Task-Gated Multi-Expert Collaboration Network for Degraded Multi-Modal Image Fusion

Yiming Sun[1], Xin Li[1], Pengfei Zhu[1], Qinghua Hu[2], Dongwei Ren[2], Huiying Xu[3], Xinzhong Zhu[3]

1. Southeast University;

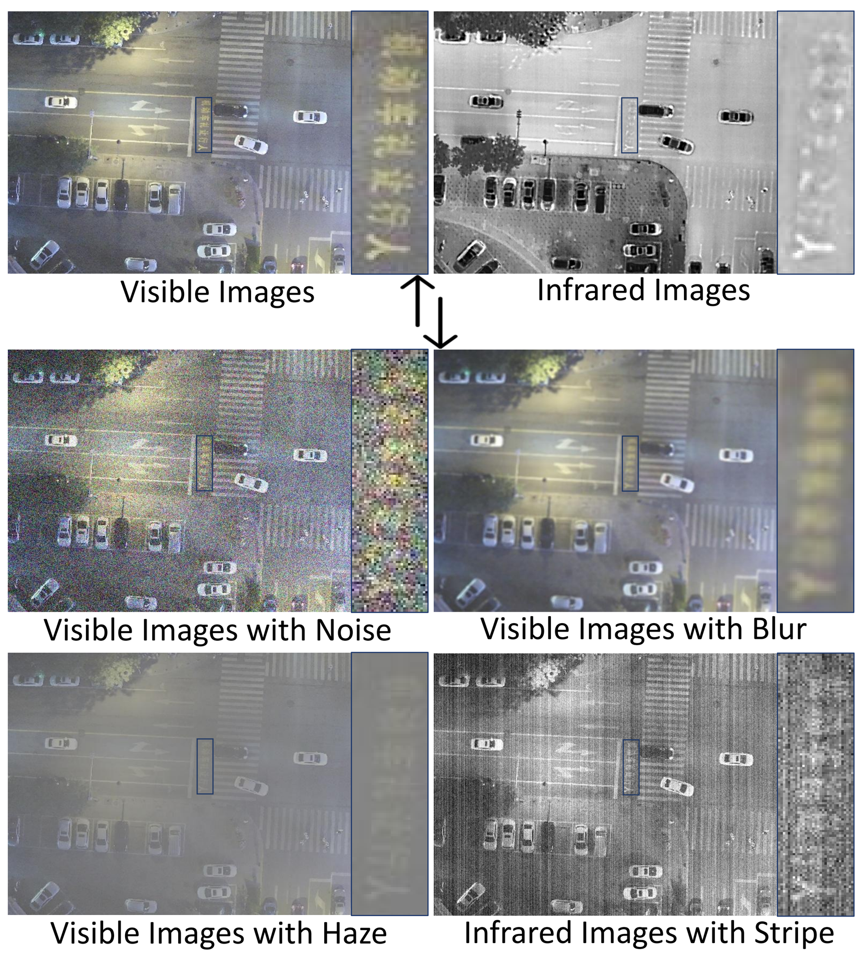
2. Tianjin University; Research Center of City Intelligence and Digital Governance, Haihe Lab of ITAI; 3. Zhejiang Normal University

