

Run VLA-Cache on the LIBERO Spatial Benchmark

Relevant Files

Evaluation

- `vla_cache_scripts`: VLA-Cache eval scripts
 - `run_of_t_with_vla_cache.sh`: VLA-Cache eval script
 - `run_of_t_without_vla_cache.sh`: VLA-Cache eval utils

VLA-Cache Implementations

- `vla_cache_scripts/src`: VLA-Cache source codes
 - `cache_utils.py`: DynamicCache for VLA-Cache
 - `modeling_llama.py`: Modeling of Transformers for VLA-Cache
 - `modeling_prismatic.py`: Huggingface module for VLA-Cache

VLA-Cache Inference Examples:

- `vla_cache_scripts/log`: VLA-Cache inference log examples (1/50 - Subset of Spatial benchmark)
 - `output_example_vla_cache_inference.log`: With VLA-Cache
 - `output_example_no_vla_cache.log`: Without VLA-Cache
- `rollouts/2025_05_23`: corresponding VLA-Cache inference example rollouts
 - `rollouts/2025_05_23/primary--2025_05_23-16_59_11--episode=*`: With VLA-Cache (Primary view)
 - `rollouts/2025_05_23/wrist--2025_05_23-16_59_11--episode=*`: With VLA-Cache (Wrist view)
 - `rollouts/2025_05_23/primary--2025_05_23-17_05_23--episode=*`: No VLA-Cache (Primary view)
 - `rollouts/2025_05_23/wrist--2025_05_23-17_05_23--episode=*`: No VLA-Cache (Wrist view)

Setup

Set up a conda environment with LIBERO environment(follow instructions of OpenVLA-OFT in [SETUP.md](#) and [LIBERO.md](#)).

Install transformers 4.47.0:

```
pip install transformers==4.47.0
```

Update transformers source code:

```
cp vla_cache_scripts/src/cache_utils.py ~/miniconda3/envs/opencvla-of-t/lib/python3.10/site-packages/transformers/cache_utils.py
cp vla_cache_scripts/src/modeling_llama.py ~/miniconda3/envs/opencvla-of-t/lib/python3.10/site-packages/transformers/models/llama/modeling_llama.py
```

Update HuggingFace modules:

```
cp vla_cache_scripts/src/modeling_prismatic.py ~/.cache/huggingface/hub/models--moojink--opencvla-7b-of-t-finetuned-libero-spatial/src/modeling_prismatic.py
```

VLA-Cache Evaluations Example

Run LIBERO-Spatial benchmark with VLA-Cache inference mode:

```
# Launch LIBERO-Spatial evals with VLA-Cache
python experiments/robot/libero/run_libero_eval.py \
  --pretrained_checkpoint moojink/opencvla-7b-of-t-finetuned-libero-spatial \
  --task_suite_name libero_spatial \
  --use_vla_cache True \
```

- output log example: [vla_cache_scripts/log/output_example_vla_cache_inference.log](#)

Run LIBERO-Spatial benchmark without VLA-Cache:

```
# Launch LIBERO-Spatial evals without VLA-Cache
python experiments/robot/libero/run_libero_eval.py \
  --pretrained_checkpoint moojink/opencvla-7b-of-t-finetuned-libero-spatial \
  --task_suite_name libero_spatial \
  --use_vla_cache False \
```

- output log example: [vla_cache_scripts/log/output_example_no_vla_cache.log](#)