

# TinyLLaMA, TinyLLaVA, TinyLLaVA Factory

余方國 博士

fangkuoyu@gmail.com

# TinyLlama, TinyLLaVA, TinyLLaVA Factory

## TinyLlama-1.1B

English | [中文](#)

[Chat Demo](#) | [Discord](#)

The TinyLlama project aims to pretrain a 1.1B Llama model on 3 trillion tokens. With some proper optimization, we can achieve this within a span of "just" 90 days using 16 A100-40G GPUs 🚀. The training has started on 2023-09-01.

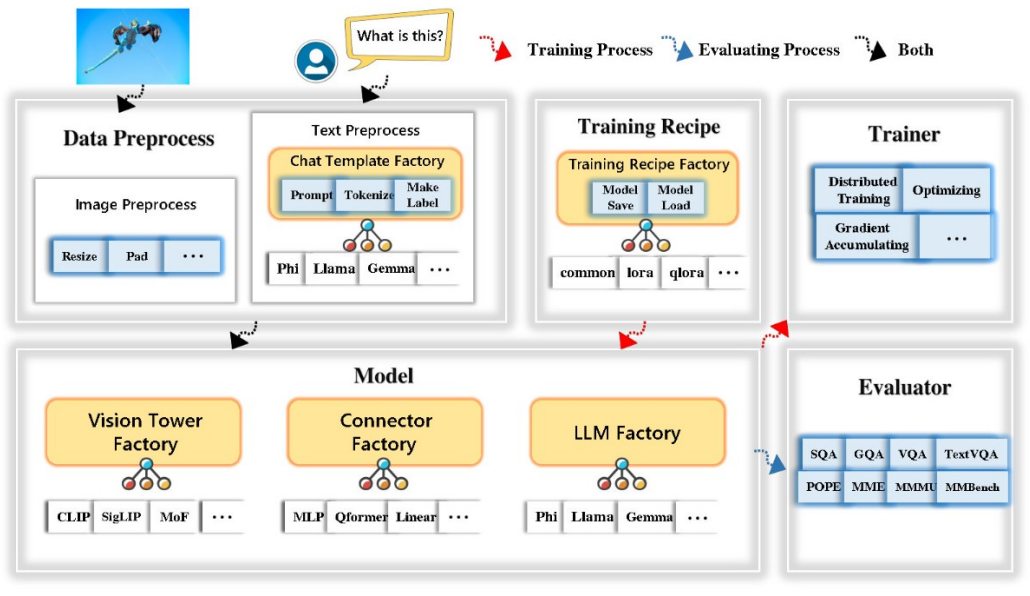


## LLM

<https://github.com/jzhang38/TinyLlama>  
[https://github.com/TinyLLaVA/TinyLLaVA\\_Factory](https://github.com/TinyLLaVA/TinyLLaVA_Factory)

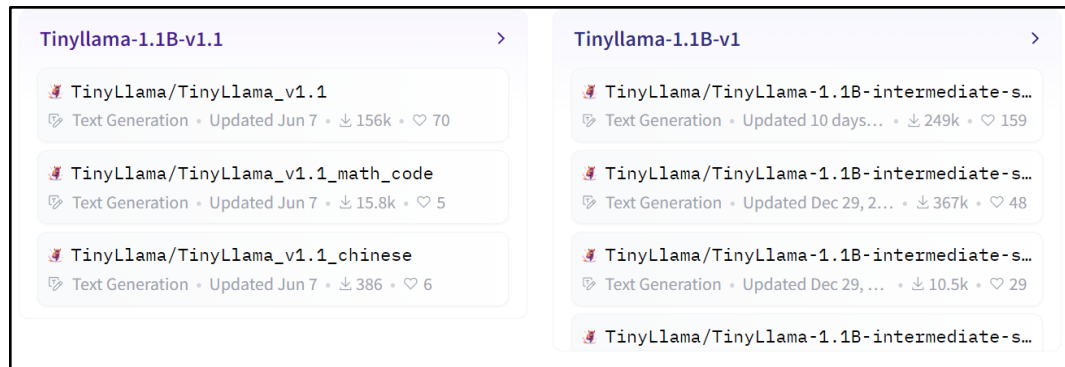
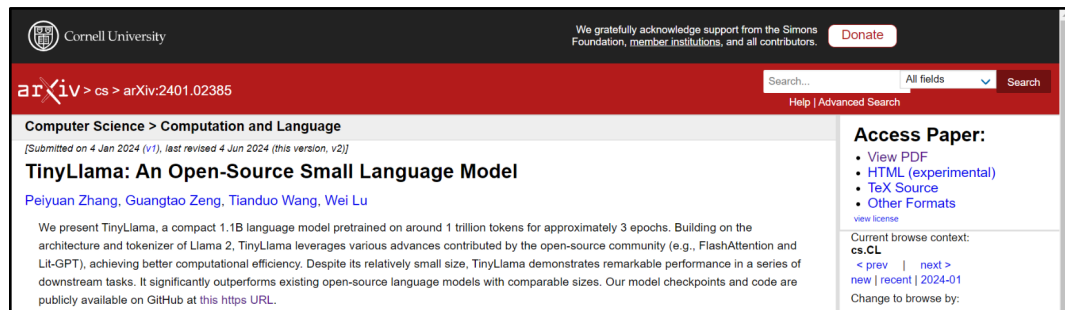
## TinyLLaVA Factory

