CEA List
Service d'Intelligence Artificielle pour le Langage et la Vision
Centre de Saclay 91191 Gif-sur-Yvette France
http://www.kalisteo.eu

Contacts Nicolas Granger
Mohamed Chaouch

E-mail nicolas.granger@cea.fr
mohamed.chaouch@cea.fr

STAGE 2022 Réf : LVA-22-S4

3D Point Cloud Perception with Transformer Models

Presentation of the host laboratory

Based in Paris-Saclay campus, CEA-LIST is one of four technological research institutes of CEA TECH, the technological research direction of CEA. Dedicated to intelligent digital systems, it contributes to the competitiveness of companies via research and knowledge transfers. The expertise and competences of the 800 research engineers and technicians at CEA-LIST help more than 200 companies in France and abroad every year on subjects categorized over 4 programs and 9 technological platforms. 21 start-ups have been created since 2003.

The "Laboratoire Vision et Apprentissage pour l'analyse de scenes" addresses computer vision subjects with a stronger emphasis on four axes:

- Recognition (detection or segmentation of objects and persons)
- Behavior analysis (action and gesture recognition, anomalous behavior of individuals or crowds)
- Smart annotation (large scale annotation of 2D and 3D data using semi-supervised methods)
- Perception and decision-making (Markovian decision processes, navigation)

The intern will join a team composed of 30 researchers (research engineers, PhD students, interns) and will be able to interact with peers working on related subjects and methods.

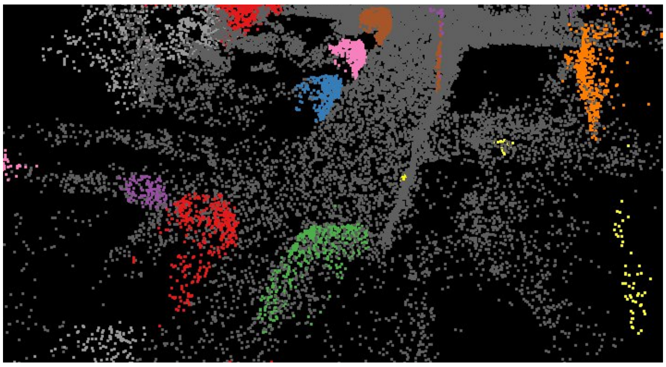

Figure 1 - Sample semantic segmentation from KITTI360 dataset

Context

LiDAR or ToF sensors evolve rapidly and become more ubiquitous, similarly to color cameras a couple of decades ago. Point clouds differ largely from images in their structure as they are largely unstructured in the form of a set of points, have a lower resolution compared to cameras for now, but provide an accurate point coordinates in 3D which helps eliminate many ambiguities related to relative and absolute positioning.

So far, the state of the art on modeling 3D point clouds has mirrored advances made on image models using one of the following strategies: up-lift a 2D model to 3D using for instance 3D convolutions [2], down-cast the 3D space to 2D via depth map projections [4] or bird-eye-views for flat scenes [3]. A third category of models addresses 3D point clouds directly as a set prediction problem [1,6].

In the mean-time Transformer models have successfully applied set-based logic onto several data modalities such as text, images, speech, and more recently point clouds. On the latter category however, we observe that the proposed solutions are still in a preliminary stage and have not yet addressed some of the challenges raised by 3D point cloud modeling. Indeed, 3D point clouds provide a sparse information within large scenes leading to computational challenges. Moreover, 3D point clouds benefit from strong localization properties that have not yet been transposed into the models.

Objectives of this internship

Starting from the current state of the art models in 3D detection, including Transformer models, the intern will research the integration of recent contributions from the field (ex: [7]) to improve the processing and training efficiency of Transformer models on 3D point clouds. In addition, this internship will seek new contributions specifically focused on the point cloud modality with the following suggested axes of research:

- Integration of occlusion hypotheses
- Stronger utilization of invariance hypotheses in model design
- Spatial coherence in time

References

[1] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017. doi: 10.1109/cvpr.2017.16.

[2] 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. doi: 10.1109/cvpr.2019.00319.

[3] 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. doi: 10.1109/CVPR.2019.01298.

[4] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "RangeNet ++: Fast and Accurate LiDAR Semantic Segmentation," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2019, Macau, SAR, China, November 3-8, 2019*, 2019, pp. 4213–4220. doi: 10.1109/IROS40897.2019.8967762.

[5] I. Misra, R. Girdhar, and A. Joulin, "An End-to-End Transformer Model for 3D Object Detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 2906–2917.

[6] 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.

[7] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable Transformers for End-to-End Object Detection," 2021. [Online]. Available: https://openreview.net/forum?id=gZ9hCDWe6ke

Keywords

object detection, semantic segmentation, deep learning, Lidar, 3D, tranformer

Required level:	Engineer, Master 2
This internship opens the possibility of pursuing a thesis and R&D engineer in our laboratory.	
Duration :	6 months
Remuneration:	between 700 € and 1300 € depending on the training.
Required Skills : - Computer vision - Machine learning (deep learning) - Shape recognition - Proficiency in programming (Python) - Mastery of a deep learning framework (in particular PyTorch)	