Tac-Valuer: Knowledge-based Stroke Evaluation in Table Tennis

Jiachen Wang*
State Key Lab of CAD&CG, Zhejiang University and Zhejiang Lab
wangjiachen@zju.edu.cn

Dazhen Deng*
State Key Lab of CAD&CG, Zhejiang University
dengdazhen@zju.edu.cn

Xiao Xie
State Key Lab of CAD&CG, Zhejiang University
xxie@zju.edu.cn

Xinhuan Shu
Department of Computer Science and Engineering, The Hong Kong University of Science and Technology
xinhuan.shu@connect.ust.hk

Yu-Xuan Huang, Le-Wen Cai
National Key Laboratory for Novel Software Technology, Nanjing University
{huangyx,cailw}@lamda.nju.edu.cn

Hui Zhang
Department of Sports Science, Zhejiang University
zhang_hui@zju.edu.cn

Min-Ling Zhang
School of Computer Science and Engineering, Southeast University
zhangml@seu.edu.cn

Zhi-Hua Zhou
National Key Laboratory for Novel Software Technology, Nanjing University
zhouzh@lamda.nju.edu.cn

Yingcai Wu†
State Key Lab of CAD&CG, Zhejiang University and Zhejiang Lab
ycwu@zju.edu.cn

ABSTRACT

Stroke evaluation is critical for coaches to evaluate players' performance in table tennis matches. However, current methods highly demand proficient knowledge in table tennis and are time-consuming. We collaborate with the Chinese national table tennis team and propose Tac-Valuer, an automatic stroke evaluation framework for analysts in table tennis teams. In particular, to integrate analysts' knowledge into the machine learning model, we employ the latest effective framework named abductive learning, showing promising performance. Based on abductive learning, Tac-Valuer combines the state-of-the-art computer vision algorithms to extract and embed stroke features for evaluation. We evaluate the design choices of the approach and present Tac-Valuer's usability through use cases that analyze the performance of the top table tennis players in world-class events.

CCS CONCEPTS

• **Computing methodologies** → **Knowledge representation and reasoning**; *Activity recognition and understanding*; • **Human-centered computing** → *Visual analytics.*

KEYWORDS

Table tennis, stroke evaluation, abductive learning

*Both authors contributed equally to this research.
†Yingcai Wu is the corresponding author.

KDD '21, August 14–18, 2021, Virtual Event, Singapore

ACM ISBN 978-1-4503-8332-5/21/08…$15.00
https://doi.org/10.1145/3447548.3467104

ACM Reference Format:
Jiachen Wang, Dazhen Deng, Xiao Xie, Xinhuan Shu, Yu-Xuan Huang, Le-Wen Cai, Hui Zhang, Min-Ling Zhang, Zhi-Hua Zhou, and Yingcai Wu. 2021. Tac-Valuer: Knowledge-based Stroke Evaluation in Table Tennis. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), August 14–18, 2021, Virtual Event, Singapore.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3447548.3467104

1 INTRODUCTION

Table tennis is a highly confrontational racket sport, where players from both sides strike the ball (the action that a player strikes the ball once is a stroke) by turn until one wins this rally. Players' techniques (patterns of single stroke) and tactics (patterns among consecutive strokes) are always the focal points of analysis [9]. Thus, many national table tennis teams adopt data analysis methods to facilitate technical and tactical analysis. In recent four years, we have worked closely with the Chinese national table tennis team, one of the top table tennis teams worldwide, and developed a data platform. It provides pre-match and post-match data support services for the team, including interactive tools for data collection [8], correlation exploration [36], and match simulation [29]. Till now, the platform has provided data support for the national team in more than 20 world-class events, e.g., World Cup, World Tour, and World Championships. However, the platform still falls short of the strong demand from coaches for the stroke evaluation. In table tennis, stroke evaluation is crucial for evaluating players' performance since the stroke is the basic unit for analysis [18]. For example, analysts evaluate a player's performance of a particular technique (e.g., topspin) by investigating whether this player can win an advantage after performing the stroke with this technique.

Methods for stroke evaluation can be divided into video-driven ones and data-driven ones. Initially, video-driven methods are straightforward and widely adopted by analysts. Analysts organize stroke videos in various ways and repeatedly examine the videos to evaluate the strokes' performances based on their knowledge [9]. Data-driven methods require analysts to collect the key stroke attributes from videos in advance. Then analysts use effective statistical indicators [34, 35] and simulation models [18, 29, 33] to evaluate strokes' performance. While these methods have helped discover in-depth insights, they are tedious and time-consuming because they rely on analysts' knowledge for video checking and data collection.

A possible solution is to train a machine learning model to extract stroke features and evaluate strokes based on the features. Training such a model requires a large number of labeled data. However, labeled data is insufficient since qualified analysts for data labeling are limited, and the labeling process is time-consuming. Instead, we have considerable unlabeled data and collection of analysts' domain knowledge. Therefore, it is highly desirable to leverage analysts' knowledge for automatic feature extraction and stroke evaluation based on limited labeled data and plentiful unlabeled data.

The challenge of proposing such a model lies in combining the analysts' knowledge and the machine learning model. It is difficult to inject symbolic knowledge into the optimization of numerical values in machine learning models. To solve this challenge, we refer to the latest effective AI framework, **AB**ductive **L**earning (ABL) [4, 37]. Unlike other frameworks (e.g., probabilistic logic program [5], statistical relational learning [14]) that make one side subsume to the other, **ABL** integrates machine learning and logical reasoning (driven by domain knowledge) in a mutually beneficial manner [37]. Furthermore, we propose a knowledge-based framework, Tac-Valuer based on **ABL** for stroke evaluation.

Tac-Valuer takes stroke videos as input, embed stroke features, and output the evaluation scores of the strokes. It consists of three steps, namely, video formalizing (**VF**), stroke embedding (**SE**), and performance rating (**PR**). Among the components, **SE** uses **ABL** to extracts and embeds the stroke features from videos into feature vectors based on analysts' knowledge. It contains an attribute recognition component (driven by neural networks) and a logical reasoning component (driven by rules of inference). The interaction between the two components can help **SE** achieve decent embedding effects with limited labeled data. The contributions of this work are as follows.

- We are the first to address the problem of automatic stroke evaluation in table tennis, benefiting table tennis analysts.
- We propose a novel framework that leverages **ABL** to achieve decent stroke evaluation results based on analysts' knowledge and limited labeled data.
- We implement our framework and conduct use cases on matches between top table tennis players. Through the cases, we obtain valuable insights into players' strokes.

