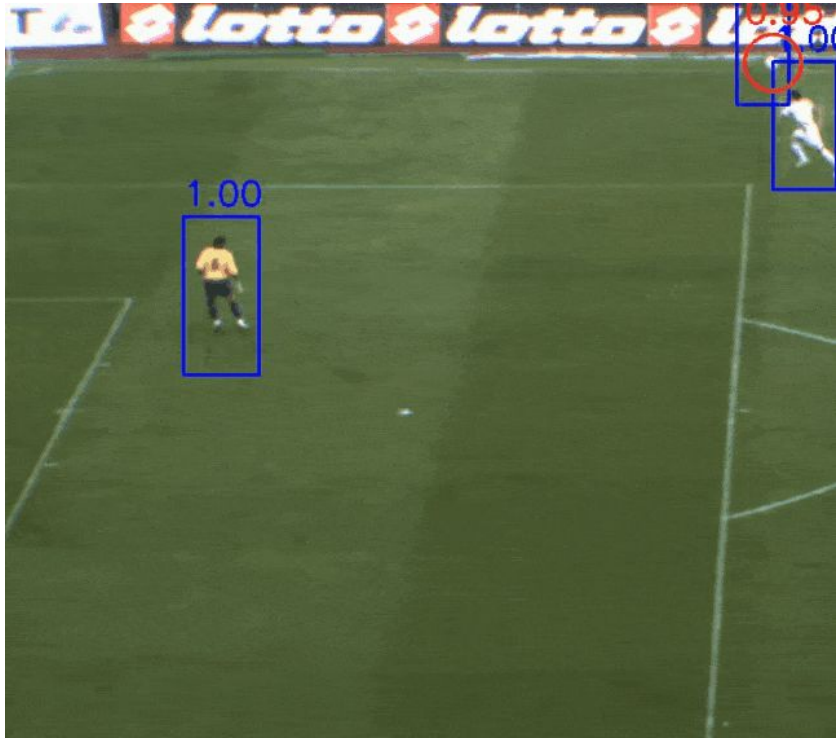
Detection of Players and Football in Broadcasted Video Stream

Soumyadip Santra
Instructor: Sujoy Kumar Biswas
Machine Learning Systems, 2021
RKMVERI, Belur

Problem Statement

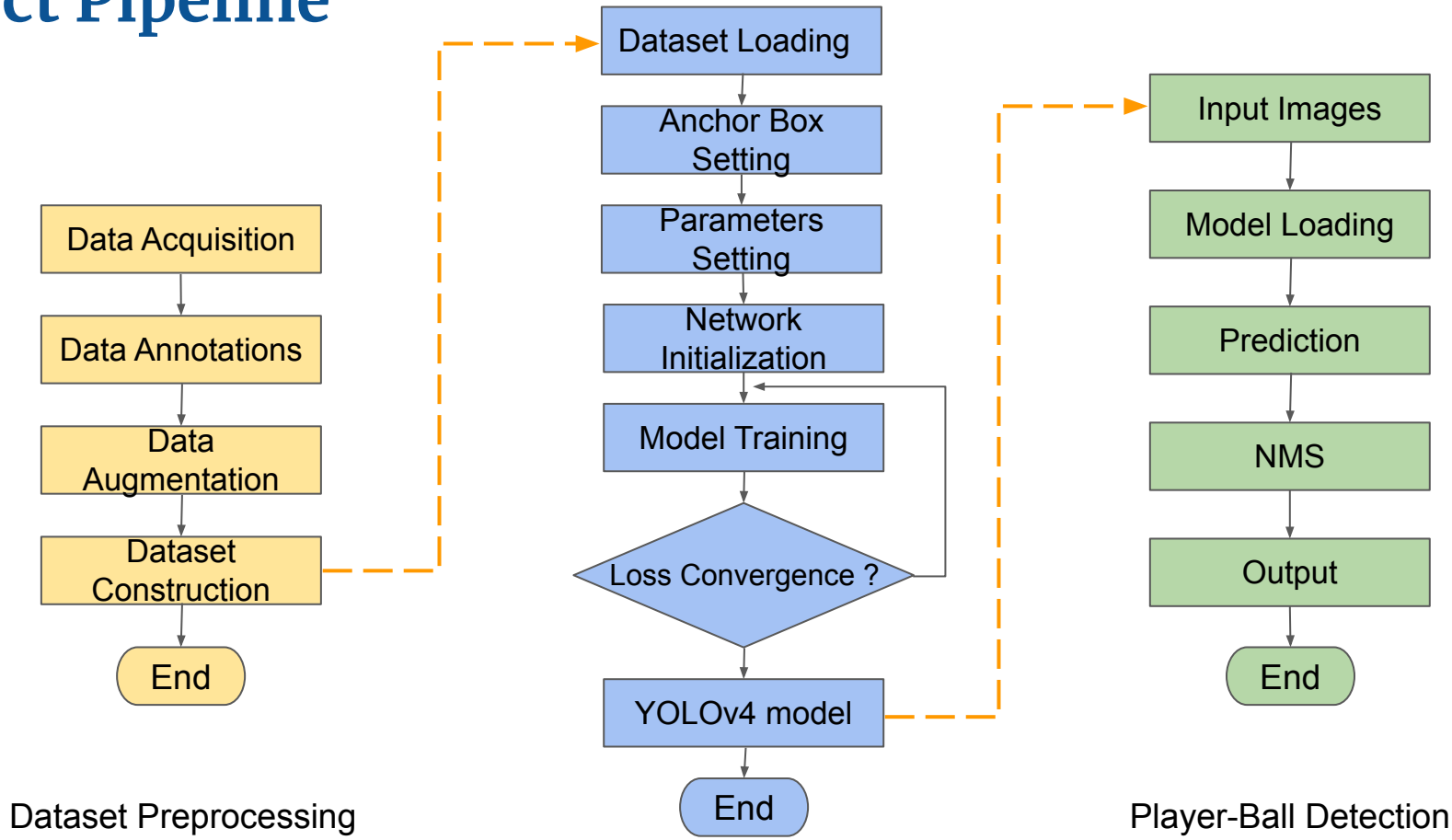
Given a broadcasted **Video** of a football match,

We aimed to develop a **Machine Learning System** using the help of **Deep Learning** and **Computer Vision** such that moving players on the pitch like **Players** and **Football** can be detected properly.



Source : https://arxiv.org/abs/1912.05445

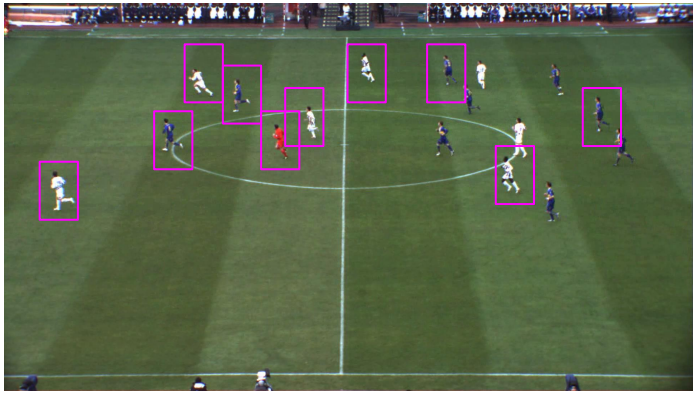
Project Pipeline

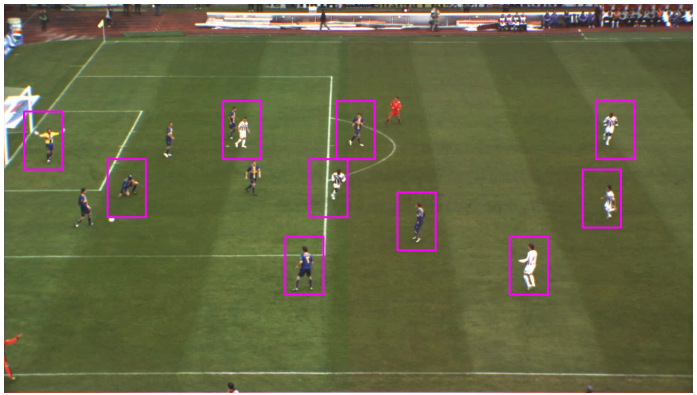


Dataset Description

ISSIA-CNR dataset :

