EgoCity Inference Toolkit

Multi-person image reasoning & urban activity analysis based on LLaVA-NeXT

Author: Xiamengwei Zhang
Affiliation: Capital Normal University, Dept. of Environmental Design
Date: April 2025
Contact: xiamengweizhang@gmail.com

📘 Overview

This repository provides inference-only tools built on top of LLaVA-NeXT, optimized for analyzing egocentric street-level video data. It integrates multi-person detection, identity deduplication, detailed behavior captioning, timestamp extraction, and diversity/utilization scoring.

Key Modules:

- 🧠 **Visual-Language Reasoning** via LLaVA (Qwen2-7B)
- 🎯 **Object Detection** with DETA (Swin-L)
- 🔁 **CLIP-based Deduplication**
- 🕒 **Timestamp & GPS Extraction** using OCR+GPT or VLM
- 📊 **Grid-level Statistics** for human type and activity diversity
- 🖼️ **Gradio UI** for interactive demonstration

📁 Project Structure

This repository is an extension of llava-next. To use this codebase:

1. First clone the original llava-next repo:

```
git clone https://github.com/llava-vl/llava-next.git
cd llava-next
```

2. Then overlay the following files and folders from egocity-inference/.

```
egocity-inference/
├── [Starting from llava-next]
├── llava/                        # (Unmodified or lightly modified)
├── ...
├── [End of llava-next files]
├── gradio_demo.py               # Gradio demo interface
├── run_model.py                 # Batch inference script
├── run_llava.py                 # Timestamp extraction (via LLaVA)
├── get_timestamp.py            # Timestamp extraction using OCR + GPT
├── get_timestamp.sh            # Shell script for batch timestamp extraction
```

```
├── tools/                        # Annotation and analysis tools
|   ├── app_bbox_1.py            # GUI tool for bounding box annotation
|   ├── import_cv2.py            # Extract frames from video
|   ├── rename_image.py          # Rename image frames
|   ├── import_pd.py             # Count people and activity types
|   ├── hill_diversity.py        # Calculate Hill diversity metrics
|   └── hill_utilization.py      # Compute utilization scores
```

⚙️ Setup

1. Environment Setup

Install the required dependencies:

```
follow the installation step from llava-next
pip install openai easyocr gradio
```

Make sure you have:

- Python 3.8+
- CUDA-enabled GPU for optimal performance
- Git LFS (for large checkpoint files)

---

2. Model Checkpoints

Download the following model weights:

- **LLaVA Qwen2-7B-OV** from Hugging Face or official LLaVA repository
- **DETA (Swin-Large)** from [jozhang97/deta-swin-large]

Place all checkpoints under a `checkpoints/` folder in the project root.

---

3. Directory Structure

Ensure the following directory structure after setup:

```
egocity-inference/
├── checkpoints/
|   ├── llava-qwen2-7b/
|   └── deta-swin-large/
├── tools/
├── *.py
├── test_datasets/
```

🚀 Inference Pipeline

1. Launch Gradio Interface

```
python gradio_demo.py
```

Upload an image → Detect all unique people → Annotate and describe them left to right.

---

2. Run Batch Inference

```
python run_model.py
```

Adjust input/output paths in the script.

---

3. Timestamp Extraction

Via OCR + GPT

```
python get_timestamp.py zhongguancun
```

Via LLaVA

```
python run_llava.py
```

🧰 Utility Tools

Supporting scripts for pre-processing and post-inference analysis.

`tools/app_bbox_1.py`

- GUI for manual annotation
- Outputs structured JSON with street type, facilities, and multi-point descriptions

`tools/import_cv2.py`

- Extracts 3 frames every 10 seconds from video
- Configurable interval and frame count

`tools/rename_image.py`

- Rename extracted frames to `0000.png`, `0001.png`, etc.

`tools/import_pd.py`

- Parses CSV/Excel files to count adults, children, elderly, and unique activity types

`tools/hill_diversity.py`

- Extracts count of 6 activity types for each image/grid

`tools/hill_utilization.py`

- Computes:
  - Shannon Diversity
  - Hill Number
  - Final Utilization Score: `person_count × diversity`

🤝 Acknowledgements

This project is built upon:

- LLaVA-NeXT
- jozhang97/deta-swin-large
- OpenAI GPT API

📜 License & Credits

- **Author:** Xiamengwei Zhang
- **Email:** xiamengweizhang@gmail.com
- **Use case:** Research-only
- **Model weights:** Refer to original repositories for licensing and usage rights