2 RELATED WORK

In this section, we review the works related to performance evaluation in sports and action recognition with domain knowledge.

2.1 Performance evaluation in sports

In table tennis, researchers have introduced many effective stroke evaluation approaches [9], which have been discussed in Section 1. Therefore, this section mainly discusses performance evaluation in other sports. In soccer, both Decroos et al. [6, 7] and Bransen et al. [1] have proposed automatic methods for valuing players' actions during matches. In basketball, Cervone et al. [2] propose a framework for valuing each moment of possession, and Sicilia et al. [24] construct a deep learning architecture to evaluate players' actions within a possession. Additionally, many similar studies have also been conducted in tennis [30–32] and ice hockey [16, 23]. These studies mainly evaluate players' actions based on position tracking data and game context data (e.g., events and box scores). While these studies are effective, they cannot be applied to stroke evaluation in table tennis for two reasons. First, the gaps between different competition rules prevent most of the aforementioned approaches, especially those for team games like ice hockey, soccer, and basketball. This gap leads to different evaluation metrics. For example, in soccer, when analysts evaluate a pass, they need to focus on the whole possession containing this pass to examine the collaboration among teammates and defense of the opponent team. However, in table tennis, there is no collaboration but confrontation in a stroke. Analysts evaluate a stroke by examining whether the stroke makes the striker at an advantage or not. Second, existing studies mainly focus on position tracking data and game context data, not considering players' motions. Therefore, a new method is required for table tennis to utilize motion data for stroke evaluation.

2.2 Action recognition with knowledge

Tac-Valuer extracts and embeds stroke features from videos by using action recognition techniques. Video-based action recognition has been a popular topic in computer vision [10, 19]. Considerable techniques have been developed to improve the efficiency and accuracy of the action recognition tasks [12, 25, 28]. To further enhance the performance of the techniques, various datasets in different fields (e.g., film, sports, music) have been introduced [15, 22, 27]. However, these techniques do not contain enough human knowledge in the machine learning models, which limits their performance and interpretability. To solve this issue, many AI frameworks, such as probabilistic logic program [5] and statistical relational learning [14], have provided solution alternatives. However, these frameworks make knowledge subsume to machine learning or vice versa. Therefore, we decide to use **ABL** to improve the state-of-the-art action recognition techniques with domain knowledge for data collection. According to existing works [4, 37], **ABL** can combine machine learning and logical reasoning with knowledge in a mutually beneficial way. Moreover, it also performs well in semi-supervised fashion, which further improves our proposal since high quality labeled table tennis data is not sufficient for supervised learning.

3 BACKGROUND AND DATA

In this section, we introduce our collaboration with the Chinese national table tennis team and the data we use in Tac-Valuer.

3.1 Collaboration with the national team

We have collaborated with the Chinese national table tennis team for four years. During the collaboration, we developed a data platform for the national team to empower the analysts in the team to collect and analyze match data (Figure 1). The platform consists of three components, database, data collection, and data analysis. The database enables the analysts to store data, such as broadcast videos, player information, and match meta information for table tennis

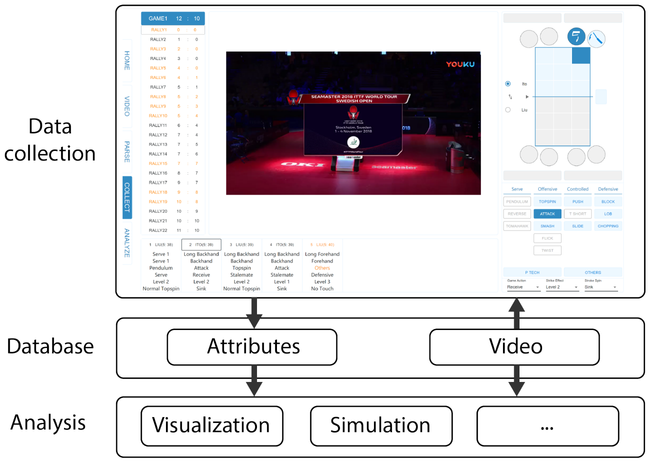

Figure 1: The structure of the data platform. It consists of three components, namely, database, data collection, and data analysis. The arrows present the data transmission in the platform.

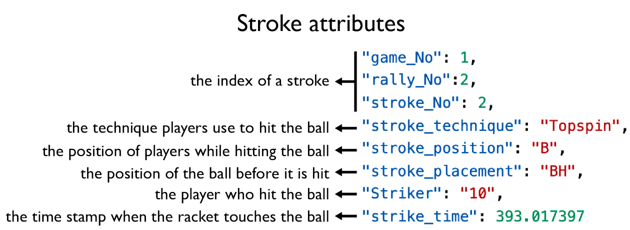

Figure 2: An example of a stroke's attributes. These attributes are manually collected by analysts.

matches with national team players involved. The data collection leverages an interactive data collection framework, EventAnchor [8], to reduce analysts' efforts on data retrieval. The data analysis supports various analysis techniques, such as visualization and simulation. The visualization part provides a comprehensive visual analytics tool, iTTVis [36], to help analysts investigate the collected data to gain new insights. The simulation part, Tac-Simur, uses a hybrid second-order Markov chain model [29] to simulate the matches against particular players to help coaches improve the training plans and playing strategies. Till now, the platform has served the teams in more than 20 top events in the world, which provides a good opportunity to design an efficient video-based framework for automatic stroke evaluation.

3.2 Data description

Tac-Valuer requires four parts of data for model training, namely, stroke videos, stroke attributes, rules of inference, and performance scores. The stroke videos are the input of Tac-Valuer. Each stroke video contains a sequence of eight frames, including a central frame when the player's racket touches the ball, four frames before this frame, and three frames after this frame. We labeled the timestamp of the central frame through crowdsourcing since labeling this data did not require proficient knowledge. We collected 9522 stroke videos from 21 matches.

The stroke attributes are used as training labels of stroke videos for the models in **SE**. In table tennis, players use rackets to strike the ball by turn. The basic observation unit in table tennis is a stroke, namely, the action that a player hits the ball once [18]. Since the recognition of stroke attributes demanded knowledge in table tennis, we asked analysts to collect this data. The data was collected through the platform [8]. Analysts needed to watch the stroke videos collected before and annotate the specific stroke attributes. Figure 2 displays an example of the stroke attribute data. The first three attributes are used to identify the temporal position of the stroke within a match. The subsequent three attributes (i.e., stroke technique, stroke position, and stroke placement) characterize the technical features of a stroke. The last attribute is derived based on the game rules. We collected the stroke attributes of all the 9522 stroke videos mentioned above.

