

DEAL: Data-Efficient Adversarial Learning for High-Quality Infrared Imaging

Zhu Liu[†], Zijun Wang[†], Jinyuan Liu[‡], Fanqi Meng[†], Long Ma[†], Risheng Liu^{†*}

[†]School of Software Technology, Dalian University of Technology

[‡]School of Mechanical Engineering, Dalian University of Technology

liuzhu@mail.dlut.edu.cn, wzijun6@gmail.com, rsliu@dlut.edu.cn

Abstract

Thermal imaging is often compromised by dynamic, complex degradations caused by hardware limitations and unpredictable environmental factors. The scarcity of high-quality infrared data, coupled with the challenges of dynamic, intricate degradations, makes it difficult to recover details using existing methods. In this paper, we introduce thermal degradation simulation integrated into the training process via a mini-max optimization, by modeling these degraded factors as adversarial attacks on thermal images. The simulation is dynamic to maximize objective functions, thus capturing a broad spectrum of degraded data distributions. This approach enables training with limited data, thereby improving model performance. Additionally, we introduce a dual-interaction network that combines the benefits of spiking neural networks with scale transformation to capture degraded features with sharp spike signal intensities. This architecture ensures compact model parameters while preserving efficient feature representation. Extensive experiments demonstrate that our method not only achieves superior visual quality under diverse single and composited degradation, but also delivers a significant reduction in processing when trained on only fifty clear images, outperforming existing techniques in efficiency and accuracy. The source code will be available at <https://github.com/LiuZhu-CV/DEAL>.

1. Introduction

Infrared imaging [1, 52], as a robust and insensitive technique for capturing thermal radiation with salient features, plays a significant role in various visual perception tasks, such as multi-modality image fusion [34, 35], small target detection [25], and semantic segmentation [28, 42]. Due to its ability to effectively capture and convey temperature information, infrared imagery demonstrates reliability and environmental adaptability in adverse conditions [45].

Unfortunately, infrared modality is inevitably characterized by complex degradation factors in real-world dynamic and extreme imaging conditions [30, 36]. These corruptions are naturally detrimental to visual human observation and downstream semantic perception tasks. Recently, numerous deep learning-based methods have been proposed, which can be roughly divided into two typical methodologies: multi-modality image fusion and single-image infrared enhancement. Various attempts for image fusion have been made to design information measurements (e.g., salience measuring [38], feature richness [53]) cross-modal attention [28, 50], and correlation-driven fusion [29]), thereby mitigating degradation through visible features. However, in extreme scenes, degradation of visible light is severe.

Specifically, in the realm of single-image infrared enhancement [4, 48, 49], learning-based methods targeting specific single degradations are widely investigated, having achieved promising breakthroughs. For instance, Generative Adversarial Networks (GANs) [44] have garnered significant attention for infrared super-resolution, incorporating recursive attention network [31], degradation model-guided learning [11], and more. Introducing supplementary information (e.g., visible modality), multi-modality fusion modules [17] and transfer learning strategies [16] have been proposed for image super-resolution. Moreover, multi-scale feature representations are widely utilized for infrared stripe denoising [54]. However, two major obstacles hinder the advancement of infrared enhancement research. (1) Scarcity of high-quality thermal data: Infrared imaging is affected by dynamic changes due to temperature-driven thermal radiation and inherent hardware limitations. Selecting appropriate models that can address various degradation factors and intensities is challenging [19, 32, 51]. (2) Lack of effective mechanisms tailored for thermal degradation: Existing methods mostly depend on models designed for visible imaging, which are not optimized for unique thermal characteristics. Our objective is to offer a comprehensive solution for complex degraded scenes, effectively utilizing limited training data.

To partially mitigate these issues, we present the first at-

*Corresponding author.

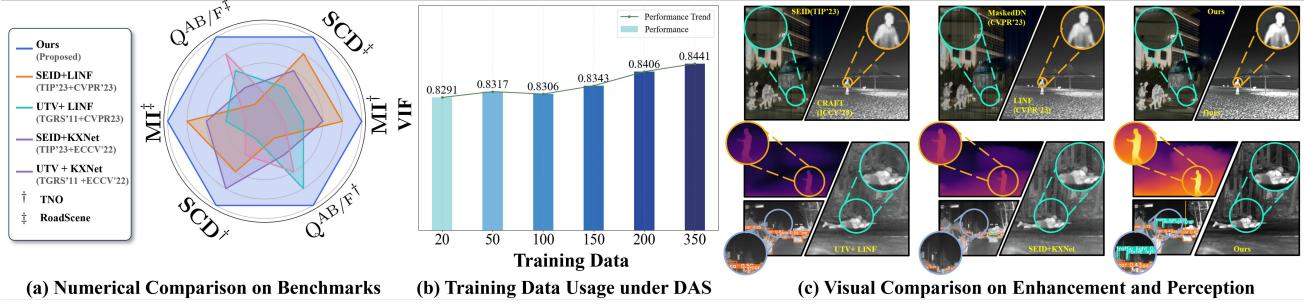


Figure 1. Comprehensive evaluation: numerical benchmark comparisons, training data efficiency of DAS, and perceptual comparisons.

tempt at establishing a unified paradigm for various thermal degradation. We formulate the learning process as an adversarial framework, separating it into dynamic degradation generation and image enhancement. The adversarial degradation aims to challenge the enhancement network by applying gradient ascent on the training objective, adaptively simulating degraded imaging with diverse levels of thermal-specific corruptions, including stripe noise, low contrast, and spatial resolution. Meanwhile, the enhancement process focuses on restoring the textural details of these degradation factors. We introduce a dynamic adversarial training strategy for progressive degradation generation and infrared image enhancement. Additionally, to capture unique infrared degradation characteristics, we propose a dual-interaction network combining paired scale transformation operations and a spiking-inspired module with dense connections. To summarize, our contributions are four-fold:

- We make the attempt to address various degradations simultaneously in infrared images. Introducing dynamic generation makes our method more applicable to the limitation of high-quality training data.
- From a learning standpoint, we design a dynamic adversarial solution, which effectively strengthens the performance of the enhancement network.
- From the architectural side, we introduce a dual interaction network to better capture degradation, fully exploring the potential of spiking characteristics with nimble parameters and efficient inference.
- Extensive experiments conducted on various degraded scenarios including single and composited degradation factors and perception tasks such as depth estimation and detection substantiate the significant advancements of proposed method, using only fifty images for training.

2. Proposed Method

2.1. Problem Formulation

In real-world thermal imaging, multiple degradations may occur simultaneously, resulting in a scarcity of high-quality training data. This situation also presents a challenge, as it

requires the selection of various learning models to address different types and intensities of corruption. Dynamically generating various thermal-specific degradations to train an all-in-one model is crucial for infrared imaging.

