Expert Systems With Applications 267 (2025) 126206

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Unlocking the power of artificial intelligence for pangolin protection: Revolutionizing wildlife conservation with enhanced deep learning models

Junjie Zhong [1], Suhang Wei [2], Qin Chen [*,3], Bing Niu [*,4]

School of Life Sciences, Shanghai University, Shanghai 200444, PR China

ARTICLE INFO

Keywords:
Yolov8
Target detection
Pangolin
Bifpn
Biological protection

ABSTRACT

Pangolins, the world's most trafficked wild mammals, face a dire need for immediate protection against illegal trade. Harnessing the cutting-edge advancements in artificial intelligence, particularly within deep learning, offers a beacon of hope for their survival. This groundbreaking study tackles this urgent call by curating a comprehensive dataset of pangolin images and pioneering a state-of-the-art pangolin detection model, built upon an enhanced Yolov8 architecture. Through innovative enhancements such as integrating bifpn into the network neck, incorporating the Triplet Attention mechanism into the backbone network, and leveraging the Slideloss loss function to prioritize challenging samples, our refined model surpasses the limitations of the original Yolov8s model. Notably, our improved model showcases a remarkable 13.6 % reduction in FLOP and 50.0 % in parameters, while elevating mAP to an impressive 87.0 % and shrinking model size to a mere 14.3 MB. These enhancements culminate in unparalleled recognition accuracy and model efficiency, empowering wildlife managers with superior tools for monitoring and identifying pangolins. This groundbreaking advancement not only facilitates the crucial task of pangolin conservation but also signifies a significant stride towards safeguarding our planet's precious biodiversity.

1. Introduction

Among mammal species, pangolins face a global threat characterized by a decline in populations due to illegal trade, establishing them as the most trafficked wild mammals worldwide (Choo, Platto, & Challender, 2022). Pangolins, being avid consumers of ants and termites (Gu, Hu, & Yu, 2023), play a crucial role in regulating forest termite infestations, contributing to their ecological significance. Beyond their ecological value, pangolins hold economic and medicinal importance. Systematic evaluations by Jin et al. (Jin et al., 2021) revealed the efficacy of pangolins in treating certain diseases; however, efforts are ongoing to develop conventional and effective treatment alternatives. Additionally, in some parts of Africa, pangolins are utilized for traditional healing or incorporated into religious rituals (Ingram, Edwards, & Manzon, 2022). Unfortunately, pangolin populations have been severely impacted by poaching and habitat destruction (X. M. Wang et al., 2022). Studies indicate a substantial trafficking of approximately 895,000 pangolins between August 2000 and July 2019, particularly from Africa to Asia (Challender, Heinrich, Shepherd, & Katsis, 2020). It is worth noting that this figure likely represents only a fraction of the actual trade volume (Phelps & Webb, 2015).

According to the Global Biodiversity Information Facility (https://www.gbif.org), pangolins primarily inhabit regions across Africa and Asia, as depicted in Fig. 1. The species and classification of living pangolins have been a subject of unresolved controversy for nearly a century (Gaudin, Emry, & Wible, 2009). Presently, scientists have categorized them into three genera, encompassing eight confirmed species.

The four African pangolin species, including the white-bellied pangolin (*Phataginus tricuspis*), black-bellied pangolin (*Phataginus tetradactyla*), Temminck's pangolin (*Smutsia temminckii*), and giant ground pangolin (*Smutsia gigantea*), belong to the *genera Phataginus* and *Smutsia*. Conversely, the four Asian pangolin species, namely the Malayan pangolin (*Manis javanica*), Chinese pangolin (*Manis pentadactyla*), Indian pangolin (*Manis crassicaudata*), and Philippine pangolin (*Manis*

* Corresponding authors.
E-mail addresses: zjunjie@shu.edu.cn (J. Zhong), wsh666@shu.edu.cn (S. Wei), chenqincc@shu.edu.cn (Q. Chen), bingniu@shu.edu.cn (B. Niu).
[1] 0000-0002-3172-5712.
[2] 0009-0003-1752-8004.
[3] 0000-0003-3089-9195.
[4] 0000-0001-7528-9855.

https://doi.org/10.1016/j.eswa.2024.126206
Received 12 March 2024; Received in revised form 4 December 2024; Accepted 16 December 2024
Available online 18 December 2024
0957-4174/© 2024 Elsevier Ltd. All rights reserved, including those for text and data mining, AI training, and similar technologies.

culionensis), fall under the genus *Manis* (Gaubert et al., 2018; Gaudin et al., 2009). Notably, *Phataginus tetradactyla* exhibits predominantly arboreal behavior, while *Manis javanica*, *Manis culionensis*, and *Phataginus tricuspis* are semi-arboreal. The remaining four species are known for their burrowing habits (Willcox et al., 2019). According to the IUCN Red List, *Manis javanica*, *Manis culionensis* and *Manis pentadactyla* are critically endangered (CR), *Phataginus tricuspis*, *Manis crassicaudata* and *Smutsia gigantea* are endangered (EN) while *Smutsia temminckii* and *Phataginus tetradactyla* are vulnerable endangered (VU). In addition, a ninth species of pangolin was found in samples confiscated in 2023, the fifth species of pangolin in Asia, and the scientific name is *Manis mysteria* (Heighton et al., 2023).

Currently, deep learning techniques have revolutionized wildlife protection by integrating UAV and satellite remote sensing, infrared cameras, and other technologies (J. Li et al., 2023). By leveraging deep learning models on drones and infrared cameras, it becomes possible to minimize human intervention, track elusive species, and streamline the process of identifying target animals from vast amounts of video footage. Furthermore, researchers are actively exploring and evaluating different models to enhance wildlife identification. For instance, Ueno *et al.* (Ueno, Kabata, Hayashi, Terada, & Yamada, 2022) enhanced the recognition of Japanese macaques by combining GoogLeNet and ResNet-18 with a sequential Bayesian filter. Their study suggested that the sequential Bayesian filter could significantly enhance the accuracy of individual recognition among Japanese macaques. Similarly, Yang *et al.* (M. L. Yang et al., 2022) gathered images of various species, including Sika deer, shaded deer, quagga, wild boar, and red-bellied pheasant, in the Shennongjia National Reserve. They compared the recognition accuracy of YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x models, concluding that YOLOv5m demonstrated superior overall performance.

The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) has categorized all pangolin species due to the rampant illegal trade and declining populations (Heinrich et al., 2016). While interventions from conservation biology have offered pangolins a glimmer of hope, there is an urgent need to enhance protective measures. The advancement of artificial intelligence, particularly deep learning object detection models, presents a promising avenue for bolstering pangolin conservation efforts. These models continuously improve by learning and analyzing vast image datasets, enabling them to discern and categorize different objects (Sharma & Mir, 2020). Leveraging this technology, we can develop pangolin recognition models capable of swiftly identifying pangolins in images or videos and assigning appropriate category labels. Such pangolin detection models can serve various purposes, including identifying pangolin presence in field camera footage and aiding in the detection of illegal trade (Cardoso et al., 2023). Moreover, they can assist customs authorities in promptly identifying pangolin-related items, thereby streamlining pangolin protection efforts and facilitating the timely detection of pangolin distribution and potential threats.

The overall process of our study is illustrated in Fig. 2. First, an image database for pangolin detection was constructed by collecting images from online resources and the iNaturalist database. The images were carefully selected to include diverse poses, angles, and environmental conditions to ensure a robust dataset for training and testing. Next, the YOLOv8s model was used as a baseline, and significant improvements were made to propose an optimized model architecture. Key enhancements were introduced to the backbone and loss functions to increase detection accuracy and efficiency, specifically tailored for pangolin detection tasks. To validate the performance of the improved model, comparative analyses were conducted with other well-established object detection models, including Faster R-CNN, SSD, and RT-DETR, as well as the latest deep learning models, such as YOLOv10 and YOLOv11. The comparisons were performed based on multiple metrics, including precision, recall, mAP, model size, and computational efficiency (GFLOPs). Through the experimental results, the advantages of the proposed model in terms of accuracy, efficiency, and robustness were verified, demonstrating its superior performance in pangolin detection tasks. Additionally, the potential application of the model in real-world wildlife conservation scenarios was highlighted, providing a practical solution for detecting and monitoring endangered species.

To ensure clarity and accessibility for readers from diverse fields, including conservationists, the study's workflow is visually summarized in Fig. 2. This diagram serves to simplify the process and make the technical aspects of the research more comprehensible.

2. Materials and methods

2.1. Data collection

We retrieved images of eight pangolin species (*Manis pentadactyla*、 *Manis crassicaudata*、 *Manis javanica*、*Manis culionensis*、*Smutsia gigantea*、 *Manis temminckii*、*Phataginus tetradactyla* and *Phataginus tricuspis*) from public databases such as iNaturalist (https://www.inaturalist.org/) and the Internet. These pangolins were subsequently categorized into Asian and African species based on their geographical distribution. The pangolin images we obtained from iNaturalist are taken by observers and uploaded to provide fuzzy geographic coordinates, which are then annotated after discussion by naturalists to ensure the reliability of the image data. The images we obtained from iNaturalist and other sources were also verified by the team members to

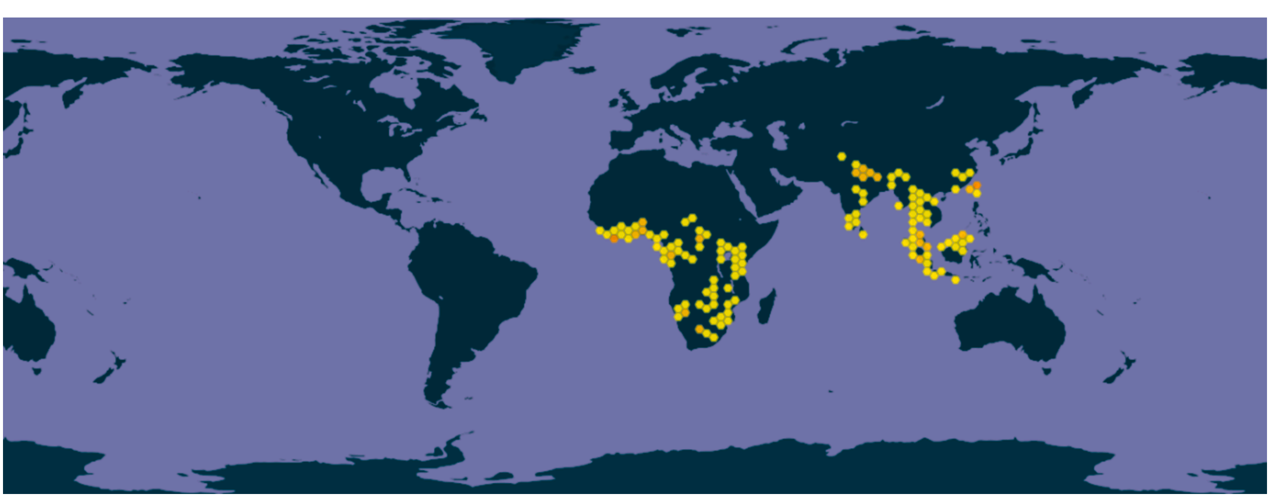


Fig. 1. Current geographical distribution of pangolins.

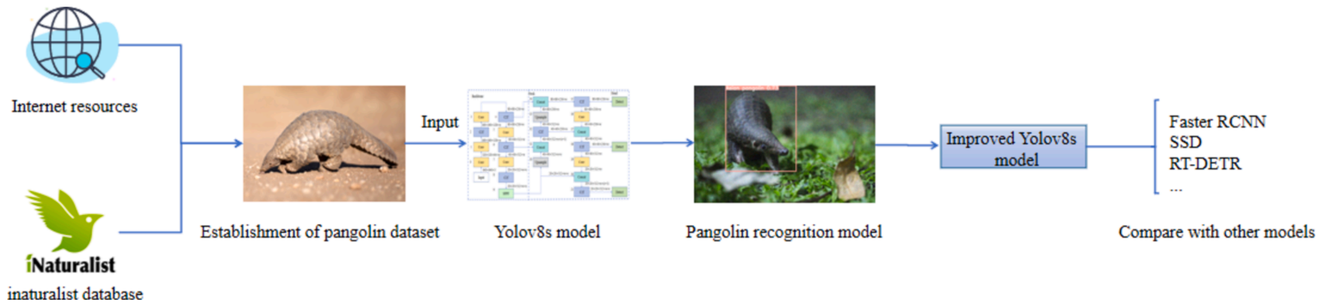


Fig. 2. A general flow chart of the research work.

ensure the reliability of the data labels. The images included different positions of pangolins under different environmental conditions, such as climbing, walking and crouching at different angles during day, night, cloudy and sunny days. These broad gestures help to improve the generalization ability of the model, ensure that the model does not over-rely on single gesture recognition, and ensure high recognition accuracy when dealing with uncertain or changeable animal gestures in the actual scene.

