Two-year Postdoc Position

3D Transformers for Autonomous Driving

Collaboration Context

A two-year postdoc position is open in the research axis “Vision and 3D Perception for Scene Understanding” of a joint Valeo / Inria context.

Scientific Context

In the same way that learning with *transformers* has somehow brought a revolution in natural language processing [Vas17], and then in image processing [Car20, Dos21], recently emerging transformers [Guo21, Zha21] for point clouds are now also starting to significantly boost the performance of 3D processing. Yet, the general concepts of transformers and the reasons for their success are not fully understood, and there are still many open ways to adapt them efficiently to 3D.

Topics

The emergence of transformers opens a wide field of research subjects that matches well with the research axis “Vision and 3D Perception for Scene Understanding”. The goal of this postdoc is to attack some of them.

1. A first objective is to investigate the specificity of transformers for 3D. It includes the question of positional encoding, which is widely different in 3D from the 1D (text) and 2D (image) perspectives. It also includes the square complexity issues with cross- and self-attention, taking into account the sparsity of point clouds. More generally, it concerns as well the time and space complexity of point cloud processing, in the perspective of real-time, embedded software.

   This study of 3D transformers will be made through downstream tasks that include object detection, semantic or panoptic segmentation, and denser depth predictions, which are key tasks for autonomous driving. We will study in particular the impact of the specific sparsity and data patterns induced by vehicle sensors. We will also consider a stream of point clouds, as available from a lidar, taking time into account in a 4D perspective.

2. Besides, we will investigate the use of transformers backbones regarding self-supervision [Sim21], which is a key approach to bring supervised learning to a new level by saving the cost of annotating large datasets. We will study new pretext tasks in 3D that transformers more specifically leverage, as well as contrastive learning techniques that are tightly linked to the attention mechanism of transformers.
   A PhD student is about to start on self-supervision for 3D at Valeo and another works on 2D supervision for 3D at Inria, though none of them focusing on transformers. The postdoc will be able to work jointly with either PhD student on self-supervision issues.

3. Moreover, we will study the transferability of learned transformer models in the perspective of domain adaptation [Vu19a, Vu19b]. In particular, we will investigate the disentangling of latent space representations, working towards domain-invariant

features by enforcing orthogonality of the domain features while enabling the discovery of exclusive task or domain features, through their realization via multi-head attention.

Another PhD student in Valeo is about to start working on 3D domain adaptation, although with a different perspective (using optimal transport). There will nonetheless be a number of collaboration opportunities between this PhD student and the postdoc regarding adapting transformer-based features.

4. A last research direction concerns multi-modality, when lidar point clouds are acquired together with camera images, to leverage the similarity and complementarity of sensor information [Jar20]. One technical subject concerns a possible interplay between the forms of transformer attention used in 2D and the kinds of attention that are and will be developed in 3D. Another more general question is the joint self-supervision from the interaction of 2D and 3D, or from cross-task representations. Last, we will study the intertwined relation of geometry and semantics through the semantic scene completion task [Rol20, Rol21].

Profile

Applicants should have defended or be finishing their PhD and have a strong publications record. They should have a solid background in computer vision (including 3D processing) and machine learning, particularly in deep learning, with strong PyTorch coding skills.

Contact

To apply, applicants should send a mail to Renaud Marlet (renaud.marlet@valeo.com) and Raoul de Charette (raoul.de-charette@inria.fr) with:

- a cover letter explaining their interest and adequacy for the postdoc topic,
- their CV/resume,
- possibly, references or recommendation letters.

References

[Car20] End-to-End Object Detection with Transformers. Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko. European Conference on Computer Vision (ECCV), 2020.

[Dos21] An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby. International Conference on Learning Representations (ICLR), 2021.

[Guo21] PCT: Point cloud transformer. Meng-Hao Guo, Jun-Xiong Cai, Zheng-Ning Liu, Tai-Jiang Mu, Ralph R. Martin, Shi-Min Hu. Computational Visual Media, 2021.

[Jar20] xMUDA: Cross-Modal Unsupervised Domain Adaptation for 3D Semantic Segmentation. Maximilian Jaritz, Tuan-Hung Vu, Raoul de Charette, Émilie Wirbel, Patrick Pérez. Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[Rol20] LMSCNet: Lightweight Multiscale 3D Semantic Completion. Luis Roldao, Raoul de Charette, Anne Verroust-Blondet. International Conference on 3D Vision (3DV), 2020

[Rol21] 3D Semantic Scene Completion: a Survey. Luis Roldao, Raoul de Charette, Anne Verroust-Blondet. International Journal of Computer Vision (IJCV), 2021.

[Sim21] Localizing Objects with Self-Supervised Transformers and no Labels. Oriane Siméoni, Gilles Puy, Huy V. Vo, Simon Roburin, Spyros Gidaris, Andrei Bursuc, Patrick Pérez, Renaud Marlet, Jean Ponce. British Machine Vision Conference (BMVC), 2021.

[Vas17] Attention Is All You Need. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Conference on Neural Information Processing Systems (NeurIPS), 2017.

[Vu19a] ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation. Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, Patrick Pérez. Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[Vu19b] DADA: Depth-aware Domain Adaptation in Semantic Segmentation. Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, Patrick Pérez. International Conference on Computer Vision (ICCV), 2019.

[Zha21] Point Transformer. Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip Torr, Vladlen Koltun. International Conference on Computer Vision (ICCV), 2021.