Night-time Wildlife Detection using YOLOv8 with Attention Modules and Custom Head

Ting-Yu Hsu - 24207795
Yanjie Shao - 24214602
ACM40960 Projects in Maths Modelling, MSc Data & Computational Science
University College Dublin

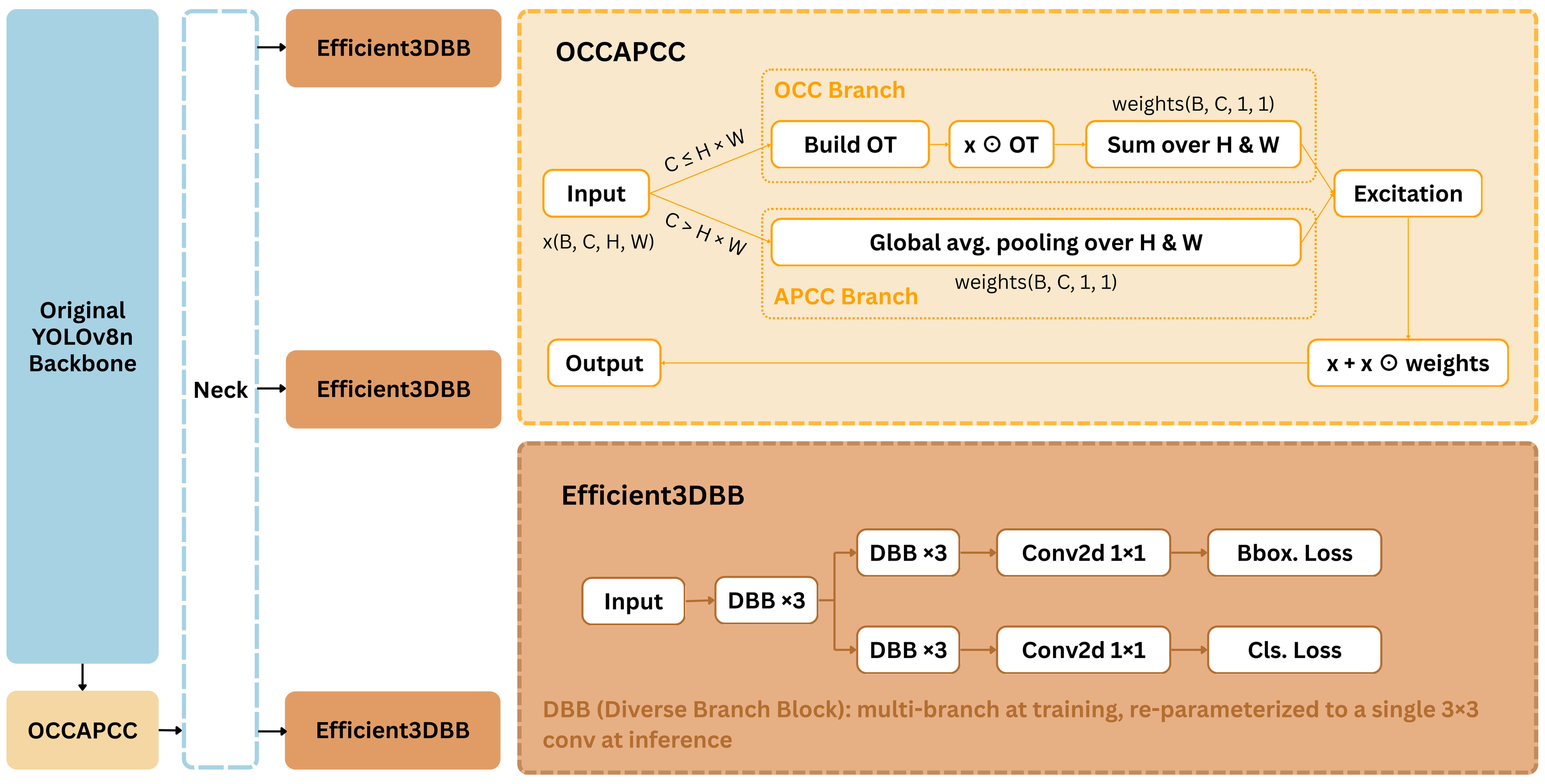
Introduction

Accurate identification of wildlife in low-light or infrared imagery is essential for ecological research and conservation, yet standard object detection models such as YOLO often experience performance degradation in nighttime conditions. Complex backgrounds, occlusion, and variations in animal scale further challenge detection accuracy.

