Soccer Event Detection

Jad YEHYA

Università degli Studi di Milano

June 27, 2023

Table of Contents

Introduction

The goal of this project was to detect events in soccer. The events we are interested in are the following:

- Corner
- Free Kick
- Red Card
- Yellow Card
- Penalty
- Substitution
- Tackle

We used different techniques and models to detect these events.

Methodology

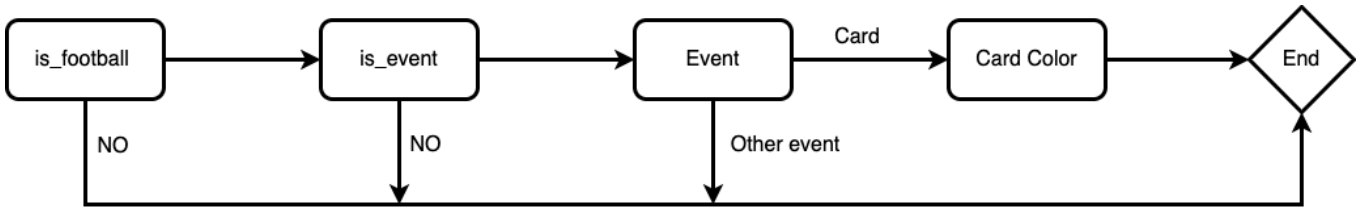
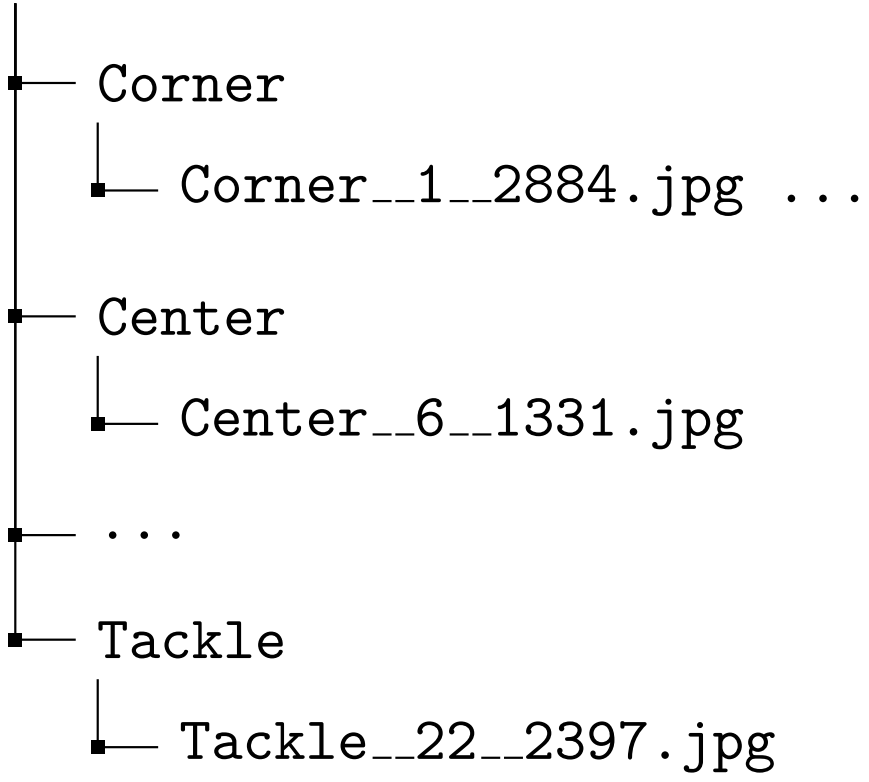


Figure: Architecture of the proposed method

We will see in detail every step of pipeline, from the detection of football to the end. We fine tuned every model except for the VQ-VAE on the data we have. It was not concievable to train all the deep learning models from scratch since they are trained on supercomputers and on a lot of data. That's why we used transfer learning.

Dataset

First step of the pipeline was to get the data and format it the right way. After all the data downloaded, we organized it the following way. We used PyTorch's `ImageFolder` class to load the data. Every element has the folder name as label.



Looking at the data

Figure: Quick look at the images

Every image is a jpeg coloured image of size 224x224.

Models

For this complicated task that we want to achieve, we divided it into **4 subtasks**. And for every subtask, we used a different model.

Vector Quantized Variational Autoencoder (VQ-VAE)

For the detection of the football, we used a VQ-VAE. The goal of this model is to encode the image into a latent space and then decode it back.

