

Research Article

Classification of Edge and Side Contacts in Table Tennis Using High-Speed Camera Images

Kazuma Sakamoto^{1*}, Reo Ishii², Iori Iwata², Riku Kaiba³, Yoshihiro Ueda¹

¹Faculty of Production Systems Engineering and Sciences, Komatsu University, Komatsu, Ishikawa 923-8511, Japan

²Graduate School of Sustainable Systems Sciences, Komatsu University, Komatsu, Ishikawa 923-8511, Japan

³Former Graduate Student, Graduate School of Sustainable Systems Science, Komatsu University, Komatsu, Ishikawa 923-8511, Japan

Email: kazuma.sakamoto@komatsu-u.ac.jp, 26211003@komatsu-u.ac.jp, 24311001@komatsu-u.ac.jp, 24211007@komatsu-u.ac.jp, yoshihiro.ueda@komatsu-u.ac.jp

*Corresponding Author

ARTICLE INFO

Article History

Received 9 December 2025

Accepted 22 September 2026

Keywords

Table tennis officiating
Edge and side classification
Instance segmentation
Overlap ratio

ABSTRACT

This research investigates an image-based method for supporting edge and side classification in table tennis using a camera arrangement intended to reduce interference with play. Ball contacts generated under controlled conditions were recorded using two synchronized high-speed cameras. Images from one camera were used for classification, whereas the second view was used to confirm the contact position and assign the ground-truth label. A frame visually judged to be closest to the contact moment was selected for each contact. The ball region was estimated using YOLO11s-based instance segmentation, and the primary evaluation used a white-line region manually specified from the visible table geometry. Each contact was classified according to the proportion of the ball region overlapping the white-line region. Five repeated holdout evaluations compared the normalized overlap ratio with the raw number of overlapping pixels and examined the influence of alternative white-line definitions. The normalized overlap ratio achieved a mean accuracy of 84.2%, with edge and side recalls of 86.0% and 82.4%, respectively. These results indicate that the local spatial relationship between the ball and the white line can support edge and side classification when an appropriate contact frame is available. Automatic contact-frame extraction and a physically consistent automatic white-line reference remain necessary for practical deployment.

© 2026 The Author(s). Published by The Society of Artificial Life and Robotics.
This is an open access article distributed under the CC BY-NC 4.0 license
(<http://creativecommons.org/licenses/by-nc/4.0/>).

1. Introduction

According to the Laws of Table Tennis, the upper surface of the table constitutes the playing surface, whereas the vertical sides of the tabletop are not included in the playing surface [1]. Therefore, a ball that contacts the upper surface near the table boundary is classified as an edge ball and constitutes a valid bounce. In contrast, a ball that contacts the vertical side is classified as a side ball and does not constitute a valid bounce. Figure 1 shows examples of edge and side contacts.

Although edge and side contacts are clearly distinguished under the rules, reliably classifying them during actual play is not always straightforward. Some contacts can be recognized relatively easily, whereas others are difficult to distinguish because the ball travels rapidly and the possible contact points on the upper and side surfaces are close to each other. The appearance of the contact also varies depending on the viewing direction and imaging conditions. Consequently, classification based on

visual observation may become uncertain when the contact position is not clearly visible.

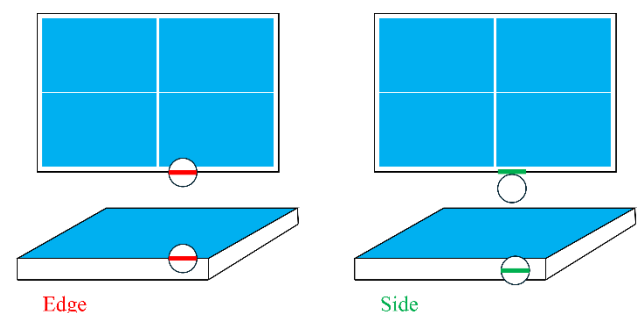


Figure 1 Examples of Edge and Side Contacts

Several approaches have been investigated to support the classification of difficult contacts in table tennis. Video-based methods track the ball across consecutive frames and analyze changes in its motion before and after contact, whereas sensor-based methods use physical responses such as impact-induced vibrations. Although

both approaches provide useful information, trajectory analysis requires stable ball tracking across multiple frames, while sensor-based systems require dedicated hardware to be installed and calibrated on the table. Therefore, an image-based classification method that does not rely on table-mounted devices may provide a more practical alternative.

In our previous research, edge and side contacts were classified using the overlap between the ball and white-line regions [2]. Both regions were estimated using YOLOv8-based instance segmentation, and classification was performed according to the raw number of overlapping pixels. The results indicated that the positional relationship between the ball and the white line could be used for edge and side classification without reconstructing the complete ball trajectory. However, the camera was positioned close to the table and could interfere with play. In addition, the effects of the classification measure and the estimated shape of the white-line region on classification performance were not examined in detail.

This paper extends that research in four respects: (1) the classification camera is positioned farther from the immediate playing area to reduce interference with play, (2) a normalized overlap ratio is compared with the raw overlap count, (3) the influence of different white-line region definitions is examined, and (4) the feasibility of applying the same classification approach to frames recorded with a consumer-grade GoPro camera is evaluated.

For the primary experiment, ball contacts were recorded using two synchronized high-speed cameras positioned at different viewpoints. Images from one camera were used for classification, whereas the video from the other camera was used to select the contact frame and assign the ground-truth label. Frames identified as being closest to the contact moment were extracted from the recorded videos and used as input images. This configuration allows the proposed method to perform classification using a single-view image while using the other viewpoint only to prepare the evaluation data.

