INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Artificial Intelligence in Padel sport: Creation of an AI System

João Rodrigo Ferreira Fernandes de Castro

Master’s in Information System Management

Supervisor:
PhD Ricardo Daniel Santos Faro Marques Ribeiro, Associate Professor, ISCTE-IUL

Co-supervisor:
PhD Rui Jorge Henriques Calado Lopes, Associate Professor, ISCTE-IUL

October, 2023

TECNOLOGIAS
E ARQUITETURA

Department of Information Science and Technologies

Artificial Intelligence in Padel sport: Creation of an AI System

João Rodrigo Ferreira Fernandes de Castro

Master’s in Information System Management

Supervisor:
PhD Ricardo Daniel Santos Faro Marques Ribeiro, Associate Professor, ISCTE-IUL

Co-supervisor:
PhD Rui Jorge Henriques Calado Lopes, Associate Professor, ISCTE-IUL

October, 2023

Acknowledgements

I would like to leave a word of gratitude to all the people involved in this project who, in one way or another, contributed to its success.

To my supervisors, Professor Rui Lopes, Professor Ricardo Ribeiro, and Professor Eliano Marques, who always helped me by indicating the correct path for this dissertation to be completed successfully.

To my parents, who were a huge support during this journey and for their support in all the most complicated moments.

To my brothers, for the sensitivity shown and for being part of this from the beginning.

To my girlfriend, who was in the same situation and was mutual support, in terms of motivation and for her affection and encouragement.

To all my family who is fundamental in my life.

To all my friends and co-workers who contributed to my personal and professional growth.

To all whom I have listed my sincere "Thank you".

Resumo

O Padel é um desporto que tem vindo a despertar a curiosidade de todos e, por isso, a procura pelo mesmo é elevada. Cada vez mais têm sido construídos campos e recintos desportivos para a prática desta modalidade. Como tal, é uma área ainda pouco explorada no que toca à tecnologia e aos modelos de previsão de *Machine Learning*. É possível gerar estatísticas num jogo de Padel, tal como acontece noutras modalidades, e esta dissertação propõe uma estratégia para desenvolver um modelo de *Machine Learning*, de modo a prever dois movimentos referentes ao jogo, nomeadamente, os *Left Volleys* e as *Bandejas*. Também é fundamental identificar a trajetória da bola para fazer uma análise pormenorizada, assim como detetar as linhas do campo de Padel, de forma a criar um sistema de Inteligência Artificial capaz de prever os movimentos feitos pelos jogadores e gerar estatísticas do mesmo.

Assim, o objetivo geral deste trabalho é conceber e desenvolver um modelo de *Machine Learning* para otimizar o jogo de Padel, que juntamente com a deteção do campo e com a identificação da bola, irá constituir o sistema de Inteligência Artificial. A metodologia divide-se em três fases. A fase exploratória que diz respeito à recolha de informação dos dados e à construção de pipelines de dados para extrair *insights*, desenvolver modelos e testar variáveis utilizando *feature engineering*. Em seguida, a fase de modelação/automação da IA para implementar no sistema. A terceira fase corresponde à análise de resultados para integrar a IA em qualquer sistema em tempo-real.

Palavras-Chave: Inteligência Artificial, *Machine Learning*, Padel, YOLO

Abstract

Padel is a sport that has been arousing everyone's curiosity and, therefore, the demand for it is high. More and more sports fields and venues have been built for the practice of this modality. As such, it is still an underexplored area in terms of Machine Learning technology and prediction models. It is possible to generate statistics in a Padel game, just like in other sports, and this dissertation proposes a strategy to develop a Machine Learning model to predict two movements relating to the game, namely *Left Volley*s and Trays. It is also essential to identify the trajectory of the ball in order to carry out a detailed analysis, as well as detect the lines of the padel court, in order to create an Artificial Intelligence system capable of predicting the movements made by the players and generating statistics on them.

Thus, the general aim of this work is to design and develop a Machine Learning model to optimize the game of Padel, which together with the detection of the court and the identification of the ball, will generate the Artificial Intelligence system. The methodology is divided into three phases. The exploratory phase involves gathering information from the data and building data pipelines to extract insights, develop models, and test variables using feature engineering. This is followed by the AI modeling/automation phase to implement in the system. The third phase is analyzing the results to integrate AI into any real-time system.

Keywords: Artificial Intelligence, Machine Learning, Padel, YOLO

Index

Tables Index

Figures Index

List of abbreviations

AI – Artificial Intelligence

ANN – Artificial Neural Network

AP – Average Precision

BB – Bounding Box

CNN – Convolutional Neural Network

CPU – Central Processing Unit

CVAT – Computer Vision Annotation Tool

DL – Deep Learning

FPS – Frames Per Second

GPU – Graphics Processing Unit

HOG – Histogram of Oriented Gradients

HPE – Human Pose Estimation

KCF – Kernel Correlation Filters

ML – Machine Learning

R-CNN – Region-based Convolutional Neural Network

SGD – Stochastic Gradient Descent

SIFT – Scale-Invariant Feature Transform

SSD – Single Shot MultiBox Detector

YOLO – You Only Look Once

Chapter 1 – Introduction

1.1. Topic Context

The sport Padel was originally created in Acapulco, Mexico, in 1969 by Enrique Corcuera. It is a sport played in pairs with rackets, using solid rackets (without net) and balls. It is usually associated with a mixture of tennis and squash due to their common characteristics. The court is rectangular, surrounded by smooth glass panels, with a metal net ten meters wide by twenty meters long, and a net in the middle of the court. The surface of the court can be made of synthetic grass, carpet, or porous concrete, and can be played on indoor, outdoor, or semi-undercover courts. Point scoring has many similarities with tennis, namely in the scoring system. The main difference is that the court is surrounded by glass, and the ball can bounce off these surfaces and then be played, in a similar way to squash.

It is currently the fastest growing sport globally because it is easy to learn and does not demand high physical and technical requirements. Furthermore, the way it is organized, by levels, allows different players to play with other players of the same level and to have balanced games. With the increasing evolution of AI in our daily lives and in the evolution of human society, it is relevant to study and understand, increasingly, what is the impact of this AI System on the world's most emerging sport. Although it is a technology that is not within everyone's reach and not easily understood, it is also a revolutionary technology that could change the way every game is played.

The optimization of the Padel game is fundamental to better understand it and be able to improve all the surrounding aspects, such as training, based on the statistics obtained. With the introduction of Computer Vision in Padel it will be possible to observe which team won the point, the number of serves that each player hit and missed, the number of times they fail when trying a specific movement, such as a *Volley* or a *Bandeja*.

1.2. Motivation and topic relevance

Technology and AI are increasingly influencing our daily lives, going from a small help to a disruptive growth, thus becoming something much more powerful than we have ever witnessed before. As such, it is crucial to understand its impact on our lives and discover its full potential. Stephen Hawking said that the successful creation of Artificial Intelligence effectively could be the greatest event in the history of our civilization or the worst. We simply don't know this yet. This means that there is still a lot to be discovered in this area, to unveil how far Artificial Intelligence may or may not go and how relevant we will or may not be in its evolution, or whether we will simply be overtaken.

Considering that Padel is the sport with the biggest growth today, it will be interesting to understand how to associate the technological component, such as Machine Learning models that will provide detailed information about each player's statistics, so that a player will know exactly what he can improve through training, to obtain better results in the game. When we talk about implementing Artificial Intelligence in sports, it is usually to help in the decision-making process, be it of the technical team, the coach, or the referee himself. In this case, it is important to realize that players are not usually so monitored, that is, they play friendly or training games to improve their skills and technique. For this reason, it will be an added value for each practitioner to have access to a detailed report at the end of each game with the number of services that he missed/kicked, the number of smashes he missed/kicked, etc.

Through this idea it is possible to advance to other types of objectives, such as, for example, drawing a profile of each player, through the Machine Learning model, to be able to understand what are the possibilities of winning or losing the game, eventually, external people can make bets based on these predictions, etc.

1.3. Questions and research goals

Given the problem considered, this work aims to optimize the game of Padel through images obtained from videos for the construction of Machine Learning models to generate statistics and understand when the ball, after a hit, was inside or outside the court, thus understanding who won the point. This is intended to answer the research question of this study:

How Artificial Intelligence can generate intelligence player performance analysis for the Padel game?

In this way, and to deepen the investigation, this dissertation has set out some general objectives and some more specific ones.

Regarding the general objectives, the main one is to build the AI system and, consequently, implement a machine learning model for detecting movements such as the *Bandeja* and the *Left Volley*. The general objectives also include detecting the Padel court, identifying its lines in order to obtain a plan, and, finally, detecting the Padel ball. Once these objectives have been implemented, it will be possible to unite all these components to form the AI system.

Once the AI system has been built, it will be possible to achieve more specific objectives, such as generating statistics for each player, like the number of times a player has hit a *Bandeja*, how many times this move has been made during the game, in which position on the court this move is most common and the same can be verified for *Left Volley*.

1.4. Methodologic approach, structure and organization of the dissertation

This dissertation has a typical quantitative approach, one that contributes to developing, validating, and generalizing results and knowledge. It considers all data to be quantifiable, translated into numbers, and information that can be classified and analyzed. The starting question plays a guiding role throughout the research and in interpreting the results, but it is important to remember that no research is "purely" quantitative or qualitative. About the structure of the research process, it is considered to be applied and applicable research, because it contributes to solving practical problems in the short and medium term, with proposals for visible applicability. A good design eliminates bias and increases the reliability of the data collected and analyzed.

To achieve the objectives intended for this dissertation, it is divided into four phases. 1) The first phase was to explore the models and algorithms best suited to this system, which in this case was YOLO, and how it could fit into the context of the game of Padel, through the classification of patterns and images. 2) The second phase was processing and cleaning the data to insert into the YOLOv8 model as input. Here we created the annotations that were associated with a particular class, either the Tray or the *Left Volley*. This is so that the model can learn to identify the movement being made in a particular frame of the video. In this way, it took a few tests to minimize the number of different annotations for a class, i.e. the aim was to identify something unique in each movement and this will make the model better at identifying each hit. 3) The third phase involves formulating the machine learning models and preparing the data for training, testing and validation in order to train the model and test it on unused data to validate whether it can identify the desired movement correctly. 4) The last phase consists of analyzing the model's results and testing them against some metrics used to determine whether they are correctly identifying the desired movements and whether they are performing well overall, identifying data that you have never seen before.

To achieve the proposed objectives and answer the questions presented, the development of this research is structured as follows.

The first chapter introduces the subject of the research and objectives of the same, as well as a brief description of the structure of the work.

In the second chapter, a literature review is carried out on some concepts about various topics such as the algorithms and models that were analyzed and studied, as well as the necessary inputs and respective requirements.

In the third chapter, reference is made to the methodology used and specifying the model used and its algorithms, to answer the questions proposed above.

In the fourth chapter, a detailed results analysis is presented, to justify the algorithms presented in the Machine Learning models.

The fifth and last chapter will present the conclusions of this study, as well as the recommendations and limitations for future work.

Chapter 2 – Literature review

2.1. Padel sport history and its movements

Sport is a physical activity that is subject to certain rules and aimed at competition. Although physical ability is one of the key factors for the result of the sport, there are other equally decisive factors, such as mental dexterity or the player's equipment. The sport also contributes to better health, where its regular practice favors and improves the prevention of psychological and biological problems. Sport is a form of entertainment for athletes and spectators. It can be practiced individually or in teams, it can be professional and paid or amateur. There are events, tournaments, and championships that also involve coaches and referees who can be professional or amateur. Sports can be divided into individual sports, team sports, fighting sports, outdoor sports, adapted sports, or confrontation sports.

Sports have been present in our society for at least 3000 years. It began to be used in training for war and hunting, thus being part of the training of ancient warriors. However, it is not completely certain which was the first year in which the Olympic games were held, the first record is from 776 BC, and these games consisted of human races, chariot races, wrestling, jumping, discus, and javelin throwing, among others [1].