Motivation

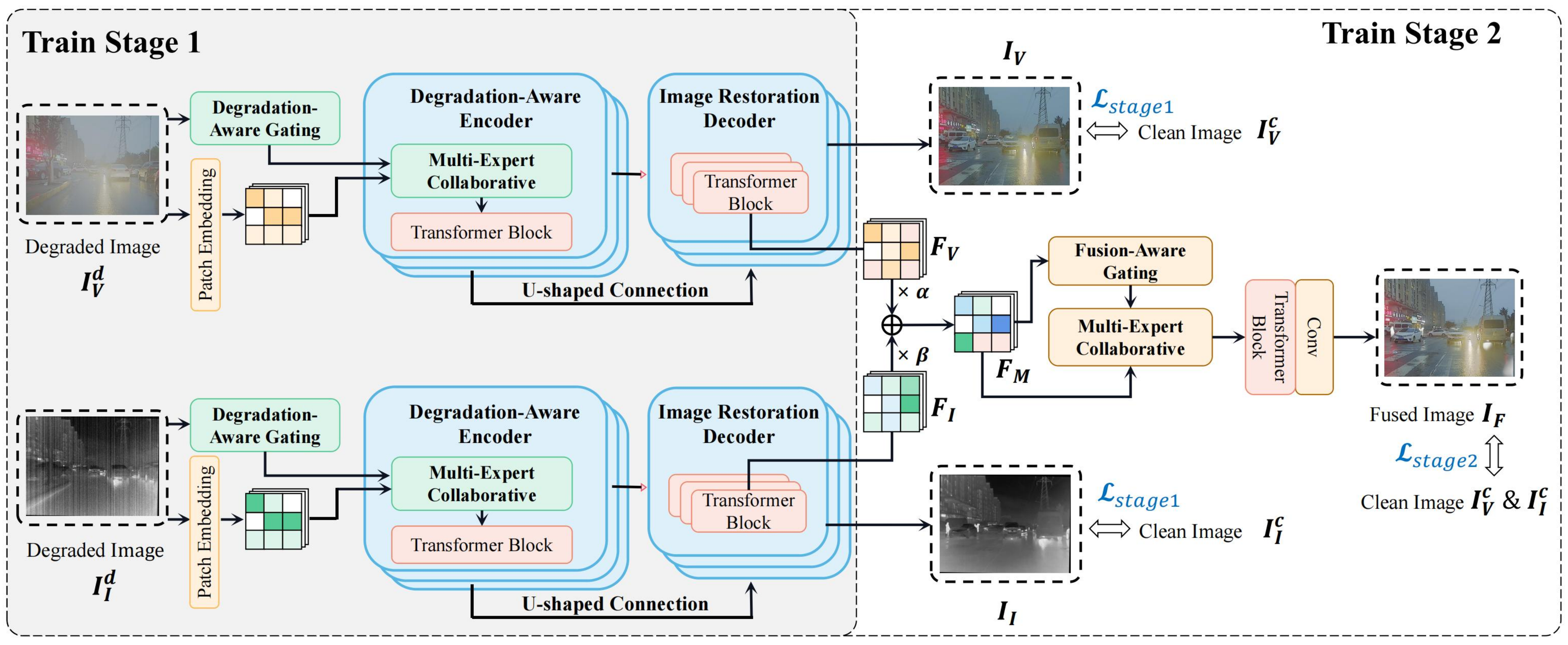
Infrared and visible image fusion aims to utilize the complementary information between the two modalities to synthesize a new image containing richer information.

Real-world imaging often suffers from degradation issues, such as noise, blur, and haze in visible imaging, as well as stripe noise in infrared imaging, which significantly degrades fusion performance.

Multi-modal image fusion models should possess robust adaptive capabilities to handle variations in different degradations to prevent fusion results from lacking crucial elements.



Framework



- **Output of Stage 1:**

$$Output = \sum_{i=1}^{N} G_{Degrad}(I_V^d)_i \cdot E_i^{Degrad}(I_V^d).$$

- **Restoration Loss:**

$$\mathcal{L}_{res} = \|I_V - I_V^c\| + \|I_I - I_I^c\|.$$

- **Gradiant Loss:**

$$\mathcal{L}_{grad} = \|sobel(I_V) - sobel(I_V^c)\| + \|sobel(I_I) - sobel(I_I^c)\|.$$

- **Stage 1 Loss:**

$$\mathcal{L}_{stage_1} = \mathcal{L}_{balance} + \mathcal{L}_{res} + \mathcal{L}_{grad}$$

