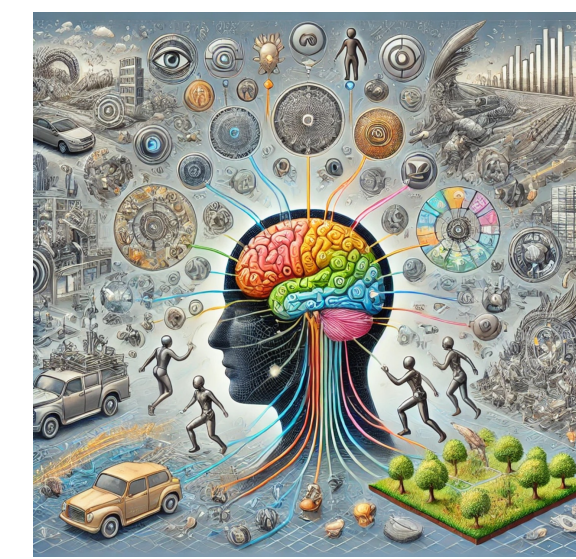


Centering the Value of Every Modality: Towards Efficient and Resilient Modality-agnostic Semantic Segmentation

Xu Zheng¹ · Yuanhuiyi Lyu¹ · jiazhou zhou¹ · Lin Wang^{1,2}
¹HKUST(GZ), ²HKUST

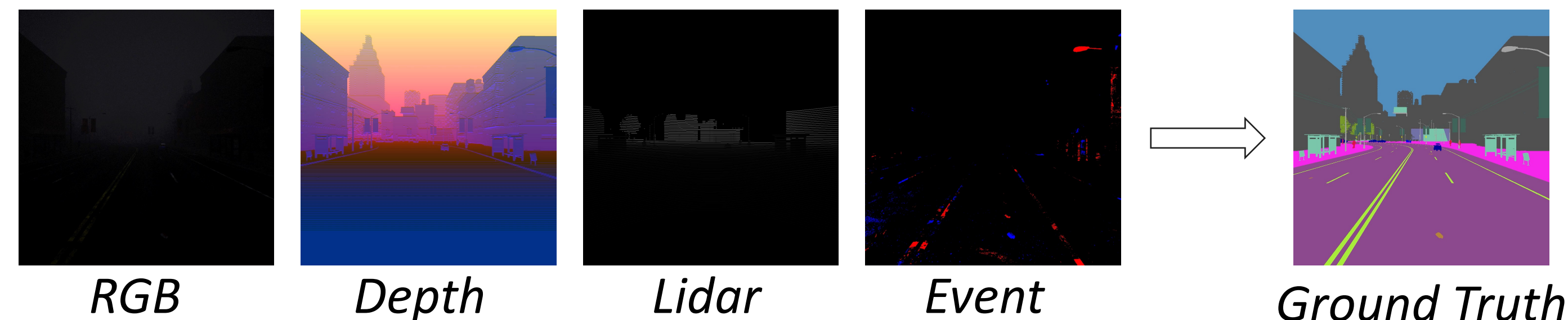
Project Page: <https://vlislab22.github.io/MAGIC/>

Why we need robust multimodal fusion?



Nature has elucidated that miscellaneous sense and processing capabilities of **visual** information are vital for the **understanding** of the sophisticated environment

