CEA List
Laboratoire de Vision et Apprentissage pour l'analyse de scène
Centre de Saclay 91191 Gif-sur-Yvette France
http://www.kalisteo.eu

Contact Nicolas Granger
Hervé Le Brogne
Tél +33 (0)1 69 08 11 32
E-mail nicolas.granger@cea.fr
herve.le-borgne@cea.fr

THESE 2022 Réf : LVA-LASTI-22-T7

Weakly Supervised methods for 3D Point Cloud Perception

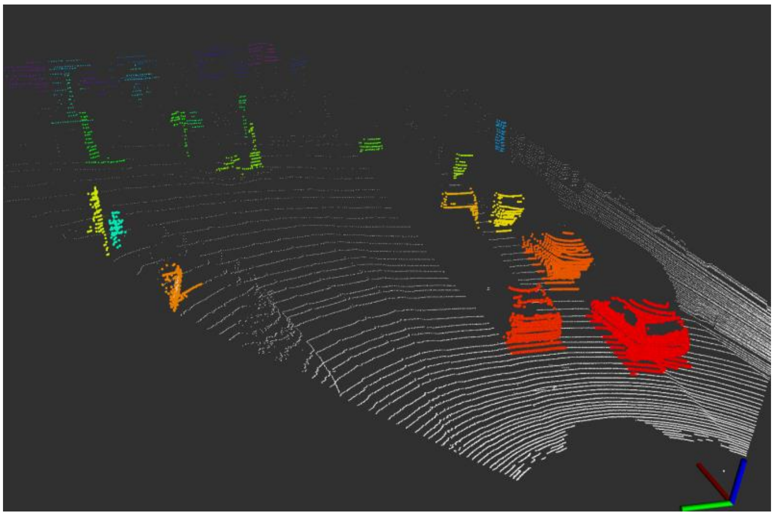

Figure 1 Sample scene from KITTI dataset [9]

Host laboratory

The PhD candidate will be supervised by Nicolas Granger, Hervé Le Borgne and Quoc-Cuong Pham from CEA LIST.

Based in Saclay (Paris region, France), the LIST Institute (CEA LIST) is one of CEA Tech three technological research institutes constituting CEA technological Division. Dedicated to smart digital systems, its mission is to achieve technological development of excellence for the industrial partners and create value.

LIST institute has more than 750 partners and every year more than a 200 partnership activities are being conducted with French and foreign industrial companies on applied research projects in four main topics: Advanced Manufacturing, Embedded systems, Data intelligence, Health and ionizing radiations. Labeled as "Carnot Institute", LIST invests every year 25% of its budget into scientific resourcing actions in order to better identify tomorrow's technological breakthroughs. CEA Tech as CEA (Commissariat à l'énergie atomique et aux énergies alternatives) Technological Research Division gathers three specialized technological research institutes among which LIST institute dedicated to digital systems.

Inside the institute, the LVA (laboratory of computer vision and machine learning for scene understanding) and LASTI (laboratory of semantic analysis of texts and images) gather around 50 persons, from PhD to researchers, working in the domain of computer vision and natural language analysis.

Context

3D point cloud LiDAR techniques have considerably improved over the course of the last few years, reaching industrial-ready status in terms of methods as well as sensors. In the context of autonomous driving and driving assistance, LiDAR data can provide a rich and reliable information about the environment. Indeed, numerous methods proposed in the literature achieve state-of-the-art performances using this data modality alone on detection [11], tracking or segmentation [12,13] tasks. These methods are often inspired by models developed for images in the visible spectrum with three mains axes for contributions: casting the 3D problem back to two dimensions [14], extending existing 2D methods to 3D [3], and leveraging properties specific to 3D LiDAR point clouds to enhance the model efficiency and quality.

CEA List
Laboratoire de Vision et Apprentissage pour l'analyse de scène
Centre de Saclay 91191 Gif-sur-Yvette France
http://www.kalisteo.eu

Contact Nicolas Granger
Hervé Le Brogne
Tél +33 (0)1 69 08 11 32
E-mail nicolas.granger@cea.fr
herve.le-borgne@cea.fr

From a scientific perspective, 3D point clouds provides a versatile environment for experimental research, as their rich data structure lends itself to original implementation of concepts also developed on other data modalities. For instance, set-based approaches or spatial attention methods, which have recently gained traction for image models, also apply quite naturally to point clouds.

Although 3D point cloud perception can be considered a mature research subject, several major scientific challenges remain nonetheless:

- Even the largest public datasets [2,9,10] cannot represent the wide variety of driving environment, as demonstrated by a few Transfer Learning experiments [4], even more so for accident-prone situation for which data is scarse. Before considering larger-scale deployment of this technology, LiDAR-based perception methods must be trained over more diverse data.
- Although robust and informative, 3D LiDAR point cloud still contain ambiguities originating from the lack of resolution at a distance and occlusions.

In parallel to the development of supervised methods, unsupervised and weakly supervised methods on text, images and videos have progressed rapidly over the last few years.

