International Journal of Science and Engineering Applications
Volume 14-Issue 03, 09 – 14, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1403.1004

Aggressive Dog Breed Recognition Based on Improved YOLOv8n

Yin Qinglin
College of Electronic Information and Electrical Engineering
Yangtze University
Jingzhou City, Hubei Province
China

Abstract: Aiming at the problems of dogs attacking, an improved YOLOv8n-based model for aggressive dog breed recognition is proposed to enhance recognition accuracy and real-time performance. First, a dataset of aggressive dog breeds containing various complex scenes and lighting conditions is constructed, covering 10 common banned dog breeds, providing high-quality samples for model training. Second, the YOLOv8n network is optimized: the GSConv module is introduced into the Backbone to significantly reduce computation and improve feature extraction capabilities; the VoV-GSCSP module is introduced into the Neck to optimize feature fusion strategies; and the EIoU loss function is adopted to accelerate model convergence and improve bounding box prediction accuracy. Experimental results show that the improved model outperforms the original YOLOv8n model in terms of accuracy, mAP50, and mAP50:95, and demonstrates higher recognition accuracy and comprehensive performance compared with other mainstream object detection models.

Keywords: aggressive dog breed recognition; YOLOv8n; GSConv; VoV-GSCSP; EIoU

1. INTRODUCTION

In recent years, with the acceleration of urbanization, the number of dog owners has been increasing continuously, and the problem of dog management has become increasingly prominent. Among them, the frequent occurrence of attacks by aggressive dogs has posed a serious threat to public safety. Traditional management of aggressive dogs mainly relies on post-incident accountability and legal prohibitions, lacking effective early warning means[1]. Against this backdrop, the use of advanced technical means to achieve rapid recognition and early warning of aggressive dogs has become the key to solving this problem. The progress of deep learning technology in the field of image recognition has provided new ideas and methods for the recognition of aggressive dog breeds.

In recent years, the development of object detection algorithms based on deep learning convolutional neural networks has been rapid. According to the number of detection stages, they are divided into two categories. One category is the two-stage detection algorithm, which needs to generate candidate regions first and then recognize the candidate regions. Therefore, there are problems such as long detection time and inability to meet real-time requirements. Specifically, there are methods such as R-CNN[2] proposed by Girshick R, Fast-RCNN[3], and Faster-RCNN[4]; The other category is the one-stage detection algorithm, specifically the YOLO[5-8] series, SSD[9-11] series, and OverFeat[12]. Among them, YOLOv8 has higher accuracy and smaller parameter quantity and model size. Therefore, this paper improves and optimizes the aggressive dog breed recognition task based on the YOLOv8n network.

In the field of dog breed recognition, Punyanuch Borwarnginn et al. used transfer learning and data augmentation techniques to fine-tune pre-trained CNNs such as MobileNetV2, InceptionV3, and NASNet. The NASNet model achieved an accuracy of 89.92% on the rotated image dataset, proving the high efficiency of deep learning in dog breed recognition[13]. Prasanth Vaidya S. et al. proposed a dog breed recognition method based on CNN, combining OpenCV, VGG16, and ResNet101 architectures. Through transfer learning and data augmentation techniques, the model accuracy was increased from 16% to 81%, which is significantly better than traditional methods[14]. Xu Pingting et al. expanded the dataset and used transfer learning techniques to improve the recognition rate of aggressive dogs using convolutional neural networks such as VGG16 and ResNet50. The final ResNet50 model achieved an accuracy rate close to 90%, realizing the high-accuracy recognition and early warning function of aggressive dogs[15]. B. Valarmathi et al. compared various deep learning algorithms (such as Xception, VGG19, NASNetMobile, etc.) and hybrid models, and found that the hybrid model of Inception-v3 and Xception achieved an accuracy rate of 92.4% on the Stanford Dog Dataset, which is better than traditional single models[16]. However, transfer learning requires high data set diversity and quality, and hybrid models have complex structures and are difficult to train. Huang Xuelei et al. improved the YOLO model for the phenomenon of illegal pet walking, using SENet feature network and SPPCSPC module to optimize feature extraction. Experiments show that the improved network can accurately recognize illegal pet walking behavior and meet the needs of practical applications[17]. Although certain progress has been made in the recognition of aggressive dog breeds in recent years, there are still only a few studies in the academic community, and existing research still faces many challenges, such as insufficient detection accuracy in complex backgrounds and difficulties in small target recognition. Shi Xinyu et al. used YOLOv8 for breed classification and combined it with the ShuffleNet-V2 twin network with ECA attention mechanism for individual recognition, optimizing the fine-grained recognition accuracy of cats and dogs in complex backgrounds[18]. Jie Hu et al. optimized the detection accuracy of traffic signs in complex backgrounds by introducing parallel deformable convolution module (PDCM), sub-pixel convolution attention module (SCAM) and GSConv module[19]. Guojun Chen et al. optimized the detection accuracy of small target melons in complex backgrounds and color changes by introducing lightweight Faster-Block, EMA module, improved detection head and α-IoU loss function[20].

