

Article

Classification of Apple Color and Deformity Using Machine Vision Combined with CNN

Dekai Qiu ^{1,†}, Tianhao Guo ^{1,†}, Shengqi Yu ¹, Wei Liu ¹, Lin Li ^{2,3}, Zhizhong Sun ^{4,5}, Hehuan Peng ^{1,*} and Dong Hu ^{1,2,3,*}

¹ College of Optical, Mechanical and Electrical Engineering, Zhejiang A&F University, Hangzhou 311300, China; 2021612021001@stu.zafu.edu.cn (D.Q.); 2022112011003@stu.zafu.edu.cn (T.G.); 2021612021002@stu.zafu.edu.cn (S.Y.); liuwei@zafu.edu.cn (W.L.)

² School of Agricultural Engineering, Jiangsu University, Zhenjiang 212013, China; lilin@ujs.edu.cn

³ Key Laboratory of Modern Agricultural Equipment and Technology, Jiangsu University, Ministry of Education, Zhenjiang 212013, China

⁴ College of Chemistry and Materials Engineering, Zhejiang A&F University, Hangzhou 311300, China; sunzz@stu.zafu.edu.cn

⁵ College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

* Correspondence: penghh@zafu.edu.cn (H.P.); 20180047@zafu.edu.cn (D.H.)

† These authors contributed equally to this work.

Abstract: Accurately classifying the quality of apples is crucial for maximizing their commercial value. Deep learning techniques are being widely adopted for apple quality classification tasks, achieving impressive results. While existing research excels at classifying apple variety, size, shape, and defects, color and deformity analysis remain an under-explored area. Therefore, this study investigates the feasibility of utilizing convolutional neural networks (CNN) to classify the color and deformity of apples based on machine vision technology. Firstly, a custom-assembled machine vision system was constructed for collecting apple images. Then, image processing was performed to extract the largest fruit diameter from the 45 images taken for each apple, establishing an image dataset. Three classic CNN models (AlexNet, GoogLeNet, and VGG16) were employed with parameter optimization for a three-category classification task (non-deformed slice-red apple, non-deformed stripe-red apple, and deformed apple) based on apple features. VGG16 achieved the best results with an accuracy of 92.29%. AlexNet and GoogLeNet achieved 91.66% and 88.96% accuracy, respectively. Ablation experiments were performed on the VGG16 model, which found that each convolutional block contributed to the classification task. Finally, prediction using VGG16 was conducted with 150 apples and the prediction accuracy was 90.50%, which was comparable to or better than other existing models. This study provides insights into apple classification based on color and deformity using deep learning methods.

Keywords: apple classification; color; convolutional neural network; deformity; machine vision



Citation: Qiu, D.; Guo, T.; Yu, S.; Liu, W.; Li, L.; Sun, Z.; Peng, H.; Hu, D. Classification of Apple Color and Deformity Using Machine Vision Combined with CNN. *Agriculture* **2024**, *14*, 978. <https://doi.org/10.3390/agriculture14070978>

Academic Editor: Maciej Zaborowicz

Received: 24 May 2024

Revised: 20 June 2024

Accepted: 21 June 2024

Published: 23 June 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/>).

1. Introduction

Apples are one of the most commonly produced fruits in China, with production exceeding 47.57 million tons in 2022 [1]. The rich content of vitamin C, fiber, and water of apples have made them important sources of human nutrition. Factors like vibrant color, ideal size, consistent shape, and flawless exterior all significantly contribute to an apple's market value and consumer preferences. Therefore, quality detection and grading of apples play a crucial role in maximizing their commercial value.

With the commercialization of agricultural products, the replacement of manual grading with automation has become increasingly important. In the past two decades, many optical sensing technologies, such as machine vision, hyperspectral or multispectral imaging, and light scattering imaging have significantly enhanced quality detection and grading of agro-products [2–5], due to the advantages of non-destructiveness, easy implementation

and low cost. Machine vision, as one of the commonly used technologies, combined with image processing and advanced modeling methods, has been widely applied for quality detection and grading of various fruits in laboratories or commercial production lines [6,7]. Zou et al. (2010) [8] developed two machine vision systems for apple quality grading based on color and defects using image feature extraction and machine learning methods. Hu et al. (2021) [9] proposed a field-based multi-feature fusion detection method for apple grading. By extracting apple size, color, and shape information and employing a support vector machine (SVM), they achieved an impressive average grading accuracy of 95.49%.

However, traditional grading methods often require image segmentation and manual feature extraction with a large scale of information loss. In recent years, deep learning has been widely researched in the field of image classification. Due to powerful learning ability and wide applicability, deep learning can well overcome the challenge of feature extraction without information loss. Due to its high accuracy and efficiency, machine vision technology coupled with deep learning has been widely used in many fields such as disease diagnosis in medicine [10,11], forest monitoring [12–14], and quality classification of agricultural products [15–17]. Several studies have been conducted on fruit grading [18–20], among which apple detection and grading based on external quality is particularly prominent. For instance, Li et al. (2020) [21] trained a shallow convolutional neural network (CNN) for classifying six types of apples, and compared the obtained results with SVM, RseNet-50, and RseNet-18. CNN achieved 92.00% accuracy in an unobscured case, which was significantly better than the other three methods. However, the accuracy of the model decreased in occluded cases with increasing occlusion. Li et al. (2021) [22] proposed a new CNN-based model for triple classification of apple defects, achieving accuracies of 98.98% and 95.33% in the validation and test sets, respectively. Their model outperformed the Google Inception v3 model and the traditional models based on the histogram of oriented gradients (HOG), gray level co-occurrence matrix (GLCM) feature merging, and SVM. Fan et al. (2020) [23] constructed a CNN model and applied it to the online detection of defective apples in a sorting machine, obtaining superior performance with an accuracy rate of 96.50%, which was better than the classification result of traditional image processing methods. Shi et al. (2022) [24] combined a pre-trained lightweight CNN with long short term memory (LSTM) and a homogenization technique to construct a spatial feature module for multi-view apple analysis, achieving a classification accuracy of 99.23%. In addition, Zeynep et al. [25] used VGG16 and AlexNet to classify apple bruising, using 1000 images of healthy apples and 500 images of bruised apples from a dataset of 500 apples. Results on the RGB dataset showed that VGG16 achieved the highest test accuracy at 86%, while AlexNet exhibited the lowest at 74.6%. When trained and tested on the NIR dataset, AlexNet, Inception v3, and VGG16 achieved accuracies of 99.33%, 100%, and 100%, respectively. Fu et al. [26] utilized GoogLeNet to classify apples, lemons, oranges, pomegranates, tomatoes, and colored peppers. Experimental results showed that GoogLeNet achieved a training accuracy of 96.88%, testing accuracy of 96%, and a training speed of 11.38 images per second. Ni et al. [27] employed the GoogLeNet model to automatically extract features from banana images and classify them using a classifier module. Research results demonstrated that the model can accurately detect banana freshness with an accuracy of 98.92%, surpassing human detection levels.

