The Second Place for X-Ray Horizontal Object Detection（IJCAI 2023）

Team: STAR

Reporter：Fuxing Leng、Chenglong Yi

1. Competition Analysis

This track is X-ray horizontal Object Detection, which needs to detect 9 types of prohibited items. The preliminary round provides 2000 training samples of horizontal bounding boxes, and the evaluation index uses AP50. In the first and second stages of the preliminary round, 500 images are provided for testing, with score weights 0.5, 757 images are provided for testing in the semi-finals. The competition task only provide a small number of training samples. At the same time, X-ray prohibited items have problems such as category imbalance, small objects, large pose, large scales, and occlusion. The above problems need to be solved in the detection algorithm.

2. Approach

The baseline used CBNet v2. We applied Swin Transformer-Large as backbone, the pipeline as shown in Figure 1.

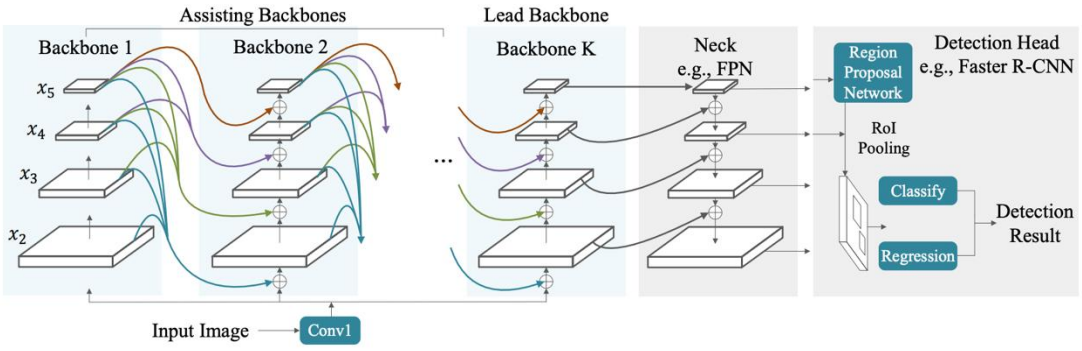

Figure 1. CBNet v2

tricks：

1. Multi-scale training: solving multi-scale problems with targets, training scales (1600, 400-1400)
2. Class balanced sampling(CBS): Solve the problem of unbalanced prohibited items, oversampling factor 0.1
3. Randomly rotate 90 degrees(randomRot90): Data samples are enhanced to improve the detection effect;
4. Autoaugment: Searched from COCO, deals with lighting, large poses and other issues;
5. Mixup: Deal with the problem of object occlusion and improve the effect of occlusion Object Detection;

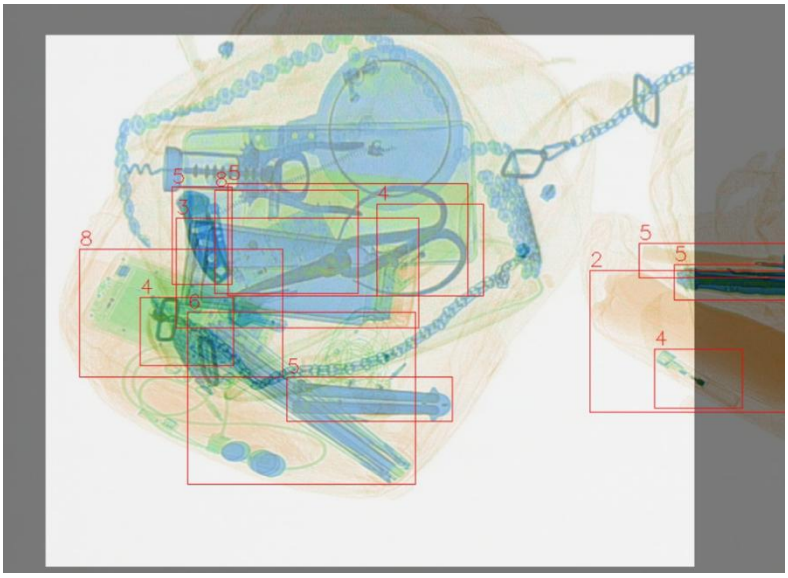

Figure 2. Mixup example

6. CopyPaste: Randomly paste the object foreground on the background, and perform balanced sampling on the target during the pasting process. It is worth mentioning that we use the bounding box of the rotated object.