2.2. Dataset preparation

Considering the potential deployment scenarios for our model, which may encompass diverse conditions encountered in the wild, including various times of day and lighting conditions, we recognized the importance of ensuring the diversity of the pangolin image range. To achieve this goal, we performed data enhancement. Data enhancement is a widely used technique in the field of computer vision, whose purpose is to augment the size and diversity of training data, thereby improving the generalization ability of detection models (J. L. Zhang et al., 2022). In order to avoid overfitting risks stemming from insufficient training data, this study adopted multiple data enhancement strategies, including rotation, horizontal mirroring and vertical mirroring to expand the pangolin data set, as shown in Fig. 3. Through the application of data enhancement techniques, we expanded the pangolin dataset, enabling the training of a more robust and accurate pangolin target recognition model while improving the robustness of the model. Each image was randomly enhanced, and a total of 1370 images were obtained, including 680 images of Asian pangolins and 690 images of African pangolins, and the data samples were relatively balanced. These images were further divided into training, validation, and test sets in a ratio of 7:1:2.

2.3. Yolov8 model and its improvements

Object detection algorithms are broadly classified into two categories: two-stage algorithms based on regions, where adjacent pixels in an image are considered, and single-stage algorithms based on regression, which generate candidate frames and then classify these frames across the network (Yi, Liu, Zhao, & Liu, 2024). Single-stage algorithms directly predict the entire screen for classification and positioning, exemplified by the Yolo algorithm introduced by (Ding & Taylor, 2016). Numerous scholars have built upon the Yolo model, introducing new modules and creating several classical models. The launch of Yolov8 by Ultralytics on January 10, 2023, represents a significant advancement in

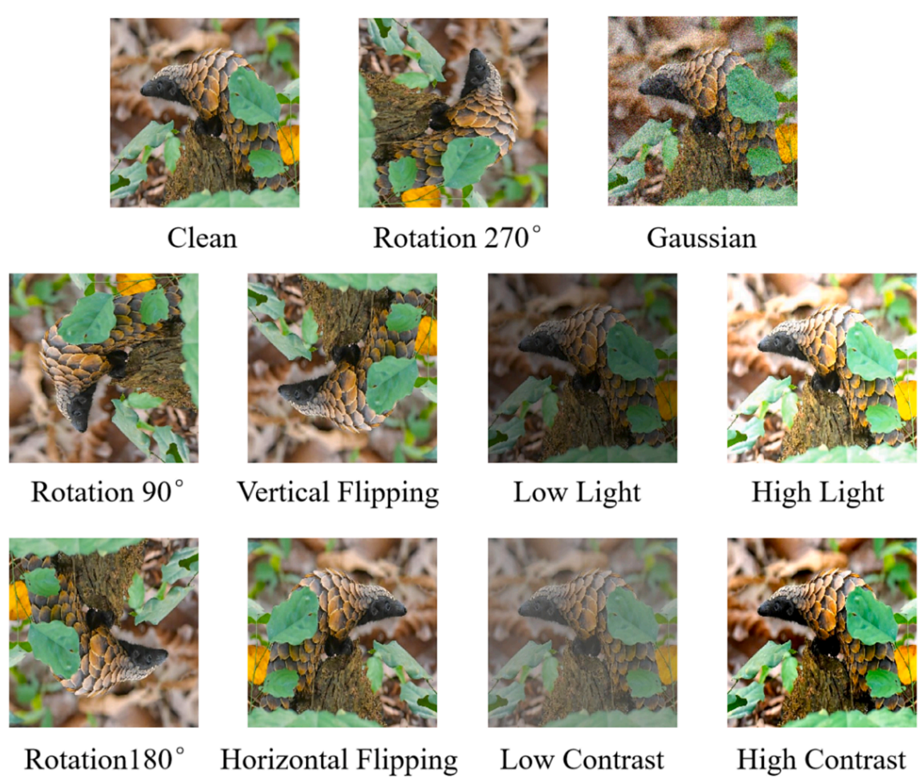


Fig. 3. Samples of the data augmentations.

this evolution (Y. T. Li, Fan, Huang, Han, & Gu, 2023). Compared to its predecessors like Yolov5 and Yolov7, Yolov8 is a cutting-edge model known for its superior detection accuracy and faster processing speed. Yolov8 comprises five versions: Yolov8n, Yolov8s, Yolov8m, Yolov8l, and Yolov8x, with Yolov8n offering the fastest detection speed and the fewest Flops and parameters among the five models. Yolov8x, on the other hand, has the slowest detection speed and a larger number of flops and parameters (Huang, Tan, Li, & Wu, 2023). To strike a balance between detection speed and accuracy, and considering deployment considerations, we selected Yolov8s as the baseline model. As depicted in Fig. 4, the Yolov8 network architecture consists of three main components: backbone, neck, and head (N. Li, Ye, Zhou, Gao, & Zhang, 2024).

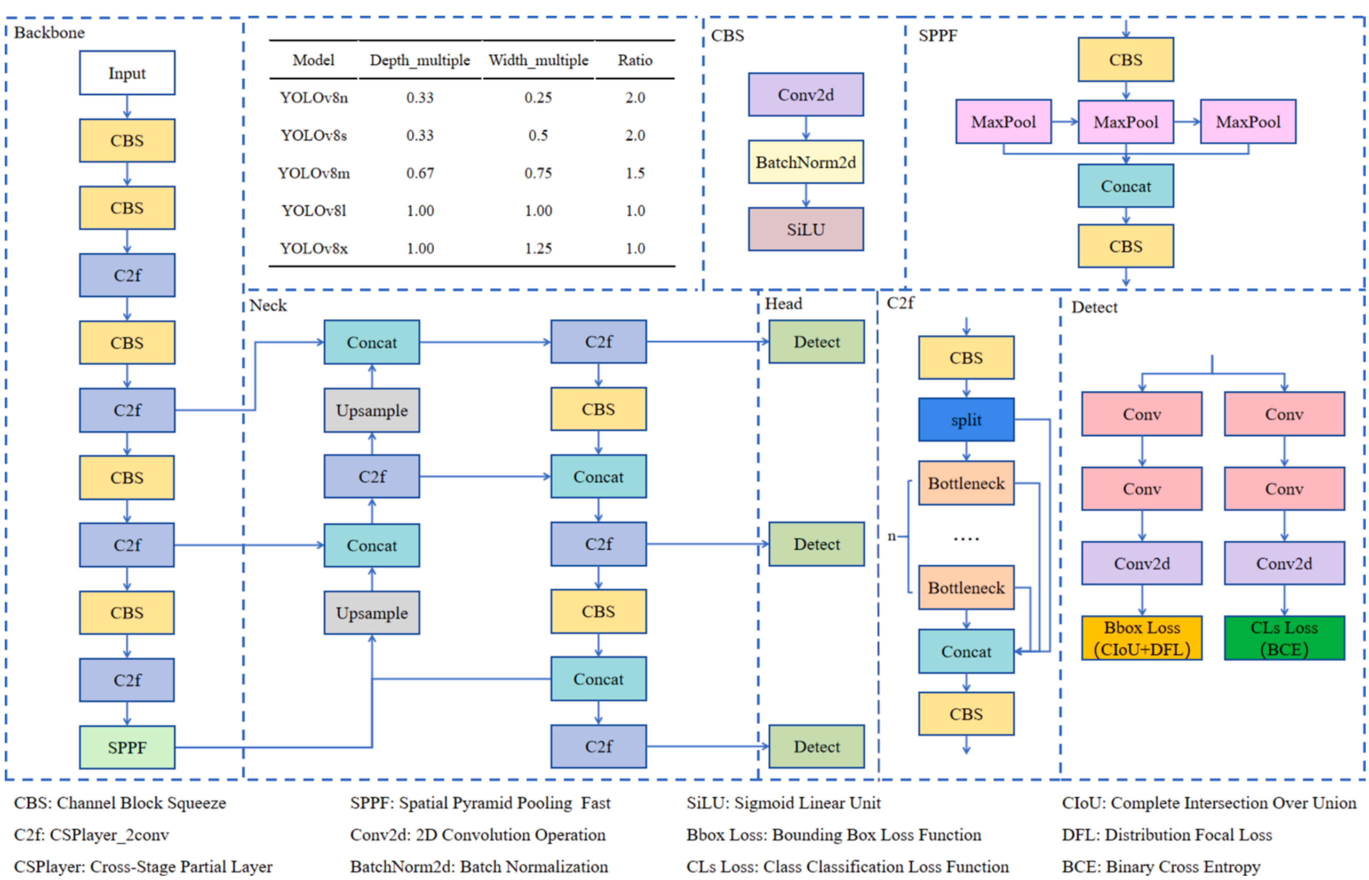
The backbone and neck modules constitute the core framework of the Yolov8 network. Input images undergo processing by multiple Conv and C2f modules to extract feature maps of varying scales. Within the Conv module, operations encompass convolution, batch normalization, and SiLU activation functions (Z. Wang, Lei, & Shi, 2023), while the C2f module, an enhanced iteration of the original C3 module, serves as the primary residual learning module (Talaat & ZainEldin, 2023). Drawing inspiration from the ELAN concept in Yolov7 (C. Y. Wang, Bochkovskiy, & Liao, 2023), the C2f module combines C3 and ELAN to form a more robust structure, enriching gradient flow information while upholding a lightweight design.

In the later stages of the backbone network, the Spatial Pyramid Fast Pooling (SPPF) module dynamically generates output of consistent size from the input feature map via pooling (G. L. Yang, Wang, Nie, Yang, & Yu, 2023). Compared to the Spatial Pyramid Pooling (SPP) framework (He, Zhang, Ren, & Sun, 2015), SPPF enhances computational efficiency and reduces latency by employing a sequence of three consecutive maximum pooling layers. Yolov8 integrates the principles of PANet (S. Liu, Qi, Qin, Shi, & Jia, 2018) and incorporates the PAN-FPN architecture into its neck components. This approach employs up-sampling and down-sampling techniques to amalgamate high-level and low-level feature maps, facilitating the amalgamation of semantic and localization features. Through this strategy, the network effectively integrates features from targets of varying scales, thereby enhancing its ability to detect objects of diverse sizes.

Yolov8 adopts a decoupled-head architecture featuring distinct branches for object classification and bounding box regression prediction, incorporating loss computation and target detection box filtering (Wang et al., 2023). The loss calculation process involves a strategy for allocating positive and negative samples and computing losses (G. L. Yang et al., 2023). Yolov8 primarily leverages the Task Aligned Assigner method, which selects positive samples based on the weighted outcomes of classification and regression scores (Tan, Pang, & Le, 2020). The loss calculation comprises two branches: classification and regression, with no objective branch. The classification branch continues to employ Binary Cross-Entropy (BCE) loss, while the regression branch utilizes the Distributed Focal Loss (DFL) (X. Li et al., 2020) and Complete Intersection over Union (CIOU) loss functions (Zheng et al., 2020).

2.4. Feature pyramid optimization

Pan-FPN introduces a bottom-up PAN structure atop FPN to compensate for lost positioning information. Despite the rich semantic and positioning information, the PAN-FPN structure still has room for further improvement (Qu et al., 2023). First, PAN-FPN structures inadequately process large-scale feature maps, potentially overlooking valuable information and resulting in degraded detection quality. Additionally, after up-sampling and down-sampling, feature maps lose some original information, leading to a relatively low reuse rate (X. Q. Wang, Gao, Jia, & Li, 2023). Therefore, there remains potential to enhance the PAN-FPN structure. To address these issues more effectively, this study reconstructs the feature fusion component of Yolov8



Model	Depth_multiple	Width_multiple	Ratio
YOLOv8n	0.33	0.25	2.0
YOLOv8s	0.33	0.5	2.0
YOLOv8m	0.67	0.75	1.5
YOLOv8l	1.00	1.00	1.0
YOLOv8x	1.00	1.25	1.0

Fig. 4. Yolov8 model structure diagram.

based on the concept of bidirectional feature pyramid network (BiFPN). The BiFPN structure was originally introduced by Google in the object detection algorithm EfficientDet (Tan et al., 2020). As illustrated in Fig. 5, BiFPN enhances semantic information by efficiently combining bidirectional cross-scale connections and weighted feature fusion. One advantage of BiFPN is its ability to expand the model's receptive field by fully utilizing high-resolution features (F. Z. Zhu, Wang, Cui, Liu, & Li, 2023). The main implementation involves treating feature graphs with a single input path without additional processing due to their low contribution. When fusing a feature map with two input paths, if the feature map has the same proportions, cross-level fusion requires introducing a new path from the backbone feature map (T. G. Li, Zhang, Li, & Zhang, 2022). This enhancement function improves the spatial information of the feature map, thereby enhancing target detection accuracy.