Padel is one of the fastest-growing sports in the world [1] in terms of numbers playing and according to the governing body, the International Padel Federation (FIP), Padel is now played by around 25 million people across the globe. It is a sport practiced with a racket, played in pairs on a rectangular court, enclosed by large glasses twenty meters long and ten meters wide, in which players are allowed to use the side and back walls [1]. The net at the center line of the court is ten meters long and should be eighty-eight centimeters high at the center and ninety-two centimeters at the end. Courts are commonly surrounded by a mixture of perspex glass or solid concrete walls, as well as wire mesh. The rear walls of the court, which need to be three meters in height, are always made of perspex glass. These walls are then topped with one meter of wire mesh. Similarly, the side walls consist of both perspex glass and wire mesh, with a height of three meters. Padel balls are very similar to standard tennis balls in look but are smaller in diameter and carry less pressure. The International Padel Federation (FIP) regulations specify that the ball must be made of yellow or white rubber and have a diameter of between 6.32 and

6.77 centimeters, while the weight must be between 56 and 59 grams. The ball must have an internal pressure of between 4.6 and 5.2 kilograms.

The rules for Padel are similar to tennis in many ways, but there are also differences here. Scoring largely follows that of a normal tennis like:

- Points are scored in the same way as in tennis, except that the ball can bounce once on the opponent's side of the court before it is returned.
- Matches are played using the best-of-three or best-of-five set format.
- The first player or team to win six games, with a lead of at least two games, wins the set.
- A tie-break is played to decide a set that is tied at six games all. The first player or team to win seven points, with a lead of at least two points, wins the tie-break.
- Games are scored the same way as in tennis, with the first player or team to win four points (with a lead of at least two points) winning the game. If the score is tied at 40-40, deuce is called. To win the game at deuce, a player or team must win two consecutive points. In professional padel, a "golden point" system is used at deuce, where the first player or team to win a single point wins the game.
- Players change sides of the court after every odd-numbered game.

Padel Court has two gates located at each end of the net. These gates remain open during gameplay, as the rules permit playing a point outside the court. The playing surface of the court can be made of concrete, synthetic grass, or carpet.

This sport has had a huge development in recent years in countries like Portugal, Spain, and Argentina. The scoring system used is the same as for tennis and the surface of the court can be of various types, from synthetic grass, carpet, or porous concrete, however, the first two are the most common [2].

The practice of Padel has shown impressive growth throughout the world, which has been steadily increasing and far above expectations since 2010 [1]. Its practice, as in other sports, is distinguished by being very demanding at the physical level, especially at the cardiovascular level, which causes substantial variations in the heart rate of its practitioners, thus leading to changes in their cardiorespiratory function during the game [1].

Besides this, the sport has many benefits, since, through its continued practice, practitioners on the one hand improve their physical condition, and on the other, they also develop physical abilities such as motor coordination, agility, precision, and muscle power.

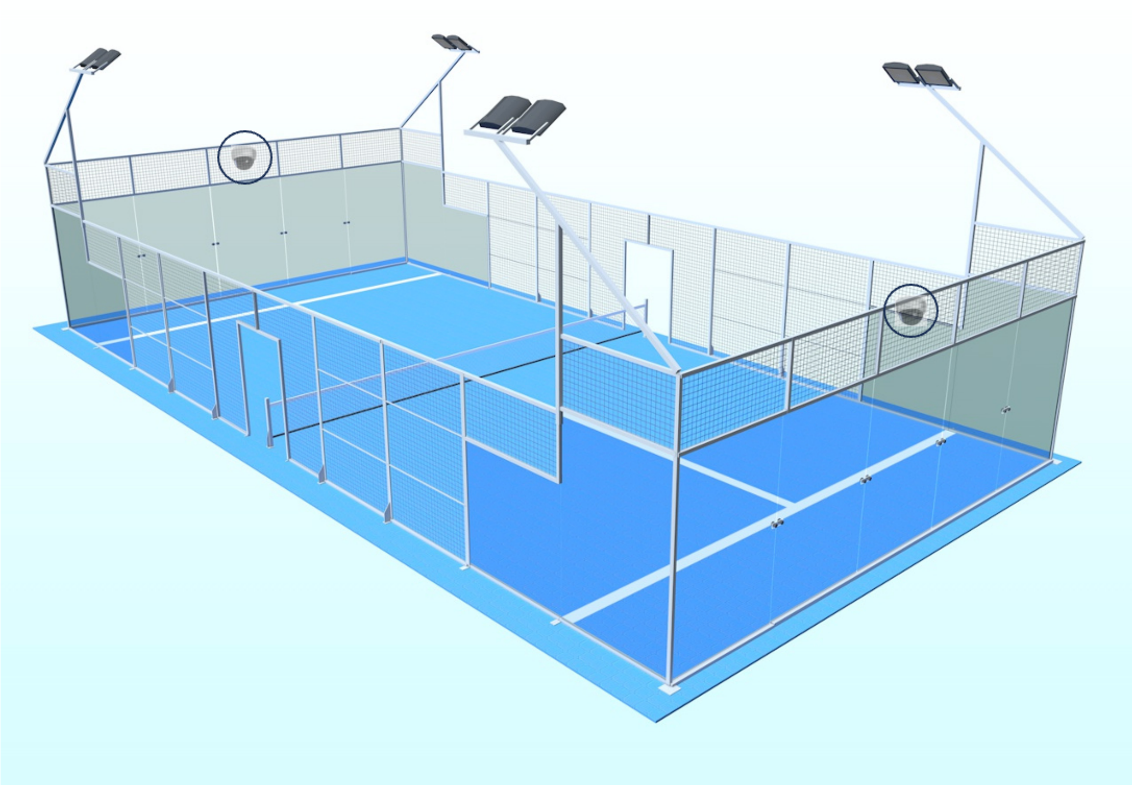

Figure 1- Illustrative image of a Padel court with two cameras

It is possible to see the two cameras on the Padel court (Figure 1). These cameras can be used to record the game and to provide real-time feedback to players and coaches.

The movements that will be covered in this dissertation are the *Left Volley* and the *Bandeja* to build the model.

The Bandeja is a sliced volley with a strong contact point. The ball is hit usually with a backspin, causing it to bounce high and deep in the opponent's court. The *Bandeja* shot is extremely adaptable, as it can be utilized to attack, defend, or transition from defense to attack. To hit a *Bandeja*, the player must be close to the net and in front of the ball. The player should next take a step forward with their opposing foot and reach up to smash the ball with a high contact point. The player should hit the ball with a combination of topspin and backspin while keeping their wrist loose and flexible [3][4].

The *Left Volley* is a forehand volley struck with the left hand. The *Left Volley* is a powerful and accurate stroke that is frequently utilized to finish points. To hit a *Left Volley*, the player must be close to the net and in front of the ball. The player should next take a step forward with the right foot and reach up to make a high contact point on the ball. The player should smash the ball with a mix of topspin and sidespin while keeping their wrist loose and supple [3][4].

The main difference between the *Bandeja* and the *Left Volley* is the type of spin that is applied to the ball. The *Bandeja* is a sliced volley, while the *Left Volley* is a forehand volley with topspin and sidespin. This difference in spin results in different trajectories for the two strokes. The *Bandeja* bounces high and deep in the opponent's court, while the *Left Volley* bounces lower and flatter. Another difference between the *Bandeja* and the *Left Volley* is the power of the shot. The *Left Volley* is a more powerful shot than the *Bandeja*. This is because the *Left Volley* is hit with topspin and sidespin, which gives the ball more power and speed [3].

This link directs readers to a video showing the *Bandeja* being hit in detail (Bandeja movement example), just as this link takes you to a video demonstration of the *Left Volley* movement (Left Volley movement example).

2.2. Artificial Intelligence and Padel game

In the 1950s, Marvin Minsky and John McCarthy, considered the fathers of Artificial Intelligence, defined it as any activity performed by a program or machine, which if it had been performed by a human being, he would have had to use intelligence to perform it [5].

The study of humans and their evolution, as well as their environment, goes back many centuries. There have always been various philosophers, scientists, and physicists who believed and struggled to clarify matter (physics), consciousness (psychology), and life (biology). Names such as Einstein, Freud, Darwin, and Skinner, among others, have studied each of these sciences of nature. Artificial Intelligence (AI) is related to all this and is based on knowledge acquired over the years through these names that built the current scientific edifice. Man has always been concerned with studying all these concepts, but beyond that, there is also a concern with the creation of artificial objects, thus emerging the sciences of the artificial. In general, human beings are afraid of the unknown, which leads to greater difficulty in accepting change and novelty. When it comes to AI, there is also this distrust and reticence to accept or understand the existence of thinking machines [5][6][7][8].

Thus, most AI systems try to replicate one of the following behaviors of human intelligence:

- Planning.
- Learning.
- Reasoning.
- Problem-solving.
- Perception.
- Movement.
- Creativity.
- Social intelligence.

Most AI systems only can focus on a single task, so it is difficult to demonstrate the full learning capacity of a human being. Due to its sensory capacity, it is possible for the human being to have a vast knowledge of its surroundings, which is what AI tries to replicate, thus being a little distant from achieving it, in its fullness [9].

We can observe different paradigms in this emerging technology, and we cannot yet say which are the most correct, but rather take advantage of each of them, using the most appropriate for each specific problem or the combination of several to achieve the optimal objective, and AI is a decision-making system [9][10].

Padel technology is still in its early phases of development, but several interesting technologies are being researched to increase the game's precision and efficiency. Angle judge technology is one of these technologies. Angle judge technology is a system that tracks the movement of the ball and the players in a Padel game using cameras and sensors. This information is then used by the algorithm to identify whether the ball has fallen in or out of bounds. There are numerous techniques to angle judgment technology. Some systems employ a single camera placed behind the baseline. Other methods rely on a network of cameras strategically placed around the court. The photos from the cameras are then used by the system to construct a 3D model of the gaming area. The 3D model is then used by the system to track the movement of the ball and the players. This information can then be used by the system to assess whether the ball has fallen in or out of bounds. Angle judge technology is still in the works, but it has the potential to transform the sport of Padel. Angle judge technology can help to reduce the number of disputed calls and make the game more fair and pleasant for everyone by enhancing the accuracy of the calls made by line judges.

This technology began to be used very recently in very specific games and only in a few cases. It is still under development and needs improvement.

2.2.1. Movement Detection

By monitoring the movement of the player's racket, an AI system can be utilized to identify the *Bandeja* and *Left Volley* strokes. A dataset of photos and videos of Padel players making these strokes can be used to train the AI system. After being trained, the AI system can recognize the *Bandeja* and *Left Volley* strokes in new videos. An AI system has the potential to be a significant tool for Padel players and coaches and can provide feedback on players' technique, track their performance over time, and design new training techniques by distinguishing the *Bandeja* and *Left Volley* strokes.

2.2.1.1 YOLOv8 Algorithm for movement detection

YOLOv8 is an object detection algorithm capable of detecting Padel movements. YOLOv8 was used to train and detect Padel movements using a dataset of images of Padel players executing the *Bandeja* and *Left Volley*. The bounding boxes of the players' movements they are executing should be tagged in the dataset. Once taught, YOLOv8 can be used to detect Padel motions in new images. To accomplish this, YOLOv8 divides the input image into a grid of cells. YOLOv8 then predicts the presence of objects and their bounding boxes for each cell. If YOLOv8 predicts the presence of a Padel player in a cell, it will also forecast the player's movement [11].

The YOLO (You Only Look Once) algorithm turns the detection model into a regression problem, which will speed up the model by using the features of the whole image instead of just a small part of the capture of that image, like other algorithms such as R-CNN. This means that YOLO can discriminate all categories of the image in a generalized way because its architecture achieves very high accuracy [11].

In the recognition process of this algorithm, the trained model is stored with parameters of each checkpoint and these parameters are read in the test phase, thus minimizing the Euclidean distance between the actual and predicted coordinates, to achieve the optimization of each layer parameter and thus the prediction is following the boundary configuration. This distance is calculated using coordinates by the rectangle function in the CV2 module, which is in TensorFlow [12].

The original YOLO algorithm uses 224 x 224 pixels before training and in detection it uses 448 x 448 pixels. When we change the classification model to the detection model, the model will adapt to the image classification [11].

