

Vision transformers

Dosovitskiy, A., et al. (2020). "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv:2010.11929.

Touvron, H., et al. (2021). "Training data-efficient image transformers & distillation through attention." arXiv:2012.12877.

Chen, R. T. Q., et al. (2021). "When Vision Transformers Outperform ResNets without Pre-training or Strong Data Augmentations." arXiv:2106.01548.

(--> arXiv platform)

LLM

Kaplan, J., et al. (2020). "Scaling Laws for Neural Language Models." arXiv:2001.08361.

Houlsby, N., et al. (2019). "Parameter-Efficient Transfer Learning for NLP." arXiv:1902.00751.

Bender, E. M., et al. (2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.

(--> arXiv, NerilPS platforms)

MLLM

Tsai, Y.-H. H., et al. (2019). "Multimodal Transformer for Unaligned Multimodal Language Sequences." arXiv:1906.00295.

"ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks." arXiv:1908.02265.

Radford, A., et al. (2021). "Learning Transferable Visual Models From Natural Language Supervision." arXiv:2103.00020.

(--> arXiv, CVPR, ICCV platforms)