Neuromorphic Drone Detection: an Event-RGB Multimodal Approach

Gabriele Magrini[1], Federico Becattini[2], Pietro Pala[1], Alberto Del Bimbo[1], Antonio Porta[3]

University of Florence – Media Integration and Communication Center (MICC)

[1][name.surname]@unifi.it [2][name.surname]@unisi.it [3][name.surname]@leonardo.com

🎯 Target

- **Drone Detection**: Struggles with lighting and fast-moving objects in RGB data.
- **Event-based Strengths**: Can captures rapid scene changes in varying lighting conditions.
- **Lack of Dataset**: Scarcity of both RGB-EV and Drone Datasets.

💡 **Multimodal Model**: Integrate both modalities for robust detection.

🔑 Key contributions

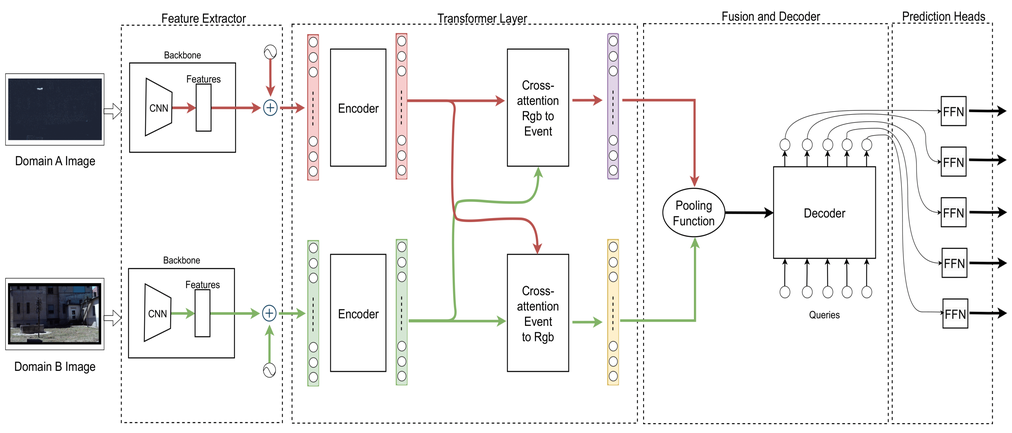
1. **Multimodal Architecture: Integrates Event-based and RGB data** for drone detection.
2. **NeRDD Dataset**: novel dataset with over **3.5 hours of spatio-temporally synchronized Event-RGB recordings**.
3. **Performance Improvements**: Demonstrates the **effectiveness of combining both modalities** for improved accuracy.

🏠 Hybrid Architectures

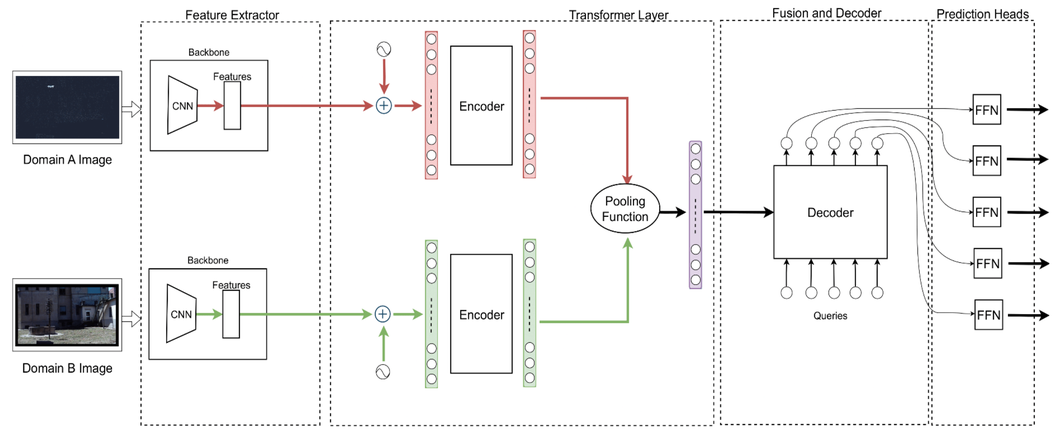
Methodology: The proposed model is based on the DETR[1] architecture. We implement diverse fusion strategies:

❑ **Asymmetric Fusion**: A **unidirectional approach** that informs one "main" modality using information form the other domain.

