ISSN: 2834-7706

Research Article

Journal of Robotics and Automation Research

AI-powered Badminton Video Detection: Enhancing Gameplay Analysis and Training

Yao Kevin Cheng[1] and Junghee Jay Kim[2]*

[1]*Arnold O.Beckman High School, Irvine, CA 92602*

[2]*dslite.ai, Yorba Linda CA 92886*

***Corresponding Author**
Junghee Jay Kim, dslite.ai, Yorba Linda CA 92886.

Submitted:2023, July 10; **Accepted**:2023, July 31: **Published**: 2023, Aug 04

Citation: Cheng, Y. K., Kim, J. J. (2023). AI-powered Badminton Video Detection: Enhancing Gameplay Analysis and Training. *J Robot Auto Res, 4(*2), 392-402.

Abstract
This research focuses on the use of AI-powered badminton video analysis to enhance gameplay analysis and training. The technology utilizes artificial intelligence and machine learning algorithms to analyze varJaccard Similarity Index aspects of badminton game footage, including player movements, shot selection, and game strategy. It provides personalized feedback and recommendations for improvement to players and helps coaches identify patterns and trends in their players' performance.

The use of lightweight models, such as YoloV5, is essential for real-time video analysis due to the need for high-speed processing. These models, including MobileNets, EfficientNets, SqueezeNet, ShuffleNet, and Tiny-YOLO, are designed to be lightweight and optimized for speed while maintaining high accuracy. The Yolo model, a one-shot learning model, is particularly suitable for real-time object detection tasks due to its impressive speed and accuracy. It uses anchor boxes and data augmentation techniques to quickly learn and recognize objects with a high level of accuracy.

The research process involved the collection of badminton match videos, preparation of datasets, and testing of the YoloV5 model's performance. Precision and Jaccard Similarity Index metrics using Jaccard similarity were used to evaluate the model's performance in detecting player positions and court boundaries. The results showed that the YoloV5 model performed better when tested on the video data with less clear image of judges in the background. The Jaccard Similarity Index metrics using Jaccard similarity demonstrated improved performance when tested on generalized video data was used to accurately evaluate the overlap between predicted and ground truth Overall, AI-powered badminton video analysis has the potential to revolutionize the way badminton is played and coached. The use of lightweight models like YoloV5 enables faster and more efficient real-time analysis, making it practical for a wide range of applications in badminton.

Keywords: Artificial Intelligence, Badminton, Analysis, Machine Learning, Deep Learning, Video Analysis, Real time Analytics

Introduction

Using computer vision and deep learning algorithms, AI-powered badminton video analysis can identify and track players and courts. It can also analyze varJaccard Similarity Index aspects of the game such as player movements [1].

This technology can provide players with personalized feedback and recommendations for improvement based on their strengths and weaknesses. It can also help coaches to identify patterns and trends in their players' performance, and adjust their training programs accordingly [2]. However, while AI video analyzers have great potential in sports like badminton, there are some challenges and potential problems that should be addressed. In some cases, such as during live game broadcasts or real-time coaching, AI video analyzers need to process video data in real-time [3]. This can be challenging due to the large amount of data and the need for high-speed processing, there is a growing need for models that can perform these tasks quickly and efficiently. As video analysis involves processing large amounts of data in real-time, it is essential for these models to be lightweight and optimized for speed.

To address this challenge, researchers have been developing lightweight models for video AI recognition, such as MobileNets, EfficientNets, SqueezeNet, ShuffleNet, and Tiny-YOLO. [4] These models are specifically designed to reduce the number of parameters and computations required while still maintaining high accuracy, making them ideal for real-time video analysis on embedded devices [5]. The YoloV5 model is one such example of a lightweight and fast one-shot learning model that has gained popularity in recent years. [6] Its one-shot learning approach allows it to quickly learn and recognize objects with a high level of accuracy, making it an attractive option for applications where speed is critical [7]. Overall, the development of lightweight models for video AI recognition is essential for the continued advancement and widespread adoption of AI video analyzers, as it enables faster and more efficient real-time analysis of video data, making it more accessible and practical for a wide range of applications [8]. Our proposed lightweight YoloV5 model showed it could manage large amount of video data more accurately with average precision.

2. Literature Review

In recent years, the application of artificial intelligence (AI) in sports has gained significant attention. One area where AI has shown promise is in the analysis of sports videos for performance improvement. Specifically, in the sport of badminton, AI-powered video analysis has become a topic of interest for players and coaches alike [9]. Several studies have explored the use of AI in analyzing badminton videos. One study utilized deep learning algorithms to detect and track badminton shuttlecocks in video footage. The results showed high accuracy in detecting shuttlecocks, which could be useful for analyzing the speed and trajectory of shots [10].

