Wildlife Camera-Trap Species Classification & Domain-Shift Study

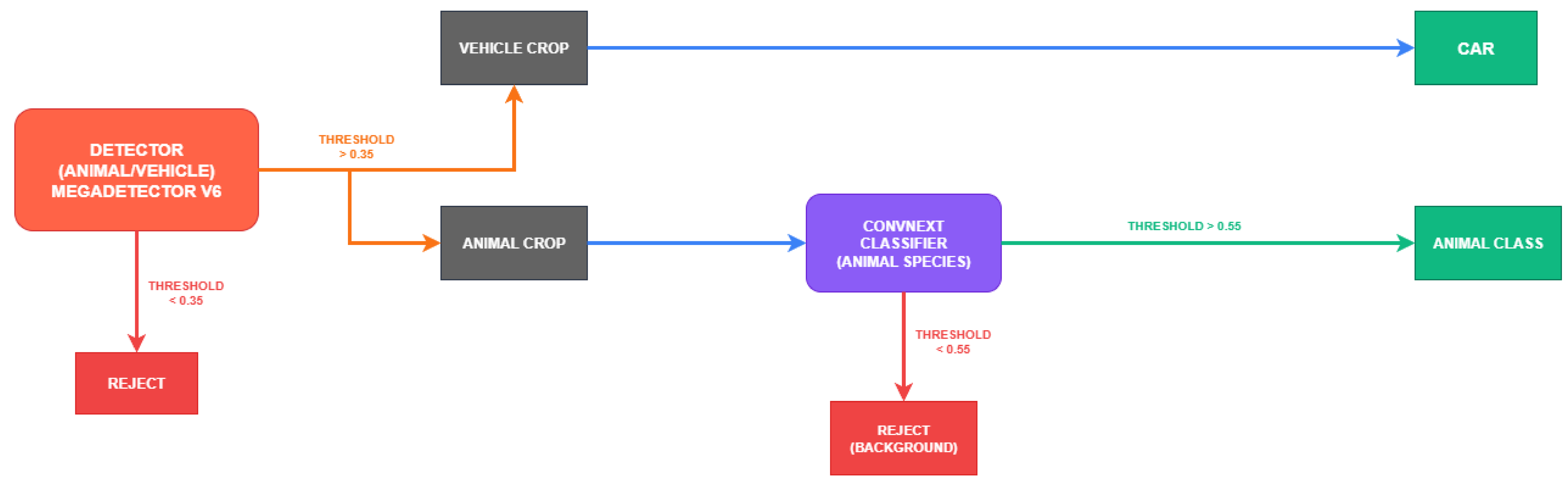
Şükrü Deniz Çilek
University College Dublin

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.
Impact: 24.2% error reduction on unseen locations while maintaining high recall for rare species.

Dataset & Challenges

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

Main Challenges:

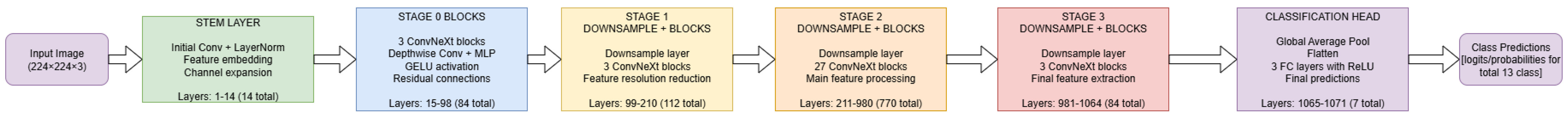
- Domain shift: performance drops on new locations
- Class imbalance: some species are very rare
- Poor conditions: night vision, motion blur, occlusion
- Environmental noise: rain, fog, camera malfunctions