This project enhances the YOLOv8 framework for nighttime wildlife detection by integrating a novel attention mechanism (**OCCAPCC**) from YOLOv8-night and a custom detection head (**Efficient3DBB**) from YOLO-MIF. Multiple YOLOv8-based variants were developed - exploring different attention modules (OCCAPCC, CBAM), insertion positions, and combinations with Efficient3DBB - trained and evaluated under identical settings on the NTLNP nighttime dataset. Results indicate that these architectural modifications can improve detection for certain species (e.g., **BlackBear**, **WildBoar**), but gains are not uniform across all classes. Efficient3DBB helps recover some performance loss when paired with OCCAPCC, though overall metrics remain close to or slightly below the YOLOv8 baseline. For reference, YOLOv11n was also tested as an additional baseline for comparison.

Methodology

Dataset. NTLNP *voc_night* subset - **10,344** infrared images of multiple animal species captured by night-vision camera traps (reflecting real-world challenges such as low contrast, background clutter, and partial occlusion).
Data Preprocessing. Converted **VOC-style** annotations to **YOLO format** with normalized coordinates, filtered for **17** target species, and **stratified** into train/val/test splits (**70% / 10% / 20%**) with an organized directory structure for YOLO training.



Model Architecture. Our approach builds on the YOLOv8n backbone with two key modifications:
OCCAPCC Attention Module: Enhances feature focus on animal regions while suppressing background noise, inserted at different backbone positions for comparison.
Efficient3DBB Detection Head: Uses Diverse Branch Blocks (DBB) during training, re-parameterized into efficient convolutions for inference, aiming to improve bounding box regression and classification in low-light conditions.

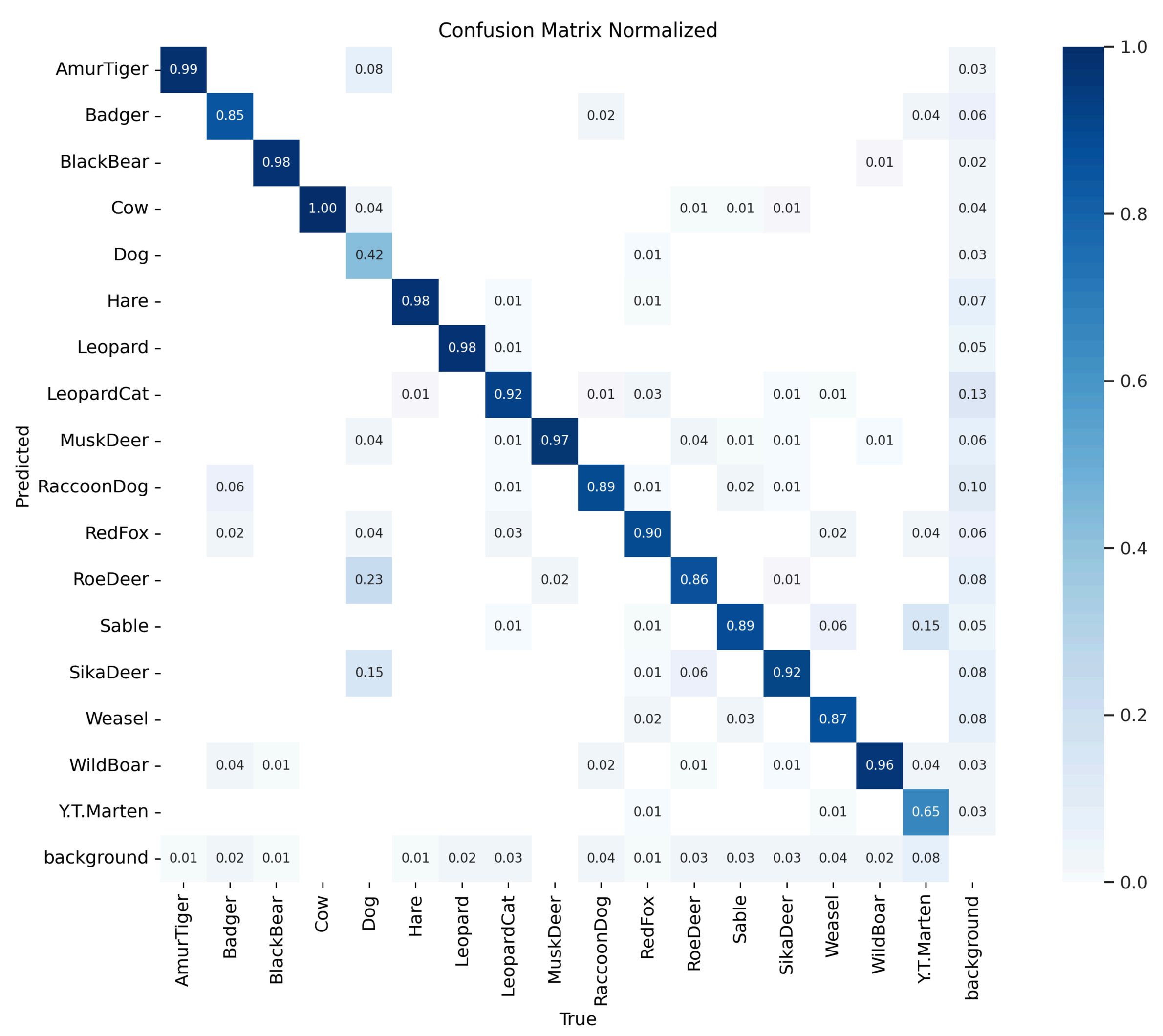
Training & Evaluation. We implemented and compared **multiple YOLOv8-based variants (baseline, OCCAPCC, CBAM, OCCAPCC+Efficient3DBB, CBAM+Efficient3DBB)** and most advanced YOLOv11n model under identical training settings. Models were evaluated using **mAP, precision, recall, and PR curves**, **focusing on detection performance across species**.