- **Output of Stage 2:**

$$Output_{Fus} = \sum_{i=1}^{N} G_{Fus}(F_M)_i \cdot E_i^{Fus}(F_M).$$

- **Pixel Loss:**

$$\mathcal{L}_{Pixel} = \|I_F - I_V^c\| + \|I_F - I_I^c\|.$$

- **Gradiant Loss and intensity Loss:**

$$\mathrm{L}_{in} + \mathrm{L}_{grad} = \|I_F - max(I_V^c, I_I^c)\| + \|sobel(I_F) - sobel(I_V^c)\| + \|sobel(I_F) - sobel(I_I^c)\|.$$

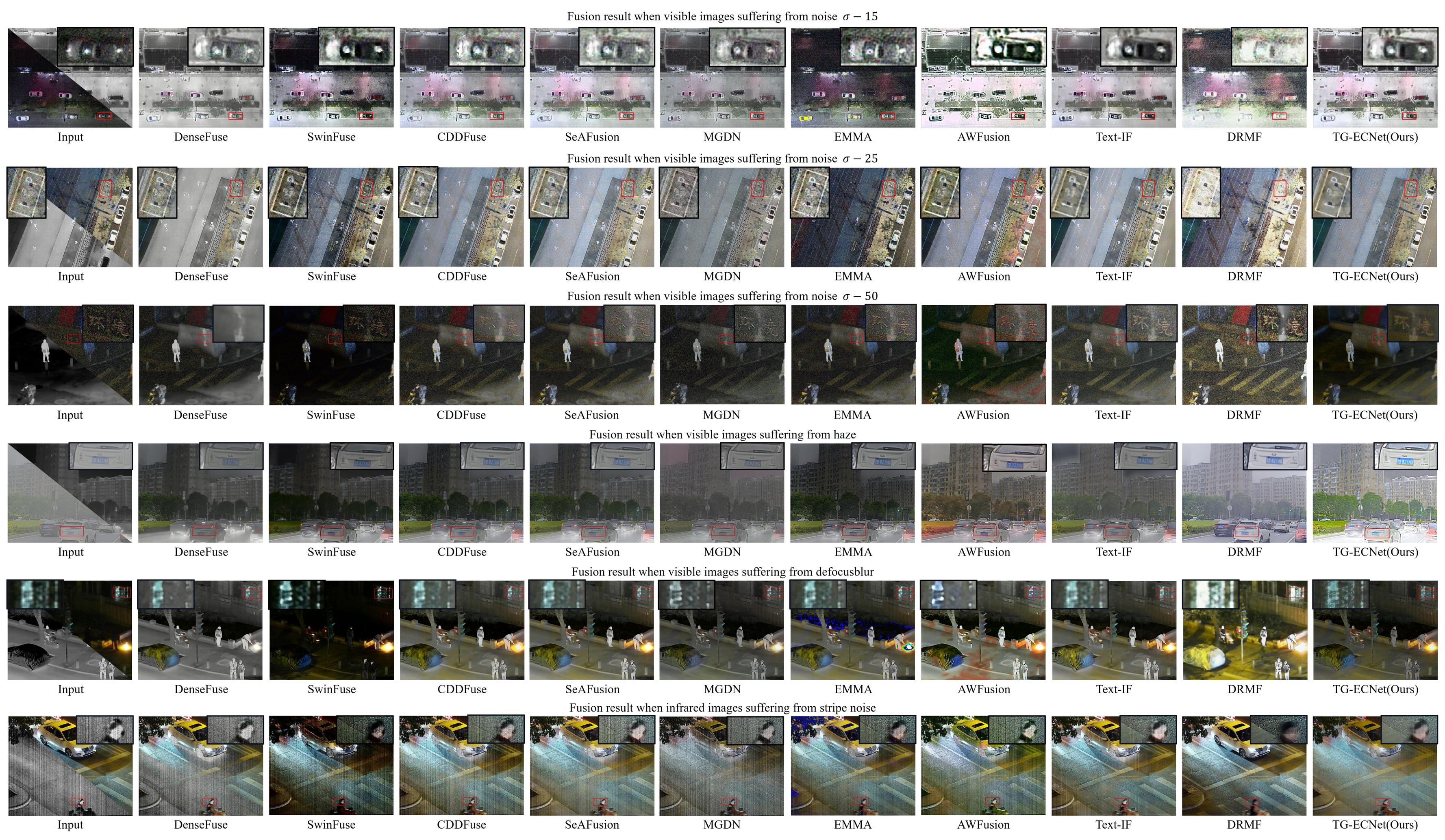
- **Stage 2 Loss:**

$$\mathcal{L}_{Stage_2} = \mathcal{L}_{Pixel} + \mathcal{L}_{balance} + \mathcal{L}_{grad} + \mathcal{L}_{in}$$

