A Systematic Investigation on end-to-end Deep Recognition of Grocery Products in the Wild

Marco Leo, Pierluigi Carcagnì, Cosimo Distante

National Research Council of Italy, Institute of Applied Sciences and Intelligent Systems - Lecce (CNR-ISASI)

Email : {marco.leo,pierluigi.carcgni,cosimo.distante}@cnr.it

INTRODUCTION

Automatic recognition of products on grocery shelf images is a new and attractive topic in computer vision and machine learning since, it can be exploited in different application areas. This paper introduces a pure end-to-end classification to make the task ready to be exploited on data acquired in different contexts and without any constraint. The proposed pipeline takes as input a raw image containing grocery products (acquired in the wild from the market shelves with multiple products in the scene even under occlusions and viewpoint changes) and gives as output the product class. Along with the manuscript, a systematic investigation of different CNN architectures is proposed. The architectures are based on convolutional neural networks for addressing the product recognition task exploiting the proposed pipeline on the huge challenging dataset in [1].

DATASET

The dataset contains natural images of raw and refrigerated grocery items (fruits, vegetables, and packages). The dataset consists of 5125 images of raw grocery items, from 81 classes, where the number of images in each class ranges from 30 to 138. It has a hierarchical structure: it takes into account three categories of items: fruits, packaged food and drinks, and vegetables. Each of these three category items contains 19, 9, and 15 coarse classes respectively. Some of the coarse classes have subcategories that lead to their fine classes.

Fruits

Packages

Vegetables

METHOD

4 families of CNN architectures have been investigate:
ResNet [2], DenseNet [3], SENet [4] and EfficientNet [5].

Architectures that achieved at least an accuracy score of 85% were considered for final classification and are reported in the experimental results section.

RESULTS

TABLE I: Results concerning 81 fine classes classification by using a single model.

Model	Acc	F1-score	Prec.	Rec.
ResNet-18	87.51	88.79	89.97	88.95
ResNet-34	89.10	89.98	90.88	90.11
ResNet-50	90.58	91.75	92.47	91.96
ResNet-101	**92.55**	**93.23**	**93.68**	**93.55**
ResNet-152	91.68	92.84	93.66	92.84
SE-ResNet-50	90.58	91.47	92.32	91.55
SE-ResNet-101	88.89	89.87	90.96	89.82
EfficientNet-b0	88.94	89.82	90.57	90.11
EfficientNet-b1	88.35	88.89	89.72	88.99
EfficientNet-b2	88.98	90.14	90.60	90.41
EfficientNet-b3	88.68	90.03	90.68	90.29
EfficientNet-b4	89.19	90.61	90.82	91.07
EfficientNet-b5	88.22	89.30	89.88	89.60
EfficientNet-b6	88.73	89.98	90.53	90.27
EfficientNet-b7	89.36	90.55	91.27	90.84
DenseNet-121	91.16	91.74	92.78	92.06
DenseNet-169	91.29	92.47	93.22	92.55
DenseNet-201	92.01	92.99	93.40	93.16

TABLE II: Results concerning 43 coarse classes classification by using a single model.

Model	Acc	F1-score	Prec.	Rec.
ResNet-18	91.54	89.02	90.41	88.69
ResNet-34	92.43	89.74	91.12	89.56
ResNet-50	93.61	91.91	92.70	92.05
ResNet-101	94.87	93.35	93.63	**93.61**
ResNet-152	94.32	92.98	93.80	92.95
SE-ResNet-50	92.97	90.98	91.98	90.86
SE-ResNet-101	92.43	89.77	91.01	89.65
EfficientNet-b0	92.01	89.38	90.00	89.53
EfficientNet-b1	92.85	91.31	92.18	91.07
EfficientNet-b2	92.64	90.75	91.47	90.70
EfficientNet-b3	92.09	90.06	90.82	90.12
EfficientNet-b4	92.43	89.42	90.01	89.68
EfficientNet-b5	92.05	89.90	91.11	89.58
EfficientNet-b6	92.13	89.68	90.74	89.60
EfficientNet-b7	92.38	89.68	91.62	89.04
DenseNet-121	94.66	92.60	92.73	93.11
DenseNet-169	93.94	92.35	93.09	92.46
DenseNet-201	**94.95**	**93.56**	**94.25**	93.49

TABLE III: Results concerning 81 fine classes classification with ensembles of models.

Model	Acc	F1-score	Prec.	Rec.
Ensemble A	92.89	93.87	94.46	93.98
Ensemble B	92.47	93.63	94.08	93.62
Ensemble C	**93.48**	**94.46**	**94.98**	**94.52**
Ensemble D	93.14	94.41	94.88	94.42
Ensemble E	92.68	93.77	94.36	93.80

TABLE IV: Results concerning 43 coarse classes classification with with ensembles of models.

Model	Acc	F1-score	Prec.	Rec.
Ensemble A	95.42	94.41	95.13	94.19
Ensemble B	95.16	93.99	94.79	93.57
Ensemble C	**95.84**	**94.88**	95.31	**94.78**
Ensemble D	95.58	94.81	**95.48**	94.59
Ensemble E	95.42	94.41	95.13	94.19

TABLE V: Comparison of the results concerning 81 fine classes and 43 coarse classes classification. Experimental outcomes reported in [1] have been considered for reference.

Net	F1-score fine classes	F1-score coarse classes
Proposed		
ResNet-101	**93.23**	89.65
Ensemble C	**94.46**	**94.88**
DenseNet-201	92.99	**93.56**
Previous		
AlexNet + SVM	72.6	78
VGG16 + SVM	73.3	76.6
DenseNet-169 + SVM	85	84
AlexNet	69.3	76.4
VGG16	73.8	74.9
DenseNet-169	80.4	82

CONCLUSIONS

Gathered classification results were very encouraging and largely outperformed previous approaches on the same dataset. This could be a useful step towards the implementation of an accurate classification system able to supply support in different assistive applications.
Future works will explore other deep learning models.
Besides, we are being collecting new images in order to extend the dataset, making it even more suitable to be exploited for modern machine learning approaches.

REFERENCES

[1] M. Klasson, C. Zhang, and H. Kjellstr¨om, "A hierarchical grocery store image dataset with visual and semantic labels," in IEEE Winter Conference on Applications of Computer Vision (WACV), 2019.
[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[3] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
[4] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.
[5] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," arXiv preprint arXiv:1905.11946, 2019.