

Precision Hardening V2 Report

A reproducible before-and-after account assembled from the required v2 results JSON, available supporting benchmark JSONs, and verified workspace artifact hashes.

DECLARED VERDICT	PASS
CLAIM BOUNDARY	The pixelation claim applies only to the two supplied screenshots, block size 14, equal nominal region counts, and the recorded local runtime. The segmentation claim uses a synthetic eight-case corpus; its pre-hardening values are summary-only because raw baseline masks were not preserved. These measurements do not establish universal image quality, aesthetic preference, SAM quality, legal compliance, market leadership, or state-of-the-art performance.
EVIDENCE TIMESTAMP	2026-07-18T08:43:29.064Z
SCHEMA	mythiril.precision-hardening/v2
INPUT PROVENANCE	output/precision-hardening-v2/results.json SHA-256 ce8384c582b392c2bc48c025c6073110 57d876ba5112900d1044fd4fb8904aef

13 BEFORE / AFTER METRICS	19 SUPPORTING JSONS CONSUMED	33 PROVENANCE RECORDS	PASS INPUT-DECLARED OUTCOME
------------------------------	---------------------------------	--------------------------	--------------------------------

Before and After

Values are copied from the v2 results JSON. A delta is calculated only when both sides are numeric and the input does not supply one. A positive delta is not labeled an improvement unless the input explicitly supplies a status or interpretation.

METRIC	BEFORE	AFTER	DELTA	STATUS	INPUT NOTE
Strong-edge energy retention	9.04%	54.05%	+45.02 pp	IMPROVED	Average across two supplied screenshots; fixed grid versus adaptive BSP at identical per-image region counts.
Strong-edge correlation	8.90%	34.87%	+25.96 pp	IMPROVED	Average correlation over strong source-edge pixels on the same two-image corpus.
Mean absolute RGB error	4.69580	2.08625	-2.60955	IMPROVED	Lower is better; average across the same equal-budget comparison.
PSNR	22.07910 dB	27.02995 dB	+4.95085 dB	IMPROVED	Average across two screenshots; higher is better for source reconstruction fidelity.
Average of per-image median runtimes	679.7680 ms	1,499.6055 ms	+819.8375 ms	TRADEOFF	Adaptive BSP is approximately 2.21x slower on this local two-image run; progressive preview and workers remain important.
Rectangle-seeded segmentation macro IoU	52.16%	74.49%	+22.33 pp	IMPROVED	Before is a preserved summary-only pre-hardening metric; after is a raw rerun over eight synthetic cases.
Difficult-case macro IoU	45.33%	70.85%	+25.52 pp	IMPROVED	Seven difficult cases; before is summary-only and after is a raw rerun.
Same-color distractor IoU	0.00%	100.00%	+100.00 pp	CORRECTED	Normalized component scoring stopped the larger disconnected distractor from winning.
Boundary-touching target IoU	21.35%	100.00%	+78.65 pp	CORRECTED	Exterior-ring background modeling handles a target that occupies the ROI border.
Loose ambiguous ROI IoU	0.00%	0.00%	+0.00 pp	UNRESOLVED	Color and connectivity alone cannot identify the intended object when the rectangle supplies insufficient semantic evidence.
Cataloged algorithms integrated	33	36	+3	EXPANDED	Adds adaptive BSP pixelation and summed-area tables, and records Euclidean morphology as integrated in production refinement.
Cross-runtime renderer cases	15	16	+1	EXPANDED	The new adaptive BSP case is byte-identical across Node, Chrome, and Edge.
Passing automated tests	403	403	0	PASS	All 403 tests pass after the new compiler, API, governance, morphology, catalog, and conformance wiring.

Methodology and Decision Gates

Report assembly method: parse the required JSON without recomputing benchmark values, load every available allow-listed or explicitly declared supporting benchmark JSON, verify declared file hashes, embed only the checked DM Sans and Manrope font instances, build to a temporary PDF, validate the font and monochrome contracts, and atomically replace the final report.