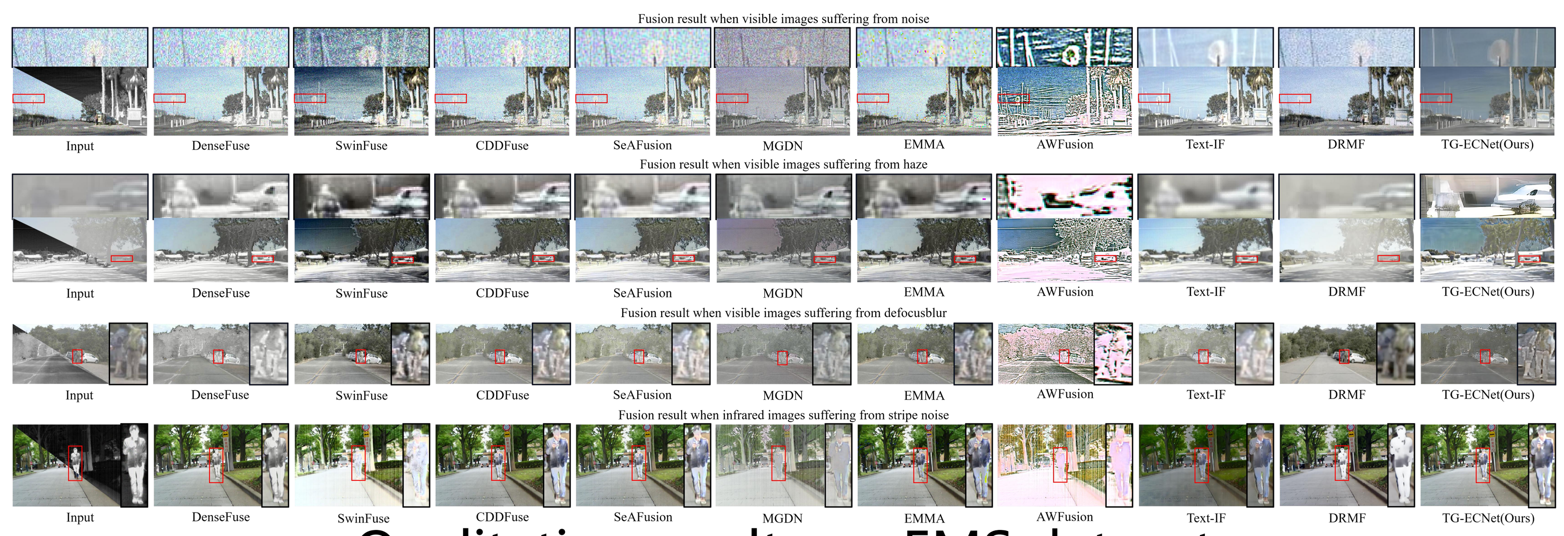
Contributions

(1) We propose a unified framework for degraded multi-modal image restoration and fusion, which bridges different tasks together through a two-stage training strategy to learn inter-task information while avoiding mutual interference, enabling all-in-one processing.
(2) We propose the task-aware gating and multi-expert collaboration module. The degradation-aware gating adapts to different degradation types and selects the optimal expert group for image restoration, while the fusion-aware gating dynamically balances the information retention between fusion and restoration tasks to achieve better fusion performance.
(3) We construct a large-scale degraded multi-modal image fusion benchmark, DeMMI-RF, which contains more than 30,000 multi-modal data of different degradation types, including those from UAVs and driving viewpoints. Results on multiple datasets validate the superior performance of the model in complex degraded scenarios and robustness for downstream applications.

