SwiFT V2: Towards Large-scale Foundation Model for Functional MRI

Jubin Choi [1], Heehwan Wang [1], Junbeom Kwon [2], Shinjae Yoo [3], Jiook Cha [1,2,4,5]

1 Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, South Korea
2 Department of Psychology, Seoul National University, Seoul, South Korea
3 Artificial Intelligence Department, Brookhaven National Laboratory, Upton, NY, USA.
4 Department of Brain and Cognitive Sciences, Seoul National University, Seoul, South Korea
5 Graduate School of Data Science, Seoul National University, Seoul, South Korea

{wnqlszoq123, dhkdgmlghks}@snu.ac.kr, kjb961013@gmail.com, sjyoo@bnl.gov
Corresponding Author: Jiook Cha (connectome@snu.ac.kr)

Abstract

Foundation models, leveraging large-scale datasets and extensive parameter counts, show unprecedented capabilities across various domains. Recent studies have explored foundation models for neuroimaging to effectively capture the complex dynamics of the human brain. However, training such models end-to-end on four-dimensional functional MRI data remains unexplored. Here, we introduce SwiFT V2, a fMRI foundation model based on the 4D Swin fMRI Transformer. We pre-trained SwiFT V2 using masked image modeling on large-scale aggregated resting-state fMRI datasets of 49,321 subjects. Especially, we trained models up to 8.8 billion parameters with maximal update parameterization technique, leading to stable and efficient scaling. We observed that these models follow neural scaling laws, where performance predictably improves with scale. Also, we showed that masked modeling pre-training enhances performance across various downstream tasks. These results validate the application of scaling principles to fMRI modeling and motivate the further development of large foundation models for neuroscience.

Keywords: foundation model; functional MRI; scaling laws; self-supervised learning

Introduction

Functional Magnetic Resonance Imaging (fMRI) offers high-dimensional spatiotemporal data for understanding brain function, but its complexity poses analytical challenges. While deep learning shows promise (Abrol et al., 2021), task-specific models often lack generalizability. Foundation models, pre-trained via self-supervision on vast datasets (Bommasani et al., 2021), learn versatile representations, transforming fields like NLP (Vaswani et al., 2017). Applying this to neuroscience (Caro et al., 2023; Malkiel, Rosenman, Wolf & Hendler, 2022) requires overcoming significant computational hurdles, especially for complex 4D fMRI data.

This work introduces SwiFT V2, based on the Swin 4D fMRI Transformer (Kim et al., 2023; Liu et al., 2021), as a scalable foundation model for fMRI. We investigate its scaling properties using self-supervised pre-training on large datasets. Key contributions include: (1) Training 4D fMRI Transformers up to 8.8B parameters via masked image modeling; (2) Using Maximal Update Parametrization (muP) for stable scaling (Yang et al., 2021); (3) Providing empirical evidence for neural scaling laws (Kaplan et al., 2020) in fMRI models; and (4) Highlighting potential for transfer learning to diverse downstream tasks in limited sample size settting.

Methods

SwiFT V2 Architecture

As shown in Figure 1, SwiFT V2 employs a SwiFT, hierarchical transformer with efficient 4D windowed self-attention, as an fMRI encoder. Patch embedding converts fMRI volumes to tokens, and spatial and temporal dimensions are merged between stages.

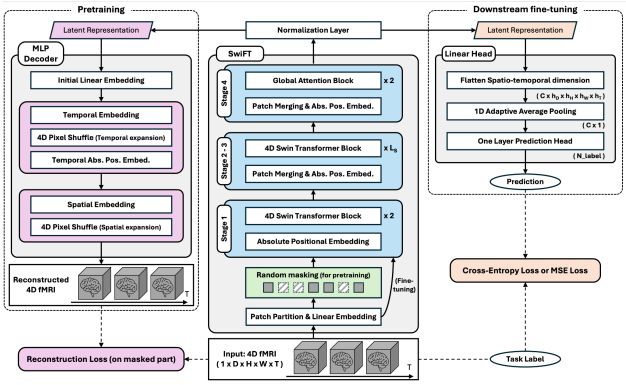

Figure 1: SwiFT V2 architecture

Pre-training and Scaling Techniques

We used self-supervised pre-training via Masked Image Modeling, similar to SimMIM (Xie et al., 2022). A fraction of tokenized input patches were masked, and the model's encoder-decoder structure was trained to reconstruct the masked token values using an L1 loss. Pre-training utilized large resting-state fMRI datasets (UK Biobank, HCP, ABCD) of 49,321 individuals.