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

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

Method

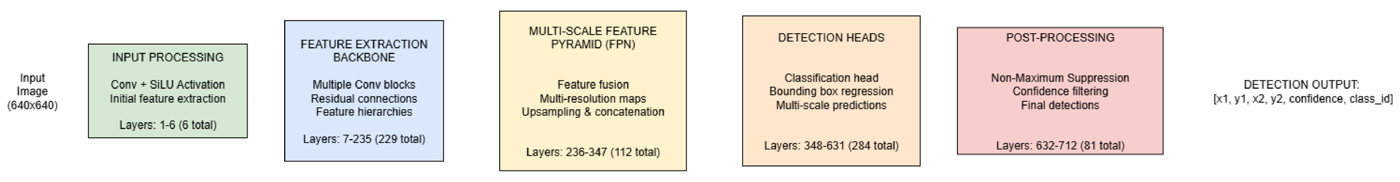
Classifier: ConvNeXt-Small



ConvNeXt-Small: stem 4×4/4 → depthwise 7×7 blocks → GAP → FC (13-way).

ConvNeXt-Small is a modern CNN with depthwise 7×7 **convolutions**, GELU activation, and MLP blocks with residual connections. A 4×4 patchify stem reduces resolution early for efficiency while preserving detail. **Large receptive fields and texture bias improve robustness** to domain shift.

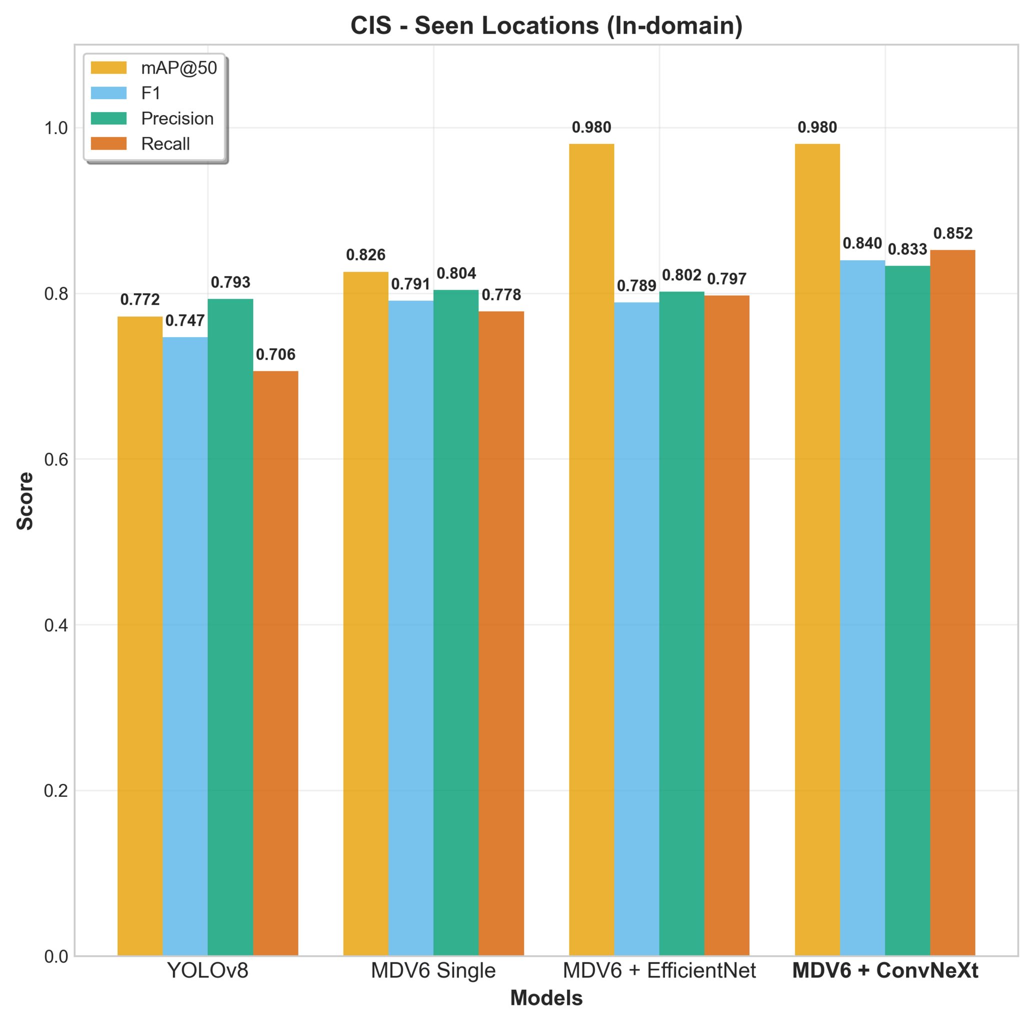
Detector: MegaDetector v6 (YOLOv9-c)

