A Foundation Model for Brain Lesion Segmentation with Mixture of Modality Experts

Xinru Zhang[1,2,*], Ni Ou[3], Berke Doga Basaran[4], Marco Visentin[4],
Mengyun Qiao[2], Renyang Gu[5], Cheng Ouyang[6],
Yaou Liu[7], Paul M. Matthews[2,8], Chuyang Ye[1,✉], Wenjia Bai[2,4,✉]

[1]School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing, China
[2]Department of Brain Sciences, Imperial College London, London, UK
[3]School of Automation, Beijing Institute of Technology, Beijing, China
[4]Department of Computing, Imperial College London, London, UK
[5]Department of Bioengineering, Imperial College London, London, UK
[6]Institute of Clinical Sciences, Imperial College London, London, UK
[7]Beijing Tiantan Hospital, Capital Medical University, Beijing, China
[8]UK Dementia Research Institute, Imperial College London, London, UK

Abstract. Brain lesion segmentation plays an essential role in neurological research and diagnosis. As brain lesions can be caused by various pathological alterations, different types of brain lesions tend to manifest with different characteristics on different imaging modalities. Due to this complexity, brain lesion segmentation methods are often developed in a task-specific manner. A specific segmentation model is developed for a particular lesion type and imaging modality. However, the use of task-specific models requires predetermination of the lesion type and imaging modality, which complicates their deployment in real-world scenarios. In this work, we propose a universal foundation model for 3D brain lesion segmentation, which can automatically segment different types of brain lesions for input data of various imaging modalities. We formulate a novel *Mixture of Modality Experts* (MoME) framework with multiple expert networks attending to different imaging modalities. A hierarchical gating network combines the expert predictions and fosters expertise collaboration. Furthermore, we introduce a curriculum learning strategy during training to avoid the degeneration of each expert network and preserve their specialisation. We evaluated the proposed method on nine brain lesion datasets, encompassing five imaging modalities and eight lesion types. The results show that our model outperforms state-of-the-art universal models and provides promising generalisation to unseen datasets.

Keywords: Foundation Model · Brain Lesion Segmentation · Mixture of Experts

---

* Work conducted as a visiting PhD student at Imperial College London, under the joint supervision of ✉corresponding authors.
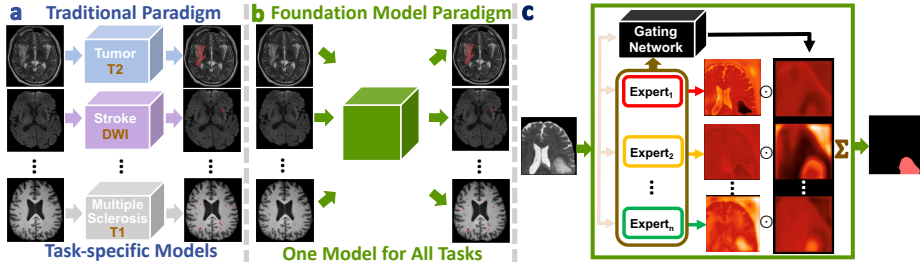
Fig. 1: Different paradigms for brain lesion segmentation: a) the traditional paradigm that trains multiple task-specific models; b) the foundation model paradigm that trains a single universal model for multiple tasks; c) the proposed *mixture of modality experts* (MoME) framework for constructing the foundation model.

1   Introduction

Brain lesion segmentation plays an essential role in neurological research and diagnosis by offering a quantitative description of pathologies [14, 34], and deep learning has greatly advanced the progress of brain lesion segmentation in recent years [3, 9, 31]. Since brain lesions tend to manifest with different characteristics due to various patient-dependent pathological alterations, the segmentation of different types of brain lesions requires the use of different *magnetic resonance imaging* (MRI) modalities [26, 27, 32], which increases the segmentation complexity. In the traditional paradigm (Fig. 1a), separate task-specific models are trained to segment specific types of brain lesions on the corresponding MRI modalities that are sensitive to the lesions [34]. This paradigm necessitates the development of numerous models for clinical analysis tasks [10, 28]. Clinicians are compelled to manually choose the optimal model for each patient-specific MRI scan based on predetermined brain lesion type criteria. This process hinders practical deployment of the *artificial intelligence* (AI) models. Contrary to this traditional paradigm, we propose a task-agnostic universal foundational model (Fig. 1b), which is capable of automatically segmenting various types of brain lesions based on inputs of different MRI modalities.

