Reliability-Aware Multimodal Fusion for Robust Semantic Segmentation in Autonomous Driving

Abdelhak Benamirouche[1,*], Lucas Deregnaucourt[2,*], Mihreteab Negash Geletu[1], Hind Laghmara[2], Remi Boutteau[2], and Jean-Philippe Lauffenburger[1]

[1]IRIMAS UR 7499, Université de Haute-Alsace, Mulhouse, France,
{abdelhak.benamirouche, mihreteab-negash.geletu,
jean-philippe.lauffenburger}@uha.fr
[2]INSA Rouen Normandie, Univ Rouen Normandie, Université Le Havre Normandie, Normandie Univ, LITIS UR 4108, F-76000 Rouen, France,
{lucas.deregnaucourt@insa-, hind.laghmara@insa-, remi.boutteau@univ-}rouen.fr
[*]Equal Contribution.

This abstract summarizes a paper published at the *2026 IEEE Intelligent Vehicles Symposium (IV)* [1].

Abstract

Robust semantic segmentation is essential for autonomous driving, yet real-world perception is strongly affected by adverse weather, sensor degradation, and partial sensor failures. While multimodal fusion is widely used to improve robustness, many approaches treat all modalities equally and do not explicitly account for their varying reliability. In this work, we present ReCoLaF, a reliability-aware conflict-guided late fusion framework for multimodal semantic segmentation under uncertainty. ReCoLaF combines evidential deep learning with Dempster–Shafer theory and introduces a two-stage trust modeling strategy: a class-wise reliability mechanism that estimates the semantic relevance of each modality, followed by a conflict-guided adjustment that reduces the contribution of inconsistent sources at the pixel level. Experiments on the DeLiVER and MUSES datasets show that ReCoLaF improves robustness under degraded sensing conditions and sensor failures, achieving the best average performance across modality-availability scenarios. These results highlight the interest of explicitly modeling both reliability and conflict for trustworthy multimodal perception in autonomous driving.

1 Summary

Semantic segmentation is a core perception task for autonomous driving, but its reliability can decrease significantly under difficult operating conditions such as fog, rain, illumination changes, or sensor failures. Multimodal perception helps overcome the limitations of individual sensors by combining complementary information from cameras, LiDAR, radar, or depth sensors. However, not all modalities are equally informative in all situations, and a robust fusion strategy should be able to account for this variability.

To address this issue, we proposed **ReCoLaF** (*Reliability-aware Conflict-guided Late Fusion*), an evidential multimodal fusion framework designed for semantic segmentation under uncertainty. The method is grounded in *Dempster–Shafer theory of evidence* [2], which enables explicit representation of uncertainty during fusion. ReCoLaF models trust in two complementary stages. First, each modality is weighted through *class-wise reliability scores*, which capture how relevant a given sensor is for predicting each semantic class. Second, a *conflict-based adjustment* is applied at the pixel level to reduce the influence of modalities that disagree with the others. The discounted evidential outputs are then fused to obtain the final segmentation.

The proposed approach was evaluated on two multimodal driving datasets: **DeLiVER** [5], a synthetic dataset with multiple weather and degradation conditions, and **MUSES** [6], a real-world dataset acquired under uncertain driving conditions. The evaluation protocol focuses on *sensor-failure robustness*: all modalities are used during training, while at inference time one or several modalities may be removed to simulate degraded sensing conditions.

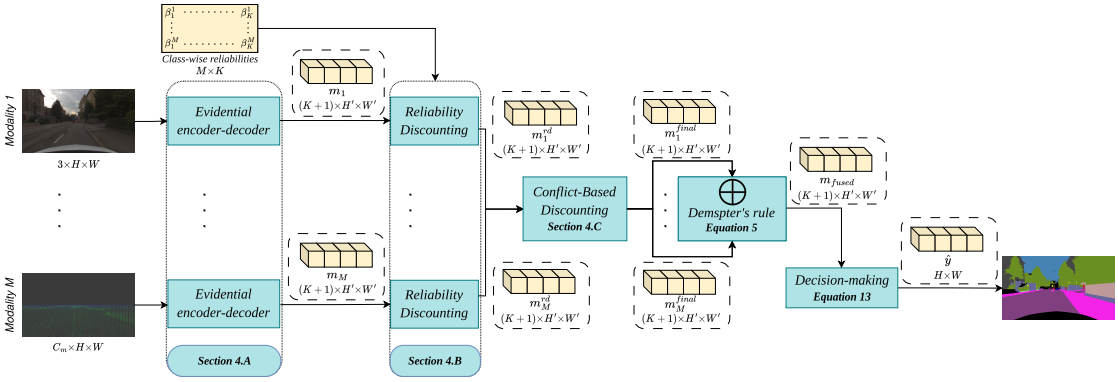


Figure 1: Overview of the proposed ReCoLaF architecture [1]. Each modality is processed by an evidential encoder-decoder, followed by class-wise reliability discounting, conflict-guided adjustment, and evidential fusion.

Results show that ReCoLaF provides the best average robustness across modality-availability scenarios. On DeLiVER, ReCoLaF reaches an average mIoU of **34.82%**, outperforming CMNeXt (30.94%), CAFuser (33.35%), and ECoLaF (32.35%). On MUSES, ReCoLaF achieves **45.85%** average mIoU, again ahead of CMNeXt (36.38%), CAFuser (42.02%), and ECoLaF (41.04%). These results indicate that explicitly modeling both semantic reliability and inter-modality disagreement leads to a more stable fusion behavior under missing or unreliable sensor conditions.

This work is particularly relevant to trustworthy and applied AI, since it addresses a key challenge in real-world autonomous systems: ensuring reliable perception despite uncertain, incomplete, or degraded observations. By integrating uncertainty representation with adaptive multimodal fusion, ReCoLaF contributes to the design of safer and more robust AI perception systems.

References

[1] L. Deregnaucourt, A. Benamirouche, M. N. Geletu, H. Laghmara, R. Boutteau, and J.-P. Lauffenburger, "Reliability-Aware Fusion for Semantic Segmentation under Sensor Degradation and Failures," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, 2026.

[2] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[3] L. Deregnaucourt, H. Laghmara, A. Lechervy, and S. Ainouz, "A conflict-guided evidential multimodal fusion for semantic segmentation," in *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, 2025, pp. 1373–1382.

[4] T. Brödermann, C. Sakaridis, Y. Fu, and L. Van Gool, "CAFuser: Condition-aware multimodal fusion for robust semantic perception of driving scenes," *IEEE Robotics and Automation Letters*, vol. 10, no. 4, pp. 3134–3141, 2025.

[5] J. Zhang, R. Liu, H. Shi, K. Yang, S. Reiss, K. Peng, H. Fu, K. Wang, and R. Stiefelhagen, "Delivering arbitrary-modal semantic segmentation," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[6] T. Brödermann, D. Bruggemann, C. Sakaridis, K. Ta, O. Liagouris, J. Corkill, and L. Van Gool, "MUSES: The multi-sensor semantic perception dataset for driving under uncertainty," in *European Conference on Computer Vision*, 2024.