Computer Umpires: A Comparative Analysis of YOLOv9 and YOLOv8 for Real-Time Ball Tracking in Table Tennis

Jingxiang Jia *

Beijing No. 101 High School, Beijing 100091, China

* Corresponding Author Email: david.jjx@outlook.com

Abstract. Real-time precise ball tracking in table tennis is of crucial importance to automatic umpiring and performance analysis, and yet remains technically challenging due to the smallness, fastness and blurriness of the moving ball. Despite improvements in existing systems, a trade-off between high precision and real time application is enigmatic, with professional tournaments still relying on human umpires and a restrictive player challenge mechanism. In this work, we examine the performance of the state-of-the-art YOLOv9 object detector for real-time table tennis ball tracking, in comparison with its predecessor YOLOv8. This work is intended to explore the compromise between precision and recall, and to make a clear recommendation for model choice in different sports analytics. The YOLOv9-C and YOLOv8 models were trained and tested using a publicly available table tennis image dataset. Both the models are trained in the same way (30 epochs; input size 640 × 640), so that the comparison is fair. Standard object detection measurements including Precision, Recall and mAP@0.5). The experiment results show the greatly improved precision (0.880 vs 0.743) and mAP@0.5 (0.613 vs 0.579) with respect to YOLOv8. On the other hand, YOLOv8 had higher recall value of 0.604 compared to 0.508 for YOLOv3, suggesting a higher number of true ball instances could have been found, however also caused a higher number of false positives. Results show that YOLOv9, with higher prediction accuracy than Fast-YOLO, is more suitable for use with low false positive tolerance systems, like automatic umpiring and rule verification. YOLOv8, on the other hand, with its higher recall should be considered for applications such as post-game trajectory analysis, which requires complete event extraction. These findings contribute significantly to a broader utilization of deep learning in high-speed sports and suggest potential future work such as incorporating temporal data to improve tracking robustness.

Keywords: Object Detection, Table Tennis, YOLOv9, Ball Tracking.

1. Introduction

Table tennis, a sport with long traditions in Europe and Asia, has rapidly grown in recent decades. Since table tennis became an Olympic sport in Seoul 1988, its influence has expanded globally, and we have seen world-class pro players emerge from North American, Latin-American, and African countries. However, with the expansion and formalization also calls for rigorous rules and umpiring. With its rapid pace and intricate movements, table tennis presents a unique challenge in the field of sports technology [1]. The main difficulties are the size (40mm diameter) and speed of the ball, which can reach up to 35 m/s in professional rallies, making it the sport where the ball moves fastest relative to the playing surface [2]. Other difficulties such as blocking of the ball and light reflections also make the ball difficult to detect, hampering tracking of the ball during play. Traditionally, high-speed cameras and complex tracking algorithms have been employed to address these challenges, but this method requires professional equipment and significant computing power to implement in real-time. As a result, tournaments still rely on human umpires, with technology used as assistance: the umpire decides if there was a rule violation, and players have two chances per match to challenge the umpire's decision and request the use of computer systems to analyze the recording of the point, a system that could be improved by fully automated, real-time tracking.

Recent progress in deep learning, especially object detection methods along the lines of YOLO (You Only Look Once), have shown promising results for real-time applications in sports. Recent work has confirmed that the latest YOLO models are also robust in the context of complex, multi-entity sports such as soccer, however, it remains unclear if this holds for table tennis. In contrast to

soccer, table tennis is played with an object much smaller than a ball, and while there may be superficial analogies to a game such as tennis, important differences in ball size, normal camera angles and the spatiotemporal nature of rallies require that we specifically examine what is occurring. Previous works of playing Table Tennis used mostly traditional machine learning, based on hand-crafted features, which don't generalize well across different lighting conditions and occlusions as well as modern deep learning detector do. This gap underscores the importance of studying implication of the latest object detectors to the specific challenges of table tennis.

The most recent version, YOLOv9, has good performance while fast and thus, it seems to be a good candidate to track table tennis balls [3]. The objective of this study is to perform a direct comparison of YOLOv9 and its reference YOLOv8 [4], with the aim of understanding the robustness of each methodology on a challenging domain. The main thrust will be to determine if the new developments in YOLOv9 leads to noticeably improved real-time detection performance when applied to table tennis-specific contexts in this period. Through systematically evaluation these works, this paper aims at promoting an accessible and efficient approach to table tennis ball tracking as well as other real-time sports tracking applications. Eventually, this might lead to better and more reliable tracking systems, opening up the high-level analysis to professional and amateur sports.