Most of the above-mentioned studies were conducted for grading fruit varieties and external defects. However, fewer studies have been conducted on apple color and deformity. The deformity index of the apple is another crucial parameter in apple grading. The apples with poor deformity indices should be picked out, while the better ones should be transferred to the market for purchase. Therefore, this study proposes a deep learning-based method for apple grading based on apple color and deformity. The three research objectives of this study are as follows: (1) Construct a dynamic machine vision system for image acquisition of apples that will capture images of apples and automatically select the one showcasing the largest diameter; (2) Perform three-classification modeling of apple deformity and color (non-deformed slice-red apple, non-deformed stripe-red apple, and

deformed apple) using classical models (AlexNet, GoogLeNet and VGG16); (3) Conduct ablation experiments on the best-in-class model for model optimization and performance improvement.

2. Materials and Methods

2.1. Machine Vision System

A self-constructed machine vision system was used for image acquisition in this study. Figure 1 shows a schematic diagram of the image acquisition system. The system mainly consists of a lifting and rotating table (PX110-100, PDV, Beijing, China) for height adjustment and apple rotation, two front strip light sources (L140-20-18, Eagle Vision Technology, Guangzhou, China), a bottom strip light source (L100-20-18, Eagle Vision Technology, Guangzhou, China), and a top ring light source (R120-75-18, Eagle Vision Technology, Guangzhou, China) for illuminating the samples, a camera (MV-SUA134GC-T, MindVision, Shenzhen, China), and a lens (MV-LD-12-3M-A, MindVision, Shenzhen, China) for image acquisition. Two front strip light sources with a maximum power of three watts were mounted on both sides of the camera vertically and symmetrically to illuminate the apple sides. Another strip light source was placed horizontally on the lower part to illuminate the bottom of the apple. Finally, a ring light source with a maximum power of three watts was used to illuminate the top of the apple. The resolution of the camera is 1024×1280 pixels, and the camera works synchronously with the rotary table controlled by the computer through a microcontroller. The rotary table is used to simulate the continuous rotation of apples during real-time grading; a rotary table continuously rotates the apples at adjustable angles controlled by a microcontroller.

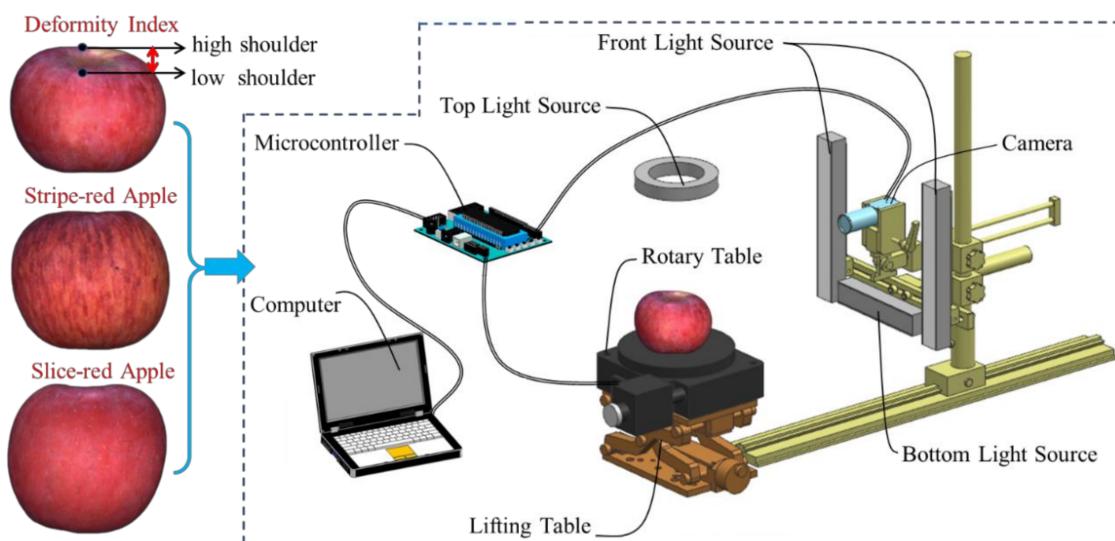


Figure 1. Schematic diagram of a self-constructed machine vision system and representative apple samples.

In this study, the system was calibrated before image acquisition. The camera's aberrations were first studied using a checkerboard grid, followed by the ColorChecker (Classic Mini 24 colors, Calibrite, America) for calibration in terms of color, to reduce the effects caused by camera distortion [28].

2.2. Image Dataset

2.2.1. Apple Samples

In this study, 1000 red 'Fuji' apples grown in Luochuan, Shanxi were purchased from the official local sales shop and used for experiments. There were 360 deformed apples and 640 non-deformed apples, including 320 stripe-red apples and 320 slice-red apples. The samples were divided into training, validation, and test sets by using random sample function mainly through the Python language to divide the raw apple data proportionally.

The specific numbers are provided in Table 1. The diameter of the apple samples was controlled within the range of 75–95 mm, and the weight of the samples was within the range of 175–300 g.

Table 1. Dataset allocation of apple samples before and after data augmentation.

Label	Training Set		Validation Set		Test Set	
	Before	After	Before	After	Before	After
Deformed apples	208	1248	52	312	100	100
Non-deformed stripe-red apples	216	1296	54	324	50	50
Non-deformed slice-red apples	216	1296	54	324	50	50
Total	640	3840	160	960	200	200

2.2.2. Image Acquisition

The image acquisition process was conducted in a dark environment, with the apples positioned at the center of the rotary table. The region of interest of the tested apple was set to 850×850 pixels, while the rotation speed of the table was 36.5 rps. Apple images obtained from the experiment underwent image processing to ensure precise measurement of apple diameter, with an accuracy tolerance of less than 0.5 mm. To minimize the number of acquired images, it was finally decided to capture the apple images at an angle of 8 degrees during rotation, achieving a balance between time cost and measurement accuracy of apple diameter. Hence, a total of 45 images were captured for each apple, and the image with the largest diameter for the apple was selected through subsequent image processing to create the dataset. The captured images were in .bmp format, with a size of 850×850 pixels. This dataset was used for classification tasks using deep learning methods.

