УДК 004.932.2:796.332 https://doi.org/10.15407/pp2025.02.054

І.Б. Івасенко, С.С. Бішир

ПІДВИЩЕННЯ ТОЧНОСТІ ВИЯВЛЕННЯ М'ЯЧА У ВІДЕО ФУТБОЛЬНИХ МАТЧІВ ЗА ДОПОМОГОЮ МЕХАНІЗМІВ УВАГИ В CNN-МОДЕЛЯХ НА ОСНОВІ FPN

Попри значний прогрес у виявленні гравців завдяки моделям глибокого навчання в спортивній аналітиці, точне розпізнавання футбольного м’яча залишається складною задачею через його малий розмір, швидкий рух, часті оклюзії та візуальну подібність до інших елементів, таких як гетри гравців, логотипи та розмітка поля. Ці обмеження значно знижують ефективність автоматизованих систем для комплексного аналізу футбольних матчів, особливо в таких задачах, як розпізнавання тактичних подій, класифікація ударів і прогнозування ігрових станів. У цій роботі запропоновано метод підвищення точності виявлення м’яча у відео футбольних матчів шляхом удосконалення наявної архітектури на основі Feature Pyramid Networks (FPN). Базова модель на основі FPN, хоча й ефективна для виявлення гравців, демонструє обмежену продуктивність у розпізнаванні дрібних об’єктів, таких як м’яч. Для вирішення цієї проблеми ми інтегрували легкі механізми уваги, які дозволяють моделі краще зосереджуватись на релевантних просторових та семантичних ознаках. Зокрема, ми впроваджуємо шари Squeeze-and-Excitation (SE) у базову мережу для переналаштування ознак на рівні каналів, а також додаємо модуль CBAM (Convolutional Block Attention Module) до голови виявлення м’яча для уточнення просторової та канальної уваги. Ці модифікації покликані покращити здатність мережі відрізняти м’яч від візуально схожих об’єктів і перевантаженого фону. Наші експерименти, проведені на наборах даних ISSIA-CNR та Soccer Player Detection, демонструють, що запропонована модель з увагою досягає кращої точності класифікації м’яча порівняно з базовим підходом, без погіршення точності виявлення гравців. Отримані результати підтверджують ефективність легких механізмів уваги в задачах виявлення дрібних об’єктів та відкривають перспективи для створення більш надійних і реалістичних систем аналізу футбольних відео у реальному часі.

Ключові слова: виявлення м’яча, глибоке навчання, аналіз футбольного відео, виявлення об’єктів, механізми уваги, Feature Pyramid Network

I.B. Ivasenko, S.S. Bishyr

ENHANCING BALL DETECTION IN FOOTBALL VIDEOS USING ATTENTION MECHANISMS IN FPN-BASED CNNS

While deep learning models have significantly advanced player detection in sports analytics, accurately identifying the football remains a persistent challenge due to its small size, rapid movement, frequent occlusions, and visual similarity to other elements such as player socks, logos, and field markings. This limitation significantly reduces the effectiveness of automated systems in comprehensively analyzing football matches, particularly in applications such as tactical event recognition, shot classification, and game state prediction. In this paper, we propose a method to improve ball detection accuracy in football videos by enhancing an existing architecture based on Feature Pyramid Networks (FPN). The original FPN-based model, although efficient for detecting large-scale players, shows limited performance in detecting small objects such as the ball. To address this, we integrate lightweight attention mechanisms to help the model focus on more relevant spatial and semantic features. Specifically, we introduce Squeeze-and-Excitation (SE) layers into the backbone of the network to perform channel-wise feature recalibration and embed a Convolutional Block Attention Module (CBAM) into the ball detection head to refine both spatial and channel-level attention. These modifications are designed to enhance the network’s ability to distinguish the ball from cluttered backgrounds and visually similar objects. Our experiments, conducted on the ISSIA-CNR and Soccer Player Detection datasets, demonstrate that the proposed attention-augmented model achieves improved ball classification accuracy compared to the baseline, with no degradation in player detection performance. These results validate the utility of lightweight attention mechanisms in the context of small object detection and provide a promising direction for more robust and real-time football video analysis systems.

