SUPPLEMENTARY MATERIAL OF EXPLORING ACTIVE 3D OBJECT DETECTION FROM A GENERALIZATION PERSPECTIVE

Yadan Luo*, **Zhuoxiao Chen***, **Zijian Wang, Xin Yu, Zi Huang, Mahsa Baktashmotlagh**
The University of Queensland, Australia

In this appendix, we discuss the prior distribution selection, motivation of the Stage 2, and evaluation division of difficulty in Sec A.1 and Sec A.2, respectively. In the rest of the supplementary material, we provide the implementation details of all baselines and the proposed approach in Sec B followed by the proof of Theorem C.1. In Sec D, the overall algorithm is summarized. Additional experimental results on KITTI (Sec E) and Waymo (Sec F) datasets are reported and analyzed. We further conducted supplemental experiments on parameter sensitivity (Sec H) and visualizations (Sec G). In the end, we leave the related work and the associated discussion in Sec I.

A APPENDIX

A.1 MORE DISCUSSIONS ON PRIOR DISTRIBUTION

In mainstream 3D detection datasets, the curated test set is commonly long-tailed distributed, with a few head classes (*e.g.*, car) possessing a large number of samples and all the rest of the tail classes possessing only a few samples. As such, the trained detector can be easily biased towards head classes with massive training data, resulting in high accuracy on head classes and low accuracy on tail classes. This suggests that for 3D detection tasks, **mean average precision (mAP)** can be a **fairer** metric of evaluation, by taking an average of all AP values per class. When the test label is uniformly distributed, mAP scores will be equal to the AP scores for all samples. This motivates us to choose the uniform distribution as the prior distribution, rather than estimating the test label distribution from the initial labeled set $\mathcal{D}_L$. In this case, the trained model tends to be more robust and resilient to the imbalanced training data, achieving higher mAP scores.

To justify the effectiveness of choosing the uniform distribution, we provide more comparisons with the SOTA active learning methods in Table 2 and Table 1, which do not take the uniform distribution as an assumption. We clearly observe that such AL methods perform poorly on **tail classes** (*e.g.*, pedestrian and cyclist), confirming that the yielded models are biased towards learning car samples.

Table 1: Performance gap (%) between different AL methods and fully supervised backbone when acquiring approximately 1% queried bounding boxes on KITTI. Gaps are calculated by subtracting the performance of a fully supervised backbone from the performance of AL methods.

	Car (↓)			Pedestrian (↓)			Cyclist (↓)			Average (↓)		
Method	EASY	MOD.	HARD	EASY	MOD.	HARD	EASY	MOD.	HARD	EASY	MOD.	HARD
LLAL	2.61	5.71	**7.16**	7.92	6.80	5.94	13.33	11.60	11.42	7.81	8.04	8.18
CORESET	4.79	6.63	9.53	16.99	14.70	13.72	7.15	12.23	11.14	9.49	11.18	11.47
BADGE	2.60	8.58	11.94	12.32	10.43	10.93	4.77	9.66	8.66	6.41	9.55	10.51
CRB	**1.58**	**5.34**	8.44	**0.09**	**1.87**	**1.09**	**1.92**	**4.50**	**3.22**	**1.05**	**3.18**	**4.25**

A.2 MORE DISCUSSIONS ON EVALUATION DIVISION OF DIFFICULTY

On the KITTI dataset, the evaluation difficulty is set based on the visual look[1] of the images, which is supposed to be unavailable for our LiDAR-based detection task. On the other hand, the Waymo dataset leverages a more reasonable and general setting of difficulty evaluation, with LEVEL 1

*Equal contribution. Correspondence to Yadan Luo <y.luo@uq.edu.au>.

[1] http://www.cvlibs.net/datasets/kitti/eval_object.php

Table 2: Performance comparisons on KITTI *val* set with different SOTA AL methods when acquiring approximately 1% queried bounding boxes. Results are reported with 3D AP with 40 recall positions. † indicates the reported performance of the backbone trained with the full labeled set (100%).

	Car			Pedestrian			Cyclist			Average		
Method	EASY	MOD.	HARD	EASY	MOD.	HARD	EASY	MOD.	HARD	EASY	MOD.	HARD
LLAL	89.95	78.65	**75.32**	56.34	49.87	45.97	75.55	60.35	55.36	73.94	62.95	58.88
CORESET	87.77	77.73	72.95	47.27	41.97	38.19	81.73	59.72	55.64	72.26	59.81	55.59
BADGE	89.96	75.78	70.54	51.94	46.24	40.98	84.11	62.29	58.12	75.34	61.44	56.55
CRB	**90.98**	**79.02**	74.04	**64.17**	**54.80**	**50.82**	**86.96**	**67.45**	**63.56**	**80.70**	**67.81**	**62.81**
PV-RCNN†	92.56	84.36	82.48	64.26	56.67	51.91	88.88	71.95	66.78	81.75	70.99	67.06

and LEVEL 2 difficulties indicating "more than five points" and "at least one point" inside labeled bounding boxes, respectively. This aligns with the design of the balance criterion (Stage 3), as the sparse point clouds or dense point clouds can be equally learned. In Table 3, we report the performance of the proposed approach with a small portion of point clouds and the fully supervised baseline reported in (Zhang et al., 2022), on the Waymo dataset. From Table 3, we can observe that the performance gap between the detectors trained with active learning (approx. 50K bounding box annotations) and fully supervised learning (approx. 8 million bounding box annotations) is smaller in LEVEL 2 (7.18% in LEVEL 2 vs 8.08% in LEVEL 1), which aligns with the balance criteria in the proposed CRB framework.

A.3 MORE DISCUSSIONS ON THE MOTIVATION OF THE STAGE 2

Our main objective of Stage 2, *i.e.*, Representative Prototype Selection is to determine a subset $\mathcal{D}^*_{S_2}$ from the pre-selected set S_1 in the last stage, by minimizing the set discrepancy in the latent feature space. However, the test features are not observable during the training phase, and it is hard to guarantee that the feature distribution can be comprehensively captured. As stated in Remark section, we focus on the features that are not learned well from the training set due to the zero training error assumption and reconsider the feature matching problem from a gradient perspective. In particular, we split the test set into two group In the gradient space: (1) seen test samples that can be easily recognized will cluster near the origin, (2) while the novel test samples will diversely distribute in the subspace. As the first group of samples have been sufficiently covered by the initiated, in this stage, we focus on finding matching with the latter group. By assuming the prior distribution of gradients follows a Gaussian distribution, finding the K-metroids is naturally a choice to mitigate the gap between mean and variance. K-metroids algorithm breaks the dataset up into groups and attempts to minimize the distance between points labeled to be in a cluster and a point designated as the center of that cluster (*i.e.*, prototype). By selecting the prototypes in the second stage, we implicitly bridge the gap between the selected set and the test set at a latent feature level.

B IMPLEMENTATION DETAILS

B.1 EVALUATION METRICS.

