Supplementary Material

1. Supplementary Implementation Details

Since the SAM2 backbone only accepts square-shaped images, we design two processing pipelines for source model training to ensure generalization across different cases: one for rectangular inputs and another for square-shaped inputs.

Rectangular Source Image. We apply a sliding window approach to segment the image into sequences of square patches, maintaining spatial continuity. These video-like sequences are then fed into the model to train the backbone, segmentation head, and memory modules. At each prediction step within an input sequence, the backbone feature integrates past memory embeddings through memory attention, allowing for more temporally consistent predictions. Subsequently, the current output mask and backbone feature are passed to the memory encoder, where the resulting embedding is stored in the memory bank for future reference.

Square Source Image. For inherently square-shaped inputs (*Stanford2D3D-Pinhole*), treating them in the same manner as rectangular images would be unnatural. In this case, a trade-off is made by directly training the source model on these squared image-mask pairs without utilizing the memory mechanism, keeping the parameters of the memory attention and memory encoder frozen. Given that panoramic images are predominantly rectangular, the memory-related modules are fine-tuned on target domain data in the UDA stage to enhance feature alignment and improve segmentation performance in the panoramic domain.

Training and Adaptation Details. During training, We introduce randomness to the window's sliding direction, including forward sliding (left-to-right) and reverse sliding (right-to-left). This helps model learn more cross-patch dependencies. For target domains in our domain adaptation, the SPan dataset contains 4096×2048 images, which include inherent black regions brought by equirectangular projection. Hence we remove these redundant areas by cropping the images and downscale them to 3072×1024. The corresponding sliding stride is 256. The DensePASS dataset contains 2048×400 images, which are resized to 3072×1024, with a corresponding sliding stride of 256. Besides, we sample 400 images from the target domain for pseudo-labels updating in each epoch.

2. Supplementary Experimental Results

2.1. Trainable Parameters

We evaluate the trainable parameters of our OmniSAM model by examining the impact of adding or removing the memory attention module. The parameter sizes for different variants of the OmniSAM model are presented in the Table 1. Without memory attention, fine-tuning via LoRA-based adaptation affects only a small subset of the model, resulting in minimal trainable parameters across all variants. The smallest (OmniSAM-T) contains 0.36 MB, while the largest (OmniSAM-L) has 0.70 MB. In contrast, the memory attention module contributes approximate 5.65 MB parameters. Specifically, OmniSAM-T increases from 0.36 MB to 6.01 MB, and OmniSAM-L grows from 0.70 MB to 6.35 MB.

Network	Trainable Param. (MB)
OmniSAM-T w/o MA	0.36
OmniSAM-S w/o MA	0.39
OmniSAM-B w/o MA	0.45
OmniSAM-L w/o MA	0.70
OmniSAM-T w/ MA	6.01
OmniSAM-S w/ MA	6.04
OmniSAM-B w/ MA	6.10
OmniSAM-L w/ MA	6.35

Table 1. Comparison of trainable parameters for different variants of OmniSAM.

2.2. Total Parameters and Computation Costs

The total parameters and computation costs of our OmniSAM variants are presented in the Table 2. The ***Tiny*** and ***Small*** variants exhibit competitive parameter counts, similar to state-of-the-art methods. While the ***Large*** (209.2M) and the ***Base*** (72.0M) variants significantly increase the model size, which may limit its real-time application. A trade-off in inference speed should be considered while choosing the model for specific task.

Network	Param. (M)	FLOPs (G)
Trans4PASS+-S	44.9	251.1
DATR-S	25.8	139.2
360SFUDA++	28.7	148.0
OmniSAM-T w/ MA	32.0	118.8
OmniSAM-S w/ MA	38.8	149.8
OmniSAM-B w/ MA	72.0	280.8
OmniSAM-L w/ MA	209.2	828.0

Table 2. Computatioal costs for networks with 1024^2 resolution input.

2.3. Ablation of λ.

We conduct experiments on OmniSAM-B in the outdoor scenario to evaluate the impact of λ. As shown in Table 3, OmniSAM-B achieves the highest score in mIoU while $\lambda = 0.1$.