The latest YOLOv7 algorithm is better than all previous object detection models and YOLO versions in speed and accuracy. It requires cheaper hardware than other neural networks and can be trained much faster on small datasets without any pre-trained weights [11].

The YOLOv8 was launched in 2023 and it is part of a series of object detection models developed to perform real-time detection in images and videos. Also known as You Only Look Once version 8, is an algorithm for object detection where it has a major advantage its speed, by dividing the image into a grid and making predictions in parallel for each cell, even in real-time and it can achieve high frame rates per second. This model is capable of detecting objects in real-time, achieving incredibly high frame rates per second. As a result, it is well-suited for this project where a quick response time is vital, like real-time video analytics. Utilizing an approach rooted in deep Convolutional Neural Networks, YOLOv8 is acclaimed for its remarkable accuracy and aptitude to identify a diverse array of objects, regardless of varying lighting and perspective circumstances. The adoption of this technique ensures exemplary performance in discerning tasks, in addition to its outstanding speed [11][13][14].

However, it is important to highlight that the choice of this object detection algorithm depended on the specific needs of this project. Other known algorithms, such as Faster R-CNN and SSD (Single Shot MultiBox Detector), also offer good performance in object detection, however, considering that one of the objectives is object detection, the YOLOv8 algorithm was chosen to develop this project and has this structure as shown in Figure 2.

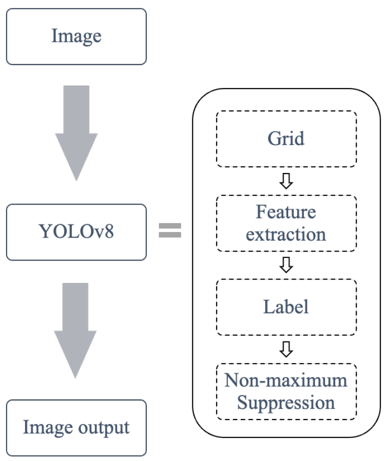

Figure 2 - Procedure of YOLOv8 model

2.2.1.2 Convolutional Neural Network

The concept of Deep Learning (DL) was first proposed by Hilton at the University of Toronto in 2006 and is currently the newest area of machine learning. Instead of using the original artificial design feature, it is a network that has better robustness and generalization capabilities [13][15][16].

It is a concept that meets Machine Learning that trains computers so that they can learn by themselves, through pattern recognition and perform tasks like humans. Deep Learning uses a specific class of algorithms, which are called neural networks.

Convolutional Neural Networks (CNN) are a type of neural network, instead of matrix multiplication, that uses the convolution operation. The architecture of these neural networks is based on alternating layers of convolution, layers that give rise to the network name, and pooling layers. Each layer will possess a set of filters, namely the kernel, which are mainly responsible for removing local features from an input. Each feature map will have a shared set of weights, which decreases the computational complexity of the network as shown in Figure 3 [17]. There is the layer responsible for the image classification process, which is the fully connected layer and which is to the classified output, that is, it is possible to recognize patterns, lines, contours, objects, etc. through this process. Considering that the type of data for this dissertation is sequential images, collected through videos, this approach will be the most effective [16].

Convolutional Neural Network has had groundbreaking results over the years in a variety of fields related to pattern recognition, from image processing to voice recognition. The most beneficial aspect of CNNs is reducing the number of parameters in Artificial Neural Network (ANN), which means that CNNs are a type of neural network that are particularly efficient for image recognition and classification tasks. They achieve this efficiency by using a technique called weight sharing, which reduces the number of parameters in the Artificial Neural Network (ANN). This makes CNNs less computationally expensive to train and more suitable for a wider range of applications.

An Artificial Intelligence technique called a Convolutional Neural Network (CNN) was developed to process and assess visual input, such as images and movies. In recent years, CNNs have revolutionized computer vision in several cutting-edge applications, including image classification, object recognition, and picture segmentation [17].

CNNs are built on convolutional layers, which process input data to extract features hierarchically. By using learnable filters or kernels, these layers collect relevant patterns and structures from the input. A CNN is capable of learning complex representations of visual data by using several convolutional layers with non-linear activation functions.

Pooling layers are widely employed in CNN designs to reduce the spatial dimensionality of the feature maps, simplify computation, and extract the most relevant properties. For example, by selecting the greatest value from each pooling zone, max pooling filter retains the most important properties.

CNNs also employ entirely connected layers, in which the neurons of each layer are interconnected to the neurons of every other layer. These layers aid in the formulation of high-level predictions or classifications based on the learned attributes [15][17].

During training, CNNs utilize a loss function such as categorical cross-entropy to calculate the difference between predicted and real labels. Optimization methods such as stochastic gradient descent (SGD) algorithm are used to decrease loss and improve performance by altering the network's weights and biases.

Transfer learning, which is commonly used with CNNs, enables pre-trained models on large datasets such as ImageNet to be adapted for specific applications. Even with limited training data, this approach may employ learned characteristics from one job to improve performance on a related but distinct one [17].

CNNs' capacity to dynamically construct hierarchical representations, tolerance to input perturbations, and remarkable performance on tough visual tasks have all contributed to their widespread adoption. CNNs have benefitted from a wide range of applications, from augmented reality and self-driving cars to face recognition software and medical image analysis.

In essence, CNNs are a class of deep-learning algorithms designed to handle visual input. Using their convolutional and pooling layers, they are very good at eliminating major features and patterns from pictures. CNNs have made significant progress in computer vision tasks, opening new opportunities in a variety of sectors, and promoting continued innovation [17].

Convolutional Neural Networks (CNNs) are a form of deep learning algorithm that excels at picture classification and object detection. CNNs function by extracting features from images and then using those characteristics to classify or recognize things within the images.

2.2.1.3 Transfer Learning and Fine-tunning technique

In an ideal world, there would be enough labeled data to learn a new domain entirely from certain data. Unfortunately, this is not always the case, and so transfer learning arises. "Transfer learning aims at improving the performance of target learners on target domains by transferring the knowledge contained in different but related source domains" [18].

Transfer Learning has been studied under different terminologies in AI, such as domain adaption, pre-training, fine-tuning, and so on. Transfer Learning refers to the process in which past experience acquired from a previous source can influence future learning and performance in a target situation [18][19].

Deep learning models train with a vast volume of data and learn model weight and bias during training. These weights are then transferred to other network models for testing. The new network model can begin with pre-trained weights.

There are various pre-trained architectures available, such as ResNet [20], AlexNet, VGG, and so on, and there are two key motivations for using this technique: to begin, avoiding greater computer power is to build enormous models on massive datasets, and training the network could take weeks. Training the new network with pre-trained weights speeds up the learning process.

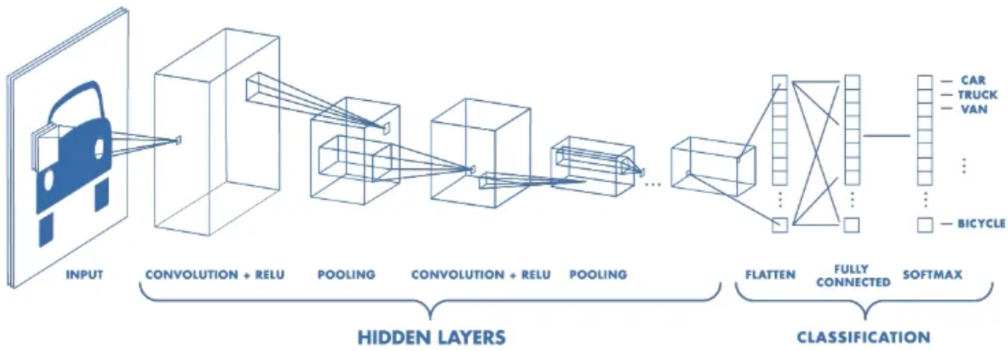

Figure 3 – CNN [13]

Transfer Learning is used to improve something from a certain domain by transferring information from a related domain. Being a technique that is used for Machine Learning models, it is also a technique used in Neural Networks, as an application of Model-Based Transfer Learning, which is the type of Transfer Learning where knowledge is in the form of pre-trained parameters, which are transferred between models. Unlike Machine Learning models that are isolated where there is a task, Transfer Learning relies on a task that has already been learned, thus being a fast, more accurate process where not as much training will be required [19][21].

This method is very effective because people can intelligently apply all the knowledge that has been previously learned, to adapt to new problems more efficiently and quickly. It is possible to learn that recognizing an orange can help in recognizing an apple, so this is a very advantageous technique when it comes to image recognition and classification [22].

A very useful technique is the multitasking learning framework that could be possible try to learn several tasks simultaneously, even if they are different and a typical approach for this multitask learning is to find the common resources to be able to benefit from each task [21].

Fine-tuning is the process of modifying the parameters of a pre-trained Convolutional Neural Network (CNN) for a new job using a smaller dataset. The goal of fine-tuning is to apply the knowledge gained by the CNN on a large-scale dataset to a new job or dataset. This technique involves retraining the model on a validation dataset, with some of the parameters of the initial model being frozen. In this work, fine tuning was used to classify padel movements. To do this, a dataset was split into training and validation with images of padel players performing different movements [12].

Fine-tuning is a strong deep-learning technique for adapting a pre-trained model to a new task or dataset. Fine-tuning the YOLOv8 model entails retraining the model on a smaller dataset while freezing some of the initial parameters [11][18][21].

The model's frozen parameters, which are responsible for feature extraction, remain unaltered during the fine-tuning process. This enables the model to apply what it has learned from a large-scale dataset during initial training.

The unfrozen parameters, on the other hand, which are responsible for predicting the position and size of objects, are updated during fine-tuning. This helps the model to adapt to the new dataset's individual properties and improve its performance on the target task [13].

By fine-tuning the YOLOv8 model, it is possible to take advantage of the generalization capabilities of the pre-trained model and adjust it to a specific object detection problem. This technique inside transfer learning can save significant time and computational resources compared to training a model from scratch and utilize the generalization capabilities of the pre-trained model and customize it for a specific object detection problem. This approach can save significant time and computational resources compared to training a model from scratch. Fine-tuning is particularly advantageous when the new dataset is small since it helps prevent overfitting and enhances the model's ability to generalize to unseen data. Furthermore, it enables to transfer of knowledge acquired from a large-scale dataset to a similar but distinct task, thereby improving the model's performance [11][23].

2.2.1.4 Human Pose Estimation

Human Pose Estimation (HPE), which has received a lot of attention in the computer vision literature, predicts the configuration of human body components from sensor input data, namely photos and videos. HPE gives geometry and kinematic data on the human body, which has been used in a variety of applications. It allows to locate the human body parts and build human body representation (e.g., body skeleton) from input data [24].

OpenPose is a free system that recognizes and tracks the 2D locations of human body joints (keypoints) in photos or movies, even when numerous persons are present. This process was used and known as human position estimation, entails identifying specific bodily markers such as the nose, shoulders, and elbows, among others. These crucial principles are critical to comprehending human body movements and positions.

OpenPose accomplishes this accomplishment by utilizing a novel deep learning architecture known as Part Affinity Fields (PAFs). PAFs connect keypoint pairs, generating limbs that depict the relationships between different bodily sections. This connection network allows OpenPose to track and evaluate the movements of several people at the same time.

Figure 4 - 2D multi-person pose estimation [25]

The goal of heatmap-based methods is to predict the approximate locations of body parts and joints, which are supervised by heatmap representation. Heatmap-based frameworks are now widely used in 2D HPE tasks. These heatmaps are formed by adding 2D Gaussian kernels on each joint's position. To be more specific, the purpose is to estimate K heatmaps H1,H2,...,HK for a total of K key points. In each key point heatmap, the pixel value Hi(x,y) reflects the likelihood that the key point is located at (x,y). [26] A 2D Gaussian centered at the ground-truth joint position generates the target (or ground-truth) heatmap. As a result, pose estimation networks are trained by reducing the difference (e.g., the Mean Squared-Error (MSE)) between predicted and target heatmaps. Heatmaps, as opposed to joint coordinates, maintain spatial position information while making the training process more fluid.