To enable stable training of billion-parameter models, we implemented muP (Yang et al., 2021). It ensures a stable activation scale even when model width increases, confirmed via coordinate checks (Figure 2), allowing efficient hyperparameter transfer (mu-transfer) from smaller to larger models (up to 8.8B).

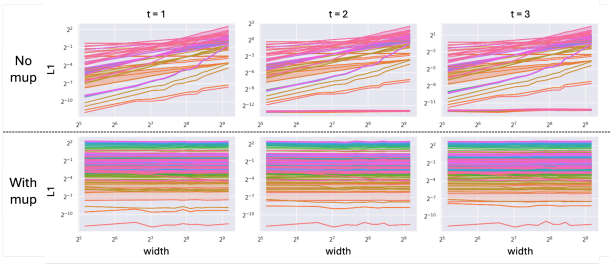

Figure 2: Maximal update parametrization results

Experiments & Results

Scaling Law Verification

We pre-trained SwiFT V2 models across sizes up to 8.8 billion parameters. Consistent with scaling laws in other domains (Kaplan et al., 2020; Zhai, Kolesnikov, Houlsby & Beyer; 2022), we observed a power-law relationship between test loss and three factors - the amount of compute, dataset size, and model size (Figure 3) using hyperparameters transferred via mu-transfer. Our largest model (8.8B), whose training is ongoing, has yet to fully converge to the performance suggested by this scaling trend. Nevertheless, increasing the model scale yields predictable improvements in learning representations from fMRI data.

Downstream Task Evaluation

To evaluate the practical utility of the learned representations, we fine-tuned pre-trained 1.2B SwiFT V2 checkpoints on various downstream tasks. Table 1 presents preliminary results on benchmark tasks using limited data samples per subject (N=8-10, 1 segment each): predicting sex, age, and cognitive scores on held-out subjects from UKB, HCP, and ABCD. While based on minimal fine-tuning data, these initial results suggest potential benefits from pre-training compared to training from scratch.

Table 1. Downstream Task Performance
(ACC: Accuracy; MAE: Mean Absolute Error)

Method	ABCD		HCP			UKB		
	Sex (ACC)	Intelligence (MAE)	Sex (ACC)	Age (MAE)	Intelligence (MAE)	Sex (ACC)	Age (MAE)	Intelligence (MAE)
From scratch	50.0	0.86	50.0	3.17	16.2	50.6	7.09	1.74
Fine-tuned	**50.2**	**0.85**	**57.6**	**3.13**	**16.1**	**56.9**	**6.44**	**1.72**

Conclusion

We scaled the SwiFT V2 4D fMRI Transformer to 8.8B parameters via masked modeling and muP. Observing neural scaling laws validates this large-scale, self-supervised approach for neuroscience. Our results show the potential of scaled learning for fMRI, suggesting that continued exploration of large foundation models is valuable for dissecting brain dynamics and enhancing predictive capabilities in the neuroscience area.