2.5. Optimization of loss function

In real-world scenarios, differentiating between Asian and African pangolins can be challenging due to various factors. To address this challenge, we incorporate Slide Loss into our method. Slide Loss assigns higher weights to challenging classification samples, enabling the network to focus more on these samples during training (Yu et al., 2022). This approach enhances the performance of challenging samples and contributes to overall network accuracy improvement. The slide loss function represents a form of adaptive weighting, where weights are dynamically adjusted based on predefined rules. The sliding function is defined as follows:

$$f(x) = \begin{cases} 1 x \leq \mu - 0.1 \\ e^{1-\mu} \mu < x < \mu - 0.1 \\ e^{1-x} x \geq \mu \end{cases} \tag{1}$$

The distinction between easy and hard samples relies on the Intersection over Union (IoU) size of the predicted bounding box and the ground truth box. To streamline hyperparameters, we compute the average IoU value across all bounding boxes as the threshold μ. Negative samples are selected for IoU values less than μ, while positive samples are chosen for IoU values greater than μ. This approach is illustrated in Fig. 6 (Tong, Zhang, Wu, Xu, & Sun, 2023).

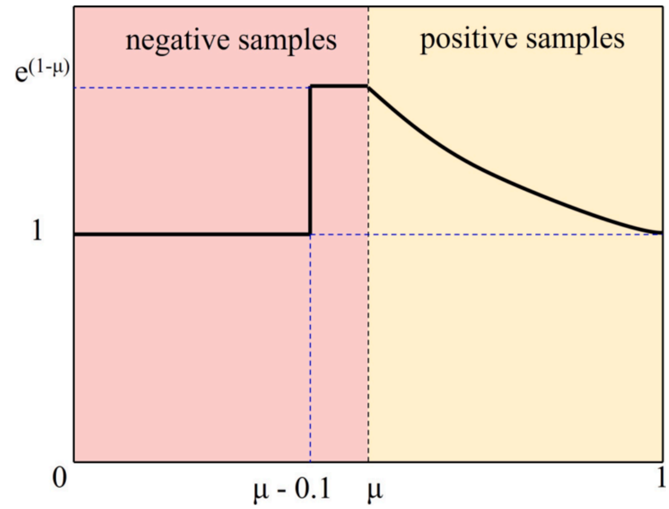


Fig. 6. The sample at the boundary is highlighted by the weighting function Slide (Yu et al., 2022).

2.6. The attention mechanism

The attention mechanism is a computational approach within deep learning that mirrors human perceptual systems, allowing models to selectively focus on crucial information while disregarding less relevant details during data processing. This mechanism empowers models to dynamically discern and prioritize different inputs, assigning varying degrees of importance to distinct components. Consequently, it enables the extraction of pertinent information from vast datasets (Niu, Zhong, & Yu, 2021).

To enhance awareness of pangolins online, we incorporated Triplet Attention (Misra, Nalamada, Arasanipalai, & Hou, 2021). This mechanism, characterized by its near-parameterless nature, facilitates multi-dimensional interaction without the need for dimensionality reduction. Triplet Attention thus captures richer features with minimal computational overhead, fostering a deeper understanding of the relationship between the target and its surroundings. Consequently, it enhances the model's ability to pinpoint target locations (Saber, Amin, Plawiak, Tadeusiewicz, & Hammad, 2022).

The structure of Triplet Attention, as illustrated in Fig. 7, comprises

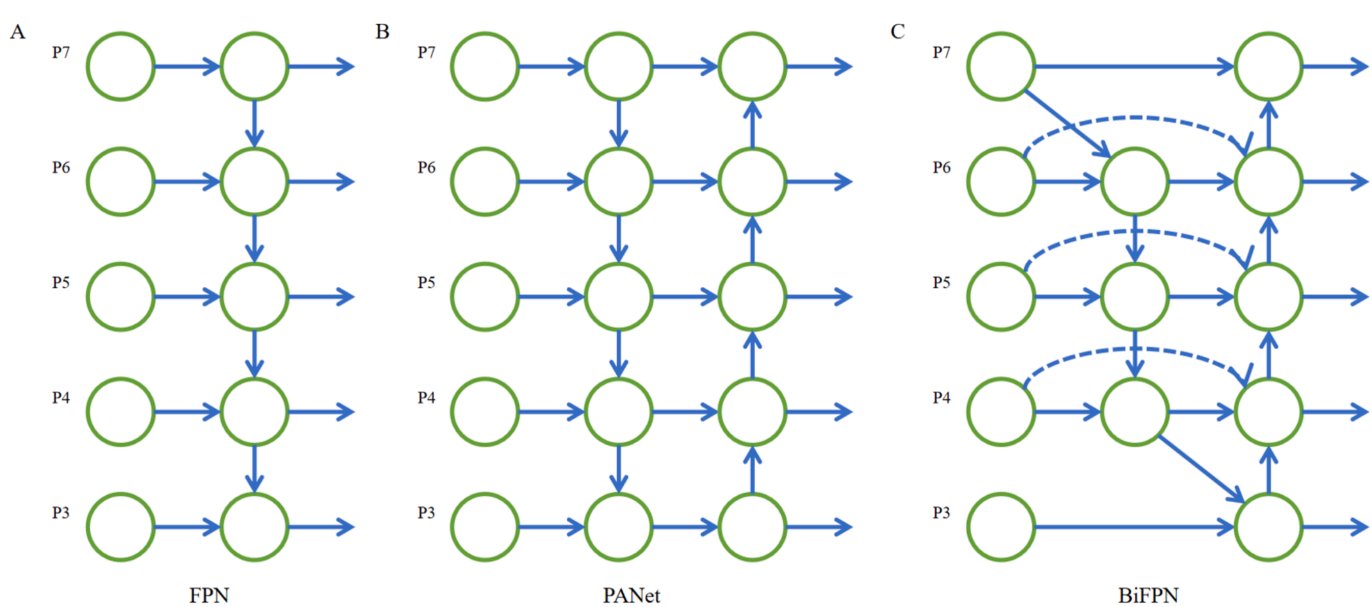


Fig. 5. The Structure of pyramid network with different features.

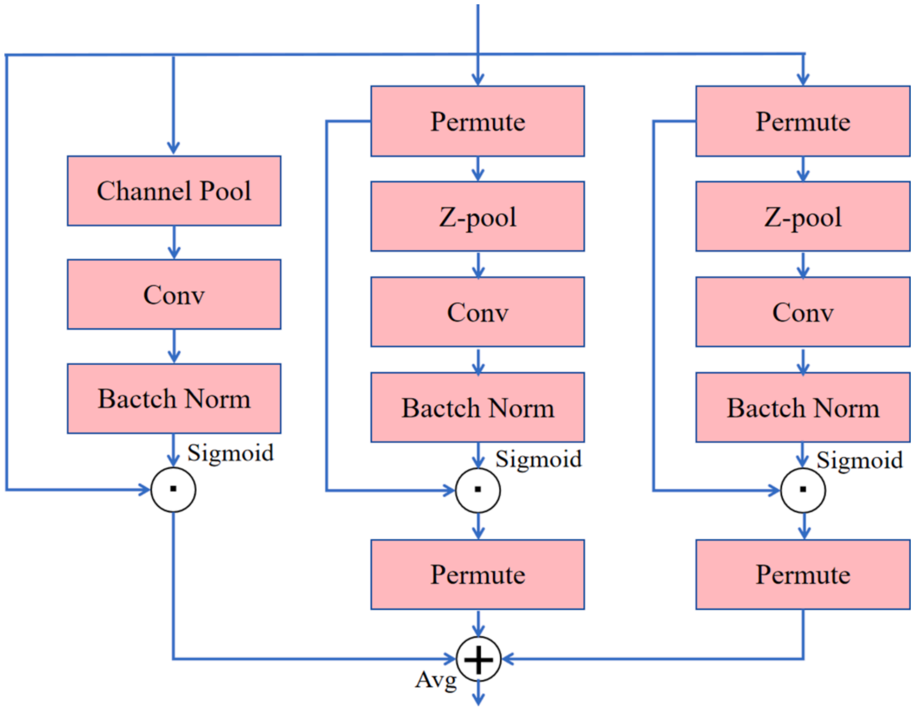
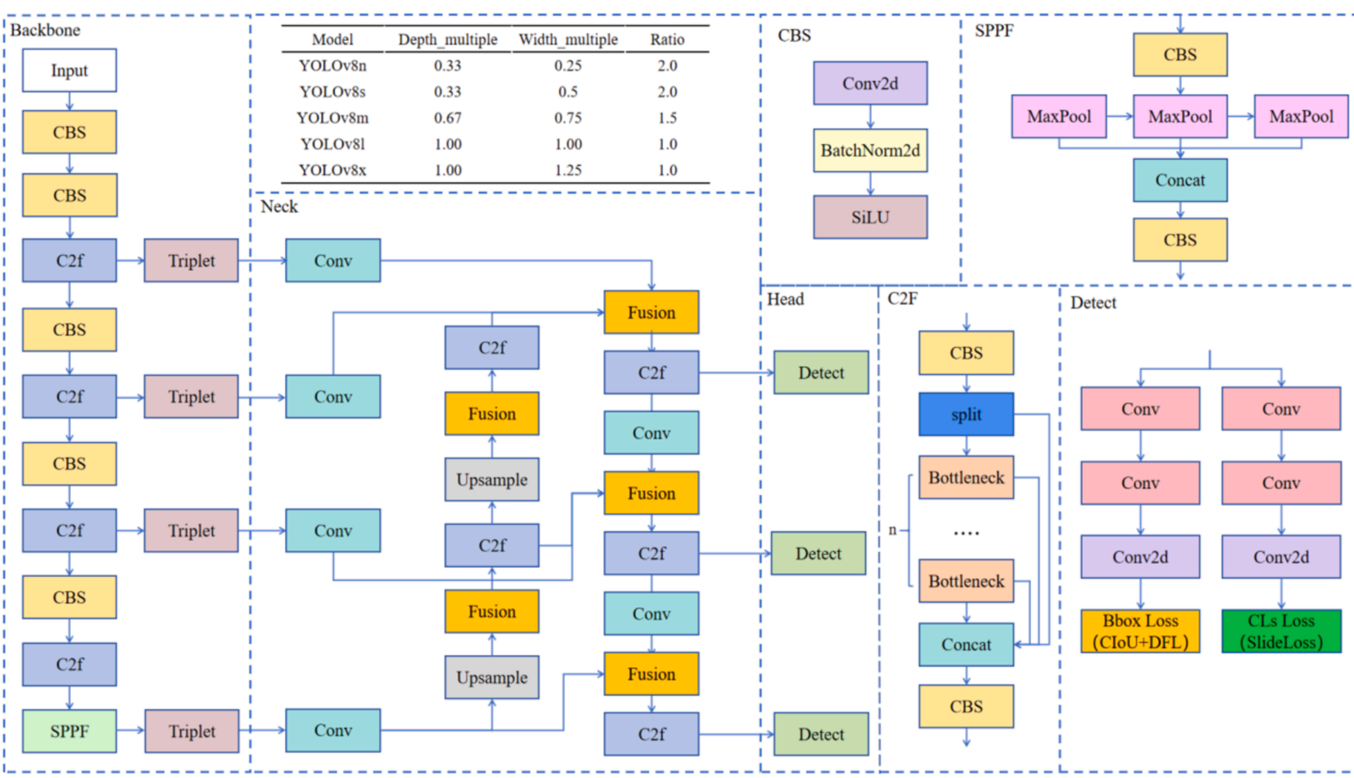


Fig. 7. The structure of Triplet Attention.

three branches aimed at capturing cross-dimensional interactions to compute attention weights. Each branch serves a distinct purpose: the top branch calculates attention weights for the channel dimension (C) and the width dimension (W), the middle branch computes attention weights for the channel dimension (C) and the height dimension (H), while the last branch captures interactions between the height (H) and width (W) dimensions (Qiang, Tao, Ye, Yang, & Xu, 2023).

Through the amalgamation of cross-dimensional information interactions, this method addresses the significant information loss inherent in traditional approaches that calculate attention weights for single dimensions. Consequently, it enables the capture of more intricate associations and dependencies in learning tasks. Finally, we integrated Triplet Attention into the backbone network, enhanced the neck component, incorporated BiFPN, and utilized the Slide loss function, as



Model	Depth_multiple	Width_multiple	Ratio
YOLOv8n	0.33	0.25	2.0
YOLOv8s	0.33	0.5	2.0
YOLOv8m	0.67	0.75	1.5
YOLOv8l	1.00	1.00	1.0
YOLOv8x	1.00	1.25	1.0

Fig. 8. The structure of Yolov8 improved model.

depicted in Fig. 8 below.

2.7. Evaluation metrics

To ensure a more accurate assessment of target detection performance, three fundamental metrics were introduced: accuracy (P), recall (R), and average accuracy (mAP) (W. Liu, Quijano, & Crawford, 2022; Wang et al., 2022).

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$AP = \int_0^1 P(R)d(R) \tag{4}$$

where N is the number of categories and AP is the area under the accuracy and recall curves.

