```
=== 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.py exists ===
-rw-r-----. 1 mdemirev mdemirev 20K Oct 22 22:18 vision_ccs.py
=== Running vision_ccs.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(
/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(
The model is automatically converting to bf16 for faster inference. If you want to disable the
automatic precision, please manually add bf16/fp16/fp32=True to
"AutoModelForCausalLM.from_pretrained".

Configuration:
  Model: Qwen/Qwen2-VL-7B-Instruct
  Samples per category:
    - object_detection: 1323
    - attribute_recognition: 3410
    - spatial_recognition: 1030
  Batch size: 40
  Cache enabled: False
  Categories: object_detection, attribute_recognition, spatial_recognition

CCS Training:
  Epochs per trial: 1000
  Random restarts: 10
  Learning rate: 0.001
  Weight decay: 0.01

######################################################################
# CATEGORY: OBJECT_DETECTION
######################################################################
LOADING DATA for category: 'object_detection'
Using 1323 samples from '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: qwen2
Device: cuda

Loading checkpoint shards:   0%|          | 0/10 [00:00<?, ?it/s]
Loading checkpoint shards:  10%|▉         | 1/10 [00:00<00:07,  1.24it/s]
Loading checkpoint shards:  20%|█▊        | 2/10 [00:01<00:06,  1.19it/s]
Loading checkpoint shards:  30%|██▋       | 3/10 [00:02<00:05,  1.24it/s]
Loading checkpoint shards:  40%|███▌      | 4/10 [00:03<00:04,  1.25it/s]
Loading checkpoint shards:  50%|████▌     | 5/10 [00:03<00:03,  1.27it/s]
Loading checkpoint shards:  60%|█████▍    | 6/10 [00:04<00:03,  1.29it/s]
Loading checkpoint shards:  70%|██████▎   | 7/10 [00:05<00:02,  1.29it/s]
Loading checkpoint shards:  80%|███████▏  | 8/10 [00:06<00:01,  1.23it/s]
Loading checkpoint shards:  90%|████████  | 9/10 [00:07<00:00,  1.15it/s]
Loading checkpoint shards: 100%|█████████| 10/10 [00:08<00:00,  1.22it/s]
Loading checkpoint shards: 100%|█████████| 10/10 [00:08<00:00,  1.23it/s]
/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(
```

```
✓ Model loaded successfully


Batches:   0%|          | 0/34 [00:00<?, ?it/s]
Batches:   3%|▎         | 1/34 [00:16<09:02, 16.45s/it]
Batches:   6%|▌         | 2/34 [00:32<08:44, 16.40s/it]
Batches:   9%|▉         | 3/34 [00:46<07:49, 15.14s/it]
Batches:  12%|█▏        | 4/34 [01:02<07:49, 15.66s/it]
Batches:  15%|█▍        | 5/34 [01:16<07:14, 14.97s/it]
Batches:  18%|█▊        | 6/34 [01:32<07:05, 15.18s/it]
Batches:  21%|██        | 7/34 [01:46<06:43, 14.96s/it]
Batches:  24%|██▎       | 8/34 [02:00<06:17, 14.51s/it]
Batches:  26%|██▋       | 9/34 [02:14<05:59, 14.39s/it]
Batches:  29%|██▉       | 10/34 [02:29<05:52, 14.67s/it]
Batches:  32%|███▏      | 11/34 [02:44<05:36, 14.62s/it]
Batches:  35%|███▌      | 12/34 [02:58<05:18, 14.48s/it]
Batches:  38%|███▊      | 13/34 [03:12<04:59, 14.26s/it]
Batches:  41%|████      | 14/34 [03:27<04:49, 14.45s/it]
Batches:  44%|████▍     | 15/34 [03:43<04:43, 14.94s/it]
Batches:  47%|████▋     | 16/34 [03:58<04:31, 15.07s/it]
Batches:  50%|█████     | 17/34 [04:11<04:07, 14.56s/it]
Batches:  53%|█████▎    | 18/34 [04:26<03:54, 14.68s/it]
Batches:  56%|█████▌    | 19/34 [04:43<03:49, 15.29s/it]
Batches:  59%|█████▉    | 20/34 [04:57<03:29, 14.94s/it]
Batches:  62%|██████▏   | 21/34 [05:12<03:13, 14.89s/it]
Batches:  65%|██████▍   | 22/34 [05:26<02:54, 14.51s/it]
Batches:  68%|██████▊   | 23/34 [05:40<02:39, 14.53s/it]
Batches:  71%|███████   | 24/34 [05:56<02:30, 15.04s/it]
Batches:  74%|███████▎  | 25/34 [06:12<02:17, 15.32s/it]
Batches:  76%|███████▋  | 26/34 [06:26<01:58, 14.86s/it]
Batches:  79%|███████▉  | 27/34 [06:43<01:47, 15.37s/it]
Batches:  82%|████████▏ | 28/34 [06:56<01:27, 14.64s/it]
Batches:  85%|████████▍ | 29/34 [07:11<01:14, 14.88s/it]
Batches:  88%|████████▊ | 30/34 [07:25<00:58, 14.53s/it]
Batches:  91%|█████████ | 31/34 [07:39<00:43, 14.46s/it]
Batches:  94%|█████████▍| 32/34 [07:52<00:27, 13.98s/it]
Batches:  97%|█████████▋| 33/34 [08:06<00:13, 13.91s/it]
Batches: 100%|██████████| 34/34 [08:08<00:00, 10.39s/it]
Batches: 100%|██████████| 34/34 [08:08<00:00, 14.36s/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(
The model is automatically converting to bf16 for faster inference. If you want to disable the
automatic precision, please manually add bf16/fp16/fp32=True to
"AutoModelForCausalLM.from_pretrained".

======================================================================
✓ 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, 4096)
  Negative: (1140, 4096)
  Labels: (1140,)

Cached to: hidden_states_cache/cache_object_detection_1323_qwen2.npz
```

