Multi-modal Vision Pre-training for Medical Image Analysis

Supplementary Material

A. Distilled Modality Template for Downstream Tasks

In this section, we will elaborate on how the distilled modality templates obtained from pre-training can be applied in downstream tasks. As shown in Fig. 5, in the downstream fine-tuning stage, the distilled modality templates are frozen. Let $\mathcal{D}_{ds} = \{(X_i, Y_i)\}_{i=1}^{M}$ denote the downstream dataset, where M represents the number of annotated samples. X_i is the multi-modal MRI input volume, and Y_i represents the corresponding label, which can be a segmentation map for segmentation tasks or a one-hot vector for classification tasks. Specifically, we randomly select m and n modalities in X_i and replace them with the corresponding modalities from $\{T_m\}_{m=1}^{S}$, obtaining two augmented copies X_i' and X_i''. The encoded features of these two copies are $\mathcal{F}_{enc}(X_i')$ and $\mathcal{F}_{enc}(X_i'')$, respectively. Since the two embeddings are representations of the same sample with different numbers of replaced modalities, we use the L2 norm to maintain semantic consistency in the feature space.

$$\mathcal{L}_{cons} = ||\mathcal{F}_{enc}(X_i') - \mathcal{F}_{enc}(X_i'')||_2 \tag{7}$$

Subsequently, the features of the two copies are decoded to the output space to calculate supervision loss with the ground-truth annotations. The overall fine-tuning loss is:

$$\begin{aligned}\mathcal{L}_{FT} = &\frac{1}{|\mathcal{B}|}\sum_{i=1}^{|\mathcal{B}|}(\mathcal{L}_{sl}(\mathcal{F}(X_i'), Y_i) + \mathcal{L}_{sl}(\mathcal{F}(X_i''), Y_i) \\ &+\lambda_{cons} * \mathcal{L}_{cons})\end{aligned} \tag{8}$$

where λ_{cons} is the weight of the consistency loss $\mathcal{L}_{cons}$ term and $\mathcal{L}_{sl}$ is the supervision loss used in segmentation or classification tasks, e.g., Dice Loss in segmentation or Cross-Entropy Loss in classification. $|\mathcal{B}|$ represents number of cases in a batch.

For the uni-modal input scenario, instead of replacing the selected modalities with distilled modality templates, we perform a partially masking strategy like Algorithm 1 where X_i is replaced with the corresponding distilled modality template. Then we randomly mask the uni-modal input volume twice to obtain two augmented copies of X_i, and the remaining procedures are the same as the aforementioned multi-modal scenario.

B. Pre-processing

B.1. Pre-training

During pre-training, data pre-processing is performed sequentially in Python based on MONAI 1.3.0 library. The orientation of the mpMRI scan is first unified to the RAS axcodes and co-registered to the same anatomical template. Subsequently, each MRI scan is resampled to an isotropic voxel spacing of $1.0mm \times 1.0mm \times 1.0mm$ using bilinear interpolation, and skull-stripping is performed as well. We linearly clip the pixel values between the 1st and 99th percentiles and re-scale them to [0, 1]. The images are then cropped into $96 \times 96 \times 96$ voxel patches centered on either foreground or background areas, to ensure that the modality-wise data distillation is learned sufficiently. We do not apply any other data augmentation techniques.

B.2. Segmentation

The input mpMRI scan is first reoriented to the RAS coordinate system, then the image spacing is adjusted to a uniform $1.0mm \times 1.0mm \times 1.0mm$ (for the ISLES22 [22] dataset it's $1.5mm \times 1.5mm \times 1.5mm$) using bilinear interpolation. Subsequently, the pixel grayscale values of the input mpMRI scan are normalized from the 5th to the 95th percentile, with each channel being adjusted to a range between 0 and 1. After cropping the foreground area of the image, we randomly crop a fixed area of $96 \times 96 \times 96$. To avoid over-segmentation, we allow the sampling center to be in the background area. Then, random mirror flipping along three axes with a probability of 0.5, random intensity offset with 0.1 offset, random intensity scaling with probability 1.0 in a scale factor of 0.1 are performed for data augmentation. For network training, we employ the AdamW optimizer [35] with an initial learning rate of 3e-4, incorporating cosine learning rate decay. Weight decay is set to 1e-3 for UNETR [19]-based models, 1e-4 for UniFormer [31] and Swin-UNETR [18]-based models, and 1e-5 for UNET3D [40]-based models. We train the network with a batch size of 3 for 500 epochs, and $\lambda_{cons}$is set to 0.1.

