HAFM: A Post-Fusion Gating Module for Haze-Aware RGB–Thermal Object Detection

Supplementary Material

Juan M. Saeteros Nick J. Arévalo Boris X. Vintimilla
ESPOL Polytechnic University, Ecuador
{jumasaet, narevalo, boris.vintimilla}@espol.edu.ec

Abstract

This document provides additional experimental results, methodological explanations, and qualitative visual analyses to support the findings presented in the main manuscript. First, the detection performance of the Haze-Aware Fusion Module (HAFM) is evaluated under clear weather conditions to demonstrate its robust gating behavior when atmospheric degradation is absent. Second, a Sim-to-Real adaptation analysis is presented, this section includes quantitative metrics and qualitative visual comparisons of both the synthetic haze generators and the resulting degradation masks on real-world data, validating the transferability of the proposed non-homogeneous generator. Finally, comprehensive implementation details of the synthetic haze pipeline are provided, including parameter sampling distributions and noise modeling strategies, followed by information regarding code availability to ensure full reproducibility of the framework.

1. Failure Modes

While the main manuscript focuses on the performance of HAFM under synthetic and real haze, it is equally important to analyze its behavior in clear conditions and consider potential failure scenarios. Two primary edge cases were identified: false positive haze detection (RGB failure) and degraded thermal inputs (IR failure). In the first scenario, the mask estimator might incorrectly flag a clear RGB region as hazy—for instance, when observing textureless surfaces like white walls or clear skies. In these instances, HAFM will route features to the thermal branch. However, detection performance is not significantly compromised because such background regions naturally lack object-specific information. Furthermore, if a target (e.g., a pedestrian) is present against a white wall, its distinct thermal and structural signatures prevent the mask estimator from confusing the object itself with fog. The second scenario involves poor-quality thermal imaging. Since HAFM dynamically balances the baseline fusion with the thermal image, a degraded IR signal could theoretically impair the network. However, due that HAFM acts as a gating mechanism working on the established baseline fusion (TarDAL[3], SeaFusion[5], etc), it serves as a lower-bound performance guarantee. In the worst case, HAFM converges to the baseline's performance rather than degrading it further. To empirically validate that HAFM does not negatively impact performance when haze is entirely absent, the models were tested on the clear images of the M3FD dataset (the original images before synthetic haze was applied). As shown in Table 1, applying HAFM to the baselines under clear weather strictly maintains or improves the original detection accuracy. Most notably, the CFMW baseline experiences a significant boost (+0.061 mAP50) even without haze, demonstrating that the module safely passes through uncorrupted features when the degradation mask $m \approx 0$ and can even refine cross-modal alignment.

Table 1. **Performance in Clear Conditions.** Evaluation on the clear images of the M3FD dataset (no haze). The integration of HAFM maintains or strictly improves the baseline metrics when no atmospheric degradation is present.

Method	P	R	mAP50	mAP50:95
SFDFusion	0.916	0.749	0.832	0.561
SFDFusion + HAFM	0.917	0.750	0.831	0.561
TarDAL	0.892	0.768	0.833	0.565
TarDAL + HAFM	0.913	0.763	0.837	0.571
SeAFusion	0.855	0.779	0.837	0.568
SeAFusion + HAFM	0.887	0.780	0.838	0.568
CFMW	0.918	0.774	0.842	0.554
CFMW + HAFM	0.858	0.825	0.903	0.593

Table 2. **Sim-to-Real Mask Detection on Real Haze.** Performance of mask estimators trained on different synthetic haze generators and tested on 222 real-haze images from M3FD. Arrows indicate whether higher (↑) or lower (↓) is better. **Red** is for the worst and **black** is for the best values per column.

Synthetic Hazer	IoU ↑	F1 ↑	Precision ↑	Specificity ↑	MCC ↑	MAE ↓
Synt. Haze [6]	**0.520**	**0.648**	**0.527**	**0.255**	0.294	0.419
HazeFlow [4]	**0.101**	**0.154**	0.648	**0.903**	**0.064**	**0.449**
Ours	0.425	0.581	**0.689**	0.802	**0.354**	**0.401**

2. Sim-to-Real Adaptation Analysis

While the reliance on synthetic data is currently unavoidable due to the absence of large-scale, paired real-haze datasets with detection ground truths—a limitation shared by state-of-the-art works such as CFMW[2]—the ultimate goal of our synthesis pipeline is to enable the model to generalize to real-world conditions. To rigorously evaluate this Sim-to-Real transferability and differentiate our synthesis approach from existing methods, we conducted a cross-generator validation experiment. The M3FD dataset (excluding the 222 manually curated real-haze images) was augmented using three different synthetic haze pipelines: our proposed non-homogeneous generator, and two reference methods from the literature (Synt. Haze [6] and HazeFlow [4]). Subsequently, three separate U-Net degradation mask estimators were trained from scratch using identical hyperparameters, one for each synthetic dataset. The quality and generalizability of these models were then tested exclusively on the 222 real-haze images from M3FD. The results, presented in Table 2, demonstrate how well each synthetic prior adapts to real atmospheric degradation.

