S07 - Label Efficient 3D Detection with Foundation Models

6-month internship @ CEA List

Internship context

Based in Saclay (Essonne), the LIST is one of the two institutes of CEA Tech, the technological research division of the CEA. Dedicated to intelligent digital systems, its mission is to carry out technological developments of excellence on behalf of industrial partners in order to create value.

Within the LIST, the Laboratory of Vision and Learning for Scene Analysis (LVA) conducts research in the field of computer vision and artificial intelligence for the perception of intelligent and autonomous systems. The laboratory's research themes include visual recognition, behavior and activity analysis, large-scale automatic annotation, and perception and decision models. These technologies are applied in major sectors such as security, mobility, advanced manufacturing, healthcare, and sports...

Missions

3D object detection is a critical component of many applications such as autonomous driving, robotics, and augmented reality, where having a precise understanding of the 3D environment is crucial. In the context of 3D object detection, **a key challenge lies in the high cost of annotating 3D bounding boxes**, making it difficult to scale supervised learning methods to new applications.

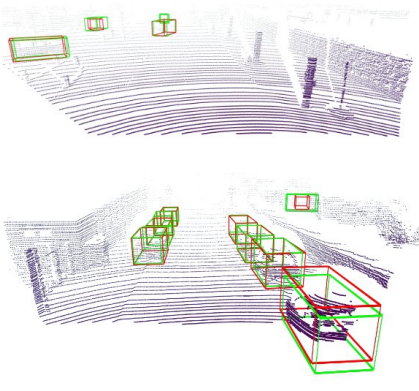
To address this, various learning paradigms such as semi-supervised[1][2], weakly supervised[3], and unsupervised domain adaptation have been proposed. These paradigms aim to reduce the need for large amounts of annotated data while maintaining or improving performance. By leveraging **minimal labeled data or even unannotated data from different domains**, these approaches help reduce the reliance on costly 3D box annotations.

Most state-of-the-art methods for these paradigms rely on a teacher-student architecture, which typically uses the teacher model to generate pseudo-labels that the student model learns from. A crucial aspect of this approach is pseudo-label filtering, which can be done using two main strategies. One strategy involves untrained heuristics, such as confidence scores produced by detection models, while the other strategy uses trained uncertainty estimation modules, which rely on a small set of annotated 3D data. Both of these approaches, however, have limitations. **Heuristics can be overly reliant on specific hyperparameters** that may overfit to the seen data, while **uncertainty estimators can be unreliable** when trained on limited data, leading to suboptimal performance.

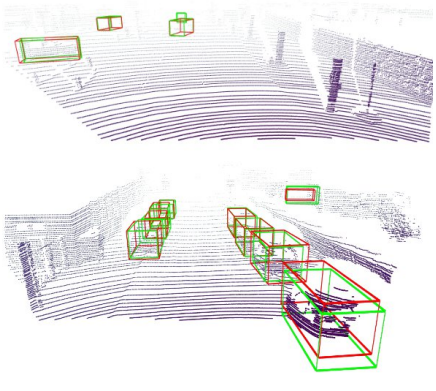
Recent breakthroughs in **2D vision-language models (VLMs)** have inspired research in 3D vision, particularly around the potential of these models for pretraining [4][5].

However, despite the promise of VLMs, there is little exploration of their use in the context of semi-supervised, weakly supervised, or unsupervised domain adaptation for 3D object detection. **Therefore we aim to fill this gap by leveraging the power of foundation models for more robust pseudo-label filtering**. This could involve using pixel features from the 2D projections of 3D points to calculate intra-object coherence, as well as neighborhood incoherence scores to ensure that objects are correctly detected and isolated. Additionally, 2D features could be used as a pretext for scene completion tasks, providing finer object contours and estimating occluded parts of detected objects.

Camera View



Iteration 0 Predictions



Final Predictions

Internship objectives

As part of our team, you will:

- Leverage vision-language models (VLMs) to enhance 3D detection performance.
- Work on innovative pseudo-labeling techniques to improve model training with minimal labeled data.
- Use 2D and 3D feature integration to improve scene understanding.
- Gain hands-on experience with deep learning frameworks and 3D vision algorithms.

References

[1] Zhao, N., Chua, T. S., & Lee, G. H. (2020). Sess: Self-ensembling semi-supervised 3d object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11079-11087).

[2] Xu, H., Liu, F., Zhou, Q., Hao, J., Cao, Z., Feng, Z., & Ma, L. (2021, September). Semi-supervised 3d object detection via adaptive pseudo-labeling. In 2021 IEEE International Conference on Image Processing (ICIP) (pp. 3183-3187). IEEE.

[3] Yao, B., Dong, L., Qiu, X., Song, K., Yan, D., & Peng, C. (2024). Uncertainty-guided Contrastive Learning for Weakly Supervised Point Cloud Segmentation. IEEE Transactions on Geoscience and Remote Sensing.

[4] Chen, Zhimin, et al. "Bridging the domain gap: Self-supervised 3d scene understanding with foundation models." Advances in Neural Information Processing Systems 36 (2024).

[5] Sirko-Galouchenko, S., Boulch, A., Gidaris, S., Bursuc, A., Vobecky, A., Pérez, P., & Marlet, R. (2024). OccFeat: Self-supervised Occupancy Feature Prediction for Pretraining BEV Segmentation Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4493-4503).

Qualifications

- Students in their 4th or 5th year of studies (M1, M2 or gap year)
- Computer vision skills
- Machine learning skills (deep learning, perception models, generative AI...)
- Python proficiency in a deep learning framework (especially TensorFlow or PyTorch)

Job-related benefits

Joining the CEA List and the LVA as an intern means:

- Joining an organization that addresses societal challenges to build the world of tomorrow.
- Working in one of the most innovative research organizations in the world (ranked in the global top 100, top 3 in France).
- Discovering a rich ecosystem where the institute creates privileged links between the industrial and academic sectors.
- Conducting research in an environment where autonomy and creativity are recognized, and where valorizing results is encouraged (publication of scientific articles, patents, and sharing of open-source code whenever possible).
- Joining a young and dynamic team made up of research engineers, PhD students, post-doctoral researchers, and interns.
- Benefiting from an internal computing infrastructure equipped with around 300 state-of-the-art GPUs.
- Receiving a stipend between €1300 and €1400 per month.
- Having the opportunity to continue with a PhD or as a research engineer after the internship.
- Having the possibility of remote work, receiving a 75% (instead of 50%) reimbursement on public transportation costs, and benefiting from the "mobili-jeune" aid to reduce rent costs...

To apply, contact the laboratory with a CV and cover letter: lva-stages@cea.fr