

=== GPU Information ===

NVIDIA A100-SXM4-40GB, 40960 MiB, 580.95.05

NVIDIA A100-SXM4-40GB, 40960 MiB, 580.95.05

NVIDIA A100-SXM4-40GB, 40960 MiB, 580.95.05

NVIDIA A100-SXM4-40GB, 40960 MiB, 580.95.05

=== Checking if vision_ccs_corrected.py exists ===

-rw-r----- 1 mdemirev mdemirev 19K Oct 23 12:45 vision_ccs_corrected.py

=== Running vision_ccs_corrected.py ===

/home/mdemirev/.local/lib/python3.11/site-packages/huggingface_hub/file_download.py:945:

FutureWarning: `resume_download` is deprecated and will be removed in version 1.0.0. Downloads always resume when possible. If you want to force a new download, use `force_download=True`.

warnings.warn(

Special tokens have been added in the vocabulary, make sure the associated word embeddings are fine-tuned or trained.

Model: llava-hf/llava-1.5-7b-hf

#####

CATEGORY: OBJECT_DETECTION

#####

=====

EXTRACTING HIDDEN STATES: OBJECT_DETECTION

=====

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

Processing 1323 samples in batches of 40

Searching in 2 image directories

LOADING MODEL: llava

Device: cuda

Loading checkpoint shards: 0% | 0/3 [00:00<?, ?it/s]

Loading checkpoint shards: 33% | 1/3 [00:02<00:05, 2.60s/it]

Loading checkpoint shards: 67% | 2/3 [00:05<00:02, 2.50s/it]

Loading checkpoint shards: 100% | 3/3 [00:07<00:00, 2.30s/it]

Loading checkpoint shards: 100% | 3/3 [00:07<00:00, 2.36s/it]

✓ Model loaded successfully

Batches: 0% | 0/34 [00:00<?, ?it/s]

Batches: 3% | 1/34 [00:13<07:30, 13.64s/it]

Batches: 6% | 2/34 [00:26<07:04, 13.27s/it]

Batches: 9% | 3/34 [00:37<06:22, 12.32s/it]

Batches: 12% | 4/34 [00:49<06:00, 12.00s/it]

Batches: 15% | 5/34 [01:00<05:39, 11.72s/it]

Batches: 18% | 6/34 [01:13<05:36, 12.00s/it]

Batches: 21% | 7/34 [01:25<05:23, 11.99s/it]

Batches: 24% | 8/34 [01:35<05:02, 11.64s/it]

Batches: 26% | 9/34 [01:47<04:50, 11.62s/it]

Batches: 29% | 10/34 [01:59<04:43, 11.82s/it]

Batches: 32% | 11/34 [02:11<04:32, 11.84s/it]

Batches: 35% | 12/34 [02:23<04:19, 11.78s/it]

Batches: 38% | 13/34 [02:32<03:51, 11.03s/it]

Batches: 41% | 14/34 [02:44<03:48, 11.41s/it]

Batches: 44% | 15/34 [02:55<03:32, 11.17s/it]

Batches: 47% | 16/34 [03:08<03:28, 11.60s/it]

Batches: 50% | 17/34 [03:19<03:13, 11.39s/it]

Batches: 53% | 18/34 [03:31<03:07, 11.69s/it]

Batches: 56% | 19/34 [03:44<03:03, 12.22s/it]

Batches: 59% | 20/34 [03:56<02:48, 12.05s/it]

Batches: 62% | 21/34 [04:08<02:36, 12.04s/it]

Batches: 65% | 22/34 [04:19<02:21, 11.80s/it]

Batches: 68% | 23/34 [04:31<02:10, 11.84s/it]