Keywords: Ball Detection, Deep learning, Football Video Analysis, Object Detection, Attention mechanisms, Feature Pyramid Network

Introduction

Ball detection plays a crucial part in the automated analysis of football matches, enabling advanced tasks such as event detection, match analysis, and performance assessment [1] [2]. However, accurate ball detection remains challenging because of its small size, fast movement, occlusions, and similar appearance to other elements, such as player socks, goalkeeper gloves, or field lines [3] [4]. While deep learning methods, especially convolutional neural networks (CNNs), have significantly advanced the state of object detection in sports analytics, existing approaches often struggle with reliably identifying the ball under diverse match conditions [5] [6].

Feature Pyramid Networks (FPN) are a promising approach to object detection in complex scenes [7]. They allow for effective multi-scale feature extraction by combining low and high-level features. A recent study proposed an FPN-based approach as an integrated ball and player detector in footage from football matches [3]. The approach demonstrated a strong performance in player detection. Nonetheless, the same approach showed comparatively lower accuracy in the ball detection tasks because of the small size, high speed, frequent occlusions, and visual similarity with other objects. As a result, even the state-of-the-art models for object detection, such as YOLO [8] and SSD [9], frequently missidentify or completely miss the detection of small, fast-moving targets [10] [11] [12]. This indicates a need for further refinement to increase the effectiveness of detecting small and fast-moving objects.

This paper addresses that specific limitation. We aim to enhance the ball detection performance of an existing FPN-based architecture by integrating lightweight attention mechanisms. Our approach is based on the recent success of applying an attention mechanism to improve the performance of small object detection in remote sensing [15], aerial imagery [16], and medical imaging [17]. Additionally, the idea of enhancing FPNs with attention is supported by the work Attentional Feature Pyramid Network (AFPN) proposed by Min et al. [18].

The remainder of this paper is organized in the following structure:

- Section 2 discusses related work on object detection in football and attention mechanisms.
- Section 3 provides the methodology, including the original architecture and our proposed enhancements.
- Section 4 describes the setup and the outcome of the experiments.
- Section 5 presents an analysis of the work.
- In conclusion, Section 6 summarizes the work and outlines future research directions.

Related Work

Ball Detection in Football Analytics. Object detection has become essential to football video analytics, helping recognize players, the ball, and key events such as shots [1] [2]. Traditional computer vision approaches relied on handcrafted features and motion tracking [5] but struggled in scenarios involving occlusion, fast motion, or cluttered backgrounds. With the help of deep learning, CNN-based methods have achieved better performance in sports analytics tasks.

Recent studies have employed architectures like YOLO [8] and SSD [9] for real-time player and ball detection. However, these models often struggle to detect small objects like the ball, especially in low-resolution frames or when the ball is partially occluded [5] [6]. The FPN-based base model used in this work represents an improvement by leveraging multi-scale feature maps, improving the detection of large and small objects [3] [19]. Despite this, the detection accuracy for the ball remained lower than for players, motivating further research into specialized enhancements.

Feature Pyramid Networks (FPN). The Feature Pyramid Network (FPN) [7], introduced by Lin et al., is a widely adopted architecture for multi-scale object detection. It enhances a backbone CNN (e.g., ResNet) by creating a

top-down pathway and lateral connections that fuse semantic-rich features from higher layers with detailed spatial features from earlier layers. FPN models are especially effective in object detection within the same image at different scales [10]. They are well-suited for complex scenes like football fields, where players and the ball vary in size and appearance. However, even with FPN's multi-scale approach, small objects like the ball can remain hard to detect due to weak spatial cues or low contrast. Some works, such as the Attentional Feature Pyramid Network (AFPN) [18], further enhance FPNs by introducing attention mechanisms to better focus on important features at multiple scales.

Attention Mechanisms in CNNs. Attention mechanisms are powerful tools that enhance feature representation in CNNs, emphasizing important information while suppressing irrelevant noise. Two modules used in our work are:

- Squeeze-and-Excitation (SE) blocks, proposed by Hu et al. [13], introduce channel-wise attention by modeling the interdependencies between feature channels. This allows the network to recalibrate the importance of different channels, leading to improved discriminative ability, especially in cluttered scenes.
- Convolutional Block Attention Module (CBAM), proposed by Woo et al. [14], extends this idea by incorporating channel and spatial attention. CBAM sequentially applies channel attention followed by spatial attention to refine the feature maps, making it particularly effective for tasks involving small and occluded objects.

