

RESEARCH

Open Access



Deep indicator for fine-grained classification of banana's ripening stages

Yan Zhang¹, Jian Lian^{1*}, Mingqu Fan^{1*} and Yuanjie Zheng^{2,3*}

Abstract

Determining banana's ripening stages is becoming an essential requirement for standardizing the quality of commercial bananas. In this paper, we propose a novel convolutional neural network architecture which is designed specifically for the fine-grained classification of banana's ripening stages. It learns a set of fine-grained image features based on a data-driven mechanism and offers a deep indicator of banana's ripening stage. The resulted indicator can help to differentiate the subtle differences among subordinate classes of bananas in ripening state. Experimental results from 17,312 images of bananas in different ripening stages show that our deep indicator achieves an accuracy superior significantly to state-of-the-art computer vision-based systems both in rough- and fine-grained classification of ripening stages no matter the bananas bear or not severe defects.

Keywords: Image processing, Machine vision, Image classification, Banana, Ripening stage

1 Introduction

Banana is one of the most consumed fruit globally. It contributes about 16% of the world's fruit production. According to FAO (Food and Agriculture Organization of the United Nations), more than 114 million tons of bananas were produced worldwide in 2014. It was reported that China ranked second among top banana-producers during recent years. Banana is also the biggest tropical fruit in China, which covered an area of 392,000 ha, production of 11,791,900 tons in 2014. China produces more than 110 banana cultivars and the Musa AAA Cavendish cv. Brazil is the most important both for Chinese national and international markets [1].

The quality control and the acceptance of consumers need an optimum ripening stage for bananas [2, 3]. In the last decades, many types of research on the ripening assessment of banana were conducted. The presented methods could be classified into two categories: the biochemical or physiochemical property-based methods and the computer vision-based techniques. A plethora of

methods were presented to understand the ripening process of bananas through various bioindicators including starch content, soluble solid content, sugar content, and firmness. The primary objective of these methods is to discover the relationship between the ripening process and the physical, biochemical, or nutritional transformations [4–7]. These methods can supply fine-grained classification results of banana during the ripening process. For example, in [4] a starch staining instrument is used to differentiate the banana maturity through the disappearance of the starch. With the accurate determination of starch content, the ripening process of banana can be divided into more than 50 stages. However, most of these methods are laborious and expensive due to the involvement of invasive or destructive techniques.

Many computer vision-based approaches were also proposed to classify the ripening stages of bananas based on the appearance of bananas and various computer based algorithms [8]. These computer vision techniques [9–11] can potentially provide an automated and non-destructive tool for the classification of ripening banana. However, none of them has been widely used due to several limitations. First, they have rarely paid attention to the fine-grained classification and are therefore incapable of differentiating subtle differences among subordinate classes of ripening bananas. Due to the limitation in using

*Correspondence: lianjianlian@163.com; fmqskd@163.com; zhengyuanvision@outlook.com

¹Department of Electrical Engineering Information Technology at Shandong University of Science and Technology, Jinan, 250031, China

²Shandong Normal University, School of Information Science and Engineering, Jinan 250358, China

Full list of author information is available at the end of the article

the skin color of banana and of the computer vision system itself, these methods are usually able to divide the ripening process into seven stages [12]. Second, these methods exploit hand-crafted features, resulting in limited performances due to the hardships in the manual design of features. Third, most of them deteriorate for lower-class bananas due to the difficulties in modeling bananas' skin defects [10].

In the field of artificial intelligence, recent advances in deep learning [13–15] have led to breakthroughs in long-standing tasks such as vision-related problems of feature extraction [16, 17], image segmentation [18–20], and image classification [21–26]. Among all these techniques, convolutional neural network (CNN) is one of the most successful methods [13] and has acquired a broad application in image classification. Recently, CNN-based models of fine-grained image classification have obtained tremendous advances in recognition of subtle difference among subordinate classes, including the traffic classification [27, 28], medical image classification [29], plant classification [30], and food classification [31].

In this paper, we present a deep indicator for fine-grained classification of the precise ripening stages of bananas based on images. It is accomplished through a novel CNN architecture designed specifically for the unique characteristics of banana appearance. The proposed CNN framework takes triple images as input, from which the triplet loss (similarity loss) is produced. Through the joint optimization of classification accuracy and similarity loss, our proposed technique can effectively learn the fine-grained feature representations from the ripening process of a banana.

The proposed technique bears at least three advantages. First, it leverages a training process to extract automatically multi-scale image features combining both the global and local features of the banana image. The mapping learned from these fine-grained features offers a function to automatically specify the ripening stages of banana given a new input image of a banana. Experimental results from a large number of 17,312 images of bananas show that the proposed CNN architecture produces an impressive classification accuracy of 94.4% at the laboratory level. Second, as the samples of bananas with imperfections are also included in the experiments, the experimental results also validate that our method can be applied to fine-grained classification of bananas at different ripening stages no matter the banana peel bears defects or not. Third, different from some existing methods [32], our indicator does not combine any information from the physiochemical or biological changes during the ripening process of banana. However, it is still able to achieve competitive classification accuracy as validated by our experiments.

