Action Recognition of Basketball Players Based on Hybrid Attention Module and Spatial Feature Pyramid Module

Zhihua Tan[1], Sheng Gao[2], Shihai Wei[1], Jingyu Zhang[2], Min Zhu[3]*

[1] School of Physical Education, Changsha University of Science and Technology, China
[2] School of Computer and Communication Engineering, Changsha University of Science and Technology, China
[3] College of Information Science and Technology, Zhejiang Shuren University, China
luck@csust.edu.cn, 22208051657@stu.csust.edu.cn, 001960@csust.edu.cn,
zhangzhang@csust.edu.cn, zhumin@zjsru.edu.cn

Abstract

Watching basketball game videos is an important reference for coaches to analyze the team's tactics. Detecting the athletes' actions on the court in real time through an object detection algorithm can help coaches find team problems and formulate solutions. Aiming at the problems of fewer basketball players' action detection datasets and the difficulty of action detection, this paper proposes a dataset of basketball players' action detection, BPAD (basketball player action dataset), and an object detection algorithm, YOLOSS (YOLOv4 SimSE and SPPFCSPCG), in which the BPAD dataset consists of 2,151 pictures, which are obtained by extracting the videos of college basketball teams' matches and annotated with the Labelimg tool, and include three categories, namely, walk, run, and defense. The YOLOSS algorithm is based on YOLOv4, with the more efficient Ghostnet as the backbone of the model, and add a new hybrid attention module SimSE and a new spatial feature pyramid pooling module SPPFCSPCG. YOLOSS can effectively improves the detection algorithm's ability to recognize the typical basketball actions of athletes in the video stream-run, defense, and walk. The recognition ability of the YOLOSS algorithm on the BPAD dataset is as high as 82.4% mAP_{50}, which can clearly express the action of each player on the court. By comparing the results of various experiments, it is proved that YOLOSS, the object detection algorithm proposed in this paper, can effectively detect the actions of basketball players.

Keywords: Basketball player action detection, Basketball player action detection dataset, Hybrid attention, Spatial pyramid pooling, Object detection

1 Introduction

Nowadays, computer technology is becoming more and more mature. Computer technology is gradually beginning to step into production and life of human beings. Big data is a popular field in computer technology, can explore the potential value of data, helping make decisions for social development [1]. Of course, there are security risks associated with data. In order to cope with the traditional methods, which are prone to information loss and leakage, blockchain technology was born [2]. At the same time, with data scales up, traditional Extreme Learning Machines (ELMs) can no longer process serial data quickly. Duan et al. [3] proposed an efficient ELM based on the Spark framework, which greatly reduces the computational cost of the learning machine. In addition to big data, computer vision is beginning to help humans solve a wide variety of problems. Duan et al. [4] proposed a multi-attribute tensor correlation neural network, MTCN, to distinguish high-level features and original features in face, which promotes the development of face recognition towards high accuracy. Pu et al. [5] proposed FUSPR, an automatic fetal ultrasound plane recognition model, which enables precise positioning and tracking fetus across frames, promoting the development of smart healthcare. The work in this paper applies computer vision to the field of basketball, intending to be useful for practitioners involved.

With the continuous improvement of living conditions and the growing interest in various sports, automatic sports video analysis in computer vision has gradually become one of the most popular research areas [6]. Basketball is currently the hottest of several ball sports, the use of digital images, multimedia, and other technologies, can be statistical shooting hot zone distribution, singles success rate and other data presented to the coaches and athletes, science to improve the technical level of the players [7]. Therefore, it is of great practical significance to apply object detection technology to automate the recognition of basketball players' actions in video to help coaches analyze the team's tactics and formulate corresponding training plans.

Most of the current basketball player action detection methods use digital image processing techniques, human detection-categorization methods, and 3D convolutional neural network methods. Digital image processing technology realizes action detection by calculating the similarity of motion volume and combining some machine learning algorithms, but the performance of this method needs to improve. The human detection-categorization method first intercepts human body images using an object

*Corresponding Author: Min Zhu; E-mail: zhumin@zjsru.edu.cn
DOI: https://doi.org/10.70003/160792642025032602006

detection algorithm, then inputs the intercepted images into a classification network to classify them. The human detection-categorization method doesn't directly apply the object detection algorithm to basketball player action detection. The 3D convolutional neural network method needs to use a tracker in advance to manually select the ROI to track the player when recognizing the athlete's action. In this paper, we propose an end-to-end automatic recognition basketball player action detection algorithm YOLOSS to address the problems in the above studies, such as the method performance needs to improve or there are too many pre-steps. The YOLOSS algorithm is built based on YOLOv4, for strengthening the ability of the YOLOSS algorithm to localize and identify, we designed a hybrid attention SimSE module and a spatial pyramid pooling SPPFCSPCG module for the YOLOSS algorithm. The SimSE module retains spatial information and channel information, strengthening the ability of YOLOSS cross-dimensional information interaction after adding a small amount of overhead. The SPPFCSPCG module adds the CSP structure and group-convolution based on the SPPF module, which strengthens the ability of the YOLOSS algorithm to focus on the relevant region while keeping the sensory field unchanged, enhancing the detection accuracy of the YOLOSS algorithm. Finally, we also propose a basketball player action dataset BPAD, which contains 1,457 training images and 694 test images, containing three categories: running, walking, and defense. Overall, the contributions of this paper are as follows:

