


Article

Apple Surface Defect Detection Method Based on Weight Comparison Transfer Learning with MobileNetV3

Haiping Si ¹, Yunpeng Wang ¹, Wenrui Zhao ¹, Ming Wang ¹, Jiazhen Song ¹, Li Wan ¹, Zhengdao Song ¹, Yujie Li ¹, Bacao Fernando ²  and Changxia Sun ^{1,*}

¹ College of Information and Management Science, Henan Agricultural University, Zhengzhou 450046, China

² NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, 1099-085 Lisbon, Portugal

* Correspondence: sunchangxia@henau.edu.cn

Abstract: Apples are ranked third, after bananas and oranges, in global fruit production. Fresh apples are more likely to be appreciated by consumers during the marketing process. However, apples inevitably suffer mechanical damage during transport, which can affect their economic performance. Therefore, the timely detection of apples with surface defects can effectively reduce economic losses. In this paper, we propose an apple surface defect detection method based on weight contrast transfer and the MobileNetV3 model. By means of an acquisition device, a thermal, infrared, and visible apple surface defect dataset is constructed. In addition, a model training strategy for weight contrast transfer is proposed in this paper. The MobileNetV3 model with weight comparison transfer (Weight Compare-MobileNetV3, WC-MobileNetV3) showed a 16% improvement in accuracy, 14.68% improvement in precision, 14.4% improvement in recall, and 15.39% improvement in F1-score. WC-MobileNetV3 compared to MobileNetV3 with fine-tuning improved accuracy by 2.4%, precision by 2.67%, recall by 2.42% and F1-score by 2.56% compared to the classical neural networks AlexNet, ResNet50, DenseNet169, and EfficientNetV2. The experimental results show that the WC-MobileNetV3 model adequately balances accuracy and detection time and achieves better performance. In summary, the proposed method achieves high accuracy for apple surface defect detection and can meet the demand of online apple grading.

Keywords: defect detection; image fusion; deep learning; transfer learning; weight comparison



Citation: Si, H.; Wang, Y.; Zhao, W.; Wang, M.; Song, J.; Wan, L.; Song, Z.; Li, Y.; Fernando, B.; Sun, C. Apple Surface Defect Detection Method Based on Weight Comparison Transfer Learning with MobileNetV3.

Agriculture **2023**, *13*, 824.

<https://doi.org/10.3390/agriculture13040824>

Academic Editor: Maciej Zaborowicz

Received: 2 March 2023

Revised: 30 March 2023

Accepted: 31 March 2023

Published: 3 April 2023



Copyright: © 2023 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

Consumer demand for fresh fruit is increasing. Apples are popular with consumers because of their good taste and rich nutrition [1]. Consumers often prefer to buy apples that are brightly colored, regular in shape, and have no visible surface scars. Therefore, grading apples according to their appearance is an important process to improve the economic efficiency of the apple industry [2]. However, the current postharvest work of fruits is still a labor-intensive task performed by manual laborers [3]. Hence, slight mechanical damage such as scratches and bruises is inevitable during the commercial postharvest processing of apples (collection, manual sorting, storage, and transportation) [4,5]. The surface defects of apples caused by these mechanical damages can lead to certain physiological changes [6], such as water loss and rot, which can lead to a shortened guarantee period and a drop in quality, which in turn can cause the loss of commercial value of apples and severely limit the increase in value of the apple industry [7]. In addition, the current fruit surface defect detection task is usually performed manually [8] and quality inspectors can lead to misjudgments due to excessive working time, resulting in a decrease in detection efficiency.

Currently, computer vision techniques based on visible imaging systems and spectroscopic techniques are most widely used in fruit surface defect detection tasks [9]. In addition, numerous researchers have achieved defect detection of high precision on visible imaging systems by implementing machine learning and deep learning techniques

for grading and defect detection of various types of fruits, overcoming the limitations of classical computational paradigms.

1.1. Machine Learning Technique on Surface Defect Detection of Fruit

Machine learning techniques have played an essential role in the development of artificial intelligence, image processing, and data analysis [10], and this technique has significant results in searching, processing, and analyzing data acquired by most sensors [11]. In recent years, some researchers proposed some technological solutions based on machine learning for the task of fruit surface defect detection and achieved better results [12]. Moallem et al. [13] proposed a machine learning-based apple grading algorithm. The authors classified apples into two categories: defective and healthy. Apple grading is performed using a classifier support vector machine (SVM) and K-nearest neighbor (KNN). Bhargava et al. [14] developed an automatic apple grading system. Different combinations of several features are considered based on the damage exposed to the apple. In this work, these features are considered inputs for training an SVM model. The system achieved maximum accuracies of 96.81% and 93.00% for both datasets using k-fold cross-validation techniques. Zhang et al. [15] constructed a computer vision system that combines automatic brightness correction, defect candidate region counting, and a weighted correlation relevance vector machine (RVM) classifier to propose a novel method for the automatic detection of defective apples. The overall detection accuracy is 95.63% for 160 samples.

Chithra et al. [16] proposed a global thresholding algorithm to segment the defective parts of apple images, and then the coefficients obtained from wavelet transform and Haar filtering are used to extract features from the segmented images. The plain Bayesian classifier is used to classify and identify the defective and intact fruits according to the extracted features. The experimental results show that the average accuracy of the plain Bayesian classifier using the global thresholding algorithm is 96.67%, which is 31.67% and 3.34% higher compared to the Otsu segmentation algorithm and the K-pur segmentation algorithm. Tan et al. [17] proposed a citrus surface defect detection method based on KF-2D-Renyi and ABC-SVM. Firstly, Kent chaos theory (RF-2D-Renyi) is introduced in the firefly algorithm, which is based on the principles of ergodicity and randomness of chaotic sequences to find the optimal threshold. Edge features, texture features, and geometric features are extracted from the obtained segmented images, and then the extracted features are input to the support vector machine classifier which is optimized by the artificial bee colony algorithm (ABC-SVM) for the recognition of defects. The experimental results indicate that the average accuracy of this method for eight types of defect recognition is 98.45%. Wang et al. [18] developed a region of interest (ROI) extraction algorithm based on background separation, luminance correction, and global threshold segmentation. After luminance correction, 0.8 times the average gray value of the apple region is used as a threshold to extract the region of interest. The SVM model is built by extracting the texture features of the region of interest. The experimental results show that the average accuracies of the classifier based on angular second moments and the classifier model based on entropy are 94.8% and 94.7%, respectively, and the detection time of a single apple is about 1.2 s.

All of the above studies basically follow the steps of image segmentation, feature extraction, and feature classification recognition to solve the task of fruit surface defect detection. However, all of these methods experience the problems of complex processes, detection accuracy cannot reach the actual production requirements, and detection time is slow.

1.2. Deep Learning Technique on Surface Defect Detection of Fruit

In recent years, with the development of deep learning and deep learning techniques have been widely used in agriculture [19]. Meanwhile, some researchers started to apply deep learning techniques to solve fruit surface defect detection tasks and achieved good detection results. Deep learning-based fruit detection methods use end-to-end deep networks, which allow for the integration of feature extraction. This approach greatly simplifies the

algorithm process. The robustness and accuracy of the detection results are significantly improved with the support of large data volumes. For example, Zhou et al. [20] modified ResNet50 to obtain WideResNet50, while the AdamW optimizer and the weighted cross-entropy loss function are used during the training process. Experiments are conducted on a self-built dataset of plum surface defects and the experimental results indicate that the average accuracy of the model to identify defects is 98.95%, and the detection time is 0.1037 s for a single plum image. Deng et al. [21] constructed a lightweight deep learning model (CDDNet) based on ShuffleNet and transfer learning for the carrot surface defect detection task. The experimental results indicate that the average accuracy of the model for carrot surface defect detection is 99.82%. Yao et al. [22] developed a YOLOv5-based kiwi surface defect detection model, which improved the original YOLOv5 by adding a small object detection layer, embedding SE attention module, adopting CIoU loss function, and transfer learning to improve the performance of YOLOv5. The experimental results indicate that the mAP@50 of this model can reach 94.7% and the detection time of a single image is about 0.1 s. Da Costa et al. [23] obtained an improved ResNet50 model by fine-tuning all layers of the ResNet50 model and performing experiments on a self-constructed dataset of tomato surface defects. The experimental results indicate that the average accuracy of the model for recognition on the test set is 91.7%.

All the above methods of fruit surface defect detection based on machine learning techniques and deep learning techniques are performed on a visible imaging system, which can effectively detect more obvious defects, such as insect spots, rot, and other defects, but the detection of slight mechanical damage is not ideal, which is due to the fact that the detection precision of slight mechanical damage is affected by the background color of the defect area.

1.3. Thermal Technique on Surface Defect Detection of Fruit

In general, according to wavelength bands, infrared spectra can be divided into near-infrared (0.75–3 μm), mid-infrared (3–6 μm), far-infrared (6–15 μm), and ultra-infrared (15–1000 μm) spectra. Infrared thermography is a non-invasive, non-contact, and non-destructive technique [24]. Infrared images obtained by thermal imaging sensors are acquired by photoelectric conversion, which transforms thermal radiation into a grayscale image that is used to describe the temperature distribution on the surface of the object. According to the need for external excitation sources, thermal infrared imaging techniques can be divided into two categories, namely active and passive thermography [25]. Active thermography requires an external heat source to heat the object, while passive thermography does not. Passive thermography is used to measure the temperature difference between the object and its surrounding environment in the natural environment. Some scholars have introduced passive thermography to the field of fruit surface defect detection.

For example, Jawale and Deshmukh [26] proposed a method for fruit bruise detection based on passive thermography and image processing. The method first uses an infrared thermal imager to acquire image data, preprocesses the acquired thermal images, extracts the features of energy, entropy, contrast, mean, correlation, standard deviation, and homogeneity from the processed images, and uses an artificial neural network (ANN) for classification and recognition based on the extracted features.

Compared to passive thermography, active thermography can detect defects that are produced on the surface and subsurface of the object, as well as the depth of the defect based on the characteristics of the thermal pulse and the change in the object temperature with time [27]. In addition, infrared thermography differs from other imaging techniques in that the technique does not need to solve the problem of inhomogeneous illumination and scattering caused by spherical fruits and only requires the stable temperature of the environment. Therefore, with its unique advantages, the infrared thermography technique is applied by many scholars in the field of fruit defect detection.

The earlier related studies including Varith et al. [28] use active thermography to detect bruises on the surface of apples, where the apples are kept in a refrigerator for 3 h before

acquiring infrared images, following which the infrared images are captured in heating and cooling processes, respectively. The accuracy of identification is 100% for apple defects under heating treatment and 66% for apple defects under the cooling process, and they observed that the heating process is more suitable for detecting fruit surface defects. With the development of infrared thermography and deep learning techniques, scholars are starting to combine both of them for fruit surface defect detection tasks; for example, Zeng et al. [29] constructed a simple infrared thermography system for studying the detection of pear bruises based on infrared thermography, and the defective parts of pears are analyzed at different days. Meanwhile, the infrared image dataset of pear surface defects required for training the deep learning model is constructed during the process of analysis. The recognition experiments are performed under the convolutional neural network model and the experimental results show that the accuracy of the model for recognition of defective and intact pears is 99.3% when the number of iterations is 20. Dong et al. [30] developed a method to detect bruises on jujube based on infrared thermography and a convolutional neural network. By constructing an infrared thermography system to build an infrared image dataset of jujube surface defects and analyzing the captured infrared images at the same time, the temperature difference at the boundary of the bruise region can be known to be in the range of 1.72–3.25 °C, and the bruise degree of jujubes can be judged based on the difference in temperature. When the DenseNet is modified, the modified DenseNet performs classification recognition experiments on the infrared image dataset of jujubes surface defects and the experimental results indicate that the recognition accuracy for jujube bruises is 99.5%.

