

Furniture Engine Test Evaluation Report

Report file: furntniture-test-evaluation.docx

Generated: 2026-09-09 14:38

Engine: furniture | Catalogue domain: furniture | Catalogue size: 161 items

Test set: 26 held-out images from asia_furnitures/test/

Confidence threshold: 45.0%

Metric: category retrieval (query filename category vs top-match category). Exact duplicate images were not in the catalogue.

1. Executive Summary

The furniture matching engine was evaluated on 26 held-out product images. Top-1 category accuracy is 10/26 (38.5%). Top-3 accuracy is 14/26 (53.8%) and top-5 is 18/26 (69.2%). All 26 queries scored above the 45.0% confidence threshold, including 16 incorrect matches — indicating poor confidence calibration.

2. Overall Metrics

Metric	Result	Rate
Top-1 category accuracy	10/26	38.5%
Top-3 category accuracy	14/26	53.8%
Top-5 category accuracy	18/26	69.2%
Above confidence threshold	26/26	100.0%

3. Methodology / Test Flow

1. Load each query image from asia_furnitures/test/.
2. Extract furniture-engine features (LSD line geometry, SIFT/affine keypoints, material signature).
3. Compare against all catalogue items using precomputed feature caches (.npz).
4. Composite confidence = $0.40 \times \text{line} + 0.35 \times \text{affine} + 0.25 \times \text{material}$.
5. Rank catalogue items by confidence; take top match.
6. Mark correct if top-match category equals query filename category (e.g. sofa_12 → sofa).
7. Compare confidence to threshold 45% (UI 'match confirmed' gate).

4. Per-Category Accuracy

Expected category	Correct / Total	Accuracy
armchair	0/3	0%
cabinet	1/4	25%
chair	2/6	33%

drawer_near_bed	0/2	0%
lamp	2/3	67%
makeup_chair	1/1	100%
photo_frame	1/2	50%
sofa	3/5	60%

5. Images That Worked (Correct Top-1 Category)

10 of 26 queries returned a top match in the correct furniture category.

✓ carbinet_48.jpg — PASS

Expected category: cabinet

Top match: Cabinet 43 (id=fur-11eee9e5, sku=SKU-FURN-CARBINET-43)

Matched category: cabinet

Confidence: 94.5% (threshold 45.0%: PASS)

Score breakdown: line=94.3%, affine=100.0% (inliers=27), material=87.0%

Top-3 runners-up: Cabinet 43 [cabinet] 94.5%; Sofa 4 [sofa] 86.3%; Cabinet 40 [cabinet] 74.2%

Query (left) vs Top Match (right):

QUERY: carbinet_48.jpg



TOP MATCH: Cabinet 43 (94.5%)



✓ chair_22.jpg — PASS

Expected category: chair

Top match: Chair 20 (id=fur-90b1bafb, sku=SKU-FURN-CHAIR-20)

Matched category: chair

Confidence: 100.0% (threshold 45.0%: PASS)

Score breakdown: line=100.0%, affine=100.0% (inliers=293), material=100.0%

Top-3 runners-up: Chair 20 [chair] 100.0%; Drawer Next To Bed 46 [drawer_next_to_bed] 84.9%; Chair 24 [chair] 83.9%

Query (left) vs Top Match (right):

QUERY: chair_22.jpg

TOP MATCH: Chair 20 (100.0%)



✓ chair_46.jpg — PASS

Expected category: chair

Top match: Chair 4 (id=fur-a918bd98, sku=SKU-FURN-CHAIR-4)

Matched category: chair

Confidence: 87.5% (threshold 45.0%: PASS)

Score breakdown: line=82.5%, affine=100.0% (inliers=37), material=78.2%

Top-3 runners-up: Chair 4 [chair] 87.5%; Drawer Next To Bed 50 [drawer_next_to_bed] 85.9%; Lamp 52 [lamp] 85.1%

Query (left) vs Top Match (right):

QUERY: chair_46.jpg

TOP MATCH: Chair 4 (87.5%)



✓ lamp_40.jpg — PASS

Expected category: lamp

Top match: Lamp 52 (id=fur-762d6d78, sku=SKU-FURN-LAMP-52)

