Breaking Degradation Coupling: A Structural Entropy–Guided Decoupled Framework and Benchmark for Infrared Enhancement

Supplementary Material

1. Preliminaries

In this section, we summarize the concepts and definitions related to the background of our work, including thermal infrared (TIR) image enhancement, Evidential deep learning, and Structural entropy.

1.1. TIR Enhancement

TIR degradations primarily manifest as low contrast, blur, and noise. These effects arise from (i) insufficient target–background temperature differences and compression introduced by radiometric calibration; (ii) point-spread function (PSF) broadening due to defocus, diffraction, turbulence, or motion; and (iii) fixed-pattern noise (FPN) together with readout/photon noise. Let I^c denote the clean reference image. The observed degraded image I^d can be modeled as

$$I^d = (\mathcal{N}_s \circ \mathcal{N}_o) \circ (\mathcal{K} * \mathrm{C}(I^c)) + \mathcal{N}_r, \tag{1}$$

where $\circ$ denotes composition operator and $*$ denotes convolution. $\mathrm{C}(\cdot)$ is a contrast-compression operator and $\mathcal{K}$ is the blur kernel/PSF. $\mathcal{N}_o$ and $\mathcal{N}_s$ represent optics-related and stripe-related FPN, respectively, and $\mathcal{N}_r$ denotes readout/photon noise. Note that a single image may contain one or multiple unknown degradations, added in a random order. Given a degraded TIR image I^d, TIR enhancement seeks to learn a mapping $f_\theta : I^d \mapsto I^r$ such that $I^r \approx I^c$.

1.2. Evidential Deep Learning

Evidential Deep Learning (EDL) [12] interprets network outputs as Dirichlet evidence parameters in subjective logic, thereby modeling prediction and uncertainty jointly while discouraging unwarranted overconfidence. For continuous targets on the interval $[0, 1]$, we can employ the binary special case of the Dirichlet—the Beta distribution. A Beta-evidence head outputs nonnegative evidence (α, β) for the presence and absence hypotheses, respectively. The mean $p = \frac{\alpha}{\alpha+\beta}$ serves as the predicted value while the total evidence $S = \alpha + \beta$ quantifies confidence. Given a target label $y \in [0, 1]$, the negative log-likelihood is minimized under the Beta model together with a KL regularizer that pulls uncertain posteriors toward the uniform prior:

$$\begin{aligned}\mathcal{L}_{EDL} = &\underbrace{-\Big[(\alpha-1)\ln y + (\beta-1)\ln(1-y) - \ln B(\alpha,\beta)\Big]}_{L_{NLL}} \\ &+ \tau \cdot \underbrace{\mathrm{KL}\big(\mathrm{Beta}(\alpha,\beta)\,\|\,\mathrm{Beta}(1,1)\big)}_{L_{KL}},\end{aligned} \tag{2}$$

where $B(\alpha, \beta)$ is the Beta function and $\mathrm{KL}(\cdot\|\cdot)$ denotes the Kullback–Leibler divergence. $\mathrm{Beta}(\cdot, \cdot)$ is a Beta distribution, $\mathrm{Beta}(1, 1)$ is the uniform prior, and τ is an annealing coefficient that increases from 0 to 1 over training.

1.3. Structural Entropy

Structural entropy (SE) [9] quantifies the uncertainty and information content of a graph. Lower SE indicates a more ordered structure. SE can be defined on coding trees of different heights to measure structural information at different orders. Given a graph $G(\mathcal{V}, \mathcal{E})$, a coding tree $\mathcal{T}$ is a hierarchical partition of the vertex set $\mathcal{V}$ that satisfies [9]:

1. Each node x in $\mathcal{T}$ is associated with a set $T_x \subseteq \mathcal{V}$. For the root node λ of $\mathcal{T}$, $T_\lambda = \mathcal{V}$. Any leaf node q in $\mathcal{T}$ is associated with a single node in G, i.e., $\mathcal{T}_q = \{v\}, v \in \mathcal{V}$
2. For each node a in $\mathcal{T}$, denote all its children as $b_1, \ldots, b_k$, then $T_{b_1}, \ldots, T_{b_k}$ is a partition of T_a.
3. For each node a in $\mathcal{T}$, denote its height as $h(a)$. Let $h(\lambda) = 0$ and $h(\bar{a}) = h(a) + 1$, where $\bar{a}$ is the parent of a. The height of $\mathcal{T}$, $h(\mathcal{T}) = \max_{a\in\mathcal{T}} h(a)$.

