International Journal of Science and Engineering Applications
Volume 13-Issue 10, 76 – 81, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1310.1016

YODE-FEIM: An Enhanced YOLOv5s Algorithm for Snake Detection in Wild Environments

Yimeng Xia
School of Electric Information and Electrical Engineering Yangtze University Jingzhou, China

Hao Luo
School of Electric Information and Electrical Engineering Yangtze University Jingzhou, China

Abstract: Addressing the current issues of limited research and low detection accuracy in snake detection, this paper proposes an improved YOLOv5s detection algorithm - YODE-FEIM. Firstly, based on YOLOv5s, the FasterNet Block is combined with the C3 module in the backbone and neck, and the EMA attention mechanism is added at the end of the module, enhancing the feature extraction capability in snake images while reducing parameter computation. Secondly, the detection head is replaced with the RT-DETR detection head, accelerating model convergence and improving detection accuracy. Finally, the Inner-MPDIou loss function is introduced, utilizing boundary regression for localization to enhance the model's detection precision. Experimental results show that on the self-made ChineseSnake dataset, the proposed YODE-FEIM model achieves a precision (P) of 92.7% and a mean average precision (mAP) of 90%, demonstrating high accuracy and providing support for snake detection in wild environments.

Keywords: Snake Dataset; Object Detection; FasterNet Block; EMA Attention Mechanism; Inner MPDIou

1. INTRODUCTION

Snakes are a type of reptile. China is the country with the most abundant snake resources in the world. Snakes have diverse habits, varying according to different species and environments. Regardless of the environment, snakes can easily capture prey with their unique sensory organs and hunting skills. However, snake bites often occur, especially in the wild. Some people cannot accurately identify the species of the snake after being bitten, do not know whether the snake is venomous, and are at a loss after being bitten due to a lack of self-help knowledge. Therefore, there is an urgent need for a method that can promptly detect snakes and provide early warnings. At present, snake identification mainly relies on manual work. However, the living environment of snakes is not stable and is affected by many factors. Moreover, snakes prefer dark and damp places and often have obstacles blocking their vision. Therefore, detecting and identifying them not only consumes a lot of time and manpower but also leads to some errors in judgment.

In recent years, deep learning has made significant progress in machine learning. Object detection methods based on deep learning have shown good performance in many fields and are widely used in many areas. At present, there is less research on snake detection and identification, and snake detection technology based on deep learning has broad application prospects in the field of computer vision. It helps to conduct more accurate research on snakes, thereby better promoting social supervision and protection of snakes, reducing the risk of people suffering from snake venom. This is not only the protection of wild snakes but also an important measure to promote the harmonious development of humans and nature.

Object detection is to find the object to be detected in the image and determine the type and location of the object. However, in practical applications, object detection is often interfered with by various factors (such as the shape of the object, lighting, object occlusion, etc.); therefore, object detection has always been a difficult point in computational vision research. Current object detection technologies are mainly divided into single-stage detection algorithms and two-stage detection algorithms. The two-stage object detection method uses multiple candidate regions as samples, and on this basis, each candidate region is detected and partitioned. R-CNN [1], SPP-Net [2], etc. are representatives of this. The single-stage is a classification method based on neural networks. This method can directly extract the attributes of the target from the network and calculate the category and coordinates of the object on this basis, but the accuracy is not high, and the calculation speed is relatively slow. YOLOv5 [3] is an algorithm that has performed well in the field of object detection in recent years. It can balance speed and accuracy and can play a good role in real-time detection.

Patel and his team's research [4] utilized deep learning technology to achieve real-time identification of snakes on the Galapagos Islands. They collected image data of different snake species on the islands and designed a deep convolutional neural network model for image classification. Experimental results showed that the model could identify different species of snakes with high accuracy, which has significant application value for ecological monitoring and biodiversity conservation. Vasmatkar and others [5] explored methods of snake identification and classification. They collected image data of different snake species and designed a recognition system based on deep learning. The system extracts image features and uses a classifier for identification, achieving accurate classification of snakes. This research provides an effective technical means for snake identification. Rajabizadeh and Rezghi [6] conducted a comparative study, exploring snake identification methods based on machine learning. They compared different image processing and machine learning techniques, including Support Vector Machines (SVM), Random Forests, etc., for snake image classification. The research results showed that deep learning methods outperform traditional machine learning algorithms in snake identification tasks. Ahmed and others [7] focused on using deep learning technology for snake classification. They designed a deep convolutional neural network model and trained and tested it on a large-scale snake image dataset. The model can accurately identify different species of snakes, including some rare and difficult-to-distinguish snake species. This research provides an efficient method for snake identification and has potential value in practical applications.