- 6 synchronized, long-shot views of the football pitch.
  - Recorded at **25 FPS** with **1920x1088** resolution
  - Total frames **17993**
    - **129154** Players
    - **8336** Balls
- 6 annotation file in .xgtf format.
  - Contains bounding box info.
    - For Player (**xtl, ytl, width, height**)
    - For Ball (**x_center, y_center**)

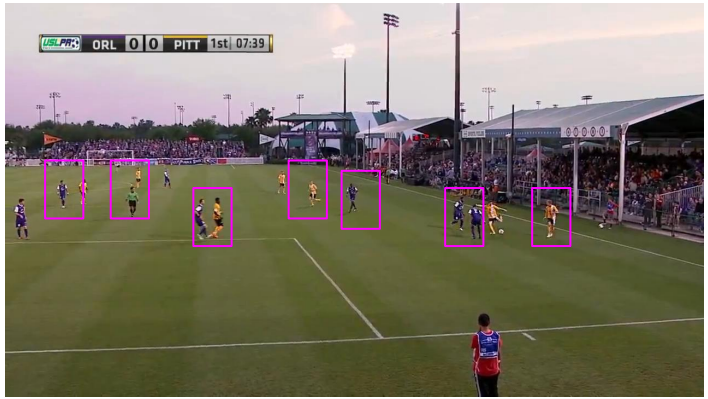


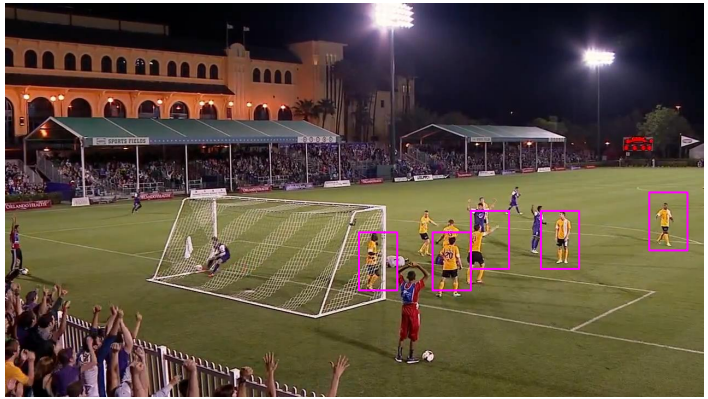


Dataset Description

SPD-BMV-2017 Dataset :

- Videos from 2 professional football match .
  - Recorded by 3 cameras at **30 FPS** with **1280x720x3** resolution.
  - Contains **2019** image frames
    - **22586** annotated Players
    - **2942** Balls
- 2 annotation file in .mt format.
  - Contains bounding box info.
    - Only for Player (**x_min, y_min, width, height**).
    - Annotated Ball using **CVAT**.
    - Ball bounding box (**x_min, y_min, width, height**)

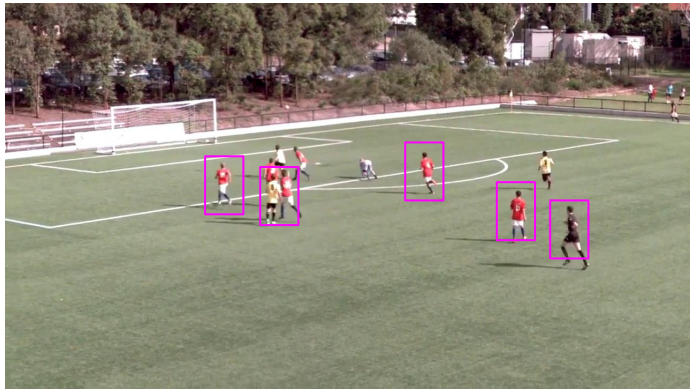


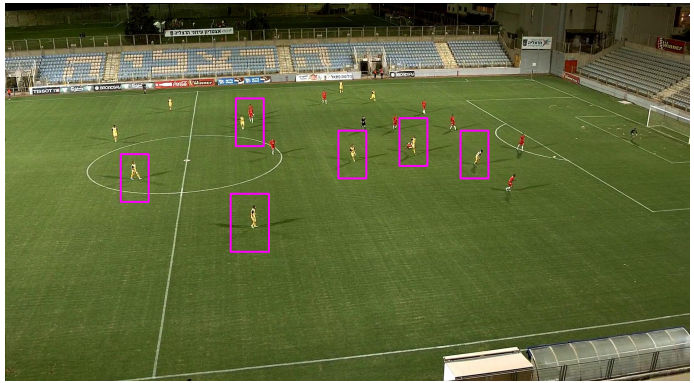


Dataset Description

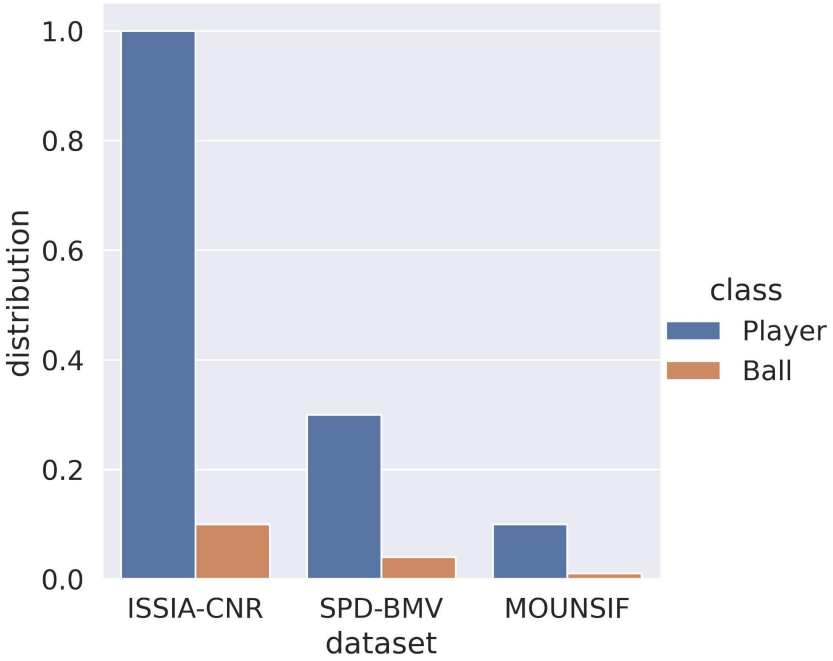
MOUNSIF Dataset :

- Videos from 2 professional football match .
  - Recorded by 5 cameras at **25 FPS** with **1280x720x3** resolution
  - Contains **500** .jpg image frames
    - **7799** annotated Players
    - **856** annotated Balls
- 2 annotation file in .txt format.
  - Contains bounding box info.
    - For player (**x_min, y_min,width , height**)
    - For Ball (**x_min, y_min,width , height**)