B.3. Classification

The data augmentation part is different from segmentation in that we resize the input image to a fixed size of $128 \times 128 \times 64$ after normalizing it to fit the training of the comparison methods. Subsequently, we randomly crop a fixed region of $96 \times 96 \times 64$ and then perform the same random data augmentation as segmentation. In the inference

https://monai.io/

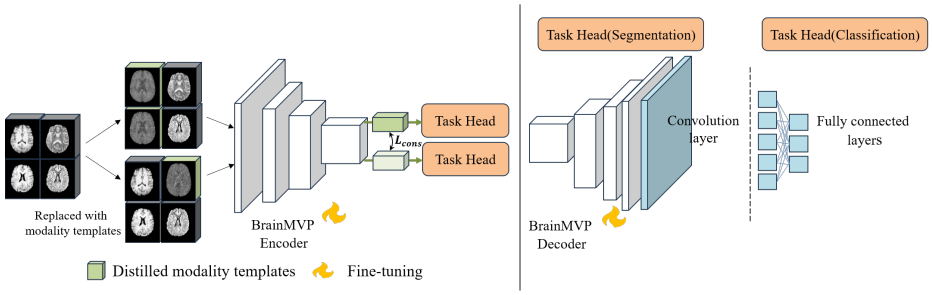

Figure 5. Modality-wise data distillation for **downstream tasks**. The input multi-modal MRI scans are randomly selected to replace a certain number of modalities with the corresponding modality templates. Then L2 norm is used to ensure feature consistency between the two replacement copies. Finally, the task head is replaced with corresponding modules based on the task type.

stage, we crop an area of $96 \times 96 \times 64$ at the center of the input image. we set the batch size to 64 considering gradient accumulation and train all networks for 200 epochs. The remaining hyper-parameters are the same as those used for segmentation.

C. Dataset Details

C.1. Pre-training datasets

BraTS2021 [2]: This dataset comprises 1,470 cases publicly available multi-sequence MRI scans, encompassing four paired modalities: T1, T1CE, T2, and FLAIR. All images have been registered and resampled to $1.0mm \times 1.0mm \times 1.0mm$. We only utilize the image data without incorporating the segmentation annotations.

BraTS2023-SSA [1] and **BraTS2023-MEN** [29]: These datasets are two of the five segmentation sub-tasks in BraTS2023 with 75 cases and 1,141 cases mpMRI, respectively. The former dataset focuses on the segmentation of brain gliomas in patients from sub-Saharan Africa, while the latter is dedicated to adult meningioma segmentation. Note that the modality type is identical to **BraTS2021 [2]**, albeit involving a different type of brain tumor.

UCSF-PDGM [6]: This dataset comprises 501 cases with various mpMRI data, from which we select six modalities–T1, T1CE, T2, FLAIR, DWI, and ADC for corresponding downstream applications.

IXI: This dataset includes 600 MR images from normal, healthy subjects with T1, T2, PD, MRA and DTI images. We select 568 cases that include all four modalities: T1, T2, PD, and MRA for pre-training and this dataset serves as a supplement to the pre-training brain dataset, specifically for normal brain cases.

C.2. Downstream datasets

We conduct a comprehensive evaluation using ten downstream datasets encompassing segmentation and classification tasks. The details are as follows:

https://brain-development.org/ixi-dataset/

Segmentation: (1) BraTS2023-PED [26]: This dataset comprises 99 publicly annotated pediatric brain glioma multi-sequence MRI scans. The annotations include Non-Enhancing Core (NEC), Edema, and Enhancing Tumor (ET). (2) BraTS2023-MET [37]: Similarlly, this dataset focuses on brain metastasis sub-region segmentation from multi-sequence MRI. It contains 238 publicly available imaging cases with four modalities: T1, T1CE, T2W, and FLAIR. (3) ISLES22 [22]: This dataset aims to segment acute to subacute ischemic stroke lesions from multi-sequence MR images (including FLAIR, DWI, and ADC). We collected 238 publicly annotated cases. (4) MRBrainS13 [36]: This dataset targets brain structure segmentation from 20 cases with three sequences: T1, T1CE, and FLAIR MR images. The segmentation targets include Cerebrospinal Fluid (CF), Gray Matter (GM), and White Matter (WM). (5) UPENN-GBM [4]: We collected 127 publicly annotated multi-sequence MR images from de novo Glioblastoma (GBM) patients, similarly focusing on segmenting three tumor subregions. (6) VSseg [41]: This dataset includes 242 cases of multi-sequence MRI data from patients with vestibular schwannoma, aiming to segment the vestibular schwannoma region.