❑ **Symmetric Fusion**: A **bidirectional approach** that informs both modalities about each other.



❑ **Pooling-Based Fusion: Combines features from both modalities** at an intermediate layer.



🗂️ NeRDD Dataset

- **Released NeRDD dataset** comprising over 3.5 hours of **synchronized Event-RGB recordings**, manually annotated with drone bounding boxes.

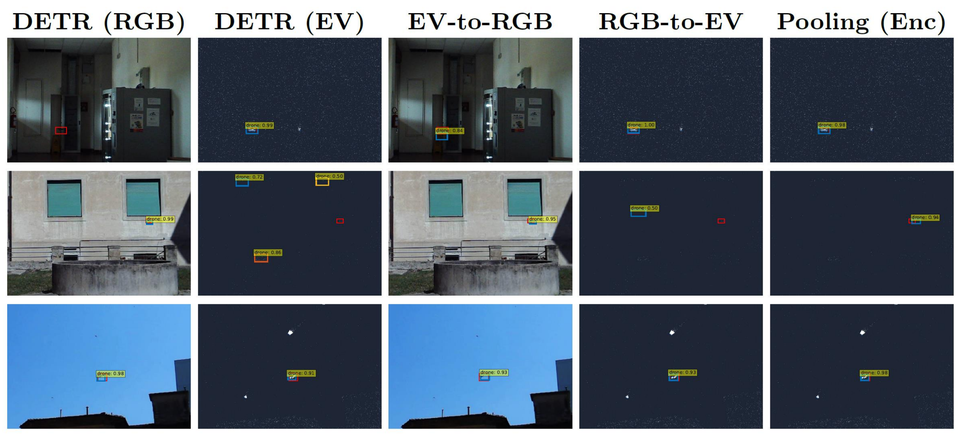
Dataset	Resolution	RGB/Event	Hours	Drone-Centric
VisEvent [2]	346 x 260	✓/✓	<5	×
EventVOT [3]	1280 x 720 (HD)	×/✓	<5	×
F-UAV-D [4]	1280 x 720 (HD)	✓/✓	0.5	✓
NeRDD(Ours)	1280 x 720 (HD)	✓/✓	3.5	✓

📊 Results

Model	AP50	AP50:95	AP75	Parameters
DETR Event	80.5	34.8	21.6	41.302.368
DETR RGB	32.7	9.1	2.0	41.302.368
Asymmetric RGB-to-EV	84.4	39.0	**27.4**	60.746.247
Asymmetric EV-to-RGB	40.8	13.0	3.8	60.746.247
Symmetric Fusion	80.9	33.6	18.7	90.869.255
Pooling (Encoder)	**85.2**	**39.3**	27.2	59.166.983

Pooling	AP50	AP50:95	AP75
Encoder	**85.2**	**39.3**	**27.2**
ResNet	84.7	35.2	17.9

Symm.	AP50	AP50:95	AP75
Encoder	**80.9**	**33.6**	**18.7**
Decoder	79.9	33.4	17.9



✅ Conclusions

- Proposed a **multimodal Event-RGB Architecture** for Drone Detection.
- Presented a novel **hybrid Event-RGB Spatio-Temporally synchronized drone dataset**.

Event-RGB integration strengthens both modalities

📝 References

[1] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.:End-to-end object detection with transformers. CoRR abs/2005.12872 (2020),https://arxiv.org/abs/2005.12872

[2] X. Wang, J. Li, L. Zhu, Z. Zhang, Z. Chen, X. Li, Y. Wang, Y. Tian, and F. Wu,"Visevent: Reliable object tracking via collaboration of frame and event flows," IEEETransactions on Cybernetics, 2023.

[3] X. Wang, S. Wang, C. Tang, L. Zhu, B. Jiang, Y. Tian, and J. Tang, "Event stream-based visual object tracking: A high-resolution benchmark dataset and a novel base-line," in Proceedings of the IEEE/CVF Conference on Computer Vision and PatternRecognition, pp. 19248–19257, 2024.

[4] J. Mandula, J. Kühne, L. Pascarella, and M. Magno, "Towards real-time fastunmanned aerial vehicle detection using dynamic vision sensors," arXiv preprintarXiv:2403.11875, 2024.