The SE of graph G on coding tree $\mathcal{T}$ is defined as:

$$\mathcal{H}^{\mathcal{T}}(G) = -\sum_{a\in\mathcal{T}, a\neq\lambda} \frac{g_a}{vol(\lambda)} \log \frac{vol(a)}{vol(\bar{a})}, \tag{3}$$

where g_a is the summation of the degrees of the cut edges of $\mathcal{T}_a$ (i.e., the weight sum of edges with exactly one endpoint in $\mathcal{T}_a$). $vol(a)$, $vol(\bar{a})$, and $vol(\lambda)$ represent the volumes, i.e., the sums of node degrees within $\mathcal{T}_a$, $\mathcal{T}_{\bar{a}}$ and $\mathcal{T}_\lambda$, respectively. The d-dimensional SE of G, defined as $\mathcal{H}^{(d)}(G) = \min_{\forall\mathcal{T}:h(\mathcal{T})=d}\{\mathcal{H}^{\mathcal{T}}(G)\}$, is realized by acquiring an optimal coding tree of height d, in which the disturbance derived from stochastic variation is minimized.

The two-dimensional structural entropy (2D-SE) measures second-order structural information, and minimizing the 2D-SE of a graph reveals the true second-order structure in disordered graphs. Let $\mathcal{P} = \mathcal{C}1, \mathcal{C}2, \ldots, \mathcal{C}k$ be a partition of G, the corresponding 2D-SE is defined as

$$\mathcal{H}^2(\mathcal{P}) = -\sum_{\mathcal{C}\in\mathcal{P}} \left(\frac{g_\mathcal{C}}{v_G} \log \frac{v_\mathcal{C}}{v_G} + \sum_{x\in\mathcal{C}} \frac{o_x}{v_G} \log \frac{o_x}{v_\mathcal{C}} \right), \tag{4}$$

where o_x is the degree of vertex x (the sum of weights of edges incident to x) and v_G is the graph volume (the sum of all vertex degrees). $g_\mathcal{C}$ and $v_\mathcal{C}$ denote the cut weight and the volume of part $\mathcal{C}$, respectively. Inspired by literature [15], the per-node 2D-SE contribution can be derived in detail as

follows:

$$
\begin{aligned}
&\mathcal{P} = \{\mathcal{C}_1, \ldots, \mathcal{C}_j, \ldots, \mathcal{C}_k\},\\
&\mathcal{P}' = \{\mathcal{C}_1, \ldots, \mathcal{C}_j \backslash \{x\}, \ldots, \mathcal{C}_k, \{x\}\},\\
&\Delta H_x = \mathcal{H}^2(\mathcal{P}') - \mathcal{H}^2(\mathcal{P})\\
&= -\sum_{\mathcal{C}' \in \mathcal{P}'} \left(\frac{g_{\mathcal{C}'}}{v_G} \log \frac{v_{\mathcal{C}'}}{v_G} + \sum_{x' \in \mathcal{C}'} \frac{o_{x'}}{v_G} \log \frac{o_{x'}}{v_{\mathcal{C}'}} \right)\\
&+ \sum_{\mathcal{C} \in \mathcal{P}} \left(\frac{g_{\mathcal{C}}}{v_G} \log \frac{v_{\mathcal{C}}}{v_G} + \sum_{x \in \mathcal{C}} \frac{o_x}{v_G} \log \frac{o_x}{v_{\mathcal{C}}} \right)\\
&= \underbrace{-\frac{g_{\mathcal{C}'_j}}{v_G} \log \frac{v_{\mathcal{C}'_j}}{v_G} - \sum_{m \in \mathcal{C}'_j} \frac{o_m}{v_G} \log \frac{o_m}{v_{\mathcal{C}'_j}}}_{\mathcal{H}^2(\mathcal{C}'_j)} \underbrace{-\frac{o_x}{v_G} \log \frac{o_x}{v_G}}_{\mathcal{H}^2(\{x\})}\\
&\underbrace{+\frac{g_{\mathcal{C}_j}}{v_G} \log \frac{v_{\mathcal{C}_j}}{v_G} + \sum_{m \in \mathcal{C}_j \backslash \{x\}} \frac{o_m}{v_G} \log \frac{o_m}{v_{\mathcal{C}_j}}}_{-\mathcal{H}^2(\mathcal{C}_j \backslash \{x\})} \underbrace{+\frac{o_x}{v_G} \log \frac{o_x}{v_{\mathcal{C}_j}}}_{-\mathcal{H}^2(x)}\\
&= -\frac{g_{\mathcal{C}'_j}}{v_G} \log \frac{v_{\mathcal{C}'_j}}{v_G} + \frac{g_{\mathcal{C}_j}}{v_G} \log \frac{v_{\mathcal{C}_j}}{v_G} - \frac{o_x}{v_G} (\log \frac{o_x}{v_G} - \log \frac{o_x}{v_{\mathcal{C}_j}})\\
&- \sum_{m \in \mathcal{C}'_k} \frac{o_m}{v_G} (\log \frac{o_m}{v_{\mathcal{C}'_j}} - \log \frac{o_m}{v_{\mathcal{C}_j}})\\
&= -\frac{g_{\mathcal{C}'_j}}{v_G} \log \frac{v_{\mathcal{C}'_j}}{v_G} + \frac{g_{\mathcal{C}_j}}{v_G} \log \frac{v_{\mathcal{C}_j}}{v_G} - \frac{o_x}{v_G} \log \frac{v_{\mathcal{C}_j}}{v_G}\\
&- \log \frac{v_{\mathcal{C}_j}}{v_{\mathcal{C}'_j}} \sum_{m \in \mathcal{C}'_j} \frac{o_m}{v_G}\\
&= -\frac{g_{\mathcal{C}'_j}}{v_G} \log \frac{v_{\mathcal{C}'_j}}{v_G} + \frac{g_{\mathcal{C}_j}}{v_G} \log \frac{v_{\mathcal{C}_j}}{v_G} - \frac{o_x}{v_G} \log \frac{v_{\mathcal{C}_j}}{v_G} - \frac{v_{\mathcal{C}'_j}}{v_G} \log \frac{v_{\mathcal{C}_j}}{v_{\mathcal{C}'_j}}.
\end{aligned}
\tag{5}
$$

2. Additional Details of Night-TIR Benchmark

Considering that nighttime scenes exhibit smaller target–background temperature differences and weaker radiative signals, thermal infrared (TIR) imagery therefore tends to have reduced contrast. To this end, we introduce a challenging nighttime TIR enhancement benchmark, Night-TIR. As shown in Figure 1, Night-TIR covers a broad spectrum of nocturnal scenes—including pedestrians, vehicles, roads, woodlands, residential buildings, high-rises, plazas, lakes, and mountainous areas. All images are captured at night using the integrated thermal camera on a DJI Matrice 4T UAV (spatial resolution 640×512, spectral band 8–14 μm and super-resolution mode supported). Flights are conducted at altitudes of 20–80 m; imaging is performed during hover and stabilized by a three-axis mechanical gimbal. To ensure data reliability, we perform rigorous manual quality control and remove low-quality samples. In total, Night-TIR comprises 647 high-quality nighttime TIR images, providing a realistic, diverse, and challenging benchmark for infrared image enhancement and compound-degradation restoration.

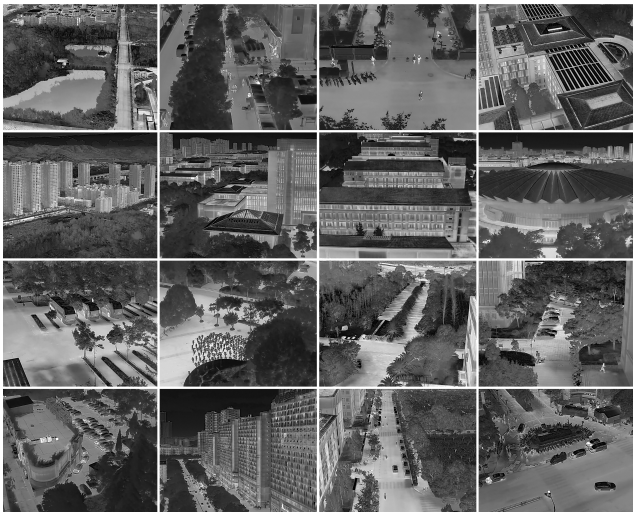

Figure 1. TIR images within Night-TIR benchmark.