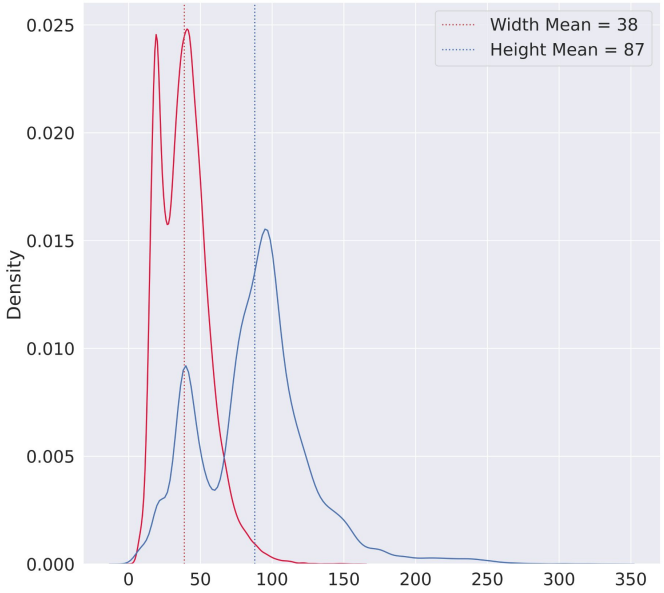




Dataset Description

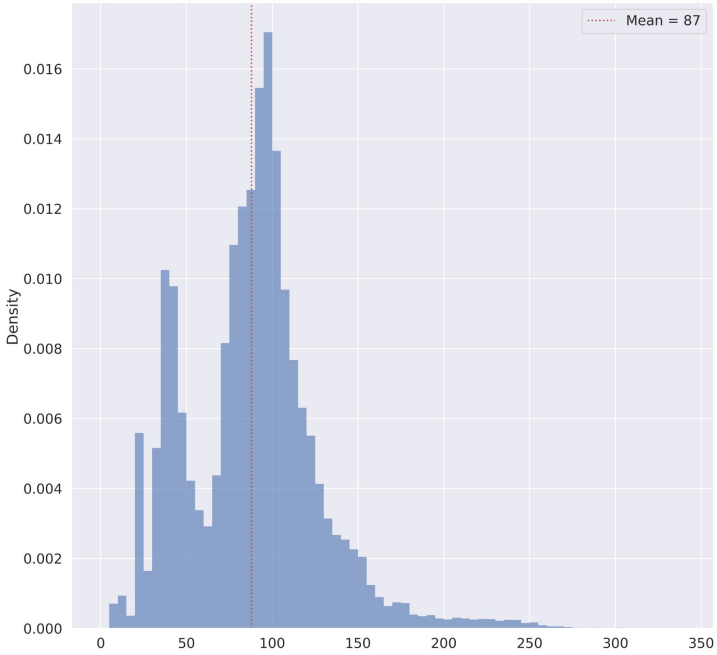


Class Distribution Plot

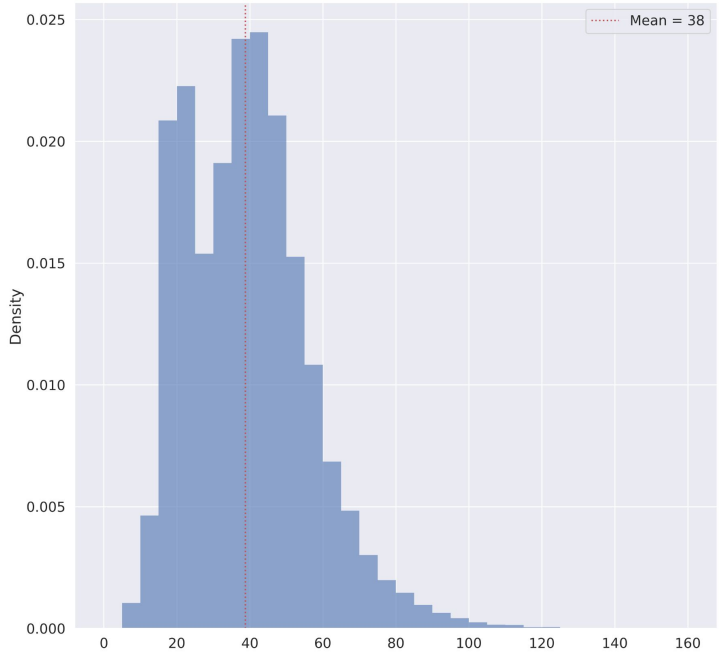


Height-Width Distribution Plot

Dataset Description



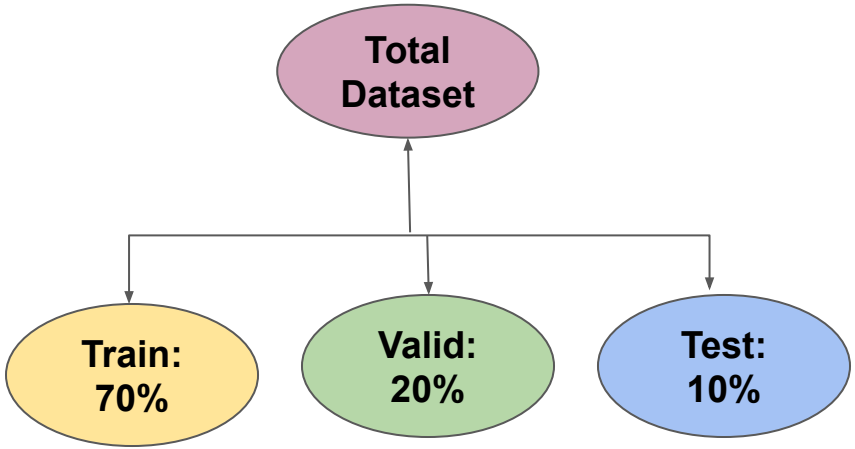
Height-Density Plot



Width-Density Plot

Dataset Construction

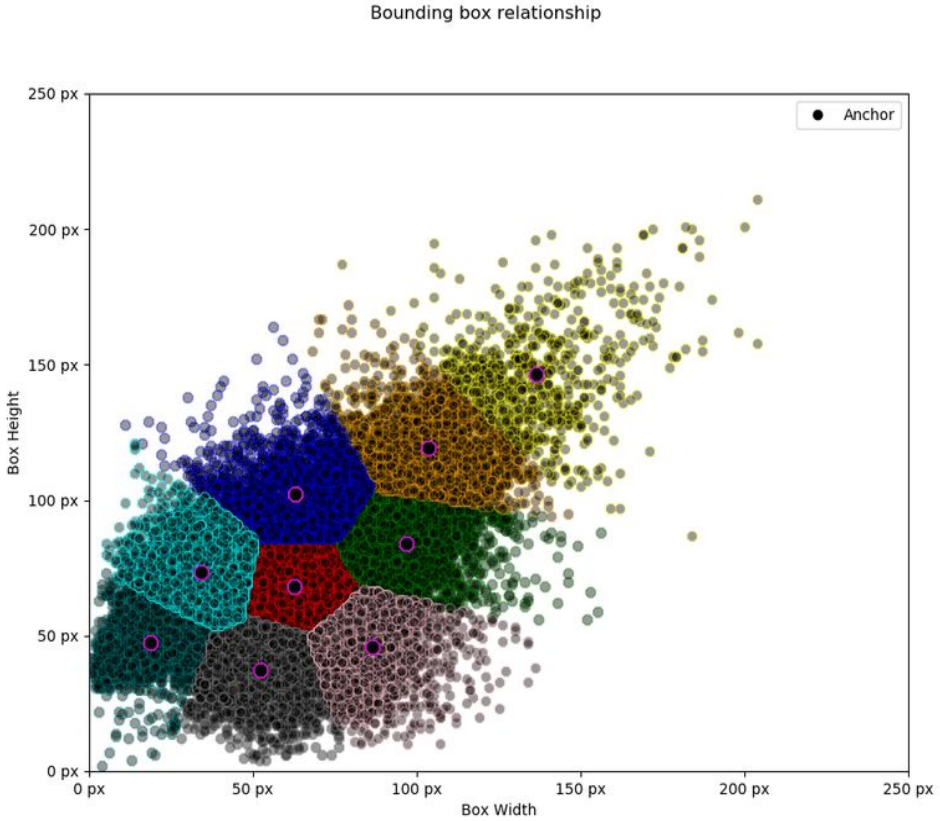
- Discarded garbage annotations.
- Extract frames from the videos using **ffmpeg** tool where each frame contains 3 color channel.
- Convert the dataset provided annotation format into model specific format -
  - (**class_id, xtl, ytl, xbr, ybr**) where player_id : **2**, ball_id : **1**
- Merge those three types of dataset and splitted into **Training, Validation** and **Testing** subset.



