```
=== Running vision_linear.py ===
The image processor of type `Qwen2VLImageProcessor` is now loaded as a fast processor by default, even if the model checkpoint was saved with a slow processor. This is a breaking change and may produce slightly different outputs. To continue using the slow processor, instantiate this class with `use_fast=False`. Note that this behavior will be extended to all models in a future release.
`torch_dtype` is deprecated! Use `dtype` instead!
Qwen/Qwen2.5-VL-7B-Instruct + Contrast Pairs + Supervised Methods

######################################################################
# 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: qwen2_5
Device: cuda

Loading checkpoint shards:   0%|          | 0/5 [00:00<?, ?it/s]
Loading checkpoint shards:  20%|██        | 1/5 [00:01<00:07,  1.87s/it]
Loading checkpoint shards:  40%|████      | 2/5 [00:03<00:05,  1.77s/it]
Loading checkpoint shards:  60%|██████    | 3/5 [00:05<00:03,  1.75s/it]
Loading checkpoint shards:  80%|████████  | 4/5 [00:07<00:01,  1.75s/it]
Loading checkpoint shards: 100%|██████████| 5/5 [00:07<00:00,  1.28s/it]
Loading checkpoint shards: 100%|██████████| 5/5 [00:07<00:00,  1.50s/it]
✓ Model loaded successfully


Batches:   0%|          | 0/34 [00:00<?, ?it/s]
Batches:   3%|▎         | 1/34 [00:09<05:13,  9.51s/it]
Batches:   6%|▌         | 2/34 [00:18<05:02,  9.45s/it]
Batches:   9%|▉         | 3/34 [00:26<04:24,  8.52s/it]
Batches:  12%|█▏        | 4/34 [00:33<04:02,  8.08s/it]
Batches:  15%|█▍        | 5/34 [00:41<03:51,  7.97s/it]
Batches:  18%|█▊        | 6/34 [00:50<03:50,  8.24s/it]
Batches:  21%|██        | 7/34 [00:58<03:40,  8.18s/it]
Batches:  24%|██▎       | 8/34 [01:05<03:25,  7.90s/it]
Batches:  26%|██▋       | 9/34 [01:13<03:17,  7.88s/it]
Batches:  29%|██▉       | 10/34 [01:21<03:11,  8.00s/it]
Batches:  32%|███▏      | 11/34 [01:30<03:06,  8.11s/it]
Batches:  35%|███▌      | 12/34 [01:38<02:58,  8.13s/it]
Batches:  38%|███▊      | 13/34 [01:44<02:41,  7.70s/it]
Batches:  41%|████      | 14/34 [01:53<02:37,  7.88s/it]
Batches:  44%|████▍     | 15/34 [02:00<02:26,  7.69s/it]
Batches:  47%|████▋     | 16/34 [02:09<02:22,  7.94s/it]
Batches:  50%|█████     | 17/34 [02:16<02:12,  7.80s/it]
Batches:  53%|█████▎    | 18/34 [02:24<02:08,  8.00s/it]
Batches:  56%|█████▌    | 19/34 [02:34<02:04,  8.33s/it]
Batches:  59%|█████▉    | 20/34 [02:41<01:53,  8.14s/it]
Batches:  62%|██████▏   | 21/34 [02:49<01:45,  8.11s/it]
Batches:  65%|██████▍   | 22/34 [02:57<01:35,  7.94s/it]
Batches:  68%|██████▊   | 23/34 [03:05<01:28,  8.03s/it]
Batches:  71%|███████   | 24/34 [03:14<01:23,  8.37s/it]
Batches:  74%|███████▎  | 25/34 [03:22<01:14,  8.31s/it]
Batches:  76%|███████▋  | 26/34 [03:30<01:05,  8.21s/it]
Batches:  79%|███████▉  | 27/34 [03:39<00:58,  8.40s/it]
Batches:  82%|████████▏ | 28/34 [03:46<00:47,  8.00s/it]
Batches:  85%|████████▌ | 29/34 [03:55<00:40,  8.07s/it]
Batches:  88%|████████▊ | 30/34 [04:02<00:31,  7.90s/it]
Batches:  91%|█████████ | 31/34 [04:10<00:23,  7.98s/it]
```

```
Batches:  94%|█████████▍ | 32/34 [04:17<00:15,  7.65s/it]
