International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

YOLOV8: An Enhanced Object Detection Model for Distance Estimation

Urvashi Verma, Anshul Kalia, Sumesh Sood

Submitted: 05/02/2024 **Revised**: 13/03/2024 **Accepted**: 19/03/2024

Abstract: The rapid evolution of deep learning has transformed computer vision, yet research on leveraging this technology for distance estimation remains limited. Such investigations could greatly benefit various applications, notably anomaly detection. This study introduces an enhanced detection model, YOLOV8-CAW, which Integrates Coordinate Attention and Wise-loU Into the YOLOV8 framework to improve detection accuracy. Incorporating a distance estimation algorithm yields comprehensive outputs, combining detection results with accurate distance calculations. Experimental results demonstrate significant performance enhancements, with improvements in recall (0.4%), precision (2.2%), and Mean Average Precision (mAP) (1.5%) within the 0.5 to 0.95 threshold range while maintaining inference speeds comparable to the baseline model on the PASCAL VOC dataset. Additionally, distance estimation achieves an approximate average accuracy of 90%, indicating promising outcomes. The effective combination of computer vision and separate estimation presents unused roads for viable applications, highlighting the potential of this approach in real-world scenarios.

Index terms: *YOLOv8-CAW, Coordinate Attention (CA) module, Wise-Intersection over Union (WIoU) loss function, Object detection, Distance estimation.*

I. Introduction

Computer vision has experienced a ground breaking evolution, largely due to the rapid progress in deep learning techniques. These headways have engaged computer vision frameworks to attain unparalleled exactness and productivity in different errands, outperforming human capabilities in a few Occasions [1].

Various applications have risen, extending from calculating question measurements to timberland fire location and viciousness discovery in recordings. However, these applications merely scratch the surface of computer vision's potential.

Recognising the Immense power of computer vision techniques, this study proposes a distance estimation approach using YOLOv8 as its foundation. Estimating distances accurately holds significant value across industries and daily life scenarios. For instance, in pipeline cleaning, pinpointing the exact location of dirt can expedite the cleaning process, maximizing productivity [2].

Similarly, detecting distances between vehicles and providing early warnings to drivers can mitigate accident risks. These practical examples motivated our exploration into enhancing distance estimation through computer vision principles.

The paper introduces YOLOV8-CAW, an enhanced architecture for object detection, building upon the original YOLOV8 model. YOLOV8-CAW Incorporates the Coordinate Attention (CA) module and replaces the Complete Intersection over Union (C-IoU) loss function with the Wise-loU (WIoU) loss function [3].

These modifications aim to improve model accuracy and training convergence without significantly increasing model size. For separate estimation, a proportion calculation strategy is utilized to appraise separations between objects and the camera based on their sizes within the genuine world versus their sizes captured in computer vision [4].

The consequent areas of the paper are organized as takes after: Segment II surveys the most recent inquire about in question discovery and remove estimation, Segment III gives nitty gritty bits of knowledge into the proposed strategy, Area IV traces the explore setup, Segment V talks about the test comes about, and at last, Segment VI offers a brief rundown of the discoveries.

II. Literature Review

The ponder found relevant to this aspect of inquire about is portrayed in this segment, it is bifurcated for way better investigation.

A. Object Detection

Protest location strategies, especially those centred around YOLO models, have been broadly investigated and upgraded by analysts pointing to boost execution and address particular challenges. Different methodologies

Department Of Computer science, Himachal Pradesh University, Shimla
Vashuverma4005@gmail.com ,ansh.kamal007@gmail.com ,sumesh64@gmail.com

have been utilized, counting relevant relationship understanding between shallow and profound layers, include combination, and consideration components [5].

For occasion, combining YOLOv4 with PANet and Crush and Excitation (SE) squares Progressed Cruel Normal Accuracy (mAP) by 2.86%. Other approaches incorporate versatile setting modules, adjusted expectation layers, and multi-scale location arrangements like Spatial Pyramid Pooling (SPP) and GSConv. These adjustments empower way better discovery of small-scale objects whereas keeping up computational effectiveness [6].