Our work offers at least four significant contributions as below:

- (I) To the best of our knowledge, this paper is the first attempt to introduce CNN into banana ripening assessment, while previously proposed methods mainly are based on traditional machine vision-based algorithms.
- (II) We propose a novel CNN architecture adapted to fine-grained-feature based representation for a subtle classification of banana's ripening stages. And the state-of-the-art methods mainly aimed at coarse classification of the ripeness of bananas.
- (III) The proposed approach is data-driven and employs an integrated system which combines a set of learning functions for both features and the feature-to-class mapping.
- (IV) Our approach performs with an impressive superiority to the state-of-the-art techniques.

The rest of this paper is organized as follows. In Section 2, we present the details of the materials we used and our approach. Section 3 contains our experimental results and discussions. In Section 4, we provide our conclusion and vision for the future.

2 Materials and methods

2.1 Fruit selection and sampling

Twenty batches of bananas (Musa AAA Cavendish) cv. Brazil were purchased from a local wholesale market. To avoid the sample variability of the samples, 197 bananas with similar size, color, and weight (200–245 g) were selected for the following analysis.

The bananas were sanitized with 1% NaOCl for 15 min, then were rinsed with distilled water and dried at ambient temperature. During the experiments, they were stored under conditions of darkness at 25 °C and 75% of relative humidity for 14 days in a stability chamber (Fig. 1) (Sailham 523000, China).

2.2 Computer vision system

A computer vision system (as shown in Fig. 2) was built to capture the images of bananas during the ripening process. This system includes a lighting system with a ring fluorescent lamp (FGR series, Wordop, China) to avoid inhomogeneous lighting. The camera (Canon EOS 760D, Japan) settings is the manual mode, exposure level 0.0, without zoom, and flash. The white balance is set to the white fluorescent light. The samples (bananas) are placed on a piece of white paper placed on a steady table. The color space applied in subtle classification of ripening stages of banana is the sRGB (standard RGB). The size of the captured images is 3200 * 2400 pixels (0.1 mm/pixel) and then stored in the server in PNG (Portable Network Graphic) format.

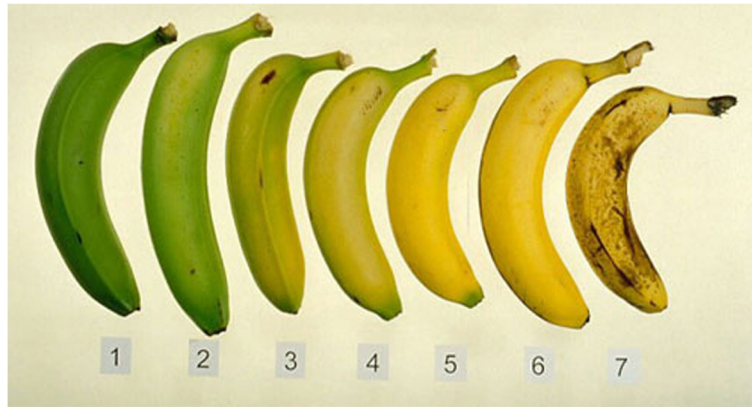


Fig. 1 Image of the bananas at selected ripening stages (from left to right). Each banana represents a stage during the ripening process

2.3 The proposed CNN architecture

CNN is typically a multilayer, hierarchical neural network but bears at least three principal factors different from a generic neural network: local receptive fields, weight sharing, and spatial pooling layers [33]. CNN employs a local receptive field rather than a global one, which is similar to the brain capturing local structure of image through constraining each neuron to depend only on a spatially local subset of the neurons in the proceeding layer. Moreover, weights are shared across different neurons in the