The rules of inference are used as the logical reasoning component of **ABL** to improve the performance of the models in **SE**. We summarized analysts' knowledge during data collection and quantified their knowledge in the manner as follows.

$$\begin{aligned} Pre\_Tech(X, \text{"Block"}) \quad &\rightarrow \quad \neg Cur\_Tech(X, \text{"Push"}) \\ &\wedge \quad \neg Cur\_Tech(X, \text{"Short"}) \\ &\wedge \quad \neg Cur\_Tech(X, \text{"Slide"}) \\ &\wedge \quad \neg Cur\_Tech(X, \text{"Block"}), \end{aligned}$$

where "$\neg$" is negation, "$\wedge$" is conjunction, and "$\rightarrow$" is implication. This example rule means if the technique of the previous stroke is "Block", then the technique of the current stroke is impossible to be "Push", "Short", "Slide", and "Block". We collected 28 rules.

The performance scores are used as training labels of stroke videos for the model in **PR**. The performance scores were manually rated by professional analysts. We hired seven analysts working for the professional table tennis teams to evaluate the performance of players' strokes in the stroke videos. We let them rate the strokes into two levels, namely, 1 for advantaged strokes and -1 for disadvantaged ones. Each stroke was rated by three different analysts. Thus, the score of each stroke was decided according to majority voting. For the strokes rated differently by the three analysts, we asked another senior analyst to re-evaluate these strokes. The senior analyst has worked for the Chinese national table tennis team for more than five years and has rich experience in stroke evaluation. We collected the scores of all the 9522 stroke videos.

4 FRAMEWORK

In this section, we first define the problem of stroke evaluation in this work and present an overview of Tac-Valuer. Then, we introduce the implementation details of Tac-Valuer.

4.1 Problem definition

Table tennis is a racket sport that requires players from both sides to interact with each other's stroke by turn. We denote a rally as $R = \{S_1, S_2, \ldots, S_i\}$, where S_i represents the video frames of the i^{th} stroke in the rally R. Then, the problem of stroke evaluation in

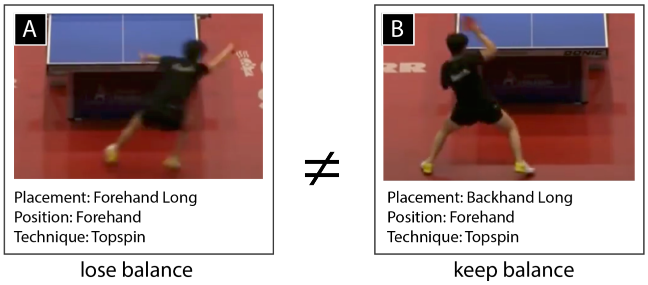

Figure 3: A player performs the strokes with similar technical attributes in different motions. The player loses balance in A and keep balance in B, which results in different performance of these two strokes.

table tennis can be defined as follows.

$$Score_i = F(S_i) \tag{1}$$

where F denotes the evaluation function and $Score_i$ represents the performance of the stroke. The last stroke of a rally is evaluated as poor by default since it belongs to the loser of this rally in our data.

A straightforward alternative to F is training an end-to-end model for stroke evaluation. Specifically, the model takes the stroke videos as input and directly outputs the evaluation scores. However, an end-to-end model lacks interpretability, which is not accepted by analysts. To yield an interpretable evaluation result, we decided to imitate analysts' evaluation process when constructing F. Generally, analysts often need to identify the stroke attributes. Besides, they also have to consider the stroke motions. Specifically, for example, in Figure 3, the effects of the two strokes hit by topspin in forehand are quite different according to the motion features of the two players. However, from the perspective of attributes, these two strokes are similar strokes since the technique and position of both strokes are identical (Figure 3). With these two kinds of data, analysts can evaluate a stroke comprehensively. Therefore, the function F needs first to extract stroke attributes and stroke motions from videos, and then, evaluates the strokes based on them. The stroke attributes are primarily depicted by stroke technique, stroke placement, and stroke position according to existing studies [9]. The stroke motions are depicted by visual features of video frames. Based on these considerations, we propose Tac-Valuer, a knowledge-based stroke evaluation framework.

4.2 Framework overview

Tac-Valuer takes the stroke videos within a rally as input, extracts and embeds stroke features from videos, and output the evaluation scores of these strokes. To accurately extract and embed the stroke features with limited labeled data, we use **ABL** to optimize pseudo-labels of unlabeled data to enhance model performance. The challenges during constructing Tac-Valuer are as follows.

- **C1: How to obtain effective stroke videos?** The stroke videos contain noise within each frame. The noise is often brought by irrelevant persons on the court. In a table tennis match video, apart from the players, the referees and other staff, such as caddies, can appear in the scene (Figure 5). The occurrence of these persons can decline the performance of the models in **SE** and **PR**.
- **C2: How to effectively extract and embed stroke features with limited labeled data?** Labeled stroke attribute data is limited due to its high demand for proficient domain knowledge. Training an effective model for extracting stroke attributes with limited labeled data is challenging. Moreover, the stroke attributes have been well defined by existing works in table tennis analysis as shown in Figure 2. However, stroke motions for stroke evaluation have not been extensively studied. Extracting stroke motions from the stroke videos and embed them with stroke attributes for evaluation is also challenging.
- **C3: How to quantify the performance of a stroke?** Current studies in table tennis mostly use the scoring rate for quantified analysis. However, this method is not rigorous since whether the player can score in this rally is related to all the strokes in this rally. A more accurate way is to ask analysts to evaluate the performance of a stroke according to the given stroke video. However, their evaluation criteria are based on their knowledge and hard to be verbalized.

Given the challenges above, we constructed Tac-Valuer into three steps (Figure 4) as follows.

- **Step 1: video formalizing (VF)**. We remove the irrelevant persons in each video frame to address **C1**. We use an object detection algorithm to detect the two players in each frame. Then we add masks to other irrelevant persons to remove their influence (Figure 4A).
- **Step 2: stroke embedding (SE)**. We leverage **ABL** to extract and embed the stroke attributes with stroke motions with limited labeled data (**C2**). As Figure 4B shows, we first use an attribute recognition component to embed strokes into feature vectors. Then we introduce a logical reasoning component constructed based on analysts' knowledge to improve pseudo-labels of unlabeled data. These labels are further used to improve the performance of the attribute recognition component iteratively.
- **Step 3: performance rating (PR)**. We train a classifier that learns analysts' evaluation in video-based methods to obtain quantified evaluation results (**C3**). We classify analysts' evaluation of strokes into two levels as training labels and train the model to classify the level of each stroke based on the embedding vectors in Step 2 (Figure 4C).

