```
=== 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_extended.py exists ===
-rw-r-----. 1 mdemirev mdemirev 20K Oct 22 19:27 vision_ccs_extended.py
=== Running vision_ccs_extended.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/Qwen-VL-Chat
  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: qwen
Device: cuda

Loading checkpoint shards:   0%|          | 0/10 [00:00<?, ?it/s]
Loading checkpoint shards:  10%|█         | 1/10 [00:00<00:08,  1.05it/s]
Loading checkpoint shards:  20%|██        | 2/10 [00:01<00:06,  1.15it/s]
Loading checkpoint shards:  30%|███       | 3/10 [00:02<00:05,  1.20it/s]
Loading checkpoint shards:  40%|████      | 4/10 [00:03<00:04,  1.22it/s]
Loading checkpoint shards:  50%|█████     | 5/10 [00:04<00:04,  1.25it/s]
Loading checkpoint shards:  60%|██████    | 6/10 [00:04<00:03,  1.25it/s]
Loading checkpoint shards:  70%|███████   | 7/10 [00:05<00:02,  1.27it/s]
Loading checkpoint shards:  80%|████████  | 8/10 [00:06<00:01,  1.19it/s]
Loading checkpoint shards:  90%|█████████ | 9/10 [00:07<00:00,  1.12it/s]
Loading checkpoint shards: 100%|██████████| 10/10 [00:08<00:00,  1.19it/s]
Loading checkpoint shards: 100%|██████████| 10/10 [00:08<00:00,  1.19it/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<09:14, 16.82s/it]
Batches:   6%|▌         | 2/34 [00:33<08:48, 16.50s/it]
Batches:   9%|▉         | 3/34 [00:46<07:51, 15.22s/it]
Batches:  12%|█▏        | 4/34 [01:01<07:34, 15.13s/it]
Batches:  15%|█▍        | 5/34 [01:15<07:04, 14.64s/it]
Batches:  18%|█▊        | 6/34 [01:31<07:00, 15.03s/it]
Batches:  21%|██        | 7/34 [01:45<06:41, 14.86s/it]
Batches:  24%|██▎       | 8/34 [01:59<06:13, 14.35s/it]
Batches:  26%|██▋       | 9/34 [02:13<05:57, 14.28s/it]
Batches:  29%|██▉       | 10/34 [02:28<05:50, 14.58s/it]
Batches:  32%|███▏      | 11/34 [02:43<05:34, 14.56s/it]
Batches:  35%|███▌      | 12/34 [02:57<05:17, 14.44s/it]
Batches:  38%|███▊      | 13/34 [03:08<04:46, 13.64s/it]
Batches:  41%|████      | 14/34 [03:23<04:40, 14.02s/it]
Batches:  44%|████▍     | 15/34 [03:36<04:20, 13.70s/it]
Batches:  47%|████▋     | 16/34 [03:52<04:15, 14.19s/it]
Batches:  50%|█████     | 17/34 [04:06<04:01, 14.19s/it]
Batches:  53%|█████▎    | 18/34 [04:21<03:50, 14.41s/it]
Batches:  56%|█████▌    | 19/34 [04:38<03:47, 15.14s/it]
Batches:  59%|█████▉    | 20/34 [04:52<03:27, 14.84s/it]
Batches:  62%|██████▏   | 21/34 [05:08<03:16, 15.15s/it]
Batches:  65%|██████▍   | 22/34 [05:21<02:56, 14.70s/it]
Batches:  68%|██████▊   | 23/34 [05:36<02:42, 14.78s/it]
Batches:  71%|███████   | 24/34 [05:52<02:32, 15.22s/it]
Batches:  74%|███████▎  | 25/34 [06:08<02:17, 15.31s/it]
Batches:  76%|███████▋  | 26/34 [06:22<01:58, 14.85s/it]
Batches:  79%|███████▉  | 27/34 [06:37<01:45, 15.07s/it]
Batches:  82%|████████▏ | 28/34 [06:51<01:27, 14.67s/it]
Batches:  85%|████████▌ | 29/34 [07:06<01:14, 14.80s/it]
Batches:  88%|████████▊ | 30/34 [07:20<00:57, 14.47s/it]
Batches:  91%|█████████ | 31/34 [07:34<00:43, 14.44s/it]
Batches:  94%|█████████▍| 32/34 [07:47<00:27, 13.96s/it]
Batches:  97%|█████████▋| 33/34 [08:01<00:13, 13.90s/it]
Batches: 100%|██████████| 34/34 [08:02<00:00, 10.14s/it]
Batches: 100%|██████████| 34/34 [08:02<00:00, 14.20s/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_qwen.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 = 0.007365
    ✓ New best probe found!