Several studies have demonstrated that integrating SE or CBAM modules into standard CNNs improves performance across tasks such as remote sensing [15] [16], image classification, object detection [10] [11] [12], and segmentation [17]. However, their application to sports analytics, particularly for small object detection in dynamic environments, has been limited. In this paper, we explore the benefits of applying SE and CBAM to enhance the ball detection capability of an FPN-based network.

Methodology

In this section, we first describe the baseline architecture (FootAndBall) [3] that serves as the foundation for our work. Then, we present the proposed modifications that involve integrating attention mechanisms — Squeeze-and-Excitation (SE) [13] and Convolutional Block Attention Module (CBAM) [14] — to improve the detection of small, challenging objects such as a ball.

Integration of SE Block in Backbone. We add a Squeeze-and-Excitation (SE) [13] module after the first, third, and fifth convolutional blocks (Conv1, Conv3, and Conv5) in the backbone. The SE block works by performing global average pooling across each channel of the feature map, creating a channel descriptor that passes through two fully connected layers with a ReLU and sigmoid activation to learn the importance of each channel. The output is used to reweight the input feature map channels:

$$F_{scaled} = F \cdot \sigma\left(W_1 \cdot ReLU\left(W_2 \cdot GAP(F)\right)\right), \quad (1)$$

where F is the input feature map GAP is global average pooling, and W_1, W_2 are the learned weights. This allows the network to focus on informative feature channels and improve the representation of small objects [13] [20]. Fig. 1 represents a diagram of the SE block.

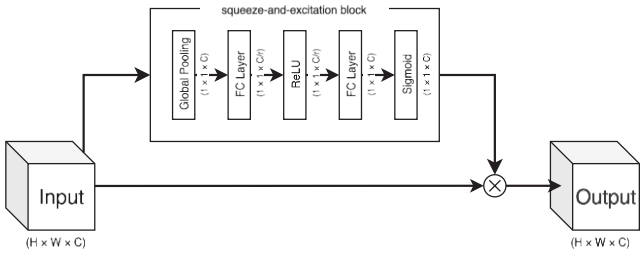

Fig. 1. Squeeze-and-excitation block

CBAM in Ball Classifier Head. The Convolutional Block Attention Module (CBAM) [14] enhances channel and spatial attention. We apply CBAM to the output feature map before the ball classification head. CBAM sequentially applies:

1. Channel attention uses average and max pooling along the spatial dimension followed by shared MLP layers.
2. Spatial attention, using a convolution over concatenated average-pooled and max-pooled feature maps across channels.

The schematic representation of the CBAM architecture is illustrated in Fig. 2.

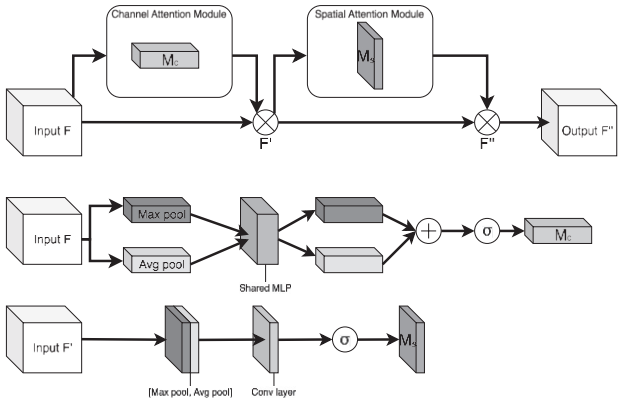

Fig. 2. Convolutional Block Attention Module (CBAM). The top diagram provides a general overview of the CBAM architecture. The middle diagram details the Channel Attention Module. The bottom diagram illustrates the Spatial Attention Module