The ball region is estimated using YOLO11s-based instance segmentation [3]. The white-line region is defined before classification and serves as a reference corresponding to the boundary of the upper playing surface. A normalized overlap ratio is calculated by dividing the number of pixels shared by the ball and white-line regions by the total number of pixels in the ball region. A contact is classified as an edge when the normalized overlap ratio is equal to or greater than a specified threshold; otherwise, it is classified as a side. Normalization by the ball-region area is used because the raw overlap count may vary depending on the apparent size of the ball and the area of the estimated ball mask.

The evaluation consists of four parts corresponding to the extensions described above. First, classification performance is evaluated using high-speed-camera images recorded from a position designed to reduce interference with play. Second, the normalized overlap ratio is

compared with the raw overlap count. Third, the influence of the white-line region definition is examined using manually specified regions, YOLO11s-estimated regions, and vertically expanded regions. Finally, the feasibility of applying the proposed classification approach to frames recorded with a consumer-grade GoPro camera is investigated.

The purpose of this research is to determine whether the local spatial relationship between the ball region and the white-line reference can support edge and side classification using a single image captured near the moment of contact. This research also evaluates a camera arrangement designed to reduce interference with play and examines how the classification measure and white-line region definition affect classification performance.

2. Related Work

Sports decision-support systems use different types of information depending on the target event. Video-based methods analyze the motion of players and balls, sensor-based methods measure physical responses generated by contact, and segmentation-based methods convert image regions into quantitative values. This section reviews these approaches and clarifies the position of the present research.

2.1. Video-Based Officiating Support in Sports

Video-based systems have been introduced to support decisions that are difficult to make through direct observation. In soccer, multi-view footage has been used to analyze foul situations [4]. In tennis, Hawk-Eye estimates ball trajectories and bounce positions from multiple cameras and compares them with court lines [5]. A monocular-camera method has also been proposed to determine whether a ball lands inside or outside the court [6].

These examples demonstrate the value of video evidence for rapid events and incidents outside the referee's direct view. Most existing systems primarily assess motion or landing locations over a comparatively broad area. Edge and side decisions in table tennis concern a much smaller zone, so the present work examines the local spatial relationship between the ball and the table boundary at contact.

2.2. Research on Table Tennis Video Analysis

Table tennis video analysis has been applied to service-fault detection, rally analysis, scoring, and event recognition. Service-fault detection methods estimate the ball and racket positions and examine the height and motion of the ball toss [7], [8]. Other methods track the ball using background subtraction and Kalman filtering [9]. Multi-view systems reconstruct three-dimensional trajectories and detect events such as bounces, net crossings, and repeated bounces [10]. Real-time tracking has also been developed for robotic table tennis [11].

Such approaches depend on ball locations obtained across consecutive frames, and reliable tracking can deteriorate when the ball moves rapidly, occupies only a limited image area, appears blurred, or becomes temporarily occluded. The present work avoids reconstructing an entire trajectory: a frame judged visually to be nearest to contact is selected, and classification is based on the spatial arrangement observed in that frame.

2.3. Classification Based on Contact Information Obtained Using Sensors

Ball contact can also be detected from physical signals. Accelerometers installed on a table or its supports have been used to estimate contact positions from vibration waveforms [12]. A practical system has also been developed to distinguish edge and side contacts by analyzing vibrations measured on the side of the table [13].

Collision signals provide direct evidence of impact and are relatively insensitive to the visual appearance of the ball. Their use nevertheless requires sensors to be attached and maintained, and a different table or mounting configuration may require recalibration. The image-based method considered here avoids dedicated contact sensors, although accurate identification of the contact frame and image regions remains necessary.

2.4. Officiating Support Using Ball Trajectories and Landing Points

Research by Fujiwara and colleagues introduced a monocular-camera approach that detects table tennis balls with Mask R-CNN and determines in-or-out decisions from their post-bounce trajectories [14], [15]. Tennis line-calling methods have also classified balls using the relationship between court lines and landing positions [16].

Ball and line information is also used in these officiating-support approaches, but their decisions are derived from temporal trajectory changes or landing coordinates. The present method instead evaluates the pixel-level intersection between the ball region and the white-line reference within a single image near contact.

2.5. Deep Learning Based Object Detection and YOLO11s

Deep learning-based methods have been developed to estimate object locations and regions in images. Faster R-CNN is a two-stage object detection method that first generates region proposals and then performs classification and localization [17]. TTNNet is a multi-task network developed for real-time analysis of table tennis videos. It performs ball detection, semantic segmentation of the table, players, and scoreboard, and event spotting for ball bounces and net hits [18].

Since the proposed method requires the shape of the ball at the pixel level rather than only its center coordinates or bounding box, instance segmentation is used to estimate the ball region. YOLO11s supports multiple computer

vision tasks, including object detection, instance segmentation, image classification, pose estimation, and oriented object detection. In this research, a YOLO11s-based instance segmentation model is used to estimate the ball region. The effect of the white-line region definition on classification performance is also examined.

2.6. Quantitative Evaluation Using Region Information

Estimated image regions have been converted into quantitative measures such as area, proportion, span, and length. In fruit analysis, detected regions have been used to estimate fruit diameter, visible ratio, and size [19], [20]. Similar approaches have been applied to quantify diseased plant regions and cracks in infrastructure [21], [22].

The present work applies the same concept of converting segmented regions into numerical information by computing the overlap between the YOLO11s-estimated ball region and the white-line reference. Here, the resulting value is used for edge and side classification rather than for measuring object dimensions or damage severity.

