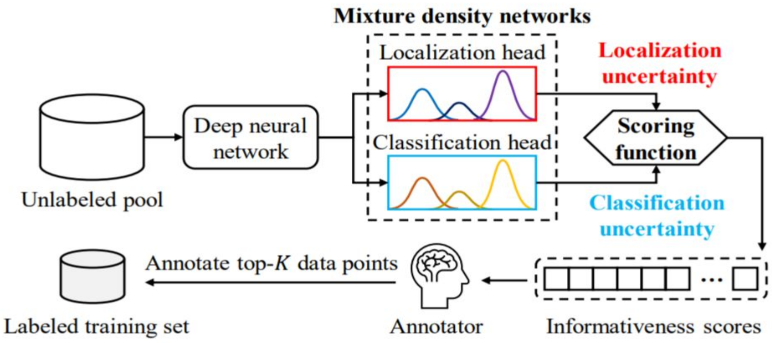
CS776A: Deep Learning For Computer Vision

- **Mid-Project Presentation: Active Learning for Deep Object Detection**

Team : “Brute Force”
Team members : Abhinav Kumar (16907018)
Nishant Kiran Valvi (21111407)

Problem Statement

- The performance of deep neural networks for object detection depends on the size of the labeled data. Creating large amount of labeled data is a very costly task.
- The goal of active learning is to reduce labeling costs by selecting only the most informative samples on a dataset [1].



Active Learning

- Active learning approaches
  - Membership query synthesis: The learner can request to query the label of any unlabeled sample in the input space, including the sample generated by the learner
  - Stream-based selective sampling: Makes an independent judgment on whether each sample in the data stream needs to query the labels of unlabeled samples
  - Pool-based: Finding the best query sample based on the evaluation and ranking
- Query Strategy
  - Uncertainty-based: Sample uncertainty ranking to select the samples to be queried
  - Deep Bayesian Active Learning: Active Learning with Bayesian CNNs
  - Density-based Methods: Finding the subset representing the distribution of the entire dataset

Related Work and Literature Survey

- We narrowed down the following papers:
  - Active Learning for Deep Object Detection via Probabilistic Modeling (2021) [1]
    - This approach uses a single model with a single forward pass and utilizes both localization and classification-based aleatoric and epistemic uncertainties.
    - This approach is based on a mixture density network that learns a Gaussian mixture model (GMM) for each of the network's outputs, i.e., localization and classification.
  - BayesOD: A Bayesian Approach for Uncertainty Estimation in Deep Object Detector (2019) [2]
    - This approach provides a Bayesian treatment for every step of the neural network inference procedure, allowing the incorporation of anchor-level and object-level priors in closed form.
    - It replaces the standard non-maximum suppression (NMS) with Bayesian inference, allowing the detector to retain all predicted information for both the bounding box and the category of a detected object instance

Related Work and Literature Survey

- We narrowed down the following papers:
  - Deep Bayesian Active Learning with Image Data (2017) [3]
    - Active learning (AL) methods generally rely on being able to learn and update models from small amounts of data. Whereas deep neural networks depend on large amounts of data.
    - Many AL acquisition functions rely on model uncertainty, yet deep learning methods rarely represent such model uncertainty. This paper addresses these shortcomings by utilizing Bayesian CNNs.
  - Localization-Aware Active Learning for Object Detection (2018) [4]
    - In this paper they present two metrics for measuring the informativeness of an object hypothesis, which allow them to leverage active learning to reduce the amount of annotated data needed to achieve a target object detection performance.
    - These metrics consider different aspects of object detection in spite that the ground truth of object locations is unknown, making these metrics suited for active learning.

Active Learning for Deep Object Detection with YOLOv3

Following the approach mentioned in [1] we apply active learning on the YOLOv3 algorithm to solve the problem statement.