This results in a refined feature map:

$$CBAM(F) = SA\big(CA(F)\big) \cdot F, \qquad (2)$$

$$CA(F) = \sigma\left(W_1\left(W_0\left(F^c_{avg}\right)\right) + W_1\left(W_0(F^c_{max})\right)\right), \qquad (3)$$

$$SA(F) = \sigma\left(f^{7\times7}\left(\left[F^s_{avg}, F^s_{max}\right]\right)\right), \qquad (4)$$

Modified Network Architecture Overview. The overall architecture remains fully convolutional and lightweight but with improved attention modeling. The SE-enhanced backbone generates richer feature maps, while the CBAM-augmented detection head improves ball localization precision. Fig. 3 illustrates the modified network architecture, where the SE modules are integrated into the first, third, and fifth convolutional blocks, and the CMAB module is integrated into the ball classification layer. A schematic comparison between the original and modified models is provided in Table 1.

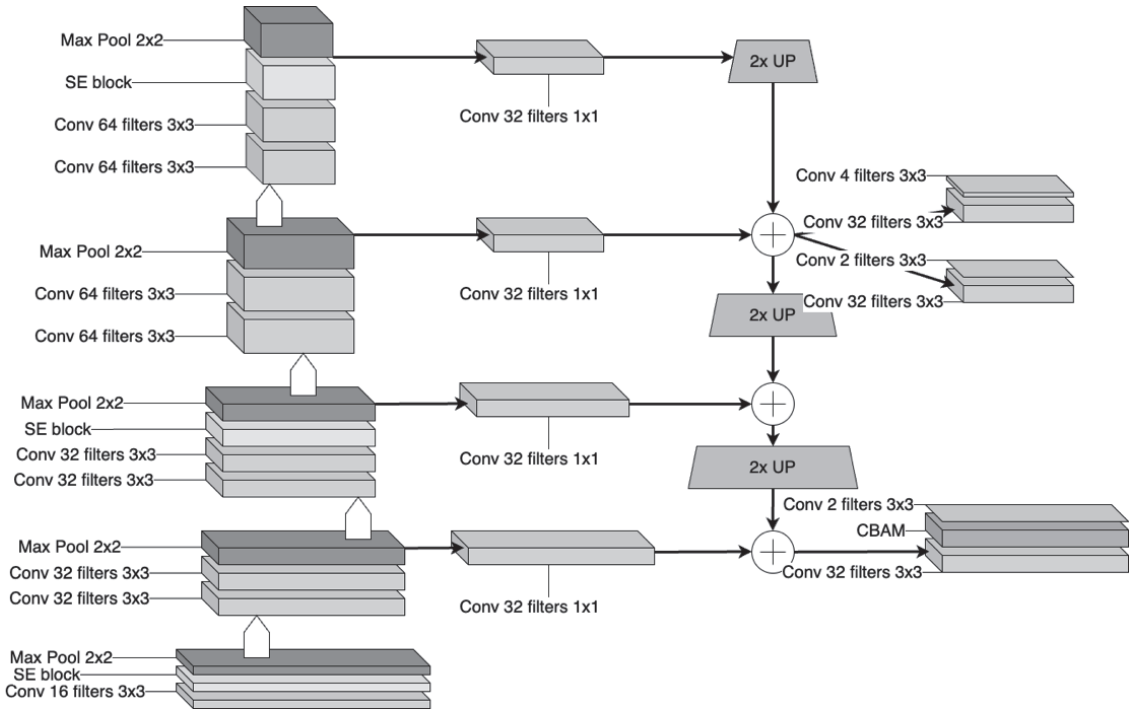

Fig. 3. The modified network architecture includes SE layers in blocks Conv1, Conv3, and Conv5, and a CBAM layer in the Ball classifier head

Table 1.