2.4. Ablation Study on Memory Mechanism

Table 4 and Table 5 present additional class-specific results of our OmniSAM in the real-world scenario. The outcomes also serve as an ablation study on the memory mechanism of the model.

λ	0	0.01	0.1	0.2	0.5	1.0
mIoU	56.49	61.64	**62.46**	61.59	61.44	60.90
Δ	-	+5.15	**+5.97**	+5.10	+4.95	+4.41

Table 3. Ablation Study of λ.

2.5. More Visualization Results

Fig. 1 and Fig. 2 present segmentation results from our proposed model and baseline methods [1, 2] across various settings. Additionally, the t-SNE visualization shown in Fig. 3 further highlights the effectiveness of our adaptation approach by illustrating distinct and informative feature representations for each semantic category.

References

[1] Xu Zheng, Tianbo Pan, Yunhao Luo, and Lin Wang. Look at the neighbor: Distortion-aware unsupervised domain adaptation for panoramic semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18687–18698, 2023. 2

[2] Xu Zheng, Peng Yuan Zhou, Athanasios V Vasilakos, and Lin Wang. 360sfuda++: Towards source-free uda for panoramic segmentation by learning reliable category prototypes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2

Method	Network	mIoU	Road	S.Walk	Build.	Wall	Fence	Pole	Tr.L	Tr.S	Veget.	Terrain	Sky	Person	Car
Source-only	OmniSAM-T w/o MA	49.79	71.91	26.12	86.42	36.34	39.73	36.29	8.27	13.04	79.47	20.92	94.16	59.10	75.53
	OmniSAM-S w/o MA	53.98	75.28	36.88	87.69	42.41	40.89	38.45	19.19	14.03	80.08	27.48	94.36	63.59	81.38
	OmniSAM-B w/o MA	55.03	76.69	38.98	88.60	42.56	48.07	40.18	20.43	14.79	81.04	29.38	94.83	60.73	79.15
	OmniSAM-L w/o MA	54.23	75.70	36.29	87.23	41.29	47.53	38.98	20.52	14.29	81.18	28.75	94.54	57.16	81.56
	OmniSAM-T w/ MA	54.99	75.48	36.30	88.60	39.74	41.35	42.29	20.32	24.57	78.35	32.43	94.23	62.5	78.65
	OmniSAM-S w/ MA	55.40	76.86	36.92	88.64	39.89	41.31	42.70	25.11	20.19	78.14	32.76	94.68	61.9	81.11
	OmniSAM-B w/ MA	56.49	74.41	43.58	87.65	**44.80**	46.91	45.17	16.24	22.24	80.15	32.41	94.86	65.31	80.67
	OmniSAM-L w/ MA	56.61	77.52	43.45	88.50	36.51	51.65	38.12	20.96	21.70	81.48	31.65	94.62	70.88	78.95
Ours	OmniSAM-T w/o MA	53.73	79.03	42.19	86.09	28.28	45.95	35.19	10.20	20.53	79.41	35.49	94.40	63.81	77.88
	OmniSAM-S w/o MA	57.03	78.99	49.57	88.77	38.48	47.47	38.77	21.84	15.81	81.12	39.32	94.72	65.36	81.22
	OmniSAM-B w/o MA	59.34	81.69	53.87	89.33	39.74	50.84	41.98	20.54	21.50	81.71	44.63	95.06	68.13	82.34
	OmniSAM-L w/o MA	59.02	83.49	56.14	88.29	34.29	52.39	38.81	23.97	19.83	**82.52**	44.84	**95.26**	61.97	**85.43**
	OmniSAM-T w/ MA	59.01	79.95	45.78	88.03	39.74	47.99	42.68	26.69	29.55	78.00	42.98	94.56	68.23	82.98
	OmniSAM-S w/ MA	60.23	80.76	46.36	89.79	44.46	48.68	**45.32**	29.33	25.15	79.51	46.28	94.41	68.90	84.06
	OmniSAM-B w/ MA	**62.46**	**84.02**	**56.23**	89.93	44.01	54.54	44.50	25.19	**33.42**	81.77	**49.16**	94.69	**71.64**	82.89
	OmniSAM-L w/ MA	61.63	82.45	53.65	**90.05**	44.00	**54.75**	43.36	**30.99**	28.27	80.04	43.59	94.48	70.70	84.88

