Bayesian Neural Networks for One-to-Many Mapping in Image Enhancement

Guoxi Huang[1], **Qirui Yang**[2], **Ruirui Lin**[1], **Zipeng Qi**[3], **David Bull**[1], **Nantheera Anantrasirichai**[1]

[1] Visual Information Laboratory, University of Bristol
[2] Tianjin University
[3] Baidu Inc
{guoxi.huang, r.lin, n.anantrasirichai, dave.bull}@bristol.ac.uk, yangqirui@tju.edu.cn, qizipeng1107@gmail.com

Abstract

In image enhancement tasks, such as low-light and underwater image enhancement, a degraded image can correspond to multiple plausible target images due to dynamic photography conditions. This naturally results in a one-to-many mapping problem. To address this, we propose a Bayesian Enhancement Model (BEM) that incorporates Bayesian Neural Networks (BNNs) to capture data uncertainty and produce diverse outputs. To enable fast inference, we introduce a BNN-DNN framework: a BNN is first employed to model the one-to-many mapping in a low-dimensional space, followed by a Deterministic Neural Network (DNN) that refines fine-grained image details. Extensive experiments on multiple low-light and underwater image enhancement benchmarks demonstrate the effectiveness of our method.

Code — https://github.com/BinCVER/BEM

1 Introduction

Image enhancement refers to the process of improving visual quality, primarily by adjusting illumination, as well as reducing noise, correcting colors, and refining structural details. The perceived quality of the enhanced image varies, as it is influenced by personal preferences and context-specific requirements. In low-light image enhancement (LLIE) and underwater image enhancement (UIE) tasks, a significant challenge arises from the *one-to-many mapping* problem, where a single degraded input image can correspond to multiple plausible target images. As illustrated in Fig. 1 (top), some reference images are unreliable due to poor visibility during image acquisition, which is often caused by challenging environments and limitations of imaging equipment.

Recent advances in deep learning have steered image enhancement towards data-driven approaches, with several models (Peng, Zhu, and Bian 2023; Cai et al. 2023; Li et al. 2021; Huang et al. 2025b) achieving state-of-the-art (SOTA) results by employing deterministic neural networks (DNNs) to learn one-to-one mappings between inputs and ground-truth images using paired datasets. For LLIE and UIE tasks in particular, the ambiguity of target images makes DNNs poorly suited to capture the inherent variability in one-to-many image pairs, as illustrated in Fig. 1 (left).

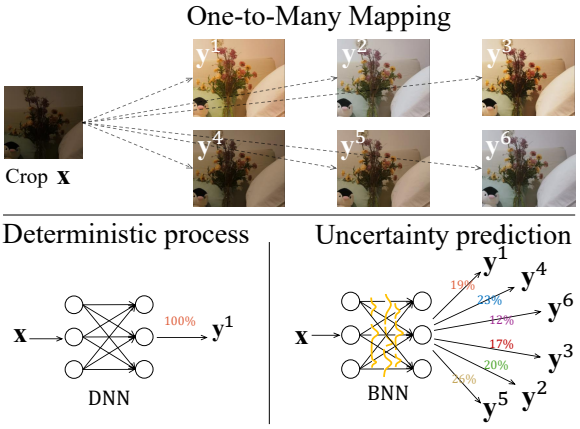

Figure 1: One-to-Many Mapping where an image crop $\mathbf{x}$ associated with multiple targets $\{\mathbf{y}^1, \ldots, \mathbf{y}^6\}$. A DNN (left) can only predict one of the targets. In contrast, a BNN (right) can produce many predictions according to a learned probability distribution.

In these enhancement tasks, ground-truth images are collected in real-world environments, which inevitably introduces low-quality targets into the training data—i.e., label noise. Such label noise is further amplified in extremely low-visibility data collection environments—particularly in challenging underwater or low-light scenes—where obtaining high-quality ground truth becomes nearly impossible. As a result, learning a deterministic mapping from an input to a noisy ground truth can jeopardize enhancement quality. However, directly discarding low-quality image pairs from training data leads to degraded performance in difficult scenes and harms the model's generalization ability.

In this paper, we use a Bayesian Neural Network (BNN) to probabilistically model the one-to-many mappings between inputs and targets: We leverage Bayesian inference to sample network weights from a learned posterior distribution, where each sampled set of weights corresponds to a distinct plausible output. Through multiple sampling processes, the model maps a single input to a distribution of possible outputs, as illustrated in Fig. 1 (right). Then, a ranking-based selection or Monte Carlo sampling is employed to ob-

tain a reliable final output.

Although BNNs show promise in capturing uncertainty across various tasks (Kendall and Cipolla 2016; Kendall, Gal, and Cipolla 2018), their potential for modeling the one-to-many mapping in image enhancement remains largely under-explored, despite clear benefits. By incorporating Bayesian inference into the enhancement process, our approach captures uncertainty in dynamic, uncontrolled environments, providing a more flexible and robust solution than deterministic models.

Applying BNNs to high-resolution image tasks poses notable challenges: 1) BNNs with high-dimensional weight spaces often suffer from underfitting (Tomczak et al. 2021), hindering their ability to learn complex mappings effectively. To address this, we propose the *Adaptive Prior*, which stabilizes training and accelerates convergence. 2) Producing multiple high-resolution outputs with a BNN incurs substantial inference latency, making real-time processing impractical. To overcome this, we propose a two-stage BNN-DNN framework (Sec. 3.2) that captures one-to-many mappings in a compact low-dimensional space, significantly reducing computational cost while maintaining high-quality predictions.

We explore the feasibility of BNNs on the LLIE and UIE tasks where the *one-to-many mapping* problem is particularly pronounced. The main contributions of this paper are summarized as follows: 1) We identify the one-to-many mapping between inputs and targets as a key bottleneck in image enhancement models for LLIE and UIE, and propose a BNN-based method to address this challenge; 2) We design a two-stage BNN-DNN framework for efficient inference, enabling low-latency prediction by avoiding the explicit generation of multiple low-quality outputs. 3) We demonstrate that our method is backbone-agnostic and can benefit from future advances in backbone architectures.

2 Background

Bayesian Deep Learning. BNNs quantify uncertainty by learning distributions over network weights, offering robust predictions (Neal 2012). Variational Inference (VI) is a common method for approximating these distributions (Blundell et al. 2015). Gal and Ghahramani (2016) simplified the implementation of BNNs by interpreting dropout as an approximate Bayesian inference method. Another line of approaches, such as Krishnan, Subedar, and Tickoo (2020), explored the use of empirical Bayes to specify weight priors in BNNs to enhance the model's adaptability to diverse datasets. These BNN approaches have shown promise across a range of vision applications, including camera relocalization (Kendall and Cipolla 2016), semantic and instance segmentation (Kendall, Gal, and Cipolla 2018). Despite these advances, BNNs remain underutilized in image enhancement tasks.

