HDCFN: Haze Distribution-aware Cross-modal Fusion Network for Infrared-guided Dense Haze Removal in UAVs

China Telecom

Cloud Computing Research Institute

Junwei Zhao, Qianchun Luo, Shiliang Zhang, Shen Gao, and Jie Wu*

ACM MM 2025

ACM multimedia

Dublin, Ireland 27-31.10.2025

Motivation

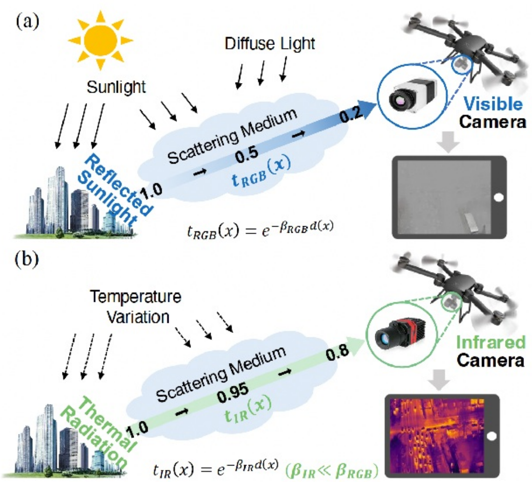
- Visible-only dehazing misses structure; infrared is haze-resistant but lacks color/texture fidelity.
- In UAV vision, small, fully obscured objects are difficult to infer from surrounding cues.

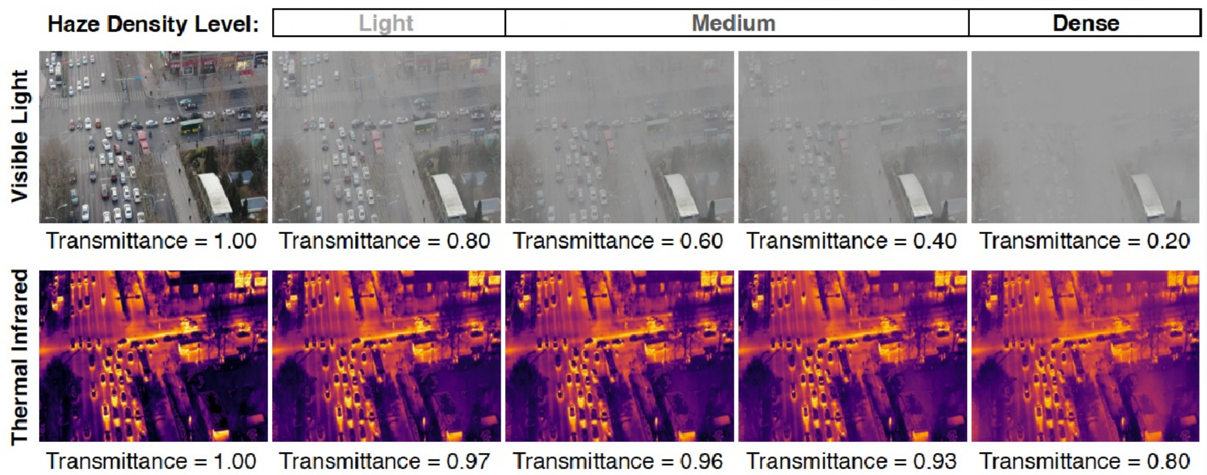
Physical Background

- **Visible light** preserves color/texture; **thermal Infrared** preserves more structure cues.
- **HDCFN** combines both for robust dehazing.

- **Visible Light imaging:** $I_{RGB}(x) = J(x) \cdot \exp(-\beta_{RGB} \cdot d(x)) + A \cdot (1-\exp(-\beta_{RGB} \cdot d(x)))$,
- **Thermal IR imaging:** $I_{IR}(x) = I_{IR}(x) \cdot \exp(-\beta_{IR} \cdot d(x)) + A \cdot (1-\exp(-\beta_{IR} \cdot d(x)))$.

