HRSNet: Hierarchical Recursive Scaling for Efficient UAV Object Detection

Bowen Yang, , Northeastern University
School of Computer Science and Engineering,

1. Problem

- **Small/Low-quality Objects:** Limited by flight altitude and sensors, objects often appear tiny, blurry, and low-contrast, breaking traditional detection methods.
- **Complex Background Noise:** Urban clutter like building shadows and moving traffic blends with targets, making separation extremely difficult.
- **Strict Real-time Demands:** Existing models sacrifice speed for accuracy, failing critical needs like emergency response and live logistics.

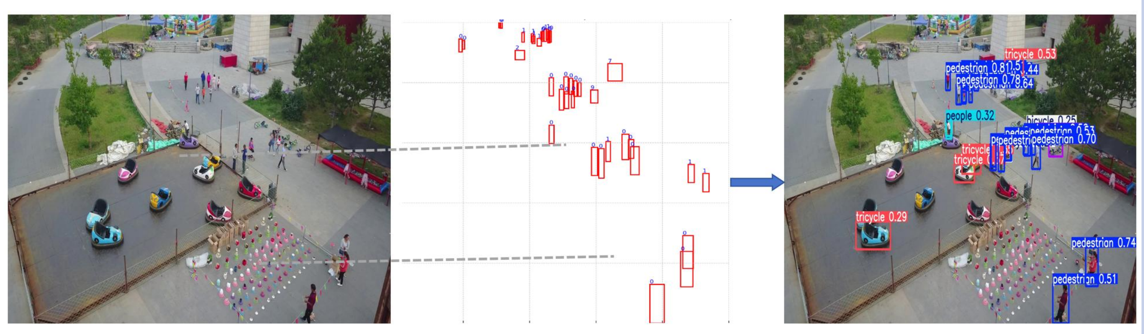
2. Background

- **UAV Technology Advancement:** The rapid development of UAV has been accelerated by the growing low-altitude economy, with expanding applications in smart transportation, disaster rescue, and infrastructure inspection.
- **Critical Role in Smart Systems:** With flexible deployment and wide coverage capabilities, UAVs serve as essential sensing nodes for smart city infrastructure and emergency response systems.
- **Detection Challenges:** UAV-based object detection faces significant technical hurdles, including identifying small parcel labels in cluttered logistics environments and tracking moving targets in complex urban settings.
- **Stringent Performance Requirements:** These applications demand detection models with exceptional accuracy, real-time processing, and robust environmental adaptability to ensure reliable performance.

