BrainMVP: Multi-modal Vision Pre-training for Medical Image Analysis

Shaohao Rui[1,2*], Lingzhi Chen[2*], Zhenyu Tang[1,2], Lilong Wang[2], Mianxin Liu[2], Shaoting Zhang[2], Xiaosong Wang[2]

[1]Shanghai Jiao Tong University   [2]Shanghai AI Laboratory

Email: ruishaohao@pjlab.org.cn; wangxiaosong@pjlab.org.cn

CVPR 2025 Highlight Paper

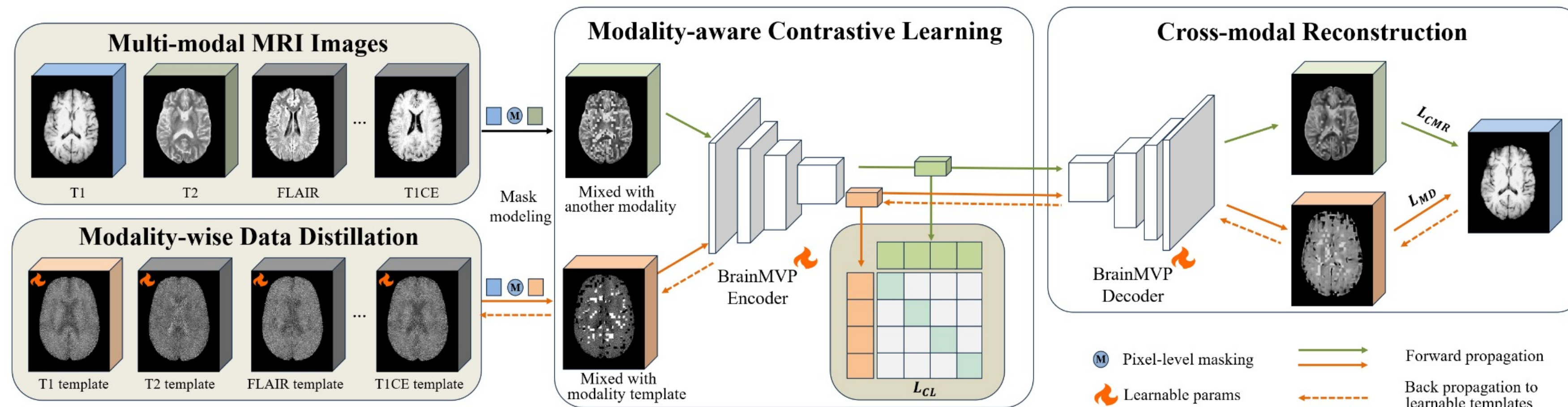Code, dataset, model and full paper available

Background

Challenges:

➤ Lack of SSL methods for leveraging strongly correlated multi-modal data from the same patient.

➤ Missing modality during pre-training: arise from acquisition complexity and hardware limits

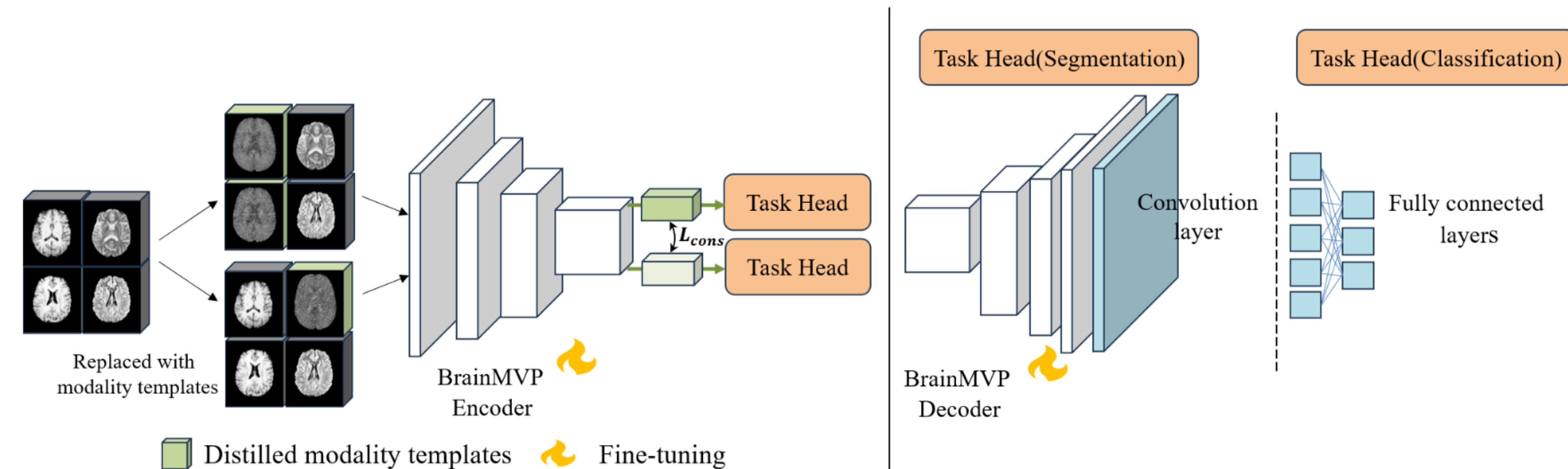➤ Lack of correlation between pre-training and downstream applications.