Foundation models have gained popularity in medical image segmentation, thanks to the success of the *Segment Anything Model* (SAM) [15] and its variants [11, 20, 29], as well as other innovative foundation models [4, 10, 28, 36] specifically developed for medical imaging. Earlier works aim to develop a universal model across all modalities, organs, and tasks. However, due to the complexity of medical imaging data and tasks, the segmentation accuracy of such a universal model is often limited [33]. Thus, foundation models focusing on narrower scopes are explored, which are designed for a particular modality or a particular organ/task [33]. For example, several models are proposed specifically for *computed tomography* (CT) image segmentation [19, 28, 30], showing promising results for anatomical structure and large tumour segmentation.

Despite these progresses, most existing foundation models may not be directly applicable to brain lesion segmentation. Most noticeably, universal brain lesion segmentation requires the capability of handling the challenge of input modality diversity, which is not addressed by many of the existing models. SAM-inspired methods, such as SAM-Med3D [29] and MedSAM [20], require human-given prompts, which can be impractical for brain lesion segmentation as the lesions can be scattered across the brain. Neuralizer [9] demonstrates the potential of a foundation model for brain imaging that generalises to new tasks by learning transferable features from a context set to the input image. It is evaluated on 2D coronal slices, which might limit its use on 3D brain lesion segmentation task. Hermes [10] is proposed with the awareness of the modality diversity, where modality information is embedded as a prior in model training. However, it does not fully explore the relationship between different modalities. The development of a universal foundational model for brain lesion segmentation with diverse input modalities is still an open problem.

In this study, we propose a *Mixture of Modality Experts* (MoME) model, a foundation model for brain lesion segmentation capable of handling diverse imaging modalities and accommodating various types of brain lesions. As shown in Fig. 1c, inspired by *Mixture of Experts* (MoE) [5, 23, 37], MoME enhances the model capacity by assembling a team of experts, where each expert is specialised to handle a particular imaging modality. A hierarchical gating network is developed to combine the expert predictions and foster collaborative expertise exploration. Moreover, to avoid the degeneration of each expert network and preserve their specialisation, we introduce a curriculum learning strategy for training the MoME model. For evaluation, we conducted comphrehensive experiments on nine brain imaging datasets comprising a total number of 6,585 annotated 3D images, with five MRI modalities and eight lesion types. The results show that MoME outperforms competing methods in multiple aspects and provides improved generalisation to unseen datasets. The codes of our method are available at https://github.com/ZhangxinruBIT/MoME.

2   Methods

2.1   Method Overview

The architecture of MoME features two major components, namely the expert networks and the gating network. Each expert network is adept in a specific domain of imaging modality, contributing to expert specialisation. The gating network functions as a decision-maker and ensembles the expert networks by dynamically adjusting the weights of their outputs with hierarchical guidance. **Regarding the training of MoME**, an anti-degeneration strategy based on curriculum learning is developed, where model training gradually transitions from specialising the role of each expert to determining the entire model through collaboration.

2.2   Detailed Design

Expert Specialisation Suppose there are N imaging modalities $\{\mathcal{M}_i\}_{i=1}^N$ of interest, where $\mathcal{M}_i$ represents the i-th modality. To address the problem of modality diversity in brain lesion segmentation, in MoME, N expert networks $\{E_i\}_{i=1}^N$ are developed, where E_i denotes the i-th expert. Each expert E_i is expected to specialise in a specific modality $\mathcal{M}_i$. Without losing generality, we assume that each expert network follows a multi-resolution encoder-decoder architecture, which is common for medical image segmentation such as in U-Net [13]. We focus on the five most widely used brain imaging modalities, including T1-weighted, T2-weighted, T1 contrast-enhanced (T1ce), FLAIR, and *diffusion weighted imaging* (DWI), and thus $N = 5$. Note that the experts here are built for 3D brain image segmentation, distinguishing them from sparse MoE architectures for 2D natural images where a higher number of experts can be afforded and routing is employed for expert selection [24, 35]. Due to the smaller number of experts, we do not explore expert routing in this work.

