Monte Carlo Dropout for Object Detection on Point Clouds

Guided Research

Supervised by:

Prof. Angela Dai

Kerem Yildirir

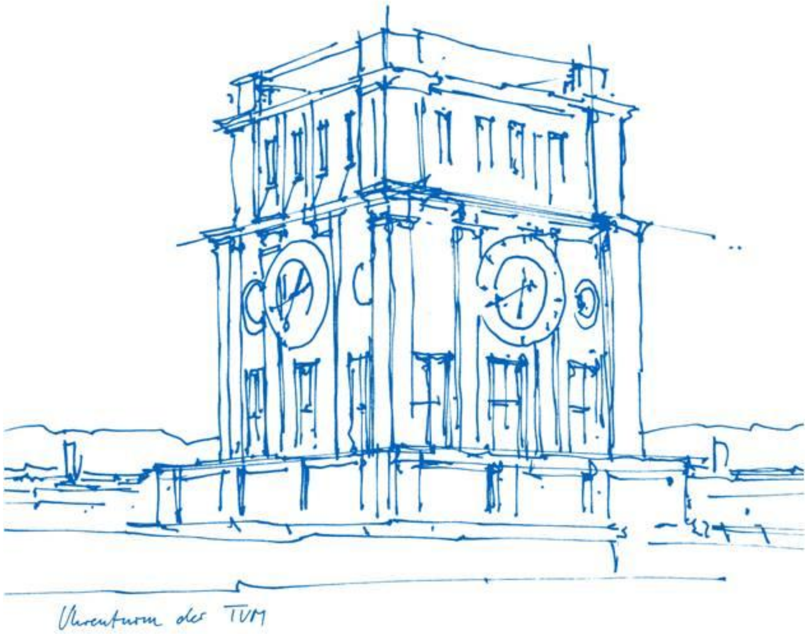
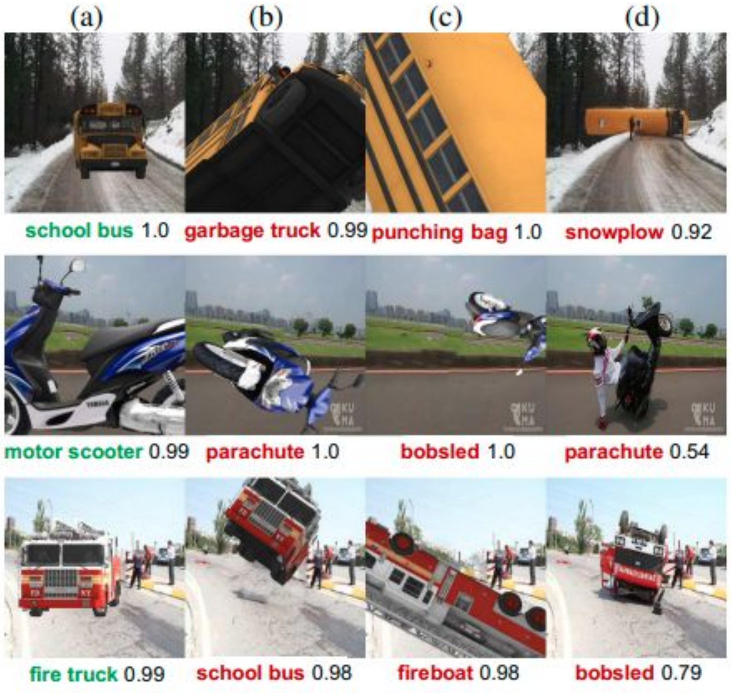


Table of Contents

Motivation

Deep learning methods can make wrong but very confident predictions



[Alcorn et al. 2018]

Motivation

Deep learning methods can make wrong but very confident predictions

Google apologizes for algorithm mistakenly calling black people 'gorillas'

Image source [CNet, 2015]

Tesla driver dies in first fatal crash while using autopilot mode

The autopilot sensors on the Model S failed to distinguish a white tractor-trailer crossing the highway against a bright sky

Image source [The Guardian, 2016]

Motivation

- Leveraging uncertainty information for 3D scene understanding is an underexplored area
- Overconfident predictions can cause catastrophic consequences, it is critical in real life scenarios to be aware of the uncertain situations.
- Train models that are able recognize when they are likely to make mistakes.

Related Work: 3D Scene Understanding

- Deep Hough Voting for 3D Object Detection in Point Clouds [Qi et al, 2019]
- Mix3D: Out-of-Context Data Augmentation for 3D Scenes [Nekrasov et al, 2021]
- RfD-Net: Point Scene Understanding by Semantic Instance Reconstruction[Nie et al, 2021]
- TransformerFusion: Monocular RGB Scene Reconstruction using Transformers [Božič et al., 2021]

Related Work: Uncertainty

Uncertainty

Epistemic Uncertainty

- Uncertainty of the model
- Can be explained away with more data

Aleatoric Uncertainty

- Uncertainty of the data, usually inherent noise from the sensor
- Cannot be explained away with more data

Related Work: Epistemic Uncertainty in Deep Learning

- In order to capture epistemic uncertainty in a neural network, we put a prior distribution (usually Gaussian) over its weights and instead of optimizing weights directly, we average over all possible weights (marginalization) during inference, these models are referred as Bayesian neural networks. [Kendall et al. 2017]

Related Work: Bayesian Neural Networks (BNN)

- Performing inference in BNN's is done by applying the Bayes' Rule for evaluating the posterior probability $p(\mathbf{W} \mid \mathbf{X}, \mathbf{Y}) = p(\mathbf{Y} \mid \mathbf{X}, \mathbf{W}) p(\mathbf{W}) / p(\mathbf{Y} \mid \mathbf{X})$
- This is nice to formulate, but often intractable to do in practice, because we need to evaluate marginal probability $p(Y|X)$ over all the parameters.
- Hence, in order to perform inference in BNN's we need to approximate true posterior.

Related Work: Monte Carlo (MC) Dropout [3]

- Existing approximation methods follow the strategy of defining a simple distribution that fits into the posterior distribution, and then optimizing the simple distribution instead of the true distribution [Kendall et al, 2017].
- [Gal et al. 2016] Has shown that performing dropout at test time is equivalent to approximating the posterior distribution with a Gaussian which minimized the distance between the distributions, and that we can easily compute the uncertainty of complex and deep models without having to tackle an intractable problem.