Visual Enhancement. DNN-based methods (Zamir et al. 2022; Malyugina et al. 2025) have been widely adopted for visual enhancement, and recent works (Bai, Yin, and He 2024; Huang et al. 2025a; Lin, Huang, and Anantrasirichai 2025) further leverage Mamba-based architectures (Gu and Dao 2023) for improved performance. In parallel, probabilistic approaches have also been introduced into this domain, where Jiang et al. (2021); Islam, Xia, and Sattar (2020) employ GAN-based frameworks for both low-light and underwater image enhancement. Wang et al. (2022) applied normalizing flow-based methods to reduce residual noise in LLIE predictions. However, its invertibility constraint limits model complexity. Zhou et al. (2024) address the limitations of conventional normalizing flows by integrating a normal-light codebook with a latent normalizing flow. Diffusion models offer high fidelity and have been widely adopted for image enhancement tasks (Hou et al. 2024; Tang, Kawasaki, and Iwaguchi 2023), but they suffer from high inference latency due to their iterative denoising process.

2.1 Preliminary

In image enhancement, the output of a network can be interpreted as the conditional probability distribution of the target image, $\mathbf{y} \in \mathcal{Y}$, given the degraded input image $\mathbf{x} \in \mathcal{X}$, and the network's weights $\mathbf{w}$, i.e., $P(\mathbf{y} \mid \mathbf{x}, \mathbf{w})$. Assuming the prediction errors follow a Gaussian distribution, the conditional probability density function of the target $\mathbf{y}$ can be modeled as a multivariate Gaussian with mean given by the neural network output $F(\mathbf{x}; \mathbf{w})$, i.e., $P(\mathbf{y} \mid \mathbf{x}, \mathbf{w}) = \mathcal{N}(\mathbf{y} \mid F(\mathbf{x}; \mathbf{w}), \mathbf{\Sigma})$.

The network weights $\mathbf{w}$ can be learned through maximum likelihood estimation (MLE). Given a dataset of image pairs $\{\mathbf{x}^i, \mathbf{y}^i\}_{i=1}^N$, the MLE of $\mathbf{w}$, denoted as $\mathbf{w}^{\mathrm{MLE}}$, is computed by maximizing the log-likelihood of the observed data:

$$\mathbf{w}^{\mathrm{MLE}} = \underset{\mathbf{w}}{\operatorname{argmax}} \sum_{i=1}^{N} \log P(\mathbf{y}^i | \mathbf{x}^i, \mathbf{w}). \tag{1}$$

By optimizing such an objective function in Eq. (1), the network $F_{\mathbf{w}}$ fits a one-to-one mapping, $F_{\mathbf{w}} : \mathcal{X} \to \mathcal{Y}$, implying that $\mathbf{y}^i \neq \mathbf{y}^j$ requires $\mathbf{x}^i \neq \mathbf{x}^j$. However, this assumption leads to mode collapse and poor expressiveness when modeling inherently one-to-many enhancement problems.

3 Method

3.1 Variational Bayesian Inference for One-to-Many Modeling

To model the one-to-many mapping, we introduce uncertainty into the network weights $\mathbf{w}$ via Bayesian estimation, yielding a posterior distribution, $\mathbf{w} \sim P(\mathbf{w}|\mathbf{y}, \mathbf{x})$. During inference, weights are sampled from this distribution to generate diverse predictions. The posterior distribution over the weights is expressed as:

$$P(\mathbf{w}|\mathbf{y}, \mathbf{x}) = \frac{P(\mathbf{y}|\mathbf{x}, \mathbf{w})P(\mathbf{w})}{P(\mathbf{y}|\mathbf{x})}, \tag{2}$$

where $P(\mathbf{y} \mid \mathbf{x}, \mathbf{w})$ is the likelihood of observing $\mathbf{y}$ given the input $\mathbf{x}$ and weights $\mathbf{w}$, $P(\mathbf{w})$ denotes the prior distribution of the weights, and $P(\mathbf{y} \mid \mathbf{x})$ is the marginal likelihood.

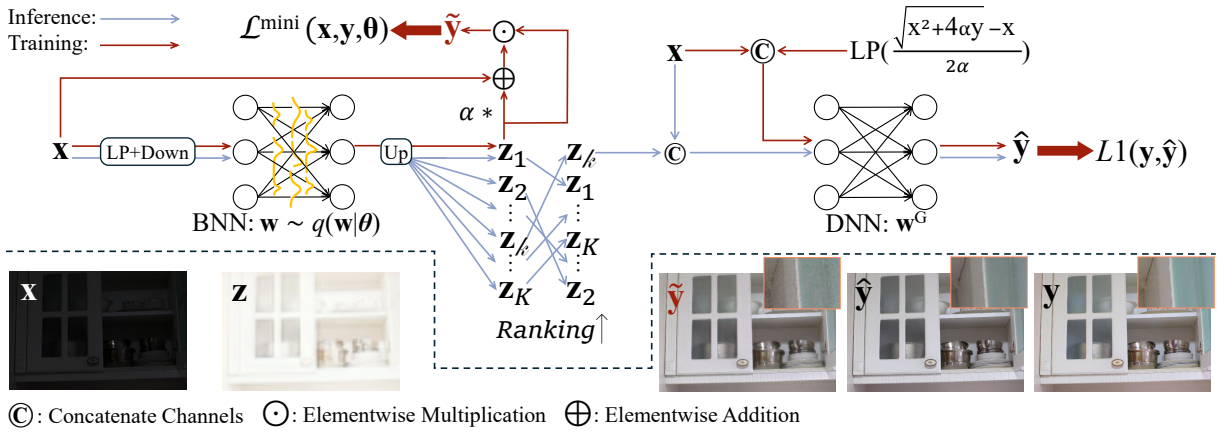

Figure 2: The two-stage pipeline. In Stage I, the BNN with weights $\mathbf{w} \sim q(\mathbf{w}|\boldsymbol{\theta})$ is trained by minimizing the minibatch loss $\mathcal{L}^{\text{mini}}$ in Eq. (6). In Stage II, the DNN with weights $\mathbf{w}^{\text{G}}$ is trained by minimizing the L1 loss, $L1(\mathbf{y}, \hat{\mathbf{y}})$. The inference process is denoted by $\rightarrow$, while the training process for each stage is indicated by $\rightarrow$.