2.2.3. Image Processing

Figure 2 illustrates the workflow of image acquisition, image processing, extraction of apple diameter, and quality classification. Firstly, the acquired color image was converted into a grayscale image to extract luminance information and simplify processing. Then Gaussian filter denoising was applied to smooth the image and eliminate noise. Next, a threshold value of 30 was applied to convert the grayscale image into a binary image. Pixel points exceeding the threshold were set as white, and those below were set as black. Subsequently, noise reduction was carried out on the obtained binary image using morphological operations to eliminate small white areas and fill small black areas through erosion and dilation operations. Next, the minimum external rectangle method was utilized to extract the minimum bounding rectangle of the apple object in the binary image, and the pixel value of the maximum diameter of the apple was determined by measuring the length of the rectangle. Finally, the image depicting the apple with the largest diameter was fed to the CNN for model training and prediction.

2.2.4. Image Data Augmentation

Data augmentation is a commonly used method in deep learning [29] that often increases the number of samples by introducing random transformations to make the dataset richer. Thus, the generalization ability and robustness of the model can be improved, and overfitting can be mitigated to improve the model's performance. In this study, various augmentation operations were performed on the original data using OpenCV in the Python environment. These operations included mirroring images, adding Gaussian noise, adding pretzel noise, reducing brightness, and random image masking. Representative augmented images of apples are displayed in Figure 3. Consequently, the training and validation sets, originally containing 800 apple images each, were increased to 4800 images. The test set remained at 200 samples. The specific sample division for training, validation and test sets before and after data augmentation is shown in Table 1.

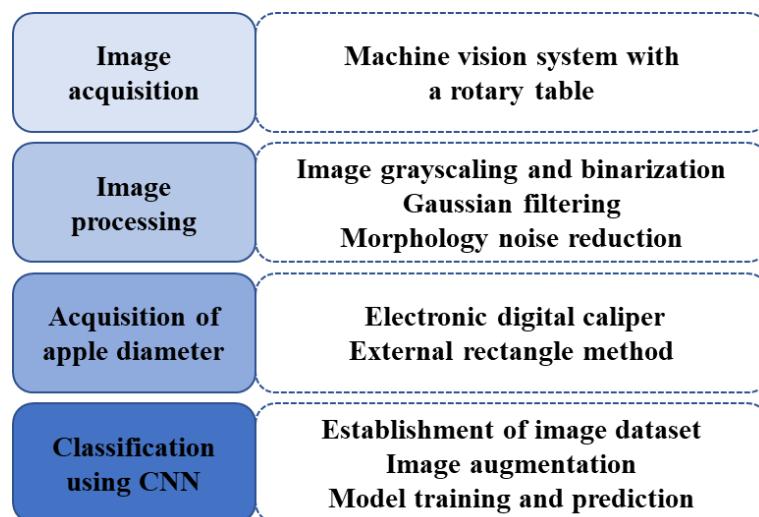


Figure 2. Flowchart of data acquisition and processing.

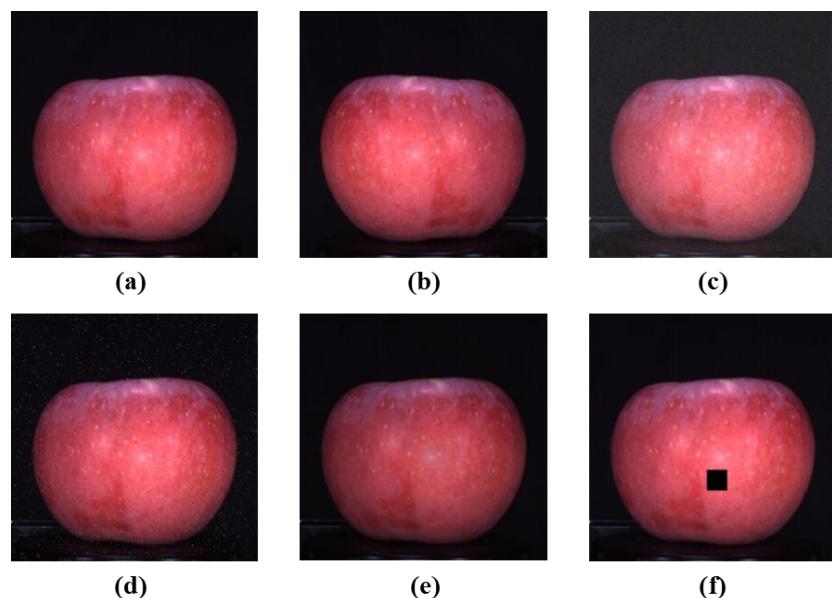


Figure 3. Representative apple images after data augmentation: (a) original image; (b) mirroring image; (c) adding Gaussian noise; (d) adding pretzel noise; (e) reducing brightness; (f) random image masking.

2.3. Grading Criteria

In this study, the apples were classified into three categories based on their color and deformity index. The categories were non-deformed slice-red apple, non-deformed stripe-red apple, and deformed apple.

2.3.1. Apple Deformity Index

The deformity index is a crucial parameter for describing apple appearance quality. It is defined as the distance between the high and the low shoulders when the apple is placed on a table, as shown in Figure 1. The reference deformity index of the apple was manually measured by employing an electronic digital caliper (G101-102-101, SNORT, Huzhou, China). In this study, apples with a deformity index greater than 10 mm were treated as deformed apples, while those with an index below 10 mm were considered non-deformed. The 10 mm threshold for deformity classification is an industry standard in China to ensure the applicability of the research results, which has been reported by official documents or websites.

2.3.2. Apple Color

The red ‘Fuji’ apples used in this study can be divided into slice-red apples and stripe-red apples according to their appearance color. Figure 1 illustrates a slice-red apple, which is usually characterized by large areas of red skin, while the stripe-red apple has thin and irregularly distributed red stripes on the skin. Consumer preferences for apple color can vary greatly. Some may associate specific colors with flavor profiles, like crispness, but taste perception is subjective. It can be observed that the slice-red and stripe-red apples differ greatly in appearance. They were mixed for model training and testing in this study, resulting in a great challenge for classification.