4.3 Video formalizing

VF takes stroke videos as input and attaches a mask to remove irrelevant persons for each frame (Figure 4A). We use Faster R-CNN [21], to detect the bounding boxes of the table and all persons in each frame. According to the relative position between the box of the table and boxes of the persons, we identify the two players in each frame. For the boxes of other persons, we use a mask to cover them. We reserve both players for two reasons. First, analysts need to investigate both players' motions. For example, when player A is performing a stroke, the other player B will anticipate the stroke and move his/her body and racket in advance. If player A's motion misleads the anticipation of player B, this stroke is usually

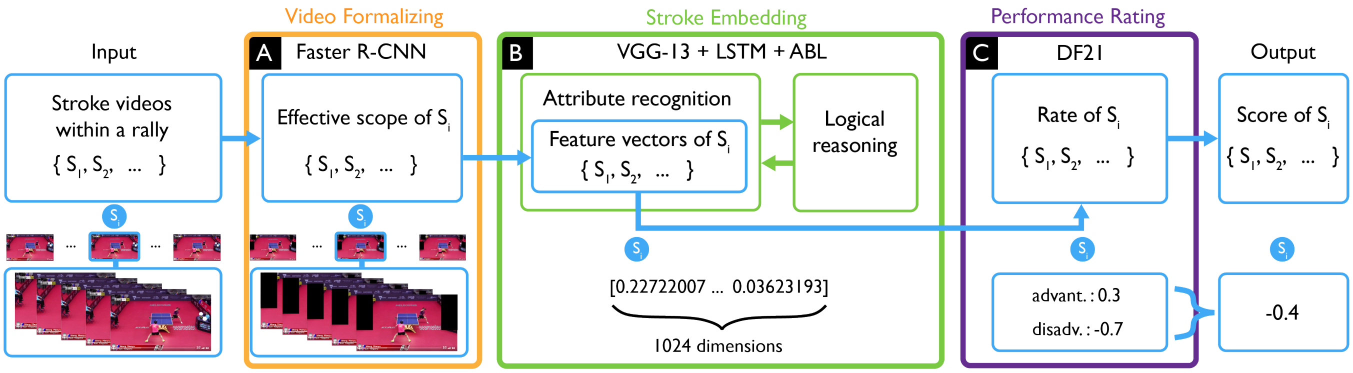

Figure 4: The structure of the framework. The input is the video frames of strokes within a whole rally. The output is the score of each stroke. The orange, green, and purple boxes represent the three steps and the blue ones represent the data flowing among different steps. Video formalizing (A) applies masks to frames to remove irrelevant persons. Stroke embedding (B) extracts and embeds stroke features from videos by using ABL. Performance rating (C) evaluates the performance of each stroke based on the feature vectors. The contents under each blue box present the example of the data format.

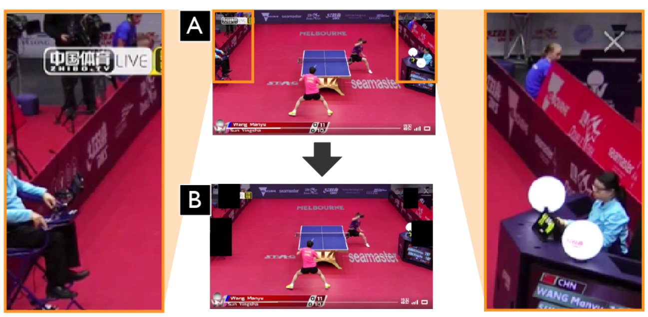

Figure 5: The details of video formalizing. (A) is the original frame with referees and staff. (B) is the processed frame where the irrelevant persons are removed by masks.

performed with high quality. Second, the motions of two players can provide more features to improve the training accuracy of stroke embedding in **SE**.

4.4 Stroke embedding

Given stroke videos, **SE** extracts the stroke features and embeds them into feature vectors. This step is based on Abductive Learning (**ABL**) and contains two components, an attribute recognition component (Figure 6A) and a logical abduction component (Figure 6B).

Attribute Recognition Component. In the attribute recognition component, the input data are stroke videos of a whole rally S_i. We first use VGG-13 [26] and LSTM [11] to extract feature vectors FV_i of each S_i. Then the feature vectors FV_i are fed to dense layers and softmax functions to get the result A_i, which contains the three technical attributes (i.e., stroke technique, stroke position, and stroke placement) (Figure 2). This component is pre-trained by labeled stroke videos. Due to limited labeled data, it may produce incorrect pseudo-labels A_i for the unlabeled stroke videos.

Logical Abduction Component. The logical abduction component contains domain knowledge rules of table tennis. Given the

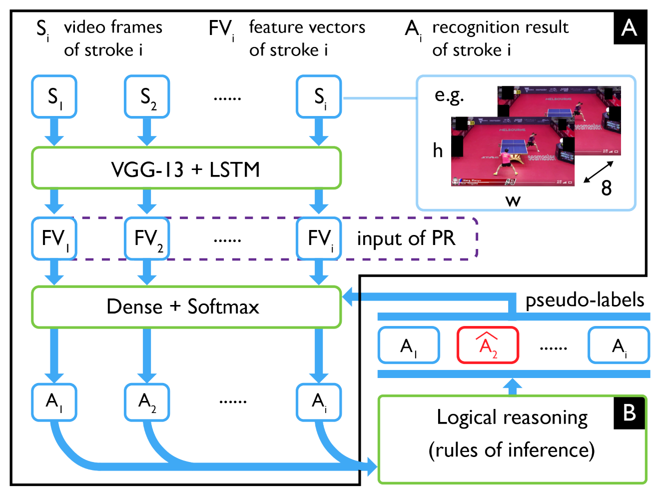

Figure 6: The structure of stroke embedding. (A) is an attribute recognition component and (B) is a logical reasoning component. The input of (A) is the unlabeled formalized stroke videos (S_i) within a whole rally. The recognized attributes (A_i) are taken as pseudo-labels of the stroke videos and input into (B). (B) revises the labels and transfers them back to (A) to improve the performance of (A) iteratively. The feature vectors (FV_i) are the final input for performance rating.

