Product Based Classification of Bulk Food Grains using Bag of Visual Words and Deep Features

Abdelmgeid Amin Ali
Computer Science Department, Faculty of Science, Minia University, Al Minia 61519, Egypt
abdelmgeid@yahoo.com

Usama Sayed Mohammed
Electrical Engineering Department, Faculty of Engineering, Assiut University, Assiut 71516, Egypt
usama@aun.edu.eg

Rehab Ragaa Nour
Computer Science Department, Faculty of Science, Minia University, Al Minia 61519, Egypt
rehab.ibrahim@mu.edu.eg

Abstract: *The goal of this research is to compare between the performance of the traditional machine learning classification algorithm using Bag of Visual Words (BoVW) method and off-the-shelf deep features extracted by VGG-19, and Inception-V3 models and trained SVMs using the extracted features. By comparing the AUC, sensitivity, and specificity of SVM with VGG-19 and Inception-V3, we can conclude that off-the-shelf deep features has an important impact on food grains image classification.*

Keywords: *Image classification, Bag of visual words, Transfer learning, Convolutional neural networks, Deep learning*

1. Introduction

The main goal of this paper is to examine two approaches, called Bag of Visual Words and Transfer Learning. We used them for classifying food grains images. Among several cereal and pulse crops that are growing and ripening around the world, corn, flax, rice, and wheat are considered important cereal crops, while on the contrary common bean, pea, lentil, and soybean top the list of pulses. They are extremely essential products for countries' economies. Production, preservation, trade, and consumption of these materials are critical to the growth of a country's agricultural economy[1].

Automatic food grains identification has abroad range of applications, including automatic packaging and transportation in post-harvest agricultural machines equipment, food automation processing by chef robots and helping visually impaired individuals and grocery checkout systems to label an unpacked food[2]. One suggested solution to raise crop yields is to apply modern automated smart farming and robots are performing varied tasks, for instance, cultivation, transplanting, and harvesting[3]. Image processing has been used in different ways to identify the crops[4].

However, Grains seeds identification is challenging because target classes are often visually quite similar. Fig. 1 demonstrates two different seeds of grains with high intra-class variance and small inter-class variance. To distinguish, we need to extract visual features with high informative visual features.

Figure 1 Examples of species with fine differences. From left to right Soy, and Pea samples.

In this paper, we explore using hand-engineered features in Bag of Visual Word approach and pretrained deep convolution neural network (CNN) approach. We reuse the feature extraction part of VGG-19 and Inception-V3 models and then we re-train the classification part with bulk grains images dataset.

The remainder of this paper is structured as follows: Section 2 presents related work on food grains classification using image processing techniques and presents an overview of previous research and the state of the art. Section 3 describes the dataset and contains the details about the proposed methodology and the investigations we carried out. Section 4 has the experiment results obtained, followed by a conclusion in Section 5.

2. Related Work

Computer vision systems in food and agricultural products industries have been applied in the areas of quality inspection, grading, insect infestation detection, damage detection and disease detection [5, 6]. These systems can provide an accurate, accelerated,, no subjective and non-destructive metrics tools[7]. More specifically, computer vision in food grains industries has been applied in the areas of sorting, grading, examining the effect of humidity, detection of defects such as insect and pest detection, microbial and fungi infection, and foreign matter[6].

Research on automatic food grains classification has been active[8]. Several studies have developed machine vision systems for class and variety identification of grains. [9] use morphological and texture features using a neural network model and a non-parametric model. They examined the benefit of combining morphological and textural features in classification, the method had accuracies >96%. Study by [10] found that morphological, color, textural and wavelet features get 96% classification accuracy. [11] has achieved 99% accuracy by using morphological and color-based models for cereal grain classification.

Later, studies focus on hand crafted features such as Scale Invariant Feature Transform (SIFT)[12, 13] and the histograms of gradient orientation (HOG)[14], which achieve high accuracy on controlled conditions for one seed on contrastive background.

At the 2012 ImageNet Large Scale Visual Recognition Competition (ILSVRC), a method using CNN called AlexNet won the championship by a large margin, and since deep learning has attracted attention, CNN has become the mainstream in image recognition[15], researchers start to apply deep CNN to this problem. In [16], Xinshao W, et al. used features extracted from PCANet. In [17] an interlacing PCA net is proposed to extract weed seeds image features for automatic identification. However, it only gives results on weed seeds dataset.