Matched category: lamp

Confidence: 88.5% (threshold 45.0%: PASS)

Score breakdown: line=83.3%, affine=100.0% (inliers=30), material=81.0%

Top-3 runners-up: Lamp 52 [lamp] 88.5%; Chair 4 [chair] 87.1%; Cabinet 42 [cabinet] 85.1%

Query (left) vs Top Match (right):

QUERY: lamp_40.jpg

TOP MATCH: Lamp 52 (88.5%)



✓ lamp_47.jpg — PASS

Expected category: lamp

Top match: Lamp 52 (id=fur-762d6d78, sku=SKU-FURN-LAMP-52)

Matched category: lamp

Confidence: 82.1% (threshold 45.0%: PASS)

Score breakdown: line=87.0%, affine=100.0% (inliers=27), material=49.1%

Top-3 runners-up: Lamp 52 [lamp] 82.1%; Chair 4 [chair] 80.8%; Sofa 1 [sofa] 76.9%

Query (left) vs Top Match (right):

QUERY: lamp_47.jpg



TOP MATCH: Lamp 52 (82.1%)



✓ makeupchair_2.jpg — PASS

Expected category: makeup_chair

Top match: Makeup Chair 5 (id=fur-29b559e0, sku=SKU-FURN-MAKEUPCHAIR-5)

Matched category: makeup_chair

Confidence: 76.6% (threshold 45.0%: PASS)

Score breakdown: line=80.3%, affine=62.7% (inliers=15), material=90.3%

Top-3 runners-up: Makeup Chair 5 [makeup_chair] 76.6%; Sofa 17 [sofa] 72.2%; Sofa 4 [sofa] 70.3%

Query (left) vs Top Match (right):

QUERY: makeupchair_2.jpg

TOP MATCH: Makeup Chair 5 (76.6%)



✓ **photoframe_12.jpg — PASS**

Expected category: photo_frame

Top match: Photo Frame 11 (id=fur-14066198, sku=SKU-FURN-PHOTOFRAME-11)

Matched category: photo_frame

Confidence: 70.9% (threshold 45.0%: PASS)

Score breakdown: line=88.9%, affine=38.0% (inliers=4), material=88.1%

Top-3 runners-up: Photo Frame 11 [photo_frame] 70.9%; Sofa 50 [sofa] 69.9%; Photo Frame 13 [photo_frame] 68.7%

Query (left) vs Top Match (right):

QUERY: photoframe_12.jpg

TOP MATCH: Photo Frame 11



✓ sofa_12.jpg — PASS

Expected category: sofa

Top match: Sofa 4 (id=fur-824e87cf, sku=SKU-FURN-SOFA-4)

Matched category: sofa

Confidence: 90.4% (threshold 45.0%: PASS)

Score breakdown: line=99.0%, affine=100.0% (inliers=62), material=63.1%

Top-3 runners-up: Sofa 4 [sofa] 90.4%; Sofa 45 [sofa] 88.3%; Drawer Next To Bed 44 [drawer_next_to_bed] 77.9%

Query (left) vs Top Match (right):



✓ sofa_15.jpg — PASS

Expected category: sofa

Top match: Sofa 16 (id=fur-cca6c582, sku=SKU-FURN-SOFA-16)

Matched category: sofa

Confidence: 87.9% (threshold 45.0%: PASS)

Score breakdown: line=98.1%, affine=74.5% (inliers=17), material=90.1%

Top-3 runners-up: Sofa 16 [sofa] 87.9%; Photo Frame 19 [photo_frame] 86.5%; Lamp 8 [lamp] 80.7%

Query (left) vs Top Match (right):



✓ sofa_7.jpg — PASS

Expected category: sofa

Top match: Sofa 10 (id=fur-94c51871, sku=SKU-FURN-SOFA-10)

Matched category: sofa

Confidence: 87.6% (threshold 45.0%: PASS)

Score breakdown: line=88.0%, affine=100.0% (inliers=33), material=69.5%

Top-3 runners-up: Sofa 10 [sofa] 87.6%; Sofa 4 [sofa] 83.8%; Cabinet 3 [cabinet] 74.8%

Query (left) vs Top Match (right):



6. Images That Did Not Work (Wrong Top-1 Category)

16 of 26 queries returned a wrong-category top match. Each case below includes the visual pair, score details, and a likely failure reason.