Multi-modal Sensing @ Condition: Fog + Underexposure



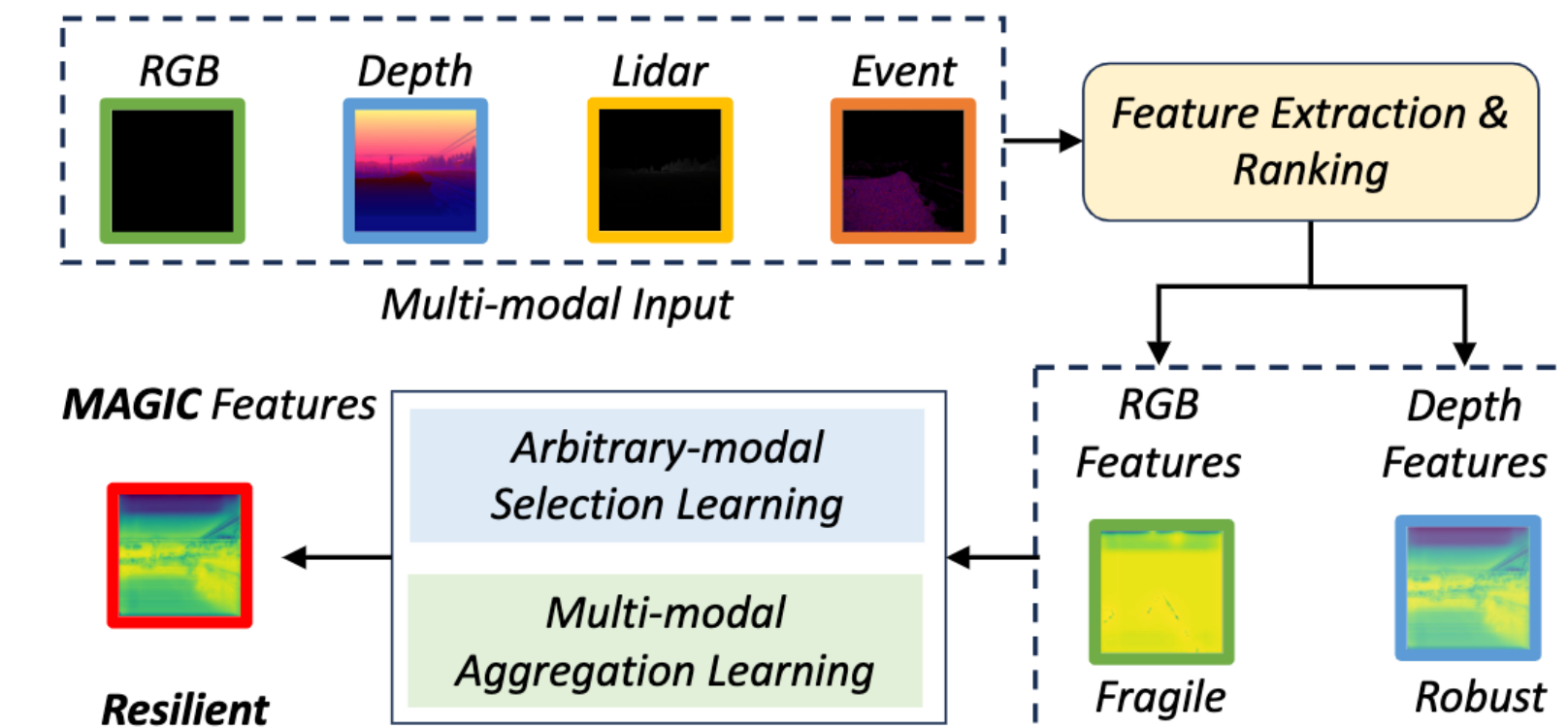
Each sensor delivers its **distinct** advantages in certain conditions.

Why we focus modality-agnostic fusion?

