```
=== Running supervised_vision_ccs.py ===
`torch_dtype` is deprecated! Use `dtype` instead!
LLaVA + Contrast Pairs + Logistic Regression

######################################################################
# CATEGORY: OBJECT_DETECTION
EXTRACTING HIDDEN STATES: OBJECT_DETECTION

Loading checkpoint shards:   0%|          | 0/3 [00:00<?, ?it/s]
Loading checkpoint shards:  33%|███▎      | 1/3 [00:05<00:11,  5.63s/it]
Loading checkpoint shards:  67%|██████▋   | 2/3 [00:11<00:05,  5.52s/it]
Loading checkpoint shards: 100%|██████████| 3/3 [00:15<00:00,  5.09s/it]
Loading checkpoint shards: 100%|██████████| 3/3 [00:15<00:00,  5.22s/it]

Batches:   0%|          | 0/34 [00:00<?, ?it/s]
Batches:   3%|▎         | 1/34 [00:06<03:31,  6.42s/it]
Batches:   6%|▌         | 2/34 [00:12<03:12,  6.03s/it]
Batches:   9%|▉         | 3/34 [00:17<02:51,  5.52s/it]
Batches:  12%|█▏        | 4/34 [00:22<02:40,  5.35s/it]
Batches:  15%|█▍        | 5/34 [00:27<02:31,  5.23s/it]
Batches:  18%|█▊        | 6/34 [00:32<02:29,  5.34s/it]
Batches:  21%|██        | 7/34 [00:38<02:23,  5.31s/it]
Batches:  24%|██▎       | 8/34 [00:42<02:14,  5.16s/it]
Batches:  26%|██▋       | 9/34 [00:47<02:08,  5.15s/it]
Batches:  29%|██▉       | 10/34 [00:53<02:05,  5.23s/it]
Batches:  32%|███▏      | 11/34 [00:58<02:00,  5.24s/it]
Batches:  35%|███▌      | 12/34 [01:03<01:54,  5.21s/it]
Batches:  38%|███▊      | 13/34 [01:07<01:42,  4.87s/it]
Batches:  41%|████      | 14/34 [01:13<01:40,  5.04s/it]
Batches:  44%|████▍     | 15/34 [01:17<01:33,  4.93s/it]
Batches:  47%|████▋     | 16/34 [01:23<01:32,  5.11s/it]
Batches:  50%|█████     | 17/34 [01:28<01:25,  5.03s/it]
Batches:  53%|█████▎    | 18/34 [01:33<01:22,  5.15s/it]
Batches:  56%|█████▌    | 19/34 [01:39<01:20,  5.37s/it]
Batches:  59%|█████▉    | 20/34 [01:44<01:14,  5.29s/it]
Batches:  62%|██████▏   | 21/34 [01:50<01:08,  5.28s/it]
Batches:  65%|██████▍   | 22/34 [01:54<01:02,  5.19s/it]
Batches:  68%|██████▊   | 23/34 [02:00<00:57,  5.21s/it]
Batches:  71%|███████   | 24/34 [02:06<00:54,  5.41s/it]
Batches:  74%|███████▎  | 25/34 [02:11<00:48,  5.41s/it]
Batches:  76%|███████▋  | 26/34 [02:16<00:42,  5.29s/it]
Batches:  79%|███████▉  | 27/34 [02:22<00:37,  5.37s/it]
Batches:  82%|████████▏ | 28/34 [02:26<00:30,  5.17s/it]
Batches:  85%|████████▌ | 29/34 [02:32<00:25,  5.20s/it]
Batches:  88%|████████▊ | 30/34 [02:37<00:20,  5.13s/it]
Batches:  91%|█████████ | 31/34 [02:42<00:15,  5.13s/it]
Batches:  94%|█████████▍| 32/34 [02:46<00:09,  4.99s/it]
Batches:  97%|█████████▋| 33/34 [02:51<00:04,  4.98s/it]
Batches: 100%|██████████| 34/34 [02:52<00:00,  3.65s/it]
Batches: 100%|██████████| 34/34 [02:52<00:00,  5.07s/it]
Successfully processed: 1140/1323