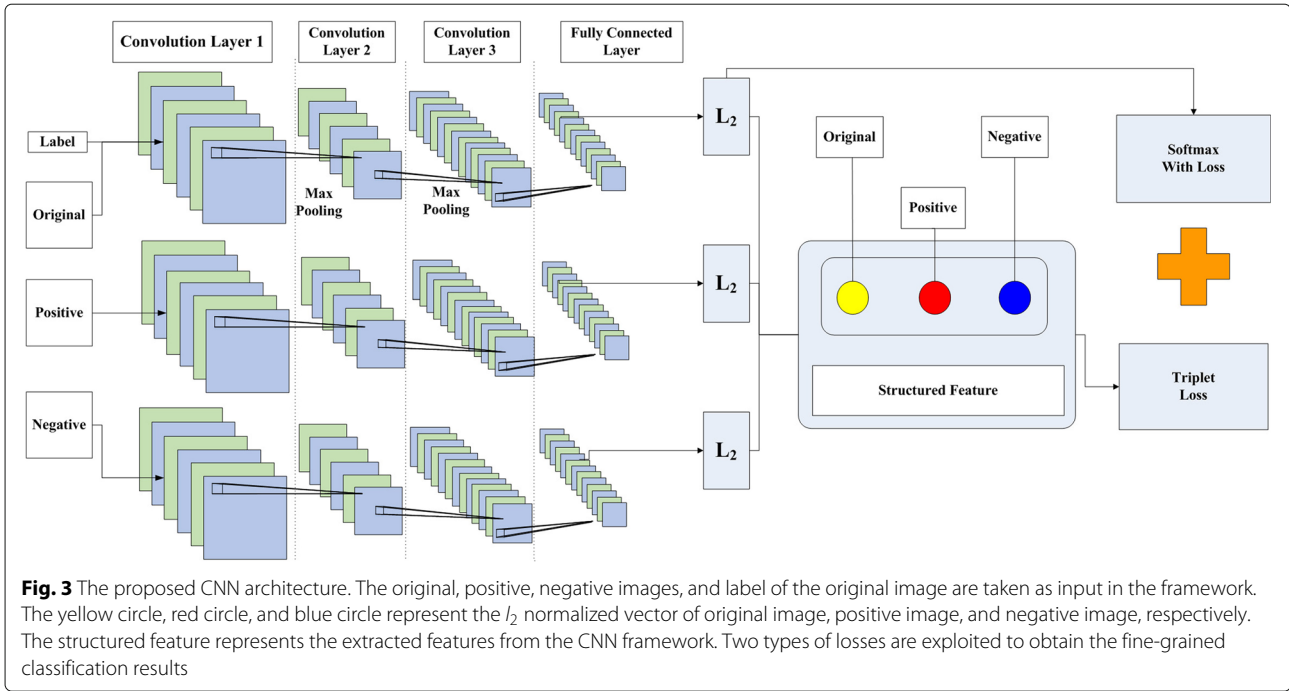
same layer, which can be translated to evaluating the same filter over all local windows of the input image. Spatial pooling in CNN is to divide the image into an array of blocks and then evaluate a pooling function over the responses in each block. The goal of pooling is to reduce the dimensionality of the convolutional responses and enforce a translational invariance (in a small degree) into the model. In the case of max pooling, the response for each block is taken to be the maximum value over all response values within the block. A typical CNN consists of multiple layers, alternating between convolution and pooling. Compared with shallow CNN architectures, deep CNN has more hidden layers. Lower layers defined as the ones closer to the input construct the low-level convolutional filters which can be thought of as providing low-level encoding of the input image. In contrast, higher layers learn more complicated structures. In CNN, stride length is used to specify the number of pixels with which the local receptive field is moved to the right (or down).

To classify the bananas at subtle ripening stages, we propose a novel architecture of CNN to predict the probability of classification of the inputted image. It is composed of convolutional layer with rectified linear unit (ReLU), max pooling and fully connected layer with ReLU, as shown in Fig. 3. For each image, a positive image (at the same ripening stage as the original image) and a negative image (not at the same ripening stage) are chosen from the captured images of bananas. Then, three images and the label of the original image are jointly inputted into our proposed CNN structure. Three parameter-sharing CNNs are presented to handle the original, positive, and negative image, respectively. The structure of the CNN is shown in Fig. 3.

- Convolutional layer 1. 48 kernels of size $3 \times 7 \times 7$ (3 represents the number of RGB channels, with a stride of 2) are applied to the input banana image in the first



Fig. 2 The proposed computer vision system. The system consists of a camera, a ring fluorescent lamp, the white paper under the banana, and a steady table



layer combined with the ReLU. A max pooling layer follows this convolutional layer.

- Convolutional layer 2. 128 kernels of size $3 \times 5 \times 5$ (with a stride of 2) are applied to the input banana image in the second layer combined with the ReLU. A max pooling layer follows this convolutional layer.
- Convolutional layer 3. 128 kernels of size $3 \times 3 \times 3$ (with a stride of 2) are applied to the input banana image in the third layer combined with the ReLU.
- Fully connected layer. 512 neurons combined with ReLU, which is used to perform high-level reasoning like neural networks.

The structured feature in Fig. 3 displays the features that could be extracted from the original image, positive image, and negative image by the proposed CNN framework. Our architecture serves as a baseline to naturally embed label structures without sacrificing the classification accuracy.

One softmax loss operation locates at end of the CNN channel for original image. The corresponding softmax loss function is shown as

$$L_s = - \sum_{i=1}^N \log (P(\omega_k | (L_i))), \quad (1)$$

where N denotes number of the input images, $P(\omega_k | L_i)$ indicates the probability of the k th image to image label L_i correctly classified as l_k .

Besides the softmax loss function used that has been exploited in other CNNs, the triplet loss is also presented in our proposed CNN architecture.

$$L_t = \frac{1}{2N} \sum_{i=1}^N \max (0, D(o_i, p_i) - D(o_i, n_i) + m), \quad (2)$$

where $D(.,.)$ is the squared Euclidean distance between two l_2 normalized vectors; o_i, p_i, n_i , respectively denote the l_2 normalized vectors from original image, positive image, and the negative image, as shown in Fig. 3; m denotes the hyper-parameter to confine the value of Eq. (2) greater than zero as the difference between the original image and the positive image is expected to be greater than the difference between the original image and the negative image.

As shown in Eq. (3), the two losses are integrated to obtained the fine-grained indicator of classification.

$$L = \lambda L_s + (1 - \lambda) L_t, \quad (3)$$

where λ denotes the weight that is used to manipulate the trade-off between the softmaxloss (L_s) and the tripletloss (L_t).