Prior Anchor box Generation

- Obtained **9 cluster centers** from heights, widths of all the bounding boxes using **K-means clustering.**
- **9 anchor box** for **3 types** of object
  - Small
  - Medium
  - Large

But What's It Needed For ?

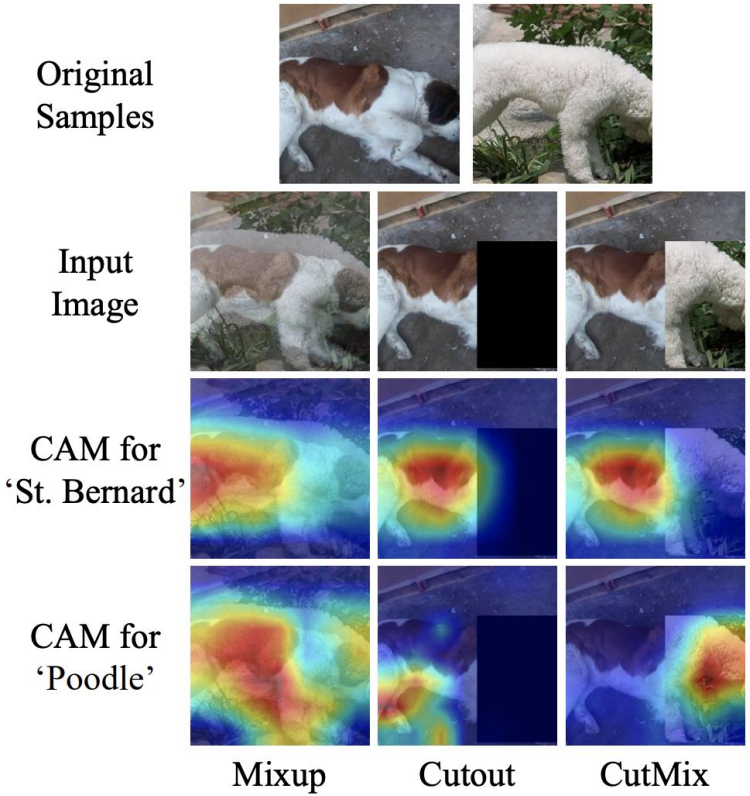


Data Augmentation Strategies

- **MixUp** : Overlaying of Image pairs proportionally with each-other.
- **CutOut** : Randomly masks square portion from images during training.
- **CutMix** : In CutMix, the cutout is replaced with a part of another image.

What's For ?

Reduces the chance of **Overfitting**.



Source: https://arxiv.org/pdf/1905.04899.pdf

Data Augmentation Strategies

MOSAIC : Combines 4 training images into one in certain ratios (instead of only two in CutMix).

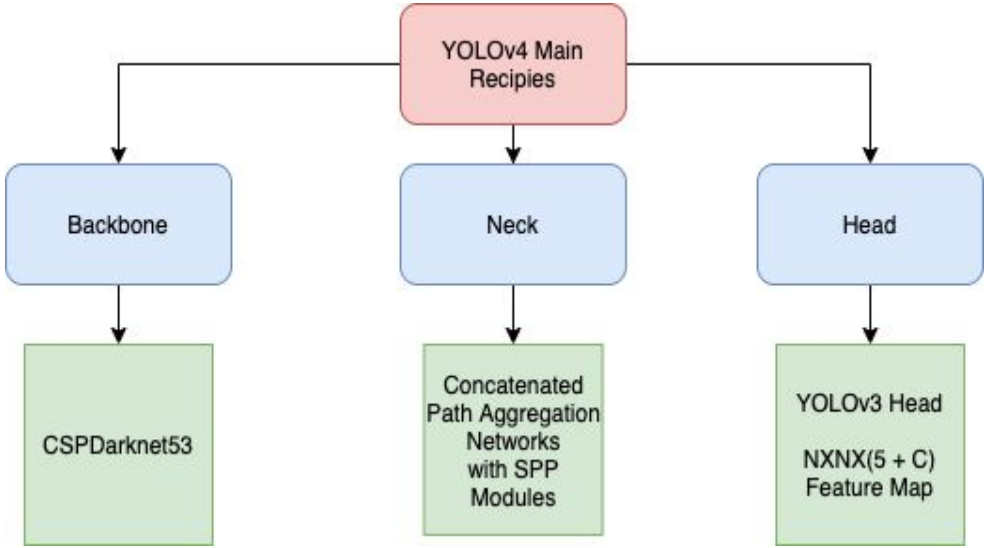
What's For ?

- Helps to learn how to identify **SMALLER** scaled objects.
- Reduces the needs for a large mini-batch size.

Source: https://arxiv.org/pdf/2004.10934.pdf

YOLOv4 Overview

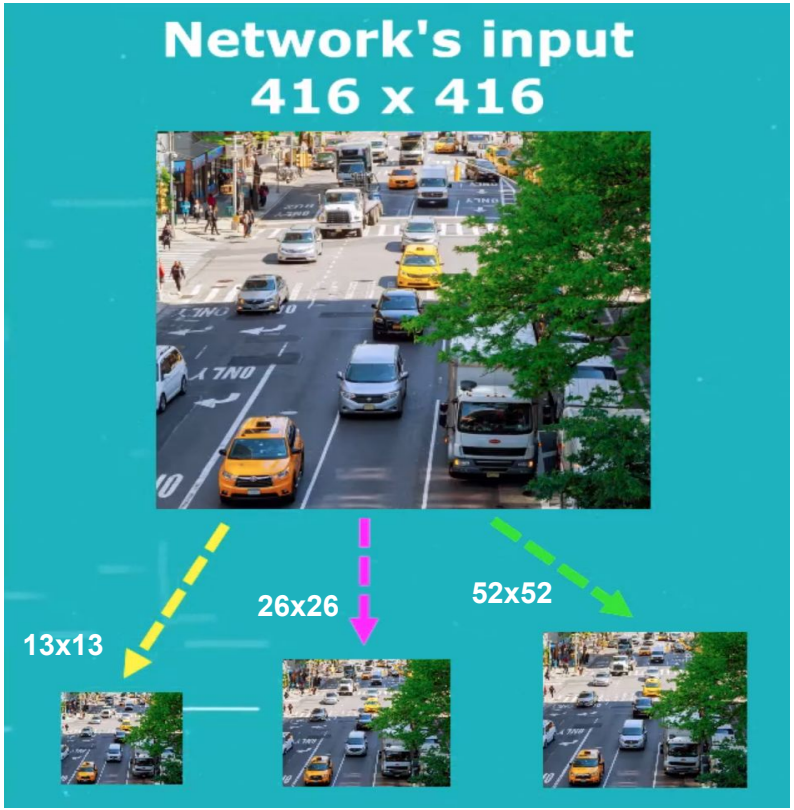
- Uses **CSPDarknet53** as **Feature Extractor**.
- **Neck** helps to add layers between the **Backbone** and the dense prediction block(**Head**).
- It's Prediction Block predicts bounding boxes in **3 scales** just like **YOLOv3**.



