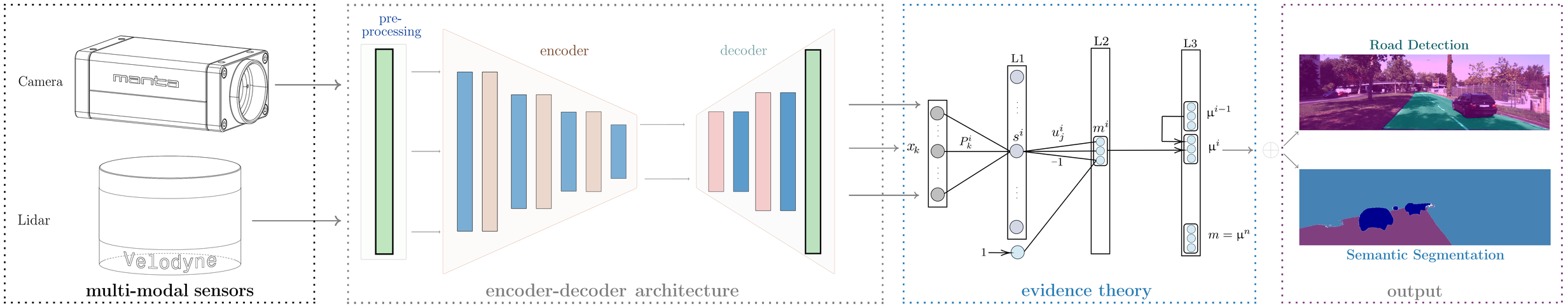
Evidential Deep-Learning-Based Multi-Modal Environment Perception for Intelligent Vehicles

M.N. GELETU [1,2,*], D-V. GIURGI [1,*], T. JOSSO-LAURAIN [1], M. DEVANNE [1], J-P. LAUFFENBURGER [1], M. M. WOGARI[2]

[1]IRIMAS - Université de Haute-Alsace, [2]AAiT - Addis Ababa University



Background - Evidence Theory

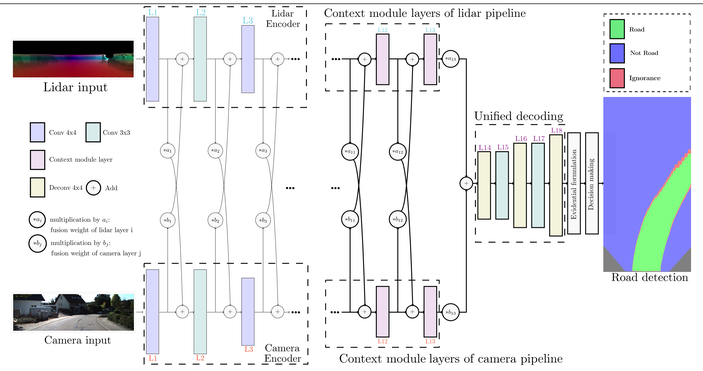
- Belief function (Bel), respectively Plausibility function (Pl):

$$\mathbf{Bel(A)} = \sum_{\mathbf{B \subseteq A}} \mathbf{m(B)}, \qquad \mathbf{Pl(A)} = \sum_{\mathbf{B \cap A \neq \emptyset}} \mathbf{m(B)} = \mathbf{1 - Bel(\bar{A})} \quad (1)$$

- 2 BBAs m_1 and m_2 can be combined by Dempster's rule defined as:

$$(\mathbf{m_1 \oplus m_2})(\mathbf{A}) = \frac{\mathbf{1}}{\mathbf{1-k}} \sum_{\mathbf{B \cap C = A}} \mathbf{m_1(B) m_2(C)}, \qquad \mathbf{k} = \sum_{\mathbf{B \cap C = \emptyset}} \mathbf{m_1(B) m_2(C)} \quad (2)$$

Evidential Cross-Fusion Architecture



Lite-CF-Evi architecture: two pipelines of **13 layers** each (lidar and camera)

