

YOLO-P1P2-CBAM: IMPROVING YOLO11 SOCCER BALL DETECTION USING EXTRA HIGH-RESOLUTION HEADS AND CBAM

Shohruh Begmatov

Tashkent University of Information Technologies named
after Muhammad al-Khwarizmi

Doctor of Philosophy (PhD) in Technical Sciences, Doctoral (DSc) student

Email: sh.begmatov@tuit.uz

ORCID: 0000-0002-2441-916X

Mukhriddin Arabboev

Tashkent University of Information Technologies named
after Muhammad al-Khwarizmi

Doctor of Philosophy (PhD) in Technical Sciences, Doctoral (DSc) student

Email: mukhriddin.arabboev@tuit.uz

ORCID: 0000-0001-5733-5889

Akhram Nishanov

Tashkent University of Information Technologies named
after Muhammad al-Khwarizmi

Doctor of Science in Technical Sciences, Professor

Email: a.nishanov@tuit.uz

ORCID: 0000-0002-5652-8977

<https://doi.org/10.5281/zenodo.20093430>

Abstract

Reliable soccer-ball detection is a key building block for sports analytics, broadcast enhancement, and robotic training systems, yet it remains challenging because the ball often occupies a very small region and appears under motion blur, occlusion, and complex backgrounds. This paper presents YOLO-P1P2-CBAM, a lightweight modification of an Ultralytics YOLO11 detector that (i) adds extra high-resolution detection heads (P1 and P2) to better preserve fine spatial details for tiny targets and (ii) injects Convolutional Block Attention Modules (CBAM) to refine multi-scale features before prediction. Experiments on a single-class soccer-ball dataset (2,474 images; 1,978/246/250 train/val/test) show that YOLO-P1P2-CBAM achieves 0.9398 mAP@0.5 and 0.6025 mAP@0.5:0.95, outperforming YOLO11n/s/m/l/x and two additional baselines trained under identical settings. Qualitative results indicate improved localization stability in cluttered scenes and at long ranges.

Keywords: soccer ball detection; small object detection; YOLO11; attention mechanisms; CBAM; feature pyramid; real-time vision

1. Introduction

Accurate ball localization enables downstream tasks such as player-ball interaction analysis, event detection (passes, shots), camera auto-framing, and autonomous robot perception [4], [5]. Unlike larger objects, a soccer ball can occupy only a few pixels in high-definition frames, causing frequent missed detections when standard detectors rely mainly on deep, low-resolution features [4], [5]. The problem is aggravated by rapid motion, partial occlusion, specular highlights, and background clutter (crowds, advertising boards, field lines) [4], [5].

Recent single-stage detectors from the YOLO family remain attractive due to their speed and deployment simplicity; however, their performance may degrade for tiny targets unless the feature pyramid and receptive-field allocation are tailored for small-object regimes [2], [3].

This work proposes YOLO-P1P2-CBAM, a practical modification of an Ultralytics YOLO11 detector designed for soccer-ball detection. The main contributions are as follows. First, an enhanced detection head introduces additional high-resolution pyramid levels (P1 and P2) to increase sensitivity to tiny objects. Second, CBAM attention blocks are integrated at multiple scales to strengthen discriminative ball features while suppressing background noise. Third, a controlled experimental comparison against YOLO11n/s/m/l/x and two additional baselines is conducted under identical training settings.

2. Related Work

Ball detection has been studied in robotics and sports analytics using both classical pipelines and deep detectors [4], [5]. Modern approaches typically rely on one-stage detectors for real-time constraints. Attention mechanisms, including channel and spatial attention, have been widely adopted to improve representation quality; CBAM is a lightweight module that sequentially applies channel and spatial attention and has demonstrated gains across vision tasks [1].

For small-object detection, several studies report that adding higher-resolution pyramid features (e.g., P2 and, in some designs, P1 heads) can increase recall for tiny targets, particularly in aerial, surveillance, and long-range imagery. Motivated by these findings, the proposed method augments YOLO11 with P1/P2 detection heads and attention-based refinement.

3. Materials and Methods

3.1 Dataset and Experimental Setup

A single-class soccer-ball detection dataset exported in YOLO format is used. The dataset contains 2,474 images with a native resolution of 1,920×1,080 pixels and a single label class (soccer ball). The data are split into 1,978 training images, 246 validation images, and 250 test images.

Figure 1 summarizes the dataset label statistics and spatial distribution.