International Journal of Science and Engineering Applications
Volume 13-Issue 10, 76 – 81, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1310.1016

Although deep learning-based snake recognition technology has made significant progress, there are still some challenges that we need to address. First, the acquisition and annotation cost of snake image data is relatively high, and the diversity and quality of the dataset directly determine the performance of the model, which is an issue that we need to focus on. Secondly, the appearance changes of snakes under different environmental conditions bring additional challenges to recognition, which requires our model to have good environmental adaptability. Finally, for some application scenarios, such as medical emergencies, we need the model to quickly and accurately give recognition results, which puts higher requirements on the real-time performance of the model. Therefore, future research should focus on improving the generalization ability of the model, dealing with recognition problems under complex environmental conditions, and developing larger-scale and higher-quality snake image datasets.

2. METHOD

2.1 YOLOv5 Model

YOLOv5 is a single-stage object detection algorithm that adds some new improvement ideas based on YOLOv4, greatly enhancing its speed and accuracy. Its network structure mainly consists of four parts: Input, Backbone, Neck, and Prediction. The input end mainly preprocesses the input data. The backbone network uses a series of convolutional layers, pooling layers, and Fast Spatial Pyramid Pooling (SPPF) to extract and fuse feature maps at different scales to improve the detection accuracy of the model. The neck network enhances the model's detection capability by fusing different features through top-down semantic information transmission and bottom-up position information transmission. The output end outputs the calculated network prediction results. Starting from YOLOv5n, the detection accuracy of the model gradually increases, but the depth and width of the model also increase in turn, leading to an increase in model complexity and the number of parameters, which affects the detection speed. The YOLOv5 model has five versions, specifically: YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x. This paper mainly uses YOLOv5s as the basic network model, and its basic network structure is shown in Figure 1.

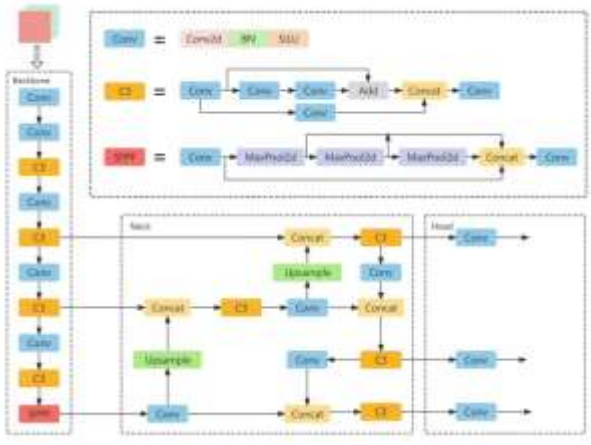

Figure 1. The structure of YOLOv5s model.

2.2 YODE-FEIM Model

To solve the problem of low accuracy in current snake image detection, this paper proposes an innovative snake detection model named YODE-FEIM, based on the YOLOv5s model. The specific network model structure is shown in Figure 2.

- This model mainly integrates FasterNet Block into the C3 module in the backbone and neck, and introduces the EMA attention mechanism at the tail. This unique design strategy enables the model to effectively integrate multi-scale features, greatly enhancing the full utilization of multi-scale features, thereby better extracting more features from the image, and further improving the detection accuracy of the model.
- In addition, we use the detection head of RT-DETR for detection. This improvement not only improves the detection accuracy of snake images, but also enhances the model's ability to recognize and locate snake features in complex environments.
- Finally, we also introduce the Inner MPDIoU loss function on this basis. This novel loss function can enhance the model's ability to locate features, further improve the model's detection performance, and make the model more reliable and stable in practical applications.

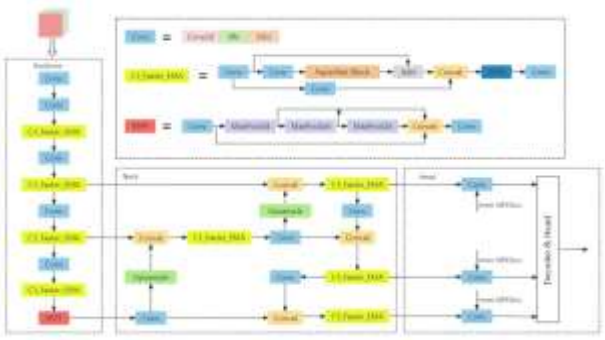

Figure 2. The structure of YODE-FEIM model.