  Trial  2/10: Loss = 0.001993
    ✓ New best probe found!
  Trial  3/10: Loss = 0.002091
  Trial  4/10: Loss = 0.002194
  Trial  5/10: Loss = 0.001888
    ✓ New best probe found!
  Trial  6/10: Loss = 0.001616
    ✓ New best probe found!
  Trial  7/10: Loss = 0.001873
  Trial  8/10: Loss = 0.002105
  Trial  9/10: Loss = 0.001942
  Trial 10/10: Loss = 0.001609
    ✓ New best probe found!

======================================================================
EVALUATION WITH BEST PROBE
======================================================================
Best loss: 0.001609

Test Results:
  Overall Accuracy: 52.8% (181/343)
  Positive samples: 61.3% (173 samples)
  Negative samples: 44.1% (170 samples)

✓ COMPLETE: object_detection → 52.8%

######################################################################
# 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: qwen
Device: cuda
```

```
Loading checkpoint shards:   0%|          | 0/10 [00:00<?, ?it/s]
Loading checkpoint shards:  10%|█         | 1/10 [00:00<00:06,  1.29it/s]
Loading checkpoint shards:  20%|██        | 2/10 [00:01<00:06,  1.26it/s]
Loading checkpoint shards:  30%|███       | 3/10 [00:02<00:05,  1.27it/s]
Loading checkpoint shards:  40%|████      | 4/10 [00:03<00:04,  1.26it/s]
Loading checkpoint shards:  50%|█████     | 5/10 [00:03<00:03,  1.27it/s]
Loading checkpoint shards:  60%|██████    | 6/10 [00:04<00:03,  1.27it/s]
Loading checkpoint shards:  70%|███████   | 7/10 [00:05<00:02,  1.27it/s]
Loading checkpoint shards:  80%|████████  | 8/10 [00:06<00:01,  1.22it/s]
Loading checkpoint shards:  90%|█████████ | 9/10 [00:07<00:00,  1.21it/s]
Loading checkpoint shards: 100%|██████████| 10/10 [00:07<00:00,  1.28it/s]
Loading checkpoint shards: 100%|██████████| 10/10 [00:07<00:00,  1.26it/s]
✓ Model loaded successfully


Batches:   0%|          | 0/86 [00:00<?, ?it/s]
Batches:   1%|          | 1/86 [00:15<21:50, 15.42s/it]
Batches:   2%|▏         | 2/86 [00:32<23:01, 16.44s/it]
Batches:   3%|▎         | 3/86 [00:48<22:17, 16.12s/it]
Batches:   5%|▍         | 4/86 [01:05<22:24, 16.39s/it]
Batches:   6%|▌         | 5/86 [01:22<22:26, 16.63s/it]
Batches:   7%|▋         | 6/86 [01:35<20:49, 15.61s/it]
Batches:   8%|▊         | 7/86 [01:51<20:30, 15.58s/it]
Batches:   9%|▉         | 8/86 [02:03<18:59, 14.60s/it]
Batches:  10%|█         | 9/86 [02:17<18:23, 14.33s/it]
Batches:  12%|█▏        | 10/86 [02:31<18:11, 14.36s/it]
Batches:  13%|█▎        | 11/86 [02:46<18:09, 14.53s/it]
Batches:  14%|█▍        | 12/86 [03:01<17:47, 14.43s/it]
Batches:  15%|█▌        | 13/86 [03:14<17:08, 14.09s/it]
Batches:  16%|█▋        | 14/86 [03:30<17:36, 14.68s/it]
Batches:  17%|█▋        | 15/86 [03:45<17:27, 14.75s/it]
Batches:  19%|█▊        | 16/86 [04:01<17:51, 15.30s/it]
Batches:  20%|█▉        | 17/86 [04:17<17:37, 15.32s/it]
Batches:  21%|██        | 18/86 [04:30<16:35, 14.65s/it]