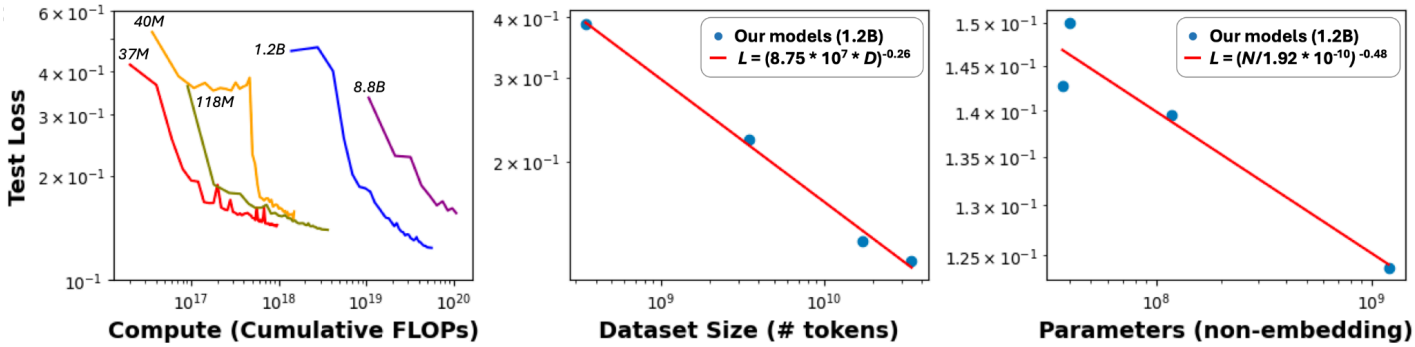

Figure 3: Neural scaling laws for SwiFT V2

Acknowledgments

An award for computer time was provided by the U.S. Department of Energy's (DOE) ASCR Leadership Computing Challenge (ALCC). This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy User Facility, using NERSC award ALCC for Leadership Computing Challenge Awards-ERCAP m4750-2024, supporting resources at the Argonne and the Oak Ridge Leadership Computing Facilities, a U.S. Department of Energy (DOE) Office of Science user facility at Argonne National Laboratory and Oak Ridge National Laboratory. Also, this work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2021R1C1C1006503, RS-2023-00266787, RS-2023-00265406, RS-2024-00421268, RS-2024-00342301, RS-2024-00435727, NRF-2021M3E5D2A01022515), by Creative-Pioneering Researchers Program through Seoul National University(No. 200-20240057, 200-20240135), by Semi-Supervised Learning Research Grant by SAMSUNG(No.A0342-20220009), by Identify the network of brain preparation steps for concentration Research Grant by LooxidLabs(No.339-20230001), by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)] by the MSIT(Ministry of Science, ICT), Korea, under the Global Research Support Program in the Digital Field program(RS-2024-00421268) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation), by the National Supercomputing Center with supercomputing resources including technical support(KSC-2023-CRE-0568), by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A3A2A02090597), by the Korea Health Industry Development Institute (KHIDI), and by the Ministry of Health and Welfare, Republic of Korea (HR22C1605), by Artificial intelligence industrial convergence cluster development project funded by the Ministry of Science and ICT(MSIT, Korea) & Gwangju Metropolitan City and by KBRI basic research program through Korea Brain Research Institute funded by Ministry of Science and ICT(25-BR-05-01).

References

Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., & Calhoun, V. (2021). Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nature communications*, *12*(1), 353.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

Caro, J. O., Fonseca, A. H. D. O., Averill, C., Rizvi, S. A., Rosati, M., Cross, J. L., ... & van Dijk, D. (2023). BrainLM: A foundation model for brain activity recordings. *bioRxiv*, 2023-09.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., ... & Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Kim, P., Kwon, J., Joo, S., Bae, S., Lee, D., Jung, Y., ... & Moon, T. (2023). Swift: Swin 4d fmri transformer. *Advances in Neural Information Processing Systems*, *36*, 42015-42037.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10012-10022).

Malkiel, I., Rosenman, G., Wolf, L., & Hendler, T. (2022, December). Self-supervised transformers for fMRI representation. In *International Conference on Medical*

Imaging with Deep Learning (pp. 895-913). PMLR.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Xie, Z., Zhang, Z., Cao, Y., Lin, Y., Bao, J., Yao, Z., ... & Hu, H. (2022). Simmim: A simple framework for masked image modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9653-9663).

Yang, G., Hu, E., Babuschkin, I., Sidor, S., Liu, X., Farhi, D., ... & Gao, J. (2021). Tuning large neural networks via zero-shot hyperparameter transfer. *Advances in Neural Information Processing Systems*, *34*, 17084-17097.

Zhai, X., Kolesnikov, A., Houlsby, N., & Beyer, L. (2022). Scaling vision transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12104-12113).