According to the above research, it is known that mechanical damages that are not obvious on visible images, such as slight scratches and bruises, are more difficult to identify on visible images due to factors such as color features and texture features on the fruit surface. Compared with the method of fruit surface defect detection based on visible images, the mechanical damages suffered on the fruit surface have more obvious features in the infrared image. Therefore, the development of infrared thermography provides a method to identify surface defects, such as slight scratches and bruises, by detecting changes in the thermal conductivity of fruit tissue and obtaining relatively better detection results.

1.4. The Contribution of this Research

Therefore, to better achieve the detection of apple surface defects, calyx, and apple stalks in this paper, a thermal infrared and visible fusion algorithm is introduced to construct a thermal infrared and visible apple surface defect image dataset, and a training strategy based on the comparative transfer of weight parameters is proposed to train MobileNetV3, which provides a new idea for apple surface defect detection. The main work of this study is described as follows:

- (1) Building a thermal infrared and visible image dataset of apple surface defects using a thermal infrared and visible apple image acquisition device.
- (2) A model training strategy for weight comparison transfer is proposed. The specific approach is to compare the pretraining weights with the default weights of the model itself, thus allowing the model to maximize the extraction of the required feature parameters from the pretraining weights during training.
- (3) Comparison experiments on different spectral datasets (VIS, IR, and VIS + IR). The effectiveness of fused image datasets is verified compared to single light image datasets.
- (4) Ablation experiments are performed on the model training strategy of weight contrast transfer. The experiments are compared with three model training strategies of no freezing (no freezing of network layers), fine tuning (freezing of network layers), and freezing (freezing of all network layers). WC-MobileNetV3 is compared with other classical convolutional neural networks to verify the superiority of WC-MobileNetV3 for the recognition accuracy of apples subject to minor mechanical damage.

2. Materials and Methods

2.1. Image Acquisition

The thermal infrared and visible apple surface defect image dataset constructed in this paper is acquired by a thermal infrared and visible apple image acquisition device built by the group of Si Haiping from Henan Agricultural University, China, as shown in Figure 1. The thermal infrared and visible apple images are acquired using a dual-light camera equipped with the Yu2 Industry Advance UAV developed and manufactured by Shenzhen DJI Innovation Technology Co., Shenzhen, China. The dual-light camera device consists of an RGB camera with a color image resolution (8000×6000 pixels) and a thermal infrared imaging camera with a thermal image resolution (640×512 pixels). The thermal imaging camera has a long infrared wavelength of 8–14 μm , a temperature measurement range of $-40\text{ }^{\circ}\text{C}$ to $550\text{ }^{\circ}\text{C}$, a temperature measurement accuracy of $\pm 2\text{ }^{\circ}\text{C}$, and a fixed-focus lens focal length of 9 mm. The specific parameters are shown in Tables 1 and 2.

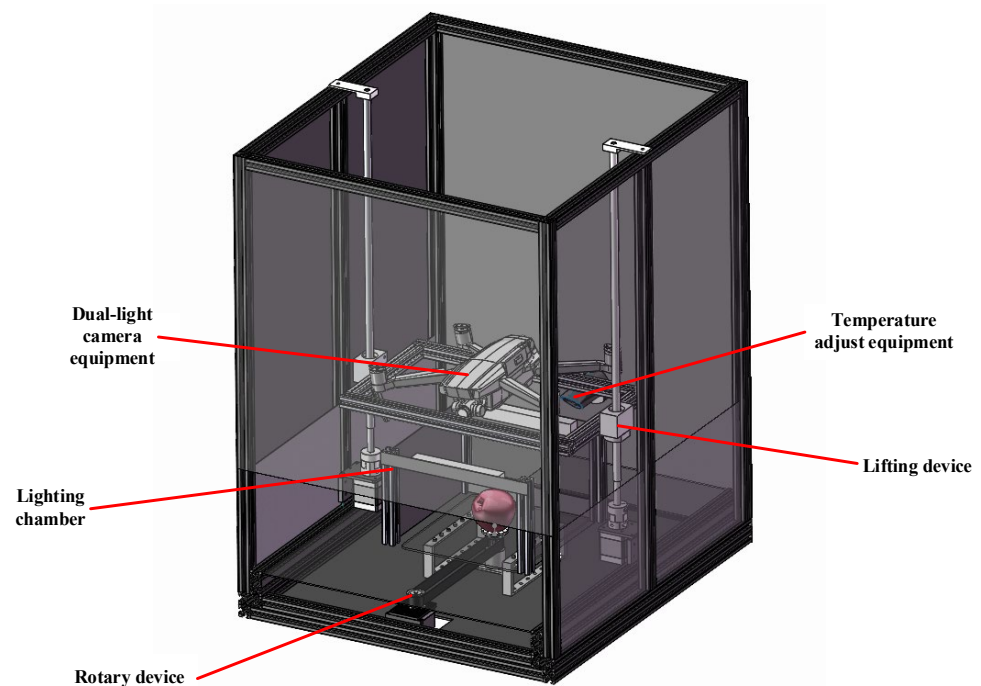


Figure 1. Dual light apple image acquisition system.

Table 1. Infrared camera.

Sensors	Lens Focal Length	Sensor Resolution	Infrared Temperature Measurement Accuracy
Uncooled VO_x Microbolometer	Fixed-focus lens focal length approx. 9 mm; equivalent focal length approx. 38 mm	640×512 @30 Hz	$\pm 2\text{ }^{\circ}\text{C}$ or $\pm 2\%$ (whichever is greater)

Table 2. Visible camera.

Image Sensor	Lens	Sensor Resolution	Infrared Temperature Measurement Accuracy
12.7 mm CMOS; effective pixels 48 million	Viewing angle: 84° ; equivalent focal length: 24 mm; aperture: $f/2.8$; focus point: 1 m to infinity	Video: 100 to 12,800 (auto) Photo: 100~1600 (auto)	$32\times$ digital zoom

In order to meet the conditions of uniform illumination and adequate heating of the test sample, a semi-enclosed chamber is built, which consisted of two LED strips and two black baffles. The LED strips are located on the front inner side and the back inner side of the semi-enclosed chamber. As the thermal properties of defective apples differ from those of intact apples, the temperature on their surface will show a non-uniform distribution under the action of an external excitation source [31]. Therefore, this study uses active external excitation loaded with thermal waves to achieve temperature difference enhancement [32], while in practical application scenarios, hot fans are used as temperature regulation modules in image acquisition devices due to their advantages such as uniform heating and long-acting distance. Furthermore, this is used in order to overcome the problem of large spatial gradients in the acquired images due to overheating caused by the close heating of the hot blower. Designing the fixed position and heating distance of the hot blower: a 500 W hot blower is fixed at the rear of the semi enclosure. After repeated tests, the distance between the hot blower and the sample surface is finally determined to be 0.4 m, with the air outlet at an angle of 45° to the floor. The arrangement of the temperature regulation module is shown in Figure 2.

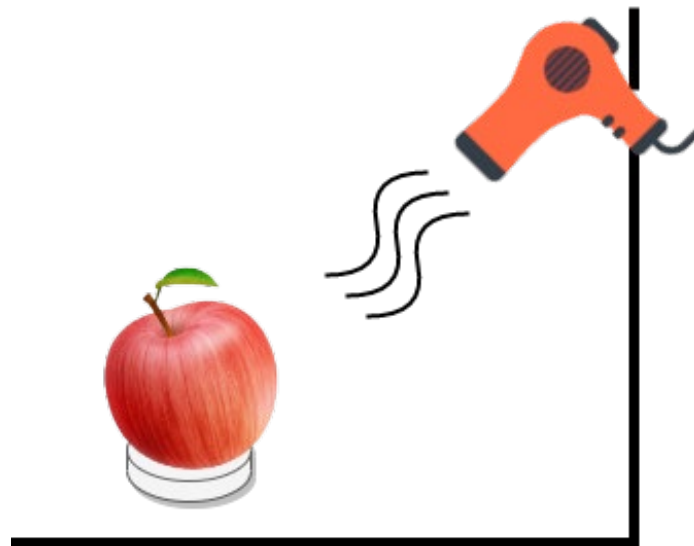


Figure 2. Temperature adjustment module.

When acquiring thermal infrared and visible images of apples, the thermal excitation source and the dual-light camera need to be as spatially separated as possible to minimize the effect of external thermal noise and to improve the contrast between the target and the background. To facilitate the adjustment of the distance between the dual-light camera and the sample surface for subsequent work, an automatic lifting device is constructed. Placing the apple sample on top of the rotating device allows the apples to be uniformly heated. In this study, the distance between the dual-light camera and the sample surface is set to 0.45 m and the sample is focused to keep the thermal infrared and visible images of the apples in sharp focus.

The lifting device consists of two 1204 screws, two couplings, two 42 stepper motors, and a camera bracket. The device adjusts the distance between the camera and the apple sample by rotating the drive motor and the distance is set to 400 mm. The rotating device consists mainly of the drive motor, two timing pulleys, the timing belt, and the apple sample support bracket. The stepper motor mounted under the apple sample holder drives the timing pulleys via the timing belt. By controlling the speed of the stepper motor, the attitude of the apple sample can be adjusted to obtain a suitable thermal infrared thermal image and a visible image.

A total of 100 “Yantai Fuji”, 20 “Akesu Fuji”, 20 “Gansu Fuji”, and 20 “Shaotong Fuji” apples are purchased from Yaoqiao Farmers’ Market in Zhengzhou City, Henan Province,

China as test samples. The main sample used in this study is “Yantai Fuji” due to its wide availability in our city and easy accessibility. Therefore, it is chosen as the main experimental sample. To investigate the effect of mechanical damage on the surface defects of the apples, 100 samples are collected at 7:00 am on the same day for a wear test. The samples are equally divided into two groups: a control group and an experimental group. Additionally, to recreate the realism of the production scenario, the samples are subjected to abrasion using sharp objects common to the production scenario, such as at the edges of plastic packaging boxes. Each group of samples was placed in a box at a temperature of 15 °C and a humidity of 65 °C. Thermal infrared and visible images of the samples are taken from 4 pm onward. A total of 600 pairs of “Yantai Red Fuji”, 120 pairs of “Akesu Fuji”, 120 pairs of “Gansu Fuji”, and 117 pairs of “Shaotong Fuji” dual light images are collected.

2.2. Image Preprocessing

To investigate the effects of different divided ratios between the training set and the test set, the experimental dataset is divided into the training set and test set according to 9:1, 8:2, 7:3, 6:4, 5:5, 4:6, 3:7, 2:8, and 1:9 and the experimental results are shown in Table 3. According to the analysis of the experimental results in Table 3, the WC-MobileNetV3 model proposed in this paper basically achieves the accuracy rate of 98% or higher when the training set is above 60% and the highest accuracy rate of 99.20% is achieved when the ratio of the training set to the test set is 8:2. In general, when the training set is above 60%, accuracy rates of various divisions are similar, and there is not much difference between the division of training set and test set. When the training set accounted for less than 60%, the accuracy declines rapidly from above 98% to 91.81% and then to 63.77%, with a faster drop. Thus, 600 pairs of image data from the thermal infrared and visible apple image dataset are divided into a training set and test set according to 8:2 in this experiment.

Table 3. Comparison of experimental results of different proportions of experimental data.