Another study utilized computer vision techniques to track player movements and identify key performance indicators (KPIs) such as shot placement, speed, and accuracy [11]. The AI-powered analysis provided insights into player performance, which could aid coaches in designing training programs and improving player technique [12]. And also There were several efforts to use AI techniques in badminton analysis, including computer vision, machine learning, and deep learning, all of which have shown great potential in automating the analysis of complex game scenarios and providing valuable insights for coaches, players, and fans. Paper “A deep learning based framework for badminton rally outcome prediction (2022)” proposed a player-independent framework to investigate the relationship between strokes and rally outcome in badminton games. To classify the rally outcome, strokes are represented by deep features extracted using CNN and fitted into LSTM [13]. Paper “Shot detection using skeleton position in badminton videos (2021)” proposed a shot detection method using the poses of a player in a badminton video sequence. In this method, the hit timing is detected by focusing on the arm movements of the player and analyzing the swing movement using skeletal information [14].

and another paper called Vision Based Automated Badminton Action Recognition Using the New Local Convolutional Neural Network Extractor (2020) proposed automated badminton action recognition from the computer vision data inputs using the deep learning pre-trained AlexNet Convolutional Neural Network (CNN) for features extraction and classify the features using supervised machine learning method which is linear Support-Vector Machine (SVM) and achieved an accuracy of 98.7% [15]. On the other hand, another paper called A Comparative Study on Deep Learning Architectures for Badminton Action Recognition (2021) compared the performance of different deep learning architectures, including ResNet, Inception-v3, and DenseNet, for badminton action recognition. The models were trained on a dataset of 1,500 videos and evaluated on a test set of 300 videos [16]. Overall, the literatures suggests that AI-powered video analysis has the potential to revolutionize the way badminton is played and coached. By providing accurate and insightful data, AI-powered analysis can help players and coaches make more informed decisions, leading to better performance and outcomes [17].

3. Methodology

3.1. Application Development

Main Structure:

1. Five parts: ai.py, image.py, Court.py, Video.py and app.py.

• ai.py :Yolo (You Only Look Once) deeplearning model is implemented in this code and it detects players within images.

• image.py : it loads video data and Yolo model detection model is applied and also court detection algorithm is applied and create detected images

• court.py : the court contours and corners are identified and then it draws them on images.

• video.py : court detected image and people detected images are combined and video is generated.

• app.py : video streams are loaded to app’s memory and become rendered for playback.

2. Video clips are all uploaded to this program and processed to app.

Here’s our flow chart. Let’s take a look at the 5 major components that we can identify.

