Football Event Detection using CNN

Andrew Thomas Costa

December 2023

Sommario

The detection of an event in images is an important tool, especially in the game of football. This paper proposes an exploration of deep learning methods in order to detect various events in a game of football using just images - for example, the detection in the difference between yellow and red cards within an image, as well as other various scenarios.

Indice

Elenco delle figure

Elenco delle tabelle

1 Introduction

Football is the most popular sport in the world. The English Premier League, for reference, broadcasts to 643 million homes and has a potential total audience of 4.7 billion viewers. In addition to this, the English Premier League alone generated £5.5 billion (GBP) in 2021/2022 alone. Given the commercial importance of football, the relative success of teams becomes even more important and this can be achieved through gaining a tactical edge on the pitch.

This paper will examine the application of different deep learning models in order to classify different scenarios from a football pitch; scenarios from which different teams can exploit in order to gain a tactical edge or simply for the sake of data collection and analysis. The traditional method would be to use manpower, however this can be costly, both temporally and financially. With the use of machine learning applications (ML) this data collection process can be automated, saving both time and money.

A Convolutional neural network stands out for its superior performance for certain tasks such as image and audio recognition as it is able to handle non-linear data, whilst also possessing the ability to understand data features with little human oversight.

2 Exploratory Data Analysis

In this dataset, the images were sourced from Soccer Event Detection Using Deep Learning, Karimi et.al (2021). The images consist of 10 different events that can take place on the football pitch, those being: Cards, Center, Corner, Free-kick, Left, Right, Penalty, Red Cards, Yellow Cards, Tackle.

The original dataset consists of a total of 60,000 images, with 6,000 images. For each train, validation and test split, there consists of 5000, 500 and 500 images respectively. Nevertheless, for this paper, the total number of images were reduced due to the absence of computational power. As a result, only 1500 images were used in the training split and 500 images for the testing set. The dataset remained balanced as each category had the same number of images. One difficulty would be that the model does not have enough images to learn the correct patterns and therefore might misclassify images more frequently.

Figura 1: Sample of Football Events

From solely observing the images in Figure 1, it will be difficult to distinguish between certain events. For example the camera angles of 'Left', 'Right' and 'Center' given the similar colour compositions/structure of their respective images. Also, the distinction between the different colour cards (Yellow, Red) will be an arduous task, as the target area in the image – the card – is very small.

3 Model Background

Convolutional Neural Networks (CNNs) will be used to tackle this classification problem. The aim of a CNN and the reason behind its success with image data falls down to its ability to systematize the idea of spatial invariance. Spatial invariance is the concept that a network can detect features/objects even if they are not accurately represented in the training set. So lets say we want to learn the characteristics of a specific picture through the processing of its image structure, a CNN will learn this representation with few parameters.

3.0.1 Activation Function

The activation function of choice is the Rectified Linear Unit (ReLU). At each layer in the model, an affine transformation is performed by applying the ReLU function, eliminating any negative values. After performing the ReLU operator a number of times, we are able to linearly separate the data

The limitations of this ReLU operator are that if a bias, in one of the 5 layers, places a data point not in the top-right quadrant (i.e. not non-negative), then the ReLU will be under performing and everything will collapse to zero. Other functions like the sigmoid cannot be used in a model with many layers due to the vanishing gradient problem.

3.0.2 Max pooling

The function of this layer is to reduce the complexity of the model whilst extracting local features. More specifically, it acquires the local features generated by convolving a filter over an image. Max Pooling does this by finding the maximum values for each 2x2 'pool' and progressively reduces the spatial size of the representation, which as aforementioned, ties into reducing the overall complexity of the model.

3.1 Optimiser

Adam is a popular optimization algorithm that combines the benefits of two other widely used optimizers: AdaGrad and RMSprop. It is known for its ability to adaptively adjust learning rates for each parameter in the model, which can be especially beneficial when dealing with complex, high-dimensional data.

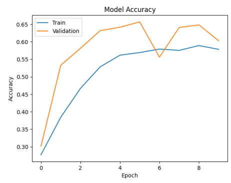
4 Models

4.1 Model 1: Architecture

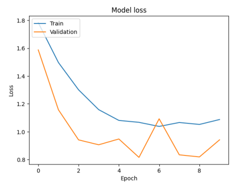
The architecture of the first implemented neural network consists of five convolutional blocks, the ReLU activation function and the Adam optimiser.