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

Dataset split (Stratified):
  Train: 797 samples (403 pos, 394 neg)
  Test:  343 samples (173 pos, 170 neg)
  Hidden dim: 4096

Probe architecture:
  Input: 4096
  Hidden: 256 → 128
  Output: 1 (probability)

Training config:
  Epochs per trial: 1000
  Number of trials: 10
  Learning rate: 0.001
  Weight decay: 0.01

======================================================================
TRAINING WITH MULTIPLE RANDOM RESTARTS
======================================================================
  Trial  1/10: Loss = 0.002176
    ✓ New best probe found!
  Trial  2/10: Loss = 0.001768
    ✓ New best probe found!
  Trial  3/10: Loss = 0.003461
  Trial  4/10: Loss = 0.001651
    ✓ New best probe found!
  Trial  5/10: Loss = 0.001662
  Trial  6/10: Loss = 0.001835
  Trial  7/10: Loss = 0.002167
  Trial  8/10: Loss = 0.001795
  Trial  9/10: Loss = 0.002389
  Trial 10/10: Loss = 0.002075

======================================================================
EVALUATION WITH BEST PROBE
======================================================================
Best loss: 0.001651

Test Results:
  Overall Accuracy: 50.4% (173/343)
  Positive samples: 58.4% (173 samples)
  Negative samples: 42.4% (170 samples)

✓ COMPLETE: object_detection → 50.4%

######################################################################
# CATEGORY: ATTRIBUTE_RECOGNITION
######################################################################
LOADING DATA for category: 'attribute_recognition'
Using 3410 samples from '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
Device: cuda

Loading checkpoint shards:   0%|          | 0/10 [00:00<?, ?it/s]
Loading checkpoint shards:  10%|▉         | 1/10 [00:00<00:07,  1.28it/s]
Loading checkpoint shards:  20%|██        | 2/10 [00:01<00:06,  1.28it/s]
```