Adversarial formulation of degradation. To establish the relationship between thermal-specific degradation generation and image enhancement, we adhere to the adversarial learning paradigm to enhance the robustness of enhancement and alleviate the amount of training data. A hierarchical mini-max optimization is defined as follows:

$$\min_{\omega} \mathcal{L}(\mathcal{N}_E(\hat{\mathbf{x}}; \omega), \mathbf{y}), \quad (1)$$

$$\text{s.t. } \begin{cases} \hat{\mathbf{x}} = \mathcal{N}_G(\mathbf{x}; \theta^*), \\ \theta^* = \arg \max_{\theta} \mathcal{L}(\mathcal{N}_E(\mathcal{N}_G(\mathbf{x}; \theta); \omega), \mathbf{y}), \end{cases} \quad (2)$$

where \mathcal{N}_G and \mathcal{N}_E represent the degradation generation with parameters θ and the enhancement network with parameter ω . \mathbf{x} , $\hat{\mathbf{x}}$, and \mathbf{y} denote the original infrared images, corrupted inputs, and the corresponding ground truths, respectively. \mathcal{L} denotes the loss functions of image enhancement. The lower-level maximization sub-problem aims at generating changeable combinations of various thermal degradation, which can significantly impair the performance of the image enhancement network. Thus, the lower-level objective is to maximize losses related to infrared enhancement. The upper-level sub-problem involves a general optimization procedure for image enhancement based on the training of generated degradation $\hat{\mathbf{x}}$, aiming to restore images from various degradation conditions with high robustness. Additionally, mini-max adversarial optimization can equip the enhancement network with a complex degradation-aware capability. Additionally, the pipeline of entire framework is depicted in Fig. 2 (a).

Initially, with regard to the degradation generation network \mathcal{N}_G , we incorporate a classifier [40] to generate weights that simulate the degradation process. This classifier takes the original infrared image \mathbf{x} as input and produces a differentiable weight matrix \mathbf{a} to control the various degrees of various degradation. Denoting one type of degra-

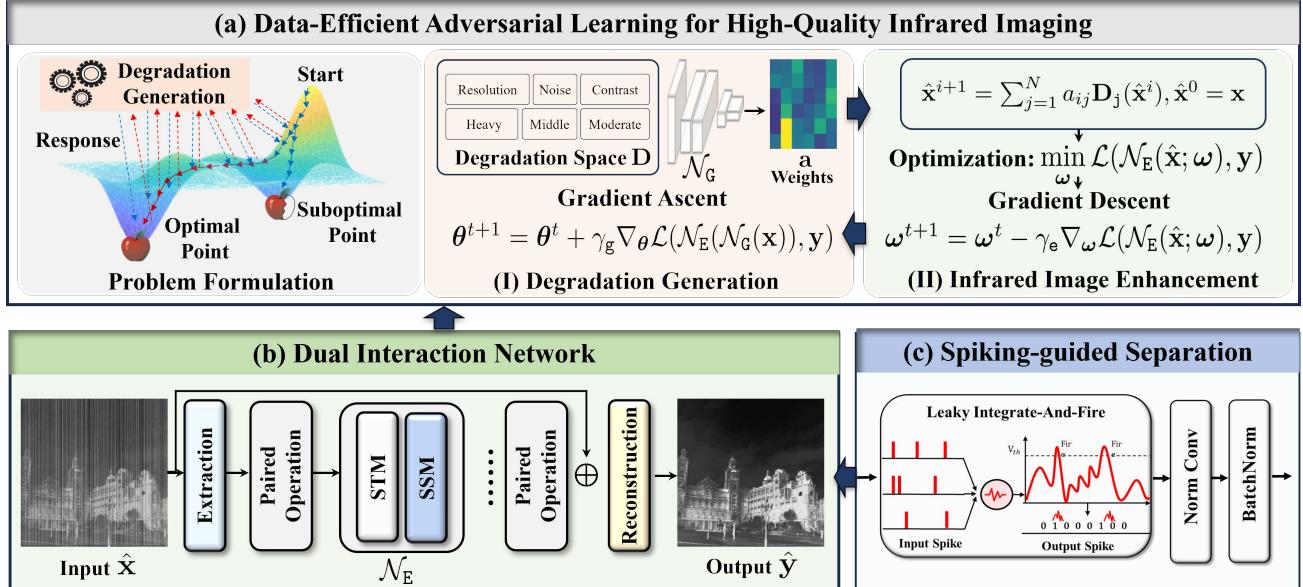


Figure 2. Schematic of the major components of proposed paradigm. We present a data-efficient adversarial learning strategy, which constructs the dynamic degradation generation to guide the image enhancement procedure with a Dynamic Adversarial Solution (DAS) at (a). The concrete architecture of dual interaction network, consisting of Scale Transform Module (STM) and Spiking-guided Separation Module (SSM) is shown at (b). Spiking-guided Separation to capture sharp intensities of thermal degradation is depicted at (c).

Algorithm 1 Dynamic Adversarial Solution (DAS).

Require: Infrared datasets with $\{\mathbf{x}, \mathbf{y}\}$, learning rate γ_e, γ_g and other necessary hyper-parameters.

- 1: % Warm starting the enhancement network.
- 2: $\omega^{t+1} = \omega^t - \gamma_e \nabla_\omega \mathcal{L}(\mathcal{N}_E(\hat{\mathbf{x}}; \omega), \mathbf{y})$;
- 3: **while** not converged **do**
- 4: % Optimization of infrared image enhancement.
- 5: Generating the degraded samples by \mathcal{N}_G and setting batch as $\{\hat{\mathbf{x}}_1, \mathbf{y}_1, \dots, \hat{\mathbf{x}}_N, \mathbf{y}_N\}$;
- 6: $\omega^{t+1} = \omega^t - \gamma_e \nabla_\omega \mathcal{L}(\mathcal{N}_E(\hat{\mathbf{x}}; \omega), \mathbf{y})$;
- 7: % Gradient update of corruption generation.
- 8: $\theta^{t+1} = \theta^t + \gamma_g \nabla_\theta \mathcal{L}(\mathcal{N}_E(\mathcal{N}_G(\mathbf{x}; \theta); \omega), \mathbf{y})$;
- 9: **end while**
- 10: **return** ω^* .

dition as \mathbf{D} , the generation can be formulated as:

$$\hat{\mathbf{x}}^{i+1} = \sum_{j=1}^N a_{ij} \mathbf{D}_j(\hat{\mathbf{x}}^i), \hat{\mathbf{x}}^0 = \mathbf{x}. \quad (3)$$

We consider three major degradations in thermal imaging, including stripe streaks, low-resolution, and low contrast. In detail, infrared striped noise typically arises due to non-uniformities in the sensor, such as the effects of diverse column responses of the focal plane array. The low resolution of infrared imaging is widely acknowledged, caused by the propagation characteristics of thermal radiation. Limited by sensor performance, temperature differences, and

signal processing algorithms, low contrast in infrared images can stem from various factors. Furthermore, for the network of infrared image enhancement \mathcal{N}_E , we propose a dual interaction network to uniformly process degradation. The detailed architecture is proposed in Section 2.2.

