OneHRL: Hybrid Reinforcement Learning for One-step Degraded Infrared and Visible Image Fusion

Hongwei Yu[1,2*], Daoqing Zha[1*], Jiawei Li[1†], Yancheng Bai[2†‡], Lei Sun[2], Xiangxiang Chu[2]

[1]University of Science and Technology Beijing [2]DreamX Team, Alibaba Group

* Equal contribution † Corresponding authors ‡ Project lead

Project Page

1 Motivation

Real-world inputs suffer composite degradations

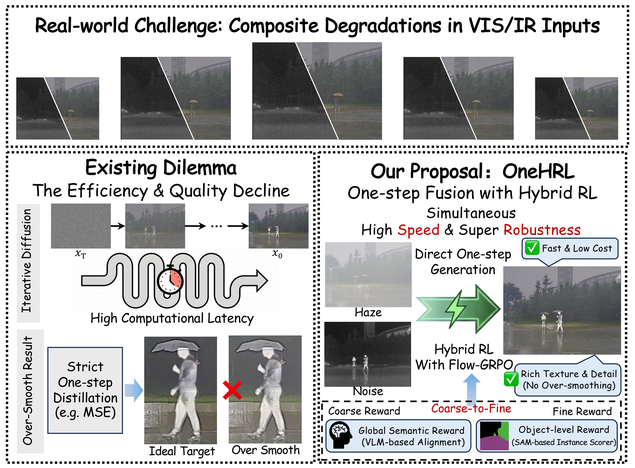
Haze, rain, blur, noise and stripes often corrupt VIS and IR images at once, yet most IVIF methods assume clean inputs or a single degradation type.

Diffusion is robust, but slow

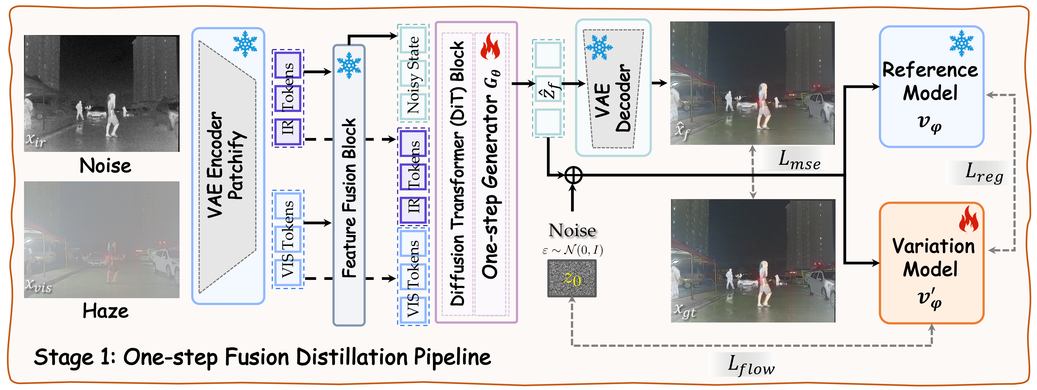
Diffusion-based fusion (e.g., DRMF, MMAIF) copes well with degradations, but iterative sampling is costly: DRMF needs 25 s per image.

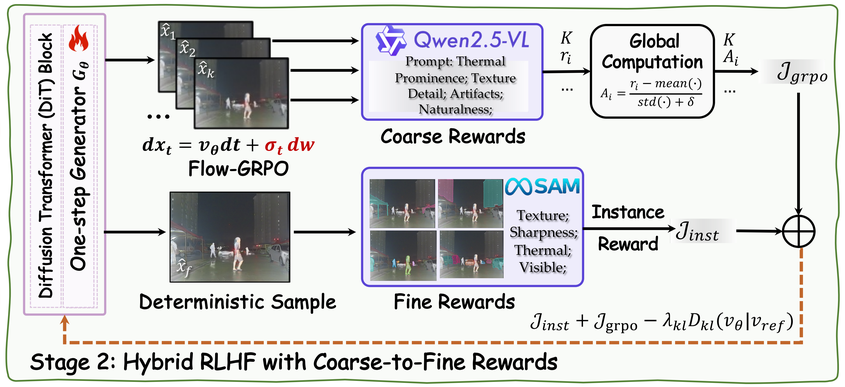
One-step distillation over-smooths

Strict distillation (e.g., MSE) over-smooths results, and its deterministic mapping of a fixed input leaves no stochasticity for sampling-based RL.



