AER 1515H Final Project

Knowledge Distillation for YOLO Object Detection

Presented by:

Matthew Teichman (1002640476), Min Woo (David) Kong (1008691435)

December 12, 2022

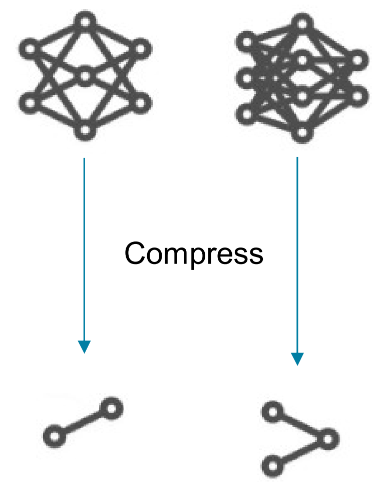
UNIVERSITY OF TORONTO

Introduction

This project is to study and develop a **knowledge distillation** algorithm that compresses a large complex model into a smaller, more efficient model.

Knowledge distillation was originally developed by Geoffrey Hinton at el. in 2015 and was since been adapted for object detection networks.

Our project utilizes **imitation learning** to train a smaller model to effectively imitate a larger model.



UNIVERSITY OF TORONTO

Motivation

1) Edge Computing

2) Mobile Application

3) Autonomous Vehicles

UNIVERSITY OF TORONTO

Knowledge Distillation Experiments

- Our knowledge distillation experiments will perform traffic light detection.

- The dataset that we will be utilizing is DriveU traffic light dataset (DTLD) which contains 230,000 labeled image in 11 different German Cities.

- We will be performing knowledge distillation on Yolov5 developed by Ultralytics, by compress larger Yolov5 models into small Yolov5 models.

UNIVERSITY OF TORONTO

YOLO Datasets Preparation

Dataset consists:

- 24,520 training images
- 800 validation images
- 4,345 testing images

Data Augmentations:

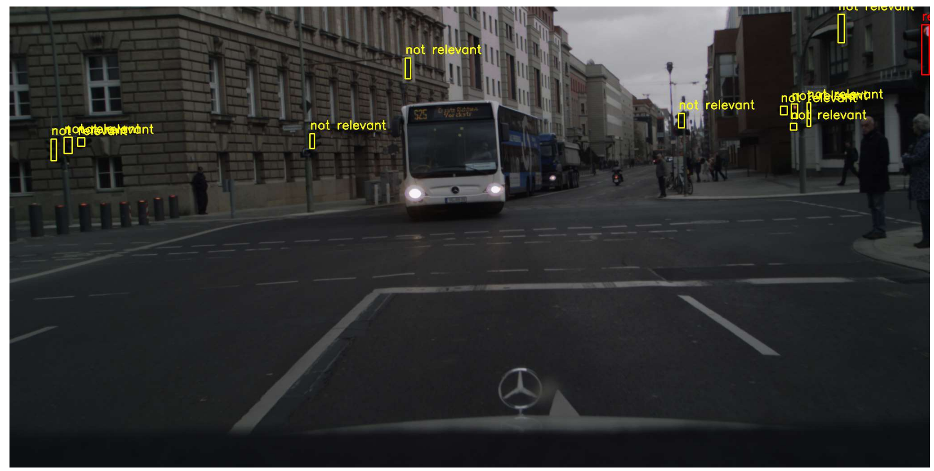
- Image Blur
- Contrast equalization

Image Size:

- Original Image: 1024x2048
- Input Image: 640x640

Labels:

- Relevant (perspective of vehicle)
- Not Relevant



UNIVERSITY OF TORONTO

YOLO Dataset Training - Teacher

- Teacher model is used using YOLO Datasets.
- All are trained to 30 epochs with settings of initial learning rate 0.01 and final learning rate 000.1
- The batch sizes were kept as large as possible.
- Image were size to 640 by 640

Model	FLOPs	mAP50	mAP50-95	Relevant mAP50	Relevant mAP50-95	Not Relevant mAP50	Not Relevant mAP50-95	Epochs
Yolov5n	4.5	36.6	15.55	45.25	18.9	14.92	4.83	30
Yolov5s	16.5	42.85	19.8	66	32.9	19.7	6.77	30
Yolov5m	49	49.95	23.75	73.35	37.8	26.5	9.505	30
Yolov5l	109.1	51.4	25.2	74.65	39.7	28.2	10.67	30
Yolov5x	205.7	53.95	27.0	76.55	41.95	31.3	12.1	30