Road Detection

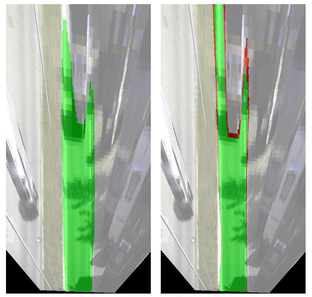
BBA Modeling - Distance to prototypes [2]:

$$\mathbf{d^i} = \|\boldsymbol{x} - \boldsymbol{p^i}\|, \ \mathbf{i = 1, \cdots, n} \quad (3)$$

$$\begin{aligned} \mathbf{m^i}(\{\omega_j\}) &= \alpha^i u_j^i \phi^i(d^i), \ j = 1, 2 \\ m^i(\Omega) &= 1 - \alpha^i \phi^i(d^i) \end{aligned} \quad (4)$$

$$\phi^{\mathbf{i}}(\mathbf{d^i}) = \exp\left(-\gamma^{\mathbf{i}}(\mathbf{d^i})^{\mathbf{2}}\right), \ \gamma^{\mathbf{i}} > \mathbf{0} \quad (5)$$

Perspective view: camera



(a) **Probabilistic**: Lite-CF (b) **Evidential**: Lite-CF-Evi2

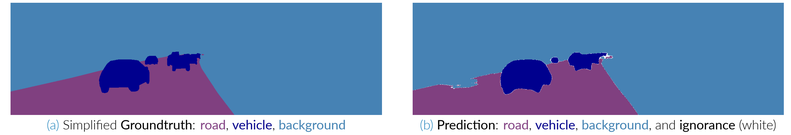
Bird-eye-view projection on road detection

The table below shows the performances in terms of mean and standard deviation among the 10-folds. **Lite-CF-Evi1** has a **15.7% reduction** in model parameters vs. the Baseline (CF). The increase in the $MaxF$ value and the associated small standard deviation can be attributed to the **evidential representation** and **fusion**.

Table: *Model performance evaluation*

Model arch.	# model param.	MaxF	PRE	REC	ER	FPS
Baseline (CF) [1], [3]	$3,246,830$	96.25 ± 0.71	96.46 ± 0.66	96.05 ± 1.06	1.34 ± 0.26	27
Lite-CF [3]	$2,737,213$	95.50 ± 0.52	95.57 ± 0.69	95.45 ± 0.74	1.61 ± 0.21	**35**
Lite-CF-Evi1	$\mathbf{2,737,066}$	$\mathbf{96.91 \pm 0.36}$	$\mathbf{96.74 \pm 0.56}$	$\mathbf{97.09 \pm 0.71}$	1.11 ± 0.14	33
Lite-CF-Evi2	$\mathbf{2,737,066}$	-	-	-	$\mathbf{0.81 \pm 0.13}$	29

Semantic Segmentation



(a) Simplified **Groundtruth**: road, **vehicle**, background (b) **Prediction**: road, **vehicle**, background, and **ignorance** (white)

- 127 images of KITTI semantic, 3 singletons: $\Omega = \{road, car, background\}$. A 4^{th} class (***ignorance***) is introduced to represent **uncertainties** (far-end points).

Conclusions

The **belief theory** introduces **ignorance** enabling cautious decision-making and multi-class versatility. The **evidential cross-fusion deep learning** model handles better the uncertainties in the **perception** environment. Performance improvements are achieved for both **road detection** and **semantic segmentation** tasks.

References

[1] Luca Caltagirone, Mauro Bellone, Lennart Svensson, and Mattias Wahde. Lidar–camera fusion for road detection using fully convolutional neural networks. *Robotics and Autonomous Systems*, 111:125–131, 2019.

[2] Thierry Denoeux. A neural network classifier based on dempster-shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 30(2):131–150, 2000.

[3] M. N. Geletu, T. Josso-Laurain, M. Devanne, M. M. Wogari, and J-P. Lauffenburger. Deep learning based architecture reduction on camera-lidar fusion for autonomous vehicles. In *2nd International Conference on Computers and Automation (CompAuto)*, pages 1–7. CPS, 2022.

Code repository:

mihreteab-negash.geletu@uha.fr

vasile.giurgi@uha.fr