pseudo-labels (technical attributes) of the attribute recognition component, which may be incorrect, it first checks whether these pseudo-labels are consistent with the rules. If the pseudo-labels are consistent with domain knowledge, it returns them without any revision. Otherwise, it explores logical abduction to revise the pseudo-labels. The revision is conducted under the principle of minimal inconsistency [4, 37] between the pseudo-labels and domain knowledge, aiming at correcting the most likely inaccurate pseudo-labels. Specifically, logical abduction assumes that some

Algorithm 1 Abductive Learning in Tac-Valuer

Input: Labeled videos (S_l, A_l); Unlabeled videos S_u; Knowledge base KB
Parameter: Epoch E
Output: Model M

1: $M \leftarrow TrainModel(S_l, A_l)$ # Pre-train model
2: **for** $e := 1 to E$ **do**
3: $\widehat{A}_u \leftarrow M(S_u)$ # Generate pseudo-labels $\widehat{A}_u$
4: $\overline{A}_u \leftarrow Abduce(KB, \widehat{A}_u)$ # Revise pseudo-labels
5: $M \leftarrow Train(M, S_l, A_l, S_u, \overline{A}_u)$ # Update M
6: **end for**
7: **return** M

pseudo-labels are incorrect and makes them "unknown" (abducible), while other pseudo-labels are fixed. Then the abduction module will abduce the most compatible labels and use them to replace the original pseudo-labels so that the revised labels are consistent with domain knowledge.

Abductive Learning (ABL). By **ABL** [4, 37], the logical abduction component leverages the knowledge base (summarized rules of inference) of table tennis and unlabeled data to improve the attribute recognition component's performance. An outline of our learning algorithm is shown in Algorithm 1. Given labeled stroke videos, unlabeled stroke videos and domain knowledge, it first uses labeled stroke videos to pre-train the deep learning models in the attribute recognition component. The pre-trained model generates pseudo-labels of unlabeled stroke videos, which may contain mistakes. Then the logical abduction component tries to revise the pseudo-labels based on domain knowledge rules. Finally, the revised pseudo-labels will be treated as ground-truth labels, and transferred back to the attribute recognition component for training. The model's performance is improved by repeating the above circulation.

4.5 Performance rating

With the feature vectors, **PR** uses DF21[1] (an efficient implementation of Deep Forest [38, 39]) as a classifier to rate the performance of a stroke. After training, the classifier can predict the probabilities of the performance, (i.e., $p_{i,ad}$ for the probability of advantaged performance and $p_{i,dis}$ for disadvantaged performance) of S_i. In this way, the final score, $Score_i$ can be defined by the expectation value of the level as follows.

$$Score_i = p_{i,ad} * 1 + p_{i,dis} * (-1) \tag{2}$$

5 EVALUATION

In this section, we evaluate our framework from two aspects. First, we conduct two experiments to justify the design decisions of Tac-Valuer. Then, we present two use cases based on the implementation of the framework. We conduct the cases with analysts in the national table tennis team and discover valuable insights.

5.1 Evaluation of design decisions

We evaluate the design choices on the critical aspect of **VF** and **PR** by comparing the performance of our methods with alternatives.

[1]https://www.lamda.nju.edu.cn/deep-forest/

Table 1: The mAP of the experiments in SE

method	label	tec	pla	pos	striker
supervised	50%	0.66	0.56	0.75	0.79
ABL	50%	**0.69**	**0.66**	**0.79**	**0.86**
supervised	100%	0.71	0.64	0.8	0.91

Table 2: The result of the experiment in PR

	FC	SVM	RF	DF21	XGBoost	CatBoost	LightGBM
Avg.	82.02	84.90	84.83	**85.16**	85.11	84.97	84.41
Std.	**1.51**	1.63	1.71	1.81	2.21	1.75	1.75

*5.1.1 Choice of **ABL**.* In **SE**, we used **ABL** to construct a semi-supervised model due to the limited labeled data. To validate the effectiveness of **ABL**, we compared **ABL** with the other two supervised conditions in this experiment. First, we used 50% labeled data with 50% unlabeled data to train the attribute recognition component with **ABL**. Then, we used 50% labeled data and 100% labeled data to train the attribute recognition component without **ABL** separately. The result of the comparison is displayed in Table 1. We added one more attribute, "striker," to evaluate the model performance because each stroke video contains two players and the model needs to figure out which player is the striker. Among the three conditions, **ABL** outperforms the other two conditions in recognizing stroke placement. For other attributes, **ABL** outperforms the supervised one with only 50% labeled data and performs nearly as well as the supervised one with 100% labeled data.

5.1.2 Choice of DF21. In **PR**, we needed a model that can accurately determine the performance level. Therefore, we tested popular classification models, including a three-layer fully connected neural network (FC), SVM [17], Random Forest (RF) [17], DF21 [38, 39], XGBoost [3], CatBoost [20], and LightGBM [13]. We used 10-fold cross validation. The result is displayed in Table 2. Although FC has the lowest standard variance, the average accuracy of DF21 (85.16%) is the highest during cross validation. We finally chose DF21 as the model in for performance rating.

5.2 Use Cases

In this section, we present three cases we conducted with two professional analysts to demonstrate the effectiveness of Tac-Valuer. Both of the two analysts have worked for the Chinese table tennis team for more than five years. They did not participate in the collection of performance scores in this work. We mainly analyzed Ito Mima, a strong opponent of the Chinese table tennis players. In the first case, we compared the evaluation of Tac-Valuer with those of traditional indicators and analysts to demonstrate the reliability and advantage of Tac-Valuer. In the second case, we focused on Ito's performance when she played against different opponents.

