

A Light-Weight CNN Based Multi-Task Architecture for Apple Maturity and Disease Classification [†]

Li Zhang ^{1,*}  and Jie Cao ^{1,2}

¹ Key Laboratory of Biomimetic Robots and Systems, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China; caojie@bit.edu.cn

² Yangtze Delta Region Academy, Beijing Institute of Technology, Jiaxing 314003, China

* Correspondence: zhanglsky@126.com

[†] Presented at the 2nd International Online Conference on Agriculture, 1–15 November 2023; Available online: <https://iocag2023.sciforum.net/>.

Abstract: Quickly and accurately judging the quality grades of apples is the basis for choosing suitable harvesting date and setting a suitable storage strategy. At present, the research of multi-task classification algorithm models based on CNN is still in the exploration stage, and there are still some problems such as complex model structure, high computational complexity and long computing time. This paper presents a light-weight architecture based on multi-task convolutional neural networks for maturity (L-MTCNN) to eliminate immature and defective apples in the intelligent integration harvesting task. L-MTCNN architecture with diseases classification sub-network (D-Net) and maturity classification sub-network (M-Net), to realize multi-task discrimination of the apple appearance defect and maturity level. Under different light conditions, the image of fruit may have color damage, which makes it impossible to accurately judge the problem, an image preprocessing method based on brightness information was proposed to restore fruit appearance color under different illumination conditions in this paper. In addition, for the problems of inaccurate prediction results caused by tiny changes in apple appearance between different maturity levels, triplet loss is introduced as the loss function to improve the discriminating ability of maturity classification task. Based on the study and analysis of apple grade standards, three types of apples were taken as the research objects. By analyzing the changes in apple fruit appearance in each stage, the data set corresponding to the maturity level and fruit appearance was constructed. Experimental results show that D-Net and M-Net have significantly improved recall rate, precision rate and F1-Score in all classes compared with AlexNet, ResNet18, ResNet34 and VGG16.



Citation: Zhang, L.; Cao, J. A Light-Weight CNN Based Multi-Task Architecture for Apple Maturity and Disease Classification. *Biol. Life Sci. Forum* **2024**, *30*, 19. <https://doi.org/10.3390/IOCAG2023-16881>

Academic Editor: Francesco Marinello

Published: 11 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: light-weight CNN; multi-task; fruit quality

1. Introduction

Fruit appearance grades are usually determined by attributes such as color, maturity, shape, texture, and size of the fruit. Manual grading faces practical difficulties such as subjective inconsistency, high cost, long time required, and insufficient number of professional grading staff during peak seasons. The appearance grade of apples has a great influence on consumers and market prices, so the task of fruit grading has become one of the most important research areas for many orchard planters and farm entrepreneurs [1–5]. Measuring the grade of apples mainly includes two important indicators of apple maturity and whether the appearance is defective. These two indicators are also important criteria for market sales [6]. As the apple ripens, the chlorophyll content inside the fruit decreases, the red pigment in the skin increases, the flesh becomes softer, starch converts into sugar, and the acidity decreases. Therefore, the maturity of an apple is determined by multiple factors, including the appearance color, hardness, starch content [7], soluble solids content, and sugar content. Chemical methods based on quantitative testing of various components

inside the fruit are commonly used to assess its maturity [8,9], but they often cause damage to the fruit, rendering it unsuitable for further sale [10]. Changes in the appearance color are a key indicator of apple ripeness and can be used to determine the maturity grade. With the development of computer image processing and machine vision technology, non-destructive automatic assessment of apple maturity can be achieved. By analyzing and processing the changes in appearance color, it is possible to discern the maturity level of the fruit. This method enables fast and accurate maturity evaluation without causing damage to the fruit. To achieve intelligent integrated operations and automate task of apple selecting by removing unripe or defective apples, this paper proposes a lightweight multitask maturity classification model called Light-Weights Multi-Task Convolutional Neural Network for Maturity (L-MTCNN). The model consists of two sub-networks: the Disease-Classification Sub Network (D-Net) and the Maturity-Classification Sub Network (M-Net). These sub-networks enable the multi-task classification of apple appearance defects and maturity levels. Furthermore, to address the issue of inaccurate predictions caused by minimal appearance changes between different maturity levels, Triplet-Loss [11] is employed as the loss function for the Maturity-Classification Sub Network (M-Net). This loss function aims to increase the distance between different maturity levels and reduce the distance within the same level. By analyzing the appearance changes of apples at various stages and utilizing industry standards, a dataset is constructed that correlates maturity levels with corresponding fruit appearance. The experimental results demonstrate that both D-Net and M-Net exhibit significant improvements in terms of recall, precision, and other metrics compared to classification models which validates that M-Net possesses high generalization capabilities.