[Open In HF](#) [Arxiv 2402.14289](#) [Arxiv 2405.11788](#) [License Apache 2.0](#) [Doc Document](#) [Demo Demo](#)



## VLMs

# TinyLlama



## Overview

- <https://arxiv.org/abs/2401.02385>
- <https://github.com/jzhang38/TinyLlama>
- <https://huggingface.co/TinyLlama>


## TRL SFTTrainer

- Fine-Tune Your Own Tiny-Llama on Custom Dataset ([YouTube](#))
- TinyLlama Colorist - fine-tuned with Color dataset ([Medium](#))

## PPO/DPO

- ?

# TinyLLaVA / TinyLLaVA Factory

 Cornell University

We gratefully acknowledge support from the Simons Foundation, member institutions, and all contributors. [Donate](#)

arXiv > cs > arXiv:2402.14289

Search... All fields Search

Help | Advanced Search

Computer Science > Machine Learning

[Submitted on 22 Feb 2024]

**TinyLLaVA: A Framework of Small-scale Large Multimodal Models**

Baichuan Zhou, Ying Hu, Xi Weng, Junlong Jia, Jie Luo, Xien Liu, Ji Wu, Lei Huang

We present the TinyLLaVA framework that provides a unified perspective in designing and analyzing the small-scale Large Multimodal Models (LMMs). We empirically study the effects of different vision encoders, connection modules, language models, training data and training recipes. Our extensive experiments showed that better quality of data combined with better training recipes, smaller LMMs can consistently achieve on-par performances compared to bigger LMMs. Under our framework, we train a family of small-scale LMMs. Our best model, TinyLLaVA-3.1B, achieves better overall performance against existing 7B models such as LLaVA-1.5 and Qwen-VL. We hope our findings can serve as baselines for future research in terms of data scaling, training setups and model selections. Our model weights and codes will be made public.

Comments: Our model weights and codes will be made public at [this https URL](https URL)

Subjects: **Machine Learning [cs.LG]**; Computation and Language (cs.CL)

Cite as: arXiv:2402.14289 [cs.LG]  
(or arXiv:2402.14289v1 [cs.LG] for this version)  
<https://doi.org/10.48550/arXiv.2402.14289>

**Submission history**  
From: Xi Weng [\[view email\]](#)  
[v1] Thu, 22 Feb 2024 05:05:30 UTC (2,308 KB)

Bibliographic Tools


Code, Data, Media

Demos

Related Papers

About arXivLabs

Bibliographic and Citation Tools

 Cornell University

We gratefully acknowledge support from the Simons Foundation, member institutions, and all contributors. [Donate](#)

arXiv > cs > arXiv:2405.11788

Search... All fields Search

Help | Advanced Search

Computer Science > Machine Learning

[Submitted on 20 May 2024]

**TinyLLaVA Factory: A Modularized Codebase for Small-scale Large Multimodal Models**

Junlong Jia, Ying Hu, Xi Weng, Yiming Shi, Miao Li, Xingjian Zhang, Baichuan Zhou, Ziyu Liu, Jie Luo, Lei Huang, Ji Wu

We present TinyLLaVA Factory, an open-source modular codebase for small-scale large multimodal models (LMMs) with a focus on simplicity of code implementations, extensibility of new features, and reproducibility of training results. Following the design philosophy of the factory pattern in software engineering, TinyLLaVA Factory modularizes the entire system into interchangeable components, with each component integrating a suite of cutting-edge models and methods, meanwhile leaving room for extensions to more features. In addition to allowing users to customize their own LMMs, TinyLLaVA Factory provides popular training recipes to let users pretrain and finetune their models with less coding effort. Empirical experiments validate the effectiveness of our codebase. The goal of TinyLLaVA Factory is to assist researchers and practitioners in exploring the wide landscape of designing and training small-scale LMMs with affordable computational resources.