Dynamic adversarial solution. We propose a dynamic adversarial solution for the training procedure, providing an optimization strategy for both Eq. (1) and Eq. (2).

In detail, directly utilizing adversarial corruptions can easily result in the non-convergence of enhancement and failure in scene restoration. Therefore, we initially warm-start the enhancement network \mathcal{N}_E , which is frozen the degradations generated by \mathcal{N}_G for several epochs based on gradient descent techniques, which can be formulated as:

$$\omega^{t+1} = \omega^t - \gamma_e \nabla_\omega \mathcal{L}(\mathcal{N}_E(\hat{\mathbf{x}}; \omega), \mathbf{y}). \quad (4)$$

Inspired by hierarchical training strategies (e.g., generative adversarial learning [16] and adversarial attack [43]), we construct a single-loop alternative solution. Since the adoption of relaxed aggregation with learnable weights \mathbf{a} , we can optimize the generation parameters θ through gradient ascent, which can be written as:

$$\theta^{t+1} = \theta^t + \gamma_g \nabla_\theta \mathcal{L}(\mathcal{N}_E(\mathcal{N}_G(\mathbf{x}; \theta); \omega), \mathbf{y}). \quad (5)$$

After obtaining the degraded images $\hat{\mathbf{x}}$, we further utilize gradient descent with Eq.(4) to update ω for the enhancement network. We construct a single-loop optimization [33, 37, 39] to update the generation and enhancement network alternately. The solution is summarized in Alg. 1.

Levels	Metrics	MaskedDN	GF	SEID	WDNN	WFAF	UTV	LRSID	Ours
Moderate	MI	3.0901	3.2738	3.2665	3.3383	3.3204	3.4257	3.1504	3.3972
	VIF	0.8714	0.9217	0.9446	0.9252	0.9130	0.9302	0.8270	0.9607
	SD	10.668	10.668	10.638	10.647	10.669	10.690	10.655	11.050
	$Q^{AB/F}$	0.4539	0.5001	0.5017	0.4937	0.4861	0.5157	0.4468	0.4523
	SCD	1.3668	1.3757	1.4278	1.3400	1.3450	1.3002	1.3509	1.4467
Heavy	MI	2.6092	3.0228	3.0241	3.1129	3.0752	3.0752	3.1374	3.2443
	VIF	0.9581	1.0483	1.0721	1.0685	1.0508	1.0854	1.0681	1.0982
	SD	10.277	10.288	10.327	10.283	10.285	10.294	10.280	10.575
	$Q^{AB/F}$	0.3731	0.5003	0.5039	0.4953	0.4870	0.5183	0.5152	0.5520
	SCD	1.4596	1.4619	1.4882	1.4198	1.4238	1.4009	1.4204	1.5116

Table 1. Numerical comparisons of the enhancement on stripe noise with seven advanced competitors.

Scale	Metric	LINF	CRAFT	HAT	ETDS	BTC	KXNet	SwinIR	SR-LUT	FeMaSR	Ours
$\times 2$	MI	2.9892	2.9731	2.9683	2.9696	2.9794	2.9839	2.9723	2.9666	2.9397	3.0536
	VIF	1.0676	1.0651	1.0646	1.0646	1.0662	1.0627	1.0651	1.0629	1.0519	1.0511
	SD	10.265	10.263	10.264	10.263	10.262	10.264	10.263	10.269	10.251	10.495
	$Q^{AB/F}$	0.5133	0.5041	0.5022	0.5026	0.5067	0.5147	0.5028	0.5028	0.4985	0.4698
	EN	6.6534	6.6572	6.6579	6.6571	6.6544	6.6574	6.6572	6.6631	6.6395	6.8261
$\times 4$	MI	3.0031	2.9958	2.9966	2.9901	3.0020	2.9738	2.9941	2.9793	2.8927	3.1162
	VIF	1.0608	1.0482	1.0504	1.0449	1.0481	1.0549	1.0482	1.0449	1.0268	1.0648
	SD	10.267	10.264	10.259	10.266	10.251	10.268	10.263	10.263	10.306	10.511
	$Q^{AB/F}$	0.529	0.5303	0.5275	0.5296	0.5284	0.5055	0.5295	0.5220	0.4774	0.4996
	EN	6.6659	6.6764	6.6772	6.6721	6.6730	6.6704	6.6767	6.6745	6.6903	6.8287

Table 2. Numerical comparisons of image super-resolution task with nine state-of-the art methods.

Discussion with augmentation techniques. Existing techniques (*e.g.*, Augmix [15], AdaAugment [56]) typically rely on static, predefined operations, such as rotation, translation, or adjusting brightness for the whole dataset, lacking adaptability to the dynamic and complex degradations. These techniques do not dynamically adjust the type and intensity of degradation based on the model’s weaknesses during training, thus falling short in enhancing model performance. In contrast, our method introduces adversarial degradation generation, treating degradation as an adaptive attack against the enhancement network.

2.2. Dual Interaction Network

We propose a dual interaction network, composed of paired operations, including scale transform and spiking-guided separation. The architecture is plotted in Fig. 2 (b).

Scale transform module. Scale transform is widely leveraged for image restoration tasks [29], which can effectively enhance the spatial structure and remove noise and artifacts. We construct a Scale Transform Module (STM) to further fine contextual information at the feature level. We utilize the up-sampling and down-sampling with one convolution operations to scale features with dense connections, following with literature [41].

Spiking-guided separation module. We observe that the pixel intensity of degradation, such as infrared stripe noise, is typically higher than in other regions of the images. This type of noise usually results from inconsistencies or calibration errors in the response to infrared wavelengths, lead-

ing to anomalous brightness or intensity levels. In Spiking Neural Networks (SNNs), information is encoded as discrete spikes (*e.g.*, binary signals), which can be leveraged to align the thermal-specific noise pattern more effectively by causing more neurons to fire. Thus, we propose a Spiking-Guided Separation Module (SSM) to isolate degraded features. As shown in Fig. 2 (c), it comprises three components: Leaky Integrate-and-Fire (LIF) neurons to convert spatial features into spike sequences, standard convolution, and threshold-dependent batch normalization. These paired operations can ensure compact parameters, low energy consumption, and effective analysis of thermal degradation.

3. Experiment Results

In this section, we first elaborate on the concrete experimental settings. Next, we provide comprehensive comparisons for single corruption. Following that, we conduct experiments under general infrared imaging benchmarks. Lastly, we perform sufficient ablation studies on learning strategies and architectures. Details of the degradation generation are provided in the supplementary material.

We used the PyTorch framework to implement models on a single V100 GPU. The degradation model was optimized using SGD, while the enhancement model utilized Adam Optimizer. The learning rates γ_e and γ_g were set to $1e^{-4}$ and $2e^{-4}$. We only trained on 50 images from M3FD [27] for 840 epochs. The total loss function \mathcal{L} in Eq. (1) and Eq. (2), composited by the measurement of pixel intensities \mathcal{L}_{pixel} and structural similarity \mathcal{L}_{SSIM} .

