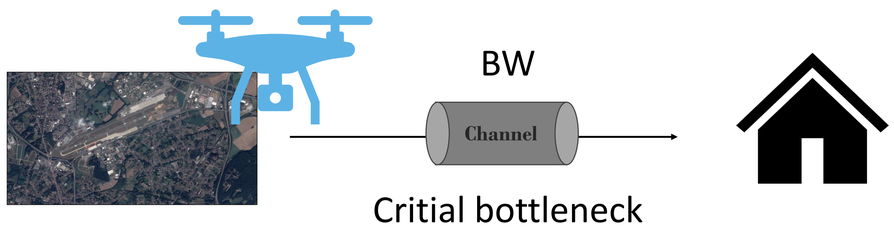
Streamlined Hybrid Annotation Framework using Scalable Codestream for Bandwith-Restriced UAV Object Detection

Paper ID: 2085

IEEE International Conference on Image Processing
27 - 30 October 2024 | Abu Dhabi, UAE

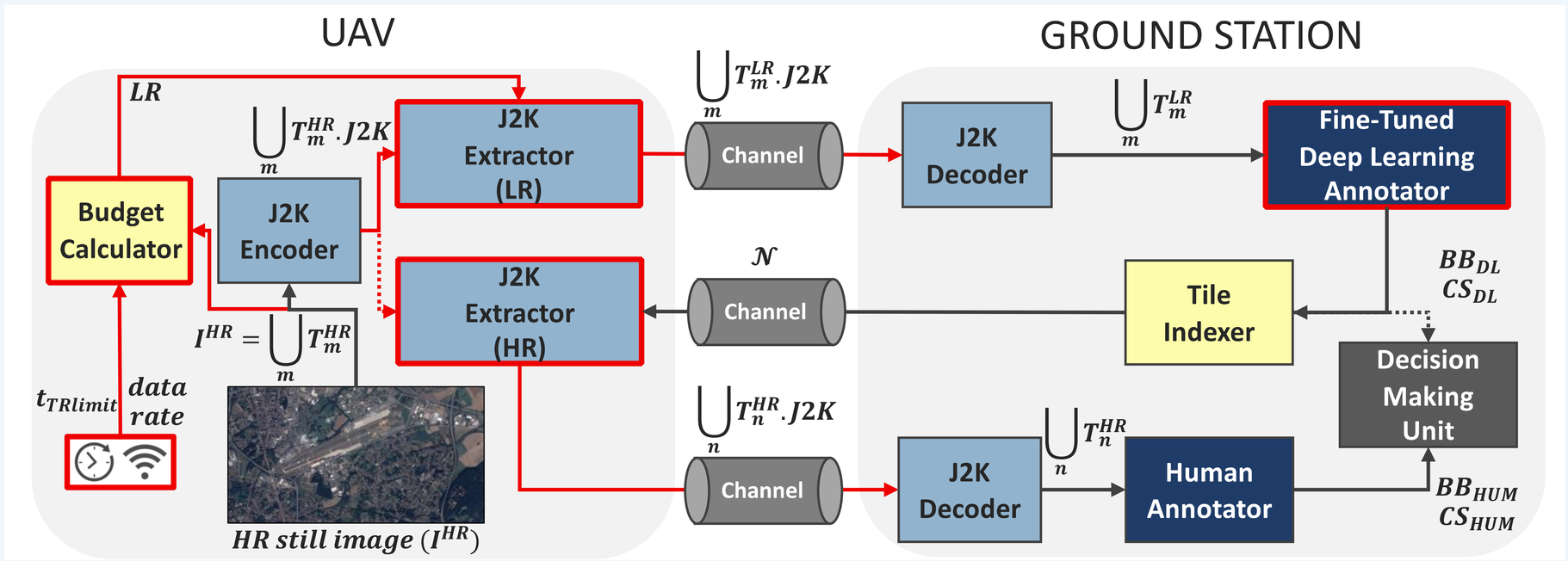
Karim El Khoury*, Tiffanie Godelaine*, Simon Delvaux, Sébastien Lugan, and Benoît Macq
ICTEAM, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

1. CONTEXT: UAVs are essential in emergency scenarios



2 challenges for decision-making:
Fast transmission and **annotation**
→ Use of **multiresolution** and **hydrid annotation** to allow **fast** and **accurate** decisions while **saving expert time**

2. METHOD: Proposed approach



3 improvements:
- **Scenario constraints are integrated** to determine the highest available resolution **LR** instead of HR (baseline)
- Only tiles selected for human annotator are sent in **HR**
- DL models are fine-tuned at different resolution levels

3. EXPERIMENTAL SETUP: Application on satellite images of airports with aircrafts

- 165 HR images (60 MP)
- Fine-tuned Yolov8 on 5 resolutions levels
- BW fixed : 176, 88 and 22 kbps
- Level of emergency fixed, determined by $t_{RSlimit}$: 3, 10 and 30 min
- 2 metrics to compare the proposed approach to the baseline : t_{RS}_*ratio* and *recall_diff*

