4. Discussion

To the problem of flying object detection, we apply transfer learning with weights learned from our generalized model to our refined model in order to achieve state-of-the-art results in this domain. We argue that our algorithm extracts better feature representations of flying objects than those seen in previous research, furthering the current state of research in this domain. Our refined model achieves a 99.1% mAP50, 98.7% Precision, and 98.8% Recall with 50 fps inference speed on the 3-class data set (drone, plane, and helicopter), surpassing models generated from previous research to a significant extent. Aydin et al. proposed a YOLOv5 instance that achieved 90.40% mAP50, 91.8% Precision, and 87.5% Recall with 31 fps inference speed trained on a data set only containing drones and birds [3]. Rozantsev et al. trained their proposed model on a data set reflective of ours, containing flying objects that occupy small portions of the input image with clustered backgrounds. They achieve an 84.9% AP on a data set containing only UAVs and 86.5% AP on a data set containing only aircraft [18]. Al-Qubaydhi et al. proposed a model utilizing the YOLOv5 framework and achieves an impressive 94.1% mAP50, 94.7% Precision, and 92.5% Recall on a dataset containing only one class of drones. [2]. Even with our exceptional results, a potential limitation of our refined model is that it was trained on a data set with a low amount of distinct environments. To address this potential generalization issue, we suggest utilizing our generalized model weights to transfer learn on a data set with higher frequency of distinct backgrounds.

5. Model Architecture

With the publication of "You Only Look Once: Unified, Real-Time Object Detection" first proposed by Redmon et al. [14] in 2015, one of the most popular object detection algorithms, YOLOv1, was first described as having a "refreshingly simple" approach [21]. At its inception, YOLOv1 could process images at 45 fps, while a variant, fast YOLO, could reach upwards of 155 fps. It also achieved high mAP compared to other object detection algorithms at the time.

The main proposal from YOLO is to frame object detection as a one-pass regression problem. YOLOv1 comprises a single neural network, predicting bounding boxes and associated class probability in a single evaluation. The base model of YOLO works by first dividing the input image into an S x S grid where each grid cell (i,j) predicts B bounding boxes, a confidence score for each box, and C class probabilities. The final output will be a tensor of shape S x S x (B x 5 + C).

5.1. YOLOv1 Overview

YOLOv1 architecture [Figure 6] consists of 24 convolutional layers followed by two fully connected layers. In the paper [14], the authors took the first 20 convolutional layers from the backbone of the network and, with the addition of an average pooling layer and a single fully connected layer, it was pre-trained and validated on the ImageNet 2012 dataset. During inference, the final four layers and 2 FC layers are added to the network; all initialized randomly.

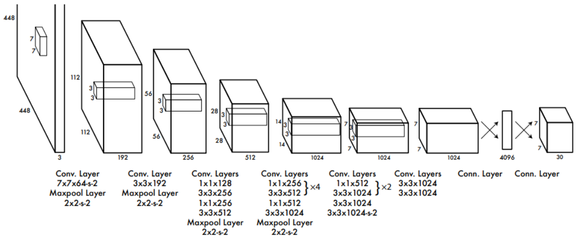

Figure 6: YOLO Architecture [14]

YOLOv1 uses stochastic gradient descent as its optimizer. The loss function, shown by Equation 5, comprises two parts: localization loss and classification loss. The localization loss measures the error between the predicted bounding box coordinates and the ground-truth bounding box. The classification loss measures the error between the predicted class probabilities and the ground truth. The λ_{coord} and λ_{noobj} are regularization coefficients that regulate the magnitude of the different components, emphasizing object localization and de-emphasizing grid cells without objects.

$$\begin{aligned}
&\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}\left[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2\right] \\
&+\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}\left[(\sqrt{w_i}-\sqrt{\hat{w}_i})^2+(\sqrt{h_i}-\sqrt{\hat{h}_i})^2\right] \\
&+\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}(C_i-\hat{C}i)^2 \\
&+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{noobj}(C_i-\hat{C}i)^2 \\
&+\sum_{i=0}^{S^2}\mathbb{1}_{i}^{obj}\sum_{c\in classes}(p_i(c)-\hat{p}_i(c))^2 \qquad (5)
\end{aligned}$$

5.2. YOLOv5 Overview

YOLOv5 [6] is an object detection model introduced in 2020 by Ultralytics, the originators of the original YOLOv1