2.1. Analysis of Mask Quality Metrics

When evaluating continuous degradation masks transferred to real-world data, observing the balance between competing metrics is crucial:

- **The Over-Prediction Problem (Synt. Haze):** At first glance, the model trained on the synthetic haze generator proposed in [6] (named Synt. Haze from now on) achieves the highest Intersection over Union (IoU) and F1-Score. However, a deeper look reveals a severely low Specificity (0.255) and lower Precision (0.527). This indicates that the model learned to aggressively over-predict haze across the entire image. While this brute-force behavior captures actual foggy regions (inflating the IoU), it generates massive false positives in clear areas, which would catastrophically route useful clear RGB features to the thermal branch in our HAFM framework.
- **The Under-Prediction Problem (HazeFlow):** Conversely, the model trained on HazeFlow [4] exhibits excellent Specificity (0.903) but catastrophically low IoU (0.101) and F1-Score (0.154). This model is overly conservative, failing to identify the majority of real-haze regions, rendering the fusion gate ineffective.
- **Balanced Sim-to-Real Transfer (Ours):** Our non-homogeneous generator provides the optimal balance. It achieves the highest Precision (0.689), ensuring that when it detects haze, it is genuinely present. Most importantly, our method achieves the highest Matthews Correlation Coefficient (MCC of 0.354). MCC is considered one of the most reliable metrics for evaluating binary classifiers under class imbalance [1], it is widely regarded as the most robust metric for imbalanced binary predictions, as it only yields a high score if the prediction obtained good results in all four confusion matrix categories (true positives, false negatives, true negatives, and false positives). Furthermore, our method achieves the lowest Mean Absolute Error (MAE of 0.401), indicating that the continuous intensity of the predicted mask most closely aligns with the actual atmospheric degradation.

These metrics confirm that explicitly modeling depth-guided, non-homogeneous irregularities (such as FBM noise and diverse fog banks) during training is essential for the network to successfully identify the complex, spatially varying nature of real-world haze. Figure 1 shows how the models trained with each synthetic haze generator perform. In (b), it can be seen that the model detected virtually the entire image as haze, which is why a higher IoU is observed in Table 2. Column (c) shows that the model trained with HazeFlow does not perform well in detecting haze in terms of opacity or shape; however, its Specificity value is the highest because the small amount it detected as "haze" was indeed real haze, although it performs poorly in the other metrics. Finally, in column (d), it is possible to see what was explained above: detection is not perfect, but it is closer to reality than the other detectors tested. Furthermore, Figure 2 shows visual examples of what synthetic haze looks like when applied.

3. Synthetic Haze Implementation Details

To ensure dataset diversity and prevent overfitting, synthetic haze parameters are dynamically sampled from uniform distributions per image (detailed in Table 3). Spatial airlight modulation is omitted ($\lambda_A = 0$) in favor of a sampled global airlight A_0 to maintain illumination consistency, this was done this way in order to avoid unrealistic illumination gradients inconsistent with depth-guided haze formation in multimodal RGB-thermal scenes. The base transmission,

Figure 1. **Qualitative comparison of Real Haze mask prediction.** Mask estimators were trained exclusively on synthetic data and tested on real-haze images from M3FD. (a) Input Real Haze, (b) Output from model trained on Synt. Haze [6], (c) Output from model trained on HazeFlow [4], and (d) Output from our model. Synt. Haze over-segments the image, while HazeFlow fails to detect most regions. Our method achieves an acceptable fidelity to the actual spatial degradation.