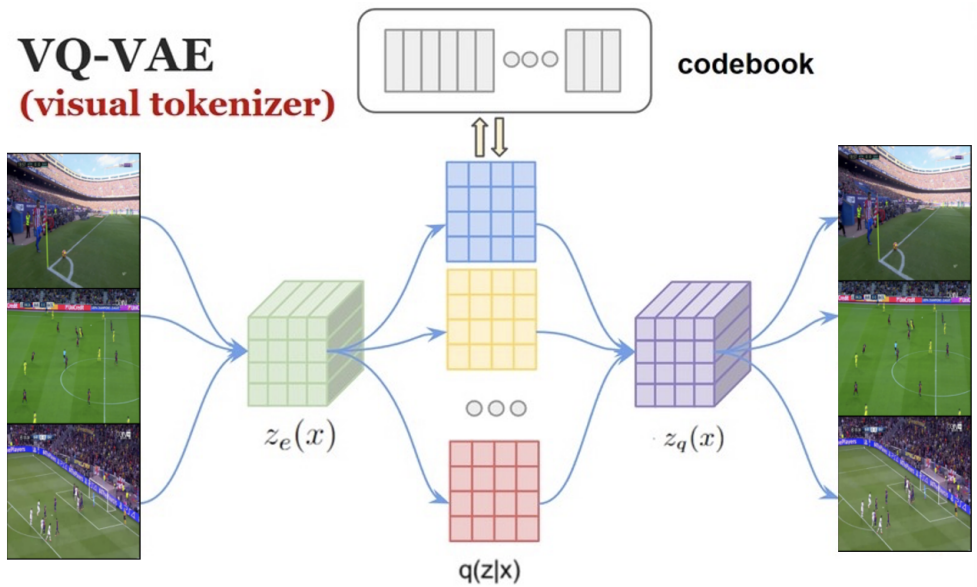
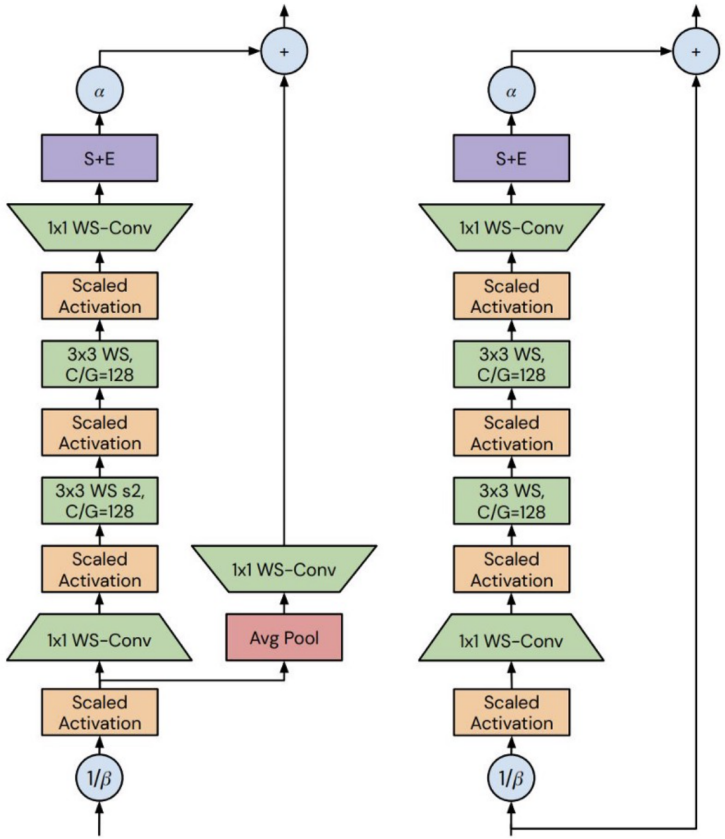


Figure: Overview of VQ-VAE model

NFNet

To detect football events froom non football events, we used a NFNet. NFNet is a new model that was introduced in 2021. It is a very powerful and efficient.



NF-Net architecture. Image taken from paper. Left-Image depicts transitional block and right-Image depicts non-transitional block.

Figure: Architecture of NFNet

Visual Transformer (ViT)

This model is the core of the pipeline. It is the model that will detect the events. We used a Visual Transformer (ViT) for this task, we saw in class the way it leverages the attention mechanism for images.

