Cross-Camera Module Training of Raw Sensor Data-Based Automotive Machine Vision: Challenges and Solutions

Supplementary Material

7. Advantages of RSDB Object Detection

As mentioned in Section 3.1, we conducted a preliminary experiment to confirm the advantages of RSDB object detection in the context of automotive machine vision. We first applied ISP in [19] to yield full color images (RGB), then trained Faster RCNN object detection model [38] to detect cars in them. We repeated the training by changing the input layer of the Faster RCNN to accept the raw sensor data instead, with otherwise identical network architecture. The result of detecting 534 car samples in the test set using an intersection over union (IOU) threshold of 0.5 is reported in Table 1. Based on the average precision, the raw data and demosaicked images yield comparable *average* precision performance. However, the precision of the object detection at high recall rates (the preferred metric for automotive applications; see Section 2.5 above) has a meaningfully improved when trained to accept raw sensor data at the input layer.

Consider the ability for the object detection neural networks to detect cars that are far away. The fish-eye/wide-angle lenses prevalent in automotive camera modules have the effect of stretching out the scene. Combined with the low pixel density (large pixel pitch, low image resolution) of automotive image sensors, the target objects occupy far fewer pixels compared to the consumer-focused image sensors (high resolution, longer focal lengths). This is particularly problematic for detecting cars that are at a distance, which is a common (and important) scenario we encounter in real-world automotive applications.

Figure 5 shows the IOU of the predicted and target boxes as a function of the pixel size of the target boxes (a proxy for the *proximity* to the target cars). Interpreting IOU as the quality of object detection, this experiment suggests that the RSDB inference trained and performed on raw images exhibit a higher IOU prediction compared to that of the same network architecture trained on demosaicked images. The advantage of RSDB processing is especially noticeable for smallest targets, when the detection performed on RGB images becomes less reliable. Hence raw sensor-based machine vision is attractive for the automotive camera modules with relatively low resolution sensors with fish-eye or wide-angle lenses.

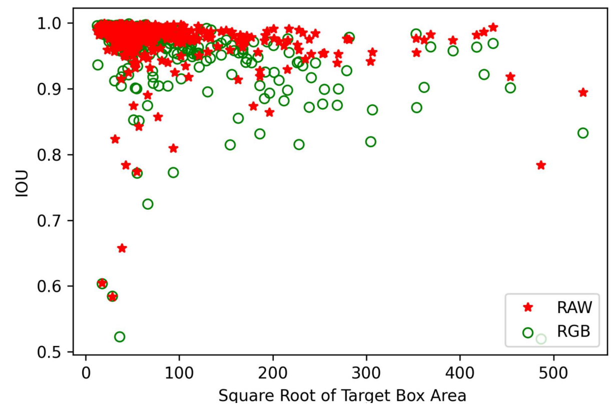
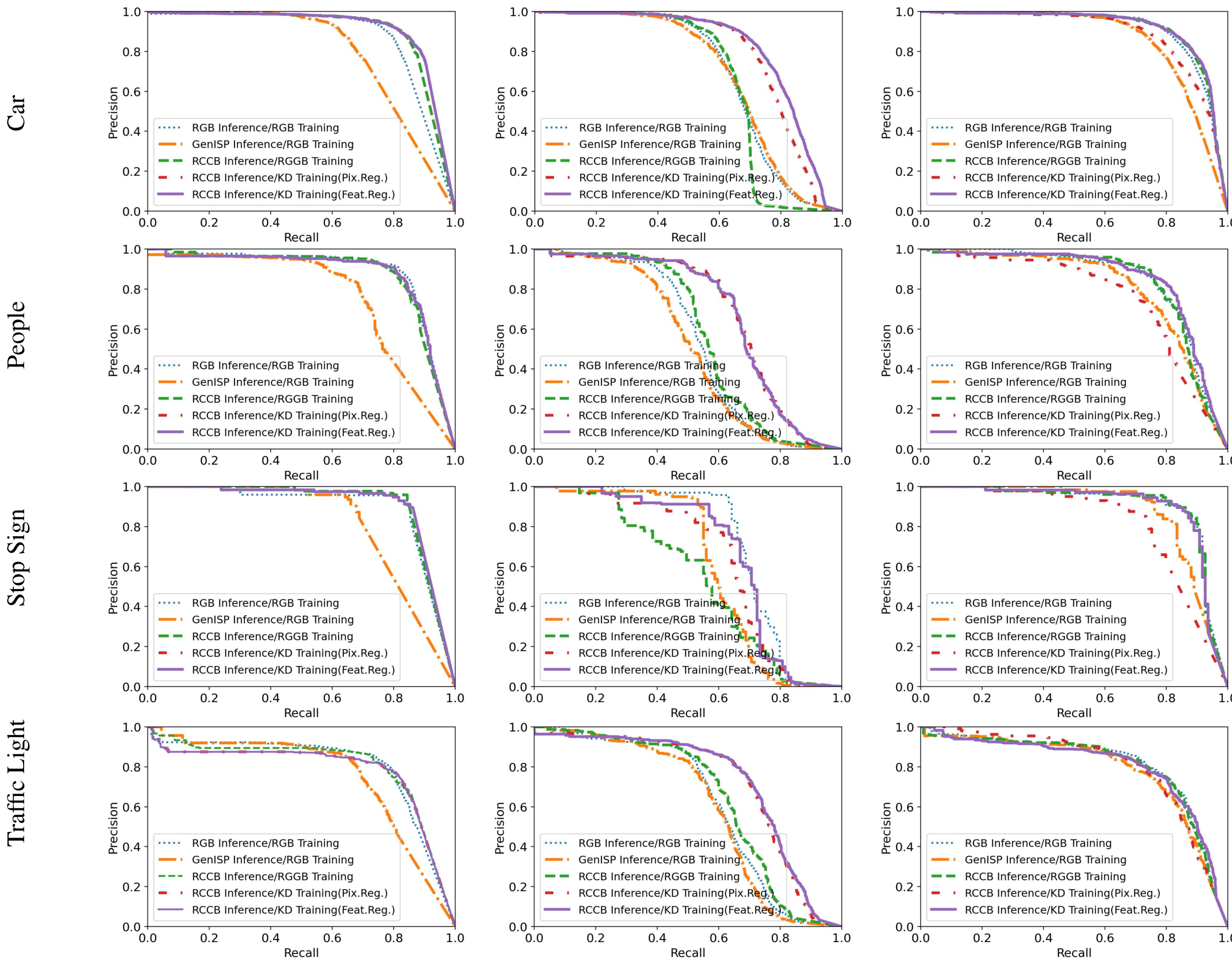


Figure 5. Prediction comparison of cars by identical neural network models trained on raw and post-ISP color (RGB) images.

8. Additional Figures for Benchmarking

Figure 6 shows the precision of the object detection models plotted as a function of recall values, based on which the statistics in Table 2 were derived. As it can be seen in Figure 6 YOLO's precision drops quickly at high recall rates. The very same trend was seen in RGGB teacher network (not shown) , meaning that it is a limitation of YOLO and is not related to cross-camera module training. Overall the proposed knowledge distillation technique with feature registration has the most consistent performance trend over the benchmark (GenISP), the baseline (ablation: RGB Input/RGB Trained), the teacher network (ablation: RCCB input/RGGB Trained), and knowledge distillation with pixel registration, especially in high recall operating point.



(a) Faster RCNN (b) YOLOv5 (c) RetinaNet

Figure 6. Performance of the object detection models on Car/People/Stop Sign/Traffic Light targets in the RCCB camera module dataset (Evaluation Set).