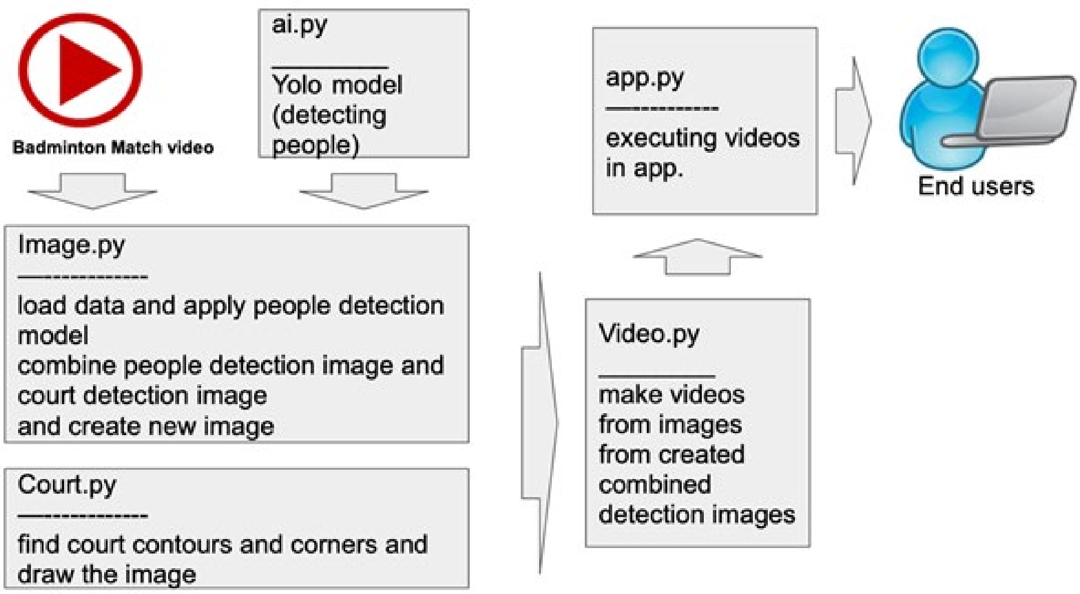

Figure 1: A.I. Application Component Diagram

3.2. Model Selection

Because A.I. video analyzer needs to be fast, lightweight models for video AI recognition are considered. Developing lightweight models for video AI recognition has been an area of active research in recent years, as it can help improve real-time performance and reduce computational requirements. Some of the lightweight models in recent years are as below. MobileNets: MobileNets are a family of lightweight deep neural network architectures that have been specifically designed for mobile and embedded devices. These models use depth-wise separable convolutions to reduce the number of parameters and computations required, while still maintaining high accuracy on tasks such as object detection and image classification [18]. EfficientNets: EfficientNets are a family of convolutional neural network architectures that have been optimized for both accuracy and efficiency. These models use a combination of depth-wise separable convolutions, linear bottlenecks, and compound scaling to achieve state-of-the-art performance on tasks such as image classification, object detection, and semantic segmentation [19].

SqueezeNet: SqueezeNet is a lightweight convolutional neural network architecture that uses 1x1 convolutions to reduce the number of parameters while still maintaining high accuracy. This model was specifically designed for real-time video analysis on embedded devices and has been shown to achieve high accuracy on tasks such as image classification and object detection [20]. ShuffleNet: ShuffleNet is a family of lightweight convolutional neural network architectures that use group convolutions and channel shuffling to reduce the number of parameters while still maintaining high accuracy. These models have been shown to achieve state-of-the-art performance on tasks such as image classification and object detection, while using significantly fewer computations than other models [21]. Tiny-YOLO: Tiny-YOLO is a lightweight version of the YOLO (You Only Look Once) object detection model, specifically designed for real-time performance on embedded devices. This model uses a combination of depth-wise separable convolutions and residual connections to achieve high accuracy while using significantly fewer computations than other object detection models [22].

We used YoloV5 because YoloV5 is a one-shot learning model that has gained popularity in recent years due to its impressive speed and accuracy. This model is specifically designed for real-time object detection tasks, and its one-shot learning approach allows it to quickly learn and recognize objects with a high level of accuracy, making it an attractive option for applications where speed is critical. YoloV5 is an object detection model that is based on a one-shot learning approach, which allows it to quickly learn and recognize objects with a high level of accuracy. It is the fifth iteration of the You Only Look Once (YOLO) series of object detection models, which were first introduced in 2015 by Joseph [23]. The YoloV5 model is designed to be both fast and accurate, making it suitable for real-time object detection tasks. It uses a convolutional neural network (CNN) architecture with a series of convolutional layers, followed by max-pooling layers, and finally fully connected layers. The architecture is optimized to minimize the number of computations required, which makes it suitable for running on resource-constrained devices such as mobile phones and embedded systems [24].

One of the key features of YoloV5 is its one-shot learning approach, which allows the model to quickly learn and recognize objects in a single pass through the network. This is achieved through the use of anchor boxes, which are pre-defined bounding boxes that the model uses to predict the location and size of objects in the image. By using anchor boxes, the model can quickly learn to recognize objects with a high level of accuracy, even when the objects have different shapes and sizes [25]. Another important feature of YoloV5 is its use of data augmentation techniques, such as random scaling, cropping, and flipping of the input images. This helps to increase the variability of the training data, which in turn improves the robustness of the model to different lighting conditions, orientations, and backgrounds [26].

Overall, YoloV5 is a state-of-the-art object detection model that is both fast and accurate, making it suitable for a wide range of real-time object detection tasks. Its one-shot learning approach and data augmentation techniques help to improve its accuracy and robustness, making it a popular choice for applications such as autonomous driving, robotics, and surveillance systems [27]. The YoloV5 architecture is a convolutional neural network (CNN) that consists of a series of convolutional layers, followed by max-pooling layers, and finally fully connected layers. The architecture is designed to minimize the number of computations required, which makes it suitable for running on resource-constrained devices such as mobile phones and embedded systems [28]. The YoloV5 architecture uses a backbone network, which is a series of convolutional layers that extract features from the input image. The backbone network is based on the CSPNet (Cross-Stage Partial Network) architecture, which is optimized for both accuracy and efficiency [29].

The output of the backbone network is fed into a neck network, which is responsible for further feature extraction and fusion. The neck network uses a combination of 3x3 and 1x1 convolutional layers to fuse the features from different scales and resolutions [30]. The output of the neck network is then passed through the detection head, which is responsible for predicting the location and class of objects in the input image. The detection head uses anchor boxes to predict the location and size of objects, and it uses a multi-label soft-margin loss function to train the model [31]. The YoloV5 architecture has several different variants, including YoloV5s, YoloV5m, YoloV5l, and YoloV5x, which differ in terms of the number of layers and the number of filters in each layer. The YoloV5s variant has the fewest number of layers and filters, while the YoloV5x variant has the most layers and filters, making it the most accurate but also the most computationally expensive [32]. Overall, the YoloV5 architecture is optimized for both speed and accuracy, making it suitable for real-time object detection tasks on a wide range of devices. Its backbone network, neck network, and detection head work together to extract features and predict object locations and classes with a high degree of accuracy [33].