3. Network Architectural Details

The proposed SEGD framework comprises an encoder, a decoder, DENet, and a set of DRMs. The encoder contains two convolutional layers followed by eight ResBlocks (each residual block comprises two convolutional layers, each followed by GroupNorm and GELU). The feature width is fixed at 64 channels. The decoder consists of four ResBlocks and two convolutional layers, producing a single-channel restored image at the output. DENet is composed of a feature extractor, a Degradation-Aware Head (DAH), and a Degradation Evidence Head (DEH). The feature extractor mirrors the encoder architecture and appends a global average pooling layer, yielding a vector of size $(B, 64)$ (where B denotes the batch size). Both DAH and DEH are three-layer, fully connected multilayer perceptrons with output dimensions $(B, 3)$ and $(B, 2)$, respectively. Each DRM adopts a RestormerBlock [18] with 32 channels. For DENet training, we adopt the same optimization configuration as the main backbone: the Adam optimizer with a batch size of 48; random cropping and flipping to produce 256×256 patches; an initial learning rate of 5×10^{-5} with a cosine-annealing schedule; and 100 training epochs. After training, DENet's parameters are frozen when training the backbone network.

4. More Results on HM-TIR and Night-TIR

To thoroughly evaluate the proposed SEGD, this section presents three additional studies: (i) qualitative comparisons with competing methods under single- and double-degradation settings; (ii) comparisons between *SEGD* and state-of-the-art single-degradation methods in the single-degradation regime; and (iii) evaluations against additional all-in-one visible image enhancement methods.