```
Loading checkpoint shards:  30%|███       | 3/10 [00:02<00:05,  1.29it/s]
Loading checkpoint shards:  40%|████      | 4/10 [00:03<00:04,  1.28it/s]
Loading checkpoint shards:  50%|█████     | 5/10 [00:03<00:03,  1.27it/s]
Loading checkpoint shards:  60%|██████    | 6/10 [00:04<00:03,  1.28it/s]
Loading checkpoint shards:  70%|███████   | 7/10 [00:05<00:02,  1.28it/s]
Loading checkpoint shards:  80%|████████  | 8/10 [00:06<00:01,  1.22it/s]
Loading checkpoint shards:  90%|█████████ | 9/10 [00:07<00:00,  1.16it/s]
Loading checkpoint shards: 100%|██████████| 10/10 [00:08<00:00,  1.23it/s]
Loading checkpoint shards: 100%|██████████| 10/10 [00:08<00:00,  1.25it/s]
✓ Model loaded successfully


Batches:   0%|          | 0/86 [00:00<?, ?it/s]
Batches:   1%|          | 1/86 [00:15<22:20, 15.78s/it]
Batches:   2%|▏         | 2/86 [00:32<22:49, 16.30s/it]
Batches:   3%|▎         | 3/86 [00:48<22:12, 16.06s/it]
Batches:   5%|▍         | 4/86 [01:02<20:54, 15.29s/it]
Batches:   6%|▌         | 5/86 [01:20<22:04, 16.35s/it]
Batches:   7%|▋         | 6/86 [01:34<20:34, 15.43s/it]
Batches:   8%|▊         | 7/86 [01:52<21:23, 16.25s/it]
Batches:   9%|▉         | 8/86 [02:03<18:53, 14.54s/it]
Batches:  10%|▉         | 9/86 [02:19<19:14, 14.99s/it]
Batches:  12%|█         | 10/86 [02:31<17:58, 14.20s/it]
Batches:  13%|█▏        | 11/86 [02:46<18:12, 14.56s/it]
Batches:  14%|█▎        | 12/86 [03:00<17:38, 14.31s/it]
Batches:  15%|█▍        | 13/86 [03:13<17:02, 14.00s/it]
Batches:  16%|█▌        | 14/86 [03:30<17:53, 14.90s/it]
Batches:  17%|█▋        | 15/86 [03:45<17:38, 14.91s/it]
Batches:  19%|█▊        | 16/86 [04:02<17:57, 15.39s/it]
Batches:  20%|█▉        | 17/86 [04:17<17:40, 15.38s/it]
Batches:  21%|██        | 18/86 [04:30<16:27, 14.52s/it]
Batches:  22%|██▏       | 19/86 [04:44<16:05, 14.41s/it]
Batches:  23%|██▎       | 20/86 [05:02<17:12, 15.64s/it]
Batches:  24%|██▍       | 21/86 [05:18<16:50, 15.55s/it]
Batches:  26%|██▌       | 22/86 [05:34<16:45, 15.72s/it]
Batches:  27%|██▋       | 23/86 [05:49<16:23, 15.62s/it]
Batches:  28%|██▊       | 24/86 [06:03<15:38, 15.13s/it]
Batches:  29%|██▉       | 25/86 [06:19<15:41, 15.44s/it]
Batches:  30%|███       | 26/86 [06:36<15:40, 15.68s/it]
Batches:  31%|███       | 27/86 [06:49<14:50, 15.09s/it]
Batches:  33%|███▎      | 28/86 [07:03<14:15, 14.74s/it]
Batches:  34%|███▎      | 29/86 [07:18<14:03, 14.79s/it]
Batches:  35%|███▍      | 30/86 [07:33<13:52, 14.87s/it]
Batches:  36%|███▌      | 31/86 [07:48<13:31, 14.76s/it]
Batches:  37%|███▋      | 32/86 [08:05<13:58, 15.53s/it]
Batches:  38%|███▊      | 33/86 [08:19<13:14, 14.99s/it]
Batches:  40%|███▉      | 34/86 [08:32<12:31, 14.45s/it]
Batches:  41%|████      | 35/86 [08:48<12:43, 14.97s/it]
Batches:  42%|████▏     | 36/86 [09:06<13:06, 15.73s/it]
Batches:  43%|████▎     | 37/86 [09:20<12:33, 15.38s/it]
Batches:  44%|████▍     | 38/86 [09:37<12:37, 15.78s/it]
Batches:  45%|████▌     | 39/86 [09:50<11:46, 15.03s/it]
Batches:  47%|████▋     | 40/86 [10:05<11:24, 14.89s/it]
Batches:  48%|████▊     | 41/86 [10:20<11:09, 14.88s/it]
Batches:  49%|████▉     | 42/86 [10:32<10:28, 14.29s/it]
Batches:  50%|█████     | 43/86 [10:47<10:17, 14.36s/it]
Batches:  51%|█████     | 44/86 [11:03<10:21, 14.81s/it]
Batches:  52%|█████▏    | 45/86 [11:17<10:03, 14.72s/it]
```

