NEURAL INFORMATION PROCESSING SYSTEMS

Deno-IF: Unsupervised Noisy Visible and Infrared Image Fusion Method

Han Xu[1] Yuyang Li[1] Yunfei Deng[1] Jiayi Ma[2] Guangcan Liu[1]*

[1] School of Automation, Southeast University, Nanjing, China

[2] Electronic Information School, Wuhan University, Wuhan, China

SOUTHEAST UNIVERSITY

WUHAN UNIVERSITY

Introduction

- **Image Fusion:**

It aims to merge the complementary information of source images and generate a single informative fused image with better scene representation.

- **Background:**

In real-world scenarios, challenging environments and the inherent constraints of cost-effective multi-modal devices often degrade image quality with significant noise, posing a critical challenge for noise-robust, high-quality image fusion.

Visible (noisy) **Scene** **Infrared (noisy)**

Motivation

➢ Separate Denoising and Fusion:

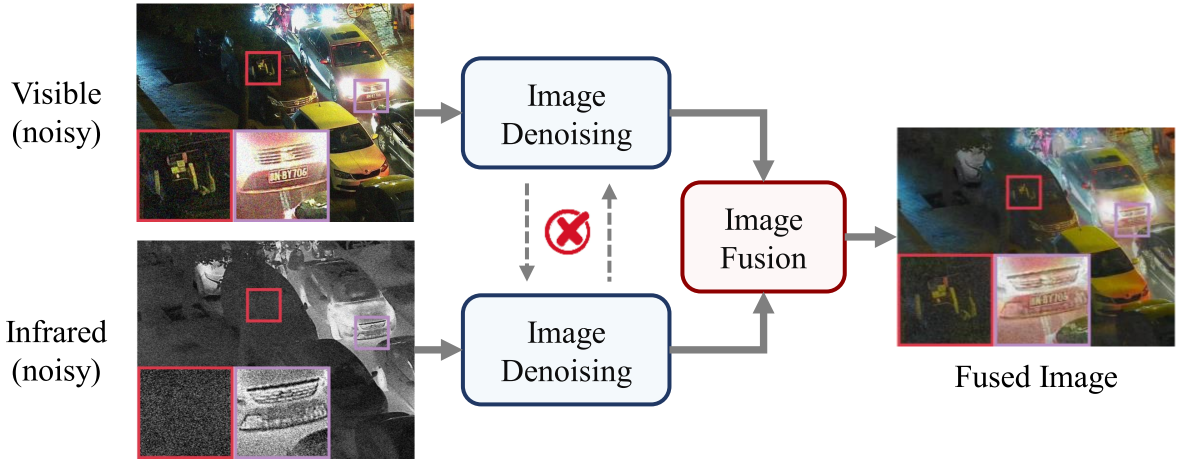
Most methods are tailored for standard scenarios and ineffective at suppressing noise. **Independent denoising-then-fusion approach:**

- Single-source denoising **ignores cross-modal** complementary information, **propagating residual** noise or artifacts.
- Disjoint framework is **computationally redundant**.

➢ Supervised Degradation-aware Fusion Methods:

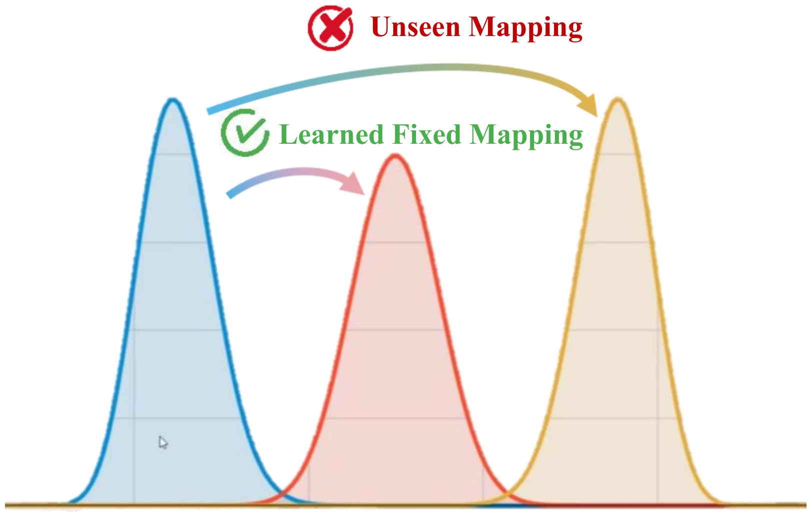
Existing degradations-aware fusion methods learn **fixed mappings** from noisy-clean source image pairs:

- Supervised approach **restricts** the denoising **efficacy and generalization**.



- Limited interaction
- Accumulated errors
- Computational Redundancy

Independent Two-Stage Approach



- Limited efficacy and generalization

Supervised Approach

Contributions

- **An unsupervised noisy visible and infrared image fusion method:**

  Without the supervision of clean data, it can still realize denoising and fusion simultaneously with fewer parameters, and is robust against various and variable noise conditions.

- **A convolutional low-rank optimization module:**

  As clean data exhibits convolutional low-rankness, we introduce the convolution nuclear norm minimization to decompose clean data from noisy inputs, providing optimization guidance for the network during training.

- **A joint denoising and fusion network:**

  It consists of intra-modal recovery and inter-modal recovery and fusion. It leverages self- and cross-modal attention to approximate guidance. A convolution matrix-based regularization loss further suppresses noise.