Motivation:

➤ Use large-scale missing modality dataset for multi-modal representation learning

➤ Learn bridging components to facilitate both pre-training and fine-tuning

(a) Multi-modal datasets

(b) Pre-training methods
Cross modal reconstruction | Modality-wise data distillation | Modality-aware contrastive learning
model | ❄ modality template

(c) Downstream applications
Segmentation 1 | Segmentation N | Classification 1 | Classification N

Methodology

➤ BrainMVP consists of three pretext tasks to facilitate the learning of cross-modality representations and correlations:

➤ **Cross-modal Reconstruction**: We aim to learn cross-modal correlations through transformation

➤ **Modality-wise Data Distillation:** Inspired by methods of dataset distillation, we learn condensed structural representations of different modalities to introduce certain bridging components to facilitate both pre-training and downstream tasks.

➤ **Modality-aware Contrastive learning**: introduce study/case-level modality invariance to the learned features.

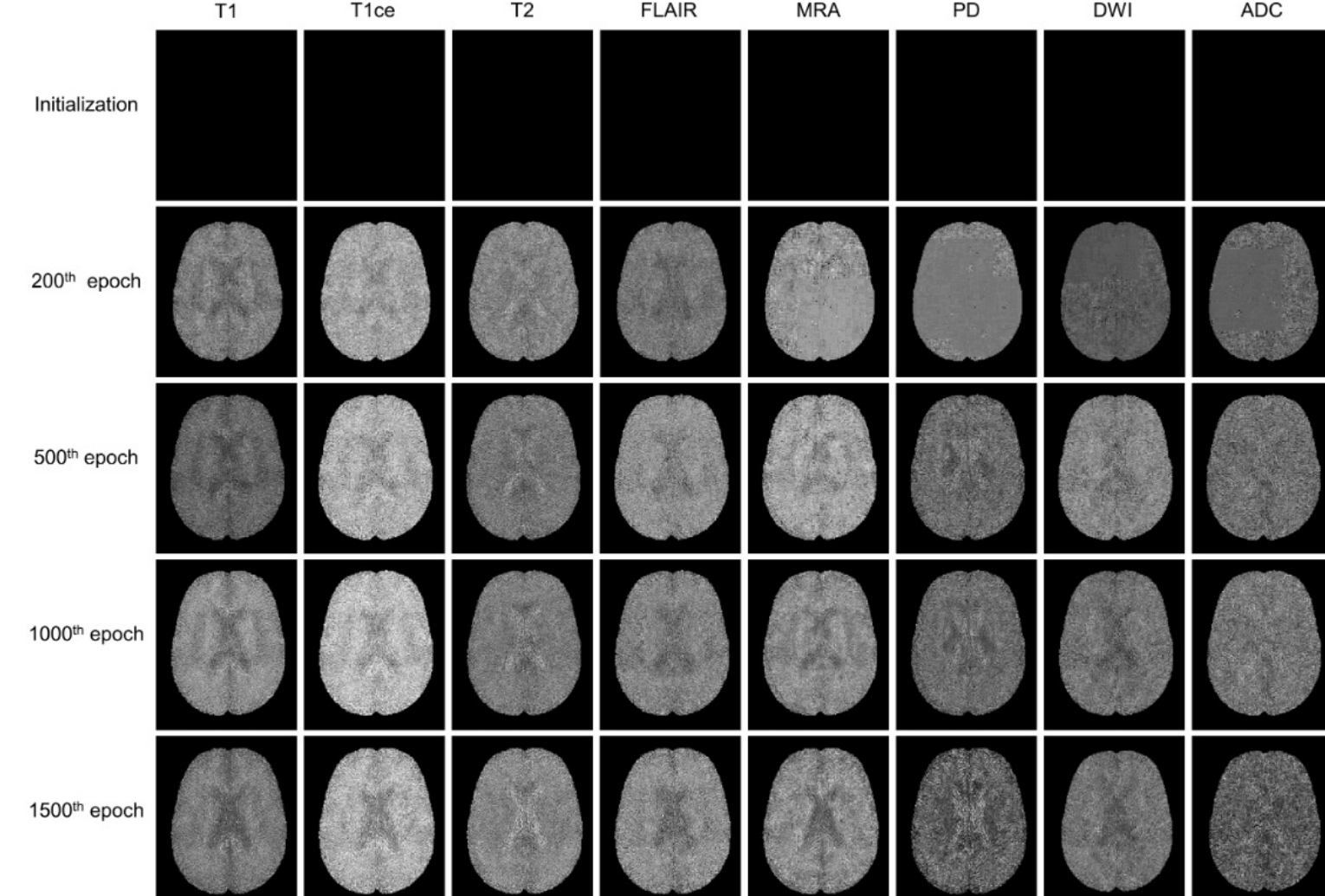### Templates Usage: mixing pre-trained templates and real modality images