2.7. Positioning of This Research

Video analysis, trajectory reconstruction, contact sensors, and deep-learning-based region estimation have all been studied for sports officiating. Within this literature, relatively little attention has been given to edge and side classification from the region relationship visible in a single table-tennis contact image.

The proposed setup uses two synchronized high-speed cameras to verify the timing and location of contact, while the image from one viewpoint is supplied to the classifier. YOLO11s estimates the ball region, and the white-line reference is provided explicitly. Edge and side decisions are then obtained from the proportion of the ball region intersecting that reference. The distinguishing feature is the use of spatial region information at contact rather than complete trajectories or vibration signals.

3. Methodology

This section describes a method for classifying edge and side contacts using a single image captured near the moment of contact. The ball region is estimated using the YOLO11s-based instance segmentation model, and its overlap ratio with a predefined white-line region is compared with a threshold. The processing flow is shown in Figure 2.

An edge contact is expected to produce a larger overlap with the white-line reference than a side contact. Camera perspective can nevertheless make a side ball appear to cross the line in the image. The decision is therefore based on the normalized amount of overlap rather than on a simple binary check for any intersection.

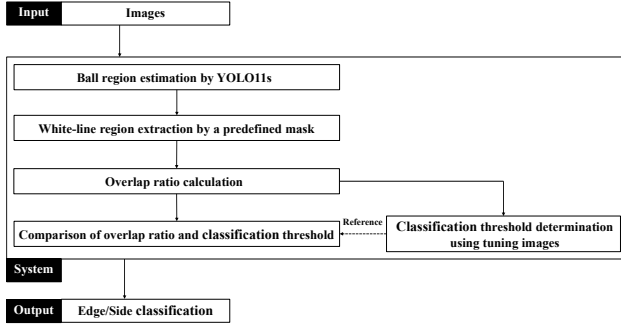


Figure 2 Processing Flow of the Proposed Method

3.1. Image Preparation

Two synchronized high-speed cameras are used to record ball contacts near the table boundary from different viewpoints. Images from one camera are used for edge and

side classification, whereas the second camera is used to confirm the contact moment and the actual contact position.

The verification video was inspected frame by frame by a single evaluator. After the ball approached the table boundary, the first frame showing a change in its motion was identified, and the immediately preceding frame was regarded as the frame closest to the contact moment. The temporally corresponding frame from the classification camera was then selected as the input image. The ground-truth label was assigned by visually examining the contact state using the frames immediately before and after the selected moment in the verification video. An example of an input image is shown in Figure 3.

This configuration allows classification to be performed using a single-view image while improving the reliability of the assigned labels.

3.2. Acquisition of the Ball and White-Line Region

The ball region is estimated using a YOLO11s-based instance segmentation model. Each input image is processed by the trained model to obtain an instance mask corresponding to the table tennis ball. Center coordinates and bounding boxes do not preserve the pixel-level shape required to evaluate the intersection between the ball and the white line. Therefore, the proposed method uses the estimated ball mask for classification.

The white-line region serves as a reference for evaluating the spatial relationship between the ball region and the boundary of the upper playing surface. Because this research focuses on the effectiveness of overlap-based classification rather than the accuracy of automatic white-line detection, the white-line region is specified before classification and supplied to the classification procedure. This setup allows the classification method to be evaluated without directly incorporating errors caused by automatic white-line detection.

Slight changes in the camera position made it difficult to apply a single white-line mask consistently to all images.

Therefore, a separate white-line region was manually delineated for each image using the same annotation rule based solely on the visible geometry of the table and white line. The ground-truth class label, estimated ball mask, and ball position were not referenced during annotation. This procedure was applied consistently to all images to avoid adjusting the white-line region according to the contact class or ball location. Automatically correcting the white-line region to compensate for camera displacement remains a subject for future research.

3.3. Overlap-Based Edge and Side Classification

After the ball and white-line regions are obtained, their spatial overlap is quantified. Let B and W denote the binary masks of the ball region and the white-line region, respectively. The normalized overlap ratio r is defined as

$$r = \frac{|B \cap W|}{|B|} \quad (1)$$

where $|B \cap W|$ denotes the number of pixels contained in both regions, and $|B|$ denotes the total number of pixels in the ball region. Thus, r represents the proportion of the estimated ball region that overlaps the white-line region. Figure 4 shows a visualization example, in which the ball region is displayed in blue, the white-line region in yellow, and the overlapping region in red.

For an image in which the ball region is successfully detected, the predicted class $\hat{y}$ is determined as follows

$$\hat{y} = \begin{cases} \text{edge}, & r \geq T \\ \text{side}, & r < T \end{cases} \quad (2)$$

where T is the classification threshold. If the ball region is not detected, the image is treated as undetected and counted as an incorrect classification in the evaluation.

The raw number of overlapping pixels may vary depending on the apparent size of the ball and the area of the ball mask estimated by YOLO11s. Normalizing the overlap by the total number of pixels in the ball region is therefore expected to reduce the influence of these variations. For this reason, the normalized overlap ratio is used as the primary classification measure.

3.4. Threshold Determination

The classification threshold is determined using only the tuning set. For the normalized overlap ratio, threshold candidates ranging from 0.000 to 1.000 are tested at intervals of 0.001. For the raw overlap count, integer thresholds ranging from 0 to the maximum overlap count observed in the tuning set are tested at one-pixel intervals.