In response to the current problems of aggressive dog breed recognition and referring to the experience and shortcomings of scholars in other object classification tasks, this study makes the following improvements based on the YOLOv8n algorithm to improve the accuracy and real-time performance of aggressive dog breed recognition and provide more effective technical support for the prevention of aggressive dog attacks:

(1) Self-made aggressive dog breed dataset: In response to the problem of insufficient existing datasets, this paper has created a high-quality dataset containing various aggressive dog breeds, covering complex natural scenes and lighting conditions, providing richer sample support for model training.

(2) Optimizing network structure: Introducing the GSConv module into the Backbone of YOLOv8n and the VoV-GSCSP module into the Neck part, which significantly reduces the model's computational amount and inference time while enhancing object detection capabilities.

(3) Improving loss function: Replacing the traditional CIoU loss function with the EIoU loss function optimizes the convergence speed of object bounding box regression, generates more accurate predicted bounding boxes, and thus further improves the model's detection accuracy.

2. AGGRESSIVE DOG BREED DATASET

To achieve efficient recognition of aggressive dog breeds, this study constructs an aggressive dog breed dataset for training and validating the improved YOLOv8n model. The images in the dataset are collected from the internet and taken on-site, covering the 10 most common aggressive dog breeds banned in urban areas of China. As shown in Figure 1, the breeds are: (a) Spanish Canary Dog, (b) Canis lupus familiaris, (c) Tibetan Mastiff, (d) Wolf Dog, (e) Central Asian Shepherd, (f) Caucasian Shepherd Dog, (g) Neapolitan Mastiff, (h) Standard Bull Terrier, (i) Pit Bull Terrier, (j) Rhodesian Ridgeback.

Figure. 1 Illustration of aggressive fog breeds dataset

The dataset includes images of aggressive dog breeds under various scenes and lighting conditions. To ensure the quality of the data and the accuracy of the annotations, all images are manually annotated using the Labelme tool. The annotation information for each image includes the dog breed category and the corresponding bounding box, which is used for the subsequent object detection task. The dataset contains 2874 images, and it is divided into training and validation sets in a ratio of 8:1.

3. IMPROVED YOLOV8N ALGORITHM

3.1 Improved Algorithm Structure

Based on different network depths and widths, YOLOv8 is divided into YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. Considering the model size, this paper selects the small-sized and high-precision YOLOv8n network. The YOLOv8n detection network mainly consists of four parts: Input, Backbone, Neck, and Head.

Input: Mosaic data augmentation is used for the input, but it is turned off in the last 10 epochs. The anchor-free mechanism directly predicts the center of the object rather than the offset of known anchor boxes, reducing the number of anchor box predictions and thus accelerating non-maximum suppression (NMS)[21]. Backbone: This part is mainly used for feature extraction and includes modules such as Conv, C2f, and SPPF. The Conv module performs convolution, batch normalization (BN), and SilU activation function operations on the input image. YOLOv8n has a newly designed C2f structure, which is the main module for learning residual features. This structure allows YOLOv8n to maintain rich gradient flow information while being lightweight. The SPPF module, also known as spatial pyramid pooling, can convert feature maps of any size into fixed-size feature vectors. Neck: The primary function of the Neck is to fuse multi-scale features and generate a feature pyramid. The structure used in the Neck is the PANet structure, which consists of two core parts: the feature pyramid network (FPN)[22] and the path aggregation network (PAN)[23]. FPN first constructs a feature pyramid by extracting feature maps from the convolutional neural network and then fuses features from different levels using upsampling and coarser-grained feature maps in a top-down manner. However, FPN alone lacks spatial information about the target. PAN complements FPN with a bottom-up structure, using a convolutional layer to fuse feature maps from different levels and accurately preserve spatial information. The combination of FPN and PAN fully integrates the network's up-down information flow, enhancing detection performance. Head: As the final prediction part, the Head outputs the class and position information of objects of different sizes based on feature maps of different dimensions.