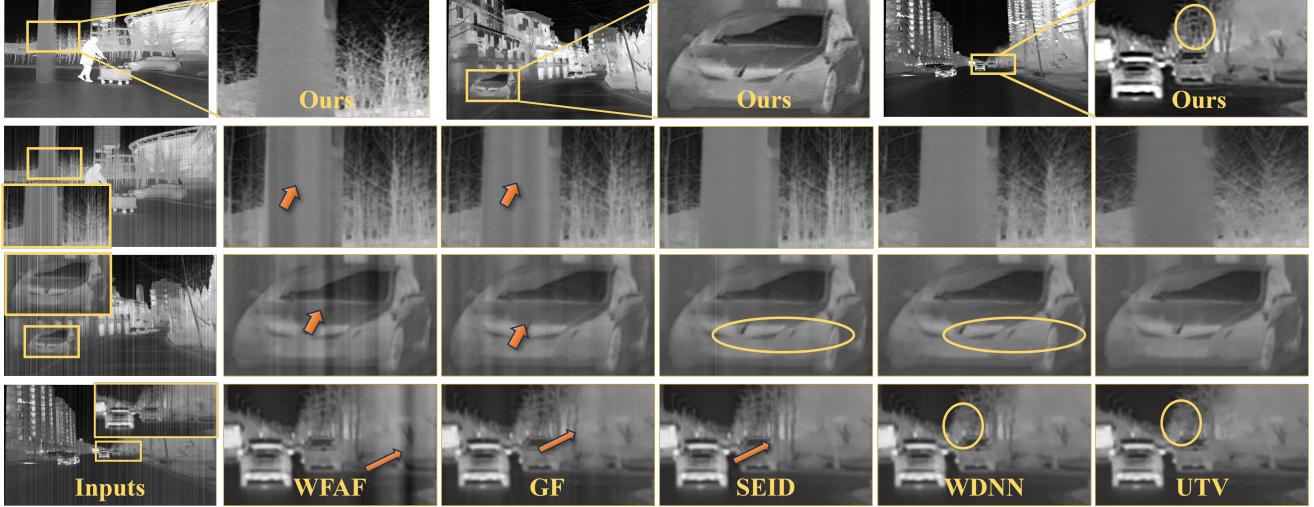


Figure 3. Visual comparisons of stripe noise removing with several advanced competitors.

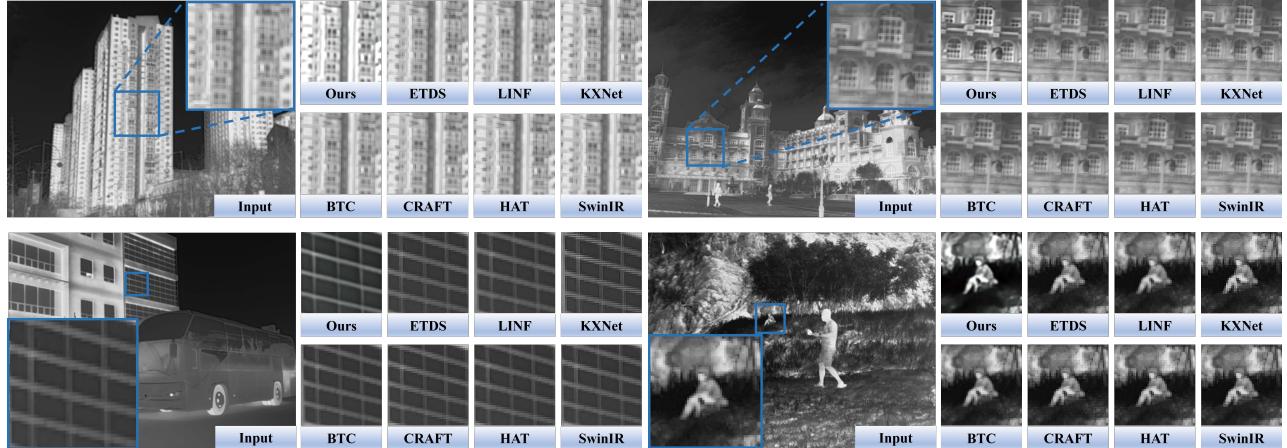


Figure 4. Infrared image super-resolution compared with several state-of-the-art methods.

The whole formulation of the criterion can be written as: $\mathcal{L}(\hat{\mathbf{y}}; \mathbf{y}) = \alpha \mathcal{L}_{\text{pixel}}(\hat{\mathbf{y}}; \mathbf{y}) + \beta \mathcal{L}_{\text{SSIM}}(\hat{\mathbf{y}}; \mathbf{y})$, where α and β are the trade-off parameters of the related loss functions. In this paper, we set α and β as 0.75 and 1.1, respectively. As for the degradation generation, we leverage the negative losses $-\mathcal{L}(\hat{\mathbf{y}}; \mathbf{y})$ with $\mathcal{L}(\hat{\mathbf{x}}; \mathbf{x})$ for optimization.

3.1. Comparisons on Single Degradation

Note that our method simultaneously addresses diverse degradation factors. It is challenging to establish appropriate ground truths and utilize widely-used reference-based metrics (e.g., PSNR and SSIM) to objectively measure performance with specialized schemes fairly. Thus we introduce the measurements of fusion for numerical evaluation.

Stripe noise restoration. In this part, we introduce seven competitive methods for comparison, including MaskedDN [9], GF [5], SEID [49], WDNN [2], WFAF [46], UTV [3], and LRSID [6] on the stripe noise

scenes. The subjective results under two degrees of stripe noise are reported in Table 1. We adopt ReCo [18] as a basic fusion method for a fair comparison. Three remarkable advantages of our proposed method can be concluded from this table. Firstly, our scheme can effectively restore the inherent characteristics of infrared images with high data fidelity with the highest MI. Secondly, our method generates visual-appealing results (higher VIF and $Q^{\text{AB/F}}$), which is consistent with the human visual perception system. Lastly, the results of our method illustrate the high contrast effects, indicated by the highest SD and SCD. Moreover, visual comparisons are depicted in Fig.3, showcasing three challenging scenes: dense textures, low contrast, and weak details. It is noticeable that GF and WDNN cannot effectively preserve the sharp structure and smooth textural details. Our method presents two key advantages: it effectively removes stripe noise without leaving visible residues,

Datasets	TNO					RoadScene				
Metrics	LINF+SEID	LINF+UTV	KXNet+SEID	KXNet+UTV	Ours	LINF+SEID	LINF+UTV	KXNet+SEID	KXNet+UTV	Ours
MI	2.2107	2.1300	2.1992	2.0724	2.3950	3.2690	3.1268	3.2625	3.1232	3.3477
SD	8.9578	8.8680	8.8822	8.7715	9.1313	10.321	10.201	10.300	10.179	10.660
$Q^{AB/F}$	0.3888	0.3562	0.3854	0.3410	0.4176	0.4243	0.4187	0.4246	0.4194	0.4470
SCD	1.3480	1.3635	1.3576	1.3643	1.4454	1.5411	1.5938	1.5433	1.5935	1.6131
EN	6.6543	6.6116	6.6393	6.6135	6.7835	6.9861	6.9466	6.9884	6.9473	7.1000

Table 3. Numerical comparisons of joint enhancement under two general benchmarks.

