

Compositional Diffusion Image Synthesis for Long-Tailed Multi-Country Road Infrastructure Assessment

Amirhossein Hassani¹, Muhammad Shahid, Marko Ševrović

*Faculty of Transport and Traffic Sciences,
University of Zagreb,
Vukelićeva Street 4, HR-10000 Zagreb, Croatia,
ahassani@fpz.unizg.hr, mshahid@fpz.unizg.hr, msevrovic@fpz.unizg.hr*

Abstract

Road safety agencies increasingly rely on automated assessments of road infrastructure. However, the image datasets used to train these systems are often highly imbalanced, fragmented, and incomplete. This issue is especially pronounced in low- and middle-income contexts where coverage is sparse, and certain types of infrastructure are rarely observed. The long-tail distribution limits the reliability of visual models precisely where high-risk environments are most critical. It also affects countries with missing attribute categories in their available imagery.

We present a controllable image synthesis framework for multi-country road-infrastructure scenes that addresses long-tail scarcity and missing country-attribute combinations. Building on recent advances in parameter-efficient adaptation of diffusion models and compositional personalization, our approach factorizes generation into (i) country or domain theme modules that capture visual conventions and acquisition characteristics and (ii) attribute modules representing discrete infrastructure properties (e.g., roadside configuration, delineation, or cross-section cues). At inference time, these modules are composed to synthesize targeted attribute combinations, including those absent from a given country's training subset, while reducing interference that commonly arises when multiple customizations are merged. To make synthetic data usable for learning and auditing, we couple generation with an automatic verification loop that filters or reranks samples using attribute checkers, producing a curated synthetic set with measurable attribute fidelity.

We evaluate the system under a benchmark that explicitly holds out selected country-attribute cells during training to test generalization to unseen combinations. Beyond realism metrics, we report attribute correctness, compositional interference diagnostics, and tail-sensitive diversity indicators to detect rare-mode collapse. Finally, we demonstrate the practical impact of targeted synthesis as a data augmentation method for downstream road-attribute recognition, focusing on improvements in rare categories that are typically underrepresented in real-world surveys. The results indicate that factorized, compositional diffusion synthesis can mitigate long-tail limitations and reduce dependence on costly new data collection, supporting more robust AI-assisted road infrastructure assessment across heterogeneous national contexts.

Keywords: Road infrastructure assessment, Class imbalance, Diffusion models, Synthetic data augmentation

¹ * Corresponding author.

E-mail address: ahassani@fpz.unizg.hr