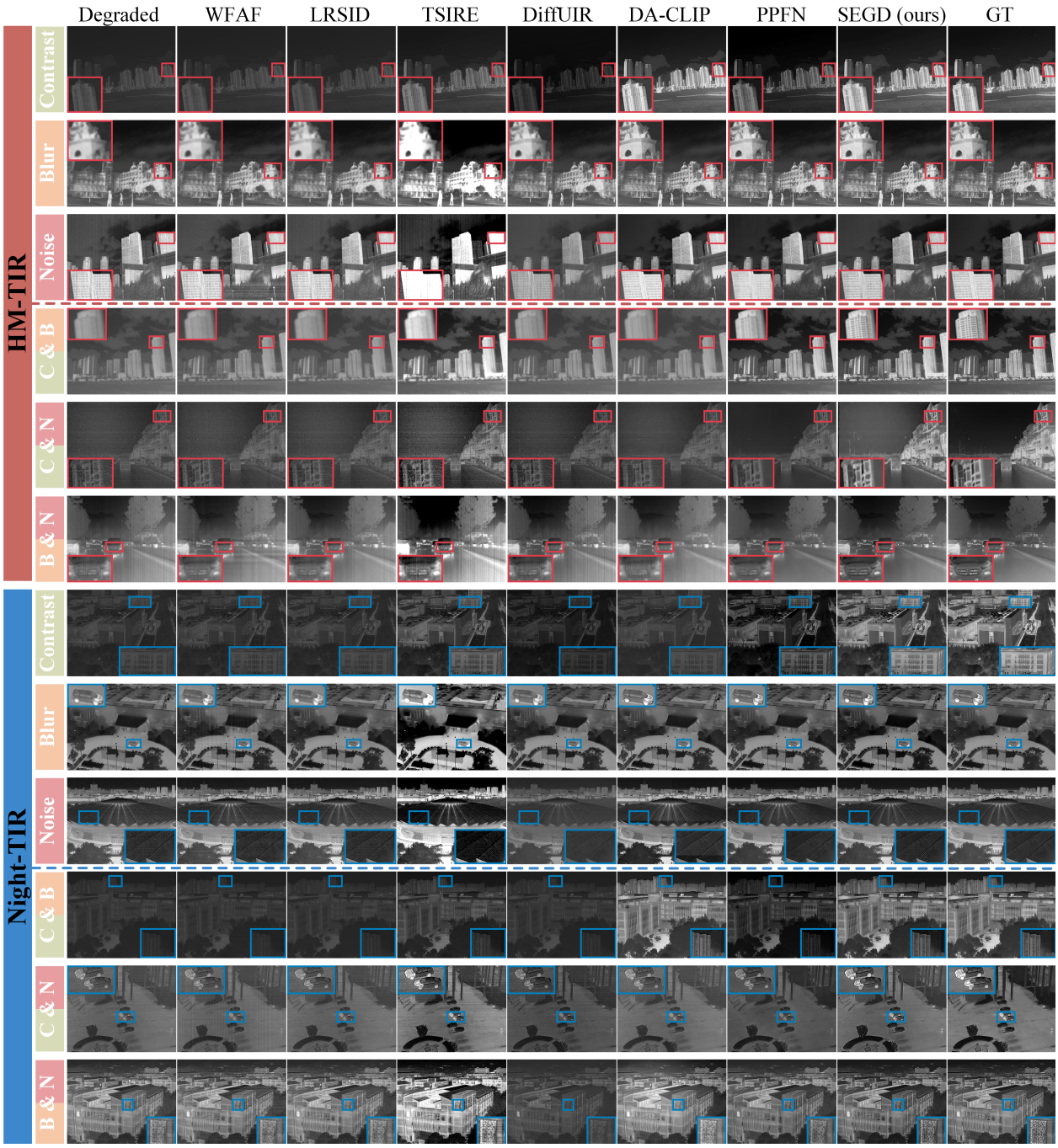

Figure 2. Additional qualitative comparison of single and double compound degradations on the HM-TIR and Night-TIR datasets.

Additional qualitative comparisons. Figure 2 presents qualitative comparisons between the proposed SEGD and competing methods under single- and double-degradation settings. PPFN performs comparably to SEGD on deblurring and denoising but lags in contrast enhancement. Under double-degradation scenarios, SEGD yields sharper edges and more faithful repetitive textures across both datasets, effectively suppresses background noise, recovers high-frequency details, and delivers more natural contrast, resulting in clearly superior overall visual quality relative to the other methods.

Results on single degradation TIR. To assess the performance of the proposed SEGD under single-degradation settings, we construct three category-specific test subsets—noise, blur, and low contrast—on both HM-TIR and Night-TIR datasets. For contrast enhancement, we compare against LLFormer [14], ReDDiT [7], SCI++ [11], and DarkIR [3]; for deblurring, we compare with MIMO-

Figure 3. Qualitative and quantitative comparisons under single-degradation settings on HM-TIR and Night-TIR.

UNet [2], Stripformer [13], FFTformer [5], and EVSSIM [6]; and for denoising, we consider CycleISP [17], AP-BSN [8], DCD-Net [19], and ScoreDVI [1]. As illustrated in Fig. 3, these methods alleviate the degradations to some extent but still suffer from issues such as insufficient contrast or local overexposure, blurred fine details, and residual noise. In contrast, SEGD consistently delivers superior results across all three tasks—achieving the highest PSNR and SSIM—while producing sharper edges, more faithful textures, and a more natural overall appearance.

Additional evaluation of all-in-One visible-image enhancement methods. As shown in Table 1, we additionally evaluate two recent all-in-one visible enhancement methods—DCPT [4] and MoCE-IR [16]. Both methods transfer poorly to TIR images, primarily due to the pronounced modality gap in imaging physics and degradation distributions between the visible and TIR domains. These findings underscore the limited effectiveness of directly porting visible methods to TIR restoration and highlight the risk of negative transfer in such cross-domain deployments.

5. Complexity Comparison

For a fair comparison, we compute and report the number of learnable parameters, inference time, and floating-point operations (FLOPs) for SEGD and all competing methods on single-degradation HM-TIR inputs at a resolution of 640×512, excluding the non–deep-learning baselines WFAF and LRSID. As summarized in Table 2, SEGD has the fewest parameters among all methods and achieves the second-fastest inference. Owing to backbone convolutions on full-resolution feature maps, SEGD does not achieve the lowest FLOPs; nevertheless, it delivers markedly superior restoration quality under complex degradations. Overall, SEGD offers a superior trade-off among inference efficiency, model size, and restoration quality.

6. Additional Ablation Studies

All experiments in this section are conducted on the HM-TIR dataset. We first examine training strategies for DENet: the DAH and DEH are trained either independently—each

Table 1. Quantitative comparison with additional visible all-in-one methods on the HM-TIR and Night-TIR datasets. The best and second-best performances for each metric are highlighted with Red and Blue backgrounds, respectively.

