Bilegt Gankhuyag

Opponent's review & evaluation (based on opponensi_lap_en.pdf)

Thesis: *Image Compression Recommendation Under Perceptual-Quality Constraints – A Multi-Criteria Decision-Support Web Application with Cross-Method Validation* (*Gankhuyag.pdf*)

Executive verdict

This is a **strong, engineering-grade, reproducible applied thesis** with a clear problem statement, a deployed artifact, and an unusually explicit validation/triangulation layer (TOPSIS + COCO Y0 + LLM benchmark + sensitivity sweep).

However, it also contains **non-trivial internal inconsistencies** (decision-space description vs corpus strategies; sensitivity "plateau" interpretation), plus **formal issues** (unresolved "[ref]" placeholders; occasional language slips). These reduce confidence in parts of the written argumentation even though the implementation narrative is convincing.

---

1) Use of AI & suspected plagiarism (separate major criterion)

AI use (disclosure & appropriateness)

- The thesis **explicitly discloses** LLM usage and separates **writing-support** from **benchmarking as verification**, with annexed conversations/prompts and claimed traceability.
- The LLM benchmark design (four prompts: rule design, per-image recommendation, critique, validation methodology) is described as **non-ground truth "independent reference reasoner"** and is treated cautiously.

Assessment: AI use is **transparent and methodologically integrated**, not hidden. That is a net positive.

Suspected plagiarism

- From the text available here, there is **no direct evidence of plagiarism** (the work provides citations, structured references, and explicit AI disclosure).
- **Important limitation:** I cannot run Turnitin/Urkund or source-code similarity checks in this environment. Therefore this is **not a formal plagiarism clearance**—only a reasoned "no obvious red flags in-document."

Assessment: *No clear signs of plagiarism*, but recommend institutional similarity screening as standard.

---

2) Scoring by the opponensi criteria (max 200)

Summary table (recommended points)

	Criterion	Max	Score	Rationale (short)
1	Topic & Objectives	40	**36**	Clear aims, RQs, target groups, measurable outputs; strong alignment with CT_00/BPROF framing.

2	Review of the Literature	40	**34**	Broad and relevant coverage (formats, metrics, MCDA, TOPSIS, COCO, GDPR, sensitivity); but some citation/formal weaknesses (placeholders).
3	Presentation of Author's Own Work	60	**52**	Substantial implementation: pipeline, metrics, Pareto, rule, TOPSIS, COCO (local+external), sensitivity, testing, deployment; but internal inconsistencies lower trust in reporting.
4	Structure, Style, Formal Aspects	40	**32**	Overall strong structure and annex discipline; reduced by unresolved "[ref]", repeated words, and at least one contradiction in results interpretation.
5	Overall Impression	20	**16**	Solid, defendable applied thesis with honest limitations; would be excellent with tighter editorial consistency and corrected contradictions.
6	**Total**	**200**	**170**	**Good** (160–179) per the provided grading scale.

Final grade band: Good (170/200)

Detailed justification by criterion

A) Topic and Objectives (36/40)

Strengths

- The topic is **relevant and well-scoped**: image compression recommendation under perceptual constraints for web use, framed as a multi-criteria decision problem.
- Objectives are **explicit, measurable, and traceable** (pipeline, decision rule, cross-method verification, sensitivity analysis, deployment/testing).
- The thesis clearly defines RQs and expected outcomes; also sets an "interpretation boundary" (what claims do/do not mean).

Why not full points

- A key **scope/description consistency** problem appears later (see "Own work"): early decision space vs later corpus strategies differ, weakening objective-to-implementation clarity.

B) Review of the Literature (34/40)

Strengths

- Coverage is **comprehensive and structured**: JPEG/WebP/AVIF, PSNR/SSIM, MCDA, Pareto, TOPSIS, COCO Y0, sensitivity analysis, LLM benchmarking, GDPR.
- The literature is not only listed but **operationalized**: each concept is linked to a later implementation chapter.
- Reference classification/discipline (T01–T16 matrix) and cross-thesis comparison are consistent with the program's expectations.

Weaknesses

- There are **unresolved citation placeholders** ("[ref]") in multiple places. This is a formal academic issue: it prevents the reader from verifying claims.
- Some quantitative claims (e.g., "citation count >40,000", "browser support %") are presented, but the *exact* sources/URLs are not always visible in the extracted text; this needs tightening for auditability.

