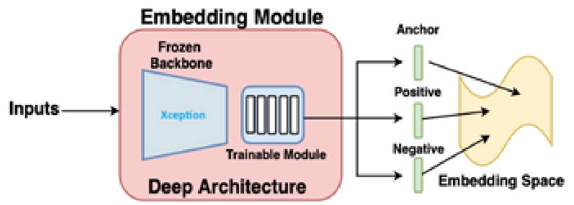
Enhanced Vegetable Classification Using Siamese Network

Jashmina[1], T., Shajini[2], M. and Ramanan[1], A.
[1]Department of Computer Science, Faculty of Science, University of Jaffna,
[2]Department of Knowledge Engineering and Communication, Faculty of Computing, University of Sri Jayewardenepura
[1]2019sp090@univ.jfn.ac.lk

Introduction

Automatically identifying and categorising grocery products through images is crucial for automated billing and inventory management. This task faces challenges like visually similar products and misplaced items. Traditional image processing methods often struggle, but Convolutional Neural Networks (CNNs) have shown promise. This study enhances vegetable classification by comparing various CNN architectures and using a Siamese network with a triplet loss function, along with the Xception model for better feature extraction. Testing on the vegetable class of the Grocery Store Dataset [2], the model achieved 86% accuracy, effectively classify similar vegetables such as cucumbers and zucchinis.

Methodology



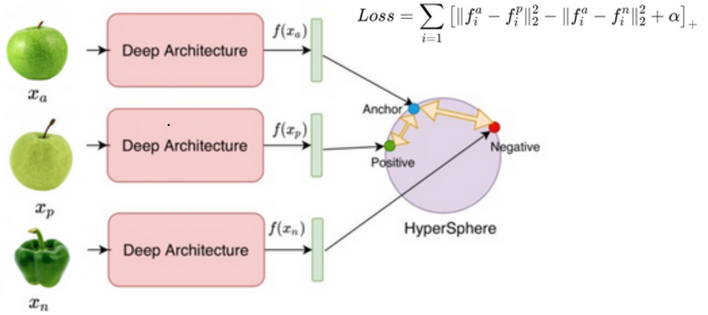
Our methodology centers on a Siamese network architecture, designed to learn effective feature representations by comparing image pairs. The Xception model, known for its computational and parameter efficiency and its ability to capture fine-grained features, is used for enhanced feature extraction. To further refine feature extraction, we added custom top layers: a flatten layer, two dense layers, two dropout layers, a batch normalisation layer, and a normalisation layer.

By combining the Siamese network architecture with the Xception model and the triplet loss function, our approach effectively addresses challenges such as visual similarity among vegetables.

Triplet Loss

Triplet loss formula

$$Loss = \sum_{i=1}^{N}\left[\|f_i^a - f_i^p\|_2^2 - \|f_i^a - f_i^n\|_2^2 + \alpha\right]_+$$



- **N is the number of triplets (anchor, positive, negative).**
- **α is the margin that enforces a minimum distance between the positive and negative pairs.**
- **[.]+ indicates that only positive values are considered (i.e., the loss is zero if the difference is negative).**

This triplet loss function calculates the distance between anchor and positive images, anchor and negative images. By minimizing the distance between similar items while maximizing the distance between dissimilar ones, this approach significantly enhances our model's ability to recognize and distinguish vegetables in our dataset. Triplet loss is effective for imbalanced data.

Experimental Setup

Dataset:
- Grocery Store Dataset (Vegetables category)

Model:
- Base Architecture: Xception model pre-trained on ImageNet
- Encoder: Modified Xception model with fully connected layers for feature extraction
- Siamese Network: Compares image triplets (Anchor, Positive, Negative) using Euclidean distance

Loss Function:
- Distance-based loss: Trained to minimize distance between Anchor and Positive, and maximize distance between Anchor and Negative

Training:
- Optimizer: Adam with learning rate : 0.001
- Original image (224 x 224) -> resized image (128 x 128)
- Batch size: 8
- Epochs: 50