Classification: (1) BraTS2018 [3]: This dataset includes a tumor subtype classification task, aiming to determine the severity grade of brain tumors from four MR modalities, labeled as HGG (High-Grade Glioma) or LGG (Low-Grade Glioma). (2) ADNI [23]: This dataset represents late-life brain disorders through Alzheimer's Disease (AD) cases. Given the importance of early diagnosis, we analyze the most recent neuroimaging scans and demographic data from 1348 subjects, labeled as mild cognitive impairment (MCI) or normal control (NC). (3) ADHD-200 [11] and (4) ABIDE-I [14]: These two datasets are utilized for early-life brain disorder studies. For ADHD-200 [11], T1-weighted MRI scans and demographic information (age and gender) are collected from 767 subjects, including 279 ADHD patients and 488 controls. ABIDE-I [14] comprises neuroimaging data from 819 subjects (327 with autism spec-

trum disorder and 492 typically developing controls) with matching imaging modalities.

The aforementioned datasets, except for MRBrainS13 [36], are randomly partitioned into training, validation, and test sets with a ratio of 6:1:3. For MRBrainS13 [36], 5 cases are used for training and the remaining 15 cases for testing. It's worth noting that the data splits for ADNI [23], ADHD-200 [11], and ABIDE-I [14] datasets are performed at the patient/case level, ensuring that scans from the same subject will not appear across different sets.

Algorithm 1 Pixel-level cross-modal masking.

Sample randomly X_{im} from X_i
Sample randomly $X_{in}(n \neq m)$ from X_i
$p_{total} \leftarrow H \times W \times D$
$p_{mask} \leftarrow 0$
while $p_{mask} < p_{total} \times p^*$ **do**
  Select randomly (x, y, z) in X_{im}
  Mask an area of size $r \times r \times r$ centered at (x, y, z)
  Fill with corresponding data from X_{in}
  $p_{mask} \leftarrow p_{mask} \bigoplus r \times r \times r$
end while
return modified X_{im}

D. HD95 Results and Visualization

In Table 5 and Table 6, we report the HD95 metric results of the pre-trained model on segmentation and classification tasks, respectively. These experimental results indicate that BrainMVP consistently exhibits smaller structural errors.

To facilitate qualitative comparison, we visualize the results obtained from MAE3D [10, 20], MG [65], GVSL [21], VoCo [53], and BrainMVP on four datasets. The visualizations are shown in Fig. 6. The visualization results indicate that our BrainMVP segmentation results are most consistent with the ground truth (GT), significantly mitigating the issues of under-segmentation and over-segmentation. As shown in Fig. 6 (a) for the NCR region boundary, BrainMVP demonstrates more accurate identification, while other methods exhibit substantial under-segmentation.

Table 5. Experimental results on datasets BraTS2023-PED [26], BraTS2023-MET [37] and ISLES22 [22]. We report the mean HD95 (↓) on each dataset.

Method	Modality	Network	BraTS2023-PED [26]				BraTS2023-MET [37]				ISLES22 [22]
			ET	TC	WT	AVG	ET	TC	WT	AVG	IS
From Scratch											
UNETR [19]	-	-	25.06	39.07	39.14	34.43	44.11	45.22	43.36	44.23	15.48
UNET3D [40]	-	-	22.48	34.02	33.07	29.86	45.68	46.85	39.93	44.15	4.43
UniFormer [31]	-	-	11.55	16.71	16.14	14.80	25.90	28.16	19.97	24.68	4.13
Swin-UNETR [18]	-	-	17.37	22.56	21.03	20.32	28.68	31.03	24.26	27.99	11.31
With General SSL											
MAE3D [10, 20]	Natural	UNETR	25.37	38.43	37.92	33.90	36.89	36.57	38.38	37.28	15.20
SimMIM [55]	Natural	UNETR	24.70	31.61	32.52	29.61	39.37	41.26	40.06	40.23	17.14
MoCoV3 [8]	Natural	UNETR	20.60	31.88	32.12	28.20	41.88	43.17	41.92	42.32	15.04
With Medical SSL											
MG [65]	CXR, CT	UNET3D	19.71	15.72	17.65	17.69	46.39	48.33	42.02	45.58	3.68
TransVW [16]	CT	UNET3D	18.36	25.42	24.67	22.82	47.85	48.06	39.41	45.11	7.93
GVSL [21]	CT	UNET3D	17.45	15.33	16.00	16.26	37.33	38.05	30.61	35.33	9.35
Swin-UNETR* [46]	MRI	Swin-UNETR	18.65	17.44	17.64	17.91	40.57	41.54	33.93	38.68	8.09
VoCo [53]	MRI	Swin-UNETR	18.98	17.21	17.16	17.78	38.52	39.79	34.73	37.68	12.22
DAE [48]	MRI	Swin-UNETR	19.33	21.41	21.71	20.82	37.63	37.37	38.74	37.91	12.50
M^3AE [33]	MRI	UNET3D	13.48	11.91	10.88	12.09	22.40	23.87	18.96	21.74	4.58
M^3AE [33]	MRI	UniFormer	16.19	15.95	19.78	17.31	25.89	28.37	24.35	26.21	2.64
BrainMVP	MRI	UNET3D	15.93	7.24	9.81	10.99	20.37	22.50	18.34	20.40	5.85
BrainMVP	MRI	UniFormer	13.93	7.88	14.56	12.12	22.60	25.88	19.83	22.77	2.69

CXR: Chest X-Ray; ET: enhancing tumor; TC: tumor core; WT: whole tumor; AVG: average; CF: Cerebrospinal Fluid; GM: Gray matter; WM: White matter; IS: Ischemic Stroke.

