

# YOLOv7: Custom Object Detection in the field of Athlete Sports

**First Author<sup>1</sup>, Second Author<sup>2</sup>, Third Author<sup>3</sup>** (Size 11 Bold)

<sup>1</sup> Department, University or college Name, State, Country Name. (Size 10 &italic)

<sup>2</sup> Department, University or college Name, State, Country Name. (Size 10 &italic)

<sup>3</sup> Department, University or college Name, State, Country Name. (Size 10 &italic)

**Abstract** - Video analytical approach in the field of athlete sport is a game changer. Object detection is the basis of any video analytical system which is powered via Computer Vision. Whenever 2-Dimensional or 3-Dimensional media are involved, predominantly regional convolutional algorithm makes into play. Here comes State-of-the-Art YOLO (You Only Look Once) algorithm surpassing SSD model and all R-CNN model descendants. This paper aims at detecting the subjects of a basketball game at higher f1-score and mAP trained through YOLOv7 model.

**Keywords** - five Video Analytics, Computer Vision, YOLO, SSD, R-CNN

## 1. Introduction

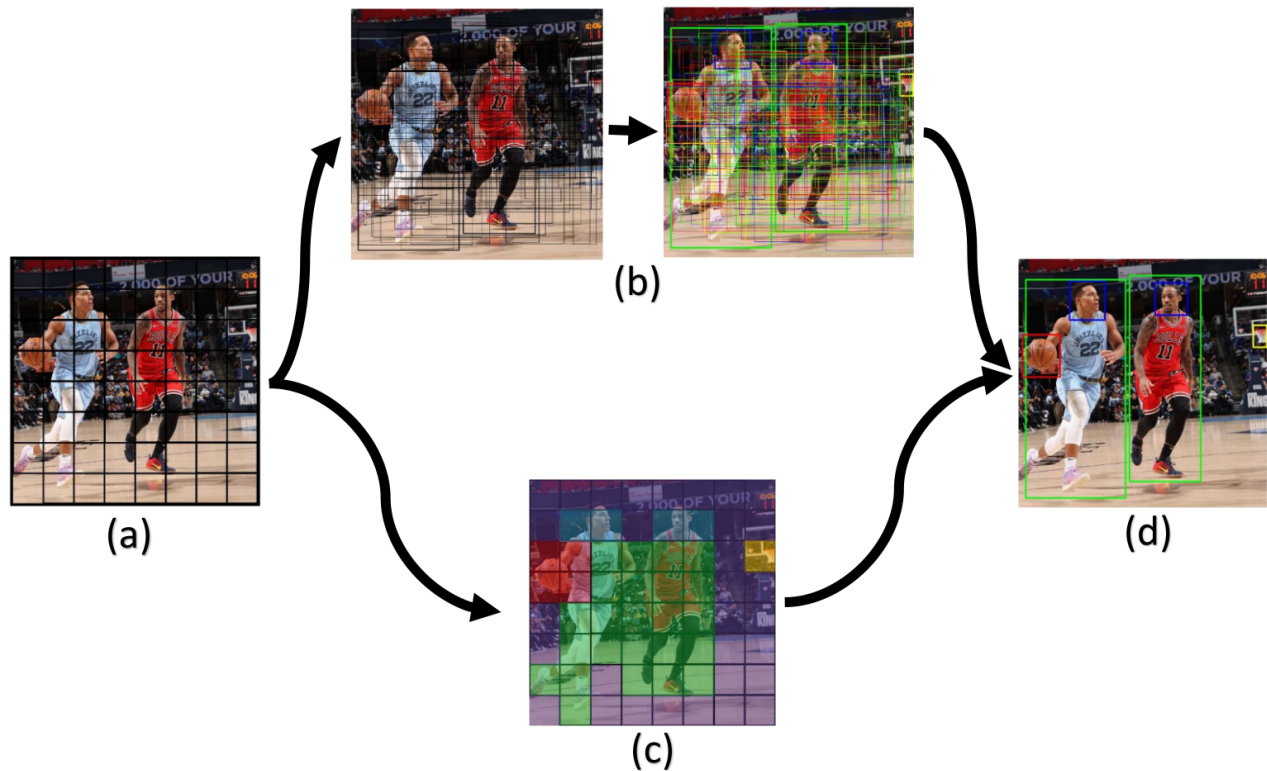
Computer Vision, a revolutionary technology that augments the field of monitoring and analytics. Deep learning approach, if implemented with right NN architecture & algorithms can achieve robust object detection model. In order to build a video analytical system for an athlete sport, the system must be able to classify and localize the core subjects in the game even in different lighting conditions, different environments and distorted media which significantly lays stronger bases for the system. YOLO is the industry leading object detection model in terms of accuracy, speed and efficiency.

YOLO algorithm covers up object classification, object localization and object detection in a single-shot/proposal-free approach.

Basically, this paper utilizes the recent advancements in the object detection algorithms in order to overcome the shortcoming in traditional algorithms which are influenced by different environments, lighting condition, distortions and motion blurs.

The concept of Transfer learning is accounted as a G-factor in the field of video analytics due to its flexibility in deployment and its minimal hardware requirement, which enables object detection in both real-time and computational-time.





(a)  $S \times S$  Grid on input (b) Bounding boxes + Confidence  
(c) Class probability map (d) Final detection

## 2. State-of-the-art Algorithm

YOLO (You Only Look Once) is an end-to-end single CNN model that brings out the  $S \times S$  grid from the input (Figure 2, (a)).

Bounding boxes are plotted thus forming a dimension cluster based on the training data which is calculated by the defined loss function. Through the concept of IoU (Intersection over Union), the Confidence of the object is estimated (Figure 2, (b)).