3.3. Testing Flowchart

The research process flow is depicted in Figure 2 to provide a clearer understanding, with each sub-process discussed in Sections 3.3.1 to 3.3.3.

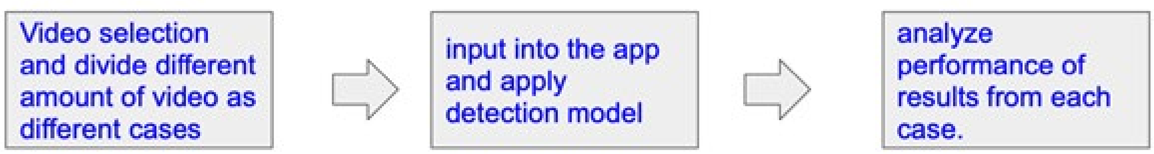

Figure 2: Testing Flowchart

3.3.1 Selecting Videos

Numerous sports videos are readily accessible on varJaccard Similarity Indexs online platforms like YouTube, blogs, and official sports websites.

- case 1 : Men’s badminton match video dataset
- case 2 : Women’s badminton match video dataset
- case 3 : Junior men’s badminton match video dataset
- case 4 : Junior women’s badminton match video dataset

This research employed Yolo to analyze badminton broadcast videos, including the Senior National Badminton Championships (Figure 3), PV Sindhu vs. Tai Tzu-Ying Women's Badminton Round of 16 at Rio 2016 (Figure 4), the Men’s under 15 at Nanjing 2014 Youth Olympic Games (Figure 5) and the Women under 15 at Princess Sirivannavari Thailand Masters 2017 (Figure 6)

Figure 3: Example of Image of Sample Video Case 1

Figure 4: Example of Image of Sample Video Case 2

Figure 5: Example of Image of Sample Video Case 3

Figure 6: Example of Image of Sample Video Case 4

3.3.2 Applying The Data To Yolo Model

Once the dataset was prepared, it was utilized to test our yolo model. When applying the data to our app model, predicted dataset were created using the prepared dataset. After the preparation phase, the performance of the our app model was evaluated using a testing dataset as different cases of videos, as presented in Table 1.

Case	Types of Model and test video	Test datasets
1	Model 1: Yolo V5 with video 2	video 1 adult men's match
2	Model 1: Yolo V5 with video 3	video 2 adult women's match
3	Model 1: Yolo V5 with video 4	video 3 men's match under 15
4	Model 1: Yolo V5 with video 1	video 4 women's match under 15

Table 1: datasets used for testing model performance

3.3.3 Analyze Performance of Results From Each Case.

Our app models were tested using different combinations of testing videos as specified in Table 1 to assess their ability to track the player's position. Once the testing was completed, the models produced an image frame with a square box indicating the level of confidence in detecting the player. To evaluate the detector's performance in each scenario, precision graphs were generated, and the average precision was calculated.

4. Result And Discussion

The performance of each model was evaluated using accuracy, with the average precision used to evaluate the model. Scores indicating the badminton player's position were shown in Figure 7, with yolo model consistently detecting the player in the same video data

