# Depth-Centric Dehazing and Depth-Estimation from Real-World Hazy Driving Video

Junkai Fan[1], Kun, Wang[1], Zhiqiang Yan[1], Xiang Chen[1], Shangbing Gao[2], Jun Li[1*] , and Jian Yang[1*]
[1]PCA Lab, Nanjing University of Science and Technology, China
[2]Huaiyin Institute of Technology, China

AAAI-25 / IAAI-25 / EAAI-25
FEBRUARY 25 – MARCH 4, 2025 | PHILADELPHIA, USA

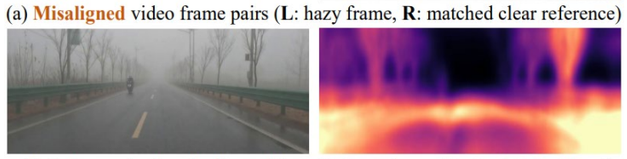
Project Page

## Motivation

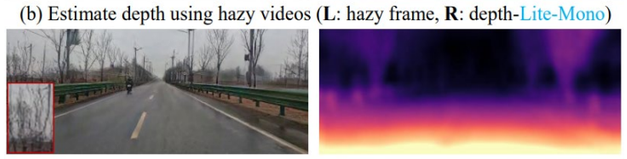
- Obtaining clear RGB and accurate dense depth ground truth is challenging in real hazy scenes.
- Self-supervised depth estimation and physics-based dehazing are inherently complementary in real-world hazy scenes.

$$\begin{cases} I_t(x) = J_t(x)e^{-\beta \boldsymbol{d}(x)} + A_\infty(1 - e^{-\beta \boldsymbol{d}(x)}) \\ J_t(x) = \mathcal{S}(J_s, y), y \sim KP_{x\to y}\boldsymbol{d}(x)K^{-1}x \end{cases} \quad (1)$$

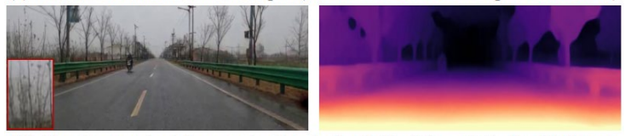
(a) **Misaligned** video frame pairs (**L**: hazy frame, **R**: matched clear reference)



(b) Estimate depth using hazy videos (**L**: hazy frame, **R**: depth-Lite-Mono)



(c) Dehaze first, then estimate depth (**L**: dehazed-DVD, **R**: depth-Lite-Mono)



(d) Simultaneously dehaze and estimate depth (**L**: dehazed, **R**: depth) (Ours)

## Depth-Centric Learning (DCL)

### Main idea:

- Based on Eq. (1), we propose a novel Depth-Centric Learning (DCL) framework to simultaneously remove haze and estimate depth from real-world hazy videos.
- We present two misaligned regularization discriminator networks, $D_{MFIR}$ and $D_{MDR}$, for enhancing constraints on high-frequency details and weak texture regions.
- In real hazy scenes, scattering does not always conform to an ideal model, as β depends not only on wavelength but also on the size and distribution of scattering particles (e.g., patchy haze). Therefore, we assume β to be a non-uniform variable.

