RacketDB: A Comprehensive Dataset for Badminton Racket Detection

Muhammad Abdul Haq[1,3][a], Shuhei Tarashima[2][b] and Norio Tagawa[1][c]

[1]*Faculty of Systems Design, Tokyo Metropolitan University, Tokyo, Japan*

[2]*Innovation Center, NTT Communications Corporation, Tokyo, Japan*

[3]*Department of Information Technology, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia*

Keywords: Badminton Racket Dataset, Annotated Sports Dataset, Object Detection Dataset.

Abstract: In this paper, we present RacketDB, a specialized dataset designed to address the challenges of detecting badminton rackets in images. This task often hindered by the lack of dedicated datasets. Existing general-purpose datasets fail to capture the unique characteristics of badminton rackets. RacketDB includes 16,608 training images, 3,175 testing images, and 2,899 validation images, all meticulously annotated to enhance object detection performance for sports analytics. To evaluate the effectiveness of RacketDB, we utilized several established object detection models, including YOLOv5, YOLOv8, DETR, and Faster R-CNN. These models were assessed based on metrics like mean average precision (mAP), precision, recall, and F1. Our results demonstrate that RacketDB significantly improves detection accuracy compared to general datasets, highlighting its potential as a valuable resource for developing advanced sports analytics tools. This paper provides a detailed description of RacketDB, the evaluation process, and insights into its application in enhancing automated detection in badminton. The dataset is available at https://github.com/muhabdulhaq/racketdb.

1 INTRODUCTION

The advancement of computer vision technologies has significantly impacted various fields, including sports analytics, where accurate object detection plays a crucial role in performance analysis, coaching, and automated game monitoring. However, the effectiveness of these technologies is highly dependent on the availability of high-quality datasets that capture the specific characteristics of target objects. Detecting rackets in badminton is crucial to evaluate player performance, refining coaching methods, and boosting viewer experience with automated highlights. Despite this need, there is a notable gap in existing datasets that specifically cater to badminton rackets, posing challenges for researchers and developers working on related applications.

General purpose datasets such as COCO (Lin et al., 2014) and PASCAL VOC (Everingham et al., 2010) have contributed significantly to progress in object detection. However, they are inadequate for specialized areas such as the detection of sports equipment. These datasets do not adequately represent the unique attributes of badminton rackets. The absence of a specialized dataset hampers the development of robust models that can reliably detect rackets under diverse conditions encountered in real-world scenarios, such as varying lighting, cluttered backgrounds, and dynamic player movements.

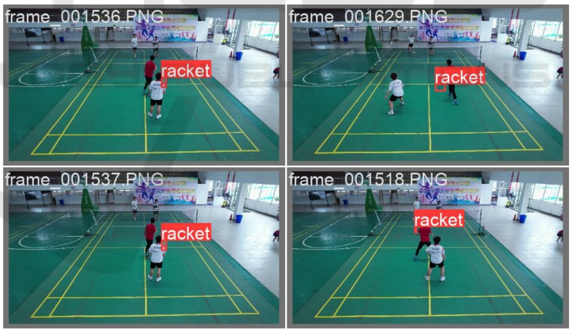

Figure 1: Examples of racket detection across multiple frames in the RacketDB dataset. The bounding boxes highlight the position of the badminton rackets.

RacketDB is introduced as a response to this challenge, offering a comprehensive dataset specifically designed for badminton racket detection. It includes 22,682 images, with detailed annotations that capture rackets in various movements. This dataset aims to

[a] https://orcid.org/0000-0002-2539-4179
[b] https://orcid.org/0009-0007-6022-2560
[c] https://orcid.org/0000-0003-0212-9265

Haq, M. A., Tarashima, S. and Tagawa, N.
RacketDB: A Comprehensive Dataset for Badminton Racket Detection.
DOI: 10.5220/0013159700003912

In *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2025) - Volume 3: VISAPP*, pages 426-433
ISBN: 978-989-758-728-3; ISSN: 2184-4321

bridge the gap between general-purpose object detection datasets and the specific needs of sports equipment detection, providing a valuable resource to the research community.