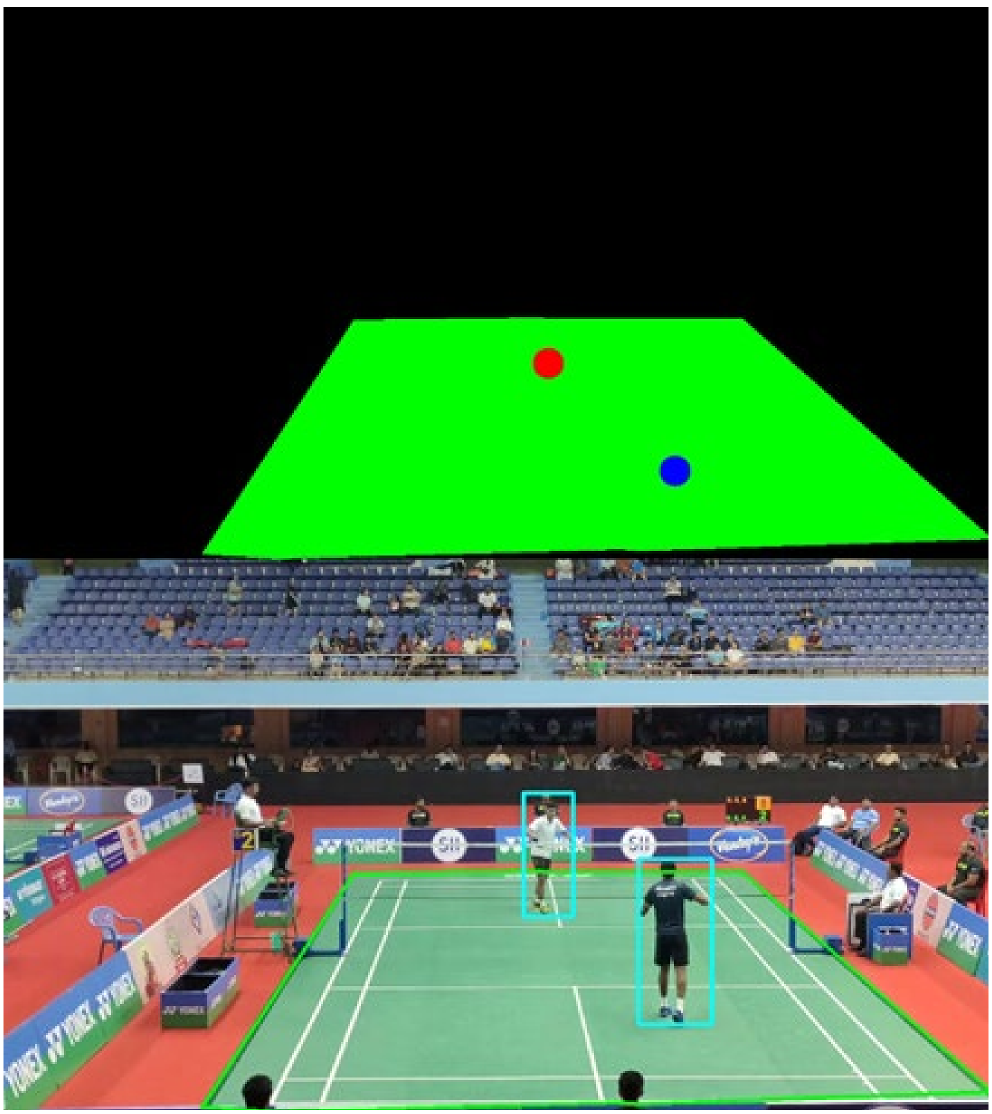

Figure 7: Example of Detected Player From Video Case 1

Average precision was evaluated to check the performance of the model. Average precision was obtained to decide what datasets model has the best performance where it is able to spot the badminton player correctly in the testing video.