Table 6. Experimental results on datasets MRBrainS13 [36], VSseg [41] and UPENN-GBM [4]. We report the mean HD95 (↓) on each dataset.

Method	Modality	Network	MRBrainS13 [36]				VSseg [41]	UPENN-GBM [4]			
			CF	GM	WM	AVG	VS	ET	TC	WT	AVG
From Scratch											
UNETR [19]	-	-	4.16	3.46	5.04	4.22	24.54	16.97	24.80	31.00	24.26
UNET3D [40]	-	-	3.24	2.91	3.70	3.28	34.36	5.30	9.34	13.31	9.32
UniFormer [31]	-	-	2.38	2.43	4.04	2.95	5.68	4.46	6.97	11.32	7.58
Swin-UNETR [18]	-	-	3.38	2.65	4.00	3.34	14.12	1.86	7.22	9.15	6.08
With General SSL											
MAE3D [10, 20]	Natural	UNETR	3.69	2.62	3.59	3.30	24.17	15.41	20.10	35.71	23.74
SimMIM [55]	Natural	UNETR	3.84	2.67	3.55	3.35	26.82	17.23	20.71	32.11	23.35
MoCoV3 [8]	Natural	UNETR	3.84	2.99	4.74	3.86	21.35	17.08	19.83	34.35	23.75
With Medical SSL											
MG [65]	CXR, CT	UNET3D	3.47	9.43	12.67	8.52	14.87	2.27	4.29	12.67	6.41
TransVW [16]	CT	UNET3D	3.81	3.45	2.93	3.40	16.83	3.36	5.73	12.95	7.35
GVSL [21]	CT	UNET3D	3.73	3.44	3.28	3.48	11.58	2.23	3.71	9.17	5.03
Swin-UNETR* [46]	MRI	Swin-UNETR	3.33	2.26	2.33	2.64	20.73	2.44	4.07	9.79	5.43
VoCo [53]	MRI	Swin-UNETR	3.14	3.88	7.87	4.96	13.26	28.50	43.05	31.51	34.35
DAE [48]	MRI	Swin-UNETR	3.07	2.27	3.36	2.90	19.84	2.24	3.90	9.56	5.23
M^3AE [33]	MRI	UNET3D	3.69	3.88	3.01	3.53	9.20	1.85	4.65	8.24	4.91
M^3AE [33]	MRI	UniFormer	1.89	2.92	4.53	3.11	9.16	4.75	6.54	9.93	7.07
BrainMVP	MRI	UNET3D	3.71	4.92	3.84	4.14	16.41	2.35	4.60	9.13	5.36
BrainMVP	MRI	UniFormer	1.53	5.60	7.02	4.72	6.00	1.48	6.66	10.59	6.24

CXR: Chest X-Ray; ET: enhancing tumor; TC: tumor core; WT: whole tumor; AVG: average; CF: Cerebrospinal Fluid; GM: Gray matter; WM: White matter; VS: Vestibular schwannoma.

References

[1] Maruf Adewole, Jeffrey D Rudie, Anu Gbdamosi, Oluyemisi Toyobo, Confidence Raymond, Dong Zhang, Olubukola Omidiji, Rachel Akinola, Mohammad Abba Suwaid, Adaobi Emegoakor, et al. The brain tumor segmentation (brats) challenge 2023: Glioma segmentation in sub-saharan africa patient population (brats-africa). *ArXiv*, 2023. 5, 2

[2] Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C Kitamura, Sarthak Pati, et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*, 2021. 2, 5

[3] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018. 2, 5, 7, 8

[4] Spyridon Bakas, Chiharu Sako, Hamed Akbari, M Bilello, A Sotiras, G Shukla, et al. Multi-parametric magnetic resonance imaging (mpmri) scans for de novo glioblastoma (gbm) patients from the university of pennsylvania health system (upenn-gbm). *The Cancer Imaging Archive (TCIA) Public Access*, 2021. 5, 6, 7, 8, 2, 3