Unfortunately, for any neural networks, the posterior in Eq. (2) cannot be calculated analytically. Instead, we can leverage variational inference (VI) to approximate $P(\mathbf{w}|\mathbf{y},\mathbf{x})$ with a more tractable distribution $q(\mathbf{w}|\boldsymbol{\theta})$. Such that, we can draw samples of weights $\mathbf{w}$ from the distribution $q(\mathbf{w}|\boldsymbol{\theta})$. As suggested by (Hinton and Van Camp 1993; Blundell et al. 2015), the variational approximation is fitted by minimizing their Kullback-Leibler (KL) divergence:

$$\begin{aligned}\boldsymbol{\theta}^\star &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathrm{KL}\left[q(\mathbf{w}|\boldsymbol{\theta}) \| P(\mathbf{w}|\mathbf{y},\mathbf{x})\right] \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \int q(\mathbf{w}|\boldsymbol{\theta}) \log \frac{q(\mathbf{w}|\boldsymbol{\theta})}{P(\mathbf{w})P(\mathbf{y}|\mathbf{x},\mathbf{w})}\, \mathrm{d}\mathbf{w} \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} -\mathbb{E}_{q(\mathbf{w}|\boldsymbol{\theta})}\left[\log P(\mathbf{y}|\mathbf{x},\mathbf{w})\right] + \mathrm{KL}\left[q(\mathbf{w}|\boldsymbol{\theta}) \| P(\mathbf{w})\right].\end{aligned} \tag{3}$$

We define the resulting cost function from Eq. (3) as:

$$\mathcal{L}(\mathbf{x},\mathbf{y},\boldsymbol{\theta}) = \underbrace{-\mathbb{E}_{q(\mathbf{w}|\boldsymbol{\theta})}\left[\log P(\mathbf{y}|\mathbf{x},\mathbf{w})\right]}_{\text{data-dependent term}} + \underbrace{\mathrm{KL}\left[q(\mathbf{w}|\boldsymbol{\theta}) \| P(\mathbf{w})\right]}_{\text{prior matching term}}, \tag{4}$$

where the data-dependent term can be interpreted as a reconstruction error, such as *L1* loss.

Let the variational posterior be a diagonal Gaussian $q(\mathbf{w}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}, \mathrm{diag}(\boldsymbol{\sigma}^2))$, parameterized by $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma})$, which defines a BNN. During each forward pass of the BNN, the network weights $\mathbf{w}$ are sampled using the reparameterization trick (Kingma 2014):

$$\mathbf{w} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \quad \text{where } \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}). \tag{5}$$

Adaptive Prior. To accelerate the convergence of Bayesian training, we adopt the idea of momentum update (He et al. 2020) to establish an adaptive prior, which has been shown to achieve faster convergence than fixed or empirical priors. Specifically, the prior $P(\mathbf{w})$ at step t is defined as:

$$P(\mathbf{w}) = \mathcal{N}(\boldsymbol{\mu}_t^{\text{EMA}}, \mathrm{diag}((\boldsymbol{\sigma}_t^{\text{EMA}})^2)),$$

where $\boldsymbol{\mu}_t^{\text{EMA}}$ and $\boldsymbol{\sigma}_t^{\text{EMA}}$ are updated via EMA from posterior parameters. This temporally adaptive prior smooths KL regularization across iterations, promoting consistent optimization. The resulting mini-batch loss follows:

$$\mathcal{L}^{\text{mini}} = \frac{1}{M}\sum_{i}^{M} \mathbb{E}_{\mathbf{w}\sim q}\|F(\mathbf{x}^i;\mathbf{w}) - \mathbf{y}^i\|_2^2 + \mathrm{KL}\left[q(\mathbf{w})\|P(\mathbf{w})\right], \tag{6}$$

After optimizing $\boldsymbol{\theta}^\star$ using Eq. (6), the BNN generates multiple distinct predictions $\{\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_K\}$ by sampling different weights $\mathbf{w}$ from $q(\mathbf{w}|\boldsymbol{\theta})$ during each forward pass.

3.2 BNN-DNN Framework

In a BNN, producing multiple high-resolution outputs can incur a high computational footprint. However, we are only interested in the highest-quality prediction among $\{\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_K\}$. To improve inference efficiency, we propose a two-stage architecture (see Fig. 2). The first stage uses a BNN to model one-to-many mappings in a low-dimensional latent space, capturing coarse structure and uncertainty. The second stage employs a DNN to reconstruct high-frequency details in the original image space. The multiple coarse outputs $\{\mathbf{z}_k\}_{k=1}^K$ from the first-stage BNN are used as proxies to identify the highest-quality prediction among $\{\hat{\mathbf{y}}\}_{k=1}^K$, eliminating the need to produce all $\{\hat{\mathbf{y}}\}_{k=1}^K$.

In Stage I, we employ low-pass filtering followed by downsampling to map the input's coarse information into a lower-dimensional space, $\mathrm{Down}(\mathrm{LP}(\mathbf{x}), r)$, where r denotes the scaling factor and LP represents a low-pass filter implemented via FFT. Subsequently, a BNN models the uncertainty in the low-dimensional coarse input information. The forward process of Stage I can be expressed as:

$$\mathbf{z} = \mathrm{Up}(F(\mathrm{Down}(\mathrm{LP}(\mathbf{x}), r); \mathbf{w})), \quad \mathbf{w} \sim q(\mathbf{w} \mid \boldsymbol{\theta}), \tag{7}$$

where $\mathrm{Up}(\cdot)$ is the bilinear upsampling operation for dimensionality matching. For a given input $\mathbf{x}$, multiple proxies $\{\mathbf{z}_k\}_{k=1}^K$ are generated via repeated forward passes in Eq. (7). The most reliable proxy $\mathbf{z}_k$ can then be automatically selected using a ranking mechanism (see Sec. 3.3). From Fig. 2, we can observe that $\mathbf{z}$ approximates the en-

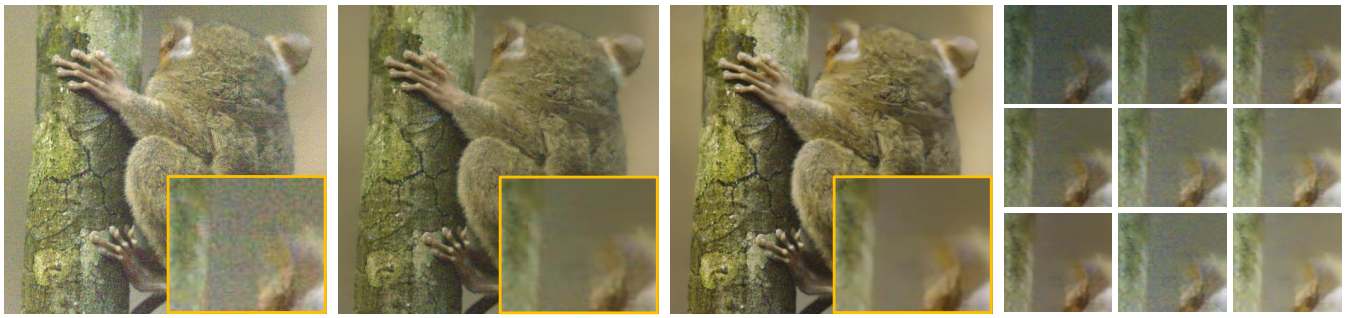

Figure 3: Visual comparisons of the DNN baseline, BEMMC, and BEMRank with CLIP-IQA. The rightmost patches highlight the diverse unselected predictions reflecting BEM's one-to-many modeling capability.