✗ armchair_46.jpg — FAIL

Expected category: armchair

Top match: Lamp 12 (id=fur-1e5d52ac, sku=SKU-FURN-LAMP-12)

Matched category (WRONG): lamp

Confidence: 82.7% (still above threshold — false positive confirmation)

Score breakdown: line=73.1%, affine=100.0% (inliers=27), material=73.7%

Top-3 runners-up: Lamp 12 [lamp] 82.7%; Lamp 15 [lamp] 81.1%; Makeup Chair 9 [makeup_chair] 80.9%

Correct category rank (within top-10): 7

Query (left) vs Incorrect Top Match (right):

QUERY: armchair_46.jpg

TOP MATCH: Lamp 12 (82.7%)



Likely reason it failed: Armchair silhouette and lamp product shots can share strong vertical edges; line-geometry score dominates and can prefer a lamp with similar rectilinear structure. Correct category first appeared at rank #7. Top match score breakdown — line: 73.1%, affine: 100.0% (inliers=27), material: 73.7%.

✗ armchair_47.jpg — FAIL

Expected category: armchair

Top match: Sofa 4 (id=fur-824e87cf, sku=SKU-FURN-SOFA-4)

Matched category (WRONG): sofa

Confidence: 71.6% (still above threshold — false positive confirmation)

Score breakdown: line=78.2%, affine=65.4% (inliers=17), material=69.6%

Top-3 runners-up: Sofa 4 [sofa] 71.6%; Cabinet 41 [cabinet] 71.3%; Sofa 17 [sofa] 68.6%

Correct category rank (within top-10): 9

Query (left) vs Incorrect Top Match (right):

QUERY: armchair_47.jpg

TOP MATCH: Sofa 4 (71.6%)



Likely reason it failed: Armchairs and sofas share cushioned upholstery and horizontal/vertical frame lines; material + structure cues overlap heavily. Correct category first appeared at rank #9. Top match score breakdown — line: 78.2%, affine: 65.4% (inliers=17), material: 69.6%.

✗ armchair_48.jpg — FAIL

Expected category: armchair

Top match: Lamp 8 (id=fur-a6803c5e, sku=SKU-FURN-LAMP-8)

Matched category (WRONG): lamp

Confidence: 91.3% (still above threshold — false positive confirmation)

Score breakdown: line=86.3%, affine=100.0% (inliers=26), material=87.1%

Top-3 runners-up: Lamp 8 [lamp] 91.3%; Photo Frame 19 [photo_frame] 76.0%; Armchair 45 [armchair] 70.3%

Correct category rank (within top-10): 3

Query (left) vs Incorrect Top Match (right):

QUERY: armchair_48.jpg

TOP MATCH: Lamp 8 (91.3%)



Likely reason it failed: Armchair silhouette and lamp product shots can share strong vertical edges; line-geometry score dominates and can prefer a lamp with similar rectilinear structure. Correct category first appeared at rank #3. Top match score breakdown — line: 86.3%, affine: 100.0% (inliers=26), material: 87.1%. Note: the correct category was in the top-3, so ranking/thresholding (not total miss) is the issue.

✗ carbinet_44.jpg — FAIL

Expected category: cabinet

Top match: Sofa 4 (id=fur-824e87cf, sku=SKU-FURN-SOFA-4)

Matched category (WRONG): sofa

Confidence: 84.1% (still above threshold — false positive confirmation)

Score breakdown: line=79.3%, affine=95.5% (inliers=26), material=76.0%

Top-3 runners-up: Sofa 4 [sofa] 84.1%; Cabinet 43 [cabinet] 73.9%; Photo Frame 15 [photo_frame] 72.3%

Correct category rank (within top-10): 2

Query (left) vs Incorrect Top Match (right):

QUERY: carbinet_44.jpg

TOP MATCH: Sofa 4 (84.1%)



Likely reason it failed: Cabinets and sofas are both large boxy rectilinear objects; LSD line histograms look similar for frontal product shots. Correct category first appeared at rank #2. Top match score breakdown — line: 79.3%, affine: 95.5% (inliers=26), material: 76.0%. Note: the correct category was in the top-3, so ranking/thresholding (not total miss) is the issue.

