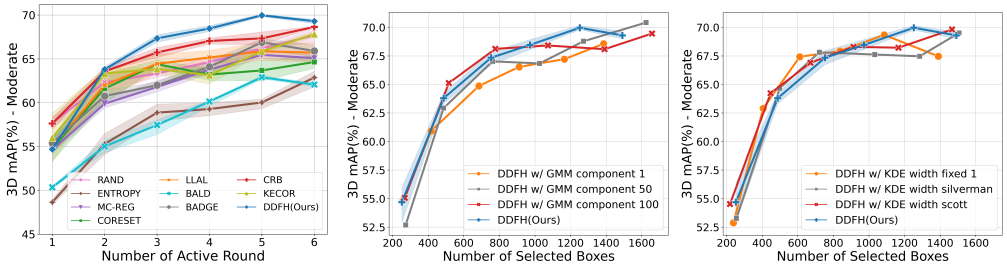
# Supplementary Material of Distribution Discrepancy and Feature Heterogeneity for Active 3D Object Detection

Table 1: Compare 3D mAP(%) scores for different SOTA apporch in KITTI Dataset when acquiring approximately 1% queried bounding boxes. † indicates the reported performance of the backbone trained with the 100% labeled set.

| Method | AVERAGE | | | CAR | | | PEDESTRIAN | | | CYCLIST | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Easy | Mod. | Hard | Easy | Mod. | Hard | Easy | Mod. | Hard | Easy | Mod. | Hard |
| CRB [1] | 79.06 | 66.49 | 61.76 | 90.81 | 79.06 | 74.73 | 62.09 | 54.56 | 48.89 | 84.28 | 65.85 | 61.66 |
| CRB(offi.) | 80.70 | 67.81 | 62.81 | 90.98 | 79.02 | 74.04 | 64.17 | 50.82 | 50.82 | 86.96 | 67.45 | 63.56 |
| KECOR [2] | 79.81 | 67.83 | 62.52 | 91.43 | 79.63 | 74.41 | 63.49 | 56.31 | 50.20 | 84.51 | 67.54 | 62.96 |
| KECOR(offi.) | 81.63 | 68.67 | 63.42 | 91.71 | 79.56 | 74.05 | 65.37 | 57.33 | 51.56 | 87.80 | 69.13 | 64.65 |
| DDFH(Ours) | **82.27** | **69.84** | **64.76** | **91.76** | **80.65** | **76.46** | **66.37** | **59.40** | **52.97** | **88.68** | **69.47** | **64.85** |
| PV-RCNN† | 81.75 | 70.99 | 67.06 | 92.56 | 84.36 | 82.48 | 64.26 | 56.67 | 51.91 | 88.88 | 71.95 | 66.78 |




(a) (b) (c)

Figure 6: (a) report 3D mAP of various AL methods on KITTI in each active round. (b-c) represents the impact of different density estimation methods and varying parameter settings on the performance of DDFH.

## A More Implementation Details

To ensure the fairness and reproducibility of our experiments, we implemented DDFH and reproduced most of the baselines based on the public ACTIVE-3D-DET toolbox. We followed all KECOR training settings, using Adam as the optimizer, and a onecycle learning scheduler with an initial learning rate of 0.01. The batch size was set to 6, and each active round was trained for 40 epochs before proceeding to a new sampling round. We used one NVIDIA RTX A6000 to complete all experiments. The runtime for an experiment on KITTI and Waymo is approximately 5 and 81 GPU hours, respectively. The model embeddings $f^e$ used in our method are extracted from the second convolutional layer in the shared block of PV-RCNN.

## B More Experimental Details

**DDFH in the KITTI Dataset.** In Fig. 6a, we present the performance of various AL methods in each active round. The number of point clouds in each active round is fixed, allowing us to compare the performance of models under conditions where they have seen the same number of scenes. Notably, KECOR's performance is below expectations given the same number of frames, indicating that KECOR does not effectively consider the diversity information of the scenes. In contrast, DDFH considers frame-level information to avoid redundant instances in similar scenes. The results show that DDFH has a significant advantage in each active round. We present more

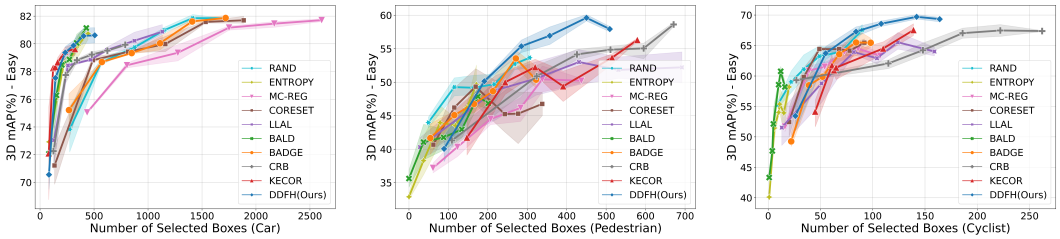



Figure 7: 3D mAP(%) of DDFH and the AL Baseline across various categories on the KITTI dataset at the moderate difficulty with PV-RCNN.

