This CVPR Workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Augmenting Pass Prediction via Imitation Learning in Soccer Simulations

Takeshi Kaneko[1], Rei Kawakami[1], Takeshi Naemura[2], Nakamasa Inoue[1]
[1]Tokyo Institute of Technology
[2]The University of Tokyo

Abstract

Pass analysis in soccer is essential for predicting players' actions and optimizing team strategies. Existing pass prediction methods involve supervised learning, which requires costly annotations about who passes where and when. We propose the use of additional synthetic data generated by a soccer simulator to overcome this challenge. Specifically, we employ imitation learning to train a policy network that mimics player behavior patterns using the data intended for prediction. This policy network, along with the simulator, is used to generate synthetic data. The generated synthetic data is then combined with real-world data to learn pass prediction by an existing model that utilizes both trajectory and video data. Experiments confirm that our approach improves the top-1 prediction accuracy of the intended pass receiver by 3.72% compared to an existing state-of-the-art method.

1. Introduction

Soccer pass prediction is the task of forecasting which player will receive the pass based on match data available up to the moment before the pass is made [13, 31]. The process of soccer pass prediction involves analyzing information such as the positions and movements of players during the match, as well as the position of the ball, to predict the next pass.

In the context of pass prediction, forecasting a player's pass selection is crucial for supporting tactical decision-making and enhancing team performance [5, 28, 30]. When analysts and coaches can efficiently understand team strategies, they contribute to improving win rates [7]. AI-enabled detailed data analysis also aids in evaluating teams and players and enhancing the quality of scouting [29], and analyzing the talents of undervalued players [27]. Accurate pass prediction assists in devising effective strategies to breach the opponent's defense and provides information for choosing the optimal pass options even under the pressure of a match [26]. Power *et al*. [23] estimated risk and reward for passes to analyze the game and the players. These studies allow teams to approach games more strategically, thereby increasing their chances of winning.

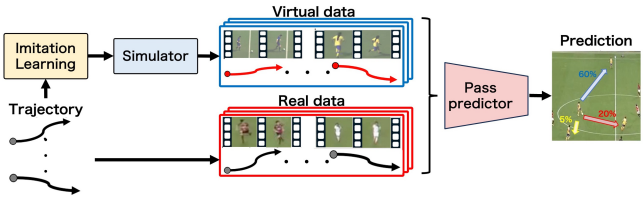


Figure 1. Overview of our method. Our method involves generating synthetic data based on real-world data. Imitation learning is performed using the trajectories of players, and the agents obtained through this process are utilized in a simulator. By utilizing both real and synthetic data simultaneously, the pass prediction deep learning model is trained to enhance video representation learning and acquire new passing scenarios. This approach facilitates a deeper understanding of game dynamics, enabling more accurate pass predictions. Ultimately, it can predict the probability of passing to each teammate.

In the field of pass prediction, there are primarily two approaches: one utilizing only trajectory data, and another that employs both trajectory and video data [13]. Although trajectory-based methods are more prevalent due to the ease of data acquisition, incorporating both trajectories and video in pass prediction offers numerous advantages. The semantic information from video data not only contributes to improved prediction accuracy but also enables the visual presentation of pass options within the video. Additionally, video data, which is rich in information, is utilized in soccer for tasks such as event detection, ball detection, and player tracking [11, 16].

However, when utilizing trajectory and video data for soccer pass prediction, the high costs of annotation and the scarcity of training data present challenges [16, 24]. Annotation requires careful consideration of multiple variables, such as player positions, movements, and the configurations of opposing teams [1]. Video data requires manually labeling pass events for all involved players, demanding video review to identify passers and receivers. High-quality pass prediction requires extensive video data covering various passing scenarios, presenting significant challenges.

In this study, we propose a method utilizing a soccer simulator to compensate for the lack of real data, as illustrated in Fig. 1. Specifically, agents are trained through imitation learning from real soccer data, and these agents are then operated within a simulator to generate realistic play scenes. The generated virtual data, combined with real data, are used to train an existing pass prediction model. The existing model utilizes both video and trajectory data [13]; extracting information such as the orientation of a player's body and the direction of their face from video, and combining it with trajectory data, allows for more accurate pass predictions. The use of synthetic data in this learning process proves to be effective.

In our experiments, we observed a 3.72% improvement in accuracy with the proposed method compared to an existing method. Additionally, we implemented a baseline model using reinforcement learning instead of imitation learning and found that imitation learning tends to select defensive passes more akin to actual play. The analysis of the experiments revealed various advantages, including enhanced learning of video features and player relationships, potentially facilitated by the proposed method.

Our study makes two significant contributions. First, it demonstrates a learning approach that utilizes a simulator and imitation learning to generate data at low cost, which can be employed to train models. This allows for covering a wider range of scenarios without solely relying on actual match data. Second, by applying this to a pass prediction model that uses both video and trajectory data, we have improved prediction accuracy. Our analysis confirmed the enhanced learning of video features and the relationships between players, and that our method can produce predictions with more variability than an existing method.

2. Related Work

Pass prediction Soccer pass prediction uses information up to the frame just before a player with the ball makes a pass, to predict which teammate the player will pass to [4, 15]. Pass prediction models mostly utilize trajectory data as input. Dauxais *et al*. [4] uses the coordinates at the moment of the pass, combining features manually for prediction with a random forest algorithm. Hubacek *et al*. [15] predicts the probability of pass destinations using Convolutional Neural Networks (CNNs).