To fairly evaluate baselines and the proposed method on KITTI dataset (Geiger et al., 2012), we follow the work of (Shi et al., 2020): we utilize Average Precision (AP) for 3D and bird eye view (BEV) detection, and the task difficulty is categorized to EASY, MODERATE, and HARD, with a rotated IoU threshold of 0.7 for cars and 0.5 for pedestrian and cyclists. The results evaluated on the validation split are calculated with 40 recall positions. To evaluate on Waymo dataset (Sun et al., 2020), we adopt the officially published evaluation tool for performance comparisons, which utilizes AP and the average precision weighted by heading (APH). The respective IoU thresh-

Table 3: Comparing the performance of detectors with active learning (AL) by CRB and fully supervised learning (FSL) on Waymo *val* set. Results (mAP %) are calculated by Waymo official evaluation metric.

Method	mAP Level 1	mAP Level 2
CRB	58.60	52.65
FSL	66.68	59.83
Gap (↓)	−8.08	−7.18

olds for vehicles, pedestrians, and cyclists are set to 0.7,
0.5, and 0.5. Regarding detection difficulty, the Waymo test set is further divided into two levels. LEVEL 1 (and LEVEL 2) indicates there are more than five inside points (at least one point) in the ground-truth objects.

B.2 IMPLEMENTATION DETAILS OF TRAINING

To ensure the reproducibility of the baselines and the proposed approach, we develop a PyTorch-based active 3D detection toolbox (attached in the supplemental material) that implements mainstream AL approaches and can accommodate most of the public benchmark datasets. For fair comparison, all active learning methods are constructed from the PV-RCNN (Shi et al., 2020) backbone. All experiments are conducted on a GPU cluster with three V100 GPUs. The runtime for an experiment on KITTI and Waymo is around 11 hours and 100 hours, respectively. Note that, training PV-RCNNon the full set typically requires 40 GPU hours for KITTI and 800 GPU hours for Waymo.

Parameter Settings. The batch sizes for training and evaluation are fixed to 6 and 16 on both datasets. The Adam optimizer is adopted with a learning rate initiated as 0.01, and scheduled by one cycle scheduler. The number of MC-DROPOUT stochastic passes is set to 5 for all methods.

Active Learning Protocols. As our work is the first comprehensive study on active 3D detection task, the active training protocol for all AL baselines and the proposed method is empirically defined. For all experiments, we first randomly select m fully labeled point clouds from the training set as the initial $\mathcal{D}_L$. With the annotated data, the 3D detector is trained with E epochs, which is then freezed to select N_r candidates from $\mathcal{D}_U$ for label acquisition. We set the m and N_r to 2.5 3% point clouds (*i.e.*, $N_r = m =$100 for KITTI, $N_r = m =$400 for Waymo) to trade-off between reliable model training and high computational costs.The aforementioned training and selection steps will alternate for R rounds. Empirically, we set $E = 30$, $R = 6$ for KITTI, and fix $E = 40$, $R = 5$ for Waymo.

B.3 IMPLEMENTATION DETAILS OF BASELINES AND CRB

In this section, we introduce more implementation details of both baselines and the proposed CRB.

CRB. In comparison with baselines as reported in Figure 2, the $\mathcal{K}_1$ and $\mathcal{K}_2$ are empirically set to $300, 200$ for KITTI and $2,000$ and $1,200$ for Waymo. The gradient maps used for RPS are extracted from the second convolutional layer in the shared block of PV-RCNN. Three dropout layers in PV-RCNN are enabled during the MC-DROPOUT and the dropout rate is fixed to 0.3 for both datasets. The number of MC-DROPOUT stochastic passes are set to 5 for all methods. In the GPCB stage, we measure the KL-divergence between the KDE PDF of the selected set and the uniform prior distribution of the point cloud density for each class. The goal of conducting a greedy search is to find the optimal subset that can achieve the minimum sum of KL divergence for all classes. Considering the high variance of KL divergence across different classes, we unify the scale of KL-divergence to $\bar{d}_c$ by applying the following function,

$$\bar{d}_c = \frac{2}{\pi} \arctan \frac{\pi}{2} d_c,$$

where d_c denotes the KL-divergence for the c-th class. To this end, the ultimate objective for greedy search is $\arg\min_{\mathcal{D}_S \subset \mathcal{D}_{S_2}} \sum_{c \in [C]} \bar{d}_c$. The normalized measurement can avoid dominance by any single class.

CORESET (Sener & Savarese, 2018). The embeddings extracted for both labeled and unlabeled data are the output from the shared block, with the dimension of 128 by 256. The CORESET adopts the furthest-first traversal for k-Center clustering strategy, which computes the Euclidean distance between each embedding pair.

LLAL (Yoo & Kweon, 2019). For implementing the loss prediction module in LLAL, we construct a two-block module that connects to two layers of the PV-RCNN, which takes multi-level knowledge into consideration for loss prediction. Particularly, each block consists of a convolutional layer with a channel size of 265 and a kernel size of 1, a batchnorm layer, and a relu activation layer. The outputs are then concatenated and fed to a fully connected layer and map to a loss score. All real loss for each training data point is saved and serves as the ground-truth to train the loss prediction module.

BADGE. According to (Ash et al., 2020), hypothetical labels for the classifier are determined by the classes with the highest predicted probabilities. The gradient matrix with the dimension 256 by 256 for each unlabeled point cloud is extracted from the last convolutional layer of the PV-RCNN's classification head and then fed into the BADGE algorithm.

C PROOF OF THEOREM 2.1

Theorem C.1. *Let $\mathcal{H}$ be a hypothesis space of Vapnik-Chervonenkis (VC) dimension d, with f and g being the classification and regression branches, respectively. The $\widehat{\mathcal{D}}_S$ and $\widehat{\mathcal{D}}_T$ represent the empirical distribution induced by samples drawn from the acquired subset $\mathcal{D}_S$ and the test set $\mathcal{D}_T$, and ℓ the loss function bounded by $\mathcal{J}$. It is proven that $\forall\ \delta \in (0,1)$, and $\forall f, g \in \mathcal{H}$, with probability at least $1-\delta$ the following inequality holds,*

$$\mathfrak{R}_T[\ell(f,g;\boldsymbol{w})] \leq \mathfrak{R}_S[\ell(f,g;\boldsymbol{w})] + \frac{1}{2}disc(\widehat{\mathcal{D}}_S, \widehat{\mathcal{D}}_T) + \lambda^* + const,$$

where $const = 3\mathcal{J}(\sqrt{\frac{\log\frac{4}{\delta}}{2N_r}} + \sqrt{\frac{\log\frac{4}{\delta}}{2N_t}}) + \sqrt{\frac{2d\log(eN_r/d)}{N_r}} + \sqrt{\frac{2d\log(eN_t/d)}{N_t}}$.