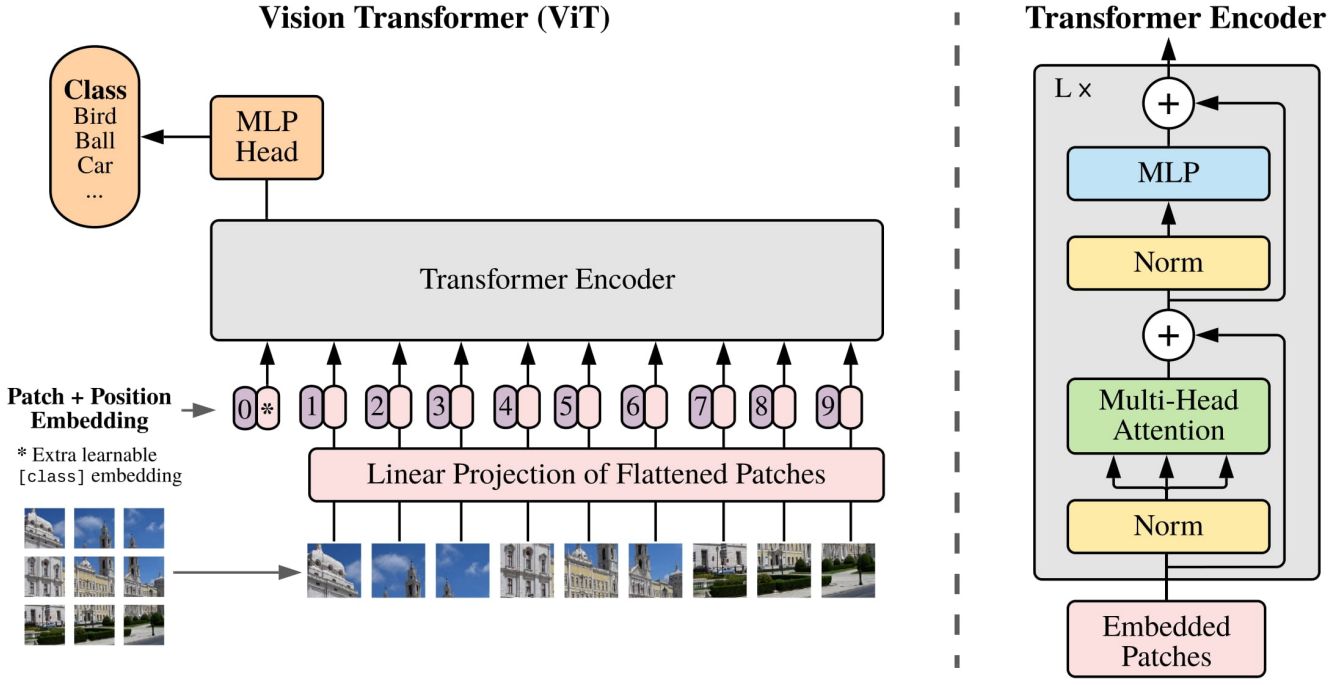


Figure: Architecture of ViT

BCNN

This last step is directly taken from the paper of the project, since they found that seperating the classification problem of the colour of the cards to an individual subtask improved the performance. We chose to use a BCNN for this task since it is used for fine grain classification.