* All experiments were repeated twice and average out

UNIVERSITY OF TORONTO

Knowledge Distillation (KD) Training - Student

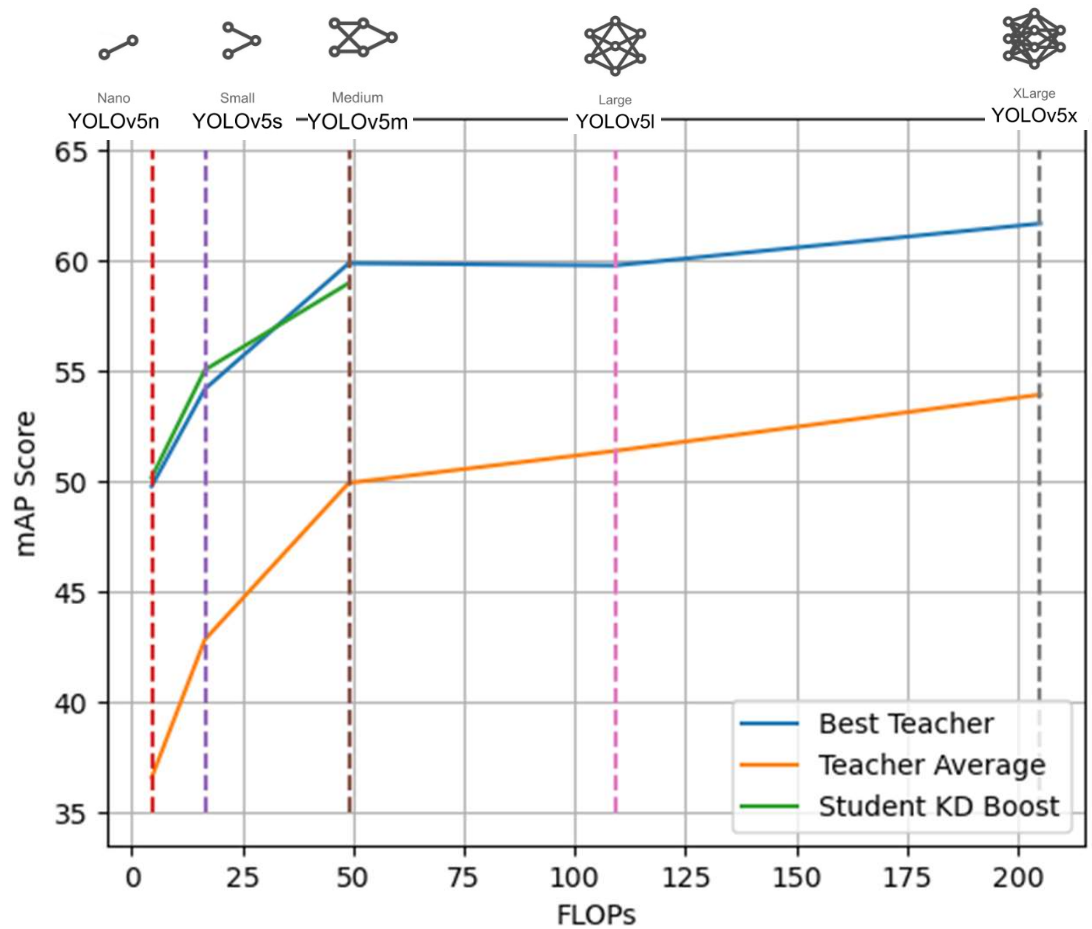
- Student model is also used using YOLO Datasets.
- All are trained to 30 epochs with settings of initial learning rate 0.01 and final learning rate 000.1
- The batch sizes were kept as large as possible.
- Image were size to 640 by 640