```
Batches:  53%|█████▎    | 46/86 [11:32<09:51, 14.78s/it]
Batches:  55%|█████▍    | 47/86 [11:47<09:37, 14.81s/it]
Batches:  56%|█████▌    | 48/86 [12:03<09:34, 15.11s/it]
Batches:  57%|█████▋    | 49/86 [12:17<09:03, 14.68s/it]
Batches:  58%|█████▊    | 50/86 [12:31<08:47, 14.65s/it]
Batches:  59%|█████▉    | 51/86 [12:47<08:40, 14.88s/it]
Batches:  60%|██████    | 52/86 [13:00<08:13, 14.52s/it]
Batches:  62%|██████▏   | 53/86 [13:17<08:19, 15.14s/it]
Batches:  63%|██████▎   | 54/86 [13:33<08:10, 15.32s/it]
Batches:  64%|██████▍   | 55/86 [13:47<07:48, 15.10s/it]
Batches:  65%|██████▌   | 56/86 [14:03<07:41, 15.38s/it]
Batches:  66%|██████▋   | 57/86 [14:19<07:32, 15.61s/it]
Batches:  67%|██████▋   | 58/86 [14:38<07:44, 16.59s/it]
Batches:  69%|██████▊   | 59/86 [14:55<07:27, 16.59s/it]
Batches:  70%|██████▉   | 60/86 [15:10<06:56, 16.04s/it]
Batches:  71%|███████   | 61/86 [15:25<06:38, 15.94s/it]
Batches:  72%|███████▏  | 62/86 [15:40<06:15, 15.65s/it]
Batches:  73%|███████▎  | 63/86 [15:54<05:46, 15.07s/it]
Batches:  74%|███████▍  | 64/86 [16:11<05:44, 15.67s/it]
Batches:  76%|███████▌  | 65/86 [16:25<05:19, 15.23s/it]
Batches:  77%|███████▋  | 66/86 [16:42<05:11, 15.60s/it]
Batches:  78%|███████▊  | 67/86 [16:55<04:43, 14.93s/it]
Batches:  79%|███████▉  | 68/86 [17:08<04:17, 14.30s/it]
Batches:  80%|████████  | 69/86 [17:22<04:00, 14.14s/it]
Batches:  81%|████████  | 70/86 [17:37<03:51, 14.44s/it]
Batches:  83%|████████▎ | 71/86 [17:52<03:38, 14.57s/it]
Batches:  84%|████████▎ | 72/86 [18:06<03:23, 14.56s/it]
Batches:  85%|████████▍ | 73/86 [18:20<03:04, 14.18s/it]
Batches:  86%|████████▌ | 74/86 [18:35<02:56, 14.70s/it]
Batches:  87%|████████▋ | 75/86 [18:50<02:41, 14.65s/it]
Batches:  88%|████████▊ | 76/86 [19:05<02:27, 14.77s/it]
Batches:  90%|████████▉ | 77/86 [19:23<02:20, 15.66s/it]
Batches:  91%|█████████ | 78/86 [19:39<02:05, 15.69s/it]
Batches:  92%|█████████▏| 79/86 [19:52<01:44, 14.96s/it]
Batches:  93%|█████████▎| 80/86 [20:06<01:29, 14.84s/it]
Batches:  94%|█████████▍| 81/86 [20:23<01:16, 15.31s/it]
Batches:  95%|█████████▌| 82/86 [20:36<00:58, 14.56s/it]
Batches:  97%|█████████▋| 83/86 [20:53<00:45, 15.27s/it]
Batches:  98%|█████████▊| 84/86 [21:07<00:29, 14.93s/it]
Batches:  99%|█████████▉| 85/86 [21:23<00:15, 15.22s/it]
Batches: 100%|██████████| 86/86 [21:27<00:00, 11.93s/it]
Batches: 100%|██████████| 86/86 [21:27<00:00, 14.97s/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(
The model is automatically converting to bf16 for faster inference. If you want to disable the
automatic precision, please manually add bf16/fp16/fp32=True to
"AutoModelForCausalLM.from_pretrained".