3 Results and discussion

3.1 Dataset and preprocessing

We collected 17,312 images of bananas at different ripening stages (the whole process is 14 days long). There are about 30 images captured for every banana everyday. To

overcome the potential over-fitting effect, we enlarge the original dataset with data augmentation methods including translations (varying from 10 to 100 pixels with a gap of 10 pixels), vertical, and horizontal reflections. After the data augmentation, the images are then resized into 256×256 . For each image, a positive image and a negative image as mentioned in Section 2.3 are chosen from the captured images to form a triplet.

3.2 Training and evaluating

We manually labeled the samples into 7 categories and 14 categories (according to the date of image capturing), respectively. Fifty percent of the images are taken as training dataset, 30% are chosen into the evaluation dataset, and the other 20% are used in the testing process. In the training process, the proposed framework is refined with the back propagation mechanism which calculates the minimization of the squared difference between the classification ground truth and the corresponding output prediction, iteratively. The training is performed on GPU of high performance and implemented in TensorFlow [34]; it takes 10^5 iterations. For each iteration, it costs about 0.5 s.

3.3 Experimental results

After the training process, we conducted experiments to testify the performance of our proposed CNN architecture. We choose several state-of-the-art image classification methods [35–38] to compare with our proposed method. Different features including single feature and combined features are exploited by these SVM-based methods, respectively.

As shown in Fig. 4, the accuracy of our method for 7 categories classification is 94.4% after 31 iterations. Meanwhile, the training loss of our method decreases to 0.15. Note that the training loss's decrease indicates the convergence of the CNN architecture.

Figure 5 shows the classification result of our approach for distinguishing the 14 ripening stages. The results demonstrate that the accuracy obtained by our method is 92.4% after 34 iterations, and the training loss of our method decreases to 0.22. We also find that the proposed method can perform well in the fine-grained classification of bananas during the ripening process.

Meanwhile, to further compare the performance of our method and state-of-the-art methods, the precision and recall were calculated with the following two equations, respectively.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (5)$$

where TP, FP, and FN denote true positive, false positive, and false negative, respectively.

As revealed by the results for 7- and 14-category classifications in Table 1, the proposed deep indicator outperforms state-of-the-art methods.

We show the feature maps extracted by our CNN architecture in Fig. 6. We noticed that the extracted features combine information from color, shape, and texture of the banana.

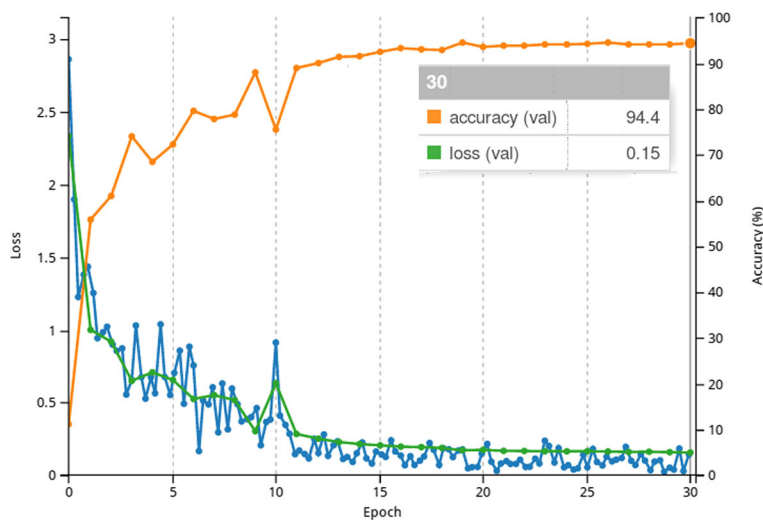


Fig. 4 The classification accuracy for seven ripening stages of banana. The orange line denotes the accuracy of classification during the training and validation process. The green line and blue line represent the loss of training and loss of validation, respectively. Epoch (the X axial) represents the iteration numbers of the classification in the evaluation. Loss and accuracy (the Y axial) represents the error and the precise rate during the evaluating process

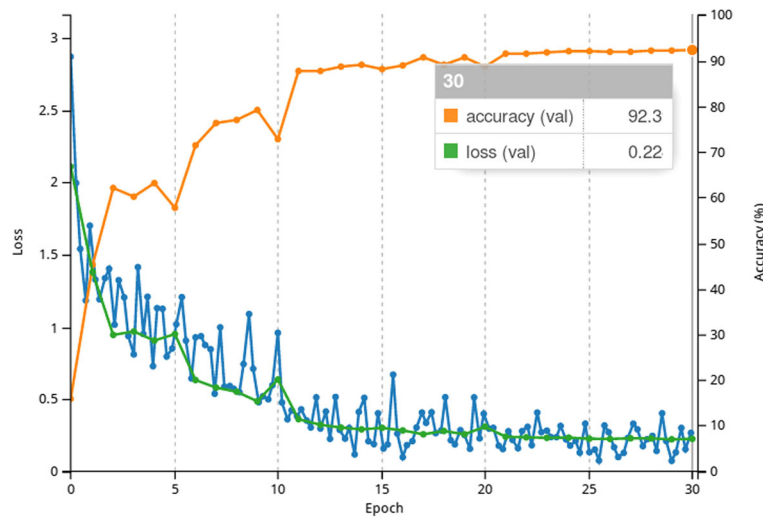


Fig. 5 The classification accuracy for 14 ripening stages of banana. The orange line denotes the accuracy of classification during the training and validation process. The green and blue lines represent the loss of training and loss of validation, respectively. Epoch (the X axial) represents the iteration numbers of the classification in the evaluation. Loss and accuracy (the Y axial) represents the error and the correctness during the evaluating process

