Does the Source of Carrier Image Affect Steganographic Detectability?

Full Midway Proposal

Abdul Moiz Akbar | Malo Coquin | Daria Gjonbalaj | Nico Muller-Spath
Jimena Narvaez del Cid | David Wicker | Nikolas Zouros

Department of Advanced Computing Sciences
Maastricht University

Project 2.2 | February 2026

Agenda

1. Motivation and problem statement
2. Research questions and verification
3. Chosen approaches
4. Experiments and validation
5. Prototype status
6. Related work positioning
7. Relation to curriculum
8. Planning and passing requirements

Motivation and Problem Statement

- Image steganography detectability depends on carrier statistics, not only embedding logic (Petitcolas et al., 1999; Cheddad et al., 2010; Fridrich & Kodovsky, 2012).
- Most steganalysis benchmarks assume camera photos with familiar noise/compression traces.
- Modern generators (Stable Diffusion, StyleGAN3) produce photorealistic images from different processes (Rombach et al., 2022; Karras et al., 2021).
- Synthetic images have measurable statistical fingerprints that may alter detector behavior (Wang et al., 2020; Corvi et al., 2023).

Central Problem

Do steganalysis methods designed and validated on photographs remain effective when the carrier is ML-generated?

Closest prior work

De et al. (2022) show AI-generated image steganography, but not a controlled real-vs-ML comparison with standardized LSB/DCT detectors.

Why This Study Matters

Security

- If synthetic carriers are harder to detect, attackers gain an easy evasion path.
- If easier, defenders get a concrete screening advantage.

Scientific gap

- Interaction between generative-model distributions and embedding distortion is largely unexplored.
- Need controlled experiments isolating carrier origin.

Practical relevance

- AI images are now common in social and communication channels.
- Practitioners need evidence on retraining/adaptation requirements.

Study scope: 2 x 2 x 3 x 2 factorial design over 1,000 images, with CPU-feasible execution in 7 weeks.

Research Questions

RQ1: Carrier origin

Does carrier-image origin (real vs ML-generated) affect detectability under identical settings?

RQ2: Payload

Does increasing payload size widen the real-vs-ML detectability gap?

RQ3: Encryption

Does encrypting payload before embedding change detectability, and does origin modify that effect?

RQ4: Embedding method

Do LSB and DCT interact differently with carrier origin in terms of detectability?

RQ5: Image quality

How are PSNR/SSIM/FSIM affected by embedding method and payload size?

Verification Criteria by RQ

RQ	Verification target	Primary analysis
RQ1	Real-vs-ML AUC difference with fixed settings.	RS and SRM+FLD + significance/effect size
RQ2	AUC-gap trend across Low/Medium/High payload.	Trend test + carrier×payload interaction
RQ3	Plain vs AES-256-CBC detectability difference.	Condition-wise AUC comparisons by method and origin
RQ4	Carrier-origin dependence on LSB vs DCT.	Two-way ANOVA interaction term
RQ5	Quality stability across conditions.	PSNR/SSIM/FSIM against target thresholds

Common statistical settings

We evaluate performance primarily with ROC-AUC and report uncertainty and effect size alongside significance, using a stricter threshold to account for multiple comparisons.

Chosen Approach: Factorial Design and Conditions

Design matrix

- **Carrier type (2):** Real, ML-generated
- **Embedding (2):** Spatial LSB, frequency-domain DCT-QIM
- **Payload (3):** Low, Medium, High
- **Detectors (2 main):** RS Analysis, SRM+FLD

Sample size

- 500 real images
- 500 ML-generated images
- 1,000 total carriers
- Full condition coverage

Controlled variable principle

Carrier origin is treated as the central independent variable; image size/format, embedding pipelines, payload levels, and detector protocols are standardized across conditions.

Datasets and Preprocessing

Real photographs (500)

- RAISE: 250 images (RAW-derived forensic-quality baseline)
- COCO: 150 images
- Flickr30k: 100 images

Sources: Dang-Nguyen et al. (2015), Lin et al. (2014), Young et al. (2014)

ML-generated images (500)

- Stable Diffusion v2.1: 250 images
- StyleGAN3: 250 images
- Prompts aligned to COCO/Flickr semantics

Sources: Rombach et al. (2022), Karras et al. (2021)

Normalization and quality gate

All images normalized to 512x512 RGB 8-bit PNG. BRISQUE <= 50 filter for generated outputs to exclude low-quality artifacts.

Embedding Methods

LSB substitution (spatial)

- PRNG-keyed pixel/channel selection
- k = 1 for low and medium payload
- k = 2 for high payload
- Optional AES-256-CBC payload encryption

DCT-QIM (frequency)

- 8x8 block DCT per channel
- Mid-frequency zigzag coefficients (10–54)
- QIM embedding:

$$C_i' = \Delta \cdot \mathrm{round}(C_i/\Delta) \pm \Delta/4$$

Method rationale

LSB and DCT represent the two canonical embedding domains, allowing direct tests of method-origin interactions under controlled payload settings.

Payload and Encryption Conditions

Level	Approx. bpp	LSB setting	Purpose
Low	≈ 0.08	$k = 1$ sparse mask	Near-threshold detectability
Medium	≈ 0.16	$k = 1$ denser mask	Baseline operating point
High	≈ 0.32	$k = 2$	Stress-test detector sensitivity

Encryption condition

Each payload level is tested in:

- Plain payload mode
- AES-256-CBC pre-encrypted mode

Why include encryption?

