
GFN (CVPR'18)	22.30	0.8800	21.55	0.8444	13.27	0.7514	17.645	0.612	0.54M
EPDN (CVPR'19)	25.06	0.9232	22.57	0.8630	15.28	0.8096	16.309	0.686	17.38M
GridDehaze (ICCV'19)	32.16	0.9836	30.86	0.9819	14.21	0.7783	21.913	0.730	0.95M
FFA-Net (AAAI'20)	35.77	0.9846	33.38	0.9804	17.32	0.8522	20.836	0.679	4.45M
KDDN (CVPR'20)	34.72	0.9845	-	-	17.27	0.8676	**25.455**	0.780	5.99M
AECR-Net (CVPR'21)	37.17	0.9901	-	-	-	-	-	-	2.61M
UDN$^-$ (ours)	37.71	0.9903	34.18	0.9851	17.64	0.8648	23.431	0.719	1.01M
UDN (ours)	**38.62**	**0.9909**	**34.92**	**0.9871**	**18.43**	**0.8854**	25.412	**0.785**	4.25M

Table 1: Quantitative comparison on three synthetic datasets and a real-world dataset in terms of PSNR, SSIM and number of parameters (#param/M-Million). The sign "-" denotes the number is unavaliable.

Results on Synthetic Datasets

In Table 1, we summarize the performance and number of parameters of our UDN and the seven SOTA approaches on SOTS indoor, SOTS outdoor, and Middlebury datasets. Our UDN achieves the best dehazing results among all methods on all datasets. In particular, compared to the previous best method AECR-Net [Wu et al. 2021] on the SOTS indoor dataset, our UDN achieves 1.3dB PSNR performance gains. Figure 5 presents a visual comparison of a hazy image from Middlebury [Cosmin Ancuti 2016]. It is observed that AOD-Net and GFN cannot successfully remove haze, and their dehazing results still suffer from serious color distortion. Although EPDN, GridDehaze, and FFA fail to restore the dense hazy area, they achieve better results in the light hazy area. Our method generates the most natural result and achieves similar colors and details to the ground truth in both the light and dense hazy regions. The PSNR and SSIM values presented in Figure 5 can also verify the superiority of UDN. To reduce the parameters of our model UDN, we also implement it by sharing all the parameters of the six UDMs. This version of UDN is denoted as UDN$^-$. The parameters of UDN$^-$ are reduced to 0.9M. Even with few parameters, UDN$^-$'s performance is still better than most of the compared methods.

Results on Real-World Hazy Scenes

Results on O-HAZE. We follow the setting of NTIRE Image Dehazing Challenge [Ancuti, Ancuti, and Timofte 2018] and evaluate the dehazing models by re-training UDN on the training set of O-HAZE [Ancuti et al. 2018]. Table 1 presents the quantitative results of our network and the state-of-the-arts. Obviously, our model outperforms other methods in terms of PSNR and SSIM. It demonstrates that UDN can achieve satisfactory results on images captured in real-world outdoor scenes. The visual comparison presented in Figure 1 also verifies it.

Results on Real Hazy Photographs. Additionally, Figure 7 shows the visual comparisons of two real-world hazy photos collected by previous works. As observed, AOD-Net, GFN, and EPDN can only remove the haze in the near scenes and fail in far and severely degraded regions. The results of FFA-Net and GridDehaze tend to leave haze or darken some regions that may be caused by the overfitting. In contrast, our UDN generates the most natural and sharp results in the whole image.

Ablation Study

We perform an ablation study to investigate the effectiveness of the loss functions and the main components of the proposed method. We first construct a **base** network as our baseline, which consists of one downsampling layer, six FEBs each with six residual blocks, and one upsampling layer. Subsequently, we construct five variants: (1) **base+UEB**: Add six UEBs into the baseline. (2) **base+UEB+Gate**: Add the gate unit (Eq. 1) into **base+UEB**. (3) **base+UEB+Gate+UFM**: Replace the residual blocks of FEBs in **base+UEB** with UFMs. The losses in **base**, **base+UEB**, **base+UEB+Gate**,

σ_E^2 σ_A^2 σ_E^2 +σ_A^2 error

Figure 8: Visualization of the uncertainty maps and prediction errors on a test image of SOTS. From top to down are the results from the 1st and 4th UEBs. From left to right are the epistemic uncertainty σ_E^2, aleatoric uncertainty σ_A^2, fused uncertainty σ_E^2 +σ_A^2 and error maps. Note that since the values of epistemic uncertainty are small, we scale them for better view.

