Progress Evaluation Report

Enhancing Visual Object Recognition in Indoor Environments Using Topologically Persistent Features

Kulathunga K.A.J.T.

210306G

Table of Contents

1. Introduction

Object recognition in indoor environments is a key challenge in robotics. While CNN-based approaches such as ResNet [2], EfficientNet [3], and Faster R-CNN [4] achieve strong accuracy in benchmark datasets, they exhibit performance drops in unseen environments due to illumination variations, background clutter, and domain shifts [6].

Samani et al. [1] proposed the use of features extracted through Topological data analysis (TDA) from segmentation maps using persistent homology. By generating sparse persistence images (PI) and amplitude descriptors, they achieved more robust recognition in novel environments compared to ResNet, EfficientNet, and domain-adaptive detection models.

The goal of this project is to reproduce this baseline and design targeted enhancements that yield measurable performance gains. Specifically, improvements will focus on (i) optimizing feature extraction and encoding, (ii) enhancing recognition network architecture, and (iii) improving training and data augmentation strategies.

2. Literature Review

2.1. Deep Learning for Recognition and Detection

CNN-based architectures such as ResNet [2] and EfficientNet [3] learn strong image features but are sensitive to distribution shifts. End-to-end detectors such as Faster R-CNN [4] and SSD [5] provide object localization and recognition jointly but degrade substantially in unseen domains.

2.2. Domain Adaptation

Adversarial feature alignment [6], CycleGAN-based domain transfer [7], and progressive adaptation [8] improve generalization but require retraining, limiting applicability in robotics where environments continually change [9,10].

2.3. Topological Data Analysis in Vision

Topological Data Analysis (TDA) captures shape-based invariants independent of lighting or texture. Persistent homology generates persistence diagrams (PDs) summarizing birth and death of features [11]. Representations such as Persistence Images (PI) [12] and amplitude descriptors [13] convert PDs into machine-learning–friendly vectors. Samani et al. [1] showed that sparse PI features outperform CNN-based features in unseen indoor environments.

2.4. Research Gap

Existing CNNs fail under domain shifts, while TDA-based recognition struggles with segmentation quality and extreme pose changes. A hybrid approach with optimized training, architectural improvements, and enhanced preprocessing can bridge this gap.

3. Methodology Outline

3.1. Baseline Framework

Following [1], the pipeline consists of:

i. Segmentation: DeepLabv3+ generates binary object segmentation maps.
ii. Topological Feature Extraction: Cubical complexes are constructed; persistence diagrams are computed.
iii. Feature Encoding: Sparse PI (QR-pivoting); amplitude descriptors from bottleneck distances [11].
iv. Recognition: Features are input to fully connected networks (3-5 layers).

3.1. Planned Enhancements

i. Hybrid Feature Encoding – Combine Sparse PI + Amplitude features into a single vector with dimensionality reduction. Improves robustness by leveraging complementary descriptors.
ii. Improved Training Strategies – Use different optimizers, cyclical learning rates, and label smoothing for better generalization.
iii. Advanced Data Augmentation – Apply shape-preserving deformations, contrast/illumination shifts, and Gaussian noise to simulate unseen conditions.
iv. Lightweight Ensemble Model – Train separate classifiers on PI and amplitude features, combine predictions via weighted voting to reduce variance.

3.2. Evaluation Plan

Datasets:

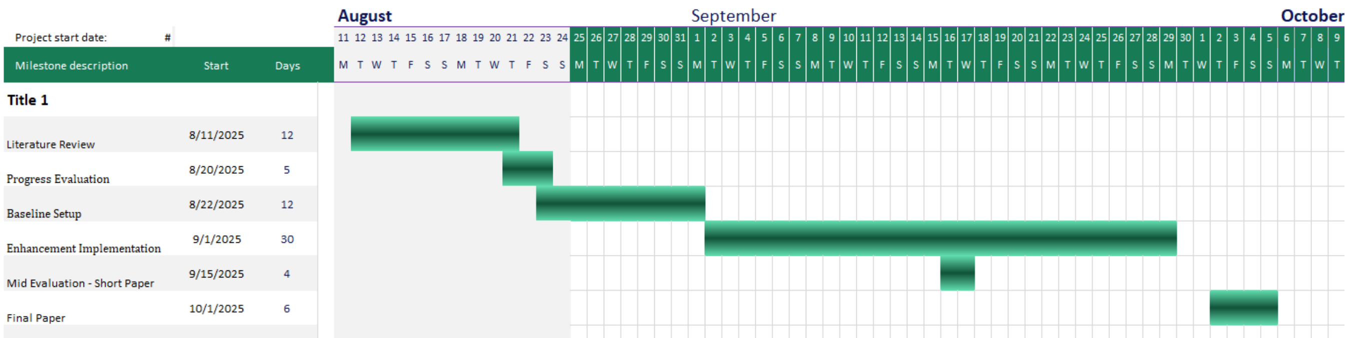
- MPEG-7 Shape Silhouette (benchmark for shape descriptors)