Later considers have too centered on diminishing show parameters and upgrading proficiency. Strategies like Joining channel consideration (CA) modules, elective spine systems, and lightweight adjustments have been proposed. These strategies point to advance show execution without essentially expanding complexity or relinquishing precision [7].

Also, headways in Neural Engineering Look (NAS) have driven to the advancement of calculations that naturally plan neural organize designs, encourage moving forward question discovery capabilities. Other Imaginative approaches incorporate gathering learning, line encoding strategies, and unsupervised learning for protest location [8].

B. Distance Estimation

Separate estimation in computer vision ranges different applications such as car and wrongdoing examination. Analysts have proposed assorted strategies, extending from Coordination remove estimation inside protest location models to leveraging stereoscopic standards and radar information. Approaches inside question discovery models include presenting devoted separate expectation vectors and misfortune capacities to empower learning and utilization of remove data amid preparing [9].

Other strategies Incorporate expanding systems with profundity estimation branches and utilizing middle-fusion strategies combining radar point clouds and RGB Pictures for more exact separate estimation, especially in independent driving scenarios.

While radar data tends to outperform image data in distance estimation, efforts are made to reduce the Information needed by leveraging 2D images. This research aims to develop a model capable of achieving better object detection performance using 2D Images, compared to existing models heavily reliant on radar data [10].

III. Methodology

A. YOLOv8 Model

The YOLOv8 model represents a significant advancement in the YOLO series, integrating improvements over previous iterations. Unlike its predecessors, YOLOv8 adopts an anchor-free detection head, directly predicting object centers for enhanced performance in detecting small or overlapping objects [11].

Additionally, it introduces the C2f module, replacing the C3 module from YOLOv5, which effectively reduces network complexity while maintaining computational speed.

YOLOV8 moreover leverages a multi-scale include combination propelled by the Way Accumulation Arrange (Container), upgrading its capacity to distinguish objects over diverse scales [12].

The C2f module improves the stream of angle data whereas maintaining a streamlined structure, as portrayed in Figure 1.

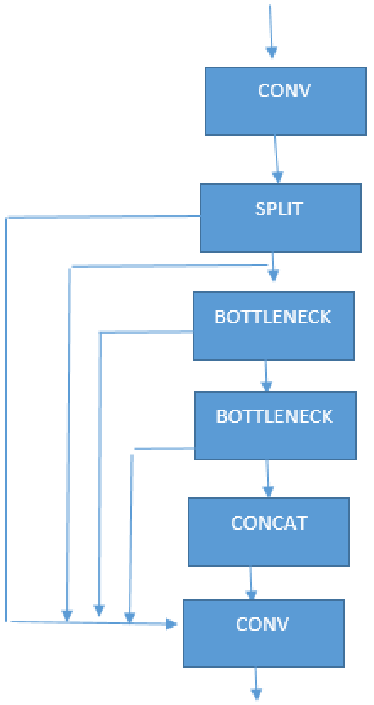


Figure – 1 C2F MECHANISM

B. Coordinated Attention Module

Chosen for efficiency and effectiveness, the Coordinated Attention (CA) module enhances the YOLOv8-CAW model by building in local information channel attention. This mechanism allows the network to focus on important areas with minimal computational cost [14]. Unlike traditional attention mechanisms, CA captures distance relationships while preserving spatial information that promotes more accurate target localization.

C. YOLOV8-CAW Model With Coordinate Attention Module

The integration of the CA module in the YOLOv8-CAW model is done after the C2f model by optimizing object detection feature maps, minimizing the computational complexity [15]. This architecture allows the model to learn and prioritize relevant features, which improves detection performance.

