Intermediate Result

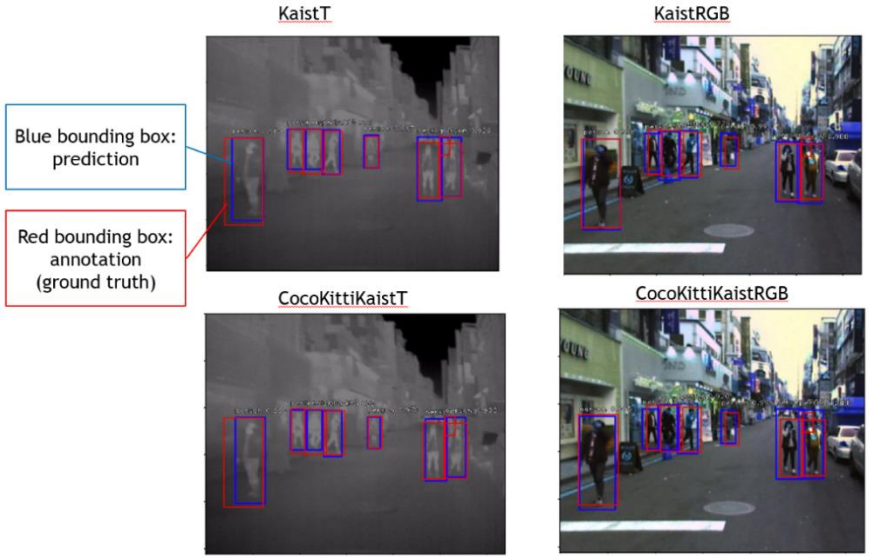


Figure 12 an example of bounding boxes (in red) predicted by four different models (kaistT, kaistRGB, cocokittikaistT, cocokittikaistRGB)

