



Hierarchical Representations for Transformers

Transformer models demonstrate notable effectiveness in numerous NLP and sequence modeling tasks. Their ability to process relatively long sequences enables them to generate lengthy, coherent outputs, such as complete paragraphs by GPT-3 or intricately organized images by DALL-E. However, these large language models are resource-intensive and expensive, restricting their practical use and accessibility. In particular, the self-attention mechanism scales very poorly with the sequence length.

Previous work has already explored the potential of hierarchical transformer architectures, for example [1, 2, 3, 4]. The goal of this project is to continue research in this direction and to come up with more capable, scalable and efficient hierarchical transformer architectures, and to evaluate the architecture on common NLP and vision tasks.

We will have weekly meetings to address questions, discuss progress and think about future ideas.



Requirements: Strong knowledge in Python and PyTorch. Strong foundation in deep learning. Previous experience with NLP and the transformer architecture is an advantage.

Interested? Please contact me for more details!

Contact

- Benjamin Estermann: besterma@ethz.ch, ETZ G60.1

References

- [1] Ze Liu et al. “Swin transformer: Hierarchical vision transformer using shifted windows”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 10012–10022.
- [2] Piotr Nawrot et al. “Hierarchical transformers are more efficient language models”. In: *arXiv preprint arXiv:2110.13711* (2021).
- [3] Chaitanya Ryali et al. “Hiera: A Hierarchical Vision Transformer without the Bells-and-Whistles”. In: *arXiv preprint arXiv:2306.00989* (2023).
- [4] Xiaosong Zhang et al. “Hivit: A simpler and more efficient design of hierarchical vision transformer”. In: *The Eleventh International Conference on Learning Representations*. 2023.