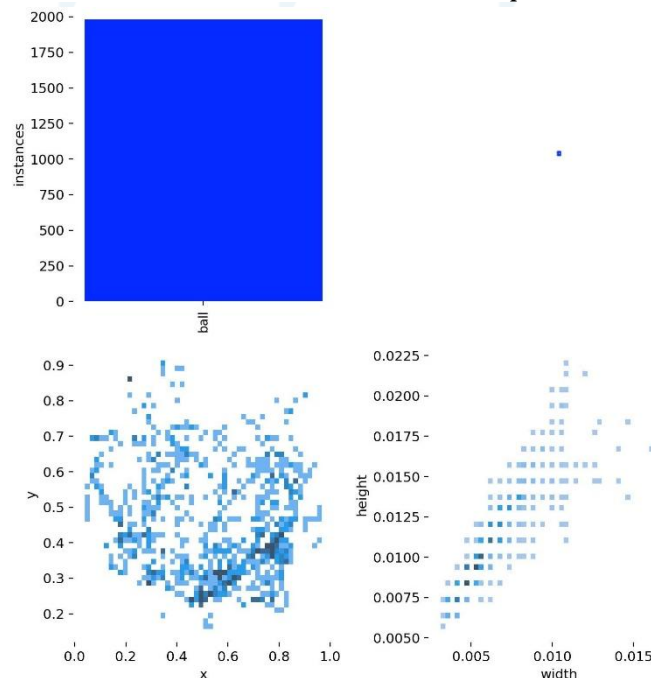


Fig. 1. Dataset label statistics and spatial distribution.

All models are trained under identical settings to enable a fair comparison. Training is performed for 1,200 epochs with image size 1,080 and batch size 8 using the AdamW optimizer

(initial learning rate 0.002, cosine-style decay factor 0.2, weight decay 0.05). Standard data augmentation is enabled (HSV jitter, random scaling/translation, mosaic, and horizontal flip). Evaluation is performed using precision, recall, mAP@0.5, and mAP@0.5:0.95 (COCO-style).

3.2 Proposed Method: YOLO-P1P2-CBAM

The proposed YOLO-P1P2-CBAM starts from a YOLO11-style backbone and head implemented in the Ultralytics framework [2], [3]. Two modifications are applied to better address the tiny-object regime:

Extra high-resolution detection heads (P1, P2): Standard YOLO heads typically predict from feature maps with progressively lower resolution (e.g., P3–P5). For small objects like a soccer ball, higher-resolution features are crucial. The head is extended by adding additional upsampling and fusion stages to create P1 and P2 feature maps, enabling predictions from five scales in total.

Multi-scale attention with CBAM: To improve robustness under clutter and motion blur, CBAM blocks are inserted before detection on multiple scales. CBAM refines features using sequential channel and spatial attention, improving the signal-to-noise ratio of ball-like patterns and suppressing irrelevant background activations [1].

4. Results

4.1 Quantitative Performance Comparison

Table 1 summarizes the best validation metrics achieved by each model during training (best epoch selected by maximum mAP@0.5:0.95). YOLO-P1P2-CBAM yields the highest mAP@0.5:0.95 (0.6025), indicating improved localization quality across IoU thresholds.

Table 1. Best validation metrics for each model (best epoch selected by maximum mAP@0.5:0.95).

Model	Best epoch	Precision	Recall	mAP@0.5	mAP@0.5:0.95
YOLO11n	707	0.9334	0.7155	0.8836	0.5796
YOLO11s	494	0.9026	0.7561	0.8838	0.5662
YOLO11m	346	0.9502	0.7751	0.9313	0.5909
YOLO11l	331	0.8985	0.7917	0.8884	0.5668
YOLO11x	482	0.9581	0.7432	0.8995	0.5734
YOLO11x-SOS	290	0.9437	0.8618	0.9412	0.5866
YOLO11x-Updated	291	0.9631	0.8491	0.9287	0.5661
YOLO-P1P2-CBAM (Proposed)	238	0.9714	0.8537	0.9398	0.6025

Figure 2 compares mAP@0.5:0.95 across all models.