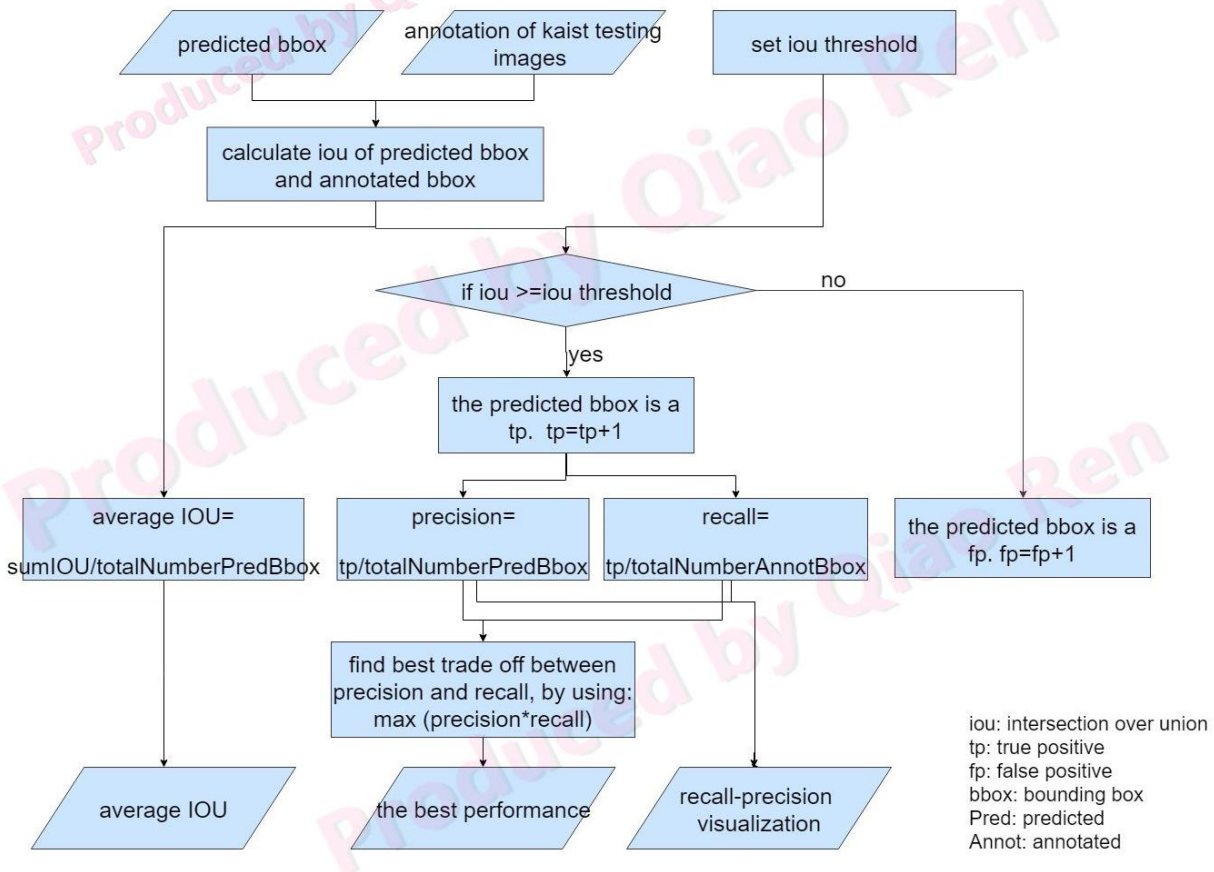


Figure 11 flow chart of how to evaluate a model by calculating average IOU, precision and recall

Figure 13 shows the classification loss and regression loss of each model. Classification loss measures whether the objects are classified correctly or not. Regression loss measures whether the coordinates of predicted bounding boxes are correct or not. Classification and regression losses are recorded during the training process. There are 50 epoches in the training process. From the images, all four models have been converged.

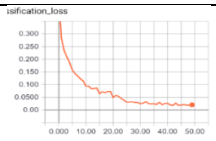
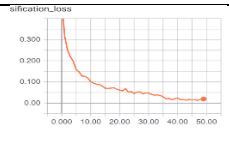
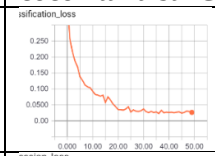
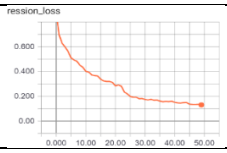
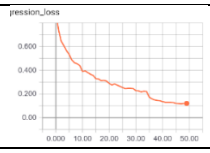
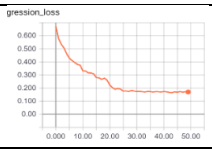
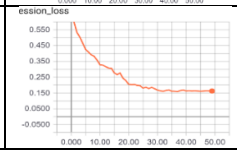
	KaistT	KaistRGB	cocokittikaistT	cocokittikaistRGB
Classification loss				
Regression loss				

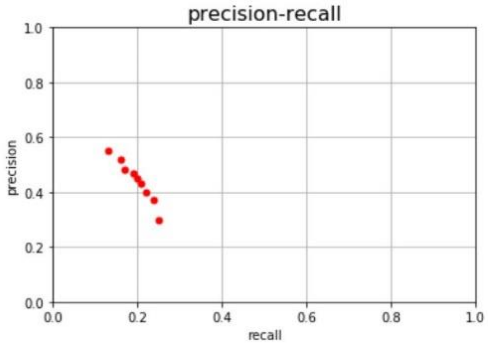
Figure 13 the classification loss and regression loss of four models (kaistT, kaistRGB, cocokittikaistT, cocokittikaistRGB)

Figure 14 shows the recall-precision graph of the four models respectively. Each graph contains 9 dots, corresponding to 9 different score threshold: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. Figure 14 shows the trade-off of precision and recall of four models respectively. The trade-off is calculated by the formula 2.

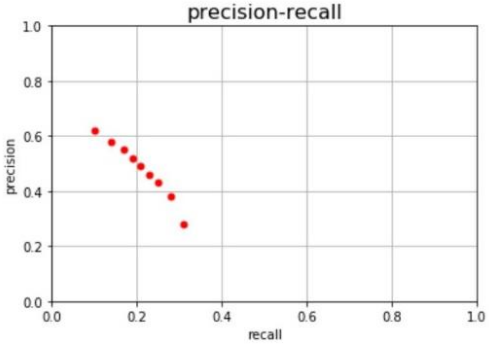
$$Rectangle\ Area = precision * recall$$

Formula 2 a formula used for evaluating the tradeoff between precision and recall