#	SUPPLIED BENCHMARK METHOD
1	Pixelation: render each supplied screenshot with fixed-grid and variance-guided BSP pixelation at block size 14, preserve the same nominal and actual region count per image, run each variant three times, and compare median runtime plus source-reconstruction and strong-edge metrics.
2	Adaptive BSP: build alpha-weighted linear-luminance summed-area moments, repeatedly choose the axis-aligned split with the greatest deterministic weighted SSE reduction, and average each final block in linear light.
3	Foreground segmentation: evaluate eight deterministic synthetic fixtures, compare against the input rectangle, repeat each algorithm output five times, and measure IoU, precision, recall, F1, and boundary F1.
4	Baseline discipline: retain the five available pre-hardening segmentation values as summary-only evidence and do not invent missing baseline precision, recall, boundary, runtime, or raw-mask data.
5	Release verification: run lint, TypeScript, the full Vitest suite, engine canaries, Node/Chrome/Edge hash conformance, optimized Next.js build, live Azure vision probe, MCP closed-loop probe, desktop/mobile UI QA, and a dedicated adaptive-BSP export smoke.

GATE	THRESHOLD	MEASURED	DECLARED STATUS
Equal pixelation region budget	adaptive regions <= fixed-grid regions for every image	equal on 2 of 2 images	PASS
Per-image strong-edge wins	adaptive retention and correlation exceed fixed grid	2 of 2 images	PASS
Deterministic algorithm outputs	repeated outputs byte-identical	pixelation true; segmentation 5 repeats per case true	PASS
Hostile segmentation macro IoU	>= 0.70	0.744925	PASS
Full automated suite	zero failures	403 of 403 tests	PASS
Cross-runtime renderer conformance	exact aggregate and per-case hashes	16 of 16 in Node, Chrome, and Edge	PASS
Optimized production build	Next.js build exits 0	18 routes generated or traced	PASS
Desktop and mobile product QA	core flows complete with no browser or layout errors	6 route-view checks; 0 errors	PASS
Adaptive BSP live UI smoke	adaptive-bsp selected and export enabled	passed on optimized localhost build	PASS

GATE	THRESHOLD	MEASURED	DECLARED STATUS
Agent interface probe	MCP closed loop succeeds	12 tools; probe ok	PASS
Live AI provider probe	strict structured vision response accepted	Azure gpt-4o passed	PASS

Bounded Claims and Limitations

DECLARED CLAIM	On the two supplied screenshots at an equal region budget, adaptive BSP pixelation improved average strong-edge energy retention by 0.450166, strong-edge correlation by 0.259648, reduced mean absolute RGB error by 2.60955, and raised PSNR by 4.95085 dB. On the synthetic eight-case rectangle-seeded corpus, the hardened estimator raised macro IoU from a summary-only 0.521608 baseline to a raw-rerun 0.744925, while the loose ambiguous ROI remained unsolved.
BOUNDARY	The pixelation claim applies only to the two supplied screenshots, block size 14, equal nominal region counts, and the recorded local runtime. The segmentation claim uses a synthetic eight-case corpus; its pre-hardening values are summary-only because raw baseline masks were not preserved. These measurements do not establish universal image quality, aesthetic preference, SAM quality, legal compliance, market leadership, or state-of-the-art performance.
VERDICT SOURCE	results.json declares PASS.
REPORT BOUNDARY	This document does not infer market leadership, universal accuracy, legal compliance, provider readiness, or state-of-the-art performance. It reports only supplied measurements and verified provenance.

#	LIMITATION
1	The adaptive pixelation benchmark contains two UI-heavy screenshots, not a broad photographic, illustration, typography, sprite, and low-resolution corpus.
2	Adaptive BSP averaged approximately 2.21 times the fixed-grid median runtime in this rerun. It is a quality mode, not yet the universal instant-preview path.
3	The pre-hardening segmentation baseline is summary-only. Raw masks and several baseline metrics were not preserved, so those missing values are not reconstructed.
4	The loose ambiguous rectangle remains at 0 IoU. A semantic prompt, additional positive/negative points, or a configured SAM provider is required when local color and connectivity evidence is non-identifying.
5	SAM 3 and SAM 2 adapters are implemented and tested, but no live SAM request was verified because FAL_API_KEY and REPLICATE_API_TOKEN are absent.
6	Firefox and WebKit were not installed; exact cross-runtime proof currently covers Node, Google Chrome, and Microsoft Edge.
7	Four catalog entries remain explicitly approximate, including the manual rectangle fallback, local rectangle-seeded estimator, SAM 2 prompt approximation, and bounded English OCR rules.
8	Secure Redact is a separate high-assurance destructive export lane. The larger product also compiles transparent pixel assets, product graphics, and contained object effects.