### Train loss:

Misaligned Reference Loss

ASM Loss | BCC Loss | $D_{MFIR}$

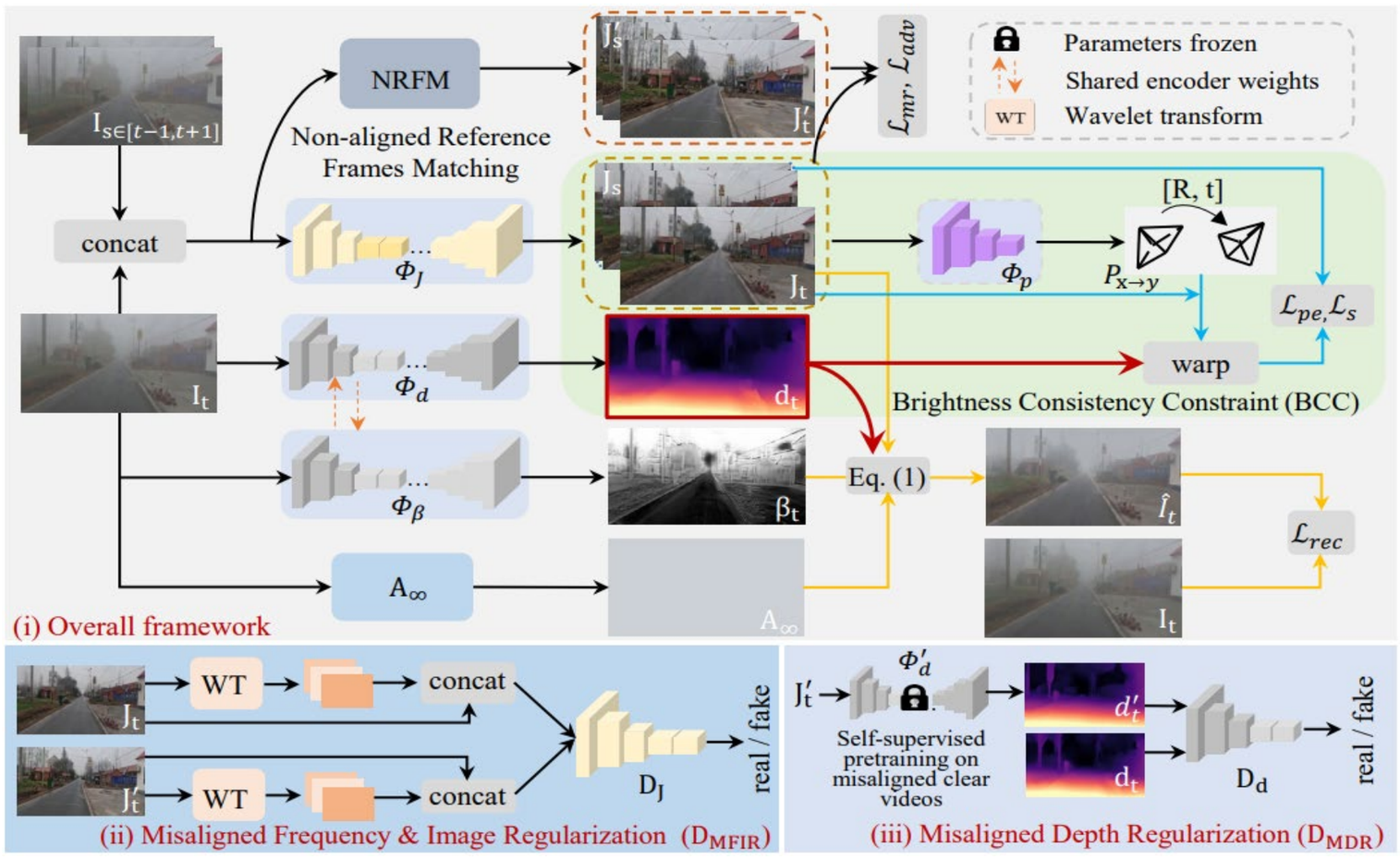
$$Loss = \eta\mathcal{L}_{rec} + \gamma\mathcal{L}_{mr} + m_a\mathcal{L}_{pe} + \xi\mathcal{L}_s + \omega_1(\mathcal{L}_D^I + \mathcal{L}_G^I) + \omega_2(\mathcal{L}_D^d + \mathcal{L}_G^d)$$

$D_{MDR}$

**Note:** Based on the brightness consistency assumption, $D_{MDR}$ regularizes the depth while also constraining the dehazing network to maintain brightness consistency across consecutive frames, thereby preventing flickering.

## Overall Framework

Our DCL framework integrates the atmospheric scattering model (ASM) and brightness consistency constraint (BCC) via shared depth prediction. $D_{MFIR}$ enhances high-frequency details in dehazed frames, while $D_{MDR}$ reduces black holes in depth maps from weakly textured regions.