Model	Teacher	mAP50	mAP50-95	Relevant mAP50	Relevant mAP50-95	Not Relevant mAP50-95	Not Relevant mAP50-95	Epochs
Yolov5n	Yolov5l	50.9	23.2	77.9	38.3	24.0	8.03	30
Yolov5s	Yolov5l	54.5	26.2	80.0	41.9	29.1	10.5	30
Yolov5m	Yolov5l	59.7	29.7	81.7	46.0	35.4	13.3	30
Yolov5n	Yolov5x	49.6	22.6	75.4	37.5	23.8	7.76	30
Yolov5s	Yolov5x	55.4	26.4	80.9	41.8	30.0	11.1	30
Yolov5m	Yolov5x	58.3	29.6	81.2	45.8	35.4	13.3	30
Yolov5n	Yolov5m	50.1	22.8	75.8	37.8	24.4	7.89	30
Yolov5s	Yolov5m	55.3	26.9	80.0	42.5	30.5	11.3	30

* All experiments were repeated twice and average out

UNIVERSITY OF TORONTO

Results from Experiments:



* FLOPs is floating point operation per second is a measure of computer performance.

UNIVERSITY OF TORONTO

Traffic Light Simulation

Average Inference Time: 285ms

Large

YOLOv5l

89 MB$_{FP16}$
10.1 ms$_{V100}$

Average Inference Time: 72ms

Small

YOLOv5s

14 MB$_{FP16}$
6.4 ms$_{V100}$

UNIVERSITY OF TORONTO

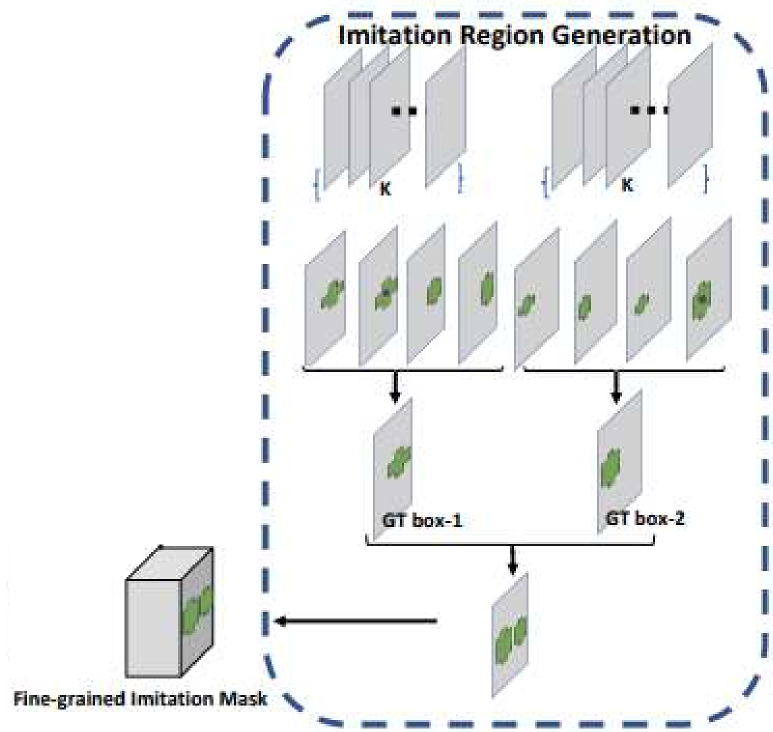
Knowledge Distillation using Imitation Learning

Implemented technique is **Fine-grained Feature Imitation** by paper “Distilling object detectors with fine-grained feature imitation” T. Wang at el.

Knowledge Distillation using Imitation Learning

Implemented technique is **Fine-grained Feature Imitation** by paper “Distilling object detectors with fine-grained feature imitation” T. Wang at el.