There are also methods that perform pass prediction using both video and trajectory data. Video data allows for the utilization of information not obtainable from direction and speed alone, such as the orientation of a player's body and face. Sanguesa *et al*. [3] uses player coordinates and information on the orientation of attacking players' bodies obtained from video to calculate the pass possibility among team players. Achieving over 70% Top-3 accuracy from more than 6,000 pass scenes, the combination with pass evaluation metrics also enables the refinement of existing pass evaluation models.

Honda *et al*. [13] proposes a method combining trajectory and video data for prediction, achieving significant accuracy improvements compared to methods that use only trajectory data. Video data is processed using a 3D Convolutional neural network (3DCNN), and trajectory data is analyzed with Long Short-Term Memory (LSTM) networks. Feature vectors are calculated for each player and fused using a Transformer encoder to predict the probability of the pass receiver.

Pass prediction using both video and trajectory data tends to achieve higher accuracy compared to methods using only trajectory data. However, the challenge lies in the difficulty of acquiring video data, which limits the quantity and variety of data available. For example, while short-distance passes tend to be more predictable, the predictive performance for long-distance passes often decreases [13].

Pass evaluation An important analytical method in soccer is pass evaluation, which quantifies the potential value of passes to the team. Machine learning-based pass evaluation methods often use information before and after the pass as input, employing manually designed features and machine learning models to output evaluation scores. Rein *et al*. [25] evaluates passes based on field area domination and the number of defenders between the goalkeeper and the ball possessor. Chawla *et al*. [6] classifies passes into good, neutral, and bad categories based on the pass's position, direction, distance, and spatiotemporal information during the match. Decroos *et al*. [8] classifies players' actions and probabilistically evaluates their contribution to scoring or conceding goals. Goes *et al*. [12] assesses passes using characteristics such as length, speed, and direction.

Some approaches incorporate the prediction of a pass's success probability into pass evaluation models, evaluating passes in conjunction with the expected value of scoring upon success. Fernández *et al*. [10] proposes the Expected Possession Value (EPV), which quantifies the potential value of shots or passes to specific locations based on the positions of all players and the ball. Power [23] *et al*. estimates the success rate of passes based on the distance and angle between players using a linear regression. Fernández and Bornn [9] uses a CNN to calculate the success rate and selection probability of passes, producing a SoccerMap that represents the success rate based on the pass location. Anzer and Bauer [2] utilizes pass length, direction, and player positions to predict the success probability of passes to each player. Liu *et al*. [20] combines deep learning and reinforcement learning to learn the spatiotemporal dynamics of players, evaluating actions based on the change in scoring opportunities for home and away teams.

Pass evaluation, particularly in identifying the optimal

pass selection during actual matches, is challenging. This is because the models are trained to quantify evaluations of passes that have already been executed. Improving the accuracy of pass predictions can contribute to better pass evaluations.

Utilization of simulation The field of pass analysis often faces challenges related to the quantity and quality of data. Due to the large size of soccer fields and the numerous players involved, resolution can become limited, leading to errors in coordinate data. Additionally, the cost of annotation means that soccer pass data are expensive to obtain, yet the quantity and quality of the trained data can significantly influence prediction outcomes.

There are examples in other sports where simulations are used to generate data, enhancing the accuracy of analysis. Newman *et al*. [22] utilizes the Madden NFL 2020 PC game as a simulator for American football formation classification, improving identification performance with CNNs and YOLO. Huang *et al*. [14] simulates a realistic badminton environment, predicting flight trajectories with the use of physical models. Google Research Football (GRF) [18] is a soccer simulation platform designed for developing and testing reinforcement learning and machine learning algorithms, based on real soccer matches and capable of simulating realistic scenarios and tactics. It has been used in the development of soccer reinforcement learning AI [19, 32].

Soccer research also employs simulations. Komorowski and Kurzejamski[17] utilizes GRF [18] for multi-camera tracking of soccer players, applying sim2real. They combine real soccer match data with simulation data and employ graph neural networks to account for the movements and interactions among players for tracking. Morra *et al*. [21] enhanced event recognition during soccer matches by utilizing a modified GRF simulator, achieving improved performance.

The GRF simulator is widely used for player detection and event recognition. However, in pass analysis, the research field is divided into two areas: robot soccer and the analysis of actual play. The former uses simulations while few papers in the latter apply simulations to the analysis of real players. Simulators have the potential to function as a complement to actual match data, and their utilization is anticipated across many applications.

3. Method

Overview Our proposed method generates diverse synthetic data that resembles real data and utilizes this for training. Thus, it comprises three steps: 1) Learning behaviors similar to real data, 2) Generating diverse synthetic data, and 3) Training with a mix of synthetic and real data.

First, there is the learning of agent behaviors. Through deep imitation learning, the model learns actions frame by frame from soccer players' match data, enabling it to mimic the movements of players. Second, we have the generation of synthetic training data. Trained agents are pitted against each other to automatically generate synthetic data, obtaining labels for video, trajectory, and actions. This allows for the expansion of the dataset. Third, training combines synthetic and actual data. Existing prediction models [13] are used to learn pass classification, but during this process, we apply weighting to both real and synthetic data to adjust the gap between them as training progresses.