2. Literature Review

Tracking and detection of moving table tennis balls in video sequences on the fly has become an important area of research in sports analytics, since it is fundamentally useful for many applications such as umpiring, performance analysis, and tactical studies. However, finding table tennis balls is a challenging task, due to their small size, fast speed, quick brake, and a lot of distracting visual clutter, such as occlusion and motion blur. There has been extensive research in the recent past with respect to such problems on geometric analysis, traditional computer vision algorithms, as well as machine learning algorithms [5]. Nevertheless, achieving robust detection in the real condition, especially in the dynamic environment of the table tennis game, is still a challenging task.

2.1. Validation of YOLOv9 in Soccer

Markappa et al. recently conducted a study using the publicly available SoccerNet dataset to compare various YOLO models, viz., YOLOv3, YOLOv5, YOLOv8, and the most recent YOLOv9, at various training epochs and performance metrics, such as mAP@50 and mAP@50–95 [6]. The results obtained large variants of YOLOv9, respectively termed YOLOv9-C and YOLOv9-E, outperformed both in detection performance: either YOLOv8-L or YOLOv8-X and earlier updates of YOLO. The true positive rate of YOLOv9, in particular, was higher than that of YOLOv8, suggesting its applicability for inference in real-world scenarios. While the above findings establish YOLOv9's robustness in a complex, multiple entity sporting environment, such as soccer, its suitability for table tennis is currently unknown. Of significance, the speed with which the ball moves relative to the playing court in table tennis Jaccard et al. is substantially higher than that in soccer. Consequently, the probability of a miss in ball detection due to high-speed movement is also considerably higher. It is, therefore, necessary to delve, through research, further to establish the suitability of YOLOv9 for the transformation of video data within novice-based sports, such as table tennis.

2.2. Traditional Tracking Methods in Table Tennis

Ji et al. explained the specific difficulties and problems that need to be solved when tracking a table tennis ball, and it is unavoidable at low-speed cameras which are cheaper but sensitive to shake and smear [7]. They integrated several conventional machine learning methods with a visual attention-mimicking segmentation model (VOCUS) to overcome the shortcomings commonly observed in real-time tracking. The VOCUS framework based on human vision attention led to saliency-based image segmentation, then Local Gradient Patterns (LGP) were applied for feature extraction and AdaBoost method was used for classification. This two-step pipeline ensured accurate

ball localization even under visual blurring or distortion in cluttered visual scenes. While this work successfully showed the feasibility based on classical machine learning methods, it relied on hand engineered features and was limited by the amount of training data and variety of scenes. Furthermore, the pipeline required a rather fixed setting and the inclusion of modern object detectors using deep learning techniques was missing. This is an exciting opportunity to improve filling this gap with more powerful deep learning-based detectors, like YOLOv9, which have superior generalization in diverse lighting conditions, better occlusion handling and higher true positive rates. It is remarkable that YOLOv9 performs well in fast sports such as soccer, the article proposes the validation of YOLOv9 on table tennis as a novel and substantial contribution to the community.

2.3. YOLO in Tennis vs. Table Tennis

Recently, Glaspey 2023 has conducted research on tennis ball detection with deep learning models and compared YOLOv8 and Mask R-CNN for high-speed sport shooting [8]. They reported precision 0.806 with YOLOv8, but a pretty mild correct classification rate of 28.98% due to it being seriously challenged by small and fast-moving objects in HD broadcast frames. Even with high confidence, YOLOv8 was not very consistent on its positive detections of the tennis ball (especially when the ball becomes only a few pixels in wide angle shots) on the consecutive frames. This limitation further shows that a comparative study between YOLOv8 and YOLOv9 models is required in order to guarantee that recent updates are increasing the correct classification rate. Although such results illustrate the potential of YOLOv8 for achieving high-quality predictions in an ideal environment, detection accuracy remains sensitive to the size of the object of interest as well as to the portion of the respective object that is occluded when it comes to small and partially occluded objects—a circumstance prevalent in both tennis and table tennis.