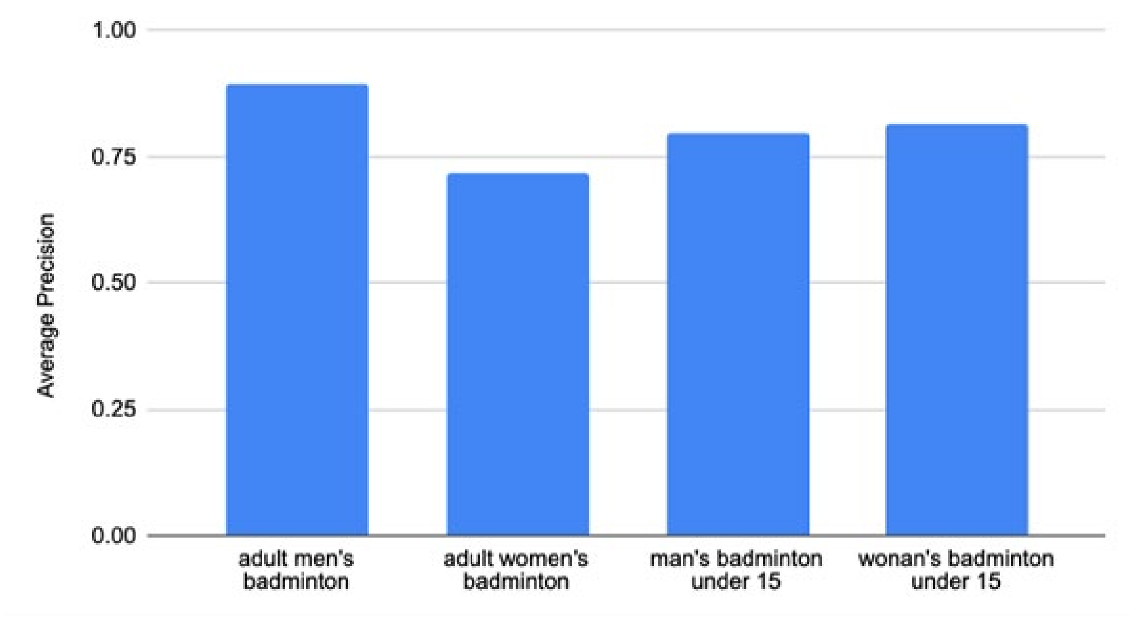

Figure 8: Average Precision vs Model

The performance of Yolo model result was shown as improved when tested on video set where less audiences are, as seen in Cases 1, 3, and 4, with the best performance achieved when tested on video with no clear judge figure (Case 1). This is because yolo model is confused at detecting players from the video when it has background figures similar to the human figures. When it has clear image of judges in the back, model would be confused if the judges are players. During testing, the model was able to accurately identify adult man match images with an accuracy rate of 89.33 %. However, when it came to junior man match images, the accuracy rate dropped to 79.52% and 81.48%. The model struggled to accurately identify matches with clear image of judges occurring simultaneously in the same frame.

This is an important consideration for real-world applications, where accurate recognition of both junior and adult images is necessary for effective video analysis and player tracking.

For adult women's match, one of the judge's figure and chair figure next to the judge confused A.I. detection by being recognized as a player often time. This makes the precision rate of adult woman's match to lower precision rate of 73.68% In addition, the Jaccard Similarity Index accuracy metric is commonly used to evaluate the performance of object detection algorithms, including badminton image detection. The Jaccard Similarity Index can find the overlapped areas between the predicted images and the ground truth images for each object in an image.

A higher Jaccard Similarity Index score indicates a better match between the predicted and ground truth bounding boxes, with a perfect score of 1 indicating complete overlap. Typically, an Jaccard Similarity Index threshold of 0.5 or higher is used to determine whether a detection is considered accurate.

In this study the Jaccard Similarity Index accuracy metric can be used to evaluate the performance of algorithms in detecting court in images or video frames. For the purpose, we used Jaccard similarity for the Jaccard Similarity Index accuracy metric. The Jaccard similarity is a measure of the similarity between two sets, and is commonly used in image processing to evaluate the Intersection over Union (Jaccard Similarity Index) metric. The Jaccard similarity is calculated as the ratio of the intersection of two sets to the union of the same sets. By using the Jaccard similarity to calculate the Jaccard Similarity Index, the accuracy of object detection algorithms can be evaluated more accurately, since it takes into account the proportion of the area covered by the predicted bounding box and ground truth bounding box. This is particularly important in cases where the bounding boxes are irregular in shape or size, since the Jaccard similarity allows for a more flexible measure of overlap.

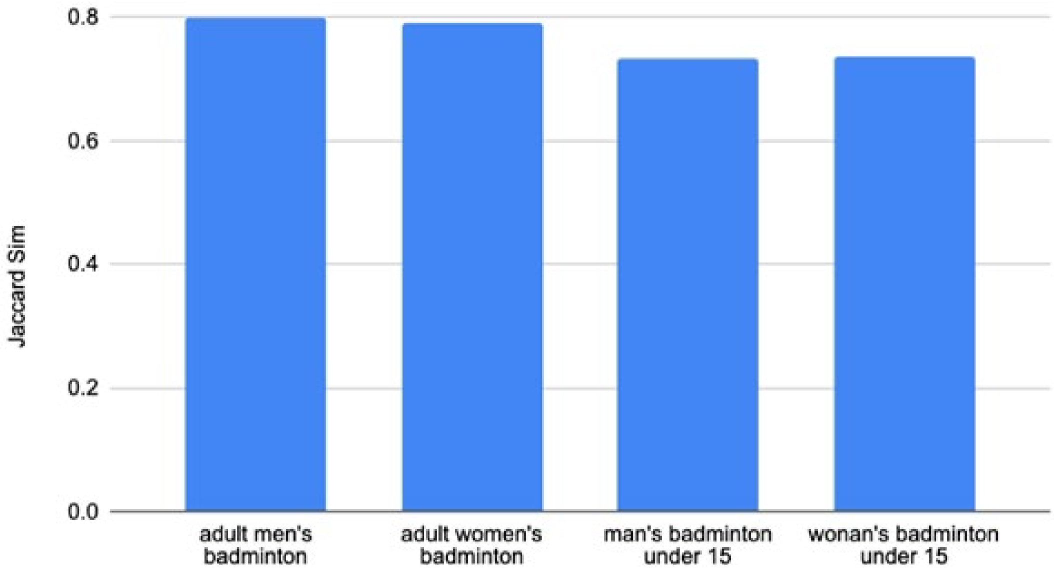

Figure 9: Jaccard Similarity Index vs. Model

Figure 9 illustrates the result of Jaccard Similarity Index showing that it performed better when tested on the clear badminton court than when tested on badminton court with audience in the video data as demonstrated in Cases 1 and 2.