Results

Experiments on segmentation and classification tasks

Method	Modality	Network	BraTS2023-PED [26]				BraTS-MET [37]				ISLES22 [22]	MRBrainS13 [36]				VSseg [41]	UPENN-GBM [4]			
			ET	TC	WT	AVG	ET	TC	WT	AVG	IS	CF	GM	WM	AVG	VS	ET	TC	WT	AVG
From Scratch																				
UNETR [19]	-		46.46	76.43	78.66	67.19	54.01	54.87	59.44	56.11	74.65	67.55	78.73	83.69	76.66	70.28	83.10	80.88	81.98	81.99
UNET3D [40]	-		47.12	81.60	83.94	70.89	56.44	58.75	62.76	59.32	80.94	70.47	73.93	82.96	75.78	69.43	85.65	88.76	86.27	86.89
UniFormer [31]	-		46.73	83.87	86.97	72.52	67.22	72.74	70.78	70.25	84.97	77.66	74.09	75.60	75.78	80.33	87.93	91.86	88.81	89.53
Swin-UNETR [18]	-		49.66	81.10	84.13	71.63	63.84	67.08	68.58	66.50	75.88	70.35	81.66	84.65	78.89	76.82	87.60	91.15	87.34	88.70
With General SSL																				
MAE3D [10, 20]	Natural	UNETR	46.55	77.08	79.32	67.65	57.45	59.19	62.06	59.57	70.43	68.30	80.57	84.69	77.86	69.57	83.66	80.42	81.86	81.98
SimMIM [55]	Natural	UNETR	45.14	76.59	78.61	66.78	54.46	55.84	58.93	56.40	69.94	68.11	80.49	84.76	77.79	69.08	83.70	81.68	82.47	82.61
MoCov3 [8]	Natural	UNETR	45.66	77.37	79.88	67.64	55.84	56.77	61.62	58.07	70.32	67.97	79.64	84.36	77.32	69.83	83.02	80.54	81.77	81.78
With Medical SSL																				
MG [65]	CXR, CT	UNET3D	47.99	**86.69**	88.41	74.36	60.11	64.05	65.43	63.19	83.53	71.40	74.71	80.41	75.51	76.33	86.64	90.58	87.03	88.08
TransVW [16]	CT	UNET3D	46.38	80.05	81.98	69.47	56.10	58.69	62.91	59.20	80.24	68.92	80.53	83.70	77.72	71.76	85.95	89.51	86.91	87.46
GVSL [21]	CT	UNET3D	49.05	84.47	86.81	73.45	62.46	66.81	67.26	65.51	80.05	69.34	75.07	82.85	75.75	72.21	87.09	91.75	87.53	88.79
Swin-UNETR* [46]	MRI	Swin-UNETR	49.07	81.74	84.13	71.65	60.60	64.56	64.53	63.23	79.55	69.67	**82.09**	**86.13**	79.30	75.55	87.24	91.46	87.28	88.66
VoCo [53]	MRI	Swin-UNETR	48.66	82.26	84.64	71.85	57.49	59.43	59.50	58.07	77.58	71.29	76.43	81.40	76.37	76.45	86.65	90.54	87.34	88.18
DAE [48]	MRI	Swin-UNETR	49.30	82.12	84.78	72.07	62.27	65.99	64.85	64.37	73.92	71.37	78.50	83.20	77.69	74.51	86.90	90.83	87.32	88.35
M³AE [33]	MRI	UNET3D	46.77	85.67	86.89	73.11	66.01	70.92	70.18	69.04	83.85	71.32	69.56	79.28	73.39	75.96	87.15	91.90	88.44	89.16
M³AE [33]	MRI	UniFormer	50.77	84.95	86.70	74.14	68.08	72.35	70.74	70.39	86.32	78.23	77.20	76.43	77.29	79.31	87.75	92.43	88.72	89.63
BrainMVP	MRI	UNET3D	47.75	85.99	**88.46**	74.07	67.24	71.27	68.63	69.05	83.31	68.88	74.60	82.66	75.38	76.02	87.30	91.87	**88.98**	89.38
BrainMVP	MRI	UniFormer	55.45	86.54	**88.41**	**76.80**	70.70	75.80	74.52	73.67	86.60	81.04	78.17	81.61	80.27	83.64	88.49	92.48	89.07	90.01

