Introduction
Monocular 3D object detection predicts 3D bounding boxes using a single monocular, typically RGB image. Identifying 3D bounding boxes is a difficult task as RGB images lack critical depth channel information.

Motivation
• Alternatives systems include LiDAR sensors that are expensive and sensitive to adverse weather.
• 3D bounding box regression can help in more accurate traffic study - analysis, data archiving, speed estimation, reconstruction, etc.

Contributions
• System to regress 3D bounding boxes of vehicles and pedestrians on the road in real-time.

Methodology
The system is divided into 2 main parts: 2D Object Detection & 3D Detector.

2D Object Detection
• We have used a single-stage feed-forward architecture of YOLO for the 2D object detection module.
• Features - skip connections, no pooling, and 3 prediction heads (processing at different spatial compressions).

3D Object Detection
• Architecture of Multi-Bin estimation takes in the cropped image patches (resized to 224x224) and outputs - orientation (local yaw) and Dimension of objects.

Experiments and Results

| Dataset | KITTI 3D object detection dataset consists of 7481 training images and 7518 test images with 3D annotations.
| Metric important for use-case | Inference time
| Evaluation metric | Intersection over Union (IoU) and Average Orientation Similarity (AOS)
| Per pose inference time PPIT (Titan-X) | • PPIT KITTI – 0.112 s
| • PPIT ARGOS – 0.098 s

Conclusion & Future Work
• We present a deep learning architecture that regresses 3D bounding boxes of objects from monocular RGB images.
• The key idea is that the perspective transformation of a 3D bounding box should fit tightly in the 2D bounding box.

Work in progress and future work
• Transforming UA-DETRAC 2D object detection dataset into a 3D detection dataset. Testing the system with IoU and AOS.
• Training the network on the dataset and testing inference time on the ARGOS Vision camera.
• Swap the Yolo 2D detector with Tiny-Yolo, SqueezeDet.
• Swap the 2D/3D constraint flow with a deep neural network for location estimation of the 3D box' central point.

References
• Arsalan Mousavian et al. “3d bounding box estimation using deep learning and geometry”