

Causal Synthetic Text Descriptions for Domain Shift Reduction in Few-Shot Classification

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: Can causal synthetic text descriptions reduce domain shift gaps in few-shot classification across Office-Home and VisDA datasets. 17 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Inferring Latent Class Statistics from Text for Robust Visual Few-Shot Learning. Research question: Can causal synthetic text descriptions reduce domain shift gaps in few-shot classification across Office-Home and VisDA datasets?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.2/10.

3 Results

4 papers retrieved. 17 claims extracted; 1 independently verified. Quality review score: 4.2/10.

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 experiments use ImageNet and iNaturalist as the two base datasets.	×	0.03
iNaturalist is a hierarchical dataset containing fine-grained classes of different species.	×	0.04
The study evaluates performance on 9 cross-domain datasets: Caltech, EuroSAT, Food, Flowers, SUN397, DTD, Pets, Cars, and	×	0.03
Visual and text features are extracted using a pre-trained CLIP ResNet50 model.	×	0.07
The CLIP ResNet50 model used was trained on the LAION400M dataset.	×	0.01
The LAION400M dataset is disjoint from the base datasets (ImageNet, iNaturalist) and the test datasets used in the exper	×	0.02
The method predicts the mean and covariance of a class distribution in the feature space from text.	✓	0.20
The method consists of a learning phase to map text to moments and an adaptation phase to adjust predictions to the down	×	0.04
The method infers a diagonal covariance matrix due to the high dimensionality of the features.	×	0.04
Text inputs for the method include contexts like 'a photo of a {class}' or visual descriptions generated by GPT3.	×	0.08
In the 1-shot setting on the Caltech dataset, the Baseline method achieved an accuracy of 62.07%.	×	0.03
In the 16-shot setting on the Caltech dataset, the Baseline method achieved an accuracy of 86.34%.	×	0.03
Zero-Shot CLIP achieved an average accuracy of 59.42% across the 9 cross-domain datasets.	×	0.05
Zero-Shot CLIP achieved an accuracy of 36.01% on the EuroSAT dataset.	×	0.02
In the 16-shot setting, the Baseline method achieved an average accuracy of 67.48% across the 9 cross-domain datasets.	×	0.04
In the 1-shot setting, the Baseline method achieved an average AUROC of 84.70% across the 9 cross-domain datasets.	×	0.05
In the 16-shot setting, the Baseline method achieved an average AUROC of 92.25% across the 9 cross-domain datasets.	×	0.04

References

- <http://arxiv.org/abs/1911.06045v3>
- <http://arxiv.org/abs/2311.14544v1>
- <http://arxiv.org/abs/2206.00092v1>