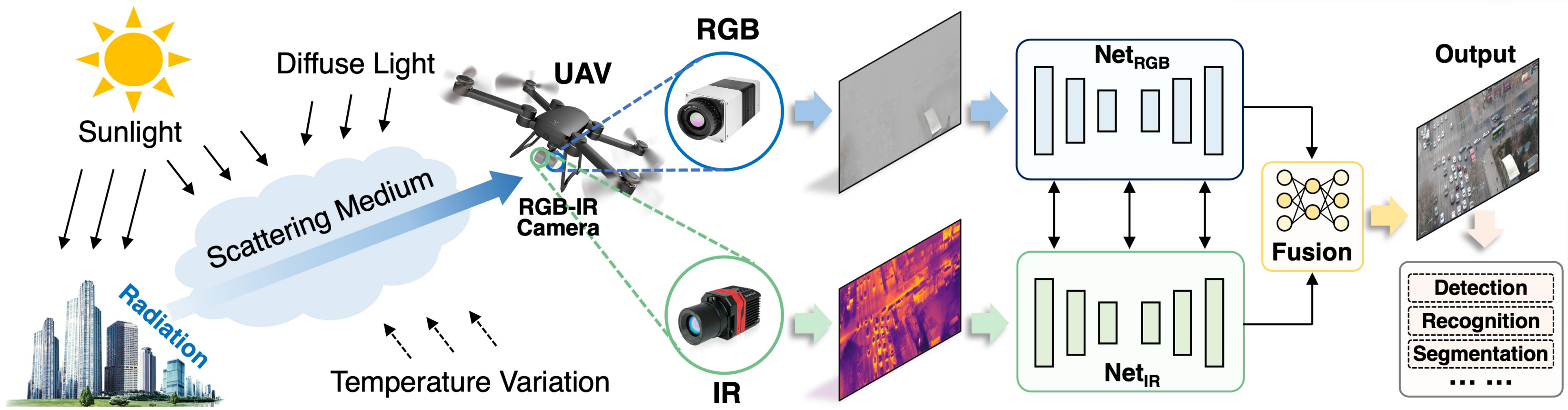
where $\beta_{IR} \ll \beta_{RGB}$ (Rayleigh scattering)





Main Contribution

- We present an original work to utilize infrared features to guide visible light dehazing in UAV scenarios, where small objects are prone to dense occlusions and dynamic haze.
- We propose the Haze Distribution-aware Cross-modal Fusion Network (HDCFN), incorporating Infrared-guided Multiscale Feature Enhancement (IMFE) and Haze Distribution-aware Cross-Modal Adaptive Fusion (CMAF).
- Extensive experiments on VTUAV, CART, and real-world UAV-captured datasets show superior PSNR/SSIM and better downstream detection.

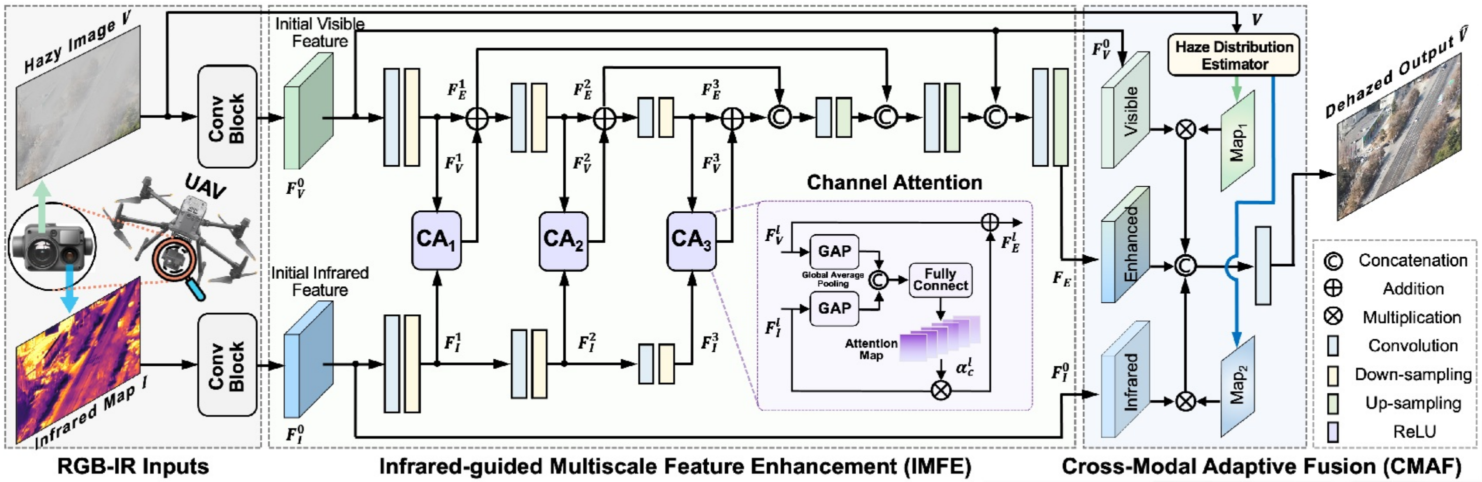


HDCFN Overview

Formulation:

- **Input:** spatially aligned visible V and infrared I.
- **Network F_θ outputs dehazed image** V': $V' = F_\theta(V, I)$.
- **Training loss:** $L = L_{rec}(V', V_{gt}) + \lambda \cdot L_{perc}(F^0_V, F^0_I)$,

  where L_{rec}: MAE; L_{perc}: MSE between initial feature maps (λ=0.1).

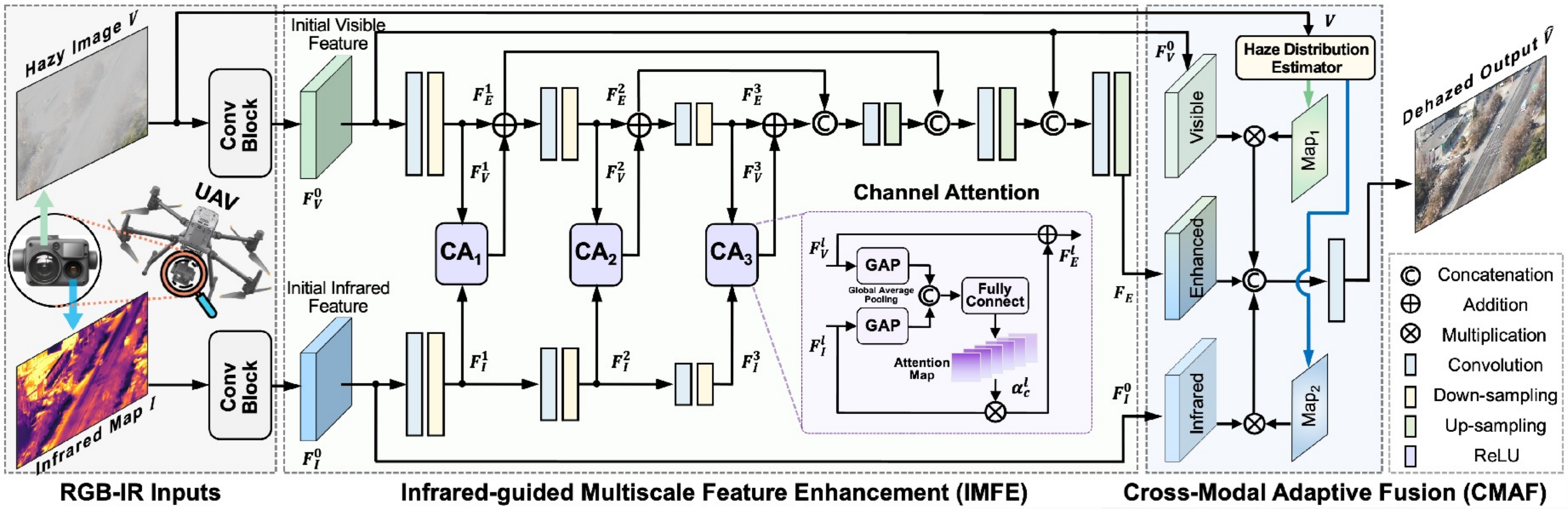


HDCFN leverages the strengths of dual modalities for robust dehazing.

HDCFN Overview

Architecture:

- **IMFE:** Multiscale extraction from RGB/IR; channel attention injects IR structure into RGB features across scales.
- **CMAF:** Haze Distribution Estimator (HDE) predicts spatial haze map.
- **Adaptive Fusion:** prioritize IR features in dense haze and RGB features in light haze, leading to better color and structure restoration.