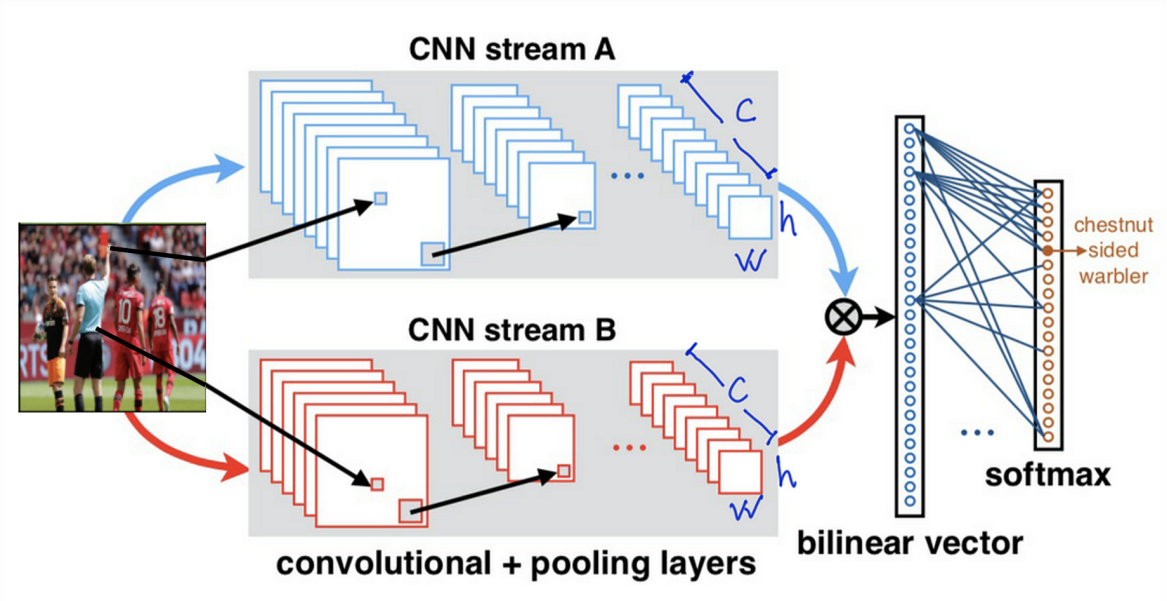


Figure: Architecture of BCNN

Other approach ideas

Even though these ideas were not used, they are still worth mentioning.

- Utilizing spatial information for the last step. We know that the cards are almost always in the top of the image, so we thought of croping the image for only the top part. But keeping the image as it is, the model can also learn this information.
- Greedy search for the cards.
- Using N small networks

Other approach ideas

Even though these ideas were not used, they are still worth mentioning.

- Utilizing spatial information for the last step.
- Greedy search for the cards. Searching for a small square/rectangle red or yellow in the top part of the image.
- Using N small networks

Other approach ideas

Even though these ideas were not used, they are still worth mentioning.

- Utilizing spatial information for the last step.
- Greedy search for the cards.
- Using N small networks We thought about initalizing N small binary networks, each one for a specific event. This idea came by trying to do a specialized model for every event.

Experimental Results

We did a lot of experiments, we will see the most important ones. First, we will see the hyperparameters that we used for the models.

Table: Hyperparameters used

Batch Size	Optimizer	Learning Rate	Loss Criterion	Transforms
64[1]	Adam	1e-3	Cross entropy	Resize to 224*224, Normalize

[1] Tests with 128 were made when it was possible for the GPU.

VQ-VAE

Reconstruction of the VQ-VAE :

Figure: Reconstruction of the VQ-VAE

We end up with a reconstruction loss of 0.082 for football images.

VQ-VAE

Reconstruction of the VQ-VAE : We end up with a reconstruction loss of 0.082 for football images.

Visualization of Reconstructed images

Figure: Reconstruction of the VQ-VAE

NFNet

The NFNet had very good results. It was implemented in order to divide again the big task into subtasks. It achieved an almost perfect accuracy on the validation set.

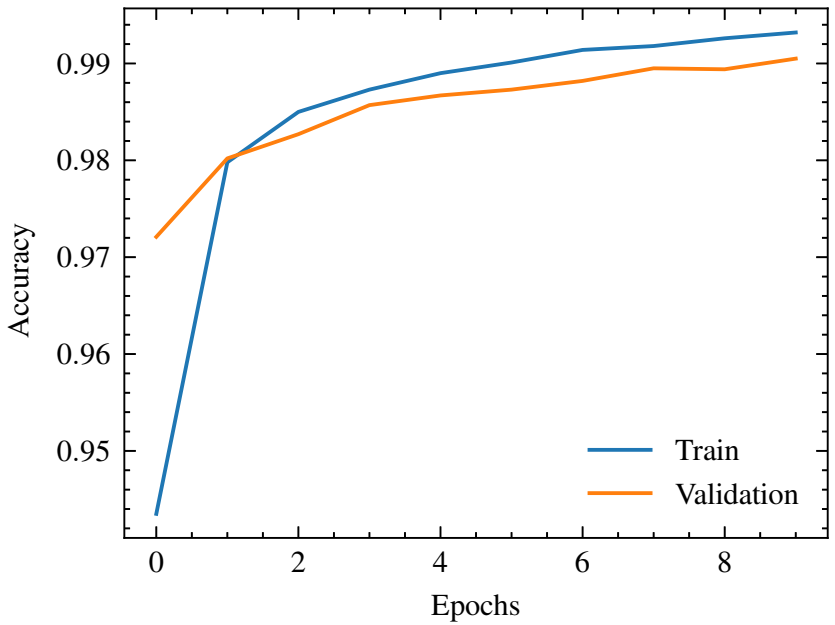


Figure: Accuracy of NFNet

ViT

This is the biggest cell of the pipeline and it is the one that took the most time to train and fine tune.