3. Project Description

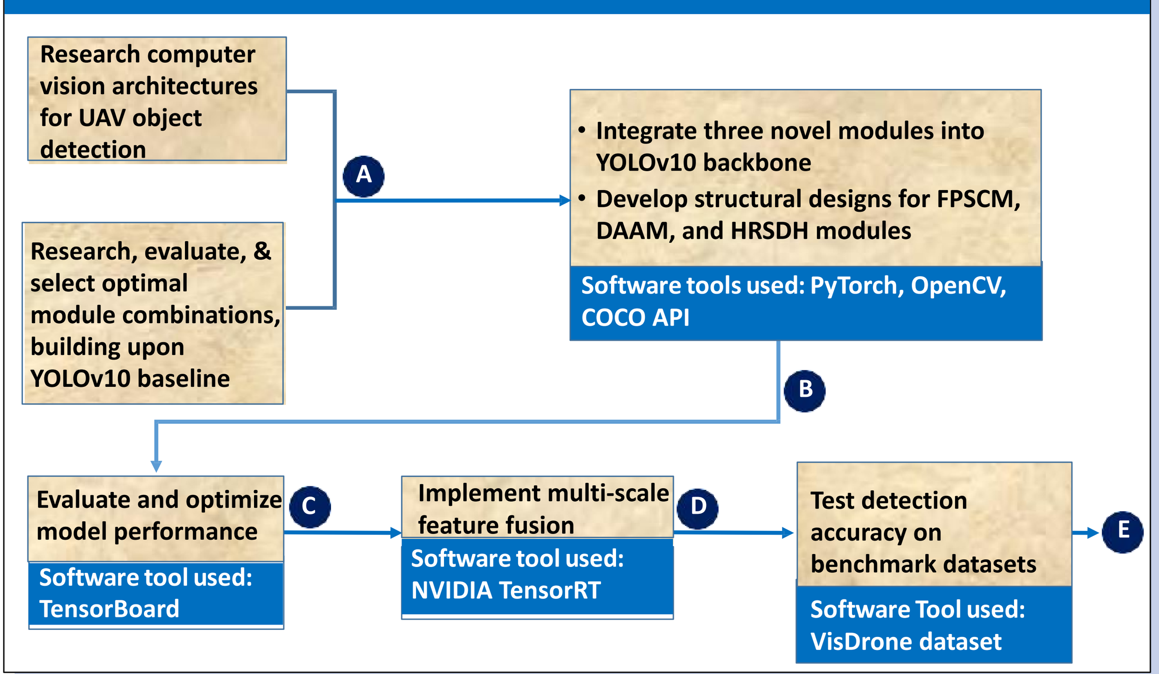
- **HRSNet is an innovative deep learning framework designed to address critical challenges in UAV-based object detection. The system specifically targets three persistent issues in aerial imagery analysis: low-contrast object recognition, sparse bounding box annotation limitations, and precise target localization in complex environments. By overcoming these barriers, HRSNet significantly enhances detection performance for applications ranging from infrastructure inspection to emergency response operations.**
- **The architecture introduces three groundbreaking modules:**
- **FPSCM (Feature Pyramid Shared Convolution Module) for robust multi-scale feature extraction**
- **DAAM (Dynamic Attention Module) that adaptively enhances object features while suppressing background noise**
- **HRSDH (Hierarchical Receptive Field Scale Detection Head) enabling precise localization across varying object sizes**

(a)Task: UAV Object Detection



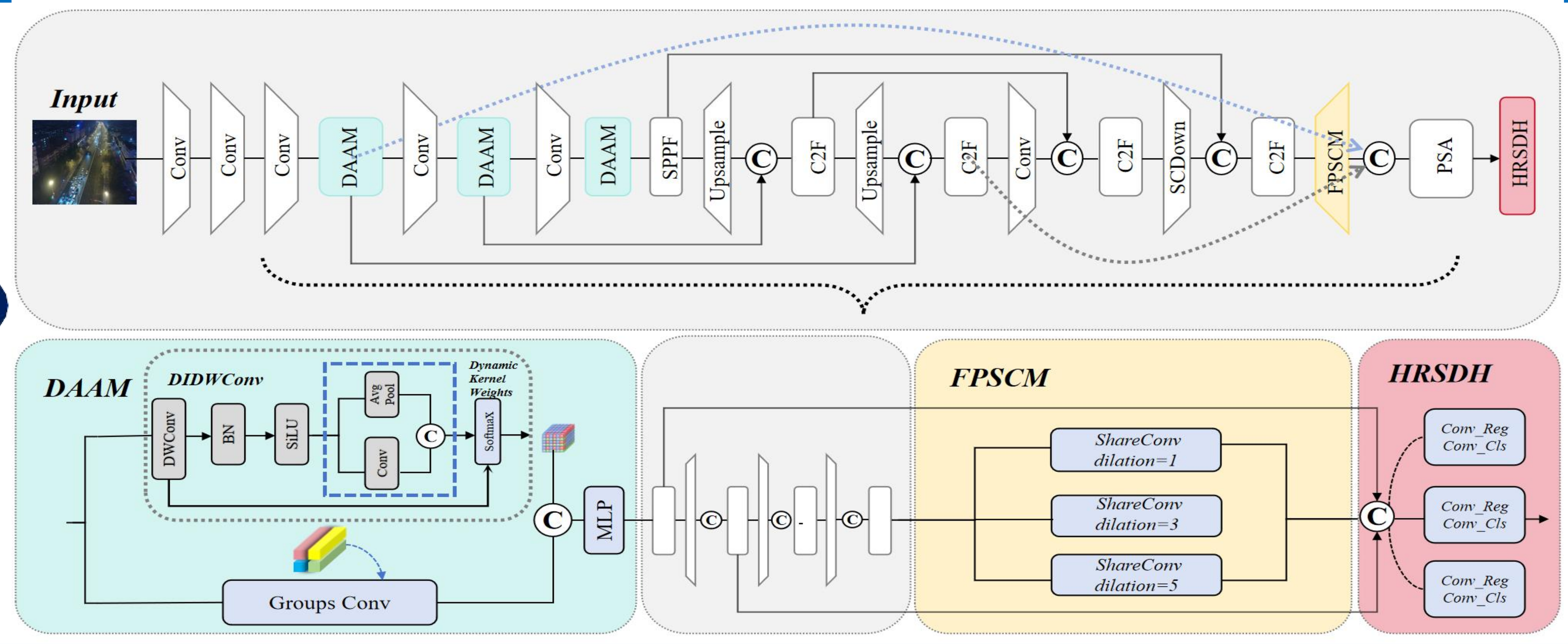
Low-Contrast UAV Objects *Sparse Bounding Box Annotation* *UAV Target Localization*