Batches: 71% ██████████ | 24/34 [04:45<02:02, 12.29s/it]
Batches: 74% ██████████ | 25/34 [04:57<01:50, 12.29s/it]
Batches: 76% ██████████ | 26/34 [05:08<01:35, 12.00s/it]
Batches: 79% ██████████ | 27/34 [05:21<01:25, 12.19s/it]
Batches: 82% ██████████ | 28/34 [05:31<01:10, 11.70s/it]
Batches: 85% ██████████ | 29/34 [05:43<00:58, 11.75s/it]
Batches: 88% ██████████ | 30/34 [05:54<00:46, 11.59s/it]
Batches: 91% ██████████ | 31/34 [06:06<00:34, 11.57s/it]
Batches: 94% ██████████ | 32/34 [06:17<00:22, 11.28s/it]
Batches: 97% ██████████ | 33/34 [06:28<00:11, 11.26s/it]
Batches: 100% ██████████ | 34/34 [06:29<00:00, 8.22s/it]
Batches: 100% ██████████ | 34/34 [06:29<00:00, 11.45s/it]

/home/mdemirev/.local/lib/python3.11/site-packages/huggingface_hub/file_download.py:945:

FutureWarning: `resume_download` is deprecated and will be removed in version 1.0.0. Downloads always resume when possible. If you want to force a new download, use `force_download=True`.

warnings.warn(
Special tokens have been added in the vocabulary, make sure the associated word embeddings are fine-tuned or trained.

=====

✓ Successfully processed: 1140/1323

X 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, 4096)

Negative: (1140, 4096)

Labels: (1140,)

Cached to: hidden_states_cache/cache_object_detection_1323_llava.npz

=====

TRAINING CCS PROBE

=====

Dataset split:

Train: 797 samples (403 pos, 394 neg)

Test: 343 samples (173 pos, 170 neg)

Hidden dim: 4096

=====

TRAINING WITH MULTIPLE RANDOM RESTARTS

=====

Trial 1/10: Loss = 0.012694

✓ New best probe found!

Trial 2/10: Loss = 0.012562

✓ New best probe found!

Trial 3/10: Loss = 0.016379

Trial 4/10: Loss = 0.015903

Trial 5/10: Loss = 0.013277

Trial 6/10: Loss = 0.015818

Trial 7/10: Loss = 0.012685

Trial 8/10: Loss = 0.015946

Trial 9/10: Loss = 0.012418

✓ New best probe found!

Trial 10/10: Loss = 0.015791

=====

EVALUATION WITH BEST PROBE

=====

Best loss: 0.012418

Test Results:

Overall Accuracy: 51.9% (178/343)
Positive samples: 59.5% (173 samples)
Negative samples: 44.1% (170 samples)

✓ COMPLETE: object_detection → 51.9%

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: llava
Device: cuda

Loading checkpoint shards: 0%| | 0/3 [00:00<?, ?it/s]
Loading checkpoint shards: 33%|██| 1/3 [00:02<00:05, 2.63s/it]
Loading checkpoint shards: 67%|██████| 2/3 [00:05<00:02, 2.53s/it]
Loading checkpoint shards: 100%|██████████| 3/3 [00:07<00:00, 2.32s/it]
Loading checkpoint shards: 100%|██████████| 3/3 [00:07<00:00, 2.39s/it]