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{5}$$

3. Results

3.1. Ablation experiment

Initially, we utilized the Yolov8s model as the foundational framework for training. We observed that the mAP50 achieved 83.1 %, with a model size of 22.5 MB, 11.1 million parameters, and 28.4 GFLOPs. Following enhancements to the Yolov8s architecture, the mAP increased to 87 %, accompanied by reductions in model size and parameter count. To validate the efficacy of our proposed enhancement approach, we conducted an ablation experiment to gain a more intuitive understanding of its effectiveness, as delineated in Table 1. Notably, each stage of the improved algorithm demonstrated enhanced performance, manifesting clear benefits. Subsequent integration of the Bifpn module led to an increase in model mAP, along with reductions in model parameters and size, indicative of Bifpn's enhancement of spatial feature map information, thereby augmenting target detection accuracy. Furthermore, the addition of the Slide Loss function resulted in further improvement in mAP, signifying heightened model attention towards challenging cases and enhanced prediction performance on difficult samples. Finally, the incorporation of Triplet Attention enabled the model to prioritize essential aspects of information processing, further elevating mAP without necessitating an increase in parameter count.

3.2. Effect of attention mechanism on model performance

In target detection, the integration of an attention mechanism aids the network in directing its focus towards critical feature components, thereby enhancing the accuracy of target localization and classification. To investigate the impact of various attention mechanisms on detection performance, several conventional attention mechanisms were compared with Triplet Attention, as delineated in Table 2.

Various types of attention mechanisms can be seamlessly integrated into the backbone network of the Yolov8 model, with TripletAttention demonstrating superior performance compared to other attention mechanisms. This highlights the significant enhancement in detection accuracy facilitated by the TripletAttention mechanism, without imposing excessive additional parameters or computational overhead.

3.3. Contrast experiment

To confirm the lightweight and superior detection performance of the proposed enhanced model, we compared it against various first-order and second-order algorithms, including the newly released YOLOv10(A. Wang et al., 2024) and YOLOv11 (Khanam & Hussain, 2024) Performance was evaluated across six metrics: precision, recall, mAP, parameter count, model size, and GFLOPs, as shown in Table 3.

In terms of accuracy, our improved YOLOv8 model outperforms most models, with the exception of RT-DETR (Lv et al., 2023), which is 0.8 % higher. However, our model achieves significantly better accuracy compared to other models while maintaining relatively low parameters and model size. For recall, models such as SSD(Wei Liu et al., 2016), Yolov3(Redmon & Farhadi, 2018), Faster RCNN(Ren, He, Girshick, & Sun, 2015), RetinaNet(Lin, Goyal, Girshick, He, & Dollár, 2017), RT-DETR(Lv et al., 2023), EfficientDet(Tan et al., 2020), Yolov9s(C.-Y. Wang, Yeh, & Liao, 2024) and Yolov10s(A. Wang et al., 2024) perform better. Despite this, our model demonstrates superior mAP50, outperforming 14 other models. It achieves 3.1 % and 1.2 % higher mAP50 than YOLOv10s and YOLOv11, respectively, and 10.2 % and 4.1 % higher than Faster R-CNN and RT-DETR, respectively.

In terms of GFLOPs, our model exhibits higher values compared to lightweight models such as YOLOv5s(Jocher et al., 2021), Improved-yolov5s(Xie et al., 2023), Yolov5S-Senet(Bhagabati, Sarma, & Bora, 2024), EfficientDet, Yolov10s and Yolov11s, with EfficientDet and Improved-YOLOv5s being 19.8 % and 17 % lower, respectively. However, our model demonstrates significantly lower GFLOPs than YOLOv3 (+129.6 %), Faster R-CNN (+345.2 %), RetinaNet (+145.1 %), RT-DETR (+78.4 %), YOLOv7 (C.-Y. Wang et al., 2023) (+80.1 %), YOLOv9s (+42.2 %), YOLOv8s (Y. T. Li et al., 2023) (+3.4 %), and SSD (+1.3 %). The higher GFLOPs in these models can be attributed to their more complex network structures and advanced feature extraction capabilities.

The improved YOLOv8 model features fewer parameters than other models, including YOLOv5s, Improved-YOLOv5s, YOLOv5S-SENet, and YOLOv10s. Notably, it has 129.7 % and 55.3 % fewer parameters than Faster R-CNN and SSD, respectively. While YOLOv5s, Improved-YOLOv5s, YOLOv5S-SENet, and YOLOv10s have slightly fewer parameters than our model (0.4 %, 3.7 %, 0.4 %, and 0.2 %, respectively), these differences are negligible given our model's superior accuracy.

Regarding model size, the wildlife recognition models proposed by Xie et al. (Xie et al., 2023) and Bhagabati et al. (Bhagabati et al., 2024) are smaller at 6.9 MB and 0.6 MB, respectively, compared to our improved YOLOv8s. However, the mAP of our model exceeds theirs by 13.9 % and 11.5 %, respectively. Additionally, our model matches the size of the model proposed by Jocher et al. (Jocher et al., 2021) is the same as that of the improved Yolov8s. The mAP of the improved Yolov8s model exceeds 13.9 % and 11.5 % of Xie et al. (Xie et al., 2023) and Bhagabati et al. (Bhagabati et al., 2024), while achieving 10.7 % higher mAP. These differences suggest that other models may prioritize reducing size at the expense of accuracy. In general, our proposed model strikes a balance between compactness, parameter efficiency, and accuracy. It is smaller, has fewer parameters, and achieves superior

Table 1
Results of ablation experiment.

Model	Precision(%)	Recall(%)	mAP50 (%)	GFLOPs	Params (M)	Model size (MB)
Yolov8s(baseline)	86.1	72.8	83.1	28.4	11.1	22.5
Yolov8s + Bifpn	82.3	78.5	84.5	25.0	7.4	14.3
Yolov8s + Bifpn + Slidloss	86.2	76.8	86.3	25.0	7.4	14.3
Yolov8s + Bifpn + Slidloss + Triplet Attention	87.6	76.1	87.0	25.0	7.4	14.3

Table 2
Comparison of attention mechanism modules.

Model	Precision(%)	Recall(%)	mAP50 (%)	GFLOPs	Params (M)	Model size (MB)
Yolov8s + Bifpn + Slidloss + TripletAttention(Misra et al., 2021)	87.6	76.1	87.0	25.0	7.4	14.3
Yolov8s + Bifpn + Slidloss + EMA(Ouyang et al., 2023)	82.9	78.9	86.3	26.0	7.4	14.3
Yolov8s + Bifpn + Slidloss + SE(Hu et al., 2018)	80.3	72.7	79.1	25.0	7.4	14.3
Yolov8s + Bifpn + Slidloss + CA(Hou, Zhou, & Feng, 2021)	81.0	73.3	81.8	25.0	7.4	14.3
Yolov8s + Bifpn + Slidloss + LSKA(Lau, Po, & Rehman, 2024)	82.1	77.4	83.5	25.9	7.7	14.3
Yolov8s + Bifpn + Slidloss + EffectiveSE(Lee & Park, 2020)	75.4	78.2	82.7	25.0	7.7	14.3
Yolov8s + Bifpn + Slidloss + SimAM(L. Yang, Zhang, Li, & Xie, 2021)	81.2	78.4	84.7	25.0	7.4	14.3

Table 3
Performance comparison of different models on datasets.

Model	Precision (%)	Recall(%)	mAP50 (%)	GFLOPs	Params (M)	Model size (MB)
Yolov8s(Y. T. Li et al., 2023)	86.1	72.8	83.1	28.4	11.1	22.5
Yolov5s(Jocher et al., 2021)	72.6	74.5	76.3	15.8	7.0	14.3
Yolov3(Redmon & Farhadi, 2018)	79.8	78.5	81.2	154.6	61.5	123.4
Faster RCNN(Ren et al., 2015)	55.6	91.5	76.8	370.2	137.1	108.2
SSD(Liu et al., 2016)	78.7	84.5	84.5	26.3	62.7	91.1
RetinaNet(Lin et al., 2017)	71.9	81.8	84.9	170.1	38.0	139.0
RT-DETR(Lv et al., 2023)	88.4	84.3	82.9	103.4	32.0	66.2
EfficientDet(Tan et al., 2020)	78.2	81.4	80.6	5.2	3.9	15.1
Improved-yolov5s(Xie et al., 2023)	73.3	72.9	73.1	8.0	3.7	7.4
Yolov5s-SENet(Bhagabati et al., 2024)	73.6	72.5	75.5	15.8	7.0	13.7
Yolov7(C. Y. Wang et al., 2023)	77.0	61.7	68.5	105.1	37.2	71.3
Yolov9s(Wang et al., 2024b)	80.3	77.3	84.8	67.2	15.5	30.3
Yolov10s(Wang et al., 2024a)	77.6	76.9	83.9	21.4	7.2	15.7
Yolov11s(Khanam & Hussain, 2024)	84.0	72.9	85.8	21.3	9.4	19.2
Improved Yolov8s model	87.6	76.1	87.0	25.0	7.4	14.3

detection performance, making it easier to deploy and apply in real-world scenarios. Therefore, the improved YOLOv8 model demonstrates clear advantages in both performance and practicality.

3.4. Detection effect and analysis

The paper proposes an enhanced model based on Yolov8s, which significantly improves the effectiveness of pangolin identification compared to the original model. The comparison of detection performance and accuracy is depicted in Fig. 9. As illustrated, Yolov8s exhibits issues of missed and false detections in pangolin identification. Conversely, the enhanced Yolov8 model enhances recognition accuracy while demonstrating higher confidence and lower rates of missed and false detections.

Image quality is affected by various factors including environmental conditions, camera quality and image processing techniques (W. J. Yang et al., 2023). Therefore, we evaluated the robustness of the model under diverse and complex environmental conditions by adjusting the parameters such as contrast, brightness and Gaussian noise to simulate the interference in different environments. As shown in Fig. 10, we compared the recognition performance of Yolov8 with that of the enhanced Yolov8s model on images exhibiting low brightness, low contrast and Gaussian noise. It can be found that in the complex environmental background, the improved Yolov8s model exhibits superior recognition capabilities. It accurately delineates the African pangolin without any erroneous detection boxes positioning the recognition box closer to the animal itself. Consequently, the improved Yolov8 model demonstrates enhanced robustness in complex environments settings.

3.5. Visualized analysis

The model was further visualized using Grad-CAM, as depicted in Fig. 11. Grad-CAM is utilized to visualize the activation regions of the network model across different target categories, offering a clear

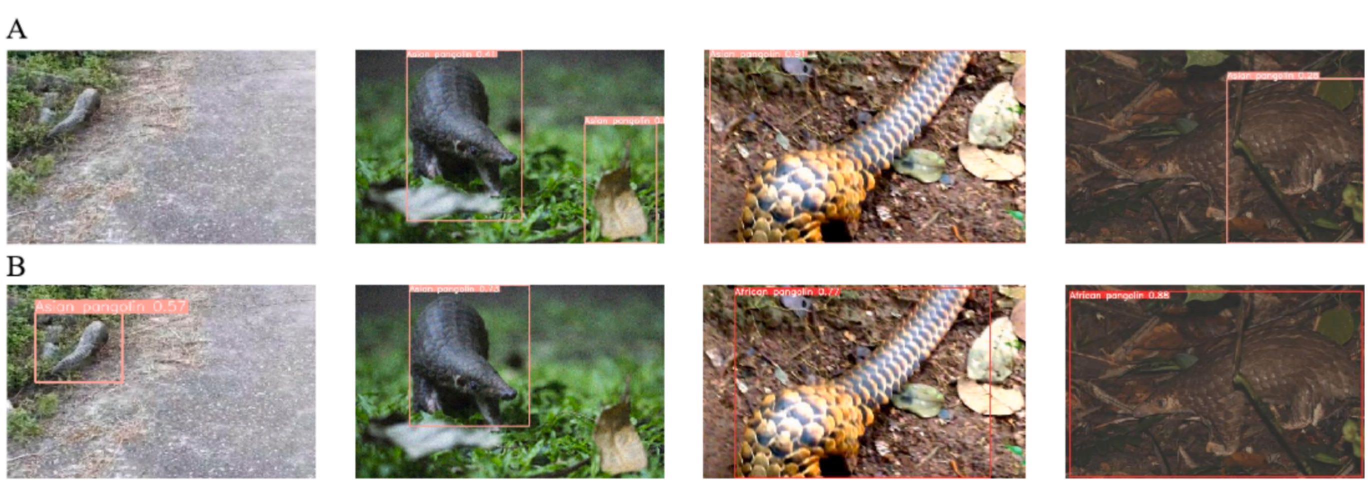


Fig. 9. Analysis of pangolin recognition results A.Yolov8s model B. Improved Yolov8s model.