4. Methods

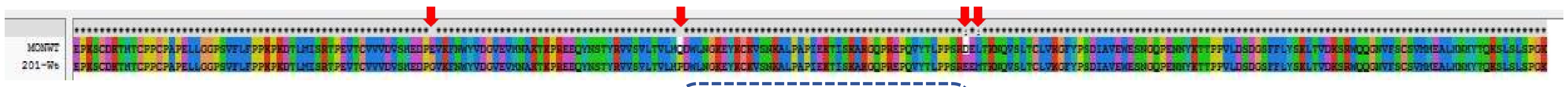


5. Model Analysis & Results

A



- HRSNet enhances the YOLOv10 architecture with three key optimizations for small object detection: (1) efficient small-object feature extraction, (2) parameter redundancy reduction, and (3) lightweight detection heads. As shown in Figure B1, the framework introduces three breakthrough modules - a parameter-shared backbone, adaptive feature fusion, and efficient detection heads - achieving superior performance while maintaining computational efficiency.
- HRSNet's modular design also allows for flexible adaptation to various UAV platforms and sensor configurations, establishing a new benchmark for efficient aerial object detection systems.



B

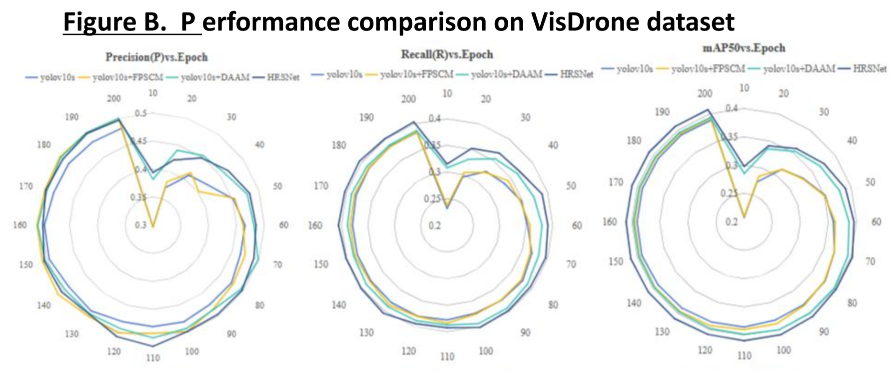
Table B. Outcome Analysis of Ablation Experiments

Dataset	Model	Para. (M)	GFLOPs	P (%)	R (%)	mAP50 (%)
VisDrone	YOLOv10s	7.22	16.5	49.8	38.3	39.1
VisDrone	+FPSCM	8.69	22.1	51.1	38.4	39.7
VisDrone	+DAAM	6.66	15.5	49.7	39.4	40.1
VisDrone	+HRSDH	7.20	18.2	49.2	37.8	39.4
VisDrone	HRSNet	7.20	18.4	50.6	40.1	41.2

The evaluation of HRSNet's core components reveals significant performance improvements over the baseline YOLOv10s architecture. The baseline model demonstrates limited performance on VisDrone, particularly showing late-stage saturation in mAP50 metrics. Integration of the FPSCM module leads to substantially improved detection performance, with accelerated mAP50 growth during mid-training phases, confirming the effectiveness of feature pyramid spatial compression techniques.The combined FPSCM+DAAM configuration achieves 40.1% mAP50, representing a 1% improvement over FPSCM alone.

