Background

- Recently, a new learning paradigm Pre-training, Finetuning and Prediction has demonstrated great effectiveness in a wide range of visual recognition tasks.

- Inspired by the advances in natural language processing, a new deep learning paradigm named Vision-Language Model Pre-training and Zero-shot Prediction has appeared.
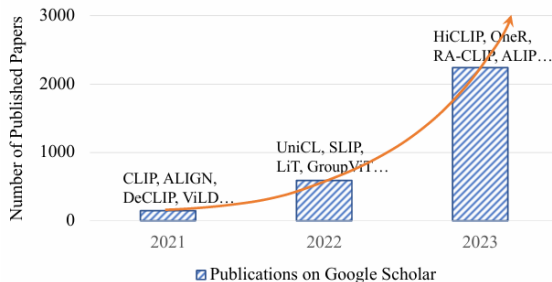

Figure 1: Number of publications on visual recognition VLMs (from Google Scholar)