Notably, $\lambda^* = \mathfrak{R}_T[\ell(f^*,g^*;\boldsymbol{w}^*)] + \mathfrak{R}_S[\ell(f^*,g^*;\boldsymbol{w}^*)]$ *denotes the joint risk of the optimal hypothesis f^* and g^*, with $\boldsymbol{w}^*$ being the model weights. N_r and N_t indicate the number of samples in the $\mathcal{D}_S$ and $\mathcal{D}_T$. The proof can be found in the supplementary material.*

Proof. For brevity, we omit the model weights $\boldsymbol{w}$ in the following proof. Based on the triangle inequality of ℓ and the definition of the discrepancy distance $disc(\cdot,\cdot)$, the following inequality holds,

$$\begin{aligned}
\mathfrak{R}_T[\ell(f,g)] &\leq \mathfrak{R}_T[\ell(f^*,g^*)] + \frac{1}{2}\mathfrak{R}_T[\ell(f,f^*)] + \frac{1}{2}\mathfrak{R}_T[\ell(g,g^*)] \\
&\leq \mathfrak{R}_T[\ell(f^*,g^*)] + \mathfrak{R}_S[\ell(f^*,g^*)] + \frac{1}{2}|\mathfrak{R}_T[\ell(f,f^*)] - \mathfrak{R}_S[\ell(f,f^*)]| \\
&+ \frac{1}{2}|\mathfrak{R}_T[\ell(g,g^*)] - \mathfrak{R}_S[\ell(g,g^*)]| \\
&\leq \mathfrak{R}_T[\ell(f^*,g^*)] + \mathfrak{R}_S[\ell(f^*,g^*)] + \frac{1}{2}disc(\mathcal{D}_S,\mathcal{D}_T) \\
&\leq \mathfrak{R}_T[\ell(f^*,g^*)] + \mathfrak{R}_S[\ell(f,g)] + \mathfrak{R}_S[\ell(f^*,g^*)] + \frac{1}{2}disc(\mathcal{D}_S,\mathcal{D}_T).
\end{aligned}$$

By defining the joint risk of the optimal hypothesis $\lambda^* = \mathfrak{R}_T[\ell(f^*,g^*)] + \mathfrak{R}_S[\ell(f^*,g^*)]$ and the Corollary 6 in (Mansour et al., 2009), we have,

$$\begin{aligned}
\mathfrak{R}_T[\ell(f,g)] &\leq \mathfrak{R}_S[\ell(f,g)] + \frac{1}{2}disc(\mathcal{D}_S,\mathcal{D}_T) + \lambda^* \\
&\leq \mathfrak{R}_S[\ell(f,g)] + \frac{1}{2}disc(\widehat{\mathcal{D}}_S,\widehat{\mathcal{D}}_T) + \lambda^* + 4q(\mathrm{Rad}_S(\mathcal{H}) + \mathrm{Rad}_T(\mathcal{H})) \\
&+ 3\mathcal{J}(\sqrt{\frac{\log\frac{4}{\delta}}{2N_r}} + \sqrt{\frac{\log\frac{4}{\delta}}{2N_t}}),
\end{aligned}$$

where N_r and N_t indicate the sample size of the selected set and the test set, respectively. q stands for the function is q-Lipschitz. As our regression loss, ℓ^{reg} is the smooth-L1 loss function and bounded by $\mathcal{J}$, q equals 1 in our case. $\mathrm{Rad}_S(\mathcal{H})$ and $\mathrm{Rad}_T(\mathcal{H})$ indicates the empirical Rademacher complexity of a hypothesis set $\mathcal{H}$ whose VC dimension is d over the selected set and the test set.

Considering the Rademacher complexity is bounded by:

$$\mathrm{Rad}_S(\mathcal{H}) \leq \sqrt{\frac{2d\log(eN_r/d)}{N_r}}, \quad \mathrm{Rad}_T(\mathcal{H}) \leq \sqrt{\frac{2d\log(eN_t/d)}{N_t}},$$

then we can rewrite the inequality as,

$$\mathfrak{R}_T[\ell(f,g)] \leq \mathfrak{R}_S[\ell(f,g)] + \frac{1}{2}disc(\widehat{\mathcal{D}}_S,\widehat{\mathcal{D}}_T) + \lambda^* + const,$$

where $const = 3\mathcal{J}(\sqrt{\frac{\log\frac{4}{\delta}}{2N_r}} + \sqrt{\frac{\log\frac{4}{\delta}}{2N_t}}) + \sqrt{\frac{2d\log(eN_r/d)}{N_r}} + \sqrt{\frac{2d\log(eN_t/d)}{N_t}}$. □

Algorithm 1 The algorithm of CRB for active 3D object detection

Inputs:
$\mathcal{D}_L$: initially labeled point clouds
$\mathcal{D}_U$: unlabeled pool of point clouds
$\mathbf{\Omega}$: oracle
B: total budget of active selection
$e(\cdot) : \mathcal{P} \rightarrow \boldsymbol{x}$: point cloud encoder of 3D detector
$f(\cdot)$: classifier of 3D detector
$g(\cdot)$: regression head of 3D detector
R: total active learning rounds

$\mathcal{D}_S \leftarrow \emptyset$
Pre-train 3D detector $\{e(\cdot), f(\cdot), g(\cdot)\}$ with $\mathcal{D}_L$ until converge
$\mathcal{D}_S \leftarrow \mathcal{D}_S \cup \mathcal{D}_L$
for $r \in [R]$ **do** ▷ For each round of active selection
  $\widehat{Y} \leftarrow f \circ e(\mathcal{D}_U)$ ▷ Get the predicted labels
  $\widehat{\mathcal{B}}, \phi \leftarrow g \circ e(\mathcal{D}_U)$ ▷ Get the predicted boxes $\widehat{\mathcal{B}}$ and box point densities ϕ
  $\bar{B} \leftarrow$ Hypothetical labels computed by Equation (5)
  $\mathcal{D}^*_{S_1} \leftarrow \textsc{Cls}(\mathcal{D}_U, \widehat{Y})$ selects $\mathcal{K}_1$ samples via Equation (2),(3),(4)
  ▷ Stage 1: Concise Label Sampling
  $\mathcal{D}^*_{S_2} \leftarrow \textsc{Rps}(\mathcal{D}^*_{S_1}, e(\mathcal{D}^*_{S_1}), \bar{B})$ selects $\mathcal{K}_2$ samples via Equation (6),(7)
  ▷ Stage 2: Representative Prototype Selection
  $\mathcal{D}^*_S \leftarrow \textsc{Gpdb}(\mathcal{D}^*_{S_2}, \widehat{\mathcal{B}}, \phi)$ selects N_r samples via Equation (8),(9)
  ▷ Stage 3: Greedy Point Cloud Density Balancing
  $\mathcal{D}^*_S \leftarrow \mathbf{\Omega}(\mathcal{D}^*_S)$ ▷ Query labels from oracles
  $\mathcal{D}_S \leftarrow \mathcal{D}_S \cup \mathcal{D}^*_S$
  $\mathcal{D}_U \leftarrow \mathcal{D}_U \backslash \mathcal{D}^*_S$
  Train 3D detector $\{e(\cdot), f(\cdot), g(\cdot)\}$ with $\mathcal{D}_S$ until converge
end for

D Algorithm Description

To thoroughly describe the procedure of active 3D object detection by the proposed CRB, we present the Algorithm 1 in detail. Firstly, the 3D detector consisting of an encoder $\{e(\cdot)$, a classifier $f(\cdot)$, and regression heads $g(\cdot)\}$ is pre-trained with a small set $\mathcal{D}_L$ of labeled point clouds. During the stage 1: Cls, the pre-trained 3D detector infers all samples from the unlabeled pool $\mathcal{D}_U$ and obtains the predicted bounding boxes $\widehat{\mathcal{B}}$, predicted box labels $\widehat{Y}$, and calculated box point densities ϕ for each point cloud. Also, the hypothetical labels $\bar{B}$ through stochastic monte-carlo sampling are computed by Equation (5) during inference. Based on the criterion of maximizing the label entropy, the set $\mathcal{D}^*_{S_1}$ containing $\mathcal{K}_1$ candidates are formed via Equations (2), (3), (4). In stage 2, we set the model to the training mode and allow the gradient back-propagation to retrieve gradients for each point cloud. Yet, the model weights will be fixed and not updated. Rps selects the set $\mathcal{D}^*_{S_2}$ of size $\mathcal{K}_2$ from the previous candidate set $\mathcal{D}^*_{S_1}$ based on the Equation (6) and (7). In stage 3, Gbps selects the set $\mathcal{D}^*_S$ of size N_r from set $\mathcal{D}^*_{S_2}$, predicted boxes $\widehat{\mathcal{B}}$ and box point densities ϕ via Equation (8), (9). The final set $\mathcal{D}^*_S$ at this round is then annotated by an oracle $\mathbf{\Omega}$ and merged with the selected set in the previous round as the training data. Notably, the selected set at the 0-th round is $\mathcal{D}_L$. When the training data is determined, we re-train the 3D detector with the merged selected set until the model is converged. We iterate the above process starting with model inference for R rounds and add N_r queried samples to the selected set $\mathcal{D}_S$ for each round.

