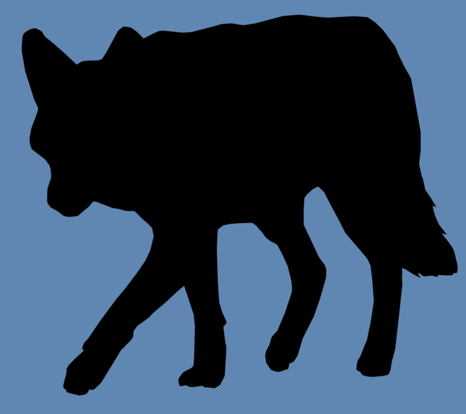
Wildlife Camera-Trap Classification

Şükrü Deniz Çilek
University College Dublin

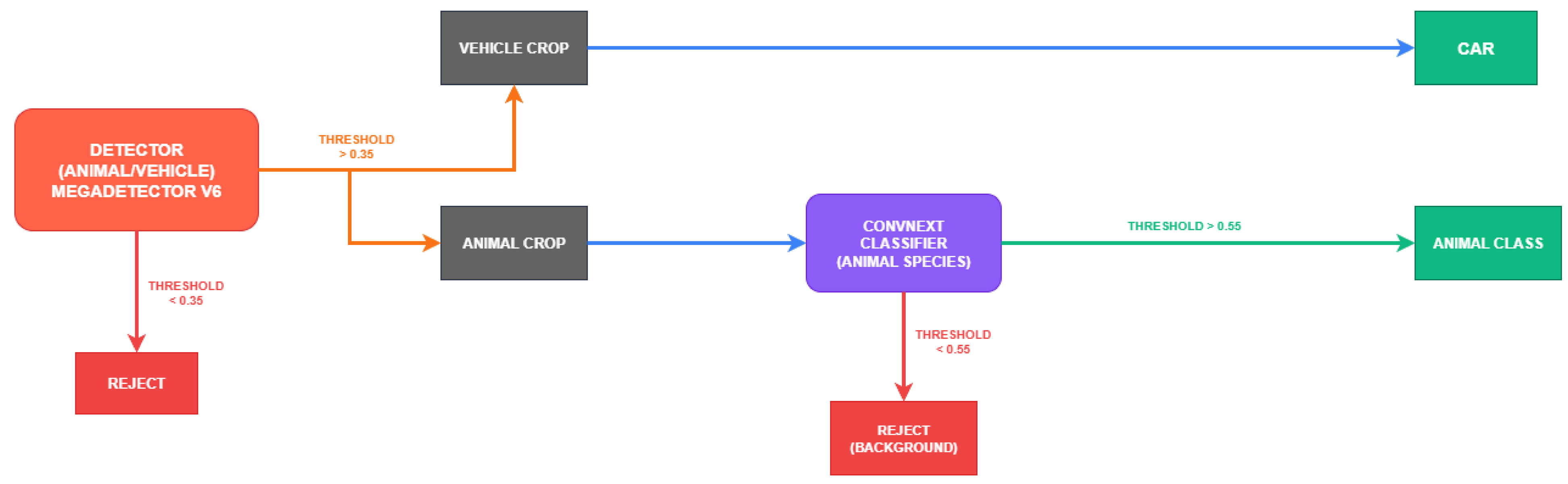


Abstract

Camera-trap imagery enables large-scale wildlife monitoring but suffers from domain shift, where models fail on unseen locations. We propose a two-stage pipeline (MegaDetector v6 + ConvNeXt-Small) for detection and classification, **achieving F1 scores of 0.84 (seen) and 0.71 (unseen)** on CCT20. Our modular design **cuts cross-domain error by 24.2%** while maintaining high recall for rare species, enabling robust and scalable conservation monitoring.

Problem Definition and Contribution

Problem: Camera trap images present unique difficulties, night-time infrared capture, motion blur, partial occlusion, weather conditions, and class imbalance. Models struggle to generalize across different camera locations due to site-specific environmental variations.