Train:Test	Accuracy/%	Precision/%	Recall/%	F1-Score (%)	Parameter (M)	T _s (ms)
9:1	98.15	98.48	98.15	98.23	4.21	33.32
8:2	99.20	99.12	99.02	99.04	4.21	24.08
7:3	98.28	98.26	98.33	98.27	4.21	21.51
6:4	98.35	98.43	98.37	98.36	4.21	17.14
5:5	97.58	97.64	97.54	97.53	4.21	15.37
4:6	96.00	96.04	96.17	95.96	4.21	14.26
3:7	96.37	96.52	96.39	96.37	4.21	13.63
2:8	91.81	92.27	91.84	91.75	4.21	14.06
1:9	63.77	73.30	63.75	62.03	4.21	13.71

The apples are divided into intact, calyx, stem, and defect apples. In addition, since calyx and stem may appear in the same view as the defective part, two additional categories are classified as calyx + defect and stem + defect. The constructed thermal infrared and visible apple surface defect datasets are divided into six categories: intact apple, defect apple, calyx, apple stem, calyx + defect, and apple stem + defect.

To make the defective part of the thermal infrared image more distinctive, histogram equalization is selected to perform image enhancement operations on the thermal infrared image. The various types of dual-light image datasets are shown in Figure 3. By using the RFN-Nest thermal infrared and visible fusion algorithm [33], the VIS + IR dataset, VIS dataset, and IR dataset are obtained for experimental use. Additionally, this study has a small sample dataset. To reduce the possibility of overfitting the apple surface defect detection model during the training process, to improve its generalization ability and to increase the learning efficiency of the model for defect features, data extensions must be performed on the dataset.

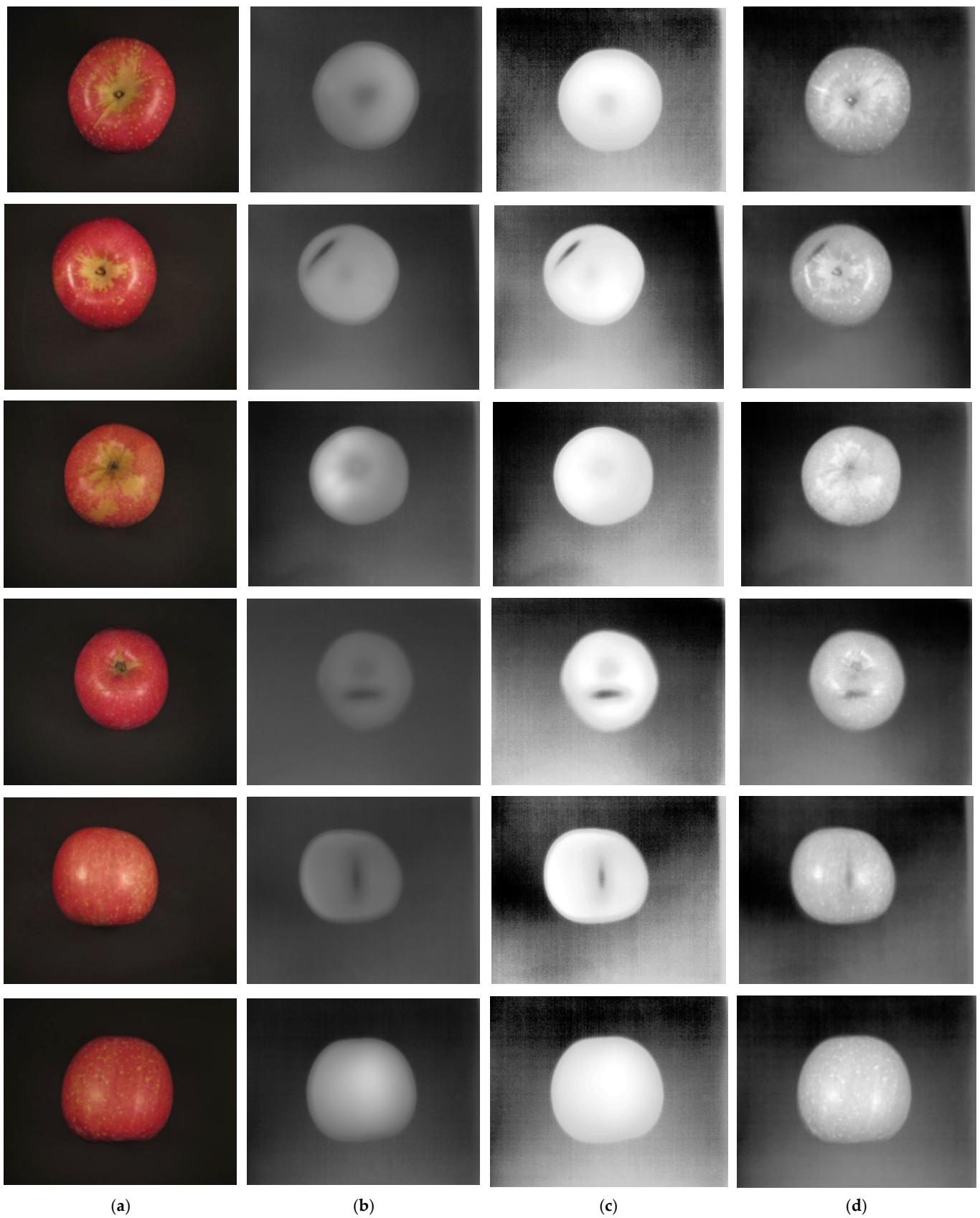


Figure 3. Infrared and visible apple surface defect image dataset types (from top to bottom, stem, stem + defect, calyx, calyx + defect, defect apple, and intact apple). (a) Visible image; (b) Thermal infrared images; (c) Enhanced thermal infrared images; (d) Fusion images.

This experiment uses the transform function with the PyTorch framework to implement data enhancement operations. Each batch is resized to 256×256 and the training samples are expanded with random cropping, random horizontal flipping, random vertical flipping, random rotation, adjustment of contrast, saturation and hue, and central cropping. The above data enhancement preprocessing operations are used for the training and test sets in this experiment.

2.3. Experimental Environment and Parameter Setting

The experimental environment in this paper is shown in Table 4. All the models are trained and tested in NVIDIA GeForce RTX 2060. The cross-entropy loss function (CE) is employed as the loss function, the Adam optimizer is used to optimize the model, the learning rate strategy is a custom tuning strategy (LambdaLR), the initial value of the learning rate is 0.0001, the batch size is 32, and the network is trained for 150 epochs, which is the detail hyperparameter setting as shown in Table 5.

Table 4. The experimental environment in this paper.

Experimental Tool	Specific Model
CPU	Intel(R) Core(TM) i7-10750H
GPU	NVIDIA GeForce RTX 2060
Operating System	Windows 10
Programming Language	Python 3.7.11
Deep Learning Framework	Pytorch 1.9.1

Table 5. Hyperparameter setting in this paper.

Hyperparameter	Detail Setting
Loss Function	The cross-entropy loss function
Optimizer	Adam
Learning Rate Strategy	LambdaLR
Initial Learning Rate Value	0.0001
Batch Size	32
The Number of Epoch	150

The learning rate represents the speed of updating the model weights, which is a relatively important parameter. In this study, different learning rates are experimentally compared to select the best performing learning rate. In this paper, learning rates of 0.1, 0.01, 0.001, and 0.0001 are tested and the model performs best when the learning rate is set to 0.0001. The accuracy and loss curves of the training process are shown in Figure 4. As the loss values in the pretraining period at learning rates of 0.1 and 0.01 are too high and too different from those at learning rates of 0.001 and 0.0001 to be compared in the figure, only the loss curves at learning rates of 0.001 and 0.0001 are compared.

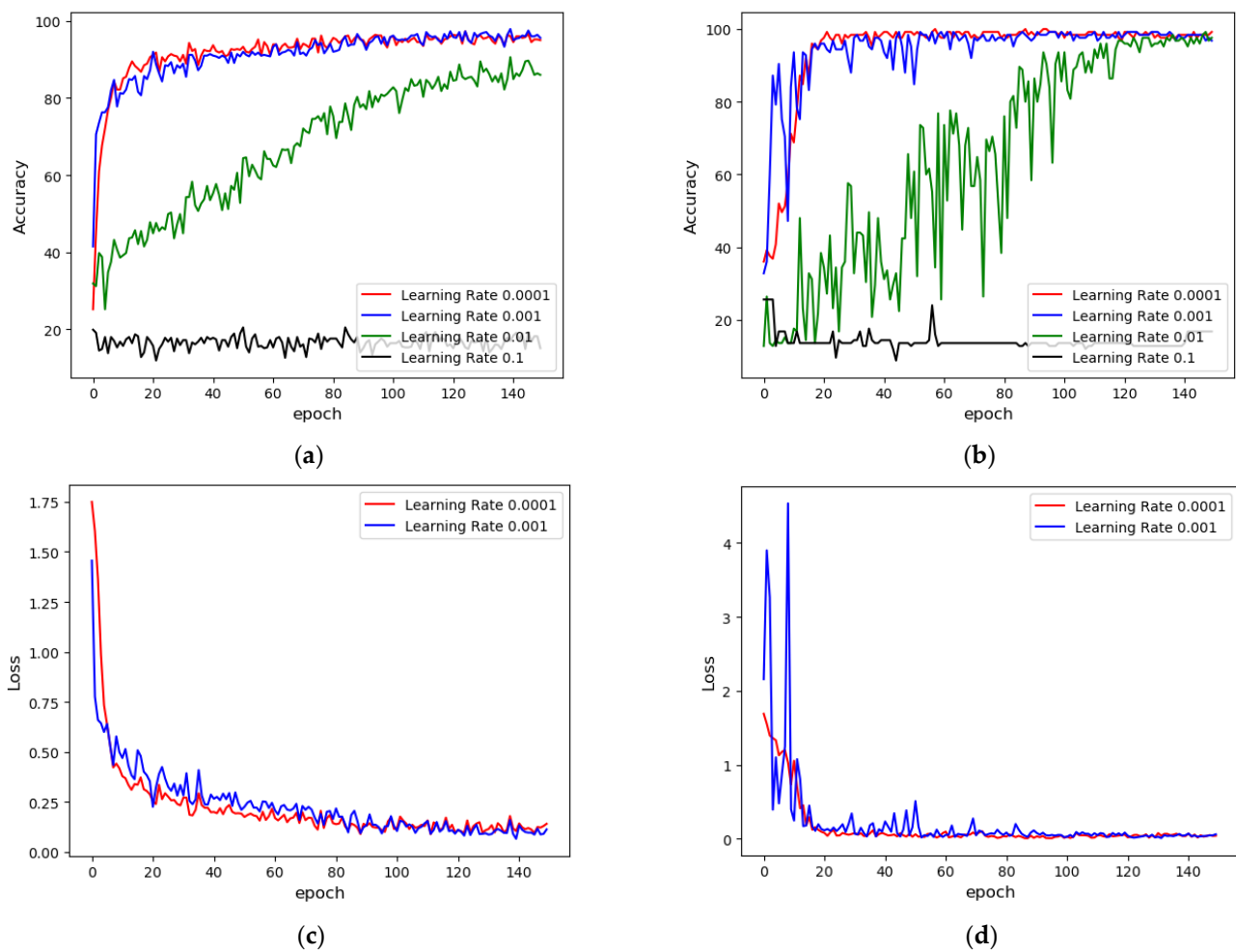


Figure 4. Comparison of accuracy and loss curves of different learning rates during training and testing. (a) Accuracy curve in the training. (b) Accuracy curve in the testing. (c) Loss curve in the training. (d) Loss curve in the testing.

2.4. MobileNetV3 Model

The Google team proposed the MobileNet-V1 model in 2017 [34] for deploying lightweight convolutional neural networks in mobile or embedded devices. Compared with traditional convolutional neural networks, depthwise separable convolution is introduced as an effective alternative to standard convolution, which significantly reduces the parameters and computation of the model with slightly lower accuracy. In 2018, the Google team proposed the MobileNet-V2 model [35] network, which introduces linear bottlenecks and inverted residual structures compared to the MobileNet-V1 network to improve the efficiency of the layer structure by exploiting the low-rank nature of the problem.

