Prediction-Guided Distillation for Dense Object Detection (Supplementary Material)

Chenhongyi Yang[1], Mateusz Ochal[2,3], Amos Storkey[2], and Elliot J. Crowley[1]

[1] School of Engineering, University of Edinburgh, UK
[2] School of Informatics, University of Edinburgh, UK
[3] School of Engineering and Physical Sciences, Heriot-Watt University, UK

1 Additional Experiments

1.1 Weak to strong distillation

We conduct experiments to show that our method is also effective when distilling a weak teacher into a stronger model. Specifically, we use ATSS [15] as the detector and use a ResNet-50 based model as teacher with a ResNet-101 based model as student. The results are shown in Table 1. We observe that the student AP is improved to 44.2 from 41.4, even surpassing the teacher performance. A possible reason for this is that the student learns from the key predictive regions where it should focus on. Then, empowered by its stronger backbone, the student is able to surpass the teacher model.

Model	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
R50 (ms, 3×)	43.6	61.8	47.4	28.5	48.1	54.3
R101 (1×)	41.4	59.8	45.2	24.2	45.8	53.8
R50→R101 (1×)	44.2	62.7	47.9	28.8	48.7	55.7

Table 1: Distillation results using our PGD on ATSS detector. Training settings are inherited from the paper: teacher is trained for 3× schedule with multi-scale input; Student and KD models are trained for 1× schedule with single-scale input.

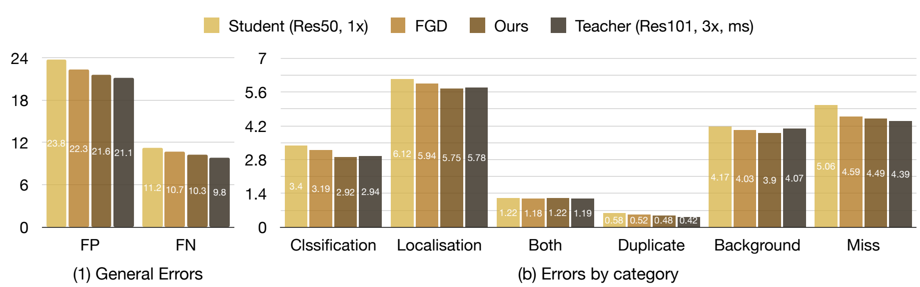

Fig. 1: TIDE error analysis on COCO *mini-val* using ATSS as the object detector.

1.2 Improvement Discussion

We used the TIDE [1] toolkit to analyse the performance improvement after KD and compare our approach with the baseline FGD. We use ATSS as the object detector and present the COCO *mini-val* evaluation analysis under 0.5 IoU threshold in Fig. 1. For the general errors, we observe that our PGD can significantly improve both false negatives and false positives over the baseline FGD. When we look into different categories of errors, we find PGD can improve most error cases, including *Classification*, *Localisation*, *Duplicate*, *Background*, and *Missing* errors. Moreover, the student model improved by our method makes fewer *Localisation* and *Background* errors than the teacher model; our method can improve a student model's localisation ability and the ability to ignore background noise. However, it does not lead to a reduction in *Both* errors, which happens when classification and localisation are both incorrect. However, as this error is similar between the teacher and plain student models, we attribute this problem to training randomness.

We also used the recently proposed LRP [12] metrics to compare our method with the baseline FGD. The results are shown in Table 2. Our method improves the student detector's LRP performance in both classification and localisation comparing with the FGD, illustrating the effectiveness of our approach in improving the student's detection ability.

Setting	AP	LRP	LRP_{loc}	LRP_{FP}	LRP_{FN}
Teacher (R101, 3×, ms)	45.5	63.5	14.1	26.2	41.0
Student (R50, 1×)	39.4	68.5	15.3	30.5	46.5
FGD	42.6	66.0	14.5	28.0	44.1
PGD (Ours)	44.2	64.6	14.2	26.1	43.0