To evaluate the effectiveness of RacketDB, we employed several established object detection models, including YOLOv5 (Jocher et al., 2020), YOLOv8 (Jocher et al., 2023), DETR (Carion et al., 2020), Faster R-CNN with ResNet50 (He et al., 2016) and ResNet101 (He et al., 2016). These models represent a range of approaches, from traditional convolutional neural networks to modern transformer-based architectures, allowing us to benchmark the dataset across diverse methodologies. Our evaluation focuses on assessing the performance of these models in terms of mean average precision (mAP), precision, recall, and F1, providing insight into the strengths and limitations of RacketDB.

In our evaluation of the RacketDB test dataset using the RacketDB model with YOLOv5, we achieved a mAP_{50} score of 0.77. Surpassing YOLOv5 trained on the COCO dataset, which achieves a mAP_{50} score 0.63 for tennis racket detection. This performance highlights the value of RacketDB in the specific task of badminton racket detection and demonstrates that specialized datasets can significantly enhance object detection models. Figure 1 illustrates examples of racket detection in the RacketDB dataset, highlighting the challenges of detecting badminton rackets across multiple frames. We evaluate RacketDB using YOLOv5, YOLOv8, DETR, and Faster R-CNN as backbone architectures, demonstrating the applicability of the RacketDB dataset in this task. Furthermore, the dataset's versatility extends beyond object detection, with potential applications in activity recognition and game strategy analysis, making it a valuable resource for sports analytics.

The contributions of this paper are threefold: (1) we introduce RacketDB, a specialized dataset for badminton racket detection, addressing a critical gap in sports analysis; (2) we evaluate the dataset using state-of-the-art object detection models, demonstrating its impact on performance; and (3) we provide a detailed analysis of the evaluation results, offering insights for future research. By making RacketDB publicly accessible, we aim to drive further advancements in sports technology, enabling more accurate and efficient detection models for badminton and potentially other sports.

2 RELATED WORK

Object detection has become an essential tool for improving performance assessment, strategy development, and automated game monitoring in sports analysis. Large-scale datasets such as COCO (Lin et al., 2014), PASCAL VOC (Everingham et al., 2010), and ImageNet (Deng et al., 2009) have driven advancements in object detection, leading to the development of state-of-the-art models such as YOLO (Redmon et al., 2016), Faster R-CNN (Ren et al., 2015), SSD (Liu et al., 2016), and DETR (Liu et al., 2016). These datasets and models have shown effectiveness across a range of applications, including sports analytics; however, they primarily cater to general-purpose object detection tasks and lack the specificity required for niche applications such as equipment detection in badminton.

General-Purpose Object Detection Datasets. COCO (Lin et al., 2014) and PASCAL VOC (Everingham et al., 2010) have set benchmarks for object detection by offering diverse categories and a large volume of labeled images. These datasets have enabled the development of robust detection models; however, their broad focus means they do not address the specific needs of sports equipment detection. Meanwhile, COCO has the "tennis racket" label and it can be applied to badminton racket detection, but its performance is poor (as demonstrated in section 5). The datasets are designed to handle a wide array of object categories, but they fall short in scenarios requiring high precision and specialized annotations for sports equipment like badminton rackets.

Sports-Specific Object Detection Datasets. Existing datasets for sports generally emphasize player detection and tracking rather than equipment detection. For instance, TennisNet (Faulkner et al., 2020) and FootballDB (Team, 2023) focus on tennis rackets and soccer-related objects, providing valuable insights into equipment detection in those sports. These datasets have demonstrated the benefits of specialized data for improving detection accuracy, highlighting the importance of domain-specific resources. However, no equivalent dataset has been developed for badminton rackets, leaving a significant gap in the field of racket sports analysis.

Research in Badminton Sport Analysis. Traditional research in badminton has concentrated on player detection, pose estimation in badminton match (Ding et al., 2024). Studies employing some neural network models have focused on shuttlecock tracking (Haq et al., 2024)(Sun et al., 2020)(Tarashima et al., 2023), offering valuable insights for coaching and performance enhancement. Despite this progress,

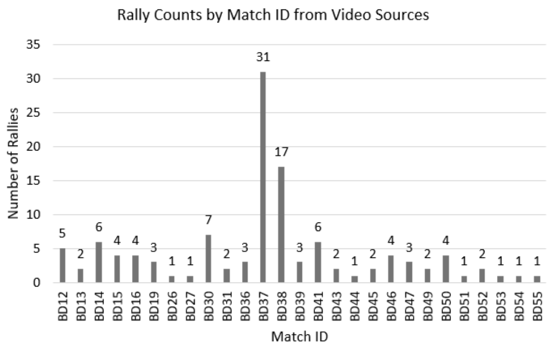

Figure 2: Distribution of rally counts by match ID from video sources.