#	LIMITATION
9	No benchmark in this iteration proves human aesthetic preference, universal design quality, Figma or Adobe adoption, legal compliance, market superiority, or state-of-the-art status.
10	Universal segmentation accuracy
11	Live SAM quality or latency
12	Aesthetic superiority over generative image systems
13	Market leadership or acquisition readiness
14	Irreversible deletion of all copies or contextual inference

04 / SUPPORTING EVIDENCE

Benchmark JSON Inventory

Every row below was parsed as JSON and hashed at report-generation time. Presence in this table is provenance, not endorsement: the declared result column is shown only when the source JSON contains an explicit boolean pass/fail field.

PATH	SCHEMA	TIMESTAMP	DECLARED RESULT	SHA-256
output/adaptive-pixelation-benchmark/results.json	mythiril.adaptive-pixelation-benchmark/v1	2026-07-18T05:10:54.566Z	not declared	59ffb56426eb5fad78667e91e11531881204664e6ba4231d82f6b4f4cf89b6c9
output/foreground-segmentation-benchmark/baseline-results.json	mythiril.foreground-segmentation-baseline-summary/v1	not supplied	not declared	0a1077863e047d1d67272560159da13f7d6adb040ee840bd9ff04b9dd450dc73
output/foreground-segmentation-benchmark/results.json	1	not supplied	not declared	2c24eebb9700db1592924ba8f30ab804cf9468605b655abee1a01642fd0392c6
output/mythiril-algorithm-conformance.json	mythiril.algorithm-conformance/v1	2026-07-18T08:18:45.404Z	not declared	f983af4bbdd3f270df1c6375895da59b048afeca133f75acab05e44f1f265e2c
output/mythiril-conformance.json	mythiril.runtime-conformance-report/v1	2026-07-18T08:17:15.124Z	PASS	9856a3da60de90ab7c662900167a117027202843f322de8781d792382105f72f
output/mythiril-evidence.json	5	2026-07-16T16:15:55.860Z	not declared	fc0c86969d0c93f60876dc3395e55457e200478c5ba043b8ef6a976e9e2fee09
output/mythiril-mcp-probe.json	not supplied	not supplied	not declared	99869a1115dd084e5b453e4b3e88b4564c872e49de6ffa4e5b781c69ca548312
output/mythiril-outcome-selection-benchmark.json	1	not supplied	not declared	5c1f4261d85e39e44ef61343ea4d40144359ebea72b502e4bb13dc03c629bcb
output/mythiril-pipeline-benchmark.json	1	2026-07-16T16:09:01.709Z	not declared	2db35a1dcbc40bebb5d856d23ebf66a1f4d12ce85054cc4a88b10dd87484c200
output/mythiril-provider-probe.json	1	2026-07-18T08:35:21.541Z	not declared	b515cd0b016c8146d965cd3cdac2fbb21924c5a1a647867e937e47b9214ebd3e
output/pdf-redaction-hardening/results.json	mythiril.pdf-redaction-precision-hardening/v1	2026-07-17T09:55:29.264480+00:00	PASS	145aec2edf7dca24c3e18479dc798f1ed9aa5918aab627d9f14d73524abc299c

PATH	SCHEMA	TIMESTAMP	DECLARED RESULT	SHA-256
output/pdf-redaction-live-test/results-matrix.json	nightowl.pdf-redaction-live-evidence/v1	2026-07-17T09:20:42.419986Z	PASS	7e5888d251942a8983383f33f6e73936d4c09b5a82f5608066f7f4c46ad62db7
output/pdf-redaction-online-sample/results.json	nightowl.online-pdf-live-test/v1	2026-07-17T04:21:15.950492Z	PASS	2e315b95eaf7b2190d5fe4883ddf970e6dd44026564da536a80f0296d4ed3ba1
output/playwright/product-ui-qa.json	not supplied	not supplied	PASS	4e278f7cc6b951893c6017c2100d9b559be5aa7a33d1c7cdbb8c1d9c0f805206
output/precision-hardening-v2/adaptive-ui-smoke.json	not supplied	not supplied	PASS	fd2481c32c261a5627eb10290025e3f4b7ee943843e298389f729e77e7cca5b8
output/precision-hardening-v2/vitest-results.json	not supplied	not supplied	not declared	01589ff7421f852745f7c3e6e9387f462549aee9b9758417b089d12c87a40fa5
output/redaction-detection-benchmark/baseline-v1.json	mythiril.local-redaction-detection-benchmark/v1	2026-07-17T08:49:24.453Z	FAIL	de1d3dd3edab4ac7e9f0010dad9d71840b063b1411aee43a1cd27f115cc38ad
output/redaction-detection-benchmark/results.json	mythiril.local-redaction-detection-benchmark/v1	2026-07-17T09:27:08.769Z	PASS	4d171a291407cabf807b537c33de3541bd94de980236c3ab251ad575353adbfe
output/user-screenshot-tests/results.json	mythiril.user-screenshot-test/v1	2026-07-16T11:47:39.946Z	not declared	9fb4014951e47fec0b55d2d7a1a588de8117db9c88be9a6e0f0b7cbc05ee5379