Dataset	HM-TIR								Night-TIR							
Degradation	Single		Double		Triple		Avg.		Single		Double		Triple		Avg.	
Metric	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCPT [4]	17.878	0.627	16.672	0.529	16.714	0.326	17.088	0.494	19.236	0.707	16.474	0.499	16.315	0.323	17.341	0.510
MoCE-IR [16]	23.257	0.737	18.061	0.537	16.561	0.324	19.292	0.533	22.013	0.746	17.573	0.503	16.244	0.320	18.610	0.522
SEGD (ours)	31.601	0.951	26.342	0.872	20.310	0.761	26.085	0.862	26.789	0.922	22.323	0.842	18.312	0.737	22.475	0.834

Table 2. Complexity comparison on parameters, FLOPs, and time. The best and second-best results for each metric are highlighted with Red and Blue backgrounds, respectively.

Methods	TSIRE	DiffUIR	DACLIP	DCPT	MoCE-IR	PPFN	SEGD
Params (M)	2.52	12.41	233.14	26.10	23.49	26.60	2.268
FLOPs (G)	77.91	494.04	660.18	704.95	446.41	704.33	589.48
Time (S)	0.01	0.43	15.32	0.27	0.32	0.65	0.19

with its own feature extractor—or jointly with a shared extractor. As shown in Table 3, independent training yields a slight increase in degradation-type classification accuracy and a small reduction in the mean absolute error (MAE) of degradation-intensity estimation; however, the overall restoration quality remains essentially unchanged (PSNR increases by only 0.046 dB and SSIM is nearly identical). Meanwhile, independent training introduces additional parameters and longer training time.

We further assess how each DRM behaves under mismatched degradations. Concretely, we deliberately permute the mapping between degradation types and DRMs. For example, assigning the contrast DRM to handle blur or noise, and analogously for the other DRMs. We evaluate three configurations: fully mismatched (all degradations processed by incorrect DRMs), partially matched (at least one correct DRM–degradation pair), and fully matched (one-to-one correspondence). As summarized in Table 4, the fully mismatched setting performs the worst; introducing any correct pairing produces a performance rebound; and the fully matched case achieves the best results. These observations indicate that each DRM's benefit is primarily specific to its designated degradation, and the gains diminish—or even vanish—when transferred across degradation types. Notably, in triple-degradation scenarios, permuting the correspondence does not affect performance because all DRMs are invoked jointly to handle the compound degradations. By contrast, in single- or double-degradation settings, mismatches between DRM and the degradation lead to clear performance drops.

Table 3. Ablation of DENet Training Strategies. The best results for each metric are highlighted with Red backgrounds.

Module	DAH		DEH	SEDA	
Metric	EM-ACC	F1	MAE ↓	PSNR	SSIM
Independent	0.966	0.995	0.031	26.131	0.862
Joint	0.957	0.991	0.034	26.085	0.862

Table 4. Impact of DRM-degradation assignment on performance. The best and second-best results for each metric are highlighted with Red and Blue backgrounds, respectively.

	Contrast Enhancement	Deblurring	Denoising	Avg. Metrics	
				PSNR	SSIM
1	NDRM	CDRM	BDRM	21.064	0.801
2	CDRM	NDRM	BDRM	24.771	0.836
3	NDRM	BDRM	CDRM	21.816	0.807
4	BDRM	CDRM	NDRM	23.275	0.833
5	CDRM	BDRM	NDRM	26.085	0.862

Finally, we validate the effectiveness of the 2D-SE–based weighted aggregation in SE-ROS by replacing it with simple averaging for the selected features. Because averaging ignores differences in feature importance, performance in the triple-degradation setting declines slightly, with PSNR/SSIM reduced by 0.212 dB and 0.007, respectively.

7. Limitations and Future Work

We follow the degradation synthesis strategy designed in PPFN [10], where TIR degradations are categorized into low contrast, blur, and noise, and training samples are generated accordingly. However, as noted in PPFN, obtaining strictly paired degraded–clean TIR images is inherently difficult, and any degradation pipeline constructed from limited priors can hardly cover the full range of more complex and diverse real-world degradations. Consequently, when encountering degradation types or atypical combinations that lie outside the training distribution, the performance of the model may still deteriorate to some extent. Nevertheless, experiments on real-world datasets such as AWMM demonstrate that the proposed SEGD still exhibits satisfactory generalization under no-reference quality metrics.