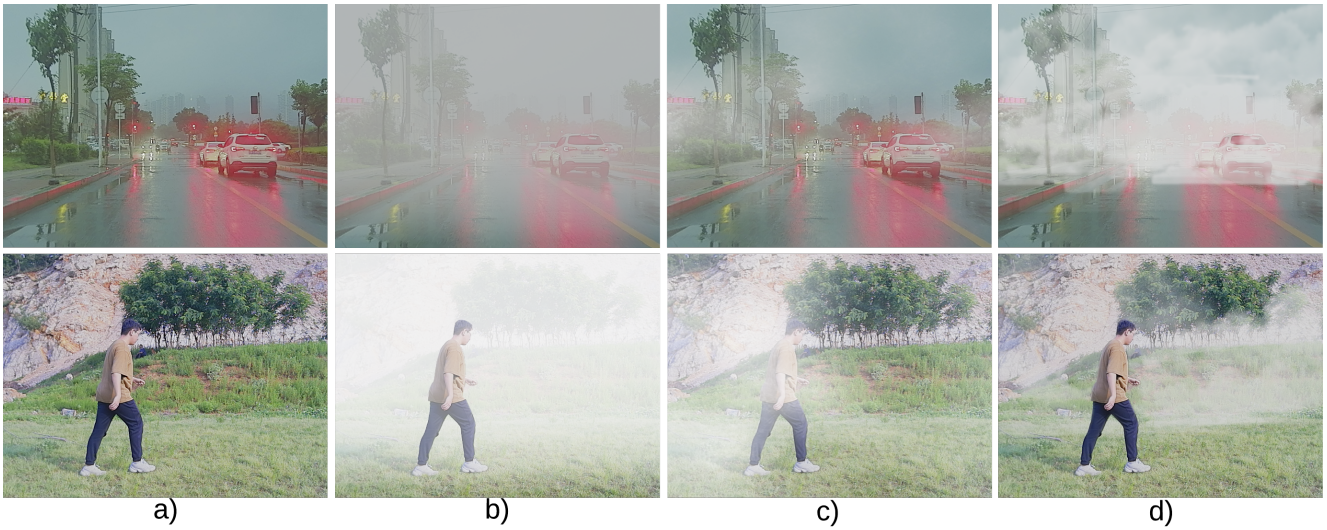

Figure 2. **Visual comparison of synthetic haze generators.** (a) Clean images from M3FD, (b) augmented with Synt. Haze [6], (c) augmented with HazeFlow [4], and (d) augmented with our proposed non-homogeneous pipeline. Our method generates diverse, cloud-like textures and spatially varying densities, which is crucial for emulating the irregularity of real-world atmospheric conditions.

$t(x) = \exp(-\beta_0\, d(x))$, incorporates an extinction coefficient β_0 that is set to zero with a 50% probability to guarantee varied fog densities between batches, this is designed to ensure that the mask estimator is also exposed to clear images during training, ensuring the estimator learns identity behavior when no degradation is present. To simulate realistic atmospheric irregularities, Gaussian-blurred Fractional Brownian Motion (FBM) noise is applied to the transmission map. This non-homogeneity is further amplified by compositing K randomly transformed and depth-gated fog sprites. Finally, a near-guard depth threshold (z_{guard}) is enforced to protect immediate foreground objects from unnatural occlusion, preserving visibility at short ranges.

4. Code Availability and Reproducibility

To facilitate future research in adverse weather RGB–Thermal object detection and to ensure the full reproducibility of our empirical results, the source code developed for this study is made publicly available. The provided repository includes the following key components of our framework:

- **Synthetic Haze Generator:** The complete depth-guided, non-homogeneous synthesis pipeline, including the im-

Table 3. Synthetic Haze Parameter Distributions. Values may be taken from a continuous range $\mathcal{U}(a, b)$ or discrete range $\mathcal{U}\{a, b\}$ per image.

Parameter	Symbol	Distribution / Range
Base Extinction	β_0	$\mathcal{U}(0.0, 0.5)$
Global Airlight	A_0	$\mathcal{U}(0.65, 0.95)$
Number of Sprites	K	$\mathcal{U}\{10, 30\}$
Sprite Alpha Gain	λ_{sprite}	$\mathcal{U}(0.4, 0.8)$
Near Guard Depth	z_{guard}	$\mathcal{U}(0.05, 0.80)$
Global Fog Factor	-	$\mathcal{U}(0.6, 1.2)$
Mask Gamma	γ_{mask}	$\mathcal{U}(0.4, 0.8)$
Mask Gain	g_{mask}	$\mathcal{U}(0.5, 1.2)$

plementation of Fractional Brownian Motion (FBM) noise and the dynamic depth-gated sprite compositing algorithms.

- **Degradation Mask Estimator:** The architecture, training routines, and pre-trained weights for the U-Net model used to predict the continuous degradation mask m.
- **HAFM Integration:** The implementation of the post-fusion gating equation and examples of its integration with the baseline object detection architectures.

The complete source code, along with instructions for dataset preparation and model evaluation, can be accessed at:
`https://github.com/jumasaet/HAFM.git`

References

[1] Davide Chicco and Giuseppe Jurman. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC genomics*, 21 (1):6, 2020. 2

[2] Haoyuan Li, Qi Hu, Binjia Zhou, You Yao, Jiacheng Lin, Kailun Yang, and Peng Chen1. Cfmw: Cross-modality fusion mamba for robust object detection under adverse weather. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2025. 2

[3] Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5792–5801, 2022. 1

[4] Junseong Shin, Seungwoo Chung, Yunjeong Yang, and Tae Hyun Kim. Hazeflow: Revisit haze physical model as ode and non-homogeneous haze generation for real-world dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6263–6272, 2025. 2, 3

[5] Linfeng Tang, Jiteng Yuan, and Jiayi Ma. Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network. *Information Fusion*, 82:28–42, 2022. 1

[6] Le-Anh Tran, Chung Nguyen Tran, Dong-Chul Park, Jordi Carrabina, and David Castells-Rufas. Toward improving robustness of object detectors against domain shift. In *2024 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*, pages 01–05, 2024. 2, 3