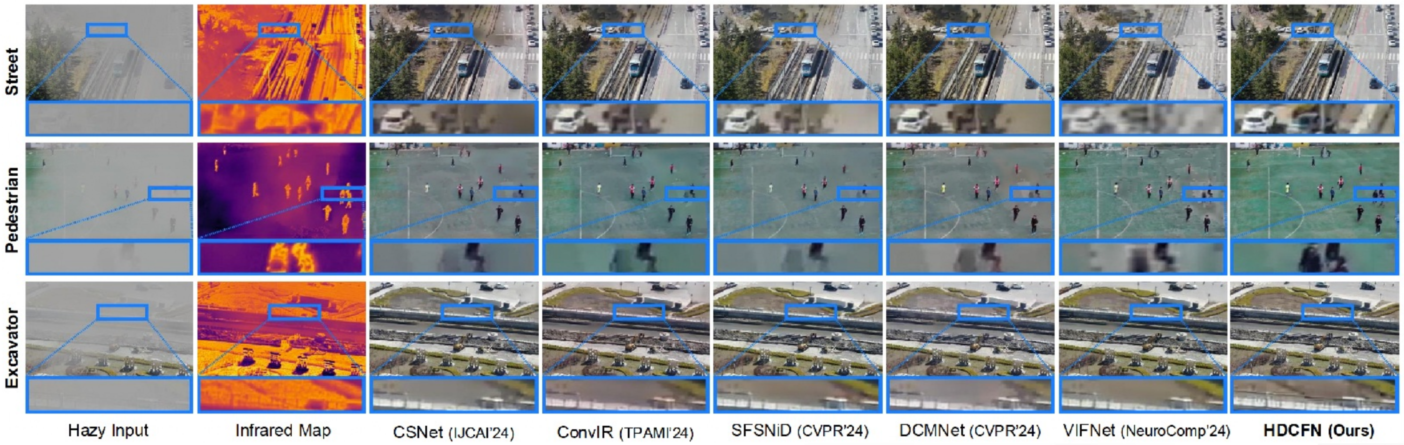


Experimental Results

On VTUAV Dataset:

- Experimental results demonstrate a clear performance advantage of our method over other methods.

- Our method not only restores the object contours but also recovers its color. These results highlight the robustness of our method in reconstructing fine details and preserving structural integrity, even in challenging conditions.