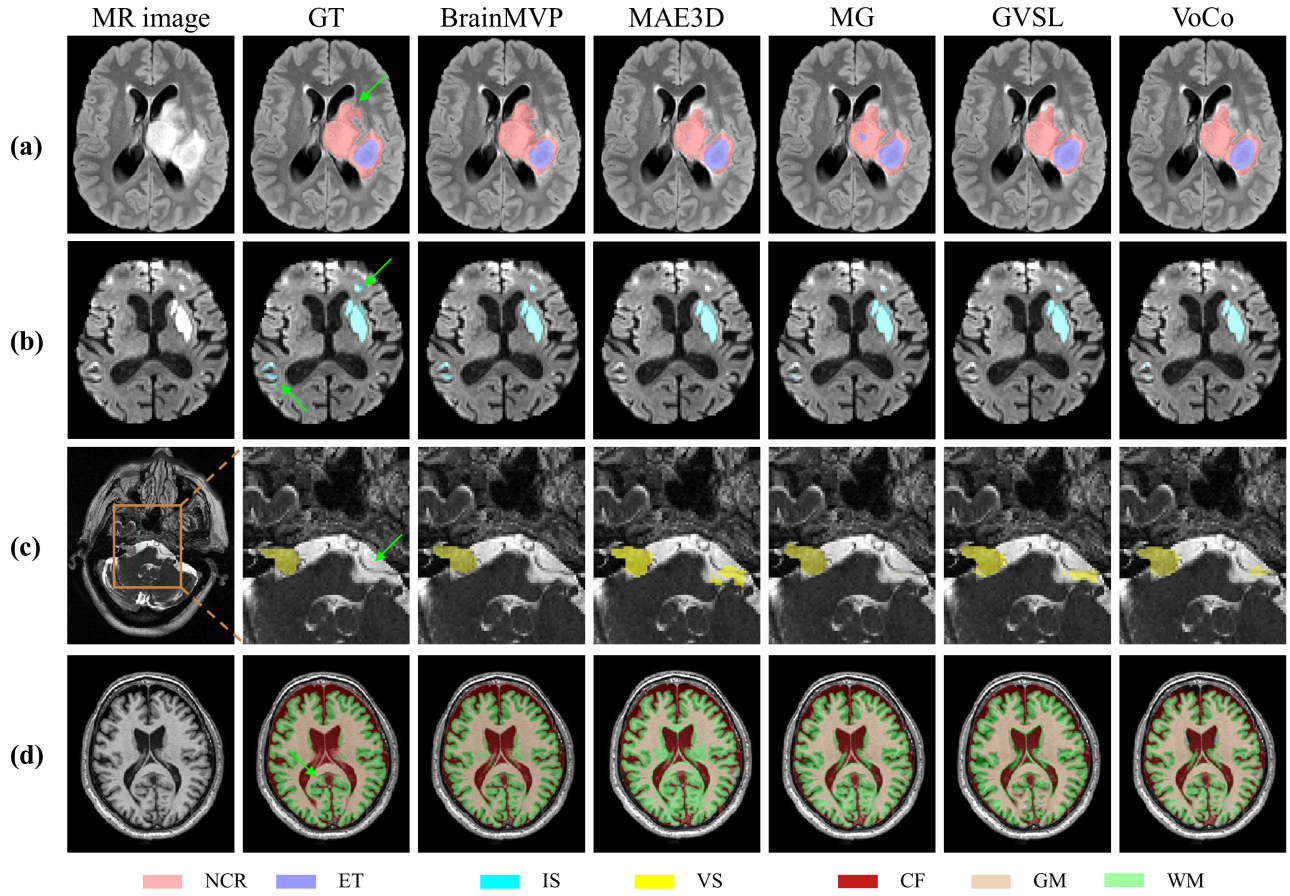

Figure 6. Visualization results of segmentation tasks. (a) BraTS2023-PED [26]: pediatric tumor subregion segmentation. NCR: necrotic tumor core; ET: enhancing tumor. (b) ISLES22 [22]: Ischemic Stroke lesion (IS) segmentation. (c) VSseg [41]: Vestibular schwannoma (VS) segmentation. (d) MRBrainS13 [36]: brain structure segmentation. CF: Cerebrospinal Fluid; GM: Gray matter; WM: White matter. GT: ground truth. The green arrows highlight the regions where BrainMVP demonstrates superior performance over other methods.

[5] Yu Cai, Hao Chen, Xin Yang, Yu Zhou, and Kwang-Ting Cheng. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. *Medical Image Analysis*, 86:102794, 2023. 1, 2

[6] Evan Calabrese, Javier E Villanueva-Meyer, Jeffrey D Rudie, Andreas M Rauschecker, Ujjwal Baid, Spyridon Bakas, Soonmee Cha, John T Mongan, and Christopher P Hess. The university of california san francisco preoperative diffuse glioma mri dataset. *Radiology: Artificial Intelligence*, 4 (6):e220058, 2022. 5, 2

[7] George Cazenavette, Tongzhou Wang, Antonio Torralba, Alexei A Efros, and Jun-Yan Zhu. Dataset distillation by matching training trajectories. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4750–4759, 2022. 3

