Soccer Image Classification with Convolutional Neural Networks (CNN)

Rishin Tiwari
University of South Florida
Tampa, FL, United States
rishintiwari@usf.edu

Saloni Birthare
University of South Florida
Tampa, FL United States
salonibirthare@usf.edu

Martin J. Zuluaga Ospina
University of South Florida
Tampa, FL United States
mzuluaga@usf.edu

***Abstract*— This project presents the development of models for automating soccer event image classification, leveraging Convolutional Neural Networks (CNNs) and Large Language Models (LLMs). The need for this model comes from the errors and inefficiencies of current practices, manually handling the classification process in the sports analysis and broadcasting industry.**

This project presents a model with automated classification of multiple soccer events, enhancing the data utility for multiple stakeholders in sports, broadcasting, betting, and other related industries.

I. Introduction

Soccer, being the most popular sport globally, generates substantial amounts of visual content from recorded and streamed events. This content can be used to analyze and classify different plays. Automatic classification of soccer events can have tremendous potential impact on different business activities and industries, enhancing sports broadcasting, coaching, referee assistance, fan engagement and betting experience.

Currently, classifying soccer events from broadcasts and streaming relies on manual analysis processes. These processes are not only extremely time-consuming, but they are prone to human error. This proposal aims to tackle these issues by using Convolutional Neural Networks (CNNs) and create a competitive deep learning model capable of classifying soccer images into predefined categories.

Automating the classification of soccer images, the model looks to streamline the process, reduce potential errors, and provide valuable insights and trigger business decisions for its stakeholders.

The specific event categories in this model include penalty kicks, free kicks, yellow cards, red cards, substitutions, tackles, and corner kicks.

These events are extremely important in the dynamics of soccer, as many of them represent a high scoring chance, affecting the game by ejecting a player or the substitution of a player. These events can be translated into player performance metrics, live betting odds, broadcast highlights and statistic records.

In this project, the primary business use case is, and purpose is to provide valuable player and team performance analysis. Soccer clubs and coaches can use image classification to analyze player and team performance during matches and training. By categorizing and analyzing soccer events, they can evaluate the overall player contributions, tactical decision, and game strategies.

Using the model to help coaches analyze game events for training and identify individual and collective strength and weaknesses in their own teams and rivals. This application can boost player development and overall team performance. This analysis can also be used to analyze other team's players performances and plan potential signings and investments.

II. Data Preparation

A. Collection

The Data collection process for this project consisted of the collection of 7000 soccer event images from different online sources. The images were sourced from different online sources, using Python's 'bing_image_downloader' library for efficient scraping and collection from videos. This eased gathering images through automated data scraping techniques to ensure a large variety of events.

The images were collected from professional soccer competition events from multiple seasons and tournaments, ensuring a realistic set of images. Regarding geographical context, the images also vary in location and players involved, ensuring diversity on the representation of the events.

Collected images were stored in a local directory for further processing. The images were thoroughly filtered, ensuring the quality and relevance of the data regarding the business problem.

B. *Exploratory Data Analysis (EDA) and Preprocessing*

During the Exploratory Data Analysis (EDA), sample images were visualized to understand the dataset's distribution. This process facilitated tackling data imbalances and identifying the preprocessing requirements of this model.

Preprocessing steps consisted of standardizing the images to enhance the model's performance. This process involved resizing all images to 32X32 pixel dimensions and normalized pixel values. Rotation, flipping, and scaling techniques were used to augment the dataset and increase diversity.

The dataset was then labeled with the corresponding event categories, which was critical for the supervised learning model. The final dataset consisted of a standardized, balanced and labeled set of 7000 images, with 1000 images for each soccer event.

III. MODEL DESIGN AND IMPLEMETATION

The CNN architecture was developed with specific layers and parameters to optimize the performance of the model. The model consists of convolutional layers for feature extraction, followed by fully connected layers for classification. Hyper parameter tuning was implemented to optimize the model and find the best architecture for this project.

A. *Final Model Architechture*

The final model architecture resulting from the hyper parameter tuning includes two convolutional layers and pooling layers, followed by a flattening step and dense layers. The first convolutional layer consists of 64 filters with a kernel size of 3x3, followed by a max pooling layer, reducing spatial space and computation in the network.

The second convolutional layer includes 128 filters, allowing the model to learn complex features in the images. This layer is followed by another max pooling layer. Then, the network flattens the output and feeds it to the final dense layer of 64 units for classification.

B. *Activation Functions and Optimizers*

- Activation functions play a significant role in the model, with convolutional layers using ReLU (Rectified Linear Unit) to introduce non-linearity and the ability to learn complex patterns. The final layer uses a SoftMax activation function to perform the multiclass classification.

- The model compilation used the Adam optimizer due to its efficiency with sparse and noisy nature of the data. The learning rate was set to 0.001 as suggested by the hyper parameter tuning, providing a good balance between speed and accuracy in convergence. The model uses categorical cross-entropy as the loss function, appropriate for multiclass classification.