D. Wise Section Of The Union (Wiou) Loss Function

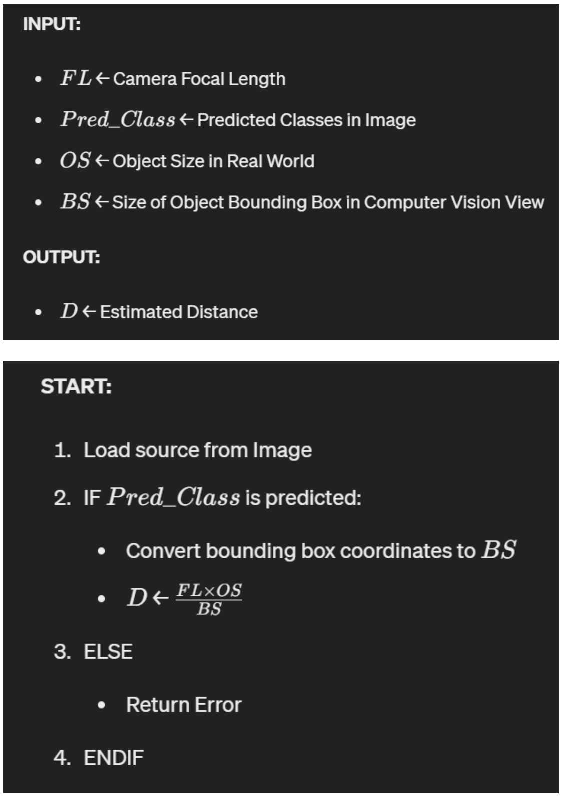
The WIoU loss function, chosen for its suitability to support range estimation tasks, improves bounding box regression for object detection. WIoU dynamically adjusts the loss coefficients for each training sample, focusing on complex samples to improve overall detection.

Wlou includes a dynamic, non-monotonic focusing mechanism and effectively balances the model's attention to different types of data, improving learning and detection accuracy.

E. Distance Estimation

Diffusion, essential for numerous applications, is accomplished through a proposed algorithm that determines the ratio of the object's actual size to the size of the bounding box Identified by the model.

This algorithm relies on the principle that the focal length yields precise distance measurements, enabling accurate distance estimation even without a predetermined focal length. Pseudocode is provided to implement the distance estimation algorithm, guaranteeing its practical application.



```
INPUT:

  • $FL$ ← Camera Focal Length
  • $Pred\_Class$ ← Predicted Classes in Image
  • $OS$ ← Object Size in Real World
  • $BS$ ← Size of Object Bounding Box in Computer Vision View

OUTPUT:

  • $D$ ← Estimated Distance

START:

  1. Load source from Image
  2. IF $Pred\_Class$ is predicted:
       • Convert bounding box coordinates to $BS$
       • $D \leftarrow \frac{FL \times OS}{BS}$
  3. ELSE
       • Return Error
  4. ENDIF
```

Figure – 3

Algorithm For Distance Estimation

This algorithm describes how to determine an object's distance from a picture. It takes into account input data such as the focal length of the camera, the anticipated classes of the picture, the actual dimensions of the object, and the size of the The method for figuring out how far one object is from another object in an image is described in this algorithm. It considers input parameters including the focal length of the camera, the image's projected classes, the object's actual dimensions, and the bounding box size from the computer vision perspective.

Then, using these parameters, the algorithm estimates the distance. It generates an error if the expected class cannot be found.

Equations:

BOUNDING BOX REGRESSION LOSS FUNCTION (IOU):

[IoU = 1 - \frac{W_i \times H_i}{S_u} \]

WISE INTERSECTION OVER UNION (WIOU) LOSS FUNCTION:

\[LWIoUv3 = r \times LWIoUv1 \]

\[r = \frac{\beta}{{\delta \alpha \beta - \delta}} \]

DISTANCE ESTIMATION:

\[Distance = \frac{{ObjectSize \times FocalLength}}{{Bounding BoxSize}} \]

\[FocalLength = \frac{{Distance \times Bounding BoxSize}}{{ObjectSize}} \]

This section highlights key advances and methods used in research and highlights the integration of innovative techniques to enhance target detection and range estimation.

Iv. Experiment Utilization

The setups and steps for training the YOLOVB-CAW model and carrying out interval estimation tests are described in the test execution section.

The PASCAL VOC 2007 and PASCAL VOC 2012 datasets, which comprise a variety of object classes such people, birds, automobiles, and chairs, are used to train the YOLOVB-CAW model. These lessons are illustrated with different examples in the training graphics.

Model training is performed on a Windows 11 operating system equipped with an Intel® Core™ i9-12900H processor and an NVIDIA GeForce RTX 3080 Ti laptop GPU with 16 GB of memory.

The programming language is Python and the CUDA® parallel computing platform is used to accelerate the GPU to accelerate the training.