[8] Xinlei Chen, Saining Xie, and Kaiming He. An empirical study of training self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9640–9649, 2021. 5, 6, 7, 3

[9] Zhihong Chen, Yuhao Du, Jinpeng Hu, Yang Liu, Guanbin Li, Xiang Wan, and Tsung-Hui Chang. Multi-modal masked autoencoders for medical vision-and-language pre-training. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 679–689. Springer, 2022. 2

[10] Zekai Chen, Devansh Agarwal, Kshitij Aggarwal, Wiem Safta, Mariann Micsinai Balan, and Kevin Brown. Masked image modeling advances 3d medical image analysis. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1970–1980, 2023. 5, 6, 7, 3

[11] ADHD-200 consortium. The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in systems neuroscience*, 6: 62, 2012. 5, 7, 2, 3

[12] Justin Cui, Ruochen Wang, Si Si, and Cho-Jui Hsieh. Scaling up dataset distillation to imagenet-1k with constant memory. In *International Conference on Machine Learning*, pages 6565–6590. PMLR, 2023. 3

[13] Zhiwei Deng and Olga Russakovsky. Remember the past:

Distilling datasets into addressable memories for neural networks. *Advances in Neural Information Processing Systems*, 35:34391–34404, 2022. 3

[14] Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014. 5, 7, 2, 3

[15] Yuhang Ding, Xin Yu, and Yi Yang. RFNet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3975–3984, 2021. 2

[16] Fatemeh Haghighi, Mohammad Reza Hosseinzadeh Taher, Zongwei Zhou, Michael B Gotway, and Jianming Liang. Transferable visual words: Exploiting the semantics of anatomical patterns for self-supervised learning. *IEEE transactions on medical imaging*, 40(10):2857–2868, 2021. 1, 2, 5, 6, 7, 3

[17] Fatemeh Haghighi, Mohammad Reza Hosseinzadeh Taher, Michael B Gotway, and Jianming Liang. Dira: Discriminative, restorative, and adversarial learning for self-supervised medical image analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20824–20834, 2022. 1, 2

[18] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R Roth, and Daguang Xu. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI Brainlesion Workshop*, pages 272–284. Springer, 2021. 5, 6, 7, 1, 3

[19] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 574–584, 2022. 5, 6, 7, 1, 3

[20] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022. 5, 6, 7, 3

[21] Yuting He, Guanyu Yang, Rongjun Ge, Yang Chen, Jean-Louis Coatrieux, Boyu Wang, and Shuo Li. Geometric visual similarity learning in 3d medical image self-supervised pre-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9538–9547, 2023. 1, 2, 5, 6, 7, 3

[22] Moritz R Hernandez Petzsche, Ezequiel de la Rosa, Uta Hanning, Roland Wiest, Waldo Valenzuela, Mauricio Reyes, Maria Meyer, Sook-Lei Liew, Florian Kofler, Ivan Ezhov, et al. Isles 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Scientific data*, 9(1): 762, 2022. 5, 6, 7, 8, 1, 2, 3, 4

[23] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008. 5, 7, 8, 2, 3

[24] Yankai Jiang, Mingze Sun, Heng Guo, Xiaoyu Bai, Ke Yan, Le Lu, and Minfeng Xu. Anatomical invariance modeling and semantic alignment for self-supervised learning in 3d medical image analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15859–15869, 2023. 1

[25] Zixuan Jiang, Jiaqi Gu, Mingjie Liu, and David Z Pan. Delving into effective gradient matching for dataset condensation. In *2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, pages 1–6. IEEE, 2023. 3

[26] Anahita Fathi Kazerooni, Nastaran Khalili, Xinyang Liu, Debanjan Haldar, Zhifan Jiang, Syed Muhammed Anwar, Jake Albrecht, Maruf Adewole, Udunna Anazodo, Hannah Anderson, et al. The brain tumor segmentation (brats) challenge 2023: Focus on pediatrics (cbtn-connect-dipgr-asnr-miccai brats-peds). *arXiv preprint arXiv:2305.17033*, 2023. 5, 6, 7, 8, 2, 3, 4

[27] Aishik Konwer, Chao Chen, and Prateek Prasanna. Magnet: Modality-agnostic network for brain tumor segmentation and characterization with missing modalities. In *International Workshop on Machine Learning in Medical Imaging*, pages 361–371. Springer, 2023. 1

[28] Aishik Konwer, Xiaoling Hu, Joseph Bae, Xuan Xu, Chao Chen, and Prateek Prasanna. Enhancing modality-agnostic representations via meta-learning for brain tumor segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21415–21425, 2023. 2