2.2.1 C3-FasterNet-EMA

In the latest version of the YOLOv5 model, the BottleneckCSP module is replaced with the C3 module. Although these two modules are fundamentally similar in structure and function, both belonging to the CSP architecture, there are differences in the choice of correction units. The C3 module contains three standard convolutional layers and multiple Bottleneck modules. The C3 module plays a key role in enhancing the receptive field, etc., however, its large number of parameters and computational complexity severely limit the detection speed of the model, making it difficult to meet the real-time requirements of snake detection tasks in the wild.

To address this challenge, this study proposes a new strategy, namely, using the lightweight network FaterNet [8], combining the C3 module with the FasterNet module, and introducing the EMA attention mechanism [9] at the end of the module, to construct the C3_Fasternet_EMA module, reduce the amount of computation, and achieve the fusion of multi-scale spatial and positional information. This method can capture the mapping relationship between each pixel point in the image, highlight the overall context in the image, thereby further enhancing the feature extraction ability of the model and improving the detection accuracy of the model. This design strategy not only improves the detection accuracy of the model but also improves the running efficiency of the model, making it better meet the requirements of real-time.

2.2.1.1 FasterNet Block

The lightweighting of the model can reduce the complexity of computation while ensuring accuracy. A new partial convolution (PConv) is proposed, which aims to improve the extraction efficiency of spatial features on the basis of reducing

storage overhead. On this basis, we will use a method that combines partial convolution with pointwise convolution to construct a new type of neural network. The basic structure of FasterNet is shown in the figure below. After the input feature map, it is first subjected to local convolution operations, then pointwise convolution, batch normalization and activation, and then pointwise convolution, to obtain richer features while reducing the computational cost.

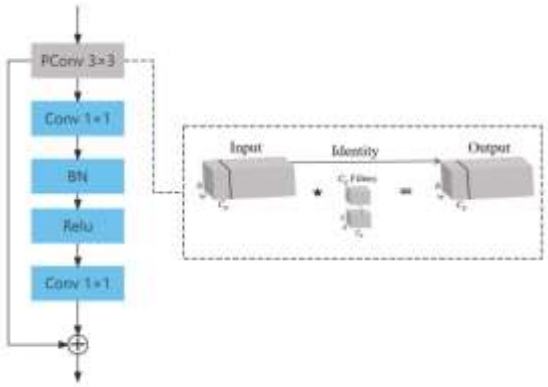

Figure 3. The structure of FasterNet block.

2.2.1.2 *EMA*

We adopted an innovative cross-space learning method and designed a multi-scale parallel subnetwork to establish short-term and long-term dependencies. We found that modeling cross-channel relationships through channel dimension reduction may have side effects on the extraction of deep visual representations. Therefore, we propose a new and efficient multi-scale attention (EMA) module. This module aims to retain information on each channel while reducing computational overhead. We reshape part of the channel into a batch dimension and group the channel dimension into multiple sub-features to evenly distribute spatial semantic features in each feature group. In addition to building local cross-channel interactions in each parallel subnetwork, we also fused two parallel output feature maps through a cross-space learning method. The EMA module encodes global information in parallel branches for channel weight recalibration, thereby enhancing the ability of feature representation. Compared with common CBAM, SA, ECA, and CA, the EMA module achieves better results and is more efficient.

In the EMA module, the input is first grouped, and then processed through different branches: one branch performs one-dimensional global pooling through 1×1 convolution, and the other extracts features through 3×3 convolution. The output features of the two branches are modulated through the sigmoid function and normalization operation, and finally merged through the cross-dimension interaction module to capture pixel-level pairwise relationships. The EMA module adopts a cross-space dimension information aggregation method, encodes the global spatial information output by the convolution branch through 2D global average pooling, and converts the output channel features of the smallest branch into the corresponding dimension shape. On the output of 2D global average pooling, the EMA module uses the SoftMax function to fit the linear transformation. Finally, the results of parallel processing are multiplied together through matrix dot product operations to obtain the first spatial attention map. The EMA attention mechanism enhances the model's ability to focus on target features, enabling the model to acquire more feature information, thereby improving the model's small target detection accuracy.