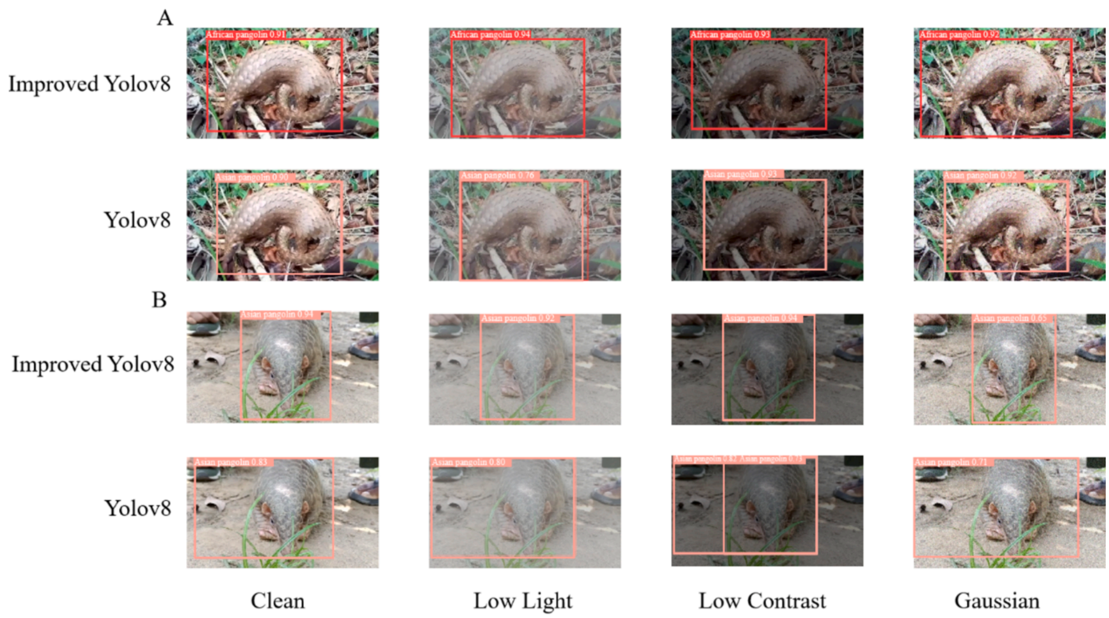


Fig. 10. Comparison of predictions of Yolov8s and improved Yolov8s in common complex scenarios A. African pangolin B. Asian pangolin.

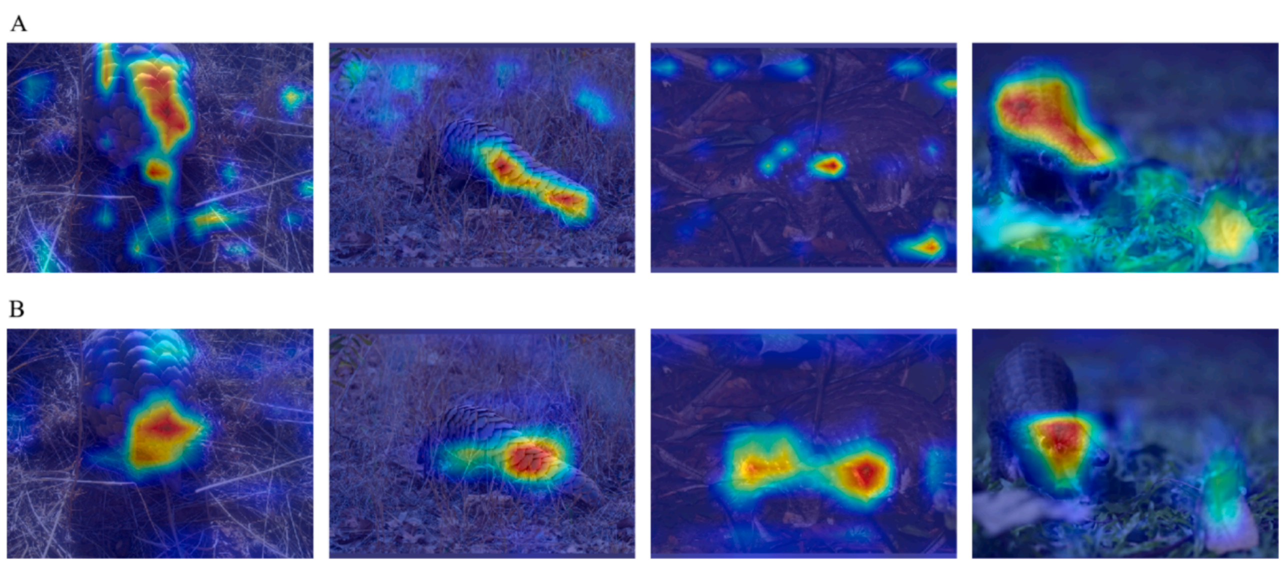


Fig. 11. Thermal map comparison results A.Yolov8s model B. Improved Yolov8s model.

depiction of the model's predictions within the image. This visualization technique aids in understanding the focal areas of the model during target detection, thereby facilitating model evaluation and enhancement. The results indicate that compared to the benchmark model, the improved model exhibits reduced thermal distribution in irrelevant regions, thus distinctly highlighting the key positions and regions of focus within the image.

In order to elucidate the decision-making process and enhance the interpretability of the model, we extracted the feature maps from Yolov8s and the improved Yolov8s network. The Yolov8s comprises 21 layers, whereas the improved yolov8s is featured by 31 layers of network. The feature maps are shown in Fig. 12. Visualizing these feature map reveals that bright colored areas potentially denote crucial recognition zones. Initially, the model extracts the overall features from the pangolin image (e.g., texture, outline, background). As the layers of the network deepen, the noise from background is filtered out and the location of the pangolin is highlighted. It can be found that the Yolov8s model erroneously identifies the disturbance as a pangolin, while the improved Yolov8s model effectively filters out the noise in the final layer, enhancing its accuracy and reliability.

3.6. Extended application

To enhance the generalizability assessment of the improved Yolov8 model, we conducted training and prediction using both the original Yolov8 model and the enhanced version on a publicly available wildlife dataset (D. Liu et al., 2023). This dataset comprises 21,987 images encompassing various wildlife species such as *Capricornis milneedwardsii, Elaphodus cephalophus, Hystrix brachyura, Rusa unicolor, Macaca thibetana, Martes flavigula, Ailuropoda melanoleuca, Pseudois nayaur, Ailurus fulgens, Rhinopithecus roxellana* and *Sus scrofa*.

As shown in Table 4, The results revealed that the improved model exhibited a 0.8 % increase in performance compared to the original algorithm when applied to the public dataset. This suggests that the enhanced model demonstrates robust performance and holds promise for application to other wildlife datasets.

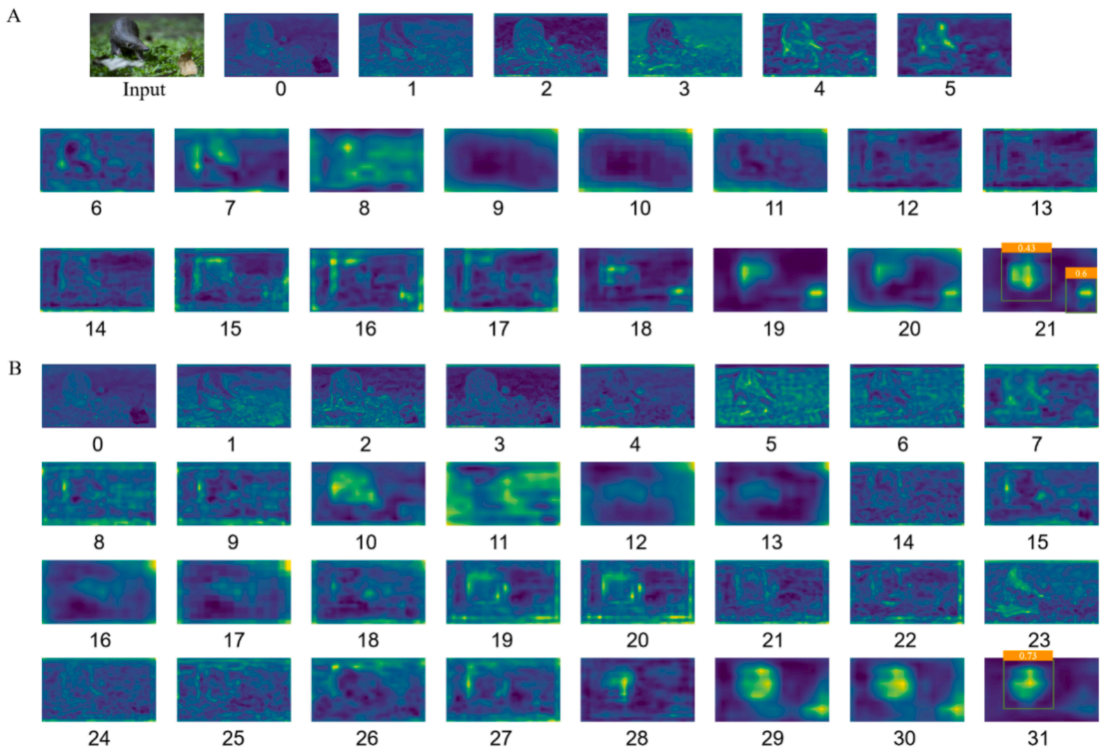


Fig. 12. Visual comparison results of feature maps A. Yolov8s model B. Improved Yolov8s model.

Table 4
Performance of Yolov8 and the improved model on a public dataset.

Model	Precision (%)	Recall(%)	mAP50 (%)	GFLOPs	Params (M)	Model size (MB)
Yolov8s	92.8	85.2	92.5	28.5	11.1	22.5
Yolov8s + Bifpn + Slidloss + TripletAttention	91.9	86.9	93.3	25.0	7.4	14.3

In addition, we aim to address the challenge of identifying rare animals from small sample datasets by utilizing a public dataset(Q. Y. Zhang et al., 2023) comprising six rare animal classes: Phayre's leaf monkey, Malabar pied hornbill, Skywalker hoolock gibbon, wreathed hornbill, red-thighed falconet, and great hornbill. Each class contains only 10 images, presenting a significant limitation in data availability. To overcome this, we applied Cycle-GAN (J. Y. Zhu, Park, Isola, & Efros, 2017) for image augmentation, effectively doubling the number of images per category to create a richer dataset, as illustrated in Fig. 13. The experimental results, summarized in Table 5, indicate that applying our improved model to this augmented dataset increased the mAP50 by 0.8 % compared to the original algorithm. These findings demonstrate that

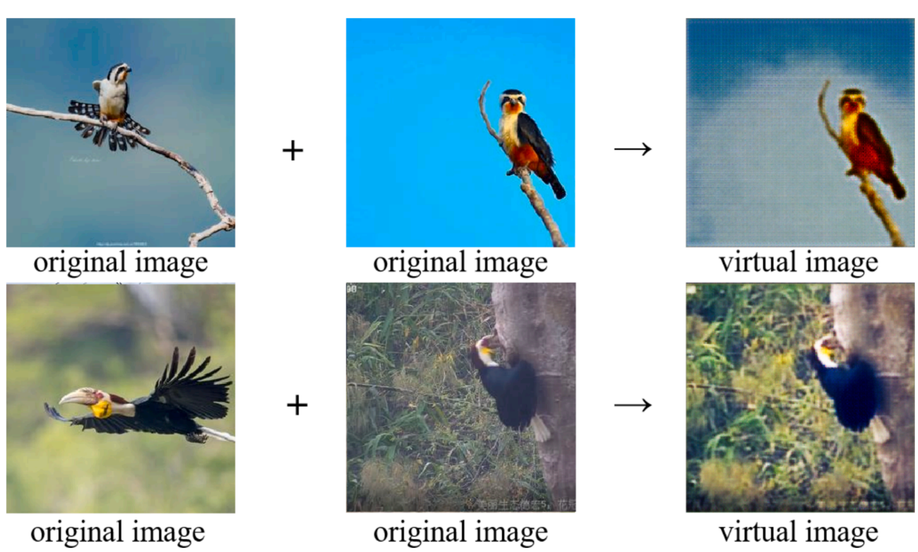


Fig. 13. style transfer diagram based on Cycle-GAN algorithm.

Table 5
Yolov8 and improved model performance on the small sample dataset.

Model	Precision (%)	Recall(%)	mAP50 (%)	GFLOPs	Params (M)	Model size (MB)
Yolov8s	90.2	87.7	89.3	28.5	11.1	22.5
Yolov8s + Bifpn + Slidloss + TripletAttention	93.9	87.6	90.1	25.0	7.4	14.3

Cycle-GAN-based data augmentation can significantly enhance the target recognition performance of rare animals in small sample scenarios. When combined with our improved model, this approach showed promising potential for addressing the challenges of rare animal identification in data-limited contexts.

4. Discussion

Pangolins, characterized as solitary and predominantly nocturnal creatures, have historically posed challenges for monitoring efforts (Khwaja et al., 2019). However, advancements in artificial intelligence (AI) and deep learning techniques have facilitated more effective pangolin surveillance. Deep learning has demonstrated remarkable proficiency in image processing tasks (LeCun, Bengio, & Hinton, 2015), prompting researchers to leverage these capabilities in constructing various wildlife recognition models. Some researchers have used deep learning to build different wildlife recognition models. For instance, Guo et al. (Guo et al., 2020) athered nocturnal images featuring coyotes, deer, possums, raccoons, and skunks, often obscured by vegetation, thereby complicating detection and classification. They devised a faster and 21 % more accurate multi-channel Area Suggestion and Classification network (VCRPCN) compared to the conventional R-CNN. Similarly, Silva et al. (de Silva et al., 2022) evaluated five CNN models (VGG16, ResNet50, InceptionV3, Xception, and Alexnet) for identifying Asian elephants, highlighting the superior performance of the Xception model, which emphasized the distinctive features of the elephants' ears. Moreover, the rapid evolution of deep learning models has led to the emergence of increasingly potent architectures, among which yolov8 stands out. In this exploratory study, we aimed to harness the potential of yolov8 for pangolin identification. Our findings underscore the considerable promise of the enhanced yolov8 model in image target detection within the realms of ecology and wildlife conservation, thereby furnishing a crucial tool for combating illicit wildlife trafficking. Notably, our improved model achieved exemplary performance with mAP values reaching 87.0 %. Furthermore, comparative analyses against other models such as SSD and Faster RCNN revealed superior accuracy levels.