Method	Modality	Network	BraTS2018 [3]			ADNI [23]			ADHD-200 [11]			ABIDE-I [14]		
			ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1
From Scratch														
UNETR [19]	-	-	0.7895	0.7817	0.6621	0.5672	0.6066	0.5645	0.6688	0.6523	0.6204	0.6121	0.5478	0.5507
UNET3D [40]	-	-	0.7368	0.7373	0.4242	0.5756	0.4966	0.3653	0.6494	0.6798	0.4265	0.6061	0.5059	0.4591
UniFormer [31]	-	-	0.7762	0.7719	0.6994	0.5546	0.6343	0.5526	0.6039	0.6387	0.5796	0.5879	0.4433	0.4292
Swin-UNETR [18]	-	-	0.7018	0.7143	0.6069	0.5672	0.5853	0.5650	0.6494	0.6950	0.6240	0.6121	0.5530	0.5596
With General SSL														
MAE3D [10, 20]	Natural	UNETR	0.7018	0.6754	0.5645	0.5756	0.5414	0.5651	0.6169	0.6489	0.5906	0.6061	0.4983	0.4591
SimMM [55]	Natural	UNETR	0.7368	0.8349	0.7077	0.6218	0.6026	0.5446	0.6234	0.6567	0.5790	0.5394	0.5819	0.5318
MoCov3 [8]	Natural	UNETR	0.7368	0.8135	0.7304	0.6092	0.5769	0.5996	0.6104	0.6265	0.6007	0.5939	0.6284	0.5890
With Medical SSL														
MG [65]	CXR, CT	UNET3D	0.7368	0.9286	0.4242	0.5756	0.5496	0.3653	0.6169	0.6980	0.6141	0.6121	0.6266	0.5892
TransVW [16]	CT	UNET3D	0.7368	0.7222	0.4242	0.4958	0.6661	0.4450	0.6818	0.7228	0.6271	0.6424	0.5292	0.5003
GVSL [21]	CT	UNET3D	0.7895	0.8516	0.7286	0.5966	0.6661	0.5959	0.6623	**0.7309**	0.6565	0.6242	0.5244	0.4701
Swin-UNETR* [46]	MRI	Swin-UNETR	0.7368	0.5032	0.4242	0.5462	0.5517	0.5461	0.6299	0.6437	0.5953	0.6303	0.4993	0.3866
VoCo [53]	MRI	Swin-UNETR	0.7368	0.5135	0.4242	0.5210	0.5407	0.5207	0.6558	0.6971	0.6413	0.5818	0.5626	0.5466
DAE [48]	MRI	Swin-UNETR	0.7719	0.8151	0.7120	0.5294	0.5666	0.5294	0.6688	0.7129	0.6548	0.6061	0.5173	0.5548
M³AE [33]	MRI	UNET3D	0.7370	0.6984	0.5915	0.6088	0.6338	0.6003	0.6364	0.7049	0.6177	0.6061	0.5453	0.4769
M³AE [33]	MRI	UniFormer	0.7895	0.8659	0.7159	0.6092	0.5352	0.5756	0.6169	0.6597	0.6028	0.5636	0.4682	0.4500
BrainMVP	MRI	UNET3D	0.7895	0.7746	0.6621	0.6555	0.6669	0.6609	0.6883	0.7245	0.6665	0.6970	0.5817	0.5275
BrainMVP	MRI	UniFormer	0.8596	0.9452	0.8324	0.6765	0.6964	0.6609	0.6883	0.7249	0.6723	0.6182	0.6329	0.5890

Templates Learning Trajectories

Visualization results comparison