Table 2: LRP analysis on COCO mini-val using ATSS as the object detector

Setting	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
RetinaNet [10]	39.1	59.1	42.3	21.8	42.7	50.2
FCOS [13]	41.5	60.7	45.0	24.4	44.8	51.6
FreeAnchor [16]	43.1	62.2	46.4	24.5	46.1	54.8
SAPD [18]	43.5	63.6	46.5	24.9	46.8	54.6
ATSS [15]	43.6	62.1	47.4	26.1	47.0	53.6
AutoAssign [17]	44.5	64.3	48.4	25.9	47.4	55.0
PAA [7]	44.8	63.3	48.7	26.5	48.8	56.3
GFL [9]	45.0	63.7	48.9	27.2	48.8	54.5
IQDet [11]	45.1	63.4	49.3	26.7	48.5	56.6
OTA [6]	45.3	63.5	49.3	26.9	48.8	56.1
GFLv2 [8]	46.0	64.1	50.2	27.6	49.6	56.5
VFNet [14]	46.0	64.2	50.0	27.5	49.4	56.9
RepPointsV2 [2]	46.0	65.3	49.5	27.4	48.9	57.3
DDOD [3]	46.7	65.3	51.1	28.2	49.9	57.9
Ours	**48.2**	**66.9**	**52.5**	**30.1**	**51.6**	**58.5**

Table 3: Comparison with state-of-the-art dense object detectors on COCO *test-dev*. All models are trained for 2× (24 epochs) with ResNet-101 as backbone.

1.3 COCO *Test-Dev* results

Knowledge distillation aims to equip a lightweight model with a strong generalisation capability. With this in mind, we compare a detector produced using our method with the most state-of-the-art dense object detectors on the COCO *test-dev* set. We used DDOD [3] as the object detector; ResNet-101 is used as the student backbone, and Res2Net-DCN [5, 4] is used as the teacher backbone. We train the model using COCO *train-2017* set with 2× training schedule (24 epochs) and multi-scale input. The results are presented in Tab 3, from which we can see that the model trained using our method achieves the highest AP, suggestion that our approach can indeed improve a detector's generalisation ability.

References

1. Bolya, D., Foley, S., Hays, J., Hoffman, J.: TIDE: A General Toolbox for Identifying Object Detection Errors. In: ECCV (2020)
2. Chen, Y., Zhang, Z., Cao, Y., Wang, L., Lin, S., Hu, H.: Reppoints v2: Verification meets regression for object detection. In: NeurIPS (2020)
3. Chen, Z., Yang, C., Li, Q., Zhao, F., Zha, Z.J., Wu, F.: Disentangle your dense object detector. In: ACM MM (2021)
4. Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y.: Deformable convolutional networks. In: ICCV (2017)
5. Gao, S., Cheng, M.M., Zhao, K., Zhang, X.Y., Yang, M.H., Torr, P.H.: Res2net: A new multi-scale backbone architecture (2021). https://doi.org/10.1109/TPAMI.2019.2938758
6. Ge, Z., Liu, S., Li, Z., Yoshie, O., Sun, J.: OTA: Optimal Transport Assignment for Object Detection. In: CVPR (2021)
7. Kim, K., Lee, H.S.: Probabilistic anchor assignment with iou prediction for object detection. In: ECCV (2020)
8. Li, X., Wang, W., Hu, X., Li, J., Tang, J., Yang, J.: Generalized Focal Loss V2: Learning Reliable Localization Quality Estimation for Dense Object Detection. In: CVPR (2021)
9. Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., Tang, J., Yang, J.: Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. In: NeurIPS (2020)
10. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV (2017)
11. Ma, Y., Liu, S., Li, Z., Sun, J.: IQDet: Instance-Wise Quality Distribution Sampling for Object Detection. In: CVPR (2021)
12. Oksuz, K., Cam, B.C., Kalkan, S., Akbas, E.: One metric to measure them all: Localisation recall precision (lrp) for evaluating visual detection tasks. TPAMI (2021). https://doi.org/10.1109/TPAMI.2021.3130188
13. Tian, Z., Shen, C., Chen, H., He, T.: FCOS: Fully convolutional one-stage object detection. In: ICCV (2019)
14. Zhang, H., Wang, Y., Dayoub, F., Sunderhauf, N.: VarifocalNet: An IoU-Aware Dense Object Detector. In: CVPR) (2021)
15. Zhang, S., Chi, C., Yao, Y., Lei, Z., Li, S.Z.: Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: CVPR (2020)
16. Zhang, X., Wan, F., Liu, C., Ji, R., Ye, Q.: Freeanchor: Learning to match anchors for visual object detection (2019)

17. Zhu, B., Wang, J., Jiang, Z., Zong, F., Liu, S., Li, Z., Sun, J.: Autoassign: Differentiable label assignment for dense object detection. In: arXiv preprint arXiv:2007.03496 (2020)
18. Zhu, C., Chen, F., Shen, Z., Savvides, M.: Soft anchor-point object detection (2020)