Meanwhile, we chose the images with severe defects as shown in Fig. 7 to testify the performance of our proposed method in relatively extreme situations. The precision and recall values generated from these images are shown in Table 2. The receiver operating characteristics curves (ROCs) of our method and state-of-the-art methods are shown in Fig. 8 and the corresponding areas under curves (AUCs) are included in Table 3. And ROC and AUC are classical measurement used to assess the classification results [39]. We use Z test to compare the statistical difference of AUC between state-of-the-art methods and our approach, as shown in Table 3.

There are mis-classifications due to the severe blemishes in these images. Most of the mis-classifications are caused by extreme viewing conditions, e.g., too many black spots appearing in the image. However, the experimental result shows that even under an extreme situation, our method can still obtain more satisfactory performance than the state-of-the-arts.

3.4 Discussion

From the results shown previously, we can see that the proposed CNN architecture presents a deep indicator (to indicate the ripeness of the banana with a deep CNN architecture) which performs accurately no matter for a rough classification or a fine-grained classification of the banana's ripening stages. This indicator outperforms state-of-the-art techniques for bananas with or without severe defects and its improvements are significant (as shown in Table 3).

Similar to human being's visual system, our proposed approach can extract the global and local features of the images of bananas. The global features combined with local features form a layout for each category of bananas. Figure 6 shows that in the first convolutional layer, the global features including shape of the banana are extracted. Then, in the two following convolutional layers, other features including color and texture are extracted hierarchically. Unlike the manually extracted features used

Table 1 Performances on classifying 7 and 14 ripening stages of banana

Methods	7 categories		14 categories	
	Precision	Recall	Precision	Recall
Gabor + SVM [35]	85.2%	83.2%	70.1%	69.1%
Wavelet + SVM [36]	86.5%	87.9%	73.4%	71.9%
Wavelet + Gabor + SVM [37]	88.2%	86.4%	77.6%	75.3%
Combined features + SVM [38]	89.2%	86.1%	82.3%	85.7%
Our method	95.6%	92.3%	93.5%	93.1%

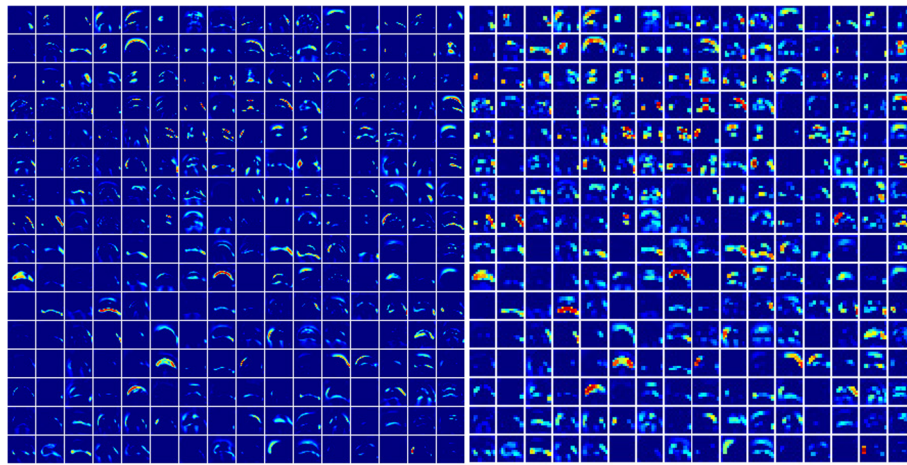


Fig. 6 (left) The feature maps extracted from the convolution layer 1. (right) The feature maps obtained after the convolution layer 3. Both the convolution layers are shown in Fig. 3

in [35–38], these features are automatically extracted by our proposed CNN architecture. As illustrated in Fig. 6, the global and local features of banana image are automatically extracted by our proposed CNN architecture. The map between the input banana image and the output classification result could be obtained through the extracted features.

Our proposed CNN framework can significantly enhance the performance of image classification by jointly optimizing the classification loss (between the label of the original image and the output classification result) and the similarity loss (between the original image, positive image, and negative image). The parameter λ in Eq. (3) that is used to balance the softmax loss and loss plays an essential role in the proposed CNN architecture. With the λ is set to 1 or 0, the performance of the structure would degenerate to softmax loss or triplet loss, respectively. According to the process of error and trial, it is reasonable to assign a value greater than 0.5 to λ , which shows that the softmax loss should be more important than triplet loss in our pro-

posed CNN architecture. To note that the introduction of triplet loss contributes substantially to the image classification according to the complementary information from both the positive and negative images. Meanwhile, it could encourage the intra-class similarity and inter-class difference at the same time. The proposed CNN architecture is suitable for characteristics of banana. Through combining the softmax loss with the newly presented triplet loss, the subtle difference between intra-category and inter-categories of bananas during different ripening stages can be differentiated from each other.