2. Image Dataset Setup

The presence of disease for fruits is an important criterion for its quality. Apple surface defects primarily result from diseases, pests, and external damages. To setup diversity dataset, this paper utilizes methods such as on-site photography and online retrieval to collect 1680 images of apples with disease. These images are manually labeled and cropped to create a defect dataset. In addition, to establish an accurate relationship between fruit appearance and maturity, on-site collection is employed to analyze the changes in apple appearance during the ripening period. Depending on the corresponding relationship between the changes in fruit epidermis and ripeness, the fruit ripeness grade was divided into four categories according to the United States Standards for Grades of Apples [12]. Ref. [12] indicated as a mature apple becomes overripe it will show varying degrees of firmness, depending upon the stage of the ripening process, hence “Hard”, “Firm”, “Firm ripe”, and “Ripe” were the four terms used for describing the different stages. Due to the appearance of fruits being highly related to its attributes, we invited experts to label these images. According to the captured image date and the appearance of each fruit, the experts classified these images into unripe, turning, ripe, and overripe. These four categories were respectively marked as grade one (G1), grade two (G2), grade three (G3), and grade four (G4) for short.

3. CNN-Based Lightweight Multi-Task Classification

To achieve integrated non-destructive maturity assessment of apples, the task is divided into two subtasks: fruit disease detection and maturity classification. For fruits with diseases, their maturity categories are not considered. For fruits without diseases, further assessment is required to determine their maturity levels.

There are limitations to using color thresholding based on image space transformation to classify fruit maturity levels, as it may not effectively capture the small differences between different levels of the same fruit type. Additionally, image processing or machine learning classification models often require appropriate feature selection based on empirical knowledge, which can lead to lower accuracy if the features are not chosen correctly. Currently, CNN-based classification models such as VGG-16 [13], ResNet [14], and AlexNet [15]

have shown excellent performance in various classification tasks. However, these models are deep and have a large number of parameters, requiring a significant amount of training data to prevent issues such as gradient explosion or overfitting. Lightweight classification models like the MobileNet [16] series have the advantage of having a shallower network structure with fewer parameters compared to the aforementioned models. However, these models are designed for single-task classification and cannot handle multiple classification tasks simultaneously. Therefore, in this section, based on the strengths and weaknesses of various classification models and the specific requirements of apple classification tasks, a lightweight multi-task CNN classification model is designed and implemented. The structure of the model is shown in Figure 1.