Currently, numerous image recognition datasets pertaining to endangered animals and wildlife exist, including the panda dataset developed by Chen et al. (Chen et al., 2020), and a comprehensive full-HD wildlife monitoring image dataset created by Zhang et al. (T. Zhang et al., 2020) However, there is a scarcity of image target recognition models specifically tailored for pangolins. While Cardoso et al. (Cardoso et al., 2023) oncentrated on pangolin-related investigations, primarily focusing on the illicit trade of pangolins and assessing their physical condition and the condition of their body parts, our study adopts a distinct approach.

In contrast to previous research, we categorize pangolins into Asian and African species, facilitating the tracking of illicit trade routes and origins. Furthermore, we take into account the practicality of deploying the model on edge devices. Notably, the yolov8 model utilized in our study is lightweight, ensuring efficiency while maintaining accuracy, thus rendering it more suitable for deployment on edge devices (Ma & Pang, 2023).

To address the challenge of recognizing difficult pangolin instances, we aimed to enhance the Yolov8 model. Our primary concern in this enhancement process was to ensure the model's lightweight nature. Hence, we incorporated the bifpn structure, which surprisingly not only maintained accuracy but also achieved lightweightness. Subsequently, we sought to enhance the model's ability to discern difficult samples and pertinent information from input data, thus bolstering its performance and generalization capacity. To this end, we introduced the Slide Loss function (Tong et al., 2023), which guides model refinement by iteratively adjusting parameters to minimize losses and enhance predictions.

The Triplet Attention mechanism comprises three parallel branches: two capture interdimensional interactions between channels and spaces, while the third constructs spatial attention, akin to CBAM. The outputs of these branches are averaged and summed to obtain the final attention mechanism weight. Triplet Attention has demonstrated excellent performance in computer vision tasks such as object detection (Misra et al., 2021; P. Zhang, Deng, & Chen, 2023). Therefore, we further integrated the Slide Loss function and Triplet Attention, and their efficacy was confirmed in enhancing model accuracy. Since Triplet Attention is nearly parameterless, and the enhanced loss function does not augment the model's parameter count, the overall model size and parameters remain unchanged, preserving its lightweight nature.

Since the introduction of attention mechanisms in RNN in 2014 (Mnih, Heess, Graves, & Kavukcuoglu, 2014), numerous attention mechanisms have emerged. Hence, we replaced various attention mechanisms to ascertain if Triplet Attention was the most effective. Results indicated that compared to the original model, the model's performance was enhanced with the incorporation of EMA, SimAM, LSKA, and Triplet Attention. The mAP of the model with Triplet Attention reached 87.0 %, surpassing other attention mechanisms. Consequently, we selected Triplet Attention to integrate into the backbone network.

Currently, object detection methods are primarily categorized into two groups: two-stage algorithms typified by the R-CNN series, and one-stage algorithms exemplified by SSD and YOLO. Two-stage algorithms generate a set of candidate boxes that potentially contain target objects, filter them using predefined rules, and then conduct detection. Conversely, one-stage algorithms directly extract features from the network to predict object categories and locations. In this study, we compared the performance of Faster R-CNN, RetinaNet, and SSD algorithms with our improved algorithm, revealing superior performance of our enhanced approach. Furthermore, the authors of Yolov8 have investigated other algorithms such as Yolov3, Yolov5, and RT-DETR. Additionally, Tan et al. (Tan et al., 2020) proposed a novel feature fusion technique, BiFPN, a bidirectional feature pyramid network with dual weighting, which was integrated into their model EfficientDet. These algorithms have demonstrated promising results in target detection tasks. Therefore, we conducted a comparative analysis between these models and our enhanced model on the pangolin dataset proposed in this study, revealing superior performance of our improved model.

To vividly illustrate the performance outcomes of the enhanced model, we conducted an in-depth analysis of the detection results, revealing that the improved model significantly enhanced recognition accuracy while exhibiting higher confidence levels, lower miss rates, and reduced error rates. Our data enhancement approach takes into account various weather and lighting scenes encountered in the real-world scenarios, demonstrating our model's robustness. However, we acknowledge the oversight regarding occlusion scenario in the data enhancement. In the future modeling, we plan to address this scenario by incorporating techniques such as occlusion data enhancement to fortify the data set and enhance the model's resilience. Moreover, utilizing the Grad-CAM method, we obtained thermal maps demonstrating the improved model's ability to focus on relevant regions while reducing thermal distribution in irrelevant regions, thereby clearly delineating key positions and regions of interest within the image. Furthermore, we extended the applicability of the model beyond the original scope to assess its effectiveness on external wildlife datasets. Remarkably, the enhanced model achieved a 0.8 % higher mean Average Precision (mAP) on the wildlife dataset compared to the original model, underscoring its versatility and effectiveness in diverse wildlife monitoring applications.

In contemporary computer vision and image classification research, few-shot learning, particularly in the context of rare animal identification, presents a significant challenge (Feng & Xiao, 2022). Due to the limited availability of image data for rare animals and the high cost of acquiring labeled samples, traditional deep learning methods often struggle to perform effectively in small-sample scenarios.

To address this issue, various data enhancement techniques have been proposed in recent years, with style transfer emerging as a highly effective approach (Tang, Zhao, Feng, & Zhao, 2022). By combining style transfer with traditional manual augmentation methods, it is possible to generate more diverse image samples, thereby improving the accuracy and generalization of classification models. In particular, style transfer techniques in visual tasks expand training datasets by altering the visual style of images while preserving their original content, enhancing the model's robustness.

Among the many style transfer methods, Cycle-GAN (Cycle-Consistent Generative Adversarial Network) stands out for its ability to transform styles between different image domains by generating new and diverse samples.(J. Y. Zhu et al., 2017) Unlike traditional image enhancement techniques, Cycle-GAN does not require paired samples for training and can produce high-quality style-transformed images, making it especially suitable for small-sample learning (Shang et al., 2023).

Leveraging these advantages, this study utilized public datasets of rare animals with limited samples and applied Cycle-GAN for image data augmentation to tackle the problem of rare animal target recognition in small-sample conditions. The results demonstrate that our proposed model, combined with Cycle-GAN-based data augmentation, achieved strong performance. This approach shows great potential for addressing the challenges of rare animal identification in data-scarce scenarios, offering a practical and innovative solution.

Collecting image datasets of pangolins presents unique challenges compared to other wildlife datasets. Firstly, the rarity of these endangered animals makes observation and data collection more challenging. Additionally, pangolins comprise eight distinct species with overlapping geographical distributions and similar physical appearances, posing difficulties in species identification for observers. The classification of pangolins into Asian and African categories in this study is influenced by previous research highlighting significant distinctions between African and Asian pangolin species (T. T. Gu et al., 2023). Moreover, the scarcity of labeled images for each of the eight species further complicates species classification efforts.

The accurate identification of pangolins is essential for monitoring and conserving pangolin biodiversity. At present, wildlife identification models have been integrated into existing wildlife monitoring systems for practical application (Dertien et al., 2023). Likewise, our pangolin recognition models hold the potential to integrate into monitoring systems, not only in camera traps (Q. Y. Zhang et al., 2023) and drones (S. K. Wang et al., 2023), but also in real-time early warning systems that detect and alert authorities to approaching poacher, and can collaborate with other wildlife recognition models. To form a comprehensive monitoring system, artificial intelligence (AI) models can be integrated with a variety of technical tools, resources, and partnerships to work together to enhance wildlife conservation and ecological monitoring.

Pangolin identification models, combined with various technologies like remote sensing, drones and camera traps, offer comprehensive ecological information for pangolin conservation. The high-resolution image data obtained by remote sensing satellites and unmanned aerial vehicles cover a wide range of areas and capture hard-to-reach geographical locations, which can accurately identify the habitat and activity area of mountain beetles, understand their distribution, habitat environment and habitat status, and provide important data support for planning protected areas and ecological monitoring. Using drones equipped with infrared cameras and thermal imaging equipment to regularly cruise specific areas, we can locate pangolins in real time at night or in difficult terrain and monitor their movements in a timely manner. Once the AI automatic alarm system identifies abnormal situations, such as illegal hunting behavior, it can automatically send an alarm signal to the protection personnel or relevant agencies to find potential threats and illegal hunting activities, so that protection measures can be taken in time. AI is able to analyze animal distribution patterns, habitat quality and food chain health based on data captured by drones, helping reserve staff to make ecological assessments and strategic adjustments.

Camera traps strategically positioned in the hot spots of pangolin activity automatically capture the images and behavior data of pangolins like triggering the shooting, providing important data support for scientific research and conservation work, while also enhancing the accuracy and robustness of the identification model. In addition, leveraging Internet-based detection technology, we can monitor pangolin trading information in real time on the Internet and e-commerce platforms, swiftly detecting and intercepting illegal trade activities. Successful identification of pangolins can determine their current range of the pangolin's habitat, guiding decision-makers in identifying preferred habitat types and critical habitats, thereby informing the planning and management of protected areas. The analysis of pangolin population distribution and behavior reveal population trend and ecological requirements, and provide scientific basis for formulating target-oriented conservation measures. Based on the identification model, we can estimate the pangolin population, analyze the rise and fall of the pangolin population, assess the effectiveness of conservation efforts, combat illegal trade and enhance habitat quality. By collecting and analyzing data on various aspects of pangolins, including habitat usage, population distribution and behavioral patterns, valuable information and insights are provided to conservation organizations and government agencies (D. Q. Yang et al., 2021). These insights shed light on key issues related to the population status of pangolin species, habitat quality and habitat connectivity, which in turn contribute to the development of effective conservation strategies and management plans. The application of AI models should not be limited to technical implementation, but should also be integrated into conservation actions and integrated with the work processes and objectives of conservation organizations. AI models can share data and analysis results with existing monitoring platforms to enhance overall monitoring capabilities, and these conservation organizations can use AI to provide analysis results to optimize their conservation strategies, allocate resources, and develop action plans. AI models and drone technology require collaboration with people in local conservation organizations for technical training and equipment operation support. By working together, local conservation organizations can better understand and use these technologies to improve management efficiency. Through smartphone apps or online platforms, the public and volunteers can participate in monitoring missions in real time, uploading wildlife data to help expand monitoring coverage and improve data quality. The AI system can review and analyze this data in real time to ensure the reliability of the information. By combining AI models with the participation of drone monitoring, wildlife conservation organizations, other technological tools such as satellite imagery, ground sensors, etc., a multi-level, all-encompassing ecological monitoring system can be built. This will not only help improve monitoring efficiency, but also provide real-time data support for decision-making, thus promoting more scientific and sustainable conservation actions.

The key innovation of this study lies in the development of an improved YOLOv8 model, which incorporates BiFPN as the neck, Triplet Attention in the backbone network, and the SlideLoss loss function. This enhanced model is designed for the target detection of pangolins and other wild animals. Currently, there is very few target detection models specifically developed for pangolins. For example, Cardoso et al. (Cardoso et al., 2023) utilized a classical target detection model for pangolin identification but did not differentiate between species within the pangolin population. In contrast, our study advances this by classifying pangolins into Asian and African species.

Additionally, the model was extended to external wildlife datasets, where it demonstrated strong performance, indicating its broader applicability to wildlife detection tasks. We also compared our proposed model with the latest YOLO versions including YOLOv10 (A. Wang et al., 2024) and Yolov11 (Khanam & Hussain, 2024), which introduced notable advancements. YOLOv10 features consistent dual allocation without NMS training, lightweight classification headers, spatial-channel decoupling down-sampling and large kernel convolution. YOLOv11 incorporates an embedded multi-head attention mechanism in the C2f module and adds two DWConvs to the classification detection header, significantly reducing parameter and computation requirements.

Our comparative experiments showed that both YOLOv10 and YOLOv11 outperformed the baseline YOLOv8 model, confirming their effectiveness. However, on our dataset, the improved YOLOv8 model proposed in this study demonstrated superior performance compared to YOLOv10 and YOLOv11, highlighting its efficiency and reliability in addressing the specific challenges of pangolin target detection.