2.4. Convolutional Neural Networks

2.4.1. AlexNet

AlexNet is a classical CNN model proposed by Alex Krizhevsky et al. [30] in 2012. AlexNet is considered to be one of the important milestones in the history of deep learning and has demonstrated excellent performance in several domains [31,32]. The AlexNet model, which usually contains 8 convolutional layers, 3 fully-connected layers, and the final output layer, is deeper and more complex in structure than the traditional neural networks. ReLU activation function and the Dropout (random deactivation) operation are often used to enhance the generalization ability of the network. The ReLU function is used to enhance the non-linear expressive power of the model, while Dropout is used to prevent overfitting. The AlexNet model also uses large convolutional kernels (11×11) and accelerates the training process with GPU clusters. In this study, the batch size, activation function, learning rate and number of epochs of the AlexNet model were set as 128, ReLU, 0.001, and 100, respectively.

2.4.2. GoogLeNet

GoogLeNet is a major contribution to the field of deep learning created by the Google team [33]. GoogLeNet uses an architecture called the “Inception” module (Figure 4) to process images at different scales and sizes for extracting feature information. Each Inception module consists of multiple parallel convolutional layers (1×1 , 3×3 , 5×5 convolutions) and a maximum pooling layer. This design excels at simultaneously capturing image features at different scales. The resulting feature maps are then combined, creating a more comprehensive representation, further improving the classification accuracy of the model. In GoogLeNet, multiple stacked Inception modules are used to process image features at different scales. To further improve the generalization ability of the model, a Dropout operation is added to each Inception module, which effectively reduces the overfitting phenomenon. GoogLeNet has been commonly applied in the field of image classification [34,35]. The batch size, activation function, learning rate, and number of epochs of the GoogLeNet model used in this study were the same as for the AlexNet.

2.4.3. VGG16

VGG16 is a deep learning model proposed in 2014 by a team of researchers in the Department of Computer Science at the University of Oxford [36]. VGG16 has a relatively simple and repetitive network structure. VGG16 employs multiple consecutive small-sized convolutional kernels and pooling kernels, which could increase the depth of the network. The use of multiple small-sized convolutional kernels allows the network to learn increasingly complex feature representations, compared to using convolutional kernels with larger receptive fields. Specifically, VGG16 consists of 13 convolutional layers and 3 fully connected layers, as shown in Figure 5. Each convolutional layer is constructed with 3×3 sized convolutional kernels, immediately followed by a maximum pooling layer with the size of 2×2 , which is used to reduce the spatial size of the feature map and extract dominant features. This stacked structure of convolutional and pooling layers makes VGG16 more expressive and capable of handling more complex image features with good generalization capabilities. Moreover, the simplicity and repetitiveness of the VGG16 network structure

make it not only easy to understand but also straightforward to implement. VGG16 has achieved impressive performance in image recognition tasks and become one of the important models in the field of deep learning-based image classification [37,38]. In this study, the batch size, activation function, learning rate, and number of epochs of the VGG16 model were exactly the same as for the AlexNet and GoogLeNet.

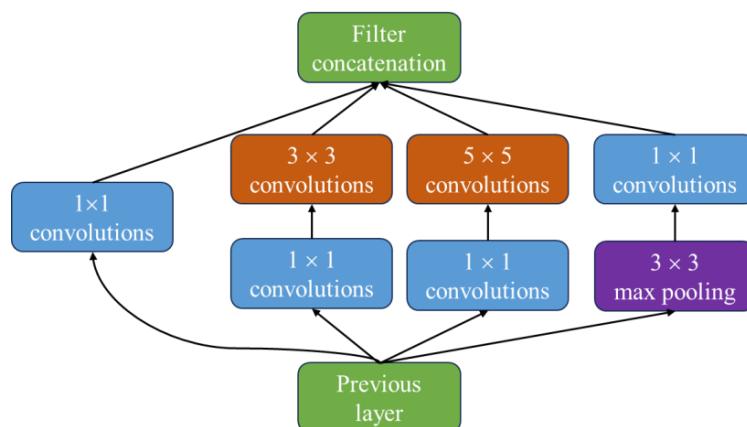


Figure 4. Diagram of Inception architecture.

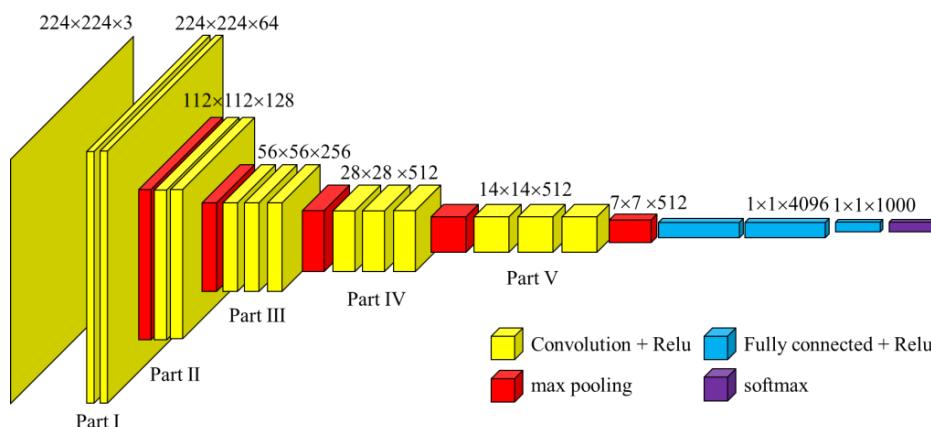


Figure 5. Diagram of VGG16 architecture.

In addition, since the VGG16 outperformed the other two models in apple classification based on experimental results, ablation experiments were conducted on the VGG16 model to assess its performance by varying the convolutional layers. In each experiment, all convolutional layers except the initial layer of each section were simultaneously removed.

2.4.4. Experimental Environment

The CNN models (AlexNet, GoogLeNet, and VGG16) used in this study were constructed in the PyCharm framework. Table 2 presents the software environment configuration for the experiments.

Table 2. Environment configuration for constructing CNN models.

Configuration	Parameter
GPU	RTX 4090
CPU	Xeon(R) Platinum 8352V
RAM	90 G
Accelerated framework	CUDA 11.0
Deep learning framework	Pytorch 1.7.0
Programming language	Python 3.8

2.5. Evaluation Indicators

This study aimed to achieve the classification of external qualities (deformity and color) of apples. Therefore, the performance of the model was evaluated using both classification accuracy, which reflects the model's generalization ability, and model complexity.

1. Precision evaluation indices

In addition to utilizing classification accuracy as a measure to evaluate the overall classification performance of the model on the entire dataset, this study also incorporated precision, recall, and F1-score for further evaluation. The equations for calculating accuracy, precision, recall, and F1-score are shown in Equations (1)–(4), respectively.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100\% \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (3)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (4)$$

where TP, FP, TN, and FN denote the number of true-positive, false-positive, true-negative, and false-negative samples, respectively.