3. Materials and Methods

This section discusses the experimental setup, including the dataset and details on the image classification approaches used. First it provides the grains images dataset is presented with the pre-processing process. Then the pre-trained features and Support Vector Machine (SVM) are presented. Finally, we show the experimental results of the classification.

3.1. Dataset

the published studies focus on determining types of food grains from one seed image [18], but in field, grocery shops, malls and other vending agencies the food grains of different types are available in boxes or containers. So, we used bulk food grains images for this study. we used the grain seeds images database provided from The Laboratory of Vision, Robotics and Imaging (VRI) in the Federal University of Parana (UFPR)[19-21]. The image dataset has 339 images from 13 distinct seeds species. The dataset used a Samsung T-65 model with a resolution of 1024 × 768. The images saved in JPG format. The figure 2 offers some samples of the dataset. This study used food crops images from grain types called barely, canola, corn, flax, lentil, pea, quinoa, sesame, and soybeans.

Figure 2 Thumbnails of the images chosen to sample 11 classes of the dataset.

Images were cropped and augmented using transformations to prevent overfitting. Overfitting means that the training data has been trained, but the unknown data (test data) has not been adapted. In other words, overfitted neural networks do not have generalization performance, so they cannot be classified by unknown data.

3.2. Bag of Visual Words Method

Bag of Visual Words (BoVW) has been used in image classification for a long time [22]. This method guided by the bag-of-words models that has been used in the analysis of documents. BoVW represents the images by a histogram. The BoVW histogram has the number of occurrences of visual words of trained codebook that were detected in the training images.

In our study, we use this method as follows. The BoVW method obtains the images of food gains through pre-processing, then extracts hand-engineered features called Speeded-up Robust Features (SURF) features [23], and finally we use K-means clustering to get the trained codebook. the dictionary histogram in the image codebook was serviced as the final feature of the image. Support vector machine (SVM) classify the final features. SVM is a method that map data in a higher dimensional space by a kernel function and selecting the maximum-margin hyper-plane dividing training data [24].

This method has two parts. the training process (Figure 3) and the classification (Figure 4).

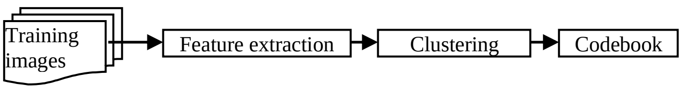


Figure 3 Codebook generation.

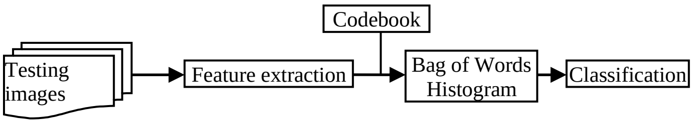


Figure 4 Classification using BoVW.

3.3. Transfer Learning Method

Convolutional neural networks (CNN) demonstrate potential for image classification tasks by learning high level features directly from the image data instead of using low level hand engineered features [15]. It has been shown that the features that CNN extracts from images have a layered structure that corresponds to the structure of the network, but this is also morphologically like the hierarchical structure found in the visual cortex of the biological brain. It should be noted that the hierarchical structure of the features of the lower layers are more universal and are common to different tasks. In other words, the features learned in one recognition task can be diverted to another recognition task, this technique is called transfer learning.

However, CNN are difficult to train from scratch due to small sample sizes produce overfitting. Therefore, we take a method called Dropout to prevent overfitting. In Dropout, when learning with a neural network, when updating to the next layer, instead of using all the units in the previous layer to find the unit in the next layer, some of the units in the previous layer are selected[25].

In this paper, we perform feature extraction from pre-trained VGG-19 and Inception-V3 CNN Models [26]. All two models were pretrained on the ImageNet dataset. then machine learning algorithms such as logistic regression, SVM, and KNN used to learn results.

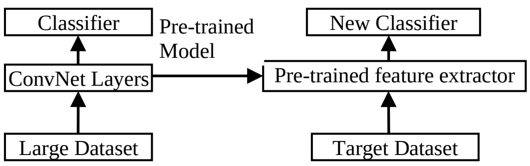


Figure 5 Transfer Learning Method.

The structure of VGG-19 and Inception-V3 is briefly presented as follows:

- *VGG-19 Model*: the structure of VGG-16 including convolutional layer and fully connected layer (16 layers in total). The number below the convolution layer represents the number of convolution filters. The size of the convolution filter is all 3x3. The fully connected layer consists of 2 layers of 4096 units and 1-layer of 1000 units for classification.
- *Inception-V3 Model* is a network that propagates losses for each set of parallel convolution layers called the Inception module.