there has been minimal attention to detecting and analyzing badminton rackets, which are crucial for a comprehensive understanding of the sport. Moreover, a tool like VIRD (Chu et al., 2022) that offers an immersive analysis of a badminton match has highlighted this gap. A coach using the system specifically recommended including racket data to enhance the accuracy of visualizations and provide deeper insights into shot mechanics. This underscores the need for a specialized dataset like RacketDB, which focuses on racket detection to complement existing models and improve the granularity of badminton performance analysis.

Limitations of Existing Datasets. The lack of specialized datasets for badminton rackets is evident when considering the broader context of sports equipment detection. Datasets like TinyImage (Torralba et al., 2008) have been used for general object recognition tasks but are less suited for sports applications due to high noise levels and lack of detail.

RacketDB's Contribution. RacketDB addresses these gaps by providing the first comprehensive dataset specifically designed for badminton racket detection. It includes annotations of capturing rackets in various environments of 22,682 images. By evaluating established object detection models such as YOLOv5, YOLOv8, DETR and Faster R-CNN, RacketDB allows for benchmarking the performance of different approaches across diverse methodologies, from traditional convolutional neural networks to transformer-based architectures. Our evaluation demonstrates that the object detection model performs effectively in detecting badminton rackets, highlighting the advantages of using the RacketDB dataset, and highlighting its potential as a valuable resource for advancing sports equipment detection in badminton. RacketDB not only fills a critical gap in badminton research but also sets the groundwork for comprehensive analysis tools that encompass some aspects of the sport.

Figure 3: The annotation process for the RacketDB dataset. Each racket is manually labeled with a bounding box to accurately capture its position and size.

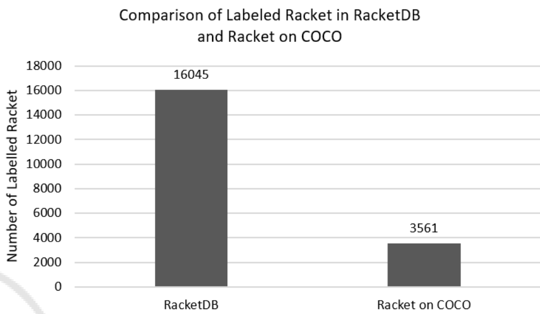

Figure 4: Comparison of labeled racket images between RacketDB and racket on COCO dataset.

3 DATASET DESCRIPTION

RacketDB is a specialized dataset developed to address the unique challenges of detecting badminton rackets in real-world scenarios. This section provides an overview of the data collection process, annotation details, dataset composition, and the format used to ensure RacketDB's applicability for training and evaluating object detection models.

3.1 Data Collection

The dataset used in this work is derived from videos originally collected for a study conducted on 2-vs-2 men's doubles badminton games among members of a college badminton club. The original data collection was approved by the Ethics Committee of Anhui Normal University (approval number [AHNU-ET2022042]) on April 14, 2022, and was conducted according to the principles of the Declaration of Helsinki, with all participants providing their signed informed consent (Ding et al., 2024).

The videos were captured using two DJI Air 2S drones (Da-Jiang Innovations Science and Technology Co., Ltd., China) to provide back views of the badminton court. The video resolution was 4K (3,840 × 2,160 pixels), and the frame rate was 30 fps. We ue

Table 1: Number of bounding boxes and rallies for training, validation, and test splits.

Split	Bounding Boxes	Rallies
Train	12,343	83
Validation	1,923	18
Test	1,779	18
Total	16,045	119

the video that conclude 27 games, resulting in a total of 119 rallies. Figure 2 presents the counts of rallies categorized by match IDs extracted from various video sources. This data provides insights into the frequency of rallies across different matches, which can be crucial for analyzing gameplay patterns and performance metrics in badminton.

For our purposes, we processed the original videos by splitting them into individual frames, which now serve as the primary data source for our analysis and model training.