E More Experimental Results on KITTI

E.1 AL performance comparisons on Easy difficulty level

In addition to the Moderate and Hard difficulties reported in the body text, we provide the additional quantitative analysis *w.r.t.* the Easy mode. Figure 1 depicts the mAP(%) variation of

the baselines against the proposed CRB with an increasing number of selected bounding boxes. The solid lines indicate the mean value from three running trials and the standard deviation are shown in the shaded area. The results indicate that with increasing annotation cost, CRB consistently achieves the highest mAP and outperforms the state-of-the-art active learning approaches on both 3D and BEV views. Note that CRB with only 1k boxes selected for annotation reaches the comparable performance of RAND that selects around 3k boxes. Other AL baselines share the same trend as the ones under the difficulties of MODERATE and HARD (reported in Figure 2 of the main paper).

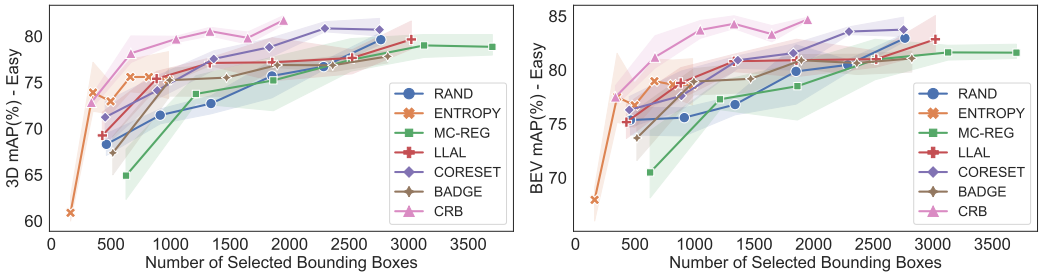

Figure 1: 3D and BEV mAP (%) of CRB and AL baselines on the KITTI *val* split at the EASY level.

E.2 AL PERFORMANCE COMPARISONS FOR EACH CLASS

To investigate the effectiveness of AL strategies on detecting specific classes, we plot the results of Cyclist and Pedestrian at all difficulty levels in Figure 2 (3D AP) and Figure 3 (BEV AP). We mainly compare three aspects: performance, annotation cost and error variance. 1)Performance: the plots in Figure 2 and Figure 3 show that the proposed CRB outperforms all state-of-the-art AL methods by a noticeable margin, for all settings of difficulty, classes and views, except at easy cyclist. This evidences that our proposed AL approach explores samples with more conceptual semantics covering test sets so that the detector tends to perform better on more challenging samples. 2) Annotation cost: all the plots consistently demonstrate that the proposed CRB reaches comparable performance while requiring very few ($\sim$1/3) annotation costs as baselines, except ENTROPY. ENTROPY takes the minimal annotation cost, yet its result is inferior, especially for difficult classes like Cyclist. 3) Variance: we observe that AP variance of CRB is lower than all baselines, which shows that our method is less sensitive to randomness and more stable to produce expected results.

E.3 AL PERFORMANCE COMPARISONS FOR EACH ACTIVE SELECTION ROUND

Figure 4 compares the performance variation of the AL baselines against the proposed CRB with the increasing percentage of queried point clouds (from 2.7% to 16.2%). The reported performance is mAP scores (%) $\pm$ the standard deviation of three trials for both 3D view (top row) and BEV view (bottom row) and all difficulty levels. We clearly observe that our method CRB consistently outperforms the state-of-the-art results, irrespective of percentage of annotated point clouds and difficulty settings. Surprisingly, when the annotation costs reaches 16.2%, RAND strategy outperforms all the baselines at the MODERATE and HARD level. This implicitly evidences that existing uncertainty and diversity-based AL strategies fail to select samples that are aligned with test cases.

F MORE EXPERIMENTAL RESULTS ON WAYMO

To explore the performance for different classes on the Waymo dataset, we plot the AP(%) variation of Cyclist and Pedestrian yielded by the baselines and CRB with increasing annotated bounding boxes in Figure 5. We present the results at two levels of difficulty officially defined by Waymo. LEVEL 1 (and LEVEL 2) indicates there are more than five inside points (at least one point) of the ground-truth objects. As can be observed by the AP curves in the plots, CRB achieves the superiror recognition accuracy when the annotation cost comes to $\sim$ 45k bounding boxes. Specifically, the AP values of CRB are boosted by the largest margin (3.1% on LEVEL 2 Cyclist and 1.6% on LEVEL 2 Pedestrian) over the best performing baseline (RAND) that takes extra cost of 5k bounding boxes

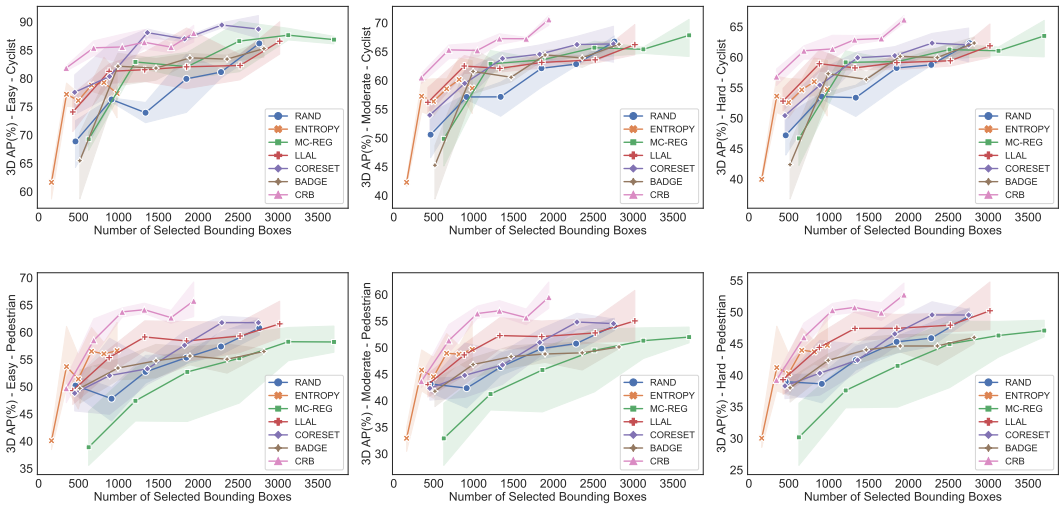

Figure 2: Detection results of different classes on the KITTI *val* set (3D view) with an increasing number of queried bounding boxes.

