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 (Presenter)

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

SNU Connectome Lab

Brookhaven National Laboratory

1. Introduction

- **The Challenge**

Analyzing the brain's complex dynamics from high-dimensional, 4D fMRI data is a major computational challenge.
Task-specific deep learning models often lack the generalizability needed for broad neuroscience applications.

- **The Opportunity**

Foundation models, pre-trained on vast datasets using self-supervision, can learn versatile representations that transform scientific fields.

- **Our Goal**

(1) To develop **SwiFT V2**, the first end-to-end, large-scale foundation model for 4D fMRI data, and (2) to investigate if its performance predictably improves with scale, following neural scaling laws.

2. Methods

- **SwiFT V2 Architecture**

The model uses SwiFT [1], a hierarchical 4D Swin Transformer as an encoder, specifically designed for 4D spatiotemporal data with an efficient windowed self-attention mechanism. An MLP decoder reconstructs masked portions of the fMRI data during the pre-training phase.

- **Pre-training Strategy**

Self-Supervised learning: We used Masked Image Modeling (MIM), where the model learns to reconstruct randomly masked-out patches of the fMRI signal (Fig. 1).

Massive Dataset: The model was pre-trained on aggregated resting-state fMRI data from more than 50,000 **individuals** across the UK Biobank [2], HCP [3], and ABCD [4] cohorts.