5.2.1 Case 1: Ito's push is not good enough. The match in this case is the final of women singles in *2019 ITTF World Tour, Swedish Open* between Ito and Chen Meng. Current methods (e.g., the three-phase

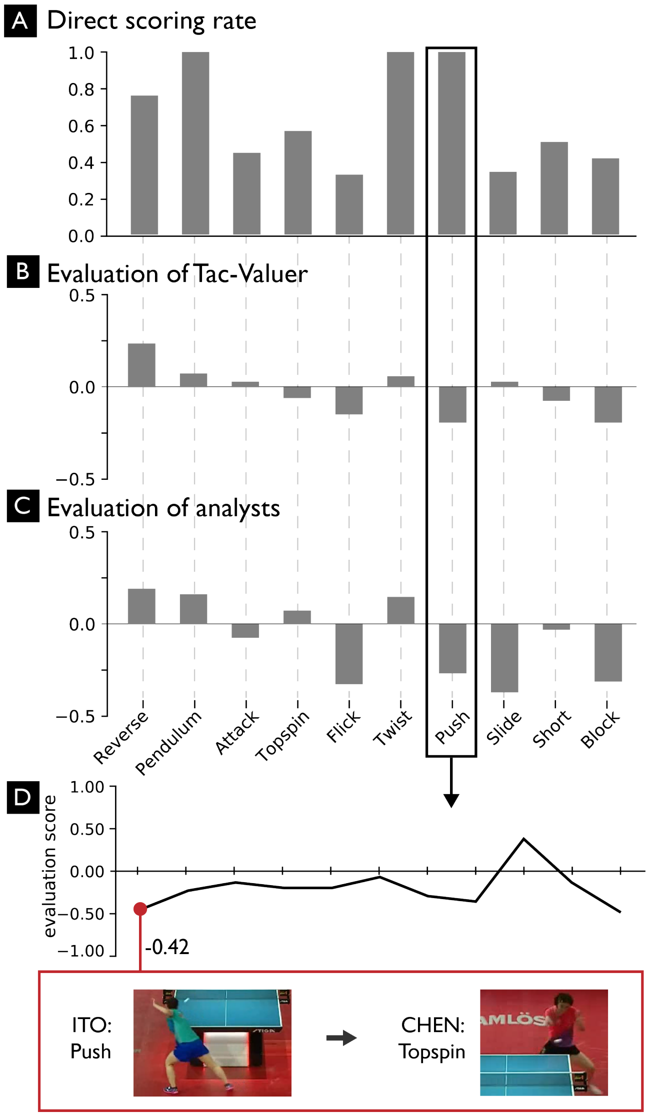

Figure 7: Case 1: (A), (B), (C) display the performance of Ito's technique evaluated by the direct scoring rates, Tac-Valuer, and analysts, respectively. (D) shows the evaluation scores of all strokes performed with push.

method [34, 35]) mostly use the direct scoring rates as an important indicator for stroke evaluation. In a table tennis match, the direct scoring rate of Player B's A-type strokes is computed as follows:

$$R = N_{win}/(N_{win} + N_{lose}) \tag{3}$$

where N_{win} refers to the number of rallies where Player B performs an A-type stroke, and his opponent fails to receive it, and N_{lose} refers to the number of rallies where Player B fails to receive his opponent's stroke using A-type stroke. In this way, it considers the A-type strokes that directly lead to the results of rallies. However, the indicator omits many strokes that can potentially affect the results. Tac-Valuer has solved this limitation.

We analyzed Ito's techniques since she uses pimpled rubber, which other players rarely use. Figure 7A, B, and C display the scoring rates, evaluation of Tac-Valuer, and evaluation of analysts of Ito's techniques, respectively. Analysts' evaluation was collected by other analysts in advance. According to Figure 7B, C, the evaluation results of Tac-Valuer almost conformed to those of analysts. However, the results of the direct scoring rate were quite different, especially the result of "Push". We examined the result of each stroke performed by "Push" (Figure 7D). Most of these strokes' scores are below 0. We examined the first stroke with the lowest evaluation score (-0.42). This stroke is the third stroke of the rally. Ito served in this rally, however, she did not take preemptive actions. Instead, she used "Push" to try to control the ball. After her "Push", Chen seized the chance to attack first by using the offensive technique, "Topspin". In such a condition, Ito totally lost the advantage in this rally. Through this case, we concluded that Ito's push is not as good as the traditional indicators presented.

5.2.2 Case 2: Ito should reduce errors in matches. The matches here include the final, the semi-final, and the quarter-final of women singles in *2019 ITTF World Tour, Swedish Open* between Ito (purple) and Chen Meng, Sun Yingsha, and Wang Manyu (orange). We first examined the stroke performance from the perspective of stroke technique (Figure 8A). We found that when playing against Sun and Wang, Ito's technique performed better than her opponents. She had fewer techniques whose scores were substantially below zero, especially when she was against Wang. Her control techniques, "Push" and "Short" were much better than Wang's. However, when playing against Chen, Ito did not display many advantages, which may explain her losing the final. We further examined the stroke performance of this match in Figure 8B.

We sum the evaluation scores of a player's strokes within a rally as the evaluation score of this player in the rally. Then, we visualize the difference between the summed scores of two players within a rally in a customized bar chart (Figure 8A). The height of the bars presents the absolute values of difference. Thus, if the evaluation score of Chen's strokes is large than that of Ito's strokes, the bar is encoded by orange and towards up, and vice versa. The dots in the middle of the chart indicate the winner of each rally. There are seven games in this match, and the numbers in orange and purple indicate the scores of the two players in each game.

We found some interesting rallies where the player whose strokes were better lost the rally. The most obvious one was the eighth rally of the second game. We investigated the strokes within this rally (Figure 8C). According to the bar chart, Ito performed all strokes well except the last one. This condition was rare and abnormal. To verify it, we examined the stroke video and found that Ito did perform the first three strokes well. After she served, she seized the chance to use topspin to attack first. However, her offense did not last for many strokes since she made a mistake when performing the last stroke, which was a pity for her.

6 DISCUSSION

Significance: We have solved a critical domain problem that heavily relies on analysts' labour and domain knowledge in an automatic manner for efficient match analysis. The significance of Tac-Valuer lies in four aspects. First, Tac-Valuer can benefit all analysts and