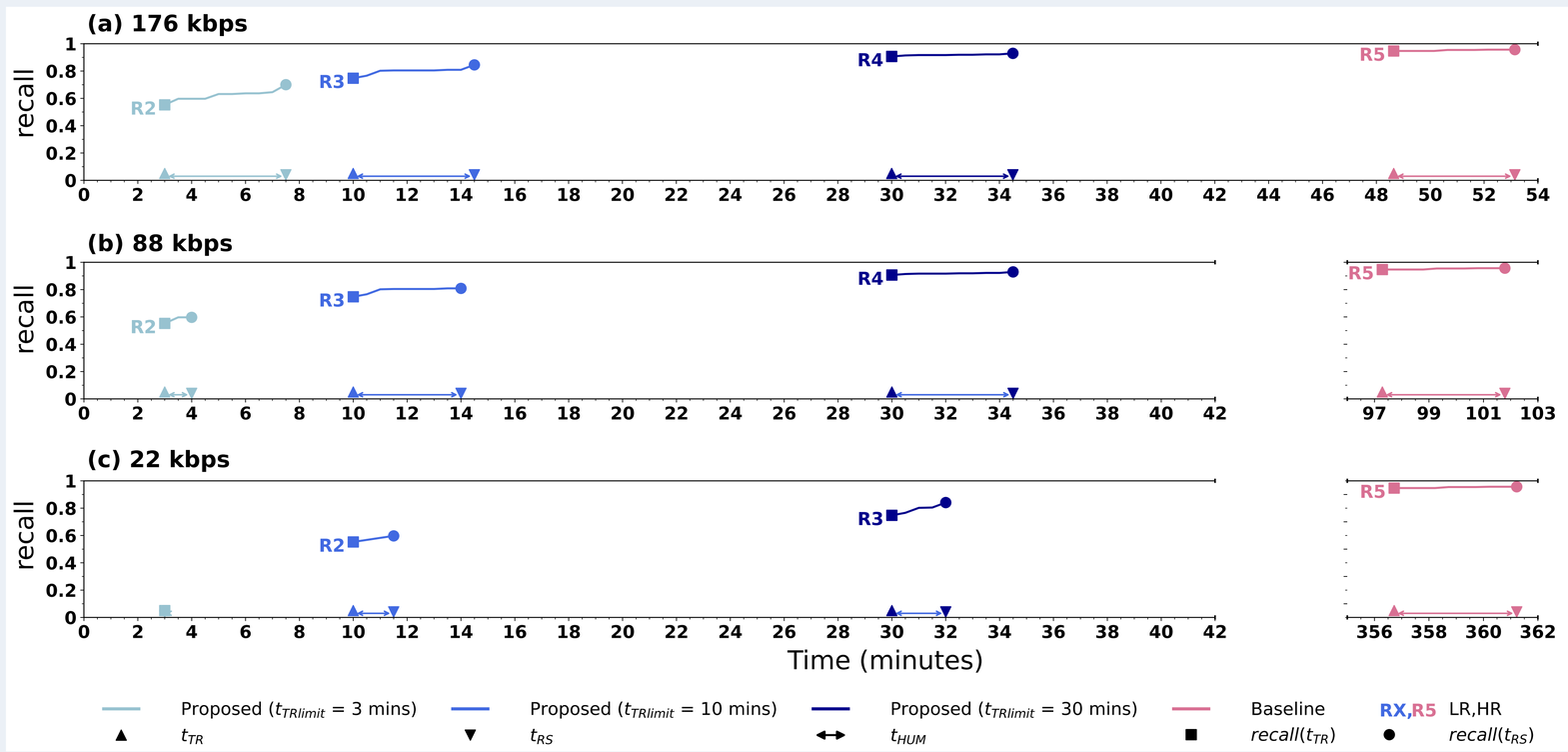
	3	10	30
176	0.257	0.112	0.028
	7.08	3.67	1.54
88	0.35	0.148	0.028
	25.44	7.27	2.95
22	/	0.359	0.115
	/	34.4	11.28

4. RESULTS: Response time is reduced by a factor of 34

176 kpbs:
Proposed transmits LR tiles LR
→ Improvement of t_{RS}_*ratio*
Baseline always transmits HR tiles → Penalty in *recall_diff* because performance of fine-tuned model at HR is better

88 kpbs:
Higher improvement of t_{RS}_*ratio* and lower decrease of *recall_diff* because budget for human annotation is lower

22 kbps:
Highest gain in t_{RS} . However, at 3min, even the lowest LR cannot be sent because the BW is too restrictive



Key References

[1] Preethy Byju, *et al.*, "Remote-sensing image scene classification with deep neural networks in jpeg 2000 compressed domain," IEEE Transactions on Geoscience and Remote Sensing,vol. 59, no. 4, pp. 3458–3472, 2021.

[2] El Khoury, *et al.*, "Improved 3d u-net robustness against jpeg 2000 compression for male pelvic organ segmentation in radiotherapy," Journal of Medical Imaging, vol. 8, no. 04, 2021.

[3] El Khoury, *et al.*, "Improving 3d lesion segmentation robustness against image compression in multiple sclerosis," in 2024 IEEE International Symposium on Biomedical Imaging, 2024, p. 1.

[4] Yamani, *et al.*, "Active learning for single-stage object detection in uav images," in Proceedings of the IEEE/CVFWinter Conference on Applications of Computer Vision (WACV), January 2024, pp. 1860–1869.

[5] Palmaerts, *et al.*, "Oriented aircraft object detector using scaled yolov4 on very high resolution satellite and synthetic datasets," in 2023 Joint Urban Remote Sensing Event (JURSE), 2023, pp. 1–4.

5. CONCLUSION

The combination of fine-tuned DL model and human annotation with multiresolution allows to speed up the decision-making process in BW limited scenarios, up to 34 times in the most restrictive case
FUTURE WORKS will focus on optimizing each component of the framework (fine-tuning of the model and tile selection strategy for human annotation)

The Pléiades images were obtained via the Pléiades 4 Belgium platform under an agreement with BELSPO and labeled by ISSeP.
T. Godelaine (tiffanie.godelaine@uclouvain.be) is supported by MedReSyst, funded by the Walloon Region and the EU Wallonie.
*Denotes equal contribution