1. Fusing arbitrary modalities is vital yet remains less explored.
2. RGB sensors struggle in certain circumstances.

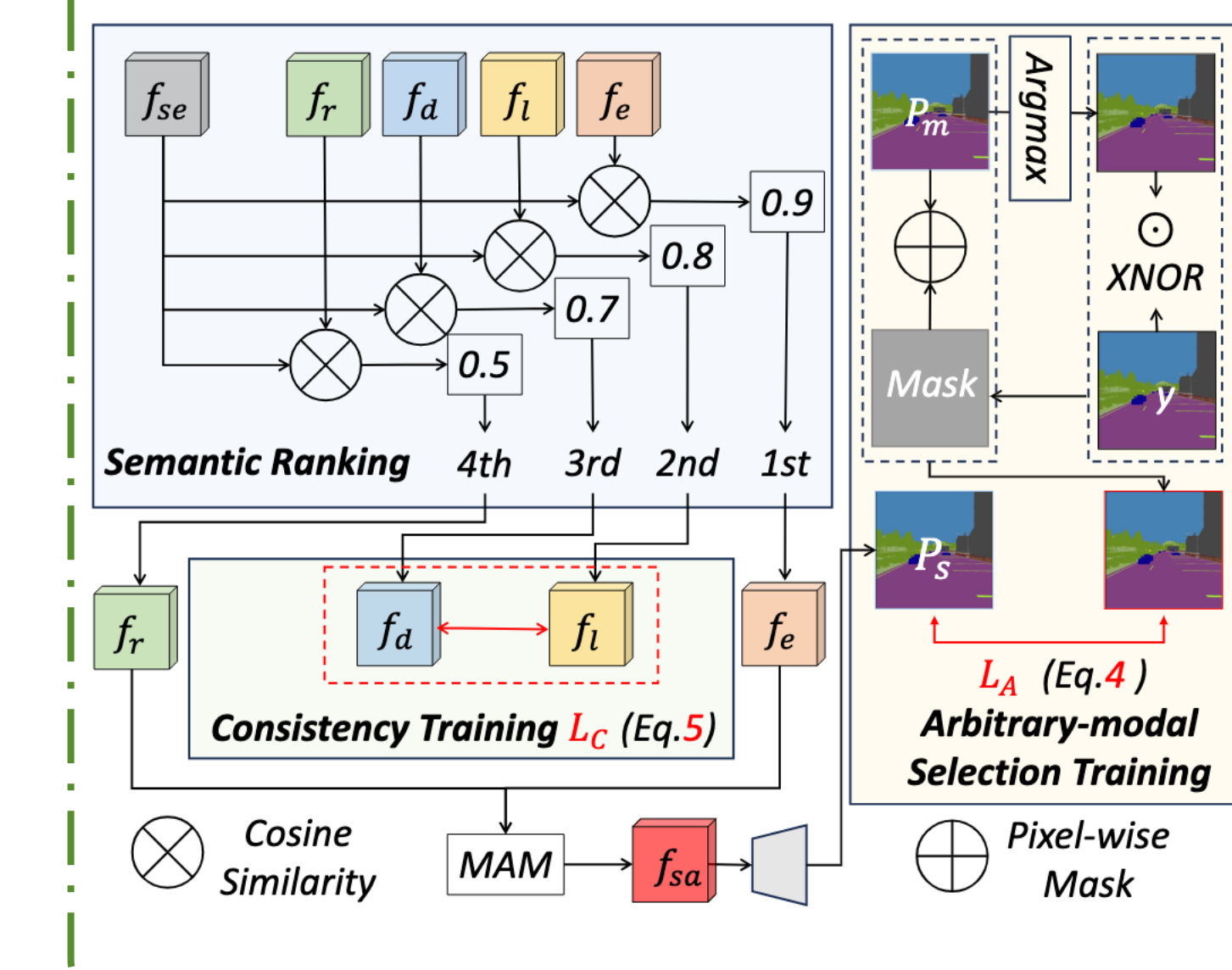
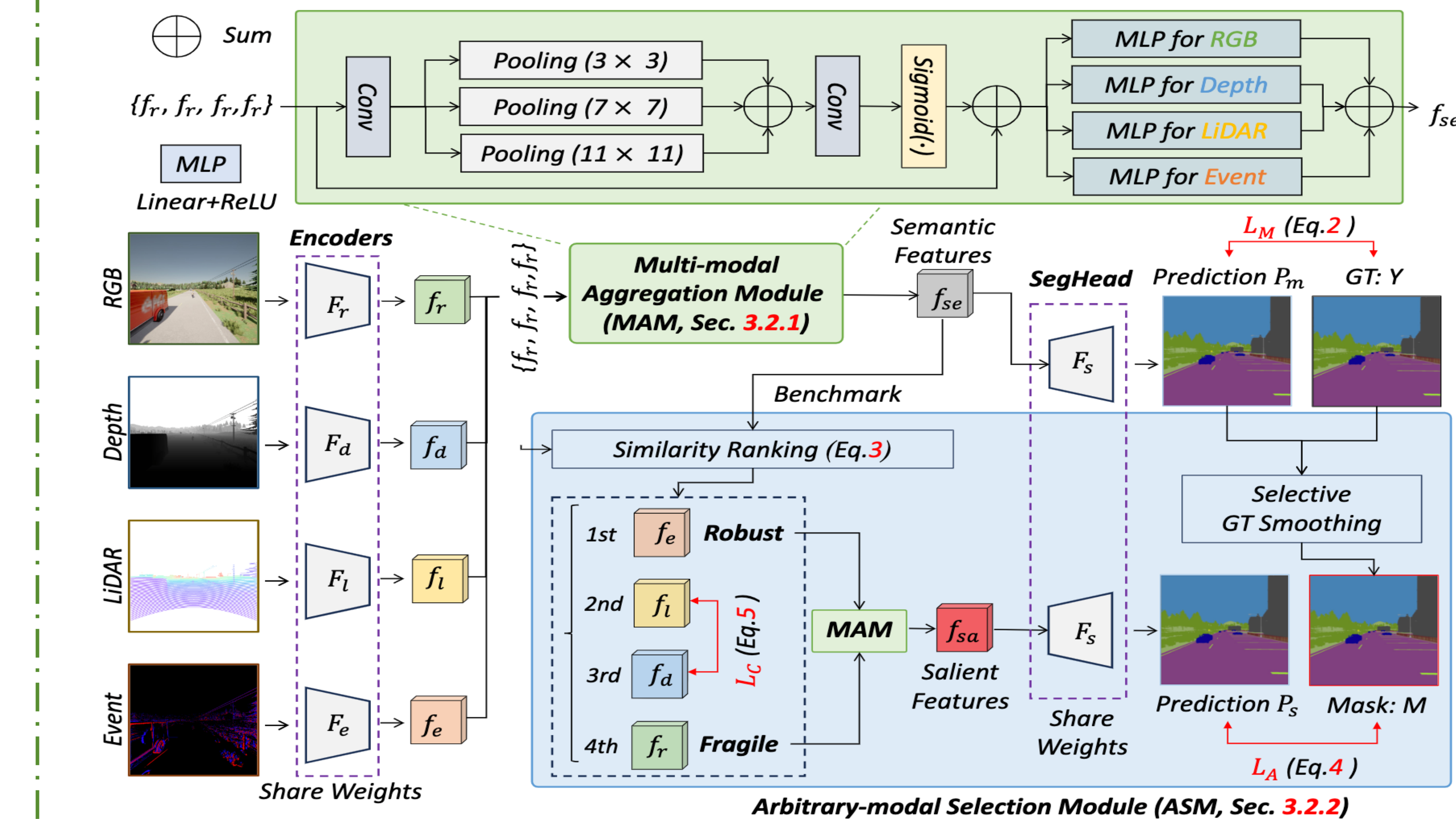
Centering the value of every modality:

1. Discern **robust** and **fragile** modalities;
2. Incorporate the most robust and fragile ones to learn a **resilient** multimodal framework.



Methodology

Overall Framework

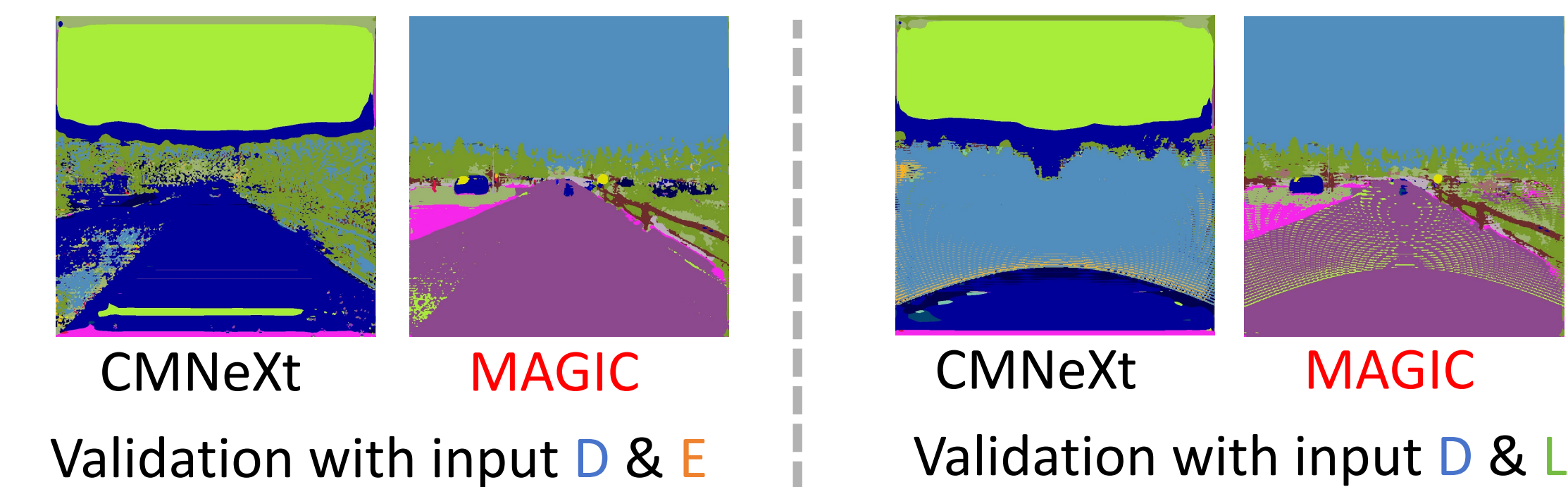
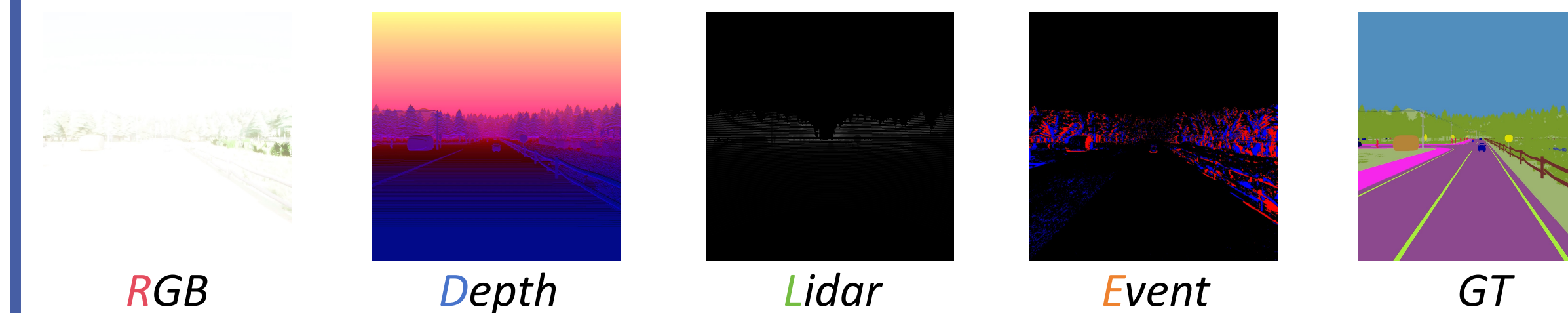


