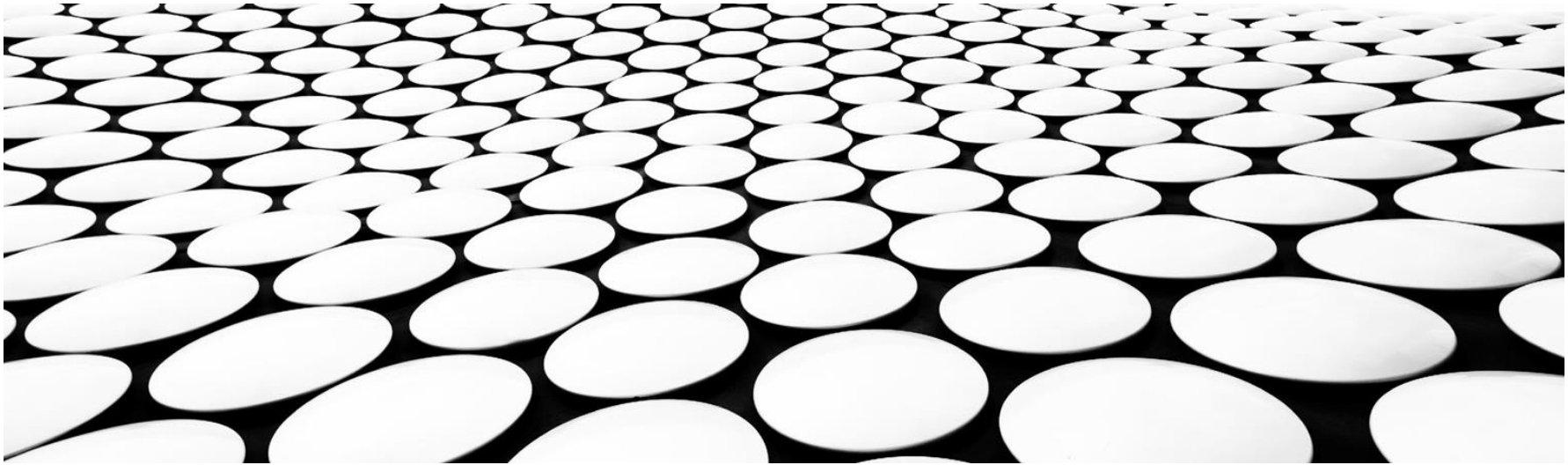
ICVGIP VISUAL DATA CHALLENGE 2020 (BIRDSAI TRACK)

TEAM BATMAN_RETURNS:
ANSHUL MITTAL (SIT, IIT-DELHI)
AMAN BHARDWAJ (SIT, IIT-DELHI)



ACKNOWLEDGEMENT

We would like to thank **Dr. Chetan Arora** for this opportunity and considerate guidance throughout the competition. We would also like extend our regards for **Ankita** and **Anupam** (From computer vision group of Dr. Chetan) for their constant support as well as answering our endless doubts. We would have stuck at evaluation had **Aradhya** and **Avinoor** (IIIT Delhi) had not helped us understand the prediction pipeline and answered our calls even after midnight.

At last we would like to appreciate **HPC team of IIT Delhi** to provide us with the V100 GPU to run our experiments and provide us with enough credits so that we could participate in the competition.

TEAM BATMAN RETURNS

Anshul Mittal

Google PhD Fellow, SIT, IIT-Delhi

Research Interests:

1. Extreme Classification
2. Extreme Vision
3. IoT Botnet Detection

Aman Bhardwaj

MS Research Scholar, SIT, IIT-Delhi

Research Interests:

1. Healthcare AI systems for Developing Countries
2. Extreme data imputation

YOLO
YOU ONLY LOOK ONCE

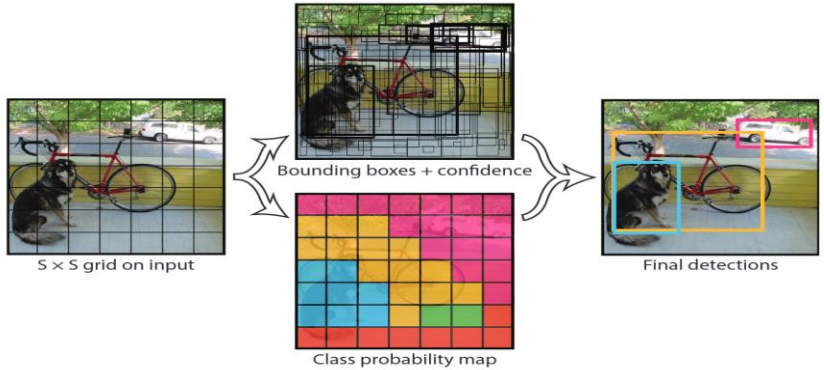
BASELINE METHOD

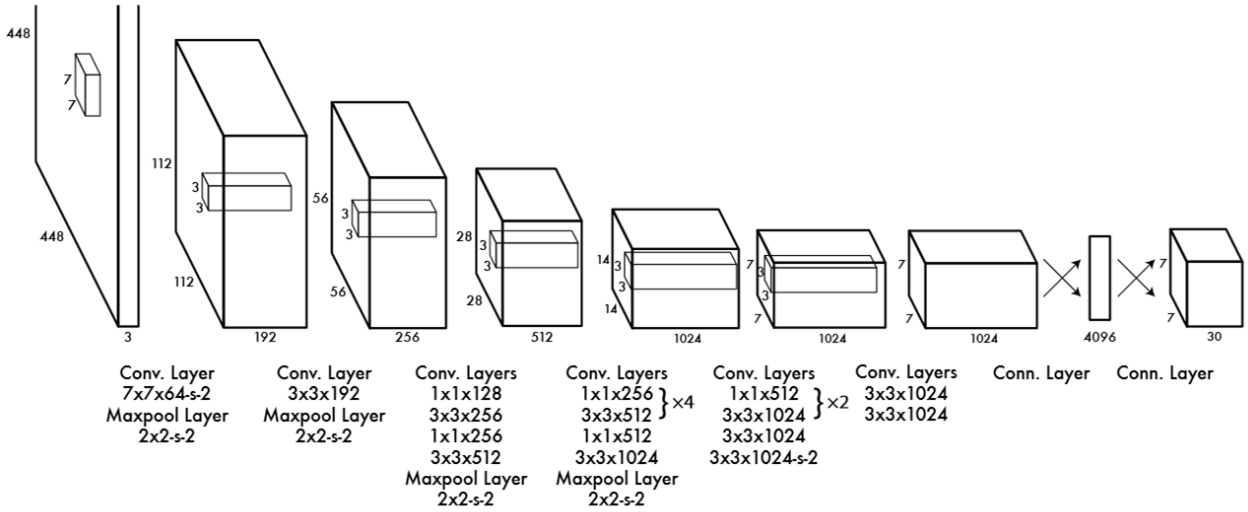
Features:

- Multi-object Detection with R-CNN.
- Bounding Box prediction with confidence level for each object

Drawbacks:

- Each grid can only detect one object. Bad for Small Objects.
- No Batch Normalization. Therefore slow.





COMPUTATIONAL SPECIFICATIONS

Infrastructure Used: NVIDIA Tesla V100 (32Gb) @ IITD HPC Facility

Total time spent on the Challenge: 5 Days

Model Train Details:

- **Model:** YOLOv5x (PyTorch)
- **Total Params:** 89.0M
- **Batch Size:** 32
- **Epochs:** 30
- **Training Time:** 15 Hrs
- **Testing Time:** 25-30 Mins

IMPROVEMENTS OVER BASELINE

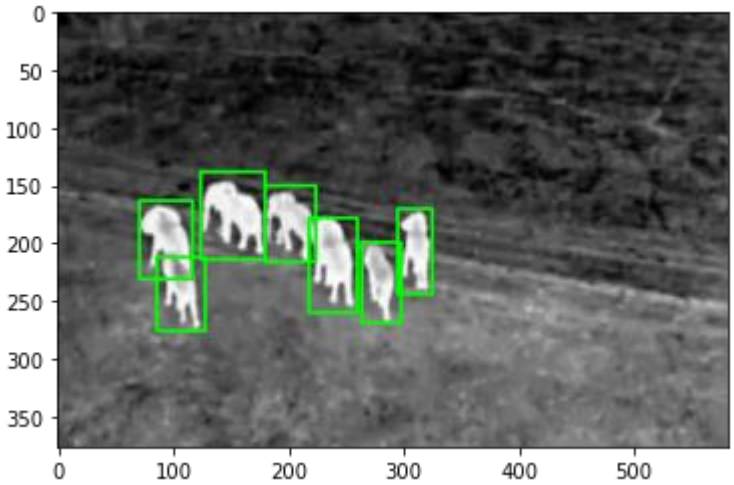
1. **Improved Model:** YOLOv5

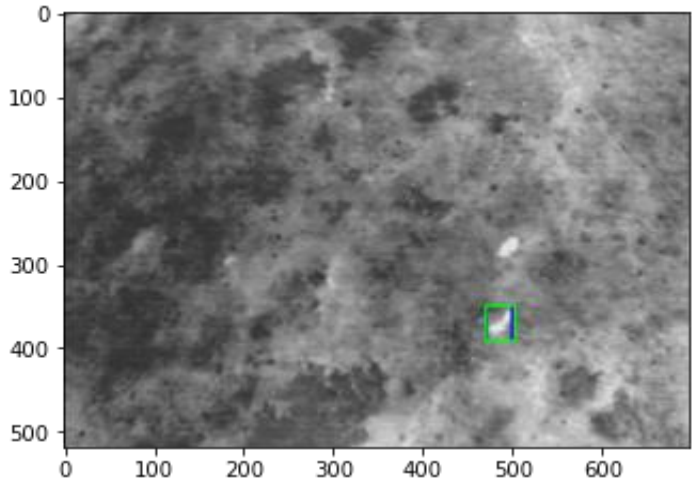
2. **Image Pre-processing:** Contrast stretching, brightness Enhancement

3. **Adaptive Thresholding based on classes: -** Class imbalanced can lead to under-confident classifier for the class with less data. Therefore to adjust for the class imbalance we used adaptive confidence thresholding such that for class (c) if bounding box is predicted with the confidence of prediction 0.5*(Training points for c)/(Total training points) then only it is kept in the prediction

4. **Bounding box buffer adjustments. -** Large objects are often under bounded by the model. Therefore to adjust of this we used weighted padding. In our experiments we found that padding of $0.01*\sqrt{(w*h)}$ leads to better results.

QUALITATIVE RESULTS





Success Case:

- **Model achieved. F1 Score for Animals Class: 0.64**

Failure Case:

- **High False Positives of Human class brought F1 to 0.13**

PS: For Computer Vision A3

WINNERS BIRDSAI TRACK

THANK YOU!