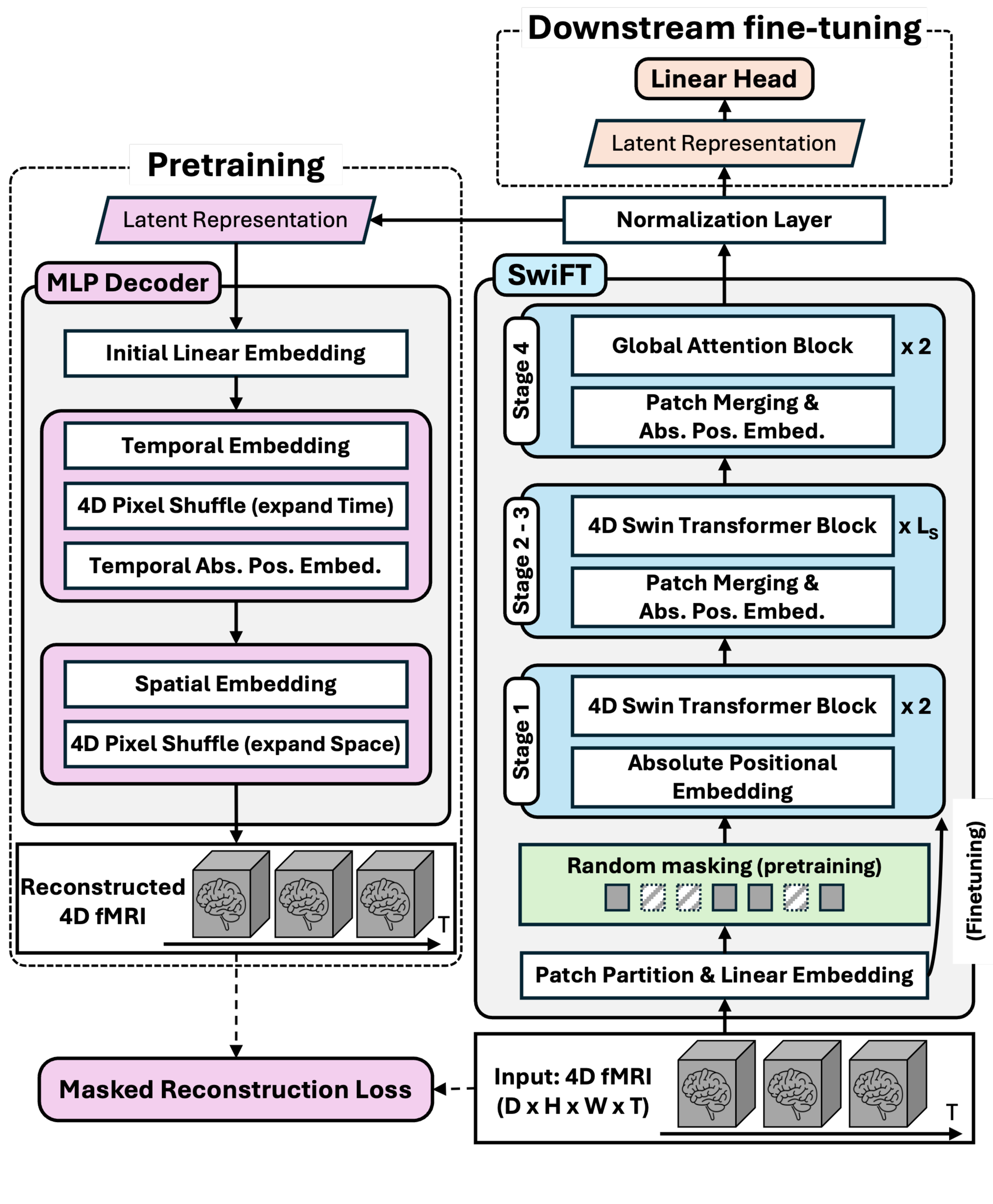

Figure 1: The SwiFT V2 architecture.
Our model consists of a 4D Swin fMRI Transformer encoder and an MLP decoder for pre-training.

3. Results & Discussions

3.1. Neural Scaling Laws in fMRI Modeling

- We successfully pre-trained SwiFT V2 models with up to **3 billion parameters**.
- Our experiments provide the first empirical evidence that fMRI foundation models adhere to **neural scaling laws** [5]. We observed a predictable, **power-law relationship** where the model's test loss consistently decreases with increases in (1) **Model Size** (number of parameters), (2) **Dataset Size** (number of tokens), and (3) **Compute** (total FLOPs) (Fig. 2).
- This finding validates that investing in larger models and datasets yields predictable performance improvements for learning representations from fMRI data.

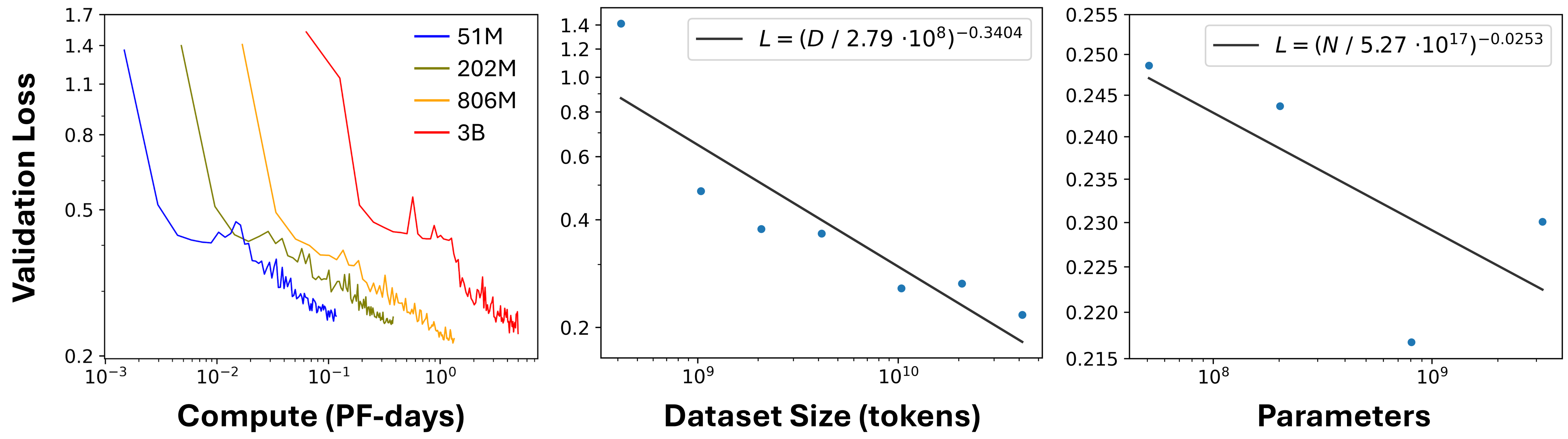

Figure 2: Neural Scaling laws in SwiFT V2. Validation loss improves predictably as a power-law function of Compute, Dataset Size, and the number of Parameters, confirming that fMRI models follow neural scaling laws