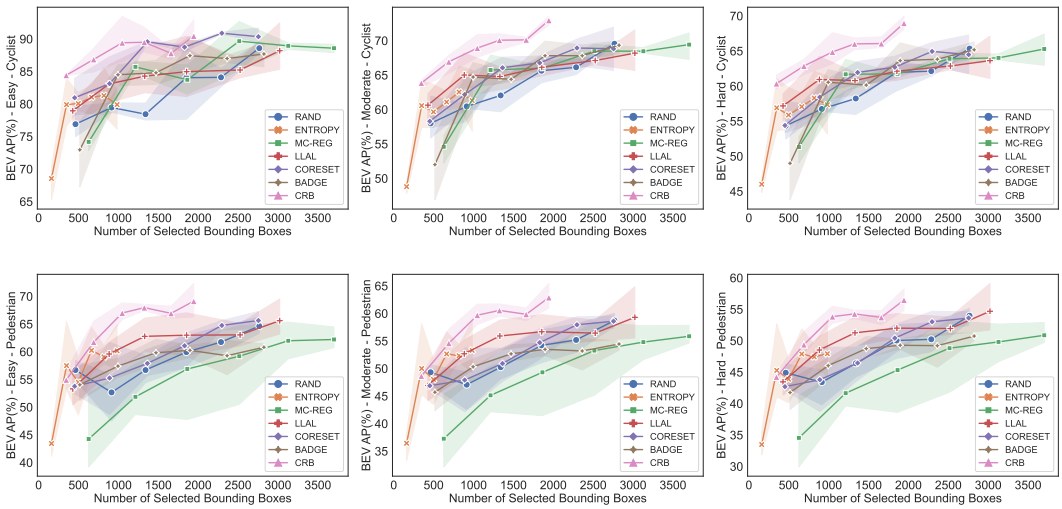

Figure 3: Detection results of different classes on the KITTI *val* set (BEV view) with increasing number of queried bounding boxes.

than ours. Surprisingly, note the results on the class of Pedestrian, the AP curves of most baselines except ENTROPY and LLAL are bounded by RAND. The AP curves of ENTROPY and LLAL are bounded by CRB with the increasing cost to 15k $\sim$ 20k bounding boxes. This confirms the CRB's superiority over compared AL baselines. Besides, the boosted margin achieved by CRB set for LEVEL 2 Pedestrian is larger than set for LEVEL 1 Pedestrian. This indicates that the samples selected by CRB matches well with the data at the time, covering more diverse samples that span different difficulties.

G ADDITIONAL QUALITATIVE ANALYSIS

To intuitively demonstrate the benefits of our proposed active 3D detection strategy, Figure 6 visualizes that the 3D detection results produced by **RAND** (bottom left) and **CRB** selection (bottom right) from the corresponding image (upper row). Both 3D detectors are trained under the budget of 1K annotated bounding boxes. False positives and corrected predictions are indicated with red and green boxes. It is observed that, under the same condition, CRB produces more accurate and more confident predictions than RAND. Specifically, our CRB yields accurate predictions for multiple pedestrians on the right sidewalk, while RAND fails. Besides, note the car parked on the left that is

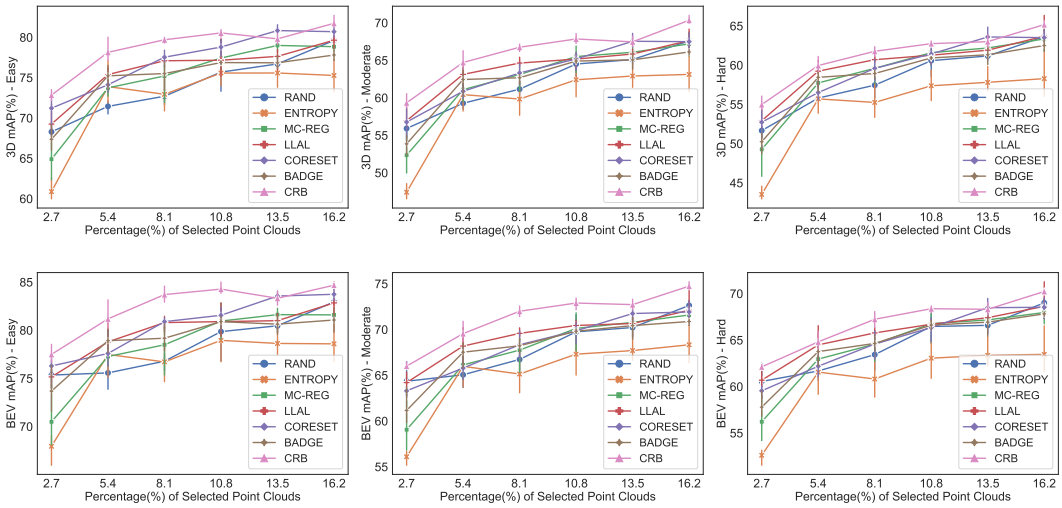

Figure 4: Results on KITTI datasets with an increasing percentage of queried point clouds.

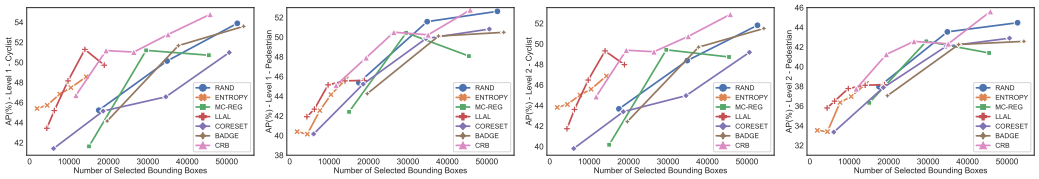

Figure 5: Results of CRB and baselines on the Waymo *val* splt for different classes at Level 2.

highlighted in the orange box in Figure 6, the detector trained with RAND produces a significantly lower confidence score (0.62) compared to our approach (0.95). This validates that the point clouds selected by CRB are aligned more tightly with the test samples.

H ADDITIONAL RESULTS FOR PARAMETER SENSITIVITY ANALYSIS

Sensitivity to Prototype Selection. To further analyze the sensitivity of performance to different prototype selection approaches, *i.e.*, GMM, K-MEANS, and K-MEANS++, we show more results on BEV views in Figure 7 (right). We again run two trials for each prototype selection method and plot the mean and the variance bars. Note that there is very little difference (1.65% in the last round) in the mAP(%) of our approach when using different prototype selection methods. This evidences that the more performance gains achieved by CRB than existing baselines do not depend on choosing the prototype selection method.

Sensitivity to Bandwidth h. Figure 7 shows additional results w.r.t the BEV views of CRB with the bandwidth h varying in $\{3, 5, 7, 9\}$. Observing the trends of four curves, CRB with the bandwidth of all values yields consistent results within the 1.7% variation. This demonstrates that the CRB is insensitive to different values set for bandwidth and can produce similar mAP(%) on BEV views.