Multi-person HPE is more complicated and challenging than single-person HPE since it must figure out the number of individuals and their locations, as well as how to organize crucial points for various persons [24].

In the context of sports, 2D Human Pose Estimation has been used in a variety of ways, including performance analysis, where HPE can be used to track athletes' movements and highlight areas where they might improve. HPE can be used to track a baseball player's swing mechanics or a track athlete's running stride. It helps with injury prevention because athletes who are at risk of injury can be identified. HPE, for example, can be used to track a runner's joint angles to identify any irregularities that could contribute to an injury. HPE can be used to monitor the progress of athletes who are recovering from injuries. For example, HPE can be used to monitor an athlete's knee range of motion after surgery and could be used to provide feedback to athletes while they are training. HPE can be also utilized in broadcasting to provide viewers with insights into the performance of athletes and it is possible to track a soccer player's speed and position on the field [24][26].

2.2.2. Identification of Padel Court lines

2.2.2.1 Canny Edge detector

The Canny edge detector is a widely used multiply-scale edge detection technique invented by John F. Canny in 1986. It allows to identify the best edge detection method for image processing. This technique consists of multiple phases that work together to recognize edges in a picture with a low error rate and few false detections. To begin, the image is denoised using a Gaussian smoothing filter to remove noise. The gradient amplitude and direction are then determined, with the gradient direction taking four angles - 0, 45, 90, and 135 degrees. Non-maximum suppression is used to remove non-edge pixels, leaving only a few tiny lines.

The Canny edge detection algorithm uses a two-threshold technique called hysteresis thresholding to distinguish between genuine edges and noise-induced distortions. This strategy establishes two separate thresholds, one high and one low. Pixels that exceed the high threshold are considered strong edges, with the highest intensity contrast and the fewest false positives. Pixels between the low and high thresholds are classified as weak edges and must be investigated further. Hysteresis thresholding evaluates the authenticity of a weak edge pixel by analyzing its surroundings. If surrounding pixels also surpass the high threshold, the weak edge pixel is determined to be genuine, with strong edges supporting it. Otherwise, it's ruled a false positive. The Canny edge detector can achieve a compromise between accurate edge identification and minimizing noise-induced false positives thanks to this two-threshold technique. The system effectively distinguishes between real edges and fake signals by prioritizing strong edges and validating weak edges based on their context. In conclusion, hysteresis thresholding successfully filters out noise and false detections, transforming the Canny edge detector into a versatile tool for accurate edge detection in image processing applications [27].

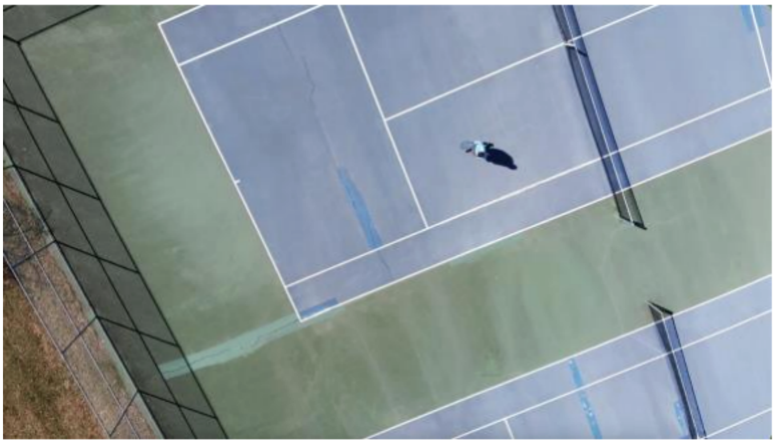

Figure 5 - Top view of a tennis court [28]

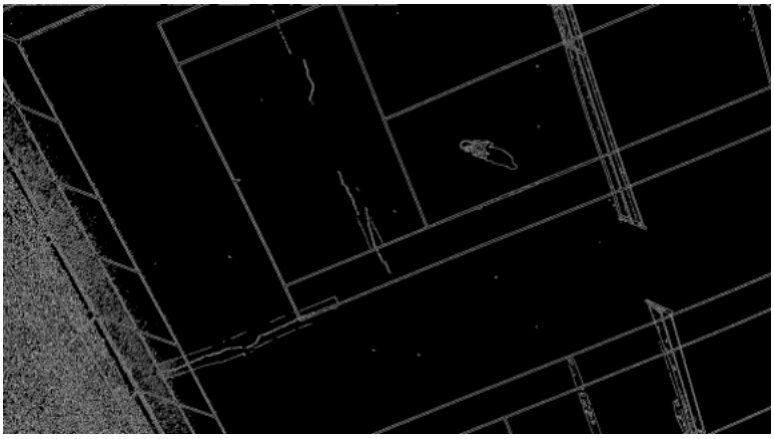

Figure 6 - Canny Edge method applied [28]

By employing a two-dimensional Gaussian function, the algorithm effectively smooths the image and eliminates any existing noise.

$$G(x,y) = \exp\left[-(x^2 + y^2) / 2\sigma^2\right] / 2\pi\sigma^2$$

where σ stands for the parameter of Gauss filter, and it controls the extend of smoothing image.

2.2.2.2. Probabilistic Hough lines transform

The original form of Hough Transform aimed to identify straight lines. To use Hough Lines Transform, processed image should be binary. (Haralick, R. M., & Shapiro, L. G. (1985)) The most common solution is to firstly grayscale the image and then to detect edges [29]. Such mask of edges can be then fetched to the Hough Lines method which should output a set of straight lines found on an image [29]. It is possible describe the line using the pair *(ρ, θ)* in polar system. The first parameter, ρ, is the shortest distance from the origin to the line (approaching the line perpendicularly). The second, θ, is the angle between x-axis and the distance line. One of the benefits of such representation is that we can describe vertical lines by ρ and θ which is impossible by using only *(a, b)* parameters in Cartesian system. Once the parameters a and b have been estimated, the line can be drawn on the image by plotting the points (x, y) that satisfy the equation $\rho = a \cos\theta + b \sin\theta$.

For a given line, we can determine specific ρ and θ. Then, the following equation is satisfied for each a, b point belonging to this line:

$$\rho = a \cos(\theta) + b \sin(\theta)$$

The Hough Lines transform begins by identifying probable edge points in the image using edge detection techniques such as the Canny edge detector. These edge points are then converted from Cartesian coordinates (x, y) to parameter space (,), which denotes the perpendicular distance from the origin to the line and denotes the angle between the x-axis and the perpendicular line [29].

Following the transformation, a parameter space accumulator, commonly represented as a 2D array or matrix, is formed. Each accumulator cell corresponds to a distinct line of the image, as specified by its values. The accumulator is initialized by setting all cells to zero.

The method iterates through all possible lines in the parameter space for each edge point, voting gradually for the lines that could potentially pass through the edge point. This voting procedure accumulates votes in the accumulator's appropriate cells [29].

Finally, the algorithm detects the lines in the image with the most votes in the accumulator, indicating the presence of strong lines. These lines are then removed and returned as the Hough Lines transform output. The Hough transform maps points in the Cartesian coordinate system to lines in a parameter space (ρ, θ).

In the context of the Hough Transform equation you provided, **θ = arctan2(y, x)** is used to calculate the angle **θ** for each point **(x, y)** in the plane. This angle is then used in the equation above for further computations in the Hough Transform algorithm, often used in image processing for detecting shapes like lines or circles.

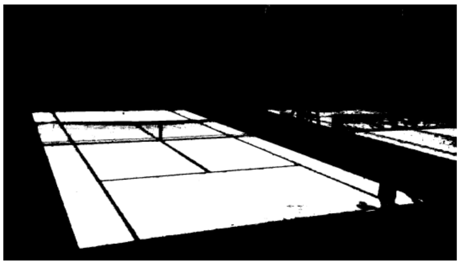

Figure 7 - Unfiltered HSV tennis court [30]

Figure 8 - Filtered HSV tennis court [30]

The result is then used to build an outline of the court using clever edge detection (Figure 7). It is possible to identify the dominant lines in the outline using the Hough transform (Figure 8), and record any intersections between the Hough lines in a separate result. This result is significantly dilated, resulting in clusters of intersections becoming connected components.

Figure 9 - Canny Edge detector [30]

Figure 9 shows the result of the applied canny edge detector, which is just an intermediate step to finalize the court boundaries.

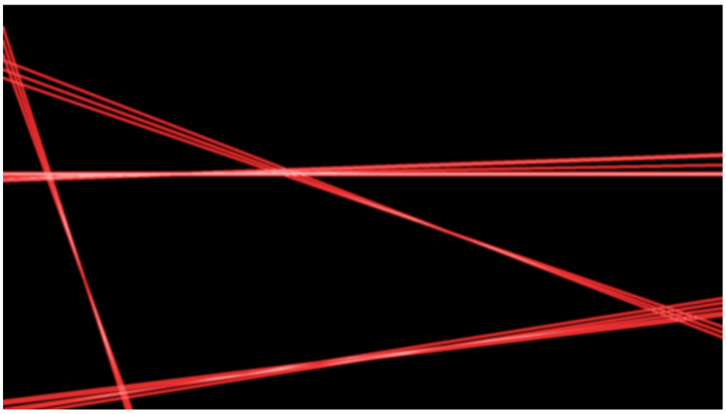

Figure 10 - Hough lines transform [30]

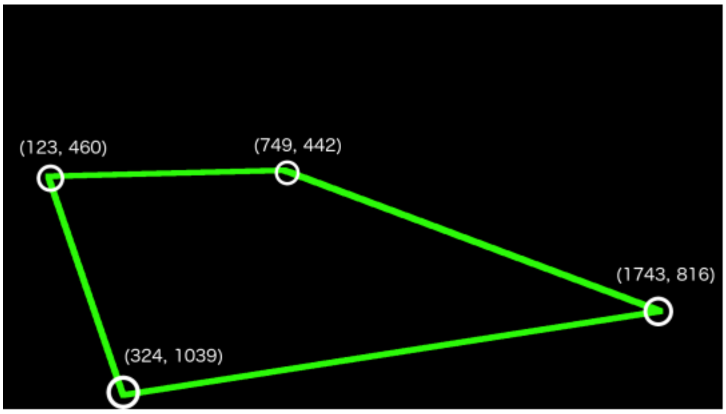

Figure 11 - Court detected after Hough transform [30]

Figure 11 shows the final result of the Hough transform when detecting the intersection points of the court.

2.2.2.3. Barrel distortion on cameras

Barrel distortion is a type of geometric distortion that can be seen in wide-angle lenses. It causes straight lines to appear curved outward from the center of the image. This occurs because wide-angle lenses have a shorter focal length compared to telephoto lenses. Consequently, the center of the image is magnified more than the edges, resulting in curved edges. When capturing images and videos of a Padel court, the camera's lens may create a curved effect, giving the impression that the rectangular court has a barrel shape. This distortion could potentially hinder the accuracy of object detection algorithms, making it challenging to precisely locate and track items on the court [31].

Barrel distortion is a non-uniform distortion that occurs along the radial direction. As one moves away from the optical center, the degree of distortion rises. Objects appear twisted, and straight lines appear bent because of this distortion. As a result, fisheye lens-based vision systems require an appropriate approach for correcting this distortion.
Several researchers have attempted to correct the fisheye lens distortion using various methods. These methods can be separated into three categories:

1. geometric projection model-based;
2. calibration pattern-based;
3. image-based or feature-based methods.

The geometric projection model connects the undistorted input scene with the distorted output image [31].

The fisheye lens was invented in 1906 by scientist Robert W. Wood to simulate the perspective of a fish underwater. It gained popularity in the 1920s when it was used in meteorology to investigate cloud forms. However, it wasn't until the 1960s that the fisheye lens became widely popular due to mass production.

There are two types of fisheye lenses which are circular and full-frame and each produces a unique effect. For cameras with a 35mm sensor or film, circular fisheye lenses typically have focal lengths ranging from 8mm to 10mm. In contrast, full-frame lenses have significantly greater focal lengths [31].
For cameras with sensor sizes smaller than 35mm, the focal length of a fisheye lens is effectively extended. To calculate the equivalent focal length, multiply the focal length

of the lens by the camera's "crop factor." A 10mm fisheye lens on a 1.5 crop factor camera, for example, would have an effective focal length of 15mm, resulting in a reduced field of vision.