Method

Problem Formulation

A pair of observed noisy visible and infrared images $\{\mathbf{V}, \mathbf{R}\}$ can be decomposed as:

$$\mathbf{V} = \mathbf{L}_v + \mathbf{S}_v, \quad \mathbf{R} = \mathbf{L}_r + \mathbf{S}_r,$$

clean data

convolutional low-rankness

noise

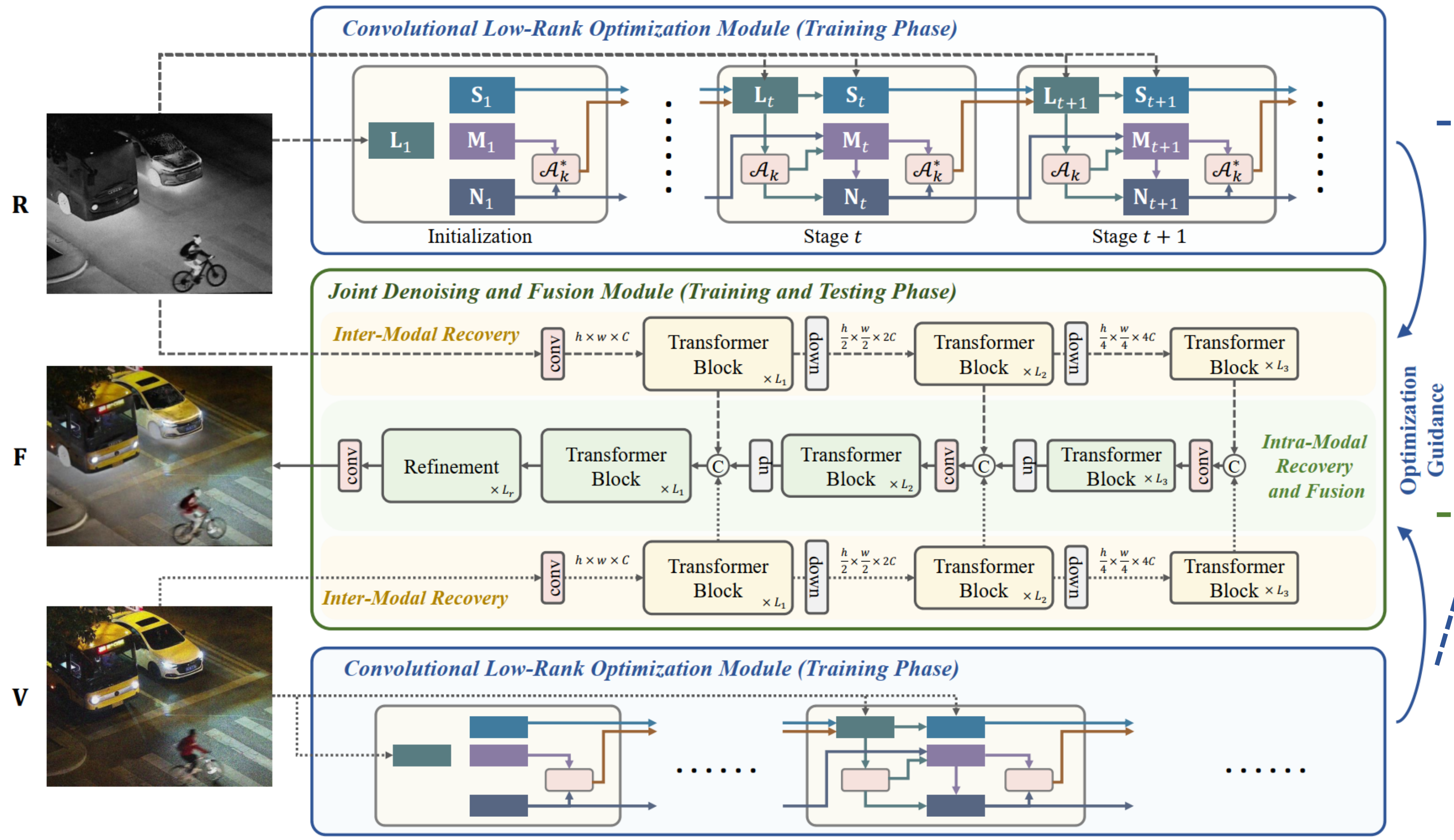
small-scale perturbations with limited overall energy and amplitude

Target: $\mathbf{V} \rightarrow$, $\mathbf{R} \rightarrow$ **joint denoising and fusion network** $\rightarrow$ high-quality fused image $\mathbf{F}$

contain vital information in the clean data $\mathbf{L}_v$, $\mathbf{L}_r$ mined from $\mathbf{V}$, $\mathbf{R}$, respectively

Method

Overall Framework



Convolutional Low-Rank Optimization Module

solve the ill-posed decomposition problem to provide precise guidance for network optimization

Joint Denoising and Fusion Module

optimized with the decomposition results; enable the joint restoration and fusion of complementary multi-modal information

Method

Convolutional Low-Rank Optimization Module

We take the decomposition of the noisy infrared image **R** as example and decomposition problem can be formulated as:

$$\min_{\mathbf{L},\mathbf{S}} \underbrace{\|\mathcal{A}_k(\mathbf{L})\|_*}_{\text{convolution nuclear norm of the clean data}} + \beta\underbrace{\|\mathbf{S}\|_F^2}_{\text{Forbenius norm of noise}}, \quad s.t. \;\; \mathbf{R} = \mathbf{L} + \mathbf{S}$$

These components can be obtained by minimizing the following energy function:

$$E(\mathbf{L},\mathbf{S}) = \|\mathbf{R}-\mathbf{L}-\mathbf{S}\|_F^2 + \alpha\|\mathcal{A}_k(\mathbf{L})\|_* + \beta\|\mathbf{S}\|_F^2$$

By releasing nuclear norm with $\|\mathbf{X}\|_* = \min_{\mathbf{A},\mathbf{B}} \frac{1}{2}\|\mathbf{A}\|_F^2 + \frac{1}{2}\|\mathbf{B}\|_F^2, \quad s.t. \;\; \mathbf{X} = \mathbf{AB},$ the energy function can be rewritten as:

$$E(\mathbf{L},\mathbf{S},\mathbf{M},\mathbf{N}) = \|\mathbf{R}-\mathbf{L}-\mathbf{S}\|_F^2 + \frac{\alpha}{2}\|\mathbf{M}\|_F^2 + \frac{\alpha}{2}\|\mathbf{N}\|_F^2 + \gamma\|\mathcal{A}_k(\mathbf{L}) - \mathbf{MN}\|_F^2 + \beta\|\mathbf{S}\|_F^2$$

