Soccer Event Classification

Greta Angolani
University of Milan
gretangolani@gmail.com

January 2024

Abstract

This paper presents an Image Classification Analysis, which is a fundamental task in vision recognition that aims at understanding and categorizing an image as a whole under a specific label. This project aims to develop a system to correctly classify test images: the model should be able to discern between soccer-related events and general images, and also classify the specific event occurring. For this purpose a custom Convultional Neural Network (CNN) will be used.

1 Introduction

Soccer is one of the most popular sports worldwide, with millions of images and videos shared daily across various platforms. The automatic analysis of sports images can have significant applications, such as improving match analysis and supporting refereeing decisions. In this context, the automatic classification of football-related images represents an interesting challenge, given the complexity of sports scenes and the variability in image conditions.

However, the classification of sports images poses several challenges. Football scenes are often dynamic and complex, with players in motion, varying lighting conditions, and overlapping elements such as spectators and referees. Additionally, images can differ significantly in terms of angles, quality, and context.

In this work, we propose an approach based on a custom Convolutional Neural Network (CNN) for the classification of football images. Our objective is to discern between soccer-related events and general images, and also classify the specific event occurring.

The remainder of this work is organized as follows: in Section 2 Dataset and Methodology are described, in Section 3 the CNN architecture is explained, as well as the classification results obtained. In Section 4 I have explored the classification result of the proposed pipeline, while in Section 5 and 6 we found conclusion and references of the present work.

2 Dataset and Methodology

The analysis is carried out on the Soccer Event Dataset (SEV) , which contains images taken from UCL and EL football. The dataset is structured as follows:

Split Groups	Folder Name	Train DS	Test DS
Split 1	Center	5500	500
Split 1	Left	5500	500
Split 1	Right	5500	500
Split 1	Other	5496	498
Split 2	Corner	5500	500
Split 2	Penalty	5500	500
Split 2	Free-Kick	5500	500
Split 2	Tackle	5500	500
Split 2	To - Substitute	5500	500
Split 2	Cards	5500	500
Split 3	Red - Cards	5500	500
Split 3	Yellow - Cards	5500	500

(a) SEV Dataset

General DS	Images	Train, Validation, Test
Event	1200	Train 60%, Validation 20%, Test 20%
Soccer	1200	Train 60%, Validation 20%, Test 20%

(b) Soccer-Event Dataset

Figure 1: Overview of SEV Dataset and Soccer-Event Dataset

As shown in image (a), the dataset was divided into three main folders, named "Split 1," "Split 2," and "Split 3," to train three different models. Meanwhile, the General Dataset (b) was used to train the first model in the pipeline. Additionally, a folder named "Other" was intentionally created (it was not part of the original dataset) by randomly selecting images from the six sub-folders of Split 2. Specifically, for the folder "Other" in the training dataset, 916 images were taken from each of the five sub-folders, resulting in a total of 5.496 images. For the folder "Other" in the test dataset, 83 images were taken from each of the five sub-folders, resulting in a total of 498 images. The structure of the proposed pipeline is the following:

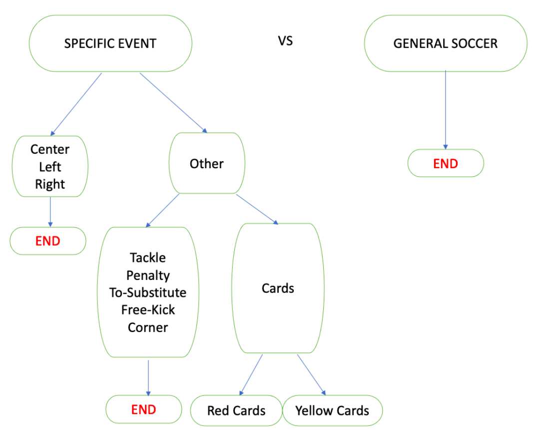


Figure 2: Pipeline

As we can see in figure [2]:

1. Model 1: aims at distinguishing between general soccer images and specific soccer event; if the prediction is "Soccer" the image doesn't proceed along the pipeline.

2. Model 2: if the predictions of Model 1 returns "Event" the images proceed to the second model for being classified into "Center", "Right", "Left", or "Other". If the model predicts "Other", the image proceeds to the third classifier, otherwise stops at the second, since not referring to an action event.

3. Model 3: this model ability is to classify the event in the correct category, Tackle, To-Substitute, Corner, Free-Kick, Penalty and Cards. If the prediction falls into the class "Cards" the image proceeds to Model 4, otherwise it stops.

4. Model 4: the last model works on the distinction of the images classified as "Cards", predicting if it display a "Red-Cards" or a "Yellow- Cards".

The purpose of adding a new folder "Other" was to train the second model to distinguishing between real Events and images that are to be considered as No-Highlights.

3 CNN classification

3.1 Pre-Processing

The first step was to remove corrupted images and check the data distribution. This is a crucial step because an imbalanced dataset can lead to overfitting or underfitting during the training process. In figure (3) and (4) we can see the distributions after the cleaning step and, as we can see, the dataset is still balanced. The remaining, non-corrupted images were resized to conform to the 224x224x3 dimensions. This resizing process was applied to the training dataset, test dataset, and the general dataset.