Two-Stage Pipeline: MegaDetector v6 → ConvNeXt Classifier

Key Innovation: Modular design with threshold-based decision points (0.35 detection, 0.55 classification) enables independent optimization and robust cross-domain deployment.

Dataset & Challenges

CCT20 Dataset: 51,000+ images, 20 locations, 13 species + vehicle

Main Challenges:

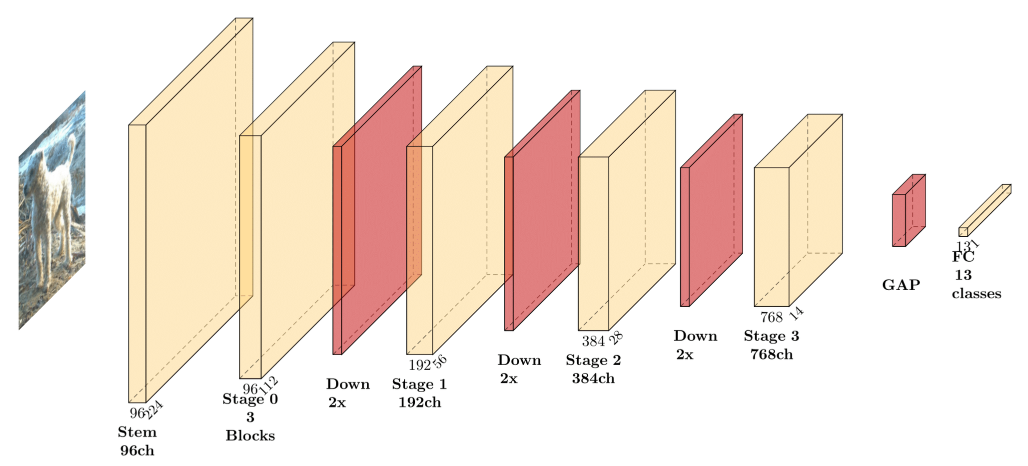
- Domain shift: performance drops on new locations
- Class imbalance: some species are very rare
- Poor conditions: night vision, motion blur, occlusion
- Environmental noise: rain, fog, camera malfunctions

Seen (train) — Bobcat, daylight, clear background.

Unseen (test) — Coyote, night IR, motion blur.