Hierarchical Gating Network Following the specialisation of expert networks, we propose a gating network G that facilitates the collaboration and conducts voxel-wise weighted aggregation among the expert outputs. Note that different imaging modalities can be correlated and provide complementary information. For example, FLAIR appears similar to T2 but with fluid being attenuated. They may both contribute to lesion segmentation with different features, and the outputs of an expert that is not specialized in the input modality may still provide complementary useful information. To enhance collaboration, the gating network is hierarchical and leverages multi-resolution information jointly.

Mathematically, given an input image $\mathbf{X}$, each expert E_i generates outputs at multiple resolution levels on the decoder side, denoted by $E_i(\mathbf{X}) = \{\mathbf{e}_i^l\}_{l=1}^L$, where $\mathbf{e}_i^l$ is the output of E_i at the l-th resolution level and L indicates the number of resolution levels. The gating network G generates a set of hierarchical voxel-wise weight maps at different resolution levels for each expert, denoted as $\{\mathbf{g}_i^l\}_{l=1}^L$. Combining the expert networks and the gating network, the final output $\mathbf{O}$ is formulated as

$$\mathbf{O} = \left\{\mathbf{o}^l\right\}_{l=1}^L \quad \text{with} \quad \mathbf{o}^l = \sum_{i=1}^N \mathbf{e}_i^l \odot \mathbf{g}_i^l, \tag{1}$$

where $\mathbf{o}^l$ is the aggregated output of all experts at the l-th level and $\odot$ represents voxel-wise multiplication. The segmentation at the l-th level is obtained by applying the softmax operator to $\mathbf{o}^l$. During training, deep supervision [13, 17] is employed to guide the segmentation at multiple resolution levels. At the inference stage, only the highest resolution level determines the final segmentation result [13, 17].

Curriculum Learning to Alleviate Expert Degeneration Although it is feasible to simply jointly train the experts and gating network, this may lead to

expert degeneration, where each expert no longer focuses on a specific modality and their activation becomes imbalanced [6, 7], reducing the capability of the whole model to handle diverse modalities. To address this issue, we propose a curriculum learning strategy for model training, which gradually transitions from specialising the experts to determining the whole model, including gating-based expert collaboration and expert refinement.

Specifically, the curriculum learning is achieved with a dynamic loss $\mathcal{L}_{\mathrm{cl}}$ that comprises the specialisation loss $\mathcal{L}_i$ and the collaboration loss $\mathcal{L}_{\mathrm{MoME}}$:

$$\mathcal{L}_{\mathrm{cl}} = f_{\mathrm{epoch}} \cdot \sum_{i=1}^{N} \mathbb{1}_{\mathbf{X} \in \mathcal{M}_i} \cdot \mathcal{L}_i + (1 - f_{\mathrm{epoch}}) \cdot \mathcal{L}_{\mathrm{MoME}}. \tag{2}$$

Here, $\mathbb{1}_{\mathbf{X} \in \mathcal{M}_i}$ is an indicator function that is one when the input image $\mathbf{X}$ belongs to the modality $\mathcal{M}_i$ and zero otherwise, and f_{epoch} is a dynamic weight that changes with the epoch iteration. f_{epoch} is designed as $f_{\mathrm{epoch}} = (1 - \mathrm{epoch}_{\mathrm{current}}/\mathrm{epoch}_{\mathrm{total}})^2$, where $\mathrm{epoch}_{\mathrm{current}}$ and $\mathrm{epoch}_{\mathrm{total}}$ are the current epoch and total number of epochs, respectively. In this way, during the early training stage, f_{epoch} is greater, which emphasises the loss $\mathcal{L}_i$ ($i \in \{1, \ldots, N\}$) for each expert E_i that matches the modality of $\mathbf{X}$. During the later training stage, f_{epoch} decreases and $\mathcal{L}_{\mathrm{MoME}}$ becomes gradually more emphasised.