Related Work: Epistemic Uncertainty in Classification

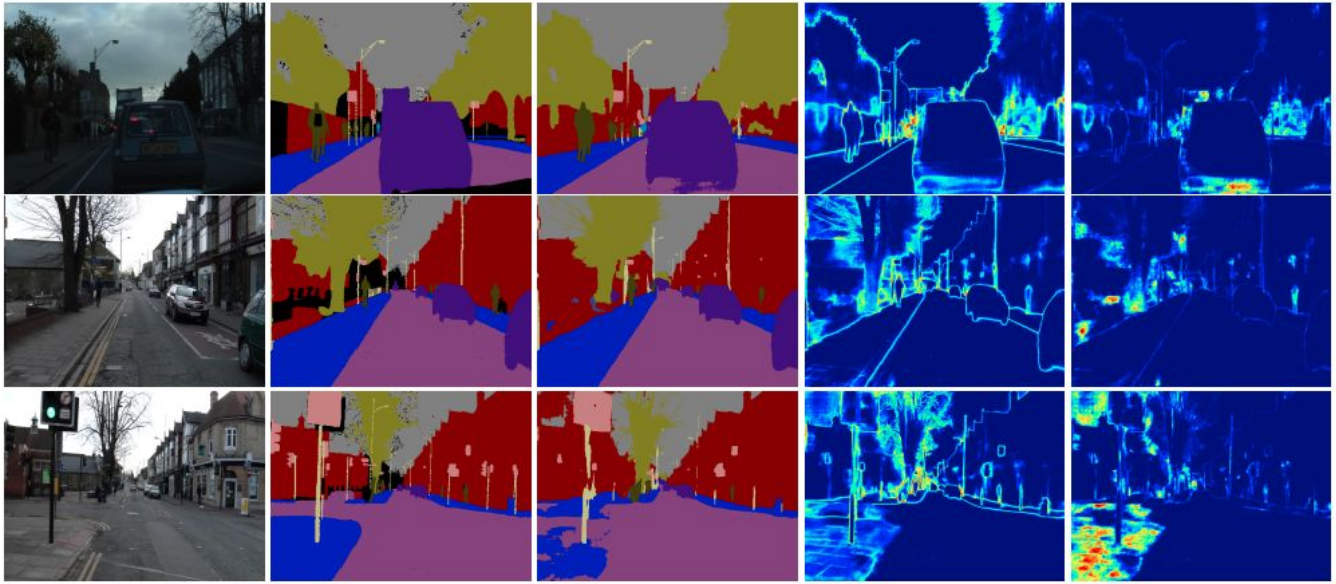
- Expected Entropy
  - High when there's ambiguity in the decisions
  - MC approximation: $\text{Expected Entropy} \sim \frac{1}{T}\sum_{t=1}^{t}(H(f_t^W(x)))$

Related Work: Epistemic Uncertainty in Classification

- Predictive Entropy
  - Entropy of the expected prediction
  - High when there's ambiguity
  - High when the predictions far away from the data

- Mutual Information
  - Mutual Information = Predictive Entropy - Expected Entropy
  - Only high when far away from data

Related Work: What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?



(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

Image source [Kendall et al., 2017]

Related Work: Neural RGB→D Sensing: Depth and Uncertainty from a Video Camera

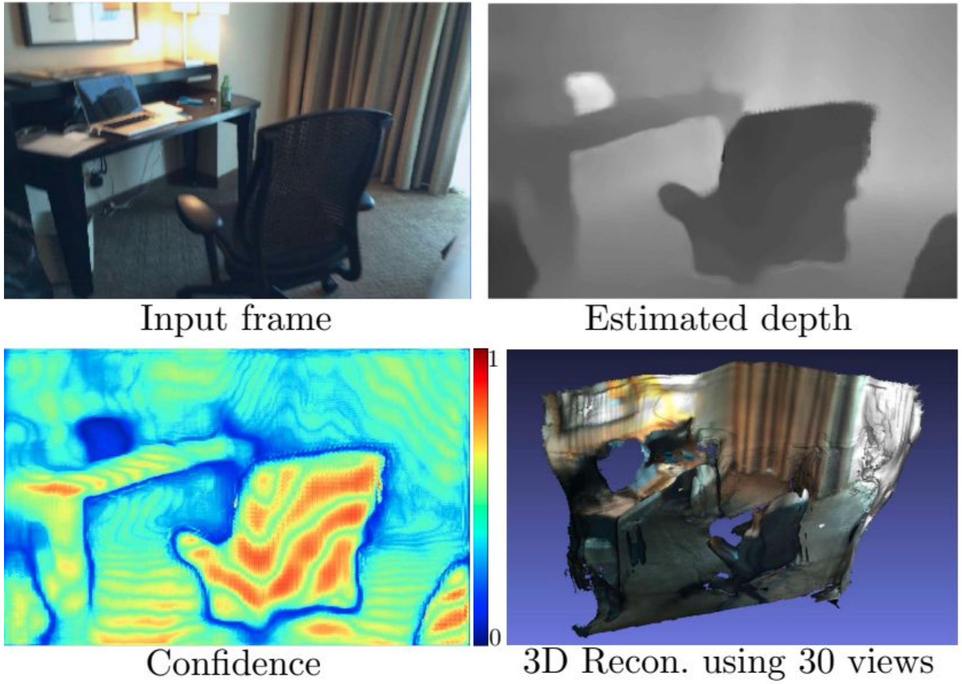


Image source [Liu et al, 2019]

Related Work: Neural RGB→D Sensing: Depth and Uncertainty from a Video Camera

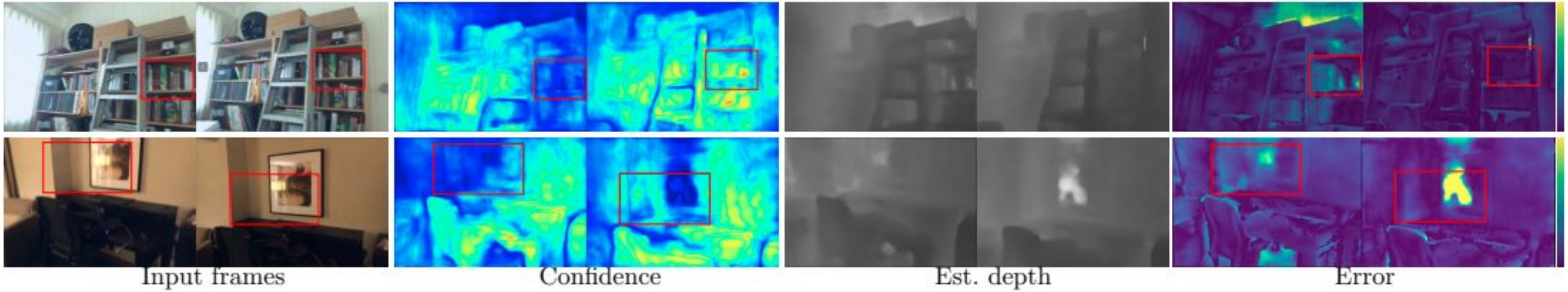


Image source [Liu et al, 2019]

Related Work: D3VO: Deep Depth, Deep Pose and Deep Uncertainty for Monocular Visual Odometry