Batches:  97%|█████████▋ | 33/34 [04:25<00:07,  7.65s/it]
Batches: 100%|██████████| 34/34 [04:26<00:00,  5.66s/it]
Batches: 100%|██████████| 34/34 [04:26<00:00,  7.83s/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: 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, 3584)
  Negative hidden states: (1140, 3584)
  Labels: (1140,)

Cached to: hidden_states_cache/cache_object_detection_1323_supervised_contrast_qwen2_5.npz

======================================================================
SUPERVISED LOGISTIC REGRESSION
======================================================================

Dataset split:
  Train: 570 samples (273 pos, 297 neg)
  Test:  570 samples (303 pos, 267 neg)
  Hidden dim: 3584

Logistic regression accuracy: 83.3%

======================================================================
SUPERVISED LINEAR PROBE
======================================================================

Results:
  Train Accuracy: 89.1%
  Test Accuracy:  85.3%
  Positive samples: 83.8% (303 samples)
  Negative samples: 86.9% (267 samples)

======================================================================
COMPARISON SUMMARY
======================================================================
  Logistic Regression:     83.3%
  Supervised Linear Probe: 85.3%

✓ COMPLETE: object_detection

######################################################################
# 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
```

```
Loading checkpoint shards:   0%|          | 0/5 [00:00<?, ?it/s]
Loading checkpoint shards:  20%|██        | 1/5 [00:01<00:07,  1.87s/it]
Loading checkpoint shards:  40%|████      | 2/5 [00:03<00:05,  1.77s/it]
Loading checkpoint shards:  60%|██████    | 3/5 [00:05<00:03,  1.75s/it]
Loading checkpoint shards:  80%|████████  | 4/5 [00:07<00:01,  1.74s/it]
Loading checkpoint shards: 100%|██████████| 5/5 [00:07<00:00,  1.28s/it]
Loading checkpoint shards: 100%|██████████| 5/5 [00:07<00:00,  1.50s/it]
✓ Model loaded successfully


Batches:   0%|          | 0/86 [00:00<?, ?it/s]
Batches:   1%|          | 1/86 [00:08<12:30,  8.83s/it]
Batches:   2%|▏         | 2/86 [00:18<13:08,  9.38s/it]
Batches:   3%|▎         | 3/86 [00:27<12:44,  9.21s/it]
Batches:   5%|▍         | 4/86 [00:35<11:45,  8.61s/it]
Batches:   6%|▌         | 5/86 [00:44<11:47,  8.74s/it]
Batches:   7%|▋         | 6/86 [00:51<11:02,  8.28s/it]
Batches:   8%|▊         | 7/86 [00:59<10:51,  8.25s/it]
Batches:   9%|▉         | 8/86 [01:06<09:54,  7.62s/it]
Batches:  10%|▉         | 9/86 [01:13<09:39,  7.53s/it]
Batches:  12%|█▏        | 10/86 [01:20<09:11,  7.26s/it]
Batches:  13%|█▎        | 11/86 [01:28<09:33,  7.64s/it]
Batches:  14%|█▍        | 12/86 [01:36<09:22,  7.60s/it]
Batches:  15%|█▌        | 13/86 [01:43<09:17,  7.63s/it]
Batches:  16%|█▋        | 14/86 [01:51<09:06,  7.60s/it]
Batches:  17%|█▋        | 15/86 [01:59<09:13,  7.79s/it]
Batches:  19%|█▊        | 16/86 [02:08<09:34,  8.20s/it]
Batches:  20%|█▉        | 17/86 [02:17<09:31,  8.28s/it]
Batches:  21%|██        | 18/86 [02:23<08:50,  7.81s/it]