But there are multiple differences between these two sports, and models trained from tennis do not necessarily adapt well for the game of table tennis. First, while in both sports balls are small, moving fast objects, the table tennis one is much smaller (40 mm diameter vs. 67 mm for tennis) and usually white, leading to different contrast conditions compared to varying table and background (scene) colors. Two, tennis coverage tends to be filled with wide-angle images shot from behind the baseline bringing the ball down to being a mere speck in high-resolution frames. In contrast, recordings of table tennis typically use close, side-angle or top views that make the object larger per pixel, but are plagued by a lot of motion blur, stemming from the fast motion and spin of the ball. Third, although the speed of tennis serves is above 200 km/h, the longer flying path yields relatively more frames to track the ball. The frame-based window of detection in table tennis is also much less (i.e., a portion of the time available in a frame is smaller for the viewer in table tennis than in baseball), due to the shorter distances and interval between shots, even if the speeds of the rotary torso and racket are only slightly less (~100 km/h) in table tennis.

Taken together, these differences are what make inference results of YOLOv8 in tennis not directly applicable for the table tennis detection task. The application of the YOLO implementation of the YOLOv8 model in tennis demonstrates the ability to localize high-speed objects in certain cases, however, for table tennis, specific calibration and even possibly architectural adjustments need to be considered.

2.4. Summary

The literature reviewed, highlights the possibilities and limitations of multiple object detection techniques for fast- paced sports scenarios. Studies on soccer and tennis have demonstrated that YOLO-based models, YOLOv8, YOLOv9 in particular, achieve solid performance for real-time detection. However, the unique spatiotemporal difficulty of table tennis (e.g., the fast speed of table tennis and the small dimension of object etc.), such as high velocity of the ball as compared to the playing field, the smaller size of the objects and the faster changes of the trajectory and so on, makes it a serious challenge that previous works only partially solve. Our work extends the previous work

by specifically investigating YOLOv9 on table tennis in an attempt to fill the gap between the sports tracking generative model and pioneering sports tracking demand of the sport analysis.

3. Methodology

3.1. Dataset

The experiment is applied to "Table Tennis Ball Detection" dataset provided in Roboflow Do these Annotated images in table tennis play under various condition are shown. Annotations are YOLO format using bounding boxes on the ball class. We split this dataset into 80% training set, 13% validation set and 7% test set in a way to have an unbiased estimate. Data augmentation methods such as change in brightness, random crop, rotation, horizontal flip, motion blur, and adding noise were used to improve the robustness of the model in different environmental conditions.

3.2. Models and Framework

The GELAN-C model used in this work is the core model of the YOLOv9 family, which is the variant of YOLOv9 for high accuracy (with a relative moderate model size). Its performance was compared with the YOLOv8 model for baseline comparison. The training and inference were performed using the Ultralytics YOLOv9 framework in PyTorch running on a Tesla T4 GPU. The model was pretrained on gelan-c with pretrained weights. pt and fine-tuned on the table tennis dataset. It adopted RepNCSPELAN, ADown and SPPELAN components to allow the spatial-aware reasoning and gradient propagation.

3.3. Training Configuration

The training configurations, detailed in Table 1, were kept constant across the training of YOLOv9 and v8 to ensure consistency in results.

Table 1. Training Hyperparameters.

Parameter	Value
Epochs	30
Batch size	16
Input image size	640×640 pixels
Optimizer	SGD
Initial learning rate	0.01
Momentum	0.937
Weight decay	0.0005
Warmup epochs	3.0
Device	CUDA: Tesla T4 GPU

In terms of augmentations, the default augmentations of Mosaic, Flip, Mixup, and Copy-Paste were applied to both YOLOv9 and v8, while Blur and CLAHE were only applied to v9 due to a technical limitation in the framework's support for the YOLOv8 model; all other parameters were kept identical to ensure a fair comparison.

3.4. Experimental Procedure

The experimental procedure followed three main steps:

1. Data Preparation: The Rob flow dataset was downloaded and preprocessed using the Ultralytics preprocessing pipeline.
2. Training: YOLOv8 and YOLOv9 models were independently trained using the same training and validation data splits.
3. Testing: Both trained models were evaluated on the unseen test set.

4. Results and Discussion

4.1. Evaluation Metrics

The standard object detection metrics were used to evaluate the performance of the YOLOv9-C (GELAN-C):

Precision (P): The ratio of the number of successful table tennis ball detections to the number of resulted detections.

Recall (R): The ratio of the true table tennis ball instances that were predictably detected by the model.

mAP@0.5: Mean average precision at Intersection over Union (IoU) threshold of 0.5, a threshold widely used to assess general object detection performance.

These statistics were calculated on 460 images and 231 annotated ball instances as a dedicated validation set.

4.2. Quantitative Results

As detailed in Table 2, a direct comparison of the models after 30 epochs of training reveals distinct performance trade-offs.

Table 2. Comparative Model Performance.

