Compositional Diffusion Image Synthesis for Long-Tailed Road Infrastructure Assessment

Initial conference pilot: rare-class synthesis under severe class imbalance in road imagery

By: Amirhossein Hassani
Faculty of Transport and Traffic Sciences, University of Zagreb

Scientific thesis

Road-attribute datasets are strongly long-tailed. A lightweight diffusion adapter can make a rare infrastructure class more visually expressible than prompt-only generation, providing a first step toward a larger multi-country synthesis framework.

IVORY
University of Zagreb
Faculty of Transport and Traffic Sciences
Funded by the European Union

Why this matters: long-tail road data limit safety-oriented AI

Rare infrastructure classes are precisely where automated assessment becomes most fragile.

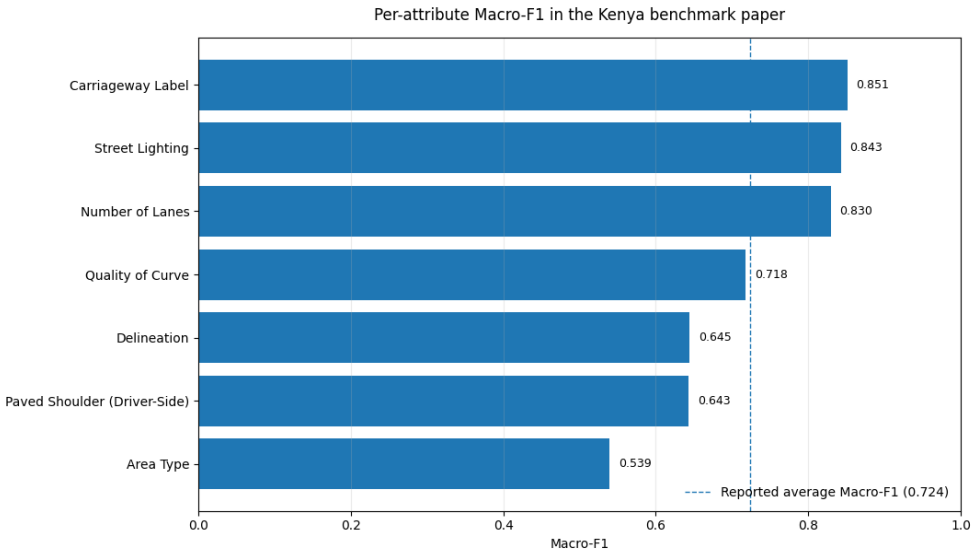
Why imbalance is a scientific problem

- Road-attribute datasets are dominated by a few frequent classes, while rare but safety-relevant conditions have very little data.
- This is amplified in smaller and LMIC-relevant collections, where unusual infrastructure types may appear only a handful of times.
- The result is a familiar failure mode: high overall accuracy, but weak macro-F1 because the model mostly learns the head classes.

Why synthesis is attractive

- A generator can target specific underrepresented classes rather than waiting for expensive new surveys.
- It also offers a controllable route toward future multi-country experiments: theme, attribute, and missing-combination synthesis.

Selected macro-F1 weaknesses from the road-attribute model [1]



These values reflect the same pattern: rare categories are underlearned even when headline accuracy looks strong.

[1] EfficientNet-Swin Transformer for Automated iRAP Road Safety Attribute Extraction in Low-and Middle-Income Countries

IVORY University of Zagreb Faculty of Transport and Traffic Sciences
Funded by the European Union

Kenya pilot attribute: severe imbalance within a single sidewalk label set

Passenger-side sidewalk labels already show the long-tail structure that motivates targeted synthesis.