C) Presentation of the Author's Own Work (52/60)

Major strengths (this is the thesis's core)

- A complete, end-to-end **implemented pipeline** is described: compression variant generation, metric measurement, Pareto filtering, lexicographic decision rule, and cross-checks (TOPSIS, COCO Y0 local + external workflow, LLM benchmark).
- Evidence orientation: corpus run, method agreement tables, sensitivity sweep (0.80–0.99), testing protocol with pass conditions, and a reproducibility package are all stated.
- Deployment is described (FastAPI + Docker + Render), plus security/privacy posture (input validation, ephemeral storage, GDPR framing, manual boundary for external solver).

Key weaknesses reducing points

1. **Decision-space inconsistency**
   - Early: the "five compression strategies" are listed as **JPEG q90/q70, WebP q80/q60, AVIF q50**.
   - Later corpus results: strategies include **jpeg_q50/q70/q85/q95, webp_q50/q75/q90, plus png_lossless** (with PNG excluded from rule evaluation).

This isn't a small typo—it affects reproducibility and interpretation (what exactly is the system recommending?).

2. **Sensitivity plateau interpretation contradiction**
   - The table shows agreement at **0.94 = 70.8%** and **0.96 = 79.2%**, implying **only 0.95** is ≥80%.
   - Yet the text also talks about a plateau roughly **[0.94, 0.96]** at ≥80% in one place, which conflicts with the shown values.

This should be corrected because it directly affects the strength of the threshold defense.

3. **Validation architecture caveat**
   - Local COCO Y0 is explicitly an **OLS approximation** and not the real MIAU solver; that's honest, but it limits the "formal-method agreement" strength unless external MIAU runs are systematically reported.

D) Structure, Style, and Formal Aspects (32/40)

Strengths

- Clear CT_00/BPROF-like structure: aims/tasks, literature, own development, discussion, conclusions, future work, annexes.
- Lots of "audit-friendly" touches: explicit units, labeled figures/tables, testing counts, explicit limitations, and annexed materials for reproducibility and AI use.

Weaknesses

- Unresolved citation placeholders ("[ref]")—formal defect.
- Minor language issues (e.g., duplicated "websites websites") contradict the "zero spelling errors" claim and reduce polish.
- The inconsistencies noted above are also **formal/structural quality issues** (internal coherence).

E) Overall Impression (16/20)

This thesis is **not just a write-up**; it is a **working, deployed decision-support system** with a strong "evidence pipeline" mindset (tests, reproducibility, structured validation methods).

The author also reports negative findings honestly (e.g., the primary rule being an outlier vs TOPSIS/COCO; threshold being load-bearing).

To reach "Excellent", it needs an editorial pass to eliminate contradictions and finalize references.

Concrete improvement list (high impact, fast fixes)

1. **Unify the decision space** (one canonical set of variants everywhere): abstract, methods, corpus, UI, and results must match.
2. **Fix the sensitivity plateau narrative** to match the table; recompute or correct the threshold-robustness claim.
3. **Replace all "[ref]" placeholders** with full citations (author/year + URL/DOI in the chosen style).
4. Add a short **"replication recipe"** in the main text (not only annex): command(s) to reproduce corpus CSVs/figures and where outputs appear.
5. If available, include **at least a small set of external MIAU COCO runs** (not only local approximation) systematically for the corpus (even 10/24 would strengthen credibility).

Suggested defense questions (to test understanding)

1. *Why lexicographic (threshold + min size) rather than weighted utility—what user preference is encoded, and who is it for?*
2. *How would your recommendation change if SSIM were replaced by LPIPS/DISTS (conceptually, not necessarily implemented)?*
3. *Explain the disagreement pattern between your rule and TOPSIS/COCO as "structural." What exactly differs in preference encoding?*
4. *What's your strongest evidence that the tool is "correct" given no formal perceptual study? Where are the limits?*

Final recommended score

170 / 200 → Good

If you want, I can also draft the **filled opponent form text** (the "General summary" + category paragraphs) in the exact tone/length you typically use for KJE defenses, based strictly on these findings.