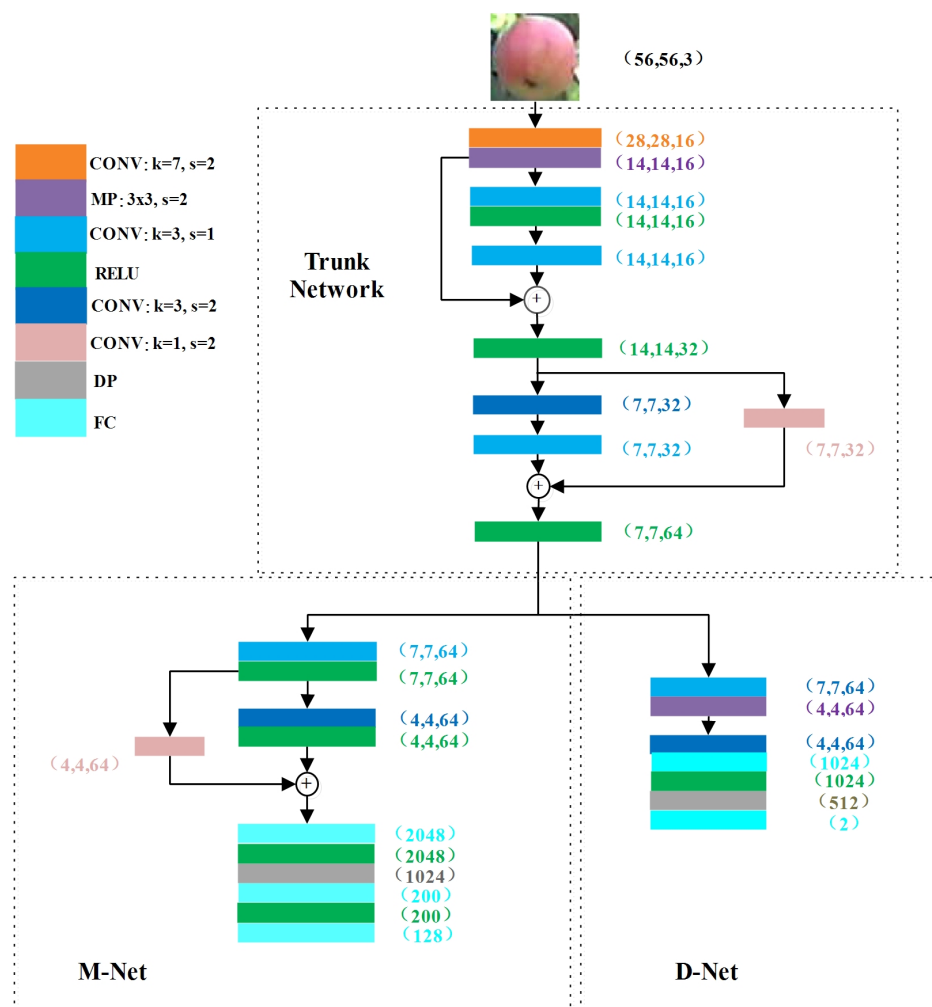


Figure 1. The structure of the multi-task classification architecture. Note: The dimensions and numbers in parentheses represent the output feature dimensions of each corresponding layer. CONV denotes the convolutional layer, where k represents the size of the convolutional kernel and s represents the stride. MP denotes the max pooling layer, where s represents the stride information. RELU represents the non-linear activation function. FLATTEN denotes the flattening operation. DP denotes the dropout processing. FC denotes the fully connected layer.

The classification model consists of a backbone network, a Disease Classification Sub-Network (D-Net), and a Maturity Classification Sub-Network (M-Net). The backbone network is responsible for extracting common feature information from the input images and providing this information to both the disease detection task and the maturity classification task. After obtaining the feature information extracted by the trunk network, feature extraction is performed for disease detection to determine whether the fruit has any

appearance disease. The maturity classification task utilizes the feature information from the trunk network to classify the maturity level of the fruit, achieving differentiation based on maturity. The following is a detailed introduction to each sub network:

1. Trunk Network: Responsible for extracting common feature information from input images,
2. D-Net: Focuses on targeted feature extraction for defect detection,
3. M-Net: Utilizes the shared features from the trunk network to classify the maturity grades of fruits.

3.1. Trunk Network

The trunk network is responsible for extracting common feature information required by both the M-Net and D-Net. Therefore, the design of this component directly affects the prediction results of the two sub-tasks and the accuracy of fruit quality classification. During the training process, as the depth of the neural network increases, there is a risk of degradation in model performance [14]. To address this issue, an improved backbone network based on the principles of residual networks is adopted. In this section, a five-layer CNN model is used as the backbone network to extract feature information. The design of the backbone network utilizes a skip connection structure to mitigate problems such as gradient explosion or over-fitting. By sharing features, the extracted feature information is simultaneously provided to both the defect classification task and the maturity classification task.

3.2. Defects Classification Sub-Network

Due to the prominent feature information of whether the fruit appearance is defective, a two-layer CNN model is designed for feature extraction in the D-Net. The loss function is defined as equation:

$$L_{det}^i = -(y_{det}^i \log(p^i) + (1 - y_{det}^i)(\log(1 - (p^i))) \quad (1)$$

3.3. Maturity Classification Sub-Network

Compared to the binary classification task of whether the fruit appearance is defective, the apple maturity classification task is more complex. Therefore, the design of M-Net is based on a model architecture with short connections [14]. Additionally, due to the small differences between different maturity grades of apples, to increase the distance between different categories and reduce the distance within the same category during model training, Triplet-loss is used as the loss function, as shown as equation:

$$L = \left[\left\| f(x_a^i) - f(x_p^i) \right\|_2^2 - \left\| f(x_a^i) - f(x_n^i) \right\|_2^2 + \alpha \right]_+ \quad (2)$$

where $f(x_a^i)$ denotes the learned feature vector, $f(x_p^i)$ denotes the feature vector of the current positive sample, and $f(x_n^i)$ denotes the feature vector of the current negative sample. $\left\| f(x_a^i) - f(x_p^i) \right\|_2^2$ and $\left\| f(x_a^i) - f(x_n^i) \right\|_2^2$ represent the Euclidean distances between the current learned feature vector and the normalized feature vectors of the positive and negative samples, respectively. The distance between the current feature and the positive sample should be smaller than the distance between the current feature and the negative sample, with q_i representing the threshold value for the difference between the positive and negative sample distances.