Model	PSNR	SSIM
base	35.53	0.9866
base+UEB	36.6	0.9884
base+UEB+Gate	36.72	0.9889
base+UEB+Gate+UFM	38.21	0.9904
base+UEB+Gate+UFM+$\mathcal{L}_{SD}$	38.39	0.9906
base+UEB+Gate+UFM+$\mathcal{L}_{USD}$	**38.62**	**0.9909**

Table 2: Ablation study on UDN with different architectures and loss functions on the SOTS indoor test set.

and **base+UEB+Gate+UFM** are all $\mathcal{L}_r^m + \lambda_p \mathcal{L}_p^m$. (4) **base+UEB+Gate+UFM+$\mathcal{L}_{SD}$**: Its loss function is similar to Eq. 7 but with $\mathcal{L}_{USD}^m$ replaced by $\mathcal{L}_{SD}^m$. (5) **base+UEB+UFM+$\mathcal{L}_{USD}$**: Its loss function is Eq. 7. We use the ITS dataset for training and SOTS indoor test set for evaluation. Their performances are summarized in Table 2.

Effectiveness of UEB. UDN learns the uncertainty maps with UEBs and employs the maps learned by UDM as guidance to focus on enhancing the uncertain representation. In Table 2, it shows that UEB achieves 1.19dB PSNR gains over the base network. Futhermore, we evaluate the effect of the two types of uncertainty. As reported in Table 3, using different types of uncertainty can all improve the performances, and using the fused uncertainty $U = \sigma_A^2 + \sigma_E^2$, both the base model and the full UDN model can achieve the best results. We further visualize the uncertainty maps and reconstruction errors from different UEBs in Figure 8. From top to down, we can observe that the uncertainty is decreased gradually, which indicates that our model can improve the representation quality. From left to right, we can see that the uncertainty maps are related to the reconstruction errors, and the larger the error, the higher the uncertainty value.

Effectiveness of UFM. UFM significantly improves the performance against base+UEB with an increase of 1.49dB PSNR. Therefore, UFM is effective in adaptively modulating the features. We also evaluate the effectiveness of the kernel K and channel-wise weight v in Table 4. It shows

Model	PSNR	SSIM	Model	PSNR	SSIM
base	35.53	0.9866	-	-	-
base+σ_E^2	35.84	0.9871	UDN+σ_E^2	37.89	0.9893
base+σ_A^2	36.23	0.9876	UDN+σ_A^2	38.21	0.9901
base +U	**36.6**	**0.9884**	UDN+U	**38.62**	**0.9909**

Table 3: Comparisions of different types of uncertainty. Due to the values of epistemic uncertainty are small, we remove the gate unit (Eq. 1) in UDN+σ_E^2.

that both the modulations can improve the performance, and using them together obtain the best result.

Model	PSNR	SSIM
base+UEB	36.72	0.9889
base+UEB+UFM$_K$	37.34	0.9895
base+UEB+UFM$_v$	37.78	0.9899
base+UEB+UFM	**38.21**	**0.9904**

Table 4: Comparisions of different types of UFM. UFM$_K$ and UFM$_v$ denote the model using only K and only v, respectivelly. UFM uses both of them.

Effectiveness of $\mathcal{L}_{USD}$. $\mathcal{L}_{SD}$ forces more self-similarity in IF_m and $\mathcal{L}_{USD}$ is derived from $\mathcal{L}_{SD}$. As reported in Table 2 with $\mathcal{L}_{SD}$ and $\mathcal{L}_{USD}$ we can obtain better dehazing results which benefit from the increasing of self-similarity. Moreover, to verify the assumption that similar pixels in the clear image should also be similar in the feature domain, we first randomly select a point in the GT image and choose its top 10 similar pixels, denoted as a set A. Then, we calculate its top 10 similar pixels in the feature domain, denoted as a set B. Lastly, we calculate the Intersection of Union (IOU) of set A and set B. The average results of 10 000 randomly selected points from 100 images are 0.815 and 0.572 with L_{USD} and without L_{USD}, respectively. It indicates that our assumption is reasonable and using it we can achieve better results.

Conclusion

This work presents an uncertainty-driven dehazing network, UDN, for obtaining a reliable and clear dehazing result. Specifically, we develop an Uncertainty Estimation Block (UEB) to predict the aleatoric and epistemic uncertainty together. With the help of estimated uncertainty maps, we propose an Uncertainty-aware Feature Modulation (UFM) block to adaptively enhance the learned features. UFM predicts a convolution kernel and channel-wise modulation coefficients conditioned on the uncertainty weighted representation. Moreover, an uncertainty-driven self-distillation loss is presented to effectively transfer the knowledge from confident representation to uncertain one and improve the feature self-similarity. Extensive experimental results on synthetic datasets and real-world images show that UDN achieves significant quantitative and qualitative improvements, outperforming the state-of-the-arts.

Acknowledgments

This work was supported by the National Key Research and Development Programme of China No.2020AAA0108301, the National Natural Science Foundation of China under Grant No.61876161 and No.62176224, and CAAI-Huawei MindSpore Open Fund.