(1) We propose a BPAD basketball player action object detection dataset, which contains 1,457 training images and 694 test images, covering most of the typical actions of athletes while walking, running, and defending.

(2) We propose a hybrid attention SimSE module, which preserves channel information and spatial information, enhancing the algorithm's ability to interact cross-dimensional information while bringing a small amount of overhead.

(3) We propose a spatial pyramid pooling SPPFCSPCG module, which adds a CSP structure and group-convolution based on the SPPF module, enhancing the ability of the YOLOSS algorithm to focus on relevant regions while keeping the sensory field unchanged.

(4) We propose an end-to-end basketball player action detection model YOLOSS. The YOLOSS model achieves a mAP_{50} of 82.4% on the BPAD dataset, which is capable of accurately detecting the actions of players on the court.

2 Relate Work

2.1 Attention Mechanism

Attentional mechanisms mimic the human visual and cognitive systems to help neural networks learn and attend to important information in the input image. Hu et al. [8] proposed the SE module, which explicitly constructs dependencies between channels, adaptively recalibrates the channel features, and significantly improves the network performance with a small increase in computational cost. Wang et al. [9] proposed an efficient channel attention module ECA, which realizes local cross-channel interactions through one-dimensional convolution, adding a small number of parameters but bringing significant performance improvement. Park et al. [10] proposed the BAM attention module, which constructs hierarchical attention by calculating an Attention Map through the channel paths and spatial paths at the bottlenecks of the module. Experiments show that using the BAM for various models are improve performance. Woo et al. [11] proposed the convolutional attention module CBAM to compute the Attention Map in channel and spatial, through multiply it by the input feature maps for adaptive feature refinement. CBAM brings negligible overhead and consistent performance improvement. Liu et al. [12] proposed the GAM attention module, which enhances the cross-dimensional interaction capability of the module by preserving spatial information and channel information, improving the performance of the model. Yuan et al. [13] proposed a dynamic pyramid attention network, consisting of a self-attentive feature pyramid network (SAFPN) and a dynamic feature map selection module (DFMS), to learn the correlation between global pixels for high-precision object detection in remote sensing images. The work in this paper is applying the proposed hybrid attention module SimSE to a real-world scenario of basketball player action recognition by combining the SimSE hybrid attention module with an object detection algorithm.

2.2 Spatial Pyramid Pooling

The spatial pyramid module solves the challenge that the input images must be of the same size in a convolutional neural network and it is a common component module of object detection algorithms. He et al. [14] proposed a spatial pyramid pooling strategy that generates a fixed-length output regardless of the size and scale of the input of the image, which solves the deformation caused by resizing the image and affecting the final results. Ge et al. [15] used the SPP module in the Anchor Free detector YOLOX to help the network aggregate contextual information, achieved very excellent results on several object detection common datasets. The Ultralytics [16] team proposed the SPPF module in the YOLOv5 algorithm, which computes faster than the SPP module and computes exactly the same results as the SPP module. Inspired by the SPP module, Chen et al. [17] proposed the ASPP module in semantic segmentation modeling, which acquires multi-scale information of objects by constructing convolutions with different receptive fields with different rates. Wang et al. [18] proposed a new spatial pyramid pooling module SPPCSPC in the YOLOv7 algorithm, although the amount of computation and the number of parameters have risen, the performance of the detector is improved. Li et al. [19] proposed the SPPFCSPC module in the YOLOv6-3.0 detection algorithm, which borrows from the SPPF module and the SPPCSPC module, improving the speed and performance compared to SPPF and SPPCSPC. Cheng et al. [20] proposed the FPP feature pyramid pooling module in the MSYOLOF detector. The FPP module improves the network's ability to obtain global information by aggregating the context information of the region, which effectively improves the detection performance of MSYOLOF. The work in this paper proposes a new spatial pyramid pooling module-SPPFCSPCG, which combines the CSP structure and group-convolution, improving the computational speed and the ability of the module to focus on correlated regions, improving the performance of the YOLOSS algorithm on the BPAD dataset.