Method

Convolutional Low-Rank Optimization Module

$$E(\mathbf{L},\mathbf{S},\mathbf{M},\mathbf{N}) = \|\mathbf{R}-\mathbf{L}-\mathbf{S}\|_F^2 + \frac{\alpha}{2}\|\mathbf{M}\|_F^2 + \frac{\alpha}{2}\|\mathbf{N}\|_F^2 + \gamma\|\mathcal{A}_k(\mathbf{L}) - \mathbf{MN}\|_F^2 + \beta\|\mathbf{S}\|_F^2$$

This problem can be solved by **iteratively addressing the subproblems** related to **L**, **S**, **M**, **N**.

With t denotes the iteration step, this problem can be partitioned into the following four **subproblems**:

During the iteration, the **close-form solutions** for these subproblems can be obtained as:

$$\mathbf{L}_t = \arg\min_{\mathbf{L}} \|\mathbf{R}-\mathbf{L}-\mathbf{S}_{t-1}\|_F^2 + \gamma\|\mathcal{A}_k(\mathbf{L}) - \mathbf{M}_{t-1}\mathbf{N}_{t-1}\|_F^2, \quad \rightarrow \quad \mathbf{L}_t = \frac{\mathbf{R}-\mathbf{S}_{t-1}+\gamma\mathcal{A}_k^*(\mathbf{M}_{t-1}\mathbf{N}_{t-1})}{\gamma+1},$$

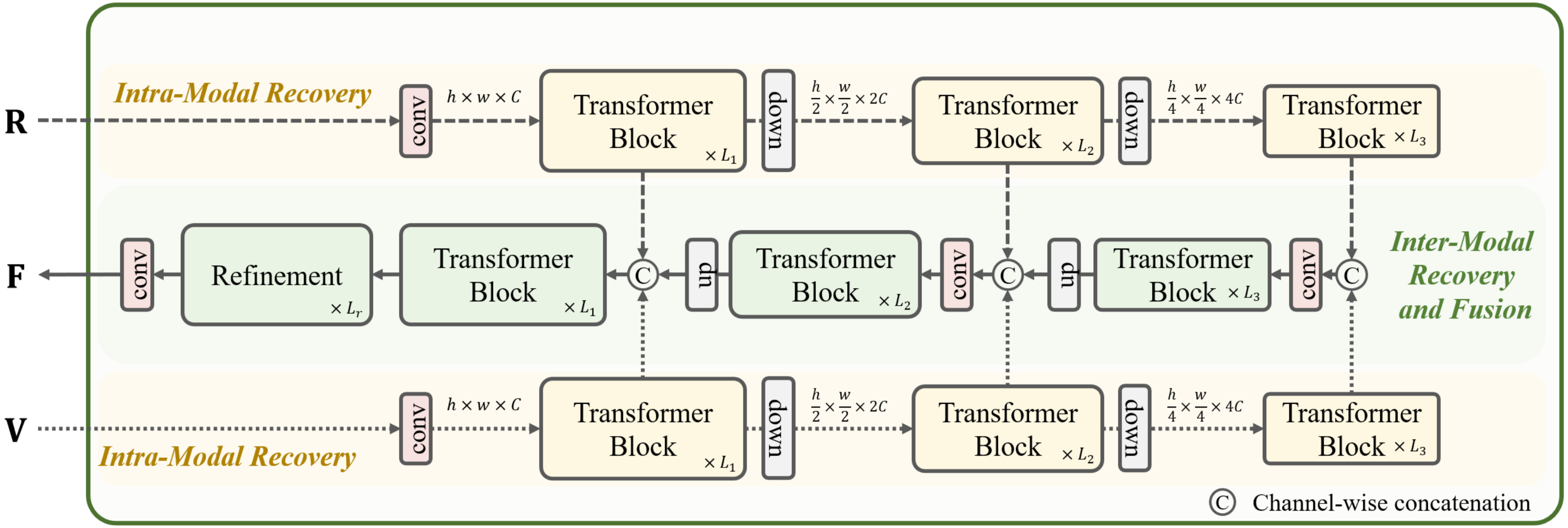
$$\mathbf{S}_t = \arg\min_{\mathbf{S}} \|\mathbf{R}-\mathbf{L}_t-\mathbf{S}\|_F^2 + \beta\|\mathbf{S}\|_F^2, \quad \rightarrow \quad \mathbf{S}_t = \frac{\mathbf{R}-\mathbf{L}_t}{\beta+1},$$

$$\mathbf{M}_t = \arg\min_{\mathbf{M}} \frac{\alpha}{2}\|\mathbf{M}\|_F^2 + \gamma\|\mathcal{A}_k(\mathbf{L}_t) - \mathbf{M}\mathbf{N}_{t-1}\|_F^2, \quad \rightarrow \quad \mathbf{M}_t = 2\gamma\mathcal{A}_k(\mathbf{L}_t)\mathbf{N}_{t-1}^\top(\alpha\mathbf{I}+2\gamma\mathbf{N}_{t-1}\mathbf{N}_{t-1}^\top)^{-1},$$

$$\mathbf{N}_t = \arg\min_{\mathbf{N}} \frac{\alpha}{2}\|\mathbf{N}\|_F^2 + \gamma\|\mathcal{A}_k(\mathbf{L}_t) - \mathbf{M}_t\mathbf{N}\|_F^2. \quad \rightarrow \quad \mathbf{N}_t = 2\gamma(\alpha\mathbf{I}+2\gamma\mathbf{M}_t^\top\mathbf{M}_t)^{-1}\mathbf{M}_t^\top\mathcal{A}_k(\mathbf{L}_t).$$