This enhancement stems from DAAM's depth-invertible structure, which simultaneously reduces parameter count while improving occlusion handling capabilities and maintaining consistent mAP50 growth beyond 180 training epochs. These experimental results highlight the synergistic benefits of combining FPSCM's spatial compression, DAAM's lightweight multi-scale modeling, and enhanced detection head architecture. The integrated approach achieves an optimal balance in remote sensing detection tasks, demonstrating improvements not only in accuracy metrics but also in training stability and convergence speed characteristics.

Figure B. Performance comparison on VisDrone dataset



(a) Precision (b) Recall (c) mAP50

Figure B1 vividly illustrates HRSNet's performance progression on the VisDrone
dataset across training epochs, showing three key improvements:
1) a sustained increase in mAP50, indicating enhanced object detection capability;
2) rising precision values, demonstrating better identification accuracy;
3) improved recall rates, reflecting greater proficiency in recognizing true objects.

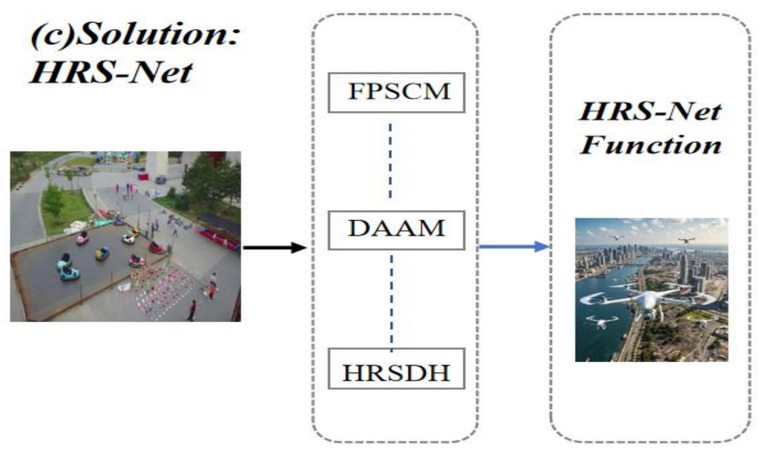
C

Table C. Performance comparison of SOTA models on VisDrone

Model	Precision (%)	Recall (%)	mAP50 (%)	Params (M)
RMVAD-YOLO [26]	46.3	35.6	33.8	10.61
YOLOv5s	49.5	37.8	38.5	9.12
YOLOv8s	48.3	38.6	38.5	9.83
YOLOv10s	49.8	38.3	39.1	7.22
YOLOv11s	49.4	38.9	39.4	9.42
TPH-YOLO	48.4	38.0	39.3	9.20
FFCA-YOLO [21]	48.5	35.8	37.0	7.32
PS-YOLO [27]	51.4	39.4	40.7	5.53
HRSNet (Ours)	**50.6**	**40.1**	**41.2**	**7.20**

As shown in Table C, HRSNet leads the other comparison models across the board in terms of key performance metrics. On the VisDrone dataset, it achieves a precision of 50.6%, a recall of 40.1%, and a mAP50 of 41.2%. These metrics significantly outperform mainstream detection models including YOLOv8, fully demonstrating HRSNet's stable detection capability for small objects in complex aerial scenarios,especially in dealing with high-density objects and multi-scale variations.repulsion energy.

Figure C. Schematic of the algorithmic flow of HRS



By processing UAV imagery through three core modules, the framework's streamlined pipeline achieves precise small-object localization with high computational efficiency.

