Robust Semantic Segmentation through Camera-3D Data Fusion for Autonomous Robots

Supervised by Riccardo Bertoglio

1 Research Proposal

In recent years, deep learning (DL) has enabled major advances in semantic segmentation for autonomous systems, providing fine-grained understanding of the driving environment. However, most vision-based models remain dangerously fragile. A critical point of failure lies in the sensitivity to appearance changes in the environment—such as lighting variations, shadows, and reflections—which can mislead RGB-based perception [1]. For example, a dark shadow on the road can be misclassified as an obstacle, potentially triggering unsafe reactions in autonomous vehicles.

The fundamental issue is that a camera-only model has no means to distinguish between an appearance change (a shadow) and a real physical change (an obstacle). To address this limitation, the research community has recently focused on **multi-modal sensor fusion**, combining complementary sensing modalities such as LiDAR (Light Detection and Ranging) and RGB cameras. Unlike cameras, LiDAR actively measures the 3D geometry of the scene, producing dense point clouds that are invariant to shadows, lighting, or color.

By fusing the reliable geometric information from LiDAR with the rich texture information from RGB cameras (see Figure 1), a deep learning model can learn to verify its own perception. It can, for instance, infer that a dark region in the image should not be considered an obstacle if no corresponding 3D structure is detected in the LiDAR signal.

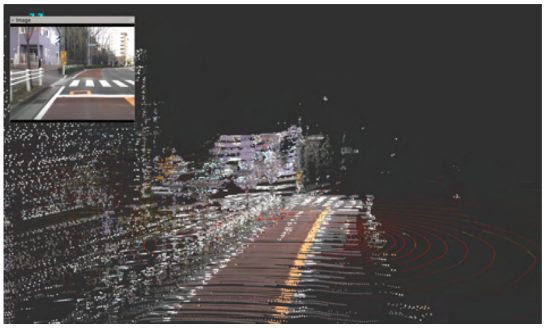

Figure 1: Fusion of 3D point cloud data and camera imagery: point cloud colored by the corresponding RGB color information. Image from [1].

This internship will explore **multi-modal fusion architectures** capable of combining LiDAR and RGB data to produce segmentation models robust to environmental appearance changes. Recent studies have shown that attention-based fusion modules, transformers, and cross-modal distillation strategies significantly improve performance in challenging lighting conditions [2], [3]. However,

key open questions remain about how to design architectures that best leverage the complementary strengths of the two modalities and maintain real-time efficiency for deployment on robotic platforms.

1.1 Objectives

The aim of this internship is to design and evaluate deep learning architectures for **robust semantic segmentation through camera-3D data fusion**. The main research hypothesis is that geometric features from 3D data can act as a reliable "physical check" against misleading visual cues, enhancing the robustness of semantic understanding in real-world scenes.

Specifically, the objectives are:

- Review state-of-the-art camera-3D data fusion methods for semantic segmentation in autonomous systems
- Implement baseline segmentation models using RGB only
- Investigate and advance the state-of-the-art of RGB–3D data fusion approaches
- Evaluate robustness to appearance variations (shadows, lighting changes) using public datasets and our own data
- Summarize findings and write a report

References

[1] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, "Perception and sensing for autonomous vehicles under adverse weather conditions: A survey," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 196, pp. 146–177, Feb. 2023.

[2] J. Gu, M. Bellone, T. Pivoňka, and R. Sell, "CLFT: Camera-LiDAR Fusion Transformer for Semantic Segmentation in Autonomous Driving," *IEEE Transactions on Intelligent Vehicles*, pp. 1–12, 2024, arXiv:2404.17793 [cs].

[3] H. Guan, C. Song, and Z. Zhang, "LiDAR-camera Cooperative Semantic Segmentation," en, *Machine Intelligence Research*, vol. 22, no. 5, pp. 956–968, Oct. 2025.

2 Research Environment

The National Research Institute for Agriculture, Food, and the Environment (INRAE) is a public research institution that ranks among the world's leading institutions in agricultural and food sciences, plant and animal sciences. INRAE addresses global challenges such as population growth,

climate change, resource scarcity, and biodiversity loss, advancing sustainable agricultural practices, quality food systems, and resource management.

The Technologies and Information Systems for Agro-Systems (TSCF) research unit at INRAE's Clermont-Ferrand center focuses on agro-equipment for ecological agricultural transitions. With 60 researchers, TSCF develops robots capable of precise, safe, and repeatable tasks in natural settings. Their work in perception and control algorithms enables these robots to perform agricultural tasks, advancing agroecology principles.

3 Required Qualifications

- Master's level engineering student / university student (niveau master 2)
- Proficiency in Python
- Foundational knowledge in Machine Learning
- Experience with ML frameworks such as PyTorch or TensorFlow
- Proficiency in English

4 Internship Conditions

- Location: INRAE/Unité TSCF, 9 avenue Blaise Pascal, 63170 Aubière
- Duration and Desired Period: 6 months (start date flexible, beginning of 2026)
- Compensation: Hourly stipend of 4.35€/h in accordance with current regulations
- Meals: Subsidized meals at the affiliated canteens
- Transportation: Partial reimbursement of public transport subscriptions

5 Applications

Please submit your application by filling out this form https://ec.europa.eu/eusurvey/runner/vision3dfusioninternship. Links to online repositories showcasing previous Python or ML projects will be highly appreciated. If your profile is shortlisted, you will be contacted regarding the next steps in the interview process.