In wildlife detection and identification, the number of samples for rare animal species is often extremely limited, and in some cases, may even include species that have never been encountered before. Zero-Shot Learning (ZSL), as a large-scale concept learning approach, addresses this challenge by learning features of new categories without requiring labeled training data (Zhao, Yin, Zhou, Cai, & Qin, 2024). It achieves this by extracting feature information from existing known samples and applying it to identify previously unseen categories. For instance, Lampert et al. (Lampert, Nickisch, & Harmeling, 2009) introduced a classical ZSL method known as Direct Attribute Prediction (DAP). In their study, they extracted attributes such as “color” and “stripe” from images of tigers and pandas and successfully used these attributes to classify zebras and bees. Inspired by such methods, we aim to explore solutions for rare animal image recognition, particularly for species where data scarcity is a significant barrier. By mining common attributes of animals, such as body size, hair type, habitat, and other defining characteristics, and integrating these with visual information and semantic descriptions, we can achieve accurate identification of rare species. This approach can effectively expand the scope of species detection and provide critical technical support for the protection of endangered species. Looking ahead, the continued advancement of attribute learning and transfer learning techniques is expected to enhance the capabilities of ZSL. These innovations hold great promise for ecological conservation, species monitoring, and biodiversity protection. By promoting greater intelligence and precision in these fields, ZSL is poised to play an increasingly vital role in addressing some of the most pressing challenges in wildlife conservation.

Additionally, increasing the availability of labeled pangolin images specific to each species could facilitate more refined classification and the development of species-specific image target detection models. While our current model has not been deployed online, our future plans include making the established model accessible to the public through mobile apps or websites. This initiative aims to empower observers and researchers to accurately identify pangolins, contributing to their conservation and monitoring efforts. By establishing an image recognition model for pangolins, we envision deploying the model at ports to aid in combating illegal pangolin trade by identifying intercepted pangolins and determining their geographical origins. Furthermore, the model can be utilized for online monitoring of illegal pangolin trade, although this would necessitate integrating textual information and contextual cues for early warning systems to mitigate the risk of false positives and missed detections. To further enhance the effectiveness of the model, we intend to collect additional tagged pangolin images for model optimization, thereby improving its performance and advancing pangolin conservation efforts. Several open-source wildlife datasets cover a wide variety of wildlife species. We can integrate these existing datasets with our pangolin dataset to build a pre-trained model, accelerating the identification process for other wildlife species, including those endangered. Additionally, as new wildlife image data becomes available, our model can gradually accumulate new knowledge and adapt to evolving environment through incremental learning techniques (Peng, Zhao, Maksoud, Wang, & Lovell, 2023), eliminating the need for complete model retraining.

Lightweight networks not only reduce the computing demands for GPS or other hardware, but also facilitate integration into various monitoring devices (J. Li et al., 2023). Thus, in the future, we aim to further enhance the model’s lightweight nature and compress its size. For example, employing model pruning techniques (Sakai et al., 2022) can reduce computational complexity and storage requirements of a model by removing redundant information from the model or reducing the size of the model, without compromising predictive power. With lightweight models at our disposal, we can deploy them into camera traps and wildlife monitoring systems for real-time monitoring of current and future wildlife. Moreover, leveraging vast amount of camera trap video and image data enables us to uncover wildlife information that may have been missed in the past. Although genomic analysis has unveiled a ninth pangolin species (T.-T. Gu et al., 2023), lacking specific image poses a challenge. Therefore, collecting more tagged pangolin images remains a priority for further refinement. Combining genomic methods with machine learning presents an opportunity to uncover and identify the characteristic genes of all nine pangolins species. In the future, merging pangolin image recognition models with semantic models will enable the construction of a more comprehensive online monitoring and early warning system. By aggregating pangolin-related image and text data, the system employs image recognition models to analyze visual content, while semantic models extract key information and semantics from the textual data. Subsequently, a public opinion system and large-scale semantic model perform further analysis and monitoring. The public opinion system can actively monitor social media, news reports and other channels in real time to collect pangolin-related information. Meanwhile, the large-scale semantic model can conduct in-depth semantic analysis and correlation mining of these data to identify key information. The online monitoring and early warning system plays a crucial role in timely detection and prevention of potential threats posed by illegal wildlife trade. In summary, through intelligent, real-time and accurate identification and detection capabilities can expedite responses and provide robust support for a wide array of wildlife intelligent monitoring and conservation initiatives.

5. Conclusion

In this study, we employ deep learning methodologies for pangolin recognition detection and propose an enhanced Yolov8 pangolin target detection model. Yolov8s integrates the Bifpn module to optimize model size and FLOP while enhancing accuracy. Furthermore, we refine the loss function to prioritize challenging cases, prompting the model to focus on difficult instances. Additionally, the introduction of the Triplet attention module into the backbone network enables the selection of crucial information while suppressing non-critical data, thereby boosting model performance.

Beyond pangolin identification, our model is extended to wildlife datasets, offering potential advancements in wildlife image recognition. This broader application scope holds promise for the conservation and monitoring of various wildlife species in open environments. Ultimately, our efforts aim to contribute to the overarching goal of wildlife conservation and research.

Funding information

This study was supported by the National Key Research and Development Program of China (2022YFC2601205).

CRediT authorship contribution statement

Junjie Zhong: Writing – original draft, Methodology. **Suhang Wei:** Visualization, Investigation. **Qin Chen:** Project administration. **Bing Niu:** Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. This study was supported by the National Key Research and Development Program of China (2022YFC2601205)].

Data availability

Data will be made available on request.

References

Bhagabati, B., Sarma, K. K., & Bora, K. C. (2024). An automated approach for human-animal conflict minimisation in Assam and protection of wildlife around the Kaziranga National Park using YOLO and SENet Attention Framework. *Ecological Informatics, 79*. https://doi.org/10.1016/j.ecoinf.2023.102398. ARTN 102398.

Cardoso, A. S., Bryukhova, S., Renna, F., Reino, L., Xu, C., Xiao, Z. X., & Vaz, A. S. (2023). Detecting wildlife trafficking in images from online platforms: A test case using deep learning with pangolin images. *Biological Conservation, 279*. https://doi.org/10.1016/j.biocon.2023.109905. ARTN 109905.

Challender, D. W. S., Heinrich, S., Shepherd, C. R., & Katsis, L. K. D. (2020). Chapter 16 - International trade and trafficking in pangolins, 1900–2019. In D. W. S. Challender, H. C. Nash, & C. Waterman (Eds.), *Pangolins* (pp. 259–276). Academic Press.

Chen, P., Swarup, P., Matkowski, W. M., Kong, A. W. K., Han, S., Zhang, Z. H., & Rong, H. (2020). A study on giant panda recognition based on images of a large proportion of captive pandas. *Ecology and Evolution, 10*(7), 3561–3573. https://doi.org/10.1002/ece3.6152

Choo, S. W., Platto, S., & Challender, D. W. S. (2022). Helping to save pangolins from extinction. *Conservation Letters, 15*(3). doi:ARTN e12871. 10.1111/conl.12871.

de Silva, E. M. K., Kumarasinghe, P., Indrajith, K. K. D. A. K., Pushpakumara, T., Vimukthi, R. D. Y., de Zoysa, K., & de Silva, S. (2022). Feasibility of using convolutional neural networks for individual-identification of wild Asian elephants. *Mammalian Biology, 102*(3), 909–919. https://doi.org/10.1007/s42991-021-00206-2

Dertien, J. S., Negi, H., Dinerstein, E., Krishnamurthy, R., Negi, H. S., Gopal, R., & Baldwin, R. F. (2023). Mitigating human-wildlife conflict and monitoring endangered tigers using a real-time camera-based alert system. *Bioscience, 73*(10), 748–757. https://doi.org/10.1093/biosci/biad076

Ding, W. G., & Taylor, G. (2016). Automatic moth detection from trap images for pest management. *Computers and Electronics in Agriculture, 123*, 17–28. https://doi.org/10.1016/j.compag.2016.02.003

Feng, J. F., & Xiao, X. X. (2022). Multiobject tracking of wildlife in videos using few-shot learning. *Animals, 12*(9). https://doi.org/10.3390/ani12091223. ARTN 1223.

Gaubert, P., Antunes, A., Meng, H., Miao, L., Peigné, S., Justy, F., & Luo, S.-J. (2018). The complete phylogeny of pangolins: scaling up resources for the molecular tracing of the most trafficked mammals on earth. *Journal of Heredity, 109*(4), 347–359. https://doi.org/10.1093/jhered/esx097

Gaudin, T. J., Emry, R. J., & Wible, J. R. (2009). The phylogeny of living and extinct pangolins (Mammalia, Pholidota) and associated taxa: a morphology based analysis. *Journal of Mammalian Evolution, 16*(4), 235–305. https://doi.org/10.1007/s10914-009-9119-9

Gu, T.-T., Wu, H., Yang, F., Gaubert, P., Heighton, S. P., Fu, Y., & Yu, L. (2023). Genomic analysis reveals a cryptic pangolin species. *Proceedings of the National Academy of Sciences, 120*(40). https://doi.org/10.1073/pnas.2304096120. e2304096120.

Gu, T. T., Hu, J. Y., & Yu, L. (2023). Evolution and conservation genetics of pangolins. *Integrative Zoology*. https://doi.org/10.1111/1749-4877.12796

Guo, Y. H., Rothfus, T. A., Ashour, A. S., Si, L., Du, C. L., & Ting, T. F. (2020). Varied channels region proposal and classification network for wildlife image classification under complex environment. *Iet Image Processing, 14*(4), 585–591. https://doi.org/10.1049/iet-ipr.2019.1042

He, K. M., Zhang, X. Y., Ren, S. Q., & Sun, J. (2015). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *Ieee Transactions on Pattern Analysis and Machine Intelligence, 37*(9), 1904–1916. https://doi.org/10.1109/Tpami.2015.2389824

Heighton, S. P., Allio, R., Murienne, J., Salmona, J., Meng, H., Scornavacca, C., & Gaubert, P. (2023). Pangolin genomes offer key insights and resources for the world's most trafficked wild mammals. *Molecular Biology and Evolution, 40*(10). https://doi.org/10.1093/molbev/msad190. msad190.

Heinrich, S., Wittmann, T. A., Prowse, T. A. A., Ross, J. V., Delean, S., Shepherd, C. R., & Cassey, P. (2016). Where did all the pangolins go? International CITES trade in pangolin species. *Global Ecology and Conservation, 8*, 241–253. https://doi.org/10.1016/j.gecco.2016.09.007

Hou, Q., Zhou, D., & Feng, J. (2021). *Coordinate attention for efficient mobile network design.*.

Hu, J., Shen, L., & Sun, G. (2018). *Squeeze-and-excitation networks*.

Huang, Y., Tan, W. Z., Li, L., & Wu, L. J. (2023). WFRE-YOLOv8s: A New Type of Defect Detector for Steel Surfaces. *Coatings, 13*(12). doi: ARTN 2011. 10.3390/coatings13122011.

Ingram, D. J., Edwards, I. B., & Manzon, A. K. (2022). Of meat and ritual: Consumptive and religious uses of pangolins in Mali. *African Journal of Ecology, 60*(2), 184–192. https://doi.org/10.1111/aje.12945

Jin, X. Y., Chua, H. Z., Wang, K. Y., Li, N., Zheng, W. K., Pang, W. T., & Zhang, J. H. (2021). Evidence for the medicinal value of Squama Manitis (pangolin scale): a systematic review. *Integrative Medicine Research, 10*(1). https://doi.org/10.1016/j.imr.2020.100486. doi:ARTN 100486.

Jocher, G., Stoken, A., Borovec, J., Changyu, L., Hogan, A., Chaurasia, A., Ye, H. (2021). ultralytics/yolov5: v4. 0-nn. SiLU activations, Weights & Biases logging, PyTorch Hub integration. *Zenodo.*.

Khanam, R., & Hussain, M. (2024). YOLOv11: An Overview of the Key Architectural Enhancements. *arXiv preprint arXiv:2410.17725*.

Khwaja, H., Buchan, C., Wearn, O. R., Bahaa-el-din, L., Bantlin, D., Bernard, H., & Challender, D. W. S. (2019). Pangolins in global camera trap data: Implications for ecological monitoring. *Global Ecology and Conservation, 20*. https://doi.org/10.1016/j.gecco.2019.e00769. ARTN e00769.

Lampert, C. H., Nickisch, H., & Harmeling, S. (2009). Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer. *Cvpr: 2009 Ieee Conference on Computer Vision and Pattern Recognition, Vols 1-4*, 951-958. Retrieved from <Go to ISI>://WOS:000279038000122.

Lau, K. W., Po, L.-M., & Rehman, Y. A. U. (2024). Large separable kernel attention: rethinking the large kernel attention design in CNN. *Expert Systems with Applications, 236*, Article 121352.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, Y., & Park, J. (2020). *Centermask: Real-time anchor-free instance segmentation*.