Metric	YOLOv9	YOLOv8
Precision	0.880	0.743
Recall	0.508	0.604
mAP@0.5	0.613	0.579

The training progression for the superior model, YOLOv9, is shown in Figure 1.

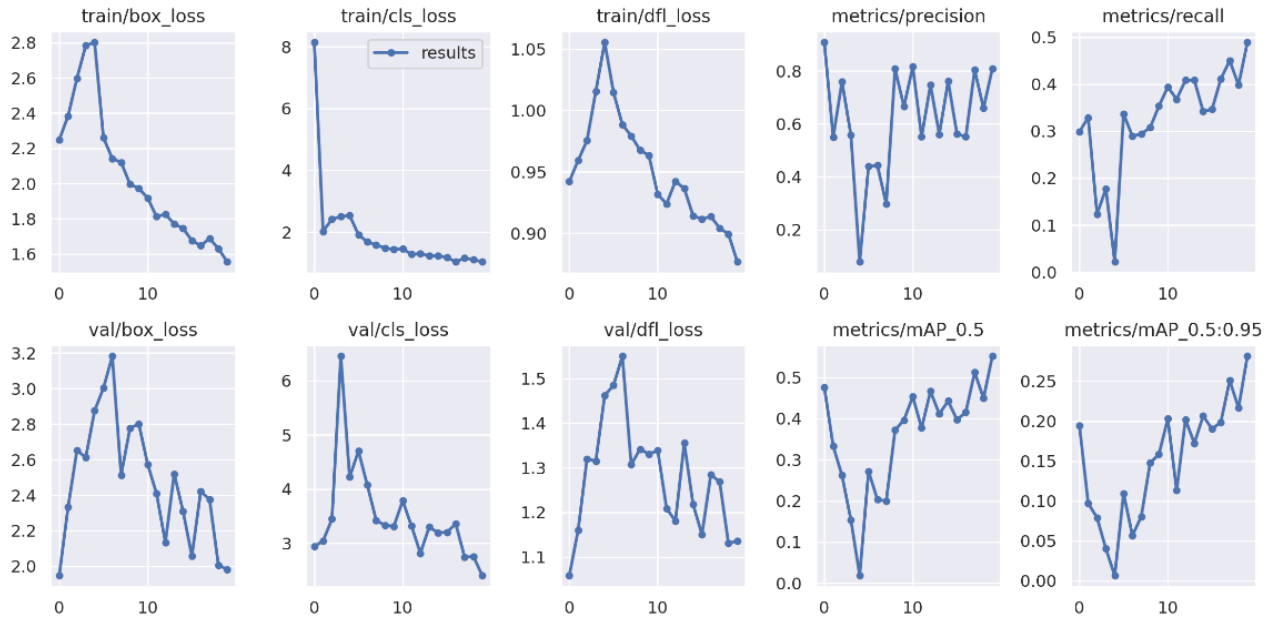

Figure 1. Training and validation performance curves for the YOLOv9-C model over 30 epochs.

For Figure 1, the top row shows metrics calculated on the training set (e.g., train/box_loss) and the bottom row shows the same measures on the validation set (e.g., val/box_loss). The steadily decrease of loss functions and the overall improvement of mAP validate that the model is learning from the training phase properly.

4.3. Discussion of Results

Based on the results from Table 1 and 2, we can clearly observe that YOLOv9 shows superior performance over YOLOv8 in terms of precision and mAP@0.5, showing better predictive value and

reliability. Nevertheless, YOLOv8 has a contains higher recall, which means it can detect more real table tennis ball contour compared to others at the price of introducing more false positives. In particular, the high precision of YOLOv9 (88%) suggests that the model produces relatively few false positives, which is important in real-time applications when false alarms need to be minimized. The model's mAP@0.5 of 0.613 also proves its strong baseline detection ability, and the mAP@0.5: 0.95 for 0.296 illustrates the difficulty of achieving a maintained high detection confidence across localization overlaps of differing degrees. However, with 50.8% recall, the model is failing to find large number of true ball instances, presumably due to difficult situations including motion blur, small object size, or partial occlusion that are common in table tennis videos.

The training curves of the YOLOv9 run (See. Fig. 1) indicate that key performance indicators are going through successful iterations:

The box/class/DFL loss decreased consistently during the training, and they all converged effectively.

mAP@05 and mAP@0.5: 0.95 increased steadily and after 30 epochs had a mean Dice coefficient of 0.613 while 0.296.