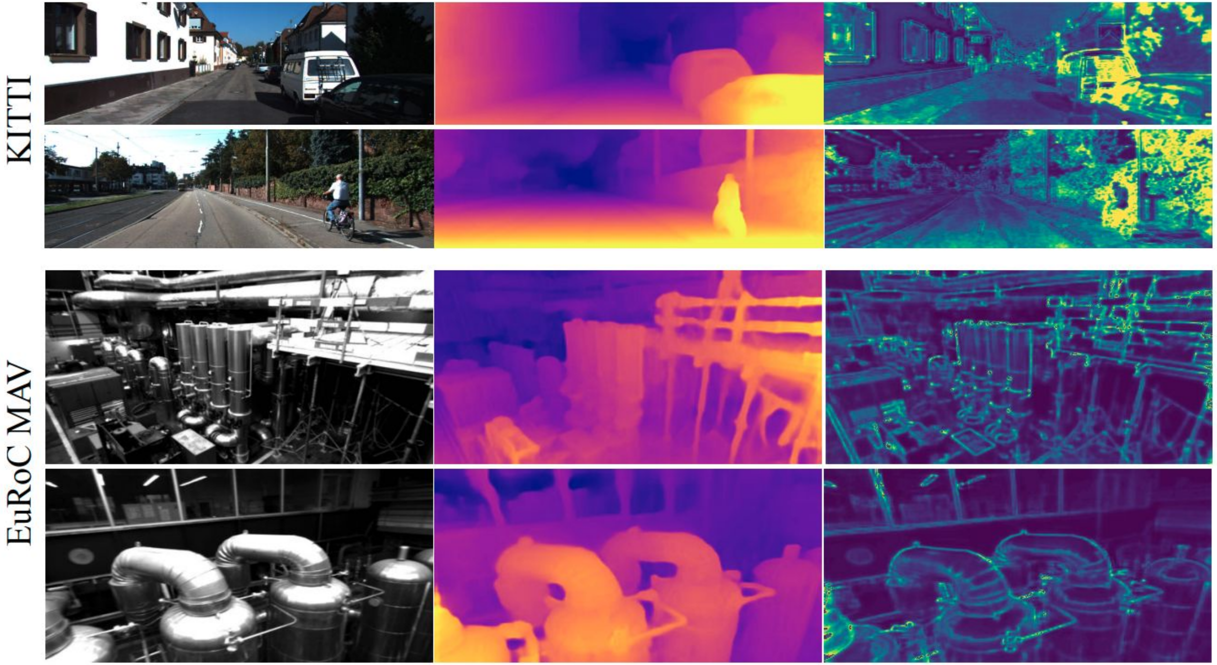


Image source [Yang et al, 2020]

Related Work: D3VO: Deep Depth, Deep Pose and Deep Uncertainty for Monocular Visual Odometry

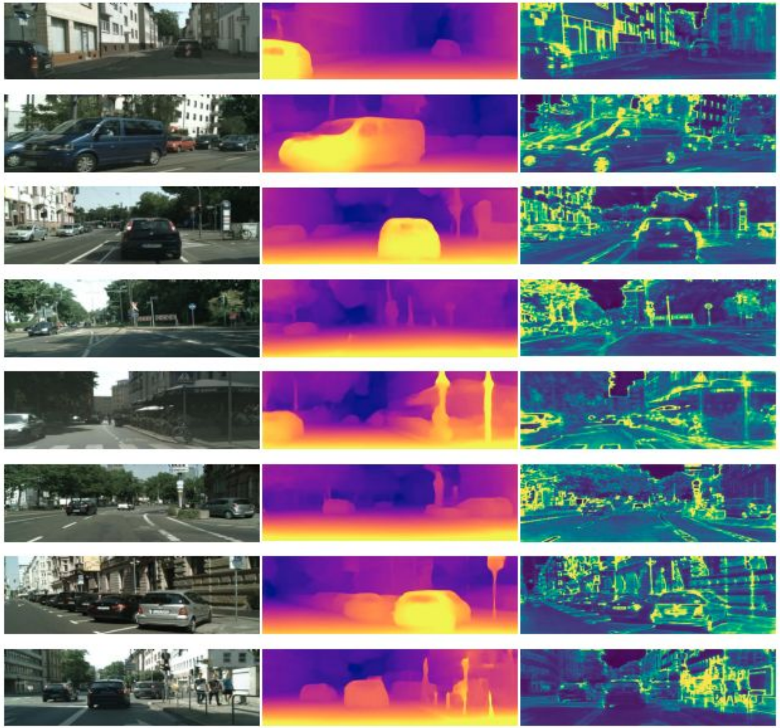
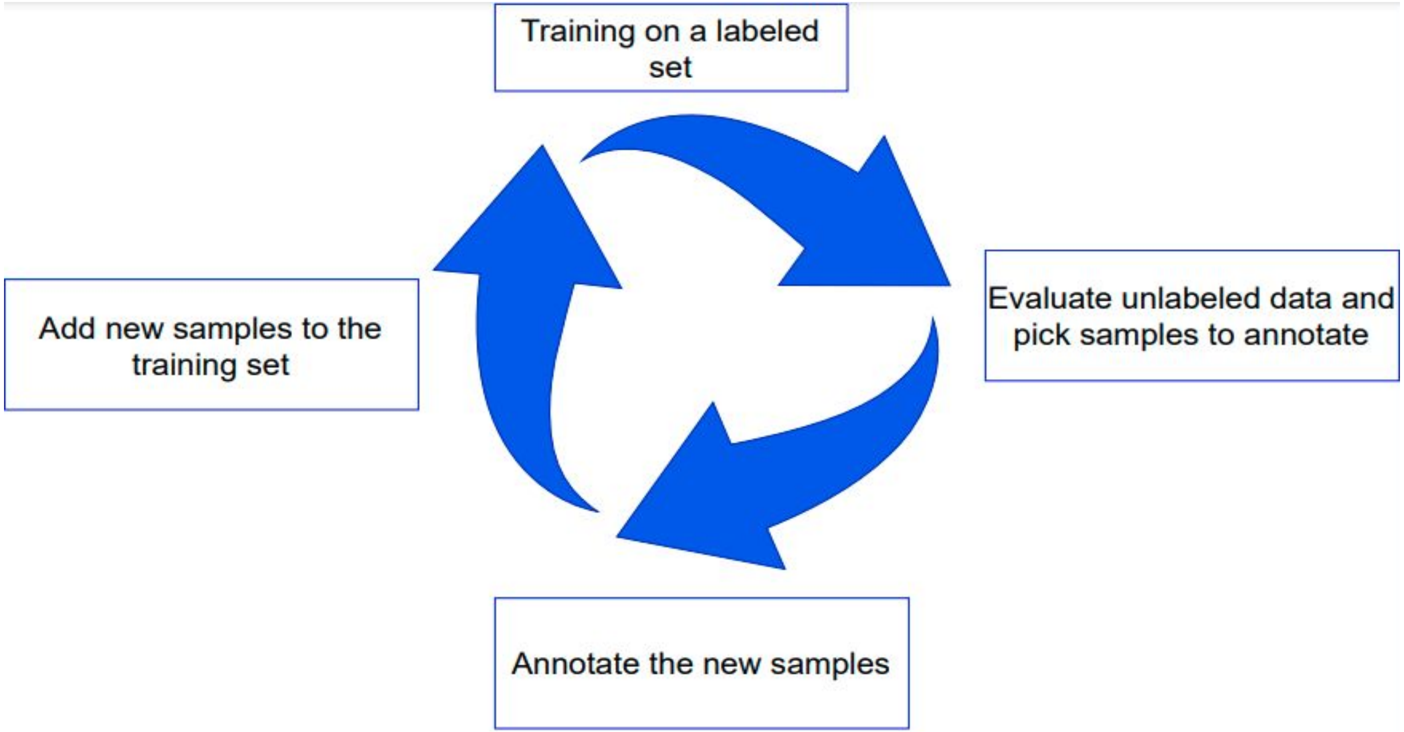


Image source [Yang et al, 2020]

Related Work: Active Learning



Related Work: Learning Loss for Active Learning [Yoo et al, 2019]

- Learning to predict what would the loss value be for an unlabeled input
- Selecting data where this prediction suggests that model is mistaken

Related Work: Bayesian Active Learning by Disagreement (BALD) [Houlsby et al, 2011]

- Sample the data with a Bayesian model
- Pick the samples with largest mutual information
- Add it to the dataset

Related Work: BatchBALD: Efficient and Diverse Batch Acquisition for Deep Bayesian Active Learning [Kirsch et al, 2019]