Fusion	Metrics	AirNet	DALE	IAT	MBLLEN	Ours
ReCo	MI	2.4599	1.7566	2.8294	2.7484	3.1636
	VIF	1.0319	0.9762	0.9980	1.0486	0.9046
	SD	9.0136	8.5928	9.3036	9.7832	10.7724
	$Q^{AB/F}$	0.4772	0.4850	0.5030	0.5357	0.5001
	EN	6.0448	6.0811	6.6712	6.5364	6.7142
LRR [24]	MI	3.1392	2.9433	3.0799	2.9304	3.0118
	VIF	0.7390	0.8086	0.8155	0.7843	0.8200
	SD	7.6213	7.9167	8.0923	7.7350	9.9614
	$Q^{AB/F}$	0.4680	0.4463	0.4915	0.5330	0.5219
	EN	6.3720	6.1191	6.4803	6.3241	6.5596

Table 4. Numerical results of low contrast enhancement tasks.

as seen in smooth surfaces, clearly enhances thermal radiation information with well-preserved structural details, such as the shapes of pedestrians and forest.

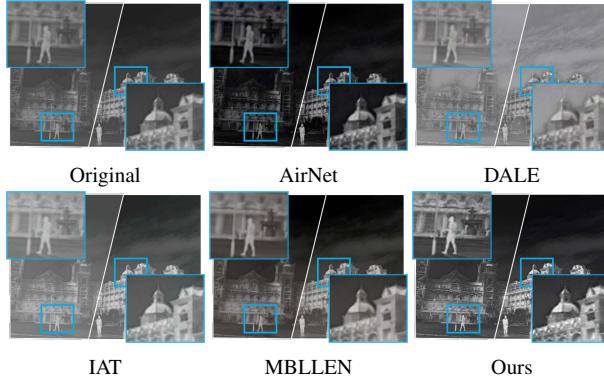


Figure 5. Visual comparisons of low-contrast enhancement with several advanced competitors under different corrupted degrees.

Image super-resolution. We compare nine latest state-of-the-art methods for image super-resolution, including LINF [57], CRAFT [22], HAT [10], ETDS [7], BTC [47], KXNet [14], SwinIR [26], SR-LUT [20], and FeMaSR [8]. The inputs are with same size. Table 2 presents numerical results at $\times 2$, $\times 4$ resolution, respectively. In this study, we replaced the SCD with Entropy (EN) to enhance the measurement of textural details (modal information) present in the features. The table demonstrates the consistently superior performance of our method across five representative metrics. The informative metrics (MI and EN) underscore our method’s ability to preserve more textural details in infrared images. Moreover, the highest SD indicates a

Data Usage	20	50	100	150	200	350
MI	3.0373	3.0891	3.0393	3.1075	3.102	3.0891
VIF	0.8291	0.9310	0.8306	0.8343	0.8406	0.8441
SD	10.3819	10.3432	10.4543	10.3321	10.3129	10.4764
$Q^{AB/F}$	0.445	0.4821	0.4454	0.4531	0.4555	0.4571
EN	6.811	6.8188	6.8014	6.8088	6.7923	6.7891

Table 5. Stripe denoising performance under diverse data usage.

substantial enhancement in contrast across different fusion methods. The qualitative comparison under four challenging scenes is depicted in Fig. 4. We can summarize two significant characteristics compared with existing methods. Firstly, most current schemes generate noticeable artifacts with blurred details caused by the bi-cubic downsampling. As shown in the third case, KXNet and CRAFT cannot effectively persevere the structures with lines. Our method can effectively restore the high-frequency textural details (*e.g.*, dense structure in the first case), which realizes the visual-appealing result and is beneficial for human observation. Furthermore, our method comprehensively considers the impacts of multi-degradation, which can significantly highlight the salient thermal objects with sharp structures (*e.g.*, street lamp in the second case). However, existing methods only consider one single degradation, which cannot effectively handle the contrast information.

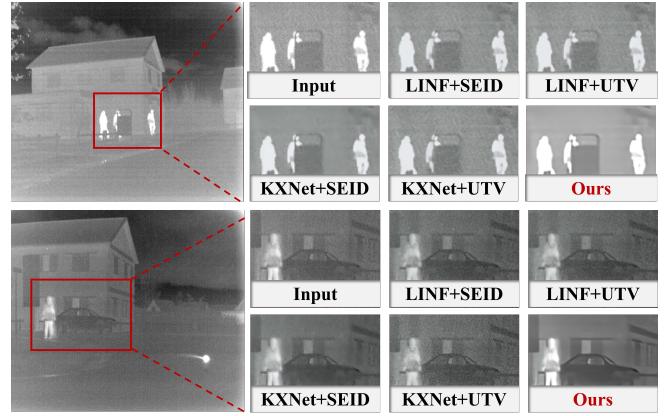


Figure 6. Infrared enhancement on two challenging scenarios.

Low-contrast enhancement. Moreover, we also compare our method under the low-contrast scenario. Due to

there are a few methods designed for the low-contrast enhancement for infrared images, we adopt several all-in-one image restoration and contrast enhancement schemes for visible images, including AirNet [23], DALE [21], IAT [12], and MBLLEN [13]. Table 4 presents the results of low-contrast enhancement compared with various methods. Remarkably, our method consistently demonstrates improvements across diverse metrics. In summary, our approach effectively preserves edge information and textural details, resulting in high-contrast observations. Furthermore, we provide two typical scenarios to illustrate the superiority in Fig. 5. Although the all-in-one method AirNet increases the contrast, it interferes with the infrared information. DALE and IAT improve the global intensity of the thermal image but introduce blurred texture details. In contrast, our scheme can effectively highlight thermal targets (e.g., pedestrians and building details) while preserving clear and sharp structures. This provides visually appealing observations that are consistent with human visual system.

3.2. Comparisons on Composed Degradation

We also demonstrate the effectiveness of our method using widely-used benchmarks TNO and RoadScene, which encompass diverse composited degradation factors. We select representative state-of-the-art methods for these corruptions to conduct a comparison, including UTV, SEID, LINF, and KXNet, uniformly enhanced by IAT and fused by ReCo. Table 3 reports the numerical results. Specifically, the optimal MI and EN values indicate that our method effectively highlights thermal information in the fusion schemes. The highest SD value illustrates that our method enhances contrast for human observation. Additionally, the highest $Q^{AB/F}$ value demonstrates our method's remarkable ability to extract textural details and edge information.