In this study, the GSConv module is introduced into the Backbone of YOLOv8n. This is an innovative lightweight convolution technique that combines the advantages of standard convolution (SC) and depthwise separable convolution (DSC), significantly enhancing feature representation capabilities while greatly reducing computational load. Specifically, GSConv mixes the channel-dense features of standard convolution with the spatially sparse features of depthwise separable convolution and further fuses information through feature rearrangement (such as Shuffle operations). This maintains low computational complexity while preserving as many hidden connections between channels as possible. Secondly, in the Neck part, the VoV-GSCSP module is introduced to further optimize feature fusion strategies and pyramid pooling structures. Based on GSConv, the VoV-GSCSP module adopts an efficient cross-stage partial network (CSP) structure, reducing feature redundancy and optimizing feature reuse rates, further lowering computational complexity and inference time. Additionally, the module's one-stage aggregation strategy makes feature fusion more efficient while maintaining model accuracy. Finally, the EIoU loss function is adopted to improve the algorithm's accuracy and performance. The improved YOLOv8n algorithm structure is shown in Figure 2.

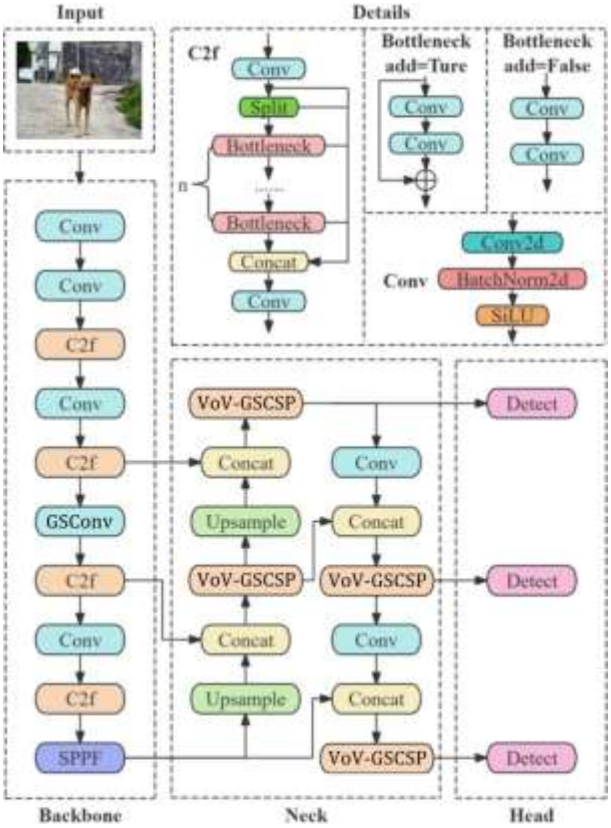

Figure. 2 Structure of the improved YOLOv8n algorithm

3.2 GSConv

In 2022, Hulin Li[24] and others proposed the GSConv module. In order to speed up predicting in the end, fed images almost always have to undergo a similar transformation process in the backbone: transferring the spatial information toward the channels step by step. Each time, the spaces dimension (the width and height) compression and channels dimension expansion of the features will cause a partial loss of semantic information. The channel-dense convolutional maximally preserves the hidden connections between each channel, but the channel-sparse convolution severs these connections completely. The GSConv preserves these connections as much as possible with lower time complexity. As show in Figure 3, the authors use the shuffle to permeate the features generated by the SC (the channel- dense convolutional) into every part of the features generated by the DSC. The shuffle is a uniform mixing and allows the information from the SC to be fully mixed into the outputs of the DSC by uniformly to exchange local feature information on different channels, without bells and whistles.

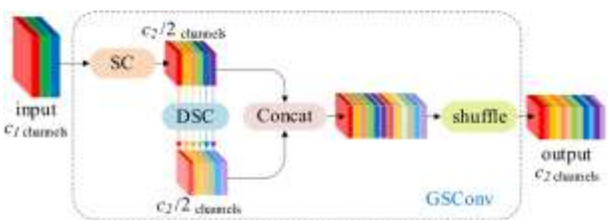

Figure. 3 Structure of the GSConv

3.3 VoV-GSCSP

The computational cost of GSConv is about 50% of the SC, but the contribution to the model learning ability is comparable to the SC. Based on GSConv, The same authors as GSConv continue to introduce the GS bottleneck, the structure showed in Figure 4 (a). Then, they investigate the generalized methods to enhance the learning ability of CNNs and use the one-shot aggregation strategy to design the efficient cross stage partial network (CSP) module, VoV-GSCSP, to reduce the computational complexity and inference time but maintain the accuracy. Figure 4 (b) shows the design structure for the VoV-GSCSP.