Table 4. Per-class results of the Cityscapes13-to-DensePASS13 scenario (* denotes the baseline)

Method	Network	mIoU	Ceiling	Chair	Door	Floor	Sofa	Table	Wall	Window
Source-only	OmniSAM-T w/o MA	66.70	91.34	62.32	32.20	93.75	37.62	69.98	80.64	65.73
	OmniSAM-S w/o MA	69.56	91.06	67.76	44.86	94.73	38.36	77.55	82.54	59.63
	OmniSAM-B w/o MA	72.32	92.21	71.40	43.93	94.47	49.28	79.28	83.08	64.94
	OmniSAM-L w/o MA	76.85	**93.86**	**73.61**	65.10	95.04	**55.51**	**83.42**	88.03	60.28
Ours	OmniSAM-T w/o MA	68.72	92.12	64.62	35.54	94.21	37.70	74.33	81.70	69.50
	OmniSAM-S w/o MA	70.65	91.46	65.41	55.10	94.56	33.88	75.61	84.53	64.66
	OmniSAM-B w/o MA	73.09	91.63	69.33	60.43	94.45	36.16	76.22	85.26	71.26
	OmniSAM-L w/o MA	78.02	93.79	71.19	**78.07**	**95.17**	47.25	81.77	**89.85**	66.51
	OmniSAM-T w/ MA	69.10	92.10	64.60	36.97	94.25	38.86	74.16	81.78	70.09
	OmniSAM-S w/ MA	70.81	91.74	66.46	66.74	94.88	12.35	77.43	86.90	69.95
	OmniSAM-B w/ MA	74.72	91.06	66.65	69.31	94.57	36.79	76.98	86.58	**75.87**
	OmniSAM-L w/ MA	**79.06**	93.25	72.12	77.97	95.00	52.08	81.82	89.62	70.58

Table 5. Per-class results of the Stanford2D3D pinhole-to-panoramic scenario (* denotes the baseline).