I RELATED WORK

Generic Active Learning. For a comprehensive review of classic active learning methods and their applications, we refer readers to (Ren et al., 2021). Most active learning approaches were tailored for image classification task, where the *uncertainty* (Wang & Shang, 2014; Lewis & Catlett, 1994; Joshi et al., 2009; Roth & Small, 2006; Parvaneh et al., 2022; Du et al., 2021; Kim et al., 2021b; Bhatnagar et al., 2021) and *diversity* (Sener & Savarese, 2018; Elhamifar et al., 2013; Guo, 2010;

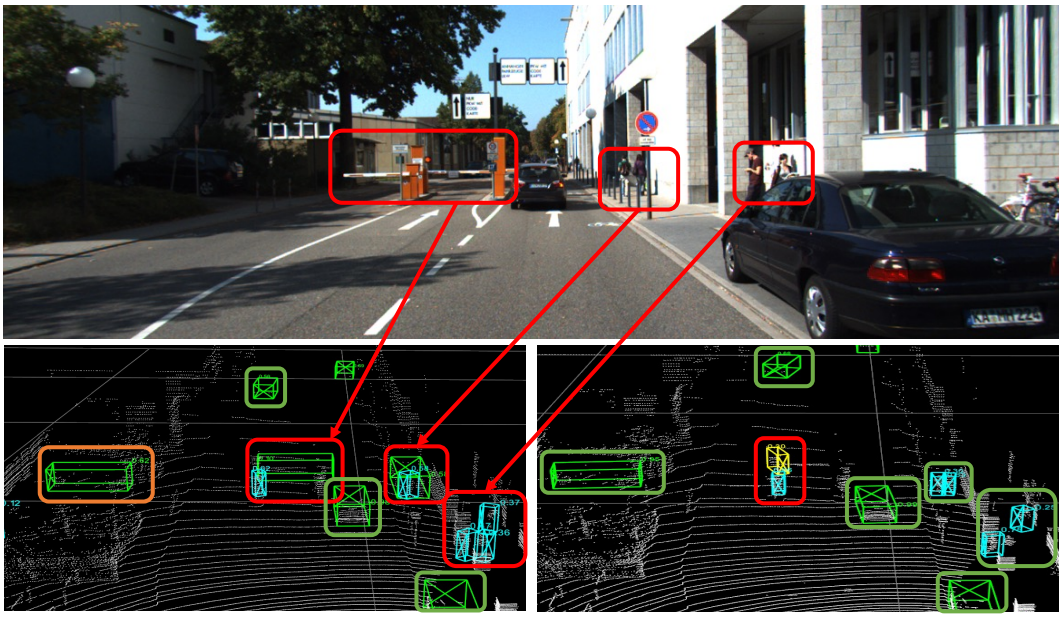

Figure 6: Another case study of active 3D detection performance of **RAND** (bottom left) and **CRB** (bottom right) under the budge of 1,000 annotated bounding boxes. False positive (corrected predictions) are highlighted in red (green) boxes. The orange box denotes the detection with low confidence.

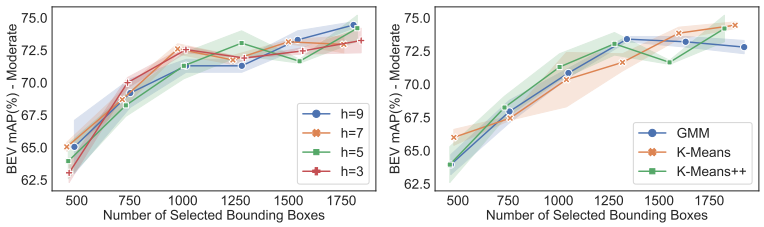

Figure 7: Performance comparison on KITTI *val* set with varying KDE bandwidth h (left) and prototype selection approaches (right) with increasing queried bounding boxes.

Yang et al., 2015; Nguyen & Smeulders, 2004; Hasan & Roy-Chowdhury, 2015; Aodha et al., 2014) of samples are measured as the acquisition criteria. The hybrid works (Kim et al., 2021a; Citovsky et al., 2021; Ash et al., 2020; MacKay, 1992; Liu et al., 2021; Kirsch et al., 2019; Houlsby et al., 2011) combine both paradigms such as by measuring uncertainty as to the gradient magnitude (Ash et al., 2020) at the final layer of neural networks and selecting gradients that span a diverse set of directions. In addition to the above two mainstream methods, (Settles et al., 2007; Roy & McCallum, 2001; Freytag et al., 2014; Yoo & Kweon, 2019) estimate the expected model changes or predicted losses as the sample importance.

Active Learning for 2D Detection. Lately, the attention of AL has shifted from image classification to the task of object detection (Siddiqui et al., 2020; Li & Yin, 2020). Early work (Roy et al., 2018) exploits the detection inconsistency of outputs among different convolution layers and leverages the query by committee approach to select informative samples. Concurrent work (Kao et al., 2018) introduces the notion of localization tightness as the regression uncertainty, which is calculated by the overlapping area between region proposals and the final predictions of bounding boxes. Other uncertainty-based methods attempt to aggregate pixel-level scores for each image (Aghdam et al., 2019), reformulate detectors by adding Bayesian inference to estimate the uncertainty (Harakeh et al., 2020) or replace conventional detection head with the Gaussian mixture model to compute aleatoric and epistemic uncertainty (Choi et al., 2021). A hybrid method (Wu et al., 2022) considers image-level uncertainty calculated by entropy and instance-level diversity measured by the similarity to the prototypes. Lately, AL technique is leveraged for transfer learning by selecting a few uncertain labeled source bounding boxes with high transferability to the target domain, where

the transferability is defined by domain discriminators (Tang et al., 2021b; Al-Saffar et al., 2021). Inspired by neural architecture searching, Tang et al. (2021a) adopted the 'swap-expand' strategy to seek a suitable neural architecture including depth, resolution, and receptive fields at each active selection round. Recently, some works augment the weakly-supervised object detection (WSOD) with an active learning scheme. In WSOD, only image-level category labels are available during training. Some conventional AL methods such as predicted probability, probability margin are explored in (Wang et al., 2022), while in (Vo et al., 2022), "box-in-box" is introduced to select images where two predicted boxes belong to the same category and the small one is "contained" in the larger one. Nevertheless, it is not trivial to adapt all existing AL approaches for 2D detection as the ensemble learning and network modification leads to more model parameters to learn, which could be hardly affordable for 3D tasks.

Active Learning for 3D Detection. Active learning for 3D object detection has been relatively under-explored than other tasks, potentially due to its large-scale nature. Most existing works (Feng et al., 2019; Schmidt et al., 2020) simply apply the off-the-shelf generic AL strategies and use hand-crafted heuristics including Shannon entropy (Wang & Shang, 2014), ensemble (Beluch et al., 2018), localization tightness (Kao et al., 2018) and MC-DROPOUT (Gal & Ghahramani, 2016) for 3D detection learning. However, the abovementioned solutions base on the cost of labelling point clouds rather than the number of 3D bounding boxes, which inherently being biased to the point clouds containing more objects. However, in our work, the proposed CRB greedily search for the unique point clouds while maintaining the same marginal distribution for generalization, which implicitly quires objects to annotate without repetition and save labeling costs.