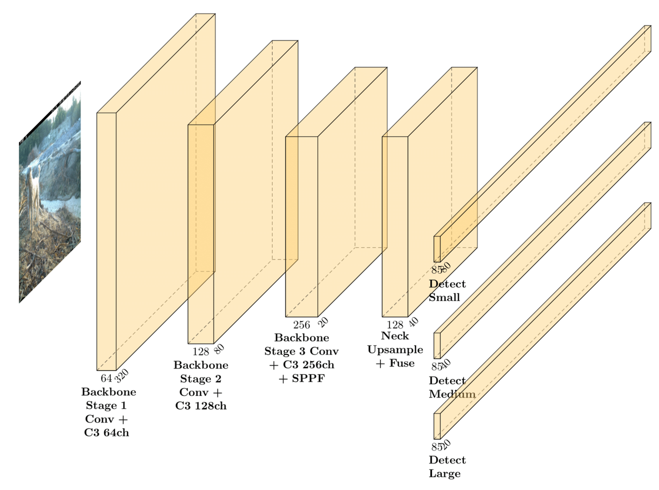
Method

ConvNeXt-Small



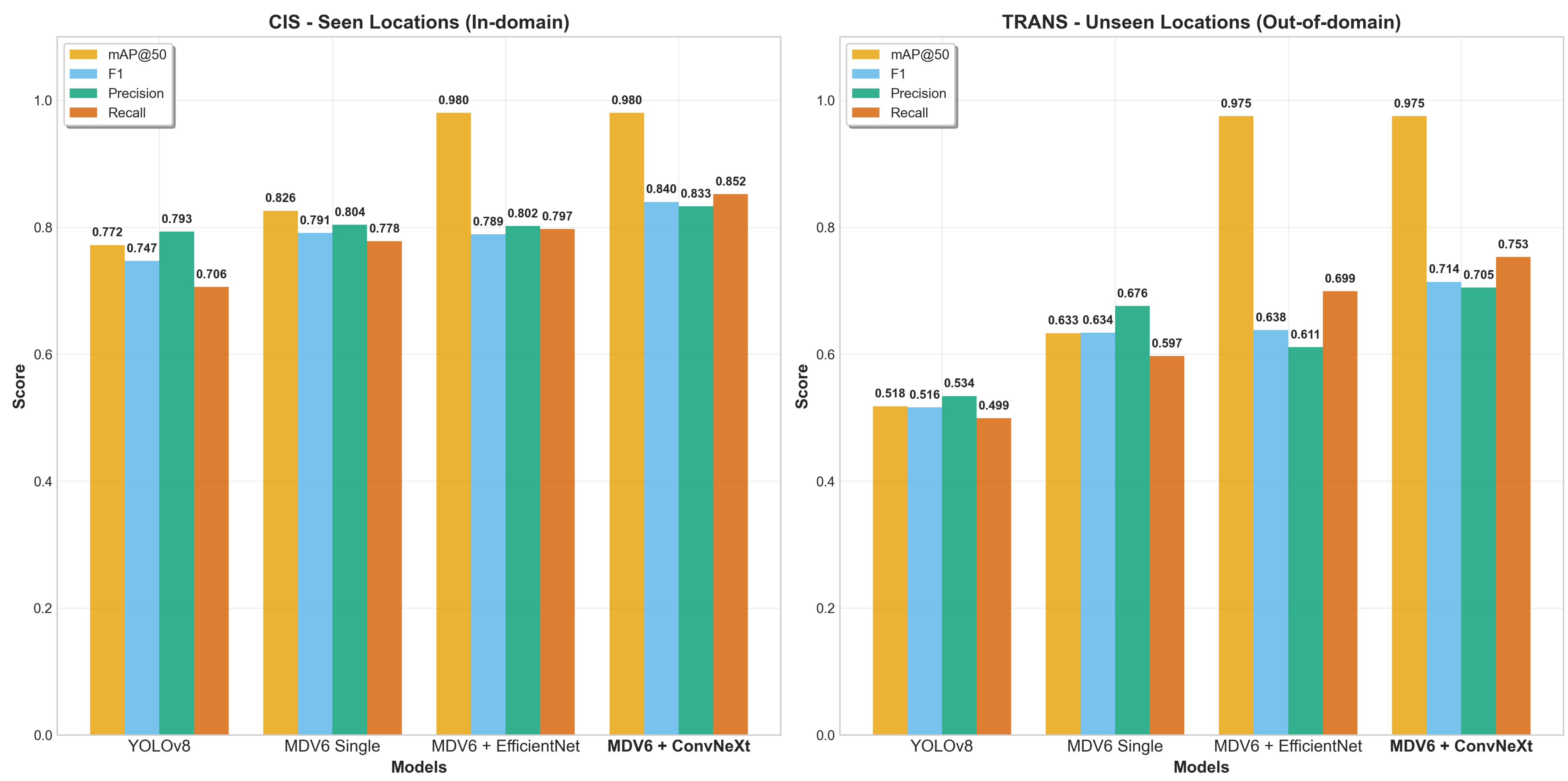
- Efficient CNN with depthwise 7×7 convolutions and MLP blocks.
- Large receptive field for robustness to domain shift.