- Similar to BALD, it is enforcing batch-aware selections

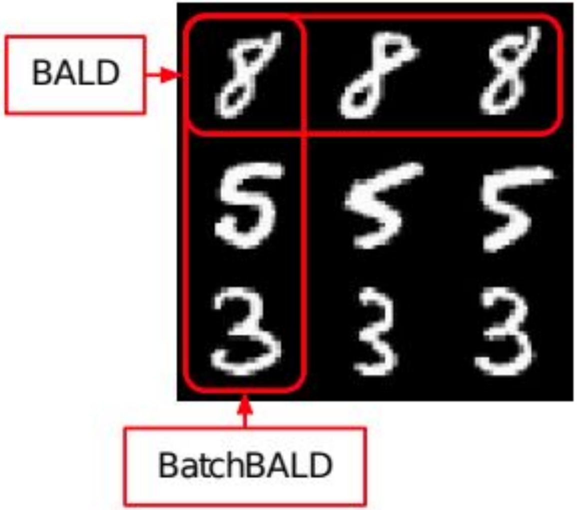


Image source [Kirsch et al, 2019]

Method: Object Detection with VoteNet [Qi et al, 2019]

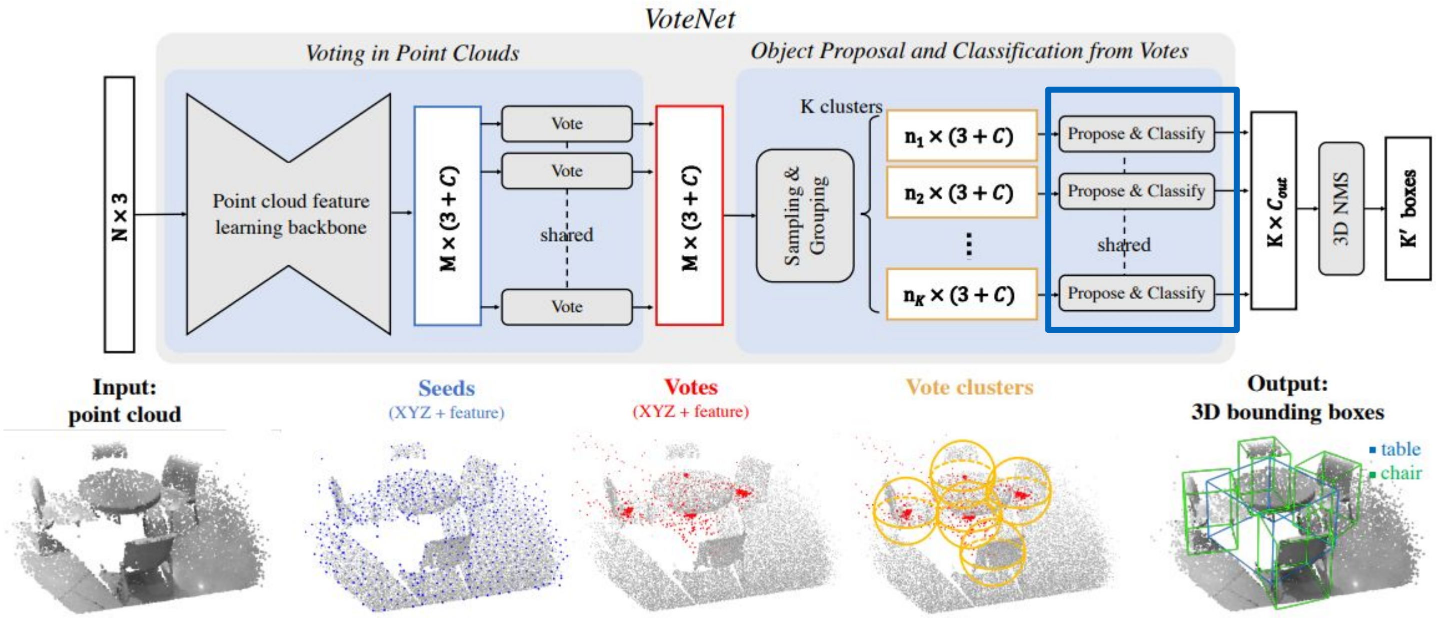
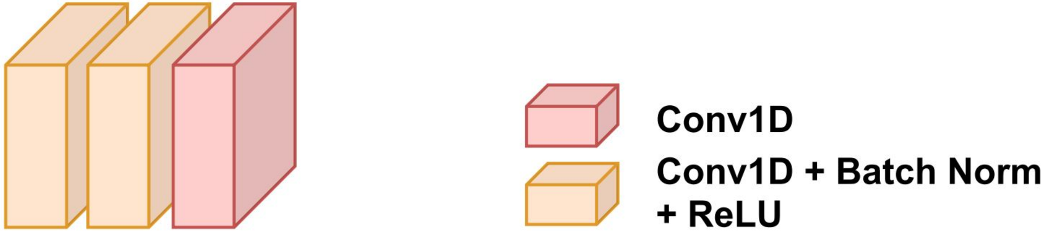


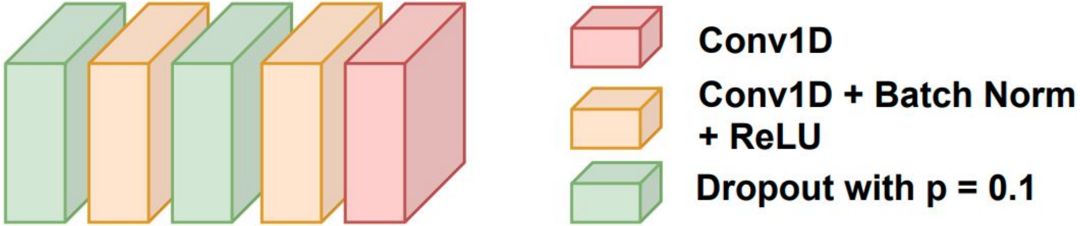
Image source [Qi et al, 2019]