Additionally, we present an illustrative visual comparison in Fig. 6 featuring four challenging scenarios. These two examples are from the TNO dataset and are corrupted by heavy stripe noise and low resolution. Combined methods (e.g., LINF+SEID, KXNet+UTV) are unable to handle the complex noise, resulting in significant textural artifacts. Other combined schemes also exhibit residual artifacts due to the hybrid degradation. Clearly, UTV-based schemes significantly degrade structure and textural details due to interference from low contrast factors. Importantly, our method effectively highlights thermal-sensitive objects while preserving sufficient textural details, representing a promising improvement in real-world thermal imaging.

3.3. Ablation Study

Validation about the amounts of training data. Our method achieves excellent performance even with a limited amount of training data in Table 5. Specifically, with smaller datasets (e.g., 20 and 50 images), metrics such as

Operations	Visual Metrics		Hardware Efficiency		
	VIF	Qabf	Runtime(ms)	FLOPs(G)	Params.(M)
Cascade-1	0.9929	0.5445	80.84	303.4	0.382
Cascade-2	0.9945	0.5545	121.6	455.6	0.572
Cascade-3	1.0001	0.5478	331.7	762.4	0.960

Table 6. Effectiveness of the paired operation.

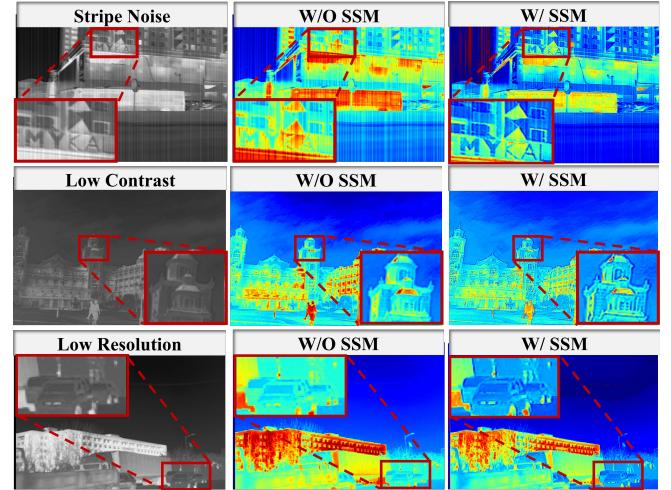


Figure 7. Feature visualization based on w/ and w/o SSM.

Degradation	Metrics	Average	All	Proposed
Stripe Noise	VIF	0.7162	0.7467	0.9310
	$Q^{AB/F}$	0.3965	0.3887	0.4821
Low Resolution	VIF	0.6698	0.6647	0.8401
	$Q^{AB/F}$	0.4105	0.3929	0.4458
Low Contrast	VIF	0.6530	0.6696	0.8941
	$Q^{AB/F}$	0.4029	0.3880	0.4924
Composited	VIF	0.7171	0.7636	0.9930
	$Q^{AB/F}$	0.3967	0.3971	0.5445

Table 7. Numerical verification for training strategies.

VIF, and $Q^{AB/F}$ already show relatively high values, indicating that our approach effectively enhances quality even with limited data.

Verification of paired operations. Table 6 demonstrates the effectiveness of paired operations in terms of visual metrics and hardware efficiency. Our method effectively balances model performance and computational efficiency by using a small number of cascades (only two paired operations in total), compared to other methods like BTC with 22.4M and KXNet with 6.51M parameters, respectively.

Effectiveness of training strategy. We compared the proposed learning strategy with two alternative strategies: averaging the mix of degradations and training with all degradations in Table 7. Clearly, the averaged aggregation significantly improve performance compared to the original joint training with random selection. Moreover, employing dy-

namic degradation generation can further enhances performance under diverse corruptions.

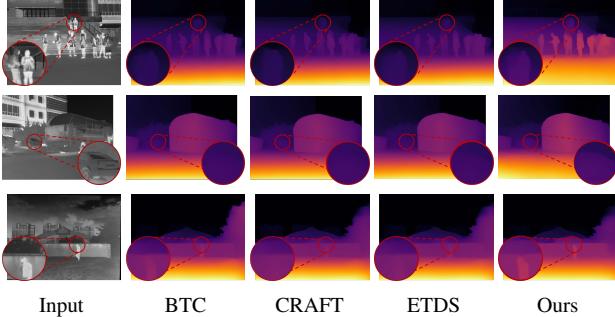


Figure 8. Depth estimation under low-resolution infrared images.

Validation of proposed architecture. We visualize the features with and without SSM in Fig. 7, demonstrating a noticeable improvement in identifying specific degradation factors. For instance, utilizing the SSM can effectively identify the stripe streaks. Furthermore, the structure of bumps and buildings also can be remarkably highlighted for the restoration. These instances demonstrate our superiority.

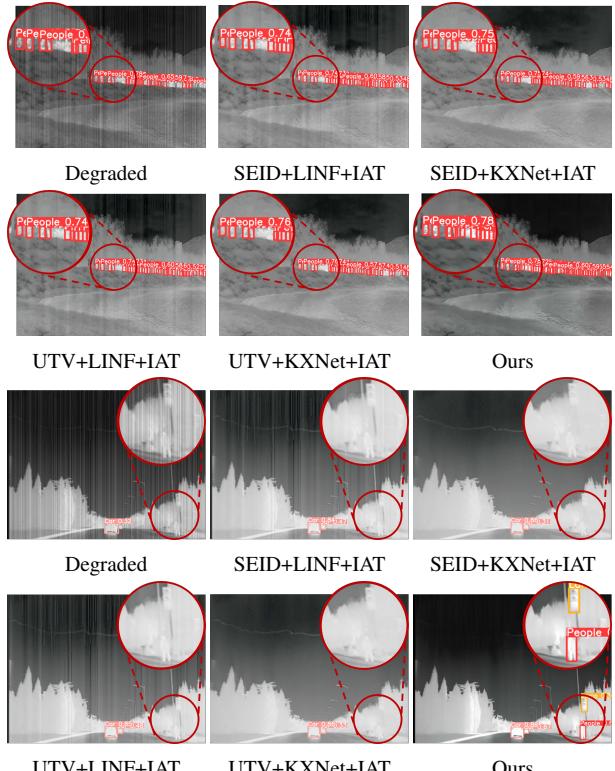


Figure 9. Object detection comparisons with several advanced competitors under three composed degradation factors.

Methods	Categories						
	People	Car	Bus	Lamp	Motor	Truck	mAP
Degraded	0.522	0.345	0.137	0.161	0.173	0.233	0.262
S+K+I	0.700	0.772	0.678	0.394	0.444	0.683	0.612
U+K+I	0.694	0.808	0.667	0.355	0.480	0.640	0.607
S+L+I	0.695	0.771	0.673	0.395	0.447	0.676	0.610
U+L+I	0.695	0.806	0.659	0.358	0.481	0.632	0.605
Ours	0.737	0.826	0.726	0.436	0.528	0.707	0.660

Table 8. Numerical results of object detection compared with four representative combinations on the M3FD dataset.