======================================================================
✓ 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, 4096)
    Negative: (3002, 4096)
    Labels: (3002,)

  Cached to: hidden_states_cache/cache_attribute_recognition_3410_qwen2.npz

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

  Dataset split (Stratified):
    Train: 2101 samples (1062 pos, 1039 neg)
    Test:  901 samples (456 pos, 445 neg)
    Hidden dim: 4096

  Probe architecture:
    Input: 4096
    Hidden: 256 → 128
    Output: 1 (probability)

  Training config:
    Epochs per trial: 1000
    Number of trials: 10
    Learning rate: 0.001
    Weight decay: 0.01

  ======================================================================
  TRAINING WITH MULTIPLE RANDOM RESTARTS
  ======================================================================
    Trial  1/10: Loss = 0.002807
      ✓ New best probe found!
    Trial  2/10: Loss = 0.002801
      ✓ New best probe found!
    Trial  3/10: Loss = 0.003460
    Trial  4/10: Loss = 0.002690
      ✓ New best probe found!
    Trial  5/10: Loss = 0.002829
    Trial  6/10: Loss = 0.002972
    Trial  7/10: Loss = 0.002774
    Trial  8/10: Loss = 0.002723
    Trial  9/10: Loss = 0.002932
    Trial 10/10: Loss = 0.002826

  ======================================================================
  EVALUATION WITH BEST PROBE
  ======================================================================
  Best loss: 0.002690

  Test Results:
    Overall Accuracy: 80.7% (727/901)
    Positive samples: 78.3% (456 samples)
    Negative samples: 83.1% (445 samples)

  ✓ COMPLETE: attribute_recognition → 80.7%

  ######################################################################
  # CATEGORY: SPATIAL_RECOGNITION
  ######################################################################
  LOADING DATA for category: 'spatial_recognition'
  Using 1030 samples from '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
```

```
Device: cuda

Loading checkpoint shards:   0%|          | 0/10 [00:00<?, ?it/s]
Loading checkpoint shards:  10%|▉         | 1/10 [00:00<00:06,  1.31it/s]
Loading checkpoint shards:  20%|█▊        | 2/10 [00:01<00:06,  1.31it/s]
Loading checkpoint shards:  30%|██▋       | 3/10 [00:02<00:05,  1.30it/s]
Loading checkpoint shards:  40%|███▌      | 4/10 [00:03<00:04,  1.28it/s]
Loading checkpoint shards:  50%|████▌     | 5/10 [00:03<00:03,  1.29it/s]
Loading checkpoint shards:  60%|█████▍    | 6/10 [00:04<00:03,  1.29it/s]
Loading checkpoint shards:  70%|██████▎   | 7/10 [00:05<00:02,  1.30it/s]
Loading checkpoint shards:  80%|███████▏  | 8/10 [00:06<00:01,  1.23it/s]
Loading checkpoint shards:  90%|████████▏ | 9/10 [00:07<00:00,  1.23it/s]
Loading checkpoint shards: 100%|█████████▏| 10/10 [00:07<00:00,  1.32it/s]
Loading checkpoint shards: 100%|██████████| 10/10 [00:07<00:00,  1.29it/s]
✓ Model loaded successfully