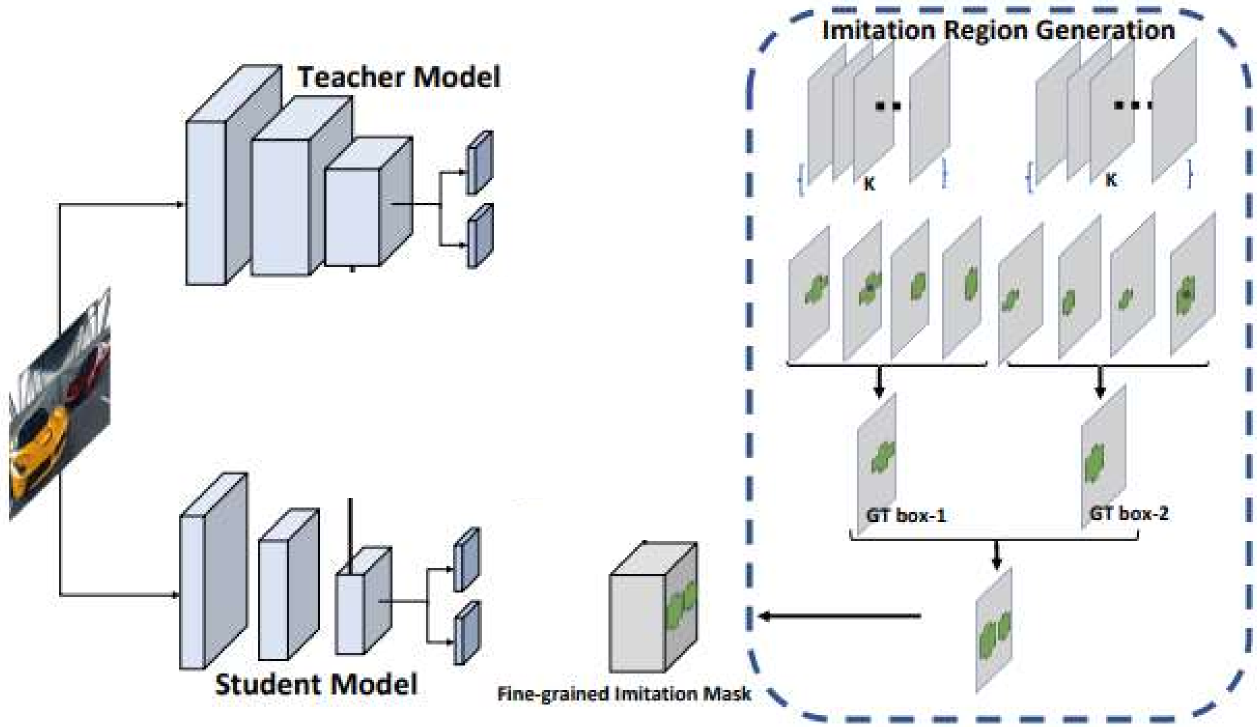
Step 1: Imitation Region Generation



UNIVERSITY OF TORONTO

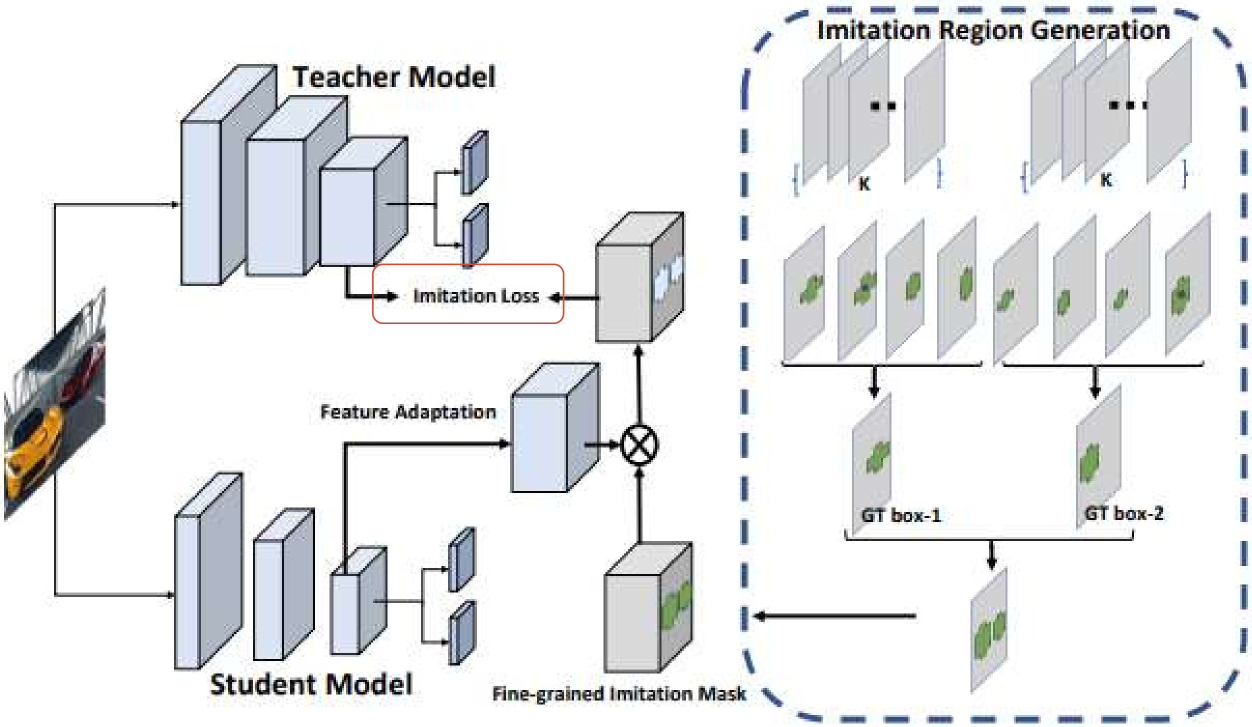
Knowledge Distillation using Imitation Learning