The Google team successively proposed the MobileNet-V3 model in 2019 [36]. MobileNet-V3 combines the deep separable convolution in MobileNet-V1 and the inverted residual structure (inverted residuals) of the linear bottleneck in MobileNet-V2, which is updimensionalized and then downdimensionalized after the deep separable convolution is followed by dimensionality reduction. The deep separable convolution serves to reduce the number of convolution kernels and accelerate the model. In addition, MobileNet-V3 also introduces the SE (squeeze and excitation) lightweight attention model [37]. Through the training process, the SE module gives each channel a weight so that the model globally serves to emphasize features with higher weights and suppress insignificant features. In addition, MobileNetV3 incorporates a new activation function, Hard-swish, which makes it much less computationally intensive and friendly to the quantization process. The basic network structure of the model is shown in Figure 5.

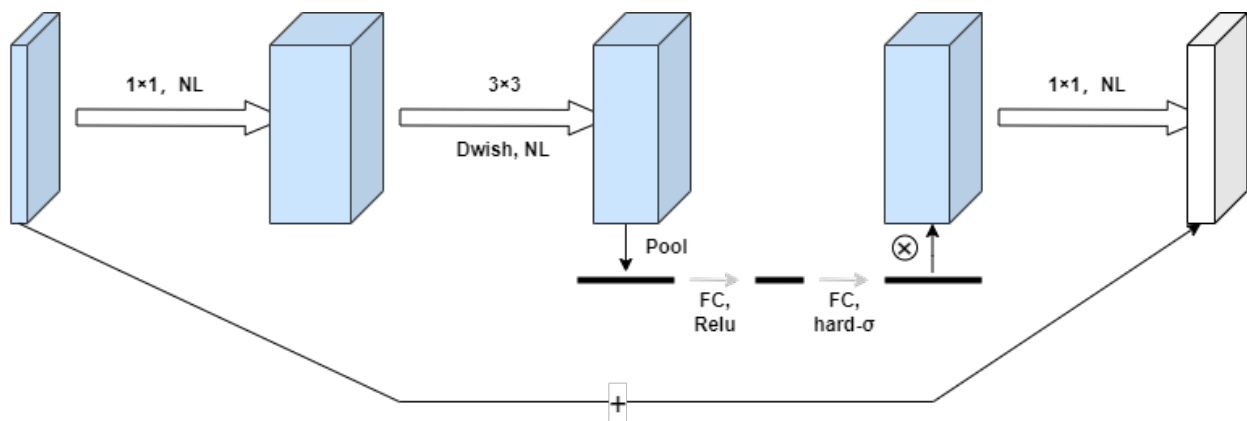


Figure 5. MobileNetV3 model network structure.

The MobileNetV3 model is suitable for mobile applications as a lightweight deep learning model. However, there is currently less research on the use of this model for apple surface defect detection, as well as the impact of realistic apple features that have suffered minor mechanical damage that is not obvious. Therefore, the model has a limited effect on the recognition of apples suffering from minor mechanical damage. To meet the task requirements for apple surface defect detection in realistic scenarios, the weight comparison transfer method proposed in this study is introduced. The MobileNetV3-Large model is chosen as the base model because this study is conducted on a small sample dataset. The specific network structure of the MobileNetV3-Large model is shown in Table 6, where NBN indicates that the BN (batch normalization) structure is not used, RE indicates the ReLU activation function, and HS indicates the activation function Hard-swish.

Table 6. Network structure of MobileNetV3-Large model.

Operator	Input Channel	Size	SE Module	NL	Step
ConvBNA, 3 × 3	3	224 × 224	-	HS	2
InvertedResidual, 3 × 3	16	112 × 112	-	RE	1
InvertedResidual, 3 × 3	16	112 × 112	-	RE	2
InvertedResidual, 3 × 3	24	56 × 56	-	RE	1
InvertedResidual, 5 × 5	24	56 × 56	✓	RE	2
InvertedResidual, 5 × 5	40	28 × 28	✓	RE	1
InvertedResidual, 5 × 5	40	28 × 28	✓	RE	1
InvertedResidual, 3 × 3	40	28 × 28	-	HS	2
InvertedResidual, 3 × 3	80	14 × 14	-	HS	1
InvertedResidual, 3 × 3	80	14 × 14	-	HS	1
InvertedResidual, 3 × 3	80	14 × 14	-	HS	1
InvertedResidual, 3 × 3	80	14 × 14	✓	HS	1
InvertedResidual, 3 × 3	112	14 × 14	✓	HS	1
InvertedResidual, 5 × 5	112	14 × 14	✓	HS	2
InvertedResidual, 5 × 5	160	7 × 7	✓	HS	1
InvertedResidual, 5 × 5	160	7 × 7	✓	HS	1
Conv2d, 1 × 1	160	7 × 7	-	HS	1
Avg Pooling, 7 × 7	960	7 × 7	-	-	1
Conv2d, 1 × 1, NBN	960	1 × 1	-	HS	1
Conv2d, 1 × 1, NBN	1280	1 × 1	-	-	1

2.5. SE Module

For deep learning models, not all extracted features are important. As a classical channel attention mechanism, SE-Net has three main components—squeeze, excitation and reweight—which are utilized to recalibrate features by learning global information and then by constructing interactions between two feature channels, stimulating features that

are important for the classification task and suppressing features that are less important for the classification task. The SENet model architecture is shown in Figure 6.

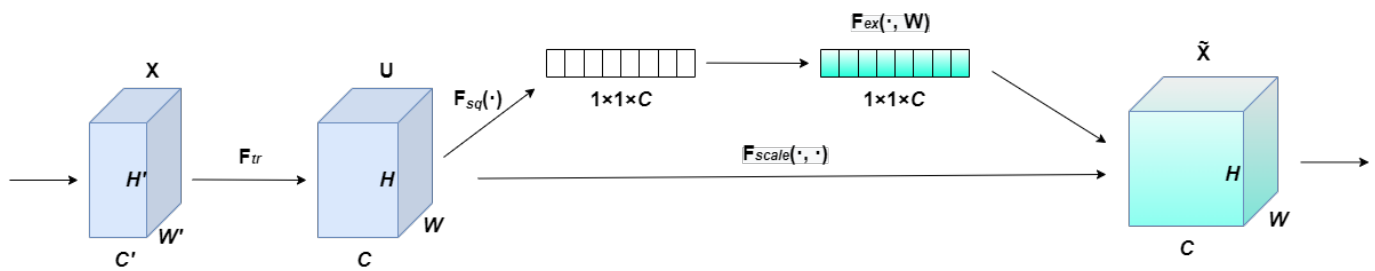


Figure 6. SE module architecture.

In the SE module of MobileNetV3, the $1 \times 1 \times C$ feature map is obtained by global averaging pooling of the dimensions of the input feature map, and then the $1 \times 1 \times C$ vector is passed through the first fully connected layer FC, whose activation function is ReLU. A vector of size $1 \times 1 \times C/r$ is obtained. The vector of size $1 \times 1 \times C$ is then passed through a second fully connected layer FC with an activation function of h-swish. A vector of size $1 \times 1 \times C$ is obtained, which is multiplied by the original corresponding channel feature matrix to complete the rescaling of the original features in the channel dimension.

2.6. Weight Comparison Transfer Learning

Convolutional neural networks are advantageous for image recognition, but they have high dataset requirements, usually in terms of the quantity and quality of image data. However, it is not always possible to obtain high quality and large amounts of data to train a model for all tasks in practice. More often than not, the datasets for image recognition tasks are small sample datasets.

Transfer learning is an important direction in deep learning. Transfer learning avoids the problem of insufficient training data in most cases, reduces the cost of using data, and enhances the usability of convolutional neural network models. The principle of transfer learning is to transfer the weights and parameters from the trained model to the model to be trained, thus allowing the new model to converge at a faster rate and to obtain the desired training results [38]. Applying the learning model from the source domain to the target domain and training the weights and parameters of the pretrained model obtained from the source domain on the new dataset allows the new model to quickly converge and reduces the model's demand for data, which to some extent solves the problem of model overfitting that may be caused by insufficient data. There are various transfer learning methods, such as no freezing, partial freezing, and freezing.

No freezing, also known as all-parameter transfer, is a model-based transfer learning method that involves sharing parameters between two models in the source and target domains to achieve an overall transfer of parameters. This method is the most common transfer learning method and is often utilized when the data characteristics of the source and target domains do not differ much. This outcome is achieved by not freezing all network layers and training the model with all parameters.

The freezing method is to freeze all the network layers in the target domain, with the exception of the fully connected layer, and to replace the fully connected layer in the original model with a fully connected layer with random weights. The model only trains the parameters of this fully connected layer. A fully connected layer is added to the model to be trained as a classifier and the pretrained weights are used as feature extractors for another task. Only the classifier parameters added at the end of the model to be trained are learned, while the other network layer parameters are kept frozen.

The partial freezing method refers to transfer training by a process of freezing some of the network layers and retraining some of them. In general, to obtain the optimal number of frozen layers, the model is frozen in a stepwise descending manner from the highest number to the lowest number of frozen layers. The significance of this approach is that it allows the

model to focus on learning feature information that is specific to the dataset during training, thus improving the overall network model’s ability to extract feature information.

The weight comparison transfer learning method proposed in this paper transfers different weight parameters to the network layer of the target model by comparing the pre-training weights obtained by the model on a large-scale dataset with the default weights of the model itself, thus enabling the model to adaptively extract effective feature parameters from the pretraining weights throughout the training process. The method takes maximum advantage of the data volume of large-scale datasets. The flow of the weight comparison transfer learning method is shown in Figure 7.

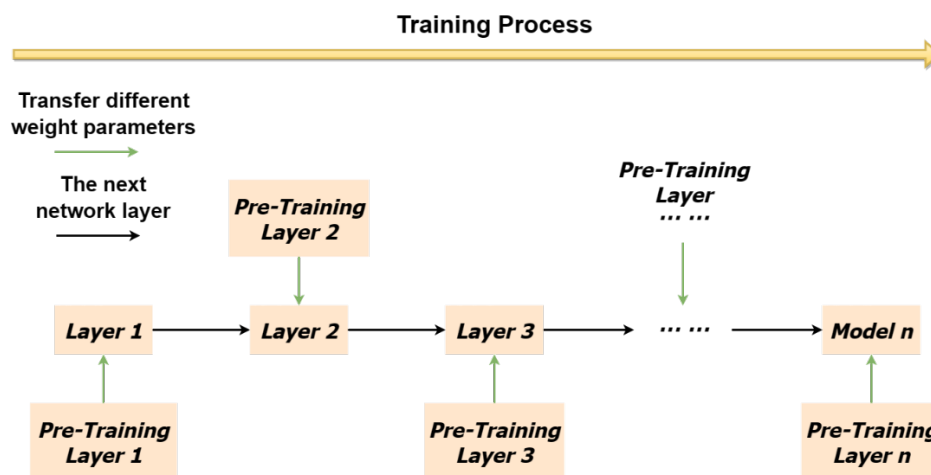


Figure 7. The process of weight comparison transfer learning method.