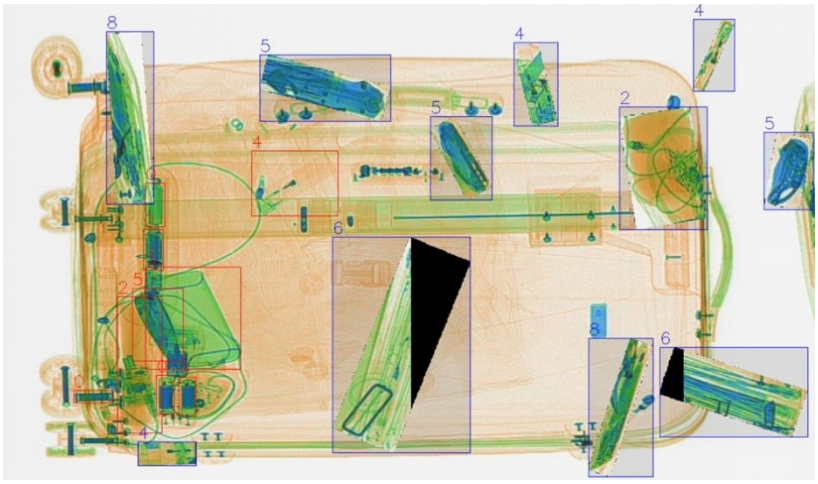

Figure 3. CopyPaste example.

7. SoftNMS: Improve recall of dense objects.
8. Weighted box fusion（WBF）: Ensemble 9 models to improve the AP50 score, including CBNetv2-Swin Large, Cascade RCNN-Hornet Large, CBNetv2-Swin Base, Cascade RCNN-Convnext Large, Detectors-ResNet101, Cascade RCNN-ResNext101_64x4d, Cascade RCNN-Res2Net101, Cascade RCNN-HRNet_w40, VFNet-ResNetxt101_64x4d, WBF IOU threshold selection 0.5. The ensemble weights of the above 9 models are [1, 0.95, 0.9, 0.85, 0.8, 0.75, 0.75, 0.75, 0.75, 0.6].

Tried but ineffective tricks: Mosaic, GridMask

Hardware Device: 8 * V100 (32G)
Environment: CUDA10.2, Pytorch 1.12.1, mmdet 2.25.0, mmcv-full 1.5.3

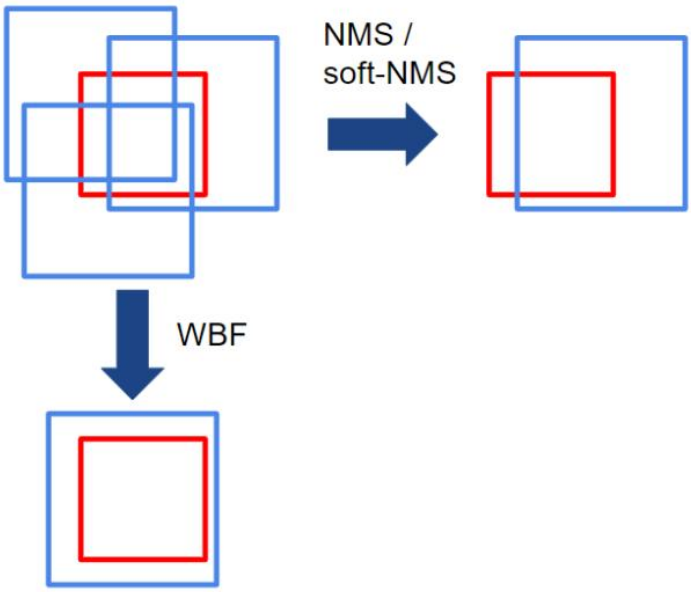

Figure 4. Schematic illustration of NMS/soft-NMS vs. WBF outcomes for an ensemble of inaccurate predictions. Blue – different models' predictions, red – ground truth.

Training details:

Dataset sampling repetition 6 times, all experimental training 12 epochs, the batch size is 8, Swin, Hornet, Convnext initial learning rate 5e-5, weight decay 0.05, ResNet, ResNext, Res2Net, HRNet initial learning rate 0.01, weight decay 1e-5, Learning Rate decay 0.1 in 8, 10 epochs, training scale (1600, 400)~ (1600, 1400)

3. Ablation Study（Best to provide）

The test defaults to using multi-scale tests ((1600, 1000), (1600, 1400), (1800, 1200), (1800, 1600)) and horizontal flip tests, and post-processing uses softNMS.