Method: Extending VoteNet

Original Proposal Module

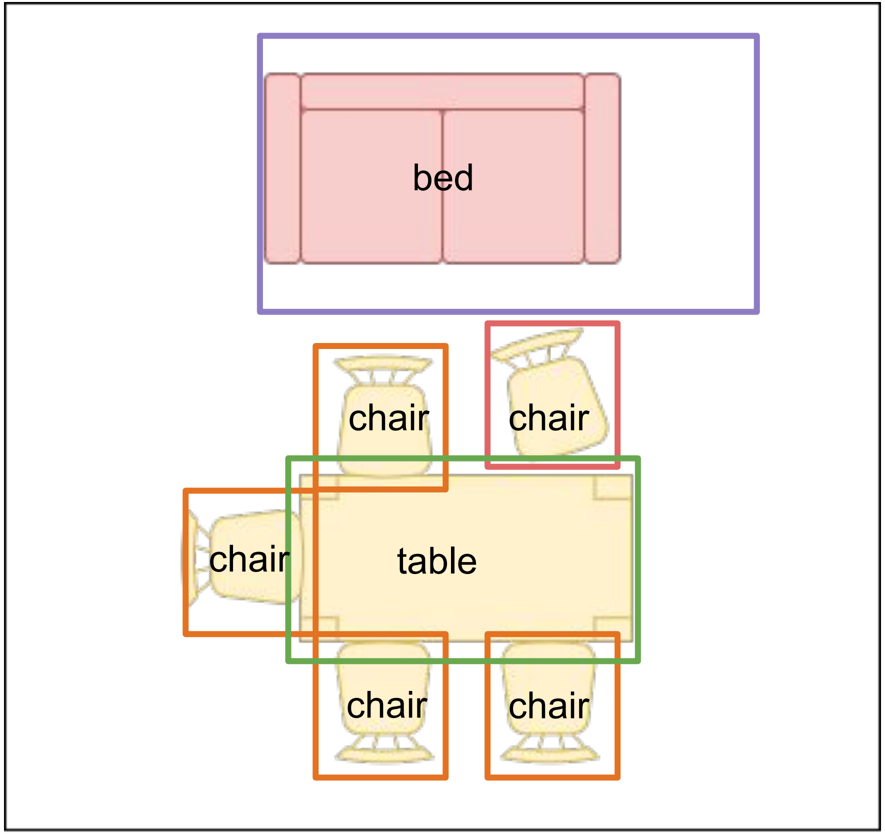


Extended Proposal Module



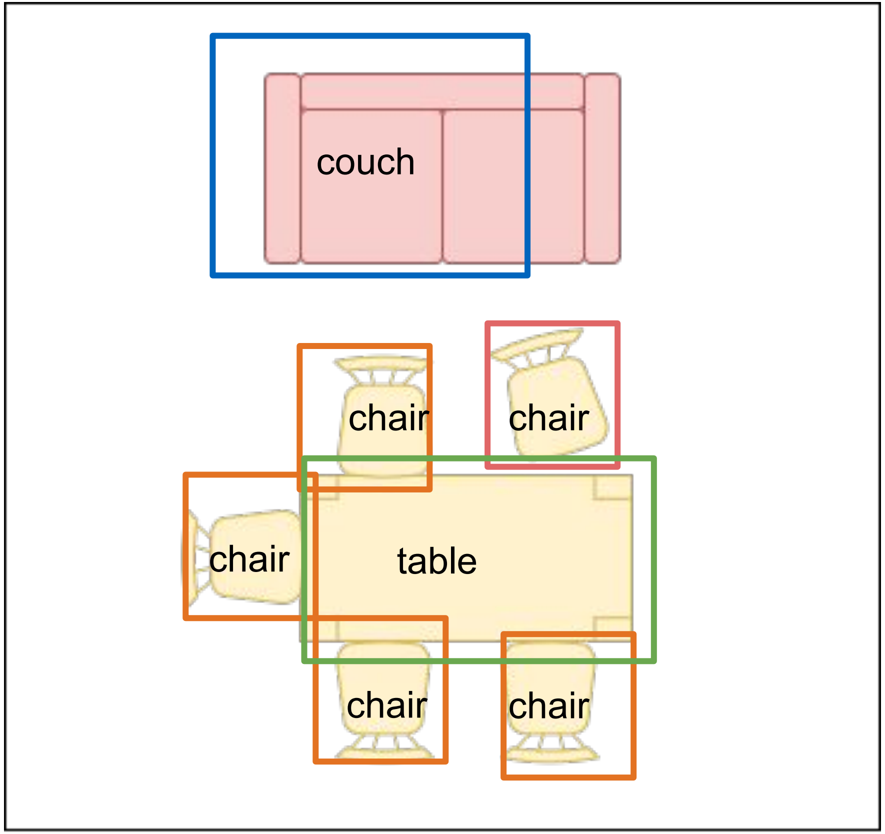
Method: Monte Carlo Dropout

Perform T forward passes with dropouts enabled



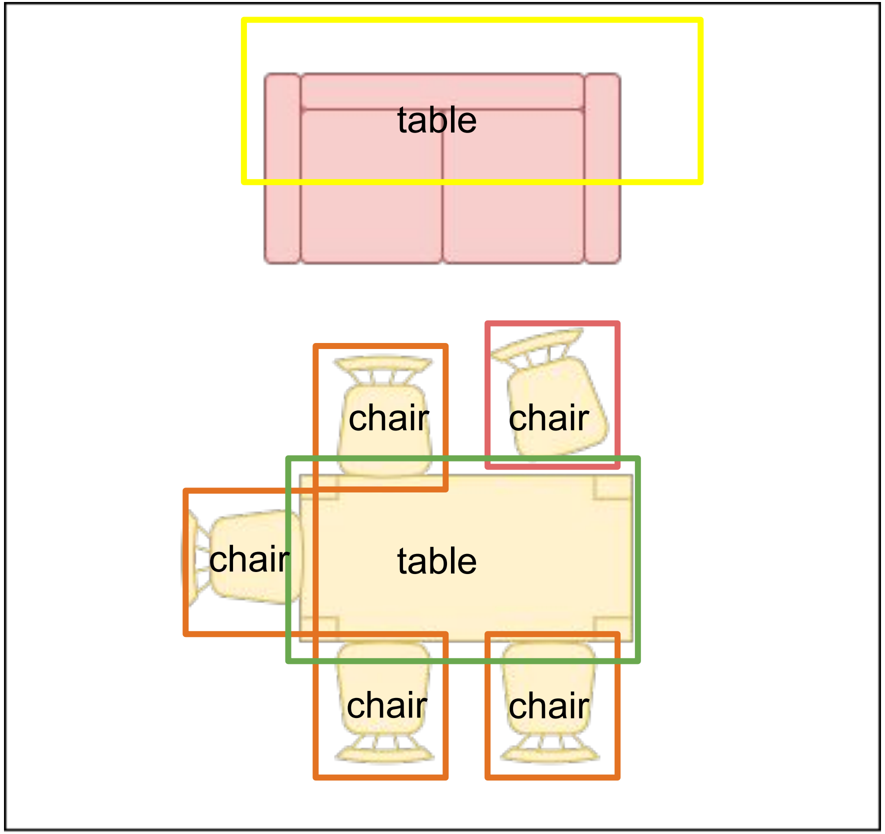
Method: Monte Carlo Dropout

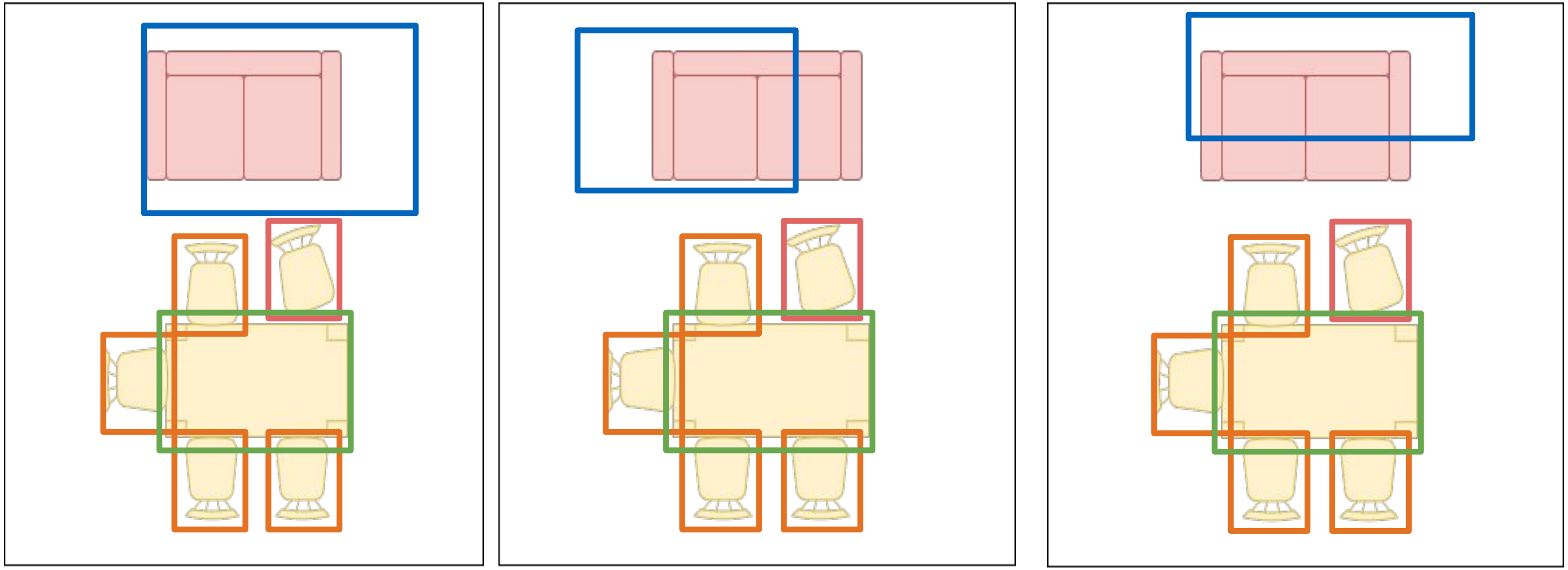
Perform T forward passes with dropouts enabled