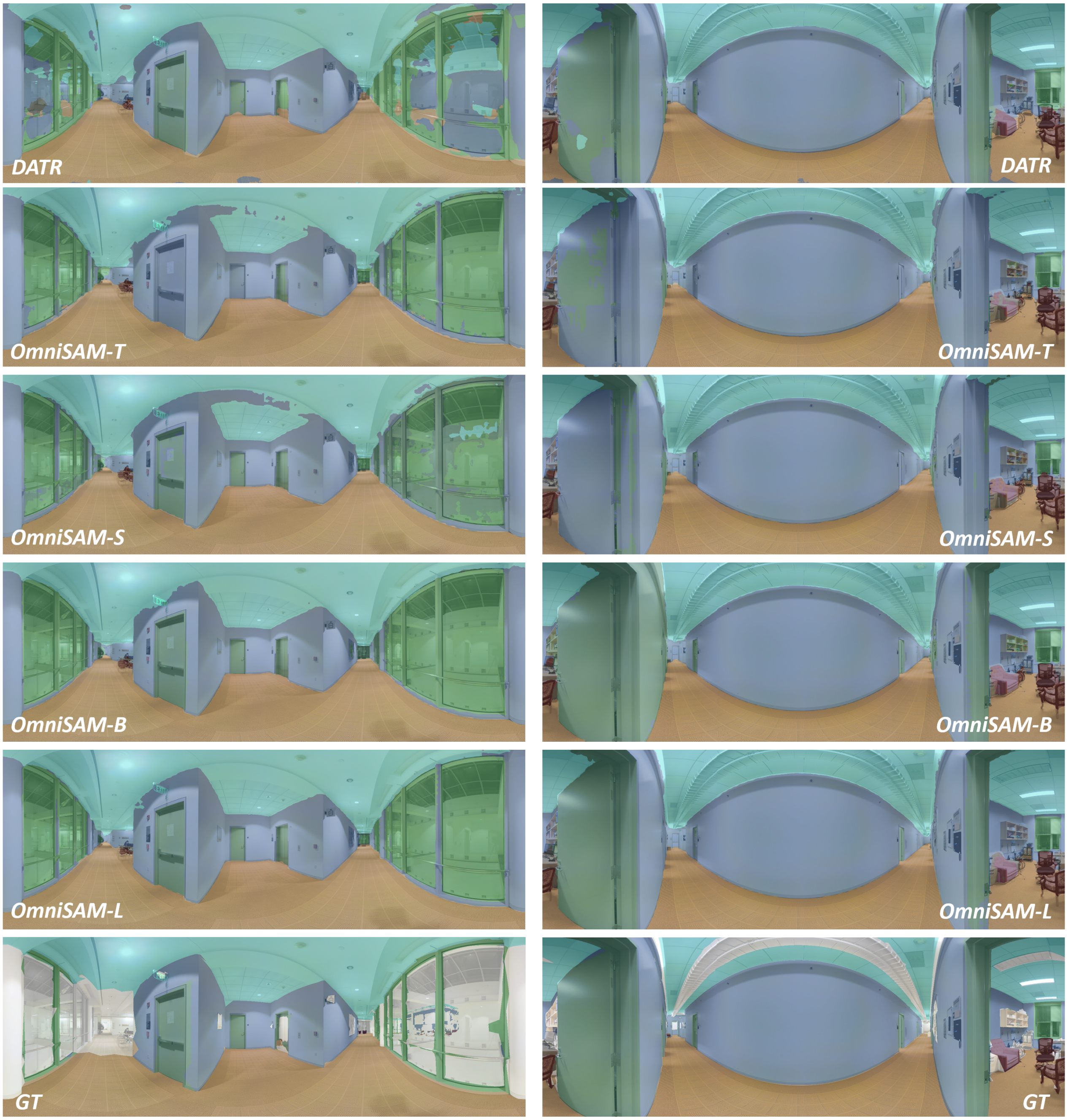
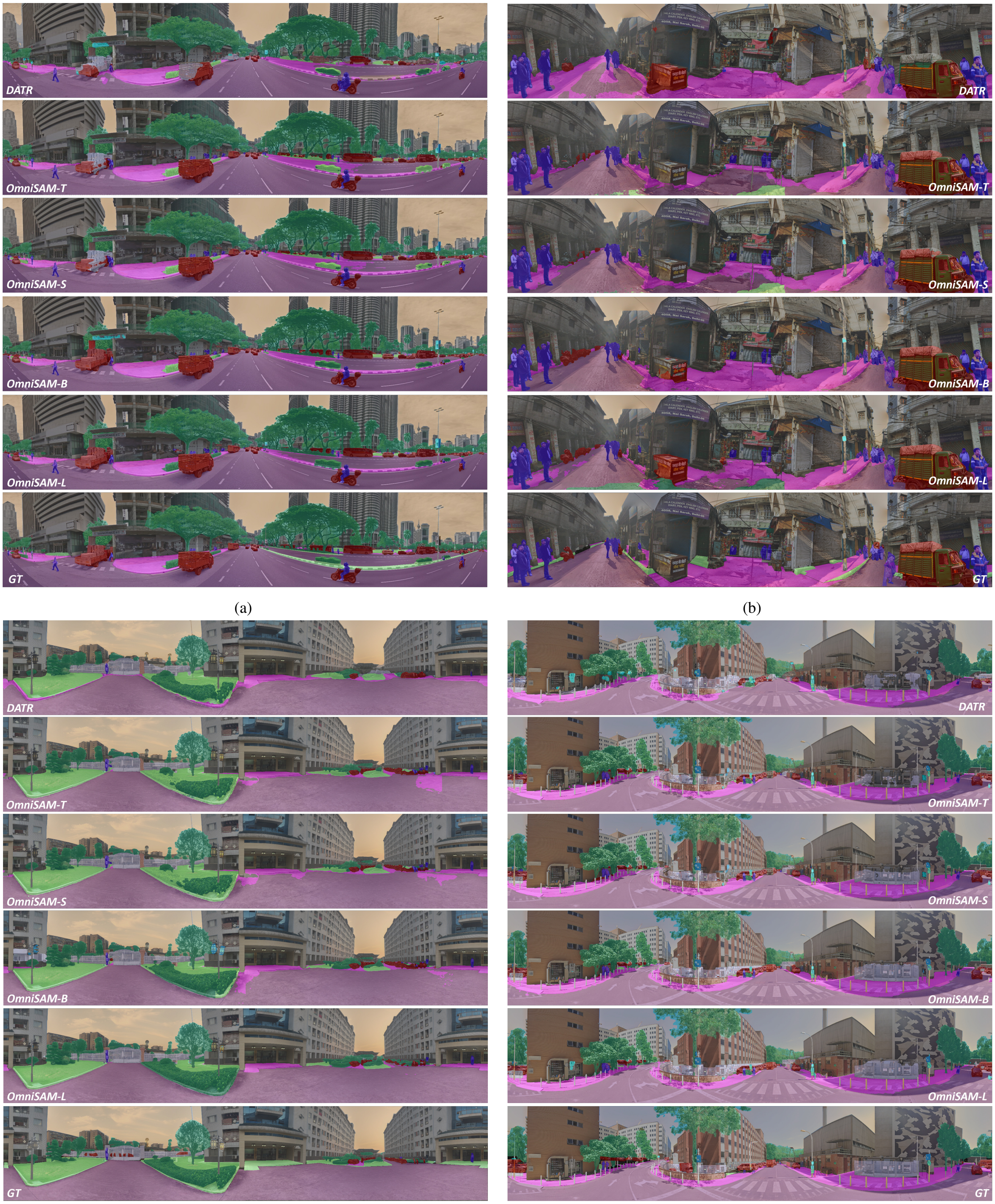


