Research Internship Proposal

Adapting Foundation Models to Study Brain Cortical Folding

Keywords: Deep learning, Multi-modal, Self-Supervised Learning, Medical Imaging, Foundation Models, Transfer Learning, Domain Adaptation, NeuroImaging

Context In recent years, foundation models, such as CLIP [10] and DINOv2 [8], have revolutionized computer vision by enabling strong performance across a wide range of tasks using large-scale pretraining on natural images. These models learn rich visual representations that are transferable across domains and tasks. However, a significant limitation of current foundation models is their lack of exposure to specialized medical data during training. In particular, neuroimaging data, such as cortical folding patterns [4] (commonly extracted from T1-w MRI scans), exhibit structural and statistical characteristics that differ from natural images. This creates a domain gap that impairs the performance of foundation models when directly applied to medical data. As such, there is a growing need to explore how these models can be adapted to new data modalities that fall outside their original training distribution.

Objectives The aim of this internship is to investigate and benchmark several recent and widely used techniques for adapting large-scale foundation models to cortical surface data. Data from several datasets have already been extracted and pre-processed, as explained in [4]. The methods to be explored include:

1. **Linear Probing** – Training a simple classifier on top of frozen foundation model features to assess their transferability.
2. **Feature Adaptation with Adapters** – Inserting lightweight adapter modules into the pretrained model to learn modality-specific representations while keeping most of the model fixed (e.g., AdapterFusion). [1, 9]
3. **Low-Rank Adaptation (LoRA)** – Efficiently fine-tuning subsets of weights via low-rank updates to achieve stronger adaptation with minimal parameter overhead. [2]
4. **Prompt Tuning and Prefix Tuning** – Modulating the input or internal activations of the model through learned prompts, requiring no changes to the model architecture. [3, 5–7, 11, 12]
5. **Full or Partial Fine-Tuning** – Adjusting some or all of the model's parameters using the medical dataset, which offers the greatest flexibility at the cost of computational efficiency and overfitting risk.

The internship will involve implementing and evaluating these methods on cortical folding pattern data, analyzing their strengths and limitations, and providing recommendations for adapting vision foundation models to medical imaging tasks.

Team This project will be carried out under the supervision of P. Gori (Télécom Paris, IPParis) and R. Guiavarch (Télécom Paris, IPParis)

Salary ∼600 euros/month.

Required background M2 in applied mathematics, statistics, computer science, engineering with a good knowledge of Python and deep learning.

How to apply Candidates are invited to send a CV to pietro.gori@telecom-paris.fr detailing their academic background and motivation.

References

[1] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., and Gelly, S. Parameter-efficient transfer learning for nlp. In *International Conference on Machine Learning (ICML)* (2019), PMLR. 1

[2] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, L., and Chen, W. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685* (2021). 1

[3] Jia, X., Tang, J., Li, Y., Yang, X., Wang, X., and Xie, L. Visual prompt tuning. In *European Conference on Computer Vision (ECCV)* (2022). 1

[4] Laval, J., et al. Towards a foundation model for cortical folding. In *Machine Learning in Clinical Neuroimaging (MLCN) - MICCAIW* (2024), vol. 15266 of *Lecture Notes in Computer Science*. 1

[5] Lester, B., Al-Rfou, R., and Constant, N. The power of scale for parameter-efficient prompt tuning. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2021). 1

[6] Li, X. L., and Liang, P. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)* (2021). 1

[7] Lu, Y., Liu, X., Zhang, Z., et al. Prompt distribution learning. In *European Conference on Computer Vision (ECCV)* (2022). 1

[8] Oquab, M., Darcet, T., Moutakanni, T., Bossard, R., Assran, M., Azab, M., Haziza, D., Traoré, I., Botbol, M., Dai, Z., et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193* (2023). 1

[9] Pfeiffer, J., Kamath, A., Rücklé, A., Cho, K., and Gurevych, I. Adapterfusion: Non-destructive task composition for transfer learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2020), Association for Computational Linguistics, pp. 5631–5641. 1

[10] Radford, A., Kim, J. W., Hallacy, J., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. Learning transferable visual models from natural language supervision. *International Conference on Machine Learning (ICML)* (2021). 1

[11] Wang, R., Zhang, J., Xu, Z., et al. Dualprompt: Complementary prompt tuning for vision transformers. In *European Conference on Computer Vision (ECCV)* (2022). 1

[12] Zhou, K., Yang, J., Loy, C. C., and Liu, Z. Learning to prompt for vision-language models. In *International Conference on Computer Vision (ICCV)* (2022). 1