2.3 Application of Computer Vision in Basketball Players Action Recognition

Now, Computer vision is widely used in basketball-related endeavors to help practitioners obtain relevant information. Zhu Xia [21] proposes a basketball player shooting action recognition method based on Gaussian mapping, which uses Gaussian mapping to extract the target features of the shooting action image. According to the target features, the image is subjected to multilevel feature decomposition and blurring to realize the shooting action recognition of the shooting player. Alexandros Iosifidis et al. [22] utilize discrete Fourier variations and compute the similarity of action volumes to achieve fast action recognition. Pan et al. [23] used Gaussian hybrid models and gradient histograms to compute motion descriptors and pose descriptors of motion blocks, and through their linear combination, basketball action recognition was implemented using the KNN algorithm. Hao et al. [24] proposed a multi-objective corner tracking algorithm by using Otsu method for grayscale feature processing and Harris corner extraction algorithm to realize the fast detection of athlete's action. Lu et al. [25] extracted the player's motion pictures by YOLOv3 algorithm, denoising the images with filters, and the denoised athlete images of upper and lower limbs are input into a neural network for action recognition. Sareshkeh et al. [26] proposed a method based on the combination of YOLO and LSTM network to detect the players in the video. The athlete images are extracted by YOLO algorithm, then the extracted images are fed into the LSTM network to realize athlete action classification. Shi et al. [27] implemented 3D athlete action detection through a 2D key-point detection algorithm with a 3D fuzzy neural network module, which is able to adapt to the fuzzy data in classification and has a better recognition of the athlete's jumping action. Xiao et al. [28] improved the problem of insufficient feature extraction by 3D convolutional neural network, proposing a method based on dynamic residual Attention Mechanism of Deep Neural Network Basketball Action Recognition, which recognizes actions with an average accuracy of more than 97%. Within the intersection of basketball player action recognition and computer vision, most of the research on athlete action recognition has used traditional digital image processing techniques, human detection-categorization methods, and 3D convolutional neural network methods. The work in this paper applies object detection algorithms to basketball player action recognition, and contributes a basketball player action object detection dataset BPAD, containing 1,457 training images and 694 test images, covering three common body movements of athletes.

3 Relevant Methods

3.1 YOLOv4 SimSE and SPPFCSPCG

The network structure of the YOLOSS (YOLOv4 SimSE and SPPFCSPCG) algorithm is shown in Figure 1. The YOLOSS model consists of the GhostBottleneck module, the SimSE module, and the SPPFCSPCG module. The GhostBottleneck module is the main module in the YOLOSS backbone network, which consists of GhostModule. GhostMudule uses convolution to compute only part of the dimensions in the feature maps, while the feature maps of the other part of the dimensions are synthesized using linear combinations, which reduces the efficiency degradation caused by computation redundant information, making the YOLOSS network more efficient.

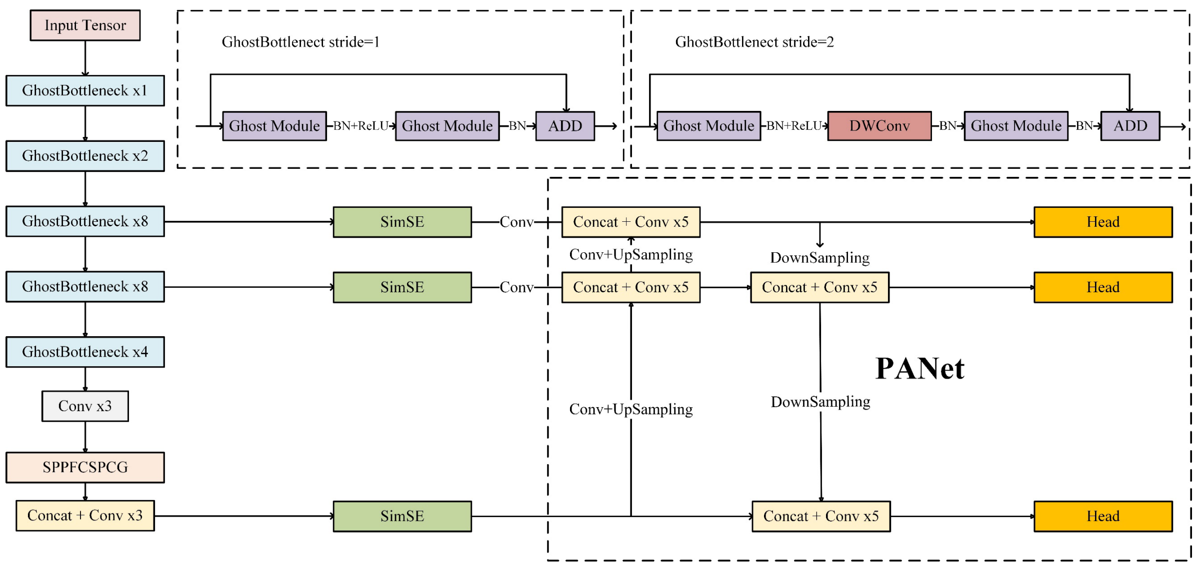


Figure 1. YOLOSS network structure diagram

The SPPFCSPCG spatial pyramid pooling module, which enables adaptive size output and combines the CSP structure and group-convolution, strengthens the ability of YOLOSS to focus on the relevant regions, improving the efficiency and accuracy of the YOLOSS algorithm. The SimSE hybrid attention module retains the spatial information and channel information in the feature map, which strengthens the module's ability to communicate cross-dimensional information and improves the algorithm's localization and recognition capabilities.