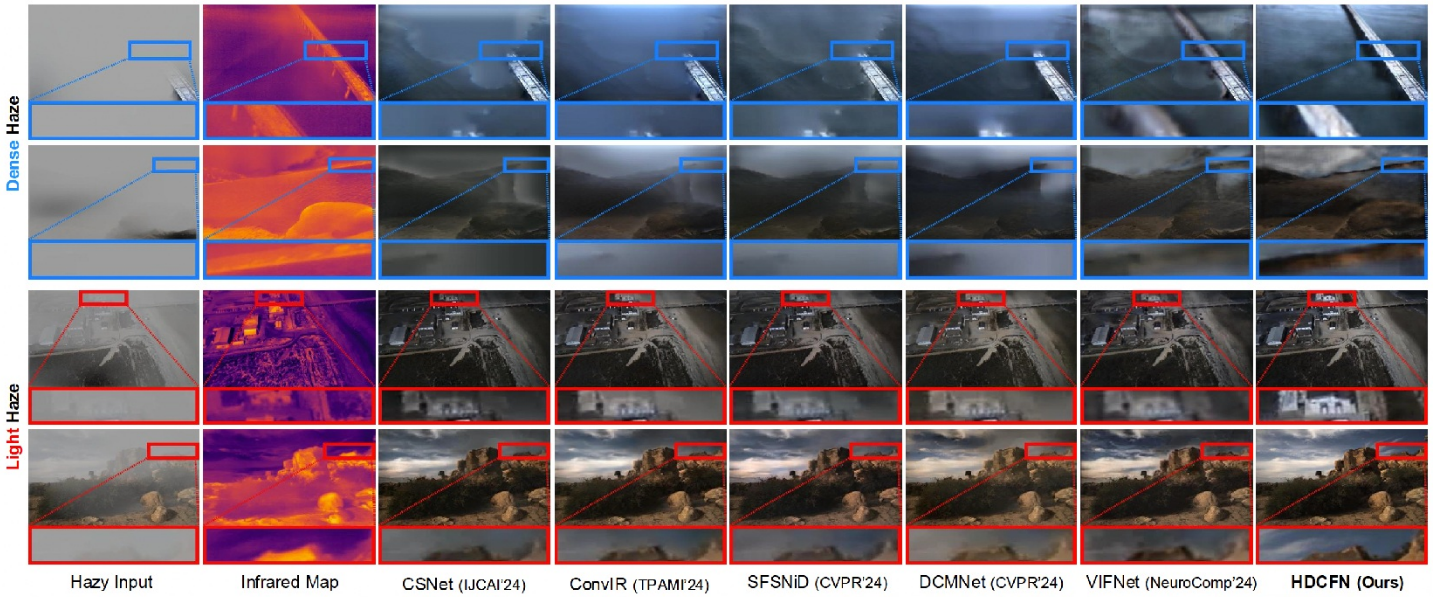
Method	Citation	Modality	Bike		Street		Car		Excavator		Pedestrian		Train		Truck		**Average**	
			PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
CSNet	IJCAI'24	RGB	25.21	0.884	25.48	0.865	25.36	0.831	25.36	0.865	27.43	0.894	25.15	0.889	27.47	0.903	25.92	0.876
ConvIR	TPAMI'24	RGB	28.07	0.891	28.29	0.869	28.16	0.832	29.13	0.872	29.32	0.892	27.45	0.891	30.16	0.907	28.64	0.879
SFSNiD	CVPR'24	RGB	26.41	0.885	25.07	0.859	26.38	0.828	26.68	0.863	27.38	0.893	24.49	0.881	26.05	0.890	26.07	0.871
DCMNet	CVPR'24	RGB+Depth	28.53	0.893	29.13	0.871	28.79	0.835	29.08	0.874	29.14	0.895	29.46	0.889	28.52	0.907	28.95	0.880
VIFNet	NeuroComp'24	RGB+IR	28.79	0.896	29.04	0.871	29.15	0.843	28.67	0.864	28.72	0.895	31.06	0.894	29.03	0.912	29.21	0.881
Ours	ACM MM'25	RGB+IR	**30.56**	**0.905**	**31.69**	**0.913**	**31.25**	**0.914**	**31.42**	**0.921**	**31.51**	**0.916**	**33.32**	**0.925**	**31.24**	**0.933**	**31.57**	**0.918**
Gain (%)	–	–	6.1%	1.0%	8.8%	4.8%	7.2%	8.4%	7.9%	5.4%	7.5%	2.3%	7.3%	3.5%	3.6%	2.3%	**8.1%**	**4.2%**

Experimental Results

On CART Dataset:

- Our method consistently surpasses competing approaches, with notable improvements observed in dense haze scenarios.

- These results highlight the effectiveness of our approach, suggesting that the integration of infrared signals enhances the model’s ability to capture essential structural and textural details obscured by haze.