- RGB-D Scenes v1 (domain robustness)

- UW-IS dataset (living room vs warehouse domain shift)

Metrics: Accuracy, weighted F1 score, recall, and precision. Baseline results from [1] will be used for comparison.

4. Project Timeline



5. References

[1] "Visual object recognition in indoor environments using topologically persistent features," IEEE Journals & Magazine | IEEE Xplore, Oct. 01, 2021. https://ieeexplore.ieee.org/abstract/document/9496193

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in Lecture notes in computer science, 2016, pp. 630–645. doi: 10.1007/978-3-319-46493-0_38.

[3] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," PMLR, May 24, 2019. https://proceedings.mlr.press/v97/tan19a.html?ref=ji

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," 2015. https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html

[5] W. Liu et al., "SSD: Single Shot MultiBox Detector," in Lecture notes in computer science, 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.

[6] Y. Chen, W. Li, C. Sakaridis, D. Dai, and V. G. Luc, "Domain Adaptive Faster R-CNN for object detection in the wild," 2018. https://openaccess.thecvf.com/content_cvpr_2018/html/Chen_Domain_Adaptive_Faster_CVPR_2018_paper.html

[7] T. Kim, M. Jeong, S. Kim, S. Choi, and C. Kim, "Diversify and Match: a domain adaptive representation learning paradigm for object detection," 2019. https://openaccess.thecvf.com/content_CVPR_2019/html/Kim_Diversify_and_Match_A_Domain_Adaptive_Representation_Learning_Paradigm_for_CVPR_2019_paper.html

[8] H.-K. Hsu et al., "Progressive domain adaptation for object detection," 2020. https://openaccess.thecvf.com/content_WACV_2020/html/Hsu_Progressive_Domain_Adaptation_for_Object_Detection_WACV_2020_paper.html
[9] "Artificial Intelligence for Long-Term Robot Autonomy: A survey," IEEE Journals & Magazine | IEEE Xplore, Oct. 01, 2018. https://ieeexplore.ieee.org/abstract/document/8421618
[10] G. Mahyari and T. Locker, "Domain adaptation for robot predictive maintenance systems," arXiv.org, Sep. 23, 2018. https://arxiv.org/abs/1809.08626
[11] "Computational homology," Google Books. https://books.google.lk/books?hl=en&lr=&id=hbcuBAAAQBAJ&oi=fnd&pg=PR7&dq=Computational+Homology&ots=NrafH54AJE&sig=NAsudk0TyglqImotgjxG-nMd3SE&redir_esc=y#v=onepage&q=Computational%20Homology&f=false
[12] H. Adams et al., "Persistence images: a stable vector representation of persistent homology," 2017. https://www.jmlr.org/papers/v18/16-337.html
[13] "SPARse-TDA: Sparse realization of topological data analysis for Multi-Way classification," IEEE Journals & Magazine | IEEE Xplore, Jul. 01, 2018. https://ieeexplore.ieee.org/abstract/document/8249544
[14] Smartslab, "GitHub - smartslab/objectRecognitionTopologicalFeatures: Code repository for methods proposed in 'Vision-Based Object Recognition in Indoor Environments using Topologically Persistent Features,'" GitHub. https://github.com/smartslab/objectRecognitionTopologicalFeatures?tab=readme-ov-file
[15] Tensorflow, "models/research/deeplab at master · tensorflow/models," GitHub. https://github.com/tensorflow/models/tree/master/research/deeplab