Table Tennis object tracking and analysis

Lakshay Baijal(2024202006), Rishabh Sahu(2024201037), S. V. Mohit Kumar(2024201010)

1 Data sources

Project Objectives and Data Needs

To track objects and analyze events in table tennis match videos, the following characteristics are desired- Data consisting of high-resolution videos capturing table tennis matches from multiple angles will help to ensure accurate object detection and event analysis. Since the detection of events such as hitting a shot, missing the ball, and hitting the ball outside the table requires that the movement of the ball in relation to the net, table, and players is observed. Thus at the occurrence of such events the ball must be clearly observed. Also, to detect objects like players, the table, and the table tennis ball, and events such as player movements, ball trajectories, and shot types, training data needs to be labeled. This poses a challenge since naturally occurring table tennis match footage does not have added labels .

Dataset Sources

To train ai models to recognize objects in match videos certain groups and organizations have created labelled datasets by manually adding annotation labels to existing table tennis match videos. Researchers who worked on TTNet[1] had created a manually labelled high definition dataset in order to train data for their model which we hereafter refer to as OpenTTGames [2].

Data Collection Techniques

Raw data for our objective can usually be found from the following sources

Video Recording : Use of high-speed cameras to capture matches from multiple angles. High frame rates are essential for tracking fast-moving objects like table tennis balls. For the purposes of labelling usually data is manually labelled by adding bounding boxes around the objects of interest. Though before this general object detection procedures can be used to reduce the load of selecting the object regions and only focus

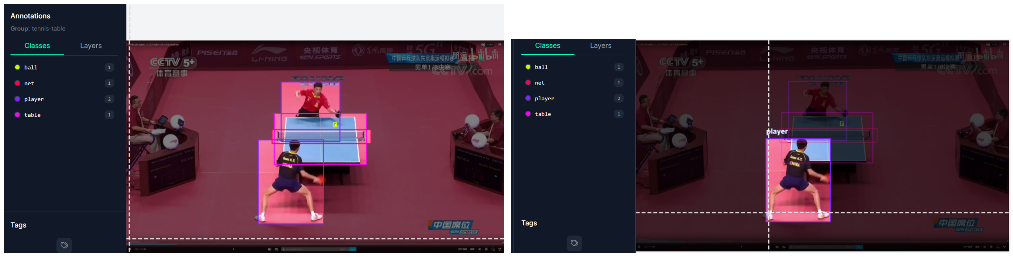
Challenges in Data Acquisition

High-Speed Object Tracking: Capturing the fast-moving table tennis ball accurately requires high-quality equipment and precise calibration. Annotation Complexity: Labeling every frame of a high-speed video is time-consuming and error-prone. Data Diversity: Ensuring sufficient representation of different playing styles, environments, and skill levels can be difficult.

2 Data Exploration

Dataset Overview

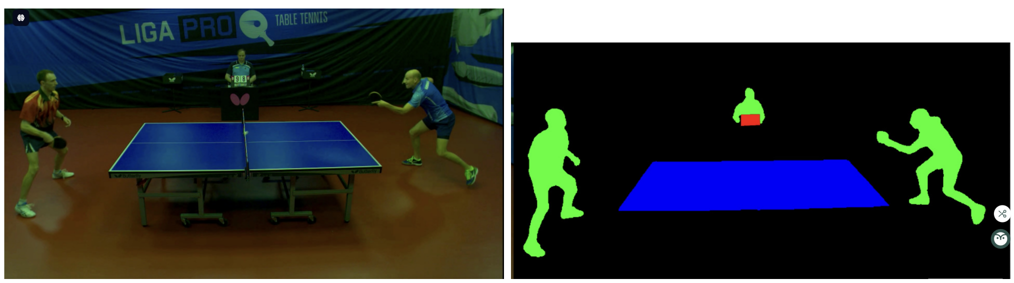
RoboFlow [3] Roboflow Universe Dataset contains annotated images extracted from table tennis matches. Each image includes bounding box annotations for key objects such as the ball, players and table. The



(a) *Bounded Boxes* (b) *Annotated Labeled Image*

Figure 1: *Example of Roboflow Dataset Image*

data set has diverse conditions like different camera angles, lighting conditions, and background variations, which is designed to capture the fast-paced nature of the sport, where the ball often appears as a small, fast-moving object.



(a) *normal image* (b) *segmentaion mask of the image*

Figure 2: *Example of OpenTTGames dataset image*

OpenTTGames [2] This is an open dataset introduced by OSAI, which aims at different tasks in Table Tennis,like ball detection,player detection, table and scorecard and fast in-game events spotting. It includes full-HD videos of table tennis games recorded at 120 fps with an industrial camera. Every video is equipped with an annotation, containing the frame numbers and corresponding targets for this particular frame: manually labeled in-game events (ball bounces, net hits, or empty event targets) and/or ball coordinates and segmentation masks, which were labeled with deep learning-aided annotation models.

Characteristics of Data

RoboFlow The dataset contains 1000 annotated images taken from real matches. Each image includes bounding box annotations for 3 object classes ball, player and table. The dataset was created on Roboflow Universe and supports multiple export formats such as YOLOv5/7/8/11. The images come from diverse environments with different lighting conditions, table colors, and camera angles. This variation improves the robustness of object detection models trained on it.

OpenTTGames The dataset contains 5 full - HD videos of 10 - 25 min duration and 7 videos for testing. Each video is accompanied with markup files having ball coordinates and events. It also contains folder with segmentation masks. The dataset contains 4271 manually entered events of 3 classes - ball bounces, net hits and empty events. The dataset doesn't have any noise like audience or any other objects, other than the players, ball, table and the match refree, as other objects other than the ones specified are blacked out and removed using deep learning models.

Relation with general model

Standard Models like Yolo locate objects by predicting bounding boxes and class labels within an image. Object Detection and Tracking apply within a specialized domain within unique challenges like

1. **Small Object Detection** Table tennis balls are very small and move quickly, very similar to the task of satellite imagery, where it is difficult to detect tiny fast-moving objects.
2. **High-Speed Motion and Occlusions** The ball can get blocked by players or the net, which test the ability to handle partial visibility and motion blur and can be compared with the general surveillance test.
3. **Contextual Relationships** In object detection task, models might simply label a person or a car. In table tennis we label player, ball and table which can have strong contextual ties as balls position relative to table and player is important. We can also find more specialized models, such as predicting the location of the bounce of balls.

2.1 Bias and Representativeness

Roboflow Dataset can be biased as we are only using professional images of table tennis matches which use fixed camera, high quality lighting, and standardized table setup. This means that the model might not work well for amateur games or in different environments. There can be fewer images of the ball as compared to players, which can make it harder for our model to spot the ball.

OpenTTGames This dataset also uses high quality videos recorded from industrial cameras in 120 fps with high quality lighing, making it inefficient for the videos with low quality and poor lighting conditions. A unique feature of this dataset is the blacking out of all elements except the players, table, and match referee, ensuring a focused analysis of the game. One limitation of this dataset is the absence of images captured from multiple angles, restricting the diversity of perspectives available for analysis.

3 References

[1] R. Voeikov, N. Falaleev, and R. Baikulov, "Ttnet: Real-time temporal and spatial video analysis of table tennis," 2020.

[2] R. Voeikov, N. Falaleev, and R. Baikulov, "Ttnet: Real-time temporal and spatial video analysis of table tennis," in *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[3] KGBUPT, "table-tennis detector dataset." `https://universe.roboflow.com/kgbupt-o5q6g/table-tennis-detector`, aug 2023. visited on 2025-03-23.