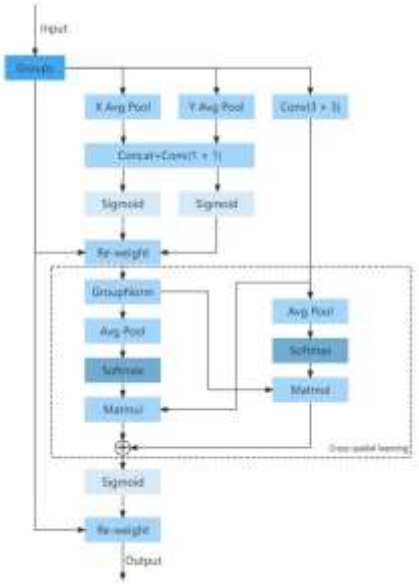

Figure 4. The structure of EMA.

2.2.2 *The detection head of RT-DETR*

RT-DETR [10] is a real-time object detection model that integrates two classic object detection methods: Transformer and DETR (Detection Transformer). The Transformer is a neural network architecture for sequence modeling, originally designed for natural language processing, however, its effectiveness in the field of computer vision has been proven. DETR is an end-to-end object detection model that transforms the object detection task into an object query problem and uses a Transformer to solve it. However, the computational cost of DETR is high, making it unsuitable for real-time application scenarios. RT-DETR inherits the structure of DETR but adopts some optimization measures to achieve real-time object detection.

Although YOLO, as a widely used object detection algorithm, its detection speed is not satisfactory, the emergence of RT-DETR provides a solution to this problem. RT-DETR can significantly improve the detection speed while ensuring detection accuracy. This is mainly due to its unique detection head design, which can effectively reduce unnecessary computational redundancy, thereby accelerating the convergence speed of model training. Therefore, this study chooses to use the detection head of RT-DETR in the hope of improving the performance of the model while also improving the efficiency of the model. This method can not only improve the practicality of the model but also improve the practicality and usability of the model while ensuring the performance of the model, thereby better serving practical applications.

2.2.3 *Loss Function - Inner MPDIoU*

This paper introduces a new bounding box regression loss function, Inner MPDIOU, to optimize the convergence speed of the network and improve the localization accuracy of snake bounding boxes. MPDIoU [11] , The full name is minimum point distance intersection to union ratio, which is a boundary box similarity measurement method based on minimum point distance. This measurement method calculates the intersection and union of two bounding boxes, directly predicting the distance between the top left and bottom right points of the bounding box and the real box, thereby simplifying the

International Journal of Science and Engineering Applications
Volume 13-Issue 10, 76 – 81, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1310.1016

similarity comparison between the two bounding boxes. This method allows the model to adapt to both overlapping and non overlapping bounding box regression, taking into account factors such as position, shape, and scale, and more accurately measuring the overlap between two bounding boxes.

Inner IoU [12] , The full name is internal intersection ratio, which is a method specifically used to measure the degree of overlap within bounding boxes. It only considers the intersection inside the bounding box and ignores the union outside, so it focuses more on the precise matching degree of the bounding box.

We combine these two methods to propose the Inner MPDIoU loss function. This new loss function considers both the shape and scale of the bounding box and the degree of overlap inside the bounding box during the calculation process, thus more accurately measuring the similarity of the bounding box. Through experiments, we found that compared with traditional loss functions, the Inner MPDIOU loss function can significantly improve the convergence speed of the model and the localization accuracy of bounding boxes, especially when dealing with snake like borders, its performance advantages are more obvious.

3. EXPERIMENTS

3.1 Experimental Environment and Parameters

The experimental environment of this paper is shown in Table 1. After preprocessing, the image size processed by the model is consistently 640×640. The batch size is set to 8, and the number of iterations is 300. The initial learning rate is 0.01, and the decay rate is 0.0005.

Table. 1 Experimental environment.

Operating system	Windows11
CPU	i5-12400F
GPU	NVIDIA GeForce RTX 3060 12G
CUDA	11.6
framework	Pytorch1.13.1
Python	3.8.0
Pycharm	2023.1.3（community）

These settings ensure that the model can effectively learn the features of the input images and converge to a good solution within a reasonable amount of time. The learning rate and decay rate are chosen to balance the speed of convergence and the stability of the learning process. The batch size and the number of iterations are set based on the computational resources available and the complexity of the task.