$\mathcal{L}_i$ is designed with deep supervision at each resolution level as

$$\mathcal{L}_i = \sum_{l=1}^{L} k^l \cdot \mathcal{L}(\sigma(\mathbf{e}_i^l), d^l(\mathbf{Y})), \tag{3}$$

where k_l is a weight for the l-th level, $\sigma(\cdot)$ is the softmax activation, $d^l(\cdot)$ is a down-sampling operator that matches the resolution of the annotation $\mathbf{Y}$ with that of the l-th level, and $\mathcal{L}(\cdot, \cdot)$ is the combination of the soft Dice loss and cross-entropy loss [13]. $\mathcal{L}_{\mathrm{MoME}}$ is formulated with the collaboration between experts as

$$\mathcal{L}_{\mathrm{MoME}} = \sum_{l=1}^{L} k^l \cdot \mathcal{L}(\sigma(\mathbf{o}^l), d^l(\mathbf{Y})). \tag{4}$$

2.3 Implementation Details

We implement each expert network and the gating network with the nnU-Net architecture [13], which is a strong segmentation backbone as shown in [13, 28]. The input to the gating network is a combination of raw images and shallow features extracted from the first encoder layers of the experts.

As each expert network is expected to specialise in one imaging modality, the expert networks are first trained separately to acquire the modality knowledge. Then, the gating network is trained jointly with the fine-tuning of the expert networks based on the dynamic curriculum learning loss Eq. 2.

All the hyperparameters (including $\{k_l\}_{l=1}^{L}$) follow the default setting in nnU-Net [13], except that 1) a low learning rate of $1e^{-3}$ is set for the dynamic curriculum learning loss [25] and 2) the number of training epochs is set to 800, a sufficient number for convergence.

3   Experiments

3.1   Dataset and Task Description

We curated six publicly available brain lesion segmentation datasets (MSSEG [8], OASIS [22], ISLES2022 [12], BraTS2021 [2], WMH2017 [16], and ATLAS2.0 [18])
and three in-house datasets (Tumour1, Tumour2, and WMHsMix), amounting to a total number of 6,585 3D images with brain lesion annotations. The dataset detail is reported in Supple. Table II. All images were registered to the MNI152 template using ANTs [1]. Then, brain masks were applied to the images for skull removal. Cropping was performed so that the size of all images was standardised to $160 \times 196 \times 160$ with a voxel size of $1 \times 1 \times 1$ mm^3. Each of the six public datasets was split into training and test sets. The MoME model was trained on the combined training sets from all six public datasets. It was then evaluated on the test sets from the public datasets (**seen datasets**) and the three in-house datasets (**unseen datasets**), which were never used for training and allowed the assessment of the generalisation of MoME.

3.2   Competing Task-specific and Foundation Models

MoME was compared with the following task-specific and foundation models.

Task-specific nnU-Nets: The nine datasets cover a variety of brain lesions and modalities. We regard the segmentation of each modality for each dataset as a distinct task [4], resulting in a total number of 17 tasks (14 tasks for seen datasets and 3 tasks for unseen datasets; task IDs listed in Supple. Table I). Thus, 17 supervised task-specific nnU-Net models were trained.

nnU-Net: A single nnU-Net [13] was trained with the combined training sets from all public datasets. It can be viewed as a baseline foundation model.

Multi-Talent [28]: Multi-Talent demonstrates superior performance in handling conflicting class definitions for multi-dataset training and shows promising lesion segmentation results. We trained it with the combined training sets.

Hermes [10]: Hermes is a foundation model that injects context prior knowledge (e.g. anatomical region and imaging modality) into the segmentation network to address the challenges of data and modality diversities. We adapted Hermes for our task and incorporated the imaging modality information as prior knowledge. Hermes was also trained with the combined training sets.

SAM-Med3D [29]: This is a SAM-based [15] model trained on a large-scale 3D medical image dataset. We fine-tuned SAM-Med3D with the combined training sets and provided a 10-point prompt [29] during the inference.

All models were trained for 800 epochs on Nvidia A100 GPUs. We used the default hyperparameters for the transformer-based SAM-Med3D. All the other models were based on nnU-Net and used its default hyperparameters.

3.3   Results

Performance on Seen Datasets Table 1 compares the average Dice score of MoME with the competing models and reports the overall GPU memory usage

Table 1: Average Dice score of MoME and the competing methods on seen datasets. The best results among the foundation models are highlighted in bold. The GPU memory usage for model training is also displayed.