For example, for men's badminton under 15, these images contained badminton match of more clear image of judges in the frame, making it confusing for the AI model to recognize and classify the image. As a result, the model achieved a lower jaccard similarity index rate of 73.44 % for the match recognition. By using separate sets of images for adult match recognition, the AI model was able to achieve high jaccard similarity index rates for both tasks. This is an important consideration for real-world applications, where accurate recognition of both junior and adult images is necessary for effective video analysis and player tracking. In summary of a more specific comparison, here are results from our AI Badminton Video Analyzer, focusing on Jaccard Similarity Index (Intersection over Union) and precision for image detection.

	Jaccard Similarity Index	Average Precision
adult men's badminton	0.7999	0.8933333333
adult women's badminton	0.7904551732	0.7154811715
man's badminton under 15	0.7344211456	0.7952380952
wonan's badminton under 15	0.7368214288	0.814814814

Table 2: summary of model performance

These result examples demonstrate that the AI Badminton Video Analyzer performed consistently well in both men's and women's matches in terms of Jaccard Similarity Index for image detection. It performed well in men's matches in terms of average precision. Women's matches in terms of average precision was relatively lower due to clear image of judge and his chair in the background. Overall, it successfully identified and localized players indicating its effectiveness in accurately detecting and classifying relevant objects in the analyzed video frames for both genders and both age groups.

5. Conclusion

The number of results obtained from this AI-powered badminton video analysis can vary depending on the specific research question or objective. However, some common performance metrics that can be analyzed using AI-powered video analysis in badminton include:

1. Player movement: The speed and trajectory of player movements, which can be analyzed to identify areas for improvement in footwork and positioning.
2. Strategy analysis: The analysis of gameplay patterns and strategies used by players and teams, which can inform coaching decisions and training strategies.

The insights gained from this analysis can be used to improve training and performance strategies, leading to better outcomes for players and teams

The use of YOLOv5 in badminton video analyzer in our work provides several benefits. Its lightweight architecture and fast processing speed make it ideal for real-time video analysis, allowing coaches and players to quickly analyze and improve their performance. With YOLOv5, badminton video analyzer can accurately detect and track players, shuttlecock, and other objects in the video frames, providing valuable insights into the players' movements and strategy. Additionally, the use of YOLOv5 can lead to the development of more sophisticated features in the badminton video analyzer, such as advanced tracking algorithms and personalized training recommendations. Overall, the incorporation of YOLOv5 in badminton video analyzer is a significant step forward in enhancing the accuracy and efficiency of video analysis in the sport of badminton.

In conclusion, AI-powered badminton video detection has the potential to greatly enhance gameplay analysis and training. Our approach showed that the AI-powered badminton video detection app can accurately detect and track players and shuttlecocks in real-time, providing valuable insights into player performance and strategy. This technology can be used by players and coaches to analyze gameplay and identify areas for improvement, as well as by broadcasters to enhance the viewing experience for fans. Additionally, this technology can be used in training programs to provide personalized feedback and guidance to players, helping them to improve their skills and achieve their goals. However, there are still challenges that must be addressed in the development and implementation of AI-powered badminton video detection. These include issues of accuracy, data privacy, and the need for robust and reliable infrastructure. Despite these challenges, the potential benefits of this technology make it an exciting area of development for the future of badminton. As the technology continues to evolve and improve, we can expect to see even more advanced gameplay analysis and training tools that will help players and coaches to achieve even greater success on the court.

