SIFT And CNN in Objects Classcification

COMP61342&41342 Cognitive Robotics And Computer Vision Assignment Report

Yuyi Yang
yuyi.yang@postgrad.manchester.ac.uk
p68393yy

Abstract—This project aims to explore the use of SIFT (Scale-Invariant Feature Transform) and CNN (Convolutional Neural Network) in object classification tasks. By implementing different approaches on different datasets, we aim to compare their performance, and understand their strengths and weaknesses of in the context of object recognition.

Index Terms—Local Feature, CNN, Object Classification

I. Introduction

In recent years, equipping robots with reliable perception capabilities has become a foundation for autonomous manipulation, navigation, and human–robot interaction. Object recognition—identifying and localizing objects from camera data—remains especially challenging in unstructured, dynamic environments where lighting, occlusion, and viewpoint changes can drastically reduce accuracy.

Traditional computer-vision pipelines, based on handcrafted local features (i.e. SIFT) and classical classifiers (e.g. SVM, Random Forest), offer interpretability and low inference latency; however, they may falter when faced with complex environment and large intra-class variability. Conversely, convolutional neural networks (CNNs) have revolutionized vision performance, automatically learning hierarchical features, but at the cost of greater computational demand and complex hyperparameter tuning.

This work systematically compares these two paradigms on the CIFAR-10 [1] and iCubWorld 1.0 [2] benchmark. This project researches how choice of feature extractor, classifier, network architecture, and training strategy affect performance of classification accuracy and inference speed, and model robustness to environmental changes.

The rest of the report is organised as follows: Section 2 describes our methodology; Section 3 reviews the experiments; Section 4 presents results and analysis; Section 5 offers a comparative discussion; and Section 6 concludes with future directions.

II. Methodology

III. Experiments

IV. Results And Discussion

V. Conclusion

References

[1] A. Krizhevsky, G. Hinton et al., "Learning multiple layers of features from tiny images.(2009)," 2009.

[2] S. Fanello, C. Ciliberto, M. Santoro, L. Natale, G. Metta, L. Rosasco, and F. Odone, "icub world: Friendly robots help building good vision data-sets," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2013, pp. 700–705.