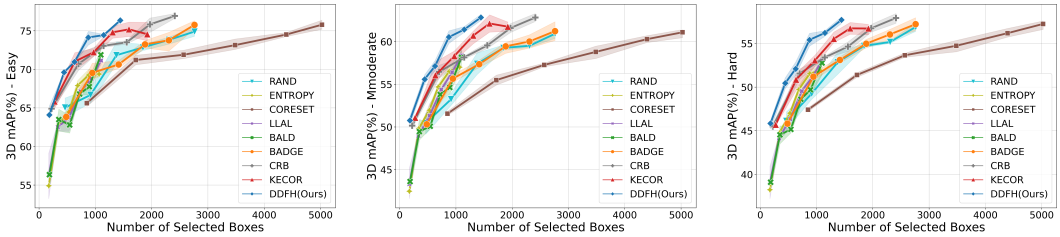



Figure 8: 3D mAP(%) of DDFH and AL baselines on the KITTI val split with SECOND.

comprehensive experimental results of DDFH on the KITTI Dataset in Fig. 7 and Fig. 8. The results in Fig. 7 indicate that DDFH with PV-RCNN has a significant advantage in all categories in KITTI, consistent with the results of Figure 4 in main text on SECOND. It is noteworthy that in the car category, some uncertainty-based methods achieve similar performance to DDFH with the same annotation cost. However, these methods fail to improve effectively in other categories, demonstrating DDFH's effectiveness in resource allocation and diversity. Fig. 8 also provides the trend of average 3D mAP for the one-stage model SECOND in different difficulties, consistent with PV-RCNN, outperforming SOTA methods in all difficulties. Further, in Table 1, we provide the performance of PV-RCNN trained on 100% labeled data, showing that DDFH's performance with only 1% of bounding box annotation is close to fully trained performance, even outperforming fully trained models in the pedestrian category.

**Ablation Study of Density Estimation.** We also test the stability and generalizability of DDFH through different density estimation methods and parameters. In Fig. 6b, we set different numbers of GMM components, specifically 1, 10 (DDFH Ours), 50, and 100. The results indicate that all experiments, except for 1 component, maintain similar effectiveness. In Fig. 6c, we use Kernel Density Estimation (KDE) to estimate the probability density and adjust different bandwidths to test the stability and generalizability of the DDFH. Silverman [3] and Scott [4] calculate bandwidth based on sample size. The results show that the performance of DDFH remains consistent and stable under different density estimation models and parameters. This is due to the distribution discrepancy focusing on distribution differences and novelty, rather than relying on highly accurate distribution estimates, thus providing sufficient robustness to noisy instances and estimation deviations.

# C Limitation

Considering that the distribution of objects in real environments is often uneven, common objects tend to occupy the majority (e.g. cars). This leads to the underestimation of less frequent categories when estimating informativeness. Therefore, the components DD, FH, and CB in DDFH reduce the impact of uneven distribution at different levels, decrease redundant annotations, and effectively balance minority categories. Although most real-world scenarios exhibit an uneven long-tail distribution, if specific situations lead to a dataset where object distribution is close to a uniform distribution, the effectiveness of DDFH might be limited due to the less apparent distribution differences. A possible solution is to incorporate indicators of uncertainty into DDFH, such as model instability, entropy, or the kernel coding rate combined with KECOR. This approach could address the mentioned limitation and is left for future research.

# References


[1] Y. Luo, Z. Chen, Z. Wang, X. Yu, Z. Huang, and M. Baktashmotlagh. Exploring active 3d object detection from a generalization perspective. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=2RwXVje1rAh.

[2] Y. Luo, Z. Chen, Z. Fang, Z. Zhang, M. Baktashmotlagh, and Z. Huang. Kecor: Kernel coding rate maximization for active 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18279–18290, 2023.

[3] B. W. Silverman. Density estimation for statistics and data analysis. 1986.

[4] D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, aug 1992. ISBN 9780470316849. doi:10.1002/9780470316849. URL http://dx.doi.org/10.1002/9780470316849.