Step 2: Forward Pass through Student and Teacher



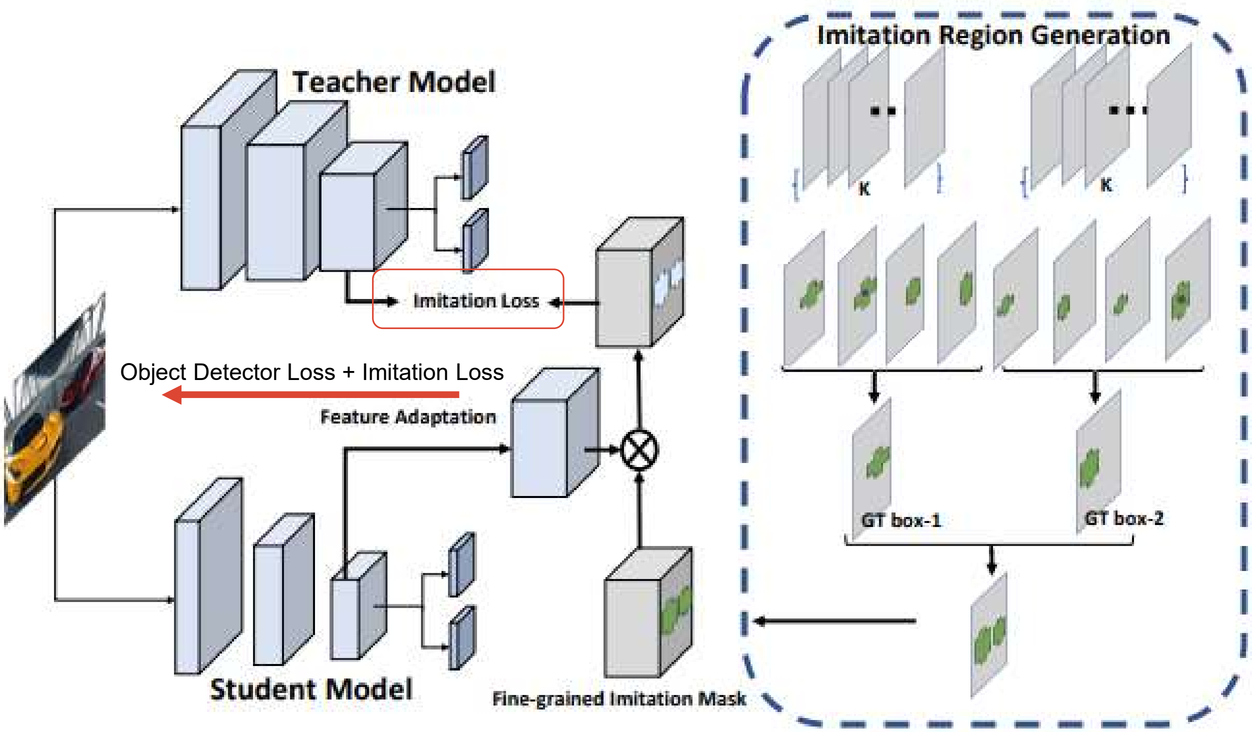
Knowledge Distillation using Imitation Learning