The training dataset was further divided into 80% for training and 20% for validation, enabling model performance monitoring and helping to prevent overfitting. For testing, the provided test set was used. Instead the general dataset, since no dedicated test set was provided, has been split into 60% for training, 20% for validation, and 20% for testing.

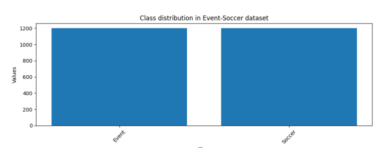
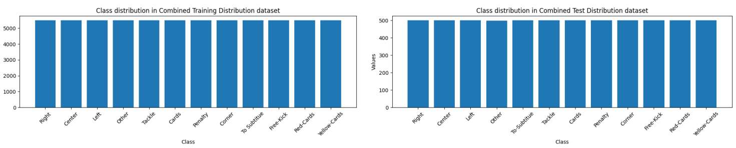


Figure 3: Soccer-Event



(a) Training set (b) Test set

Figure 4: Model 1: Train and Test set distributions

3.2 CNN models

To address the task of classifying images, four models based on Convolutional Neural Networks (CNNs) were designed, each tailored to a specific classification problem. The models follow a systematic approach to extract relevant visual features from the input images and use these features to distinguish between the target classes.

Initially, all input images are resized to a fixed shape and normalized, scaling pixel values between 0 and 1. This preprocessing step ensures efficient and stable data handling across all models. Optionally, data augmentation techniques, such as random flipping, rotation, and zoom, are applied to increase dataset variability and improve the generalization of the models.

The core of each model consists of a series of convolutional and pooling layers. The convolutional layers use filters with a ReLU activation function to detect increasingly complex visual patterns, starting with basic edges and progressing to higher-level features. After each convolutional block, max-pooling layers reduce the dimensions of the feature maps, focusing on the most significant details while lowering the computational load.

Once feature extraction is complete, the data is flattened into a single vector and passed through a dense layer with 256 units. This dense layer, equipped with ReLU activation, captures high-level abstract representations of the features. Regularization techniques, such as L2 regularization and dropout, are applied to mitigate overfitting and improve the models' ability to generalize to unseen data.

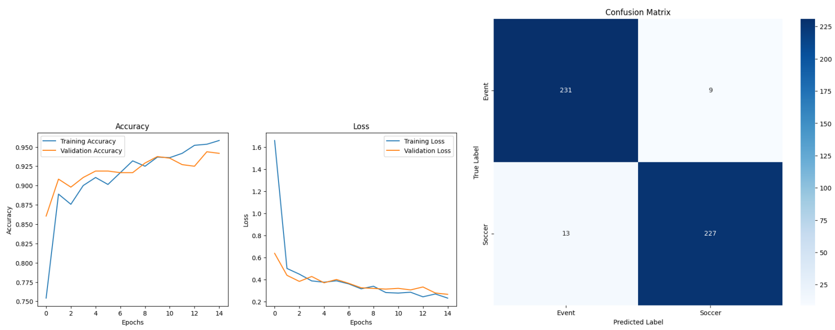
Each model is finalized with an output layer tailored to its specific task:

General Classifier: Outputs a single probability using a sigmoid activation function, classifying images as either "Soccer Image" or "Event Image." No Highlights Classifier: Utilizes a softmax activation function in the output layer to classify images into one of four categories: "Center," "Left," "Right," or "Other." Event Classifier: Outputs probabilities for six distinct classes: "Corner," "Tackle," "Cards," "Free-kick," "Penalty," or "Substitute," using a softmax activation function. Card Classifier: Focuses on distinguishing between "Red-Cards" and "Yellow-Cards" with a single output neuron and a sigmoid activation function.

3.3 Classification Results

3.3.1 Model 1: *Event - Soccer*

Figures 3 depict the training history and confusion matrix for this classification, reaching a 95% of accuracy on the test set.

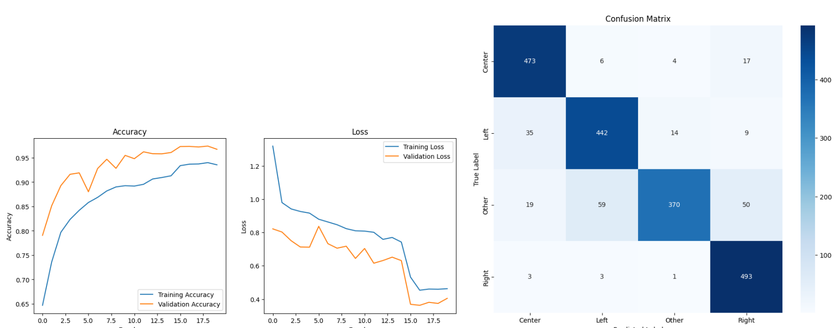


(a) Training History (b) Confusion Matrix

Figure 5: Model 1: Event - Soccer

3.3.2 Model 2: *Right - Left - Center - Other category*

The results obtained are satisfactory with a test accuracy of 88.98%. It works very well for Center Left and Right, while for Other category some images are wrongly classified. Analyzing the dataset we can see how "Left" and "Right" classes depict corner kicks, free kicks, or tackles, rather than no-highlights game-play situations. This adds some confusion to the classification, as the "Other" category is composed of images sourced from the six main event categories, leading to overlap and misclassification.



