YOLO Based Solution for Congested Traffic Detection

Automatic Dhaka traffic detection on optical images

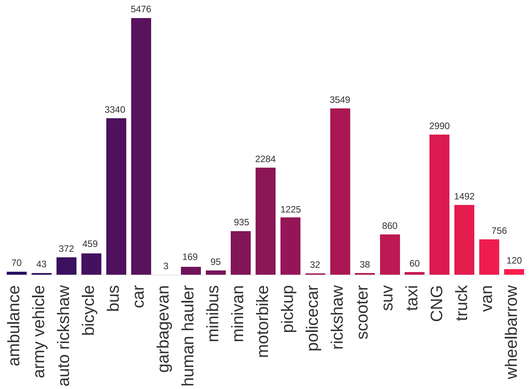
Md Awsafur Rahman, Najibul Haque Sarker, Zaber Ibn Abdul Hakim, Bishmoy Paul

Introduction

The over congested Dhaka Traffic scenario creates complicated challenges for conventional traffic detection systems. We aim to provide a viable solution in regard to these problems. A thorough inspection of the said challenges and the exploration of possible solutions are presented in this poster.

Challenges

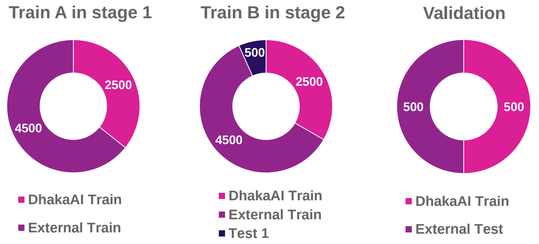
- Mislabeled Images
- Class Imbalance
- Insufficient Samples for Some Classes
- Day Night Imbalance
- Highly Congested Vehicles
- Different Camera Angles
- Variance in Aspect Ratio between train and test Images



Methodology

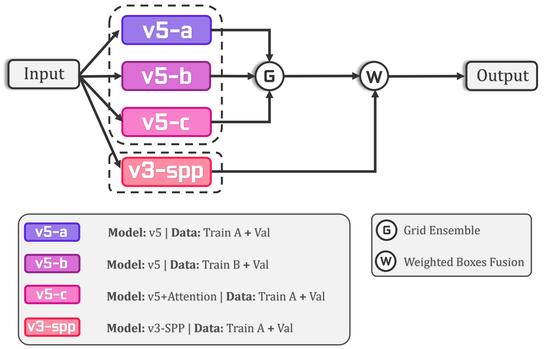
- **Data Preprocessing:** Images containing a large number of wrong labels were removed. Furthermore, there were some duplicate images which were also removed.
- **External Dataset:** Class imbalance, insufficient samples, and day-night imbalance were mitigated by using an external dataset. Furthermore, new unseen environmental cases (rain, fog etc) were introduced which made the model more robust.
- **Attention Infused YOLO Neck:** Due to high traffic congestion, there were overlapping vehicles in many images in which conventional models were performing poorly. Providing contextual information of adjacent pixels to the model should help in discerning and identifying a complete vehicle in these cases. Thus, attention modules were used to make a customized version of the YOLO neck.
- **Multiscale and Augmentation:** The provided train dataset contained images of different scales. For that reason, multiscale training was introduced with the intuition of making the model adept at handling variation in scales. Augmentations like translation, flipping, rotation, scaling, and shearing were used to make the same image look different every time the model encountered it. Mosaic and Mixup were used to infuse different scenarios in a single image which made the model robust in handling unseen cases.
- **Stretched Training:** All the test images were resized to 1024x1024 without maintaining the aspect ratio. Vanilla YOLO data pipeline resizes the images maintaining aspect ratio which creates a huge difference between train and test distribution. To overcome this situation, train images were resized without maintaining aspect ratio which resulted in better accuracy.

Data Pipeline



Modeling

- **YOLOv5:** Gave higher accuracy and FPS compared to other larger models. Thus, it was chosen as the core model.
- **YOLOv5(with Attention Neck):** Neck was infused with attention blocks for more contextual information during feature extraction.
- **YOLOv3:** Gave comparable accuracy to v5. It was able to perform well for some cases where v5 had low accuracy.
- **Ensemble:** Combined predictions of multiple models to improve overall score. Grid ensemble merged grid predictions without the risk of increase in false positives.



Comparison

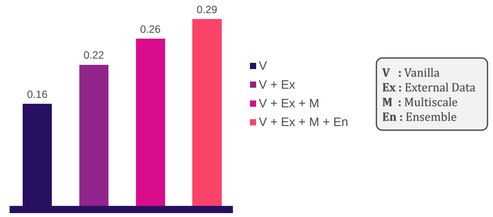


Figure 1: Cumulative Score of Different Methods

Result

Figure 1 encompasses our results succinctly. The Vanilla YOLOv5 was able to get an accuracy of **0.16** which was increased to **0.22** when the external dataset was included. Moreover, the introduction of multiscale training resulted in further improvement which brought the cumulative score to **0.26**. Analogous improvement was also noticed for YOLOv3-SPP. Finally, ensembling all the models boosted the overall result to **0.29**.

Conclusion

A YOLO based solution is proposed to efficiently detect vehicles in Dhaka using YOLOv5 and YOLOv3 models. Multiscale training, External Data, and Attention Block were also incorporated to enhance models' performance further. Experimental results obtained from extensive simulations suggest that it can a very effective choice for vehicle detection in Dhaka even in a highly congested scenario. Moreover, low inference time (32FPS) makes it very suitable for Real-Time applications too.