✓ Model loaded successfully

Batches: 0%| | 0/86 [00:00<?, ?it/s]
Batches: 1%| | 1/86 [00:12<17:47, 12.56s/it]
Batches: 2%|| | 2/86 [00:26<18:32, 13.24s/it]
Batches: 3%|| | 3/86 [00:39<18:10, 13.14s/it]
Batches: 5%|| | 4/86 [00:50<17:08, 12.54s/it]
Batches: 6%|| | 5/86 [01:04<17:18, 12.82s/it]
Batches: 7%|| | 6/86 [01:15<16:22, 12.28s/it]
Batches: 8%|| | 7/86 [01:28<16:18, 12.39s/it]
Batches: 9%|| | 8/86 [01:36<14:39, 11.27s/it]
Batches: 10%|| | 9/86 [01:48<14:27, 11.27s/it]
Batches: 12%|| | 10/86 [01:58<13:53, 10.97s/it]
Batches: 13%|| | 11/86 [02:10<14:10, 11.35s/it]
Batches: 14%|| | 12/86 [02:21<13:55, 11.29s/it]
Batches: 15%|| | 13/86 [02:32<13:36, 11.19s/it]
Batches: 16%|| | 14/86 [02:44<13:26, 11.21s/it]
Batches: 17%|| | 15/86 [02:56<13:39, 11.54s/it]
Batches: 19%|| | 16/86 [03:09<14:05, 12.07s/it]
Batches: 20%|| | 17/86 [03:22<14:03, 12.23s/it]
Batches: 21%|| | 18/86 [03:32<13:11, 11.64s/it]
Batches: 22%|| | 19/86 [03:44<13:00, 11.64s/it]
Batches: 23%|| | 20/86 [03:56<13:07, 11.93s/it]
Batches: 24%|| | 21/86 [04:09<13:10, 12.15s/it]
Batches: 26%|| | 22/86 [04:22<13:16, 12.45s/it]
Batches: 27%|| | 23/86 [04:35<13:08, 12.52s/it]
Batches: 28%|| | 24/86 [04:46<12:32, 12.14s/it]
Batches: 29%|| | 25/86 [04:59<12:42, 12.50s/it]
Batches: 30%|| | 26/86 [05:12<12:32, 12.54s/it]
Batches: 31%|| | 27/86 [05:23<11:57, 12.17s/it]
Batches: 33%|| | 28/86 [05:34<11:24, 11.80s/it]
Batches: 34%|| | 29/86 [05:47<11:22, 11.97s/it]
Batches: 35%|| | 30/86 [05:58<10:52, 11.66s/it]
Batches: 36%|| | 31/86 [06:10<10:47, 11.78s/it]