hanced illumination. The first-stage prediction $\tilde{\mathbf{y}}$ in the original resolution is computed as:

$$\tilde{\mathbf{y}} = (\mathbf{x} + \alpha\mathbf{z}) \odot \mathbf{z}, \tag{8}$$

where α is a small scalar and $\odot$ is element-wise multiplication. Compared to simpler formulations, such as $\mathbf{x} + \mathbf{z}$ or $\mathbf{x} \odot \mathbf{z}$, Eq. (8) reduces the risk of blurring fine textures or amplifying noise in $\mathbf{x}$. We note that $\tilde{\mathbf{y}}$ plays a key role in the ranking-based inference in Sec. 3.3.

In Stage II, we employ a DNN G to enhance the fine-grained details in the input. The forward process can be expressed as:

$$\hat{\mathbf{y}} = G([\mathbf{x}, \mathbf{z}]; \mathbf{w}^{\mathrm{G}}), \tag{9}$$

where $\mathbf{w}^{\mathrm{G}}$ represents the weights of the second-stage model, $[\cdot, \cdot]$ denotes the concatenation operation along the channel dimension. When training the second-stage DNN, we replace the predicted coarse information $\mathbf{z}$ with its ground truth, $\mathrm{LP}\left(\frac{\sqrt{\mathbf{x}^2+4\alpha\mathbf{y}}-\mathbf{x}}{2\alpha}\right)$, which is the explicit solution of Eq. (8) when $\tilde{\mathbf{y}}$ is replaced with $\mathbf{y}$ and $\mathbf{z}$ is treated as the unknown variable. This strategy is critical, as it prevents mode collapse—where diverse predictions from the first-stage BNN are undesirably regressed into a single output by the second-stage DNN.

The Backbone. For both the first- and second-stage models, we adopt the same backbone network but use different input and output layers. In the first stage, we construct a BNN by converting all layers in the backbone to their Bayesian counterparts via Eq. (5). The backbone follows an encoder-decoder UNet design. For the basic blocks, we consider both Transformers (Vaswani et al. 2017) and Mamba (Gu and Dao 2023), demonstrating the broad applicability of our methods across the two primary backbone architectures. We provide more details in the Supplementary Materials.

3.3 Inference under Uncertainty

As described in Algorithm 1, our method enables two inference modes: a ranking-based selection and Monte Carlo (MC) sampling. In the first stage, the BNN generates K latent candidates $\{\mathbf{z}_k\}_{k=1}^{K}$, which are computed in parallel. For ranking-based inference, an image quality assessment metric, $\mathsf{IQA}(\cdot)$, is applied to score the intermediate predictions $\{\tilde{\mathbf{y}}_k\}_{k=1}^{K}$ from Eq. (8). The latent code $\mathbf{z}^*$ corresponding to the highest-ranked $\tilde{\mathbf{y}}_k$ is selected and passed to the second stage for refinement, yielding the final output.

Algorithm 1: Inference

Require: Input $\mathbf{x}$, BNN F, DNN G
  for $k = 1$ to K **do**
    $\mathbf{w}_k \leftarrow \boldsymbol{\mu} + \boldsymbol{\sigma} \circ \boldsymbol{\epsilon}_k$, where $\boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
    $\mathbf{z}_k \leftarrow F(\mathrm{Down}(\mathrm{LP}(\mathbf{x}), r); \mathbf{w}_k)$ *Stage I*
  end for
  if Mode = *Monte Carlo* **then**
    $\mathbf{z}^* \leftarrow \frac{\mathbf{z}_1+\mathbf{z}_2+\cdots+\mathbf{z}_K}{K}$
  else
    $\mathbf{z}^* \leftarrow \underset{\mathbf{z}_k \in \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_K\}}{\mathrm{argmax}} \mathsf{IQA}((\mathbf{x} + \alpha\mathbf{z}_k) \circ \mathbf{z}_k)$
  end if
  $\hat{\mathbf{y}} \leftarrow G([\mathbf{x}, \mathrm{Up}(\mathbf{z}^*)]; \mathbf{w}^{\mathrm{G}})$ *Stage II*
Ensure: $\hat{\mathbf{y}}$

For MC inference, we aggregate the K latent samples $\{\mathbf{z}_1, \ldots, \mathbf{z}_K\}$ by computing their mean $\mathbf{z}^*$, which is subsequently passed to the second-stage DNN to generate the final prediction $\hat{\mathbf{y}}$. This process approximates the posterior expectation over the BNN's stochastic predictions. Since the BNN implicitly captures label noise through weight uncertainty, averaging multiple samples serves to marginalize out the randomness induced by noisy supervision, resulting in more stable and noise-suppressed outputs.

We denote the variant that uses ranking-based inference as BEM$_{\mathrm{Rank}}$ and the one that applies Monte Carlo sampling as BEM$_{\mathrm{MC}}$. The ranking mode can be instantiated with various no-reference IQA metrics, including CLIP-IQA, NIQE, UIQM, and UCIQE.

As illustrated in Fig. 4, CLIP-based ranking tends to favor brighter, high-contrast outputs that align with semantic and perceptual cues learned from natural images, while NIQE emphasizes statistical naturalness. Fig. 3 further shows that both BEM$_{\mathrm{Rank}}$ and BEM$_{\mathrm{MC}}$ generate visually enhanced results with minimal noise, whereas the deterministic DNN baseline exhibits notable residual artifacts. Due to its averaging nature, BEM$_{\mathrm{MC}}$ produces conservative outputs with

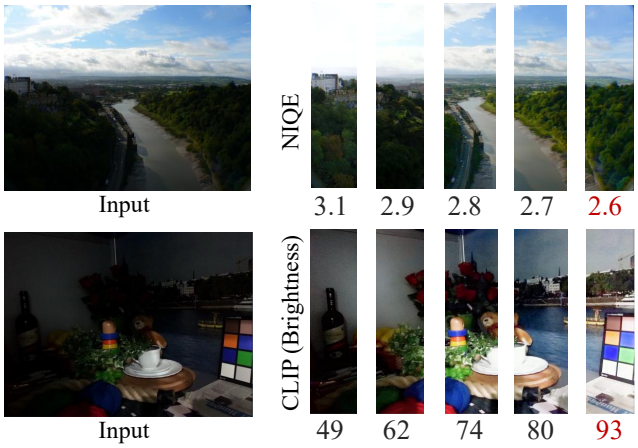

Figure 4: One-to-many mapping from input to outputs. The predictions are sorted by CLIP-IQA and NIQE.