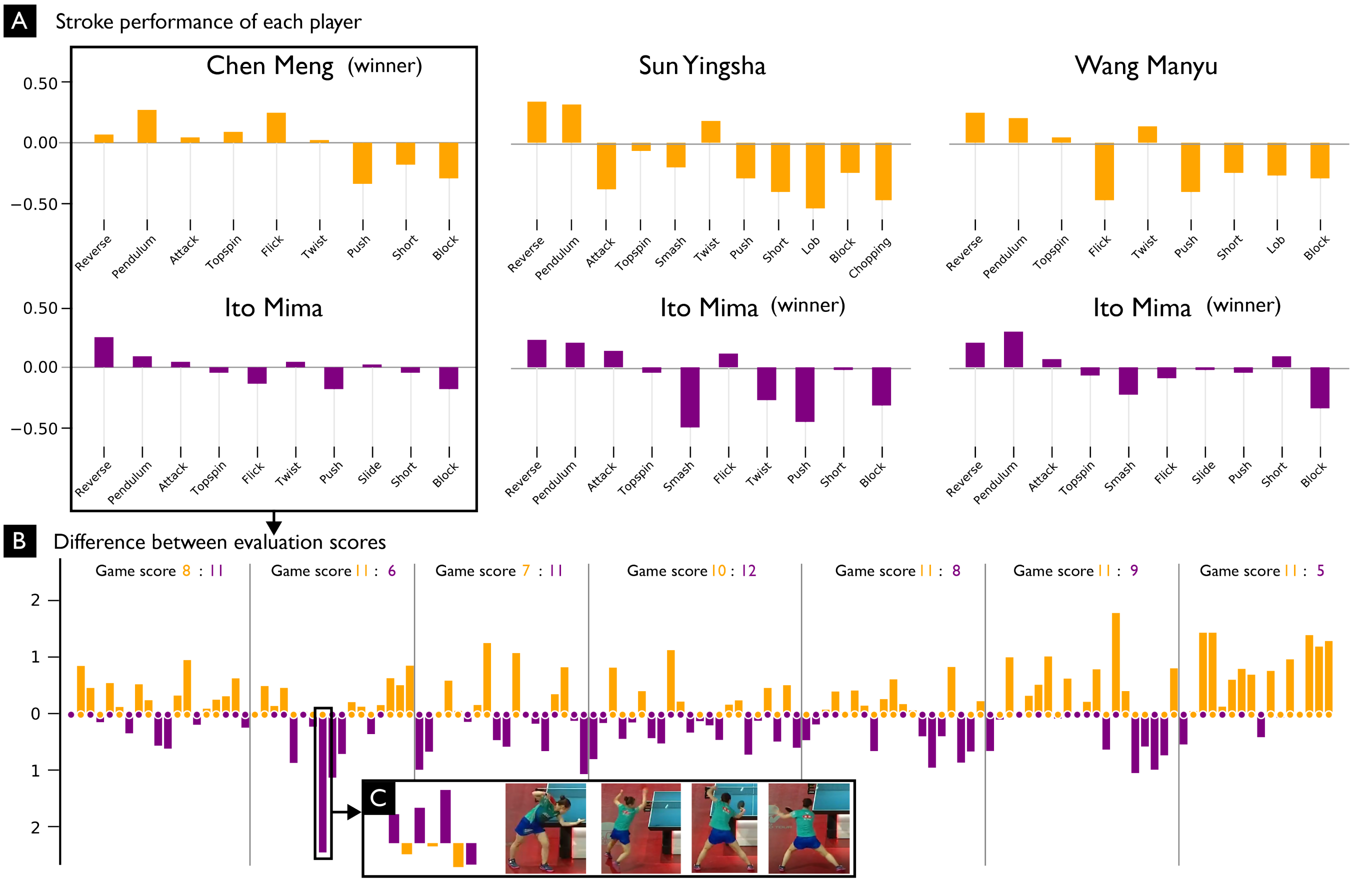

Figure 8: (A) presents the performance of different players in three matches. (B) visualizes summary difference between the evaluation scores of two players within each rally. (C) is a rally where Ito's performance is better than Chen's but Ito lost.

table tennis teams. It can automatically evaluate strokes without extra input or interactions. Second, Tac-Valuer can be implemented with a small amount of labeled data by using abductive learning. It combines machine learning models and analysts' knowledge to improve model performance with unlabeled data, alleviating analysts' burden on data annotation. Third, Tac-Valuer facilitates knowledge mining. The insights from the cases not only verify some empirical knowledge but also renew analysts' findings discovered by traditional methods. Fourth, Tac-Valuer is efficient enough for real-time analysis. It finishes the stroke evaluation of a match in half an hour, making real-time stroke evaluation becomes practical.

Insights: We obtain two insights. First, the interpretability of a machine learning model is important for domain experts and often the key for successful deployment. During our collaboration with the Chinese national table tennis team, the coaches and players were always concerned with the interpretability of our methods. They would not accept the model's results unless the results are interpretable. Second, a semi-supervised learning framework can benefit the domain. The application of machine learning models to a domain problem often requires a large amount of labeled data. The labeling process demands proficient domain knowledge and is time-consuming and error-prone. A semi-supervised learning framework can greatly alleviate the burden of data labeling.

Limitation: The lack of real-world high-quality videos is a major limitation of this work. The effectiveness of our work could largely depend on the quality of the training match videos. For instance, the low resolution and slow frame rate of existing broadcast match videos could affect the performance of stroke embedding. To address this issue, we have adopted data augmentation to extend our training data in video resolution, brightness, contrast, and rotation. In the future, we plan to extend our training dataset by incorporating more real-world high-quality videos from different sources.

Generalizability: Tac-Valuer is not limited to table tennis stroke evaluation. It can easily be extended to other similar racket sports, such as tennis and badminton by following our data format to prepare the training data and rules of inference. Moreover, given the use of **ABL**, analysts can achieve a decent result with limited labeled training data, which lowers the cost of extension.

7 CONCLUSION

In this work, we establish the problem of stroke evaluation in table tennis. To solve this problem, we introduce a knowledge-based framework, Tac-Valuer for automatic stroke evaluation. Tac-Valuer is the first attempt to extract and embed stroke features from videos based on analysts' knowledge for comprehensive stroke evaluation. Besides, Tac-Valuer enables decent model performance with limited labeled training data. We evaluate the usefulness and priority of the framework by two use cases and discover valuable insights approved by analysts in the Chinese national table tennis team.

In the future, we plan to improve the framework from two aspects. First, we will try to improve the robustness of the framework. We plan to expand the types of training data so that we can cover sufficient types of videos. Second, we plan to refer to the concept of causality and use causal models to construct the relationships among multiple strokes in the framework. Third, we plan to extend our framework to other similar sports like tennis.

REFERENCES

[1] Lotte Bransen, Pieter Robberechts, Jan Van Haaren, and Jesse Davi. 2019. Choke or Shine? Quantifying Soccer Players' Abilities to Perform Under Mental Pressure. In *MIT Sloan Sports Analytics Conference*.

[2] Dan Cervone, Alexander D'Amour, Luke Bornn, and Kirk Goldsberry. 2014. POINTWISE: Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data. In *Proceedings of MIT Sloan Sports Analytics Conference*.

[3] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 785–794.

[4] Wang-Zhou Dai, Qiuling Xu, Yang Yu, and Zhi-Hua Zhou. 2019. Bridging machine learning and logical reasoning by abductive learning. In *Advances in Neural Information Processing Systems 32*. 2811–2822.