2. Complexity assessment indicators

In deep learning models, the number of required parameters usually represents the model's complexity. The higher the number of parameters in a model, the more complex the model is. The model size is the amount of storage space occupied by a deep learning model, which is usually used to characterize the model's complexity. Another indicator of model complexity is inference time. This refers to the time a model takes to predict a sample or a batch of samples.

3. Results and Discussion

3.1. System Evaluation and Image Processing

Since the camera aberrations were small and had negligible effect on the experiment, aberrations were partly corrected for in this study. The machine vision system was eventually color-corrected due to the significant differences in the captured apple colors. Imatest software (version 24.1, China) was used to conduct reproduction analysis of the color card images before and after correction. The maximum value of color difference was 20.9 before correction and 13.4 after correction, and the color of the corrected apple image was closer to that of the apple itself. This study demonstrates the effectiveness of camera color correction. By correcting color variations, the system was able to effectively restore the apple color and extract the diameter of the apple using image processing methods. Figure 6 shows some representative images after color correction in the dataset. It can be observed that the larger the deformity index, the worse the symmetry of the apple image. Additionally, the stem area of deformed apples is often more easily identifiable within the images.

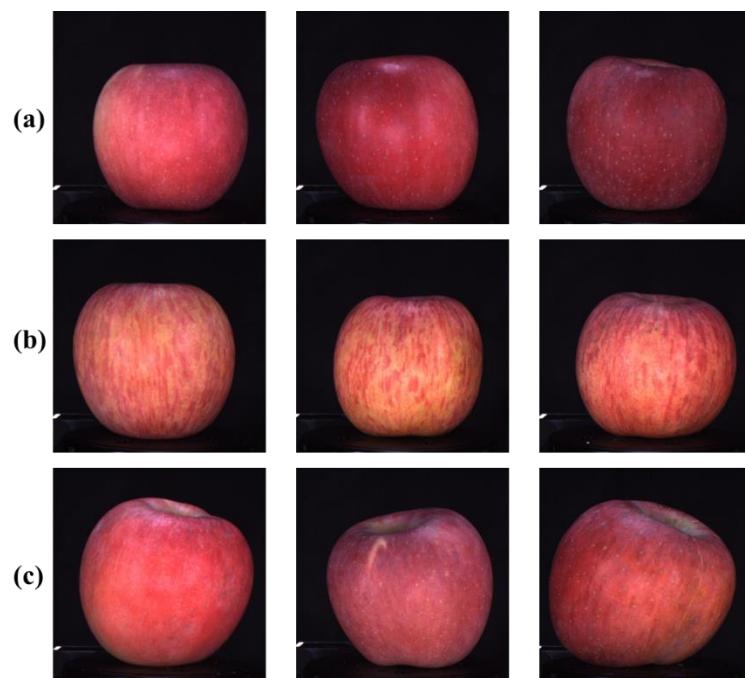


Figure 6. Representative apple images in the dataset: (a) non-deformed slice-red apples with deformity indices of 2, 5, and 8; (b) non-deformed stripe-red apples with deformity indices of 2, 5, and 8; and (c) deformed apples with deformity indices of 12, 15, and 18. The deformity index signifies the measurement value in millimeters of deviation; visually, the larger the deformity index, the more noticeable the deformity.

3.2. Performance Comparison

Three classical models (AlexNet, GoogLeNet, and VGG16) were employed for the classification task. Before feeding the models for training, the acquired apple images were resized from 850×850 pixels to 224×224 pixels. The classification performance of the above three models was comprehensively analyzed and compared using several metrics, as detailed in Section 2.5. The accuracy variation curves of the training and validation sets for the three models are shown in Figure 7.

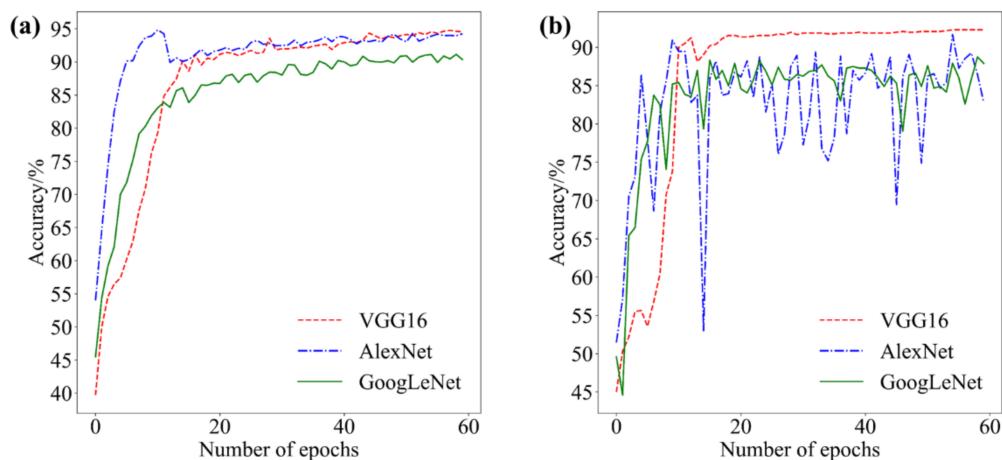


Figure 7. Accuracy curves for the training (a) and validation (b) sets of three classical models.

It can be seen from Figure 7 that VGG16 and AlexNet have comparable results in the training set, with accuracies of 94.84% and 94.78% (Table 3), respectively. GoogLeNet has much worse results, with an accuracy of 91.15%. In the validation set, VGG16 has much higher accuracy (92.29%) than AlexNet (91.66%) and GoogLeNet (88.96%). Moreover, the

VGG16 model is relatively more stable during the training process. The reason for this could be that the VGG16 model has more parameters, which allows it to fit the training data better, resulting in better performance. In contrast, GoogLeNet and AlexNet have relatively fewer parameters, resulting in a relatively poor performance. However, the VGG16 has the largest number of model parameters, exceeding AlexNet's by more than two-fold and GoogLeNet's by ten-fold. Moreover, the VGG16 model has the slowest inference speed and the largest model size. Therefore, the following study further evaluated the VGG16 model and attempted to speed up inference by eliminating some of the convolutional layers to reduce the parameters.

Table 3. Best classification results of the three models in the validation set.