Skipped (missing/error): 183/1323

There are skipped images: 000000262227.jpg, 000000262440.jpg, 000000262440.jpg, 000000262682.jpg,
000000262682.jpg, 000000262682.jpg, 000000139684.jpg, 000000000632.jpg, 000000000632.jpg,
000000000632.jpg

Extracted shapes:
  Positive hidden states: (1140, 4096)
  Negative hidden states: (1140, 4096)
```

```
  Labels: (1140,)

Cached to: hidden_states_cache/cache_object_detection_1323_supervised_contrast_llava.npz
TRAINING SUPERVISED LOGISTIC REGRESSION

Dataset split:
  Train: 570 samples (273 pos, 297 neg)
  Test:  570 samples (303 pos, 267 neg)
  Hidden dim: 4096
Logistic regression accuracy: 0.7859649122807018

✓ COMPLETE: object_detection → 78.6%

######################################################################
# CATEGORY: ATTRIBUTE_RECOGNITION
EXTRACTING HIDDEN STATES: ATTRIBUTE_RECOGNITION

Loading checkpoint shards:   0%|          | 0/3 [00:00<?, ?it/s]
Loading checkpoint shards:  33%|███▎      | 1/3 [00:02<00:04,  2.29s/it]
Loading checkpoint shards:  67%|██████▋   | 2/3 [00:04<00:02,  2.17s/it]
Loading checkpoint shards: 100%|██████████| 3/3 [00:06<00:00,  1.99s/it]
Loading checkpoint shards: 100%|██████████| 3/3 [00:06<00:00,  2.05s/it]

Batches:   0%|          | 0/86 [00:00<?, ?it/s]
Batches:   1%|          | 1/86 [00:05<07:54,  5.58s/it]
Batches:   2%|▏         | 2/86 [00:11<08:10,  5.84s/it]
Batches:   3%|▎         | 3/86 [00:17<07:59,  5.78s/it]
Batches:   5%|▍         | 4/86 [00:22<07:33,  5.53s/it]
Batches:   6%|▌         | 5/86 [00:28<07:37,  5.65s/it]
Batches:   7%|▋         | 6/86 [00:33<07:11,  5.39s/it]
Batches:   8%|▊         | 7/86 [00:38<07:09,  5.43s/it]
Batches:   9%|▉         | 8/86 [00:42<06:26,  4.96s/it]
Batches:  10%|▉         | 9/86 [00:47<06:21,  4.96s/it]
Batches:  12%|█         | 10/86 [00:52<06:05,  4.81s/it]
Batches:  13%|█▏        | 11/86 [00:57<06:14,  5.00s/it]
Batches:  14%|█▎        | 12/86 [01:02<06:08,  4.98s/it]
Batches:  15%|█▍        | 13/86 [01:07<06:00,  4.94s/it]
Batches:  16%|█▌        | 14/86 [01:12<05:55,  4.94s/it]
Batches:  17%|█▋        | 15/86 [01:17<06:01,  5.09s/it]
Batches:  19%|█▊        | 16/86 [01:23<06:12,  5.32s/it]
Batches:  20%|█▉        | 17/86 [01:29<06:11,  5.38s/it]
Batches:  21%|██        | 18/86 [01:33<05:47,  5.12s/it]
Batches:  22%|██▏       | 19/86 [01:38<05:42,  5.12s/it]
Batches:  23%|██▎       | 20/86 [01:44<05:46,  5.25s/it]
Batches:  24%|██▍       | 21/86 [01:49<05:47,  5.34s/it]
Batches:  26%|██▌       | 22/86 [01:55<05:50,  5.48s/it]
Batches:  27%|██▋       | 23/86 [02:01<05:46,  5.50s/it]
Batches:  28%|██▊       | 24/86 [02:06<05:30,  5.33s/it]
Batches:  29%|██▉       | 25/86 [02:11<05:34,  5.48s/it]
Batches:  30%|███       | 26/86 [02:17<05:30,  5.51s/it]