2.7. WC-MobileNetV3

To enhance the applicability of MobileNetV3 for apple surface defect detection tasks and improve the accuracy of apple surface defect detection in this paper, a training strategy of weight comparison transfer is introduced to MobileNetV3 to obtain the final WC-MobileNetV3 model. The flow of the proposed WC-MobileNetV3 model in this paper is shown in Figure 7. The core idea of the model is to transfer adaptively the effective features of each layer of the model, thus reducing the training complexity and improving the model training effect. First, MobileNetV3 trained on the ImageNet dataset is obtained with pretraining weight parameters. Second, the new MobileNetV3 model is trained by loading the thermal infrared and visible apple surface defect fusion image dataset. During the training process, the pretraining weights of each network layer obtained from the model on the ImageNet dataset are compared with the default weights of each network layer extracted from the model itself, and the different weight parameters are transferred to the corresponding network layers of the training model. This approach enables an efficient transfer of the pretrained weight parameters. Last, the apple surface defect detection model is obtained on the fused thermal infrared and visible apple surface defect image dataset.

The training procedures for WC-MobileNetV3 are shown below in Algorithm 1.

Algorithm 1: Training Strategy-Weight Comparison Transfer Learning with MobileNetV3**Input:** Input the fused image dataset of infrared and visible images of apple surface defect**Output:** WC-MobileNetV3 Model

- 1 The pretraining weight parameters are obtained from the MobileNetV3 trained on the ImageNet dataset;
- 2 Loading the fused image dataset of infrared and visible images of apple surface defect for the training of a new MobileNetV3 Model;
- 3 for W_i and W_i^{pre} do;
- 4 If $W_1^{pre} == W_1$;
- 5 Continue;
- 6 else;
- 7 $W_1 = W_1^{pre}$;
- 8 $i = i + 1$;
- 9 End for.

The general architecture of the WC-MobileNetV3 model is shown in Figure 8.

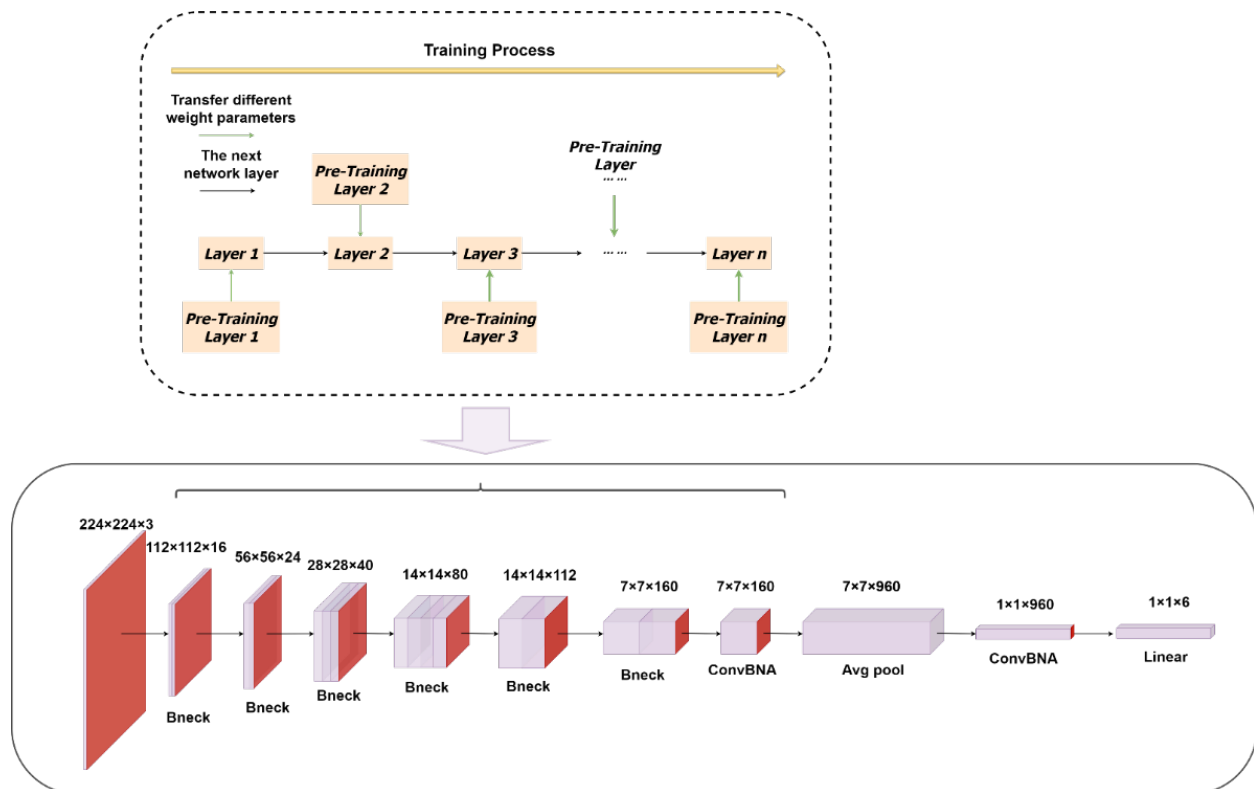


Figure 8. WC-MobileNetV3 overall architecture.

3. Experimental Results

3.1. Experimental Evaluation Indices

In this experiment, the accuracy rate is used as the main evaluation index for the results of apple surface defect detection. To further analyze the performance of the model, accuracy, recall, time spent on a single image (T_s), F1-Score, and number of parameters (M) are selected as evaluation metrics. These evaluation metrics are used to evaluate the performance of the apple surface defect detection model.

3.1.1. Accuracy

The accuracy rate represents the ratio of the number of samples correctly identified to the total number of samples in the classification recognition task. The formula for calculating the accuracy rate is shown in Equation (1),

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

where TP denotes the number of true-position samples, FP denotes the number of false-position samples, FN denotes the number of false-negative samples, and TN denotes the number of true-negative samples. When the number of different sample categories in the dataset is not homogeneous, accuracy cannot be the only criterion for evaluation and other metrics are needed to support the assessment.

3.1.2. Precision

Accuracy represents the ratio of samples that the model correctly predicts and is indeed correct to all samples that the model correctly predicts in the classification recognition task. The formula is shown in Equation (2),

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

3.1.3. Recall

Recall represents the ratio of samples that the model predicts to be correct but are indeed correct to all correct samples in the classification recognition task. The formula is shown in Equation (3),

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

3.1.4. F₁-Score

The F1-score is the summed average of precision and recall. Precision indicates the ability of the model to discriminate between two negative samples and the value of precision is proportional to the ability of the model to discriminate between two negative samples. Recall is the ability of the model to discriminate between two positive samples and the value of recall is proportional to the ability of the model to discriminate between two positive samples. The F1-score is a combination of precision and recall, and its value is proportional to the robustness of the model. The formula for calculating the F1-score is shown in Equation (4),

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

3.1.5. Time Spent on a Single Image (T_s)

T_s is an important evaluation metric to assess the performance of a model, which represents the ratio between the time taken by the model to make predictions on the test set and the number of image data in the test set. T_s is calculated as shown in Equation (5)

$$T_s = \frac{\text{Total test time}}{\text{Total number of test images}} \quad (5)$$

3.1.6. Parameter (M)

The number of parameters in a model is a comprehensive evaluation of the model's performance. The number of parameters is generally the sum of the number of parameters of all network layers of the model, which can be shown by Equation (6)

$$\text{Parameter} = \sum_{i=1}^L n_{i+1} \quad (6)$$

where L denotes the number of all network layers of the model, n_i denotes the number of parameters of the previous network layer, and n_{i+1} denotes the number of parameters of the current network layer. The number of parameters affects the speed of the model run and the size of the model. Thus, this paper includes the number of parameters as one of the indicators for evaluation of the model.

3.2. Impact of Image Augmentation

Figure 9 depicts the various stages of the training process with and without data augmentation. Figure 9 indicates that the model without data augmentation underwent an overfitting phenomenon. While the accuracy on the training set is 100%, the accuracy on the test set is approximately 90%. Additionally, the loss value of the training process is 0 and gradient disappearance occurs. The overall training process is not ideal. The model trained with data augmentation, however, did not suffer from overfitting and had an accuracy of 96.4% on the training set and 100% on the test set. The overall results are superior to those of the model without data augmentation. The experiments demonstrated that the data enhancement operation resulted in faster convergence and fit of the model and higher generalization ability.

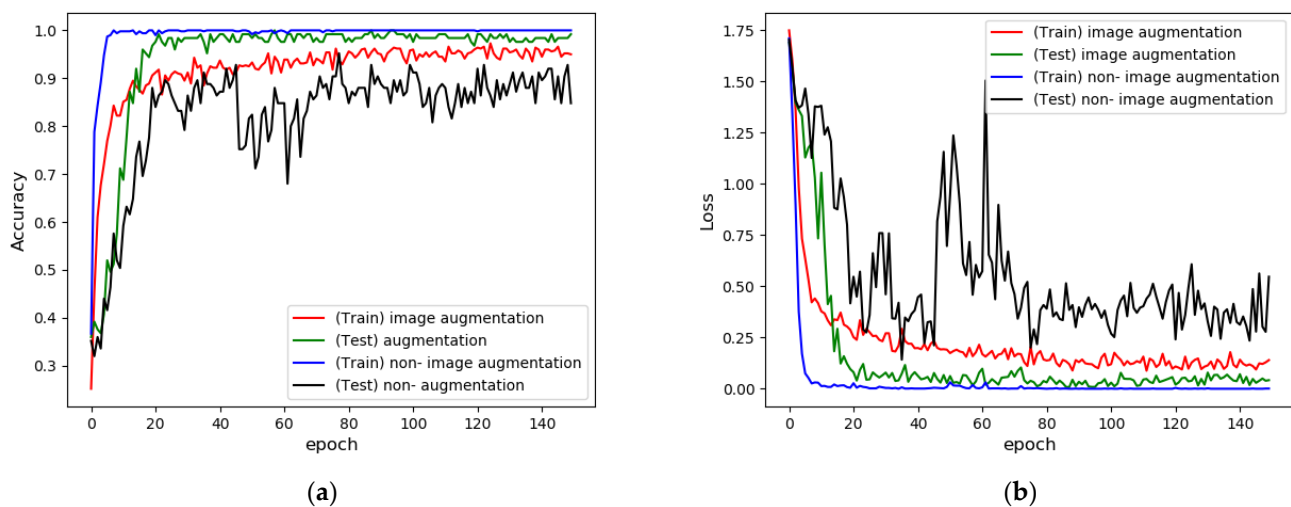


Figure 9. Accuracy and loss curve comparison with or without data augmentation operations. (a) The accuracy curve comparison with or without data augmentation operations; (b) The loss curve comparison with or without data augmentation operations.

3.3. Weighted Comparison Transfer Learning Ablation Experiment

To verify the effectiveness of the transfer learning method of weight comparison in the model training process, this study conducted ablation experiments on the test set. The results obtained are shown in Table 7. We discovered that the model with the introduction of weight-contrast transfer learning achieved an accuracy of 99.2%. The accuracy of the model with the introduction of transfer learning improved by 16% compared to the model without the introduction of weighted contrast transfer learning. In addition, the introduction of the transfer learning method allows the model to converge at a faster rate and achieve the desired results in a shorter training time. The model with the introduction of transfer

learning methods had the best performance with 99.2% accuracy, 99.12% precision, 99.02% recall, and 99.04% F1-score. These results show that the weight-contrast transfer learning method effectively enhances the detection of surface defects in apples.

Table 7. Ablation experiments to prove the effectiveness of weight comparison transfer learning.