2 Method: One-step Fusion with Hybrid RL





STAGE 1

Degraded Input as Noisy State

A frozen Feature Fusion Block fuses the degraded VIS/IR latents into an intermediate noisy state on the generative trajectory; the one-step generator rectifies it to the clean manifold in a single pass (MSE + VSD), taking 0.4 s vs. 25 s per image for DRMF.

$$\hat{x}_f = \mathcal{D}\big(G_\theta(z_f^d, z_{vis}^d, z_{ir}^d)\big)$$

$$G_\theta(\cdot) = z_f^d + \lambda_1 \, v_\theta(z_f^d, z_{vis}^d, z_{ir}^d)$$

$$\mathcal{L}_{total} = \mathcal{L}_{mse} + \lambda_{reg} \, \mathcal{L}_{reg}$$

STAGE 2 · COARSE

ODE → SDE via Flow-GRPO

A fixed input makes one-step generation deterministic and blocks RL exploration. Flow-GRPO rewrites the ODE as an SDE that injects controlled noise while preserving marginals; K samples are scored by Qwen2.5-VL with a CoT prompt and normalized into group advantages.

$$dz_t = \left[v_\theta(\cdot) + \frac{\sigma_t^2}{2t}\big(z_f^d + (1-t)\, v_\theta(\cdot)\big) \right] dt + \sigma_t \, dw$$

$$A_i = \frac{r_i - \text{mean}(\{r_1, \ldots, r_K\})}{\text{std}(\{r_1, \ldots, r_K\}) + \delta}$$

STAGE 2 · FINE

SAM-based Instance Reward

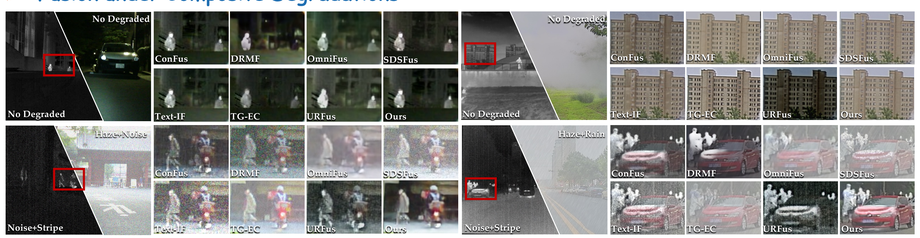
SAM extracts instance masks, and each object is scored on thermal prominence, local sharpness, information entropy, IR visibility and SSIM. With the VLM reward and a KL penalty to the reference model, it enforces global naturalness and object-level fidelity.

$$\mathcal{J}_{inst} = \frac{1}{N} \sum_{j=1}^{N} \sum_{k \in \mathcal{K}} S_k(m_j), \qquad \mathcal{K} = \{tp, ls, ie, iv, ssim\}$$

$$\mathcal{J}_{total} = \mathcal{J}_{GRPO} + \lambda_{inst} \, \mathcal{J}_{inst} - \lambda_{kl} \, \mathcal{D}_{KL}$$

3 Experiments

Fusion under Composite Degradations

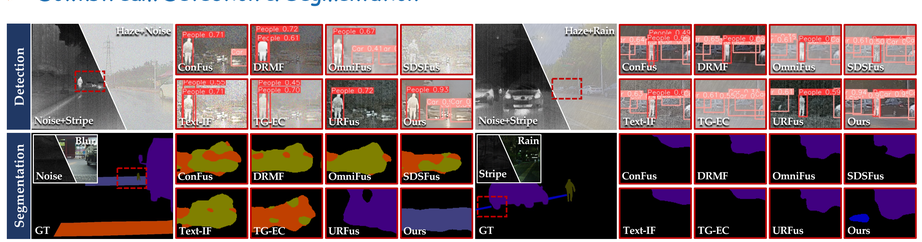


(a) MSRS (b) M3FD

Method	M3FD					MSRS				
	CC↑	SCD↑	MI↑	VIF↑	MSE↓	CC↑	SCD↑	MI↑	VIF↑	MSE↓
ConFus	0.492	1.337	2.363	0.487	0.049	0.562	0.920	1.569	0.351	0.126
DRMF	0.453	1.304	1.091	0.225	0.048	0.558	1.021	1.137	0.320	0.049
OmniFus	0.418	0.779	2.335	0.386	0.112	0.552	0.734	1.577	0.342	0.182
SDSFus	0.311	0.846	0.730	0.216	0.048	0.435	0.752	0.861	0.243	0.149
Text-IF	0.465	0.833	2.156	0.443	0.118	0.568	0.824	1.387	0.333	0.219
TG-EC	0.432	1.059	2.753	0.454	0.044	0.443	0.740	0.825	0.234	0.097
URFus	0.416	0.810	2.497	0.420	0.117	0.429	0.648	0.906	0.258	0.222
OneHRL (Ours)	**0.519**	**1.544**	**3.088**	**0.498**	**0.039**	**0.611**	**1.527**	**2.384**	**0.465**	**0.033**