Comments: Our codebase is made public at [this https URL](https URL) with documentation available at [this https URL](https URL)

Subjects: **Machine Learning [cs.LG]**

Cite as: arXiv:2405.11788 [cs.LG]  
(or arXiv:2405.11788v1 [cs.LG] for this version)  
<https://doi.org/10.48550/arXiv.2405.11788>

**Submission history**  
From: Lei Huang [\[view email\]](#)  
[v1] Mon, 20 May 2024 05:11:02 UTC (1,372 KB)

Bibliographic Tools

Code, Data, Media

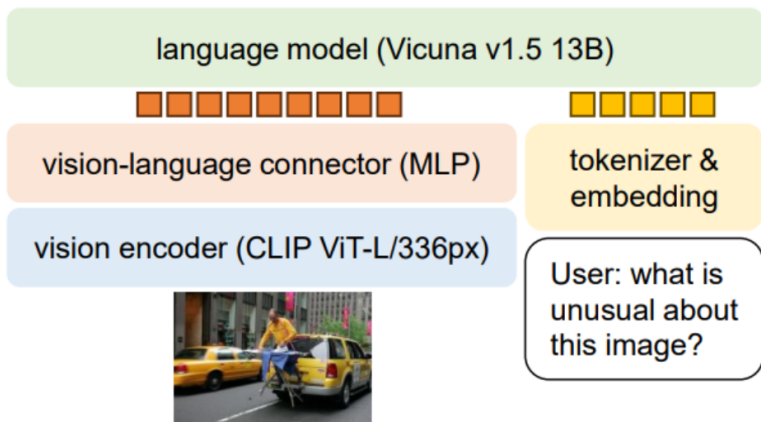
Demos

Related Papers

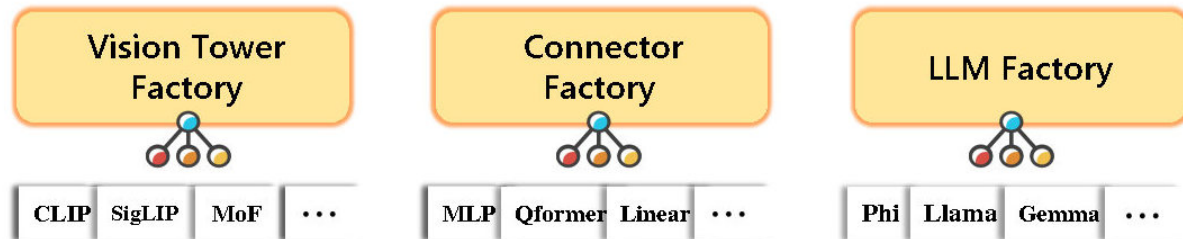
About arXivLabs

# LLaVA / TinyLLaVA Factory

## LLaVA




## TinyLLaVA Factory

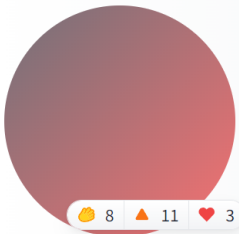


Module	Name	Module	Name
Small-scale LLM	OpenELM-450M (Mehta et al., 2024)	Vision Tower	CLIP (Radford et al., 2021)
	TinyLlama-1.1B (Zhang et al., 2024)		SigLIP (Zhai et al., 2023)
	StableLM-2-1.6B (Jonathan Tow, 2023)		DinoV2 (Oquab et al., 2023)
	Qwen-1.5-1.8B (Jinze Bai, 2024)		MoF (Tong et al., 2024)
	Gemma-2B (Team et al., 2024)		
	Phi-2-2.7B (Mojan Javaheripi, 2023)		
Connector	Identity	Training Recipe	Frozen/Full/Partially Tune LoRA/QLoRA (Hu et al., 2021; Dettmers et al., 2024)
	Linear		
	MLP		
	Q-former (Li et al., 2023)		
	Resampler (Awadalla et al., 2023)		

# TinyLLaVA Models

 **Hugging Face**


[Models](#) [Datasets](#) [Spaces](#) [Posts](#) [Docs](#) [Solutions](#) [Pricing](#) [⌵](#)





8 11 3


**Baichuan Zhou**  
bczhou


Follow




 13 followers · 2 following


 <https://baichuanzhou.github.io/>

 [baichuanzhou](#)