Weight Comparison Transfer Learning	Accuracy/%	Precision/%	Recall/%	F1-Score (%)	Parameter (M)	T _s (ms)
×	83.2	84.44	84.98	83.65	4.21	24.42
√	99.2	99.12	99.02	99.04	4.21	24.08

3.4. Comparison Experiments of Transfer Learning Methods

The MobileNetV3 model is trained by freezing the network layers before the fully connected layer of the MobileNetV3 model and unfreezing the network layers of the MobileNetV3 model layer by layer, starting from deep to shallow, to train the model with different numbers of frozen layers. The test results are shown in Table 8. By analyzing the experimental results of Table 6, the number of parameters of the model starts to increase as the number of frozen layers decreases. At the same time, the evaluation indexes of the model (accuracy, precision, recall, F1-score, and parameter) start to increase and the overall performance of the model becomes better and better. When the model is unfrozen to only the first four network layers, the model achieves the best performance with 96.8% of accuracy, 96.45% of precision, 96.6% of recall, 96.48% of F1-score, 6.71 M of parameters, and 22.77 ms of T_s. Since the number of frozen layers is different, and thus the model learns feature information differently, the model with the best overall performance (the model with the first four network layers frozen) is selected as the fine-tuned model for this study.

Table 8. Comparison of experimental results on the test set for models with different freezing layers.

Freeze the Number of Layers of the Model	Accuracy/%	Precision/%	Recall/%	F1-Score (%)	Parameter (M)	T _s (ms)
1	96.80	96.31	96.66	96.47	6.72	23.21
2	96.80	96.45	96.60	96.48	6.72	24.12
3	96.80	96.45	96.60	96.48	6.72	24.73
4	96.80	96.45	96.60	96.48	6.71	22.77
5	96.00	95.52	95.62	95.53	6.70	23.79
6	95.20	94.67	94.70	94.53	6.68	22.72
7	96.00	95.52	95.62	95.53	6.66	23.65
8	96.00	95.52	95.62	95.53	6.63	22.91
9	95.20	94.67	94.64	94.57	6.59	23.12
10	95.20	94.67	94.64	94.57	6.56	22.60
11	95.20	94.67	94.70	94.53	6.53	23.61
12	94.40	93.78	94.64	94.12	6.31	23.40
13	93.60	93.37	94.12	93.55	5.93	23.16
14	89.60	89.57	90.06	89.67	5.50	22.94
15	89.60	89.04	90.66	89.44	4.70	22.66

Table 9 shows a comparison of the results of different transfer learning methods, including no freezing (no freezing of network layers), partial freezing (freezing of some network layers), freezing (freezing of all network layers), and a comparison of weights. The analysis of Table 7 shows that when some of the network layers are frozen, the model has the best overall performance with an accuracy of 96.2%; when no network layers are frozen, the model has a better overall performance with an accuracy of 95.2%; and when all network layers are frozen, the model has the worst overall performance with an accuracy of only 84.0%. Freezing (freezing all network layers) is the least effective among the first three transfer learning approaches because the pretraining weights are obtained on the ImageNet dataset. The original recognition task is for 1000 objects, which is a poor

difference compared to the dataset employed in this study. Therefore, directly loading the weight parameters of the model without updating the parameters of the model's network layers does not give good results. No freezing (without freezing the network layers) performs better than freezing (freezing all network layers) because this training method has the most training parameters. However, some of the network layers are poorly learned, which affects the final recognition effect; partial freezing (freezing some of the network layers) works best because the model learns useful feature information better by removing some of the network layers that are poorly learned and avoids the interference of useless feature information.

Table 9. Comparison of experimental results of different transfer learning methods on the test set.

Transfer Learning Methods	Accuracy/%	Precision/%	Recall/%	F1-Score (%)	Parameter (M)	T _s (ms)
No freezing of network layers	95.2	94.67	94.64	94.57	6.72	24.52
Freezing part of the network layers	96.80	96.45	96.60	96.48	6.71	22.77
Freezing of all network layers	84.0	84.02	85.42	83.87	3.75	24.11
Weight comparison transfer learning	99.2	99.12	99.02	99.04	4.21	24.08

However, the above three transfer learning approaches do not adaptively transfer useful training parameters from the pretraining weights. The weight comparison transfer approach achieves adaptive transfer of weight parameters by filtering the training parameters from the pretraining weights. The model thus works best and can be considered the best performing model.

3.5. Thermal, Visible and Fused Image Dataset Comparison Experiment

To verify the effectiveness of the thermal infrared and visible light image fusion algorithm in the process of apple surface defect detection, the experimental results of the proposed detection method in this section on the fused image dataset are compared with the experimental results on the visible apple image dataset and the thermal infrared apple image dataset. The experimental results are shown in Table 10. Compared to the experimental results for detecting visible apple images and thermal infrared images alone, on fused images, the accuracy improved by 0.8% and 3.02%, the precision improved by 0.64% and 2.73%, the recall improved by 0.62% and 3.58%, and the F1-score improved by 0.64% and 3.23%, respectively. There is also a better performance in the single recognition time. The comparative experiments indicate that the proposed apple surface defect detection model in this paper is superior to visible or thermal infrared images alone in terms of fused images.

Table 10. Comparison of experimental results on different spectral datasets.

Data Type	Accuracy/%	Precision/%	Recall/%	F1-Score (%)	Parameter (M)	T _s (ms)
VIS	98.4	98.48	98.4	98.4	4.21	22.33
IR	96.18	96.39	95.44	95.81	4.21	22.81
Fused	99.2	99.12	99.02	99.04	4.21	24.08

3.6. Comparison Experiments of Different Models

To verify the recognition ability of the proposed detection model (WC-MobileNetV3) in this paper. AlexNet, ResNet50, DenseNet169, and EfficientNetV2 are selected for comparison experiments and compared under the same experimental conditions. The experimental results are shown in Table 11.

Table 11. Comparison of experimental results of different convolutional neural network models on the test set.

Model	Accuracy/%	Precision/%	Recall/%	F1-Score (%)	Parameter (M)	T _s (ms)
AlexNet	73.60	76.58	78.82	74.93	58.31	22.83
DenseNet169	96.00	95.94	95.50	95.59	12.49	57.48
ResNet50	91.20	91.11	90.99	90.87	23.52	34.97
EfficientNetV2	90.40	90.73	89.90	90.00	52.87	36.70
WC-MobileNetV3	99.20	99.12	99.02	99.04	4.21	24.08

According to the analysis carried out in Table 9, the detection method proposed in this paper has advantages in all evaluation metrics compared to the other four convolutional neural networks. The recognition accuracy is 3.2% higher than the best results of the other models, the number of parameters is 8.28 M less, and the recognition time of a single image is only 1.25 ms more than that of AlexNet. Simultaneously, good experimental results are achieved in terms of accuracy, recall, and F1-Score. The experimental results show that the WC-MobileNetV3 model proposed in this study can identify the most correct samples compared to other models. In terms of overall performance, WC-MobileNetV3 can be considered the best-performing model. Compared with the other three classical convolutional neural network models, the DenseNet169 model has the highest accuracy, precision, recall, and F1-score but is the highest in single image recognition time consumption. AlexNet has the highest number of parameters among these four models and only has the lowest single image recognition time consumption but the lowest in all other evaluation metrics. The above analysis reveals that the WC-MobileNetV3 model is superior. The model can achieve a balance among accuracy, number of parameters, and time consumption for single image recognition and can meet the requirements for apple surface defect detection.

In addition, in order to further compare the WC-MobileNetV3 model with four classical convolutional neural network models and three models based on different transfer learning methods, the ROC (Receiver Operating Characteristic) curve is additionally chosen as one of the evaluation metrics for the model. The ROC curve is chosen because there is a category imbalance in the actual dataset, which will lead to more negative samples than positive samples (and vice versa), while ROC is not affected by the test data and can directly evaluate the classifier when the area under curve (AUC) is larger; it indicates the better performance of the classifier and the AUC is calculated by Equation (7),

$$AUC = 1 - \frac{\sum_{i \in \text{positiveClass}} \text{rank}_i - \frac{M(1-M)}{2}}{M \times N} \quad (7)$$

where M denotes the number of positive samples and N denotes the number of negative samples. Rank can indicate the number of such combinations that can yield positive samples.

The ROC curves of the eight models are shown in Figure 10. According to the comparison of the eight ROC curves, WC-MobileNetV3 has a better performance compared with the other seven models, but the overall performance of all eight models is limited, which is caused by insufficient training data, and the subsequent research will expand the dataset to train the WC-MobileNetV3 model with more superior performance.

The performance of the five convolutional neural network models on the test set is visualized in this study in the form of confusion matrices. The MobileNetV3 models are visualized under different transfer learning methods to further analyze the performance of the MobileNetV3 models with weight comparison transfer. The confusion matrix generally refers to the analysis matrix used in machine learning to summarize the classification and prediction results of a model. Each row of the confusion matrix in this section represents the predicted data for the category and each column represents the actual data for the category. Each value in the confusion matrix is the probability of the data in the column category being predicted as a row category, with the value at the diagonal position indicating the

probability of a correct prediction. The confusion matrices for the five models are shown in Figure 10.

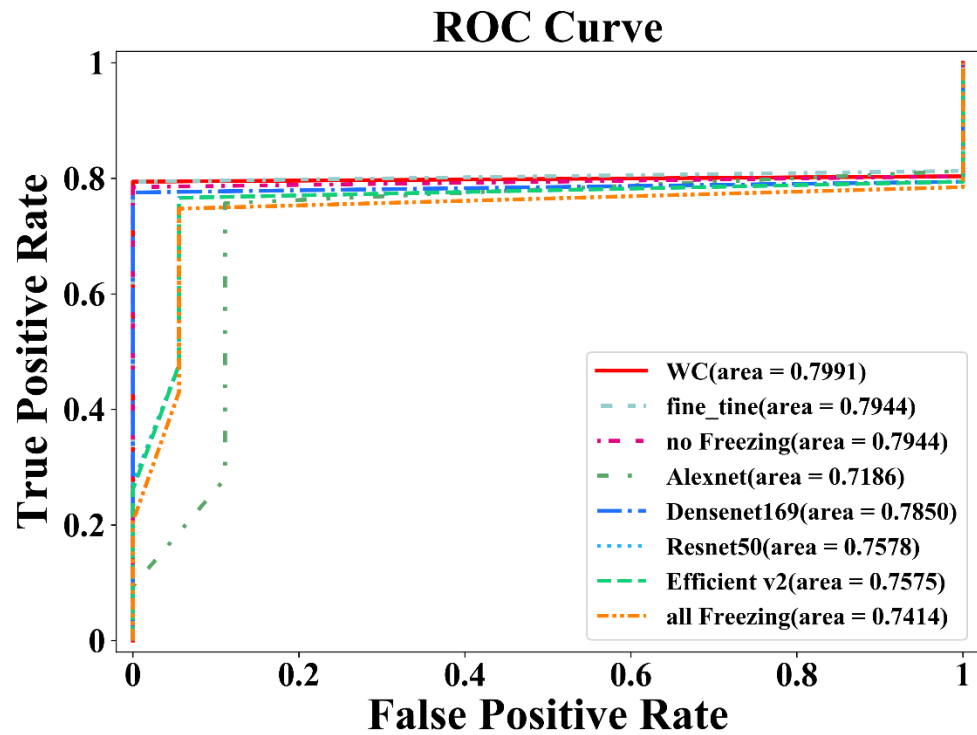


Figure 10. The ROC curve of 8 models.

As shown in Figure 11, the overall recognition effect of the MobileNetV3 model with weight contrast transfer is superior to that of the other models. The model achieves a high recognition accuracy of 100% for stem, calyx, intact apple, defect apple, and calyx + defect but an accuracy of 94.7% for stem + defect. These results are mainly attributed to the high similarity between calyx and stem, while the defective part is not sufficiently well characterized. Similar problems are identified in other models. Overall, the MobileNetV3 model with weight contrast transfer learned more valuable feature information, which ultimately improved the detection results.