3.2. Pre-training Boosts Performance, but Larger Isn't Always Better

Our comprehensive downstream evaluation reveals two key findings:

- **Pre-training is highly effective** (Fig. 3). Our fine-tuned 51M model consistently outperforms its from-scratch counterpart across seven diverse neuroscience tasks.
- **Optimal model size is highly task-dependent**, challenging the simple "larger is always better" assumption (Table 1).
  - Smaller models (51M) excelled at in-distribution UKB tasks like depression and intelligence prediction.
  - Medium-to-large models (806M & 3B) were superior on challenging, out-of-distribution clinical datasets. The 806M model was best for HCP intelligence and pain prediction, while the 3B model was optimal for MDD treatment response.

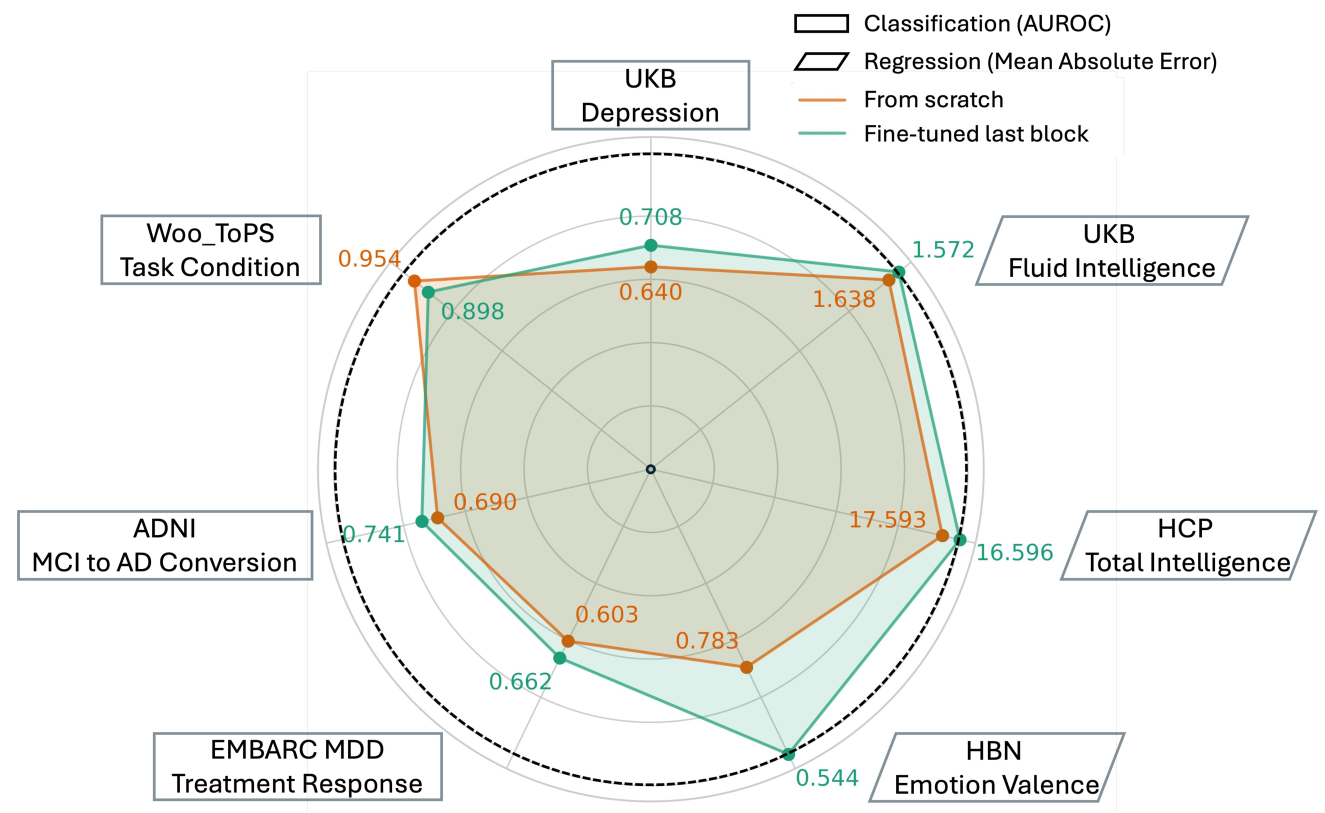

Figure 3: Preliminary downstream performance of the 51M parameter model. Fine-tuning shows potential benefits over training from scratch on several benchmark tasks. Best fine-tuned models from several configurations are reported. (**"Fine-tuned last block"** means that we only trained the last two layers of the pre-trained backbone, about 30% of the total parameters)

This complex relationship suggests the most effective model scale depends on the specific neurobiological features required by the downstream task, a critical direction for future research.

Table 1: Impact of Model Scale on Downstream Task Performance

Method		UKB		HCP	EMBARC [6]	Woo_ToPS [7]
		Depression (AUROC ↑)	Total Intelligence (MAE ↓)	Fluid Intelligence (MAE ↓)	MDD treatment response (AUROC ↑)	Task condition prediction (AUROC ↑)
51M	From scratch	0.640 ± 0.07	1.638 ± 0.06	17.593 ± 0.16	0.603 ± 0.06	0.954 ± 0.05
	Fine-tuned last block	**0.708 ± 0.01**	**1.577 ± 0.01**	16.691 ± 0.01	0.611 ± 0.06	0.898 ± 0.04
202M	Fine-tuned last block	0.483 ± 0.03	1.583 ± 0.01	16.659 ± 0.02	0.590 ± 0.05	0.467 ± 0.04
806M	Fine-tuned last block	0.620 ± 0.03	1.584 ± 0.01	**16.646 ± 0.00**	0.629 ± 0.06	**0.962 ± 0.03**
3B	Fine-tuned last block	0.572 ± 0.01	1.585 ± 0.01	16.720 ± 0.01	**0.637 ± 0.03**	0.913 ± 0.04