Initial model training parameters, including input size, set, times, and learning rates, are set to ensure efficient convergence and optimization. Additionally, an IoU threshold is defined to determine the degree of overlap between the ground truth and predicted bounding boxes.

The YOLOv8-CAW model uses an adaptive optimization strategy that uses the AdamW optimizer initially for fast convergence and switches to stochastic gradient descent (SGD) after a certain number of iterations to ensure stable training. Comparison tests are performed to validate the distance estimation algorithm.

Three classes representing small, medium and large objects are selected from the PASCAL VOC dataset. Test samples are collected when photographing these objects in a laboratory environment from 1, 2 and 3 meters away, which ensures a variation of angles and objects.

Examples of these tests are chairs, people and cars. In general, the implementation of the experiment involves selecting a data set, configuring model training, and collecting test samples for distance estimation, which provides a thorough evaluation of the performance of the YOLOv8-CAW model.

IV. Experiment Outcome

In the Experimental Results section, a comprehensive evaluation of the YOLOv8-CAW model is performed, focusing on various metrics such as recall, precision, average precision (mAP), and reasoning time. These metrics are necessary to evaluate the model identification performance and computational efficiency.

First, the experiment outlines the calculation formulas for recall, precision, and mAP, which are key measures for object recognition tasks.

Recall measures the percentage of correctly predicted positive samples, while precision measures the percentage of correctly predicted positive samples out of all predicted positive samples.

Mean Average Accuracy (mAP) aggregates AP values from different classes and provides an overall assessment of model performance. The YOLOv8-CAW model, which integrates the coordinate attention (CA) module and the Wise-IoU (WIoU) loss function, is compared with several different configurations, including the basic YOLOv8 model with and without WIoU and the inclusion.

WIoU with and without CA module. Figure 4 summarizes the performance comparison, highlighting the model's mapping, recall, precision, and execution time in different settings.

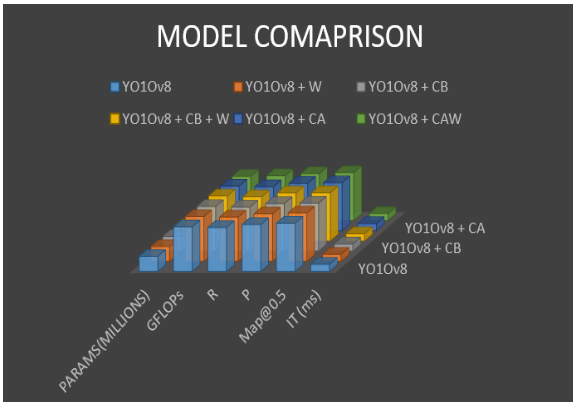


Figure – 4 MODEL COMPARISON

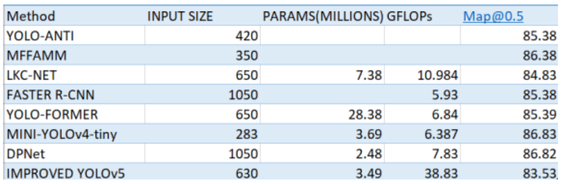
The proposed YOLOv8-CAW model consistently shows superior map-wise performance compared to other configurations.

Although the recovery performance of the model is slightly lower compared to some configurations (such as the combination of CBAM and WIoU), it maintains a higher accuracy and achieves a higher mAP value.

In addition, the proposed model maintains a low hell time, which ensures efficient computer processing. In addition, heatmap analysis provides an overview of the focus of attention of different models using Eigen-Cam imaging. The YOLOv8 model integrated with the CA module was found to have better feature map attention compared to other configurations.

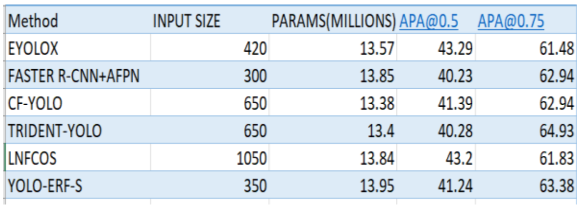
To verify the significance of the proposed model, a comparison with existing studies was performed using the same parameters for the PASCAL VOC and MS-COCO datasets.