The proposed method offers two advantages. First, the expanded data set facilitates the learning of features, improving the accuracy of predicting the receiving player of a pass. Second, soccer simulations generate diverse play styles, including new passing scenarios, enhancing the accuracy of pass predictions. Additionally, simulators reduce errors compared to manual coordinate setting or person detection from videos, offering a significant advantage.

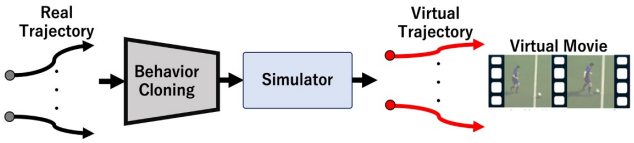


Figure 2. Diagram of behavior cloning and data generation. This diagram shows the process using machine learning and simulation. Initially, a CNN model learns from the actual players' trajectories through behavior cloning. The learned behavior patterns are applied to a simulator to generate synthetic videos and trajectories. This creates data of diverse plays that are difficult to obtain from real matches, ultimately resulting in the creation of virtual data. This sequence of processes enables the generation of a diverse dataset that contributes to improving the accuracy of pass prediction models.

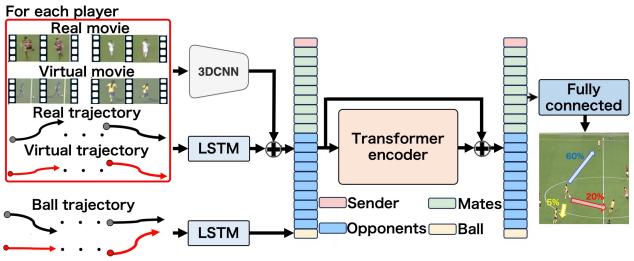


Figure 3. Diagram of the pass prediction model. Real and synthetic videos of players, along with trajectory data and the ball's trajectory data, undergo feature extraction through 3DCNN and LSTM networks. These features are then fed into a transformer encoder, which analyzes the relationships between multiple entities. Through this analysis, probabilistic predictions are made from the passer to the receiving player. Finally, the outcome of the pass is outputted by a fully connected layer.

Behavior cloning We utilize behavior cloning from imitation learning to teach agents movements close to those of soccer players, as illustrated in Fig. 2. Our goal is for agents to mimic the actions of players, using trajectory data from soccer matches as input to build a model that predicts action labels in future frames.

The policy function for agent behavior decision is replicated using a model that modifies the CNN used for reinforcement learning in GRF. Since the policy function needs to be executed within the simulator to determine actions, CNN is appropriate due to its balance of execution speed and processing capability.

The policy function, π, is represented by

$$\pi(a|s) = \frac{e^{f(a(s))}}{\sum_{a' \in A} e^{f(a'(s))}}. \quad (1)$$

The policy function π, as shown in Eq. (1), represents the probability of selecting action a_t given the state s_t at time t, and serves as a function to determine which action the agent should take. The function $f_{a_{t_k}}(s_t)$ represents the evaluation function for taking action a_t under state s_t, where n denotes the number of types of actions.

The input to the policy function is trajectory data. For the purpose of action cloning, pairs of trajectory data and associated player event data are prepared. Using this data, supervised learning is applied to predict action labels with a CNN, based on the player's position and action in the corresponding frame. The labels for the data need to be adjusted to conform to the definitions used in the simulator. The agent is trained to predict the true label as the action with the highest probability in each state.

The action labels include movements in 8 directions and actions such as sprinting. In this paper, we define 19 types of action labels and manually set appropriate speeds and distances. The details of the action labels are available in the supplementary material. Specifically, the input to the CNN encoder is a time series of minimaps showing the positions of all players at time t for the past m frames. The output of the CNN is the action label for the agent at time $t + 1$.

Data generation for simulation The generation of synthetic data utilizes the GRF simulator. This paper details the method of data generation within the simulator, as well as the specifics of the simulator's configuration.

Data generation is performed as follows: Within the simulator, when a pass command occurs, the passer and the receiver are identified, and a pass scene is generated. The start frame of the pass is defined as the moment when the pass command is first input, followed by the detection of contact between the ball and a player. Only intentional passes are captured, excluding accidental passes. The moment when the receiver catches the ball is defined as when contact with the ball is confirmed.

Based on the frame in which a pass is detected, the n frames from the preceding t seconds are saved as data for the pass scene, and the trajectory of the ball and the video are extracted. The ID of the receiver is annotated as the correct label in the data. This provides time-series data up to the start frame of the pass and training data on which player will receive the pass.

The simulator settings are as follows: Both action cloning and pass prediction models require trajectories of the ball and players. This information is the x and y axis coordinates of each player's 3D position within the simulation. The video data for the pass prediction model is also generated from the simulator. The virtual camera is configured in terms of angle of view, position, and orientation to ensure the entire soccer field is within view and all players are visible in the video. Special care is taken to ensure that, particularly at the moment a pass is made, all players, except the goalkeeper, are visible in the video.

The pass prediction model requires a time series of frames cropped with bounding boxes for each player. Since the simulator's coordinates are defined in world coordinates, a coordinate transformation is necessary to obtain the bounding boxes of players in the image. This involves using quaternions to perform an inverse affine transformation from 3D geometric coordinates to the camera origin's 3D coordinates, followed by perspective projection transformation and aspect ratio adjustment.

The overall image size of the scene is set to the maximum possible to ensure stable output, with careful consideration to prevent the resolution of the video within each player's bounding box from becoming too low.