Model	Training Accuracy (%)	Validation Accuracy (%)	Loss	Parameter ($\times 10^7$)	Model Size (MB)	Infer Time (ms)
AlexNet	94.78	91.66	0.73	6.1	217	4.81
VGG16	94.84	92.29	0.52	13.8	512.22	5.77
GoogLeNet	91.15	88.96	0.65	1.3	39.3	4.91

3.3. Testing Results

To assess the generalizability of the VGG16 model, an independent test set of 200 apples was utilized. This set comprised 50 non-deformed stripe-red apples, 50 non-deformed slice-red apples, 50 deformed apples with a deformity index greater than 12, and another 50 deformed apples with a deformity index between 10 and 12. Figure 8a shows the confusion matrix results of the VGG16 model on the test dataset. The testing results of stripe-red and slice-red apples were 98.00%, which is better than the results of VGG16 on the validation set (96.91%). However, the accuracy for classifying deformed apples in the test set was only 83.00%, which was relatively close to 82.69% in the validation set. Further analyses demonstrated that most of the misjudgments occurred in the deformed apples with a deformity index between 10 and 11 (66.00%). These apples closely resembled non-deformed apples in appearance. Conversely, the model achieved a high accuracy of 98.00% for classifying deformed apples with a deformity index greater than 11. These results demonstrate that the VGG16 has good classification performance for apples with a deformity index greater than 11, while it is less effective for apples with a deformity index between 10 and 11. The reason for this may be that the largest-diameter image of the deformed apples with small deformity indices did not characterize their deformities well, resulting in poor recognition by the model. Figure 8 shows that the classification accuracy of the VGG16 model in the test and validation sets is 90.50% and 92.29%, respectively.

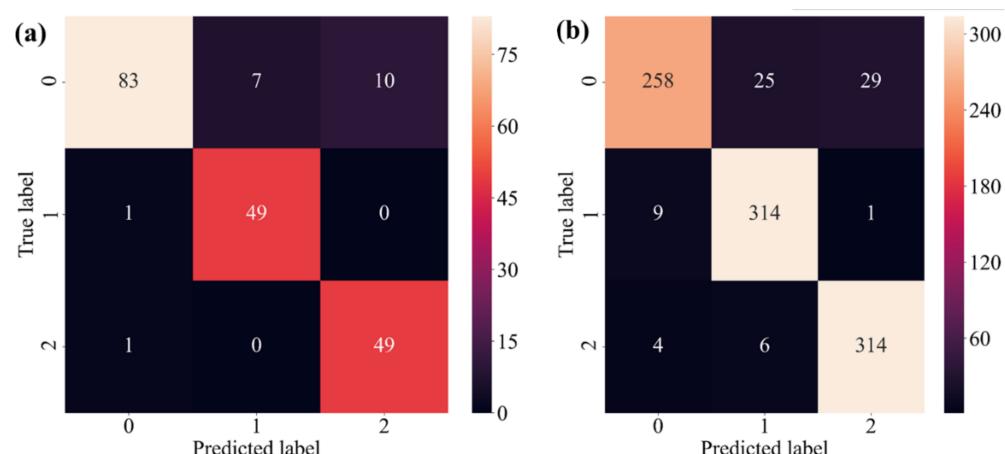


Figure 8. Confusion matrixes of the VGG16 in the test set (a) and validation set (b): 0 is the deformed apple, 1 is the non-deformed stripe-red apple, and 2 is the non-deformed slice-red apple.

Figure 8b shows the confusion matrix of the original VGG16 for the validation set, while other model performance metrics derived from the confusion matrix are summarized in Table 4. Because of the excessive discrimination errors of deformed apples, the precision of the stripe-red and slice-red apples is significantly lower than that of the deformed apples. The F1-score values of approximately 94% indicate that the classification performance of the model for stripe-red and slice-red apples is better.

Table 4. Evaluation metrics of the VGG16 model based on the confusion matrix of the validation set. 0, 1 and 2 denote the deformed apples, non-deformed stripe-red apples, and non-deformed slice-red apples, respectively.

Category	Precision (%)	Recall (%)	F1-Score (%)
0	95.21	82.69	88.54
1	91.03	96.91	93.90
2	91.30	96.91	94.02

3.4. Ablation Experiment

In this study, ablation experiments were conducted to understand the contribution of each convolutional layer in the VGG16 model to the classification performance (referring to Figure 5). In each experiment, all convolutional layers except the first one in each part were eliminated at one time. The results of the ablation experiments are summarized in Table 5, where VGG16-1, 2, 3, 4, and 5 are denoted as convolutional elimination for the first, second, third, fourth, and fifth parts of the model, respectively. Experiments have shown that by modifying the convolutional kernels, pooling kernels, and strides to reduce the features more quickly, and by removing the fully connected layers, the VGG16 model can be improved in terms of size and time efficiency.

Table 5. Performance of the VGG16 model with different convolutional layers removed in the validation set.

Model	Accuracy (%)	Parameter ($\times 10^7$)	Model Size (MB)	Infer Time (ms)
VGG16	92.29	13.8	512.22	5.77
VGG16-1	86.35	13.8	512.08	5.56
VGG16-2	89.68	13.8	511.66	5.68
VGG16-3	89.58	13.7	507.72	5.65
VGG16-4	90.52	13.4	494.22	5.71
VGG16-5	88.54	13.4	494.22	5.73

It can be seen from Table 5 that the VGG16 model has different degrees of accuracy degradation after eliminating each convolutional layer. The accuracy of VGG16-1 is the lowest, indicating that the eliminated convolutional layer of this layer has the greatest impact on the classification accuracy. The convolutional layer removed in VGG16-1 was in a more forward position in the original model. Removing this layer may result in a larger loss of deeper feature representations and a larger impact on the dependencies between each layer. However, ablation experiments revealed minimal changes in the model parameters, size, and inference time, after removing individual convolutional layers. This suggests that each convolutional layer in the original VGG16 model contributes unique feature extraction capabilities, making them irreplaceable for optimal performance.