Comparison of original and modified network architectures

Block	FootAndBall layers	Modified Network Architecture layers	Output size
Conv1	16 filters 3x3 MaxPool 2x2	16 filters 3x3 SE block MaxPool 2x2	w/2, h/2, 16
Conv2	32 filters 3x3 32 filters 3x3 MaxPool 2x2	32 filters 3x3 32 filters 3x3 MaxPool 2x2	w/4, h/4, 32
Conv3	32 filters 3x3 32 filters 3x3 MaxPool 2x2	32 filters 3x3 32 filters 3x3 SE block MaxPool 2x2	w/8, h/8, 32
Conv4	64 filters 3x3 64 filters 3x3 MaxPool 2x2	64 filters 3x3 64 filters 3x3 MaxPool 2x2	w/16, h/16, 64
Conv5	64 filters 3x3 64 filters 3x3 MaxPool 2x2	64 filters 3x3 64 filters 3x3 SE block MaxPool 2x2	w/32, h/32, 64
1x1Conv1	32 filters 1x1	32 filters 1x1	w/4, h/4, 32
1x1Conv2	32 filters 1x1	32 filters 1x1	w/8, h/8, 32
1x1Conv3	32 filters 1x1	32 filters 1x1	w/16, h/16, 32
1x1Conv4	32 filters 1x1	32 filters 1x1	w/32, h/32, 32
Ball classifier	32 filters 3x3 2 filters 3x3 Sigmoid	32 filters 3x3 CBAM 2 filters 3x3 Sigmoid	w/4, h/4, 1
Player classifier	32 filters 3x3 2 filters 3x3 Sigmoid	32 filters 3x3 2 filters 3x3 Sigmoid	w/16, h/16, 1
BBox regressor	32 filters 3x3 4 filters 3x3	32 filters 3x3 4 filters 3x3	w/16, h/16, 4

Loss Function. We adopt the same loss function as in the original FootAndBall model, consisting of:

- Binary cross-entropy losses for ball and player classification.
- Smooth L1 loss for bounding box regression, as used in SSD [9] [21].

Let L_b, L_p, L_{bbox} represent the ball classification loss, player classification loss, and player bounding box loss, respectively. The total loss is computed as:

$$L = \frac{1}{N}\left(\alpha L_b + \beta L_p + L_{bbox}\right) \qquad (5)$$

where α and β are weighting coefficients, and N is the number of examples in a batch.

Experiments

In this section, we describe the experimental configuration used to evaluate the effectiveness of the proposed modifications to the FootAndBall architecture. We assess the performance of our proposed architecture, which integrates SE and CBAM modules, and compare it with the original model.

Datasets. We used the same two datasets that were used in the baseline study:

- ISSIA-CNR Soccer Dataset [5]: Contains 20,000 annotated frames from professional matches recorded using six synchronized Full HD cameras. Each frame is labeled with ball positions and player bounding boxes.
- Soccer Player Detection Dataset [22]: Composed of 2,019 images captured from two professional football matches, annotated with over 22,000 player locations. Ball positions are not annotated in this dataset.

As in the original paper, we split each dataset into 80% for training data and 20% for evaluation [3]. Both datasets contain a range of challenges like motion blur, occlusions, and background clutter.

Implementation details. We implemented the model in PyTorch and trained it using Adam optimizer [23] with a 4-step learning rate scheduler. The initial learning rate was set to 0.001 and decreased by a factor of 10 at the 10^{th}, 25^{th}, 50^{th}, and 75^{th} epochs. This gradual decay enabled the model to converge quickly in the early training phases and allowed a fine-grained adjustment in later stages. The training was performed on the NVIDIA RTX 4000 Ada Generation GPU. The hyperparameters for the training are represented in Table 2.

Table 2.

Training hyperparameters

Optimizer	Adam
Initial learning rate	0.001
Learning rate decay	x0.1 at epochs 10, 25, 50, and 75
Epochs	100
Batch size	16

To enhance the generalization, we applied data augmentation techniques, including random cropping and flipping [24].

Evaluation metrics. We use the Average Precision (AP) metric, a standard object detection metric described in the Pascal VOC challenge [25]. Ball detection AP is computed based on maximum values in the confidence map matching the ground truth position. Player detection is calculated based on predicted bounding boxes with an Intersection over Union (IoU) threshold of 0.5. We also report model size (number of trainable parameters) to evaluate efficiency.

Results. Table 3 compares the original model with the proposed enhanced version. We compare the Average Precision for Ball and Player detection on the ISSIA-CNR dataset and player detection on the Soccer Player Detection dataset.

Table 3.