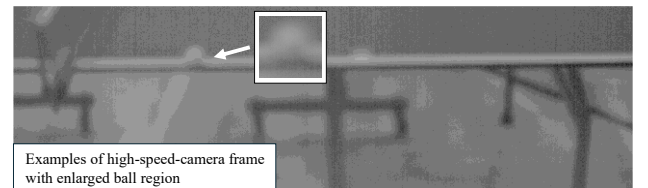


Figure 3 Example of a High-Speed Camera Frame with an Enlarged Ball Region

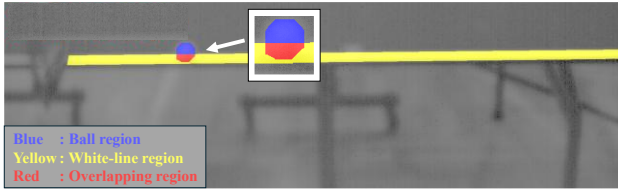


Figure 4 Visualization of the Ball Region, White-Line Region, and Overlapping Region

For each decision variable, the threshold yielding the highest classification accuracy on the tuning set is selected. When multiple thresholds achieve the same accuracy, the threshold with the highest mean of the edge and side recalls is preferred. If multiple candidates remain, the threshold with the smallest absolute difference between the edge and side recalls is selected. If the candidates are still tied, the smallest threshold is selected. The same tie-breaking rule is applied to both the normalized overlap ratio and the raw overlap count. The selected threshold is then applied unchanged to the evaluation set.

4. Experimental Design and Performance Assessment

This section presents the experimental design and performance assessment. Classification is evaluated with high-speed-camera images, followed by a comparison between the normalized overlap and the raw overlap count. Additional experiments examine the effect of the white-line definition and test whether the same principle can be applied to GoPro images.

4.1. Recording Environment

Two monochrome HAS-EF high-speed cameras manufactured by DITECT were used for recording [23]. The image resolution was set to 1600×480 pixels, and the frame rate was 1000 fps. The cameras were connected using BNC cables and a synchronization master cable. Recording was started simultaneously with a trigger button. The recording environment is shown in Figure 5. Camera 1, which was used for classification, was positioned at a horizontal distance of 2.28 m from the table edge, with the lens center at a height of 1.63 m above the floor. The physical tilt angle of the camera was not measured separately. Instead, the elevation angle defined by the horizontal direction and the line connecting the camera lens center to the table edge was 20.9° . Camera 2, which was used for contact confirmation, was positioned at a horizontal distance of 2.82 m from the table edge, with the lens center at a height of 1.06 m above the floor. The angle between the horizontal and the line connecting the camera lens center and the table edge was 6.07° . The ball launching machine was positioned at a horizontal distance of 1.46 m from the table edge, and its height was set to 0.94 m. Lens distortion correction was not applied to the recorded images before calculating the overlap ratio.

A JOOLA iPong Pro table tennis training robot was used to launch the balls [24]. The iPong Pro used in this experiment is shown in Figure 6. The height of the training

robot was set to 0.94 m. The height and distance of the training robot from the table, as well as its launch direction and operating settings, were adjusted so that the balls contacted areas near the boundary between the upper playing surface and the side surface. The use of the training robot reduced the variability associated with balls struck by a human player and enabled repeated trials under consistent conditions. An example of the robot settings is shown in Figure 7.

4.2. Images Used for Classification

The synchronized videos recorded by the two high-speed cameras were visually inspected, and frames judged to represent the moment when the ball contacted the table boundary were extracted. Images from one camera were used for classification, whereas the video from the other camera was used to verify the contact moment and position. Each image was assigned an edge or side ground-truth label according to the confirmed contact position.

A total of 140 images were used for edge and side classification, consisting of 70 edge images and 70 side images. None of these 140 images was included in the training, validation, or test sets used to develop the YOLO11s-based ball segmentation model. For each repetition, 20 images from each class were randomly selected as the tuning set, and the remaining 50 edge and 50 side images were used as the evaluation set. The random division and evaluation were repeated five times.

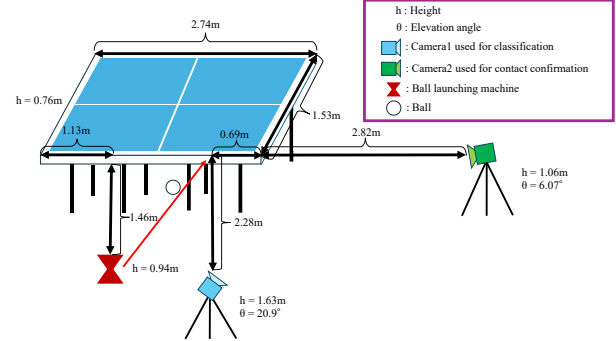


Figure 5 Recording Environment



Figure 6 iPong Pro



Figure 7 Control Settings used for the iPong Pro

4.3. Ball Segmentation Model

For the high-speed-camera images, a YOLO11s-based instance segmentation model was used to estimate the ball region. The dataset contained 400 annotated images. It was split in a 7:2:1 ratio into 280 training, 80 validation, and 40 test images.

Training used 1280×1280 -pixel inputs, four images per batch, and an upper limit of 200 epochs. Early stopping was triggered after 60 epochs without improvement. The resulting model detected the ball in every image used for classification.

4.4. Threshold Selection and Evaluation Metrics

Repeated holdout evaluation was performed using different data divisions. For a given repetition, the classification threshold was selected using the tuning data and was then kept fixed when evaluating the remaining images. Threshold candidates from 0.000 to 1.000 were tested at intervals of 0.001, and the value that produced the highest tuning accuracy was selected.

Classification accuracy, edge recall, and side recall were calculated. Accuracy represents the proportion of all images classified correctly, while edge recall and side recall represent the proportions of correctly classified images within each class. The final results are reported as the mean and standard deviation over five trials.

