



The Motivation

- Can **VLMs** perceive finegrained distortions

Input Image

Distorted Image



Vision Language Model
"What Distortion affect this image?"

Human Majority 65.7%
Best Human 71.48%

Best VLM 61.9%

- VLMs excel at semantics but fall short on low-level perception

VLMs are deployed in **content moderation**, **restoration & quality monitoring** — yet their distortion perception has never been systematically benchmarked.

The Benchmark

27 Distortion Types

13,500 MCQ Questions

5 Severity Levels

Distortion Examples



Gaussian Blur



4 Choice MCQ-Format

- Q: Which distortion is applied to this image?
- A) Gaussian blur – Level 1 (Very mild)
 - B) Lens blur – Level 3 (Moderate)
 - C) Gaussian blur – Level 3 (Moderate) ✓
 - D) JPEG – Level 5 (Very strong)

Each distortion has 5 human calibrated severity levels per type

Insights

- Thinking variants hurt.
- Base models show a U-shaped severity curve.
- Easiest distortion: Gaussian blur (67% avg) · Hardest: denoising artifacts (<27%)
- Scaling is non-monotonic
- GPT-5.4 lands mid-table (51.6%), below the top 3 open-weight models
- Qwen3.5-4B (smallest in family) still beats GPT-5.4
- Severity is harder than type — even humans nail distortion type 83.6% of the time but drop to 69.9% on severity level

Results