Batches:  22%|██▏       | 19/86 [04:44<16:15, 14.55s/it]
Batches:  23%|██▎       | 20/86 [05:00<16:31, 15.02s/it]
Batches:  24%|██▍       | 21/86 [05:16<16:22, 15.12s/it]
Batches:  26%|██▌       | 22/86 [05:34<17:02, 15.98s/it]
Batches:  27%|██▋       | 23/86 [05:49<16:34, 15.78s/it]
Batches:  28%|██▊       | 24/86 [06:05<16:21, 15.82s/it]
Batches:  29%|██▉       | 25/86 [06:21<16:12, 15.94s/it]
Batches:  30%|███       | 26/86 [06:37<15:50, 15.83s/it]
Batches:  31%|███▏      | 27/86 [06:51<15:07, 15.38s/it]
Batches:  33%|███▎      | 28/86 [07:05<14:28, 14.98s/it]
Batches:  34%|███▎      | 29/86 [07:20<14:12, 14.96s/it]
Batches:  35%|███▍      | 30/86 [07:34<13:47, 14.77s/it]
Batches:  36%|███▌      | 31/86 [07:49<13:27, 14.69s/it]
Batches:  37%|███▋      | 32/86 [08:06<13:54, 15.44s/it]
Batches:  38%|███▊      | 33/86 [08:20<13:10, 14.91s/it]
Batches:  40%|███▉      | 34/86 [08:34<12:42, 14.67s/it]
Batches:  41%|████      | 35/86 [08:50<12:51, 15.12s/it]
Batches:  42%|████▏     | 36/86 [09:08<13:16, 15.92s/it]
Batches:  43%|████▎     | 37/86 [09:22<12:39, 15.50s/it]
Batches:  44%|████▍     | 38/86 [09:37<12:16, 15.33s/it]
Batches:  45%|████▌     | 39/86 [09:50<11:31, 14.71s/it]
Batches:  47%|████▋     | 40/86 [10:05<11:14, 14.66s/it]
Batches:  48%|████▊     | 41/86 [10:21<11:14, 14.99s/it]
Batches:  49%|████▉     | 42/86 [10:33<10:26, 14.23s/it]
Batches:  50%|█████     | 43/86 [10:48<10:23, 14.50s/it]
```

```
Batches:  51%|█████     | 44/86 [11:02<10:03, 14.37s/it]
Batches:  52%|█████▏    | 45/86 [11:17<09:55, 14.54s/it]
Batches:  53%|█████▎    | 46/86 [11:32<09:41, 14.53s/it]
Batches:  55%|█████▍    | 47/86 [11:47<09:34, 14.74s/it]
Batches:  56%|█████▌    | 48/86 [12:03<09:32, 15.06s/it]
Batches:  57%|█████▋    | 49/86 [12:17<09:08, 14.83s/it]
Batches:  58%|█████▊    | 50/86 [12:31<08:45, 14.61s/it]
Batches:  59%|█████▉    | 51/86 [12:47<08:42, 14.92s/it]
Batches:  60%|██████    | 52/86 [13:01<08:15, 14.58s/it]
Batches:  62%|██████▏   | 53/86 [13:19<08:38, 15.70s/it]
Batches:  63%|██████▎   | 54/86 [13:35<08:22, 15.71s/it]
Batches:  64%|██████▍   | 55/86 [13:50<08:04, 15.63s/it]
Batches:  65%|██████▌   | 56/86 [14:05<07:39, 15.31s/it]
Batches:  66%|██████▋   | 57/86 [14:22<07:43, 15.98s/it]
Batches:  67%|██████▋   | 58/86 [14:38<07:28, 16.01s/it]
Batches:  69%|██████▊   | 59/86 [14:56<07:24, 16.46s/it]
Batches:  70%|██████▉   | 60/86 [15:10<06:49, 15.75s/it]
Batches:  71%|███████   | 61/86 [15:26<06:38, 15.95s/it]
Batches:  72%|███████▏  | 62/86 [15:41<06:15, 15.66s/it]
Batches:  73%|███████▎  | 63/86 [15:55<05:48, 15.16s/it]
Batches:  74%|███████▍  | 64/86 [16:11<05:37, 15.34s/it]
Batches:  76%|███████▌  | 65/86 [16:28<05:33, 15.89s/it]
Batches:  77%|███████▋  | 66/86 [16:44<05:16, 15.84s/it]
Batches:  78%|███████▊  | 67/86 [16:58<04:51, 15.33s/it]
Batches:  79%|███████▉  | 68/86 [17:10<04:18, 14.35s/it]