Source: https://files.ai-pool.com/a/7697d5bc15ad2b6d6bb1c3a86cc792cb.png

Detections at 3 Scales

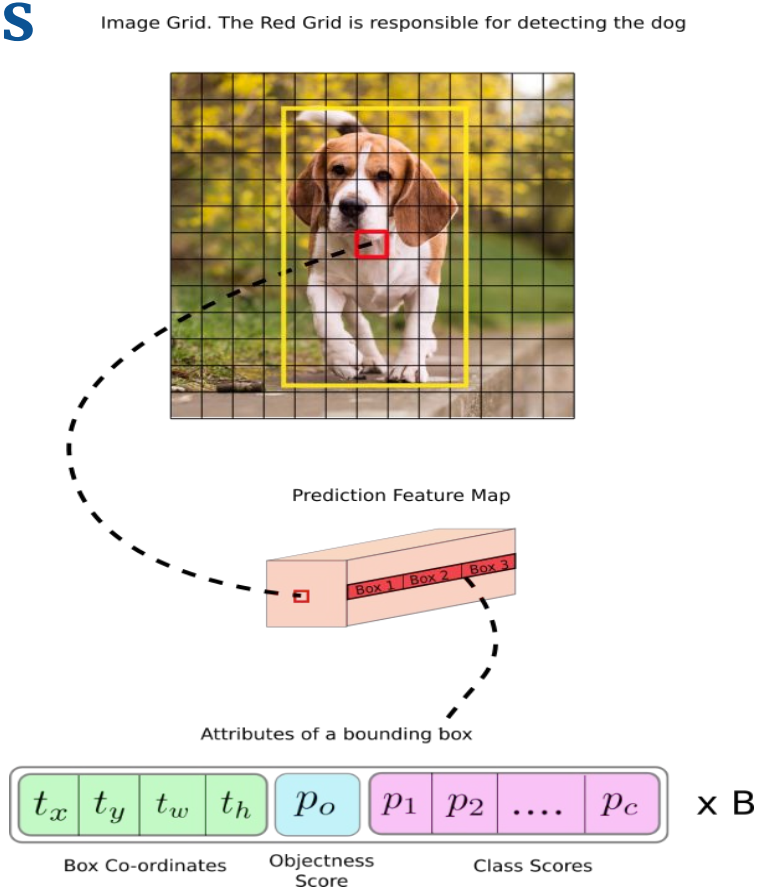
- **Downsample** the image at three separate places at the network.
- For large object detection:
  - Strides: 32
  - Output: **13x13**
- For medium object detection:
  - Strides: 16
  - Output: **26x26**
- For small object detection:
  - Strides: 8
  - Output: **52x52**



Thus It Performs Better at Detecting Smaller Objects Like Soccer Player or Ball in Aerial Images.

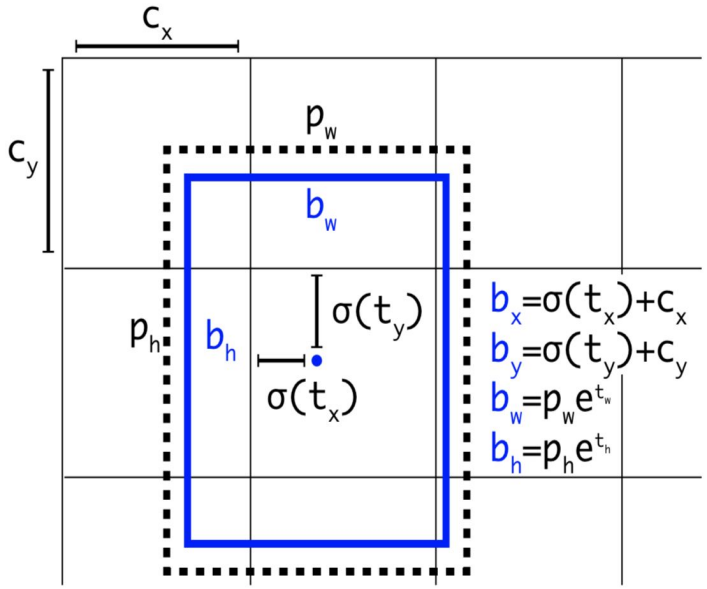
Feature Maps In Output Layers

- At each scale, every grid can predict 3 bounding boxes using 3 prior anchor boxes values.
- For instance, 416x416 image is downsample to 13x13 blocks, then the **RED** block predicts 255 values or **3*(4+1+80)** values.
- **Box-coordinates**: tx, ty are **offsets** and w,h are width-heights.
- **Objectness Score**: Confidence score of whether this block contains the center of any object in the actual image.
- **Class Probabilities**: Probabilities of the detected object belonging to a particular class.



Role of Prior Anchor Boxes

- Model gives (t_x, t_y, t_w, t_h) as bounding box information.
- **Center Coordinates:** Pass (t_x, t_y) to a **sigmoid** function, then add the top-left coordinates(C_x, C_y) to predict the actual coordinates(b_x, b_y) of the bounding box.
- **Bounding Box Dimension:** Dimensions of the bounding box are predicted by applying a log-space transformation to (t_w, t_h), then multiplying with an anchor box dim (p_w, p_h).
- Now (b_x, b_y, b_w, b_h) are actual bounding box coordinates.



Source:inverseai.com/media/blog_uploads/2020/12/09/image-20201209205805-1.png

Multiple Bounding Box for Same Object

- Need to keep the one with highest confidence score.
- **Non-Max Suppression :**

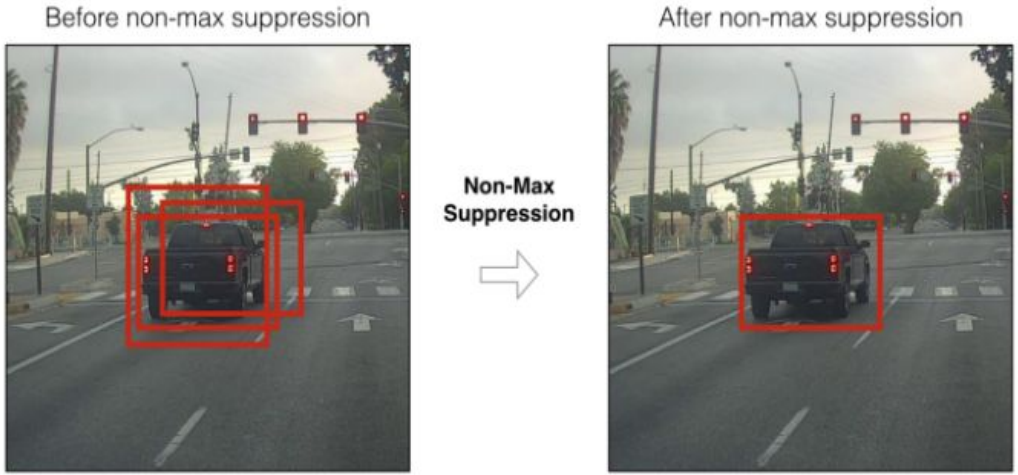
Step 1: Select the box with highest objectiveness score

Step 2: Compare the overlap (**intersection over union**) of this box with other boxes

Step 3: Remove the bounding boxes with overlap (**intersection over union**) >50%

Step 4: Move to the next highest objectiveness score.