3.2 Ghostnet Backbone

The Ghostnet [29] network model has a Top-1 correctness rate of 75.7% on the ImageNet classification task, in which the core of the Ghostnet network is the GhostModule. The difference between GhostModule and regular convolution is shown in Figure 2. Due to the generality of GhostModule, it is used to replace any convolutional operation in any convolutional neural network, which is very lightweight and efficient.

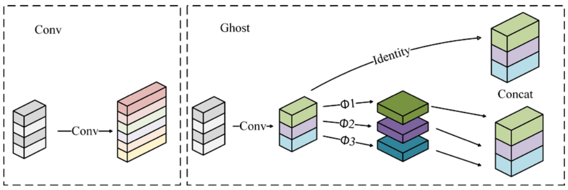


Figure 2. Computational principles of ordinary convolution and GhostModule

The results of the visualization of the feature maps by the first residual block of the Resnet50 network show that there are some feature maps that are highly similar, and these feature maps lead to a very high amount of redundant information, leading to some completely unnecessary computations. GhostModule first uses 1*1 convolution to compress the number of channels of the feature map, then performs a series of simple linear computations on each of the original feature maps to obtain the Ghost feature map. Finally, stitches them all together to form a new feature map. The inference procedure for the theoretical acceleration ratio r_s and the theoretical parameter compression ratio r_c for GhostModule and ordinary convolution is as follows:

$$\begin{aligned} r_s &= \frac{n \cdot h' \cdot w' \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot h' \cdot w' \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot h' \cdot w' \cdot d \cdot d} \\ &= \frac{c \cdot k \cdot k}{\frac{1}{s} \cdot c \cdot k \cdot k + \frac{s-1}{s} \cdot d \cdot d} \approx \frac{s \cdot c}{s+c-1} \approx s \end{aligned} \tag{1}$$

$$r_c = \frac{n \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot d \cdot d} \approx \frac{s \cdot c}{s+c-1} \approx s \tag{2}$$

where n is the number of convolution kernels, h' and w' are the size of the feature maps after convolution computation, c is the number of channels, k is the convolution sum size, and s is the number of phantom feature maps obtained, it can be seen that the computational speed and parameter sizes of GhostModule is $1/s$ of the normal convolution.

The structure of the GhostBottleneck network is shown in Figure 1. YOLOSS algorithm takes full advantage of GhostModule, and GhostBottleneck is divided into different phases according to the size of the input feature map. Except the last Ghostbottleneck of each stage which is stride=2, all other Ghostbottlenecks are applied with stride=1. The GhostBottlenecks are utilized to form the backbone network of YOLOSS to improve the detection efficiency of the algorithm.

3.3 Hybrid Attention SimSE

SimSE Hybrid Attention is a new attention module we propose. It combines SE channel attention and SimAM parameter-free attention, which preserves the spatial information and channel information of the feature map, strengthens the ability of the YOLOSS algorithm to communicate cross-dimensional information, and can effectively improve the accuracy of detecting players' positions and classifying their actions through the efficient combination of SE attention and SimAM attention.

SimAM Attention starts from neuroscience theory, modeled after human brain neurons, assesses the importance of each neuron. Information-rich neurons will show different power from ordinary nerves and inhibit peripheral neurons. SimAM distinguishes the importance of neurons and calculates attention weights through energy function, and the formula of SimAM augmented features is as follows:

$$sigmoid(x) = \frac{1}{1+e^{-x}} \tag{3}$$

$$\tilde{X} = sigmoid(\frac{1}{E}) \times X \tag{4}$$

where X is the input feature map, $\tilde{X}$ is the output feature map, and E is the result of the energy function calculation, the inverse of the result of the energy function calculation is normalized to the interval of 0~1 by the *sigmoid* function, and then multiplied with the input feature map to enhance the weight of the important features.