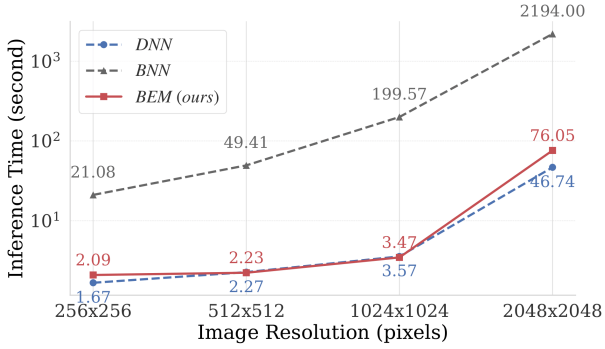

Figure 5: Inference speed on an Nvidia RTX 4090.

smoothed details, whereas BEM_{Rank} often yields higher-contrast results with enhanced perceptual sharpness. Although automatic ranking improves robustness and efficiency, users may also manually select their preferred enhancement from multiple candidates when speed is not a primary concern.

Inference Speed. Algorithm 1 avoids redundant sampling, resulting in a substantial reduction in inference latency. Fig. 5 compares the inference time of a conventional BNN, a standard DNN, and the proposed two-stage BEM. BEM achieves runtime comparable to the DNN and delivers a 22× speedup over the BNN when processing 512^2 images.

4 Experiments

Datasets. For LLIE, we evaluate our method on the paired LOL-v1 (Wei et al. 2018) and LOL-v2 (Yang et al. 2021) datasets, as well as the unpaired LIME (Guo, Li, and Ling 2016), NPE (Wang et al. 2013), MEF (Ma, Zeng, and Wang 2015), DICM (Lee, Lee, and Kim 2013), and VV (Vonikakis, Kouskouridas, and Gasteratos 2018) datasets. For UIE, we evaluate our method on the paired UIEB-R90 (Li et al. 2019) dataset, along with unpaired

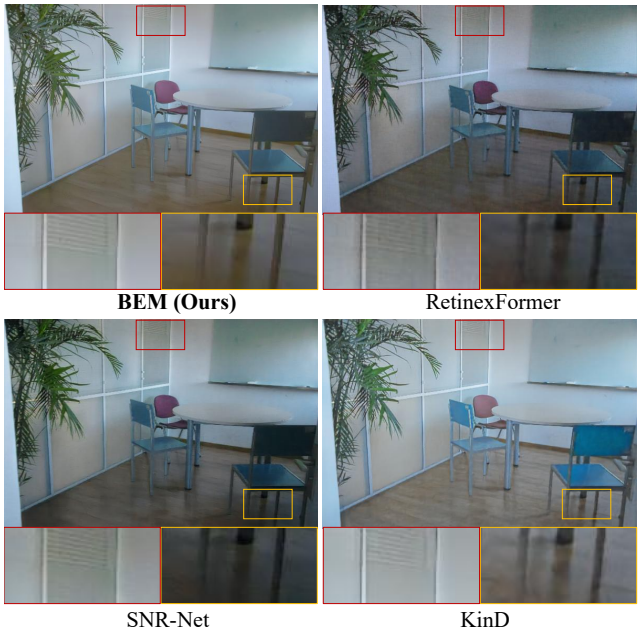

Figure 6: Visual comparisons on the LOL dataset.

datasets including C60 and U45 (Li, Li, and Wang 2019).

Experimental Settings. All models are trained using the Adam optimizer, with an initial learning rate of 2×10^{-4}, decayed to 10^{-6} following a cosine annealing schedule. The first- and second-stage models are trained for 300K and 150K iterations, respectively, on 128×128 inputs with a batch size of $M = 8$. Unless stated otherwise, the downscale factor r in Eq. (7) is set to $\frac{1}{16}$, K to 25, α in Eq. (8) to 0.025, and the adopted backbone architecture is Mamba.

4.1 Evaluation

For paired test sets with reliable ground-truth images, we evaluate enhancement quality using full-reference metrics, including PSNR, SSIM, and LPIPS. In unpaired or real-world scenarios where reference images are unavailable, we report no-reference scores using NIQE, UIQM, and UCIQE.

We compare against various types of probabilistic models and leading DNN-based methods, including normalizing flows (Wang et al. 2022; Zhou et al. 2024), GANs (Jiang et al. 2021; Cong et al. 2023; Islam, Xia, and Sattar 2020), diffusion models (Hou et al. 2024), variational autoencoders (VAEs) (Fu et al. 2022), as well as strong deterministic baselines (Zhang, Zhang, and Guo 2019; Zamir et al. 2022; Xu et al. 2022; Cai et al. 2023; Bai, Yin, and He 2024).

Full-reference. We conduct full-reference comparisons on LLIE (LOL-v1/v2) and UIE (UIEB-R90), as reported in Tab. 1 and Tab. 2 (middle). Our BEM, equipped with either a Transformer or Mamba backbone, achieves superior performance across all metrics and datasets, highlighting its robustness and generalization across diverse conditions. In contrast to prior methods that struggle to balance perceptual quality (e.g., LPIPS) and pixel-level fidelity (e.g., SSIM), BEM achieves higher SSIM and lower LPIPS simultane-

