S02 – Object detection by parametrization using elementary 3D shapes

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...

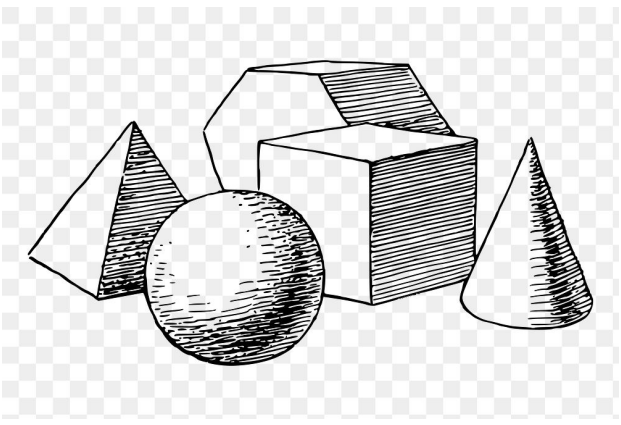
Missions

Computer vision is currently at a turning point, with the convergence of visual, audio and, above all, textual modalities. This convergence is making it increasingly possible to benefit from or pool the knowledge generated by these different modalities. This knowledge is also becoming 'fundamental', meaning that it transgresses the ontologies of databases and seeks to move towards a holistic representation of their field of expertise. We're interested here in object detection.

To interrogate these fundamental models, it is necessary to use a prompt, initially textual [1], but these prompts are becoming increasingly elaborate and multi-modal [2]. For object detection, Open Vocabulary detection is gaining ground with the emergence of methods such as [3] that are based on textual prompts. However, it is not always easy or practical to describe an object in text only. Describing it using an image [4], a diagram or a CAD model may be more relevant. We will focus here on CAD models because they are increasingly used in many fields, starting with industry.

A thesis from the laboratory has already investigated the possibility of using CAD models as detector parameters to define the concepts to be located in images. Based on a CLIP-type representation [5] obtained by contrastive matching with the image, textual and CAD model representation [6], a proof of zero-shot detection using CAD models was produced. This representation tends to excessively erase the intra-class diversity of a semantic concept. In contrast, there are approaches that seek to estimate the pose of any object without tolerance to the deformation of the CAD model supplied in relation to the instance present in the image, using learning approaches [7] or 3D geometric matching after estimating a dense 3D descriptor of the scene from an image using a neural network [8].

It will also be necessary to graduate the response of the detector according to the type of deformation, more or less tolerable, of the CAD model in relation to the object present in the scene. The aim will be to teach a network how to find any previously defined 3D shape in an image, following the example of the old 'template matching' algorithms [9] but this time with all the contribution and power of neural networks. To achieve this, we will be focusing on simple shapes and orchestrating them into more complex shapes, in order to be able to address rigid objects first and then possibly articulated objects.



Can we detect the objects in the image on the right from a given composition of elementary shapes as a neural network prompt?

Références :
[1] T. B. Brown et al., « Language Models are Few-Shot Learners », 2020, arXiv: arXiv:2005.14165
[2] L. Yuan et al., « Florence: A New Foundation Model for Computer Vision », 2021, arXiv
[3] T. Cheng, L. Song, Y. Ge, W. Liu, X. Wang, et Y. Shan, « YOLO-World: Real-Time Open-Vocabulary Object Detection », 2024, CVPR
[4] Y. Zang, W. Li, K. Zhou, C. Huang, et C. C. Loy, « Open-Vocabulary DETR with Conditional Matching », 2022, ECCV
[5] A. Radford et al., « Learning Transferable Visual Models From Natural Language Supervision », 2021, arXiv
[6] D. Hegde, J. M. J. Valanarasu, et V. M. Patel, « CLIP goes 3D: Leveraging Prompt Tuning for Language Grounded 3D Recognition », 2023, arXiv
[7] Y.Xiao, Y.Du, R.Marlet, « PoseContrast: Class-Agnostic Object Viewpoint Estimation in the Wild with Pose-Aware Contrastive Learning », 2021, arXiv
[8] G.Pitteri, A.Bugeau, S.Ilic, V.Lepetit, « 3D Object Detection and Pose Estimation of Unseen Objects in Color Images with Local Surface Embeddings », 2020, arXiv
[9] R. Brunelli, « Template Matching Techniques in Computer Vision: Theory and Practice », Wiley, ISBN 978-0-470-51706-2, 2009

Internship objectives

- Develop zero-shot object detector algorithms using breakthrough neural network technology
- Analyze and summarize existing algorithms then propose realistic development plans.
- Identify and make effective use of the computing resources needed
- Keep environmental constraints in mind and integrate the development of solutions into a responsible approach.
- Collaborate and participate in the exchange of knowledge within the laboratory in order to cultivate the laboratory's scientific excellence.
- Promote the work carried out through communication media, demonstrators and/or scientific publications.

Qualifications

- Students in their 4th or 5th year of studies (M1, M2 or gap year)
- Computer vision skills and machine learning skills (deep learning, perception models, generative AI...)
- Python proficiency in a deep learning framework (especially TensorFlow or PyTorch)
- Experience in using 3D modeling or rendering frameworks (e.g. Blender, Unity, etc)
- Significant research experience involving the development of an innovative solution, quantitative evaluation and comparison with the state of the art is a plus
- Experience of communication and scientific publication with a reading committee is a plus

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