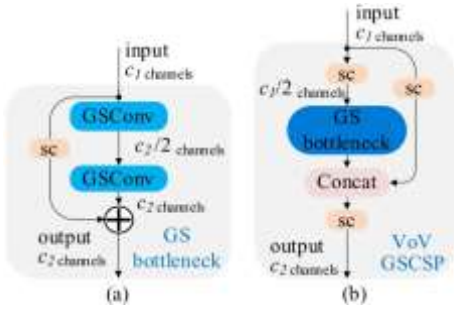

Figure. 4 Structure of the VoV-GSCSP

3.4 EIoU Loss Function

In neural networks, the loss function is indispensable. In object detection algorithms, the Intersection over Union (IoU) is commonly used to measure the overlap between predicted boxes and ground-truth boxes. The loss function used in YOLOv8n is derived from IoU, known as CIoU[25]. The traditional IoU calculation only considers the positional information of the bounding boxes, while CIoU normalizes the Euclidean distance between the centers of two bounding boxes, penalizing distant bounding boxes to more accurately measure the similarity between them.

The EIoU Loss function[26] further refines this by decomposing the aspect ratio and incorporating the differences in width and height to calculate the dimensions of the target and anchor boxes. It also introduces Focal Loss to address the imbalance between easy and hard samples, significantly accelerating convergence. This optimization allows for better differentiation between overlapping and occluded objects, reducing the probability of mismatches and enhancing the robustness and stability of object detection algorithms in complex scenarios.

Therefore, this paper adopts the EIoU Loss function, which introduces the scale information of bounding boxes and considers their overlapping parts, resulting in more accurate and reliable calculations.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Experimental Platform

The hardware environment for the training and validation of the aggressive dog breed dataset in this study is as follows: The CPU is the 12th Gen Intel(R) Core(TM) i5-12400F, with 16G of memory and an NVIDIA GeForce RTX 3060 Ti graphics card featuring 8G of video memory. The operating system is 64-bit Windows 11. The compilation platform is PyCharm, and the programming is implemented in Python 3.8. The training framework is PyTorch, with a version of torch 2.0.0+cu117. The experimental parameters are shown in Table 1, and any parameters not mentioned in this paper are set to the default values provided by the official YOLOv8.

Table 1. Experimental Parameters

Parameter Name	Parameter Value
Optimizer	SGD
Batch_Size	Good
Epochs	64
Imges_Size	200
Learning rate	640
Amp	TRUE

4.2 Evaluation Metrics

In this experiment, the model is evaluated using Precision, Recall, and mean Average Precision (mAP). Precision is the ratio of true positives (TP) to all predicted positives (TP + FP), and the calculation formula is shown in Equation (1):

$$P = \frac{TP}{TP+FP} \quad (1)$$

Recall is the ratio of true positives (TP) to all actual positives (TP + FN), and the calculation method is shown in Equation (2):

$$R = \frac{TP}{TP+FN} \quad (2)$$

In mean Average Precision (mAP), m represents the mean, and AP@0.5 refers to the average precision for a class of samples when the IoU threshold in the confusion matrix is set to 0.5. mAP@0.5 is the average of the Precision values for all classes of samples, reflecting the trend of the model's Precision with respect to Recall. A higher value indicates that the model can maintain high Precision at high Recall levels. mAP@0.5:0.95 represents the average mAP over different IoU thresholds (from 0.5 to 0.95 with a step of 0.05). The calculation method is shown below:

$$AP@0.5 = \frac{1}{n}\sum_{i=1}^{n} P_i = \frac{1}{n}P_1 + \frac{1}{n}P_2 + \cdots + \frac{1}{n}P_n \quad (3)$$

$$mAP@0.5 = \frac{1}{C}\sum_{k=1}^{C} AP@0.5_k \quad (4)$$

$$mAP@0.5{:}0.95 = \frac{1}{10}mAP@0.5 + \frac{1}{10}mAP@0.55 + \cdots + \frac{1}{10}mAP@0.95 \quad (5)$$