Tables 1, 2, 3 and 4 show performance comparisons, showing maps achieved with different models. The proposed YOLOV8-CAW model outperforms other models on both datasets, demonstrating its effectiveness in detection tasks.



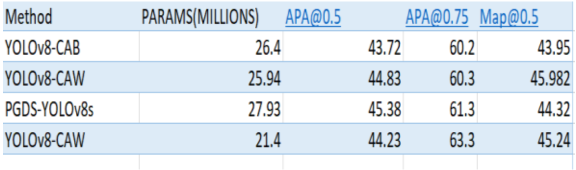
Method	INPUT SIZE	PARAMS(MILLIONS)	GFLOPs	Map@0.5
YOLO-ANTI	420			85.38
MFFAMM	350			86.38
LKC-NET	650	7.38	10.984	84.83
FASTER R-CNN	1050		5.93	85.38
YOLO-FORMER	650	28.38	6.84	85.39
MINI-YOLOv4-tiny	283	3.69	6.387	86.83
DPNet	1050	2.48	7.83	86.82
IMPROVED YOLOv5	630	3.49	38.83	83.53

Table 1 COMPARISON OF MODELS ON THE PASCAL VOC TEST 2007 DATASET



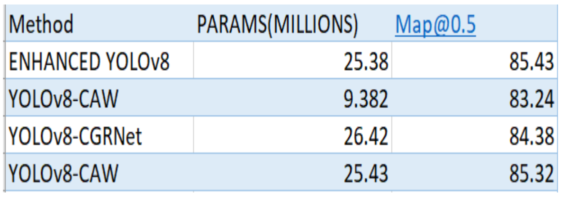
Method	INPUT SIZE	PARAMS(MILLIONS)	APA@0.5	APA@0.75
EYOLOX	420	13.57	43.29	61.48
FASTER R-CNN+AFPN	300	13.85	40.23	62.94
CF-YOLO	650	13.38	41.39	62.94
TRIDENT-YOLO	650	13.4	40.28	64.93
LNFCOS	1050	13.84	43.2	61.83
YOLO-ERF-S	350	13.95	41.24	63.38

Table 2 COMPARISON OF MODELS ON THE MS-COCO val2007 DATASET



Method	PARAMS(MILLIONS)	APA@0.5	APA@0.75	Map@0.5
YOLOv8-CAB	26.4	43.72	60.2	43.95
YOLOv8-CAW	25.94	44.83	60.3	45.982
PGDS-YOLOv8s	27.93	45.38	61.3	44.32
YOLOv8-CAW	21.4	44.23	63.3	45.24

Table 3 COMPARISON OF MODELS WITH IDENTICAL HYPERPARAMETERS ON THE MS-COCO DATASET



Method	PARAMS(MILLIONS)	Map@0.5
ENHANCED YOLOv8	25.38	85.43
YOLOv8-CAW	9.382	83.24
YOLOv8-CGRNet	26.42	84.38
YOLOv8-CAW	25.43	85.32

Table 4 COMPARISON OF MODELS WITH IDENTICAL HYPERPARAMETERS ON THE PASCAL VOC DATASET.

Confusion matrices and precision recall (PR) curves are also included to gim a thorough understanding of the model's performance on the two datasets.

The confusion matrix displays the model's accuracy for various classes, and the PR plot displays the ratio of precision to recall.

Finally, by classifying objects according to distance and class, the test validates the distance estimation technique.

The sample images and observed distances for 1, 2, and 3 meters are shown in Figures 5, 6, and 7. This approach reduces differences between classes and distances while showcasing the estimates' precision.

