

# A Coarse to Fine Detection Method for Prohibited Object in X-ray Images Based on Progressive Transformer Decoder

Anonymous Authors

## 1 DETAILED DESCRIPTION OF SOME PARTS IN EXPERIMENTAL RESULTS AND ANALYSIS

### 1.1 Robustness Analysis

To validate the robustness of the method in this paper, the training loss curve was compared with that of DAB DETR. Figure 6 shows the comparison of the training loss curves of the method in this paper and DAB DETR, where the first and second lines represent regression loss, and the third and fourth lines represent classification loss. (a) represents the overall regression loss, (b)-(f) represent the regression loss for each category. (g) represents the overall classification loss, and (h)-(l) represent the classification loss for each category. The red line represents the method of this paper, and the blue line represents DAB DETR.

It can be seen that whether it is regression loss or classification loss, the method in this paper can converge rapidly and maintain a stable training state. For regression loss, although DAB DETR can also converge quickly, the oscillation range is large in the later training process, and the robustness is poor. As for the classification loss, the advantages are more obvious, not only the convergence is rapid, but also the training process is more stable in the later stage, especially for each class of object. On the contrary, DAB DETR maintains a large oscillation amplitude while converging slowly, which fully demonstrates the excellent learning and generalization ability of our method.

### 1.2 Effect of Different CTFF Parameters on Model Performance

Different parameters are used during the coarse detection and fine detection stages. To verify the impact of these parameters on model performance, we conducted more detailed ablation experiments. The experimental results are described below.

**1.2.1 Effect of Different  $\alpha$  in Coarse Detection Stage.** In order to reduce the computational complexity as much as possible and ensure higher detection accuracy, in the coarse detection stage, this paper uses the reduction coefficient  $\alpha$  to reduce the size of the feature map. The larger the value of  $\alpha$ , the less information is lost. In this paper, we tested three values, namely 1/3, 1/2, and 2/3. See Table 7 for the results.

It can be seen from the table that when  $\alpha = 1/3$ , about 61% of the objects can be correctly detected only through the coarse detection stage. As the reduction coefficient increases, the detection accuracy gradually increases as the features become richer. When  $\alpha = 1/2$ , the number of objects detected in the coarse detection stage tends to be stable, and the rest of the complex objects should be sent to the next stage for fine detection.

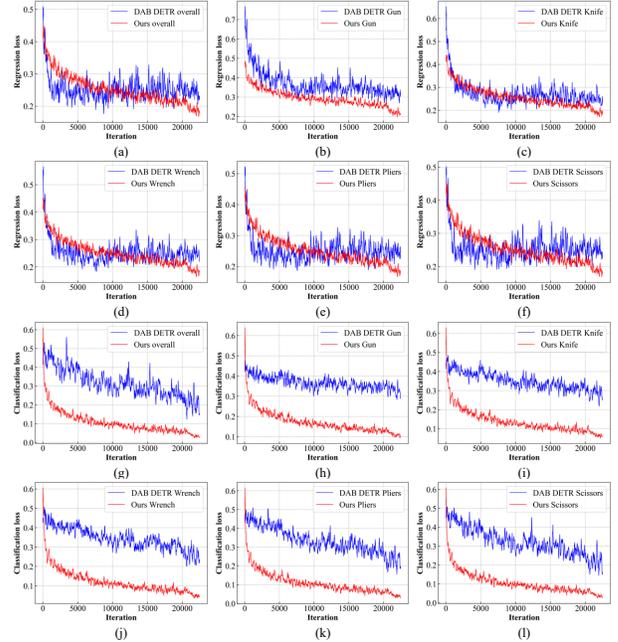


Figure 6: Comparison of training loss curves of different methods.

Table 7: Effect of Different  $\alpha$  Values on Model Performance.  $\alpha$  denotes reduction coefficient, N.cd denotes the number that passes the coarse detection.