C. *Training*

The model was trained on 10 epochs and several callbacks were established to enhance the training process. Model checkpoints were included to save the best model version according to validation accuracy, and early stopping to halt training with a patience of five epochs aiming to prevent from overfitting.

This specific model design ensures that the model is not only tailored to the specific characteristics of the soccer event dataset but is also reliable and efficient in classifying new images accurately.

IV. EVALUATIONS AND TESTING

To assess the performance of the model, a comprehensive evaluation using unseen test data was implemented. The primary metrics used to evaluate the model were accuracy and the precision-recall balance.

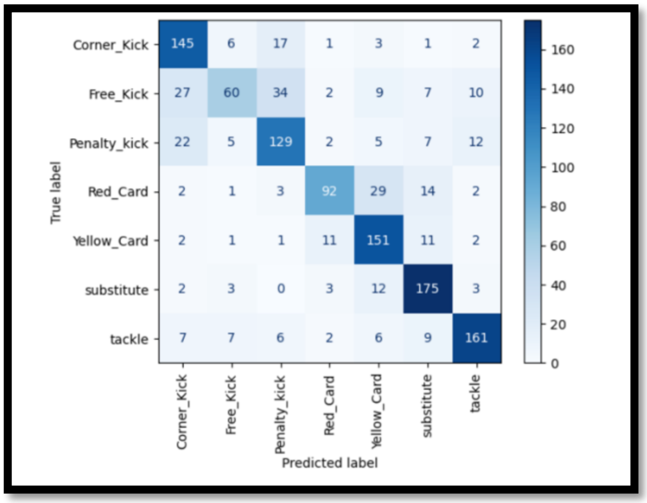
The following Evaluation Report resulted from the model:

	precision	recall	f1-score	support
Corner_Kick	0.7005	0.8286	0.7592	175
Free_Kick	0.7229	0.4027	0.5172	149
Penalty_kick	0.6789	0.7088	0.6935	182
Red_Card	0.8142	0.6434	0.7188	143
Yellow_Card	0.7023	0.8436	0.7665	179
substitute	0.7812	0.8838	0.8294	198
tackle	0.8385	0.8131	0.8256	198
accuracy			0.7459	1224
macro avg	0.7484	0.7320	0.7300	1224
weighted avg	0.7490	0.7459	0.7384	1224

The evaluation and results presented the following findings:

- The overall accuracy of the model is 0.7484, indicating that it correctly classified 74.84% images in the test dataset.
- The model achieved the highest precision for 'Red_Card' (0.8142) and 'tackle' (0.8385), indicating that it correctly predicted this classes without many false positives.
- 'substitute' and 'Free_Kick' achieved precisions of 0.7812 and 0.7229, respectively, indicating a good performance identifying these events.
- 'Corner_Kick' and 'Yellow_Card' achieved relatively high precision values (0.7005 and 0.7023, respectively), indicating a decent performance in identifying these events.
- 'Penalty_Kick' achieved the lowest precision (0.6789), indicating that the model struggled in correctly identifying these events without false positives.

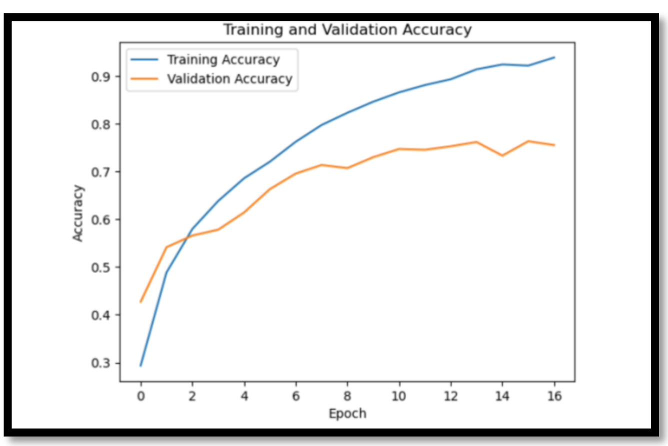
The following confusion matrix resulted from the evaluation and test of the model:



The confusion matrix provided a detailed breakdown of the model's performance across the different classes, which include 'Penalty Kick', 'Free Kick', 'Corner Kick', 'Red Card', 'Yellow Card', 'Substitution', and 'Tackle'.

The confusion matrix proved that the model had a robust ability to identify ‘Tackle’, ‘Substitute’, ‘Yellow_Card’ and ‘Corner_Kick’ events, which was reflected in a high precision and recall for these classes. However, there were challenges with classes that had similar contextual features, such as distinguishing between 'Free Kick' and 'Penalty Kick'. This insight from the confusion matrix has pointed towards areas where the model might need further tuning and improvement, particularly in differentiating between events that occur in closely related contexts within the game.