[29] Dominic LaBella, Maruf Adewole, Michelle Alonso-Basanta, Talissa Altes, Syed Muhammad Anwar, Ujjwal Baid, Timothy Bergquist, Radhika Bhalerao, Sully Chen, Verena Chung, et al. The asnr-miccai brain tumor segmentation (brats) challenge 2023: Intracranial meningioma. *arXiv preprint arXiv:2305.07642*, 2023. 5, 2

[30] Guang Li, Ren Togo, Takahiro Ogawa, and Miki Haseyama. Dataset distillation using parameter pruning. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 107(6):936–940, 2024. 3

[31] Kunchang Li, Yali Wang, Junhao Zhang, Peng Gao, Guanglu Song, Yu Liu, Hongsheng Li, and Yu Qiao. Uniformer: Unifying convolution and self-attention for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 5, 6, 7, 1, 3

[32] Weibin Liao, Haoyi Xiong, Qingzhong Wang, Yan Mo, Xuhong Li, Yi Liu, Zeyu Chen, Siyu Huang, and Dejing Dou. Muscle: Multi-task self-supervised continual learning to pre-train deep models for x-ray images of multiple body parts. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 151–161. Springer, 2022. 1, 2

[33] Hong Liu, Dong Wei, Donghuan Lu, Jinghan Sun, Liansheng Wang, and Yefeng Zheng. M^3ae: multimodal representation learning for brain tumor segmentation with missing modal-

ities. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1657–1665, 2023. 1, 2, 3, 5, 6, 7

[34] Noel Loo, Ramin Hasani, Alexander Amini, and Daniela Rus. Efficient dataset distillation using random feature approximation. *Advances in Neural Information Processing Systems*, 35:13877–13891, 2022. 3

[35] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 5, 1

[36] Adriënne M Mendrik, Koen L Vincken, Hugo J Kuijf, Marcel Breeuwer, Willem H Bouvy, Jeroen De Bresser, Amir Alansary, Marleen De Bruijne, Aaron Carass, Ayman El-Baz, et al. Mrbrains challenge: online evaluation framework for brain image segmentation in 3t mri scans. *Computational intelligence and neuroscience*, 2015(1):813696, 2015. 5, 6, 2, 3, 4

[37] Ahmed W Moawad, Anastasia Janas, Ujjwal Baid, Divya Ramakrishnan, Leon Jekel, Kiril Krantchev, Harrison Moy, Rachit Saluja, Klara Osenberg, Klara Wilms, et al. The brain tumor segmentation (brats-mets) challenge 2023: Brain metastasis segmentation on pre-treatment mri. *ArXiv*, 2023. 5, 6, 7, 8, 2, 3

[38] Timothy Nguyen, Roman Novak, Lechao Xiao, and Jaehoon Lee. Dataset distillation with infinitely wide convolutional networks. *Advances in Neural Information Processing Systems*, 34:5186–5198, 2021. 3

[39] Narinder Singh Punn and Sonali Agarwal. Bt-unet: A self-supervised learning framework for biomedical image segmentation using barlow twins with u-net models. *Machine Learning*, 111(12):4585–4600, 2022. 1

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 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*, pages 234–241. Springer, 2015. 5, 6, 7, 1, 3

[41] Jonathan Shapey, Aaron Kujawa, Reuben Dorent, Guotai Wang, Alexis Dimitriadis, Diana Grishchuk, Ian Paddick, Neil Kitchen, Robert Bradford, Shakeel R Saeed, et al. Segmentation of vestibular schwannoma from mri, an open annotated dataset and baseline algorithm. *Scientific Data*, 8(1): 286, 2021. 5, 6, 7, 8, 2, 3, 4

[42] Junjie Shi, Li Yu, Qimin Cheng, Xin Yang, Kwang-Ting Cheng, and Zengqiang Yan. M^2ftrans: Modality-masked fusion transformer for incomplete multi-modality brain tumor segmentation. *IEEE Journal of Biomedical and Health Informatics*, 2023. 2

[43] Aron S Talai, Jan Sedlacik, Kai Boelmans, and Nils D Forkert. Utility of multi-modal mri for differentiating of parkinson's disease and progressive supranuclear palsy using machine learning. *Frontiers in Neurology*, 12:648548, 2021. 2

[44] Aiham Taleb, Christoph Lippert, T Klein, and M Nabi. Self-supervised learning for medical images by solving multimodal jigsaw puzzles. *Ieee Transactions on Medical Imaging*, 12729:661–673, 2017. 1, 3

[45] Aiham Taleb, Christoph Lippert, Tassilo Klein, and Moin Nabi. Multimodal self-supervised learning for medical image analysis. In *International conference on information processing in medical imaging*, pages 661–673. Springer, 2021. 3

[46] Yucheng Tang, Dong Yang, Wenqi Li, Holger R Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. Self-supervised pre-training of swin transformers for 3d medical image analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20730–20740, 2022. 1, 2, 5, 6, 7, 3