4 Conclusions

In this paper, we present a deep indicator of banana's ripening stages based on a novel CNN architecture, which offers a unique tool for achieving an automated and non-destructive fine-grained classification of banana maturity. The proposed deep indicator integrates the capabilities of accurate fine-grained classification and non-invasive examination. The former can be currently achieved by



Fig. 7 Two exempling images with severe defects

Table 2 Performances on images of bananas with severe defects

Methods	Fine-grained classification	
	Precision	Recall
Gabor + SVM [35]	65.6%	63.7%
Wavelet + SVM [36]	71.3%	69.7%
Wavelet + Gabor + SVM [37]	75.7%	73.2%
Combined features + SVM [38]	80.4%	77.4%
Our method	92.2%	90.8%

Table 3 The AUCs and AUC group testing performance of the comparing methods

Methods	AUC	Z ($p = 0.05$)
Gabor + SVM [35]	0.62	8.71
Wavelet + SVM [36]	0.69	6.22
Wavelet + Gabor + SVM [37]	0.72	6.35
Combined features + SVM [38]	0.76	5.73
Our method	0.91	-

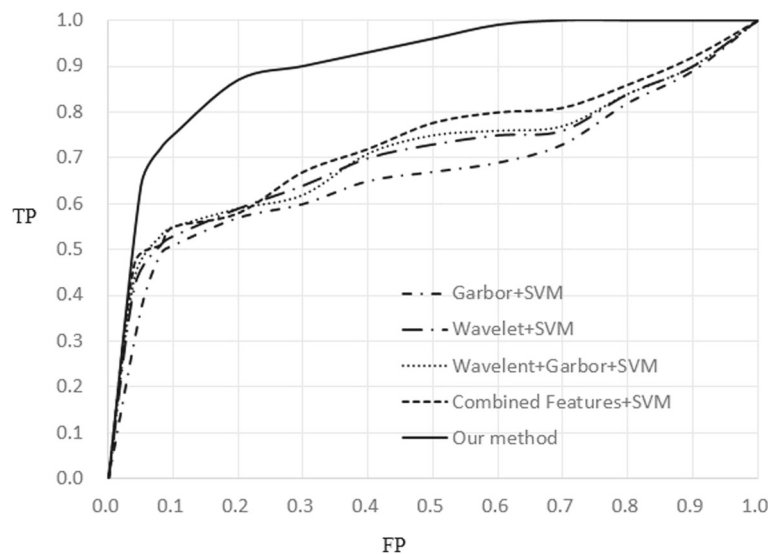
Z represents the group testing value between each state-of-the-art method and our method with Z test

bioindicators but not the computer vision systems. The latter is an advantage of current computer vision systems but not the bioindicators. In the proposed CNN architecture, three parameter-sharing CNNs are exploited to handle each of the three input images: the original image, positive image, and negative image. At the end of the CNN framework, the structured feature of the triplet input could be obtained. Then, a softmax loss integrated with a triplet loss are presented to implement the fine-grained classification. To evaluate the performance of our method, we take advantage of a large image dataset which consists of 17,312 images from bananas with or without defects.

This paper offers several contributions. First of all, this is probably the first attempt to introduce deep learning strategy into the fine-grained classification of bananas during different ripening stages. Secondly, the triplet loss submitted by our method positively affects the performance of image classification. To the best of our knowledge, this is also an early application of the

similarity between the original image, positive image, and negative image into the CNN classifier. Thirdly, similar to human being's visual system our proposed CNN framework can extract the multi-scale features including both the global and local features of the images of bananas. Finally, our approach performs with an obvious superiority to the state-of-the-art image classification techniques.

In our future works, we will delve into the construction of different CNN architectures and explore the process of implicit feature extraction within CNN. Meanwhile, we will research on the applications of our proposed CNN-based image classification method in other tasks, e.g., the classification of fruits, medical image analysis [40], and industrial products. Furthermore, to leverage on the multiple modality images including the RGB and the infrared, we would continue to study the application of CNN in multiple modality image processing [41].

**Fig. 8** ROCs generated by the comparing methods for fine-grained classification of bananas with flaws

Abbreviations

AUC: Area under curve; CNN: Convolutional neural network; FAO: Food and Agriculture Organization of the United Nations; PNG: Portable network graphic; ReLU: Rectified linear unit; ROC: Receiver operating characteristics; sRGB: Standard RGB

Acknowledgements

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

Funding

This work was supported by the Natural Science Foundation of China (NSFC) (61572300), Natural Science Foundation of Shandong Province in China (ZR2014FM001), Taishan Scholar Program of Shandong Province in China (TSHW201502038), and SDUST Excellent Teaching Team Construction Plan JXTD20160512.

Availability of data and materials

We can provide the data.