4.3 Ablation Experiment

To verify the effectiveness of the proposed model improvements for the aggressive dog breed dataset, an ablation study was designed to analyze and compare the following eight configurations:1)Original YOLOv8n model; 2)YOLOv8n model with GSConv introduced in the Backbone; 3)YOLOv8n model with VoV-GSCSP introduced in the Neck; 4)YOLOv8n model with the loss function changed to EIoU Loss Function; 5)YOLOv8n model with GSConv in the Backbone and VoV-GSCSP in the Neck; 6)YOLOv8n model with GSConv in the Backbone and the loss function changed to EIoU Loss Function; 7)YOLOv8n model with VoV-GSCSP in the Neck and the loss function changed to EIoU Loss Function; 8)YOLOv8n model with GSConv in the Backbone, VoV-GSCSP in the Neck, and the loss function changed to EIoU Loss Function.Under the same experimental conditions, experiments were conducted on the aggressive dog breed dataset presented in this paper, and the results are shown in Table 2.

Table 2. Results of Ablation Experiment

model	GSConv	VoV-GSCSP	EIoU	P/%	R/%	mAP0.5/%	mAP0.5:0.95/%
YOLOv8n	×	×	×	91.6	92.6	95.6	87
	√	×	×	92.2	92.5	95	87.2
	×	√	×	91.8	92.4	95.8	87.4
	×	×	√	93.1	91.6	96	87.8
	√	√	×	91.1	90.9	96.1	88.6
	√	×	√	93.5	92.2	94.8	86.5
	×	√	√	92.8	92.6	96.2	88.1
	√	√	√	**94.5**	**92.6**	**96.6**	**88.9**

Based on the results of the ablation study, the GSConv, VoV-GSCSP, and EIoU modules have a significant impact on the performance improvement of the YOLOv8n model. From the experimental data, when the EIoU module is introduced alone, the model's precision (P/%) and mAP50 metrics show the most significant improvement, reaching 93.1% and 96%, respectively. This indicates that EIoU plays an important role in enhancing detection accuracy. Additionally, the VoV-GSCSP module, when used alone, also has a positive effect on the mAP50 and mAP50:95 metrics, increasing them to 95.8% and 87.4%, respectively. This demonstrates its advantage in optimizing feature fusion and overall detection performance. When all three modules are integrated into the YOLOv8n model, the model's comprehensive performance reaches its best, with precision (P/%) increasing to 94.5%, mAP50 increasing to 96.6%, and mAP50:95 increasing to 88.9%. This indicates that there is a good synergy between the GSConv, VoV-GSCSP, and EIoU modules, which can significantly enhance the model's detection accuracy and robustness. This combination optimizes the model's detection capabilities across different thresholds, enabling it to perform with higher adaptability and reliability in complex scenarios.

4.4 Comparison Experiment

To objectively demonstrate the advantages of the improved model, a comparison experiment was conducted between the improved model and other mainstream models currently in use. The models involved in the comparison include Faster-RCNN, SSD, OverFeat, YOLOv5, YOLOX, YOLOv7, YOLOv8n, and the model proposed in this paper. All models were trained and validated using the aggressive dog breed dataset presented in this paper. The comparison results of the models are shown in Table 3. Compared with other models, the precision of the proposed model is increased by 21.3%, 12%, 25.9%, 6.3%, 4.9%, 4.3%, and 2.9% respectively. The recall rate is increased by 27.3%, 14.5%, 27%, 6.2%, 5.3%, 3.5%, and 0% respectively. The mAP50 is increased by 23.4%, 15.4%, 27.5%, 8.8%, 8.5%, 4.7%, and 1% respectively. The mAP50:95 is increased by 46.8%, 33.6%, 50.1%, 23.2%, 20%, 15.6%, and 1.8% respectively. The proposed model outperforms other existing object detection models in all evaluation metrics, indicating that it has higher precision and better comprehensive performance in the task of aggressive dog breed recognition.