Active Learning for 3D Semantic Segmentation. The adoption of active learning techniques has successfully reduced the significant burden of point-by-point human labeling in large-scale point cloud datasets. Super-point (Shi et al., 2021) is introduced to represent a spectral clustering containing points which are most likely belonging to the same category, then only super-points with high score are labeled at each round. An improved work Shao et al. (2022) further encoded the super-points with a graph neural network, where the edges denote distance between super-points, and then projects the super-point features into the diversity space to select the most representative super-points. Another streaming of work (Wu et al., 2021) is to obtain point labels for uncertain and diverse regions to prevent the high cost of labeling the entire point cloud. Although semantic segmentation and object detection are different vision tasks, both can benefit from active learning to substantially alleviate the manual labelling cost.

Connections to Semi-supervised Active Learning. Aiming at unifying unlabeled sample selection and model training, the concept of semi-supervised active learning (Drugman et al., 2016; Rhee et al., 2017; Sinha et al., 2019; Gao et al., 2020; Liu et al., 2021; Kim et al., 2021a;b; Zhang & Plank, 2021; Guo et al., 2021; Caramalau et al., 2021; Citovsky et al., 2021; Elezi et al., 2022; Gudovskiy et al., 2020) has been raised. (Drugman et al., 2016) combines the semi-supervised learning (SSL) and active learning (AL) for speech understanding that leverages the confidence score obtained from the posterior probabilities of decoded texts. (Sener & Savarese, 2018) incorporated a Ladder network for SSL during AL cycles, while the performance gains are marginal compared to the supervised counterpart. (Sinha et al., 2019) trained a variational adversarial active learning (VAAL) model with both labeled and unlabeled data points, where the discriminator is able to estimate how representative each sample is from the pool. (Elezi et al., 2022) proposed a combined strategy for training 2D object detection, which queries samples of high uncertainty and low robustness for supervised learning and takes full advantage of easy samples via auto-labeling. As our work is under the umbrella of the pool-based active learning, accessible unlabeled data are not used for model training in our setting, hereby the semi-supervised active learning algorithms were not considered in experimental comparisons.

REFERENCES

Hamed H. Aghdam, Abel Gonzalez-Garcia, Joost van de Weijer, and Antonio M. Lopez. Active learning for deep detection neural networks. In *Proc. International Conference on Computer Vision (ICCV)*, pp. 3672–3680, 2019.

Ahmed Al-Saffar, Alina Bialkowski, Mahsa Baktashmotlagh, Adnan Trakic, Lei Guo, and Amin M. Abbosh. Closing the gap of simulation to reality in electromagnetic imaging of brain strokes via

deep neural networks. *IEEE Transactions on Computational Imaging*, 7:13–21, 2021.

Oisin Mac Aodha, Neill D. F. Campbell, Jan Kautz, and Gabriel J. Brostow. Hierarchical subquery evaluation for active learning on a graph. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 564–571, 2014.

Jordan T. Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. Deep batch active learning by diverse, uncertain gradient lower bounds. In *Proc. International Conference on Learning Representations (ICLR)*, 2020.

William H. Beluch, Tim Genewein, Andreas Nürnberger, and Jan M. Köhler. The power of ensembles for active learning in image classification. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9368–9377, 2018.

Shubhang Bhatnagar, Sachin Goyal, Darshan Tank, and Amit Sethi. PAL : Pretext-based active learning. In *Proc. British Machine Vision Conference (BMVC)*, pp. 195. BMVA Press, 2021.

Razvan Caramalau, Binod Bhattarai, and Tae-Kyun Kim. Sequential graph convolutional network for active learning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9583–9592, 2021.

Jiwoong Choi, Ismail Elezi, Hyuk-Jae Lee, Clément Farabet, and Jose M. Alvarez. Active learning for deep object detection via probabilistic modeling. In *Proc. International Conference on Computer Vision (ICCV)*, pp. 10244–10253, 2021.

Gui Citovsky, Giulia DeSalvo, Claudio Gentile, Lazaros Karydas, Anand Rajagopalan, Afshin Rostamizadeh, and Sanjiv Kumar. Batch active learning at scale. In *Proc. Annual Conference on Neural Information Processing (NeurIPS)*, pp. 11933–11944, 2021.

Thomas Drugman, Janne Pylkkönen, and Reinhard Kneser. Active and semi-supervised learning in ASR: benefits on the acoustic and language models. In Nelson Morgan (ed.), *Interspeech Annual Conference of the International Speech Communication Association*, pp. 2318–2322, 2016.

Pan Du, Suyun Zhao, Hui Chen, Shuwen Chai, Hong Chen, and Cuiping Li. Contrastive coding for active learning under class distribution mismatch. In *Proc. International Conference on Computer Vision (ICCV)*, pp. 8907–8916, 2021.

Ismail Elezi, Zhiding Yu, Anima Anandkumar, Laura Leal-Taixe, and Jose M Alvarez. Not all labels are equal: Rationalizing the labeling costs for training object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14492–14501, 2022.

Ehsan Elhamifar, Guillermo Sapiro, Allen Y. Yang, and S. Shankar Sastry. A convex optimization framework for active learning. In *Proc. International Conference on Computer Vision (ICCV)*, pp. 209–216, 2013.

Di Feng, Xiao Wei, Lars Rosenbaum, Atsuto Maki, and Klaus Dietmayer. Deep active learning for efficient training of a lidar 3d object detector. In *Proc. Intelligent Vehicles Symposium, (IV)*, pp. 667–674, 2019.

Alexander Freytag, Erik Rodner, and Joachim Denzler. Selecting influential examples: Active learning with expected model output changes. In *Proc. European Conference on Computer Vision (ECCV)*, pp. 562–577, 2014.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proc. International Conference on Machine Learning (ICML)*, volume 48, pp. 1050–1059, 2016.

Mingfei Gao, Zizhao Zhang, Guo Yu, Sercan Ömer Arik, Larry S. Davis, and Tomas Pfister. Consistency-based semi-supervised active learning: Towards minimizing labeling cost. In *Proc. European Conference on Computer Vision (ECCV)*, volume 12355, pp. 510–526, 2020.

Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3354–3361, 2012.

Denis A. Gudovskiy, Alec Hodgkinson, Takuya Yamaguchi, and Sotaro Tsukizawa. Deep active learning for biased datasets via fisher kernel self-supervision. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9038–9046, 2020.

Jiannan Guo, Haochen Shi, Yangyang Kang, Kun Kuang, Siliang Tang, Zhuoren Jiang, Changlong Sun, Fei Wu, and Yueting Zhuang. Semi-supervised active learning for semi-supervised models: Exploit adversarial examples with graph-based virtual labels. In *Proc. International Conference on Computer Vision (ICCV)*, pp. 2876–2885. IEEE, 2021.

Yuhong Guo. Active instance sampling via matrix partition. In *Proc. Annual Conference on Neural Information Processing (NeurIPS)*, pp. 802–810, 2010.

Ali Harakeh, Michael Smart, and Steven L. Waslander. Bayesod: A bayesian approach for uncertainty estimation in deep object detectors. In *Proc. International Conference on Robotics and Automation (ICRA)*, pp. 87–93, 2020.

Mahmudul Hasan and Amit K. Roy-Chowdhury. Context aware active learning of activity recognition models. In *Proc. International Conference on Computer Vision (ICCV)*, pp. 4543–4551, 2015.

Neil Houlsby, Ferenc Huszar, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. *CoRR*, abs/1112.5745, 2011.

Ajay J. Joshi, Fatih Porikli, and Nikolaos Papanikolopoulos. Multi-class active learning for image classification. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2372–2379, 2009.

