



# Self-Supervised Vision-Language Part-whole Induction with Structure Modeling

Studying how physical objects and biomass are composed and organised is one of the main tasks for artificial intelligence. It is widely recognised that there is a hierarchical structure inside both natural and artificial system. Therefore, it would be important and interesting to induce regular patterns of the underlying structures in the nature or artificial modality such as vision and language.



In this project, we aim to induce the cognitive grammar and hierarchically assemble constituent in both vision and language. We will migrate constituency parsing methods from the domain of natural language processing to computer vision, and explore whether the joint framework will be beneficial or not.

**Requirements:** Strong motivation, knowledge in deep learning, or a solid background in machine learning. Experience with PyTorch is an advantage. We will have weekly meetings to address questions, discuss progress and think about future ideas.

**Interested? Please contact us for more details!**

## Contact

- Fengyu Cai: [fengyu.cai@tu-darmstadt.de](mailto:fengyu.cai@tu-darmstadt.de)
- Zhao Meng: [zhmeng@ethz.ch](mailto:zhmeng@ethz.ch), ETZ G60.1