Method

Joint Denoising and Fusion Module



- **Intensity Loss:**

$$\mathcal{L}_{in} = \|\mathbf{F}^y - \max(\mathbf{L}_v^y, \mathbf{L}_r)\|_1$$

- **Gradient Loss:**

$$\mathcal{L}_g = \|\nabla \mathbf{F}^y - [\mathbf{m}_g \cdot \nabla \mathbf{L}_v^y + (1-\mathbf{m}_g) \cdot \nabla \mathbf{L}_r]\|_1$$

- **Chrominance Loss:**

$$\mathcal{L}_{chr} = \|\mathbf{F}^{Cb} - \mathbf{L}_v^{Cb}\|_1 + \|\mathbf{F}^{Cr} - \mathbf{L}_v^{Cr}\|_1$$

- **Convolutional Low Rank Regularization Loss:**

$$\mathcal{L}_{rank} = \|\mathcal{A}_k(\mathbf{F})\|_*$$

Experiments and Results

- **Datasets:** LLVIP dataset, M3FD dataset
- **Patch size:** 128×128
- **Convolutional Low-Rank Optimization Module:**

  $k_1, k_2 = 12$, $k_3 = 2$, $m, n = 256$. Iteration $T = 30$.

  $\alpha = 200\mathbb{E}[|\nabla(x) - \nabla G(x)|]$, $\beta = 2$, $\gamma = 80\mathbb{E}[|\nabla(x) - \nabla G(x)|]$
- **Joint Denoising and Fusion Module:**

  Hyper-parameters: $\lambda = 1e3$, $\eta = 30$, κ equals during the training phase