4. Experiments

The main configuration of the experiment equipment used in this study includes an Intel i5-7500 CPU @ 3.20 GHz processor, 8 GB of RAM, and a dedicated graphics card NVIDIA GeForce GTX1060 with 3 GB of memory. The operating system is Ubuntu 16.04,

the software development environment is PyTorch, and the programming language used is Python.

4.1. Evaluation Standard

The terms of precision, recall, and F_1 -score, and accuracy were taken as the evaluation indexes to verify our proposed network extensively and quantitatively. The corresponding formulas are shown as equations:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F_1 = \frac{2TP}{2TP + FP + FN} \quad (5)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (6)$$

Specifically, the TP , TN , FP , and FN represent the relationship between the observed value and the predicted value, as shown in Table 1.

Table 1. Instructions for the elements of TP , TN , FP and FN .

Label Name	Predicted Positive	Predicted Negative
True Positive	TP	FN
True Negative	FP	TN

4.2. Model Training

The proposed network model consists of a backbone network, M-Net, and D-Net. However, the datasets used for training M-Net and D-Net are completely different, making it challenging to train both networks simultaneously. Considering that the classification task of M-Net is more complex than that of D-Net, this study adopts a two-step training approach. First, the M-Net subnetwork model is trained independently. After completing the training of M-Net, the backbone network parameters and M-Net parameters are fixed, and then the D-Net network is trained.

5. Results

5.1. D-Net Results

The evaluation results of the D-Net model compared to the classical network models AlexNet, ResNet-18, ResNet-34, and VGG-16 in terms of precision, recall, F_1 -score, and accuracy for different categories. The table shows the performance of the D-Net in comparison to the other models. The evaluation results of the model for disease and normal category are shown in Table 2 and Table 3, respectively.

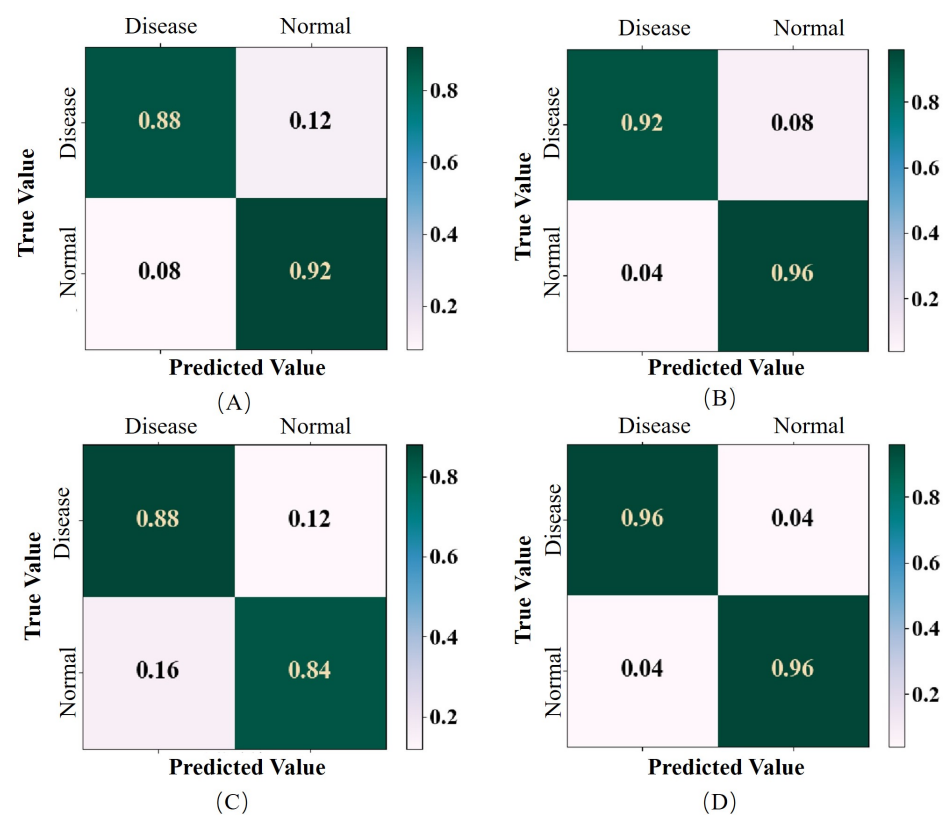
Table 2. The classification results of defect fruits by D-Net.