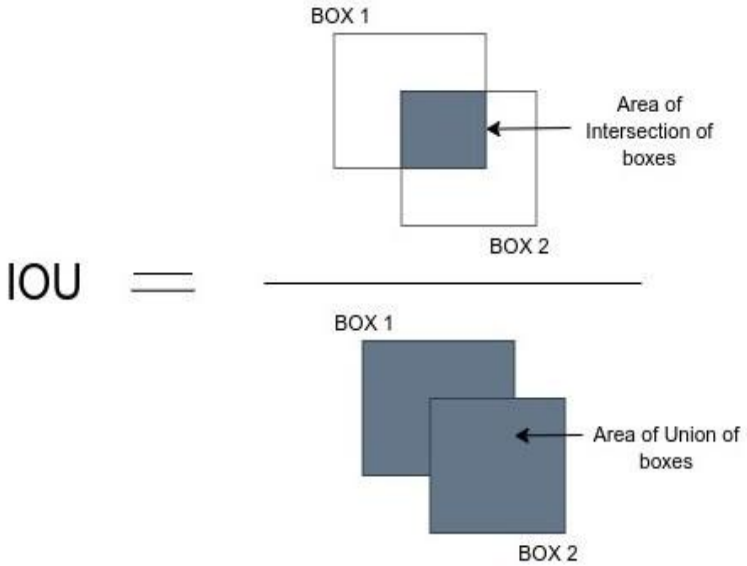
Step 5: Finally, repeat steps 2-4



https://www.inverseai.com/media/blog_uploads/2020/12/20/nms.bmp

Intersection Over Union

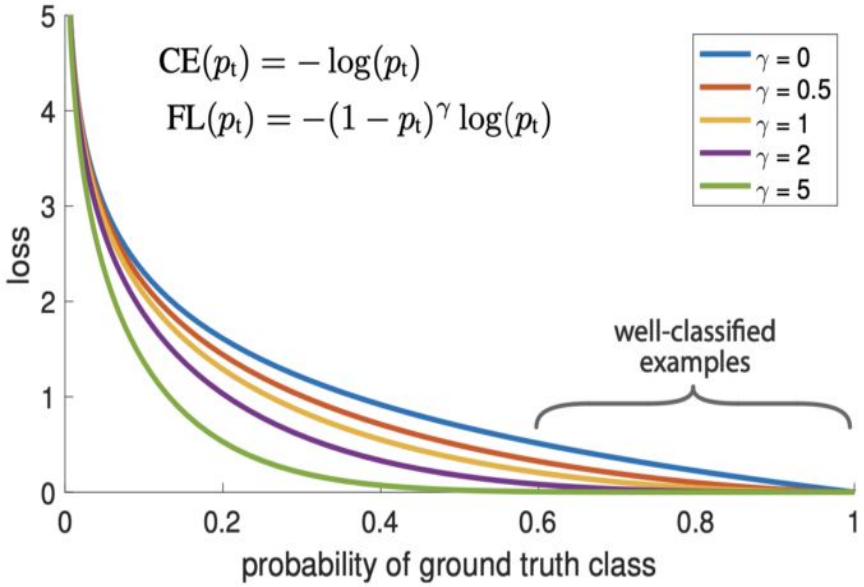
- Describe the **extent of overlap** of two boxes.
- The greater the region of overlap, the greater the **IOU**.
- Used in **NMS**, which eliminates multiple boxes that surround the same object.



*https://miro.medium.com/max/468/1*r0o3vX-x979Q84_lbJWS_g.jpeg*

Focal Loss in YOLOv4

- Works well when extreme imbalance between foreground and background.
- Based on the **cross-entropy loss** by introducing a $(1 - p_t)^{\gamma}$ coefficient.
- Focus the importance on the correction of misclassified examples.



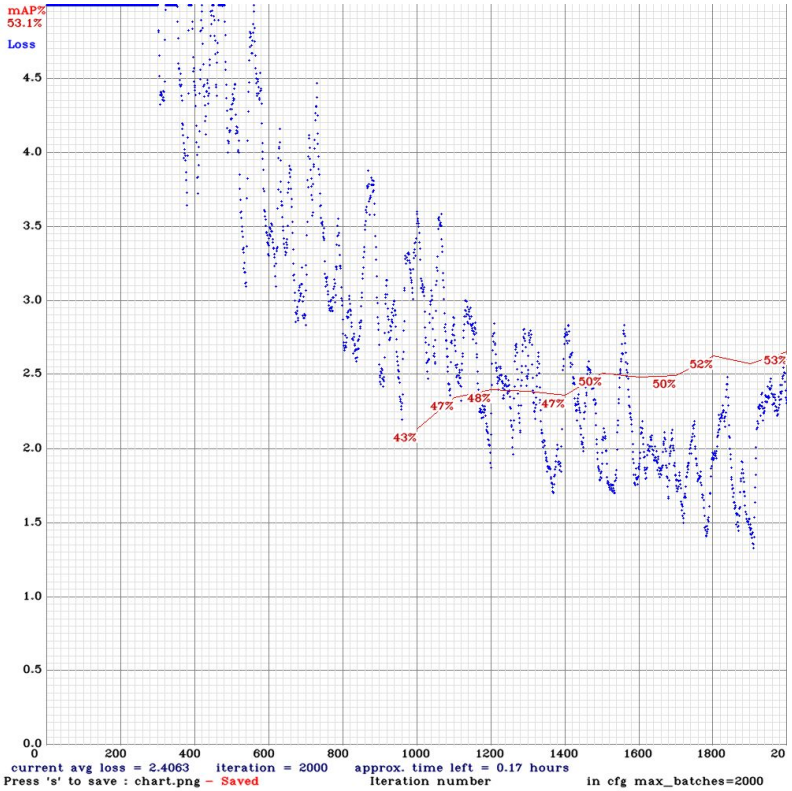
*https://miro.medium.com/max/700/1*kD5xdrktQit8zOkvYJqVIA.png*

Inference Using YOLOv4 DarkNet Framework

ISSIA-CNR Dataset :

	True Positive	False Positive	Average Precision
Player	1039	92	92.85%
Ball	7	25	13.28%

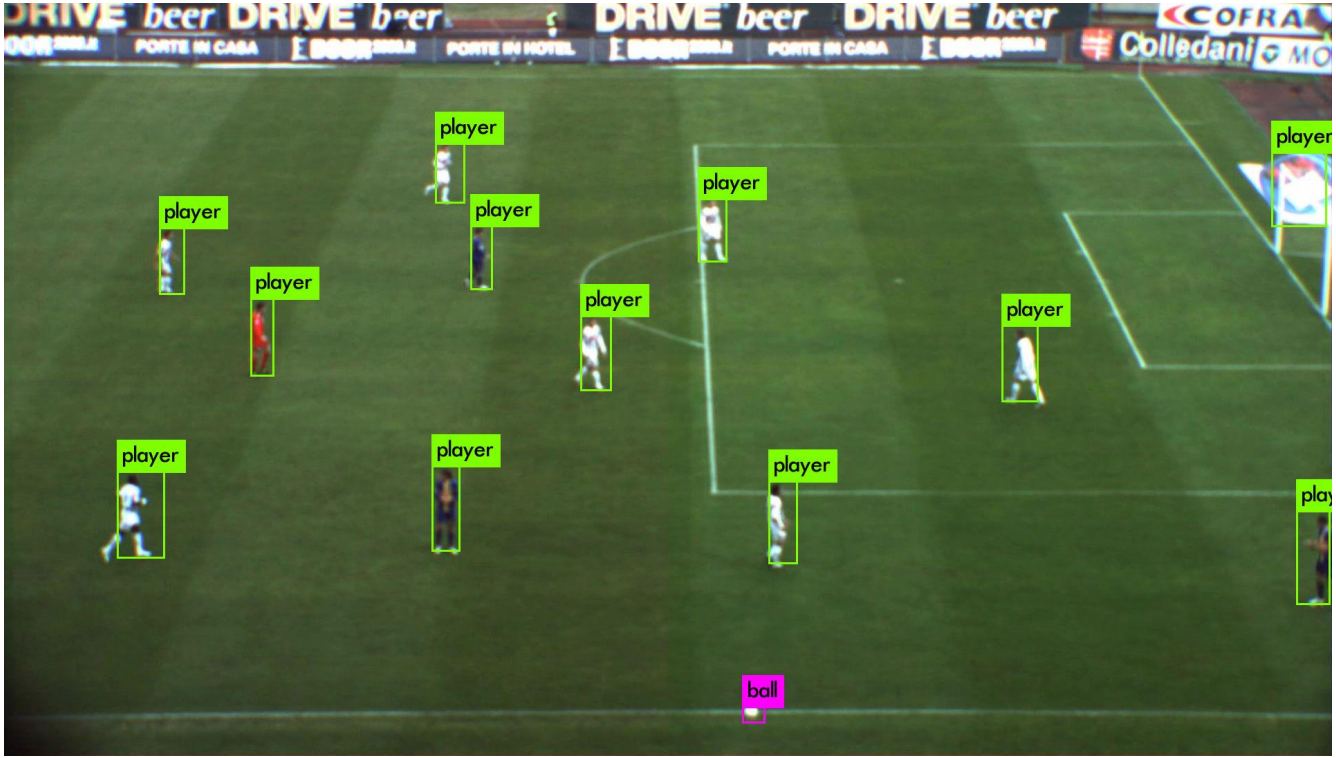
Precision	Recall	F1-score	mAP
0.90	0.90	0.90	0.53



