LLaVA-1.5

LLaVA , LLaVA-1.5 Technical Report

Overview

Developer: University of Wisconsin-Madison, Microsoft Research

Release Date: May 2024 (v2)

Type: An improved, data-efficient baseline for LLMs focused on visual instruction tuning.

Architecture

Language Model (LLM): Vicuna-v1.5 (7B and 13B variants)

Vision Encoder: CLIP ViT-L/336px

Training Data & Methodology

Total Instruction Tuning Data: 665k samples

Data: A mixture of LLava-Instruct, ShareGPT, VQAv2, GQA, TextCaps, Visual Genome, and RefCOCO datasets

Training Method: A two-stage protocol: vision-language alignment pretraining and visual instruction tuning