Batches:  80%|████████  | 69/86 [17:24<04:02, 14.28s/it]
Batches:  81%|████████▏ | 70/86 [17:39<03:51, 14.47s/it]
Batches:  83%|████████▎ | 71/86 [17:55<03:44, 14.99s/it]
Batches:  84%|████████▎ | 72/86 [18:10<03:26, 14.72s/it]
Batches:  85%|████████▍ | 73/86 [18:23<03:06, 14.34s/it]
Batches:  86%|████████▌ | 74/86 [18:39<02:57, 14.83s/it]
Batches:  87%|████████▋ | 75/86 [18:54<02:43, 14.88s/it]
Batches:  88%|████████▊ | 76/86 [19:09<02:29, 14.90s/it]
Batches:  90%|████████▉ | 77/86 [19:23<02:12, 14.68s/it]
Batches:  91%|█████████ | 78/86 [19:39<02:00, 15.01s/it]
Batches:  92%|█████████▏| 79/86 [19:52<01:41, 14.49s/it]
Batches:  93%|█████████▎| 80/86 [20:07<01:28, 14.67s/it]
Batches:  94%|█████████▍| 81/86 [20:23<01:14, 14.88s/it]
Batches:  95%|█████████▌| 82/86 [20:37<00:58, 14.59s/it]
Batches:  97%|█████████▋| 83/86 [20:53<00:45, 15.09s/it]
Batches:  98%|█████████▊| 84/86 [21:08<00:30, 15.02s/it]
Batches:  99%|█████████▉| 85/86 [21:23<00:15, 15.23s/it]
Batches: 100%|██████████| 86/86 [21:30<00:00, 12.76s/it]
Batches: 100%|██████████| 86/86 [21:30<00:00, 15.01s/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_qwen.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 = 0.002722
    ✓ New best probe found!
  Trial  2/10: Loss = 0.003134
  Trial  3/10: Loss = 0.003382
  Trial  4/10: Loss = 0.002780
  Trial  5/10: Loss = 0.002698
    ✓ New best probe found!
  Trial  6/10: Loss = 0.002746
  Trial  7/10: Loss = 0.002921
  Trial  8/10: Loss = 0.003231
  Trial  9/10: Loss = 0.002793
  Trial 10/10: Loss = 0.002733

======================================================================
EVALUATION WITH BEST PROBE
======================================================================
Best loss: 0.002698

Test Results:
  Overall Accuracy: 74.9% (675/901)
  Positive samples: 86.4% (456 samples)
  Negative samples: 63.1% (445 samples)

✓ COMPLETE: attribute_recognition → 74.9%

######################################################################
# 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: qwen
Device: cuda

Loading checkpoint shards:   0%|          | 0/10 [00:00<?, ?it/s]
Loading checkpoint shards:  10%|█         | 1/10 [00:00<00:07,  1.27it/s]
Loading checkpoint shards:  20%|██        | 2/10 [00:01<00:06,  1.29it/s]
Loading checkpoint shards:  30%|███       | 3/10 [00:02<00:05,  1.28it/s]
Loading checkpoint shards:  40%|████      | 4/10 [00:03<00:04,  1.27it/s]
Loading checkpoint shards:  50%|█████     | 5/10 [00:03<00:03,  1.27it/s]
Loading checkpoint shards:  60%|██████    | 6/10 [00:04<00:03,  1.26it/s]
Loading checkpoint shards:  70%|███████   | 7/10 [00:05<00:02,  1.26it/s]
Loading checkpoint shards:  80%|████████  | 8/10 [00:06<00:01,  1.23it/s]
Loading checkpoint shards:  90%|█████████ | 9/10 [00:07<00:00,  1.21it/s]
Loading checkpoint shards: 100%|██████████| 10/10 [00:07<00:00,  1.29it/s]