Results and Discussion

Detection Example (BlackBear, Nighttime)

(a) YOLOv8 baseline (b) YOLOv8+OCCAPCC (c) YOLOv8+OCCAPCC+Efficient3DBB

Confusion Matrix – YOLOv8n + OCCAPCC + Efficient3DBB



Model Performance Comparison (All models)

Model Variant	mAP@0.5	mAP@0.5:0.95	Precision	Recall
Baseline YOLOv8n	0.9690	0.8478	0.9536	0.9090
Baseline YOLOv11n	0.9786	0.8621	0.9609	0.9537
YOLOv8n + OCCAPCC (end)	0.9194	0.7476	0.9074	0.8515
YOLOv8n + OCCAPCC (index 8)	0.9174	0.7500	0.8968	0.8334
YOLOv8n + CBAM	0.9405	0.7669	0.8842	0.8897
YOLOv8n + OCCAPCC + Efficient3DBB	0.9336	0.7657	0.9131	0.8641
YOLOv8n + CBAM + Efficient3DBB	0.9416	0.7739	0.9271	0.8931

Detection Example (BlackBear):
- **Baseline** produces **duplicate boxes** with lower confidence (0.26 / 0.64).
- **+OCCAPCC** suppresses duplicates and **raises confidence** to **0.91**.
- **+OCCAPCC+Efficient3DBB** shows a **similar bounding box placement** to +OCCAPCC, with confidence slightly lower (**0.89**), indicating that Efficient3DBB maintains the detection quality achieved by OCCAPCC without introducing notable spatial changes.