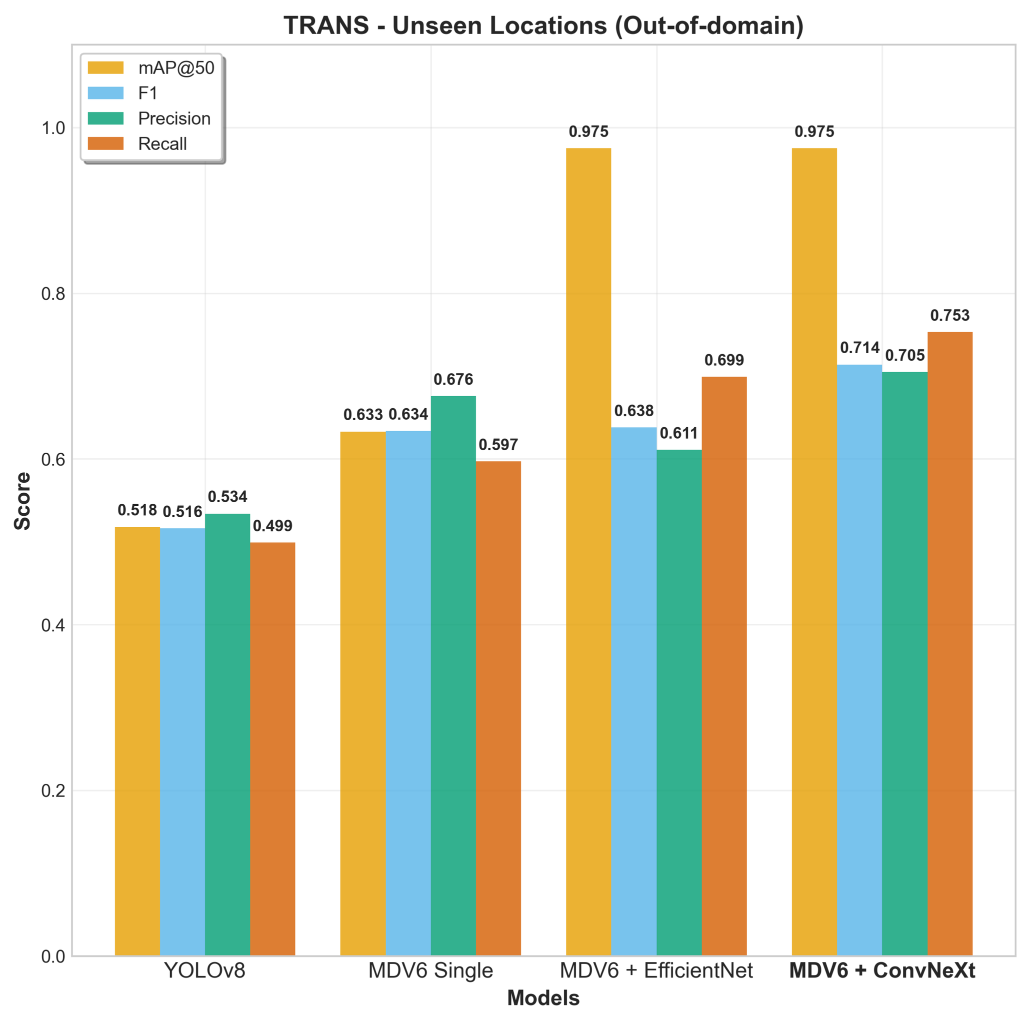


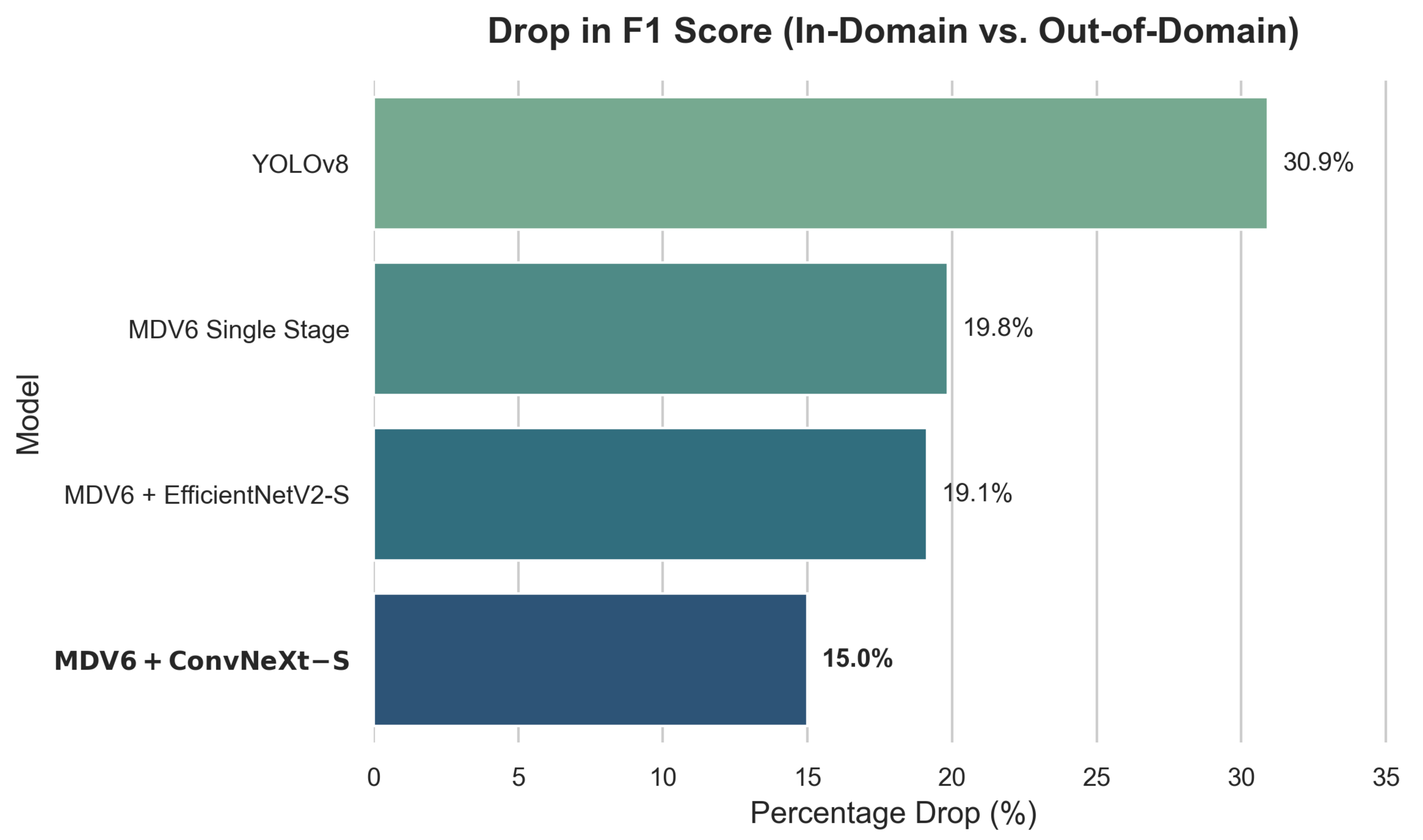
YOLO-style backbone + FPN, SPPF, and DetS/DetM/DetL heads.