Method	Type	LOL-v1			LOL-v2-real			LOL-v2-syn		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Without GT Mean										
KinD	DNN	19.66	0.820	0.156	18.06	0.825	0.151	17.41	0.806	0.255
Restormer	DNN	22.43	0.823	0.147	18.60	0.789	0.232	21.41	0.830	0.144
SNR-Net	DNN	24.61	0.842	0.151	21.48	0.849	0.157	24.14	0.928	0.056
RetinexFormer	DNN	25.16	0.845	0.131	22.80	0.840	0.171	25.67	0.930	0.059
RetinexMamba	DNN	24.03	0.827	0.146	22.45	0.844	0.174	25.89	0.935	0.054
Mamba | BEM_{MC} (ours)	BNN	23.07	0.851	0.089	20.13	0.858	0.107	24.57	0.938	0.043
With GT Mean										
LLFlow	Flow	25.13	0.872	0.117	26.20	0.888	0.137	24.81	0.919	0.067
EnlightenGAN	GAN	17.48	0.652	0.322	18.64	0.677	0.309	16.57	0.734	-
GlobalDiff	DM	27.84	0.877	0.091	28.82	0.895	0.095	28.67	0.944	0.047
GLARE	Flow	27.35	**0.883**	0.083	28.98	0.905	0.097	29.84	0.958	-
Transformer | BEM_{Rank} (ours)	BNN	28.24	0.881	0.077	32.54	0.917	0.072	32.36	0.962	0.030
Transformer | BEM_{MC} (ours)	BNN	27.22	0.879	0.075	30.86	0.905	0.069	30.21	0.944	0.035
Mamba | BEM_{Rank} (ours)	BNN	28.80	0.884	0.069	32.66	0.915	0.060	32.95	0.964	0.026
Mamba | BEM_{MC} (ours)	BNN	**28.30**	0.881	0.072	**31.41**	**0.912**	**0.064**	30.58	0.958	0.033

Table 1: Full-reference evaluation on LOL-v1 and v2. The best results are in **bold**, while the second-best are underlined. Results in gray indicate the upper bound performance of BEM and are not directly comparable to the other results.

Method	Type	UIEB-R90		C60		U45	
		PSNR ↑	SSIM ↑	UIQM ↑	UCIQE ↑	UIQM ↑	UCIQE ↑
Ucolor	DNN	20.13	0.877	2.482	0.553	3.148	0.586
PUIE-MP	VAE	21.05	0.854	2.524	0.561	3.169	0.569
Restormer	DNN	23.82	0.903	2.688	0.572	3.097	0.600
FUnIEGAN	GAN	19.12	0.832	2.867	0.556	2.495	0.545
PUGAN	GAN	22.65	0.902	2.652	0.566	-	-
U-Shape	DNN	20.39	0.803	2.730	0.560	3.151	0.592
Semi-UIR	DNN	22.79	0.909	2.667	**0.574**	3.185	0.606
WFI2-Net	DNN	23.86	0.873	-	-	3.181	0.619
BEM_{MC}	BNN	**23.92**	**0.913**	2.694	0.537	3.188	0.591
BEM_{Rank}	BNN	25.62	0.940	**2.931**	0.567	**3.406**	**0.620**

Table 2: Full-reference evaluation (left) on R90, and no-reference evaluations (right) on C60 and U45.

Method	Type	DICM	LIME	MEF	NPE	VV
KinD	DNN	5.15	5.03	5.47	4.98	4.30
ZeroDCE	DNN	4.58	5.82	4.93	4.53	4.81
RUAS	DNN	5.21	4.26	3.83	5.53	4.29
LLFlow	Flow	4.06	4.59	4.70	4.67	4.04
PairLIE	DNN	4.03	4.58	4.06	4.18	3.57
RFR	DNN	3.75	3.81	3.92	4.13	-
GLARE	Flow	3.61	4.52	3.66	4.19	-
CIDNet	DNN	3.79	4.13	3.56	3.74	3.21
BEM_{MC}	BNN	3.77	3.94	3.22	3.85	2.95
BEM_{Rank}	BNN	**3.55**	**3.56**	**3.14**	**3.72**	**2.91**

Table 3: No-reference NIQE↓ evaluation, compared to previous methods (Yan et al. 2025; Guo et al. 2020; Liu et al. 2021; Fu et al. 2023b,a)

ously. For full-reference evaluation, we report BEM_{Rank} as an upper-bound performance of our method, since its first-stage ranking selects the candidate that is closest to the reference image under the chosen metric.

No-reference. We further evaluate LLIE performance on five unpaired datasets using no-reference metrics. As shown in Tab. 3, both BEM_{MC} and BEM_{Rank} achieve lower (i.e., better) NIQE scores than other methods. Fig. 4 shows that outputs with lower NIQE scores generally exhibit more natural illumination and avoid overexposure, indicating that BEM_{Rank} with NIQE-based ranking can reliably identify high-quality predictions in LLIE. Tab. 2 (right) further demonstrates that BEM achieves the best or competitive performance on the two unpaired UIE benchmarks.

Qualitative Results. As shown in Fig. 6, our BEM better preserves fine details and structural textures, as evidenced by the zoomed-in regions, compared to existing methods. Fig. 7 further compares our method with representative UIE approaches, including Tang, Kawasaki, and Iwaguchi (2023); Fu et al. (2022); Islam, Xia, and Sattar (2020); Cong et al. (2023); Fu et al. (2022); Li et al. (2019); Jiang et al. (2023); Li et al. (2021); Tang, Kawasaki, and Iwaguchi (2023); Huo, Li, and Zhu (2021); Zhang et al. (2022); Huang et al. (2023); Liu, Li, and Ding (2024); Li et al. (2023); Peng, Cao, and Cosman (2018); Shen et al. (2023); Han et al. (2021). Visual comparisons suggest that BEM restores color and illumination more faithfully, producing more natural-looking outputs, particularly in challenging scenes such as C60.

Statistical Analysis of Uncertainty Fig. 8 shows the score distributions of predictions from BEM. The spread in these distributions reflects the predictive variability inherent in one-to-many modeling. Notably, a portion of the samples receives low scores under multiple metrics, which indicates the presence of noisy or low-quality labels in the training set. This observation supports the inclusion of uncertainty