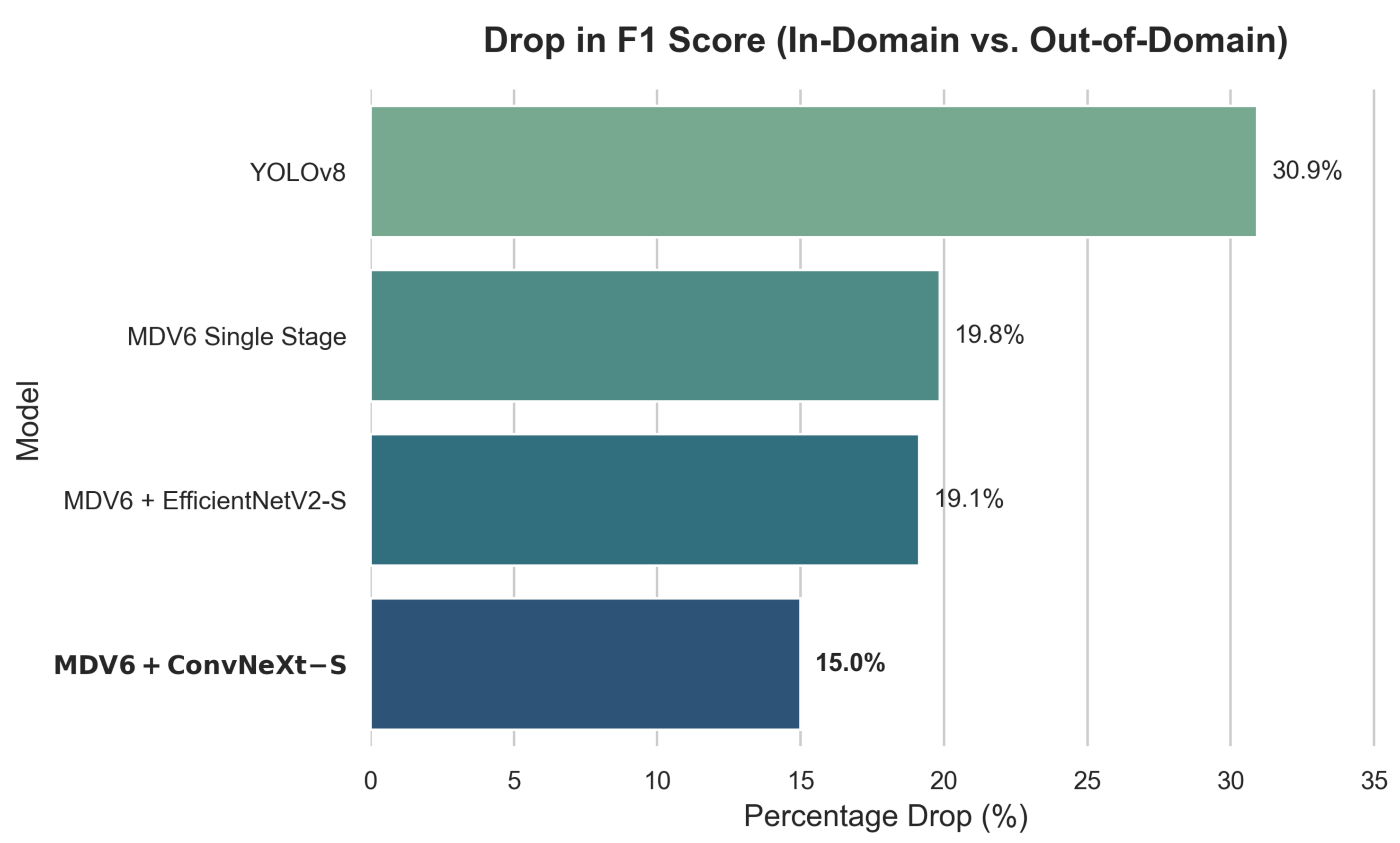
MegaDetector v6 (YOLOv9-c)