4.5. Classification Experiment Using High-Speed Camera Images

The proposed method was first evaluated using ball regions estimated by YOLO11s and white-line regions manually defined for each image. Because slight camera shifts occurred during recording, a common white-line region was not applied to all images. Table 1 summarizes the five repetitions. The mean classification accuracy was $84.2 \pm 2.3\%$, while edge recall and side recall were $86.0 \pm 6.0\%$ and $82.4 \pm 1.7\%$, respectively. The selected thresholds ranged from 0.350 to 0.370, with a mean of 0.362 ± 0.011 . These results suggest that edge and side contacts can be distinguished to some extent when an appropriate image of the contact moment is available.

4.6. Comparison of Classification Criteria

The overlap ratio was compared with the raw number of overlapping pixels. The results obtained using the pixel count are shown in Table 2. The mean accuracy was $81.4 \pm 1.8\%$, while edge recall and side recall were $84.4 \pm 10.1\%$ and $78.4 \pm 11.5\%$, respectively. The mean selected threshold was 810.0 ± 119.0 pixels.

The normalized overlap produced a mean accuracy of $84.2 \pm 2.3\%$, exceeding the result obtained from the raw pixel count. Because a pixel count varies with the apparent ball size and the predicted ball-region area, it may fluctuate even under similar contact conditions. Expressing the intersection as a fraction of the entire ball region is expected to moderate these effects. Under the present conditions, the normalized overlap was therefore the more suitable decision variable.

4.7. Effects of White-Line Region Definitions

The effect of the white-line definition was examined by comparing manually defined regions with regions estimated using YOLO11s. A separate segmentation network for the white line was trained using 250 annotated images. The data were divided into training, validation, and test sets at a ratio of 7:2:1. Training was performed using input images of 1280×1280 pixels, a batch size of four, a maximum of 200 epochs, and early stopping after 30 epochs without improvement.

Table 1 Classification Results of the Proposed Method

Trial	Threshold	Accuracy [%]	Edge Recall [%]	Side Recall [%]
1	0.350	84.0	86.0	82.0
2	0.350	88.0	96.0	80.0
3	0.370	82.0	80.0	84.0
4	0.370	83.0	84.0	82.0
5	0.370	84.0	84.0	84.0
Mean	0.362	84.2	86.0	82.4
$\pm$ SD	± 0.011	± 2.3	± 6.0	± 1.7

Table 2 Classification Results using the Number of Overlapping Pixels

Trial	Threshold [pix]	Accuracy [%]	Edge Recall [%]	Side Recall [%]
1	780	81.0	82.0	80.0
2	690	79.0	98.0	60.0
3	780	82.0	84.0	80.0
4	1,010	81.0	70.0	92.0
5	790	84.0	88.0	80.0
Mean	810.0	81.4	84.4	78.4
$\pm$ SD	± 119.0	± 1.8	± 10.1	± 11.5

The results obtained using the YOLO11s-estimated white-line regions are shown in Table 3. The mean accuracy was $89.8 \pm 2.3\%$, while the edge and side recalls were $95.6 \pm 3.8\%$ and $84.0 \pm 7.1\%$, respectively. Although these values were higher than those obtained using the manually defined regions, visual inspection showed that the estimated white-line regions often extended beyond the actual boundaries of the white line. This observation

indicates that the higher classification accuracy did not necessarily correspond to a more accurate representation of the physical white line.

To examine the influence of this expansion, the manually defined white-line regions were dilated by 0, 3, 5, and 10 pixels in the vertical direction of the image. In Table 4, “Dilation” denotes the amount of vertical dilation applied to the manually defined white-line region. Table 4 summarizes the results of these dilation tests. A dilation of 3 pixels increased the mean accuracy from $84.2 \pm 2.3\%$ to $87.8 \pm 1.8\%$. The accuracies for dilations of 5 and 10 pixels were $84.8 \pm 0.8\%$ and $75.4 \pm 2.5\%$, respectively. Edge recall increased as the white-line region became wider, whereas side recall decreased to $51.2 \pm 5.4\%$ at a dilation of 10 pixels. These results suggest that excessive vertical dilation causes more side contacts to be classified as edge contacts.

Table 3 Classification Results using White-Line Detection with YOLO11s

Trial	Threshold	Accuracy [%]	Edge Recall [%]	Side Recall [%]
1	0.420	90.0	96.0	84.0
2	0.370	86.0	100.0	72.0
3	0.460	90.0	90.0	90.0
4	0.440	92.0	98.0	86.0
5	0.460	91.0	94.0	88.0
Mean ± SD	0.430 ± 0.037	89.8 ± 2.3	95.6 ± 3.8	84.0 ± 7.1

Table 4 Classification Results with Dilation of the Manually Defined White-Line Region

Dilation [pix]	Threshold	Accuracy [%]	Edge Recall [%]	Side Recall [%]
0	0.362 ± 0.011	84.2 ± 2.3	86.0 ± 6.0	82.4 ± 1.7
3	0.420 ± 0.030	87.8 ± 1.8	92.0 ± 6.9	83.6 ± 5.2
5	0.464 ± 0.025	84.8 ± 0.8	91.2 ± 7.7	78.4 ± 7.3
10	0.470 ± 0.035	75.4 ± 2.5	99.6 ± 0.9	51.2 ± 5.4

4.8. Evaluation of Classification Feasibility Using GoPro Images

The feasibility of using a general-purpose camera was examined with a GoPro HERO12 [25]. The area near the table boundary was recorded at a resolution of 2.7K and a frame rate of 240 fps. The videos were converted into individual frames, and images visually judged to be close to the contact moment were selected. All images were cropped to 1600×480 pixels using the same crop position. The white-line definition was equivalent to that used in the high-speed-camera experiment. A separate YOLO11s-based ball-segmentation model was trained for the GoPro images under conditions equivalent to those used for the high-speed-camera model. The GoPro model was trained

using 400 annotated images, none of which were among the 60 images used for edge and side classification.