Figure 1. Visualizations on Stanford2D3D-Panoramic dataset for different variants of OmniSAM.



(a)

(b)

(c)

(d)

Figure 2. Visualizations on DensePASS dataset dataset for different variants of OmniSAM.

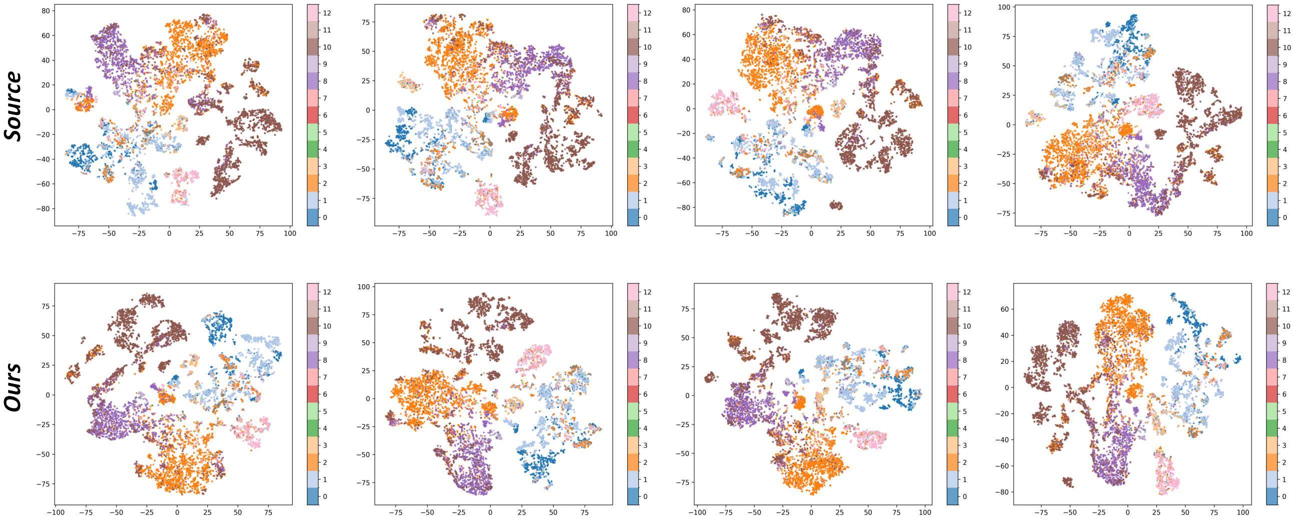


Figure 3. t-SNE Visualizations on DensePASS of Cityscapes-to-DensePASS.