Step 3: Feature Map Extraction to Calculate Imitation Loss



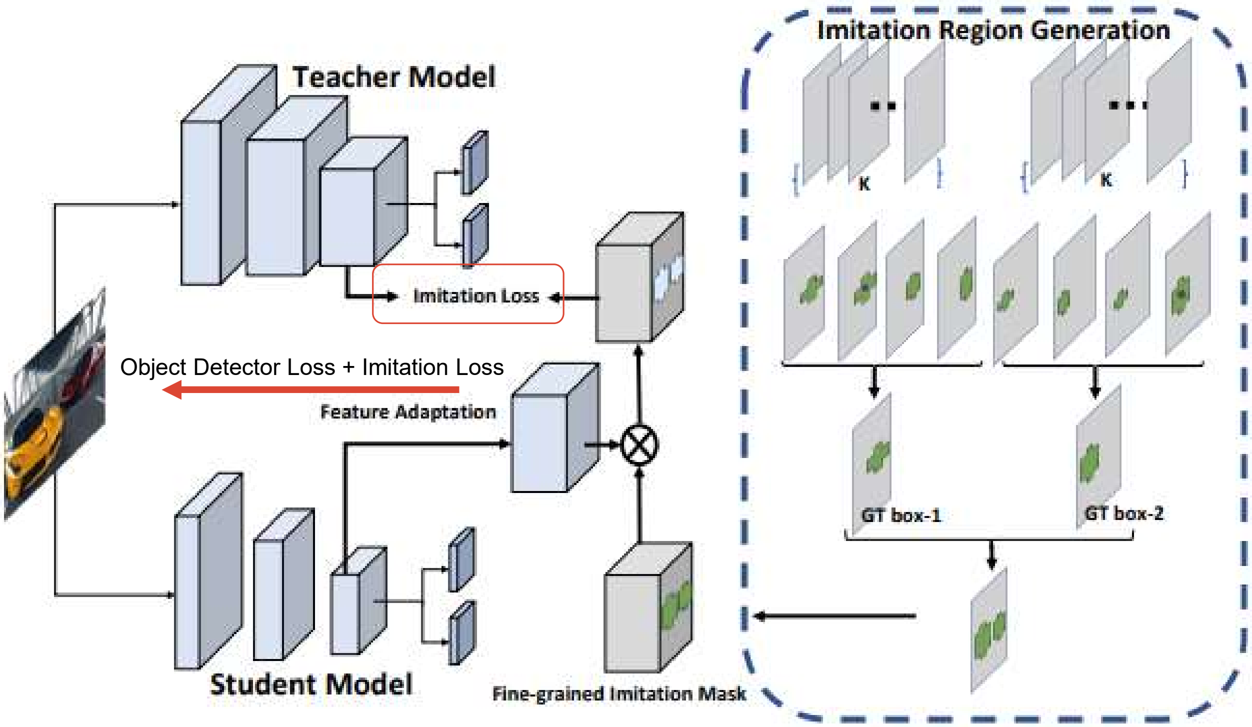
Knowledge Distillation using Imitation Learning

Step 4: Backward Propagation Detector Loss + Imitation Loss



Knowledge Distillation using Imitation Learning

Step 4: Backward Propagation Detector Loss + Imitation Loss



Knowledge Distillation using Imitation Learning

Details of Fine-grained Feature Imitation

- Student model minimizes the following objective function:

$$l = \sum_{c=1}^{C} (f_{\text{adap}}(s)_{ijc} - t_{ijc})^2,$$

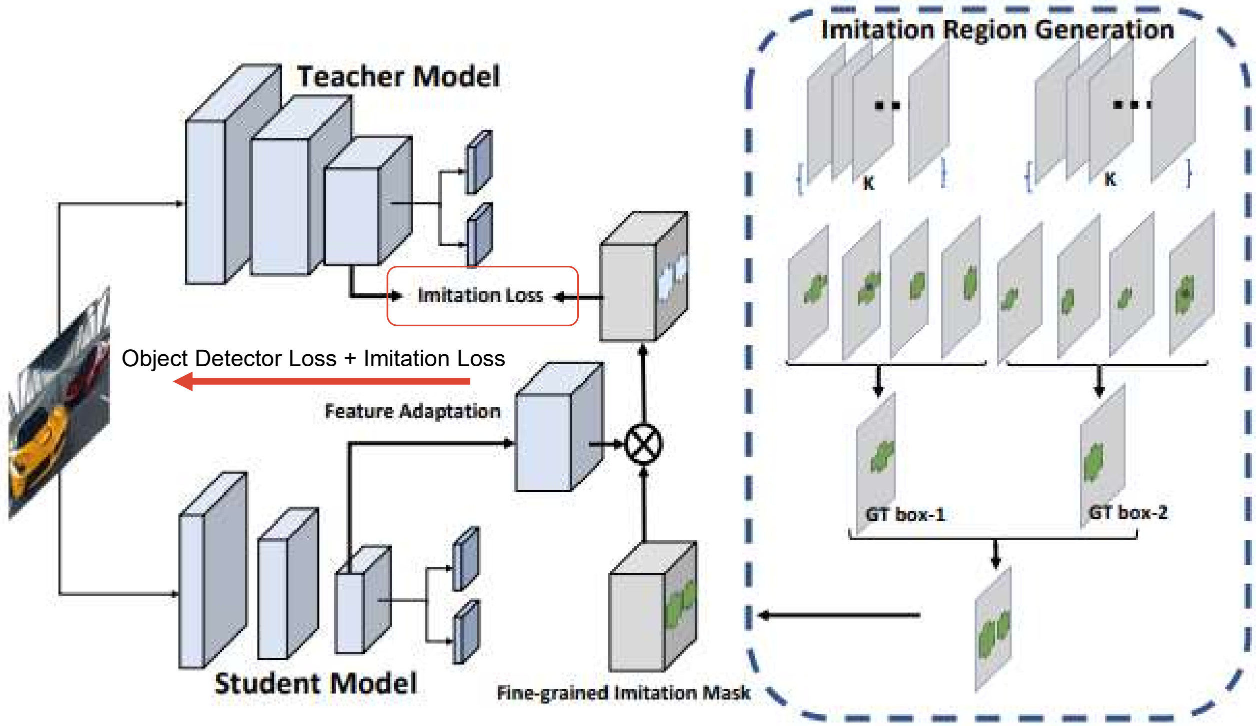
- Limitation Loss:

$$L_{imitation} = \frac{1}{2N_p} \sum_{i=1}^{W} \sum_{j=1}^{H} \sum_{c=1}^{C} I_{ij} (f_{\text{adap}}(s)_{ijc} - t_{ijc})^2,$$

$$\text{where } N_p = \sum_{i=1}^{W} \sum_{j=1}^{H} I_{ij}.$$

- Putting everything together, overall training loss:

$$L = L_{gt} + \lambda L_{imitation},$$



UNIVERSITY OF TORONTO

Conclusions:

1) Repeatability of Results

2) Additional training of teacher

3) Improved generalization of student

4) Teacher must be pretrained

5) Incorporate hint learning

6) Experiment with more advanced neural networks such as Object Detection Transformers

UNIVERSITY OF TORONTO

References:

Papers:

[1] T. Wang, L. Yuan, X. Zhang, and J. Feng, “Distilling object detectors with fine-grained feature imitation,” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[2] L. Yuan, X. Zhang, and J. Feng, “Learning Efficient Object Detection Models with Knowledge Distillation,” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[3] G. Hinton, O. Vinyals, J. Dean, “Distilling the Knowledge in a Neural Network,” 2014 (NIPS) Deep Learning Workshop. 2015

[4] A. Fregin, J. Muller, U. Krebel, and K. Dietmayer, “The DriveU Traffic Light Dataset: Introduction and comparison with existing datasets,” 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018.

Code Repositories Used:

Repository	Link	Description
YOLOV5	https://github.com/ultralytics/yolov5	Unofficial implementation of YOLOv5 used to build and train using distilled knowledge learning.
Knowledge Distillation	https://github.com/wonbeomjang/yolov5-knowledge-distillation	Implementation of knowledge distillation between YOLO models, we planning to use code to develop knowledge distillation training loop.
Knowledge Distillation	https://github.com/tranleanh/mobilenets-ssd-pytorch	Use the code and to incorporated changes to knowledge distillation training loop

UNIVERSITY OF TORONTO