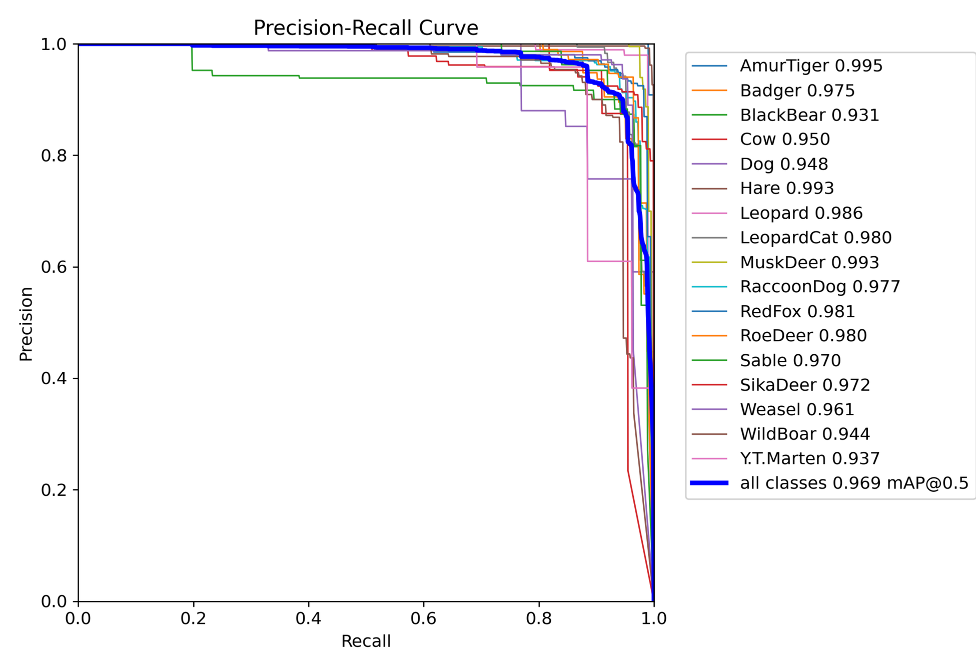
Precision–Recall Curves:
- **Improved classes**: **BlackBear** (AP@0.5 ↑ from **0.931 → 0.982/0.985**), **WildBoar** (↑), and in **+OCCAPCC** also **Cow** (↑).
- **Degraded classes**: **Dog** (↓ to **0.489**, partially recovered to **0.666** with Efficient3DBB), **Y.T. Marten** (↓ **0.937 → 0.750/0.850**), plus moderate drops for **RoeDeer, Sable, Badger, LeopardCat, etc.**
- **Most species** (e.g., **AmurTiger, Leopard, MuskDeer**) remain high in all variants with only minor fluctuations.

Confusion Matrix Analysis (Column-Normalized, +OCCAPCC+Efficient3DBB):
Interpretation: Columns = true classes; diagonal = per-class recall
- **High recall (>0.9)**: e.g. **AmurTiger, BlackBear, Leopard, MuskDeer, WildBoar** - reliably detected when present.
- **Low recall**: e.g. **Dog, Y.T.Marten** - more prone to misclassification.
- Off-diagonal entries show main misclassification targets; high "background" values = missed detections.

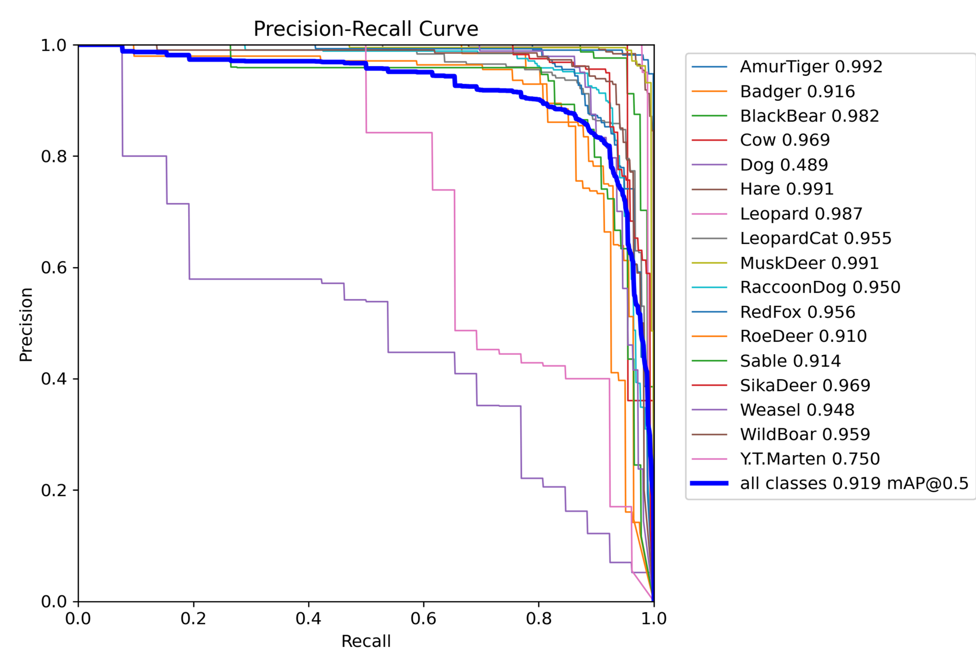
Overall performance:
On NTLNP-nighttime, the **YOLOv8n baseline** is strong (mAP@0.5 **0.969**, mAP@0.5:0.95 **0.848**, P **0.954**, R **0.909**). Adding attention and a custom head yields class-dependent changes rather than uniform gains:
- **+OCCAPCC**: mAP@0.5 **0.919**, mAP@0.5:0.95 **0.748**, P **0.907**, R **0.852**.
- **+OCCAPCC+Efficient3DBB**: partial recovery (mAP@0.5 **0.934**, mAP@0.5:0.95 **0.766**, P **0.913**, R **0.864**).