Batches:  31%|███       | 27/86 [02:22<05:15,  5.35s/it]
Batches:  33%|███▎      | 28/86 [02:27<05:01,  5.20s/it]
Batches:  34%|███▎      | 29/86 [02:32<04:59,  5.26s/it]
Batches:  35%|███▍      | 30/86 [02:37<04:46,  5.12s/it]
Batches:  36%|███▌      | 31/86 [02:42<04:43,  5.16s/it]
Batches:  37%|███▋      | 32/86 [02:48<04:52,  5.42s/it]
Batches:  38%|███▊      | 33/86 [02:53<04:39,  5.28s/it]
Batches:  40%|███▉      | 34/86 [02:58<04:25,  5.11s/it]
Batches:  41%|████      | 35/86 [03:04<04:31,  5.33s/it]
Batches:  42%|████      | 36/86 [03:10<04:34,  5.49s/it]
```

```
Batches:  43%|████▎       | 37/86 [03:15<04:26,  5.43s/it]
Batches:  44%|████▍       | 38/86 [03:20<04:17,  5.37s/it]
Batches:  45%|████▌       | 39/86 [03:25<04:04,  5.21s/it]
Batches:  47%|████▋       | 40/86 [03:30<04:00,  5.22s/it]
Batches:  48%|████▊       | 41/86 [03:36<03:57,  5.27s/it]
Batches:  49%|████▉       | 42/86 [03:40<03:41,  5.04s/it]
Batches:  50%|█████       | 43/86 [03:45<03:39,  5.10s/it]
Batches:  51%|█████       | 44/86 [03:50<03:33,  5.09s/it]
Batches:  52%|█████▏      | 45/86 [03:56<03:30,  5.14s/it]
Batches:  53%|█████▎      | 46/86 [04:01<03:27,  5.18s/it]
Batches:  55%|█████▍      | 47/86 [04:06<03:24,  5.25s/it]
Batches:  56%|█████▌      | 48/86 [04:12<03:24,  5.39s/it]
Batches:  57%|█████▋      | 49/86 [04:17<03:14,  5.26s/it]
Batches:  58%|█████▊      | 50/86 [04:22<03:07,  5.21s/it]
Batches:  59%|█████▉      | 51/86 [04:28<03:06,  5.32s/it]
Batches:  60%|██████      | 52/86 [04:33<02:57,  5.22s/it]
Batches:  62%|██████▏     | 53/86 [04:38<02:56,  5.36s/it]
Batches:  63%|██████▎     | 54/86 [04:44<02:54,  5.45s/it]
Batches:  64%|██████▍     | 55/86 [04:49<02:47,  5.40s/it]
Batches:  65%|██████▌     | 56/86 [04:55<02:40,  5.36s/it]
Batches:  66%|██████▋     | 57/86 [05:00<02:39,  5.51s/it]
Batches:  67%|██████▋     | 58/86 [05:06<02:36,  5.61s/it]
Batches:  69%|██████▊     | 59/86 [05:12<02:34,  5.72s/it]
Batches:  70%|██████▉     | 60/86 [05:17<02:24,  5.54s/it]
Batches:  71%|███████     | 61/86 [05:23<02:19,  5.59s/it]
Batches:  72%|███████▏    | 62/86 [05:29<02:12,  5.53s/it]
Batches:  73%|███████▎    | 63/86 [05:33<02:03,  5.36s/it]
Batches:  74%|███████▍    | 64/86 [05:39<02:00,  5.46s/it]
Batches:  76%|███████▌    | 65/86 [05:44<01:52,  5.36s/it]
Batches:  77%|███████▋    | 66/86 [05:50<01:49,  5.45s/it]
Batches:  78%|███████▊    | 67/86 [05:55<01:40,  5.27s/it]
Batches:  79%|███████▉    | 68/86 [05:59<01:30,  5.00s/it]
Batches:  80%|████████    | 69/86 [06:04<01:24,  5.00s/it]
Batches:  81%|████████    | 70/86 [06:10<01:21,  5.12s/it]
Batches:  83%|████████▎   | 71/86 [06:15<01:18,  5.21s/it]
Batches:  84%|████████▎   | 72/86 [06:20<01:12,  5.17s/it]