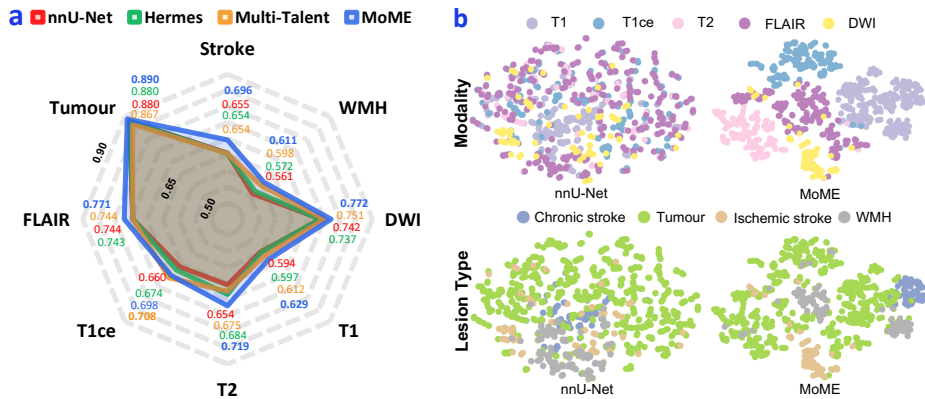| Dice	Foundation Model					Task-specific
	nnU-Net	SAM-Med3D	Multi-Talent	Hermes	MoME	nnU-Nets
Image-level	0.8007	0.7322	0.7948	0.8017	**0.8204**	0.8202
Task-level	0.6654	0.5249	0.6825	0.6720	**0.7026**	0.7099
Dataset-level	0.6561	0.5083	0.6681	0.6603	**0.6938**	0.6950
GPU memory usage (GB)	7.5	32.5	11.9	7.5	38.2	104.4

Fig. 2: More detailed analysis of the MoME result on seen datasets. a) A radar chart that compares the average Dice score of foundation models from the perspectives of different modalities and lesion types. b) t-SNE plots of latent spaces for nnU-Net and MoME, where each dot represents a brain image.

for training each model. To mitigate the bias caused by the varying numbers of images in different test sets, we report the Dice score with different levels of averaging: across the 14 seen tasks (task-level), across 6 seen datasets (dataset-level), and across 1,230 test images (image-level). MoME outperforms the other foundation models, for example, achieving 2% to 4% higher Dice scores than nnU-Net, Multi-Talent, and Hermes. It outperforms SAM-Med3D by a larger margin of 7% to 15%. This is possibly because SAM-Med3D requires point prompts to indicate a foreground region, which may not be suitable for brain lesion segmentation, where lesions can be numerous, small, and scattered across the brain, such as *white matter hyperintensity* (WMH) lesions [3,16]. The performance of MoME is comparable to task-specific nnU-Nets. However, the latter requires the training of 14 separate models, and its total amount of GPU memory consumption is about three times that of MoME.

A more detailed analysis of the results is shown in Fig. 2. Fig. 2a presents a comprehensive comparison between the foundation models from 8 more per-

Table 2: Average Dice scores on unseen datasets. The top scores are highlighted in bold.

Datasets	Foundation Model				Task-specific nnU-Nets
	nnU-Net	SAM-Med3D	Hermes	MoME	
Tumour1	0.8358	0.8339	0.8446	**0.8545**	0.8518
Tumour2	0.7959	0.8146	0.7968	**0.8293**	0.8157
WMHsMix	0.7032	0.5479	0.6968	0.7015	**0.7508**

Table 3: Ablation studies of MoME. The best scores are in bold.

Modality experts		✔	✔	✔	✘
Hierarchical gating		✔	✔	✘	✘
Curriculum learning		✔	✘	✘	✘
Seen Datasets	Image-level	**0.8204**	0.8142	0.8117	0.8027
	Task-level	**0.7026**	0.6969	0.6878	0.6529
	Dataset-level	**0.6938**	0.6852	0.6777	0.6488
Unseen Datasets		**0.7951**	0.7889	0.7912	0.7814

spectives, reporting the average Dice score for different modalities and different lesion types. For 7 out of the 8 perspectives, MoME (blue curve) outperforms the other foundation models (SAM-Med3D is not shown due to relatively lower performance but reported in Supple. Table I). Fig. 2b analyses the latent features obtained at the bottleneck of the gating network in MoME and the single nnU-Net using t-SNE plots [21]. It shows that MoME achieves a more discriminative clustering for images of different modalities and lesion types compared to the nnU-Net, which may be relevant to its improved performance.