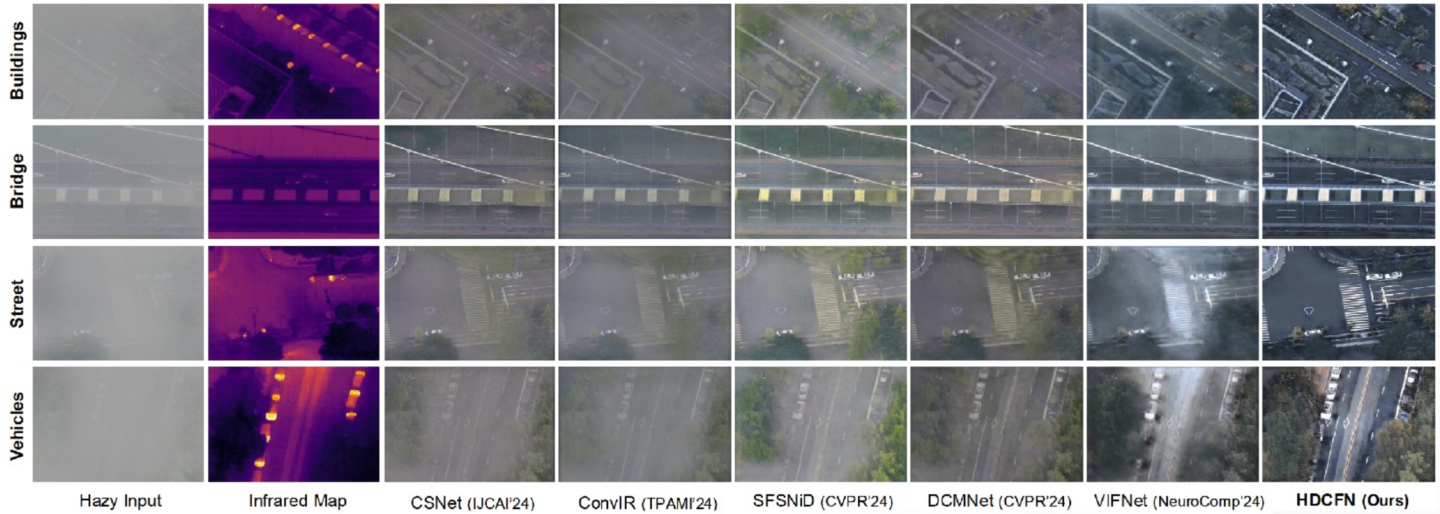


Method	Modality	Bridge		Lake		River		Rock		**Average**	
		Dense	Light	Dense	Light	Dense	Light	Dense	Light	Dense	Light
CSNet	RGB	20.80/0.795	22.74/0.865	20.87/0.821	25.42/0.907	17.32/0.807	22.24/0.882	19.93/0.756	23.68/0.867	19.73/0.795	23.52/0.880
ConvIR	RGB	21.39/0.805	26.15/0.872	21.42/0.831	27.20/0.910	19.85/0.830	25.25/0.891	21.48/0.768	26.65/0.875	21.04/0.808	26.31/0.887
SFSNiD	RGB	21.23/0.802	26.43/0.873	22.13/0.830	27.41/0.908	18.97/0.829	23.31/0.886	20.95/0.765	25.99/0.871	20.82/0.806	25.78/0.884
DCMNet	RGB+Depth	21.43/0.803	26.48/0.881	22.39/0.832	27.82/0.915	21.41/0.842	25.30/0.896	20.76/0.769	27.19/0.875	21.49/0.810	26.69/0.891
VIFNet	RGB+IR	22.12/0.809	26.54/0.881	23.14/0.837	28.25/0.917	21.37/0.833	25.84/0.897	22.79/0.773	28.53/0.879	22.36/0.813	27.29/0.893
Ours	RGB+IR	**23.51/0.834**	**28.15/0.894**	**24.86/0.896**	**29.43/0.935**	**23.78/0.869**	**27.98/0.907**	**24.37/0.837**	**29.40/0.911**	**24.13/0.859**	**28.74/0.912**
Gain (%)	–	6.3%/3.1%	6.1%/1.5%	7.4%/7.0%	4.2%/2.0%	11.0%/3.2%	8.3%/1.1%	6.9%/8.3%	3.0%/3.6%	**7.9%/5.7%**	**5.3%/2.1%**

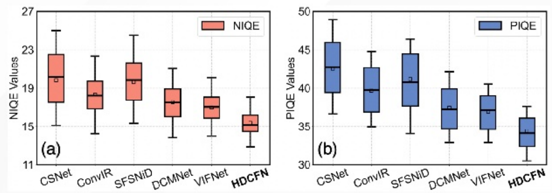
Experimental Results

On CityUAV Dataset:

- We evaluate the proposed method on the real-world UAV-captured dataset.
- HDCFN maintains structural coherence and contrast, effectively mitigating haze artifacts while preserving realistic color tones, resulting in dehazed images that are visually appealing and accurate to the original scene.



Metric	CSNet	ConvIR	SFSNiD	DCMNet	VIFNet	**HDCFN**	Gain
NIQE ↓	19.824	18.335	19.640	17.527	16.983	**15.379**	9.4%
PIQE ↓	42.532	39.682	41.168	37.419	36.926	**34.407**	6.8%



Conclusion

Conclusion:

- This paper proposes an infrared-guided image dehazing method for UAV scenarios, which exploits the complementary strengths of visible and infrared modalities.
- This approach integrates a multiscale feature enhancement framework with a cross-modal adaptive fusion module, facilitating recovery of fine-grained details and dynamic adaptation to varying haze distributions.

Future Work:

- Joint calibration & alignment between RGB and IR.
- On-board real-time UAV deployment.

Thanks for watching！

中国电信
CHINA TELECOM