Application of Behavioral Cloning on FIFA

Puranjay Rajvanshi
MS Computer Science
University of Southern California
Los Angeles, USA
pranjvans@usc.edu

Sangeeth Koratten
MS Computer Science
University of Southern California
Los Angeles, USA
koratten@usc.edu

Shelly Mehta
MS Computer Science
University of Southern California
Los Angeles, USA
shellyme@usc.edu

Rishika Verma
MS Computer Science
University of Southern California
Los Angeles, USA
rverma@usc.edu

Vandit Maheshwari
MS Computer Science
University of Southern California
Los Angeles, USA
vanditma@usc.edu

Aviral Upadhyay
MS Computer Science
University of Southern California
Los Angeles, USA
aviralup@usc.edu

***Abstract*— As the world of machine learning grows, bots are seeming to perform and execute all the actions that a human can do. With numerous instances of the former already stated in the world, another one would be creation of an artificial intelligent bot that mimics the actions and the controls that humans perform via the game controller for the game FIFA to best of its ability. The work focused to clearly distinguish various items present on the field like the ball, other players and the goalpost using convolutional neural networks (CNN) as well as choosing the most suitable option of movement from the given options on the game controller via Long Short Term Memory [1] (LSTM).**

I. Introduction

Artificial intelligence is defined as a study of rational agents. A rational agent is one which is capable of doing expected actions to maximize its performance measure on the basis of percept sequence and its built-in knowledge base. In this circumstance, the built-in knowledge base would be the colossal set of training data with snippets of the field with players and the ball in different locations and positions. The percept sequence in this case are the actions performed by the humans (in layman terms, the buttons pressed on the remote controller) to play the game of FIFA virtually.

FIFA, short for Fédération Internationale de Football Association was founded in 1904 to oversee international competition among the national associations of 8 countries in the beginning, and now contains 6 confederations and 211 national associations. Played by approximately 250 million players in over 200 countries, makes it the world's most popular sport. The game is played on a rectangular field called a pitch with a goal at each end. Each team comprises of 11 players (excluding substitutes), one of which is a goalkeeper. Players are not allowed to touch the ball with hands or arms while it is in play except the goalkeeper. Other players mainly use their feet to strike or pass the ball hence the name Foot-ball. The objective of the game is to outscore your opponents by moving the ball past the opposing goal line. These guidelines are accompanied by a set of rules and restriction making it more than a game of athletics and strength.

II. Literature Survey

There has been a lot of research on artificial intelligence in which the bot is given a set of rules or actions to learn from. Here, the model simply trains itself by imitating a human playing the game, thereby introducing the concept of behavioral cloning [2]. Behavioral cloning is a method by which human sub cognitive skills can be captured and reproduced in a computer program. While in action, the decisions made by human in different situations are recorded and then presented to these as input to the learning bot. For the detection of objects part, computer vision technology is used. Object detection draws bounding boxes around detected items thus allowing location of those objects [3][4]. This technique is vital to track the physical movement of the different articles in the game and to track other moving objects such as the ball while simulating the soccer game. An example-based learning approach is used where a model of an object class is derived implicitly from a set of training images. This ensures that the model utilizes the algorithm which is specialized to a specific domain related to the surroundings. The field of object detection has grown throughout the years and paved its way to single shot detection [5]. This technique uses a single shot to detect multiple objects present in an image using multibox. At the time of prediction, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Since it eliminates proposal generation and subsequent

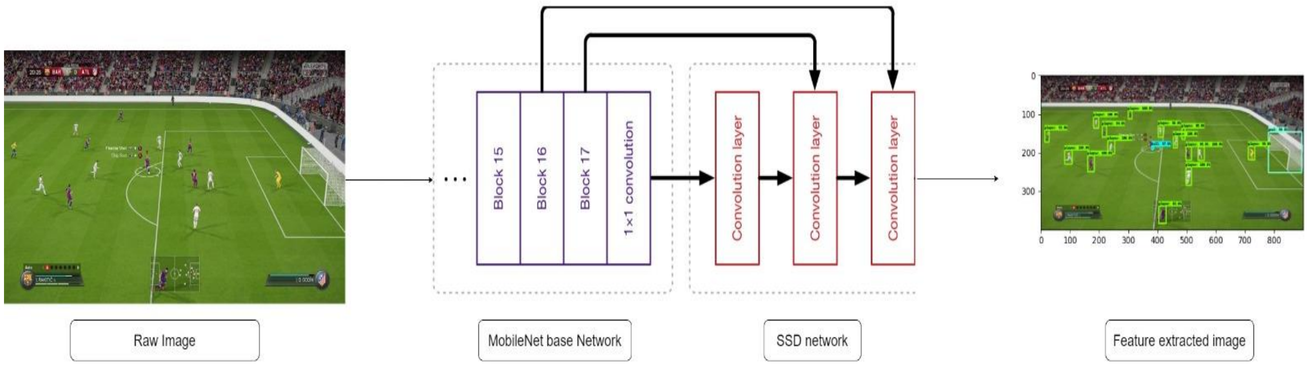

Fig. 1. Pipeline for Object Detection

pixel or feature resampling stages and encapsulates all computation in a single network, it is easy to train and integrate into the system. Also, the world is now well versed with the concept of CNNs. It is considered as a machine learning algorithm used to implement image recognition and image classification. Though this concept has been invented some time back now, it is still highly used for object detection. The model takes each input image and passes it through a series of convolutional layers with filters(kernels), pooling and fully connected layers and then applies SoftMax function to classify an object with some probabilistic values between 0 and 1. There are multiple types of CNN used for a variety of reasons. MobileNet are convolutional neural network architectures whose number of trainable parameters can be controlled by two hyperparameters – width and resolution multiplier. Although using MobileNet instead of single shot detection decreases average detection accuracy, it provides more speed relatively. Using MobileNet architecture is key while designing a game playing intelligent agent keeping in mind the constrained hardware settings.