[5] Luc De Raedt and Kristian Kersting. 2008. Probabilistic inductive logic programming. In *Probabilistic Inductive Logic Programming*. Springer, 1–27.

[6] Tom Decroos, Lotte Bransen, Jan Van Haaren, and Jesse Davis. 2019. Actions Speak Louder than Goals: Valuing Player Actions in Soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1851–1861.

[7] Tom Decroos, Jan Van Haaren, Vladimir Dzyuba, and Jesse Davis. 2017. STARSS: A Spatio-Temporal Action Rating System for Soccer. In *In ECML/PKDD 2017 Workshop on Machine Learning and Data Mining for Sports Analytics*.

[8] Dazhen Deng, Jiang Wu, Jiachen Wang, Yihong Wu, Xiao Xie, Zheng Zhou, Hui Zhang, Xiaolong Zhang, and Yingcai Wu. 2021. EventAnchor: Reducing Human Interactions in Event Annotation of Racket Sports Videos. In *CHI Conference on Human Factors in Computing Systems*.

[9] Michael Fuchs, Ruizhi Liu, Ivan Malagoli Lanzoni, Goran Munivrana, Gunter Straub, Sho Tamaki, Kazuto Yoshida, Hui Zhang, and Martin Lames. 2018. Table Tennis Match Analysis: A Review. *Journal of Sports Sciences* 36, 23 (2018), 2653–2662.

[10] Samitha Herath, Mehrtash Harandi, and Fatih Porikli. 2017. Going deeper into action recognition: A survey. *Image and vision computing* 60 (2017), 4–21.

[11] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[12] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 2012. 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence* 35, 1 (2012), 221–231.

[13] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems 30*. 3146–3154.

[14] Daphne Koller, Nir Friedman, Sašo Džeroski, Charles Sutton, Andrew McCallum, Avi Pfeffer, Pieter Abbeel, Ming-Fai Wong, Chris Meek, Jennifer Neville, et al. 2007. *Introduction to statistical relational learning*. The MIT Press.

[15] Hildegard Kuehne, Hueihan Jhuang, Estíbaliz Garrote, Tomaso Poggio, and Thomas Serre. 2011. HMDB: a large video database for human motion recognition. In *International conference on computer vision*. 2556–2563.

[16] Guiliang Liu and Oliver Schulte. 2018. Deep Reinforcement Learning in Ice Hockey for Context-Aware Player Evaluation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. 3442–3448.

[17] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12, Oct (2011), 2825–2830.

[18] Mark Pfeiffer, Hui Zhang, and Andreas Hohmann. 2010. A Markov Chain Model of Elite Table Tennis Competition. *International Journal of Sports Science and Coaching* 5, 2 (2010), 205–222.

[19] Ronald Poppe. 2010. A survey on vision-based human action recognition. *Image and vision computing* 28, 6 (2010), 976–990.

[20] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. 2018. CatBoost: unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems 31*. 6638–6648.

[21] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, Vol. 1. 91–99.

[22] Mikel D Rodriguez, Javed Ahmed, and Mubarak Shah. 2008. Action MACH a spatio-temporal Maximum Average Correlation Height filter for action recognition. In *IEEE conference on computer vision and pattern recognition*. 1–8.

[23] Kurt Routley and Oliver Schulte. 2015. A Markov Game model for valuing player actions in ice hockey. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*. 782–791.

[24] Anthony Sicilia, Konstantinos Pelechrinis, and Kirk Goldsberry. 2019. DeepHoops: Evaluating Micro-Actions in Basketball Using Deep Feature Representations of Spatio-Temporal Data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2096–2104.

[25] Karen Simonyan and Andrew Zisserman. 2014. Two-Stream Convolutional Networks for Action Recognition in Videos. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Vol. 1. 568–576.

[26] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[27] Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. 2012. UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402* (2012).

[28] Heng Wang and Cordelia Schmid. 2013. Action recognition with improved trajectories. In *International conference on computer vision*. 3551–3558.

[29] Jiachen Wang, Kejian Zhao, Dazhen Deng, Anqi Cao, Xiao Xie, Zheng Zhou, Hui Zhang, and Yingcai Wu. 2020. Tac-Simur: Tactic-based Simulative Visual Analytics of Table Tennis. *IEEE Transactions on Visualization and Computer Graphics*. 26, 1 (2020), 407–417.

[30] Xinyu Wei, Patrick Lucey, Stuart Morgan, Peter Carr, Machar Reid, and Sridha Sridharan. 2015. Predicting Serves in Tennis Using Style Priors. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2207–2215.

[31] Xinyu Wei, Patrick Lucey, Stuart Morgan, and Sridha Sridharan. 2013. Predicting Shot Locations in Tennis Using Spatiotemporal Data. In *Proceedings of 2013 International Conference on Digital Image Computing: Techniques and Applications*. 1–8.

[32] Xinyu Wei, Patrick Lucey, Stuart Morgan, and Sridha Sridharan. 2013. "Sweet-Spot": Using Spatiotemporal Data to Discover and Predict Shots in Tennis. In *Proceedings of MIT Sloan Sports Analytics Conference*.

[33] S. Wenninger and Martin Lames. 2016. Performance Analysis in Table Tennis - Stochastic Simulation by Numerical Derivation. *International Journal of Computer Science in Sport* 15, 1 (2016), 22–36.

[34] Huanqun Wu. 1963. The statistical method of techniques and tactics in table tennis. *The Material of Sport Science and Technique*, 21 (1963), 18–20.

[35] Huanqun Wu, Zhenbiao Li, Z. Tao, W. Ding, and J. Zhou. 1989. Methods of Actual Strength Evaluation and Technical Diagnosis in Table Tennis Match. *Journal of National Research Institute of Sports Science* 1 (1989), 32–41.

[36] Yingcai Wu, Ji Lan, Xinhuan Shu, Chenyang Ji, Kejian Zhao, Jiachen Wang, and Hui Zhang. 2018. iTTVis: Interactive Visualization of Table Tennis Data. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 709–718.

[37] Zhi-Hua Zhou. 2019. Abductive learning: towards bridging machine learning and logical reasoning. *Science China Information Sciences* 62, 7 (2019), 76101:1–76101:3.

[38] Zhi-Hua Zhou and Ji Feng. 2017. Deep Forest: Towards an alternative to deep neural networks. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. 3553–3559.

[39] Zhi-Hua Zhou and Ji Feng. 2019. Deep forest. *National Science Review* 6, 1 (2019), 74–86.