M3FD: Haze+Rain (VIS) & Noise+Stripe (IR); MSRS: Haze+Noise (VIS) & Noise+Stripe (IR). Orange: best; underline: second best.

Downstream Detection & Segmentation

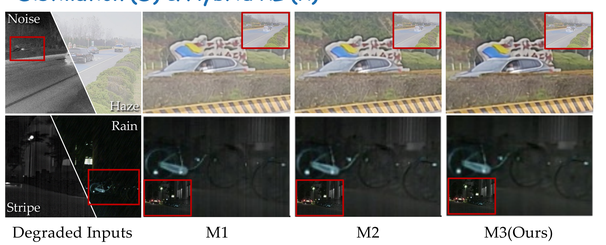


Method	mAP@.5		mIoU	
	N-S + H-N	N-S + H-R	Noise + Blur	Stripe + Rain
ConFus	$0.196_{\downarrow 0.459}$	$0.457_{\downarrow 0.198}$	$0.527_{\downarrow 0.039}$	$0.441_{\downarrow 0.125}$
DRMF	$0.165_{\downarrow 0.413}$	$0.367_{\downarrow 0.211}$	$0.521_{\downarrow 0.027}$	$0.385_{\downarrow 0.164}$
OmniFus	$0.160_{\downarrow 0.440}$	$0.249_{\downarrow 0.351}$	$0.496_{\downarrow 0.029}$	$0.449_{\downarrow 0.076}$
SDSFus	$0.081_{\downarrow 0.405}$	$0.175_{\downarrow 0.311}$	$0.407_{\downarrow 0.153}$	$0.374_{\downarrow 0.185}$
Text-IF	$0.156_{\downarrow 0.503}$	$0.285_{\downarrow 0.374}$	$0.513_{\downarrow 0.056}$	$0.390_{\downarrow 0.180}$
TG-EC	$0.134_{\downarrow 0.535}$	$0.231_{\downarrow 0.438}$	$0.508_{\downarrow 0.031}$	$0.352_{\downarrow 0.187}$
URFus	$0.172_{\downarrow 0.474}$	$0.153_{\downarrow 0.493}$	$0.275_{\downarrow 0.276}$	$0.401_{\downarrow 0.150}$
OneHRL (Ours)	**$0.470_{\downarrow 0.159}$**	**$0.568_{\downarrow 0.061}$**	**$0.601_{\downarrow 0.018}$**	**$0.609_{\downarrow 0.010}$**

YOLOv5 / DeepLabV3+ on fused images; subscript = drop vs. clean input. N-S: Noise+Stripe, H-N: Haze+Noise, H-R: Haze+Rain.

4 Ablation

Distillation (D) & Hybrid RL (R)



Data	Model	D	R	CC↑	SCD↑	MI↑	VIF↑	MSE↓
M3FD	M1	✗	✗	0.431	1.395	2.876	0.511	0.128
	M2	✓	✗	0.462	1.477	3.009	0.534	0.086
	M3 (Ours)	✓	✓	**0.520**	**1.544**	**3.310**	**0.573**	**0.059**
MSRS	M1	✗	✗	0.573	1.445	2.316	0.555	0.046
	M2	✓	✗	0.591	1.537	2.764	0.632	0.051
	M3 (Ours)	✓	✓	**0.620**	**1.618**	**2.982**	**0.705**	**0.032**

M1: DiT baseline with 5-step sampling; M2 = M1 + D; M3 = M2 + R.

Coarse Reward & KL Constraint

Removed component	Metric	w/o	OneHRL
Flow-GRPO (coarse reward)	mAP@.5	0.426	**0.561**
KL constraint	mIoU	0.493	**0.601**

Detection under Noise (VIS) + Stripe (IR); segmentation under Blur (VIS) + Noise (IR). w/o Flow-GRPO keeps only the instance reward.