3.2 Annotation Process

The RacketDB dataset was manually annotated using the Computer Vision Annotation Tool (CVAT) (Sekachev et al., 2020). The annotation process involved several key steps to ensure the accuracy and consistency of the dataset. CVAT was selected for its user-friendly interface and robust features suitable for bounding box annotations. Figure 3 illustrates how each image in the dataset was annotated with bounding boxes around badminton rackets, all labeled with the tag ”racket” to maintain standardization and clarity in object detection.

The annotations were performed manually by the author. This involved carefully drawing bounding boxes around each visible racket, ensuring that the boxes were as precise and accurate as possible. The author rechecked all annotations to confirm their accuracy and consistency. This review included verifying that the bounding boxes properly enclosed the rackets and making adjustments where necessary.

During the annotation process, there were many images where the racket was not visible, often due to the player’s position, especially during pre-defense movements when the racket was facing away from the camera. Additionally, short rallies led to a high frequency of serve poses where the racket was not fully visible in front of the camera. Consequently, the number of bounding boxes is lower than the total number of images. Initially, we attempted to annotate all visible rackets, including those partially obscured behind the net, but this approach degraded detection performance and produced poor results. To improve accuracy, we removed annotations for rackets behind the net, which enhanced the model’s performance by focusing on clearly visible racket instances and avoiding the negative impact of ambiguous or hidden objects.

By manually annotating the dataset and implementing a detailed quality control process, the RacketDB dataset offers high-quality annotations essential for evaluation of object detection models.

3.3 Dataset Composition

The RacketDB dataset is organized into three subsets to support comprehensive training and evaluation of object detection models: training, validation, and test sets. The dataset is split into 70% for training, 15% for validation, and 15% for testing. The RacketDB has total 16,045 labeled rackets. The training set has 12,343 bounding boxes in 83 rallies. The validation set comprises 1,923 bounding boxes in 18 rallies, while the test set contains 1,779 bounding boxes in 18 rallies. These splits ensure a balanced distribution for model training and evaluation, providing comprehensive coverage across various scenarios for accurate racket detection, as detailed in Table 1.

Compared to COCO, which includes 3,561 labeled instances of rackets, RacketDB is significantly larger and more specialized, containing annotated instances of badminton rackets (See Figure 4). This size advantage is crucial for improving detection accuracy, as it allows for better model training and performance in racket detection tasks. The diversity and scale of RacketDB provide a robust foundation for object detection specific to badminton.

In bounding box dimensions, the average width and height in the training set are 16.3 ± 7.8 pixels and 29.0 ± 12.3 pixels, respectively. In the validation set, the average width is 17.3 ± 8.7 pixels, while the average height is 27.5 ± 10.5 pixels. The total area of bounding boxes, reflecting the detected racket sizes, is 486.3 ± 343.8 pixels^2 in the training set and 493.1 ± 343.4 pixels^2 in the validation set. These averages and standard deviations, presented in Table 2, highlight the consistency in racket size and the variability across the dataset splits. The slightly larger width standard deviation in the validation set and the consistent area standard deviations across both splits ensure that the dataset represents a diverse yet balanced range of bounding box sizes for effective training and evaluation.

Lastly, the ratio of bounding box dimensions to the frame size remains relatively stable across both training and validation sets. In the training set, the bounding box width occupies approximately $2.5 \pm 1.2\%$ of the frame, and the height accounts for $4.5 \pm 1.9\%$, with the total area covering around $0.1 \pm 0.1\%$

Table 2: Bounding box size metrics and size ratios to frame size with averages and standard deviations for training and validation splits.

Metric	Train	Validation
Width (pixels)	16.3 ± 7.8	17.3 ± 8.7
Height (pixels)	29.0 ± 12.3	27.5 ± 10.5
Area (pixels^2)	486.3 ± 343.8	493.1 ± 343.4
Width Ratio (%)	2.5 ± 1.2	2.7 ± 1.4
Height Ratio (%)	4.5 ± 1.9	4.3 ± 1.6
Area Ratio (%)	0.1 ± 0.1	0.1 ± 0.1

of the frame. Similarly, in the validation set, the width ratio is $2.7 \pm 1.4\%$, the height ratio is $4.3 \pm 1.6\%$, and the area ratio is $0.1 \pm 0.1\%$. These standard deviations reflect the natural variation in racket sizes and positions within the frames. Overall, the ratios, summarized in Table 2, indicate that racket size, in relation to the overall frame, remains consistent across the dataset splits, ensuring uniformity and reliability in the bounding box annotations.