Iteration-Loss, mAP Plot

Inference Using YOLOv4 DarkNet Framework

Detection Sample on ISSIA_CNR Dataset

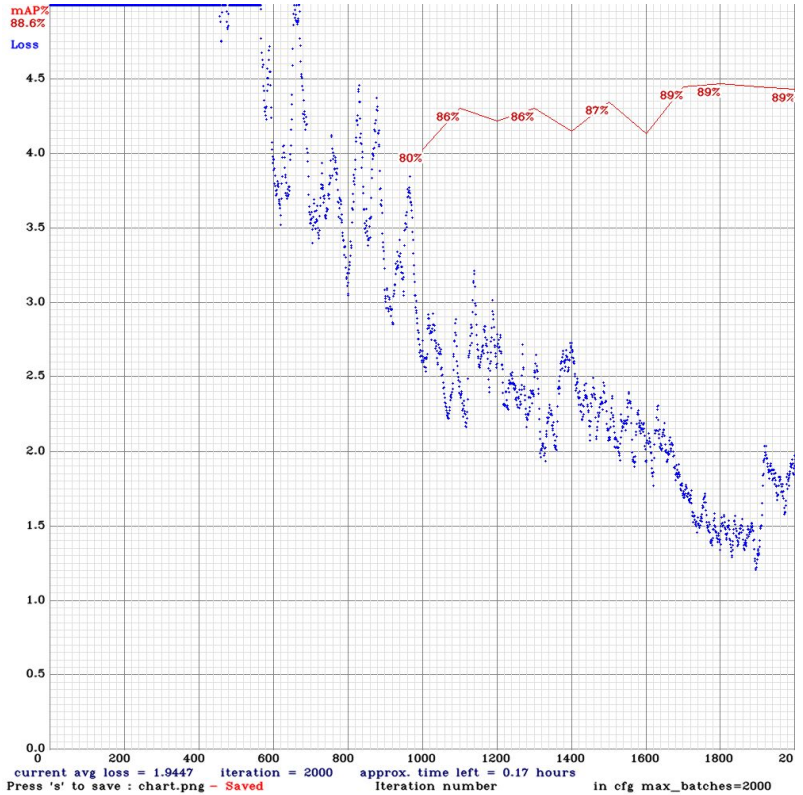


Inference Using YOLOv4 DarkNet Framework

SPD-BMV Dataset :

	True Positive	False Positive	Average Precision
Player	1075	74	98.35%
Ball	59	22	78.78%

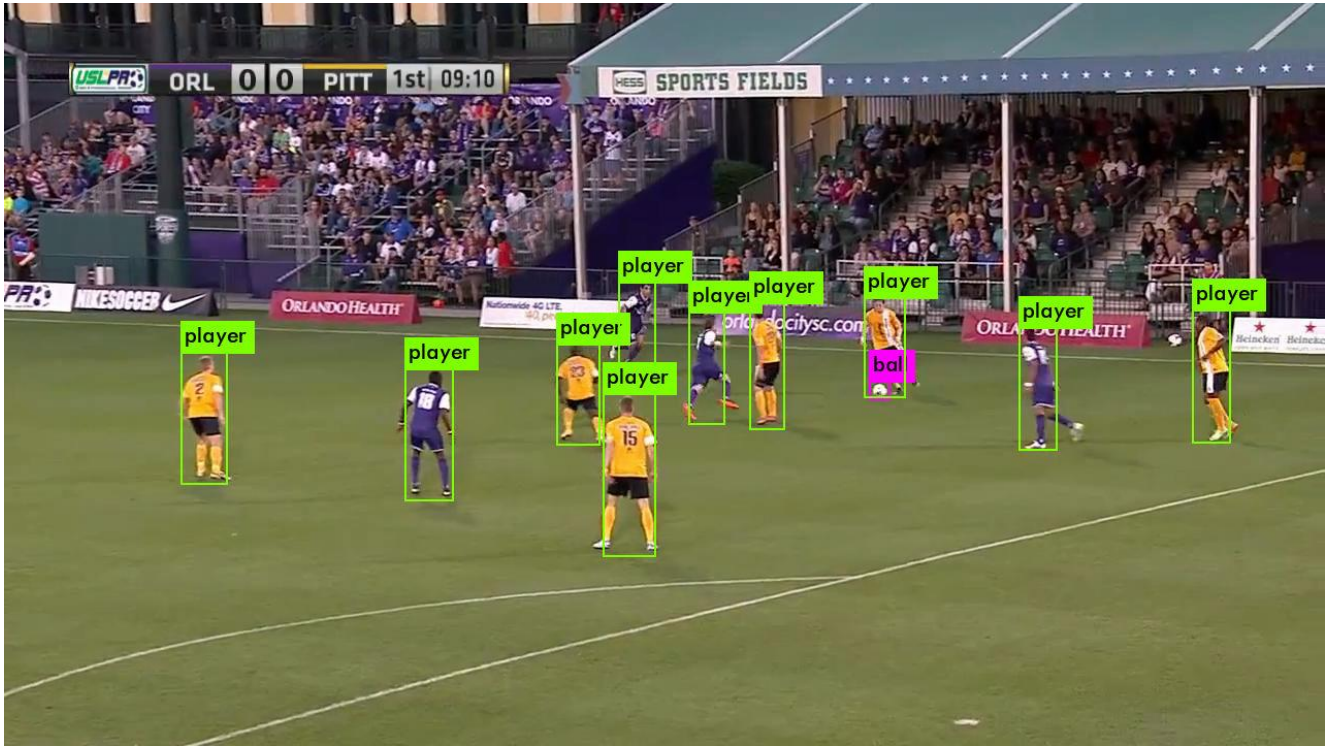
Precision	Recall	F1-score	mAP
0.92	0.97	0.94	0.8