The dataset consisted of 30 edge images and 30 side images. For each split, 10 edge and 10 side images were assigned to the tuning set, and 20 edge and 20 side images were assigned to the evaluation set. Five different splits were evaluated. The results are summarized in Table 5. The mean accuracy was $90.0 \pm 2.5\%$, while edge recall and side recall were $95.0 \pm 5.0\%$ and $85.0 \pm 9.4\%$, in that order. These results indicate that overlap-based classification was possible for the selected GoPro images.

Table 5 Classification Results using GoPro Images

Trial	Threshold	Accuracy [%]	Edge Recall [%]	Side Recall [%]
1	0.170	87.5	100.0	75.0
2	0.140	87.5	100.0	75.0
3	0.200	92.5	95.0	90.0
4	0.230	92.5	90.0	95.0
5	0.220	90.0	90.0	90.0
Mean ± SD	0.192 ± 0.037	90.0 ± 2.5	95.0 ± 5.0	85.0 ± 9.4

Because the GoPro recorded at 240 fps, the ball displacement between consecutive frames was larger than that in the 1000 fps high-speed camera recordings. A visually selected image may therefore have been captured immediately before or after the actual contact. Severe motion blur was also observed during fast ball motion in the GoPro recordings, and the ball sometimes appeared as multiple separated regions. Examples of the GoPro and high-speed camera images are shown in Figure 8 and Figure 9, in that order. Although classification was possible for selected GoPro images, high-speed cameras were considered more suitable for reliably capturing the contact moment during high-speed play.

4.9. Discussion

The experiment using high-speed camera images and manually defined white-line regions achieved a mean accuracy of 84.2%. This result suggests that the positional relationship between the ball and the white line can be used to distinguish edge and side contacts when an appropriate image of the contact moment is available. The segmented ball region also enables the relationship with the white line to be evaluated at the pixel level, which is difficult when the ball is represented only by its center coordinate or bounding box. Since small differences in the contact position near the table boundary may affect the classification result, preserving the shape and extent of the ball region is considered useful for this task.

Using the raw overlap count yielded a mean accuracy of 81.4%, whereas normalization by the ball-region area increased the value to 84.2%. The unnormalized count changes with both the apparent ball size and the area predicted by YOLO11s. A ratio referenced to the entire ball region is less dependent on those variations, which may explain the observed improvement. This benefit is not

independent of segmentation quality, because overestimation or underestimation of the ball region also changes the ratio. Stable ball-region estimation therefore remains necessary.

YOLO11s-estimated white-line regions unexpectedly produced higher accuracy than the manually specified references. Inspection of the estimated white-line regions revealed that the network frequently included pixels outside the physical white line. The numerical gain should thus not be interpreted as evidence of more accurate line localization. A wider reference region increased the measured overlap and biased the decision toward the edge class.

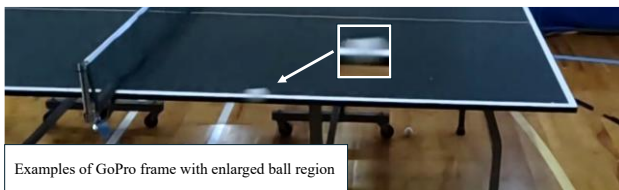


Figure 8 Example of a GoPro Frame with an Enlarged Ball Region

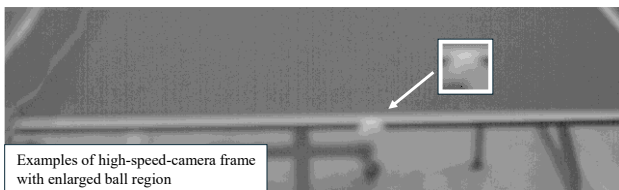


Figure 9 Example of a High-Speed Camera Frame with an Enlarged Ball Region

The dilation experiment supports this interpretation. A 3-pixel enlargement improved accuracy, but the benefit did not continue as the reference became wider. At 10 pixels, edge recall rose while side recall fell markedly, showing that excessive enlargement converts additional side cases into edge decisions. The higher score obtained with the YOLO11s-estimated white-line region is therefore not a desirable improvement if it results from a physically inaccurate reference. The white-line reference has a specific physical role: it represents the boundary between the upper playing surface and the side surface and should not be expanded merely to maximize accuracy. Future work should retain that physical meaning while automatically compensating for small camera shifts.

Although the GoPro experiment produced a higher score than the high-speed-camera experiment, this result does not indicate that the GoPro provided superior imaging performance. The GoPro frames were visually selected from moments when the ball could be confirmed near contact, which may have made the dataset less challenging. In addition, the GoPro recorded at 240 fps, resulting in larger ball displacement between consecutive frames than in the 1000-fps high-speed-camera recordings. During rapid ball motion, some GoPro frames contained elongated or fragmented ball regions due to motion blur. Therefore,

the result demonstrates the feasibility of the proposed method for the selected GoPro frames, rather than equivalent performance across a wide range of high-speed playing conditions.

