S05/2026 - Image Editing of Complex Visual Scene via Natural Language

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 its 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.

Missions

This internship focuses on the emerging field of **natural language-guided image editing**, specifically targeting the **generation and modification of complex scenes based on verbal descriptions**. The candidate will work on designing and implementing novel methods that can interpret natural language to manipulate or generate detailed images representing multifaceted scenarios (e.g., crowd scenes, cityscapes, interactions between multiple objects).

This project presents several key challenges, including:

- **Scene Complexity**: Managing multiple objects and their relationships in a scene adds significant complexity. The goal is to **maintain coherence and accuracy** in the edited images, even when the scenes described involve intricate interactions between various elements.
- **Multimodal Integration**: Successfully combining linguistic and visual inputs to obtain visual outputs, is a complex problem requiring seamless interaction between natural language processing (NLP) and computer vision models.

The objectives of this internship are to:

- Investigate current methods for natural language-based image generation and editing of complex scenes (in particular for the numerality and geometric positioning aspects);
- Develop an innovative approach for editing complex scenes using natural language descriptions;
- Demonstrate significant improvements in the accuracy and detail of generated images;
- Contribute to academic research through potential publications and/or patents.

Source : *IMAGHarmony: Controllable Image Editing with Consistent Object Quantity and Layout, arXiv:2506.01949*

"Four Eggs" → *"Four Cupcakes"*

Qualifications

- Students in their 5th year of studies (M2)
- Computer vision skills
- Machine learning skills (deep learning, LLM, VLM, generative AI…)
- Python proficiency in a deep learning framework (especially PyTorch or TensorFlow)

Job-related benefits

Join CEA List and LVA as an intern to:

- Work in one of the most innovative research organizations in the world, addressing societal challenges to build the world of tomorrow
- Discover a rich ecosystem: privileged connections between the industrial and academic sectors
- Conduct research autonomously and creatively: encouragement to valorize results (scientific articles, patents, open-source codes...)
- Join a young and dynamic team
- Benefit from an internal computing infrastructure with more than 300 state-of-the-art GPUs
- Receive a stipend between €1300 and €1400 per month
- Have the opportunity to continue with a PhD or as a research engineer after the internship
- Have the possibility of remote work, receive a 75% reimbursement on public transportation costs, and benefit from the "mobili-jeune" aid to reduce rent costs…

To apply, contact the laboratory with a CV and cover letter: angelique.loesch@cea.fr; romaric.audigier@cea.fr
More details on https://kalisteo.cea.fr/index.php/apply-for-a-job/