SE Attention Mechanism is a channel attention mechanism, the operation process of the SE Attention Mechanism is shown in Figure 3, which employs the two-step operation of Squeeze and Excitation. SE attention is able to perform channel feature enhancement on the input feature maps without changing the dimensions of the input feature maps. If the feature map U is of size $H \times W \times C$, the Squeeze operation compresses the full spatial information into dimensions through a global pooling layer, generating a feature vector of size $1 \times 1 \times C$. The Excitation operation utilizes the feature vectors aggregated in the previous Squeeze operation to learn the information in the channel through the fully connected layer and activation function, and finally the weight vector obtained from the Excitation operation is used to assign weights to the feature map U. The specific formula is as follows:

$$\tilde{x}_c = Fscale(u_c, s_c) = s_c u_c \quad \textbf{(5)}$$

where s_c is the weight vector obtained after the Squeeze and Excitation two-step operations, u_c is the feature map of the corresponding dimension on the feature map U, and $\tilde{x}_c$ is the corresponding individual output obtained from the computation.

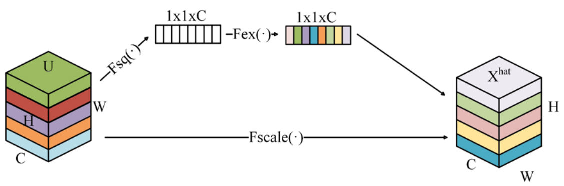


Figure 3. SE calculation schematic

The calculation process of the SimSE module is shown in Figure 4, the SimSE module is modeled after the BAM attention mechanism, and the feature maps are inputted into the SimAM attention mechanism module and the SE attention mechanism module respectively to replace the original spatial attention module and the channel attention module in the BAM structure, and the weight coefficients are multiplied by the original feature maps to obtain the new feature maps. The calculation formula of the SimSE module is as follows:

$$Atten = SimAM(X) + SE(X) \quad \textbf{(6)}$$

$$Y = X + sigmoid(Atten) \quad \textbf{(7)}$$

where X is the input feature map and Y is the output feature map, the outputs of the SimAM module and SE module are summed up and then normalized to the 0~1 interval by the sigmoid function and then summed up with the input feature map to get the enhanced feature map.

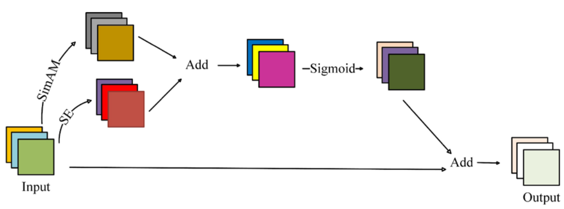


Figure 4. SimSE calculation schematic

The SimSE module is able to retain channel information and spatial information while bringing in a small number of parameters, enhancing the algorithm's ability to interact cross-dimensional information, improving the accuracy of the algorithm's detection.

3.4 SPPFCSPCG Space Pyramid Pooling

The SPPFCSPCG module is inspired by the SimCSPSPPF module in YOLOv6 [19], and based on this modules, we take advantage of the high efficiency of group -convolutional computation to improve the efficiency of the YOLOSS algorithm detection.

Group-convolution was first applied in neural networks, mainly used to solve the problem of insufficient memory. Now, it is widely used in a variety of lightweight models, the computational process of group-convolution is shown in Figure 5, compared with the conventional convolution, the number of parameters needed for group-convolution is only $1/G$, and the specific inference formula is as follows:

$$r = \frac{k \cdot k \cdot C1 \cdot C2}{k \cdot k \cdot C1 \cdot C2 \cdot G} = \frac{1}{G} \quad \textbf{(8)}$$

where k is the size of the convolution kernel, $C1$ is the number of input channels, $C2$ is the number of output channels, and G is the number of groups in the grouped convolution.

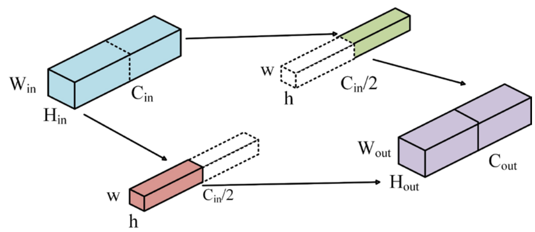


Figure 5. Schematic diagram of grouped convolutional computation

The network structure of the SimCSPSPPF module is shown in Figure 6. The SimCSPSPPF module is an optimization of the previous spatial pyramid pooling module by the YOLOv6 object detection algorithm. Compared with the SPPF module in YOLOv5, it is modeled after the CSP structure in CSPnet, which divides the original input into two dimensions, performs the convolution operation separately, so that the model learns more features. And the CSP structure can reduce the computation amount, enhancing the model's characterization ability for complex object, so as to achieve a balance between speed and accuracy.