Mixed learning with data in pass prediction This paper utilizes the model by Honda *et al*. [13], which employs both video and trajectory data for pass prediction, due to its superior performance compared to using trajectory data alone, as illustrated in Fig. 3. Data generated in the simulator and real data are combined with weighting, and both are used for supervised learning in the pass prediction model.

4. Experiment

4.1. Dataset

The real data on players' trajectories and video comprises successful passes from 25 home games played by Kashima Antlers, Urawa Red Diamonds, and FC Tokyo in J1 league in Japan. The dataset consists of tracking data, which includes the positions of all players on the field during the games, and wide-angle video footage. The position coordinates were acquired using a high-precision tracking system and manually corrected by experts, provided by Data Stadium Inc. The tracking data also includes event annotations and player action labels assigned by experts. The video data is in wide-angle format. We extracted information for 20

players, excluding goalkeepers, from this dataset, as the focus is solely on field players.

Tracking coordinates range within $(0, 0) \leq (x, y) \leq (5250, 3400)$, with a position coordinate sampling rate of 25 Hz. The video resolution is 1920×1080, with a sampling rate of 30 Hz. Only successful pass scenes were selected for analysis, with scene lengths set between 1.0 and 5.0 seconds. The total number of scenes is 15,586, of which 10,911 scenes were allocated to the training set, 1,559 scenes to the validation set, and 3,116 scenes to the test set.

For the purposes of utilizing both video and trajectory data, the dataset was processed accordingly. The ball's position information was estimated through linear interpolation from pass and shot event data and was resampled to 30 Hz for synchronization with the video. The coordinates of players and the ball were normalized to ensure visual consistency within the range $(-1, -1) \leq (x, y) \leq (1, 1)$ and were horizontally flipped as necessary to maintain uniformity in the attacking direction. To extract trajectory features, 150 frames (spanning 5 seconds) depicting the movement of each player and the ball were used, incorporating a broad temporal context. Scenes with less than 150 frames were supplemented with zero padding. For improved memory use efficiency and accelerated learning speed, the video data's sampling rate was set to 15 Hz. Cropped images of players were extracted from clips consisting of 15 frames per second, with each frame resized to 100×100 pixels.

4.2. Synthetic Dataset Processing

The synthetic dataset consists of simulated soccer match data generated using a Gaussian Random Field (GRF) approach. The simulator is a 3D soccer simulator, designed to replicate the soccer match environment with high fidelity. In the simulation, realistic actions are possible, including considerations of players' body axes and orientations, and adjusting the force of kicking the ball to match that of real human physical capabilities.

In the simulator, an 11-versus-11 match format was employed, mirroring the flow of actual games. For each match, the data output includes the x, y coordinates of the ball and 20 field players, excluding the goalkeeper, along with match video from an overhead perspective that captures all players. The simulation matches were structured with 45-minute halves and 5 minutes of additional time, aligning with real soccer games, resulting in a total game time of 50 minutes. Simulation videos were produced at a resolution of 3840×2160 to ensure a stable full-screen display, with a sampling rate set to 30 Hz to match the real dataset.

For the passing scenes, the video and coordinate data from the 5 seconds leading up to the pass are included, with both the video and coordinate data flipped to maintain consistency in the direction of attack. Due to the soccer court being symmetrically identical on both sides, this procedure ensures data consistency and simplifies conditions. Furthermore, when the ball is located at the $(x, y, 0)$ coordinates, the camera's field of view (FOV) is set to $32.12 - 0.025y$ degrees, capturing the scene from a diagonal overhead perspective at the coordinates $(0.85x, 0.75y - 84.12, 47.27)$ meters. In the synthetic dataset, similar to the real dataset, 10,911 scenes were utilized for training purposes only and not for validation or testing. This approach is intentional, as the primary objective of this research is to assess the model's accuracy against real data.

4.3. Implementation of Behavior Cloning Model

The behavior cloning model used 19 types of actions as the correct labels and predicted the sender's action at time $t+1$ using the minimap information from frames $t-4$ to t.

The behavior cloning model was trained as a 19-category multiclass classification using a CNN as the encoder model. Details on the 19-type labels and the CNN model are elaborated in the supplementary material. The batch size was set to 128, with a learning rate of $1.2e-4$ applied. The total number of epochs was set to 2,000, with an early stopping criterion that ended training if no performance improvement was observed for 40 epochs. The Adam optimization was used, with parameters set to $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1e-8$. The total number of training samples used in this study was 12,544.

For comparison, a method was also implemented that trained the agent using Proximal Policy Optimization (PPO), a reinforcement learning approach. The reward function was defined by a gain or loss of points by scoring a goal, with a score of ± 1.

4.4. Implementation of the Pass Prediction Model and Mixed Data Training

The processed trajectory and video data were incorporated into the prediction model by Honda *et al*. [13], as depicted in Fig. 3. This deep learning model consists of a 3D CNN for extracting features from video, an LSTM for extracting features from trajectories, and a Transformer for understanding the relationships between features. Initially, features were extracted from trajectory coordinates using LSTM. The features from the intermediate layers were then embedded into a 64-dimensional vector. The ball's trajectory features were used independently due to the absence of corresponding image features. The 3D CNN utilized a portion of ResNet3D-6 for video feature extraction. Features of the 20 players and the ball were input into a Transformer encoder, with input-output integration performed via residual connections. This encoder had a four-layer structure, each layer featuring four heads, with parameters optimized experimentally. Features corresponding to potential receivers were fed into a fully connected layer and converted into pass reception probabilities using the softmax function.