Batches:  85%|████████▍   | 73/86 [06:25<01:05,  5.07s/it]
Batches:  86%|████████▌   | 74/86 [06:31<01:03,  5.26s/it]
Batches:  87%|████████▋   | 75/86 [06:36<00:57,  5.26s/it]
Batches:  88%|████████▊   | 76/86 [06:41<00:53,  5.31s/it]
Batches:  90%|████████▉   | 77/86 [06:46<00:46,  5.21s/it]
Batches:  91%|█████████   | 78/86 [06:52<00:42,  5.36s/it]
Batches:  92%|█████████▏  | 79/86 [06:57<00:36,  5.20s/it]
Batches:  93%|█████████▎  | 80/86 [07:02<00:31,  5.22s/it]
Batches:  94%|█████████▍  | 81/86 [07:08<00:26,  5.32s/it]
Batches:  95%|█████████▌  | 82/86 [07:12<00:20,  5.12s/it]
Batches:  97%|█████████▋  | 83/86 [07:18<00:16,  5.34s/it]
Batches:  98%|█████████▊  | 84/86 [07:23<00:10,  5.27s/it]
Batches:  99%|█████████▉  | 85/86 [07:29<00:05,  5.40s/it]
Batches: 100%|██████████| 86/86 [07:31<00:00,  4.25s/it]
Batches: 100%|██████████| 86/86 [07:31<00:00,  5.24s/it]
/gpfs/home6/mdemirev/snellius/venv/lib/python3.11/site-
packages/sklearn/linear_model/_logistic.py:473: ConvergenceWarning: lbfgs failed to converge after
100 iteration(s) (status=1):
STOP: TOTAL NO. OF ITERATIONS REACHED LIMIT

Increase the number of iterations to improve the convergence (max_iter=100).
```

```
You might also want to scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
Successfully processed: 3002/3410
Skipped (missing/error): 408/3410

There are skipped images: 000000393282.jpg, 000000393282.jpg, 000000393282.jpg, 000000393469.jpg,
000000000285.jpg, 000000262440.jpg, 000000262440.jpg, 000000262440.jpg, 000000262440.jpg,
000000131386.jpg

Extracted shapes:
  Positive hidden states: (3002, 4096)
  Negative hidden states: (3002, 4096)
  Labels: (3002,)

Cached to: hidden_states_cache/cache_attribute_recognition_3410_supervised_contrast_llava.npz
TRAINING SUPERVISED LOGISTIC REGRESSION

Dataset split:
  Train: 1501 samples (745 pos, 756 neg)
  Test:  1501 samples (773 pos, 728 neg)
  Hidden dim: 4096
Logistic regression accuracy: 0.7341772151898734

✓ COMPLETE: attribute_recognition → 73.4%

######################################################################
# CATEGORY: SPATIAL_RECOGNITION
EXTRACTING HIDDEN STATES: SPATIAL_RECOGNITION

Loading checkpoint shards:   0%|          | 0/3 [00:00<?, ?it/s]
Loading checkpoint shards:  33%|███▎      | 1/3 [00:02<00:04,  2.33s/it]
Loading checkpoint shards:  67%|██████▋   | 2/3 [00:04<00:02,  2.21s/it]
Loading checkpoint shards: 100%|██████████| 3/3 [00:06<00:00,  2.02s/it]
Loading checkpoint shards: 100%|██████████| 3/3 [00:06<00:00,  2.09s/it]

Batches:   0%|          | 0/26 [00:00<?, ?it/s]
Batches:   4%|▍         | 1/26 [00:05<02:22,  5.71s/it]
Batches:   8%|▊         | 2/26 [00:11<02:19,  5.79s/it]
Batches:  12%|█▏        | 3/26 [00:15<01:56,  5.08s/it]