$\alpha$	N.cd	mAP	Inference Time(s)	GFLOPs
1/3	61%	90.31	0.0414	152.11
1/2	79%	92.39	0.0545	173.85
2/3	81%	92.41	0.0735	195.88

**1.2.2 The Number of Large Regions Selected During the Fine Detection Stage.** During the fine detection stage, more feature information is utilized for the detection of complex prohibited objects. Thus, the greater the number of large areas selected, the higher the detection accuracy, but at the cost of increased complexity. To achieve a balance between complexity and accuracy, we conducted ablation experiments on the number of large areas selected, denoted as  $N$ . The results of these experiments are presented in Table 8.

It can be seen from the table that as the number of large areas selected increases, the detection accuracy also increases. When  $N$  is greater than 4, the detection accuracy tends to be flat, while the computational complexity continues to increase, which shows that for X-ray images, choosing 4 large regions can achieve accurate detection of prohibited object. Therefore, in order to achieve a

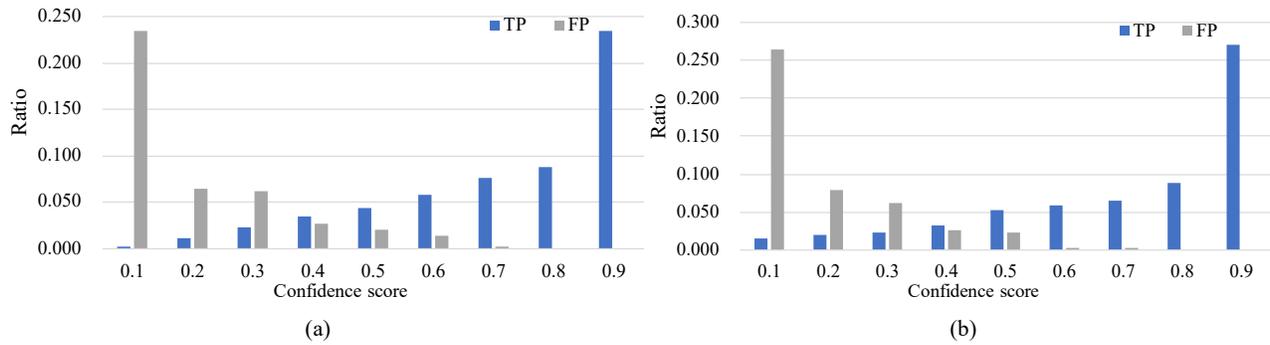


Figure 7: The distribution of TP and FP at different confidence score.

Table 8: Effect of Different Number of Large Regions on Model Performance.  $N$  denotes the number of large areas selected.

$N$	mAP	Inference Time(s)	GFLOPs
2	86.47	0.0418	155.12
3	90.90	0.0489	162.98
4	92.39	0.0545	173.85
5	92.43	0.0723	190.21
6	92.48	0.0883	201.11
7	92.48	0.1001	224.98
8	92.49	0.1318	266.10
ALL	92.49	0.1552	298.32

compromise between precision and complexity,  $N = 4$  is set in this paper.

### 1.3 Effect of different PTD parameters on model performance

To determine the appropriate thresholds, when not using PTD, the ratio of True Positives (TP) and False Positives (FP) to all predicted boxes is calculated based on the output from the fourth Transformer Decoder, as shown in Figure 7(a). It is evident that nearly all the output bounding boxes are TP when the confidence score is greater than 0.7. This implies that there is no need to continue refining the bounding box through subsequent processes when the detector's confidence score exceeds 0.7. Hence, the first threshold is set at 0.7.

After introducing a shunt mechanism in the fourth Transformer Decoder, the ratio of TP and FP to all predicted boxes is recalculated based on the output from the fifth Transformer Decoder, as depicted in Figure 7(b). For the fifth Transformer Decoder, it can be observed that nearly all the detection boxes are TP when the confidence score exceeds 0.6. Therefore, in order to protect the high-score queries from the low-score queries, the second threshold value is set to 0.6.