Table 1. Prediction accuracy comparison. PPO RL* is our implementation where agents learn policy by manually designed reward.

Methods	Top-1	Top-3	Top-5	Loss
Honda et al. [13]	62.28	91.92	97.79	1.025
PPO RL*	65.72	92.95	97.86	0.9745
Cloning (Proposed)	**66.00**	**93.20**	**97.95**	0.9681

In the proposed model, training was conducted using both real and synthetic data simultaneously, leveraging deep learning-based predictive models. A batch size of 24 was employed for training. Both models utilized the ADAM optimizer with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$, and the learning rate was set to $1e-4$. Training utilized the cross-entropy loss function, and testing was performed using top-k accuracy. To prevent overfitting, an early stopping strategy was adopted, selecting the model with the highest top-1 accuracy on the validation data.

We adjusted the hyperparameters of the loss function and the scheduling of the learning rate when training with a combination of real and synthetic datasets. This optimization of the model's training process enabled effective utilization of both real and synthetic data. We conducted adjustments on the hyperparameters related to the weighting of the loss function when using real and synthetic data simultaneously. Furthermore, in scheduling the learning rate, an appropriate scheduling method was selected, taking into account the characteristics of both real and synthetic data.

$$L_{\text{total}} = \alpha L_{\text{real}} + \beta L_{\text{virtual}} \tag{2}$$

Eq. (2) represents the total loss when training with both real and synthetic data combined. α and β are the weight parameters, which were fixed to be time-invariant. For the purpose of an ablation study, implementations where the weight parameters varied over epochs linearly, sinusoidally, and according to a sigmoid function were also explored, and each was compared with the application of their respective hyperparameters. Details on the scheduling of these loss functions are elaborated in supplementary material.

4.5. Results

Experimental results are presented in Tab. 1. Our method improved accuracy in both behavior cloning and reinforcement learning compared to the conventional method cited in [13]. Notably, in behavior cloning, our approach increased Top-1 accuracy by 3.72%, Top-3 by 1.28%, and Top-5 by 0.16%, and also reduced loss on the test data. An increase in accuracy was also observed in comparison with reinforcement learning. Since behavior cloning involves learning the actions of players from data, this improvement in accuracy is attributed to the effectiveness of our method in leveraging the data.

Table 2. Accuracy comparison when adjusting the scheduling of the mixing ratio between real and synthetic data.

Methods	Top-1	Top-3	Top-5
PPO RL - Fixed	24.76	60.05	80.34
PPO RL - Linear	65.72	93.01	97.92
PPO RL - Sinusoid	64.30	93.33	97.82
PPO RL - Sigmoid	63.89	93.08	98.02
Cloning - Fixed	66.00	93.20	97.95
Cloning - Linear	65.46	92.78	97.46
Cloning - Sinusoid	61.96	92.24	97.85
Cloning - Sigmoid	63.41	92.08	97.31

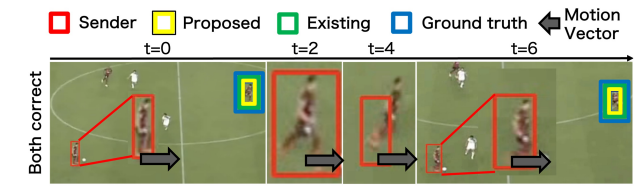
Additionally, to investigate the impact of the presence and level of strategy on accuracy, comparative experiments were conducted using synthetic data data generated with action strategies based on the simulator's built-in AI and random action strategies based on a uniform distribution. *Our behavior cloning largely surpasses those based on generated data with simple agents.* The results are detailed in supplementary material.

Next, as part of an ablation study, we present the results of experiments that involved changing the learning scheduling between real and synthetic data in Tab. 2. We conducted a comparative analysis of different learning scheduling methods (fixed, linear, sinusoidal, and sigmoid). The results showed that linear scheduling achieved the highest Top-1 accuracy in reinforcement learning, while fixed scheduling yielded the highest Top-1 accuracy in behavior cloning. For reinforcement learning methods, the highest accuracies were observed with linear scheduling for Top-1, sinusoidal scheduling for Top-3, and sigmoid scheduling for Top-5, indicating that time-dependent scheduling changes were most effective in achieving the highest accuracy.

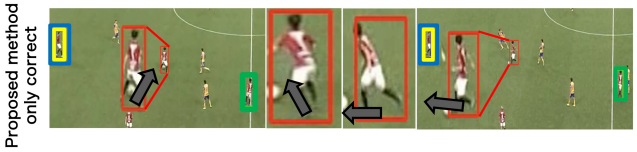
Reinforcement learning agents often exhibit significantly different strategies between generated and real data in soccer scenes, potentially causing a gap in play styles between the datasets. On the other hand, when using behavior cloning, there is a likelihood of more similar strategies in soccer scenes between generated and real data, suggesting a smaller gap in play styles.