3.2 Data Set

Considering the scarcity of publicly available snake datasets and the diversity of snake species, field shooting is not only costly but also poses significant safety risks. Therefore, this study chose a more economical and safe way to collect data. We collected and organized images of nearly ten common snake species in China from the web page (https://www.flickr.com/), and built a brand new dataset specifically for Chinese snakes, named ChineseSnake. This dataset contains a total of 4027 snake images, covering ten species of snakes, aiming to provide more abundant and comprehensive data support for snake image recognition research. To better train and validate the model, we divided the dataset in a 7:3 ratio, with 70% of the images used for training and 30% of the images used for validation. The specific composition of the dataset is shown in Table 2.

Figure 5. Data set.

Table. 2 Composition of the ChineseSnake Dataset.

Snake Species	Number of Images(sheet)
Hydrophiinae	52
Trimeresurus albolabris	498
Protobothrops	399
Rhabdophis subminiatus	458
Ophiophagus hannah	424
Daboia siamensis	386
Deinagkistrodon	371
Bungarus multicinctus	488
Bungarus fasciatus	283
Sinomicrurus macclellandi	218

International Journal of Science and Engineering Applications
Volume 13-Issue 10, 76 – 81, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1310.1016

Total	4027

3.3 Experimental Results

The experimental results of this paper show that the model has achieved a precision (P) of 0.93 and a recall rate (R) of 0.87. The specific data is shown in Table 3. This means that the model can identify snakes with high accuracy, but the recall rate is slightly lower, and there may be some cases of missed detection. The mean average precision (mAP) is 0.90 under the IoU threshold of 0.5, and 0.74 in the IoU threshold range of 0.5 to 0.95. This indicates that the model performs well under more lenient matching standards (IoU=0.5); however, the performance decreases under stricter matching standards.

Table. 3 Snake Species Detection Performance.

Snake Species	P	R	mAP 0.5	mAP 0.5:0.95
all	0.93	0.87	0.90	0.74
Hydrophiinae	0.92	0.88	0.94	0.74
Trimeresurus albolabris	0.93	0.86	0.90	0.78
Protobothrops	0.98	0.97	0.97	0.78
Rhabdophis subminiatus	0.94	0.90	0.93	0.74
Ophiophagus hannah	0.99	0.89	0.92	0.86
Daboia siamensis	0.94	0.81	0.86	0.67
Deinagkistrodon	0.85	0.83	0.84	0.69
Bungarus multicinctus	0.87	0.70	0.74	0.59
Bungarus fasciatus	0.87	0.94	0.97	0.86
Sinomicrurus macclellandi	0.97	0.92	0.95	0.72

These results suggest that our model is highly effective at identifying snakes in images, but there is still room for improvement, particularly in terms of recall and performance under stricter IoU thresholds. Future work will focus on addressing these issues, with the aim of developing a model that is both highly accurate and robust to variations in the input data.

The model performs well in identifying most snake species, especially in identifying Protobothrops and Ophiophagus hannah. However, for some snake species, such as Daboia siamensis, Deinagkistrodon, and Bungarus multicinctus, the model's recall rate is relatively low, and it may be necessary to further optimize the model to improve the recognition performance of these categories. In addition, the performance drop of the model under stricter IoU thresholds indicates that there may be room for improvement in the model's precise matching. Figure 6 shows the comparison between the original YOLOv5s and the improved YODE FEIM in snake detection results.

In future work, we will focus on improving the model's performance on these challenging species and on enhancing its ability to accurately match bounding boxes. We will also consider incorporating additional information, such as the snake's habitat or behavior, to help the model distinguish between similar species. Furthermore, we will explore the use of more advanced techniques, such as attention mechanisms or transformer architectures, to further improve the model's performance.