3.4 Data Format and Accessibility

The RacketDB dataset is available in multiple widely-used annotation formats to ensure broad compatibility and ease of use with various object detection frameworks and tools. The available formats include:

- **COCO 1.0:** A structured and versatile format commonly used in object detection, supporting a range of deep learning frameworks (Lin et al., 2014).
- **CVAT for images 1.1:** Compatible with the Computer Vision Annotation Tool (CVAT), which was used during the annotation process, allowing easy management and editing of annotations.
- **Datumaro 1.0:** A flexible format that facilitates conversions between different dataset types, aiding in data handling and manipulation tasks.
- **Open Images V6 1.0:** Suitable for those using the Open Images dataset format, providing structured annotations for object detection (Kuznetsova et al., 2020).
- **PASCAL VOC 1.1:** A traditional format used in the PASCAL VOC challenges, widely adopted for object detection and segmentation tasks.
- **YOLO 1.1:** Designed for YOLO (You Only Look Once) models, this format is optimized for use with YOLO-based detection pipelines (Redmon, 2016).
- **YOLOv8 Detection 1.0:** A format specific to the YOLOv8 models, ensuring compatibility with the latest version of the YOLO object detection series (Jocher et al., 2023).

RacketDB provides formats that facilitate easy integration into research workflows across various platforms and applications.

4 EVALUATION METHODOLOGY

To assess the effectiveness of RacketDB, we evaluated several established object detection models, including YOLOv5, YOLOv8, DETR and Faster R-CNN. These models were chosen for their diverse architectural approaches, ranging from convolutional neural networks (CNN) to transformer-based frameworks, providing a comprehensive evaluation across different detection paradigms.

4.1 Model Selection

The models selected for evaluation represent a broad spectrum of object detection architectures:

4.1.1 YOLOv5 and YOLOv8

These models are part of the YOLO (You Only Look Once) family, known for their balance of speed and accuracy in real-time object detection. YOLOv5 utilizes a CNN-based architecture with optimized feature extraction and detection heads (Jocher et al., 2020), while YOLOv8 introduces further refinements in feature fusion and model efficiency (Jocher et al., 2023).

4.1.2 DETR (DEtection TRansformers)

DETR leverages a transformer-based architecture with self-attention mechanisms to directly predict object bounding boxes and class labels. This model departs from traditional CNN approaches by utilizing a set-based prediction, which simplifies the detection pipeline and offers a novel approach to object detection tasks (Carion et al., 2020).

4.1.3 Faster R-CNN

Faster R-CNN are region-based convolutional neural network models widely used for object detection tasks. We use Faster R-CNN with ResNet50 and ResNet101 as backbone networks. ResNet50 and ResNet101 differ in depth, with 50 and 101 layers, respectively, providing insights into the impact of network depth on detection performance (Ren et al., 2015). We use Faster R-CNN with ResNet50, referred to as FRCNN (R50), and Faster R-CNN with ResNet101, referred to as FRCNN (R101).

(a) Example detection result using COCO pretrained model

(b) Example detection result using RacketDB pretrained model

Figure 5: Comparison between detection result using YOLOv5 on COCO pretrained model (a) and RacketDB model(b). Because there is no badminton racket in COCO, we use tennis racket detection when using COCO pretrained model.

Table 3: Hyperparameters used for training all models on the RacketDB dataset.

Hyperparameter	Value
Epochs	100
Batch size	256
Image size	640
Optimizer	Adam
Learning rate (initial)	0.01
Momentum	0.937
Weight decay	0.0005
Augmentation (AutoAugment)	randaugment
Flipping (Horizontal)	true (0.5)

4.2 Training Setup

All models were trained using the same hyperparameters (as shown in Table 3) to ensure a fair evaluation across different architectures. The training was conducted for 100 epochs for each model, with identical configurations applied uniformly to maintain consistency in the evaluation process.

The training environment was set up with Python 3.10.13 and PyTorch 2.0.1+cu117, leveraging CUDA for GPU acceleration. The hardware used for training included two NVIDIA RTX 6000 Ada Generation GPUs, providing substantial computational power for handling the large-scale training tasks. The processing was further supported by an Intel® Xeon® Gold 6326 CPU @ 2.90GHz with 32 cores. The operating system used was Ubuntu 22.04.4 LTS.