  Optimizer: Adam Optimizer

  Learning Rate: 2e-4 with exponential decay

  Number of Blocks: $L_1, L_2 = 4$, $L_3, L_r = 2$

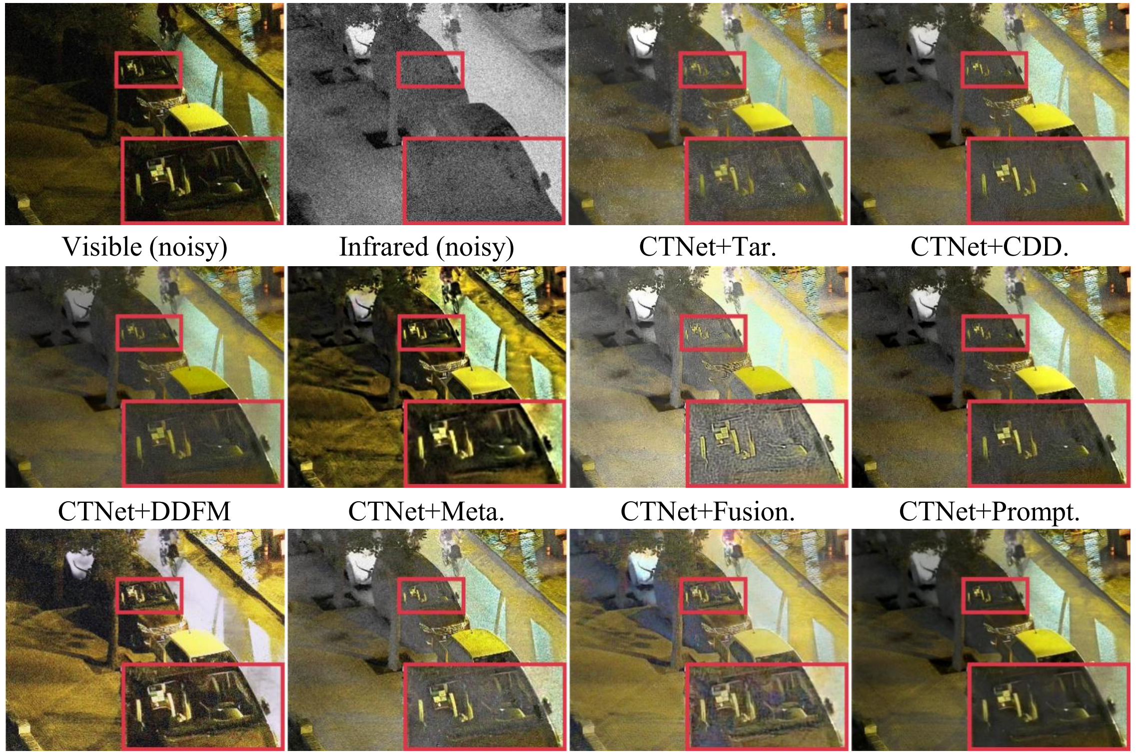
  Platform: NVIDIA 3090 GPU

Experiments and Results

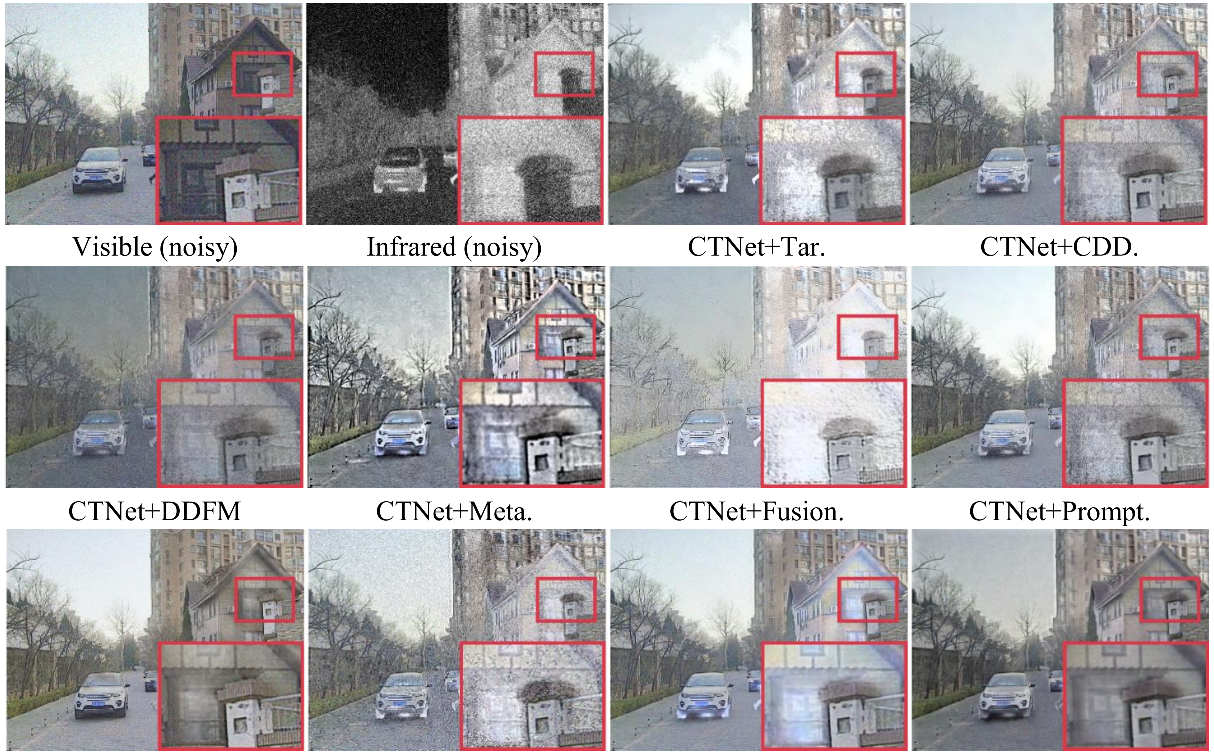
Qualitative Comparison

1. **Joint** denoising and fusion **avoids residual noise** in **denoising-then-fusion approaches** caused by limited pre-denoising performance.
2. Infer clean data from noisy source images in **an unsupervised manner**, show **generalization** across various noise types and levels.
3. The inference on clean data is based on **convolutional low-rankness**, **avoiding excessive distortion** of meaningful contents.