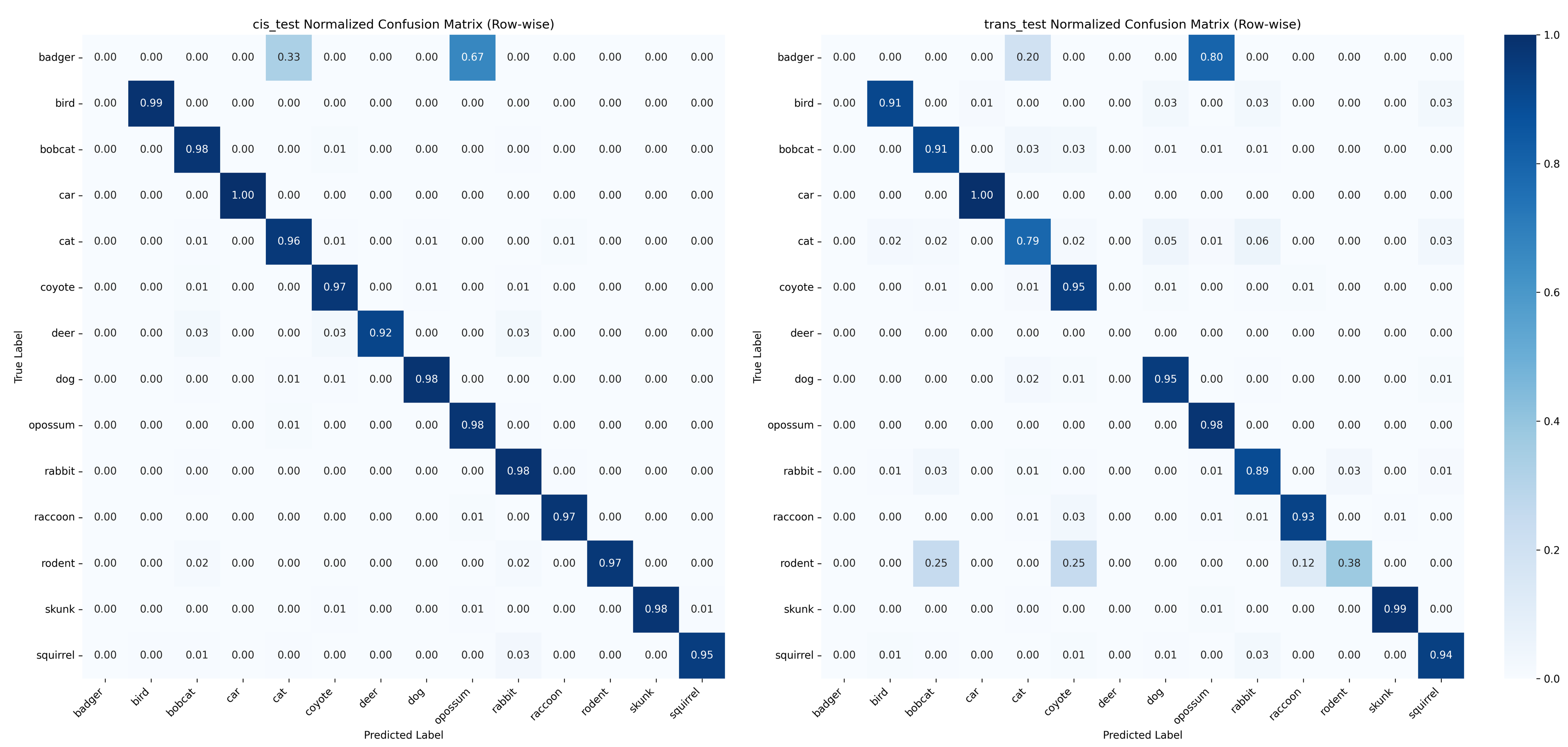
- YOLO-style detector with FPN, SPPF, and 3-scale heads.
- Trained for binary animal vs. vehicle detection with high recall.

Experiments & Results





Detector (MDV6) achieves high accuracy in separating animals from vehicles: CIS – 97.9% precision, 97.3% recall (F1 = 97.6%); TRANS – 96.4% precision, 95.6% recall (F1 = 96.0%).



CIS (seen) — row-normalized **TRANS (unseen)** — row-normalized

Future Work & Conclusion

- Two-stage MDV6 + ConvNeXt-Small reduces domain shift and maintains high precision/recall on unseen locations.
- Enhance rare-class detection via active learning on hard negatives & tail species.
- Leverage photo metadata (e.g., time of day, location) for context-aware classification.
- Experiment on the full CCT dataset for broader generalization.

Web App **GitHub Repository**

References

1. Beery, S., et al. (2018). Recognition in Terra Incognita. *ECCV*.
2. Norouzzadeh, M. S., et al. (2018). Auto ID of wild animals. *PNAS*.
3. Microsoft AI for Earth (2023). MegaDetector v6. GitHub.
4. Liu, Z., et al. (2022). ConvNeXt. *CVPR*.