Evaluation:
- Accuracy calculated based on whether the model correctly classified similar (Anchor-Positive) and different (Anchor-Negative) pairs

Dataset

Dataset: Grocery Store Dataset [2]
#Classes: 3

#Images : 5125
#Training : 2640
#Validation : 296
#Testing : 2189

Vegetable class
#Training : 609
#Validation : 85
#Testing : 562

This study employs a deep learning approach for vegetable classification using the publicly available Grocery Store Dataset [2], which includes vegetables, fruits, packages classes and our study considers only vegetable class which includes 14 vegetable subclasses: Asparagus, Aubergine, Cabbage, Carrots, Cucumber, Ginger, Leek, Brown-Cap Mushroom, Onion, Pepper, Potato, Red Beet, Tomato, and Zucchini.

Test Results

Performance Comparison on the Grocery Store Dataset [2]

Study	Model	Accuracy
Klasson *et al* (2019)	VGG16 (138,357,544 parameters)	73.3%
Klasson *et al* (2019)	DenseNet169 (14.3 million parameters)	85.0%
Klasson *et al* (2020)	VCCA (Random 10 classes)	86.8%
Sahoo *et al* (2023)	Transformer	81.3%
Our Model	Xception (22 million parameters)	86.0%
	Precision : 78.69%	Recall : 93.96%

Conclusion

This study presents an advancement in vegetable classification using deep learning. By employing a Siamese network with a triplet loss function, we effectively addressed challenges in distinguishing visually similar products. The Xception-based encoder provided strong feature extraction, and the model demonstrated improved classification accuracy. Test accuracy confirm that the Siamese network is effective in visual similarity tasks, offering a promising approach for small datasets.

References

[1] F. Schroff, D. Kalenichenko and J. Philbin, "Facenet: a unified embedding for face recognition and clustering", 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), doi:https://doi.org/10.1109/cvpr.2015.7298682, (pp.35-40).
[2] Grocery Store Dataset:https://github.com/marcusklasson/GroceryStoreDataset
[3] M. Klasson, C. Zhang and H. Kjellström, "A Hierarchical Grocery Store Image Dataset With Visual and Semantic Labels", 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), doi:10.1109/WACV.2019.00058, (pp.491-500).
[4] A. Tonioni and L. Stefano, "Domain invariant hierarchical embedding for grocery products recognition", 2019 Computer Vision and Image Understanding, doi: 10.48550/arXiv.1902.00760, (pp.81-92).
[5] E. Gothai, S. Bhatia, A. Alabdali, D. Sharma, B. Raj and P. Dadheech "Design Features of Grocery Product Recognition Using Deep Learning", 2022 Intelligent Automation and Soft Computing, doi: 10.32604/iasc.2022.026264, (pp.1231-1246).
[6] J. Domingo, R. Aparicio and L. Rodrigo, "Cross Validation Voting for Improving CNN Classification in Grocery Products", 2022 in IEEE Access, doi:10.1109/ACCESS.2022.3152224, (pp.20913-20925).
[7] A. Tonioni, E. Serra and L. Stefano, "A deep learning pipeline for product recognition on store shelves," 2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS), doi: 10.1109/IPAS.2018.8708890, (pp.25-31).
[8] A. Sinha, S. Banerjee and P. Chattopadhyay, "An Improved Deep Learning Approach For Product Recognition on Racks in Retail Stores", 2022 arXiv preprint arXiv:2202.13081, doi: 10.48550/arXiv.2202.13081.
[9] M. Klasson, C. Zhang and H. Kjellström. "Using Variational Multi-view Learning for Classification of Grocery Items", 2020 Patterns ,doi:10.1016/j.patter.2020.100143, (pp.2666-3899).
[10] S. Sahoo, S. Chalasani, A. Joshi and K. Iyer, "Enhancing Classification with Hierarchical Scalable Query on Fusion Transformer," 2023 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, doi: 10.1109/ICCE56470.2023.10043496, (pp. 1-6).