Table 1. Phase 1 ablation study

Method	CBS	randomRot90	Autoaugment	Mixup	CopyPaste	SoftNMS	AP50
CBNetv2_Swin Large							94.0(phase1-1)
-							94.3(phase1-2)

In order to speed up the training speed, Cascade RCNN + res2net101 was used in the semi-finals to further verify the effectiveness of the tricks. Ensemble 6 models (CBNetv2-Swin Large, Cascade RCNN-Hornet Large, Cascade RCNN-Convnext Large, Detectors-ResNet101, Cascade RCNN-ResNext101_64x4d, Cascade RCNN-Res2Net101), AP50 can reach 90.5, ensembles all 9 models(weights: [1, 0.95, 0.9, 0.85, 0.8, 0.75, 0.75, 0.75, 0.75, 0.6]), AP50 can reach 91.0.

Table 2. Phase 2 ablation study

Method	CBS	randomRot90	Autoaugment	Mixup	CopyPaste	SoftNMS	WBF	AP50
Cascade RCNN_Res2Net101								83.1
-								84.3
-								84.7
-								85.4
-								85.8
-								86.1
-								86.8
CBNetv2_Swin Large								88.9
6 models								90.5
9 models								91.0

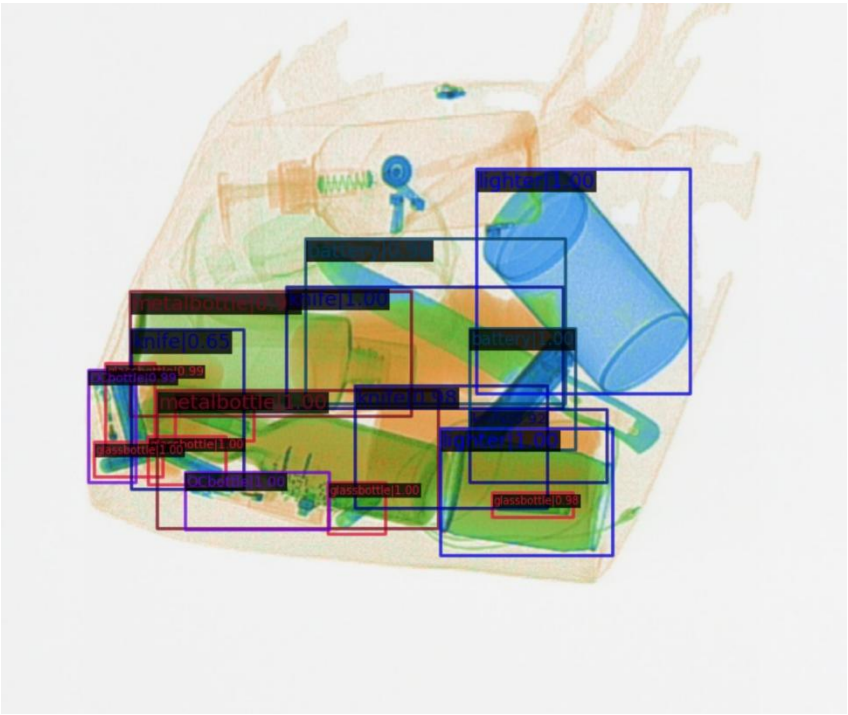

Figure 5. Detection result

4. Reference

1. Liang, T., Chu, X., Liu, Y., Wang, Y., Tang, Z., Chu, W., Chen, J., Ling, H.: Cbnetv2: A composite backbone network architecture for object detection. arXiv preprint arXiv:2107.00420 (2021)
2. Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
3. Zoph, B., Cubuk, E.D., Ghiasi, G., Lin, T.Y., Shlens, J., Le, Q.V.: Learning data augmentation strategies for object detection. In: European Conference on Computer Vision. pp. 566–583. Springer (2020)
4. Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.
5. Liu Z, Mao H, Wu C Y, et al. A convnet for the 2020s[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 11976-11986.
6. Rao Y, Zhao W, Tang Y, et al. Hornet: Efficient high-order spatial interactions with recursive gated convolutions[J]. Advances in Neural Information Processing Systems, 2022, 35: 10353-10366.
7. Qiao S, Chen L C, Yuille A. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 10213-10224.