X carbinet_45.jpg — FAIL

Expected category: cabinet

Top match: Drawer Next To Bed 40 (id=fur-23b5b67b, sku=SKU-FURN-DRAWERNEXTTOBED-40)

Matched category (WRONG): drawer_next_to_bed

Confidence: 77.6% (still above threshold — false positive confirmation)

Score breakdown: line=77.1%, affine=69.5% (inliers=17), material=89.5%

Top-3 runners-up: Drawer Next To Bed 40 [drawer_next_to_bed] 77.6%; Cabinet 41 [cabinet] 76.6%; Sofa 4 [sofa] 75.5%

Correct category rank (within top-10): 2

Query (left) vs Incorrect Top Match (right):

QUERY: carbinet_45.jpg

TOP MATCH: Drawer Next To Bed 40 (



Likely reason it failed: Cabinets and bedside drawers are both boxy storage furniture with similar edge orientation distributions. Correct category first appeared at rank #2. Top match score breakdown — line: 77.1%, affine: 69.5% (inliers=17), material: 89.5%. Note: the correct category was in the top-3, so ranking/thresholding (not total miss) is the issue.

X carbinet_50.jpg — FAIL

Expected category: cabinet

Top match: Photo Frame 19 (id=fur-d3d811f4, sku=SKU-FURN-PHOTOFRAME-19)

Matched category (WRONG): photo_frame

Confidence: 87.7% (still above threshold — false positive confirmation)

Score breakdown: line=98.2%, affine=80.7% (inliers=20), material=80.8%

Top-3 runners-up: Photo Frame 19 [photo_frame] 87.7%; Sofa 45 [sofa] 84.8%; Drawer Next To Bed 47 [drawer_next_to_bed] 84.4%

Correct category rank (within top-10): 5

Query (left) vs Incorrect Top Match (right):

QUERY: carbinet_50.jpg

TOP MATCH: Photo Frame 19 (87.7%)



Likely reason it failed: Framed rectangular panels (doors/faces vs picture frames) produce similar outer-edge geometry under line detection. Correct category first appeared at rank #5.
Top match score breakdown — line: 98.2%, affine: 80.7% (inliers=20), material: 80.8%.

✗ chair_50.jpg — FAIL

Expected category: chair

Top match: Cabinet 41 (id=fur-c75463b2, sku=SKU-FURN-CARBINET-41)

Matched category (WRONG): cabinet

Confidence: 85.7% (still above threshold — false positive confirmation)

Score breakdown: line=72.7%, affine=100.0% (inliers=39), material=86.2%

Top-3 runners-up: Cabinet 41 [cabinet] 85.7%; Cabinet 42 [cabinet] 82.4%; Cabinet 3 [cabinet] 80.7%

Correct category rank (within top-10): 6

Query (left) vs Incorrect Top Match (right):

QUERY: chair_50.jpg

TOP MATCH: Cabinet 41 (85.7%)



Likely reason it failed: Chairs with strong rectangular backs and cabinets with panel faces can produce overlapping line geometry and sparse keypoints. Correct category first appeared at rank #6. Top match score breakdown — line: 72.7%, affine: 100.0% (inliers=39), material: 86.2%.