III. Methodology

To achieve our end result of the project, the tasks are divided into different segments. One for recognition of various different kind of objects i.e. feature extraction and one for deciding the most appropriate in-game action.

A. *Data Collection and Preprocessing*

For the purpose of the project, the model is trained on different images of the game environment. To fulfil this, setting up the environment is key. The downloaded desktop version of FIFA 18 is what is used as the environment system. Further, various time spaced screenshots of the former are collected and stored in .jpg formats. To record in-person actions, key-press simulation is used to communicate the output that needs to be taken. The image is cropped to obtain desired frame of the game window.

B. *Object Detection*

This is the first segment which pertains to distinguishing various articles present on the field in each game window. The Object Detection API of TensorFlow is used to get the bounding boxes for each object on all the images. The labelImg library supported by Tensorflow is used for the above task. Installation is done using PIP. The labelImg library creates a separate window which helps to manually create bounding boxes for object and also helps in annotation. This tool helps us to create training images. After the annotation is completed, for each training image, a corresponding .xml file is created which holds the object tag along with its xmax, ymax, xmin and ymin coordinates. Moreover, the .xml files are then converted into .csv format files which are in turn used to generate tfrecord files thereby, being passed as input to the CNN model. The CNN model used is the MobileNet CNN. MobileNet is an already trained CNN model which is part of the Object Detection API. The output layer is updated to accommodate the output for the given training image dataset. Hyperparameters are tuned to obtain better results on the development dataset.

IV. Preliminary Results

With the completion of the first segment which caters to object detection of snippets of the game field, Figure 1 shows the outline of a starting image till the point of bounding boxes that annotate the players, the ball and the goalpost. Figure 2 depicts the localization loss between the predicted box and the ground-truth box parameters.

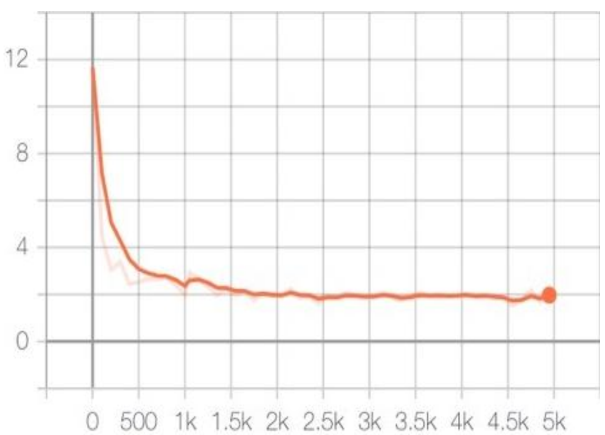

Fig. 2. Loss function over 5000 epochs

A. *Challenges*

Though there are many libraries in today's world that make tedious tasks like object detection simpler, understanding Object Detection APIs still remains a difficult task. The complexity of installation and then integration with the code base with these tools is higher and sometimes challenging.

Object Detection algorithms are still very sensitive to good frames per second (fps). Thus, on a normal machine, object detection for videos is a sphere which could require some improvement. The limited amount of annotated data currently available for object detection proves to be another substantial hurdle. Class imbalance proves to be an issue for classification problems and object detection is no exception.

V. Future Work

For the current time being, only the first chunk of the project, which pertains to observing and spotting objects correctly, has come to completion. The focus is now on the second stage of the project where the files of detected objects from the training data will be fed into recurrent neural networks, in this case LSTMs. This step will be used to make the model learn the appropriate choices of moves that can be performed during the play.

References

[1] Danilo Sorano1, Fabio Carrara2, Paolo Cintia,1Fabrizio Falchi2, and Luca Pappalardo2 ; Automatic Pass Annotation from Soccer Video Streams Based on Object Detection and LSTM

[2] Robert G. Abbott; Behavioural Cloning for Simulator Validation

[3] Juarez Monteiro_1, Nathan Gavenskiy1, Roger Granada_, Felipe Meneguzziz and Rodrigo Barrosz ; Augmented Behavioral Cloning from Observation

[4] Okihisa UTSUMI, Koichi MIURA, Ichiro IDE, Shuichi SAKAI, Hidehiko TANAKA; AN OBJECT DETECTION METHOD FOR DESCRIBING SOCCER GAMES FROM VIDEO

[5] Wei Liu1, Dragomir Anguelov2, Dumitru Erhan3, Christian Szegedy3, Scott Reed4, Cheng-Yang Fu1, Alexander C. Berg1; SSD: Single Shot MultiBox Detector