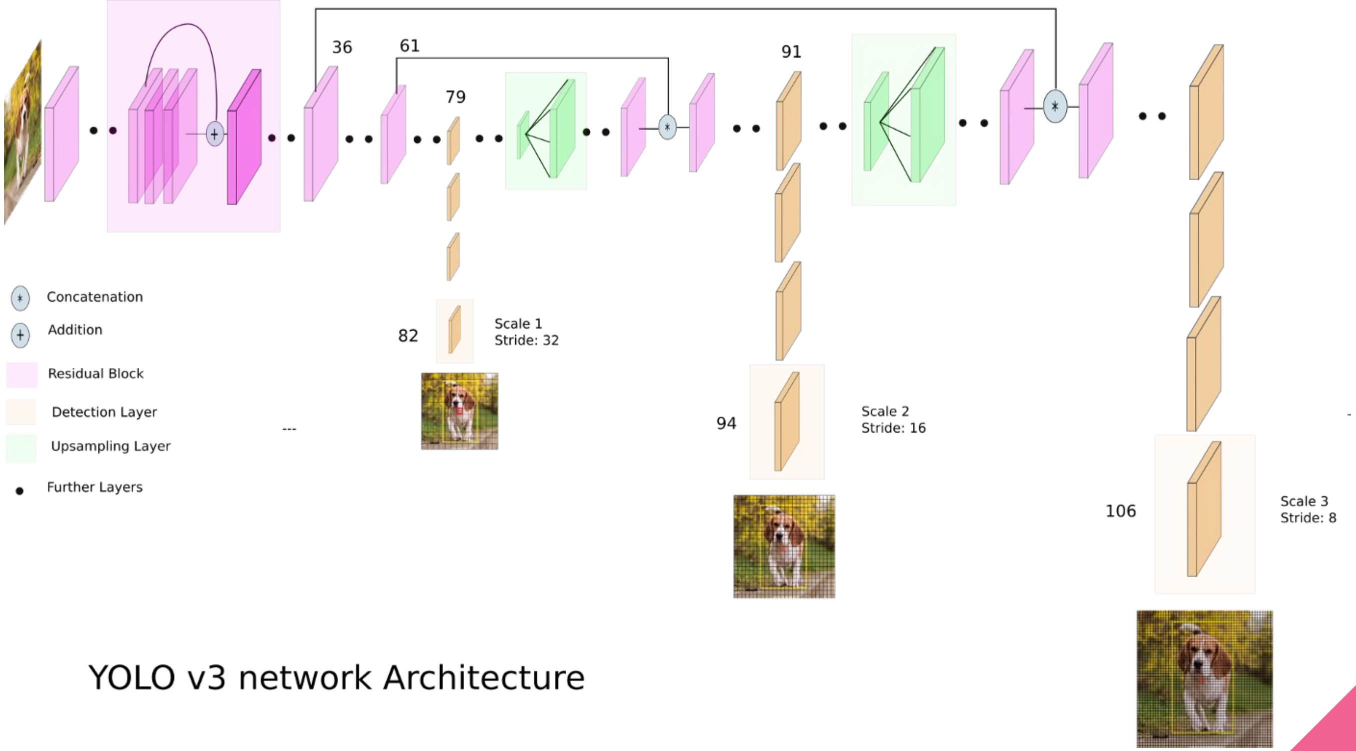
1. Firstly, we take the COCO dataset [8] . We divide it into two parts, one for training and one as an unlabeled pool of data.
2. We then train the YOLOv3 [9] algorithm on *M* data points on Kaggle.
3. We now use the trained model in an online fashion, i.e., it sees the unlabeled data in an incremental way and it selects the most informative *K* data points from it.
4. We use scoring functions to calculate the informativeness of an image by utilizing bounding box uncertainty and class probability to select these *K* data points.
5. Selected images are added to the training set and the model gets re-trained on the new updated training dataset.
6. We compare the performance of our model with that of YOLOv3 trained on full dataset.