✗ chair_6.jpg — FAIL

Expected category: chair

Top match: Makeup Chair 6 (id=fur-f3b710ed, sku=SKU-FURN-MAKEUPCHAIR-6)

Matched category (WRONG): makeup_chair

Confidence: 80.5% (still above threshold — false positive confirmation)

Score breakdown: line=66.7%, affine=90.8% (inliers=24), material=88.3%

Top-3 runners-up: Makeup Chair 6 [makeup_chair] 80.5%; Chair 24 [chair] 79.9%; Chair 4 [chair] 74.3%

Correct category rank (within top-10): 2

Query (left) vs Incorrect Top Match (right):

QUERY: chair_6.jpg

TOP MATCH: Mal



Likely reason it failed: Very close categories — both chairs. Matcher lacks an explicit semantic class head, so near-duplicates of makeup chairs can outrank dining chairs. Correct category first appeared at rank #2. Top match score breakdown — line: 66.7%, affine: 90.8% (inliers=24), material: 88.3%. Note: the correct category was in the top-3, so ranking/thresholding (not total miss) is the issue.

X chair_7.jpg — FAIL

Expected category: chair

Top match: Cabinet 43 (id=fur-11eee9e5, sku=SKU-FURN-CARBINET-43)

Matched category (WRONG): cabinet

Confidence: 83.0% (still above threshold — false positive confirmation)

Score breakdown: line=86.8%, affine=89.0% (inliers=22), material=68.6%

Top-3 runners-up: Cabinet 43 [cabinet] 83.0%; Lamp 52 [lamp] 77.5%; Armchair 51 [armchair] 76.3%

Correct category rank (within top-10): 5

Query (left) vs Incorrect Top Match (right):

QUERY: chair_7.jpg

TOP MATCH: Cabinet 43 (83.0%)



Likely reason it failed: Chairs with strong rectangular backs and cabinets with panel faces can produce overlapping line geometry and sparse keypoints. Correct category first appeared at rank #5. Top match score breakdown — line: 86.8%, affine: 89.0% (inliers=22), material: 68.6%.

✗ chair_9.jpg — FAIL

Expected category: chair

Top match: Sofa 1 (id=fur-ae515c2b, sku=SKU-FURN-SOFA-1)

Matched category (WRONG): sofa

Confidence: 81.5% (still above threshold — false positive confirmation)

Score breakdown: line=84.7%, affine=94.4% (inliers=24), material=58.3%

Top-3 runners-up: Sofa 1 [sofa] 81.5%; Photo Frame 15 [photo_frame] 78.3%; Cabinet 49 [cabinet] 75.4%

Correct category rank (within top-10): 8

Query (left) vs Incorrect Top Match (right):

QUERY: chair_9.jpg

TOP MATCH: Sofa 1 (81.5%)



Likely reason it failed: Seating family confusion: shared horizontal seat/back lines and fabric material signatures. Correct category first appeared at rank #8. Top match score breakdown — line: 84.7%, affine: 94.4% (inliers=24), material: 58.3%.

✗ drawernearbed_11.jpg — FAIL

Expected category: drawer_near_bed

Top match: Photo Frame 15 (id=fur-3d7289d3, sku=SKU-FURN-PHOTOFRAME-15)

Matched category (WRONG): photo_frame

Confidence: 86.7% (still above threshold — false positive confirmation)

Score breakdown: line=85.4%, affine=100.0% (inliers=29), material=70.0%

Top-3 runners-up: Photo Frame 15 [photo_frame] 86.7%; Drawer Next To Bed 45 [drawer_next_to_bed] 84.0%; Sofa 4 [sofa] 82.7%

Correct category rank (within top-10): Not in top-10

Query (left) vs Incorrect Top Match (right):

QUERY: drawernearbed_11.jpg



TOP MATCH: Photo Frame 15 (86.7%)



Likely reason it failed: Drawer fronts and picture frames are both flat rectangular surfaces; line geometry + material finish correlate strongly. Correct category did not appear in the top-10 ranked results. Top match score breakdown — line: 85.4%, affine: 100.0% (inliers=29), material: 70.0%.

X drawernearbed_5.jpg — FAIL

Expected category: drawer_near_bed

Top match: Photo Frame 19 (id=fur-d3d811f4, sku=SKU-FURN-PHOTOFRAME-19)

Matched category (WRONG): photo_frame

Confidence: 92.4% (still above threshold — false positive confirmation)

Score breakdown: line=95.3%, affine=100.0% (inliers=55), material=77.1%

Top-3 runners-up: Photo Frame 19 [photo_frame] 92.4%; Cabinet 42 [cabinet] 89.9%; Sofa 4 [sofa] 86.8%

Correct category rank (within top-10): Not in top-10

Query (left) vs Incorrect Top Match (right):

QUERY: drawernearbed_5.jpg

TOP MATCH: Photo Frame 19 (92.4%)



Likely reason it failed: Drawer fronts and picture frames are both flat rectangular surfaces; line geometry + material finish correlate strongly. Correct category did not appear in the top-10 ranked results. Top match score breakdown — line: 95.3%, affine: 100.0% (inliers=55), material: 77.1%.