3.5. Discussion

In this study, the VGG16 model achieved comparable classification accuracy for apple color and deformity compared to some existing studies for apple grading (Table 6). For example, Li et al. (2020) [21] used a shallow CNN model to classify apple varieties without occlusion and achieved an accuracy of 92.00%, while Li et al. (2021) [22] and Fan et al. (2020) [23] used CNNs for classifying apple defects and achieved an overall accuracy of

95.33% and 92.15%, respectively. Shallow CNNs have good applicability for single-feature classification, while for multi-feature classification tasks, some classical deep CNNs have better results. Ji et al. (2023) [39] improved the YOLOv5s by adding some modules to the model and applied the improved model to classify apples based on color, shape, diameter, and defect. The average accuracy of the final model for apple quality classification was 94.46%. In this study, the AlexNet, GoogLeNet and VGG16 models were used for classifying deformed apples in conjunction with the color feature. Ultimately, VGG16 was found to be the most effective, with results of 92.29% and 90.50% for the validation and test sets, respectively. The reason for this is that VGG16 has more convolutional layers and fewer but larger fully connected layers compared to AlexNet and GoogLeNet, making it more complex and deeper. Therefore, VGG16 can better learn and represent image features, which helps to improve performance in classification and detection tasks. Achieving comparable performance to published literature validates the feasibility of the VGG16 model for classifying apples based on color and deformed features.

Table 6. Comparison of our results with those of other studies on apple classification.

Feature	Model	Accuracy (%)	Reference
Diameter, defect	Multi-view spatial network	99.24	Shi et al. (2022) [24]
Variety	shallow CNN	92.00	Li et al. (2020) [21]
Defect	CNN	95.33	Li et al. (2021) [22]
Color, shape, diameter, defect	Improved YOLOv5s	94.46	Ji et al. (2023) [39]
Defect	CNN	92.15	Fan et al. (2020) [23]
Color, deformity	VGG16	92.29	This study

Compared to existing studies, this research addresses the challenge of classifying features of deformed, stripe-red, and slice-red apples. By integrating deep learning techniques, an efficient model has been established. It should be noted that the features for classifying apples are different between our study and the other literature. The feature of deformity is challenging to identify, compared to the features of shape, defect, and variety, and thus resulting in relatively low accuracy. Future work for improving the classification of deformed apples could include using a 3D scanner to obtain the deformity index of apples and enhancing the accuracy of classification based on the original dataset. In the future, research can also be directed to a lightweight CNN model for simplifying the model architecture and further improving the classification accuracy and efficiency. Furthermore, future work can explore the adoption of more advanced deep learning networks, such as using YOLO series networks, generative adversarial networks (GAN), or transformers.

4. Conclusions

In this study, a machine vision system was constructed to acquire apple images. By applying image processing, the system then automatically selected the image with the largest diameter from 45 images obtained for each apple. Three classical deep learning models, AlexNet, GoogLeNet, and VGG16, were utilized for achieving a three-category classification task of apples based on their color and deformity features. The results demonstrated that the VGG16 exhibited the most optimal classification results, achieving an accuracy of 94.84% and 92.29% on the training and validation sets, respectively. The model's generalizability was evaluated using a separate test set of two hundred apples. It achieved an overall accuracy of 90.50%, with most misjudgments primarily concentrated in the deformed category. Ablation experiments indicated that all convolutional layers within the VGG16 model contribute to its strong classification performance. In future research, lightweight models can be explored to accelerate the classification task, as well as to further improve the accuracy of classifying apples based on color and deformity, using more advanced deep learning networks.

Author Contributions: Conceptualization, D.Q. and D.H.; methodology, D.Q. and T.G.; software, D.Q.; validation, T.G. and S.Y.; resources, W.L.; data curation, D.Q., T.G. and S.Y.; writing—original draft preparation, D.Q. and T.G.; writing—review and editing, W.L., Z.S., L.L., H.P. and D.H.; supervision, H.P. and D.H.; funding acquisition, H.P. and D.H. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the financial support provided by the National Natural Science Foundation of China (No. 32371987), Natural Science Foundation of Zhejiang Province (No. LY24C130001), Program of the Key Laboratory of Modern Agricultural Equipment and Technology (Jiangsu University), Ministry of Education, P.R. China (No. MAET202303), and the Research Project of Zhejiang A&F University (No. 2023LFR151).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

1. National Bureau of Statistics of China in 2023. Available online: <https://www.stats.gov.cn/sj/ndsj/2023/indexch.htm> (accessed on 30 April 2023).
2. Sun, Z.; Hu, D.; Xie, L.; Ying, Y. Detection of early stage bruise in apples using optical property mapping. *Comput. Electron. Agric.* **2022**, *194*, 106725. [[CrossRef](#)]
3. Liu, T.; He, J.; Yao, W.; Jiang, H.; Chen, Q. Determination of aflatoxin B1 value in corn based on Fourier transform near-infrared spectroscopy: Comparison of optimization effect of characteristic wavelengths. *LWT* **2022**, *164*, 113657. [[CrossRef](#)]
4. Shi, Y.; Wang, Y.; Hu, X.; Li, Z.; Huang, X.; Liang, J.; Zhang, X.; Zheng, K.; Zou, X.; Shi, J. Nondestructive discrimination of analogous density foreign matter inside soy protein meat semi-finished products based on transmission hyperspectral imaging. *Food Chem.* **2023**, *411*, 135431. [[CrossRef](#)] [[PubMed](#)]
5. Xu, Q.; Cai, J.-R.; Zhang, W.; Bai, J.-W.; Li, Z.-Q.; Tan, B.; Sun, L. Detection of citrus Huanglongbing (HLB) based on the HLB-induced leaf starch accumulation using a home-made computer vision system. *Biosyst. Eng.* **2022**, *218*, 163–174. [[CrossRef](#)]
6. Zhang, B.; Huang, W.; Li, J.; Zhao, C.; Fan, S.; Wu, J.; Liu, C. Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Res. Int.* **2014**, *62*, 326–343. [[CrossRef](#)]
7. Hu, D.; Jia, T.; Sun, X.; Zhou, T.; Huang, Y.; Sun, Z.; Zhang, C.; Sun, T.; Zhou, G. Applications of optical property measurement for quality evaluation of agri-food products: A review. *Crit. Rev. Food Sci. Nutr.* **2023**, 1–21. [[CrossRef](#)] [[PubMed](#)]
8. Zou, X.; Zhao, J.; Li, Y.; Holmes, M. In-line detection of apple defects using three color cameras system. *Comput. Electron. Agric.* **2010**, *70*, 129–134.
9. Hu, G.; Zhang, E.; Zhou, J.; Zhao, J.; Gao, Z.; Sugirbay, A.; Jin, H.; Zhang, S.; Chen, J. Infield Apple Detection and Grading Based on Multi-Feature Fusion. *Horticulturae* **2021**, *7*, 276. [[CrossRef](#)]
10. Song, T.-H.; Sanchez, V.; Daly, H.E.; Rajpoot, N.M. Simultaneous cell detection and classification in bone marrow histology images. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 1469–1476. [[CrossRef](#)]
11. Zhou, J.; Wu, Z.; Jiang, Z.; Huang, K.; Guo, K.; Zhao, S. Background selection schema on deep learning-based classification of dermatological disease. *Comput. Biol. Med.* **2022**, *149*, 105966. [[CrossRef](#)]
12. Zhang, C.; Xia, K.; Feng, H.; Yang, Y.; Du, X. Tree species classification using deep learning and RGB optical images obtained by an unmanned aerial vehicle. *J. For. Res.* **2021**, *32*, 1879–1888. [[CrossRef](#)]
13. Chen, Y.; Wang, J. Deep learning for crown profile modelling of *Pinus yunnanensis* secondary forests in Southwest China. *Front. Plant Sci.* **2023**, *14*, 1093905. [[CrossRef](#)] [[PubMed](#)]
14. Sun, X.; Xu, L.; Zhou, Y.; Shi, Y. Leaves and twigs image recognition based on deep learning and combined classifier algorithms. *Forests* **2023**, *14*, 1083. [[CrossRef](#)]
15. Feng, J.; Hou, B.; Yu, C.; Yang, H.; Wang, C.; Shi, X.; Hu, Y. Research and validation of potato late blight detection method based on deep learning. *Agronomy* **2023**, *13*, 1659. [[CrossRef](#)]
16. Hu, D.; Qiu, D.; Yu, S.; Jia, T.; Zhou, T.; Yan, X. Integration of optical property mapping and machine learning for real-time classification of early bruises of apples. *Food Bioprocess Technol.* **2023**, 1–12. [[CrossRef](#)]
17. Tao, K.; Wang, A.; Shen, Y.; Lu, Z.; Peng, F.; Wei, X. Peach flower density detection based on an improved CNN incorporating attention mechanism and multi-scale feature fusion. *Horticulturae* **2022**, *8*, 904. [[CrossRef](#)]
18. Shuprajhaa, T.; Raj, J.M.; Paramasivam, S.K.; Sheeba, K.; Uma, S. Deep learning based intelligent identification system for ripening stages of banana. *Postharvest Biol. Technol.* **2023**, *203*, 112410. [[CrossRef](#)]
19. Deng, L.; Li, J.; Han, Z. Online defect detection and automatic grading of carrots using computer vision combined with deep learning methods. *LWT* **2021**, *149*, 111832. [[CrossRef](#)]