05 / PROVENANCE

Artifact Hashes and Run Identity

ROLE	PATH	BYTES	STATUS	SHA-256
adaptive BSP implementation	src/lib/engine/adaptive-pixelation.ts	12,013	verified	9baee0e6f8c94949cb81577aacc3f8332ac0787861a66aa75c8f24f0b477aba5
adaptive pixelation benchmark	output/adaptive-pixelation-benchmark/results.json	5,640	verified	59ffb56426eb5fad78667e91e11531881204664e6ba4231d82f6b4f4cf89b6c9
adaptive screenshot 151814	output/adaptive-pixelation-benchmark/screenshot-2026-07-08-151814-adaptive-bsp-14.png	238,069	verified	d4f2753853177a76771a734e323d7f913e2bd75311521bc2f2f3728d07a16851
adaptive screenshot 164113	output/adaptive-pixelation-benchmark/screenshot-2026-07-08-164113-adaptive-bsp-14.png	67,227	verified	ca42fe0ad33b202d71a7c8db7eb691290d529ecadb0093003e39053faaf9b8e
adaptive UI proof screenshot	output/playwright/adaptive-structure-aware-desktop-final.png	382,680	verified	eb0893ce31994dc7b7ddc17ab41349608945436c6e6754df0f77c49cd9e6c87e
adaptive UI smoke result	output/precision-hardening-v2/adaptive-ui-smoke.json	239	verified	fd2481c32c261a5627eb10290025e3f4b7ee943843e298389f729e77e7cca5b8
agent compiler integration	src/lib/agent/design-compiler.ts	42,876	verified	b1ceb2940067e1d82abc6ebd1435ac2ee346033c81f3b98463eb4c1f876bbbbc
algorithm catalog conformance	output/mythiril-algorithm-conformance.json	129,930	verified	f983af4bbdd3f270df1c6375895da59b048afeca133f75acab05e44f1f265e2c
algorithm inventory	src/lib/agent/algorithm-catalog.ts	29,921	verified	6b1bf1025064b57510c096e8deac6feec61b12ae04c9732cec33fba3cae3270
cross-runtime golden corpus	src/lib/agent/conformance-corpus.ts	8,654	verified	9a9cc7f63c47635acab57366d86699b3f822ca00c85ae35608feed74d9178f92