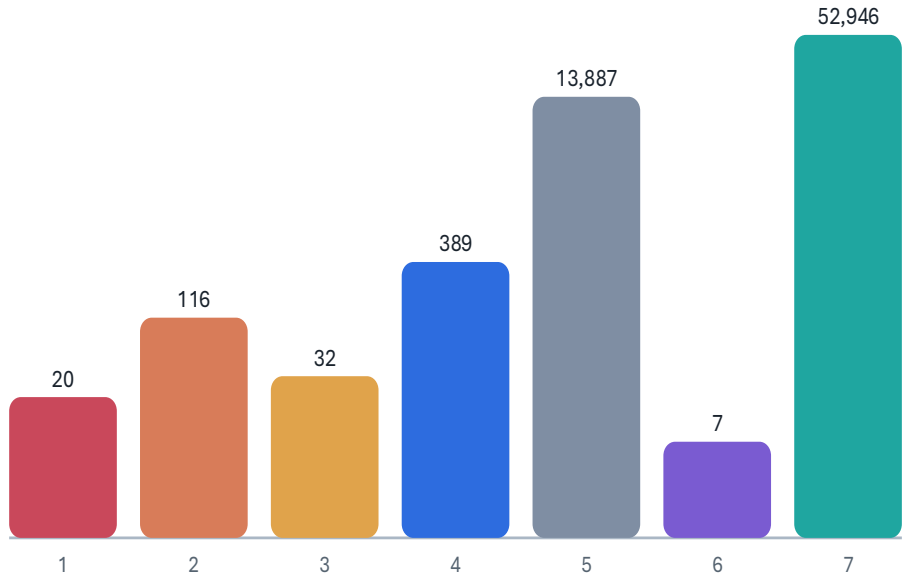
Pilot attribute and label meanings

Target rare class	1 = sidewalk with barrier
Reference class	4 = sidewalk 0m to <1m from road
Other dominant labels	5 = none, 7 = informal path 0m to <1m
Conference pilot	one dataset, one attribute, one rare class

Counts used for the pilot story

Class 1 total	20 images
Class 1 train / holdout	16 / 4
Class 4 reference subset	24 images
Practical meaning	the target concept is genuinely low-resource

Kenya passenger-side sidewalk distribution



Counts shown on log scale for readability

Class 1 is tiny; class 4 is a sidewalk reference class, not the head label. The pilot therefore tests a truly low-shot barrier concept.

IVORY
University of Zagreb Faculty of Transport and Traffic Sciences
Funded by the European Union

Broader framework and current pilot scope

The abstract describes the larger research programme; the conference result isolates the first empirical step.

Broader framework

Long-tail road data

Country / theme control

Attribute control

→ *compositional synthesis of rare or missing infrastructure configurations*

- Target use case: generate underrepresented road attributes while preserving country-specific visual context.
- Long-term evaluation: missing combinations, attribute fidelity, and eventual augmentation for downstream road-safety models.

Current conference result

- One dataset (Kenya), one attribute, one rare class.
- Question: can a small LoRA help express a rare barrier concept better than prompting alone?
- This is a feasibility study for the generative component.

Future Direction

country-style composition • missing-cell benchmarking • downstream augmentation

IVORY University of Zagreb Faculty of Transport and Traffic Sciences
Funded by the European Union

Pilot setup: rare sidewalk-class synthesis with a lightweight LoRA

The experiment isolates one rare concept under extremely low-data conditions.

Data and split

Dataset	Kenya road imagery
Attribute	Sidewalk – passenger-side
Rare class	1 = sidewalk with barrier
Reference class	4 = sidewalk 0m to <1m
Pilot split	16 train / 4 holdout / 24 reference
Preprocessing	rectangular 384×288 with letterbox

Model and adapter

Base model	Stable Diffusion v1.5
Adapter	one LoRA, rank 8
Target modules	U-Net attention projections
Checkpointing	steps 300 / 600 / 900, best chosen at 300

Controlled comparison design

Rare-class subset (16 images) → **LoRA fine-tuning (rare concept only)** → **A/B/C generation with fixed seeds**

Prompt protocol

Training caption: “dashcam road scene, sidewalk with barrier on passenger side”

Generation uses scene-aware prompts, not label-only prompts.