The existence of "barrel distortion," causes subjects in the center of the frame to look bulging and straight lines to curve considerably, which is a significant disadvantage of fisheye lenses. This style of image is known as "curvilinear." Fisheye lenses, unlike standard wide-angle lenses, feature an extreme angle of view that cannot be rectified to provide a "rectilinear" image in which straight lines appear straight and the scene's perspective appears natural [31].

However, for many photographers, barrel distortion with fisheye lenses is a desirable effect rather than a drawback. The unusual and intriguing distortions generated by fisheye lenses are frequently the reason photographers pick them in the first place [31].

There are three major ways to remove barrel distortion from an image as shown in Figure 12:

- Correction by hand: This procedure is the most precise, but it is also the most time-consuming.

- Automatic correction: While this method is more efficient than human correction, it is possible that the software will not correct all distortion, especially in complicated photos.

- Hardware correction entails the use of a lens specifically designed to correct distortion. This is the most precise procedure, but it is also the most expensive.

Adobe Photoshop, GIMP, and LibRaw are among the software packages that can be used to repair barrel distortion.

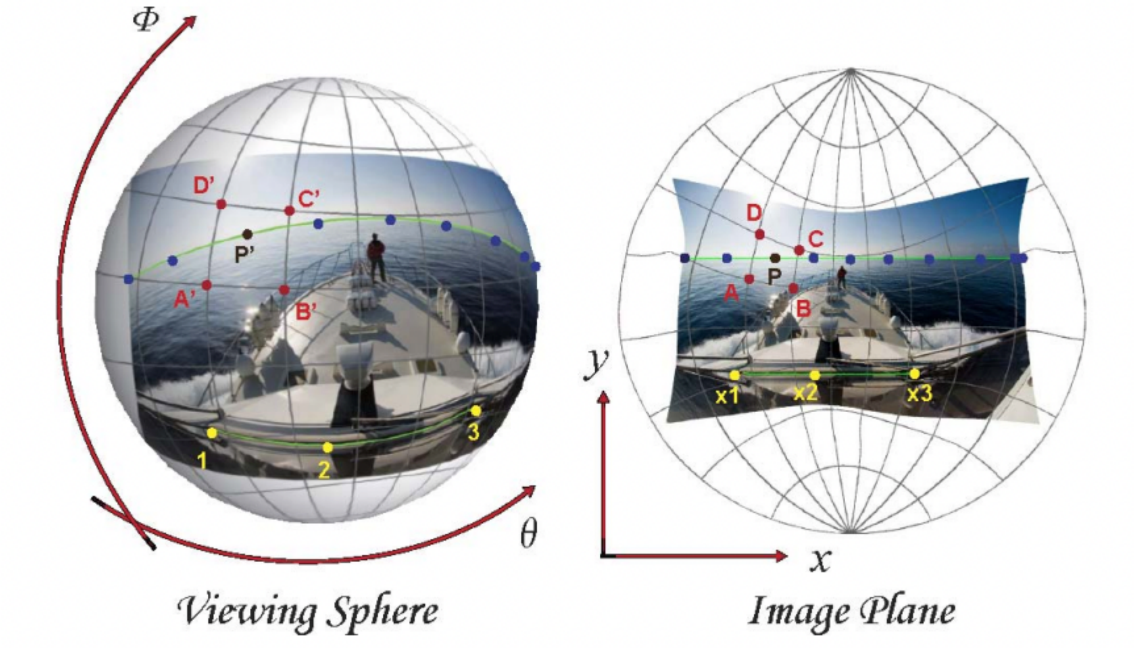

Figure 12 - Example of correction distortion [32]

2.2.3. Padel ball tracking

Object tracking is a basic computer vision problem that entails locating and following a specific object within a video stream. Kernel Correlation Filters (KCF) are an efficient and frequently used object tracking method. KCF makes use of the correlation filter idea, which is trained to recognize the target item and predict its position in consecutive frames [33].

The architecture is divided into feature extraction, training, and online tracking.
During the feature extraction stage, the algorithm uses discriminative characteristics to represent the target object and its surrounding region. Color Names, Histogram of Oriented Gradients (HOG), and Scale-Invariant Feature Transform (SIFT) are all popular options. These characteristics capture the appearance of the object and give a solid representation for later processing [34].

In training the KCF algorithm learns a correlation filter that can distinguish between target and background regions during the training stage. It creates a training set by removing positive and negative samples from the starting frame. Positive samples are for the target object, whereas negative samples are for the background. The filter coefficients are then optimized by minimizing the difference between the desired and actual response [34].

Online tracking is performed by the algorithm once the correlation filter has been trained by applying the filter to successive frames. It produces a response map by converging the filter with the feature representation of each frame. The response map's peak matches the target object's estimated position. In each frame, the algorithm modifies the filter's coefficients to respond to changes in appearance and position.

Chapter 3 – Architecture

3.1 AI system

In this dissertation, different concepts have been addressed, which together have enabled us to answer the research question presented in chapter 1.

Initially, based on videos taken from friendly Padel matches, it was possible to divide these videos into frames, thus obtaining a set of sequential images.

After this division of frames, the Human Pose Estimation technique with OpenPose algorithm was used to detect the key points for estimating the points of the human body. The connection between these points is known as a pair, and what is assumed to identify it in the image is the pair that represents the shortest distance between them.

These images were used to annotate the technical movement being performed and associate it with a particular class.

The annotations were created in a tool called CVAT (Computer Vision Annotation Tool), where it was possible download the images and annotate the part of the image that matters and then download them and get two outputs. One of them is with the coordinates of labels and the other one is with the images annotated.

These annotations were used to obtain the training data that served as input for the YOLOv8 model and a dataset was created with them, which was then divided into training, validation, and test data.

Once the data preparation was complete, the YOLOv8 model's fine-tuning technique was used, in which the model's weights were used to train with the data obtained. In this dissertation, the YOLOv8 model, an advanced object recognition algorithm known for its accuracy and efficiency, was used to properly assess player movements on the Padel court. A fine-tuning approach was used to customize the model to the complexities of Padel games.

First, the 2D multi-person pose estimation was implemented to join the points of each body part, such as the shoulders, the knees, the neck, etc, and automatically draw the human skeleton on the persons in each frame. The aim of implementing this technique was to insert it into the machine learning model as a feature that would make it easier for the model to recognize the movement in question.

This approach required creating bounding boxes around each movement of players, allowing the model to properly learn and predict these movements. The model was instructed to identify and predict each movement by carefully annotating player trajectories and applying these bounding boxes, giving a thorough understanding of player dynamics on the Padel court. This innovative technique not only advances the integration of AI into sports analysis, but it also has the potential to improve Padel training strategies and player performance evaluation.

Videos of some friendly matches played were used, which were separated into 30 frames per second (FPS). In this way, it was possible to create the training data, through the bounding boxes, which have the shape of a rectangle and are associated with a certain class. They help the model learn to identify which movement is being performed. The model was trained on CPU, so it took longer than normal, which was approximately 4 hours.

Once the model had been trained, it was possible to move on to the next stage, which was to detect the lines of the Padel court.

At first, stereo vision technique was used, because the field has a camera at each end as represented in Fig. 1, so that the field can be planned and all the blind spots that arise can be detected, but this technique requires calibration of both cameras.

At this stage, the concepts of the Canny Edge detector and the Hough Lines Transform were introduced to plan the Padel court and realize where the ball intersected with the plan (court).

When detecting the Padel court, the Canny Edge detection and Hough Lines techniques were used to detect the boundaries of the court and identify its lines, in order to obtain a plan of the court and be able to get the coordinates of the ball about the plane of the Padel court.

First of all, the image was converted to grayscale because the Canny Edge detector works better this way and is a popular edge detection algorithm that is used to identify the edges in an image. It works by first smoothing the image to remove noise, then were applied a filter to remove noise from the image.

The Hough lines method in OpenCV is used to recognize lines in images. It operates by partitioning the image's Hough space into a grid of cells. Each cell in the graphic indicates a possible line. The algorithm then scans the grid for cells with a high density of points. A cell is considered a line if it has more points than a specific threshold.

Subsequently, the ball had to be identified and for this the ball was annotated in the same tool (CVAT) so that the model could learn that the ball should be classified as a padel ball, which is similar to a tennis ball. However, this technique was not successful, as the cameras did not have a very high resolution and the ball was difficult to identify.

Another technique was applied to identify the ball, in which a class was added to the already trained YOLOv8 model, which was the tennis ball, due to the availability of annotations of it. In this way, the model was trained with this extra class on the GPU, because it would have had to train with all the other classes simultaneously, but it was not possible to identify it due to the poor quality of the cameras.

For the Padel game optimisation system to work properly and generate statistics, it is essential to create conditions after the model has been trained and integrated with the court and the ball, in which it is mandatory for the system to identify the point in a given movement through these created conditions.

The YOLOv8 model goes through a pre-processing procedure, extracts characteristics, makes object detection predictions, uses non-maximum suppression to reduce duplications, and returns the final bounding boxes with the corresponding classes and confidences. This procedure is repeated for all the images cells, resulting in object detection throughout the scene.

Figure 13 shows the process of detecting the Padel court.

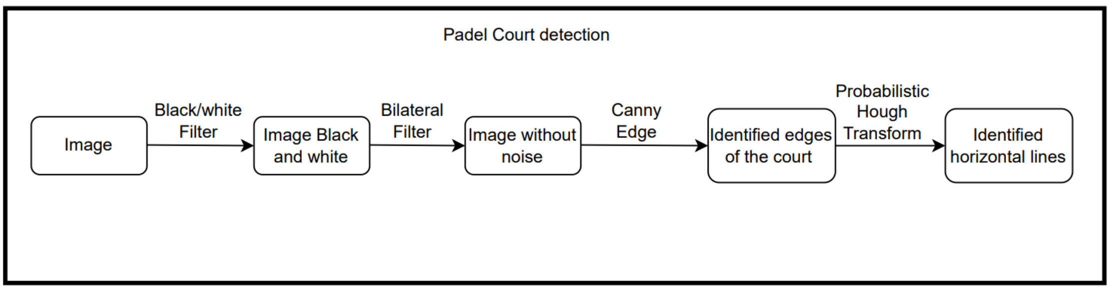

Figure 13 - Padel Court detection flowchart

Figure 14 shows a way of data creation to put in the model in which a set of input images is inserted into the CVAT tool and then the images containing the movement to be identified are annotated. After this process, two outputs are generated, specifically two datasets, one with the annotated image and the other in a text file with the coordinates relating to the annotation.

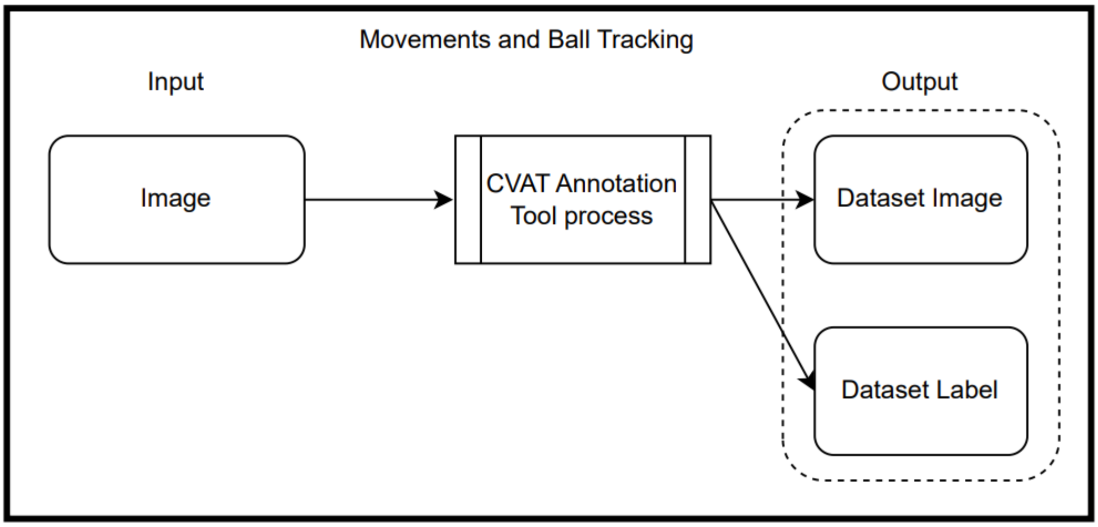

Figure 14 - Padel Ball tracking flowchart

The AI system, as illustrated in Figure 15, is formed independently at an early stage. Court detection, ball detection and the YOLOv8 model with movement detection are carried out in different processing units, requiring the use of 1 GPU for each process. The output of each of these processes will be stored in a database, more specifically in tables, which will then all be joined together to generate the main table and once this table is built, it will be the basis for generating all the desired statistics.