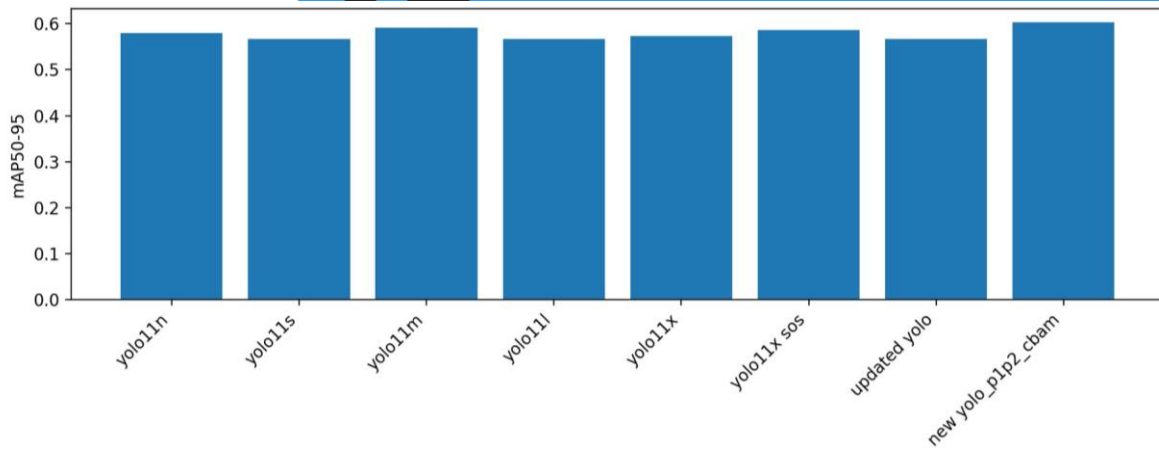


Fig. 2. mAP@0.5:0.95 comparison across models.

Compared with YOLO11m (0.5909 mAP@0.5:0.95), YOLO-P1P2-CBAM improves mAP@0.5:0.95 by 0.0116. Relative to YOLO11x, the proposed method improves mAP@0.5 by 0.0403 and mAP@0.5:0.95 by 0.0291, while also improving recall. These gains are consistent with the expectation that extra high-resolution prediction heads and attention-based feature refinement benefit tiny-object detection.

Figure 3 shows the training and validation curves for YOLO-P1P2-CBAM.

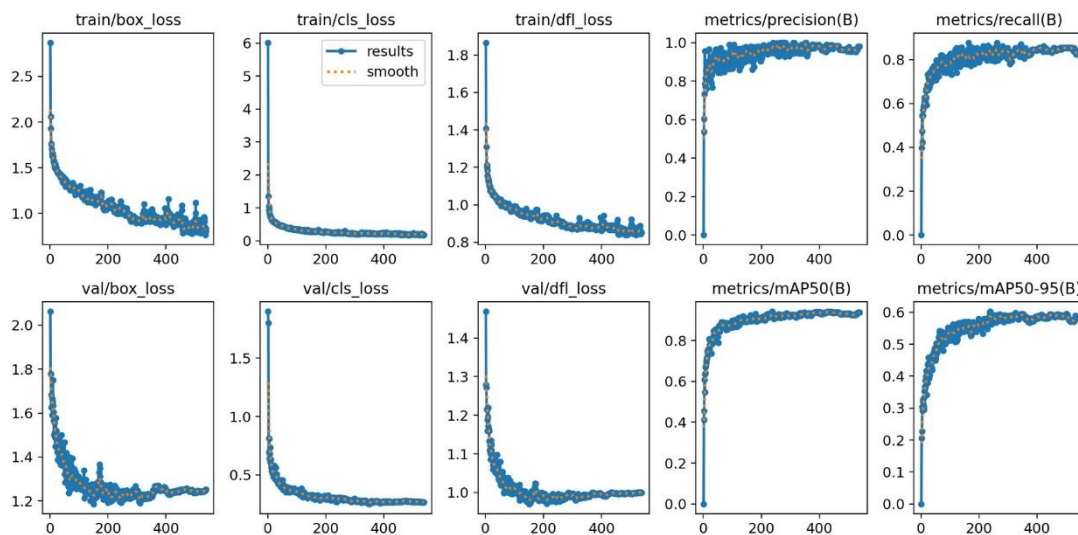


Fig. 3. Training and validation curves for YOLO-P1P2-CBAM (losses and detection metrics).

Figure 4 presents the precision–recall curve of YOLO-P1P2-CBAM.