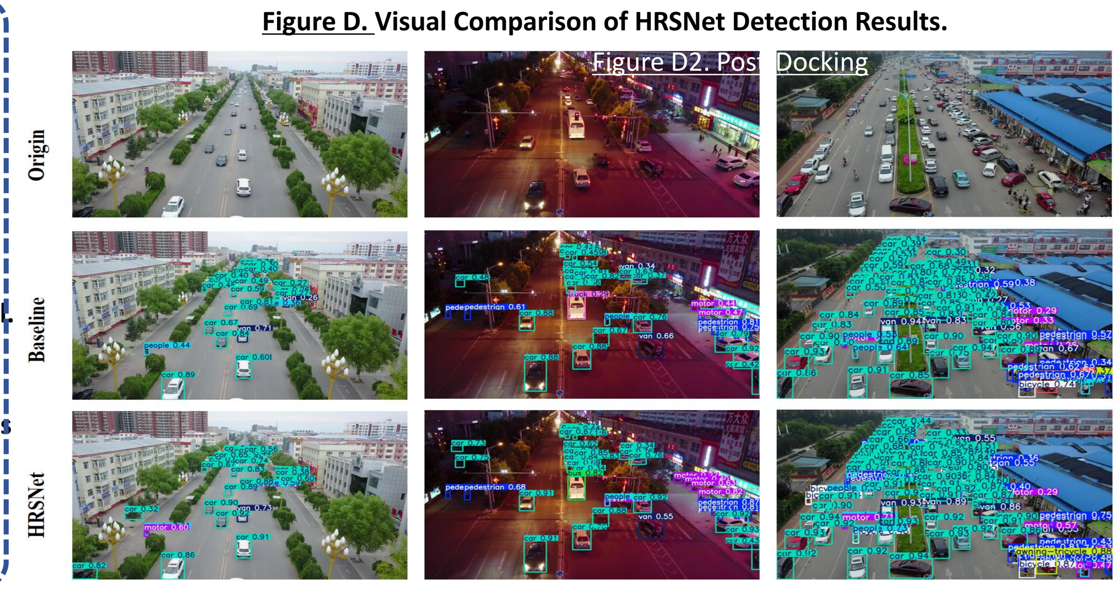
D

To demonstrate HRSNet's effectiveness in small object detection for industrial applications, we present visual comparisons with the YOLOv10s baseline. Figure D illustrates HRSNet's superior capability in identifying and localizing challenging small defects such as micro scratches, solder balls, and component misalignments that are critical in precision manufacturing.

The comparison reveals several key advantages of HRSNet. While the baseline YOLOv10s struggles with missed detections and inaccurate localization of sub millimeter defects, HRSNet consistently achieves reliable detection and precise bounding box regression. This performance improvement is particularly evident in cases of clustered defects and low-contrast anomalies where traditional detectors typically fail.

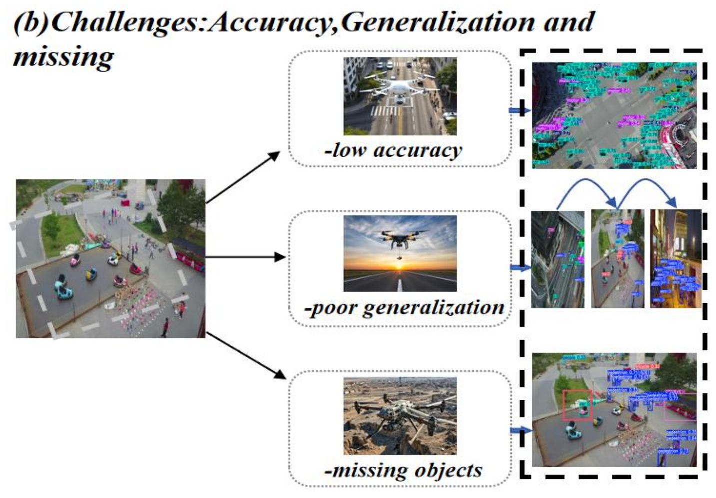
These visual results confirm HRSNet's enhanced small object detection capability through its improved feature representation and multi-scale processing. The network demonstrates robust performance across various challenging industrial scenarios, making it particularly suitable for high-precision quality control applications in electronics manufacturing where detecting microscopic defects is essential for ensuring product reliability. The superior detection accuracy combined with maintained computational efficiency positions HRSNet as a practical solution for automated visual inspection systems.