The basic block in Inception-v3 is shown in Figure 6.

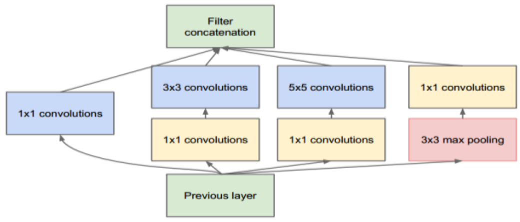


Figure 6 Original Inception module.

according to classes in the dataset we changed the final layer to eleven neurons.

Table 1 presents Vgg-16 architecture and Inceptionv3 architecture. The Vgg-16 network has 13 convolutional layers (blocks). Each block consists of 3 or 3 convolutional layers. the Inceptionv3 has 5 convolutional layers and 9 inception modules, which contain between 4 and 10 convolutional layers each. The last layers of the networks are fully connected with 1000 neurons.

Table 1 **Vgg-16 Architecture (left) and Inceptionv3 Architecture(right).**

Layers	Filter size	Number N of planes (filters)	Layers	Filter size	Number N of planes (filters)
2 × conv2D	3 × 3	64	conv2D	3 × 3	32
Max pooling: F = 2 and S = 2			conv2D	3 × 3	32
2 × conv2D	3 × 3	128	conv2D	3 × 3	64
Max pooling: F = 2 and S = 2			Max pooling: F = 3 and S = 2		
3 × conv2D	3 × 3	256	conv2D	1 × 1	80
Max pooling: F = 2 and S = 2			conv2D	3 × 3	192
3 × conv2D	3 × 3	512	Max pooling: F = 3 and S = 2		
Max pooling: F = 2 and S = 2			3 × inception modules		
3 × conv2D	3 × 3	512	4 × inception modules		
Max pooling: F = 2 and S = 2			2 × inception modules		
4096 nodes fully connected			Average pooling: F = 8		
4096 nodes fully connected			1000 nodes fully connected with softmax activation		
1000 nodes fully connected with softmax activation					

4. Implementation Details and Results

We make the experiments on an Intel(R) Core(TM) i5-4200M CPU @ 2.50GHz 2.50 GHz with windows OS.

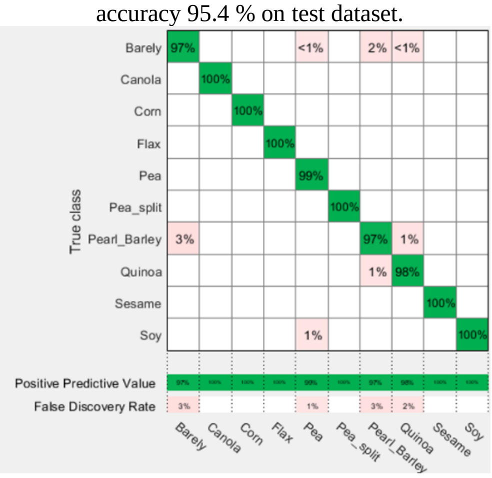
4.1. Bag of Visual Words Method

- *Preparing images datasets:* the dataset was split into two subsets called the training set and testing dataset. 70% images for training dataset and the remainder 30% images for test datasets. we use 1000 images for each category of food grains type. The training set has 750 images and testing set has 250 images for each of the food grains type.
- *Training:* After converting color images to gray images, key points are densely sampled from the gray images in the training set, we extracted SURF features and generated a visual-word dictionary of 500 words using k-means. Finally, histograms of visual words represent images in the dataset, and this is used as feature vectors for classification. We used multi-class classification method, an ensemble of binary SVM classifiers called Error-Correcting Output Codes framework

(ECOC) and 5-fold cross-validation accuracy was used to prevent overfitting.

- *Testing:* To evaluate the classifier using test dataset. We display the confusion matrix in the table 1. A confusion matrix is a cross-table of prediction and correct classes. This is useful for evaluating the identification performance. The average accuracy rate was 95.4% on test dataset. There are false positive predictive values in barely, pea, and quinoa images.

Table 2 Confusion matrix to evaluate BoVW, with average accuracy 95.4 % on test dataset.