Batches:   0%|          | 0/26 [00:00<?, ?it/s]
Batches:   4%|▍         | 1/26 [00:15<06:37, 15.91s/it]
Batches:   8%|▊         | 2/26 [00:32<06:26, 16.09s/it]
Batches:  12%|█▏        | 3/26 [00:44<05:29, 14.31s/it]
Batches:  15%|█▌        | 4/26 [00:58<05:17, 14.43s/it]
Batches:  19%|█▉        | 5/26 [01:17<05:31, 15.81s/it]
Batches:  23%|██▎       | 6/26 [01:32<05:10, 15.51s/it]
Batches:  27%|██▋       | 7/26 [01:47<04:51, 15.34s/it]
Batches:  31%|███       | 8/26 [02:01<04:31, 15.09s/it]
Batches:  35%|███▍      | 9/26 [02:15<04:09, 14.70s/it]
Batches:  38%|███▊      | 10/26 [02:30<03:54, 14.66s/it]
Batches:  42%|████▏     | 11/26 [02:46<03:49, 15.33s/it]
Batches:  46%|████▌     | 12/26 [03:01<03:29, 14.98s/it]
Batches:  50%|█████     | 13/26 [03:14<03:10, 14.62s/it]
Batches:  54%|█████▍    | 14/26 [03:26<02:46, 13.85s/it]
Batches:  58%|█████▊    | 15/26 [03:40<02:31, 13.81s/it]
Batches:  62%|██████▏   | 16/26 [03:55<02:19, 13.98s/it]
Batches:  65%|██████▌   | 17/26 [04:08<02:05, 13.93s/it]
Batches:  69%|██████▉   | 18/26 [04:23<01:52, 14.07s/it]
Batches:  73%|███████▎  | 19/26 [04:38<01:40, 14.34s/it]
Batches:  77%|███████▋  | 20/26 [04:53<01:28, 14.76s/it]
Batches:  81%|████████  | 21/26 [05:06<01:11, 14.21s/it]
Batches:  85%|████████▍ | 22/26 [05:21<00:57, 14.46s/it]
Batches:  88%|████████▊ | 23/26 [05:35<00:42, 14.26s/it]
Batches:  92%|█████████▏| 24/26 [05:51<00:29, 14.70s/it]
Batches:  96%|█████████▌| 25/26 [06:05<00:14, 14.40s/it]
Batches: 100%|██████████| 26/26 [06:16<00:00, 13.59s/it]
Batches: 100%|██████████| 26/26 [06:16<00:00, 14.49s/it]

======================================================================
✓ 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, 4096)
  Negative: (880, 4096)
  Labels: (880,)

Cached to: hidden_states_cache/cache_spatial_recognition_1030_qwen2.npz

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

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

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

Probe architecture:
  Input: 4096
  Hidden: 256 → 128
  Output: 1 (probability)

Training config:
  Epochs per trial: 1000
  Number of trials: 10
  Learning rate: 0.001
  Weight decay: 0.01

======================================================================
TRAINING WITH MULTIPLE RANDOM RESTARTS
======================================================================
  Trial  1/10: Loss = 0.001899
    ✓ New best probe found!
  Trial  2/10: Loss = 0.002005
  Trial  3/10: Loss = 0.002461
  Trial  4/10: Loss = 0.001867
    ✓ New best probe found!
  Trial  5/10: Loss = 0.002131
  Trial  6/10: Loss = 0.002034
  Trial  7/10: Loss = 0.001817
    ✓ New best probe found!
  Trial  8/10: Loss = 0.002062
  Trial  9/10: Loss = 0.002096
  Trial 10/10: Loss = 0.001940

======================================================================
EVALUATION WITH BEST PROBE
======================================================================
Best loss: 0.001817

Test Results:
  Overall Accuracy: 74.0% (196/265)
  Positive samples: 79.4% (126 samples)
  Negative samples: 69.1% (139 samples)

✓ COMPLETE: spatial_recognition → 74.0%

======================================================================

Final Results:
  object_detection         : 50.4%
  attribute_recognition    : 80.7%
  spatial_recognition      : 74.0%

  Average                  : 68.4%

======================================================================

=== Job finished at Wed Oct 22 23:25:23 CEST 2025 with exit code: 0 ===
```