MegaDetector v6 uses a YOLOv9-c backbone with multi-scale FPN, SPPF, and three detection heads for varying animal sizes. **Trained as a binary detector (animal vs. vehicle)** to maximize recall on animals while keeping high precision. Exported to ONNX for efficient batch processing in large-scale inference.

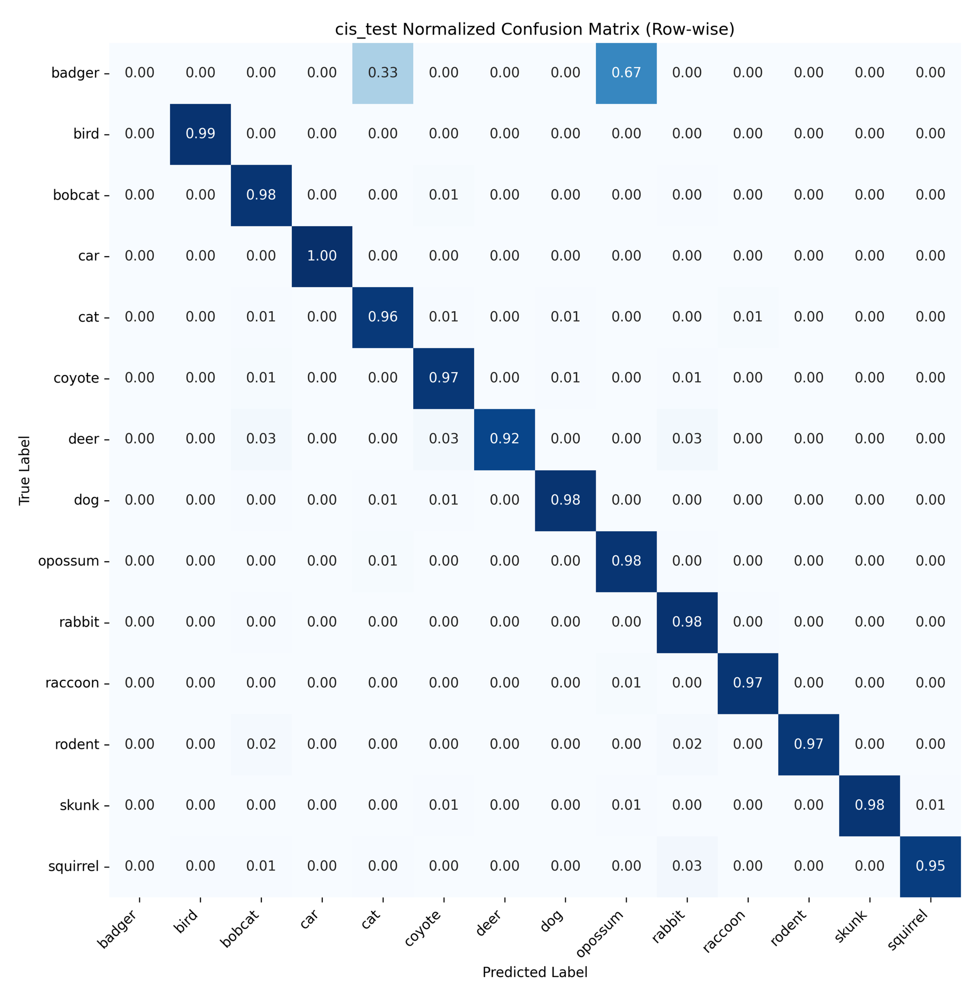
Experiments & Results







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 QR Code

GitHub QR Code

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*.