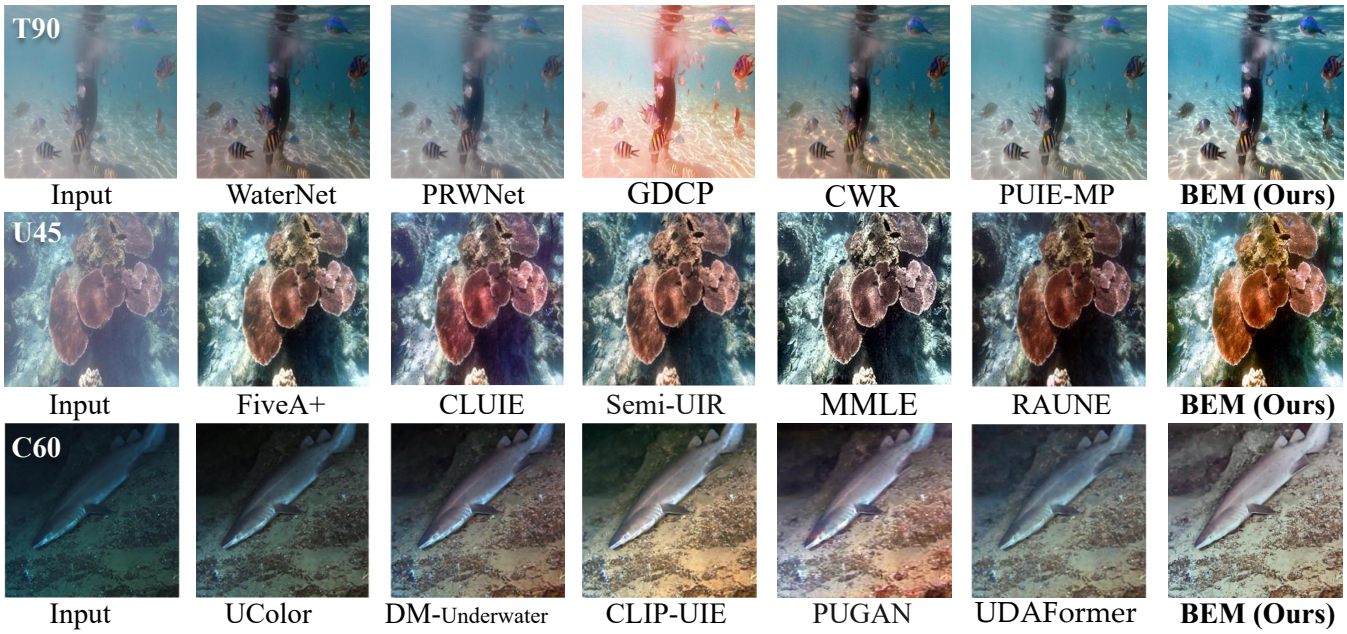

Figure 7: Visual comparisons on the R90, C60 and U45 datasets. Best viewed when zoomed in.

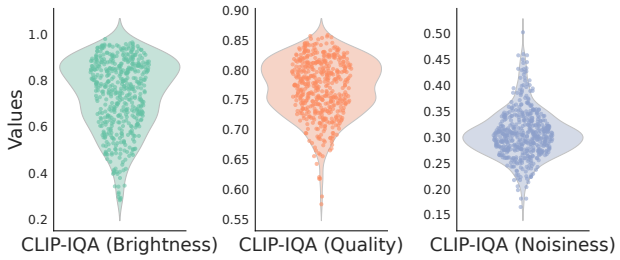

Figure 8: Score distributions of 500 predictions from BEM across three CLIP-IQA metrics.

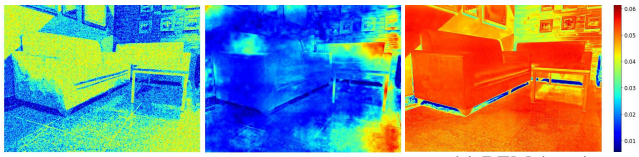

Figure 9: Visualization of pixel-wise output variability from 500 samples for: (a) VAE, (b) Diffusion, and (c) our BEM.

modeling to better accommodate imperfect supervision.

BEM vs. Diffusion Model and VAE. To assess the one-to-many modeling capacity of different probabilistic approaches, we visualize the pixel-wise output variability of the VAE, the diffusion model, and our BEM under identical inputs. As shown in Fig. 9, BEM produces substantially higher output variability while still preserving localized structural consistency, particularly along object boundaries where the predicted uncertainty remains low. In comparison, both the conditional VAE and the diffusion model

Figure 10: Visualization of BEM predictions. The input image is from LSRW (Hai et al. 2023).

display markedly lower variability, revealing a trade-off between generative diversity and structural fidelity. These observations indicate that BEM can maintain sample diversity without sacrificing geometric alignment, which is essential for modeling one-to-many mappings in ill-posed enhancement tasks. Fig. 10 further shows that the diverse enhanced outputs of BEM exhibit consistent structural appearance.

5 Conclusion

We presented the Bayesian Enhancement model (BEM) to address the one-to-many challenge in image enhancement, which is identified as a key limitation in previous data-driven models. An *Adaptive Prior* is introduced to support stable and efficient Bayesian training, while the BNN–DNN design enables fast inference. Experiments across multiple benchmarks show clear improvements over existing methods, highlighting the benefits of Bayesian modeling for ambiguous enhancement tasks.

Acknowledgments

This work was supported by the UKRI MyWorld Strength in Places Program (SIPF00006/1) and the EPSRC ECR International Collaboration Grants (EP/Y002490/1). We acknowledge Humble Bee Films for providing visual content.

References

Bai, J.; Yin, Y.; and He, Q. 2024. Retinexmamba: Retinex-based Mamba for Low-light Image Enhancement. *arXiv preprint arXiv:2405.03349*.

Blundell, C.; Cornebise, J.; Kavukcuoglu, K.; and Wierstra, D. 2015. Weight uncertainty in neural network. In *International conference on machine learning*, 1613–1622. PMLR.

Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; and Zhang, Y. 2023. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12504–12513.

Cong, R.; Yang, W.; Zhang, W.; Li, C.; Guo, C.-L.; Huang, Q.; and Kwong, S. 2023. Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators. *IEEE Transactions on Image Processing*, 32: 4472–4485.

Fu, H.; Zheng, W.; Meng, X.; Wang, X.; Wang, C.; and Ma, H. 2023a. You do not need additional priors or regularizers in retinex-based low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18125–18134.

Fu, Z.; Wang, W.; Huang, Y.; Ding, X.; and Ma, K.-K. 2022. Uncertainty inspired underwater image enhancement. In *European conference on computer vision*, 465–482. Springer.

Fu, Z.; Yang, Y.; Tu, X.; Huang, Y.; Ding, X.; and Ma, K.-K. 2023b. Learning a simple low-light image enhancer from paired low-light instances. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 22252–22261.

Gal, Y.; and Ghahramani, Z. 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, 1050–1059. PMLR.

Gu, A.; and Dao, T. 2023. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.

Guo, C.; Li, C.; Guo, J.; Loy, C. C.; Hou, J.; Kwong, S.; and Cong, R. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1780–1789.

Guo, X.; Li, Y.; and Ling, H. 2016. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2): 982–993.

Hai, J.; Xuan, Z.; Yang, R.; Hao, Y.; Zou, F.; Lin, F.; and Han, S. 2023. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation*, 90: 103712.

Han, J.; Shoeiby, M.; Malthus, T.; Botha, E.; Anstee, J.; Anwar, S.; Wei, R.; Petersson, L.; and Armin, M. A. 2021. Single underwater image restoration by contrastive learning. In *2021 IEEE international geoscience and remote sensing symposium IGARSS*, 2385–2388. IEEE.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9729–9738.