4. Applications

Depth estimation. We initially demonstrate the effectiveness of our method in the context of depth estimation using low-resolution infrared images. Fig. 8 illustrates the outcomes of depth estimation using the fundamental depth model, DepthAnything [55] under three scenarios. It is notable that employing existing enhancement techniques fails to accurately predict depth for distant objects, as evidenced by the shapes of pedestrian and car. In contrast, our approach effectively highlights salient objects, significantly improving depth estimation accuracy.

Object detection. We demonstrate the compatibility of our approach for object detection task based on YoLoV5 [27]. The numerical results are compared with various combined schemes in Table 8. Our method achieves consistently optimal precision across all categories, demonstrating a significant improvement of almost 8% in accuracy compared to the second-best method. Furthermore, we select two challenging scenes, one featuring small dense objects and another with a low-contrast scenario in Fig. 9. The first scene includes tiny dense targets, combined with several challenging degradation factors. Existing methods are unable to accurately detect dense pedestrians due to blurred object structures resulting from the denoising procedure. In contrast, our method can successfully detect all small and dense objects. The second scene presents a challenging low-contrast scenario, where existing methods fail to effectively highlight targets and shapes, leading to detection failures. In contrast, our method can effectively detect all targets, including pedestrians and traffic lamps.

5. Conclusion

This paper proposed a data-efficient adversarial learning scheme to address dynamic complex corruptions in high-quality thermal imaging. Based on the dynamic adversarial solution, the degradation generation module effectively learns to simulate thermal-specific corruption, addressing the scarcity of high-quality thermal data. A dual interaction architecture was proposed, integrating scale transform and spiking network for thermal-specific features. Extensive experiments illustrated the superiority of our schemes.

References

- [1] Fanglin Bao, Xueji Wang, Shree Hari Sureshbabu, Gautam Sreekumar, Liping Yang, Vaneet Aggarwal, Vishnu N Boddeti, and Zubin Jacob. Heat-assisted detection and ranging. *Nature*, 619(7971):743–748, 2023. 1
- [2] Arnaud Barral, Pablo Arias, and Axel Davy. Fixed pattern noise removal for multi-view single-sensor infrared camera. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1669–1678, 2024. 5
- [3] Marouan Bouali and Saïd Ladjal. Toward optimal destriping of modis data using a unidirectional variational model. *IEEE Transactions on geoscience and remote sensing*, 49(8):2924–2935, 2011. 5
- [4] Lijing Cai, Xiangyu Dong, Kailai Zhou, and Xun Cao. Exploring video denoising in thermal infrared imaging: Physics-inspired noise generator, dataset and model. *IEEE TIP*, 2024. 1
- [5] Yanpeng Cao, Michael Ying Yang, and Christel-Loic Tisse. Effective strip noise removal for low-textured infrared images based on 1-d guided filtering. *IEEE transactions on circuits and systems for video technology*, pages 2176–2188, 2015. 5
- [6] Yi Chang, Luxin Yan, Tao Wu, and Sheng Zhong. Remote sensing image stripe noise removal: From image decomposition perspective. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7018–7031, 2016. 5
- [7] Jiahao Chao, Zhou Zhou, Hongfan Gao, Jiali Gong, Zhengfeng Yang, Zhenbing Zeng, and Lydia Dehbi. Equivalent transformation and dual stream network construction for mobile image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14102–14111, 2023. 6
- [8] Chaofeng Chen, Xinyu Shi, Yipeng Qin, Xiaoming Li, Xiaoguang Han, Tao Yang, and Shihui Guo. Real-world blind super-resolution via feature matching with implicit high-resolution priors. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1329–1338, 2022. 6
- [9] Haoyu Chen, Jinjin Gu, Yihao Liu, Salma Abdel Magid, Chao Dong, Qiong Wang, Hanspeter Pfister, and Lei Zhu. Masked image training for generalizable deep image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1692–1703, 2023. 5
- [10] Xiangyu Chen, Xiantao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22367–22377, 2023. 6
- [11] Xiaohui Chen, Lin Chen, Lingjun Chen, Peng Chen, Guan-qun Sheng, Xiaosheng Yu, and Yaobin Zou. Modeling thermal infrared image degradation and real-world super-resolution under background thermal noise and streak interference. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024. 1
- [12] Ziteng Cui, Kunchang Li, Lin Gu, Shenghan Su, Peng Gao, Zhengkai Jiang, Yu Qiao, and Tatsuya Harada. You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. *arXiv preprint arXiv:2205.14871*, 2022. 7
- [13] Jianhua Wu Feifan Lv, Feng Lu and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. 2018. 7
- [14] Jiahong Fu, Hong Wang, Qi Xie, Qian Zhao, Deyu Meng, and Zongben Xu. Kxnet: A model-driven deep neural network for blind super-resolution. In *European Conference on Computer Vision*, pages 235–253. Springer, 2022. 6
- [15] Dan Hendrycks, Norman Mu, Ekin D Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. Augmix: A simple data processing method to improve robustness and uncertainty. *arXiv preprint arXiv:1912.02781*, 2019. 4
- [16] Yongsong Huang, Zetao Jiang, Rushi Lan, Shaoqin Zhang, and Kui Pi. Infrared image super-resolution via transfer learning and psrgan. *IEEE Signal Processing Letters*, 28: 982–986, 2021. 1, 3
- [17] Yongsong Huang, Tomo Miyazaki, Xiaofeng Liu, and Shinichiro Omachi. Infrared image super-resolution: Systematic review, and future trends. *arXiv preprint arXiv:2212.12322*, 2022. 1
- [18] Zhanbo Huang, Jinyuan Liu, Xin Fan, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Reconet: Recurrent correction network for fast and efficient multi-modality image fusion. In *European conference on computer Vision*, pages 539–555, 2022. 5
- [19] Zhiying Jiang, Zengxi Zhang, Jinyuan Liu, Xin Fan, and Risheng Liu. Multispectral image stitching via global-aware quadrature pyramid regression. *IEEE Transactions on Image Processing*, 2024. 1
- [20] Younghyun Jo and Seon Joo Kim. Practical single-image super-resolution using look-up table. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 691–700, 2021. 6
- [21] Dokyeong Kwon, Guisik Kim, and Junseok Kwon. Dale: Dark region-aware low-light image enhancement. *arXiv preprint arXiv:2008.12493*, 2020. 7
- [22] Ao Li, Le Zhang, Yun Liu, and Ce Zhu. Feature modulation transformer: Cross-refinement of global representation via high-frequency prior for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12514–12524, 2023. 6
- [23] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17452–17462, 2022. 7
- [24] Hui Li, Tianyang Xu, Xiao-Jun Wu, Jiwen Lu, and Josef Kittler. Lrrnet: A novel representation learning guided fusion network for infrared and visible images. *IEEE transactions on pattern analysis and machine intelligence*, 2023. 6
- [25] Xingyuan Li, Jinyuan Liu, Zhixin Chen, Yang Zou, Long Ma, Xin Fan, and Risheng Liu. Contourlet residual for prompt learning enhanced infrared image super-resolution. In *European Conference on Computer Vision*, pages 270–288, 2024. 1

