E²SAM A Pipeline for Efficiently Extending SAM's Capability on Cross-Modality Data via Knowledge Inheritance

Sundingkai Su, Mengqiu Xu, Kaixin Chen, Ming Wu, Chuang Zhang

BMVC 2023

Contributions

- Proposing a pipeline to efficiently extending SAM's[1] capability on cross-modality.
- No need to introduce additional data and any annotations since the model can be trained unsupervised.
- Evaluation on two Multimodal datasets
  - SFDD-H8[2]
  - RGB-D SHIFT[3]

Analyzing

Data		Model	CSI	mIou	mAcc	allAcc
SFDD-H8	C3*-1024	Rand	34.66	65.72	72.63	96.83
		SAM	57.76 (+23.10)	77.88 (+12.16)	86.27 (+13.64)	98.06 (+1.23)
	C16-1024	Rand	52.18	74.86	85.40	74.86
		SAM-ReSam	59.78 (+7.60)	78.88 (+4.02)	**89.69 (+3.42)**	98.04 (+0.43)
		SAM-RePE	56.96 (+4.15)	77.46 (+2.78)	85.79 (+0.39)	98.02 (+0.41)
		Ours	**61.57 (+9.39)**	**79.91 (+5.05)**	87.52 (+2.21)	**98.29 (+0.66)**
	C3*-512	SAM	**60.26**	78.81	**90.06**	98.00
	C16-512	Ours	59.92	**79.01**	87.78	**98.15**
SHIFT	C3-1024	Rand	–	63.72	73.92	94.85
	C3-1024	SAM	–	73.15 (+9.43)	82.31 (+8.39)	96.56 (+1.71)
	C4-1024	Rand	–	78.17	86.44	98.04
	C4-1024	Ours	–	**81.78 (+3.62)**	**89.25 (+2.81)**	**98.59 (+0.55)**

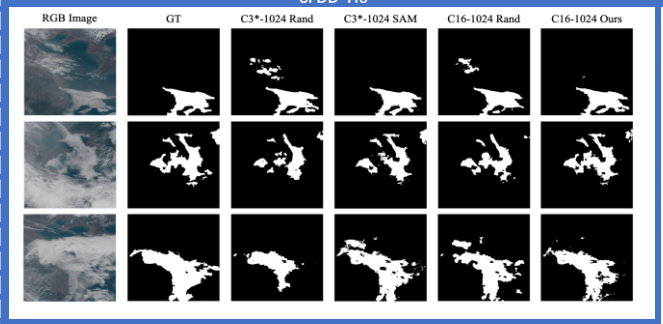
Model		All	Daytime	Dawn/dusk	Night
All-C4	Rand	78.17	78.00	79.41	77.20
All-C4	Ours	81.78 (+3.62)	81.58 (+3.58)	82.51 (+3.10)	81.12 (+3.92)
Daytime-C4	Rand	76.03	73.50	63.62	55.05
Daytime-C4	Ours	76.57 (+0.54)	77.86 (+4.18)	77.52 (+13.90)	73.50 (+18.45)
Dawn/dusk-C4	Rand	61.82	60.25	63.78	62.88
Dawn/dusk-C4	Ours	73.58 (+11.76)	73.01 (+12.76)	74.24 (+10.46)	73.61 (+10.73)
Night-C4	Rand	52.88	43.63	59.36	64.58
Night-C4	Ours	72.79 (+19.91)	71.49 (+27.86)	73.01 (+13.38)	74.21 (+9.63)

Tests under different light conditions

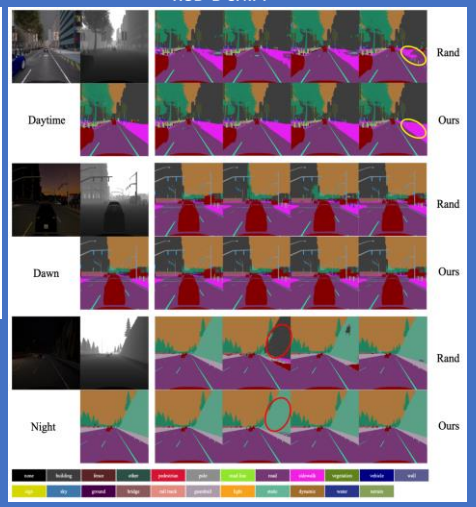
To demonstrate the effectiveness of our proposed method, we design two sets of controlled experiments by comparing the input of different modalities and the advantages of inheriting the powerful capabilities of SAM. The quantitative results are as shown in tables, where 'Rand' indicates random initialization and 'SAM-' means performing different operations based on SAM as the pre-trained model.