Manual selection of the frame closest to the contact moment may involve an error of approximately ± 1 frame. In the 1000-fps high-speed-camera recordings, one frame corresponds to 1ms, whereas in the 240-fps GoPro recordings, one frame corresponds to approximately 4.17ms. Therefore, assuming a similar ball velocity, the ball displacement associated with a one-frame selection error is expected to be approximately 4.17 times as large at 240 fps as at 1000 fps. Because the proposed method classifies each contact according to the overlap ratio between the ball region and the white-line region in a single image near the contact moment, selecting a frame immediately before or after the actual contact may change the calculated overlap ratio. In particular, when the overlap ratio is close to the classification threshold, even a small change in the ball position may change the classification result. For example, if the overlap ratio of an edge contact becomes smaller than that at the actual contact moment, it may fall below the classification threshold and be classified as a side contact. Conversely, an increase in the overlap ratio may cause a side contact to be classified as an edge contact. In addition, if such frame-selection errors are included in the tuning data, the overlap ratios of both edge and side contacts may become more variable, which may make the classification threshold less stable. Therefore, the effect of a ± 1 -frame selection error is expected to be greater in the 240-fps GoPro recordings than in the 1000-fps high-speed-camera recordings. However, because the present study did not explicitly evaluate frames shifted by ± 1 frame, the resulting change in classification accuracy cannot be quantified from the current results.

The principal obstacle to a practical decision-support system is the manual selection of the contact frame. The current pipeline automates classification only after a human has chosen the image to be examined. Contact candidates must ultimately be identified directly from video. One possible route is to estimate the ball trajectory and search for changes in displacement, velocity magnitude, and heading across successive frames.

Once trajectory information is available, restricting it to frame selection may leave useful evidence unused. Motion before and after impact may itself help distinguish an edge contact from a side contact. Future experiments should compare three alternatives: overlap-only classification, trajectory-only classification, and a fusion of both. In the fused design, trajectory changes could locate candidate frames, after which the overlap ratio and near-contact changes in speed and direction would be evaluated together. This would combine spatial evidence of the contact location with temporal evidence from the surrounding motion.

Generalization should be tested across changes in camera position, viewing angle, illumination, ball speed,

scene composition, and table appearance. Because the high-speed cameras used here produce monochrome images, ball color is not the central issue. More relevant are illumination-driven changes in brightness and contrast, which may alter ball-region estimation. The highest priority is automatic contact-frame extraction, alongside more reliable ball segmentation, a stable physically meaningful white-line reference, broader environmental evaluation, and comparison of overlap-based and trajectory-based evidence.

5. Conclusion

This research examined an image-based method for supporting edge and side classification in table tennis. The proposed method estimates the ball region using YOLO11s-based instance segmentation and measures the proportion of the ball region that overlaps a predefined white-line region. The calculated overlap ratio is compared with a threshold selected from tuning data to classify each contact as either edge or side.

The experiment using high-speed camera images and manually defined white-line regions achieved a mean accuracy of 84.2%, with an edge recall of 86.0% and a side recall of 82.4%. When the raw number of overlapping pixels was used, the mean accuracy was 81.4%. The higher mean accuracy obtained with the overlap ratio suggests that normalization may reduce the influence of differences in apparent ball size and estimated ball-region area.

Automatic white-line estimation with YOLO11s yielded a higher classification score than manual specification, but the predicted white-line regions commonly extended outside the physical line. The enlarged reference increased overlap values and favored edge decisions. A similar pattern appeared in the dilation test: a slight enlargement improved accuracy, whereas excessive enlargement sharply reduced side recall. The numerical improvement from the YOLO11s-estimated white-line region should therefore not be interpreted as accurate white-line detection or as an inherently preferable result. The reference must remain consistent with the actual boundary between the upper and side surfaces.

Selected GoPro images produced a mean accuracy of 90.0%, but the 240-fps frame rate made the exact impact frame harder to obtain during fast shots. Motion blur also distorted or split the apparent ball region in some images. The experiment consequently demonstrates classification feasibility for usable selected frames, rather than superiority of a general-purpose camera over high-speed imaging.

Manual selection of the contact frame remains the main limitation. A practical decision-support system will require automatic detection of contact candidates from frame-to-frame variation in ball location and motion. Because trajectory information may also improve the final decision, future evaluation should compare overlap-only, trajectory-only, and fused classification schemes.

Robustness must also be examined under varied camera positions, viewing angles, illumination, ball speeds, backgrounds, and table appearances. With monochrome high-speed images, changes in brightness and contrast are more relevant than ball color because they can affect ball-region estimation. Subsequent work will therefore emphasize automatic contact-frame extraction, a stable physically grounded white-line reference, improved ball segmentation, and evaluation across a broader range of recording conditions.