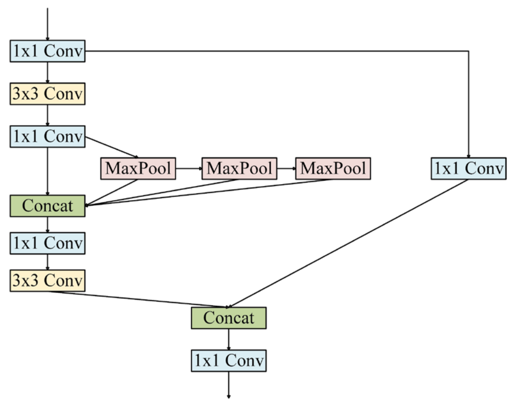


Figure 6. SimCSPSPPF network structure diagram

The network structure of the SPPFCSPCG spatial pyramid pooling module is shown in Figure 7. The SPPFCSPCG module combines the advantages of the two methods mentioned above, replacing all ordinary convolution operations in spatial pyramid pooling with g=4 group-convolutions, which can significantly improve the inference speed of the detection algorithm.

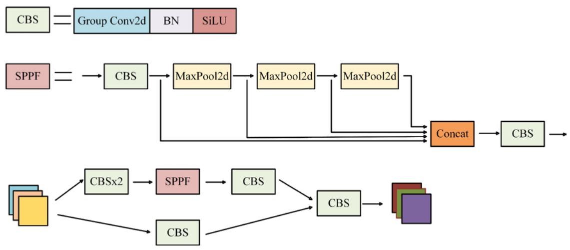


Figure 7. CSPSPPFCG network structure diagram

The SPPFCSPCG spatial pyramid pooling module enhances the detection efficiency of the YOLOSS algorithm, strengthens the ability of YOLOSS to focus on relevant regions, improving the accuracy of the algorithm's localization and identification.

4 Experimental Results and Analysis

4.1 Experimental Data and Experimental Environment

The experimental environment is a NVIDIA A30 with a graphics card with 24GB of video memory, an Intel(R) Xeon(R) Silver 4314 CPU @ 2.40GHz, a memory size of 64GB, a system with Ubuntu 20.04.1, a CUDA version of 11.3, a programming language of Python 3.8, and a deep learning framework of Pytorch version 1.12.1, and all the algorithm results were tested on this experimental environment. Specific experimental training data are shown in Table 1.

Table 1. Experimental training data

Model	Batch	Optimizer	Learning rate
YOLOSS	32	Adam	1e-2
YOLOv4	32	Adam	1e-2
YOLOv5-s	32	Adam	1e-2

The evaluation metrics used in the experiments are AP and mAP, which are commonly used in object detection to measure the performance of models. AP denotes the average accuracy of the detector for a single category, and mAP averages the AP from the dimensions of the categories and evaluates the multiclassification performance of the detector. Higher metrics denote better classification and localization of the detector.

The experimental dataset is the BPAD basketball player action detection dataset, the data comes from the video of a basketball game of a college basketball team, extracting the video and saving a picture every 5 frames, collecting 2151 pictures and labeling them with the Labelimg tool. In order to maintain a clear and concise sports scene as much as possible, the referee kickoffs, coaches and players discussions, and other irrelevant scenes are excluded. The data only contains the sports actions of the players running, defending, and walking on the field. The sample images of the BPAD dataset are shown in Figure 8.

Figure 8. Example image of the BPAD dataset

4.2 Experimental Results and Visualization Analysis

For this experiment, we compared the AP of YOLOSS, YOLOv4, and YOLOv5-s on BPAD data for the three categories as well as mAP for IoU=0.5. As can be seen in Table 2, YOLOSS obtained the highest scores in recognizing the Walk and Run categories, however, it performed poorly in recognizing the Defense category, but overall, YOLOSS had the highest recognition accuracy of all algorithms for all categories.

Table 2. Experimental results of YOLOSS with advanced algorithms

Model	Walk	Run	Defense	mAP
YOLOSS	**87.83%**	**73.29%**	86.09%	**82.4%**
YOLOv4	85.94%	68.15%	**87.45%**	81.39%
YOLOv5-s	81.52%	39.89%	58.04%	59.82%

As shown in Figure 9 and Figure 10, the left side shows the prediction results of the YOLOv5-s algorithm and the right side shows the prediction results of the YOLOSS algorithm. In Figure 9, YOLOv5-s incorrectly predicts players in Run state as Walk state. In Figure 10, YOLOv5-s incorrectly predicts the motion state of the attacking player and fails to detect the defending player, resulting in misdetections and missed detections.