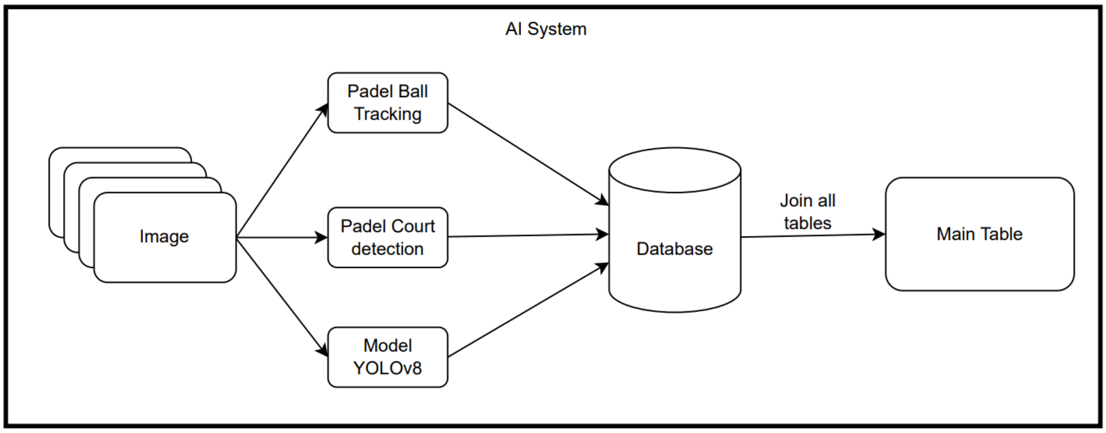

Figure 15 - AI System flowchart

The statistics generated by the AI system will be presented in the Main table. This table has the following format:

Table 1 - Output from AI System - Main table

ID	Frame Timestamp	Frame ID	Movement	Ball Position X	Ball Position Y	Court Detection X	Court Detection Y	Player ID
1	16:00:01	1	Left Volley	1	2	15	20	1
2	16:00:02	2	Left Volley	1,1	2,1	16,5	21	1
3	16:20:12	3	Bandeja	5,76	-8,34	86,4	-83,4	3
4	16:20:13	4	Bandeja	5,77	-8,35	86,55	-83,5	3
5	17:03:32	5	Left Volley	14,03	9,23	210,45	92,3	4

A table of logs as can be seen in Table 1 is a data structure that stores information about events that have occurred in a system. This is possible because the AI system is designed to monitor and record events that occur in the system. The table of logs can be used to troubleshoot problems in the system, to identify trends in the data, and to improve the performance of the system and generate statistics for each player or each move. It's possible that this table shows zero moves, since at certain points in the game no moves are being made.

3.2 Data collection for Model YOLOv8

After developing the Human Pose Estimation concept on CPU, it is possible to see the following results, as shown in figure 16.

Figure 16 - Example of body skeleton applied

These results were introduced as a feature in the Machine Learning model and made it easier for the model to identify the movements.

Then, videos of some of the games were downloaded into the computer. A code was then used to separate the frames, which were 30 frames per second in this case. In this way, bounding boxes were created in all the images in which the *Bandeja* or *Left Volley* movement was executed, as shown in Figure 17. This has been done in order to be able to identify the hit from different perspectives and from different people.

Figure 17 - Bounding box of Bandeja on CVAT tool

One hundred and five images were used to train the model for *Left Volley* and another one hundred and five for *Bandeja*. There was little data because there was a need to reduce the annotations on the movement and only identify what differentiated each specific movement and seventy-three images were used for *Bandeja* and seventy-four images were used for *Left Volley* for the test data.

3.3 Court Detection

In figure 18, we can see the image of the court that was used to carry out the entire Padel court detection process.

Figure 18 - Padel Court

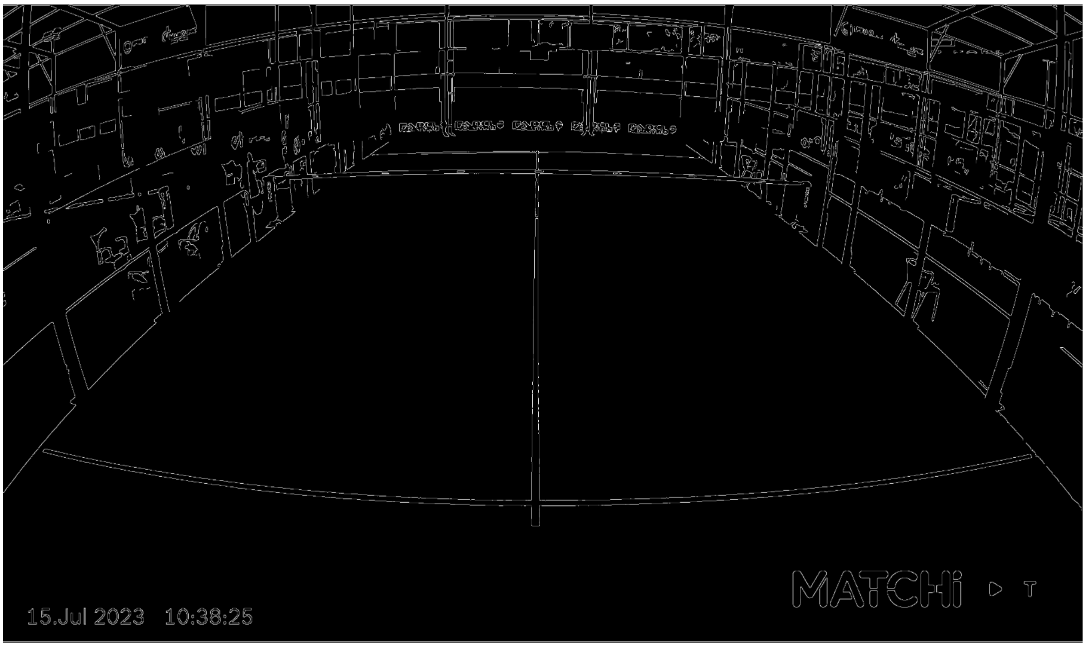

Figure 19 - Canny Edge detector effect in Padel court

For the edges, the threshold range set for the canny edge detector is 100 for the lower threshold and 200 for the upper threshold and it was applied a bilateral filter to reduce noise in the image and this thresholds were the best choose the edge pixels as it is possible to see in Figure 19.

In this study was applied a minLineLength = 10 and a maxLineGap = 5 and a threshold of 255, as shown in figure 18.

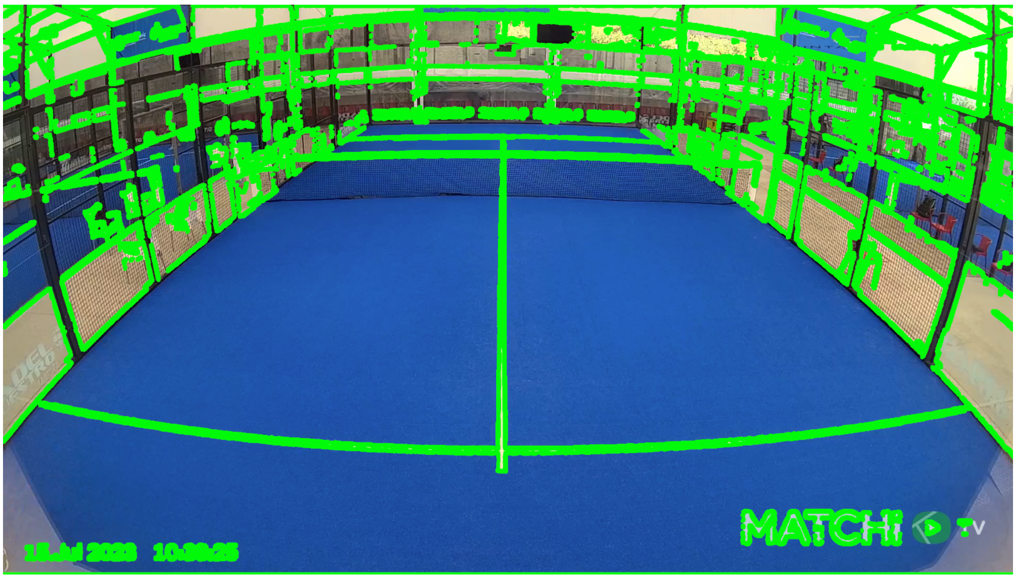

Figure 20 - Probabilistic Hough lines transform in Padel court

Then was created a Black image with a specific size of 1624x2880 pixels using NumPy, as shown in Figure 21. This black image is represented as a NumPy array with a data type of 'np.uint8', which is the standard data type for image pixels. After this a function was defined to check whether a line is horizontal or vertical based on a condition: $|(x2 - x1)| > 2 \times |(y2 - y1)|$, where x1 and y1 are the coordinates of one endpoint of the line and x2 and y2 are the coordinates of the other endpoint of the line. If this condition is true, it categorizes the line as horizontal, otherwise, it categorizes it as vertical.

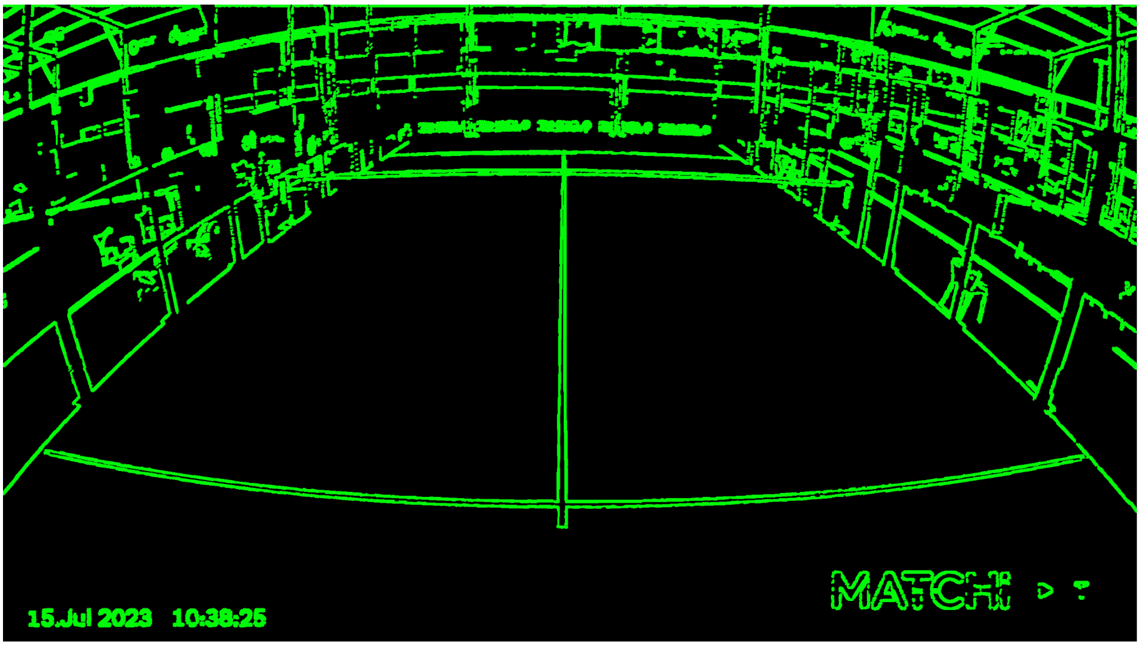

Figure 21 - black image measuring 1624x2880 pixels, represented as a 'np.uint8' NumPy array.