X lamp_41.jpg — FAIL

Expected category: lamp

Top match: Cabinet 3 (id=fur-08db82b6, sku=SKU-FURN-CARBINET-3)

Matched category (WRONG): cabinet

Confidence: 69.3% (still above threshold — false positive confirmation)

Score breakdown: line=55.0%, affine=100.0% (inliers=29), material=49.4%

Top-3 runners-up: Cabinet 3 [cabinet] 69.3%; Cabinet 2 [cabinet] 66.2%; Sofa 4 [sofa] 65.8%

Correct category rank (within top-10): 4

Query (left) vs Incorrect Top Match (right):

QUERY: lamp_41.jpg

TOP MATCH: Cabinet 3 (69.3%)



Likely reason it failed: If lamp shade/base features are weak, large cabinet edges can dominate composite score via line weight (40%). Correct category first appeared at rank #4. Top match score breakdown — line: 55.0%, affine: 100.0% (inliers=29), material: 49.4%.

✗ photoframe_16.jpg — FAIL

Expected category: photo_frame

Top match: Cabinet 4 (id=fur-18a30320, sku=SKU-FURN-CARBINET-4)

Matched category (WRONG): cabinet

Confidence: 91.5% (still above threshold — false positive confirmation)

Score breakdown: line=88.3%, affine=100.0% (inliers=27), material=84.5%

Top-3 runners-up: Cabinet 4 [cabinet] 91.5%; Sofa 45 [sofa] 82.1%; Cabinet 49 [cabinet] 82.1%

Correct category rank (within top-10): 9

Query (left) vs Incorrect Top Match (right):

QUERY: photoframe_16.jpg

TOP MATCH: Cabinet 4 (91.5%)



Likely reason it failed: Rectangular framed objects — frames vs cabinet doors — are hard to separate with geometry-first scoring. Correct category first appeared at rank #9. Top match score breakdown — line: 88.3%, affine: 100.0% (inliers=27), material: 84.5%.

✗ sofa_43.jpg — FAIL

Expected category: sofa

Top match: Photo Frame 18 (id=fur-04f800f8, sku=SKU-FURN-PHOTOFRAME-18)

Matched category (WRONG): photo_frame

Confidence: 81.9% (still above threshold — false positive confirmation)

Score breakdown: line=86.7%, affine=82.2% (inliers=21), material=73.7%

Top-3 runners-up: Photo Frame 18 [photo_frame] 81.9%; Drawer Next To Bed 44 [drawer_next_to_bed] 79.4%; Drawer Next To Bed 45 [drawer_next_to_bed] 75.8%

Correct category rank (within top-10): 4

Query (left) vs Incorrect Top Match (right):

QUERY: sofa_43.jpg

TOP MATCH: Photo Frame 18 (81.9%)



Likely reason it failed: When sofa texture keypoints are weak, large rectangular frame-like regions in the image can win on line geometry. Correct category first appeared at rank #4. Top match score breakdown — line: 86.7%, affine: 82.2% (inliers=21), material: 73.7%.

✗ sofa_47.jpg — FAIL

Expected category: sofa

Top match: Cabinet 42 (id=fur-a711c6b6, sku=SKU-FURN-CARBINET-42)

Matched category (WRONG): cabinet

Confidence: 86.9% (still above threshold — false positive confirmation)

Score breakdown: line=86.3%, affine=100.0% (inliers=31), material=69.5%

Top-3 runners-up: Cabinet 42 [cabinet] 86.9%; Cabinet 41 [cabinet] 84.8%; Drawer Next To Bed 47 [drawer_next_to_bed] 83.7%

Correct category rank (within top-10): 6

Query (left) vs Incorrect Top Match (right):



Likely reason it failed: Both large boxy furniture; frontal product photos yield similar structural histograms. Correct category first appeared at rank #6. Top match score breakdown — line: 86.3%, affine: 100.0% (inliers=31), material: 69.5%.