Method: Monte Carlo Dropout

Perform T forward passes with dropouts enabled

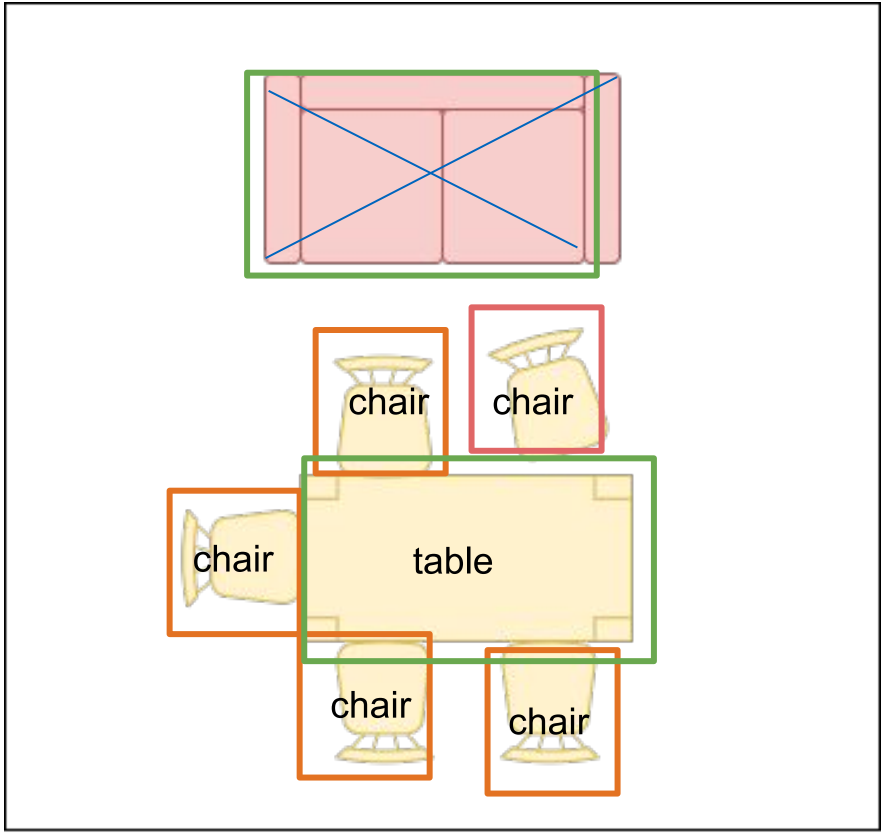




- Gather the MC samples
- Compute epistemic uncertainty

Method: Improving Detection Performance

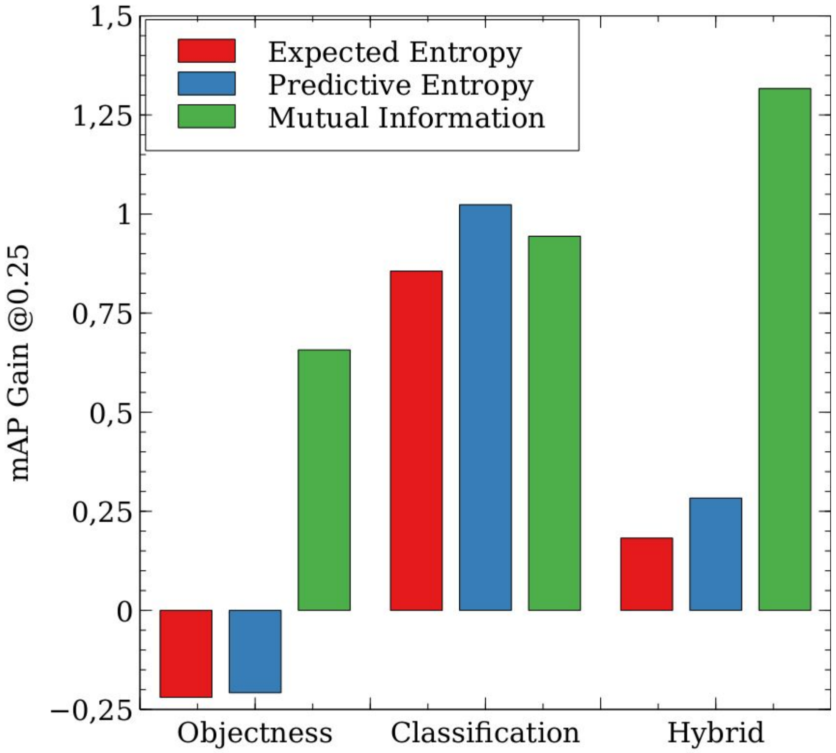
Use uncertainty as a weight for the proposal

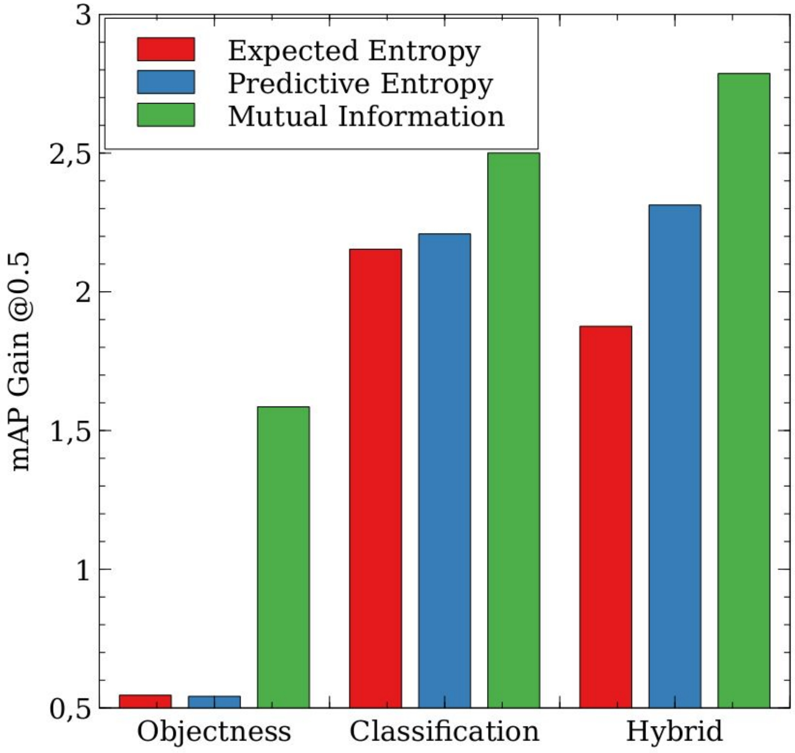


Network is not sure!

Rejected!