Batches:  22%|██▏       | 19/86 [02:31<08:39,  7.76s/it]
Batches:  23%|██▎       | 20/86 [02:40<08:49,  8.02s/it]
Batches:  24%|██▍       | 21/86 [02:48<08:52,  8.19s/it]
Batches:  26%|██▌       | 22/86 [02:57<08:51,  8.30s/it]
Batches:  27%|██▋       | 23/86 [03:05<08:48,  8.39s/it]
Batches:  28%|██▊       | 24/86 [03:13<08:25,  8.16s/it]
Batches:  29%|██▉       | 25/86 [03:22<08:31,  8.39s/it]
Batches:  30%|███       | 26/86 [03:31<08:30,  8.51s/it]
Batches:  31%|███       | 27/86 [03:39<08:10,  8.31s/it]
Batches:  33%|███▎      | 28/86 [03:46<07:51,  8.12s/it]
Batches:  34%|███▎      | 29/86 [03:55<07:46,  8.19s/it]
Batches:  35%|███▍      | 30/86 [04:02<07:26,  7.97s/it]
Batches:  36%|███▌      | 31/86 [04:10<07:21,  8.03s/it]
Batches:  37%|███▋      | 32/86 [04:20<07:41,  8.55s/it]
Batches:  38%|███▊      | 33/86 [04:28<07:18,  8.27s/it]
Batches:  40%|███▉      | 34/86 [04:35<07:02,  8.12s/it]
Batches:  41%|████      | 35/86 [04:45<07:11,  8.47s/it]
Batches:  42%|████▏     | 36/86 [04:54<07:12,  8.66s/it]
Batches:  43%|████▎     | 37/86 [05:02<06:56,  8.51s/it]
Batches:  44%|████▍     | 38/86 [05:10<06:40,  8.34s/it]
Batches:  45%|████▌     | 39/86 [05:17<06:20,  8.10s/it]
Batches:  47%|████▋     | 40/86 [05:25<06:12,  8.09s/it]
Batches:  48%|████▊     | 41/86 [05:33<06:01,  8.04s/it]
Batches:  49%|████▉     | 42/86 [05:40<05:36,  7.66s/it]
Batches:  50%|█████     | 43/86 [05:49<05:38,  7.87s/it]
Batches:  51%|█████     | 44/86 [05:56<05:22,  7.68s/it]
Batches:  52%|█████▏    | 45/86 [06:04<05:23,  7.89s/it]
Batches:  53%|█████▎    | 46/86 [06:13<05:22,  8.06s/it]
```

```
Batches:  55%|█████▌     | 47/86 [06:21<05:15,  8.09s/it]
Batches:  56%|█████▌     | 48/86 [06:30<05:16,  8.34s/it]
Batches:  57%|█████▋     | 49/86 [06:37<04:59,  8.10s/it]
Batches:  58%|█████▊     | 50/86 [06:45<04:48,  8.01s/it]
Batches:  59%|█████▉     | 51/86 [06:54<04:46,  8.18s/it]
Batches:  60%|██████     | 52/86 [07:01<04:34,  8.07s/it]
Batches:  62%|██████▏    | 53/86 [07:10<04:31,  8.24s/it]
Batches:  63%|██████▎    | 54/86 [07:19<04:28,  8.38s/it]
Batches:  64%|██████▍    | 55/86 [07:27<04:17,  8.32s/it]
Batches:  65%|██████▌    | 56/86 [07:35<04:08,  8.28s/it]
Batches:  66%|██████▋    | 57/86 [07:44<04:06,  8.50s/it]
Batches:  67%|██████▋    | 58/86 [07:53<04:02,  8.65s/it]
Batches:  69%|██████▊    | 59/86 [08:02<03:57,  8.80s/it]
Batches:  70%|██████▉    | 60/86 [08:10<03:41,  8.51s/it]
Batches:  71%|███████    | 61/86 [08:19<03:34,  8.59s/it]
Batches:  72%|███████▏   | 62/86 [08:27<03:25,  8.55s/it]
Batches:  73%|███████▎   | 63/86 [08:35<03:10,  8.26s/it]
Batches:  74%|███████▍   | 64/86 [08:44<03:05,  8.42s/it]
Batches:  76%|███████▌   | 65/86 [08:52<02:55,  8.36s/it]
Batches:  77%|███████▋   | 66/86 [09:00<02:46,  8.34s/it]
Batches:  78%|███████▊   | 67/86 [09:08<02:37,  8.27s/it]
Batches:  79%|███████▉   | 68/86 [09:15<02:21,  7.84s/it]
Batches:  80%|████████   | 69/86 [09:23<02:14,  7.89s/it]
Batches:  81%|████████   | 70/86 [09:31<02:07,  7.95s/it]
Batches:  83%|████████▎  | 71/86 [09:40<02:01,  8.13s/it]
Batches:  84%|████████▎  | 72/86 [09:47<01:51,  7.98s/it]