References

Ancuti, C.; Ancuti, C. O.; and Timofte, R. 2018. NTIRE 2018 Challenge on Image Dehazing: Methods and Results. In *CVPR Workshops*.

Ancuti, C. O.; Ancuti, C.; Timofte, R.; and Vleeschouwer, C. D. 2018. O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images. In *CVPR, NTIRE Workshop*.

Athiwaratkun, B.; Finzi, M.; Izmailov, P.; and Wilson, A. G. 2019. There Are Many Consistent Explanations of Unlabeled Data: Why You Should Average. In *ICLR*.

Berman, D.; Avidan, S.; et al. 2016. Non-local image dehazing. In *CVPR*.

Berman, D.; Treibitz, T.; and Avidan, S. 2016. Non-local Image Dehazing. In *CVPR*, 1674–1682.

Cai, B.; Xu, X.; Jia, K.; Qing, C.; and Tao, D. 2016. DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Trans. Image Processing*, 25(11): 5187–5198.

Chen, J.; Wen, S.; and Chan, S. G. 2021. Joint Demosaicking and Denoising in the Wild: The Case of Training Under Ground Truth Uncertainty. In *AAAI*.

Cosmin Ancuti, C. D. V., Codruta O. Ancuti. 2016. D-HAZY: a dataset to evaluate quantitatively dehazing algorithms. ICIP.

Deng, J.; Dong, W.; Socher, R.; Li, L. J.; and Li, F. F. 2009. ImageNet: a Large-Scale Hierarchical Image Database. In *CVPR*.

Gal, Y.; and Ghahramani, Z. 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In Balcan, M.; and Weinberger, K. Q., eds., *ICML*.

He, K.; Sun, J.; and Tang, X. 2009. Single image haze removal using dark channel prior. In *CVPR*.

Johnson, J.; Alahi, A.; and Fei-Fei, L. 2016. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*.

Kendall, A.; and Gal, Y. 2017. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In *NeurIPS, 2017*.

Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; and Wang, Z. 2018. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1): 492–505.

Liu, X.; Ma, Y.; Shi, Z.; and Chen, J. 2019. GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing. In *ICCV*.

Liu, Y.; Zhao, G.; Gong, B.; Li, Y.; Raj, R.; Goel, N.; Kesav, S.; Gottimukkala, S.; Wang, Z.; Ren, W.; and Tao, D. 2018. Improved Techniques for Learning to Dehaze and Beyond: A Collective Study. *CoRR*, abs/1807.00202.

Nathan Silberman, P. K., Derek Hoiem; and Fergus, R. 2012. Indoor Segmentation and Support Inference from RGBD Images. In *ECCV*.

Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; and Jia, H. 2020. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. In *AAAI*.

Qu, Y.; Chen, Y.; Huang, J.; and Xie, Y. 2019. Enhanced Pix2pix Dehazing Network. In *CVPR*.

Ren, W.; Liu, S.; Zhang, H.; Pan, J.; Cao, X.; and Yang, M. 2016. Single Image Dehazing via Multi-scale Convolutional Neural Networks. In *ECCV*, 154–169.

Ren, W.; Zhang, J.; Xu, X.; Ma, L.; Cao, X.; Meng, G.; and Liu, W. 2018. Deep video dehazing with semantic segmentation. *IEEE Transactions on Image Processing*, 28(4): 1895–1908.

Sim, H.; Ki, S.; Choi, J.-S.; Seo, S.; Kim, S.; and Kim, M. 2018. High-Resolution Image Dehazing With Respect to Training Losses and Receptive Field Sizes. In *CVPR Workshops*.

Tan, R. T. 2008. Visibility in bad weather from a single image. In *CVPR*.

Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; and Ma, L. 2021. Contrastive Learning for Compact Single Image Dehazing. In *CVPR*.

Yasarla, R.; and Patel, V. M. 2019. Uncertainty Guided Multi-Scale Residual Learning-Using a Cycle Spinning CNN for Single Image De-Raining. In *CVPR*.

Zhang, L.; Qi, G.-J.; Wang, L.; and Luo, J. 2019a. AET vs. AED: Unsupervised Representation Learning by Auto-Encoding Transformations Rather Than Data. In *CVPR*.

Zhang, Z.; Romero, A.; Muckley, M. J.; Vincent, P.; Yang, L.; and Drozdzal, M. 2019b. Reducing Uncertainty in Undersampled MRI Reconstruction With Active Acquisition. In *CVPR*.

Zheng, E.; Yu, Q.; Li, R.; Shi, P.; and Haake, A. R. 2021. A Continual Learning Framework for Uncertainty-Aware Interactive Image Segmentation. In *AAAI*.

Zhu, Q.; Mai, J.; and Shao, L. 2015. A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior. *IEEE Trans. Image Processing*, 24(11): 3522–3533.