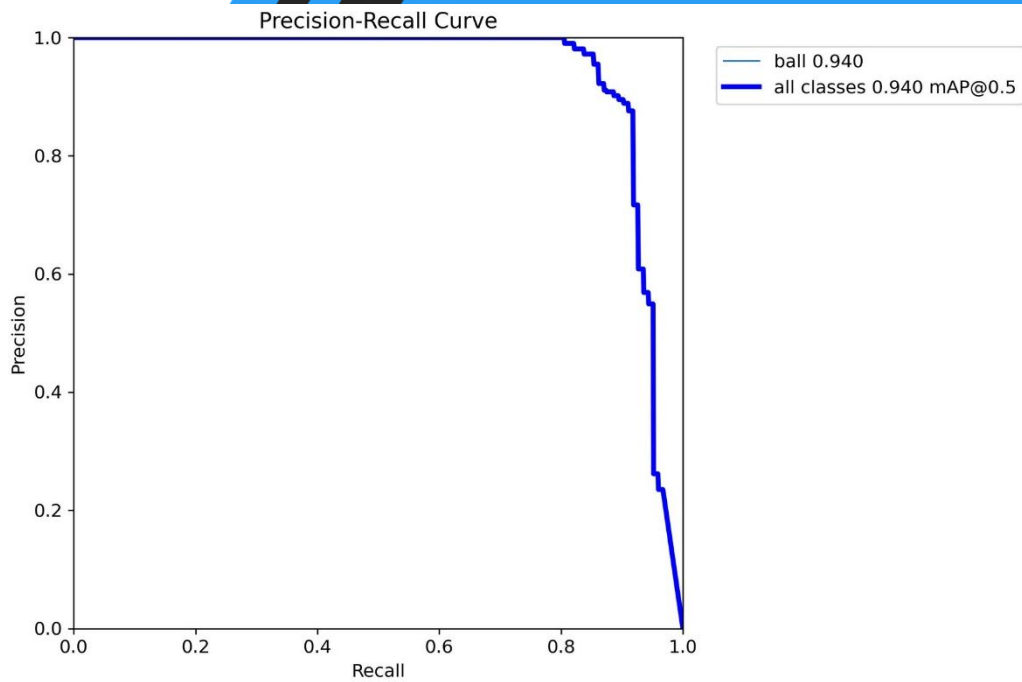


Fig. 4. Precision–Recall curve for YOLO-P1P2-CBAM. Figure 5 provides a normalized confusion matrix visualization.

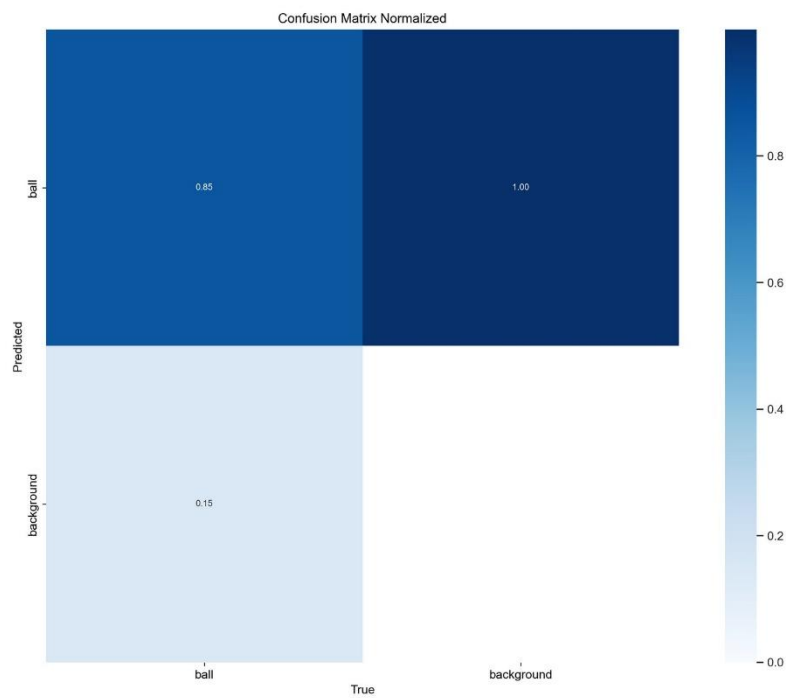


Fig. 5. Normalized confusion matrix visualization for YOLO-P1P2-CBAM.

Figure 6 illustrates representative validation examples with ground-truth (left) and predictions (right).



Fig. 6. Example validation image with ground-truth labels (left) and model predictions (right).

5. Discussion

The results indicate that adding extra high-resolution prediction heads and attention-based feature refinement improves tiny-object detection performance on the soccer-ball dataset. The largest qualitative gains are observed in cluttered long-range scenes, where the ball occupies a small region and background textures can cause false positives.

6. Conclusion

This paper introduced YOLO-P1P2-CBAM, a practical enhancement of an Ultralytics YOLO11 detector for soccer-ball detection. By adding extra high-resolution detection heads and inserting CBAM attention at multiple scales, the proposed method improves mAP@0.5 and mAP@0.5:0.95 over YOLO11n/s/m/l/x baselines trained under the same conditions. Future work will focus on reducing computational cost (e.g., pruning or distillation) while preserving small-object gains, and on validating generalization across additional match conditions and camera viewpoints.

Adabiyotlar, References, Литературы:

1. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in Proc. ECCV, 2018.
2. Ultralytics, "YOLO11 Documentation," 2024. [Online]. Available: <https://docs.ultralytics.com/models/yolo11/>
3. Ultralytics, "ultralytics/ultralytics (GitHub repository)," 2024. [Online]. Available: <https://github.com/ultralytics/ultralytics>
4. J. Komorowski, G. Kurzejamski, and G. Sarwas, "DeepBall: Deep Neural-Network Ball Detector," arXiv:1902.07304, 2019.
5. D. Barry et al., "xYOLO: A Model for Real-Time Object Detection in Humanoid Soccer on Low-End Hardware," arXiv:1910.03159, 2019.