Chieh-Chi Kao, Teng-Yok Lee, Pradeep Sen, and Ming-Yu Liu. Localization-aware active learning for object detection. In *Proc. Asian Conference on Computer (ACCV)*, pp. 506–522, 2018.

Kwanyoung Kim, Dongwon Park, Kwang In Kim, and Se Young Chun. Task-aware variational adversarial active learning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8166–8175, 2021a.

Yoon-Yeong Kim, Kyungwoo Song, JoonHo Jang, and Il-Chul Moon. LADA: look-ahead data acquisition via augmentation for deep active learning. In *Proc. Annual Conference on Neural Information Processing (NeurIPS)*, pp. 22919–22930, 2021b.

Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. In *Proc. Annual Conference on Neural Information Processing (NeurIPS)*, pp. 7024–7035, 2019.

David D. Lewis and Jason Catlett. Heterogeneous uncertainty sampling for supervised learning. In *Proc. International Conference on Machine Learning (ICML)*, pp. 148–156, 1994.

Haohan Li and Zhaozheng Yin. Attention, suggestion and annotation: A deep active learning framework for biomedical image segmentation. In Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz (eds.), *Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 12261, pp. 3–13, 2020.

Zhuoming Liu, Hao Ding, Huaping Zhong, Weijia Li, Jifeng Dai, and Conghui He. Influence selection for active learning. In *Proc. International Conference on Computer Vision (ICCV)*, pp. 9254–9263, 2021.

David J. C. MacKay. Information-based objective functions for active data selection. *Journal of Neural Computation*, 4(4):590–604, 1992.

Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Domain adaptation: Learning bounds and algorithms. In *Proc. Conference on Learning Theory (COLT)*, 2009.

Hieu Tat Nguyen and Arnold W. M. Smeulders. Active learning using pre-clustering. In Carla E. Brodley (ed.), *Proc. International Conference on Machine Learning (ICML)*, 2004.

Amin Parvaneh, Ehsan Abbasnejad, Damien Teney, Gholamreza (Reza) Haffari, Anton van den Hengel, and Javen Qinfeng Shi. Active learning by feature mixing. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12237–12246, 2022.

Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B. Gupta, Xiaojiang Chen, and Xin Wang. A survey of deep active learning. *ACM Computing Survey*, 54(9):40, 2021.

Phill-Kyu Rhee, Enkhbayar Erdenee, Shin Dong Kyun, Minhaz Uddin Ahmed, and SongGuo Jin. Active and semi-supervised learning for object detection with imperfect data. *Cognition System Research*, 45:109–123, 2017.

Dan Roth and Kevin Small. Margin-based active learning for structured output spaces. In *Proc. European Conference on Machine Learning (ECML)*, pp. 413–424, 2006.

Nicholas Roy and Andrew McCallum. Toward optimal active learning through monte carlo estimation of error reduction. In *Proc. International Conference on Machine Learning (ICML)*, pp. 441–448, 2001.

Soumya Roy, Asim Unmesh, and Vinay P. Namboodiri. Deep active learning for object detection. In *Proc. British Machine Vision Conference (BMVC)*, pp. 91, 2018.

Sebastian Schmidt, Qing Rao, Julian Tatsch, and Alois C. Knoll. Advanced active learning strategies for object detection. In *Proc. Intelligent Vehicles Symposium, (IV)*, pp. 871–876, 2020.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. In *Proc. International Conference on Learning Representations (ICLR)*, 2018.

Burr Settles, Mark Craven, and Soumya Ray. Multiple-instance active learning. In *Proc. Annual Conference on Neural Information Processing (NeurIPS)*, pp. 1289–1296, 2007.

Feifei Shao, Yawei Luo, Ping Liu, Jie Chen, Yi Yang, Yulei Lu, and Jun Xiao. Active learning for point cloud semantic segmentation via spatial-structural diversity reasoning. In *Proc. International Conference on Multimedia (MM)*, pp. 2575–2585, 2022.

Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. PV-RCNN: point-voxel feature set abstraction for 3d object detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10526–10535, 2020.

Xian Shi, Xun Xu, Ke Chen, Lile Cai, Chuan Sheng Foo, and Kui Jia. Label-efficient point cloud semantic segmentation: An active learning approach. *CoRR*, abs/2101.06931, 2021.

Yawar Siddiqui, Julien Valentin, and Matthias Nießner. Viewal: Active learning with viewpoint entropy for semantic segmentation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9430–9440, 2020.

Samarth Sinha, Sayna Ebrahimi, and Trevor Darrell. Variational adversarial active learning. In *Proc. International Conference on Computer Vision (ICCV)*, pp. 5971–5980, 2019.

Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2443–2451, 2020.

Fuhui Tang, Chenhan Jiang, Dafeng Wei, Hang Xu, Andi Zhang, Wei Zhang, Hongtao Lu, and Chunjing Xu. Towards dynamic and scalable active learning with neural architecture adaption for object detection. In *Proc. British Machine Vision Conference (BMVC)*, 2021a.

Ying-Peng Tang, Xiu-Shen Wei, Borui Zhao, and Sheng-Jun Huang. Qbox: Partial transfer learning with active querying for object detection. *Journal of IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2021b. doi: 10.1109/TNNLS.2021.3111621.

Huy V Vo, Oriane Siméoni, Spyros Gidaris, Andrei Bursuc, Patrick Pérez, and Jean Ponce. Active learning strategies for weakly-supervised object detection. In *Proc. European Conference on Computer Vision (ECCV)*, pp. 211–230, 2022.

Dan Wang and Yi Shang. A new active labeling method for deep learning. In *Proc. International Joint Conference on Neural Networks (IJCNN)*, pp. 112–119, 2014.

Xiao Wang, Xiang Xiang, Baochang Zhang, Xuhui Liu, Jianying Zheng, and QingLei Hu. Weakly supervised object detection based on active learning. *Journal of Neural Processing Letters*, pp. 1–15, 2022.

Jiaxi Wu, Jiaxin Chen, and Di Huang. Entropy-based active learning for object detection with progressive diversity constraint. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9387–9396. IEEE, 2022.

Tsung-Han Wu, Yueh-Cheng Liu, Yu-Kai Huang, Hsin-Ying Lee, Hung-Ting Su, Ping-Chia Huang, and Winston H. Hsu. Redal: Region-based and diversity-aware active learning for point cloud semantic segmentation. In *Proc. International Conference on Computer Vision (ICCV)*, pp. 15490–15499, 2021.

Yi Yang, Zhigang Ma, Feiping Nie, Xiaojun Chang, and Alexander G. Hauptmann. Multi-class active learning by uncertainty sampling with diversity maximization. *International Journal of Computer Vision*, 113:113–127, 2015.

Donggeun Yoo and In So Kweon. Learning loss for active learning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 93–102, 2019.

Mike Zhang and Barbara Plank. Cartography active learning. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (eds.), *Proc. Findings of the Association for Computational Linguistics (EMNLP)*, pp. 395–406. Association for Computational Linguistics, 2021.

Yifan Zhang, Qingyong Hu, Guoquan Xu, Yanxin Ma, Jianwei Wan, and Yulan Guo. Not all points are equal: Learning highly efficient point-based detectors for 3d lidar point clouds. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18953–18962, 2022.