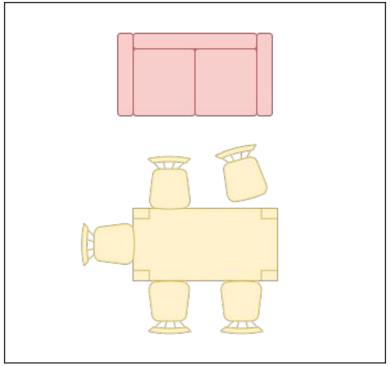
Results: Performance Boost in Detections



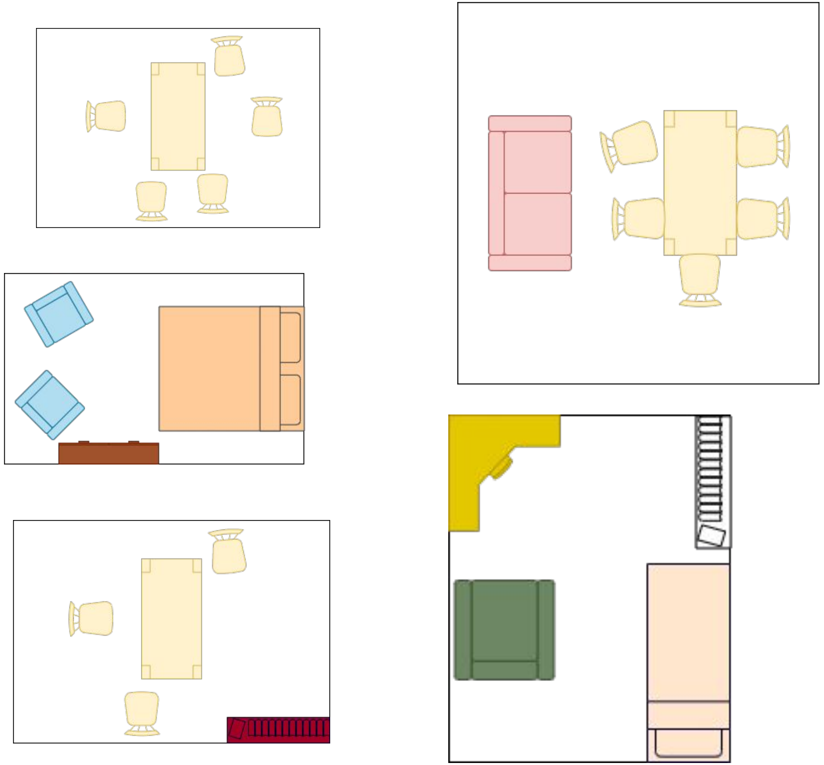


Method: Active Learning

Labeled Train Set

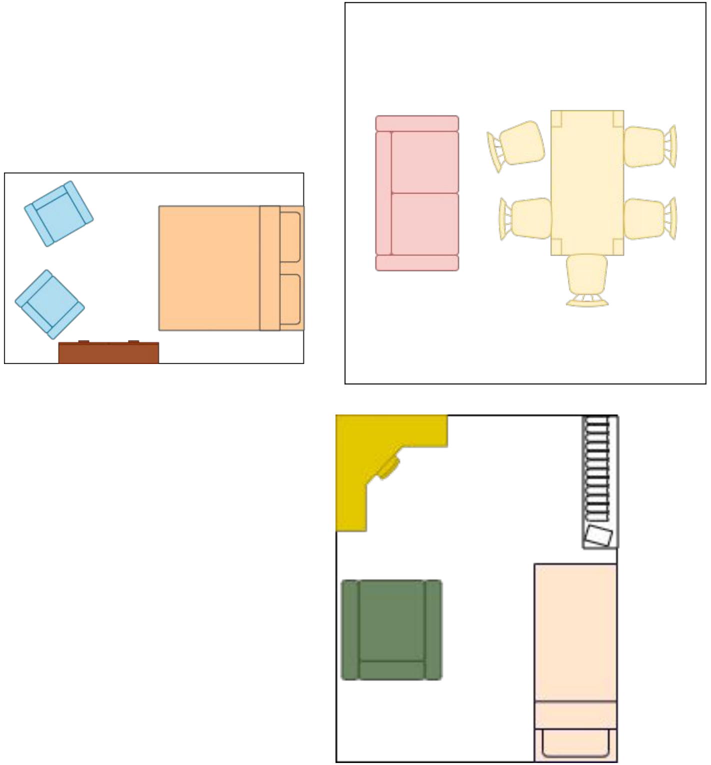


Unlabeled data

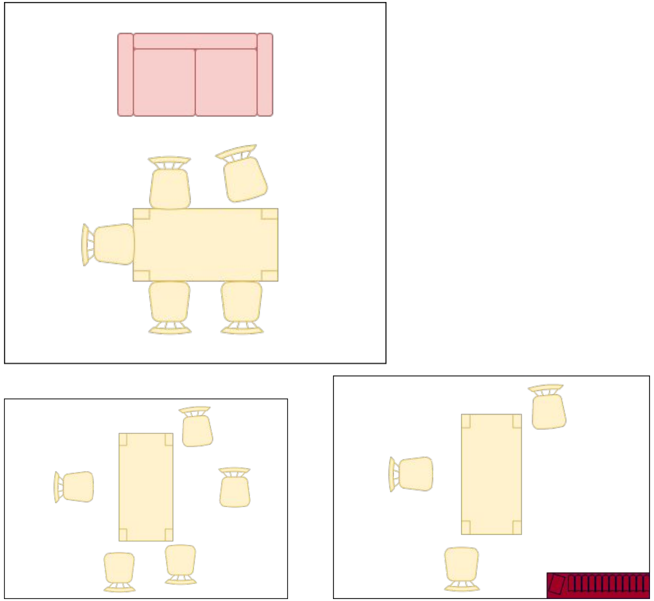


Method: Random Selection

Unlabeled data



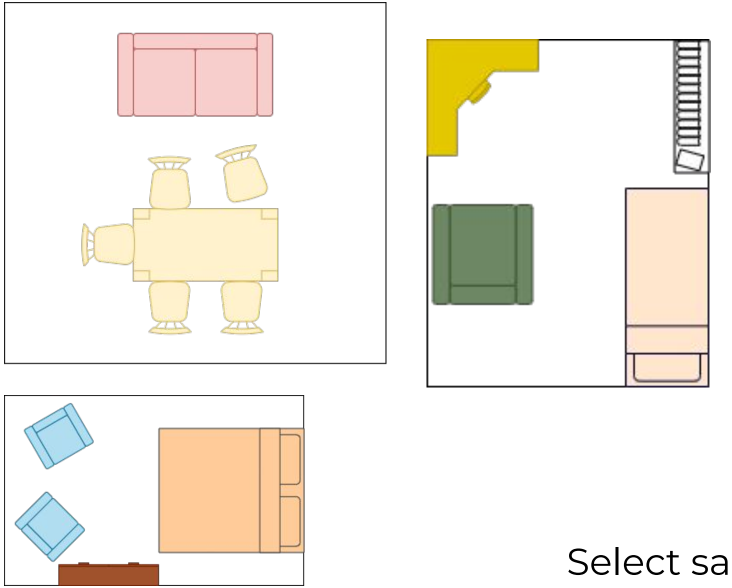
Labeled Train Set

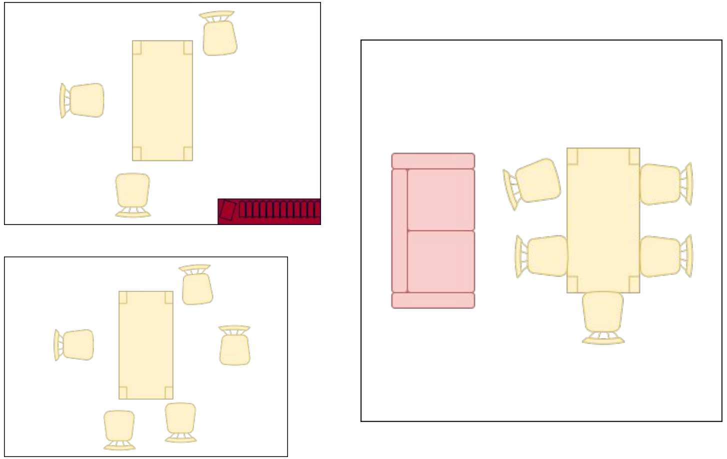


Method: Uncertainty Selection

Unlabeled data

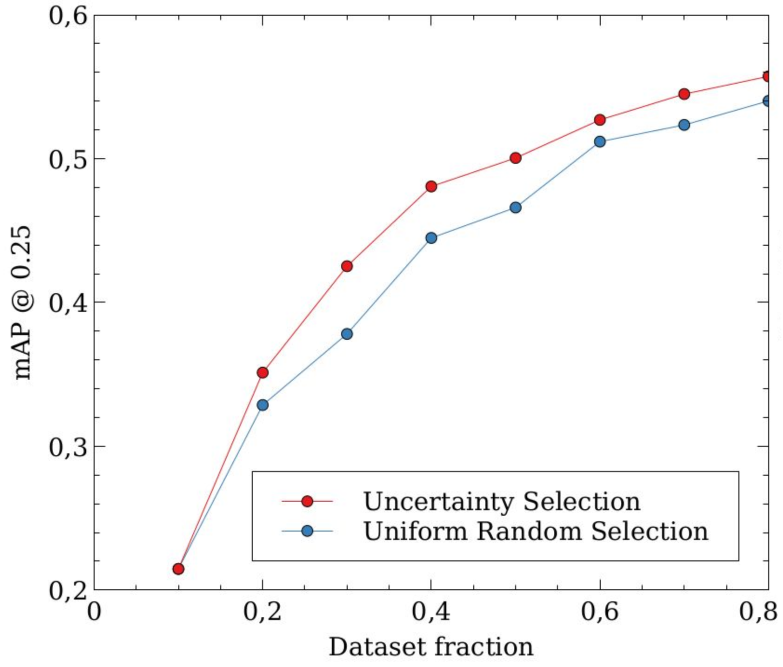
Labeled Train Set

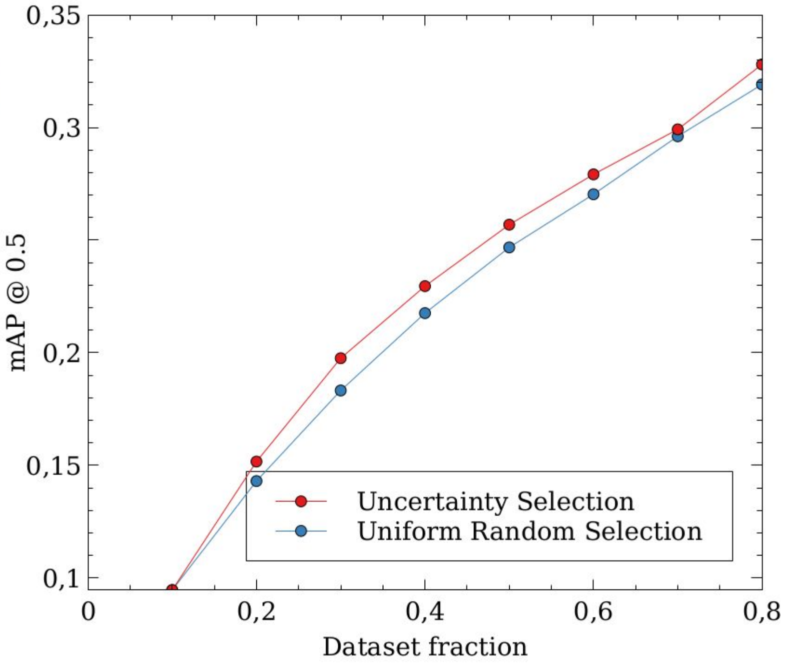




Select samples with high epistemic uncertainty!

Results: Active Learning





Conclusion & Future Work

- Even with simple techniques, we are able get visible improvements
- More complex uncertainty methods to be explored
- Aleatoric uncertainty in point clouds
- More advanced active learning schemes to be tried
- More complex weighting schemes can be tried

Appendix

Uncerainty Type	mAP @0.25	mAP @ 0.5
Native	0.57087	0.3339318
Objectness	0.57647	0.3449952
Classification	0.5795556	0.3570348
Hybrid	**0.5848642**	**0.3620454**

Uncertainty Type	AR @0.25	AR @ 0.5
Native	0.57087	0.4993466
Objectness	0.57647	0.5056164
Classification	0.5795556	0.5171316
Hybrid	**0.5848642**	**0.5173844**

Appendix

Dropout probability	mAP @ 0.25	mAP @ 0.5
p=0	57	**35**
p=0.1	**57.5**	34.5
p=0.2	55.5	30
p=0.3	51	23
p=0.5	46	15

References

- [1] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision?, 2017
- [2] Michael A. Alcorn, Qi Li, Zhitao Gong, Chengfei Wang,Long Mai, Wei-Shinn Ku, and Anh Nguyen. Strike (with) a pose: Neural networks are easily fooled by strange poses of familiar objects. CoRR , abs/1811.11553, 2018.