Batches:  85%|████████▍  | 73/86 [09:55<01:42,  7.85s/it]
Batches:  86%|████████▌  | 74/86 [10:04<01:38,  8.18s/it]
Batches:  87%|████████▋  | 75/86 [10:12<01:30,  8.22s/it]
Batches:  88%|████████▊  | 76/86 [10:21<01:23,  8.33s/it]
Batches:  90%|████████▉  | 77/86 [10:29<01:13,  8.16s/it]
Batches:  91%|█████████  | 78/86 [10:38<01:07,  8.43s/it]
Batches:  92%|█████████▏ | 79/86 [10:45<00:57,  8.17s/it]
Batches:  93%|█████████▎ | 80/86 [10:53<00:48,  8.14s/it]
Batches:  94%|█████████▍ | 81/86 [11:02<00:41,  8.27s/it]
Batches:  95%|█████████▌ | 82/86 [11:09<00:31,  7.81s/it]
Batches:  97%|█████████▋ | 83/86 [11:18<00:24,  8.21s/it]
Batches:  98%|█████████▊ | 84/86 [11:26<00:16,  8.13s/it]
Batches:  99%|█████████▉ | 85/86 [11:34<00:08,  8.31s/it]
Batches: 100%|██████████| 86/86 [11:37<00:00,  6.56s/it]
Batches: 100%|██████████| 86/86 [11:37<00:00,  8.11s/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: 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, 3584)
  Negative hidden states: (3002, 3584)
  Labels: (3002,)

Cached to: hidden_states_cache/cache_attribute_recognition_3410_supervised_contrast_qwen2_5.npz

======================================================================
SUPERVISED LOGISTIC REGRESSION
======================================================================

Dataset split:
  Train: 1501 samples (745 pos, 756 neg)
  Test:  1501 samples (773 pos, 728 neg)
  Hidden dim: 3584

Logistic regression accuracy: 79.1%

======================================================================
SUPERVISED LINEAR PROBE
======================================================================

Results:
  Train Accuracy: 85.5%
  Test Accuracy:  83.1%
  Positive samples: 81.9% (773 samples)
  Negative samples: 84.3% (728 samples)

======================================================================
COMPARISON SUMMARY
======================================================================
  Logistic Regression:     79.1%
  Supervised Linear Probe: 83.1%

✓ COMPLETE: attribute_recognition

######################################################################
# 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

Loading checkpoint shards:   0%|          | 0/5 [00:00<?, ?it/s]
Loading checkpoint shards:  20%|██        | 1/5 [00:01<00:07,  1.88s/it]
Loading checkpoint shards:  40%|████      | 2/5 [00:03<00:05,  1.78s/it]
Loading checkpoint shards:  60%|██████    | 3/5 [00:05<00:03,  1.76s/it]
Loading checkpoint shards:  80%|████████  | 4/5 [00:07<00:01,  1.76s/it]
Loading checkpoint shards: 100%|██████████| 5/5 [00:07<00:00,  1.29s/it]
Loading checkpoint shards: 100%|██████████| 5/5 [00:07<00:00,  1.51s/it]
✓ Model loaded successfully


Batches:   0%|          | 0/26 [00:00<?, ?it/s]
Batches:   4%|▍         | 1/26 [00:08<03:38,  8.76s/it]
Batches:   8%|▊         | 2/26 [00:17<03:37,  9.04s/it]
Batches:  12%|█▏        | 3/26 [00:24<02:59,  7.81s/it]
Batches:  15%|█▌        | 4/26 [00:32<02:55,  7.99s/it]
Batches:  19%|█▉        | 5/26 [00:40<02:50,  8.13s/it]
Batches:  23%|██▎       | 6/26 [00:49<02:42,  8.14s/it]