Most important control: Groups B and C use the same rare-class prompt; only C loads the LoRA.

Checkpoint metrics (FID / KID / CLIP) were used only to select the best checkpoint under extreme low-data conditions.

IVORY University of Zagreb Faculty of Transport and Traffic Sciences
Funded by the European Union

Controlled comparison: prompt-only versus prompt + LoRA

The key comparison keeps the rare-class prompt fixed and changes only the adapter.

Group A
reference prompt only

Prompt asks for a near-road sidewalk
(reference sidewalk condition).

Group B
rare-class prompt only

Prompt asks for the rare barrier class,
but no adaptation is loaded.

Group C
rare-class prompt + LoRA

Same prompt as Group B,
with the rare-class LoRA loaded.

What is controlled?

- A, tests a reference sidewalk prompt.
- B, tests whether prompting alone can express the rare barrier concept.
- C, isolates the LoRA effect because prompt text is identical to B.

Exact rare-class prompt used in B and C

“Realistic dashcam road scene, forward-facing road, daytime, natural colors, sidewalk with barrier on passenger side”

Interpretation guide

- If B and C differ, the change is due to adaptation rather than prompt engineering.
- This is the core reason the pilot is scientifically interpretable.

IVORY University of Zagreb Faculty of Transport and Traffic Sciences
Funded by the European Union

Initial result: the barrier concept becomes visible only after adaptation

The rare prompt alone remains visually close to the reference condition; the LoRA introduces the missing concept.

Group A

Reference sidewalk prompt only

Group B

Rare-class prompt only

Observed in pilot: A and B look broadly similar

Group C

Same rare-class prompt + LoRA

Observed in pilot: visible barriers appear in this condition

Core observation

The prompt-only rare-class condition (B) did not reliably separate itself from the reference condition (A). When the rare-class LoRA was added (C), barrier structures became visibly more explicit.

IVORY University of Zagreb Faculty of Transport and Traffic Sciences
Funded by the European Union

Real Reference Samples (Privacy Checked)

Class 4 (sidewalk 0 to <1 m from road)

Class 1 (sidewalk with barrier)

IVORY
University of Zagreb Faculty of Transport and Traffic Sciences
Funded by the European Union

Pilot training details and checkpoint selection

Training recipe

Base model	Stable Diffusion V1-5
Image size	384 × 288
Batch size	1
Learning rate	1e-4
LoRA rank	8
Max steps	900

Checkpoint selection (internal, not headline evaluation)

Checkpoint	FID	KID	CLIP
step_000300	**331.79**	**0.1358**	**29.08**
step_000600	462.07	0.2329	24.73
final_000900	440.54	0.2282	24.44

Checkpoint 300 was used for the final A/B/C generation because it was the best checkpoint under extreme low-data conditions.

These values were not used as the main scientific result; they only supported model selection.

IVORY University of Zagreb Faculty of Transport and Traffic Sciences
Funded by the European Union

Takeaways and path to the journal paper

The pilot is narrow by design, but it validates the rare-class generation component the larger framework needs.

What this pilot supports

- Long-tail road imagery creates a real rare-class learning problem, not merely a reporting artifact.
- A lightweight LoRA trained on only 16 images can make a rare “sidewalk with barrier” concept more visually expressible than prompt-only generation.
- This gives a concrete feasibility result for the generative core of the larger project.

What it does not support yet

- No multi-country theme transfer yet.
- No missing country×attribute benchmark yet.
- No augmentation study for downstream classifiers yet.

Next steps toward the full framework

- Add country/theme control and compose it with attribute control.
- Evaluate missing combinations explicitly rather than only a rare class.
- Test whether verified synthetic samples improve rare-class recognition in downstream road-safety models.

IVORY University of Zagreb Faculty of Transport and Traffic Sciences
Funded by the European Union