Iteration-Loss, mAP Plot

Inference Using YOLOv4 DarkNet Framework

Detection Sample on SPD-BMV Dataset

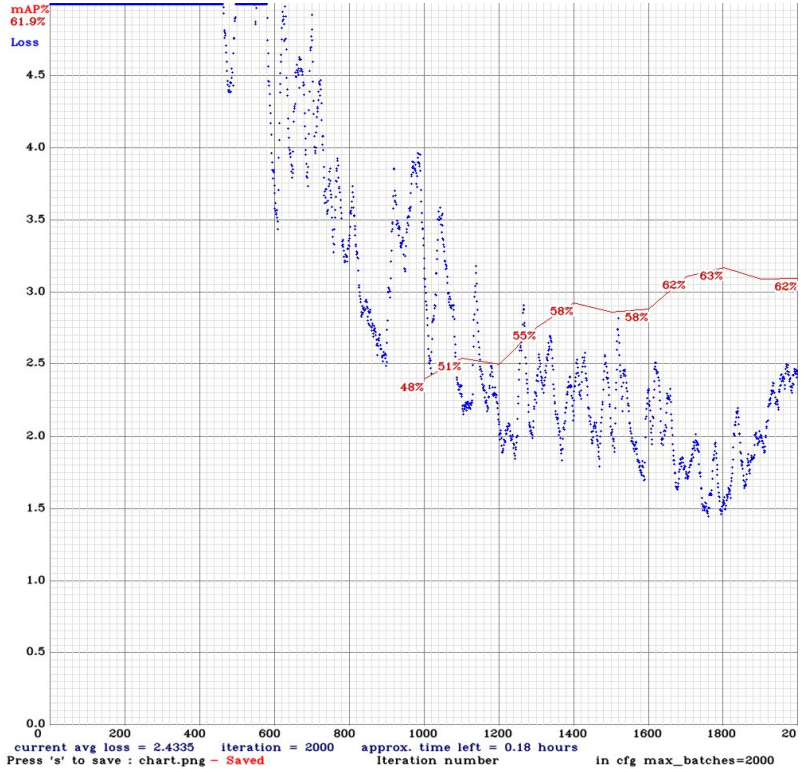


Inference Using YOLOv4 DarkNet Framework

MOUNSIF Dataset :

	True Positive	False Positive	Average Precision
Player	1450	164	91.34%
Ball	36	60	34.42%

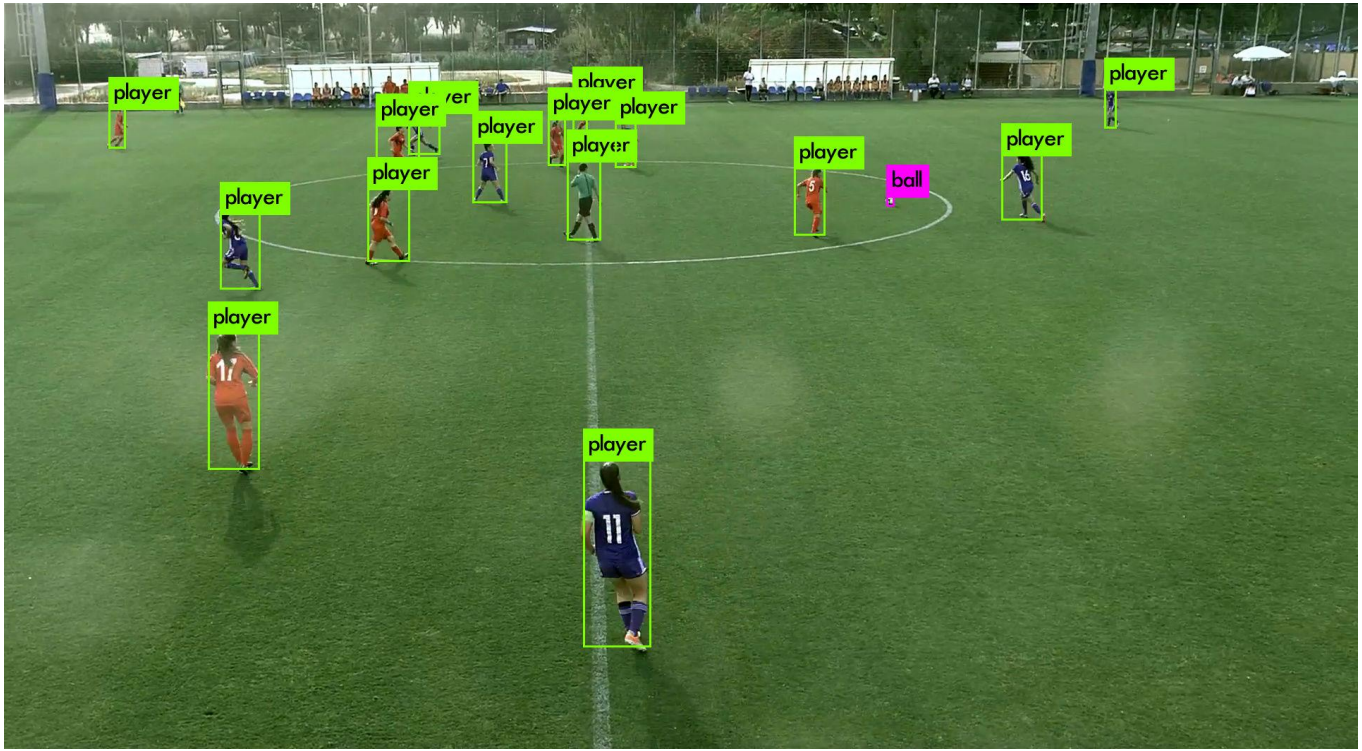
Precision	Recall	F1-score	mAP
0.87	0.89	0.88	0.61



Iteration-Loss, mAP Plot

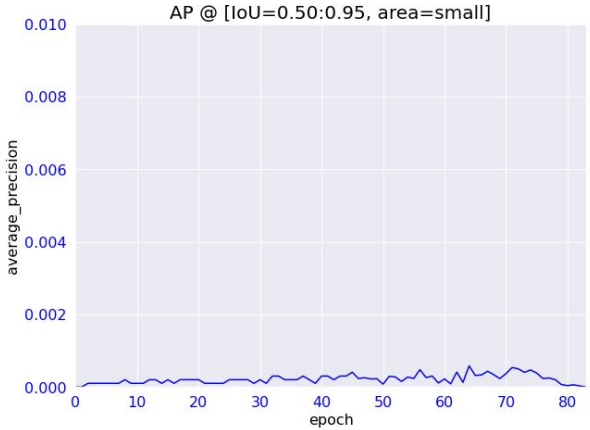
Inference Using YOLOv4 DarkNet Framework

Detection Sample on MOUNSIF Dataset

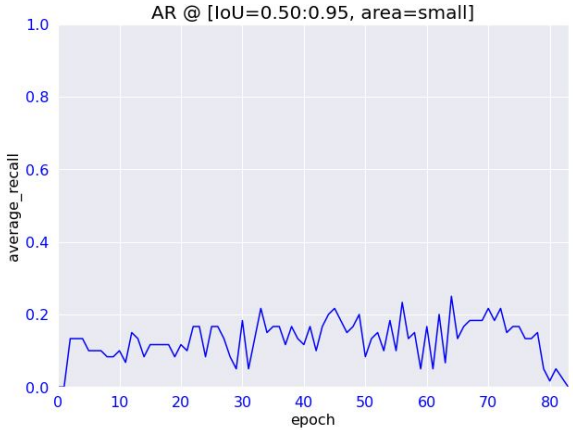


Training with PyTorch Implementation

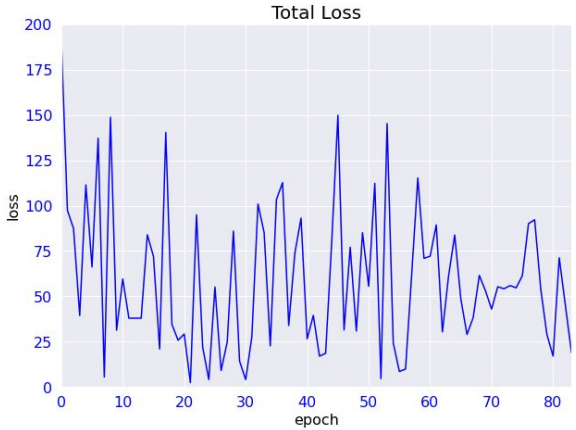
Trained for 80 epochs,
And It's a Complete Disaster.



Epoch-AP Plot



Epoch-AR Plot



Epoch-Loss Plot