Data visualisation

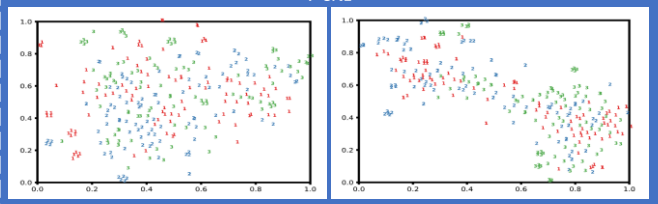
SFDD-H8



RGB-D SHIFT



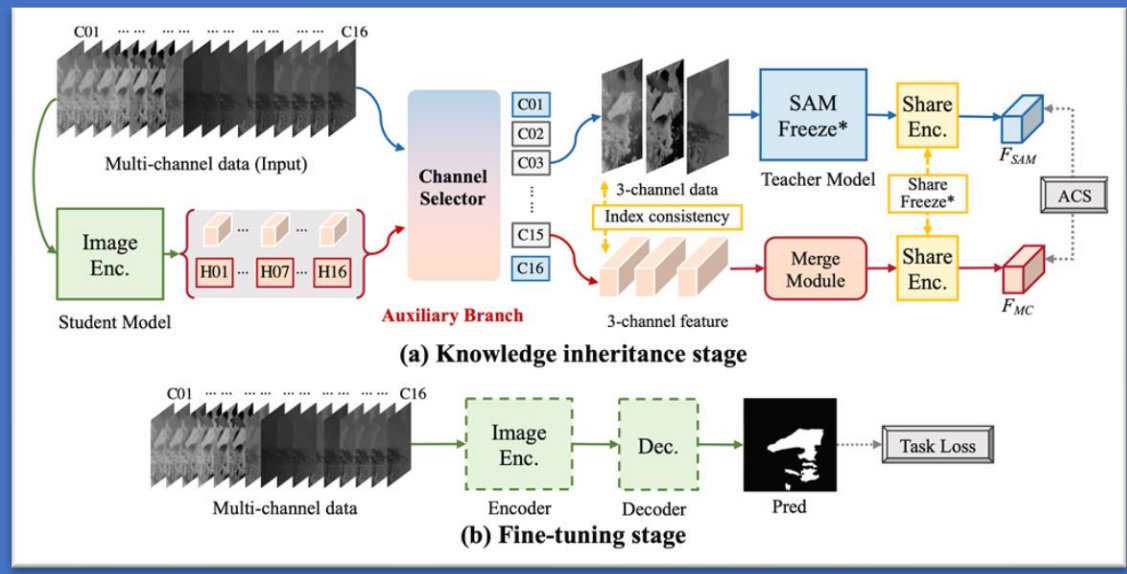
T-SNE



Our approach

Rand init

Method



Conclusion

In this paper, in order to realize the capability expansion of SAM based on cross-modality data,

we propose a universal two-stages pipeline, which is the knowledge inheritance stage for inheriting SAM capability, and the fine-tuning stage for better downstream adaptation. An auxiliary branch including a Channel Selector and Merge Module is designed to separate different cross-modality to achieve feature alignment. It is worth mentioning that we do not need to lead into additional data and labels during the knowledge inheritance process, reducing the cost of collecting and annotating data. Through meticulous experiments and visualization results, it can be demonstrated that our method can efficiently inherit the ability of SAM on cross-modality data without compressing data

References

1. Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. arXiv preprint arXiv:2304.02643, 2023.
2. Bin Huang, Ming Wu, Shuyue Sun, Wei Zhao, Zhanbei Cui, and Cheng Lv. Sea fog monitoring method based on deep learning satellite multi-channel image fusion (in chinese). Meteorological Science and Technology, 49(6):823–829, 2021.
3. Tao Sun, Mattia Segu, Janis Postels, Yuxuan Wang, Luc Van Gool, Bernt Schiele, Federico Tombari, and Fisher Yu. Shift: a synthetic driving dataset for continuous multitask domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 21371–21382, 2022.