Model	Accuracy	Recall	Precision	F1-Score
AlexNet	0.90	0.92	0.88	0.90
ResNet-18	0.94	0.96	0.92	0.94
ResNet-34	0.94	0.92	0.96	0.94
VGG-16	0.86	0.85	0.88	0.86
D-Net	0.96	0.96	0.96	0.96

Table 3. The classification results of normal fruits by D-Net.

Model	Accuracy	Recall	Precision	F1-Score
AlexNet	0.90	0.92	0.88	0.90
ResNet-18	0.94	0.92	0.96	0.94
ResNet-34	0.94	0.96	0.92	0.94
VGG-16	0.86	0.88	0.84	0.86
D-Net	0.96	0.96	0.96	0.96

According to the results, for the classification task of appearance diseases, the D-Net model achieves higher accuracy compared to AlexNet, ResNet-18, ResNet-34, and VGG-16, with improvements of 6%, 2%, 2%, and 10%, respectively. This indicates that the proposed D-network outperforms the classical network models in terms of accuracy. To further evaluate the model, the performance of D-Net in terms of recall, precision, and F1-score is analyzed for both disease and normal classes. For the disease classification task, D-Net shows improvements in recall of 4%, 0%, 4%, and 11% compared to AlexNet, ResNet-18, ResNet-34, and VGG-16, respectively. The precision also improves by 8%, 4%, 0%, and 8% for the respective models. For the normal classification task, D-Net achieves improvements in recall of 8%, 4%, 0%, and 8% and in precision of 4%, 0%, 4%, and 12% compared to AlexNet, ResNet-18, ResNet-34, and VGG-16, respectively. These results demonstrate that the proposed D-Net model achieves higher accuracy for the binary classification task of appearance classification task, and it also highlights the superior performance of ResNet-18 and ResNet-34 compared to AlexNet and VGG-16. The probably reason for these results are due to the residual structure for D-network, which are more effective for extracting image features. The distribution of values in the confusion matrix is shown in Figure 2.

**Figure 2.** Confusion matrix results for bi-classification of disease from D-Net. In particular, (A), (B), (C) and (D) are the results of Alexnet, Resnet, VGG-16 and D-net, respectively.

As the confusion matrix shown, it can be observed that the AlexNet, ResNet-18, VGG-16, and D-Net network models achieve good results in predicting the binary task of defect and non-defect classes. All the models are able to accurately discriminate whether a fruit has a disease or not with high accuracy. In particular, D-Net achieves a prediction accuracy of 0.96, which is higher than the other network models, indicating its superior performance.

5.2. M-Net Results

We evaluated the proposed M-Net with AlexNet, ResNet18, ResNet34, and VGG16 on maturity classification task. The evaluation includes average accuracy and the indicator of recall, precision, and F1-score for each maturity grade, as shown in Table 4.

Table 4. The classification results of maturity fruits by M-Net.

Grade Level	Indicators	AlexNet	ResNet-18	ResNet-34	VGG-16	M-Net
G1	Recall	0.74	0.76	0.78	0.65	0.91
G1	Precision	0.68	0.76	0.72	0.60	0.84
G1	F1-score	0.71	0.76	0.75	0.63	0.87
G2	Recall	0.50	0.63	0.62	0.44	0.76
G2	Precision	0.56	0.60	0.64	0.48	0.76
G2	F1-score	0.53	0.61	0.63	0.46	0.76
G3	Recall	0.68	0.76	0.72	0.67	0.81
G3	Precision	0.68	0.76	0.72	0.64	0.88
G3	F1-score	0.68	0.76	0.72	0.65	0.85
G4	Recall	0.92	0.88	0.88	0.85	0.96
G4	Precision	0.88	0.92	0.92	0.88	0.96
G4	F1-score	0.90	0.90	0.90	0.86	0.96
-	Accuracy	0.70	0.76	0.75	0.65	0.86

