

Multimodal Semantic Segmentation Challenge

Bellaterra, Spain Nov. 12 – 14, 2025

Baseline Results:

The baseline results are obtained by selecting state-of-the-art semantic segmentation models from existing literatures for quantitative performance assessment using multimodal RS data from the **ISASeg**. Specifically, the selected baseline model is U-Net [1] and Mask2former [2] with two testing formats (single-modal input and concatenation-based multimodal input).

The winning solution of this competition should at least outperform the listed baseline results in order to show the superior capability of the newly developed semantic segmentation models with ISASeg multimodal datasets.

Baseline results for Dataset ISASeg:

	Modality	Validation set			Test set		
		mF1	mIoU	OA	mF1	mIoU	OA
U-Net	RGB	83.41	73.64	93.30	78.23	67.38	90.98
	RGB+S2	83.70	74.00	93.38	79.58	68.95	91.58
	RGB+S1	83.59	73.81	93.07	79.49	68.75	91.18
	RGB+S1+S2	83.79	74.15	93.59	79.10	68.49	91.83
Mask2former	RGB	83.46	73.74	93.48	78.54	67.86	91.66
	RGB+S2	84.82	75.47	93.94	80.76	70.43	92.46
	RGB+S1	84.26	74.77	93.88	79.46	68.93	92.08
	RGB+S1+S2	84.93	75.01	93.95	80.63	70.25	92.33

Reference:

- [1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.
- [2] Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., and Girdhar, R. (2022). Masked-attention mask transformer for universal image segmentation. Paper presented at: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.