Scoring Function

1. Score of an image is calculated by
   Avg(box_confidence * box_class_probs)

2. We ranked based on score and selected *K* images with minimum score for next iteration.

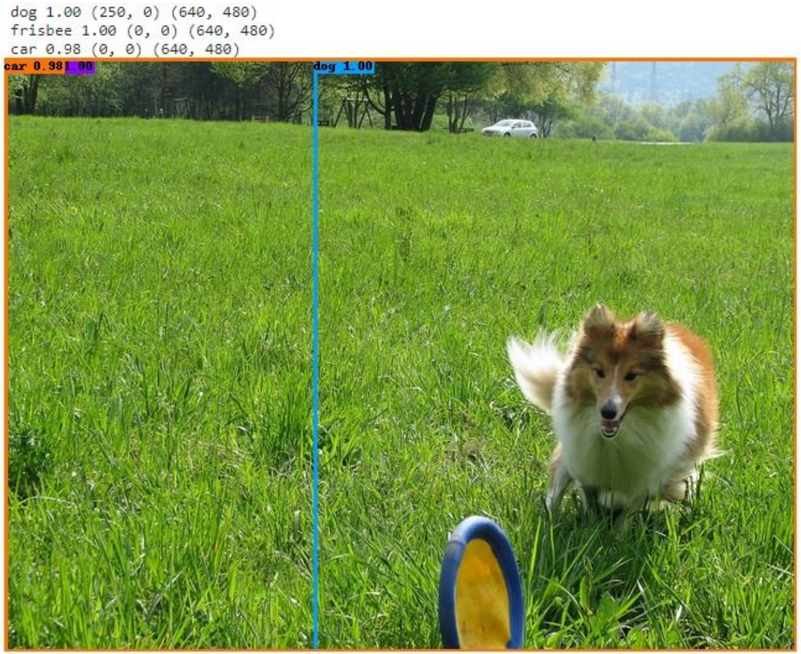
Active Learning for Deep Object Detection with YOLOv3

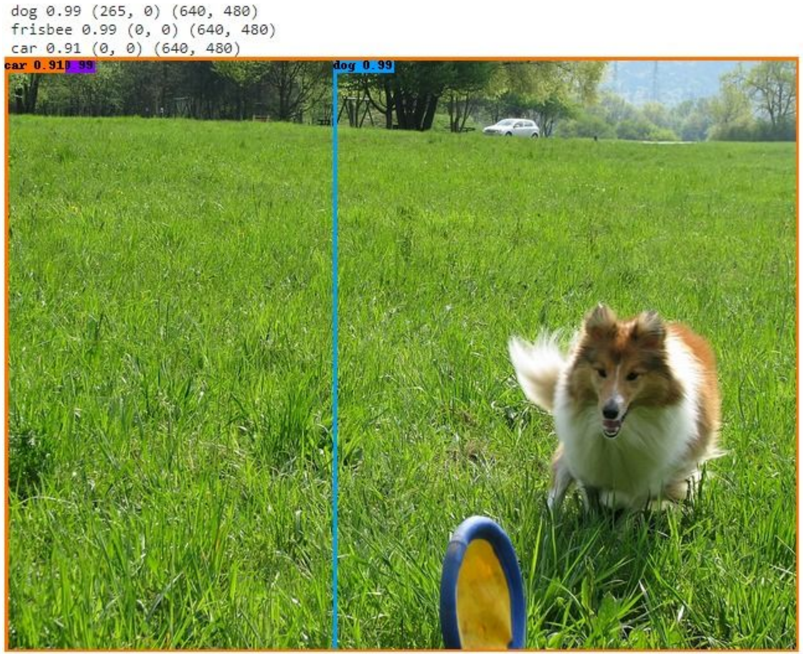


YOLO v3 network Architecture

Src:
https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b

Current Results





Pre trained model

Active learning model

Current Results

Pre trained model

Active learning model

Plan Ahead

- We will try some of the following different modeling techniques for calculating informativeness of an image in the coming weeks:
    a. Modeling bounding box coordinates and class label via probabilistic distributions such as Gaussian Mixture Models by changing the output of YOLOv3 architecture ([1]).
    b. Replacing standard non-maximum suppression (NMS) with Bayesian inference, allowing the detector to retain all predicted information for both the bounding box and the category of a detected object instance ([2]).
    c. Modeling using Bayesian Active Learning i.e., using Acquisition Functions to select data to learn from ([3]).
    d. Modeling using Variational Inference ([5]).
    e. We are looking for more scoring functions to select *K* most informative images.

Issues Faced

- We first ran the code for Active Learning for Deep Object Detection via Probabilistic Modeling, which requires a lot of GPU resources. We reduced batch size so that GPU memory won't get full. Even then the code required 60+ hours to run.

- Running the github codes for these papers is a challenging task due to limited resources and therefore we are writing our own active learning code using YOLOv3.

References

1. Active Learning for Deep Object Detection via Probabilistic Modeling
2. BayesOD: A Bayesian Approach for Uncertainty Estimation in Deep Object Detectors
3. Deep Bayesian Active Learning with Image Data
4. Localization-Aware Active Learning for Object Detection
5. Variational Adversarial Active Learning
6. A Survey of Deep Active Learning
7. A survey on active learning and human-in-the-loop deep learning for medical image analysis
8. Microsoft COCO: Common Objects in Context
9. YOLOv3: An Incremental Improvement

Thank You!