According to the results, compared to AlexNet, ResNet-18, ResNet-34, and VGG-16, the M-Net model shows a significant improvement in average accuracy. Furthermore, M-Net outperforms AlexNet, ResNet-18, ResNet-34, and VGG-16 by evaluating the model's performance in terms of recall, precision, and F1-score for each maturity grade. Specifically, as for G1, M-Net improves the F1-score by 16%, 11%, 12%, and 24% respectively. As for G2, the improvements are 23%, 15%, 13%, and 30% respectively. For G3, the improvements are 17%, 9%, 13%, and 20% respectively. And for G4, the improvements are 6%, 6%, 6%, and 10% respectively. These results indicate that M-Net achieves significant improvements in classifying different maturity grades. The distribution of values in the confusion matrix for each maturity grade is shown in Figure 3.

As the results shown, for the ripeness classification task of Fuji apple, all models perform the best in predicting G4, followed by maturity G3 and G1, and the most challenging predictions are for maturity G2. The possible reason for this phenomenon is that maturity G4 has distinct features compared to the other three grades, while G2 shares similar characteristics with G1 and G3, leading to confusion and affecting the final prediction results.

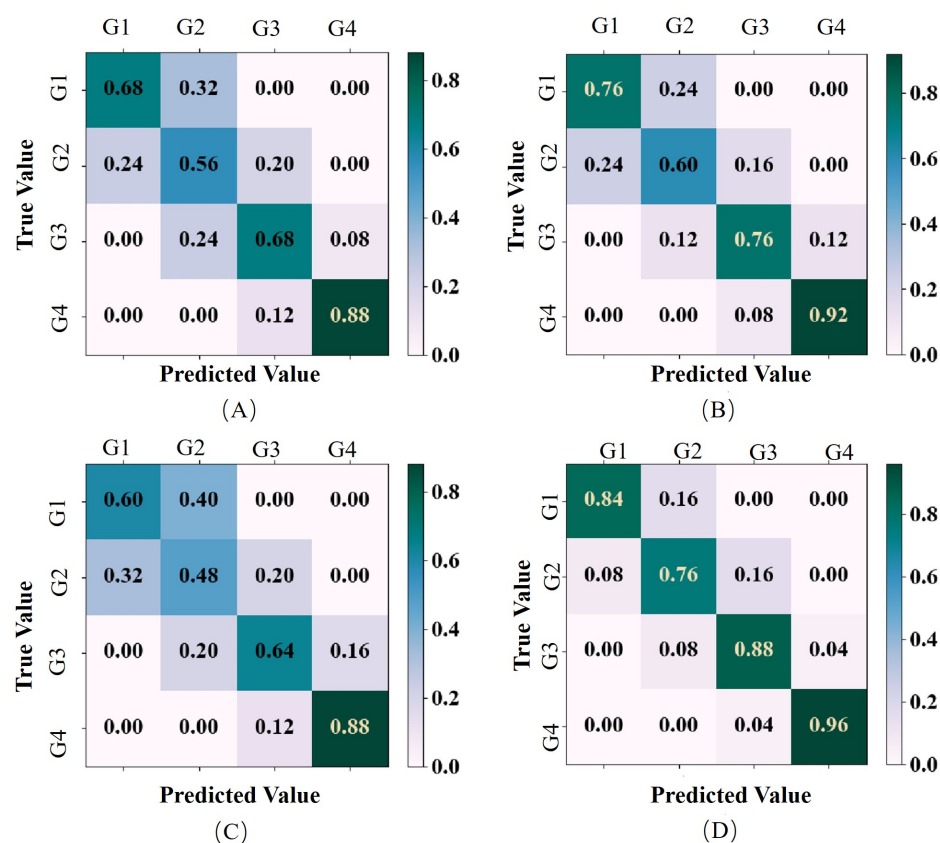


Figure 3. Confusion matrix of ripeness classification results from M-Net. In particular, (A), (B), (C) and (D) are the results of Alexnet, Resnet, VGG-16 and D-net, respectively.