References

1. International Table Tennis Federation, "ITTF Statutes 2025 (clean version)," 2025, https://documents.ittf.sport/sites/default/files/public/2025-02/2025_ITTF_Statutes_clean_version.pdf/ (accessed July 13, 2026).
2. R. Ishii, K. Sakamoto, I. Iwata, R. Kaiba, and Y. Ueda, "Basic Research on Automating Edge and Side Determination in Table Tennis," *Proceedings of the 2026 International Conference on Artificial Life and Robotics (ICAROB 2026)*, pp. 316–322, 2026.
3. Ultralytics, Ultralytics YOLO11, <https://docs.ultralytics.com/models/yolo11/> (accessed July 13, 2026).
4. Jan Held, Anthony Cioppa, Silvio Giancola, Abdullah Hamdi, Bernard Ghanem, Marc Van Droogenbroeck; VARS: Video Assistant Referee System for Automated Soccer Decision Making from Multiple Views, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2023, pp. 5086–5097.
5. B. S. Bal and G. Dureja, Hawk eye: A logical innovative technology use in sports for effective decision making, *Sport Sci. Rev.*, vol. 21, nos. 1–2, Apr (2012), pp. 107–119.
6. Y. Chen, S. Cai, Y. Wang, and J. Yan, Monocular visual analysis for electronic line calling of tennis games, in *Image and Graphics, ICIG 2021, Lecture Notes in Computer Science*, vol. 12888, Springer, Cham, 2021, pp. 528–538.
7. C.-H. Hung, A research of automatic and real-time table tennis fault serve detection system, *Sports*, vol. 6, no. 4, (2018), art. 158.
8. C.-H. Hung, Image judgment auxiliary system for table tennis umpiring under low light conditions, *Smart Sci.*, vol. 7, no. 1, (2019), pp. 39–46.
9. X. Li and L. Guo, Motion object detection model for electronic referee scoring in table tennis events, *PLOS ONE*, vol. 20, no. 3, 2025.
10. P. Wong, H. Myint, L. Dooley, and A. Hopgood, A multi-view automatic table tennis umpiring framework, *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, vol. 238, no. 4, 2024, pp. 339–351.
11. S. Gómez-González, Y. Nemmour, B. Schölkopf, and J. Peters, Reliable real-time ball tracking for robot table tennis, *Robotics*, vol. 8, no. 4, 2019.
12. K. Lu, J. Zhai, J. Wang, Z. Cheng, Y. He, and Z. Liao, Self-powered acceleration sensors arrayed by swarm intelligence for table tennis umpiring system, *PLOS ONE*, vol. 17, no. 10, 2022.
13. Rallys Editorial Department, "Could AI revolutionize table tennis officiating? Development of 'Daiwa 1mm,' an AI-based edge classification support system," *Rallys*, May 27, 2023,

https://rallys.online/read/230526_kanazawaport_daiwa/ (accessed July 13, 2026)

14. Y. Fujihara, X. Kong, A. Tanaka, H. Nishikawa, and H. Tomiyama, In/Out judgement by ball tracking in table tennis, in Proceedings of the 5th International Symposium on Advanced Technologies and Applications in the Internet of Things (ATAIT 2023), CEUR Workshop Proceedings, vol. 3459, 2023, pp. 115–126.
15. K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2961–2969, 2017.
16. Y. Liu and X. Sun, Design and image research of tennis line examination based on machine vision analysis, Computational Intelligence and Neuroscience, 2021.
17. S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, Advances in Neural Information Processing Systems, vol. 28, 2015.
18. R. Voelkov, N. Falaleev, and R. Baikulov, TTNNet: Real-time temporal and spatial video analysis of table tennis, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020.
19. J. Giménez-Gallego, J. Martínez-del-Rincon, J. D. Gonzalez-Teruel, H. Navarro-Hellín, P. J. Navarro, and R. Torres-Sanchez, “On-tree fruit image segmentation comparing Mask R-CNN and Vision Transformer models: Application in a novel algorithm for pixel-based fruit size estimation,” Computers and Electronics in Agriculture, vol. 222, Article 109077, 2024.
20. J. Gené-Mola, M. Ferrer-Ferrer, E. Gregorio, P. M. Blok, J. Hemming, J.-R. Morros, J. R. Rosell-Polo, V. Vilaplana, and J. Ruiz-Hidalgo, “Looking behind occlusions: A research on amodal segmentation for robust on-tree apple fruit size estimation,” Computers and Electronics in Agriculture, vol. 209, Article 107854, 2023.
21. C. Zhao, C. Li, X. Wang, X. Wu, Y. Du, H. Chai, T. Cai, H. Xiang, and Y. Jiao, “Plant Disease Segmentation Networks for Fast Automatic Severity Estimation Under Natural Field Scenarios,” Agriculture, vol. 15, Article 583, 2025.
22. Z. Li, C. Yin, and X. Zhang, “Crack Segmentation Extraction and Parameter Calculation of Asphalt Pavement Based on Image Processing,” Sensors, vol. 23, Article 9161, 2023.
23. DITECT Co., Ltd., “HAS-EF,” https://www.ditect.co.jp/products/camera/has_EF.html, (accessed July 13, 2026).
24. JOOLA, “iPong Pro,” <https://www.ipong.net/joomla/index.php/ipong-models/ipong-pro>, (accessed July 13, 2026).
25. GoPro, Inc., “HERO12 Black,” <https://gopro.com/ja/jp/shop/cameras/hero12-black/CHDHX-121-master.html>, (accessed July 13, 2026).

Authors Introduction

ASST. Prof. Kazuma Sakamoto



He received the B.S., M.S., and Ph.D. degrees in Informatics from Kansai University, Japan, in 2015, 2018, and 2021, respectively. Since 2021, he has been an assistant professor at the Faculty of Production Systems Engineering and Sciences, Komatsu University, Japan. His research interests include web mining, natural language processing, image processing, and sports informatics.

Mr. Reo Ishii



He received the B.S. degree in Production Systems Engineering and Sciences from Komatsu University, Japan, in 2026, where he is currently pursuing the M.S. degree at the Graduate School of Sustainable Systems Science. His research interests include computer vision and pattern recognition.

Mr. Iori Iwata



He received the M.S. degree from the Graduate School of Sustainable Systems Science at Komatsu University, Japan, in 2024, where he is currently pursuing the Ph.D. degree at the Graduate School of Sustainable Systems Science. His research interests include image recognition, tactical analysis, and sports informatics.

Mr. Riku Kaiba



He received the B.S. degree in Production Systems Engineering and Sciences from Komatsu University, Japan, in 2024, where he also received his M.S. degree from the Graduate School of Sustainable Systems Science in 2026.

Prof. Yoshihiro Ueda



He received the Ph.D. degree in Mathematical Information Science from Kanazawa University, Japan, in 2001. Since 2021, he has been a professor at the Faculty of Production Systems Engineering and Sciences, Komatsu University, Japan. His research interests include productivity improvement, labor saving, automation using data science, and human centric system development.