7. Cross-Cutting Findings

- Overall top-1 category accuracy is low at 38.5% on held-out (non-duplicate) product photos.
- The engine is designed for near-duplicate / same-item matching under viewpoint change, not open-set furniture classification.
- Confidence threshold (45%) is too low for this dataset: 100% of queries exceeded it, including wrong matches often at 80–90%+.

- Line geometry weight (40%) favors any rectilinear furniture (cabinets, frames, sofas, drawers), driving many cross-category confusions.
- Worst categories: armchair (0%) and drawer_near_bed (0%). Best: makeup_chair, lamp, sofa.
- chair ↔ makeup_chair and cabinet ↔ drawer/frame confusions show the model lacks a semantic category head.
- Top-5 accuracy (69.2%) is much higher than top-1, suggesting ranking/discrimination is the bottleneck more than total feature failure.

8. Recommendations

8. Raise furniture confidence_threshold (e.g. 75–85%) and/or require margin between #1 and #2 scores.
9. Add a lightweight category classifier (or CLIP text-image prior) before fine-grained geometric matching.
10. Evaluate with same-item distorted queries (sandbox) to measure true duplicate-match performance separately from category retrieval.
11. Balance catalogue class counts or add hard-negative mining for cabinets/frames/drawers.
12. Reduce line_weight or normalize scores per category if category-level retrieval remains a goal.

9. Appendix — Full Result Table

#	Query	Expected	Top Match	Conf%	Result
1	armchair_46.jpg	armchair	Lamp 12	82.7	FAIL
2	armchair_47.jpg	armchair	Sofa 4	71.6	FAIL
3	armchair_48.jpg	armchair	Lamp 8	91.3	FAIL
4	carbinet_44.jpg	cabinet	Sofa 4	84.1	FAIL
5	carbinet_45.jpg	cabinet	Drawer Next To Bed 40	77.6	FAIL
6	carbinet_48.jpg	cabinet	Cabinet 43	94.5	PASS
7	carbinet_50.jpg	cabinet	Photo Frame 19	87.7	FAIL
8	chair_22.jpg	chair	Chair 20	100.0	PASS
9	chair_46.jpg	chair	Chair 4	87.5	PASS
10	chair_50.jpg	chair	Cabinet 41	85.7	FAIL
11	chair_6.jpg	chair	Makeup Chair 6	80.5	FAIL
12	chair_7.jpg	chair	Cabinet 43	83.0	FAIL

13	chair_9.jpg	chair	Sofa 1	81.5	FAIL
14	drawernearbed_11.jpg	drawer_near_bed	Photo Frame 15	86.7	FAIL
15	drawernearbed_5.jpg	drawer_near_bed	Photo Frame 19	92.4	FAIL
16	lamp_40.jpg	lamp	Lamp 52	88.5	PASS
17	lamp_41.jpg	lamp	Cabinet 3	69.3	FAIL
18	lamp_47.jpg	lamp	Lamp 52	82.1	PASS
19	makeupchair_2.jpg	makeup_chair	Makeup Chair 5	76.6	PASS
20	photoframe_12.jpg	photo_frame	Photo Frame 11	70.9	PASS
21	photoframe_16.jpg	photo_frame	Cabinet 4	91.5	FAIL
22	sofa_12.jpg	sofa	Sofa 4	90.4	PASS
23	sofa_15.jpg	sofa	Sofa 16	87.9	PASS
24	sofa_43.jpg	sofa	Photo Frame 18	81.9	FAIL
25	sofa_47.jpg	sofa	Cabinet 42	86.9	FAIL
26	sofa_7.jpg	sofa	Sofa 10	87.6	PASS

End of report. Previews are side-by-side (query | top match). Evaluation used FurnitureMatcher with cached catalogue features; no catalogue items were modified.