CausalMixFT Synthetic Data Augmentation for Stable Code Generation Across Languages

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: Does the application of CausalMixFT-style synthetic data augmentation reduce performance variance in code generation tasks across different programming language domains compared to traditional data. 12 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: A survey of synthetic data augmentation methods in computer vision. Research question: Does the application of CausalMixFT-style synthetic data augmentation reduce performance variance in code generation tasks across different programming language domains compared to traditional data augmentation methods?.

2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.2/10.

3 Results

9 papers retrieved. 12 claims extracted; 0 independently verified. Quality review score: 2.2/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The FlyingThings3D dataset has proven effective in training deep learning models for optical flow and scene flow tasks.	×	0.04
Neural rendering aims to realize the scene rendering process using deep learning models.	×	0.06
Neural rendering can be accomplished in both forward and backward directions.	×	0.03
In the forward direction of neural rendering, 2D images are generated from 3D scenes and additional scene parameters.	×	0.04
In the backward direction of neural rendering, the pixel image is translated into a realistic 3D scene.	×	0.07
The rendering process is inherently non-differentiable, which constrains its incorporation in deep neural networks.	×	0.04
Differentiable rendering is a method to address the non-differentiability of the rendering process.	×	0.04
Point clouds have low memory requirements but low accuracy of scene topology information.	×	0.02
Voxel representations are more accurate with less processing and simplicity but have a high memory footprint.	×	0.03
Mesh representations provide more grounding (i.e., physics-aware scene representation) but have high computational cost	×	0.01
Multimodal representations have high resolution and are more robust to visual artifacts but are more complex and have hi	×	0.02
Implicit (NN) representations are naturally differentiable and have low memory requirements but lack grounding.	×	0.01

References

- http://arxiv.org/abs/2510.21391v1
- http://arxiv.org/abs/1402.0087v1
- http://arxiv.org/abs/2403.10075v2