Technical Report: AI-Generated Real Estate Image Detection

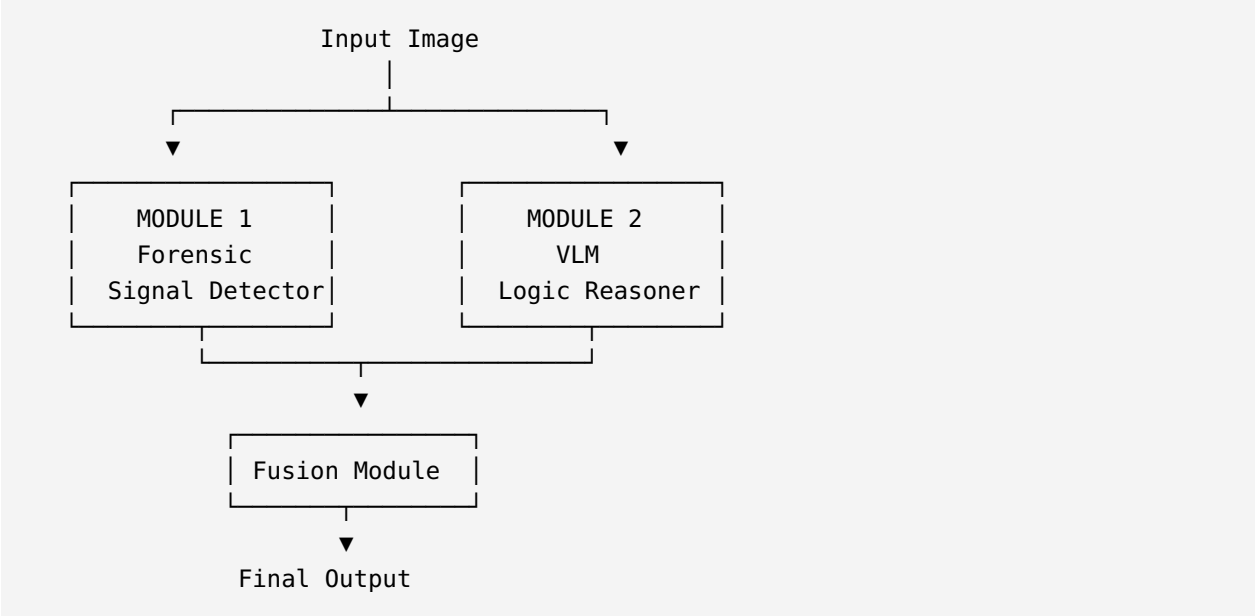
Digital Integrity Challenge - Track B: Real Estate & Commercial Integrity

---

1. Introduction

As generative AI becomes mainstream, detecting manipulated real estate images is critical for protecting buyers from deceptive virtual staging and AI-generated property photos. This report presents our dual-module system combining pixel-level forensic analysis with semantic VLM reasoning.

System Overview:



---

2. Module 1: Forensic Signal Detector (Pixel-Level)

The Forensic Signal Detector identifies low-level technical anomalies through three primary analysis techniques:

2.1 Texture Consistency Analysis

Objective: Detect "unnatural smoothness" in wall textures, floors, and surfaces.

- **Local Variance Analysis:** Computes variance across 16×16 pixel patches to identify unnaturally uniform regions
- **Laplacian Uniformity:** Measures texture detail distribution across the image

- **Rich/Poor Texture Contrast:** Compares high-detail regions with low-detail regions — AI images show less natural contrast

2.2 Compression Discrepancy Analysis

Objective: Identify if objects (furniture, fixtures) were digitally "spliced" into a scene.

- **Error Level Analysis (ELA):** Recompresses image at quality 90 and measures pixel-wise differences; spliced regions show different error levels
- **DCT Block Analysis:** Examines JPEG compression artifacts for inconsistencies
- **Double Compression Detection:** Identifies images saved multiple times with different quality settings

2.3 Frequency Domain Analysis (FFT)

Objective: Find the mathematical "fingerprint" left by AI upscalers and generators.

- **Periodic Artifact Detection:** Scans for artifacts at periods 2, 4, 8, and 16 pixels — characteristic signatures of diffusion models
- **Mid-High Frequency Analysis:** AI-generated images show distinct patterns in mid-to-high frequency bands

2.4 Additional Forensic Signals

Signal	Purpose	Weight
Noise Analysis	Detect unnatural noise patterns	18%
Sharpness	Identify over-sharpening or blur	16%
Texture Consistency	Unnatural smoothness	16%
FFT Analysis	Frequency fingerprints	15%
ELA	Compression discrepancies	12%
Color Consistency	Channel correlation anomalies	6%

---

3. Module 2: VLM Logic Reasoner (Semantic-Level)

The VLM Logic Reasoner provides "human-in-the-loop" style reasoning using **Qwen2-VL** (72B/7B with automatic fallback).

3.1 Physics Check

- Do furniture shadows match the apparent light source direction?
- Are window light patterns consistent with shadow angles?
- Do reflective surfaces show physically accurate reflections?

3.2 Structural Integrity Check

- Do cabinets merge unnaturally into walls?
- Are room proportions architecturally plausible?
- Is furniture scale consistent with room dimensions?

3.3 Natural Language Explanation

The VLM outputs a 2-sentence explanation of detected red flags.

Example: "The window reflection shows a different room layout than pictured. Shadow direction on the sofa does not match the light source."

---

4. Fusion Strategy

Adaptive late fusion with dynamic weighting based on signal confidence:

Condition	Forensic Weight	VLM Weight
Base weights	55%	45%
Strong pixel anomaly	80%	20%
VLM uncertain	85%	15%
VLM high confidence	40%	60%

Decision Thresholds: - < 0.40: Likely Authentic - 0.40 - 0.60: Uncertain - ≥ 0.60: Likely Manipulated

---

5. Results

Test Dataset: 15 authentic + 15 AI-generated images

Threshold	Accuracy	Precision	Recall
0.45	63.3%	66.7%	53.3%
0.50	66.7%	100%	33.3%

Score Distribution: Authentic mean: 0.416 | AI-generated mean: 0.457