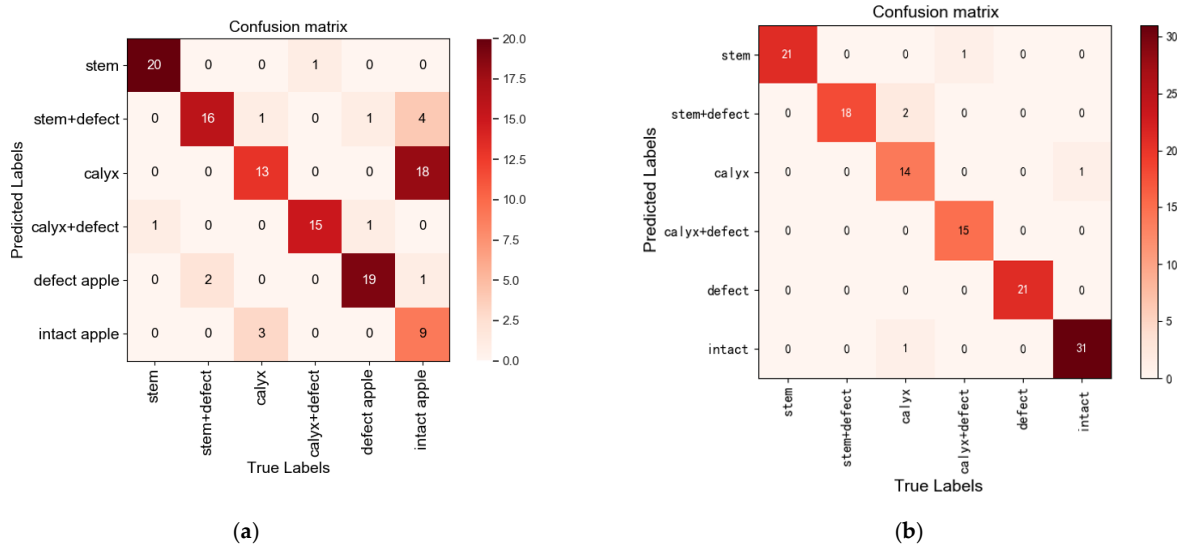


Figure 11. Cont.

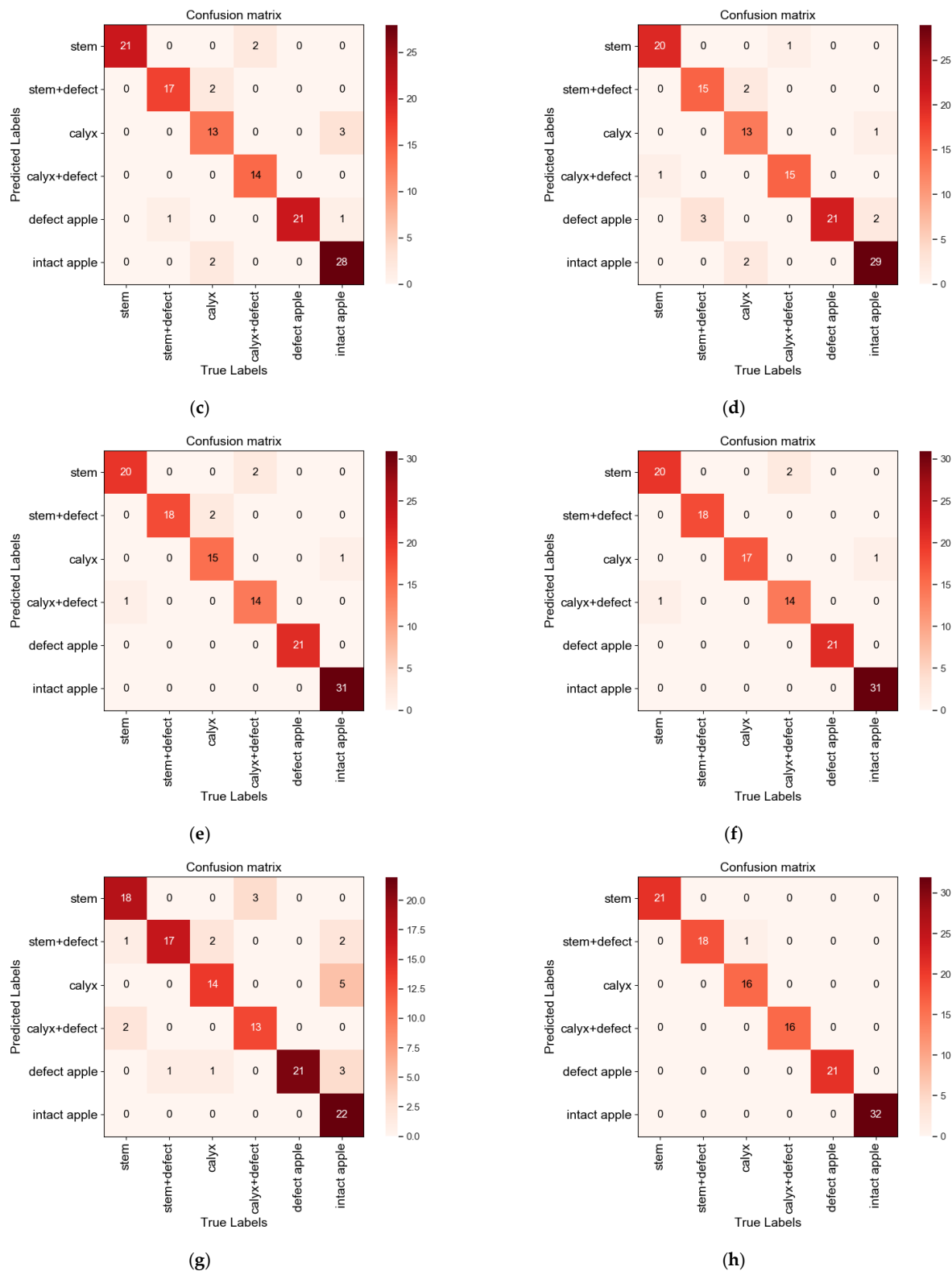


Figure 11. Confusion matrix visualization for 8 models. (a) AlexNet Confusion Matrix; (b) DenseNet Confusion Matrix; (c) ResNet50 Confusion Matrix; (d) EfficientV2 Confusion Matrix; (e) MobileNetV3 Confusion Matrix under 100% Transfer (no freezing of network layers); (f) MobileNetV3 Confusion Matrix under Fine-tuning (freezing part of the network layer); (g) MobileNetV3 Confusion Matrix under Feature Extraction (freeze all network layers); (h) MobileNetV3 Confusion Matrix under Weight Comparison.

3.7. Comparative Experiments of WC-MobileNetV3 Model on Different Varieties of Apples

Due to the differences in the appearance characteristics (color and texture) of apples of different varieties and production areas, the WC-MobileNetV3 model is used to detect defects in other apple varieties, such as “Akesu Fuji”, “Gansu Fuji”, and “Shaotong Fuji”. “Akesu Fuji”, “Gansu Fuji”, and “Shaotong Fuji” apple varieties are detected. The experimental results are shown in Tables 12 and 13.

Table 12. The accuracy of WC-MobileNetV3 model for detecting different apple varieties.

	Ground Truth							True Positive						
	Total	Intact	Stem	Stem + Defect	Defect	Calyx	Calyx + Defect	Total	Intact	Stem	Stem + Defect	Defect	Calyx	Calyx + Defect
Akesu Fuji	120	20	20	20	20	20	20	106	20	17	16	20	13	20
Gansu Fuji	120	20	20	20	20	20	20	104	20	20	17	20	7	20
Shaotong Fuji	117	20	17	20	20	20	20	101	19	14	20	20	8	20

Table 13. The experimental results of WC-MobileNetV3 model for different apple varieties.

	Accuracy/%	Precision/%	Recall/%	F1-Score (%)	Parameter (M)	T _s (ms)
Akesu Fuji	88.33	89.95	88.33	88.32	4.21	25.69
Gansu Fuji	86.67	92.06	86.67	85.67	4.21	24.54
Shaotong Fuji	86.32	90.78	86.23	85.40	4.21	24.22

The average accuracy of the WC-MobileNetV3 model is 87.11% and the results for “Akesu Fuji”, “Gansu Fuji” and “Shaotong Fuji” apples are acceptable. However, the experimental results do not achieve very good results. This is due to the insufficient fusion data for model training. The training dataset of the WC-MobileNetV3 model only uses “Yantai Fuji”, but “Yantai Fuji” already contains the common appearance characteristics of apples. With the effect of data enhancement and weight comparison transfer method, the model can still recognize similar apples efficiently to a certain extent, especially for intact and defect, but the recognition of defects in calyx is poor, which is due to the high similarity between calyx + defect and calyx, as well as the small training data of calyx and calyx + defect classes. The model does not learn enough feature information.

To improve the performance of the WC-MobileNetV3 model for detecting surface defects of other varieties of apples, the most direct improvement method is to put the apple images of these varieties into the training set, thus improving the generalization of the training set. Secondly, the amount of data within the training set can also be expanded, which can then improve the detection performance and generalization ability of the WC-MobileNetV3 model, so subsequent studies will construct larger sample datasets to be used for model training.

4. Discussion

This paper uses a dataset of fused thermal infrared and visible images to provide a new idea for the detection of surface defects in apples. Among the current methods for apple detection of surface defects, no experiments have been conducted for fused thermal infrared and visible images. The use of image fusion to extract features requires better enhancement of the IR target and preservation of the details of the visible image. Ma et al. [39] showed that fused thermal infrared and visible images can significantly improve the efficiency of the algorithm. This paper conducted comparative experiments on IR, VIS, and IR + VIS. The experimental results show that the strategy is effective and informative. However, a good fused image needs to contain the rich texture features of the visible image and the defective targets in the thermal infrared image [40]. Due to the inadequate performance of existing dual-light cameras, the quality of the acquired IR and visible images is limited, which to a certain extent, limits the accuracy of detecting defects on the apples’ surface. The accuracy in the recognition of apple stem + defect is only 94.7%. Therefore, the quality of the acquired images needs to be improved. Notably, proper preprocessing is very important.

Redundant information can exist in data acquisition. In this case, preprocessing can be effective in improving the quality of the fused images [41].

In addition, this paper proposes a weight-contrast transfer learning method. Compared with traditional transfer methods, weight-contrast transfer learning methods are adaptive in extracting feature parameters. Common transfer learning methods are no freezing (without freezing any network layer), partial freezing (freezing some network layers) and freezing (freezing all network layers) [42]. Our weight comparison transfer method is compared with these methods. By conducting ablation experiments on the test set, it is determined that the above three transfer learning approaches could not adaptively transfer useful training parameters from the pretrained weights. In contrast, the weight comparison transfer method proposed in this paper achieves adaptive transfer of weight parameters by screening the training parameters from pretrained weights. The model has better accuracy and convergence speed. Therefore, the learning method based on weight contrast transfer is feasible for enhancing the detection of surface defects on apples.

Ji et al. [43] achieved a multiclass average accuracy of 94.43% for apple recognition based on the improved MobileNetV3 model. The running time for recognition is 0.051 s per image. MobileNetV3 reduces the number of parameters and computation while ensuring accuracy. In this paper, we choose to use a weight contrast transfer training strategy for the MobileNetV3 model. The pretraining weights obtained by MobileNetV3 on the ImageNet dataset are compared with the default weights of the model itself during the training process, allowing the model pretraining weights to extract the required feature parameters for each network layer. The final model that we propose is WC-MobileNetV3. To verify the superiority of this model, it is compared with the convolutional neural networks AlexNet, ResNet50, DenseNet169, and EfficientNetV2 under the same experimental conditions. The model achieves a balance between accuracy and detection time, which is well suited to the needs for apple surface defect detection.

