How Many Visual Levers Drive Urban Perception?

Interventional Attribution via Localised Counterfactual Edits

Jason Tang[1]; Stephen Law[1]. [1]University College London

CVPR JUNE 3-7, 2026 DENVER COLORADO

BACKGROUND & MOTIVATION

A score-changing edit is not automatically evidence.

Generative editing can shift preceived scores. But the change may come from intendented cue - or by confounds introduced during editing, such as changed lighting, layout, or other visual cues.

What should change? | Where should it change? | Would the edited scene still be the same place?

Why conterfactuals still need auditing

Saliency Map Show where the model looks.	✗ Not an edit.
SHAP / Concept attribution Identifies correlational visual cues.	✗ Not intervention-tested
Full-scene counterfactuals Generates alternative scenes	✗ Changes too much at once.

None of these tell us whether a single, localised, valid edit can plausibly shift perception.

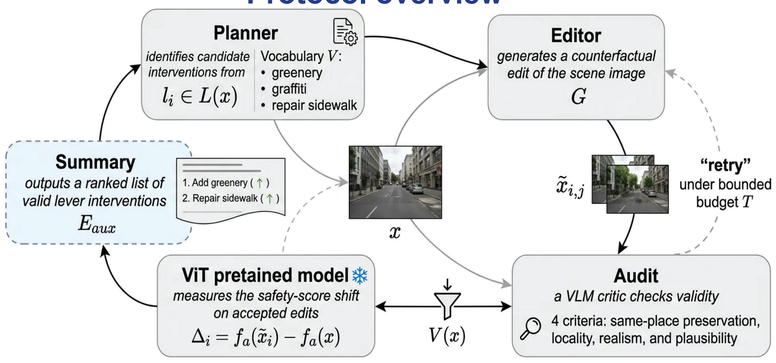
The motivation

Localised	Plausible	Same-place preserving	Auditable
Change only what matters.	Realistic & context-appropriate.	Keep the scene identity	Checked before we interpret.

We treat each edit as an auditable intervention hypothesis, the strongest lever is the largest score shift among edits that remain valid.

VALIDITY-GATED INTERVENTIONAL ATTRIBUTION

Protocol overview



Lever intervention (the unit of explanation)

A lever $l = (c, s, d, \tau)$ specifies:

Concept c	Scene support s	Direction d	Edit template τ
The urban cue to change (e.g., lane marking)	Grounded region or structure in the scene (e.g., road edge)	Change to add / remove / repair / improve	Instruction that preserves place & geometry

Example lever: same intended lever, different evidence status

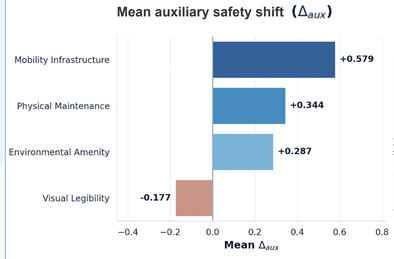
✓ VALID EDIT — Same place ▪ Local ▪ Realistic ▪ Plausible

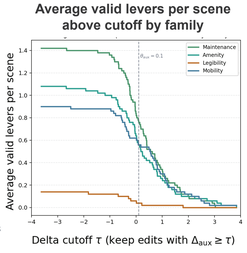
✗ INVALID EDIT — Confounded changes

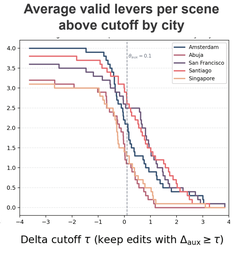
An edit is scored and retained only if it passes all validity checks (same-place preservation, locality, realism, and plausiblity).

PILOT STUDY & RESULTS

Average valid levers per scene above cut off by family







TAKEAWAY

What we claim	What we do NOT claim	What this enables
A protocol for ranking visual levers only after edits become admissible evidence.	Proxy shifts are not human-validated causal effects	A cheaper and better-prioritized path toward human evaluation and real-world urban design interventions.