ROLE	PATH	BYTES	STATUS	SHA-256
cross-runtime renderer conformance	output/mythiril-conformance.json	7,542	verified	9856a3da60de90ab7c662900167a117027202843f322de8781d792382105f72f
desktop and mobile product QA	output/playwright/product-ui-qa.json	1,175	verified	4e278f7cc6b951893c6017c2100d9b559be5aa7a33d1c7cddb8c1d9c0f805206
fixed-grid screenshot 151814	output/adaptive-pixelation-benchmark/screenshot-2026-07-08-151814-fixed-grid-14.png	199,022	verified	4583ef8c8edef1957f1d52cbf1cf7c1f1ccd6e1201802e082676d87d3af12f65
fixed-grid screenshot 164113	output/adaptive-pixelation-benchmark/screenshot-2026-07-08-164113-fixed-grid-14.png	34,824	verified	3f6d958a519523528287d7528afd5c017d51750e9df434f8dca46c8261afc580
full automated test report	output/precision-hardening-v2/vitest-results.json	140,653	verified	01589ff7421f852745f7c3e6e9387f462549aee9b9758417b089d12c87a40fa5
hostile segmentation rerun	output/foreground-segmentation-benchmark/results.json	28,964	verified	2c24eebb9700db1592924ba8f30ab804cf9468605b655abee1a01642fd0392c6
live provider probe	output/mythiril-provider-probe.json	2,175	verified	b515cd0b016c8146d965cd3cdac2fbb21924c5a1a647867e937e47b9214ebd3e
MCP closed-loop probe	output/mythiril-mcp-probe.json	47,443	verified	99869a1115dd084e5b453e4b3e88b4564c872e49de6ffa4e5b781c69ca548312
production mask-refinement integration	src/lib/engine/mask-refinement.ts	21,221	verified	929908ae428f164c65e99a016509372534015b9f852f1121dc379d6bbb45da77
rectangle-seeded segmentation implementation	src/lib/engine/foreground-segmentation.ts	28,318	verified	8e6dd104a39e406f6d7c18452f39717c1b2407dc0974c89d860a940a50e32e02
report generator source	scripts/report/generate_precision_hardening_v2_report.py	54,014	verified	a5072c0dd7a677571835ab67eb0ab9be58653d18710562c0a202fa22d897adcd
required precision-hardening v2 results	output/precision-hardening-v2/results.json	15,300	verified	ce8384c582b392c2bc48c025c607311057d876ba5112900d1044fd4fb8904aef
summary-only segmentation baseline	output/foreground-segmentation-benchmark/baseline-results.json	1,158	verified	0a1077863e047d1d67727560159da13f7d6adb040ee840bd9ff04b9dd450dc73
summed-area implementation	src/lib/engine/summed-area.ts	4,888	verified	3b1638695a941d9d918ce7c14afa187f856d8eb989f7ea7438cdba0082e27510
supporting benchmark JSON	output/mythiril-evidence.json	295,935	verified	fc0c86969d0c93f60876dc3395e55457e200478c5ba043b8ef6a976e9e2fee09
supporting benchmark JSON	output/mythiril-outcome-selection-benchmark.json	32,562	verified	5c1f4261d85e39e44ef61343ea4d40144359eaea72b502e4bb13dc03c629bcbb
supporting benchmark JSON	output/mythiril-pipeline-benchmark.json	5,659	verified	2db35a1dcbc40bebb5d856d23ebf66a1f4d12ce85054cc4a88b10dd87484c200
supporting benchmark JSON	output/pdf-redaction-hardening/results.json	195,491	verified	145aec2edf7dca24c3e18479dc798f1ed9aa5918aab627d9f14d73524abc299c
supporting benchmark JSON	output/pdf-redaction-live-test/results-matrix.json	75,382	verified	7e5888d251942a8983383f33f6e73936d4c09b5a82f5608066f7f4c46ad62db7
supporting benchmark JSON	output/pdf-redaction-online-sample/results.json	16,149	verified	2e315b95eaf7b2190d5fe4883ddf970e6dd44026564da536a80f0296d4ed3ba1
supporting benchmark JSON	output/redaction-detection-benchmark/baseline-v1.json	69,520	verified	de1d3dd3edab4ac7e9f0010dad9d71840b063b1411aee43a1cd27f115cc38ad
supporting benchmark JSON	output/redaction-detection-benchmark/results.json	90,964	verified	4d171a291407cabf807b537c33de3541bd94de980236c3ab251ad575353adbfe

ROLE	PATH	BYTES	STATUS	SHA-256
supporting benchmark JSON	output/user-screenshot-tests/results.json	8,691	verified	9fb4014951e47fec0b55d2d7a1a588de8117db9c88be9a6e0f0b7cbc05ee5379

FIELD	VALUE
Evidence generated	2026-07-18T08:43:29.064Z
Report generated	2026-07-18T08:55:47.734365Z
Schema	mythiril.precision-hardening/v2
Git head	31a4c7c66a8b7bfe580ed7d4499a60ed3a9e807c
Git dirty	True
Git status SHA-256	2a0bc5fd832fe67b71198b1b7794edb34a15557a52730f24d94c9ac4220ec033
Required input	output/precision-hardening-v2/results.json
Report output	output/pdf/mythiril-1-precision-hardening-v2-report.pdf
Font source	output/report-fonts: DM Sans Regular/Bold and Manrope Regular/Bold

The final PDF SHA-256 cannot be embedded inside the PDF without changing the file being hashed. The generator prints that hash after atomic publication. Re-run with the same input artifacts to reproduce the report content; PDF byte identity may still depend on the installed ReportLab stack.