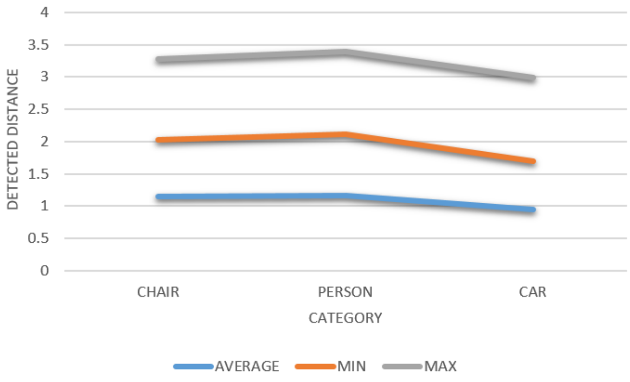


Figure – 5 DISTANCE IDENTIFIED IN 1 METER

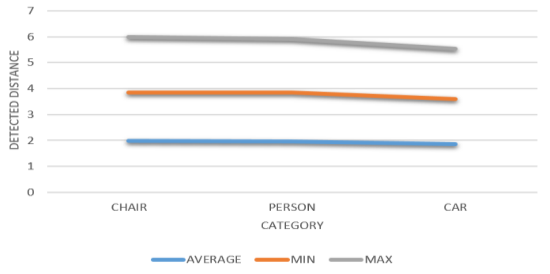


Figure – 6 DISTANCE IDENTIFIED IN 2 METER

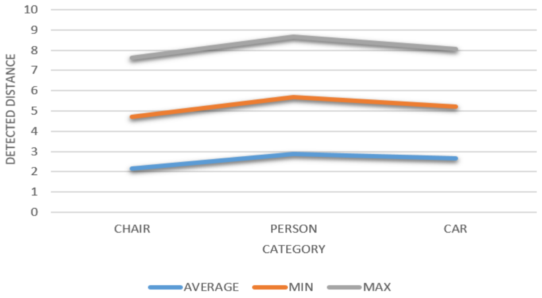


Figure – 7 DISTANCE IDENTIFIED IN 3 METER

In conclusion, the experimental results demonstrate the YOLOVB-CAW model's higher performance and computational efficiency when compared to other configurations and models, confirming the model's effectiveness in identifying targets and estimating distances.

V. Conclusion

This paper presents the YOLOVB-CAW demonstrate, an progressed protest acknowledgment framework. It includes a Facilitate Mindfulness (CA) module and substitutes Astute Crossing point over Union (WloU) for the conventional union-based approach (C-loU). The objective of bringing down computer complexity is to make strides discovery precision and preparing proficiency, especially when deciding the separations between cameras and objects. The YOLOVB engineering permits the CA component to coordinated area data into the channel, permitting the arrange to prioritize imperative areas with diminished computational overhead. Besides, the Wiou misfortune work alters the misfortune coefficient for each preparing test in arrange to improve the relapse box and, as a result, the location execution. The proposed remove estimation approach utilizes the central length concept to calculate the proportion between the object's measure and the bounding box measure that the demonstrate recognizes. The test comes about appear an impressive enhancement in review, exactness, and normal accuracy (mAP) within the limit extend of 0.5-0.95, whereas the choice speed is comparable to the first show. With an normal separate estimation exactness of almost 90%, the proposed strategy appears promising comes about in real-world applications. This inquiry about creates computer vision innovation and helps to handle challenges in a few areas and standard of living by giving precise separate estimations and effective identifying capabilities.

Conflict of Interest

No conflict of interest is associated.

References

[1] J. Ai, Z. Qu, Z. Zhao, Y. Zhang, J. Shi and H. Yan, "An SAR Target Classification Algorithm Based on the Central Coordinate Attention Module," in *IEEE*

Sensors Journal, vol. 24, no. 2, pp. 1941-1952, 15 Jan.15, 2024, doi: 10.1109/JSEN.2023.3338218.

[2] M. Zhao, G. Zuo and G. Huang, "Collaborative Learning of Deep Reinforcement Pushing and Grasping based on Coordinate Attention in Clutter," *2022 International Conference on Virtual Reality, Human-Computer Interaction and Artificial Intelligence (VRHCIAI)*, Changsha, China, 2022, pp. 156-161, doi: 10.1109/VRHCIAI57205.2022.00034.

[3] Y. Ren, X. Jiang, T. Qi, J. Li, M. Yan and X. Feng, "Low-Illumination Image Enhancement Based on End-to-End Network Using Attention Module," *2023 2nd International Conference on Image Processing and Media Computing (ICIPMC)*, Xi'an, China, 2023, pp. 9-14, doi: 10.1109/ICIPMC58929.2023.00009.