Arbitrary-modal Selection Module (ASM)

1. Cross-modal Semantic Similarity Ranking adeptly discern the **robust** modalities from the **fragile** ones at the feature level.
2. Cross-modal Semantic Consistency Training impose **semantic consistency training** between the remaining features, since the captured semantics of a scene are identical across modalities.

Results

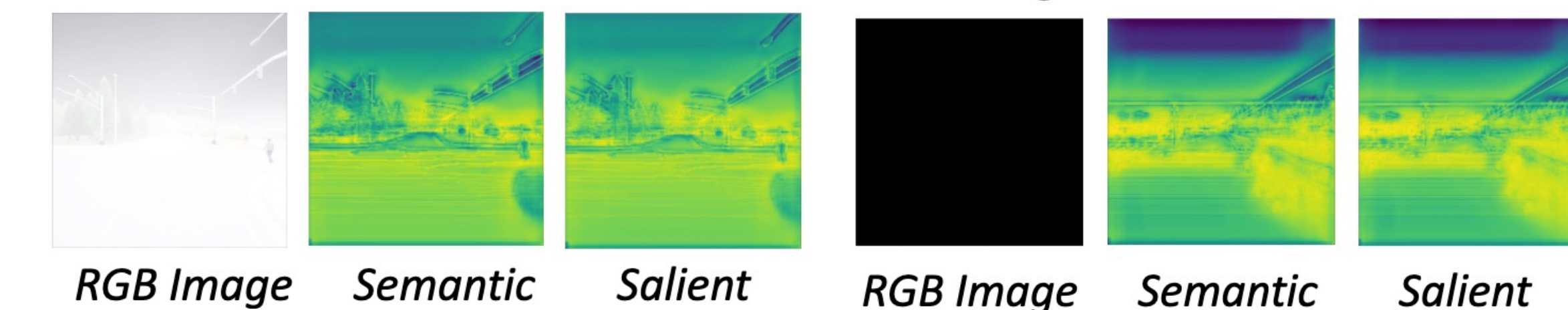
Multi-modal data @ Conditions: Sun + Underexposure



Method	Backbone	R	D	E	L	RD	RE	RL	DE
CMNeXt	Seg-B2	3.76	0.81	1.00	0.72	50.33	13.23	18.22	21.48
MAGIC	Seg-B0	32.60	55.06	0.52	0.39	63.32	33.02	33.12	55.16
	Seg-B2	41.97	57.59	0.40	0.37	67.65	41.93	42.00	57.62

Method	Backbone	DL	EL	RDE	RDL	REL	DEL	RDEL	Mean
CMNeXt	Seg-B2	3.83	2.86	66.24	66.43	15.75	46.29	66.30	25.25
MAGIC	Seg-B0	55.17	0.26	63.37	63.36	33.32	55.26	63.40	40.49
	Seg-B2	57.60	0.27	67.66	67.65	41.93	57.63	67.66	44.66

Validation with arbitrary-modal inputs.



Visualization of both semantic and salient features.