 **AI & ML interests**

Computer Vision

 **Organizations**

 Collections 1

**TinyLLaVA** >

TinyLLaVA: A Framework of Small-scale Large Multimodal Models

● bczhou/TinyLLaVA-3.1B

🔗 Text Generation • Updated Mar 25 • ↓ 921 • ♥ 24


● bczhou/TinyLLaVA-2.0B

🔗 Image-Text-to-Text • Updated Jul 26 • ↓ 158 • ♥ 5

● bczhou/TinyLLaVA-1.5B

🔗 Image-Text-to-Text • Updated Jun 14 • ↓ 404 • ♥ 16

● bczhou/tiny-llava-v1-hf

 Papers 2

**A Comprehensive Benchmark for Evaluating Large Multimodal in Multi-View Urban Scenarios**

Baichuan Zhou<sup>1</sup>, Haotie Yang<sup>1</sup>, Daining Chen<sup>1</sup>, Junyan Ye<sup>1,2</sup>, Tianyi Dai<sup>1</sup>, Jihua Ye<sup>1</sup>, Xiangyang Zhang<sup>1</sup>, Dehua Liu<sup>1</sup>, Conghui He<sup>1,2</sup>, Weiguo Li<sup>1</sup>

<sup>1</sup>Huawei AI Lab (Shanghai), <sup>2</sup>Shanghai Jiao Tong University

<sup>1</sup>Shanghai Jiao Tong University, <sup>2</sup>Shanghai Jiao Tong University

[baichuan@sjtu.edu.cn, haotie@sjtu.edu.cn, daining@sjtu.edu.cn, junyan@sjtu.edu.cn, tianyi@sjtu.edu.cn, jihua@sjtu.edu.cn, xiangyang@sjtu.edu.cn, dehua@sjtu.edu.cn, conghui@sjtu.edu.cn, weiguo@sjtu.edu.cn]

Abstract

State-of-the-art Large Multimodal Models (LMMs) have demonstrated their capabilities in various domains, with only a few specifically focusing on urban environments.

Figure 1: Overview of the benchmark.

**LLaVA: A Framework of Small-scale Large Multimodal Models**

baichuan@sjtu.edu.cn, haotie@sjtu.edu.cn, daining@sjtu.edu.cn, junyan@sjtu.edu.cn, tianyi@sjtu.edu.cn, jihua@sjtu.edu.cn, xiangyang@sjtu.edu.cn, dehua@sjtu.edu.cn, conghui@sjtu.edu.cn, weiguo@sjtu.edu.cn

Abstract

Large Language Models (LLMs) have demonstrated their capabilities in various domains, with only a few specifically focusing on urban environments.

Figure 2: Overview of the framework.

# TinyLLaVA Factory : Finetuning & Customize

📖 README   📄 Apache-2.0 license   ✎   ☰

## Contents

- 📰 [News](#)
- 🔑 [Takeaways](#)
- [Contents](#)
- [Installation and Requirements](#)
  - [Upgrade to the latest code base](#)
- [Get Started](#)
  - [1. Data Preparation](#)
  - [2. Train](#)
  - [3. Evaluation](#)
- [Model Zoo](#)
  - [Trained Models](#)
    - [Model Performance](#)
  - [Legacy Models](#)
- [Launch Demo Locally](#)
  - [Gradio Web Demo](#)
  - [CLI Inference](#)
  - [Quick Inference Scripts](#)
- [Custom Finetune](#)
- [Customize Your Own Large Multimodal Models](#)
  - [LLM](#)
  - [Vision Tower](#)
  - [Connector](#)
- [Acknowledgement](#)
- [Contact](#)
- 📄 [Citation](#)
- ❤️ [Community efforts](#)

Supervised Finetuning  
No PPO/DPO

Customize Components for New VLM

# TinyLLaVA Factory : Finetuning & Customize

📖 README   📄 Apache-2.0 license

## Contents

- 📰 News
- 🔑 Takeaways
- Contents
- Installation and Requirements
  - [Upgrade to the latest code base](#)
- Get Started
  - [1. Data Preparation](#)
  - [2. Train](#)
  - [3. Evaluation](#)
- Model Zoo
  - [Trained Models](#)
    - [Model Performance](#)
  - [Legacy Models](#)
- [Launch Demo Locally](#)
  - [Gradio Web Demo](#)
  - [CLI Inference](#)
  - [Quick Inference Scripts](#)
- [Custom Finetune](#)
- [Customize Your Own Large Multimodal Models](#)
  - [LLM](#)
  - [Vision Tower](#)
  - [Connector](#)
- [Acknowledgement](#)
- [Contact](#)
- 📄 Citation
- ❤️ Community efforts

**Warning: Linux not Windows**

Supervised Finetuning  
No PPO/DPO

Customize Components for New VLM



# TinyLLaMA, TinyLLaVA, TinyLLaVA Factory

余方國 博士

fangkuoyu@gmail.com