4.2 Model 1: Results

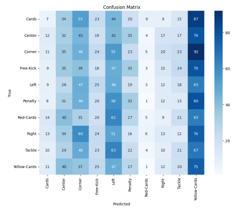
The following graphs show the learning behavior over the training period. One can see that the model is slightly overfitting as the validation accuracy is higher than the training accuracy. The confusion matrix (c) shows that the model fails to distinguish between yellow cards and other types of events.



(a) Model Accuracy



(b) Model Loss



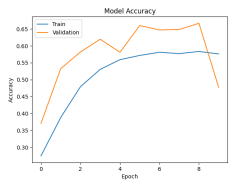
(c) Confusion Matrix

4.3 Model 2: Architecture

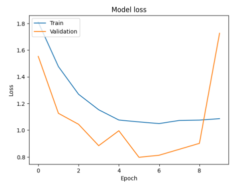
The sole difference between this model and the previous one is that the convolutional blocks contain more filters to extract a higher number of abstractions from the images and therefore better distinguish between yellow cards and other events.

4.4 Model 2: Results

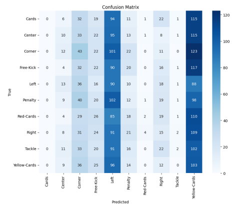
The model accuracy and loss for the second model shows how the validation accuracy and validation loss both decrease and increase respectively at the eighth epoch, another sign that overfitting still persists. The third and final model will look to rectify this overfitting.



(d) Model Accuracy



(e) Model Loss



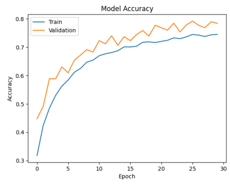
(f) Confusion Matrix

4.5 Model 3: Architecture

As the models still fail to handle the overfitting problem, a final model iteration would see to reduce the complexity of the model whilst giving it more time to train over the dataset, so that it can identify the patterns within the dataset that will allow it to generalise more effectively.

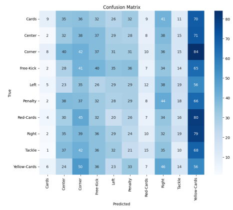
4.6 Model 3: Results

The model shows a smoother training accuracy and loss curve when simplifying the model type and increasing the training epochs. In addition to this, it follows the validation curves for both the accuracy and loss graphs very closely, indicating a reduction in underfitting. Nevertheless, despite these improvements, the confusion matrix shows that the model still struggles to distinguish between yellow cards and other types of football events.



(g) Model Accuracy

(h) Model Loss



(i) Confusion Matrix

5 Conclusion

In conclusion, this study delves into the application of deep learning models, particularly Convolutional Neural Networks (CNNs), in order to classify various football scenarios that can occur on the pitch, Recognising the significance of gaining a tactical edge in football, whilst simultaneously automating the data collection process, this research explores the feasibility of automating data collection and analysis through machine learning applications; specifically focusing on the cost-effective and efficient nature of CNNs.

The dataset underlines the challenges of distinguishing between specific football events, due to the similarity in pictures and the absence of large computing resources. Despite trying to rectify the resource issue by reducing the size of the dataset, the model still failed to properly and accurately distinguish between each football event.

Three models were proposed during this research, each one iteratively refining the architecture to mitigate overfitting and correctly classify each football event. Model 1 demonstrates overfitting, particularly in the confusion matrix where yellow cards are frequently misclassified. Model 2 increases the number of filters to improve abstraction but still exhibits overfitting. Model 3, whilst showing smoother training curves and reduced overfitting, continues to struggle in distinguishing yellow cards from other events.

In summary, the research underscores the challenges and complexities of employing deep learning models for football event classification. The iterative model development reveals improvements in training behaviour, but the persistent difficulty in distinguishing between specific events, such as yellow cards, highlights the need for further refinement and exploration in future research endeavours.

We declare that this material, which I/We now submit for assessment, is entirely my/our own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my/our work. I/We understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. This assignment, or any part of it, has not been previously submitted by me/us or any other person for assessment on this or any other course of study.