After this, a function was defined to calculate the Euclidean distance (length) between two points and was made a filter to the original lines to detect only the horizontal lines based on two criteria: it checks if a line is categorized as horizontal and if its length is greater than 40 pixels and the result is in Figure 22.

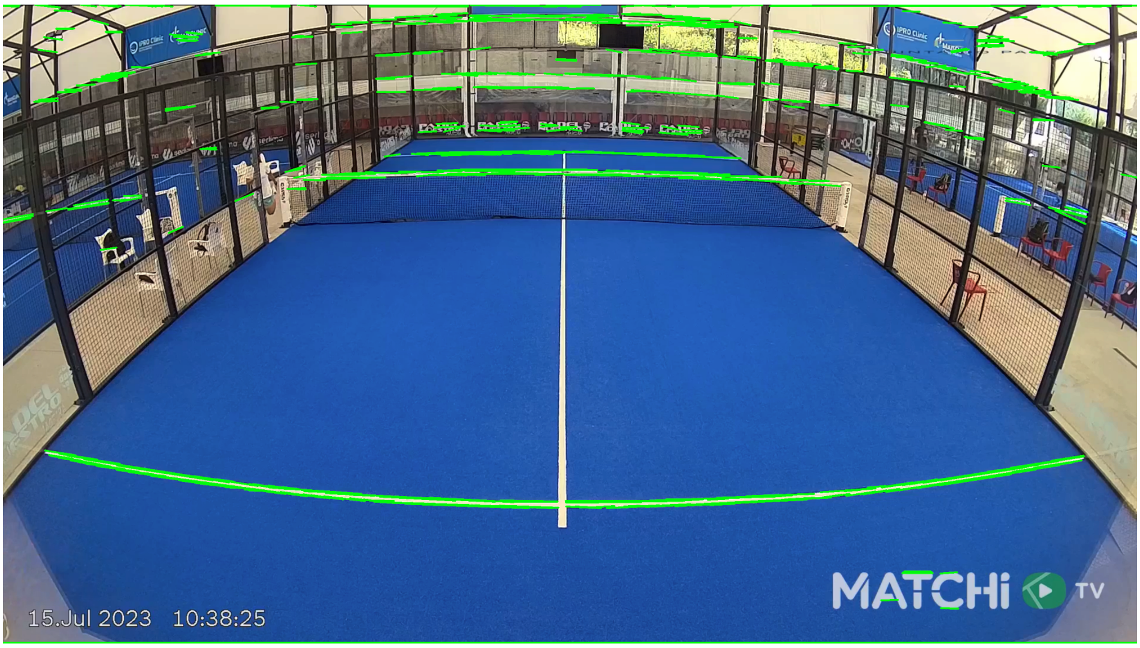

Figure 22 - Image filtered with only horizontal lines

Basically, the aim of detecting the field lines is to identify the points of intersection between the lines and to be able to obtain information from all the points in the field. This is so that it is later possible to identify the coordinates of each player and the ball itself on the pitch.

3.4 Ball tracking techniques

To detect the ball, there were three different approaches.

The first was the simplest, in which, when processing the data, annotations were also made to the Padel ball so that the model could recognize it. This would allow the model to learn to recognize the Padel ball and identify it, as well as the *Bandeja* and *Left volley* movements but it didn't work because the ball was too small in the image and couldn't be detected.

The second approach was more complex, inserting another class into the YOLOv8 model, the tennis ball. This was because was given a folder with various annotations of a tennis ball and it was possible to train the model with these annotations and then identify the Padel ball, which has the same shape and colour. The problem here was that it would be necessary to re-train all the classes of the algorithm and therefore additional computing, such as GPU, was required and it was not possible to obtain this resource.

The third and final approach was to use the KCF algorithm and the main idea behind correlation filter tracking is to determine an optimal image filter to ensure filtration with the input image provides the desired result. Because the expected response is often of a Gaussian form centered at the target point, the score drops as the distance increases.

The filter is trained on instances of the target patch that have been translated (shifted). During testing, the filter's response is examined, and the maximum yields the target's new position. The filter is educated in real time and updated with each frame so that the tracker can adjust to moderate target changes.

The correlation filter tracker has a significant advantage in terms of calculation efficiency. The rationale for this is that the computation may be done quickly in the Fourier domain. As a result, the tracker runs in near-real time, at hundreds of frames per second.

Small objects like Padel ball frequently lack the distinguishing characteristics required for accurate tracking. Because of their small size, there is less information accessible for feature extraction, making it challenging for tracking algorithms to distinguish them from the background.

As little items move around the scene, they can exhibit substantial scale fluctuations. Traditional tracking algorithms, such as the Kernelized Correlation Filter (KCF), may struggle to successfully adjust to these changes, resulting in tracking failures.

Small objects within the frame may not provide enough contextual information. Because tracking algorithms typically operate within a fixed-size search window, maintaining object identity might be difficult when the item occupies only a small section of the image.

The image's spatial resolution is critical to tracking accuracy. Small objects may not have enough pixels dedicated to their depiction, limiting the algorithm's capacity to effectively follow them.

Chapter 4 – Results presentation and analysis

4.1. Metrics

As already mentioned, the movements trained were the *Left Volley* and the *Bandeja*. The metrics used to assess the model's performance were the Mean Average Precision (mAP 50-95) and the Intersection over Union (IoU).

Mean Average Precision (mAP) is a statistic for evaluating object detection models such as Fast R-CNN, YOLO, Mask R-CNN, and others. The mean of Average Precision (AP) values is computed for recall values ranging from 0 to 1.

The mAP is calculated by calculating the Average Precision (AP) for each class and then averaging it over many classes.

The mathematical formula for calculating mAP is as below:

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} APk \, ,$$

$$APk = the\ AP\ of\ class\ k$$

$$n = the\ number\ of\ classes$$

The mAP considers both false positives (FP) and false negatives (FN) and reflects the trade-off between precision and recall. Because of this feature, mAP is a good metric for most detection applications.

Intersection over Union denotes the overlap of anticipated bounding box coordinates with ground truth box coordinates. Higher IoU implies that the anticipated bounding box coordinates are like the ground truth box coordinates.

The mathematical formula for calculating IoU is as below:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

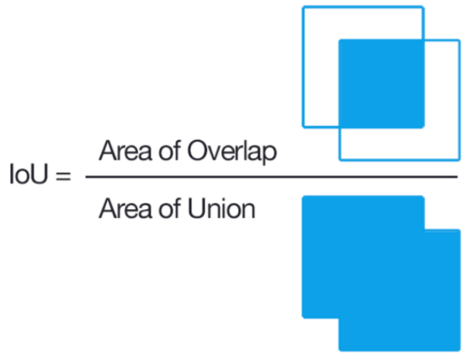

Figure 23 - Illustrative image of Intersection over Union metric

The mAP (50-95) metric is used to assess the performance of object detection systems. It is computed by averaging the precision-recall curves across ten distinct Intersection over Union (IoU) thresholds ranging from 0.5 to 0.95.

4.2. Results

The mAP (50-95) results for *Left Volley* is 73,5% and *Bandeja* is 84,5% which shows that YOLOv8 can recognize both stroke styles with good accuracy. The mAP (50-95) for *Bandeja* is slightly higher than the mAP (50-95) for *Left Volley*, indicating that YOLOv8 detects *Bandeja* movement better.

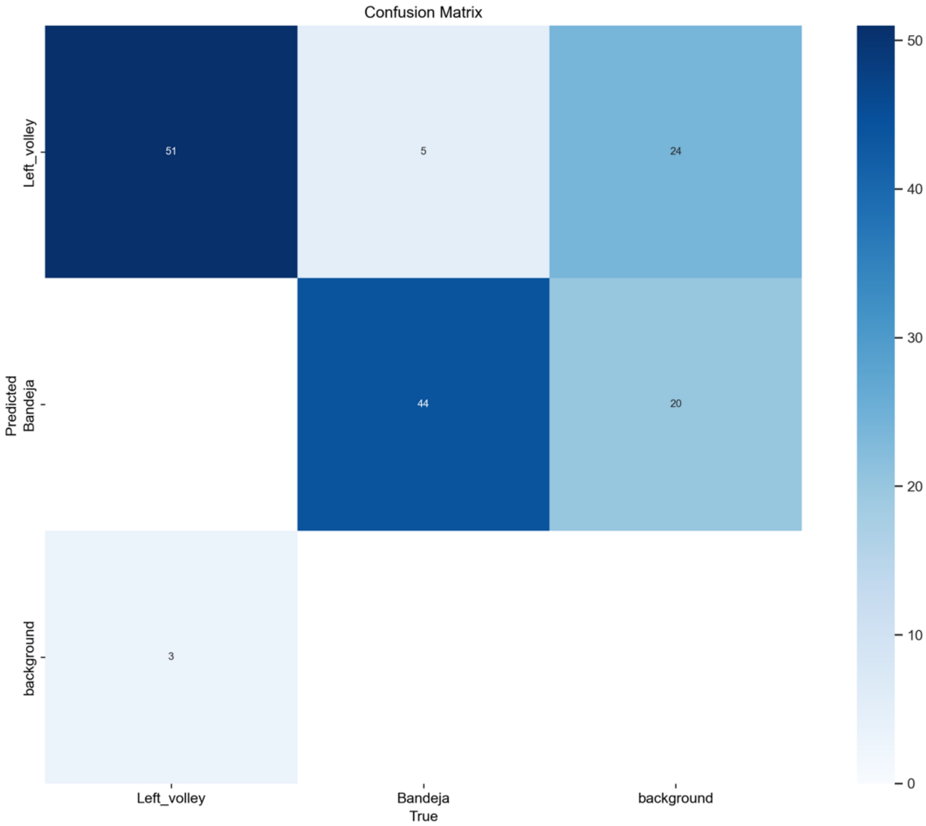

Figure 24 - Confusion Matrix of YOLOv8 in the test dataset

In Figure 24 we can see the confusion matrix on the test data, which was 147 images. The percentage of correct predictions is referred to as accuracy. The accuracy in this case is 65%, which implies that YOLOv8 accurately identified 65% of the items in the test dataset. Precision is the proportion of correct positive predictions. In this scenario, the precision is 68%, which means that 68% of the objects predicted by YOLOv8 were present in the image. Recall is the percentage of all objects correctly identified in an image. Here the recall is 92%, indicating that YOLOv8 recognized 92% of the total number of images.

The YOLOv8 object detection results are generally good. The precision and recall are both quite high, and the accuracy is reasonable. This shows that YOLOv8 can detect most objects in photos with relatively few false positives or false negatives. There is, however, still potential for development. The precision could be slightly enhanced, while the accuracy could be raised. This could be accomplished by training the model on additional data or by employing a new training approach.

Overall, YOLOv8 is a promising object detection system with numerous applications. YOLOv8's accuracy, precision, and recall are expected to increase more in the future.

There are two main reasons why the results aren't higher in precision and accuracy, namely the lack of training data inserted into the model and the other reason why the *Bandeja* has better results than the *Left Volley* is that the *Bandeja* is an easier movement to identify visually and makes it easier for the YOLOv8 to identify this movement more easily.

The YOLOv8 object identification results for the *Left Volley* and *Bandeja* strokes are quite good overall. Both types of strokes have high mAP (50-95) ratings, indicating that YOLOv8 can detect both types of strokes with good accuracy.

As we can see in Figure 2, the image preparation step (Grid) is critical since it guarantees that the input image is in a format that YOLOv8 recognizes. The resizing step ensures that the image is the same size as the photos used to train YOLOv8. The normalizing stage ensures that the image's pixel values are within a defined range. This helps to keep the CNN from saturating and ensures that it can learn the image's features properly.

Then the CNN identifies the patterns in the image during the feature extraction process. CNN is a deep learning model composed of many layers of neurons. Each layer of neurons learns to recognize different types of visual features. The CNN's earliest layers learn to recognize simple characteristics like edges and corners. The CNN's later layers learn to recognize increasingly complicated features, such as object forms and textures.