Authors' contributions

All authors take part in the discussion of the work described in this paper. The author JL puts forward the main idea. The author YZ wrote the first version of the paper and did part experiments of the paper. JL and MF revised the paper in different version of the paper. The contributions of the proposed work are mainly in two aspects: (1) to the best of our knowledge, our work is the first one to apply the convolutional neural network for the discrimination of banana's ripening stages. In this paper, we propose a new convolutional neural network architecture designed specifically for analyzing the banana images. Experimental results show that our system performs accurately in fine-grained classification of banana's ripening stages, which is beyond the ability of the related state-of-the-art computer vision techniques. At the same time, our system outperforms related state-of-the-art computer vision techniques in rough classification of banana maturity. (2) The novelty of our method attributes to our proposed approach outperforms related state-of-the-art computer vision techniques in rough classification of banana maturity. All authors read and approved the final manuscript.

Authors' information

Yan Zhang is now a doctor student in Shandong University of Science and Technology. Her research interest includes information management, machine vision, and deep learning.

Jian Lian is now an instructor in Shandong University of Science and Technology. His interest includes machine learning and image processing. Mingqu Fan is currently a professor at Shandong University of Science and Technology. His research is in the fields of image analysis, computer vision, and computational photography.

Yuanjie Zheng is currently a professor at Shandong Normal University. His research is in the fields of medical image analysis, translational medicine, computer vision, and computational photography.

Ethics approval and consent to participate

This study was approved by the ethics committee at Shandong University of Science and Technology (Jinan, China).

Competing interests

These no potential competing interests in our paper. And all authors have seen the manuscript and approved to submit to your journal. We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Electrical Engineering Information Technology at Shandong University of Science and Technology, Jinan, 250031, China. ²Shandong Normal University, School of Information Science and Engineering, Jinan 250358, China. ³Key Lab of Intelligent Computing and Information Security in Universities of Shandong, Institute of Life Sciences, Shandong Provincial Key

Laboratory for Distributed Computer Software Novel Technology, Key Lab of Intelligent Information Processing, Jinan 250358, China.

Received: 9 April 2018 Accepted: 24 May 2018

Published online: 08 June 2018

References

- FAOSTAT, Food and, Agricultural Organization, Geneva (2014). <http://www.fao.org/statistics/en/>
- D Wu, DW Sun, Colour measurements by computer vision for food quality control - a review. *Trends Food Sci. Technol.* **29**(1), 5–20 (2012)
- A Sanaeifar, SS Mohtasebi, M Ghasemivarnamkhasti, H Ahmadi, J Lozano, Development and application of a new low cost electronic nose for the ripeness monitoring of banana using computational techniques (pca, lda, simca, and svm). *Czech J. Food Sci.* **32**(32), 538–548 (2014)
- SM Blankenship, DD Ellsworth, RL Powell, A ripening index for banana fruit based on starch content. *Horttechnology.* **3**(3), 338–339 (1993)
- M Soltani, R Alimardani, M Omid, Prediction of banana quality during ripening stage using capacitance sensing system. *Australian J. Crop Sci.* **4**(6), 443–447 (2010)
- PP Subedi, KB Walsh, Assessment of sugar and starch in intact banana and mango fruit by swinr spectroscopy. *Postharvest Biol. Technol.* **62**(3), 238–245 (2011)
- JM Cardoso, RDS Pena, Hygroscopic behavior of banana (musa ssp.aaa) flour in different ripening stage. *Food Bioproducts Process.* **92**(1), 73–79 (2014)
- JFS Gomes, R Vieira, FR Leta, Colorimetric indicator for classification of bananas during ripening. *Sci. Hortic.* **150**, 201–205 (2013)
- F Mendoza, JM Aguilera, Application of image analysis for classification of ripening bananas. *J. Food Sci.* **69**(9), E471–E477
- GC Bora, D Lin, P Bhattacharya, SK Bali, R Pathak, Application of bio-image analysis for classification of different ripening stages of banana. *J. Agric. Sci.* **7**(2), 152 (2015)
- J Hou, Y Hu, L Hou, K Guo, T Satake, Classification of ripening stages of bananas based on support vector machine. *Int. J. Agric. Biol. Eng.* **8**(6), 99–103 (2015)
- HW Von Loesecke, *Bananas*. (Interscience Publishers, 1949), pp. 595–599
- Y LeCun, L Bottou, Y Bengio, P Haffner, Gradient-based learning applied to document recognition. *Proc. IEEE.* **86**(11), 2278–2324 (1998)
- GE Hinton, S Osindero, Y-W Teh, A fast learning algorithm for deep belief nets. *Neural Comput.* **18**(7), 1527–1554 (2006)
- Y Bengio, P Lamblin, D Popovici, H Larochelle, et al., Greedy layer-wise training of deep networks. *Adv. Neural Inf. Process. Syst.* **19**, 153 (2007)
- K Simonyan, A Zisserman, Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. is a standard citation format of arXiv.com
- Y Jia, E Shelhamer, J Donahue, S Karayev, J Long, R Girshick, S Guadarrama, T Darrell, in *Proceedings of the ACM International Conference on Multimedia*. Caffe: Convolutional architecture for fast feature embedding (ACM, Orlando, 2014), pp. 675–678
- SC Turaga, JF Murray, V Jain, F Roth, M Helmstaedter, K Briggman, W Denk, HS Seung, Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Comput.* **22**(2), 511–538 (2010)
- A Prasoon, K Petersen, C Igel, F Lauze, E Dam, M Nielsen, in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network (Springer, Nagoya, 2013), pp. 246–253
- J Long, E Shelhamer, T Darrell, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Fully convolutional networks for semantic segmentation (IEEE, Los Alamitos, 2015), pp. 3431–3440
- A Krizhevsky, I Sutskever, GE Hinton, in *Advances in neural information processing systems*. Imagenet classification with deep convolutional neural networks (MIT Press, Cambridge, 2012), pp. 1097–1105
- R Socher, B Huval, B Bath, CD Manning, AY Ng, in *Advances in Neural Information Processing Systems*. Convolutional-recursive deep learning for 3d object classification (MIT Press, Cambridge, 2012), pp. 665–673
- K Simonyan, A Vedaldi, A Zisserman, Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*. is a standard citation format of arXiv.com
- A Karpathy, G Toderici, S Shetty, T Leung, R Sukthankar, L Fei-Fei, in *Proceedings of the IEEE conference on Computer Vision and Pattern*