Loading checkpoint shards: 100%|██████████| 10/10 [00:07<00:00,  1.27it/s]
✓ Model loaded successfully


Batches:   0%|          | 0/26 [00:00<?, ?it/s]
Batches:   4%|▍         | 1/26 [00:15<06:34, 15.78s/it]
Batches:   8%|▊         | 2/26 [00:31<06:24, 16.02s/it]
Batches:  12%|█▏        | 3/26 [00:43<05:21, 14.00s/it]
Batches:  15%|█▌        | 4/26 [00:58<05:14, 14.30s/it]
Batches:  19%|█▉        | 5/26 [01:12<05:02, 14.40s/it]
Batches:  23%|██▎       | 6/26 [01:30<05:07, 15.38s/it]
Batches:  27%|██▋       | 7/26 [01:45<04:49, 15.25s/it]
Batches:  31%|███       | 8/26 [01:59<04:32, 15.12s/it]
Batches:  35%|███▍      | 9/26 [02:12<04:05, 14.42s/it]
Batches:  38%|███▊      | 10/26 [02:28<03:54, 14.67s/it]
Batches:  42%|████▏     | 11/26 [02:43<03:41, 14.75s/it]
Batches:  46%|████▌     | 12/26 [02:58<03:28, 14.87s/it]
Batches:  50%|█████     | 13/26 [03:11<03:07, 14.41s/it]
Batches:  54%|█████▍    | 14/26 [03:23<02:44, 13.73s/it]
Batches:  58%|█████▊    | 15/26 [03:38<02:34, 14.03s/it]
Batches:  62%|██████▏   | 16/26 [03:52<02:19, 13.93s/it]
Batches:  65%|██████▌   | 17/26 [04:06<02:07, 14.20s/it]
Batches:  69%|██████▉   | 18/26 [04:20<01:52, 14.04s/it]
Batches:  73%|███████▎  | 19/26 [04:36<01:41, 14.53s/it]
Batches:  77%|███████▋  | 20/26 [04:51<01:28, 14.78s/it]
Batches:  81%|████████  | 21/26 [05:05<01:12, 14.50s/it]
Batches:  85%|████████▍ | 22/26 [05:20<00:58, 14.63s/it]
Batches:  88%|████████▊ | 23/26 [05:34<00:43, 14.46s/it]
Batches:  92%|█████████▏| 24/26 [05:49<00:29, 14.63s/it]
Batches:  96%|█████████▌| 25/26 [06:03<00:14, 14.45s/it]
Batches: 100%|██████████| 26/26 [06:14<00:00, 13.48s/it]
Batches: 100%|██████████| 26/26 [06:14<00:00, 14.41s/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_qwen.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 = 0.002054
    ✓ New best probe found!
  Trial  2/10: Loss = 0.001985
    ✓ New best probe found!
  Trial  3/10: Loss = 0.001870
    ✓ New best probe found!
  Trial  4/10: Loss = 0.002061
  Trial  5/10: Loss = 0.002127
  Trial  6/10: Loss = 0.002020
  Trial  7/10: Loss = 0.001855
    ✓ New best probe found!
  Trial  8/10: Loss = 0.001994
  Trial  9/10: Loss = 0.001891
  Trial 10/10: Loss = 0.001901

======================================================================
EVALUATION WITH BEST PROBE
======================================================================
Best loss: 0.001855

Test Results:
  Overall Accuracy: 75.5% (200/265)
  Positive samples: 67.5% (126 samples)
  Negative samples: 82.7% (139 samples)

✓ COMPLETE: spatial_recognition → 75.5%

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

Final Results:
  object_detection         : 52.8%
  attribute_recognition    : 74.9%
  spatial_recognition      : 75.5%

  Average                  : 67.7%

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

=== Job finished at Wed Oct 22 20:31:46 CEST 2025 with exit code: 0 ===
```