[4] K. -C. Wang *et al.*, "CA-Wav2Lip: Coordinate Attention-based Speech To Lip Synthesis In The Wild," *2023 IEEE International Conference on Smart Computing (SMARTCOMP)*, Nashville, TN, USA, 2023, pp. 1-8, doi: 10.1109/SMARTCOMP58114.2023.00018.

[5] H. Liu, N. Zhang, T. Tian and J. Tian, "Mafe-Net:Multi-Scale Adaptive Feature Enhancement Network for Infrared Weak Vehicle Targets Detection," *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, Pasadena, CA, USA, 2023, pp. 6604-6607, doi: 10.1109/IGARSS52108.2023.10282461.

[6] Y. Wu, J. Li and J. Yang, "Using Improved DeepLabV3+ for Complex Scene Segmentation," *2023 IEEE 6th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, Shenyang, China, 2023, pp. 855-860, doi: 10.1109/AUTEEE60196.2023.10408693.

[7] Y. Wang, C. Cao, Y. Li, Q. Dong, H. Li and J. Sun, "Radiofrequency Fingerprint Feature Extraction and Recognition Using a Coordinate Attention-Guided Deep Residual Shrinkage Network," *2023 International Conference on Networking and Network Applications (NaNA)*, Qingdao, China, 2023, pp. 551-557, doi: 10.1109/NaNA60121.2023.00097.

[8] W. Sheng, S. Liu and P. Liu, "Speech noise reduction algorithm based on CA-DCDCCRN," *2023 2nd International Joint Conference on Information and Communication Engineering (JCICE)*, Chengdu, China, 2023, pp. 151-156, doi: 10.1109/JCICE59059.2023.00039.

[9] X. Xiang, D. Tian, N. Lv and Q. Yan, "FCDNet: A Change Detection Network Based on Full-Scale Skip Connections and Coordinate Attention," in *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022, Art no. 6511605, doi: 10.1109/LGRS.2022.3184179.

[10] H. Zhang, A. Xiong, L. Lai, C. Chen and J. Liang, "AMME-YOLOv7: Improved YOLOv7 Based on Attention Mechanism and Multiscale Expansion for Electric Vehicle Driver and Passenger Helmet Wearing Detection," *2023 IEEE International Conference on Smart Internet of Things (SmartIoT)*, Xining, China, 2023, pp. 223-227, doi: 10.1109/SmartIoT58732.2023.00039.

[11] S. Jia, X. Zhang and W. Han, "Audio-Visual Speech Enhancement Based on Multiscale Features and Parallel Attention," *2024 23rd International Symposium INFOTEH-JAHORINA (INFOTEH)*, East Sarajevo, Bosnia and Herzegovina, 2024, pp. 1-6, doi: 10.1109/INFOTEH60418.2024.10495981.

[12] J. Liao, J. Wu, L. Zhu and H. Kang, "A Pavement Cracks detection algorithm based on CCA-YOLOv5s," *2023 35th Chinese Control and Decision Conference (CCDC)*, Yichang, China, 2023, pp. 471-476, doi: 10.1109/CCDC58219.2023.10327095.

[13] Z. Deng, Y. Li, S. He, Y. Wang and X. Wang, "A High-Resolution Human Pose Estimation Method with Coordinate Attention," *2022 9th International Conference on Digital Home (ICDH)*, Guangzhou, China, 2022, pp. 299-306, doi: 10.1109/ICDH57206.2022.00053.

[14] Y. Ren, X. Jiang, T. Qi, J. Li, M. Yan and X. Feng, "Low-Illumination Image Enhancement Based on End-to-End Network Using Attention Module," *2023 2nd International Conference on Image Processing and Media Computing (ICIPMC)*, Xi'an, China, 2023, pp. 9-14, doi: 10.1109/ICIPMC58929.2023.00009.

[15] J. Ai, Z. Qu, Z. Zhao, Y. Zhang, J. Shi and H. Yan, "An SAR Target Classification Algorithm Based on the Central Coordinate Attention Module," in *IEEE Sensors Journal*, vol. 24, no. 2, pp. 1941-1952, 15 Jan.15, 2024, doi: 10.1109/JSEN.2023.3338218.