4.2. Off-the-shelf deep features Method

- *Preparing images datasets*: We apply normalization by subtracting the mean of all images from each image. Additionally, rescale each image to 224-by-224 and 299-by-299 so that its dimensions correspond to the original VGG-19 and Inception-V3 models input requirements.

  Training: Pretrained model weights were provided using Keras framework. we replaced the original 1000 dimensions fully connected layer with a 11 dimensions fully connected layer. In CNN as feature extractor, the activations of the last fully-connected layer are outputted as features. 4096 features extracted from each image using VGG-19 and 2048 features extracted from each image using Inception-V3. Models were built with the output features using several machine learning algorithms such as logistic regression, SVM, and KNN.

- *Evaluation:* through 10-fold cross-validation and random sampling using a training set size of 80% yielded ROC AUC scores ranging from 0.993 to 0.999 using VGG-19 model and ROC AUC scores

Table 3 Comparison of classification test accuracies achieved with VGG-19 and Inception-V3.

	VGG-19					Inception-V3				
Method	AUC	CA	F1	Precision	Recall	AUC	CA	F1	Precision	Recall
kNN	0.993	0.942	0.942	0.945	0.942	0.998	0.977	0.977	0.977	0.977
SVM	0.998	0.955	0.956	0.959	0.955	1	0.992	0.992	0.992	0.992
Logistic Regression	0.999	0.987	0.987	0.987	0.987	1	0.993	0.993	0.993	0.993

5. Conclusion

In this study, a novel approach to feature extraction from bulk food grains image samples using both Bag of Visual Word approach and Off-the-shelf deep features: using pre-trained neural network as feature extractor, and then training a traditional classifier (SVM) using the extracted features.

These results suggest that Off-the-shelf deep features from deep learning neural networks, such as VGG-19 and Inception-V3, could yield considerable improvements in food grains image analysis over what could be achieved by use of previously proposed method. Looking at the confusion matrix, we believe the main causes for misclassification is that food grains images have fine differences between species.

References

[1] Zhai Z., et al., “A Mission Planning Approach for Precision Farming Systems Based on Multi-Objective Optimization,” *Sensors*, vol. 18, no. 6, 2018, pp. 1795.

[2] Lillford P. and Hermansson A.-M., “Global missions and the critical needs of food science and technology,” *Trends in Food Science & Technology*, vol. 111, 2021, pp. 800-811; DOI https://doi.org/10.1016/j.tifs.2020.04.009.

[3] Braun A.-T., et al., “Farming in the Era of Industrie 4.0,” *Procedia CIRP*, vol. 72, 2018, pp. 979-984; DOI https://doi.org/10.1016/j.procir.2018.03.176.

[4] Sharma R., et al., “A systematic literature review on machine learning applications for sustainable agriculture supply chain performance,” *Computers & Operations Research*, vol. 119, 2020, pp. 104926; DOI https://doi.org/10.1016/j.cor.2020.104926.

[5] Too E.C., et al., “A comparative study of fine-tuning deep learning models for plant disease identification,” *Computers and Electronics in Agriculture*, 2018; DOI https://doi.org/10.1016/j.compag.2018.03.032.

[6] Manavalan R., “Automatic identification of diseases in grains crops through computational approaches: A review,” *Computers and Electronics in Agriculture*, vol. 178, 2020, pp.

105802; DOI https://doi.org/10.1016/j.compag.2020.105802.

[7] Zhang B., et al., "Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review," *Food Research International*, vol. 62, 2014, pp. 326-343; DOI https://doi.org/10.1016/j.foodres.2014.03.012.

[8] Velesaca H.O., et al., "Computer vision based food grain classification: A comprehensive survey," *Computers and Electronics in Agriculture*, vol. 187, 2021, pp. 106287; DOI https://doi.org/10.1016/j.compag.2021.106287.

[9] Paliwal J., et al., "Comparison of a Neural Network and a Non-parametric Classifier for Grain Kernel Identification," *Biosystems Engineering*, vol. 85, no. 4, 2003, pp. 405-413; DOI https://doi.org/10.1016/S1537-5110(03)00083-7.

[10] Choudhary R., et al., "Classification of cereal grains using wavelet, morphological, colour, and textural features of non-touching kernel images," *Biosystems Engineering*, vol. 99, no. 3, 2008, pp. 330-337; DOI https://doi.org/10.1016/j.biosystemseng.2007.11.013.