6. Conclusions

This paper proposes a lightweight CNN based architecture to achieve task of apple maturity grade and fruit appearance defects. The model proposed a lightweight trunk network to extract feature information, share the weights to the D-Net and M-Net sub-networks, which improving the utilization of features and reducing the number of parameters. To address the issue of low accuracy caused by tiny changes between different ripeness grades, Triplet-Loss function is introduced as the loss function for M-Net, enlarging the feature distance between different ripeness grades and reducing the feature distance for the same grade. To enhance the model's generalization ability in practical applications, this paper constructs an apple maturity dataset based on the study of appearance changes in multiple apple varieties during the ripening process. Experimental results demonstrate that the proposed L-MTCNN outperforms the compared models in multi-classification task.

Author Contributions: Conceptualization, L.Z. and J.C.; methodology, L.Z.; software, J.C.; investigation, J.C.; writing—original draft preparation, L.Z.; writing—review and editing, L.Z.; project administration, J.C.; funding acquisition, L.Z. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 32302456) and Beijing Nature Science Foundation of China (No. 4222017).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tu, S.; Xue, Y.; Zheng, C.; Qi, Y.; Wan, H.; Mao, L. Detection of passion fruits and maturity classification using Red-Green-Blue Depth images. *Biosyst. Eng.* **2018**, *175*, 156–167. [[CrossRef](#)]
2. Villaseñor-Aguilar, M.J.; Botello-Álvarez, J.E.; Pérez-Pinal, F.J.; Cano-Lara, M.; León-Galván, M.F.; Bravo-Sánchez, M.G.; Barranco-Gutiérrez, A.I. Fuzzy classification of the maturity of the tomato using a vision system. *J. Sens.* **2019**, *2019*, 3175848. [[CrossRef](#)]
3. Faisal, M.; Albogamy, F.; Elgibreen, H.; Algabri, M.; Alqershi, F.A. Deep learning and computer vision for estimating date fruits type, maturity level, and weight. *IEEE Access* **2020**, *8*, 206770–206782. [[CrossRef](#)]
4. Baeten, J.; Donné, K.; Boedrij, S.; Beckers, W.; Claesen, E. Autonomous fruit picking machine: A robotic apple harvester. In Proceedings of the Field and Service Robotics: Results of the 6th International Conference, Chamonix, France, 9–12 July 2007; Springer: Berlin/Heidelberg, Germany, 2008; pp. 531–539.
5. Muscato, G.; Prestifilippo, M.; Abbate, N.; Rizzuto, I. A prototype of an orange picking robot: Past history, the new robot and experimental results. *Ind. Robot. Int. J.* **2005**, *32*, 128–138. [[CrossRef](#)]
6. Zhang, B.; Zhang, M.; Shen, M.; Li, H.; Zhang, Z.; Zhang, H.; Zhou, Z.; Ren, X.; Ding, Y.; Xing, L.; et al. Quality monitoring method for apples of different maturity under long-term cold storage. *Infrared Phys. Technol.* **2021**, *112*, 103580. [[CrossRef](#)]
7. Menesatti, P.; Zanella, A.; D’Andrea, S.; Costa, C.; Paglia, G.; Pallottino, F. Supervised multivariate analysis of hyper-spectral NIR images to evaluate the starch index of apples. *Food Bioprocess Technol.* **2009**, *2*, 308–314. [[CrossRef](#)]
8. Pathange, L.P.; Mallikarjunan, P.; Marini, R.P.; O’Keefe, S.; Vaughan, D. Non-destructive evaluation of apple maturity using an electronic nose system. *J. Food Eng.* **2006**, *77*, 1018–1023. [[CrossRef](#)]
9. Chagné, D.; Lin-Wang, K.; Espley, R.V.; Volz, R.K.; How, N.M.; Rouse, S.; Brendolise, C.; Carlisle, C.M.; Kumar, S.; De Silva, N.; et al. An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. *Plant Physiol.* **2013**, *161*, 225–239. [[CrossRef](#)] [[PubMed](#)]
10. Lunadei, L.; Galleguillos, P.; Diezma, B.; Lleó, L.; Ruiz-Garcia, L. A multispectral vision system to evaluate enzymatic browning in fresh-cut apple slices. *Postharvest Biol. Technol.* **2011**, *60*, 225–234. [[CrossRef](#)]
11. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.
12. Usda, U.; Ams, A. *United States Standards for Grades of Apples*; USDA Publication: Washington, DC, USA, 2002.
13. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
15. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 84–90. [[CrossRef](#)]
16. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.