Precision stayed relatively high and stable, with a gradual increase in recall after an initial drop (a sign that the model is indeed getting better at generalizing beyond easy examples)

These phenomena confirm that YOLOv9 GELAN-C can learn effective representations for the recognition of fast-moving small objects (e.g., a table tennis ball and pulse tagged as PUL) in real-life match scenarios.

4.4. Analysis and Implications

The findings show that YOLOv9 is able to detect the ball with a higher precision, which is important in applications such as umpiring aid and rule violation detection where a low number of false positives is crucial. This improved accuracy makes a direct water on the tree of umpire fallibility and the bounds of play challenges interpret, as per the Introduction, a decisive instrument to review tight situations such as net or edge balls. While its lower recall is a limitation in coverage that could be compensated by the high confidence of the model and the clarity in predictions. On the other hand, as a "jack of all trades", YOLOv8 may be more appropriate for tracking applications that require comprehensiveness, e.g., ball trajectory analysis, in the sense that the cost of losing a detection is worse than some occasional noise.

The training curves indicate that additional performance gains may be achieved by:

- Continuing training after 30 epochs, as the mean Average Precision (mAP) and recall were increasing at epoch 30.
- More hyperparameter tuning especially on loss balance and learning rate schedules.
- Adding tracking over time or post-processing to handle frames missed under low recall.

In the realization of this, a hybrid solution could be imagined: the usage of YOLOv9 for trustworthy, high criticality (i.e., net touch or service faults) detections, and YOLOv8 for the tracking of the ball-frame by frame during a longer rally.

4.5. Limitations and Future Work

The results of the study might have been influenced by numerous problems in study design and implementation. Note that single-GPU training and inference on an NVIDIA Tesla T4 may have imposed limited batch size and slowed convergence for larger models such as YOLOv9-C. With more powerful GPU(s) it may be possible to use larger input sizes and train for more epochs. Both models were trained by only 30 epochs, longer training can be not only generalized better but also increase the recall and mAP scores that open a line for future work. However, the Roboflow dataset is small and domain-specific with problems, such as mislabeled frames and small environmental variation, that make it difficult for the model to ultimately generalize. These constraints are indicative of resource constrained Ness and experimental scope decisions disclosing of a way forward that can be used to improve future performances.

5. Conclusion

This study established the relative efficacy of YOLOv9 as against YOLOv8 at detecting table tennis balls within a real-time video sequence. A performance trade-off between the two models was realized; while YOLOv9 delivered better precision (0.880 vs. 0.743), as well as superior mAP@0.5 (0.613 vs. 0.579) YOLOv8 was better at doing so in recalled measures which implies that it was more aggressive in pinpointing occurrences of the ball but with greater noise. Practically, this result implies that the choice of model depends on the application; high precision is suitable for high-stakes applications such as automated umpiring since high precision implies a low rate of false positives that are quite costly. In turn, higher recall is the choice for application in the type of domain where detailed performance measures are sought since higher recall implies capturing every possible detection. This research quantifies the trade-off and thus provides a clear framework for deploying the models in fast-paced sports, bringing us closer to more accessible and reliable solutions of sports technology.

References

[1] Myint H, Wong P, Dooley L, et al. Tracking a table tennis ball for umpiring purposes. 14th IAPR International Conference on Machine Vision Applications, 2015: 170 - 173.

[2] Tang H, Mizoguchi M, Toyoshima M. Speed and spin characteristics of the 40mm table tennis ball. Table Tennis Sciences, 2002, 4: 278 - 284.

[3] Wang C Y, Yeh I H, Mark Liao H Y. Yolov9: Learning what you want to learn using programmable gradient information. European conference on computer vision, 2024: 1 - 21.

[4] Sohan M, Sai Ram T, Rami Reddy C V. A review on yolov8 and its advancements. International Conference on Data Intelligence and Cognitive Informatics, 2024: 529 - 545.

[5] Lin H I, Yu Z, Huang Y C. Ball tracking and trajectory prediction for table-tennis robots. Sensors, 2020, 20 (2): 333.

[6] Markappa P, O'Leary C, Lynch C. A review of YOLO models for soccer-based object detection. Sixth International Conference on Intelligent Computing in Data Sciences, 2024: 1 - 7.

[7] Ji Y F, Zhang J W, Shi Z H, et al. Research on real–time tracking of table tennis ball based on machine learning with low-speed camera. Systems Science & Control Engineering, 2018, 6 (1): 71 - 79.

[8] Glaspey J. Tennis ball detection with computer vision. California Polytechnic State University, San Luis Obispo, 2023.