- [3] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Maria Florina Balcan and Kilian Q. Weinberger, editors, Proceedings of The 33rd International Conference on Machine Learning , volume 48 of Proceedings of Machine Learning Research , pages 1050–1059, New York, New York, USA, 20–22 Jun 2016. PMLR.
- [4] https://www.cnet.com/tech/services-and-software/google-apologizes-for-algorithm-mistakenly-calling-black-people-gorillas/
- [5] https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot-death-self-driving-car-elon-musk
- [6] Chao Liu, Jinwei Gu, Kihwan Kim, Srinivasa Narasimhan, and Jan Kautz. Neural rgb->d sensing: Depth and uncertainty from a video camera
- [7] Nan Yang, Lukas von Stumberg, Rui Wang, and Daniel Cremers. D3vo: Deep depth, deep pose and deep uncertainty for monocular visual odometry, 2020.
- [8]Kirsch, A., van Amersfoort, J., and Gal, Y., “BatchBALD: Efficient and Diverse Batch Acquisition for Deep Bayesian Active Learning”, , 2019.
- [9]D. Yoo and I. S. Kweon, "Learning Loss for Active Learning," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 93-102, doi: 10.1109/CVPR.2019.00018.
- [10] Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. arXiv preprint arXiv:1112.5745, 2011.
- [11] Charles R Qi, Or Litany, Kaiming He, and Leonidas J Guibas. Deep hough voting for 3d object detection in point clouds. In Proceedings of the IEEE International Conference on Computer Vision , 2019.
- [12] Nekrasov, Alexey, et al. "Mix3D: Out-of-Context Data Augmentation for 3D Scenes." *arXiv preprint arXiv:2110.02210* (2021).
- [13] Nie, Yinyu et al. “RfD-Net: Point Scene Understanding by Semantic Instance Reconstruction.” *CVPR* (2021).
- [14] Božič, Aljaž, et al. "TransformerFusion: Monocular RGB Scene Reconstruction using Transformers." *arXiv preprint arXiv:2107.02191* (2021).