Hinton, G. E.; and Van Camp, D. 1993. Keeping the neural networks simple by minimizing the description length of the weights. In *Proceedings of the sixth annual conference on Computational learning theory*, 5–13.

Hou, J.; Zhu, Z.; Hou, J.; Liu, H.; Zeng, H.; and Yuan, H. 2024. Global structure-aware diffusion process for low-light image enhancement. *Advances in Neural Information Processing Systems*, 36.

Huang, G.; Lin, R.; Li, Y.; Bull, D.; and Anantrasirichai, N. 2025a. BVI-Mamba: video enhancement using a visual state-space model for low-light and underwater environments. In *Machine Learning from Challenging Data 2025*, volume 13460, 74–81. SPIE.

Huang, G.; Wang, H.; Seymour, B.; Kovacs, E.; Ellerbrock, J.; Blackham, D.; and Anantrasirichai, N. 2025b. Visual enhancement and 3D representation for underwater scenes: a review. *arXiv e-prints*, arXiv–2505.

Huang, S.; Wang, K.; Liu, H.; Chen, J.; and Li, Y. 2023. Contrastive semi-supervised learning for underwater image restoration via reliable bank. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 18145–18155.

Huo, F.; Li, B.; and Zhu, X. 2021. Efficient wavelet boost learning-based multi-stage progressive refinement network for underwater image enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1944–1952.

Islam, M. J.; Xia, Y.; and Sattar, J. 2020. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2): 3227–3234.

Jiang, J.; Ye, T.; Bai, J.; Chen, S.; Chai, W.; Jun, S.; Liu, Y.; and Chen, E. 2023. Five A^+ Network: You Only Need 9K Parameters for Underwater Image Enhancement. *British Machine Vision Conference (BMVC)*.

Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; and Wang, Z. 2021. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30: 2340–2349.

Kendall, A.; and Cipolla, R. 2016. Modelling uncertainty in deep learning for camera relocalization. In *2016 IEEE international conference on Robotics and Automation (ICRA)*, 4762–4769. IEEE.

Kendall, A.; Gal, Y.; and Cipolla, R. 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7482–7491.

Kingma, D. P. 2014. Auto-encoding variational bayes. *International Conference on Learning Representations (ICLR)*.

Krishnan, R.; Subedar, M.; and Tickoo, O. 2020. Specifying weight priors in bayesian deep neural networks with empirical bayes. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 4477–4484.

Lee, C.; Lee, C.; and Kim, C.-S. 2013. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE transactions on image processing*, 22(12): 5372–5384.

Li, C.; Anwar, S.; Hou, J.; Cong, R.; Guo, C.; and Ren, W. 2021. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing*, 30: 4985–5000.

Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; and Tao, D. 2019. An underwater image enhancement benchmark dataset and beyond. *IEEE transactions on image processing*, 29: 4376–4389.

Li, H.; Li, J.; and Wang, W. 2019. A fusion adversarial underwater image enhancement network with a public test dataset. *arXiv preprint arXiv:1906.06819*.

Li, K.; Wu, L.; Qi, Q.; Liu, W.; Gao, X.; Zhou, L.; and Song, D. 2023. Beyond single reference for training: Underwater image enhancement via comparative learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(6): 2561–2576.

Lin, R.; Huang, G.; and Anantrasirichai, N. 2025. Dynamic Weight-based Temporal Aggregation for Low-light Video Enhancement. *arXiv preprint arXiv:2510.09450*.

Liu, R.; Ma, L.; Zhang, J.; Fan, X.; and Luo, Z. 2021. 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*, 10561–10570.

Liu, S.; Li, K.; and Ding, Y. 2024. Underwater Image Enhancement by Diffusion Model with Customized CLIP-Classifier. *arXiv preprint arXiv:2405.16214*.

Ma, K.; Zeng, K.; and Wang, Z. 2015. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11): 3345–3356.

Malyugina, A.; Huang, G.; Ruiz, E.; Leslie, B.; and Anantrasirichai, N. 2025. Marine Snow Removal Using Internally Generated Pseudo Ground Truth. *arXiv preprint arXiv:2504.19289*.

Neal, R. M. 2012. *Bayesian learning for neural networks*, volume 118. Springer Science & Business Media.

Peng, L.; Zhu, C.; and Bian, L. 2023. U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing*.

Peng, Y.-T.; Cao, K.; and Cosman, P. C. 2018. Generalization of the dark channel prior for single image restoration. *IEEE Transactions on Image Processing*, 27(6): 2856–2868.

Shen, Z.; Xu, H.; Luo, T.; Song, Y.; and He, Z. 2023. UDAformer: Underwater image enhancement based on dual attention transformer. *Computers & Graphics*, 111: 77–88.

Tang, Y.; Kawasaki, H.; and Iwaguchi, T. 2023. Underwater image enhancement by transformer-based diffusion model with non-uniform sampling for skip strategy. In *Proceedings of the 31st ACM International Conference on Multimedia*, 5419–5427.

Tomczak, M.; Swaroop, S.; Foong, A.; and Turner, R. 2021. Collapsed variational bounds for Bayesian neural networks. *Advances in Neural Information Processing Systems*, 34: 25412–25426.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Vonikakis, V.; Kouskouridas, R.; and Gasteratos, A. 2018. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, 77: 9211–9231.

Wang, S.; Zheng, J.; Hu, H.-M.; and Li, B. 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing*, 22(9): 3538–3548.

Wang, Y.; Wan, R.; Yang, W.; Li, H.; Chau, L.-P.; and Kot, A. 2022. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 2604–2612.

Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep retinex decomposition for low-light enhancement. *British Machine Vision Conference (BMVC)*.

Xu, X.; Wang, R.; Fu, C.-W.; and Jia, J. 2022. SNR-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 17714–17724.

Yan, Q.; Feng, Y.; Zhang, C.; Pang, G.; Shi, K.; Wu, P.; Dong, W.; Sun, J.; and Zhang, Y. 2025. Hvi: A new color space for low-light image enhancement. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 5678–5687.

Yang, W.; Wang, W.; Huang, H.; Wang, S.; and Liu, J. 2021. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30: 2072–2086.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5728–5739.

Zhang, W.; Zhuang, P.; Sun, H.-H.; Li, G.; Kwong, S.; and Li, C. 2022. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Transactions on Image Processing*, 31: 3997–4010.

Zhang, Y.; Zhang, J.; and Guo, X. 2019. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, 1632–1640.

Zhou, H.; Dong, W.; Liu, X.; Liu, S.; Min, X.; Zhai, G.; and Chen, J. 2024. GLARE: Low Light Image Enhancement via Generative Latent Feature based Codebook Retrieval. *Proceedings of the European conference on computer vision (ECCV)*.