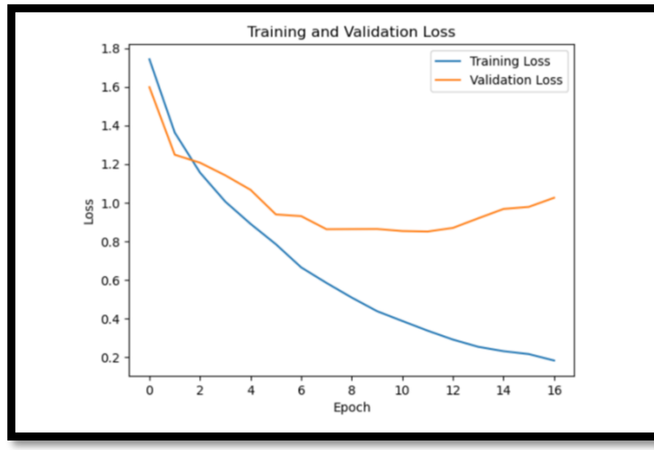
The following training and validation accuracy and loss graphs were analyzed:



Training and Validation Accuracy (Second Graph):

The Training Accuracy consistently increases, which is a sign that the model is effectively learning from the training data.

The Validation Accuracy increases until around 6 epochs, after which it plateaus and slightly decreases. This trend indicates the model has reached its capacity in generalizing from the given feature set and architecture without further tuning or data augmentation.



Training and Validation Loss:

The Training Loss starts relatively high and decreases sharply within the first few epochs, then continues to decrease at a slower pace. This is typical behavior showing that the model is learning from the training data.

The Validation Loss also decreases initially, suggesting that the model was generalizing well initially. However, after around 6 epochs, the Validation Loss starts to increase slightly. This could be an indication of overfitting, where the model is learning the training data too well, including its noise and inaccuracies, and thus performs worse on unseen validation data.

In performance metrics, the model achieved an accuracy score that accurately reflects its ability to classify most test images correctly. This was complemented by the precision, recall, and F1-scores for each category, which were derived from the confusion matrix and indicated a balanced performance across the board.

While the model demonstrates reliable performance across most classes, there is room for improvement, particularly in classes with lower precision and recall values such as ‘Free_Kick.’ Fine-tuning the model architecture, optimizing hyperparameters, and increasing the diversity and size of the training data could lead to improvements in performance. Additionally, further analysis of misclassifications and potential class imbalances could provide valuable insights for model refinement.

V. Conclusion and future Directions

The project successfully developed a sophisticated soccer event image classification system. The project continued with a demonstration of the powerful constructive collaboration between Convolutional Neural Networks (CNNs) and Large Language Models like LLaVA.

Large Language Vision Assistant (LLaVA) was implemented to classify and generate textual descriptions of soccer events. This integration not only improved the accuracy of image classification but also enriched the image context through descriptive text, enhancing user engagement and value. This model allows coaches and analysts to comprehend the classified event's context by receiving a description of the play and players involved and techniques or conditions that can be practiced during training.

The CNN model, trained on a diverse dataset of 7,000 soccer event images, demonstrated robust classification capabilities across seven categories: Corner Kick, Free Kick, Penalty Kick, Red Card, Yellow Card, Substitution, and Tackle. The use of advanced features like double convolutional layers and optimal hyperparameters like a learning rate of 0.001, along with techniques such as ReLU activation and Adam optimizer, were critical in achieving an important level of model performance.

The application of LLaVA for generating image descriptions provided a deeper understanding of the visual content, enabling more nuanced interactions with the users. The personalized soccer event image recommender system developed later in the project leveraged these descriptions to match user preferences with relevant images, thus providing a tailored browsing experience.

Looking ahead, there are several opportunities for further research and development:

- Expanding the Dataset: Incorporating more images, especially from under-represented events, could help in reducing class imbalance and improving model accuracy.
- Model Refinement: Experimenting with more complex architectures like ResNet or Inception networks, or incorporating techniques such as transfer learning, could enhance the model's robustness and accuracy.
- Real-time Classification: Developing capabilities for real-time image classification could extend the system's applicability to live sports broadcasting and online streaming platforms.
- Integration with AR/VR: The system could be integrated into augmented reality (AR) or virtual reality (VR) platforms to provide immersive experiences for soccer training and entertainment.
- Broader Sports Application: Expanding the system to include other sports could open new markets and applications, further leveraging the developed technologies.

This project lays a solid foundation for future innovations in sports technology and artificial intelligence, promising to revolutionize how sports events are experienced by fans and professionals.

References

[1] Karimi, A., Toosi, R., & Akhaee, M. A.(2021). “Soccer Event Detection Using Deep Learning”. 2021.

[2] Olav Andre Nergård Rongved, Markus Stige , Steven Alexander Hicks, Vajira Lasantha Thambawita, Cise Midoglu, Evi Zouganeli 1, Dag Johansen 4, Michael Alexander Riegler and Pål Halvorsen. (2021). “Automated Event Detection and Classification in Soccer”. 2021.