[11] Mebatsion H.K., et al., "Automatic classification of non-touching cereal grains in digital images using limited morphological and color features," *Computers and Electronics in Agriculture*, vol. 90, 2013, pp. 99-105; DOI https://doi.org/10.1016/j.compag.2012.09.007.

[12] Olgun M., et al., "Wheat grain classification by using dense SIFT features with SVM classifier," *Computers and Electronics in Agriculture*, vol. 122, 2016, pp. 185-190.

[13] Wafy M., et al., "Identification of weed seeds species in mixed sample with wheat grains using SIFT algorithm," *Proc. Computer Engineering Conference (ICENCO), 2013 9th International*, IEEE, 2013, pp. 11-14.

[14] Chen T., et al., "Seed Identification of Gramineous Grass Using Local Similarity Pattern and Linear Discriminant Analysis," *The Open Cybernetics & Systemics Journal*, vol. 11, no. 1, 2017.

[15] Krizhevsky A., et al., "Imagenet classification with deep convolutional neural networks," *Proc. Advances in neural information processing systems*, 2012, pp. 1097-1105.

[16] Xinshao W. and Cheng C., "Weed seeds classification based on PCANet deep learning baseline," *Proc. Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2015 Asia-Pacific*, IEEE, 2015, pp. 408-415.

[17] Liu C. and Cai C., "Weed Seeds Identification Based on Interlacing PCA Net."

[18] Kozłowski M., et al., "Varietal classification of barley by convolutional neural networks," *Biosystems Engineering*, vol. 184, 2019, pp. 155-165; DOI https://doi.org/10.1016/j.biosystemseng.2019.06.012.

[19] Laroca R., et al., "A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector," *ArXiv e-prints*, February 01, 2018 2018; https://ui.adsabs.harvard.edu/#abs/2018arXiv180209567L.

[20] Maruyama T.M., et al., "Automatic classification of native wood charcoal," *Ecological Informatics*, vol. 46, 2018, pp. 1-7; DOI https://doi.org/10.1016/j.ecoinf.2018.05.008.

[21] Cavalin P. and Oliveira L.S., "A Review of Texture Classification Methods and Databases," *Proc. 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*, 2017, pp. 1-8.

[22] Yang J., et al., "Evaluating bag-of-visual-words representations in scene classification," *Proc. Proceedings of the international workshop on Workshop on multimedia information retrieval*, ACM, 2007, pp. 197-206.

[23] Bay H., et al., "Speeded-up robust features (SURF)," *Computer vision and image understanding*, vol. 110, no. 3, 2008, pp. 346-359.

[24] Foody G.M. and Mathur A., "A relative evaluation of multiclass image classification by support vector machines," *IEEE Transactions on geoscience and remote sensing*, vol. 42, no. 6, 2004, pp. 1335-1343.

[25] Oquab M., et al., "Learning and transferring mid-level image representations using convolutional neural networks," *Proc. Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1717-1724.

[26] Szegedy C., et al., "Going deeper with convolutions," *Proc. Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.

[27] Simonyan K. and Zisserman A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

Abdelmgeid A. Ali received the PhD degree in Computer Science in 1996. Currently he works as a Professor in Computer Science Department, Minia University, El Minia, Egypt. He has published over 80 research papers in prestigious international journals, and conference proceedings. He has supervised over 60 Ph.D. and M. Sc. Students. Prof Ali is a member of the International Journal of Information Theories and Applications (ITA). Prof Ali interests are Information Retrieval, Software Engineering, Image Processing, Data security, metaheuristics, IOT, Digital Image Steganography, Data Warehousing.

Usama S. Mohammed received his B.Sc. and M. Sc. degrees from Assiut University, in 1985 and 1993, respectively, and his Ph.D. degree from Czech Technical University in Prague, Czech Republic, in 2000. From November 2001 to April 2002, he was a post-Doctoral Fellow with the Faculty of Engineering, Czech Technical University in Prague, Czech Republic. He authored and co-authored more than 120 scientific papers. He has been selected for the inclusion in 2010 Edition of the USA-Marquis Who's Who in the World. His research interests include telecommunication technology, wireless technology, wireless Networks, image coding, statistical signal processing, blind signal separation, and video coding.

.

Rehab R. Nour received her B.Sc. and M. Sc. degrees from Minia University, in 1999 and 2010, respectively. She is working toward the Ph.D. at Minia university, her research area is in computer vision.