Figure D. Visual Comparison of HRSNet Detection Results.



6. Conclusions

E



- **HRSNet is an advanced UAV detection framework that tackles small object challenges through three innovations: FPSCM for multi-scale fusion, DAAM for irregular object modeling, and Dynamic Sparse Convolution for efficient feature learning. On VisDrone benchmark, it achieves superior accuracy in dense scenarios while maintaining real-time performance. Future work will explore deformable attention for enhanced ultra-dense detection.**

7. Discussion

- **The proposed HRSnet effectively addresses the challenges of small-object localization in UAV imagery through its optimized three-module architecture.**
- **By integrating multi-scale feature extraction with efficient spatial attention, the network achieves high detection accuracy while remaining computationally lightweight. Experimental results demonstrate that HRSnet outperforms existing methods in handling cluttered backgrounds and scale variations, proving its robustness in real-world aerial imaging scenarios.**
- **The streamlined design ensures practical deployment even on resource-constrained platforms, making it a viable solution for UAV-based applications.**

8. Future Direction

- **Future research could focus on enhancing HRSNet's generalization ability across diverse UAV platforms and imaging conditions.**
- **Exploring self-supervised or weakly supervised learning strategies may reduce the reliance on large annotated datasets. Additionally, integrating real-time adaptive mechanisms for dynamic environments could further improve the model's robustness in practical deployments.**

9. References

1. **Y. Wang et al.**, *"UAV-EnERF: Text-Driven UAV Scene Editing with Neural Radiance Fields"*, **IEEE Transactions on Geoscience and Remote Sensing**, 2024.
2. **X. Hu et al.**, *"SPNet: Spectral Patching End-to-End Classification Network for UAV-Borne Hyperspectral Imagery"*, **IEEE Transactions on Geoscience and Remote Sensing**, 2022.
3. **J. Xu et al.**, *"YOLOOW: A Spatial Scale Adaptive Real-Time Object Detection Neural Network for Open Water Search and Rescue from UAV Aerial Imagery"*, **IEEE Transactions on Geoscience and Remote Sensing**, 2024.
4. **Y. Guo et al.**, *"Crater-DETR: A Novel Transformer Network for Crater Detection Based on Dense Supervision and Multiscale Fusion"*, **IEEE Transactions on Geoscience and Remote Sensing**, 2024.
5. **Y. Hui et al.**, *"WSA-YOLO: Weak-Supervised and Adaptive Object Detection in Low-Light Environment for YOLOv7"*, **IEEE Transactions on Instrumentation and Measurement**, 2024.
6. **H. Shen et al.**, *"Object Detection Deployed on UAVs for Oblique Images by Fusing IMU Information"*, **IEEE Geoscience and Remote Sensing Letters**, 2022.
7. **H.-I. Liu et al.**, *"A Denoising FPN with Transformer R-CNN for Tiny Object Detection"*, **IEEE Transactions on Geoscience and Remote Sensing**, 2024.
8. **Y. Zhang et al.**, *"FFCA-YOLO for Small Object Detection in Remote Sensing Images"*, **IEEE Transactions on Geoscience and Remote Sensing**, 2024.
9. **T. Zhao et al.**, *"Scene-YOLO: A One-Stage Remote Sensing Object Detection Network with Scene Supervision"*, **IEEE Transactions on Geoscience and Remote Sensing**, 2025.
10. **P. Zhu et al.**, *"Detection and Tracking Meet Drones Challenge"*, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 2021.

10. Acknowledgements

We sincerely thank Dr. Tianxin Han and Dr. Qing Dong for their invaluable guidance and support throughout this research. Special gratitude goes to Dr. Gang Wu for his expert insights and critical contributions to the development of the HRSnet framework, particularly in optimizing its hierarchical recursive scaling architecture. We also acknowledge the broader research team for their collaborative efforts in advancing this work.