Generalisation to Unseen Datasets Table 2 reports the segmentation performance on the three unseen in-house datasets for evaluating model generalisability. Multi-Talent is not included as it cannot be directly applied to unseen datasets. Again, MoME outperforms other competing foundation models. On two unseen datasets (Tumour1 and Tumour2), it even exceeds the task-specific nnU-Nets, which can be considered the upper-bound benchmark as they had access to the training data from the unseen datasets. On the other unseen dataset (WMHsMix), MoME still performs relatively well and achieves an average Dice score of 0.7015, where the dataset contains diverse WMH lesions of five diseases (see Supple. Table II).

Ablation Studies Ablation studies were performed on the seen and unseen datasets to verify the contributions of the designs in MoME with results presented in Table 3. For the modality experts, ✘/✔ denotes randomly initialised experts versus specialised experts based on the imaging modality; for the hierarchical gating, ✘/✔ denotes the basic gating without hierarchical structure versus our hierarchical gating; and for the curriculum learning, ✘/✔ denotes the application of Eq. (4) versus Eq. (2) during model training. When the three contributions are sequentially removed, the segmentation performance tends to decrease gradually, which confirms the benefits of these contributions.

Experts Degeneration We further investigated the issue of expert degeneration by comparing the expert output, weighted by gating, without and with the proposed curriculum learning strategy for avoiding degeneration. The detailed results are given in Supple. Fig. I. The figure shows that without the curriculum

learning strategy, certain experts can become inactive, losing their modality-specific expertise. The proposed curriculum learning strategy ensures that the experts remain sensitive to modality-specific information.

4   Conclusion

We have developed a novel universal foundation model MoME for brain lesion segmentation, where a mixture of modality experts is incorporated to address the challenge of modality diversity. MoME enables collaboration between experts with hierarchical gating and mitigates expert degeneration with a curriculum learning strategy. Extensive experimental results on nine public and in-house datasets and 17 tasks show that it outperforms competing foundation models across different modalities and lesion types. It is also more efficient than task-specific models, providing a flexible way for model deployment.

Disclosure of Interests The authors declare no competing interests.

References

1. Avants, B.B., Tustison, N.J., Song, G., et al.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage **54**(3) (2011)
2. Baid, U., Ghodasara, S., Bilello, M., et al.: The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv:2107.02314 (2021)
3. Basaran, B.D., Zhang, W., Qiao, M., et al.: LesionMix: A lesion-level data augmentation method for medical image segmentation. In: MICCAI Workshop on Data Augmentation, Labelling, and Imperfections (2023)
4. Butoi, V.I., Gonzalez Ortiz, J.J., Ma, T., et al.: UniverSeg: Universal medical image segmentation. International Conference on Computer Vision (2023)
5. Chen, T., Chen, X., Du, X., et al.: AdaMV-MoE: Adaptive multi-task vision mixture-of-experts. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2023)
6. Chen, Z., Shen, Y., Ding, M., et al.: Mod-Squad: Designing mixtures of experts as modular multi-task learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023)
7. Chi, Z., Dong, L., Huang, S., et al.: On the representation collapse of sparse mixture of experts. Advances in Neural Information Processing Systems **35** (2022)