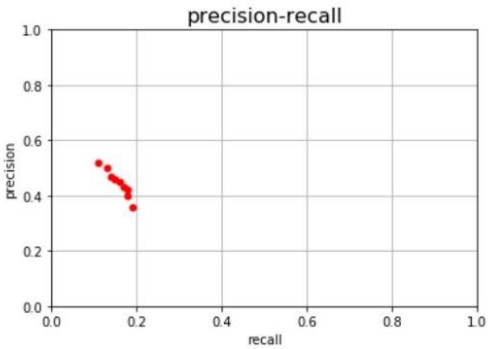
a)kaistT



b)cocokittikaistT



c)kaist RGB



d)cocokittikaist RGB

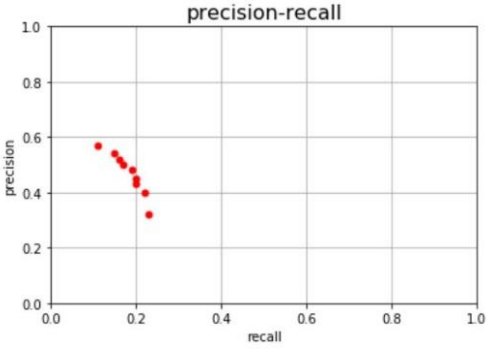
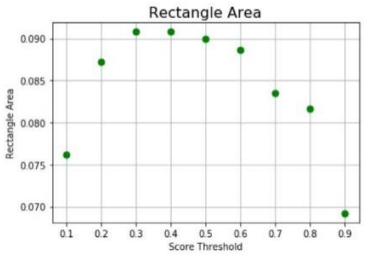
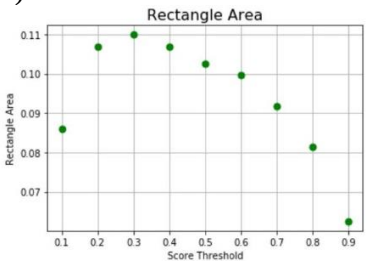


Figure 14 precision-recall graph of four models respectively (Each graph contains 9 dots, corresponding to 9 different score threshold: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)

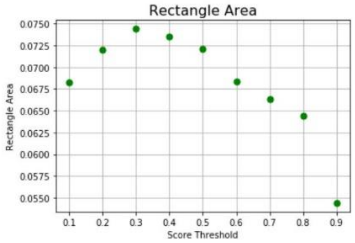
a)kaistT



b)cocokittikaistT



c)kaist RGB



d)cocokittikaist RGB

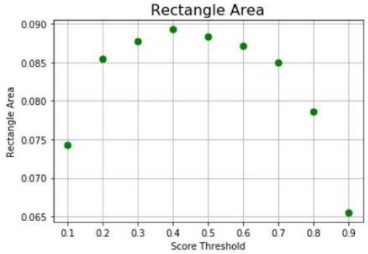


Figure 15 trade-off of precision and recall of four models respectively (Each graph contains 9 dots, corresponding to 9 different score threshold: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)

Figure 16 visualizes a comparison among the recall-precision graphs of the four models. The actual output are dots. The line that connects the dots are drew manually, in order to enhance the visualization. The big dot in each line shows the best performance with the maximum rectangle area of each model. Figure 17 visualize the comparison of average IOU of all four models. Each score threshold has a corresponding average IOU.

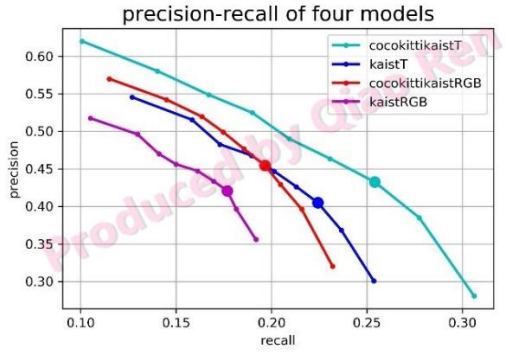


Figure 16 Precision-recall of four models

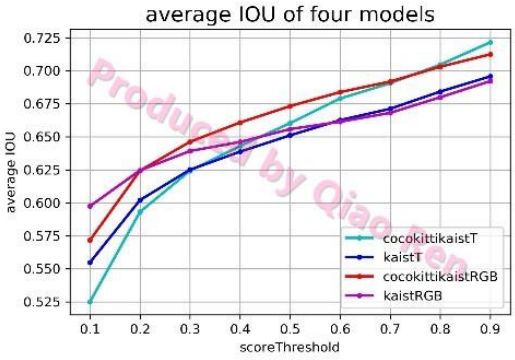


Figure 17 Average IOU of four models

The best performance of each model

Name of a model	Iou threshold	Score threshold	Recall * Precision	Average IOU
kaistT	0.7	0.3	0.0908	0.6251
cocokittikaistT	0.7	0.3	0.1099	0.6243
Kaist RGB	0.7	0.3	0.0744	0.6392
Cocokittikaist RGB	0.7	0.4	0.0893	0.6608

Table 1 The best recall-precision performance of each single-channel model (by calculating the max area of rectangle: precision*recall)

The multiplication result of recall and precision indicates the best model for each dataset. Therefore, it is concluded that the best model of dataset kaistT is model cocokittikaistT (a finetuned model). Because recall*precision of cocokittikaistT (0.1099) is larger than the value of kaistT (0.0908). The best model of the dataset kaistRGB is the model cocokittikaistRGB (a finetuned model). Because recall*precision of cocokittikaistRGB (0.0893) is larger than the value of kaistRGB (0.0744).