References

1. Haq, M. A., Tarashima, S., & Tagawa, N. (2022, August). Heatmap visualization and badminton player detection using convolutional neural network. In 2022 International Electronics Symposium (IES) (pp. 627-631). IEEE.
2. Naik, B. T., & Hashmi, M. F. (2021). Ball and player detection & tracking in soccer videos using improved yolov3 model.
3. Kamble, P. R., Keskar, A. G., & Bhurchandi, K. M. (2019). A deep learning ball tracking system in soccer videos. Opto-Electronics Review, 27(1), 58-69.
4. Zhang, J., Jin, J., Ma, Y., & Ren, P. (2023). Lightweight object detection algorithm based on YOLOv5 for unmanned surface vehicles. Frontiers in Marine Science, 9, 1058401.
5. Agarwal, P., & Alam, M. (2020). A lightweight deep learning model for human activity recognition on edge devices. Procedia Computer Science, 167, 2364-2373.
6. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
7. Yu, H., Mineyev, I., Varshney, L. R., & Evans, J. A. (2022). Learning from One and Only One Shot. arXiv preprint arXiv:2201.08815.
8. Cuiping, C. (2021, January). Badminton video analysis based on player tracking and pose trajectory estimation. In 2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) (pp. 471-474). IEEE.
9. Ghosh, I., Ramamurthy, S. R., Chakma, A., & Roy, N. (2022). Decoach: Deep learning-based coaching for badminton player assessment. Pervasive and Mobile Computing, 83, 101608.
10. Cao, Z., Liao, T., Song, W., Chen, Z., & Li, C. (2021). Detecting the shuttlecock for a badminton robot: A YOLO based approach. Expert Systems with Applications, 164, 113833.
11. Monier, E., Wilhelm, P., & Rückert, U. (2009). A computer vision based tracking system for indoor team sports. In The fourth international conference on intelligent computing and information systems.
12. Schaefer, D., Asteroth, A., and Ludwig, M. (2015). “Training plan evolution based on training models,” in 2015 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA) (New York, NY: IEEE), 1–8. doi: 10.1109/INISTA.2015.7276739
13. Tan, Y. E., See, J., Abdullah, J., Wang, L. K., & Ban, K. W. (2022, November). A deep learning based framework for badminton rally outcome prediction. In 2022 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) (pp. 1-5). IEEE.
14. Yoshikawa, Y., Shishido, H., Suita, M., Kameda, Y., & Kitahara, I. (2021, March). Shot detection using skeleton position in badminton videos. In International Workshop on Advanced Imaging Technology (IWAIT) 2021 (Vol. 11766, pp. 312-317). SPIE.
15. Rahmad, N. A., As’ ari, M. A., Ibrahim, M. F., Sufri, N. A. J., & Rangasamy, K. (2020). Vision based automated badminton action recognition using the new local convolutional neural network extractor. In Enhancing Health and Sports Performance by Design: Proceedings of the 2019 Movement, Health & Exercise (MoHE) and International Sports Science Conference (ISSC) (pp. 290-298). Springer Singapore.
16. Zhang, J., Jin, J., Ma, Y., & Ren, P. (2023). Lightweight object detection algorithm based on YOLOv5 for unmanned surface vehicles. Frontiers in Marine Science, 9, 1058401.
17. Ghosh, I., Ramamurthy, S. R., Chakma, A., & Roy, N. (2022). Decoach: Deep learning-based coaching for badminton player assessment. Pervasive and Mobile Computing, 83, 101608.
18. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
19. Mingxing Tan , Quoc V. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, arXiv:1804.04861, pages 1-8
20. Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360.
21. Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile

devices. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6848-6856).
22. Khokhlov, I., Davydenko, E., Osokin, I., Ryakin, I., Babaev, A., Litvinenko, V., & Gorbachev, R. (2020, May). Tiny-YOLO object detection supplemented with geometrical data. In 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring) (pp. 1-5). IEEE.
23. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
24. Horvat, M., & Gledec, G. (2022). A comparative study of YOLOv5 models performance for image localization and classification. In Central European Conference on Information and Intelligent Systems (pp. 349-356). Faculty of Organization and Informatics Varazdin.
25. Ieamsaard, J., Charoensook, S. N., & Yammen, S. (2021, March). Deep learning-based face mask detection using yolov5. In 2021 9th International Electrical Engineering Congress (iEECON) (pp. 428-431). IEEE.
26. Nelson, J., & Solawetz, J. (2020). Responding to the Controversy about YOLOv5. Roboflow: Roboflow news.
27. Sharma, V. (2020). Face mask detection using yolov5 for COVID-19 (Doctoral dissertation, California State University San Marcos).
28. Sukkar, M., Kumar, D., & Sindha, J. (2021, July). Real-time pedestrians detection by YOLOv5. In 2021 12th international conference on computing communication and networking technologies (ICCCNT) (pp. 01-06). IEEE.
29. Nelson, J., & Solawetz, J. (2020). Responding to the Controversy about YOLOv5. Roboflow: Roboflow news.
30. Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo algorithm developments. Procedia Computer Science, 199, 1066-1073.
31. Xu, Y., Sun, F., & Wang, L. (2022). YOLOv5-PD: A Model for Common Asphalt Pavement Defects Detection. Journal of Sensors, 2022.
32. Yang, G., Feng, W., Jin, J., Lei, Q., Li, X., Gui, G., & Wang, W. (2020, December). Face mask recognition system with YOLOV5 based on image recognition. In 2020 IEEE 6th International Conference on Computer and Communications (ICCC) (pp. 1398-1404). IEEE.
33. Karthi, M., Muthulakshmi, V., Priscilla, R., Praveen, P., & Vanisri, K. (2021, September). Evolution of yolo-v5 algorithm for object detection: automated detection of library books and performace validation of dataset. In 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES) (pp. 1-6). IEEE.