Evaluation results of the original model in comparison with the enhanced model

Model	Ball AP	Player AP (ISSIA)	Mean AP	Player AP (SPD)	Params
FootAndBall	0.909	0.921	0.915	0.885	199K
SE + CBAM	0.927	0.917	0.922	0.871	200K

Our final model with SE and CBAM blocks shows the highest ball detection accuracy, outperforming the baseline by 2% AP gain. Player detection performance is also maintained at the same level. Despite the added attention layers, the model remains lightweight with a slightly increased number of parameters. Fig. 4 illustrates a comparative analysis of classification outcomes between the original and proposed models, highlighting instances where the original results are inadequate. In contrast, the proposed model successfully classifies the ball, demonstrating its enhanced efficacy.

Fig. 4. Comparison of ball classification results: the top row displays failed classifications from the original model, while the bottom row illustrates successful classifications from the proposed model

Discussion

The experimental results show that the integration of Squeeze-and-Excitation (SE) [13] and Convolutional Block Attention Module (CBAM) [14] blocks into the FootAndBall architecture improves the performance of the model for the task of ball detection. This is a significant advancement, as accurate ball detection remains one of the most complicated tasks in football video analysis owing to its small size, frequent occlusions, motion blur, and visual similarity to player gear and background elements [3] [4].

Adding SE blocks in the backbone enhances the model's ability to emphasize informative feature channels while suppressing less relevant ones. This aligns with prior findings that SE improves model sensitivity to subtle visual cues in cluttered scenes [13] [20]. In our case, the SE-enhanced backbone produces stronger features for ball detection. Similar channel-wise recalibration strategies have also proven effective in other small object detection contexts, such as traffic sign detection [11].

Including CBAM in the ball detection head applies channel and spatial attention. It allows the model to focus on small spatial re-

gions with high semantic relevance, such as regions that contain fast-moving objects. This spatial attention appears to help to distinguish false positives like white socks, pitch lines, or advertisements from the ball distractors, which is one of the frequent issues in the baseline model.

Combining SE and CBAM yields the highest accuracy, confirming their complementary nature. SE enhances global channel interactions during feature extraction, while CBAM introduces localized attention refinements before detection [14]. Similar hybrid attention strategies have succeeded in medical image analysis [17] and aerial image object detection [15] [16], where high-level semantics and spatial precision are critical.

Despite the additional attention layers, the proposed model remains comparably small and capable of real-time performance. This echoes trends in lightweight attention integration found in mobile-focused detection models like MobileNetV3 [26]. Our enhancements increased detection accuracy without a significant trade-off in model size.

However, some challenges remain. The model occasionally fails in edge cases involving heavy occlusion or extreme motion blur, conditions common in real-world sports footage. Fig. 5 illustrates challenging frames where the model either could not detect the ball or incorrectly identified it in its absence. Because our system processes frames independently, it cannot exploit temporal continuity to reinforce uncertain predictions. Techniques such as temporal feature aggregation or recurrent modules have been shown to improve consistency [27] [28] in video-based detection tasks and could be beneficial here.

(a) *(b)*

Fig. 5. Examples of model misidentifications, showing a false positive detection (a) and missed detections (b) of the ball

Overall, the results support our hypothesis that attention mechanisms considerably enhance the detection of small, context-sensitive objects in sports videos. The proposed approach balances accuracy and computational efficiency, making it suitable for real-time sports analytics systems.

Conclusion

This paper presents an enhanced deep learning architecture for joint player and ball detection in football match videos. Building on the original FootAndBall model, we introduce two attention mechanisms — Squeeze-and-Excitation (SE) [13] and Convolutional Block Attention Module (CBAM) [14] — to enhance the accuracy of ball detection, a task known to be difficult due to its small size, high motion, and frequent occlusion [4].

By integrating SE blocks into the feature extraction backbone, we enabled the network to adaptively recalibrate channel-wise feature responses adaptively, enhancing its discriminative power in complex scenes [13] [20]. Additionally, incorporating CBAM into the ball detection head improved the network's ability to focus on relevant spatial regions, significantly increasing its precision in identifying the ball amidst cluttered backgrounds. We also proposed a 4-step learning rate schedule, which helped improve training stability and convergence over time.