Gaussian noise — LLVIP dataset



Gaussian noise — M3FD dataset

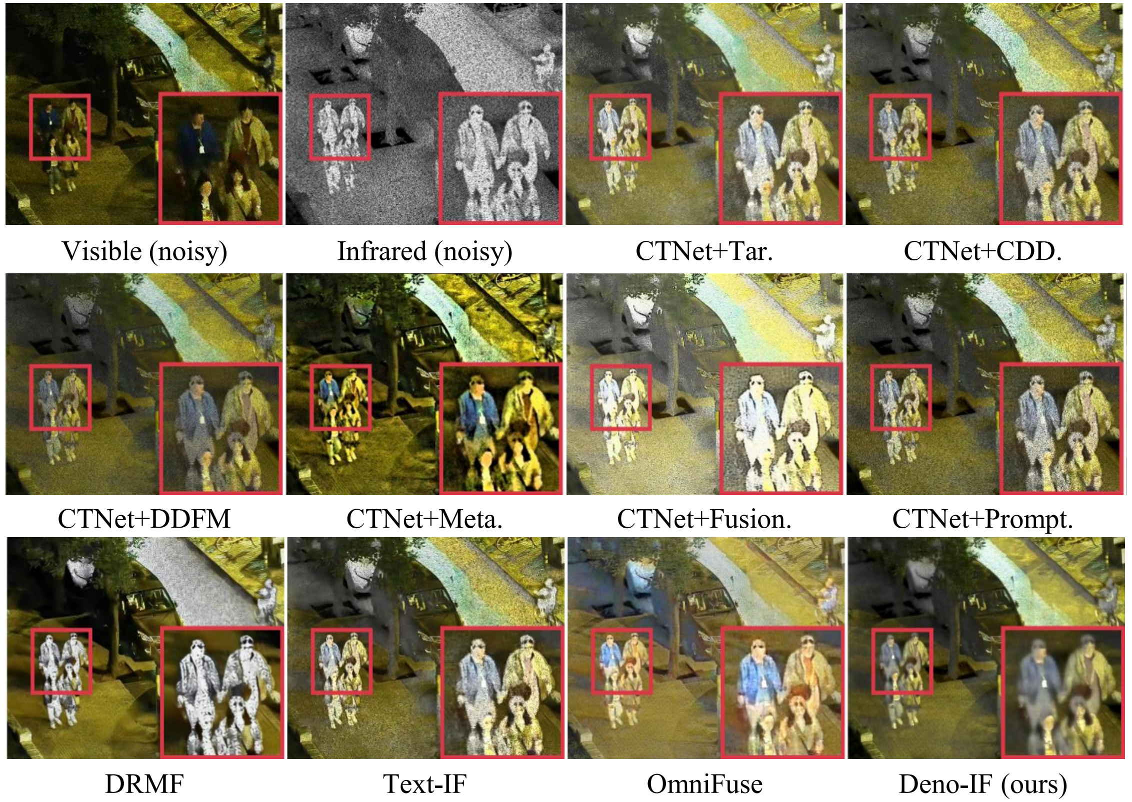


Experiments and Results

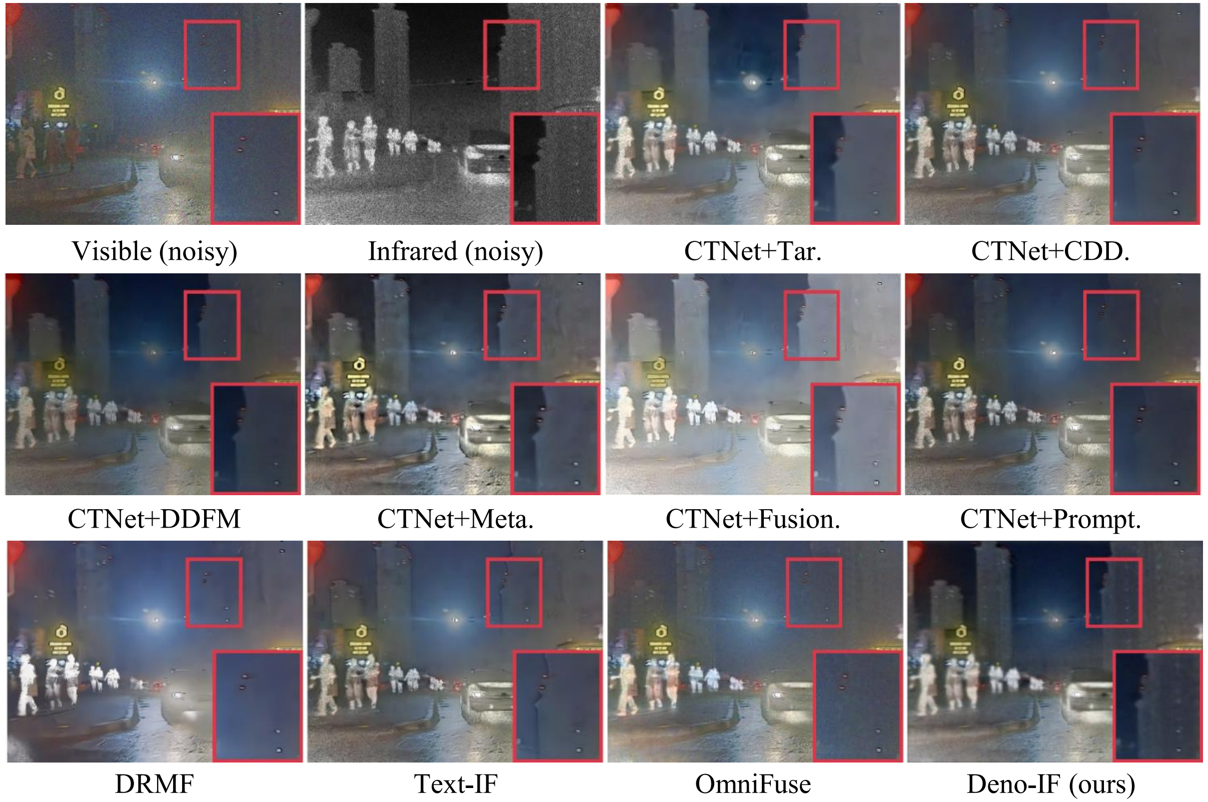
Qualitative Comparison

1. **Joint** denoising and fusion **avoids residual noise** in **denoising-then-fusion approaches** caused by limited pre-denoising performance.
2. Infer clean data from noisy source images in **an unsupervised manner**, show **generalization** across various noise types and levels.
3. The inference on clean data is based on **convolutional low-rankness**, **avoiding excessive distortion** of meaningful contents.

Speckle noise — LLVIP dataset



Speckle noise — M3FD dataset



Experiments and Results

Quantitative Comparison

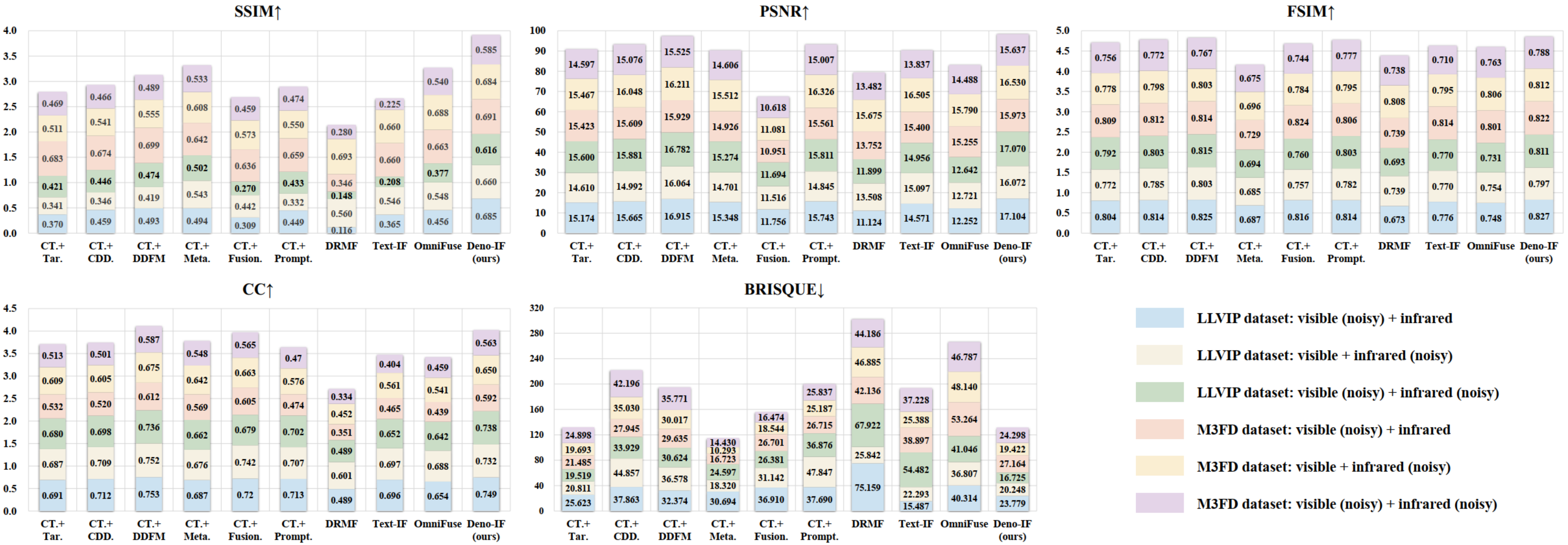
- Similarity-based **full-reference** metrics: SSIM, PSNR, feature similarity index (FSIM) and correlation coefficient (CC)
- Quality-based **no-reference** metric: BRISQUE