Batches: 37%		32/86 [06:23<11:07, 12.36s/it]
Batches: 38%		33/86 [06:35<10:38, 12.05s/it]
Batches: 40%		34/86 [06:45<10:04, 11.62s/it]
Batches: 41%		35/86 [06:59<10:17, 12.12s/it]
Batches: 42%		36/86 [07:12<10:23, 12.48s/it]
Batches: 43%		37/86 [07:24<10:05, 12.36s/it]
Batches: 44%		38/86 [07:36<09:49, 12.27s/it]
Batches: 45%		39/86 [07:47<09:19, 11.89s/it]
Batches: 47%		40/86 [07:59<09:08, 11.92s/it]
Batches: 48%		41/86 [08:11<09:02, 12.06s/it]
Batches: 49%		42/86 [08:22<08:28, 11.55s/it]
Batches: 50%		43/86 [08:34<08:22, 11.68s/it]
Batches: 51%		44/86 [08:45<08:09, 11.67s/it]
Batches: 52%		45/86 [08:57<08:02, 11.77s/it]
Batches: 53%		46/86 [09:09<07:52, 11.82s/it]
Batches: 55%		47/86 [09:22<07:47, 11.98s/it]
Batches: 56%		48/86 [09:35<07:46, 12.28s/it]
Batches: 57%		49/86 [09:46<07:22, 11.97s/it]
Batches: 58%		50/86 [09:58<07:07, 11.87s/it]
Batches: 59%		51/86 [10:10<07:03, 12.09s/it]
Batches: 60%		52/86 [10:21<06:42, 11.85s/it]
Batches: 62%		53/86 [10:34<06:42, 12.20s/it]
Batches: 63%		54/86 [10:47<06:37, 12.43s/it]
Batches: 64%		55/86 [10:59<06:20, 12.28s/it]
Batches: 65%		56/86 [11:11<06:05, 12.19s/it]
Batches: 66%		57/86 [11:25<06:03, 12.52s/it]
Batches: 67%		58/86 [11:38<05:57, 12.75s/it]
Batches: 69%		59/86 [11:52<05:51, 13.02s/it]
Batches: 70%		60/86 [12:03<05:27, 12.61s/it]
Batches: 71%		61/86 [12:16<05:17, 12.69s/it]
Batches: 72%		62/86 [12:28<05:01, 12.57s/it]
Batches: 73%		63/86 [12:40<04:40, 12.20s/it]
Batches: 74%		64/86 [12:53<04:33, 12.41s/it]
Batches: 76%		65/86 [13:04<04:15, 12.17s/it]
Batches: 77%		66/86 [13:17<04:07, 12.39s/it]
Batches: 78%		67/86 [13:28<03:47, 11.97s/it]
Batches: 79%		68/86 [13:38<03:24, 11.35s/it]
Batches: 80%		69/86 [13:49<03:12, 11.31s/it]
Batches: 81%		70/86 [14:02<03:05, 11.60s/it]
Batches: 83%		71/86 [14:14<02:57, 11.81s/it]
Batches: 84%		72/86 [14:25<02:44, 11.75s/it]
Batches: 85%		73/86 [14:36<02:29, 11.54s/it]
Batches: 86%		74/86 [14:50<02:23, 11.98s/it]
Batches: 87%		75/86 [15:02<02:11, 11.99s/it]
Batches: 88%		76/86 [15:14<02:01, 12.12s/it]
Batches: 90%		77/86 [15:25<01:47, 11.90s/it]
Batches: 91%		78/86 [15:38<01:37, 12.24s/it]
Batches: 92%		79/86 [15:49<01:23, 11.86s/it]
Batches: 93%		80/86 [16:01<01:11, 11.92s/it]
Batches: 94%		81/86 [16:14<01:00, 12.14s/it]
Batches: 95%		82/86 [16:25<00:46, 11.70s/it]
Batches: 97%		83/86 [16:38<00:36, 12.21s/it]
Batches: 98%		84/86 [16:50<00:24, 12.01s/it]
Batches: 99%		85/86 [17:03<00:12, 12.31s/it]
Batches: 100%		86/86 [17:06<00:00, 9.66s/it]
Batches: 100%		86/86 [17:06<00:00, 11.94s/it]

```
/home/mdemirev/.local/lib/python3.11/site-packages/huggingface_hub/file_download.py:945:
FutureWarning: `resume_download` is deprecated and will be removed in version 1.0.0. Downloads
always resume when possible. If you want to force a new download, use `force_download=True`.
  warnings.warn(
Special tokens have been added in the vocabulary, make sure the associated word embeddings are
fine-tuned or trained.
```

```
=====
✓ Successfully processed: 3002/3410
```

```
X 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, 4096)
```

```
Negative: (3002, 4096)
```

```
Labels: (3002,)
```

```
Cached to: hidden_states_cache/cache_attribute_recognition_3410_llava.npz
```

```
=====
TRAINING CCS PROBE
=====
```

```
Dataset split:
```

```
Train: 2101 samples (1062 pos, 1039 neg)
```

```
Test: 901 samples (456 pos, 445 neg)
```

```
Hidden dim: 4096
```

```
=====
TRAINING WITH MULTIPLE RANDOM RESTARTS
=====
```

```
Trial 1/10: Loss = 0.036114
```

```
✓ New best probe found!
```

```
Trial 2/10: Loss = 0.037136
```