Our experiments on the ISSIA-CNR [5] and Soccer Player Detection [22] datasets demonstrated that the proposed attention-based enhancements lead to notable improvements in detection accuracy, particularly for the ball, increasing the accuracy by 2%, while maintaining real-time inference speed and model efficiency. These results validate the effectiveness of lightweight attention modules in sports video analysis systems.

While the proposed model achieved strong results, several opportunities for further improvement exist, such as temporal modeling. Our current approach operates on single frames, without leveraging temporal consistency. Incorporating temporal information through optical flow, frame-level feature aggregation, or recurrent networks (e.g., ConvLSTM or 3D CNNs) could enhance robust-

ness, especially in motion blur or occlusion scenarios [27] [28].

Overall, our results highlight that attention mechanisms are a promising avenue for improving small-object detection in sports analytics. The proposed system offers a solid foundation for future research and real-world applications in football match analysis by combining architectural innovation with efficiency considerations.

References

1. Bialkowski, P. Lucey, P. Carr, Y. Yue, S. Sridharan and I. Matthews, "Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data," in *2014 IEEE International Conference on Data Mining*, December 2014. doi: 10.1109/ICDM.2014.133
2. M. Manafifard, H. Ebadi and H. Moghaddam, "A Survey on Player Tracking in Soccer Videos. Computer Vision and Image Understanding," *Computer Vision and Image Understanding,* vol. 159, pp. 19-46, June, 2017. doi: 10.1016/j.cviu.2017.02.002
3. J. Komorowski, G. Kurzejamski and G. Sarwas, "FootAndBall: Integrated Player and Ball Detector," in *15th International Conference on Computer Vision Theory and Applications*, pp. 47-56, Valletta, Malta, January, 2020. doi: 10.5220/0008916000470056
4. P. Kamble, A. Keskar and K. Bhurchandi, "A deep learning ball tracking system in soccer videos," Opto-Electronics Review, vol. 27, no. 1, pp. 58-69, March, 2019. doi: 10.1016/j.opelre.2019.02.003
5. T. D'Orazio, M. Leo, N. Mosca, P. Spagnolo and P. L. Mazzeo, "A Semi-automatic System for Ground Truth Generation of Soccer Video Sequences," in Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, Genova, Italy, September, 2009. doi: 10.1109/AVSS.2009.69
6. T. Wang and T. Li, "Deep Learning-Based Football Player Detection in Videos," Computational Intelligence and Neuroscience, pp. 1-8, 2022. doi: 10.1155/2022/3540642
7. T. -Y. Lin, P. Dollár, R. Girshick, K. He, H. B. and S. Belongie, "Feature Pyramid Networks for Object Detection," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936-944, Honolulu, HI, USA, 2017. doi: 10.1109/CVPR.2017.106
8. J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv:1804.02767, 2018. doi: 10.48550/arXiv.1804.02767
9. W. Liu, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu and A. Berg, "SSD: Single Shot MultiBox Detector," in European Conference on Computer Vision, pp 21–37, 2016. doi: 10.1007/978-3-319-46448-0_2
10. Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li and S. Hu, "Traffic-Sign Detection and Classification in the Wild," in Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June, 2016. doi: 10.1109/CVPR.2016.232
11. Y. Chen, J. Wang, Z. Dong, Y. Yang, Q. Luo and M. Gao, "An Attention Based YOLOv5 Network for Small Traffic Sign Recognition," in IEEE 31st International Symposium on Industrial Electronics (ISIE), Anchorage, AK, USA, June, 2022. doi: 10.1109/ISIE51582.2022.9831717
12. S. Du, W. Pan, N. Li, S. Dai, B. Xu, H. Liu, C. Xu and X. Li, "TSD‐YOLO: Small traffic sign detection based on improved YOLO v8," ET Image Processing, vol. 18, June, 2024. doi: 10.1049/ipr2.13141
13. J. Qu, Z. Tang, L. Zhang, Y. Zhang and Z. Zhang, "Remote Sensing Small Object Detection Network Based on Attention Mechanism and Multi-Scale Feature Fusion," Remote Sensing, vol. 15, p. 2728, May, 2023. doi: 10.3390/rs15112728
14. J. Rabbi, N. Ray, M. Schubert, S. Chowdhury and D. Chao, "Small-Object Detection in Remote Sensing Images with End-to-End Edge-Enhanced GAN and Object Detector Network," Remote Sensing, vol. 12, p. 1432, April, 2020. doi: 10.3390/rs12091432
15. O. Oktay, J. Schlemper, L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Hammerla, B. Kainz, B. Glocker and D. Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas," arXiv:1804.03999, April, 2018. doi: 10.48550/arXiv.1804.03999
16. K. Min, G.-H. Lee and S.-W. Lee, "Attentional feature pyramid network for small object detection," Neural Networks, vol. 155, p. 439–450, November, 2022. doi: 10.1016/j.neunet.2022.08.029
17. V. Renò, N. Mosca, R. Marani, M. Nitti and E. Stella, "Convolutional Neural Networks Based Ball Detection in Tennis Games," in IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, June, 2018. doi: 10.1109/CVPRW.2018.00228