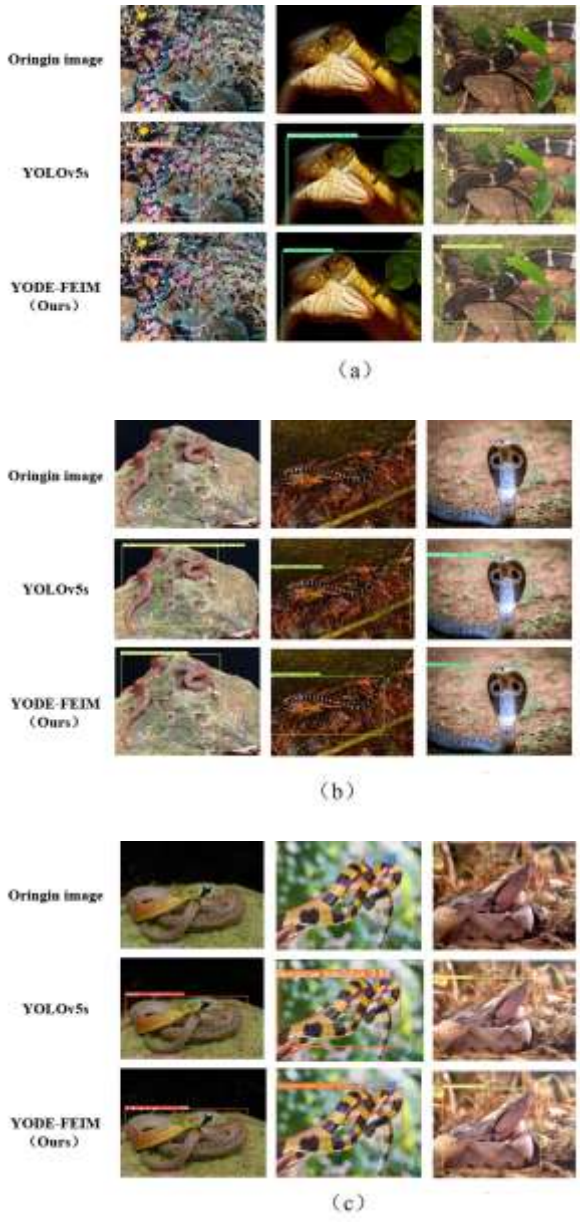

Figure 6. Comparison of experimental results.

4. CONCLUSION

To solve the problem of low accuracy in detecting snakes in the wild, this study made deep improvements on the basis of the YOLOv5s model and proposed a new detection model named YODE-FEIM. We first organically combined the FasterNet

module with the C3 module in YOLOv5s in the hope of enhancing the feature extraction capability of the model. At the same time, we introduced the EMA attention mechanism at the end of the model. This mechanism can further enhance the ability to extract target features in the image, thereby improving the recognition accuracy of the model. We use the RT-DETR detection head for detection. This detection head is efficient and accurate, which can further improve the detection performance of the model. Finally, we introduced the Inner MPDIou loss function. This loss function can effectively reduce the positioning error, thereby improving the positioning accuracy of the model. The experimental results on the homemade dataset show that the accuracy of the model algorithm we proposed has been significantly improved. However, its real-time performance and robustness are slightly insufficient, which is where we need to further improve in future research. In future work, we will consider improving the real-time detection speed, expanding the dataset, and performing detection under other complex scenarios to enhance the generalization ability of the model.

5. REFERENCES

[1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[2] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

[3] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv5: Improved real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 1450-1459.

[4] Patel A, Cheung L, Khatod N, et al. Revealing the unknown: Real-time recognition of Galápagos snake species sing deep learning[J]. Animals, 2020, 10(5): 806.

[5] Vasmatkar M, Zare I, Kumbla P, et al. Snake species identification and recognition[C]//2020 IEEE Bombay Section Signature Conference (IBSSC). IEEE, 2020: 1-5.

[6] Rajabizadeh M, Rezghi M. A comparative study on image-based snake identification using machine learning[J]. Scientific reports, 2021, 11(1): 19142. 4.

[7] Ahmed K, Gad M A, Aboutabl A E. Snake species classification using deep learning techniques[J]. Multimedia Tools and Applications, 2024, 83(12): 35117-35158.James A P, Mathews B,

[8] Chen J, Kao S, He H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 12021-12031.

[9] Ouyang D, He S, Zhang G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[10] Zhao Y, Lv W, Xu S, et al. Detrs beat yolos on real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 16965-16974.

[11] Ma S, Xu Y. Mpdiou: a loss for efficient and accurate bounding box regression[J]. arXiv preprint arXiv:2307.07662, 2023.

[12] Zhang H, Xu C, Zhang S. Inner-IoU: more effective intersection over union loss with auxiliary bounding box[J]. arXiv preprint arXiv:2311.02877, 2023.