```
Trial 3/10: Loss = 0.036104
```

```
✓ New best probe found!
```

```
Trial 4/10: Loss = 0.035442
```

```
✓ New best probe found!
```

```
Trial 5/10: Loss = 0.034705
```

```
✓ New best probe found!
```

```
Trial 6/10: Loss = 0.034712
```

```
Trial 7/10: Loss = 0.034171
```

```
✓ New best probe found!
```

```
Trial 8/10: Loss = 0.037336
```

```
Trial 9/10: Loss = 0.034188
```

```
Trial 10/10: Loss = 0.036584
```

```
=====
EVALUATION WITH BEST PROBE
=====
```

```
Best loss: 0.034171
```

```
Test Results:
```

```
Overall Accuracy: 76.8% (692/901)
```

```
Positive samples: 75.7% (456 samples)
```

```
Negative samples: 78.0% (445 samples)
```

```
✓ COMPLETE: attribute_recognition → 76.8%
```

```
#####
# 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: llava
Device: cuda

Loading checkpoint shards: 0%| | 0/3 [00:00<?, ?it/s]
Loading checkpoint shards: 33%| | 1/3 [00:02<00:05, 2.58s/it]
Loading checkpoint shards: 67%| | 2/3 [00:05<00:02, 2.49s/it]
Loading checkpoint shards: 100%| | 3/3 [00:07<00:00, 2.29s/it]
Loading checkpoint shards: 100%| | 3/3 [00:07<00:00, 2.35s/it]
✓ Model loaded successfully

Batches: 0%| | 0/26 [00:00<?, ?it/s]
Batches: 4%| | 1/26 [00:12<05:21, 12.85s/it]
Batches: 8%| | 2/26 [00:26<05:14, 13.10s/it]
Batches: 12%| | 3/26 [00:35<04:24, 11.50s/it]
Batches: 15%| | 4/26 [00:47<04:17, 11.69s/it]
Batches: 19%| | 5/26 [00:59<04:07, 11.79s/it]
Batches: 23%| | 6/26 [01:11<03:58, 11.93s/it]
Batches: 27%| | 7/26 [01:24<03:49, 12.06s/it]
Batches: 31%| | 8/26 [01:36<03:36, 12.03s/it]
Batches: 35%| | 9/26 [01:46<03:16, 11.55s/it]
Batches: 38%| | 10/26 [01:58<03:06, 11.67s/it]
Batches: 42%| | 11/26 [02:10<02:57, 11.85s/it]
Batches: 46%| | 12/26 [02:22<02:45, 11.80s/it]
Batches: 50%| | 13/26 [02:33<02:30, 11.55s/it]
Batches: 54%| | 14/26 [02:43<02:12, 11.06s/it]
Batches: 58%| | 15/26 [02:54<02:02, 11.15s/it]
Batches: 62%| | 16/26 [03:06<01:51, 11.20s/it]
Batches: 65%| | 17/26 [03:17<01:41, 11.25s/it]
Batches: 69%| | 18/26 [03:28<01:30, 11.29s/it]
Batches: 73%| | 19/26 [03:41<01:21, 11.61s/it]
Batches: 77%| | 20/26 [03:53<01:11, 11.93s/it]
Batches: 81%| | 21/26 [04:04<00:57, 11.55s/it]
Batches: 85%| | 22/26 [04:16<00:47, 11.80s/it]
Batches: 88%| | 23/26 [04:28<00:34, 11.66s/it]
Batches: 92%| | 24/26 [04:40<00:23, 11.86s/it]
Batches: 96%| | 25/26 [04:51<00:11, 11.68s/it]
Batches: 100%| | 26/26 [05:01<00:00, 10.96s/it]
Batches: 100%| | 26/26 [05:01<00:00, 11.58s/it]

=====

✓ Successfully processed: 880/1030
X 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, 4096)
Negative: (880, 4096)
Labels: (880,)

Cached to: hidden_states_cache/cache_spatial_recognition_1030_llava.npz

=====

TRAINING CCS PROBE

Dataset split:

Train: 615 samples (294 pos, 321 neg)
Test: 265 samples (126 pos, 139 neg)
Hidden dim: 4096

TRAINING WITH MULTIPLE RANDOM RESTARTS

Trial 1/10: Loss = 0.021389
✓ New best probe found!
Trial 2/10: Loss = 0.023533
Trial 3/10: Loss = 0.022456
Trial 4/10: Loss = 0.022393
Trial 5/10: Loss = 0.019901
✓ New best probe found!
Trial 6/10: Loss = 0.022371
Trial 7/10: Loss = 0.020647
Trial 8/10: Loss = 0.024282
Trial 9/10: Loss = 0.022482
Trial 10/10: Loss = 0.022303

EVALUATION WITH BEST PROBE

Best loss: 0.019901

Test Results:

Overall Accuracy: 75.8% (201/265)
Positive samples: 81.7% (126 samples)
Negative samples: 70.5% (139 samples)

✓ COMPLETE: spatial_recognition → 75.8%

Final Results:

object_detection	: 51.9%
attribute_recognition	: 76.8%
spatial_recognition	: 75.8%
Average	: 68.2%

=== Job finished at Thu Oct 23 13:29:43 CEST 2025 with exit code: 0 ===