Qwen2.5 Run

Working directory: /gpfs/home1/tgok/test/Qwen

PyTorch: 2.1.2
Transformers: 4.49.0
NumPy: 1.25.1

Model: Qwen/Qwen2.5-VL-7B-Instruct

##

CATEGORY: OBJECT_DETECTION

##

⚠ Cache disabled (use_cache=False). Extracting new...

Processing 1323 samples in batches of 40
Searching in 2 image directories
LOADING MODEL: qwen2_5
Device: cuda
✓ Model loaded successfully

==
✓ Successfully processed: 1140/1323

✗ Skipped (missing/error): 183/1323

First 10 skipped: 000000262227.jpg, 000000262440.jpg, 000000262440.jpg, 000000262682.jpg, 000000262682.jpg, 000000262682.jpg, 000000139684.jpg, 000000000632.jpg, 000000000632.jpg, 000000000632.jpg...

Extracted shapes:
Positive: (1140, 3584)
Negative: (1140, 3584)
Labels: (1140,)

Cached to: hidden_states_cache_final/cache_object_detection_1323_qwen2_5.npz

==
TRAINING CCS PROBE
==

Dataset split:
Train: 797 samples (403 pos, 394 neg)
Test: 343 samples (173 pos, 170 neg)
Hidden dim: 3584

==
TRAINING WITH MULTIPLE RANDOM RESTARTS
==
Trial 1/10: Loss = 0.000699
Trial 2/10: Loss = 0.000698
Trial 3/10: Loss = 0.000645
Trial 4/10: Loss = 0.000769
Trial 5/10: Loss = 0.000729
Trial 6/10: Loss = 0.000767
Trial 7/10: Loss = 0.000664
Trial 8/10: Loss = 0.000658
Trial 9/10: Loss = 0.000670
Trial 10/10: Loss = 0.000438
EVALUATION WITH BEST PROBE
Best loss: 0.000438

Test Results:
Overall Accuracy: 84.8% (291/343)
Positive samples: 86.1% (173 samples)
Negative samples: 83.5% (170 samples)

✓ COMPLETE: object_detection → 84.8%

##

CATEGORY: ATTRIBUTE_RECOGNITION

##

==
EXTRACTING HIDDEN STATES: ATTRIBUTE_RECOGNITION
==
⚠ Cache disabled (use_cache=False). Extracting new...

Processing 3410 samples in batches of 40
Searching in 2 image directories
LOADING MODEL: qwen2_5
Device: cuda
✓ Model loaded successfully

==
✓ Successfully processed: 3002/3410
✗ Skipped (missing/error): 408/3410

First 10 skipped: 000000393282.jpg, 000000393282.jpg, 000000393282.jpg,
000000393469.jpg, 000000000285.jpg, 000000262440.jpg, 000000262440.jpg,
000000262440.jpg, 000000262440.jpg, 000000131386.jpg...

Extracted shapes:
Positive: (3002, 3584)
Negative: (3002, 3584)
Labels: (3002,)

Cached to: hidden_states_cache_final/cache_attribute_recognition_3410_qwen2_5.npz

==
TRAINING CCS PROBE
==

Dataset split:
Train: 2101 samples (1062 pos, 1039 neg)
Test: 901 samples (456 pos, 445 neg)
Hidden dim: 3584

TRAINING WITH MULTIPLE RANDOM RESTARTS

==

Trial 1/10: Loss = 0.004829

Trial 2/10: Loss = 0.005072

Trial 3/10: Loss = 0.002456

Trial 4/10: Loss = 0.005236

Trial 5/10: Loss = 0.004712

Trial 6/10: Loss = 0.003517

Trial 7/10: Loss = 0.002941

Trial 8/10: Loss = 0.004980

Trial 9/10: Loss = 0.005689

Trial 10/10: Loss = 0.005867

EVALUATION WITH BEST PROBE

Best loss: 0.002456

Test Results:

Overall Accuracy: 85.5% (770/901)

Positive samples: 83.1% (456 samples)

Negative samples: 87.9% (445 samples)

✓ COMPLETE: attribute_recognition → 85.5%

##

CATEGORY: SPATIAL_RECOGNITION

##

==

EXTRACTING HIDDEN STATES: SPATIAL_RECOGNITION

==

⚠ Cache disabled (use_cache=False). Extracting new...

Processing 1030 samples in batches of 40

Searching in 2 image directories

LOADING MODEL: qwen2_5

Device: cuda

✓ Model loaded successfully

==

✓ Successfully processed: 880/1030

✗ Skipped (missing/error): 150/1030

First 10 skipped: 000000393282.jpg, 000000000285.jpg, 000000262682.jpg, 000000000632.jpg, 000000262895.jpg, 000000043816.jpg, 000000043816.jpg, 000000043816.jpg, 000000043816.jpg, 000000000785.jpg...

Extracted shapes:
Positive: (880, 3584)
Negative: (880, 3584)
Labels: (880,)

Cached to: hidden_states_cache_final/cache_spatial_recognition_1030_qwen2_5.npz

==
TRAINING CCS PROBE
==

Dataset split:
Train: 615 samples (294 pos, 321 neg)
Test: 265 samples (126 pos, 139 neg)
Hidden dim: 3584

==
TRAINING WITH MULTIPLE RANDOM RESTARTS
==
Trial 1/10: Loss = 0.001303
Trial 2/10: Loss = 0.001452
Trial 3/10: Loss = 0.001069
Trial 4/10: Loss = 0.001542
Trial 5/10: Loss = 0.001071
Trial 6/10: Loss = 0.001612
Trial 7/10: Loss = 0.001348
Trial 8/10: Loss = 0.000917
Trial 9/10: Loss = 0.001572
Trial 10/10: Loss = 0.001419
EVALUATION WITH BEST PROBE
Best loss: 0.000917

Test Results:
Overall Accuracy: 79.6% (211/265)
Positive samples: 72.2% (126 samples)
Negative samples: 86.3% (139 samples)

✓ COMPLETE: spatial_recognition → 79.6%

```
================================================================
```

Final Results:

object_detection : 84.8%

attribute_recognition : 85.5%

spatial_recognition : 79.6%

Average : 83.3%

```
================================================================
```

= Job finished at Fri Oct 24 20:24:33 CEST 2025 with exit code: 0 =