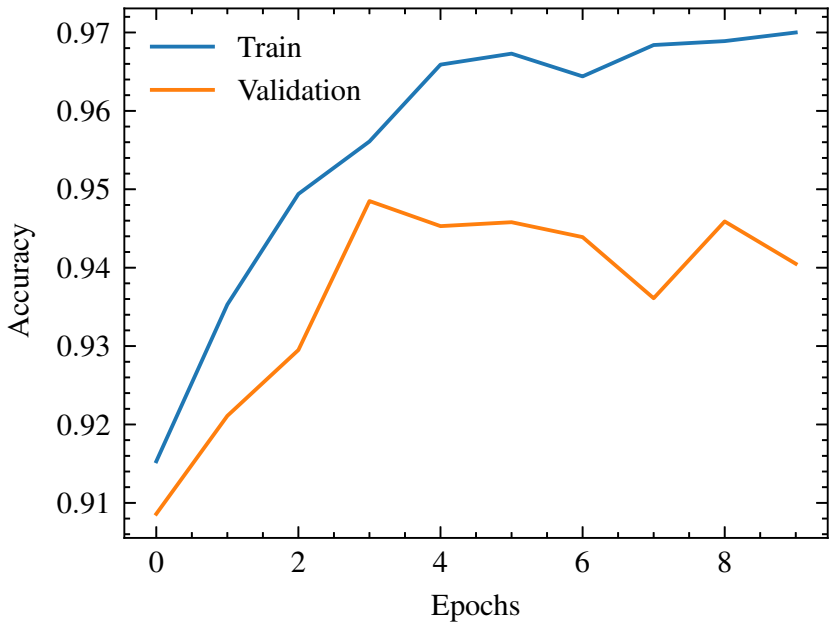


Figure: Accuracy of ViT

Fine-grain module

This fine grain classification for the colour of the cards is the hardest task as of today. It is still falling behind in terms of accuracy. It is also not the most stable, the accuracy is not always increasing etc.

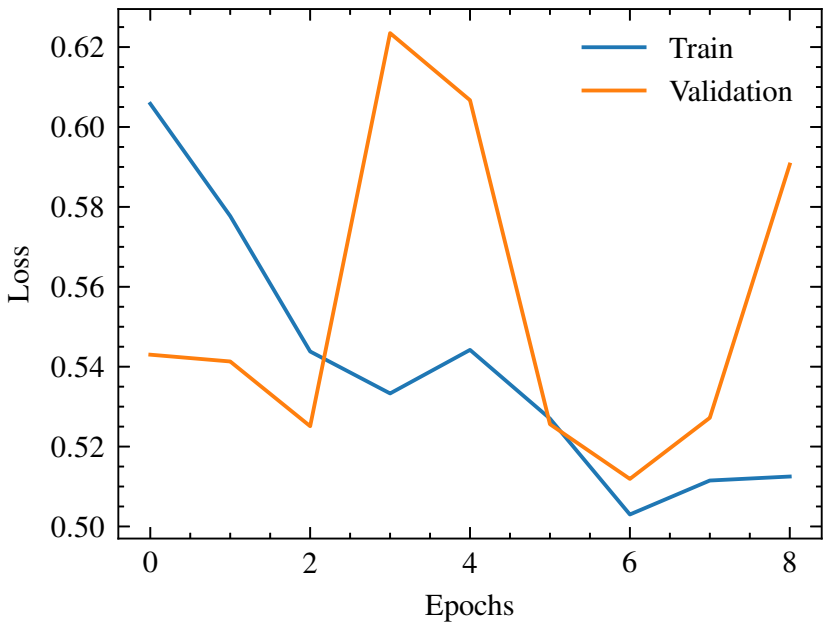


Figure: Accuracy of Fine-grain module

Final results

Table: Final results

Task	Accuracy	Precision	Recall	F1-score
Detecting the type of event (ViT)	0.935	0.879	0.878	0.84
Final	0.28	0.27	0.281	0.24

Conclusion

- We have seen that the pipeline is working and that it is possible to detect the type of event in a football match.
- We have seen that the pipeline is not perfect and that there is still improvements to be made.

Conclusion

- We have seen that the pipeline is working and that it is possible to detect the type of event in a football match.
- We have seen that the pipeline is not perfect and that there is still improvements to be made.
- First project in comptuer vision, in deep learning, in PyTorch.
- Very enriching project, read a lot of papers about the state of the art in computer vision.