Batches:  15%|█▌        | 4/26 [00:21<01:53,  5.17s/it]
Batches:  19%|█▉        | 5/26 [00:26<01:49,  5.22s/it]
Batches:  23%|██▎       | 6/26 [00:31<01:45,  5.29s/it]
Batches:  27%|██▋       | 7/26 [00:37<01:41,  5.33s/it]
Batches:  31%|███       | 8/26 [00:42<01:35,  5.31s/it]
Batches:  35%|███▍      | 9/26 [00:47<01:26,  5.12s/it]
Batches:  38%|███▊      | 10/26 [00:52<01:22,  5.16s/it]
Batches:  42%|████▏     | 11/26 [00:57<01:18,  5.24s/it]
Batches:  46%|████▌     | 12/26 [01:03<01:13,  5.22s/it]
Batches:  50%|█████     | 13/26 [01:07<01:06,  5.11s/it]
Batches:  54%|█████▍    | 14/26 [01:12<00:58,  4.89s/it]
Batches:  58%|█████▊    | 15/26 [01:17<00:54,  4.92s/it]
Batches:  62%|██████▏   | 16/26 [01:22<00:49,  4.94s/it]
Batches:  65%|██████▌   | 17/26 [01:27<00:44,  4.96s/it]
Batches:  69%|██████▉   | 18/26 [01:32<00:39,  4.97s/it]
Batches:  73%|███████▎  | 19/26 [01:37<00:35,  5.10s/it]
Batches:  77%|███████▋  | 20/26 [01:43<00:31,  5.24s/it]
Batches:  81%|████████  | 21/26 [01:47<00:25,  5.08s/it]
Batches:  85%|████████▍ | 22/26 [01:53<00:20,  5.18s/it]
Batches:  88%|████████▊ | 23/26 [01:58<00:15,  5.14s/it]
Batches:  92%|█████████▏| 24/26 [02:03<00:10,  5.23s/it]
```

```
Batches:  96%|█████████▌| 25/26 [02:08<00:05,  5.15s/it]
Batches: 100%|██████████| 26/26 [02:12<00:00,  4.83s/it]
Batches: 100%|██████████| 26/26 [02:12<00:00,  5.11s/it]
/gpfs/home6/mdemirev/snellius/venv/lib/python3.11/site-
packages/sklearn/linear_model/_logistic.py:473: ConvergenceWarning: lbfgs failed to converge after
100 iteration(s) (status=1):
STOP: TOTAL NO. OF ITERATIONS REACHED LIMIT

Increase the number of iterations to improve the convergence (max_iter=100).
You might also want to scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
Successfully processed: 880/1030
Skipped (missing/error): 150/1030

There are skipped images: 000000393282.jpg, 000000000285.jpg, 000000262682.jpg, 000000000632.jpg,
000000262895.jpg, 000000043816.jpg, 000000043816.jpg, 000000043816.jpg, 000000043816.jpg,
000000000785.jpg

Extracted shapes:
  Positive hidden states: (880, 4096)
  Negative hidden states: (880, 4096)
  Labels: (880,)

Cached to: hidden_states_cache/cache_spatial_recognition_1030_supervised_contrast_llava.npz
TRAINING SUPERVISED LOGISTIC REGRESSION

Dataset split:
  Train: 440 samples (214 pos, 226 neg)
  Test:  440 samples (206 pos, 234 neg)
  Hidden dim: 4096
Logistic regression accuracy: 0.7340909090909091

✓ COMPLETE: spatial_recognition → 73.4%

Final Results:
  object_detection         : 78.6%
  attribute_recognition    : 73.4%
  spatial_recognition      : 73.4%

  Average                  : 75.1%
=== Job finished at Fri Oct 24 02:21:39 CEST 2025 with exit code: 0 ===
```