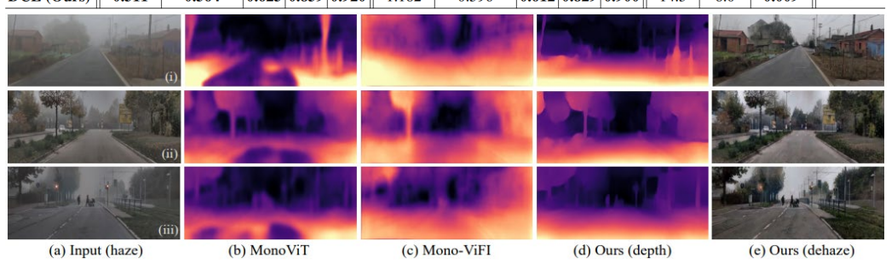

## Experimental Results and Ablation Study

### Quantitative dehazing results on three real hazy video datasets

| Data Settings | Methods | Data Type | GoProHazy FADE ↓ | GoProHazy NIQE ↓ | DrivingHazy FADE ↓ | DrivingHazy NIQE ↓ | InternetHazy FADE ↓ | InternetHazy NIQE ↓ | Params (M) | FLOPs (G) | Inf. time (S) | Ref. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Unpaired | DCP | Image | 1.0415 | 7.4165 | 1.1260 | 7.4455 | 0.9229 | 7.4899 | - | - | 1.39 | CVPR'09 |
| | RefineNet | Image | 1.1454 | 6.1837 | 1.0223 | 6.5959 | 0.8535 | 6.7142 | 11.38 | 75.41 | 0.105 | TIP'21 |
| | CDD-GAN | Image | 0.7797 | 6.0691 | 1.0072 | 6.1968 | 0.8166 | 6.1969 | 29.27 | 56.89 | 0.082 | ECCV'22 |
| | $D^4$ | Image | 1.5618 | 6.9302 | 0.9556 | 7.0448 | 0.6913 | 7.0754 | **10.70** | **2.25** | 0.078 | CVPR'22 |
| Paired | PSD | Image | 0.9081 | 6.7996 | 0.9479 | 6.3381 | 0.8100 | 6.1401 | 33.11 | 182.5 | 0.084 | CVPR'21 |
| | RIDCP | Image | 0.7250 | 5.2559 | 0.9187 | 5.3063 | 0.6564 | 5.4299 | 28.72 | 182.69 | 0.720 | CVPR'23 |
| | PM-Net | Video | 0.7559 | 4.6274 | 1.0509 | 4.8447 | 0.7696 | 5.0182 | 151.20 | 5.22 | 0.277 | ACMM'22 |
| | MAP-Net | Video | 0.7805 | 4.8189 | 1.0992 | 4.7564 | 1.0595 | 5.5213 | 28.80 | 8.21 | 0.668 | CVPR'23 |
| Non-aligned | NSDNet | Image | 0.7197 | 6.1026 | 0.8670 | 6.3558 | 0.6595 | 4.3144 | 11.38 | 56.86 | **0.075** | arXiv'23 |
| | DVD | Video | 0.7061 | 4.4473 | 0.7739 | 4.4820 | 0.6235 | 4.5758 | 15.37 | 73.12 | 0.488 | CVPR'24 |
| | **DCL (Ours)** | Video | **0.6914** | **3.4412** | **0.7380** | **3.5329** | **0.6203** | **3.5545** | 11.38 | 56.86 | **0.075** | - |

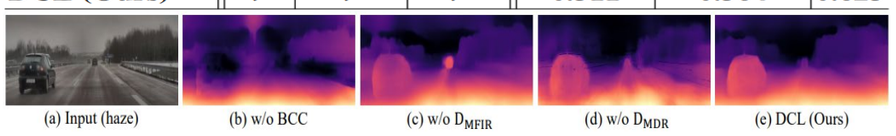

### Quantitative depth estimation results on DENSE-Fog dataset

| Method | DENSE-Fog (light) abs Rel↓ | RMSE log↓ | $\delta_1$↑ | $\delta_2$↑ | $\delta_3$↑ | DENSE-Fog (dense) abs Rel↓ | RMSE log↓ | $\delta_1$↑ | $\delta_2$↑ | $\delta_3$↑ | Params (M) | FLOPs (G) | Inf. time (S) | Ref. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MonoDepth2 | 0.418 | 0.475 | 0.499 | 0.735 | 0.847 | 1.045 | 0.632 | 0.530 | 0.771 | 0.864 | 14.3 | 8.0 | **0.009** | ICCV'19 |
| MonoViT | 0.393 | 0.454 | 0.464 | 0.728 | 0.858 | 0.992 | 0.611 | 0.512 | 0.779 | 0.876 | 78.0 | 15.0 | 0.045 | 3DV'22 |
| Lite-Mono | 0.417 | 0.473 | 0.402 | 0.687 | 0.853 | 0.954 | 0.604 | 0.469 | 0.756 | 0.886 | **3.1** | **5.1** | 0.013 | CVPR'23 |
| RobustDepth | 0.316 | 0.370 | 0.611 | 0.828 | 0.913 | **0.605** | **0.515** | 0.563 | 0.798 | 0.881 | 14.3 | 8.0 | **0.009** | ICCV'23 |
| Mono-ViFI | 0.369 | 0.459 | 0.408 | 0.704 | 0.864 | 0.609 | 0.528 | 0.489 | 0.771 | 0.883 | 14.3 | 8.0 | **0.009** | ECCV'24 |
| **DCL (Ours)** | **0.311** | **0.364** | **0.623** | **0.839** | **0.920** | 1.182 | 0.596 | **0.612** | **0.829** | **0.900** | 14.3 | 8.0 | **0.009** | - |




### Ablation visualization of core module on DENSE-Fog (light)

| Method | BCC | $D_{MFIR}$ | $D_{MDR}$ | Abs Real↓ | RMSE log↓ | $\delta_1$↑ |
|---|---|---|---|---|---|---|
| DCL w/o BCC | | ✓ | ✓ | 0.636 | 0.569 | 0.439 |
| DCL w/o $D_{MFIR}$ | ✓ | | ✓ | 0.320 | 0.366 | 0.621 |
| DCL w/o $D_{MDR}$ | ✓ | ✓ | | 0.340 | 0.392 | 0.562 |
| **DCL (Ours)** | ✓ | ✓ | ✓ | **0.311** | **0.364** | **0.623** |




### Quantitative comparison different β types on DENSE-Fog (light)

| Shape of $\beta$ | Type | Abs Rel↓ | RMSE log↓ | $\delta_1$↑ |
|---|---|---|---|---|
| (1, 1, 1) | Constant | 0.325 | 0.371 | 0.621 |
| **(1, 192, 640) (Ours)** | Non-uniform | **0.311** | **0.364** | **0.623** |