All models were trained with the same set of parameters, ensuring that the evaluation focused solely on the differences in model architectures rather than variations in training conditions.

5 RESULT AND DISCUSSION

In this section, we present the evaluation results for the various object detection models applied to the RacketDB dataset. We analyze how effectively each model learned to identify and classify badminton rackets, revealing their strengths and weaknesses in practical scenarios. Additionally, we assess the mean Average Precision (mAP) scores to provide a comprehensive measure of the models' accuracy and reliability in detecting rackets. This evaluation underscores the effectiveness of RacketDB as a specialized dataset for enhancing sports equipment detection in badminton.

5.1 Detection Performance

The performance such as Precision, Recall, and F1 Score (Table 4). YOLOv5 achieves a high Precision of 0.90 and Recall of 0.72, with an F1 Score of 0.80, indicating strong, balanced detection capabilities. YOLOv8 shows similar performance, with a slightly lower Precision of 0.89 but the same Recall and F1 Score. DETR balances Precision (0.57) and Recall (0.79) moderately, with an F1 Score of 0.66. In terms of speed, ResNet50 leads with 83 FPS, followed by YOLOv5 (76 FPS) and YOLOv8 (68 FPS), while DETR prioritizes accuracy at only 20 FPS. Faster R-CNN with ResNet50 achieving a Recall of 0.85 but a lower Precision of 0.59, resulting in F1 Score of 0.70, while Faster R-CNN with ResNet101 shows slightly lower Recall (0.84) and Precision (0.53), with F1 Score of 0.65.

Table 4: Precision, Recall, F1, and FPS of different models on RacketDB.

Model	Precision	Recall	F1	FPS
YOLOv5	0.90	0.72	0.80	76
YOLOv8	0.89	0.72	0.80	68
DETR	0.57	0.79	0.66	20
FRCNN (R50)	0.59	0.85	0.70	83
FRCNN (R101)	0.53	0.84	0.65	55

Table 5: mAP Metrics of different models on RacketDB.

Model	mAP_{50}	$mAP_{50:95}$
YOLOv5	0.77	0.48
YOLOv8	0.78	0.47
DETR	0.70	0.35
FRCNN (R50)	0.78	0.47
FRCNN (R101)	0.77	0.43

While the performance metrics show promising results, various challenges contribute to detection failures in the RacketDB dataset. Some typical cases of detection failure include:

Occlusion. Detectors often fail when rackets are partially hidden by players, other rackets, or objects in the environment. For example, when players engage in rallies, the action may block the view of the racket.

Lighting Variability. Changes in lighting conditions such as glare, shadows, or dim lighting can affect how the racket is perceived. For instance, a brightly lit court may create reflections that obscure the racket's shape.

Motion Blur. Fast-paced actions in badminton can result in motion blur, making it difficult for detectors to accurately identify and localize the racket.

5.2 mAP Performance Analysis

The mean Average Precision (mAP) metric evaluates a model's effectiveness in object detection. Table 5 summarizes these metrics for models tested on RacketDB. YOLOv5 achieves mAP_{50} and $mAP_{50\text{-}95}5$ scores of 0.77 and 0.48, respectively, showing strong detection precision. YOLOv8 performs similarly, with an mAP_{50} of 0.78 and $mAP_{50\text{-}95}$ of 0.47. DETR achieves moderate scores of 0.70 (mAP_{50}) and 0.35 ($mAP_{50\text{-}95}$), indicating challenges in precise localization. Faster R-CNN with ResNet50 show comparable mAP_{50} scores 0.78 but slightly lower $mAP_{50\text{-}95}$ scores 0.47, when Faster R-CNN with ResNet101 show mAP_{50} scores 0.77 and $mAP_{50\text{-}95}$ scores 0.43, reflecting variability in handling complex detections. Notably, YOLOv5 trained on RacketDB achieves an mAP_{50} of 0.77, significantly outperforming YOLOv5 trained with COCO, which achieves only 0.63 for tennis racket detection, as shown in Table 6. This emphasizes the importance of specialized datasets like RacketDB for badminton racket detection.

Figure 6: Detection results on the TrackNet dataset, showcasing racket detection in videos with different backgrounds and environments.