(Woo_ToPS: Tonic Pain Signature fMRI data from Woo and Lee et al. 2021[7]; MDD: Major Depressive Disorder; AUROC: Area Under the Receiver Operating Characteristic; MAE: Mean Absolute Error)

4. Conclusion

We successfully scaled a 4D fMRI foundation model, SwiFT V2, to 3 billion parameters and provided the first empirical validation of neural scaling laws for fMRI models. Our comprehensive downstream evaluations demonstrate that while pre-training provides a significant performance boost, the optimal model scale is highly task-dependent. These results highlight the potential of large, self-supervised models to advance neuroscience and open up new avenues for investigating how to best transfer learned knowledge to solve specific scientific and clinical problems.

Acknowledgment

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 NERSC, ALCF, and OLCF. This work was also supported by the National Research Foundation of Korea (NRF) and other grants.

References

[1] 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.
[2] Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., ... & Collins, R. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine, 12(3), e1001779.
[3] Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., & Wu-Minn HCP Consortium. (2013). The WU-Minn human connectome project: an overview. Neuroimage, 80, 62-79.
[4] Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., ... & Dale, A. M. (2018). The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. Developmental cognitive neuroscience, 32, 43-54.
[5] 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.
[6] Trivedi, M. H., McGrath, P. J., Fava, M., Parsey, R. V., Kurian, B. T., Phillips, M. L., ... & Weissman, M. M. (2016). Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. Journal of psychiatric research, 78, 11-23.
[7] Lee, J. J., Kim, H. J., Čeko, M., Park, B. Y., Lee, S. A., Park, H., ... & Woo, C. W. (2021). A neuroimaging biomarker for sustained experimental and clinical pain. Nature medicine, 27(1), 174-182.