4.6. Analysis

Examples of successes and failures are presented in Fig. 4. The red box represents the player passing the ball, green indicates predictions by the existing method cited in [13], and yellow represents predictions by the proposed method. In examples where both methods were correct, the player consistently faces the right side of the field, and the correct receiver is located diagonally to the upper right. In this case, the absence of change in body orientation is believed to have led to correct predictions by both methods. In examples where only the proposed method was correct, the



(a) An example where both methods predicted correctly



(b) An example where our method only predicted correctly

Figure 4. Comparison of pass prediction using existing and proposed methods. In the upper sequence, both methods share accurate predictions, with the Sender (highlighted in red) consistently facing to the right. In the lower sequence, only the proposed method makes an accurate prediction, where the Sender initially faces the top-right but changes orientation over time to eventually face left. This observation suggests that the proposed method excels in capturing significant changes in the player's body orientation to make accurate predictions.

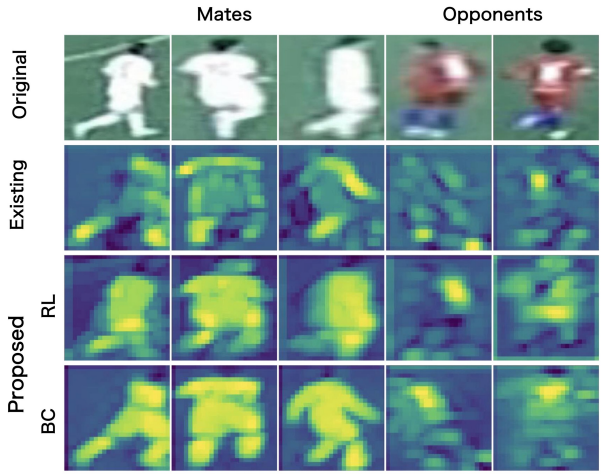


Figure 5. Visualization of the maximum average heatmap in the intermediate layers of the 3DCNN for each player's footage. Since these are the heatmaps with the maximum average for each method, the steps are different for each. The existing method presents ambiguous shapes and body orientations, whereas the proposed method clearly delineates shapes and body orientations, particularly in behavior cloning, where it is most distinct.

initial orientation of the player making the pass is towards the back of the screen, followed by a counterclockwise rotation of body orientation, eventually facing the left side of the field. The correct receiver is located on the left side of the screen. While the existing method predicted a pass to a player on the right side based on the initial orientation, the proposed method accurately predicted the receiver based on the final body orientation. This suggests that the proposed method takes into consideration significant changes in the body axis of players in motion.

To assess the learning effectiveness of visual information, we analyzed the feature maps that exhibited the highest activation in the first layer of the 3DCNN. The heatmap is shown in Fig. 5. The left side shows three ally players, and the right side shows two enemy players. Across all methods, activations were higher for ally players, indicating that visual information about allies is prioritized.

Using synthetic data resulted in higher activations for characters, and the proposed method for behavior cloning clearly delineated the shapes of ally players with high activation levels. The shapes of enemy players were also accurately recognized, and the areas of high activation matched their orientations. This suggests that the proposed method may effectively learn from visual information.

We conducted a visualization and comparison of the scaled attention map in the final layer of the Transformer. The attention map is shown in Fig. 6. In the figure, player 0 is the one passing the ball, players 1 to 9 are potential ally receivers, players 10 to 19 are opponents, and number 20 represents the ball, with player 3 at the bottom of the image being the correct receiver. Using the proposed behavior cloning method, the correlation between number 3 and 20 in the final layer of the attention map reached a maximum value of 0.35, suggesting a high level of relevance between player 3 and the ball. Additionally, player 10, the closest opponent, showed high relevance with many other players, indicating that this opponent affected numerous players. In contrast, high relevance scores for players 3 and 10 were not observed with the compared method [13]. These results suggest that the proposed method is more capable of capturing the intricate relationships between players.

We analyzed the distribution of successfully predicted passes between the existing method [13] and the proposed behavior cloning method. The results are presented in Tab. 3. We compared the average pass length, the circular mean of passes with the right direction as 0 degree, and the circular variance.

The circular mean is introduced to investigate the trend in the direction of passes. To calculate it, following circular statistics, we compute the unit vectors corresponding to these angles, take their average, and then convert this average back into degrees. When angles are distributed, the mean is close to 0 and takes some value if there is similarity in the estimated passes. Similarly, the circular variance is introduced for the purpose of grasping the consistency and degree of dispersion of passes. We calculate the distance from the origin of the mean direction, and subtract this value from 1. This metric takes values from 0 to 1, with higher values indicating that passes are more distributed.

Mathematically, the circular mean $\bar{\theta}$ and the circular

Table 3. Comparison of pass distributions of proposed (BC) and Honda et al. [13].

	Mean length (SD) (m)	Circular mean (deg)	Circular variance
Correct prediction by both	14.10 (5.868)	-54.19	0.7752
Correct prediction only by proposed (BC)	13.80 (6.979)	-20.58	0.7207
Correct prediction only by [13]	13.08 (6.784)	1.570	0.6657
Neither predicts correctly	16.37 (9.309)	20.31	0.5200