Table 6: Performance of YOLOv5 on the COCO dataset

Model	mAP_{50} (COCO)	$mAP_{50\text{-}95}$ (COCO)
YOLOv5	0.63	0.39

We also evaluated our model on videos featuring different backgrounds and environments, such as those found in the TrackNet dataset (Sun et al., 2020). This dataset primarily consists of professional badminton singles matches annotated with shuttlecock locations. As shown in Figure 6, our model successfully detects rackets in these videos, demonstrating the generalizability of the model trained on RacketDB.

6 CONCLUSION & FUTURE WORK

This paper introduces RacketDB, a specialized dataset for badminton racket detection, evaluated using object detection models such as YOLOv5, YOLOv8, DETR and Faster R-CNN. Results demonstrate that RacketDB enhances detection accuracy and reliability, with YOLOv5 achieving the best Precision and F1 Score, followed closely by YOLOv8. Faster R-CNN showed high Recall but struggled with false positives, while DETR performed reasonably but lagged in efficiency.

RacketDB's strong mAP_{50} highlight its value for sports analytics and object detection research. It addresses challenges in racket detection and provides a solid foundation for future advancements. Planned enhancements include support for rotated bounding boxes, specialized neural network architectures, and expansion into broader sports applications like activity recognition and game strategy analysis. These developments aim to enable deeper insights into player performance and tactics.

REFERENCES

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). End-to-end object detection with transformers. *European conference on computer vision*, pages 213–229.

Chu, X., Xie, X., Ye, S., Lu, H., Xiao, H., Yuan, Z., Chen, Z., Zhang, H., and Wu, Y. (2022). Tivee: Visual exploration and explanation of badminton tactics in immersive visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 28:118–128.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255.

Ding, N., Takeda, K., Jin, W., Bei, Y., and Fujii, K. (2024). Estimation of control area in badminton doubles with pose information from top and back view drone videos. *Multimedia Tools and Applications*, 83:24777–24793.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338.

Faulkner, H. et al. (2020). Tenniset: A dataset for dense fine-grained event recognition, localisation and description. *arXiv preprint arXiv:2006.14236*.

Haq, M. A., Tarashima, S., and Tagawa, N. (2024). Shuttlecock detection using residual learning in u-net architecture. *JOIV : International Journal on Informatics Visualization*.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Jocher, G. et al. (2020). Yolov5: An improved version of yolov4. *arXiv preprint arXiv:2006.14236*.

Jocher, G. et al. (2023). Yolov8: A state-of-the-art object detection model. *arXiv preprint arXiv:2301.00503*.

Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., Duerig, T., and Ferrari, V. (2020). The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *International Journal of Computer Vision*, 128:1956–1981.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. *European conference on computer vision*, pages 740–755.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. *European conference on computer vision*, pages 21–37.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.

Redmon, J. S. D. R. G. A. F. (2016). (yolo) you only look once. *Cvpr*, 2016-December:779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

Sekachev, B., Manovich, N., Zhiltsov, M., Zhavoronkov, A., Kalinin, D., Hoff, B., TOsmanov, Kruchinin, D., Zankevich, A., DmitriySidnev, Markelov, M., Johannes222, Chenuet, M., a andre, telenachos, Melnikov, A., Kim, J., Ilouz, L., Glazov, N., Priya4607, Tehrani, R., Jeong, S., Skubriev, V., Yonekura, S., vugia truong, zliang7, lizhming, and Truong, T. (2020). opencv/cvat: v1.1.0.

Sun, N. E., Lin, Y. C., Chuang, S. P., Hsu, T. H., Yu, D. R., Chung, H. Y., and Ik, T. U. (2020). Tracknetv2: Efficient shuttlecock tracking network. *Proceedings - 2020 International Conference on Pervasive Artificial Intelligence, ICPAI 2020*, pages 86–91.

Tarashima, S., Haq, M. A., Wang, Y., and Tagawa, N. (2023). Widely applicable strong baseline for sports ball detection and tracking. In *2023 British Machine Vision Conference*.

Team, F. (2023). Footballdb: A comprehensive database for soccer analytics. *Journal of Sports Analytics*.

Torralba, A., Fergus, R., and Freeman, W. T. (2008). 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 30(11):1958–1970.