(a) Training History (b) Confusion Matrix

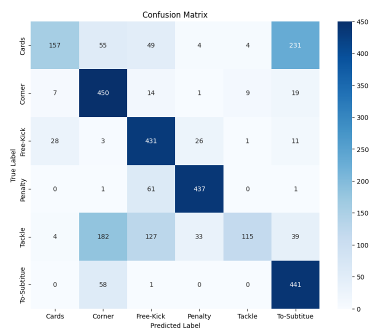
Figure 6: Model 2: Center - Left - Other - Right

3.3.3 Model 3: *Tackle - Free Kick - Penalty - Corner - To Substitute - Cards*

The accuracy achieved by Model 3 is 67.69%. These classes were the most difficult to classify, as there are some differences between the images in the training and test datasets. For example, for the "Tackle" class, most of the images in the training set resemble the one shown in Figure 6(a), while in the test dataset, some images are similar to the one in Figure 6(b). Instead, for the "Cards" class, the images in the training dataset have a higher resolution of the action, whereas the images in the test set have lower quality. This could have led to incorrect classification.

In summary, Model 3 performed reasonably well, although its performance, especially for the 'Tackle' and 'Cards' classes, could be improved.

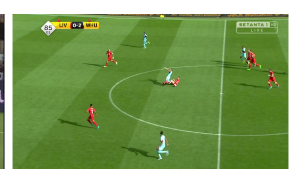
(a) Training History



(b) Confusion Matrix

Figure 7: Model 3: Cards - Corner - Free Kick - Penalty - To Sobstitute - Tackle

(a) Train dataset: Tackle



(b) Test dataset: Tackle

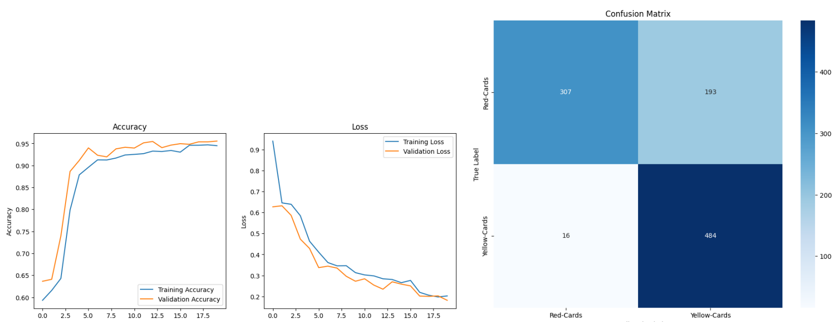
Figure 8: Images in comparison: Tackle

(a) Train dataset: Tackle (b) Test dataset: Tackle

Figure 9: Images in comparison: Cards

3.3.4 Model 4: ***Red Cards - Yellow Cards***

The accuracy for Model 4 is 79.10%, which is a good result. Moreover, the model performed well for the "Yellow-Cards" class, where most of the images were correctly classified. However, the classification for "Red-Cards" was not as good, as seen in the confusion matrix in Figure 8(b).



(a) Training History (b) Confusion Matrix

Figure 10: Model 4: Red Cards - Yellow Cards

4 Pipeline classification result

The trained models were concatenated into a pipeline at this step. Below are some images, downloaded from the website, that have been classified. As we can see, for all the selected images, the results are correct.

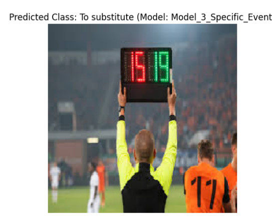
(a) True Label: Corner

(b) True Label: Free-Kick

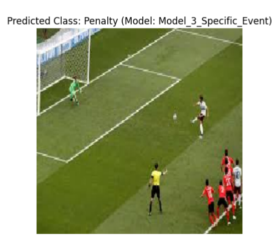
(a) True Label: Red-Cards

(b) True Label: Yellow-Cards

(a) True Label: General Soccer



(b) True Label: To-Substitute



(a) True Label: Penalty

(b) True Label: Tackle

5 Conclusion

In conclusion, the results obtained in this work are very good and satisfactory. Having greater similarity between the images in the training set and the test set, in terms of quality and zoom level, could have potentially led to even better results. Additionally, analyzing a larger dataset or using pre-trained models could further improve performance. However, this was not the primary focus of the project, which prioritized building a custom architecture and analyzing its performance. These approaches, nonetheless, present valuable opportunities for future work.

6 References

Karimi, A., Toosi, R., Akhaee, M. A. (2021). Soccer event detection using deep learning

Tensorflow tutorials

Research on sports image classification method based on SE-RES-CNN model

AI in Sports: Transforming the Game for Players and Fans