20. Sun, Z.; Xie, L.; Hu, D.; Ying, Y. An artificial neural network model for accurate and efficient optical property mapping from spatial-frequency domain images. *Comput. Electron. Agric.* **2021**, *188*, 106340. [[CrossRef](#)]
21. Li, J.; Xie, S.; Chen, Z.; Liu, H.; Kang, J.; Fan, Z.; Li, W. A shallow Convolutional Neural Network for apple classification. *IEEE Access* **2020**, *8*, 111683–111692. [[CrossRef](#)]
22. Li, Y.; Feng, X.; Liu, Y.; Han, X. Apple quality identification and classification by image processing based on convolutional neural networks. *Sci. Rep.* **2021**, *11*, 16618. [[CrossRef](#)] [[PubMed](#)]
23. Fan, S.; Li, J.; Zhang, Y.; Tian, X.; Wang, Q.; He, X.; Zhang, C.; Huang, W. On line detection of defective apples using computer vision system combined with deep learning methods. *J. Food Eng.* **2020**, *286*, 110102. [[CrossRef](#)]
24. Shi, X.; Chai, X.; Yang, C.; Xia, X.; Sun, T. Vision-based apple quality grading with multi-view spatial network. *Comput. Electron. Agric.* **2022**, *195*, 106793. [[CrossRef](#)]
25. Ünal, Z.; Kızıldeniz, T.; Özden, M.; Aktaş, H.; Karagöz, Ö. Detection of bruises on red apples using deep learning models. *Sci. Hortic.* **2024**, *329*, 113021. [[CrossRef](#)]
26. Fu, Y.; Song, J.; Xie, F.; Bai, Y.; Zheng, X.; Gao, P.; Wang, Z.; Xie, S. Circular fruit and vegetable classification based on optimized GoogLeNet. *IEEE Access* **2021**, *9*, 113599–113611.
27. Ni, J.; Gao, J.; Deng, L.; Han, Z. Monitoring the change process of banana freshness by GoogLeNet. *IEEE Access* **2020**, *8*, 228369–228376. [[CrossRef](#)]
28. McCamy, C.S.; Marcus, H.; Davidson, J.G. A color-rendition chart. *J. Appl. Photogr. Eng.* **1976**, *2*, 95–99.
29. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
30. Krizhevsky, A.; Ilya, S.; Geoffrey, E.H. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*; Neural Information Processing Systems Foundation: La Jolla, CA, USA, 2012.
31. Gayathri, D.; Kishore, B.; Senthikumar, C. Feature analysis and classification of maize crop diseases employing AlexNet-inception network. *Multimed. Tools Appl.* **2024**, *83*, 26971–26999.
32. Ni, J.; Gao, J.; Li, J.; Yang, H.; Hao, Z.; Han, Z. E-AlexNet: Quality evaluation of strawberry based on machine learning. *J. Food Meas. Charact.* **2021**, *15*, 4530–4541. [[CrossRef](#)]
33. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
34. Swarup, C.; Singh, K.U.; Kumar, A.; Pandey, S.K.; Varshney, N.; Singh, T. Brain tumor detection using CNN, AlexNet & GoogLeNet ensembling learning approaches. *Electron. Res. Arch.* **2023**, *31*, 2900–2924.
35. Yang, L.; Yu, X.; Zhang, S.; Long, H.; Zhang, H.; Xu, S.; Liao, Y. GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases. *Comput. Electron. Agric.* **2023**, *204*, 107543. [[CrossRef](#)]
36. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
37. Liu, Z.; Wu, J.; Fu, L.; Majeed, Y.; Feng, Y.; Li, R.; Cui, Y. Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion. *IEEE Access* **2020**, *8*, 2327–2336. [[CrossRef](#)]
38. Yang, H.; Ni, J.; Gao, J.; Han, Z.; Luan, T. A novel method for peanut variety identification and classification by Improved VGG16. *Sci. Rep.* **2021**, *11*, 15756. [[CrossRef](#)] [[PubMed](#)]
39. Ji, W.; Wang, J.; Xu, B.; Zhang, T. Apple Grading based on multi-dimensional view processing and deep learning. *Foods* **2023**, *12*, 2117. [[CrossRef](#)]

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.