Li, J., Xu, W. K., Deng, L. M., Xiao, Y., Han, Z. Z., & Zheng, H. Y. (2023). Deep learning for visual recognition and detection of aquatic animals: a review. *Reviews in Aquaculture, 15*(2), 409–433. https://doi.org/10.1111/raq.12726

Li, N., Ye, T. R., Zhou, Z. H., Gao, C. M., & Zhang, P. (2024). Enhanced YOLOv8 with BiFPN-SimAM for precise defect detection in miniature capacitors. *Applied Sciences-Basel, 14*(1), 429. https://doi.org/10.3390/app14010429

Li, T. G., Zhang, Y. Z., Li, Q. Q., & Zhang, T. C. (2022). AB-DLM: an improved deep learning model based on attention mechanism and BiFPN for driver distraction behavior detection. *Ieee Access, 10*, 83138–83151. https://doi.org/10.1109/Access.2022.3197146

Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., & Yang, J. (2020). Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection. *Advances in Neural Information Processing Systems, 33*, 21002–21012.

Li, Y. T., Fan, Q. S., Huang, H. S., Han, Z. G., & Gu, Q. (2023). A Modified YOLOv8 detection network for UAV aerial image recognition. *Drones, 7*(5). https://doi.org/10.3390/drones7050304. doi:ARTN 304.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). *Focal loss for dense object detection.*

Liu, D., Hou, J., Huang, S., Liu, J., He, Y., Zheng, B., & Zhang, J. (2023). *LoTE-Animal: A Long Time-span Dataset for Endangered Animal Behavior Understanding. Paper presented at the 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*.

Liu, S., Qi, L., Qin, H. F., Shi, J. P., & Jia, J. Y. (2018). Path Aggregation Network for Instance Segmentation. *Ieee/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr), 2018*, 8759–8768. https://doi.org/10.1109/Cvpr.2018.00913

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016, 2016//). *SSD: Single Shot MultiBox Detector.* Paper presented at the Computer Vision – ECCV 2016, Cham.

Liu, W., Quijano, K., & Crawford, M. M. (2022). YOLOv5-tassel: detecting tassels in RGB UAV imagery With improved YOLOv5 based on transfer learning. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15*, 8085–8094. https://doi.org/10.1109/Jstars.2022.3206399

Lv, W., Xu, S., Zhao, Y., Wang, G., Wei, J., Cui, C., Liu, Y. (2023). Detrs beat yolos on real-time object detection. *arXiv preprint arXiv:2304.08069*.

Ma, M. Y., & Pang, H. L. (2023). SP-YOLOv8s: An improved YOLOv8s model for remote sensing image tiny object detection. *Applied Sciences-Basel, 13*(14). https://doi.org/10.3390/app13148161. doi:ARTN 8161.

Misra, D., Nalamada, T., Arasanipalai, A. U., & Hou, Q. B. (2021). Rotate to Attend: Convolutional Triplet Attention Module. *2021 Ieee Winter Conference on Applications of Computer Vision Wacv 2021*, 3138-3147. doi:10.1109/Wacv48630.2021.00318.

Mnih, V., Heess, N., Graves, A., & Kavukcuoglu, K. (2014). Recurrent Models of Visual Attention. *Advances in Neural Information Processing Systems 27 (Nips 2014), 27*. Retrieved from <Go to ISI>://WOS:000452647102103.

Niu, Z. Y., Zhong, G. Q., & Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing, 452*, 48–62. https://doi.org/10.1016/j.neucom.2021.03.091

Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., & Huang, Z. (2023). *Efficient Multi-Scale Attention Module with Cross-Spatial Learning*.

Peng, C., Zhao, K., Maksoud, S., Wang, T. R., & Lovell, B. C. (2023). DIODE: dilatable incremental object detection. *Pattern Recognition, 136*. https://doi.org/10.1016/j.patcog.2022.109244. ARTN 109244.

Phelps, J., & Webb, E. L. (2015). "Invisible" wildlife trades: Southeast Asia's undocumented illegal trade in wild ornamental plants. *Biological Conservation, 186*, 296–305. https://doi.org/10.1016/j.biocon.2015.03.030

Qiang, H., Tao, Z. X., Ye, B., Yang, R. X., & Xu, W. Y. (2023). Transmission line fault detection and classification based on improved YOLOv8s. *Electronics, 12*(21). https://doi.org/10.3390/electronics12214537. doi:ARTN 4537.

Qu, Y., Wan, B. Y., Wang, C., Ju, H. J., Yu, J. B., Kong, Y. K., & Chen, X. C. (2023). Optimization algorithm for steel surface defect detection based on PP-YOLOE. *Electronics, 12*(19). https://doi.org/10.3390/electronics12194161. doi:ARTN 4161.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems, 28*.

Saber, S., Amin, K., Plawiak, P., Tadeusiewicz, R., & Hammad, M. (2022). Graph convolutional network with triplet attention learning for person re-identification. *Information Sciences, 617*, 331–345. https://doi.org/10.1016/j.ins.2022.10.105

Sakai, Y., Iwakawa, A., Tabaru, T., Inoue, A., Kawaguchi, H., & Ieee. (2022, Aug 21-25). *Automatic Pruning Rate Derivation for Structured Pruning of Deep Neural Networks.* Paper presented at the 26th International Conference on Pattern Recognition / 8th International Workshop on Image Mining - Theory and Applications (IMTA), Montreal, CANADA.

Shang, C., Zhao, H. K., Wang, M. L., Wang, X. L., Jiang, Y., & Gao, Q. (2023). Individual identification of cashmere goats via method of fusion of multiple optimization. *Computer Animation and Virtual Worlds, 34*(2). doi:ARTN e2048. 10.1002/cav.2048.

Sharma, V. K., & Mir, R. N. (2020). A comprehensive and systematic look up into deep learning based object detection techniques: a review. *Computer Science Review, 38*. https://doi.org/10.1016/j.cosrev.2020.100301. ARTN 100301.

Talaat, F. M., & ZainEldin, H. (2023). An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Computing & Applications*. https://doi.org/10.1007/s00521-023-08809-1

Tan, M., Pang, R., & Le, Q. V. (2020). *Efficientdet: Scalable and efficient object detection.*.

Tang, J. X., Zhao, Y. Q., Feng, L. Q., & Zhao, W. X. (2022). Contour-Based Wild Animal Instance Segmentation Using a Few-Shot Detector. *Animals, 12*(15). doi:ARTN 1980. 10.3390/ani12151980.

Tong, Q., Zhang, E. M., Wu, S. T., Xu, K. H., & Sun, C. (2023). A real-time detector of chicken healthy status based on modified YOLO. *Signal Image and Video Processing, 17*(8), 4199–4207. https://doi.org/10.1007/s11760-023-02652-6

Ueno, M., Kabata, R., Hayashi, H., Terada, K., & Yamada, K. (2022). Automatic individual recognition of Japanese macaques from sequential images. *Ethology, 128*(5), 461–470. https://doi.org/10.1111/eth.13277

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*.

Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv preprint arXiv:2402.13616*.

Wang, C. S., Sun, W., Wu, H. R., Zhao, C. J., Teng, G. F., Yang, Y. R., & Du, P. F. (2022). A low-altitude remote sensing inspection method on rural living environments based on a modified YOLOv5s-ViT. *Remote Sensing, 14*(19). https://doi.org/10.3390/rs14194784. doi:ARTN 4784.

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *2023 Ieee/Cvf Conference on Computer Vision and Pattern Recognition* (pp. 7464–7475). https://doi.org/10.1109/Cvpr52729.2023.00721

Wang, G., Chen, Y. F., An, P., Hong, H. Y., Hu, J. H., & Huang, T. E. (2023). UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors, 23*(16). https://doi.org/10.3390/s23167190. doi:ARTN 7190.

Wang, S. K., Cao, X. W., Wu, M. Q., Yi, C. B., Zhang, Z., Fei, H., & Yang, P. S. (2023). Detection of pine wilt disease using drone remote sensing imagery and improved YOLOv8 algorithm: a case study in Weihai, China. *Forests, 14*(10). https://doi.org/10.3390/f14102052

Wang, X. M., Janssens, G. P. J., Xie, C. G., Xie, B. W., Xie, Z. G., He, H. J., & Xu, J. (2022). To save pangolins: a nutritional perspective. *Animals, 12*(22). https://doi.org/10.3390/ani12223137. doi:ARTN 3137.

Wang, X. Q., Gao, H. B., Jia, Z. M., & Li, Z. J. (2023). BL-YOLOv8: an improved road defect detection model based on YOLOv8. *Sensors, 23*(20). https://doi.org/10.3390/s23208361. doi:ARTN 8361.

Wang, Z., Lei, L. F., & Shi, P. B. (2023). Smoking behavior detection algorithm based on YOLOv8-MNC. *Frontiers in Computational Neuroscience, 17*. https://doi.org/10.3389/fncom.2023.1243779. ARTN 1243779.

Willcox, D., Nash, H. C., Trageser, S., Kim, H. J., Hywood, L., Connelly, E., & Challender, D. W. S. (2019). Evaluating methods for detecting and monitoring pangolin (Pholidata: Manidae) populations. *Global Ecology and Conservation, 17*. https://doi.org/10.1016/j.gecco.2019.e00539. ARTN e00539.

Xie, Y. H., Jiang, J. Z., Bao, H., Zhai, P. H., Zhao, Y., Zhou, X. Y., & Jiang, G. S. (2023). Recognition of big mammal species in airborne thermal imaging based on YOLO V5 algorithm. *Integrative Zoology, 18*(2), 333–352. https://doi.org/10.1111/1749-4877.12667

Yang, D. Q., Ren, G. P., Tan, K., Huang, Z. P., Li, D. P., Li, X. W., & Xiao, W. (2021). An adaptive automatic approach to filtering empty images from camera traps using a deep learning model. *Wildlife Society Bulletin, 45*(2), 230–236. https://doi.org/10.1002/wsb.1176

Yang, G. L., Wang, J. X., Nie, Z. L., Yang, H., & Yu, S. Y. (2023). A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention. *AGRONOMY-BASEL, 13*(7). https://doi.org/10.3390/agronomy13071824

Yang, L., Zhang, R.-Y., Li, L., & Xie, X. (2021). *SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks*. Paper presented at the Proceedings of the 38th International Conference on Machine Learning, Proceedings of Machine Learning Research. https://proceedings.mlr.press/v139/yang21o.html.

Yang, M. L., Zhang, X., Guo, Y., Yu, X. W., Hou, Y. A., & Gao, J. J. (2022). Recognition of Wild Animals Using Infrared Camera Images Based on YOLOv5. *Laser & Optoelectronics Progress, 59*(12). doi:Artn 1215015. 10.3788/Lop202259.1215015.

Yang, W. J., Wu, J. C., Zhang, J. L., Gao, K., Du, R. H., Wu, Z., & Li, D. W. (2023). Deformable convolution and coordinate attention for fast cattle detection. *Computers and Electronics in Agriculture, 211*. https://doi.org/10.1016/j.compag.2023.108006. ARTN 108006.

Yi, H., Liu, B., Zhao, B., & Liu, E. H. (2024). Small Object Detection Algorithm Based on Improved YOLOv8 for Remote Sensing. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17*, 1734–1747. https://doi.org/10.1109/Jstars.2023.3339235

Yu, Z., Huang, H., Chen, W., Su, Y., Liu, Y., & Wang, X. (2022). Yolo-facev2: A scale and occlusion aware face detector. *arXiv preprint arXiv:2208.02019.*.

Zhang, J. L., Chen, L. J., Ouyang, B., Liu, B. B., Zhu, J. H., Chen, Y. J., & Wu, D. F. (2022). PointCutMix: regularization strategy for point cloud classification. *Neurocomputing, 505*, 58–67. https://doi.org/10.1016/j.neucom.2022.07.049

Zhang, P., Deng, H. W., & Chen, Z. (2023). RT-YOLO: a residual feature fusion triple attention network for aerial image target detection. *Cmc-Computers Materials & Continua, 75*(1), 1411–1430. https://doi.org/10.32604/cmc.2023.034876

Zhang, Q. Y., Yi, X. L., Guo, J. L., Tang, Y. D., Feng, T., & Liu, R. (2023). A few-shot rare wildlife image classification method based on style migration data augmentation. *Ecological Informatics, 77*. https://doi.org/10.1016/j.ecoinf.2023.102237. ARTN 102237.

Zhang, T., Liu, L. C., Zhao, K., Wiliem, A., Hemson, G., & Lovell, B. (2020). Omni-supervised joint detection and pose estimation for wild animals. *Pattern Recognition Letters, 132*, 84–90. https://doi.org/10.1016/j.patrec.2018.11.002

Zhao, D. D., Yin, H. P., Zhou, H., Cai, L., & Qin, Y. (2024). A zero-sample fault diagnosis method based on transfer learning. *Ieee Transactions on Industrial Informatics, 20*(10), 11542–11552. https://doi.org/10.1109/Tii.2024.3405634

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). *Distance-IoU loss: Faster and better learning for bounding box regression*.

Zhu, F. Z., Wang, Y. Y., Cui, J. Y., Liu, G. X., & Li, H. L. (2023). Target detection for remote sensing based on the enhanced YOLOv4 with improved BiFPN. *Egyptian Journal of Remote Sensing and Space Sciences, 26*(2), 351–360. https://doi.org/10.1016/j.ejrs.2023.04.003

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Ieee International Conference on Computer Vision (Iccv), 2017*, 2242–2251. https://doi.org/10.1109/Iccv.2017.244