For completeness, other ablations: +CBAM mAP@0.5 **0.941**; +CBAM+Efficient3DBB mAP@0.5 **0.942**, +OCCAPCC (index 8) mAP@0.5 **0.917**.

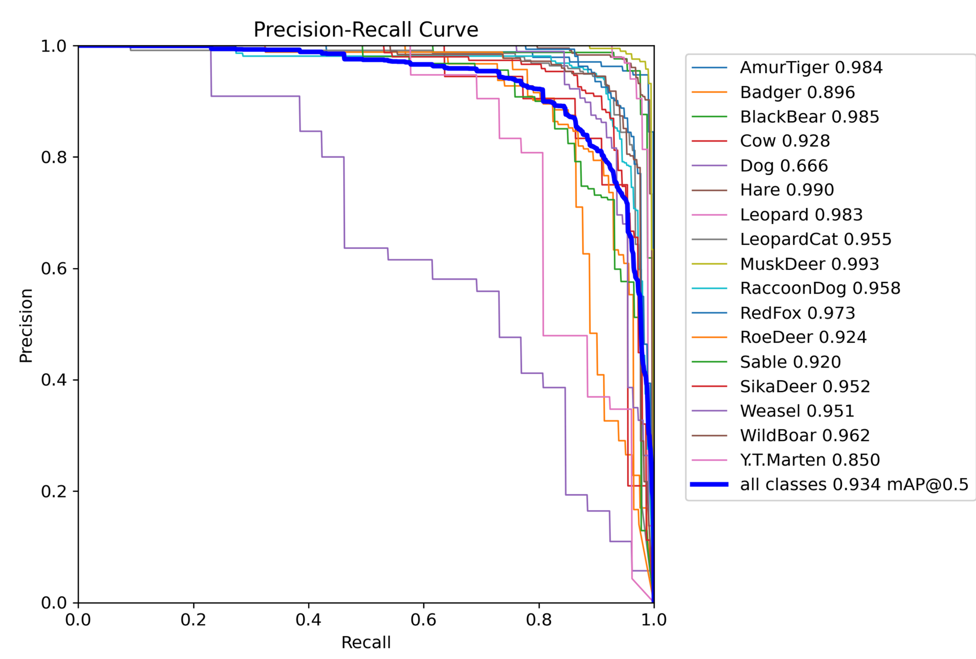
Precision–Recall Curves



(a) YOLOv8 baseline



(b) YOLOv8+OCCAPCC



(c) YOLOv8+OCCAPCC+Efficient3DBB

Conclusion

- Architectural tweaks yield class-dependent gains - notably improving detection of some species, but do not generalize uniformly across all species.
- Adding Efficient3DBB to OCCAPCC partially recovers performance but remains below the YOLOv8 baseline in overall mAP@0.5, mAP@0.5:0.95, precision and recall.
- The class-specific nature of gains suggests selective use of attention (per-class or per-feature-scale) rather than a global insertion.
- While attention modules and custom heads can enhance detection for certain species, backbone upgrades (e.g., YOLOv11n) deliver the most consistent improvement for nighttime wildlife detection. The YOLOv11n baseline achieves the best overall result (mAP@0.5 = **0.979**), maintaining high performance across nearly all species.

References

[1] NTLNP Dataset. [2] Ultralytics Document. [3] Woo, S., Park, J., Lee, J.-Y., Kweon, I. CBAM: Convolutional Block Attention Module. 2018.
[4] Wang, T., Ren, S., Zhang, H. *Nighttime wildlife object detection based on YOLOv8-night*. Electronics Letters, 60(15), 2024.
[5] Wan, D., Lu, R., Hu, B., Yin, J., Shen, S., Xu, T., Lang, X. *YOLO-MIF: Improved YOLOv8 with Multi-Information Fusion for object detection in gray-scale images.* Advanced Engineering Informatics, 62(B), 2024.

Scan the QR-code or visit our Github repo