Gaussian	LLVIP Dataset					M3FD Dataset				
Metrics	**SSIM↑**	**PSNR↑**	**FSIM↑**	**CC↑**	**BRISQUE↓**	**SSIM↑**	**PSNR↑**	**FSIM↑**	**CC↑**	**BRISQUE↓**
CT.+Tar.	0.421±0.169	15.600±1.464	0.792±0.024	0.680±0.082	19.519±8.577	0.469±0.143	14.597±2.781	0.756±0.032	0.513±0.195	24.898±9.713
CT.+CDD.	0.446±0.206	15.881±1.366	0.803±0.028	0.698±0.072	33.929±20.019	0.466±0.140	15.076±2.789	0.772±0.031	0.501±0.226	42.196±14.175
CT.+DDFM	0.474±0.178	16.782±1.138	**0.815±0.022**	0.736±0.077	30.624±16.596	0.489±0.165	15.525±2.513	0.767±0.051	**0.587±0.191**	35.771±13.177
CT.+Meta.	0.502±0.099	15.274±1.193	0.694±0.045	0.662±0.083	24.597±12.281	0.533±0.132	14.606±2.650	0.675±0.053	0.548±0.194	**14.430±5.869**
CT.+Fusion.	0.270±0.135	11.694±0.624	0.760±0.042	0.679±0.078	26.381±11.446	0.459±0.094	10.618±1.040	0.744±0.047	0.565±0.158	16.474±11.822
CT.+Prompt.	0.433±0.197	15.811±1.331	0.803±0.029	0.702±0.073	36.876±19.142	0.474±0.125	15.007±2.679	0.777±0.033	0.470±0.243	25.837±13.808
DRMF	0.148±0.081	11.899±1.242	0.693±0.031	0.489±0.156	67.922±12.462	0.280±0.121	13.482±1.628	0.738±0.043	0.334±0.219	44.186±11.117
Text-IF	0.208±0.069	14.956±1.125	0.770±0.025	0.652±0.071	54.482±10.614	0.225±0.089	13.837±2.509	0.710±0.055	0.404±0.209	37.228±8.241
OmniFuse	0.377±0.080	12.642±0.727	0.731±0.026	0.642±0.089	41.046±3.748	0.540±0.098	14.488±2.119	0.763±0.031	0.459±0.211	46.787±6.876
Deno-IF	**0.616±0.064**	**17.070±1.426**	0.811±0.023	**0.738±0.074**	**16.725±4.091**	**0.585±0.081**	**15.637±2.899**	**0.788±0.029**	0.563±0.205	24.298±7.095

Speckle	LLVIP Dataset					M3FD Dataset				
Metrics	**SSIM↑**	**PSNR↑**	**FSIM↑**	**CC↑**	**BRISQUE↓**	**SSIM↑**	**PSNR↑**	**FSIM↑**	**CC↑**	**BRISQUE↓**
CT.+Tar.	0.423±0.073	15.125±1.464	0.776±0.024	0.687±0.072	13.375±9.263	0.553±0.129	15.358±2.393	0.775±0.039	0.553±0.206	13.950±9.283
CT.+CDD.	0.419±0.096	15.527±1.515	0.789±0.026	0.681±0.063	45.497±13.496	0.539±0.145	15.906±2.520	0.789±0.041	0.558±0.215	25.231±14.022
CT.+DDFM	0.474±0.086	16.191±1.218	**0.808±0.022**	0.723±0.069	32.897±12.572	0.584±0.125	16.232±2.259	0.806±0.047	0.641±0.157	18.320±11.767
CT.+Meta.	0.538±0.072	15.102±1.220	0.697±0.043	0.669±0.075	10.860±9.531	0.615±0.115	15.606±2.526	0.710±0.046	0.611±0.164	**10.412±6.587**
CT.+Fusion.	0.338±0.055	11.909±0.694	0.737±0.032	0.683±0.066	29.446±7.250	0.516±0.106	10.950±0.807	0.774±0.034	0.612±0.150	21.178±9.460
CT.+Prompt.	0.381±0.085	15.212±1.313	0.787±0.026	0.681±0.066	48.286±12.442	0.522±0.146	15.881±2.447	0.784±0.041	0.521±0.244	21.837±12.825
DRMF	0.475±0.047	12.417±1.377	0.722±0.021	0.567±0.137	21.406±10.566	0.649±0.100	14.147±2.561	0.790±0.026	0.423±0.246	43.180±17.155
Text-IF	0.400±0.054	13.953±1.119	0.759±0.019	0.665±0.060	27.072±9.667	0.554±0.128	15.941±2.402	0.780±0.037	0.496±0.235	18.908±10.381
OmniFuse	0.479±0.052	12.859±0.775	0.741±0.017	0.665±0.074	39.696±4.430	0.585±0.104	15.221±2.089	0.767±0.033	0.523±0.213	44.850±7.131
Deno-IF	**0.633±0.052**	**16.528±1.370**	0.805±0.022	**0.738±0.065**	**9.280±5.918**	**0.656±0.075**	**16.525±2.800**	**0.813±0.026**	**0.648±0.155**	16.514±8.460