[47] Ekin Tiu, Ellie Talius, Pujan Patel, Curtis P Langlotz, Andrew Y Ng, and Pranav Rajpurkar. Expert-level detection of pathologies from unannotated chest x-ray images via self-supervised learning. *Nature Biomedical Engineering*, 6(12): 1399–1406, 2022. 1, 2

[48] Jeya Maria Jose Valanarasu, Yucheng Tang, Dong Yang, Ziyue Xu, Can Zhao, Wenqi Li, Vishal M Patel, Bennett Landman, Daguang Xu, Yufan He, et al. Disruptive autoencoders: Leveraging low-level features for 3d medical image pre-training. *arXiv preprint arXiv:2307.16896*, 2023. 1, 2, 3, 5, 6, 7

[49] Fuying Wang, Yuyin Zhou, Shujun Wang, Varut Vardhanabhuti, and Lequan Yu. Multi-granularity cross-modal alignment for generalized medical visual representation learning. *Advances in Neural Information Processing Systems*, 35:33536–33549, 2022. 2

[50] Hu Wang, Yuanhong Chen, Congbo Ma, Jodie Avery, Louise Hull, and Gustavo Carneiro. Multi-Modal Learning With Missing Modality via Shared-Specific Feature Modelling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15878–15887, 2023. 2

[51] Kai Wang, Bo Zhao, Xiangyu Peng, Zheng Zhu, Shuo Yang, Shuo Wang, Guan Huang, Hakan Bilen, Xinchao Wang, and Yang You. Cafe: Learning to condense dataset by aligning features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12196–12205, 2022. 3

[52] Tongzhou Wang, Jun-Yan Zhu, Antonio Torralba, and Alexei A Efros. Dataset distillation. *arXiv preprint arXiv:1811.10959*, 2018. 2, 3

[53] Linshan Wu, Jiaxin Zhuang, and Hao Chen. Voco: A simple-yet-effective volume contrastive learning framework for 3d medical image analysis. *arXiv preprint arXiv:2402.17300*, 2024. 1, 3, 5, 6, 7

[54] Yutong Xie, Jianpeng Zhang, Lingqiao Liu, Hu Wang, Yiwen Ye, Johan Verjans, and Yong Xia. Refs: A hybrid pre-training paradigm for 3d medical image segmentation. *Medical Image Analysis*, 91:103023, 2024. 3

[55] Zhenda Xie, Zheng Zhang, Yue Cao, Yutong Lin, Jianmin Bao, Zhuliang Yao, Qi Dai, and Han Hu. Simmim: A simple framework for masked image modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9653–9663, 2022. 5, 6, 7, 3

[56] Xiangyi Yan, Junayed Naushad, Chenyu You, Hao Tang, Shanlin Sun, Kun Han, Haoyu Ma, James S Duncan, and Xiaohui Xie. Localized region contrast for enhancing self-supervised learning in medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 468–478. Springer, 2023. 1, 2

[57] Chuyan Zhang, Hao Zheng, and Yun Gu. Dive into the details of self-supervised learning for medical image analysis. *Medical Image Analysis*, 89:102879, 2023. 1

[58] Bo Zhao and Hakan Bilen. Dataset condensation with differentiable siamese augmentation. In *International Conference on Machine Learning*, pages 12674–12685. PMLR, 2021. 3

[59] Bo Zhao and Hakan Bilen. Dataset condensation with distribution matching. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6514–6523, 2023. 3

[60] Bo Zhao, Konda Reddy Mopuri, and Hakan Bilen. Dataset condensation with gradient matching. *arXiv preprint arXiv:2006.05929*, 2020. 2, 3

[61] Hong-Yu Zhou, Chixiang Lu, Sibei Yang, Xiaoguang Han, and Yizhou Yu. Preservational learning improves self-supervised medical image models by reconstructing diverse contexts. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3499–3509, 2021. 1, 2

[62] Hong-Yu Zhou, Chixiang Lu, Chaoqi Chen, Sibei Yang, and Yizhou Yu. A unified visual information preservation framework for self-supervised pre-training in medical image analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 3

[63] Tongxue Zhou, Su Ruan, and Haigen Hu. A literature survey of mr-based brain tumor segmentation with missing modalities. *Computerized Medical Imaging and Graphics*, 104: 102167, 2023. 2

[64] Yongchao Zhou, Ehsan Nezhadarya, and Jimmy Ba. Dataset distillation using neural feature regression. *Advances in Neural Information Processing Systems*, 35:9813–9827, 2022. 2

[65] Zongwei Zhou, Vatsal Sodha, Jiaxuan Pang, Michael B Gotway, and Jianming Liang. Models genesis. *Medical image analysis*, 67:101840, 2021. 1, 2, 5, 6, 7, 3