It isolates whether message-bit structure contributes to detectability beyond carrier-level distortion.

Steganalysis Detectors and Validation Metrics

Detector set

- **RS Analysis** (training-free statistical baseline)
- **SRM+FLD** (feature-based classical ML detector)
- χ^2 **attack** as supplementary LSB check

References: Fridrich et al. (2001), Westfeld and Pfitzmann (1999), Fridrich and Kodovsky (2012)

Primary metrics

- ROC-AUC (primary)
- Accuracy at Youden's J
- Equal Error Rate
- FPR at 5% FNR

Statistical analysis plan

Two-way ANOVA (carrier x method; payload covariate), Wilcoxon pairwise tests, effect sizes, and Bonferroni-adjusted significance threshold.

Experiment Plan I (RQ1 and RQ2)

Exp. 1 – Carrier origin effect (RQ1)

- Apply RS and SRM+FLD across all payload levels and methods.
- Compare AUC for real vs ML-generated carriers.
- Decide with Wilcoxon significance and effect size.

Exp. 2 – Payload sensitivity (RQ2)

- Track real-vs-ML AUC gap across Low, Medium, High payload.
- Test monotonic trend (Spearman) and carrier x payload interaction.
- Decide whether payload amplifies origin-dependent detectability.

Experiment Plan II (RQ3 and RQ4)

Exp. 3 – Encryption effect (RQ3)

- Compare plain vs AES-encrypted payload AUC per carrier and method.
- Test whether encryption effect differs by carrier origin.

Exp. 4 – Method interaction (RQ4)

- Run two-way ANOVA on SRM AUC with carrier origin and method factors.
- Confirm whether the detectability gap depends on LSB vs DCT.

Experiment Plan III (RQ5)

Exp. 5 – Image quality (RQ5)

- Compute PSNR, SSIM, and FSIM for each stego condition versus cover.
- Check whether quality remains within target ranges across methods and payload levels.

Interpretation policy

Null findings remain valid outcomes; all major results reported with confidence intervals and effect sizes.

Prototype Status

Vertical prototype (algorithm depth)

Core components validated on a small test subset:

- LSB embedding/extraction
- DCT-QIM embedding
- RS Analysis
- SRM feature extraction/classification

Horizontal prototype (integration breadth)

Integrated end-to-end run on a mixed 50-image subset:

- 25 real + 25 ML-generated
- Medium payload setting
- Interface and output checks before full 1,000-image run

Readiness: Prototype checks reduce integration risk before full-scale execution.

Related Work Landscape

Classical embedding and detection foundations

Prior literature centers on post-hoc embedding in photographic carriers (LSB and DCT-QIM) and classical detection baselines (RS, χ^2, SRM).

Generative and coverless steganography

Recent approaches embed during generation or by content selection/generation rather than post-hoc modification. Our design instead uses ML-generated images as passive carriers.

Closest prior work

De et al. (2022) demonstrates AI-generated-image secret sharing, but not a controlled real-vs-ML comparison under identical LSB/DCT pipelines with ROC-AUC detectability endpoints.

Related Work: Our Positioning

Cross-domain and synthetic-image forensics

- Camera-domain shifts have been studied in steganalysis.
- Real-vs-synthetic shifts are broader because generation processes differ.
- Synthetic-image forensics reports distinct traces in generated images.

Classical vs deep steganalysis

- Deep methods can achieve strong accuracy.
- Classical SRM+FLD + RS chosen for interpretability and CPU feasibility.
- Matches cryptography/steganography project scope.

Positioning: Controlled evaluation of how carrier origin changes detectability under matched embedding conditions.

Relation to Curriculum

Cryptography and Steganography

LSB and DCT-QIM embedding, payload encryption (AES-256-CBC), and detectability-focused reasoning.

Machine Learning

Feature-based steganalysis with SRM representations and FLD-style linear classification.

Research Methods

Factorial design, hypothesis testing, ANOVA/Wilcoxon analysis, effect sizes, and controlled significance correction.

Algorithm Design and Implementation

Block-wise transforms, coefficient-level embedding logic, and reproducible experiment orchestration in Python.

Planning

Phase windows

- **Phase 2 (Implementation):** 30 Mar – 15 May 2026
- **Phase 3 (Completion):** 25 May – 12 Jun 2026

Phase 2 workstreams

- Dataset construction and ML generation (Weeks 1–2)
- LSB/DCT/AES pipeline implementation (Weeks 2–3)
- Detection, analysis, and writing (Weeks 3–7)

Phase 3 completion focus

- Finish remaining implementation
- Verify and rerun experiments
- Finalize slides, poster, and paper deliverables

Minimal Passing Requirements

Approach minimum

- One encryption algorithm
- Two embedding methods total
- One spatial-domain and one frequency-domain method

RQ minimum

- At minimum, answer:
- **RQ3 (Encryption)**
- **RQ4 (Embedding-method interaction)**

Definition of done

Minimum deliverable is a reproducible result set satisfying the above approach and RQ thresholds.

Thank you

Questions and Discussion

Midway proposal document: `docs/proposals/midway_proposal_final.tex`