Batches:  27%|██▋       | 7/26 [00:57<02:35,  8.16s/it]
Batches:  31%|███       | 8/26 [01:05<02:26,  8.15s/it]
Batches:  35%|███▍      | 9/26 [01:12<02:12,  7.82s/it]
```

```
Batches:  38%|███▊      | 10/26 [01:20<02:07,  7.95s/it]
Batches:  42%|████▏      | 11/26 [01:28<02:00,  8.01s/it]
Batches:  46%|████▌     | 12/26 [01:37<01:52,  8.03s/it]
Batches:  50%|█████     | 13/26 [01:44<01:42,  7.85s/it]
Batches:  54%|█████▍     | 14/26 [01:51<01:30,  7.53s/it]
Batches:  58%|█████▊     | 15/26 [01:58<01:23,  7.57s/it]
Batches:  62%|██████▏    | 16/26 [02:06<01:15,  7.58s/it]
Batches:  65%|██████▌    | 17/26 [02:14<01:08,  7.65s/it]
Batches:  69%|██████▉    | 18/26 [02:22<01:01,  7.68s/it]
Batches:  73%|███████▎   | 19/26 [02:30<00:54,  7.82s/it]
Batches:  77%|███████▋   | 20/26 [02:38<00:47,  7.98s/it]
Batches:  81%|████████   | 21/26 [02:45<00:38,  7.74s/it]
Batches:  85%|████████▍  | 22/26 [02:54<00:31,  7.95s/it]
Batches:  88%|████████▊  | 23/26 [03:02<00:23,  7.94s/it]
Batches:  92%|█████████▏ | 24/26 [03:10<00:16,  8.13s/it]
Batches:  96%|█████████▌| 25/26 [03:18<00:07,  7.91s/it]
Batches: 100%|██████████| 26/26 [03:24<00:00,  7.32s/it]
Batches: 100%|██████████| 26/26 [03:24<00:00,  7.85s/it]
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, 3584)
  Negative hidden states: (880, 3584)
  Labels: (880,)

Cached to: hidden_states_cache/cache_spatial_recognition_1030_supervised_contrast_qwen2_5.npz

======================================================================
SUPERVISED LOGISTIC REGRESSION
======================================================================

Dataset split:
  Train: 440 samples (214 pos, 226 neg)
  Test:  440 samples (206 pos, 234 neg)
  Hidden dim: 3584

Logistic regression accuracy: 69.5%

======================================================================
SUPERVISED LINEAR PROBE
======================================================================

Results:
  Train Accuracy: 84.1%
  Test Accuracy:  76.8%
  Positive samples: 69.9% (206 samples)
  Negative samples: 82.9% (234 samples)

======================================================================
COMPARISON SUMMARY
======================================================================
  Logistic Regression:     69.5%
  Supervised Linear Probe: 76.8%

✓ COMPLETE: spatial_recognition

======================================================================
FINAL RESULTS SUMMARY
======================================================================

Category                        LogReg       Linear Probe
```

```
  ------------------------------------------------------------------------
    object_detection                  83.3%       85.3%
    attribute_recognition             79.1%       83.1%
    spatial_recognition               69.5%       76.8%
  ------------------------------------------------------------------------
    Average                           77.3%       81.7%

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

  === Job finished at Fri Oct 24 22:39:43 CEST 2025 with exit code: 0 ===
```