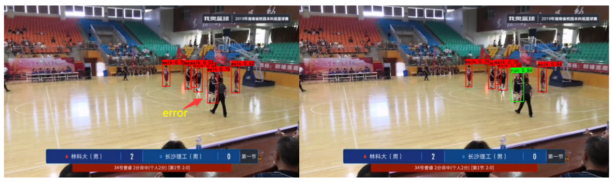


Figure 9. YOLOv5-s vs YOLOSS prediction1

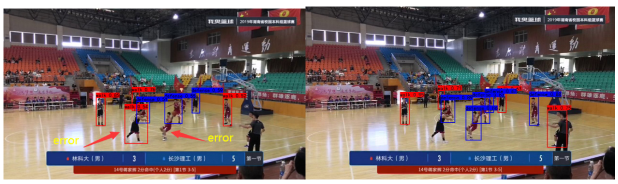


Figure 10. YOLOv5-s vs YOLOSS prediction2

As shown in Figure 11 and Figure 12, the left side is the heat map of YOLOv5-s, and the right side is the heat map of YOLOSS, from which we can find out the reasons for YOLOv5-s misdetections and omissions in prediction. YOLOv5-s focuses on hotspot regions that are very scattered, while the YOLOSS algorithm improves the ability of the model to focus on key regions on the basis of the SimSE module and the SPPFCSPCG module, hot regions are on the athletes, so it does not result in misdetections and omissions, improving the detection accuracy of the model.

Figure 11. YOLOv5-s vs YOLOSS heat map1

Figure 12. YOLOv5-s vs YOLOSS heat map2

5 Conclusion

In this paper, to solve the problems of difficult basketball player action detection and fewer detection dataset, we propose a basketball player action detection dataset BPAD and an end-to-end automatic algorithm YOLOSS for analyzing player actions, and achieving advanced results. In the YOLOSS algorithm, we introduced the hybrid attention SimSE module and the SPPFCSPCG spatial pyramid pooling module, which effectively improve the localization and recognition ability of the YOLOSS algorithm, and can accurately detect the basketball player's actions after training on the BPAD dataset. In the future, our work will focus on collecting images of athletes' actions, expanding the number and categories of the dataset, further improving the YOLOSS to enhance the ability of detection, and adding specialized sports information such as shooting hotspot areas.

References

[1] J. Wang, Y. Yang, T. Wang, R. S. Sherratt, J. Zhang, Big data service architecture: a survey, *Journal of Internet Technology*, Vol. 21, No. 2, pp. 393-405, March, 2020.

[2] J. Zhang, S. Zhong, T. Wang, H.-C. Chao, J. Wang, Blockchain-based Systems and Applications: A Survey, *Journal of Internet Technology*, Vol. 21, No. 1, pp. 1-14, January, 2020.

[3] M. Duan, K. Li, X. Liao, K. Li, A Parallel Multiclassification Algorithm for Big Data Using an Extreme Learning Machine, *IEEE transactions on neural networks and learning systems*, Vol. 29, No. 6, pp. 2337-2351, June, 2018.

[4] M. Duan, K. Li, K. Li, Q. Tian, A Novel Multi-task Tensor Correlation Neural Network for Facial Attribute Prediction, *ACM Transactions on Intelligent Systems and Technology*, Vol. 12, No. 1, pp. 1-22, February, 2021.

[5] B. Pu, K. Li, S. Li, N. Zhu, Automatic Fetal Ultrasound Standard Plane Recognition Based on Deep Learning and IIoT, *IEEE Transactions on Industrial Informatics*, Vol. 17, No. 11, pp. 7771-7780, November, 2021.

[6] W. Wu, M. Xu, Q. Liang, L. Mei, Y. Peng, Multi-camera 3D ball tracking framework for sports video, *IET Image Processing*, Vol. 14, No. 15, pp. 3751-3761, December, 2020.

[7] Z. Hu, Application of Digital Image Analysis and Processing Technology in Basketball in the Era of Big Data, *2021 International Conference on Big Data*, Artificial Intelligence and Risk Management (ICBAR), Shanghai, China, 2021, pp. 51-54.

[8] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, Salt Lake City, Utah, USA, 2018, pp. 7132-7141.

[9] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-Net: Efficient channel attention for deep convolutional neural networks, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 11531-11539.

[10] J. Park, S. Woo, J.-Y. Lee, I. S. Kweon, *Bam: Bottleneck attention module*, arXiv, July, 2018. https://arxiv.org/abs/1807.06514

[11] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, Cbam: Convolutional block attention module, *Proceedings of the European conference on computer vision (ECCV)*, Munich, Germany, 2018, pp. 3-19.

[12] Y. Liu, Z. Shao, N. Hoffmann, *Global attention mechanism: Retain information to enhance channel-spatial interactions*, arXiv, December, 2021. https://arxiv.org/abs/2112.05561

[13] H. Yuan, H. Zhou, Z. Cai, S. Zhang, R. Wu, Dynamic Pyramid Attention Networks for multi-orientation object detection, *Journal of Internet Technology*, Vol. 23, No. 1, pp. 79-90, January, 2022.

[14] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 37, No. 9, pp. 1904-1916, September, 2015.