Adapting them to 3D point clouds [7,8] would help take advantage of large unannotated datasets which are more easily acquired. Among these techniques, temporal consistency-based approaches seem particularly suitable for 3D point cloud since they might be combined with spatial consistency more naturally than what can be done with 2D data.

Combining LiDAR data over time has not yet been investigated thoroughly [1,3,6] and might also help to lift occlusion ambiguities by merging multiple points of views in the case of a moving sensor, and also increase point density. Fusing the data from other sensors, in particular cameras, is also expected to bring substantial improvements.

Practical interests

In terms of use-cases, driving has been a strong basis for research on 3D point cloud based perception, along with indoor scene analysis. But the domain expands rapidly as sensors are improving and becoming more ubiquitous.

By taking advantage of physical properties specifically available on 3D point cloud data, this thesis can contribute to improve unsupervised learning and generalization properties of Deep Learning models. In the long term, this thesis can contribute beyond the scope of driving environment with applications to medical data, aerial imagery for agriculture or cartography, safety in industrial environments, etc.

Objectives

The main goal of this thesis is to improve generalization and robustness properties of 3D point cloud perception models. Two complementary directions of research can help toward this objective:

- Leveraging unsupervised and weakly supervised methods such as the ones developed in other application domains.
- Proposing new approaches that take advantage of 3D point cloud properties, for instance temporal or spatial consistency and the availability of a complementary data modality from camera sensors.

Contributions need not focus on a specific end-task (detection, tracking, segmentation, etc.), however, the driving use-case is identified as the most suitable in terms of prior art, available data, and relevance.

References

[1] W. Luo, B. Yang, and R. Urtasun, "Fast and Furious: Real Time End-to-End 3D Detection, Tracking and

Motion Forecasting with a Single Convolutional Net," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, Jun. 2018, pp. 3569–3577.
[2] H. Caesar et al., "nuScenes: A multimodal dataset for autonomous driving," arXiv preprint arXiv:1903.11027, 2019.
[3] C. Choy, J. Gwak, and S. Savarese, "4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks," presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2019.
[4] C. Saltori, S. Lathuiliere, N. Sebe, E. Ricci, and F. Galasso, "SF-UDA3D: Source-Free Unsupervised Domain Adaptation for LiDAR-Based 3D Object Detection," in 2020 International Conference on 3D Vision (3DV), Nov. 2020, pp. 771–780.
[5] S. Xie, J. Gu, D. Guo, C. R. Qi, L. Guibas, and O. Litany, "PointContrast: Unsupervised Pre-training for 3D Point Cloud Understanding," in Computer Vision – ECCV 2020, Cham, 2020, pp. 574–591.
[6] T. Yin, X. Zhou, and P. Krähenbühl, "Center-based 3D Object Detection and Tracking," arXiv:2006.11275 [cs], Jun. 2020.
[7] S. A. Baur, D. J. Emmerichs, F. Moosmann, P. Pinggera, B. Ommer, and A. Geiger, "SLIM: Self-Supervised LiDAR Scene Flow and Motion Segmentation," in 2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021, 2021, pp. 13106–13116.
[8] H. Wang, Q. Liu, X. Yue, J. Lasenby, and M. J. Kusner, "Unsupervised Point Cloud Pre-training via Occlusion Completion," in 2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021, 2021, pp. 9762–9772.
[9] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, Jun. 2012, pp. 3354–3361. doi: 10.1109/CVPR.2012.6248074.
[10] S. Ettinger et al., "Large Scale Interactive Motion Forecasting for Autonomous Driving : The Waymo Open Motion Dataset," in 2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021, 2021, pp. 9690–9699. doi: 10.1109/ICCV48922.2021.00957.
[11] X. Chen, S. Shi, B. Zhu, K. C. Cheung, H. Xu, and H. Li, "MPPNet: Multi-Frame Feature Intertwining with Proxy Points for 3D Temporal Object Detection." arXiv, 2022.
[12] J. Xu, R. Zhang, J. Dou, Y. Zhu, J. Sun, and S. Pu, "RPVNet: A Deep and Efficient Range-Point-Voxel Fusion Network for LiDAR Point Cloud Segmentation," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Oct. 2021, pp. 16024–16033.
[13] X. Zhu et al., "Cylindrical and Asymmetrical 3D Convolution Networks for LiDAR Segmentation," arXiv:2011.10033 [cs], Nov. 2020, Accessed: Dec. 16, 2020.
[14] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast Encoders for Object Detection From Point Clouds," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, Jun. 2019, pp. 12689–12697.

Qualifications :	Engineering degree, Master 2
Contract duration :	3 years
Salary :	Between 1800 € and 2000 €

Required skills :

- Prior experience in machine learning and particularly in deep learning.
- Proficiency in Python and a deep learning framework is required.
- Good spoken and written communication skills in English. Speaking French is optional but may help for the everyday life during the PhD.
- Knowledge of computer vision, natural language processing, signal processing or physics is a plus
- Being co-author of a submitted or published paper in the field is a plus.