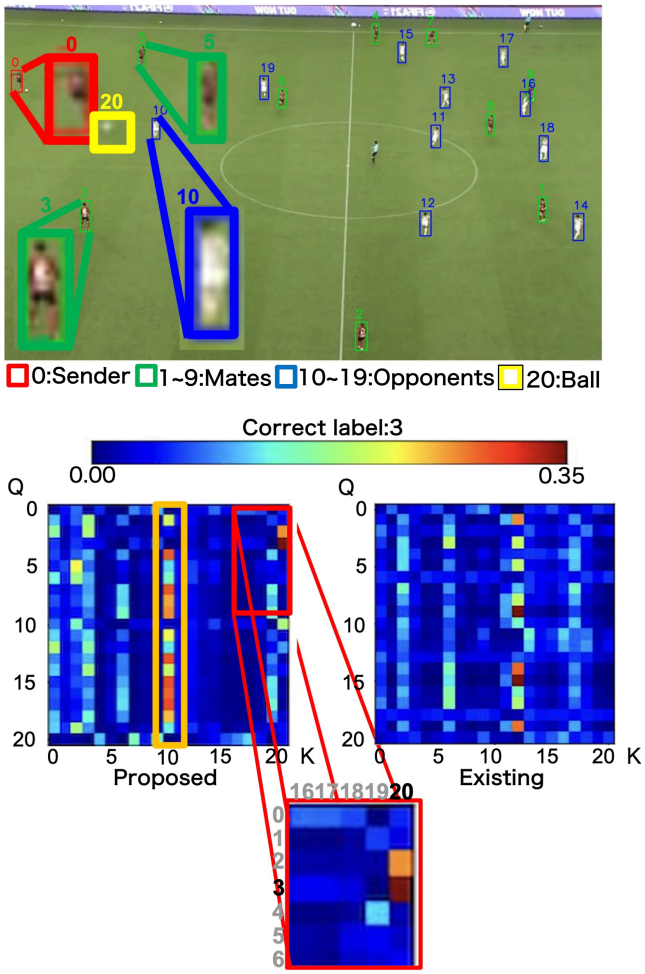


Figure 6. The scaled attention map from the final layer of the transformer is presented, where the vertical axis represents queries and the horizontal axis keys. Higher values indicate stronger correlations between queries and keys. The proposed method shows the highest correlation of 0.35 between the query for player 3, the correct receiver, and the key for ball 20. The key for player 10, being closest to 100,000 players, is considered to have generated relations with many players, suggesting it played a role in obstructing the passing course and attracting attention from other players.

variance v are calculated as shown in Eq. (3).

$$\bar{\theta} = \arctan2\left(\frac{1}{N}\sum_{i=1}^{N}\sin(\theta_i), \frac{1}{N}\sum_{i=1}^{N}\cos(\theta_i)\right) \tag{3a}$$

$$v = 1 - \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}\cos(\theta_i)\right)^2 + \left(\frac{1}{N}\sum_{i=1}^{N}\sin(\theta_i)\right)^2} \tag{3b}$$

where N is the number of passes and θ_i is the angle of each pass.

As shown in Tab. 3, cases where only the proposed method succeeded show a longer average distance and a larger circular variance compared to cases where only the existing method succeeded. This suggests that while having a similar distribution of pass distances as the existing method, the proposed method can accurately predict a more diverse range of passes. Additionally, the average pass angle was -20.58 degrees, a value between the successes of the existing method and those of both methods. The circular variance was 0.7207, which is larger than that of the successes of the existing method alone. This suggests that the proposed method is capable of capturing and accurately predicting a wider variety of pass angles.

5. Conclusion

We have presented a method for enhancing pass prediction performance by employing behavior cloning to instruct agents in policies, generating synthetic data, and integrating this with real data for the training process. As a result, in pass prediction, the accuracy improved by 3.72% for Top1, 1.28% for Top3, and 0.16% for Top5. Additionally, The qualitative analysis demonstrated that the approach is more effective than an existing method in scenarios including changes in the player's body axis and complex pass scenes. This is attributed to the effective learning of player's video representations and the relationships between players.

However, this study is limited to specific simulation data and real match data. Therefore, there is room for further research on improving generalization capabilities across different datasets. In the future, further accuracy improvements could be considered by applying inverse reinforcement learning, combining with pass evaluation, and incorporating more detailed data for each scene.

Acknowledgement

We would like to thank DataStadium Inc. for kindly providing the trajectory and annotation data.

References

[1] Sara Akan and Songül Varlı. Use of deep learning in soccer videos analysis: survey. *Multimedia Systems*, 29(3):897–915, 2023. 1

[2] Gabriel Anzer and Pascal Bauer. Expected passes: Determining the difficulty of a pass in football (soccer) using spatio-temporal data. *Data mining and knowledge discovery*, 36(1):295–317, 2022. 2

[3] Adria Arbues-Sanguesa, Adrian Martin, Javier Fernandez, Coloma Ballester, and Gloria Haro. Using player's body-orientation to model pass feasibility in soccer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2020. 2

[4] Ulf Brefeld, Jesse Davis, Jan Van Haaren, and Albrecht Zimmermann. Predicting pass receiver in football using distance based features. *Machine Learning and Data Mining for Sports Analytics*, pages 145–151, 2019. 2

[5] Ali Cakmak, Ali Uzun, and Emrullah Delibas. Computational modeling of pass effectiveness in soccer. *Advances in Complex Systems*, 21(03n04):1850010, 2018. 1

[6] Sanjay Chawla, Joël Estephan, Joachim Gudmundsson, and Michael Horton. Classification of passes in football matches using spatiotemporal data. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, 3(2):1–30, 2017. 2

[7] Paolo Cintia, Fosca Giannotti, Luca Pappalardo, Dino Pedreschi, and Marco Malvaldi. The harsh rule of the goals: Data-driven performance indicators for football teams. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–10. IEEE, 2015. 1