- Recognition. Large-scale video classification with convolutional neural networks (IEEE, Los Alamitos, 2014), pp. 1725–1732
25. W Hu, Y Huang, L Wei, F Zhang, H Li, Deep convolutional neural networks for hyperspectral image classification. *J. Sensors*. **2015**, 1–12 (2015)
 26. E Ahn, A Kumar, J Kim, C Li, et al., in *IEEE, International Symposium on Biomedical Imaging. X-ray image classification using domain transferred convolutional neural networks and local sparse spatial pyramid[C]* (IEEE, 2016), pp. 855–858
 27. S Yu, Z Song, S Su, W Li, Y Wu, W Zeng, A novel dataset generating method for fine-grained vehicle classification with cnn. *Int. J. Database Theory Appl.* **9**(6), 45–52 (2016)
 28. J Fang, Y Zhou, Y Yu, S Du, Fine-grained vehicle model recognition using a coarse-to-fine convolutional neural network architecture. *IEEE Trans. Intell. Trans. Syst.* **PP**(99), 1–11 (2016)
 29. HH Vo, A Verma, in *IEEE International Symposium on Multimedia*. New deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space (IEEE, Los Alamitos, 2016), pp. 209–215
 30. N Sunderhauf, C Mccool, B Upcroft, P Tristan, Fine-grained plant classification using convolutional neural networks for feature extraction. *Proc. Congress Faons*. **37**(2), 123–30 (2014)
 31. K Yanai, Y Kawano, in *IEEE International Conference on Multimedia & Expo Workshops*. Food image recognition using deep convolutional network with pre-training and fine-tuning (IEEE, Los Alamitos, 2015), pp. 1–6
 32. N Velezrivera, J Blasco, J Chanonaperez, G Calderondominguez, MDJ Pereaflares, I Arzatevazquez, S Cubero, RR Farrerarebollo, Computer vision system applied to classification of “manila” mangoes during ripening process. *Food Bioprocess Technol.* **7**(4), 1183–1194 (2013)
 33. Y Lecun, in *Connectionism in Perspective*. Generalization and network design strategies (Elsevier, Amsterdam, 1989)
 34. M Abadi, A Agarwal, P Barham, E Brevdo, Z Chen, C Citro, GS Corrado, A Davis, J Dean, M Devin, et al., Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467* (2016). is a standard citation format of arXiv.com
 35. Z Sun, G Bebis, R Miller, in *International Conference on Digital Signal Processing, vol. 2*. On-road vehicle detection using gabor filters and support vector machines (IEEE, Los Alamitos, 2002), pp. 1019–1022
 36. Z Sun, G Bebis, Miller R, in *International Conference on Control, Automation, Robotics and Vision, vol 3*. Quantized wavelet features and support vector machines for on-road vehicle detection[C] (IEEE, Singapore, 2002), pp. 1641–1646
 37. Z Sun, G Bebis, R Miller, in *The IEEE, International Conference on Intelligent Transportation Systems, 2002. Proceedings*. Improving the performance of on-road vehicle detection by combining Gabor and wavelet features[C] (IEEE, Los Alamitos, 2002), pp. 130–135
 38. X Wen, L Shao, W Fang, Y Xue, Efficient feature selection and classification for vehicle detection. *IEEE Trans. Circ. Syst. Video Technol.* **25**(3), 508–517 (2015)
 39. A Jiménez-Valverde, Insights into the area under the ROC curve (AUC) as a discrimination measure in species distribution modelling[J]. **21**(4), 498–507 (2011)
 40. X Ren, Y Zheng, Y Zhao, et al., Drusen segmentation from retinal images via supervised feature learning[J]. *IEEE Access*. **PP**(99), 1–1 (2017)
 41. L Jian, Y Zheng, W Jiao, et al., Deblurring sequential ocular images from multi-spectral imaging (MSI) via mutual information[J]. *Med. Biol. Eng. Comput.* **6**(6), 1–7 (2017)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)