There are some possible explanations for YOLOv8's high recall and low accuracy in this example, such as the training data being skewed toward particular categories of objects. As a result, YOLOv8 may be better at recognizing specific things while being worse at detecting others. The test data could be more difficult to interpret than the training data. As a result, YOLOv8 may be less accurate on test data. The training data may be overfitted by YOLOv8 and could make YOLOv8 less accurate with new data.

It is critical to address these potential concerns to increase the accuracy of YOLOv8. This could be accomplished by gathering additional training data representative of the types of objects that YOLOv8 will be used to detect.

You can also see that many of the predictions are made against a neutral background, which means that the model does not predict wrong movements. Practically every prediction the model makes about the move to be executed is correct, so it's possible to say that despite having many false positives, it's a reliable model for detecting the *Bandeja* and *Left Volley*. This is another reason why the Recall metric is so high, relative to the other two measures.

Chapter 5 – Conclusions and future research

5.1. Main conclusions

Nowadays, technology has had a huge impact on sport in general and the way it is practiced. It's important to realise that this is the fastest growing sport in terms of players and that many players are growing up and moving towards becoming professional players. As this is a little explored area, the integration of Artificial Intelligence into Padel game in this dissertation allows us to explore various concepts and techniques that allow players and coaches to evolve in their decision-making.

Building an AI system to optimize Padel game is a complex process that takes time to develop and get the details right so that statistics are generated as accurately as possible, and it was with the help of the Quinta do Padel club, which made its videos available in order to build the input data for this system.

This master's thesis explored on many aspects of this undertaking, such as data collection and preprocessing, model selection, and evaluation. The successful development of a padel movement detection model shows the power of machine learning in improving our understanding of sports performance. Coaches, athletes, and sports enthusiasts can gain valuable insights into player behavior, tactics, and overall game dynamics by leveraging modern technology to collect and analyze movement data, and the choice of data sources and the quality of data preprocessing play a critical role in the accuracy and reliability of the model. Model construction requires the selection of a suitable machine learning algorithm as well as feature engineering. Investigating other algorithms, such as convolutional neural networks (CNNs). The Padel movement detection model goes beyond the sport. It can be utilized for a variety of things, such as player performance analysis, opponent reconnaissance, and even fan engagement via augmented reality experiences and live statistics.

Using Canny edge detection with Hough Lines Transformation to detect a Padel court marks a significant improvement in computer vision and sports facility recognition. The effectiveness of these algorithms in reliably detecting the limits and lines of a Padel court from photographs or video footage was proved in this master's thesis, contributing to the broader field of image processing and object detection, even if it's not completely finished, it's a technique that's essential for building the AI system for the Padel game.

Despite substantial effort and careful investigation for this master's thesis, properly tracking a Padel ball using these strategies proved elusive. Several reasons contributed to the failure to achieve the anticipated result. For starters, monitoring fast-moving objects in a dynamic and unpredictable Padel gaming environment presented various obstacles. Because of the ball's high speed, unpredictable bounces, and the presence of players and equipment, traditional computer vision and tracking algorithms struggled to handle complicated tracking scenarios. Furthermore, due to the camera resolution, the integration of the KCF algorithm, which is known for its success in object tracking, did not produce the expected results. It's important to realize that identifying the ball is a process that requires GPU computing, such as adding a class to YOLOv8 that can identify the ball.

It’s a process that requires time to explore several techniques on a variety of subjects and it's necessary to have a considerable volume of annotated images to train the movement identification model. Sometimes, when a technique for detecting the padel court or even the ball doesn't work, you have to do the whole process all over again, with the exploration phase and then the implementation. However, it's a system that greatly benefits both the sport and its practice, as well as the evolution of techniques used in computer vision.

Due to the shortage of studies in this particular area, this study provides a very important contribution both on a professional level and on a practical and academic level.

5.2. Research limitations

This research has some limitations. The quality and quantity of available data can significantly impact the performance and generalizability of Machine Learning models. Building powerful AI models frequently necessitates a large amount of computational capacity. Computing power and access to specific hardware constraints may limit the complexity and size of your AI system. Creating, training, and fine-tuning machine learning models takes time. The time limits imposed by the dissertation timeline may limit the scope and depth of this research.

5.3. Future research proposals

For future work, the following proposals are presented:

For future work, it is important to train a new model with more data than that used so that the model's performance is more accurate in detecting movements, creating more bounding boxes in different images.

For the implementation and identification of the Padel court it would be interesting to explore the concept of stereo vision, since there are two cameras on opposite sides, to be able to calibrate them to obtain all the blind spots. This would be a good technique because the resolution of the cameras isn't very precise and it doesn't allow us to automatically calculate the furthest points of the Padel court.

Regarding tracking the Padel ball, it could be beneficial to use Tracknet, which is a deep learning network, to track the tennis ball from broadcast videos in which the ball images are small, since in the data in question the ball is very small and therefore difficult to identify.

It would be very beneficial to be able to exploit the lens distortion of field cameras in future work, as they contain the barrel effect or fisheye effect, the field is not in the best condition to detect their lines.

References

[1] J. C. Ibáñez, B. J. S.-A. Martínez, S. G. Benítez, and M. Echegaray, “Evolución del pádel en España en función del género y edad de los practicantes. / Evolution of padel in Spain according to practitioners’ gender and age.,” *Cultura, Ciencia y Deporte*, vol. 12, no. 34, 2017.

[2] P. M. BORREGO, O. FERNANDES, J. ALEGRETE, N. BATALHA, S. VİLLAFAİNA, and J. A. PARRAÇA, “VARIABILIDADE DA FREQUÊNCIA CARDÍACA NO PADEL AMADOR,” *Sağlık Akademisi Kastamonu*, vol. 7, no. Special Issue, 2022, doi: 10.25279/sak.1138161.

[3] J. I. Priego, J. O. Melis, S. Llana-Belloch, P. Pérezsoriano, J. C. G. García, and M. S. Almenara, “Padel: A Quantitative study of the shots and movements in the high-performance,” *Journal of Human Sport and Exercise*, vol. 8, no. 4, 2013, doi: 10.4100/jhse.2013.84.04.

[4] B. J. Sánchez-Alcaraz, D. Muñoz, F. Pradas, J. Ramón-Llin, J. Cañas, and A. Sánchez-Pay, “Analysis of serve and serve-return strategies in elite male and female padel,” *Applied Sciences (Switzerland)*, vol. 10, no. 19, 2020, doi: 10.3390/APP10196693.

[5] W. Ertel, *Introduction to Artificial Intelligence (Undergraduate Topics in Computer Science)*. 2017.

[6] P. Wang, “On Defining Artificial Intelligence,” *Journal of Artificial General Intelligence*, vol. 10, no. 2, 2019, doi: 10.2478/jagi-2019-0002.

[7] D. Gunning and D. W. Aha, “DARPA’s explainable artificial intelligence program,” *AI Mag*, vol. 40, no. 2, 2019, doi: 10.1609/aimag.v40i2.2850.

[8] M. Brady, “Artificial intelligence and robotics,” *Artif Intell*, vol. 26, no. 1, 1985, doi: 10.1016/0004-3702(85)90013-X.

[9] B. Y. E. Brynjolfsson and A. McAfee, “Artificial intelligence for real,” *Harv Bus Rev*, vol. July, no. 1, 2017.

[10] R. Geetha and D. Bhanu Sree Reddy, “Recruitment through artificial intelligence: A conceptual study,” *International Journal of Mechanical Engineering and Technology*, vol. 9, no. 7, 2018.

[11] S. Wang, L. Niu, and N. Li, “Research on Image Recognition of Insulators Based on YOLO Algorithm,” in *2018 International Conference on Power System Technology, POWERCON 2018 - Proceedings*, 2019. doi: 10.1109/POWERCON.2018.8602149.

[12] K. Yanai and Y. Kawano, “Food image recognition using deep convolutional network with pre-training and fine-tuning,” in *2015 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2015*, 2015. doi: 10.1109/ICMEW.2015.7169816.

[13] D. Han, Q. Liu, and W. Fan, “A new image classification method using CNN transfer learning and web data augmentation,” *Expert Syst Appl*, vol. 95, 2018, doi: 10.1016/j.eswa.2017.11.028.

[14] A. I. Khan and S. Al-Habsi, “Machine Learning in Computer Vision,” in *Procedia Computer Science*, 2020. doi: 10.1016/j.procs.2020.03.355.

[15] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep Learning for Computer Vision: A Brief Review,” *Computational Intelligence and Neuroscience*, vol. 2018. 2018. doi: 10.1155/2018/7068349.

[16] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *Proceedings of 2017 International Conference on*

Engineering and Technology, ICET 2017, 2018. doi: 10.1109/ICEngTechnol.2017.8308186.

[17] O. Abdel-Hamid, A. R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE Trans Audio Speech Lang Process*, vol. 22, no. 10, 2014, doi: 10.1109/TASLP.2014.2339736.

[18] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10. 2010. doi: 10.1109/TKDE.2009.191.

[19] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1. 2021. doi: 10.1109/JPROC.2020.3004555.

[20] A. Sai Bharadwaj Reddy and D. Sujitha Juliet, "Transfer learning with RESNET-50 for malaria cell-image classification," in *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019*, 2019. doi: 10.1109/ICCSP.2019.8697909.

[21] S. T. Krishna and H. K. Kalluri, "Deep learning and transfer learning approaches for image classification," *International Journal of Recent Technology and Engineering*, vol. 7, no. 5, 2019.

[22] S. Gao, I. W. H. Tsang, L. T. Chia, and P. Zhao, "Local features are not lonely - Laplacian sparse coding for image classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010. doi: 10.1109/CVPR.2010.5539943.

[23] A. Mihanpour, M. J. Rashti, and S. E. Alavi, "Human Action Recognition in Video Using DB-LSTM and ResNet," in *2020 6th International Conference on Web Research, ICWR 2020*, 2020. doi: 10.1109/ICWR49608.2020.9122304.

[24] C. Zheng *et al.*, "Deep Learning-based Human Pose Estimation: A Survey," *ACM Comput Surv*, vol. 56, no. 1, 2024, doi: 10.1145/3603618.

[25] Amithesh Gupta, "A Comprehensive Understanding of Human Pose Estimation," Feb. 2022.

[26] Q. Dang, J. Yin, B. Wang, and W. Zheng, "Deep learning based 2D human pose estimation: A survey," *Tsinghua Sci Technol*, vol. 24, no. 6, 2019, doi: 10.26599/TST.2018.9010100.

[27] M. Ackermann *et al.*, "Unresolved Gamma-Ray Sky through its Angular Power Spectrum," *Phys Rev Lett*, vol. 121, no. 24, 2018, doi: 10.1103/PhysRevLett.121.241101.

[28] Hamza Mahmood, "CRACK DETECTION AND MEASUREMENT IN TENNIS COURTS: AN EDGE DETECTION APPROACH," University of Delaware, 2022.

[29] R. O. Duda and P. E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures," *Commun ACM*, vol. 15, no. 1, 1972, doi: 10.1145/361237.361242.

[30] K. F. T. F. Megan Fazio, "Tennis Ball Tracking: 3D Trajectory Estimation using Smartphone Videos," *Department of Electrical Engineering, Stanford University*.

[31] J. Wei, C. F. Li, S. M. Hu, R. R. Martin, and C. L. Tai, "Fisheye video correction," *IEEE Trans Vis Comput Graph*, vol. 18, no. 10, 2012, doi: 10.1109/TVCG.2011.130.

[32] C.-F. L. S.-M. H. M. I. R. R. M. and C.-L. T. Jin Wei, "Fisheye Video Correction," *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 18, NO. 10, OCTOBER 2012*, 2012.

[33] S. Yadav and S. Payandeh, “Critical Overview of Visual Tracking with Kernel Correlation Filter,” *Technologies*, vol. 9, no. 4. 2021. doi: 10.3390/technologies9040093.
[34] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “High-speed tracking with kernelized correlation filters,” *IEEE Trans Pattern Anal Mach Intell*, vol. 37, no. 3, 2015, doi: 10.1109/TPAMI.2014.2345390.