Experiments and Results

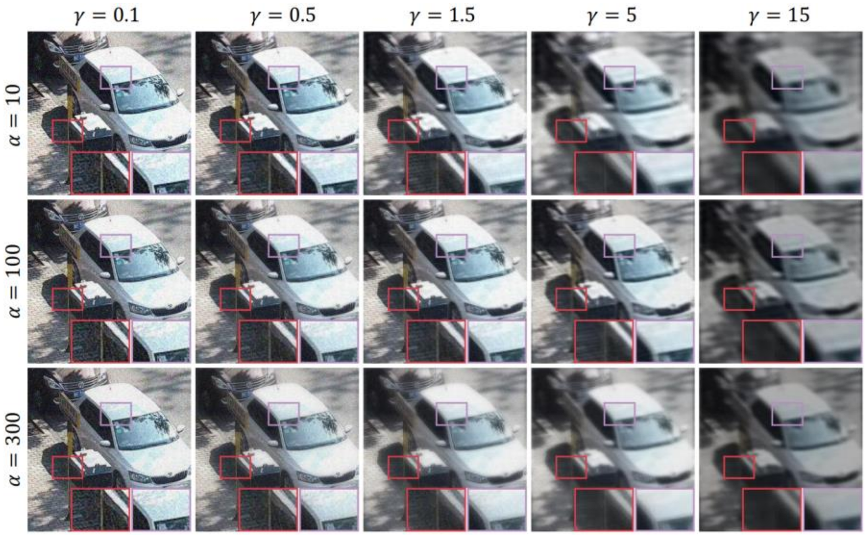
Quantitative Comparison



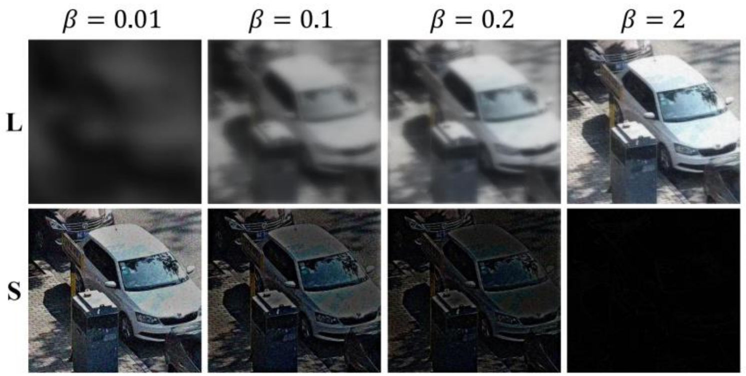
Experiments and Results

Parameter Analysis, Ablation Study, and Different-Level Performance

Parameter Analysis

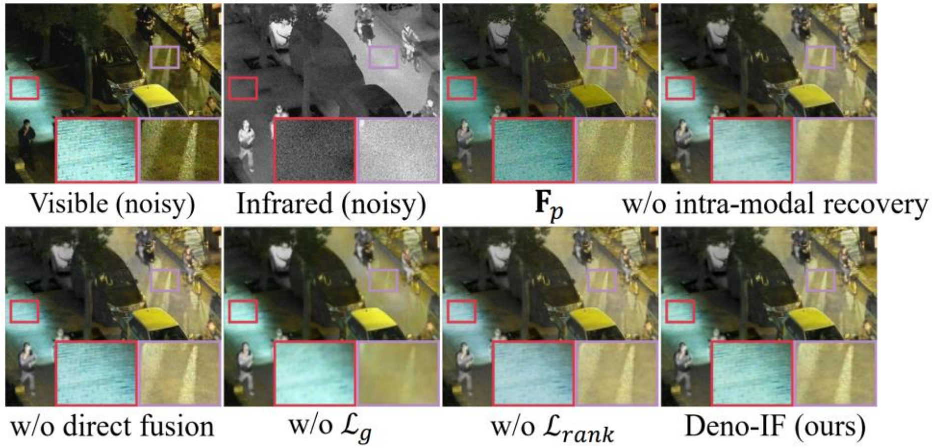


(a) Impact of α, γ on the estimation of clean component $\mathbf{L}$



(b) Impact of β on decomposition when $\alpha = 100, \gamma = 1.5$

Ablation Study



Metrics	SSIM↑	PSNR↑	FSIM↑	CC↑
w/o intra-modal	0.606	16.960	0.799	0.730
w/o direct fusion	0.613	16.965	0.797	0.730
w/o $\mathcal{L}_g$	0.600	16.160	0.789	0.685
w/o $\mathcal{L}_{rank}$	0.611	16.554	0.802	0.727
Deno-IF (ours)	**0.616**	**17.070**	**0.811**	**0.738**

Results on Dealing with Different-Level Noise

	PSNR↑				SSIM↑				FSIM↑				CC↑				BRISQUE↓			
σ	DRMF	Text.	Omn.	Deno.	DRMF	Text.	Omn.	Deno.	DRMF	Text.	Omn.	Deno.	DRMF	Text.	Omn.	Deno.	DRMF	Text.	Omn.	Deno.
10	11.997	16.141	12.522	16.878	0.309	0.431	0.484	0.675	0.716	0.814	0.747	0.818	0.557	0.702	0.667	0.740	26.855	33.060	41.297	15.364
20	11.874	15.736	12.551	17.097	0.182	0.239	0.427	0.650	0.705	0.793	0.738	0.816	0.528	0.681	0.657	0.738	58.717	45.843	41.610	16.493
30	11.949	14.433	12.701	17.199	0.170	0.148	0.384	0.632	0.700	0.737	0.732	0.813	0.510	0.629	0.648	0.735	66.016	58.087	40.411	17.989
40	11.935	14.182	12.803	17.262	0.138	0.109	0.350	0.617	0.688	0.725	0.728	0.809	0.467	0.604	0.636	0.733	74.548	61.441	38.531	20.514
50	11.719	13.477	12.848	17.242	0.106	0.084	0.318	0.602	0.675	0.693	0.722	0.805	0.423	0.562	0.624	0.732	75.814	66.258	35.604	24.256

Thanks for watching !

Deno-IF: Unsupervised Noisy Visible and Infrared Image Fusion Method

Han Xu[1] Yuyang Li[1] Yunfei Deng[1] Jiayi Ma[2] Guangcan Liu[1*]

[1] School of Automation, Southeast University, Nanjing, China [2] Electronic Information School, Wuhan University, Wuhan, China

Email: xu_han@seu.edu.cn