18. J. Hu, L. Shen and G. Sun, "Squeeze-and-Excitation Networks," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, June, 2018. doi: 10.1109/CVPR.2018.00745
19. S. Woo, J. Park, J.-Y. Lee and I. Kweon, "CBAM: Convolutional Block Attention Module," in European Conference on Computer Vision (ECCV), Munich, Germany, 2018.
20. H. Li, P. Xiong, J. An and L. Wang, "Pyramid Attention Network for Semantic Segmentation," 10.48550/arXiv.1805.10180, pp. 3-19, September, 2018. doi: 10.1007/978-3-030-01234-2_1
21. R. Girshick, "Fast R-CNN," in 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, December, 2015. doi: 10.1109/ICCV.2015.169
22. K. Lu, J. Chen, J. Little and H. He, "Light Cascaded Convolutional Neural Networks for Accurate Player Detection," 10.48550/arXiv.1709.10230, September, 2017. doi: 10.48550/arXiv.1709.10230
23. D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in International Conference on Learning Representations, December, 2014. doi: 10.48550/arXiv.1412.6980
24. C. Shorten and T. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," Journal of Big Data, vol. 6, no. 60, July, 2019. doi: 10.1186/s40537-019-0197-0
25. M. Everingham, L. Van Gool, C. Williams, J. Winn and A. Zisserman, "The Pascal Visual Object Classes (VOC) challenge," International Journal of Computer Vision, vol. 88, pp. 303-338, 2010. doi: 10.1007/s11263-009-0275-4
26. A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. Le and H. Adam, "Searching for MobileNetV3," 10.48550/arXiv.1905.02244, pp. 1314-1324, Seoul, Korea (South), 2019. doi: 10.1109/ICCV.2019.00140
27. L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang and L. Van Gool, "Temporal Segment Networks for Action Recognition in Videos," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 11, pp. 2740-2755, November, 2019. doi: 10.1109/TPAMI.2018.2868668
28. A. Kompella and R. Kulkarni, "A semi-supervised recurrent neural network for video salient object detection," Neural Computing and Applications, pp. 2065–2083, vol. 33, no. 6, March 2021. doi: 10.1007/s00521-020-05081-5

Одержано: 20.05.2025
Внутрішня рецензія отримана: 30.05.2025
Зовнішня рецензія отримана: 30.05.2025

Про авторів:

[1,2]Івасенко Ірина Богданівна,
доктор технічних наук,
старший науковий співробітник,
професор
e-mail: iryna.b.ivasenko@lpnu.ua
https://orcid.org/0000-0003-3795-9779

[2]Бішир Сергій Сергійович,
аспірант першого року навчання,
Національний університет
«Львівська політехніка»
e-mail: serhii.s.bishyr@lpnu.ua
https://orcid.org/0009-0009-1008-9292

Місце роботи авторів:
[1]Фізико-механічний інститут
Ім. Г. В. Карпенка НАН України
Тел.: +3(032) 263-30-88
79060, м. Львів, вул. Наукова 5,
e-mail: pminasu@ipm.lviv.ua

[2]Національний університет
«Львівська політехніка»
Тел.: +3(8032) 258-22-82,
79013, м. Львів, вул. Степана Бандери, 12,
e-mail: coffice@lpnu.ua, com.centre@lpnu.ua