- [26] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 6
- [27] Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5802–5811, 2022. 4, 8
- [28] Jinyuan Liu, Zhu Liu, Guanyao Wu, Long Ma, Risheng Liu, Wei Zhong, Zhongxuan Luo, and Xin Fan. Multi-interactive feature learning and a full-time multi-modality benchmark for image fusion and segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8115–8124, 2023. 1
- [29] Jinyuan Liu, Runjia Lin, Guanyao Wu, Risheng Liu, Zhongxuan Luo, and Xin Fan. Coconet: Coupled contrastive learning network with multi-level feature ensemble for multi-modality image fusion. *International Journal of Computer Vision*, 132(5):1748–1775, 2024. 1, 4
- [30] Jinyuan Liu, Guanyao Wu, Zhu Liu, Di Wang, Zhiying Jiang, Long Ma, Wei Zhong, and Xin Fan. Infrared and visible image fusion: From data compatibility to task adaption. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1
- [31] Qing-Ming Liu, Rui-Sheng Jia, Yan-Bo Liu, Hai-Bin Sun, Jian-Zhi Yu, and Hong-Mei Sun. Infrared image super-resolution reconstruction by using generative adversarial network with an attention mechanism. *Applied Intelligence*, 51: 2018–2030, 2021. 1
- [32] Risheng Liu, Xin Fan, Minjun Hou, Zhiying Jiang, Zhongxuan Luo, and Lei Zhang. Learning aggregated transmission propagation networks for haze removal and beyond. *IEEE transactions on neural networks and learning systems*, 30(10):2973–2986, 2018. 1
- [33] Risheng Liu, Shichao Cheng, Yi He, Xin Fan, Zhouchen Lin, and Zhongxuan Luo. On the convergence of learning-based iterative methods for nonconvex inverse problems. *IEEE transactions on pattern analysis and machine intelligence*, 42(12):3027–3039, 2019. 3
- [34] Risheng Liu, Jinyuan Liu, Zhiying Jiang, Xin Fan, and Zhongxuan Luo. A bilevel integrated model with data-driven layer ensemble for multi-modality image fusion. *IEEE transactions on image processing*, 30:1261–1274, 2020. 1
- [35] Risheng Liu, Zhu Liu, Jinyuan Liu, and Xin Fan. Searching a hierarchically aggregated fusion architecture for fast multi-modality image fusion. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 1600–1608, 2021. 1
- [36] Risheng Liu, Zhiying Jiang, Shuzhou Yang, and Xin Fan. Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing*, 31:4922–4936, 2022. 1
- [37] Risheng Liu, Zhu Liu, Pan Mu, Xin Fan, and Zhongxuan Luo. Optimization-inspired learning with architecture augmentations and control mechanisms for low-level vision. *IEEE Transactions on Image Processing*, 32:6075–6089, 2023. 3
- [38] Risheng Liu, Zhu Liu, Jinyuan Liu, Xin Fan, and Zhongxuan Luo. A task-guided, implicitly-searched and metainitialized deep model for image fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1
- [39] Risheng Liu, Zhu Liu, Wei Yao, Shangzhi Zeng, and Jin Zhang. Moreau envelope for nonconvex bi-level optimization: A single-loop and hessian-free solution strategy. *ICML*, 2024. 3
- [40] Wenyu Liu, Gaofeng Ren, Runsheng Yu, Shi Guo, Jianke Zhu, and Lei Zhang. Image-adaptive yolo for object detection in adverse weather conditions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1792–1800, 2022. 2
- [41] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7007–7016, 2019. 4
- [42] Zhu Liu, Jinyuan Liu, Guanyao Wu, Long Ma, Xin Fan, and Risheng Liu. Bi-level dynamic learning for jointly multi-modality image fusion and beyond. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 1240–1248, 2023. 1
- [43] Zhu Liu, Jinyuan Liu, Benhuang Zhang, Long Ma, Xin Fan, and Risheng Liu. Paif: Perception-aware infrared-visible image fusion for attack-tolerant semantic segmentation. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 3706–3714, 2023. 3
- [44] Zhu Liu, Jinyuan Liu, Guanyao Wu, Zihang Chen, Xin Fan, and Risheng Liu. Searching a compact architecture for robust multi-exposure image fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(7):6224–6237, 2024. 1
- [45] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5637–5646, 2022. 1
- [46] Beat Münch, Pavel Trtik, Federica Marone, and Marco Stampanoni. Stripe and ring artifact removal with combined wavelet–fourier filtering. *Optics express*, 17(10):8567–8591, 2009. 5
- [47] Byeonghyun Pak, Jaewon Lee, and Kyong Hwan Jin. B-spline texture coefficients estimator for screen content image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10062–10071, 2023. 6
- [48] Lingfei Song and Hua Huang. Fixed pattern noise removal based on a semi-calibration method. *IEEE TPAMI*, 45(10): 11842–11855, 2023. 1
- [49] Lingfei Song and Hua Huang. Simultaneous destriping and image denoising using a nonparametric model with the em algorithm. *IEEE TIP*, 32:1065–1077, 2023. 1, 5
- [50] Di Wang, Jinyuan Liu, Xin Fan, and Risheng Liu. Unsupervised misaligned infrared and visible image fusion via cross-

- modality image generation and registration. *arXiv preprint arXiv:2205.11876*, 2022. 1
- [51] Di Wang, Jinyuan Liu, Risheng Liu, and Xin Fan. An interactively reinforced paradigm for joint infrared-visible image fusion and saliency object detection. *Information Fusion*, 98: 101828, 2023. 1
- [52] Yadong Wang, Darui Jin, Junzhang Chen, and Xiangzhi Bai. Revelation of hidden 2d atmospheric turbulence strength fields from turbulence effects in infrared imaging. *Nature Computational Science*, 3(8):687–699, 2023. 1
- [53] Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, and Haibin Ling. U2fusion: A unified unsupervised image fusion network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):502–518, 2020. 1
- [54] Kai Xu, Yaohong Zhao, Fangzhou Li, and Wei Xiang. Single infrared image stripe removal via deep multi-scale dense connection convolutional neural network. *Infrared Physics & Technology*, 121:104008, 2022. 1
- [55] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. *arXiv preprint arXiv:2401.10891*, 2024. 8
- [56] Suorong Yang, Peijia Li, Xin Xiong, Furao Shen, and Jian Zhao. Adaaugment: A tuning-free and adaptive approach to enhance data augmentation. *arXiv preprint arXiv:2405.11467*, 2024. 4
- [57] Jie-En Yao, Li-Yuan Tsao, Yi-Chen Lo, Roy Tseng, Chia-Che Chang, and Chun-Yi Lee. Local implicit normalizing flow for arbitrary-scale image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1776–1785, 2023. 6