From a design standpoint, SEGD is highly extensible: to support additional degradation types, one can simply append the corresponding DRMs without modifying the over-

all framework. However, increasing the number of degradation types also introduces a non-negligible cost: as the number of types grows, the candidate restoration paths expand combinatorially, which in turn increases the number of trainable parameters and the training time. By contrast, the impact on the SE-ROS strategy is relatively minor, since the underlying linear-time 2D-SE minimization algorithm [15] incurs very low inference overhead on small graphs, and all quantities admit dynamic updates. Thus, we can perform constant-time updates of each node's 2D-SE contribution.

Overall, compared with visible-image enhancement, TIR enhancement is equally important yet remains relatively underexplored. In future work, we plan to: (i) construct more comprehensive and physically consistent TIR degradation models so that synthesized training data better cover real-world degradations; and (ii) develop restoration paradigms that accommodate unknown or even open-world degradations, enabling stable and reliable performance under extreme or previously unseen conditions.

References

[1] Jun Cheng, Tao Liu, and Shan Tan. Score priors guided deep variational inference for unsupervised real-world single image denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12937–12948, 2023. 4

[2] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4641–4650, 2021. 4

[3] Daniel Feijoo, Juan C Benito, Alvaro Garcia, and Marcos V Conde. Darkir: Robust low-light image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10879–10889, 2025. 3

[4] JiaKui Hu, Lujia Jin, Zhengjian Yao, and Yanye Lu. Universal image restoration pre-training via degradation classification. In *Proceedings of the International Conference on Learning Representations*, pages 1–16, 2025. 4, 5

[5] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5886–5895, 2023. 4

[6] Lingshun Kong, Jiangxin Dong, Jinhui Tang, Ming-Hsuan Yang, and Jinshan Pan. Efficient visual state space model for image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12710–12719, 2025. 4

[7] Guanzhou Lan, Qianli Ma, Yuqi Yang, Zhigang Wang, Dong Wang, Xuelong Li, and Bin Zhao. Efficient diffusion as low light enhancer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21277–21286, 2025. 3

[8] Wooseok Lee, Sanghyun Son, and Kyoung Mu Lee. Ap-bsn: Self-supervised denoising for real-world images via asymmetric pd and blind-spot network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17725–17734, 2022. 4

[9] Angsheng Li and Yicheng Pan. Structural information and dynamical complexity of networks. *IEEE Transactions on Information Theory*, 62(6):3290–3339, 2016. 1

[10] Jinyuan Liu, Zihang Chen, Zhu Liu, Zhiying Jiang, Long Ma, Xin Fan, and Risheng Liu. Enhancing infrared vision: Progressive prompt fusion network and benchmark. In *Advances in Neural Information Processing Systems*, pages 1–24. Curran Associates, Inc., 2025. 5

[11] Long Ma, Tengyu Ma, Chengpei Xu, Jinyuan Liu, Xin Fan, Zhongxuan Luo, and Risheng Liu. Learning with self-calibrator for fast and robust low-light image enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 3

[12] Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In *Advances in Neural Information Processing Systems*, pages 1–11, 2018. 1

[13] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *Proceedings of the European Conference on Computer Vision*, pages 146–162. Springer, 2022. 4

[14] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2654–2662, 2023. 3

[15] Yantuan Xian, Pu Li, Hao Peng, Zhengtao Yu, Yan Xiang, and Philip S Yu. Community detection in large-scale complex networks via structural entropy game. In *Proceedings of the 2025 ACM Web Conference (WWW)*, pages 3930–3941. ACM, 2025. 1, 6

[16] Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yuedong Tan, Danda Pani Paudel, Yulun Zhang, and Radu Timofte. Complexity experts are task-discriminative learners for any image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12753–12763, 2025. 4, 5

[17] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2696–2705, 2020. 4

[18] 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. 2

[19] Yunhao Zou, Chenggang Yan, and Ying Fu. Iterative denoiser and noise estimator for self-supervised image denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13265–13274, 2023. 4