Experimental Results



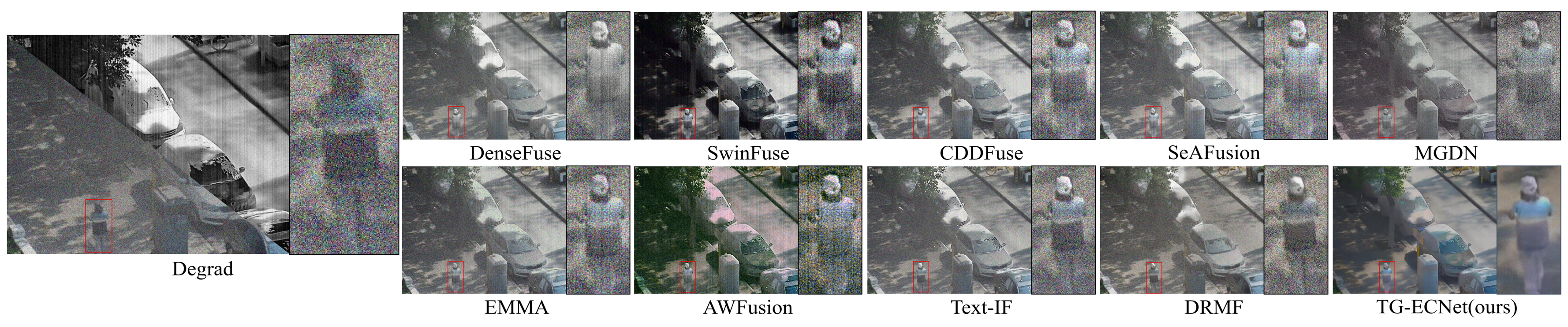
Qualitative results on DeMMI-RF datasets