[15] Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun, *Yolox: Exceeding yolo series in 2021*, arXiv, August, 2021. https://arxiv.org/abs/2107.08430

[16] Ultralytics, YOLOv5, 2020-6-9, Available from: https://github.com/ultralytics/yolov5.

[17] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 40, No. 4, pp. 834-848, April, 2018.

[18] C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023, pp. 7464-7475.

[19] C. Li, L. Li, Y. Geng, H. Jiang, M. Cheng, B. Zhang, Z. Ke, X. Xu, X. Chu, *Yolov6 v3.0: A full-scale reloading*, arXiv, January, 2023. https://arxiv.org/abs/2301.05586

[20] G. Cheng, X. Yong, X. Lyu, T. Zeng, X. Wang, J. Chen, X. Li, MSYOLOF: Multi-input-single-output encoder network with tripartite feature enhancement for object detection,

Proceedings of the 2023 5th International Conference on Pattern Recognition and Intelligent Systems, Shenyang, China, 2023, pp. 36-44.

[21] Z. Xia, A recognition method of basketball players' shooting action based on Gaussian mapping, *International Journal of Reasoning-based Intelligent Systems*, Vol. 15, No. 2, pp. 105-110, March, 2023.

[22] A. Iosifidis, A. Tefas, I. Pitas, Multi-view action recognition based on action volumes, fuzzy distances and cluster discriminant analysis, *Signal Processing*, Vol. 93, No. 6, pp. 1445-1457, June, 2013.

[23] Z. Pan, C. Li, Robust basketball sports recognition by leveraging motion block estimation, *Signal Processing: Image Communication*, Vol. 83, Article No. 115784, April, 2020.

[24] Z. Hao, X. Wang, S. Zheng, Recognition of basketball players' action detection based on visual image and Harris corner extraction algorithm, *Journal of Intelligent & Fuzzy Systems*, Vol. 40, No. 4, pp. 7589-7599, April, 2021.

[25] Z. Lu, Algorithm of Basketball Posture Motion Feature Extraction Based on Image Processing Technology, *2022 International Conference on Frontiers of Artificial Intelligence and Machine Learning (FAIML)*, Hangzhou, China, 2022, pp. 219-223.

[26] S. B. Khobdeh, M. R. Yamaghani, S. K. Sareshkeh, Basketball Action Recognition Based on the Combination of YOLO and a Deep Fuzzy LSTM Network, Vol. 80, No. 3, pp. 3528-3553, February, 2024.

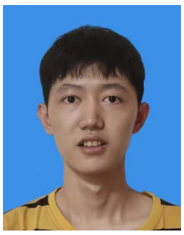
[27] Y. Shi, Z. Hu, Recognizing the take-off action pattern of basketball players based on fuzzy neural network system, *Journal of Ambient Intelligence and Humanized Computing*, Vol. 14, No. 2, pp. 1071-1081, February, 2023.

[28] J. Xiao, W. Tian, L. Ding, Basketball Action Recognition Method of Deep Neural Network Based on Dynamic Residual Attention Mechanism, *Information*, Vol. 14, No. 1, Article No. 13, January, 2023.

[29] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, C. Xu, Ghostnet: More features from cheap operations, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 1577-1586.

Biographies

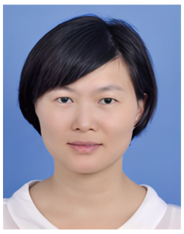
Zhihua Tan received the B.S. and M.S. degree from Hunan Normal University, Changsha, China, in 2002, 2016 respectively. Now, he is an associate professor at Changsha University of Science and Technology. His current research interests include sports training and sports humanities and sociology.



Sheng Gao received the B.S. degree from Chongqing University of Science and Technology, Chongqing, China, in 2022. He is a postgraduate student in Changsha University of Science and Technology, Changsha, China. His research interests include object detection, and deep learning.

Shihai Wei received the B.S. and M.S. degree from Wuhan Sports University, Wuhan, China, in 2000, 2010 respectively. Now, he is a lecturer at Changsha University of Science and Technology. His current research focuses on sports statistics.

Jingyu Zhang received the Ph.D. degree in Computer Science and Technology from Shanghai Jiao Tong University in 2017. He is currently a Distinguish Associate Professor at the School of Computer & Communication Engineering, Changsha University of Science and Technology, China. His research interests include computer architecture, big data and blockchain.



Min Zhu received B.S. degree from Nanjing University of Posts and Telecommunications, China in 2002, M.S. degree from Beijing University of Posts and Telecommunications, China in 2005, and received Ph.D. degree in Nanjing University of Posts and Telecommunications, China in 2018. Now, she works at College of Information Science and Technology, Zhejiang Shuren University as lecturer. Her research interests mainly include routing protocol and optimization algorithm design.