8. Commowick, O., Istace, A., Kain, M., et al.: Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. Scientific Reports **8**(1) (2018)
9. Czolbe, S., Dalca, A.V.: Neuralizer: General neuroimage analysis without retraining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023)
10. Gao, Y., Li, Z., Liu, D., et al.: Training like a medical resident: Universal medical image segmentation via context prior learning. IEEE/CVF Conference on Computer Vision and Pattern Recognition (2024)
11. Gong, S., Zhong, Y., Ma, W., et al.: 3DSAM-adapter: Holistic adaptation of SAM from 2D to 3D for promptable medical image segmentation. arXiv:2306.13465 (2023)
12. Hernandez Petzsche, M.R., de la Rosa, E., Hanning, U., et al.: ISLES 2022: A multicenter magnetic resonance imaging stroke lesion segmentation dataset. Scientific Data **9**(1) (2022)
13. Isensee, F., Jaeger, P.F., Kohl, S.A., et al.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2) (2021)
14. Kamnitsas, K., Ledig, C., Newcombe, V.F., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis **36** (2017)
15. Kirillov, A., Mintun, E., Ravi, N., et al.: Segment anything. ICCV (2023)
16. Kuijf, H.J., Biesbroek, J.M., De Bresser, J., et al.: Standardized assessment of automatic segmentation of white matter hyperintensities and results of the WMH segmentation challenge. IEEE Transactions on Medical Imaging **38**(11) (2019)
17. Lee, C.Y., Xie, S., Gallagher, P., et al.: Deeply-supervised nets. In: Artificial Intelligence and Statistics (2015)
18. Liew, S.L., Anglin, J.M., Banks, N.W., et al.: A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. Scientific Data **5**(1) (2018)
19. Liu, J., Zhang, Y., Chen, J.N., et al.: Clip-driven universal model for organ segmentation and tumor detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2023)
20. Ma, J., He, Y., Li, F., et al.: Segment anything in medical images. Nature Communications **15**(1) (2024)
21. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. Journal of Machine Learning Research **9**(11) (2008)
22. Marcus, D.S., Fotenos, A.F., Csernansky, J.G., et al.: Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. Journal of Cognitive Neuroscience **22**(12) (2010)
23. Ou, Y., Yuan, Y., Huang, X., et al.: Patcher: Patch transformers with mixture of experts for precise medical image segmentation. In: MICCAI (2022)
24. Puigcerver, J., Riquelme, C., Mustafa, B., et al.: From sparse to soft mixtures of experts. arXiv:2308.00951 (2023)
25. Rajbhandari, S., Li, C., Yao, Z., et al.: DeepSpeed-MoE: Advancing mixture-of-experts inference and training to power next-generation AI scale. In: International Conference on Machine Learning (2022)
26. Schmidt, G.P., Wintersperger, B., Graser, A., et al.: High-resolution whole-body magnetic resonance imaging applications at 1.5 and 3 Tesla: a comparative study. Investigative Radiology **42**(6) (2007)

27. Shah, A.H., Snelling, B., Bregy, A., et al.: Discriminating radiation necrosis from tumor progression in gliomas: a systematic review what is the best imaging modality? Journal of Neuro-Oncology **112** (2013)
28. Ulrich, C., Isensee, F., Wald, T., et al.: MultiTalent: A multi-dataset approach to medical image segmentation. In: MICCAI (2023)
29. Wang, H., Guo, S., Ye, J., et al.: SAM-Med3D. arXiv:2310.15161 (2023)
30. Wasserthal, J., Breit, H.C., Meyer, M.T., et al.: TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. Radiology: Artificial Intelligence **5**(5) (2023)
31. Wood, D.A., Kafiabadi, S., Al Busaidi, A., et al.: Deep learning models for triaging hospital head MRI examinations. Medical Image Analysis **78** (2022)
32. Wu, O., Christensen, S., Hjort, N., et al.: Characterizing physiological heterogeneity of infarction risk in acute human ischaemic stroke using MRI. Brain **129**(9) (2006)
33. Zhang, S., Metaxas, D.: On the challenges and perspectives of foundation models for medical image analysis. Medical Image Analysis **91** (2024)
34. Zhang, X., Liu, C., Ou, N., et al.: CarveMix: a simple data augmentation method for brain lesion segmentation. NeuroImage **271** (2023)
35. Zhang, Y., Cai, R., Chen, T., et al.: Robust mixture-of-expert training for convolutional neural networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2023)
36. Zhou, Y., Chia, M.A., Wagner, S.K., et al.: A foundation model for generalizable disease detection from retinal images. Nature **622**(7981) (2023)
37. Zhu, J., Zhu, X., Wang, W., et al.: Uni-perceiver-MoE: Learning sparse generalist models with conditional MoEs. Advances in Neural Information Processing Systems **35** (2022)