Qualitative results on EMS dataset

Comparison of average quantitative performance of single task

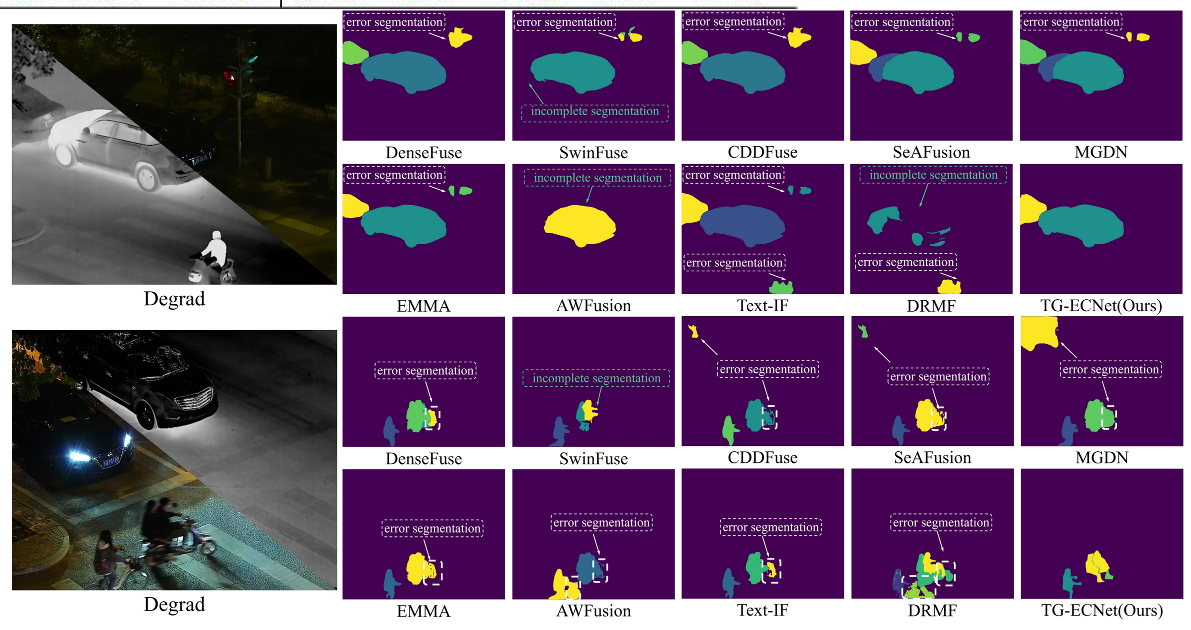
Methods	Our Dataset					EMS Dataset				
	CC	MSE	PSNR	N_{abf}	MS-SSIM	CC	MSE	PSNR	N_{abf}	MS-SSIM
DenseFuse (Li & Wu, 2018)	0.5185	0.0923	29.5794	0.0863	0.2291	0.5018	0.1090	28.9558	0.0715	0.3313
SwinFuse (Wang et al., 2022b)	0.5279	0.0928	29.5190	0.1180	0.2418	0.5002	0.1717	28.1347	0.1176	0.1396
CDDFuse (Zhao et al., 2023a)	0.5286	0.0787	29.8441	0.1116	0.2359	0.5005	0.1120	28.8962	0.0933	0.3197
SeAFusion (Tang et al., 2022a)	0.5288	0.0904	29.5892	0.1170	0.2265	0.5013	0.1115	28.9057	0.0874	0.3187
MGDN (Guan et al., 2023)	0.5279	0.0669	30.2216	0.1195	0.2431	0.4985	0.1314	28.7083	0.0924	0.1763
EMMA (Zhao et al., 2024)	0.5198	0.0821	29.7471	0.1266	0.2384	0.4996	0.1132	28.8699	0.0966	0.3057
AWFusion (Li et al., 2024b)	0.5265	0.0979	29.4554	0.1429	0.2073	0.5013	0.2221	27.5428	0.2796	0.1114
Text-IF (Yi et al., 2024)	0.5309	0.0880	29.5656	0.0804	0.2379	0.5007	0.1115	28.9034	0.0800	0.3210
DRMF (Tang et al., 2024)	0.5249	0.0883	29.5077	0.1203	0.2179	0.4991	0.1261	28.6382	0.0739	0.3158
TG-ECNet(Ours)	0.5340	0.0570	30.5822	0.0385	0.2972	0.5035	0.0738	29.7962	0.0405	0.4273



Qualitative comparisons on multi degradations

Quantitative comparisons on multi degradations

Methods	multi-tasks(noise+haze+defocusblur+stripe)				
	CC	MSE	PSNR	N_{abf}	MS-SSIM
DenseFuse [4]	0.5225	0.0950	29.2285	0.1240	0.1740
SwinFuse [33]	0.5075	0.0940	29.4485	0.1010	0.2540
CDDFuse [3]	0.5225	0.0880	29.4665	0.1330	0.1790
SeAFusion [6]	0.5220	0.0940	29.2615	0.1240	0.1640
MGDN [29]	0.5200	0.0790	29.6820	0.1570	0.1750
EMMA [5]	0.5195	0.0820	29.5760	0.1580	0.1600
AWFusion [31]	0.5190	0.0990	29.1365	0.1970	0.1360
Text-IF [2]	0.5205	0.0970	29.1655	0.1640	0.1620
DRMF [1]	0.4960	0.0900	29.3640	0.2390	0.1560
TG-ECNet(Ours)	0.5245	0.0630	30.2200	0.0110	0.2870



The segmentation resuls of Grounded-SAM

Acknowledgements

This work was supported in part by the National Science and Technology Major Project under Grant 2022ZD0116500, in part by the National Natural Science Foundation of China under Grants 62436002, 62476195, 61925602, U23B2049, and 62222608, in part by Tianjin Natural Science Funds for Distinguished Young Scholar under Grant 23JCJQJC00270, and in part by Zhejiang Provincial Natural Science Founda tion of China under Grant LD24F020004.

Contact Information

Homepage

Dr. Yiming Sun
Southeast University
Email: sunyiming@seu.edu.cn

WeChat

West Exhibition Hall B2-B3 #W-205