[8] Tom Decroos, Lotte Bransen, Jan Van Haaren, and Jesse Davis. Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1851–1861, 2019. 2

[9] Javier Fernández and Luke Bornn. Soccermap: A deep learning architecture for visually-interpretable analysis in soccer. In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part V*, pages 491–506, 2021. 2

[10] Javier Fernández, Luke Bornn, and Dan Cervone. Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. In *13th MIT Sloan Sports Analytics Conference*, 2019. 2

[11] Silvio Giancola, Mohieddine Amine, Tarek Dghaily, and Bernard Ghanem. Soccernet: A scalable dataset for action spotting in soccer videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 1711–1721, 2018. 1

[12] Floris R Goes, Matthias Kempe, Laurentius A Meerhoff, and Koen APM Lemmink. Not every pass can be an assist: a data-driven model to measure pass effectiveness in professional soccer matches. *Big data*, 7(1):57–70, 2019. 2

[13] Yutaro Honda, Rei Kawakami, Ryota Yoshihashi, Kenta Kato, and Takeshi Naemura. Pass receiver prediction in soccer using video and players' trajectories. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3502–3511, 2022. 1, 2, 3, 4, 5, 6, 7, 8

[14] Li-Chun Huang, Nai-Zen Hsueh, Yen-Che Chien, Wei-Yao Wang, Kuang-Da Wang, and Wen-Chih Peng. A reinforcement learning badminton environment for simulating player tactics (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 16232–16233, 2023. 3

[15] Ondřej Hubáček, Gustav Šír, and Filip Železný. Deep learning from spatial relations for soccer pass prediction. *Machine Learning and Data Mining for Sports Analytics*, pages 159–166, 2019. 2

[16] Paresh R Kamble, Avinash G Keskar, and Kishor M Bhurchandi. A deep learning ball tracking system in soccer videos. *Opto-Electronics Review*, 27(1):58–69, 2019. 1

[17] Jacek Komorowski and Grzegorz Kurzejamski. Graph-based multi-camera soccer player tracker. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2022. 3

[18] Karol Kurach, Anton Raichuk, Piotr Stańczyk, Michał Zając, Olivier Bachem, Lasse Espeholt, Carlos Riquelme, Damien Vincent, Marcin Michalski, Olivier Bousquet, et al. Google research football: A novel reinforcement learning environment. In *Proceedings of the AAAI conference on artificial intelligence*, pages 4501–4510, 2020. 3

[19] Chenghao Li, Tonghan Wang, Chengjie Wu, Qianchuan Zhao, Jun Yang, and Chongjie Zhang. Celebrating diversity in shared multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 34:3991–4002, 2021. 3

[20] Guiliang Liu, Yudong Luo, Oliver Schulte, and Tarak Kharrat. Deep soccer analytics: learning an action-value function for evaluating soccer players. *Data Mining and Knowledge Discovery*, 34:1531–1559, 2020. 2

[21] Lia Morra, Francesco Manigrasso, and Fabrizio Lamberti. Soccer: Computer graphics meets sports analytics for soccer event recognition. *SoftwareX*, 12:100612, 2020. 3

[22] Jacob Newman, Andrew Sumsion, Shad Torrie, and Dah-Jye Lee. Automated pre-play analysis of american football formations using deep learning. *Electronics*, 12(3):726, 2023. 3

[23] Paul Power, Hector Ruiz, Xinyu Wei, and Patrick Lucey. Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1605–1613, 2017. 1, 2

[24] Keerthana Rangasamy, Muhammad Amir As'ari, Nur Azmina Rahmad, Nurul Fathiah Ghazali, and Saharudin Ismail. Deep learning in sport video analysis: a review. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 18(4):1926–1933, 2020. 1

[25] Robert Rein, Dominik Raabe, and Daniel Memmert. "which pass is better?" novel approaches to assess passing effectiveness in elite soccer. *Human movement science*, 55:172–181, 2017. 2

[26] Markel Rico-González, José Pino-Ortega, Amaia Méndez, Filipe Clemente, and Arnold Baca. Machine learning application in soccer: a systematic review. *Biology of sport*, 40 (1):249–263, 2023. 1

[27] Pieter Robberechts, Maaike Van Roy, and Jesse Davis. unxpass: Measuring soccer player's creativity. In *Proceedings*

of the 29th ACM SIGKDD conference on knowledge discovery and data mining, pages 4768–4777, 2023. 1

[28] William Spearman, Austin Basye, Greg Dick, Ryan Hotovy, and Paul Pop. Physics-based modeling of pass probabilities in soccer. In *Proceeding of the 11th MIT Sloan Sports Analytics Conference*, 2017. 1

[29] Łukasz Szczepański and Ian McHale. Beyond completion rate: evaluating the passing ability of footballers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(2), 2016. 1

[30] Vincent Vercruyssen, Luc De Raedt, and Jesse Davis. Qualitative spatial reasoning for soccer pass prediction. In *CEUR Workshop Proceedings*. CEUR-WS. org, 2016. 1

[31] Xinyu Wei, Patrick Lucey, Stephen Vidas, Stuart Morgan, and Sridha Sridharan. Forecasting events using an augmented hidden conditional random field. In *Computer Vision–ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part IV 12*, pages 569–582. Springer, 2015. 1

[32] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, 35:24611–24624, 2022. 3