Table 3. Results of Comparison Experiment

model	P/%	R/%	mAP50/%	mAP50:95/%
Faster-RCNN	73.2	65.3	73.2	42.1
SSD	82.5	78.1	81.2	55.3
OverFeat	68.6	65.6	69.1	38.8
YOLOv5	88.2	86.4	87.8	65.7
YOLOX	89.6	87.3	88.1	68.9
YOLOv7	90.2	89.1	91.9	73.3
YOLOv8n	91.6	92.6	95.6	87.1
proposed	**94.5**	**92.6**	**96.6**	**88.9**

5. CONCLUSION

This paper presents an improved YOLOv8n-based detection model for the important and challenging task of aggressive dog breed recognition. First, a dataset of aggressive dog breeds covering various complex scenes and lighting conditions was constructed to provide rich sample support for model training. On this basis, several improvements were made to the YOLOv8n network: the GSConv module was introduced into the Backbone to effectively reduce computation and enhance feature extraction capabilities; the VoV-GSCSP module was introduced into the Neck to further optimize feature fusion strategies; and the EIoU loss function was adopted to accelerate model convergence and improve bounding box prediction accuracy. Ablation and comparison experiments were conducted to verify the effectiveness of the improved model. The results show that the improved model significantly outperforms the original YOLOv8n in terms of precision, mAP50, and mAP50:95. Moreover, compared with models such as Faster-RCNN, SSD, and YOLOv5, the proposed model demonstrates higher precision and better comprehensive performance in all evaluation metrics.

The improved model can be widely deployed in urban surveillance systems, such as community security systems, animal shelters, pet hospitals, and mobile law enforcement devices for law enforcement officers, to provide security for community residents, medical staff, and law enforcement personnel and enhance the efficiency and safety of aggressive dog management. In the future, the dataset will be further expanded to cover as many aggressive dog breeds as possible to achieve more efficient and accurate recognition of aggressive dog breeds.

6. REFERENCES

[1] Chen Qianya. Current Challenges and Countermeasures in Urban Dog Management [J]. China Working Dog Industry, 2024(3): 9-12.

[2] AGRAWAL P, GIRSHICK R, MALIK J. Analyzing the performance of multilayer neural networks for object recognition[C]//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part VII 13. Springer International Publishing, 2014: 329-344.

[3] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[4] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[6] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[7] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[8] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[9] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands October, 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[10] FU C Y, LIU W, RANGA A, et al. Dssd: Deconvolutional single shot detector[J]. arXiv preprint arXiv: 1701.06659, 2017.

[11] LI Z, ZHOU F. FSSD: feature fusion single shot multibox detector[J]. arXiv preprint arXiv:1712.00960, 2017.

[12] SERMANET P, EIGEN D, ZHANG X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks[J]. arXiv preprint arXiv:1312.6229, 2013.

[13] Borwarnginn P, Kusakunniran W, Karnjanapreechakorn S, et al. Knowing your dog breed: identifying a dog breed with deep learning[J]. International Journal of Automation and Computing, 2021, 18: 45-54.

[14] Vaidya P, Mouli C, Kishore R. A novel Dog Breed Identification using Convolutional Neural Network[J]. PriMera Scientific Engineering, 2022.

[15] Xu Pingting, Zheng Jiaqi. Aggressive Dog Detection and Alert System Based on Deep Learning Algorithms [J]. Software, 2023, 44(03): 164-166.

[16] Valarmathi B, Gupta N S, Prakash G, et al. Hybrid deep learning algorithms for dog breed identification–a comparative analysis[J]. IEEE Access, 2023.

[17] Huang Xuele, Zhai Xinyu. Identification of Illegal Pet Walking Based on YOLO Feature Detection Model [J]. Software Engineering and Applications, 2023(1): 113-123.

[18] Shi Xinyu, Fang Hongsu. Fine-Grained Cat and Dog Recognition Method Based on Deep Learning [J]. Shandong Industrial Technology, 2024(06): 49-56.

[19] Hu J, Wang Z, Chang M, et al. Psg-yolov5: A paradigm for traffic sign detection and recognition algorithm based on deep learning[J]. Symmetry, 2022, 14(11): 2262.

International Journal of Science and Engineering Applications
Volume 14-Issue 03, 09 – 14, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1403.1004

[20] Chen G, Hou Y, Cui T, et al. YOLOv8-CML: A lightweight target detection method for Color-changing melon ripening in intelligent agriculture[J]. Scientific Reports, 2024, 14(1): 14400.

[21] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS -improving object detection with one line of code[C]//Proceedings of the IEEE international conference on computer vision. 2017: 5561-5569.

[22] LIN T Y, DOLLáR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[23] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[24] Li H, Li J, Wei H, et al. Slim-neck by GSConv: a lightweight-design for real-time detector architectures[J]. Journal of Real-Time Image Processing, 2024, 21(3): 62.

[25] Lin T. Focal Loss for Dense Object Detection[J]. arXiv preprint arXiv:1708.02002, 2017.

[26] Zhang Y F, Ren W, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.