The research in this paper focuses mainly on the surface defects of apples caused by mechanical damage mainly scrapes and scratches. However, due to the special characteristics of the fused images of infrared and visible images, there is still obvious feature information on the fused images for common natural defects on the fruit surface such as rots and insect spots. In the subsequent research, we consider expanding the number of defect types and realizing the identification of defect classes in the process of detection, and giving high-quality suggestions related to subsequent fruit cultivation and management. In addition, the quality of the acquired images is limited due to the low precision of the infrared camera in the dual-light camera used in this study, but the experiments in Section 3.5 also prove the effectiveness of the infrared and visible image fusion technique for apple surface defect detection. In future research, a new type of dual-light camera will be constructed using a visible industrial camera and thermal infrared industrial camera with suitable precision to realize the acquisition of high-quality infrared and visible images. The RFN-Nest algorithm achieves the fusion of infrared and visible images of apples with a nice fusion effect, but it still has many deficiencies for fruits with high-speed movement on the sorting assembly line. In future research, the idea of infrared and visible image fusion is considered to be introduced into the object detection model to achieve real-time image fusion and defect detection. In addition, more high-quality infrared and visible images of apples will be acquired for training the deep learning models in subsequent studies, in the expectation of obtaining models with better generalization and detection performance for the task of detecting surface defects for most apple varieties.

5. Conclusions

Apples with surface defects can cause economic losses in the apple industry. Therefore, implementing the detection of surface defects on apples is an effective strategy for reducing economic losses. In this paper, a MobileNetV3 model based on weighted contrast transfer is proposed for the detection of apple surface defects. The detection method is tested on self-constructed visible, thermal infrared and fused image datasets for comparison. The

experimental results of the MobileNetV3 model with weight contrast transfer are better. The recognition accuracy is high for stem, calyx, intact apple, defect apple, and calyx + defect, all reaching 100%, and 94.7% for stem + defect. Therefore, the proposed method is fully suitable for the accurate identification of apples that have suffered minor mechanical damage. The method also has the potential to be extended to other fruits.

In future research, the number of types of natural defects will be expanded and a higher performance infrared and visible image acquisition system will be constructed to obtain higher quality dual-light image data to ultimately improve the effectiveness of apple surface defect detection. At the same time, an end-to-end dual-channel object detection model will be constructed to achieve real-time fusion and defect detection of dual-light images, laying the technical foundation for the construction of a dual-channel online sorting system.

Author Contributions: Conceptualization, H.S. and Y.W.; methodology, H.S. and Y.W.; software, Y.W.; validation, Y.W., W.Z. and M.W.; formal analysis, Y.W. and W.Z.; investigation, Y.W., J.S. and B.F.; resources, H.S., Y.W. and L.W.; data curation, H.S. and Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, W.Z. and Y.W.; visualization, Y.W., Z.S. and Y.L.; supervision, H.S. and C.S.; project administration, H.S. and C.S.; funding acquisition, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the Henan Province Key Science-Technology Research Project under Grant No. 232102520006, the National Science and Technology Resource Sharing Service Platform Project under Grant No. NCGRC-2020-57.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable. Our image datasets are self-built.

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

References

- Lu, Y.; Lu, R. Non-Destructive Defect Detection of Apples by Spectroscopic and Imaging Technologies: A Review. *Trans. ASABE* **2017**, *60*, 1765–1790. [[CrossRef](#)]
- 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](#)]
- Wang, Z.; Jin, L.; Wang, S.; Xu, H. Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system. *Postharvest Biol. Technol.* **2022**, *185*, 111808. [[CrossRef](#)]
- Hussein, Z.; Fawole, O.A.; Opara, U.L. Preharvest factors influencing bruise damage of fresh fruits—A review. *Sci. Hortic.* **2018**, *229*, 45–58. [[CrossRef](#)]
- Li, J.B.; Luo, W.; Wang, Z.L.; Fan, S.X. Early detection of decay on apples using hyperspectral reflectance imaging combining both principal component analysis and improved watershed segmentation method. *Postharvest Biol. Technol.* **2019**, *149*, 235–246. [[CrossRef](#)]
- Nturambirwe, J.F.I.; Hussein, E.A.; Vaccari, M.; Thron, C.; Perold, W.J.; Opara, U.L. Feature Reduction for the Classification of Bruise Damage to Apple Fruit Using a Contactless FT-NIR Spectroscopy with Machine Learning. *Foods* **2023**, *12*, 210. [[CrossRef](#)]
- Li, J.; Karkee, M.; Zhang, Q.; Xiao, K.H.; Feng, T. Characterizing apple picking patterns for robotic harvesting. *Comput. Electron. Agric.* **2016**, *127*, 633–640. [[CrossRef](#)]
- Tan, W.Y.; Sun, L.J.; Yang, F.; Che, W.K.; Ye, D.D.; Zhang, D.; Zou, B.R. The feasibility of early detection and grading of apple bruises using hyperspectral imaging. *J. Chemom.* **2018**, *32*, e3067. [[CrossRef](#)]
- Baneh, N.M.; Navid, H.; Kafashan, J. Mechatronic components in apple sorting machines with computer vision. *J. Food Meas. Charact.* **2018**, *12*, 1135–1155. [[CrossRef](#)]
- Lu, Y.; Lu, R. Detection of Surface and Subsurface Defects of Apples Using Structured-Illumination Reflectance Imaging with Machine Learning Algorithms. *Trans. ASABE* **2018**, *61*, 1831–1842. [[CrossRef](#)]
- Nturambirwe, J.F.I.; Opara, U.L. Machine learning applications to non-destructive defect detection in horticultural products. *Biosyst. Eng.* **2020**, *189*, 60–83. [[CrossRef](#)]
- Dhiman, B.; Kumar, Y.; Kumar, M. Fruit quality evaluation using machine learning techniques: Review, motivation and future perspectives. *Multimed. Tools Appl.* **2022**, *81*, 16255–16277. [[CrossRef](#)]
- Moallem, P.; Serajoddin, A.; Pourghassem, H. Computer vision-based apple grading for golden delicious apples based on surface features. *Inf. Process. Agric.* **2017**, *4*, 33–40. [[CrossRef](#)]

14. Bhargava, A.; Bansal, A. Machine learning based quality evaluation of mono-colored apples. *Multimed. Tools Appl.* **2020**, *79*, 22989–23006. [[CrossRef](#)]
15. Zhang, B.H.; Huang, W.Q.; Gong, L.; Li, J.B.; Zhao, C.J.; Liu, C.L.; Huang, D.F. Computer vision detection of defective apples using automatic lightness correction and weighted RVM classifier. *J. Food Eng.* **2015**, *146*, 143–151. [[CrossRef](#)]
16. Chithra, P.L.; Henila, M. Apple fruit sorting using novel thresholding and area calculation algorithms. *Soft Comput.* **2021**, *25*, 431–445. [[CrossRef](#)]
17. Tan, A.J.; Zhou, G.X.; He, M.F. Surface defect identification of Citrus based on KF-2D-Renyi and ABC-SVM. *Multimed. Tools Appl.* **2021**, *80*, 9109–9136. [[CrossRef](#)]
18. Wang, B.; Yin, J.Q.; Liu, J.J.; Fang, H.G.; Li, J.S.; Sun, X.; Guo, Y.M.; Xia, L.M. Extraction and classification of apple defects under uneven illumination based on machine vision. *J. Food Process. Eng.* **2022**, *45*, e13976. [[CrossRef](#)]
19. Andrew, J.; Eunice, J.; Popescu, D.E.; Chowdary, M.K.; Hemanth, J. Deep Learning-Based Leaf Disease Detection in Crops Using Images for Agricultural Applications. *Agronomy* **2022**, *12*, 2395.
20. Zhou, C.X.; Wang, H.H.; Liu, Y.; Ni, X.Y.; Liu, Y. Green Plums Surface Defect Detection Based on Deep Learning Methods. *IEEE Access* **2022**, *10*, 100397–100407. [[CrossRef](#)]
21. Deng, L.M.; Li, J.; Han, Z.Z. Online defect detection and automatic grading of carrots using computer vision combined with deep learning methods. *LWT—Food Sci. Technol.* **2021**, *149*, 111832. [[CrossRef](#)]
22. Yao, J.; Qi, J.M.; Zhang, J.; Shao, H.M.; Yang, J.; Li, X. A Real-Time Detection Algorithm for Kiwifruit Defects Based on YOLOv5. *Electronics* **2021**, *10*, 1711. [[CrossRef](#)]
23. Da Costa, A.Z.; Figueroa, H.E.H.; Fracarolli, J.A. Computer vision based detection of external defects on tomatoes using deep learning. *Biosyst. Eng.* **2020**, *190*, 131–144. [[CrossRef](#)]
24. Unay, D.; Gosselin, B. Automatic defect segmentation of “Jonagold” apples on multi-spectral images: A comparative study. *Postharvest Biol. Technol.* **2006**, *42*, 271–279. [[CrossRef](#)]
25. Mahanti, N.K.; Pandiselvam, R.; Kothakota, A.; Ishwarya, S.P.; Chakraborty, S.K.; Kumar, M.; Cozzolino, D. Emerging non-destructive imaging techniques for fruit damage detection: Image processing and analysis. *Trends Food Sci. Technol.* **2022**, *120*, 418–438. [[CrossRef](#)]
26. Jawale, D.; Deshmukh, M. Real time automatic bruise detection in (Apple) fruits using thermal camera. In Proceedings of the 2017 International Conference on Communication and Signal Processing (ICCSPP), Chennai, India, 6–8 April 2017.
27. He, Y.; Deng, B.; Wang, H.; Cheng, L.; Zhou, K.; Cai, S.; Ciampa, F. Infrared machine vision and infrared thermography with deep learning: A review. *Infrared Phys. Technol.* **2021**, *116*, 103754. [[CrossRef](#)]
28. Varith, J.; Hyde, G.M.; Baritelle, A.L.; Fellman, J.K.; Sattabongkot, T. Non-contact bruise detection in apples by thermal imaging. *Innov. Food Sci. Emerg. Technol.* **2003**, *4*, 211–218. [[CrossRef](#)]
29. Zeng, X.; Miao, Y.; Ubaid, S.; Gao, X.; Zhuang, S. Detection and classification of bruises of pears based on thermal images. *Postharvest Biol. Technol.* **2020**, *161*, 111090. [[CrossRef](#)]
30. Dong, Y.-Y.; Huang, Y.-S.; Xu, B.-L.; Li, B.-C.; Guo, B. Bruise detection and classification in jujube using thermal imaging and DenseNet. *J. Food Process. Eng.* **2022**, *45*, e13981. [[CrossRef](#)]
31. Jianmin, Z.; Qixian, Z.; Juanjuan, L.; Dongdong, X. Design of On-line Detection System for Apple Early Bruise Based on Thermal Properties Analysis. In Proceedings of the 2010 International Conference on Intelligent Computation Technology and Automation, Changsha, China, 11–12 May 2010.
32. Baranowski, P.; Mazurek, W.; Witkowska-Walczak, B.; Slawinski, C. Detection of early apple bruises using pulsed-phase thermography. *Postharvest Biol. Technol.* **2009**, *53*, 91–100. [[CrossRef](#)]
33. Li, H.; Wu, X.J.; Kittler, J. RFN-Nest: An end-to-end residual fusion network for infrared and visible images. *Inf. Fusion* **2021**, *73*, 72–86. [[CrossRef](#)]
34. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H.J.A. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
35. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
36. Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for MobileNetV3. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.
37. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [[CrossRef](#)]
38. Saranya, N.; Srinivasan, K.; Kumar, S.K.P. Banana ripeness stage identification: A deep learning approach. *J. Ambient Intell. Humaniz. Comput.* **2022**, *13*, 4033–4039. [[CrossRef](#)]
39. Ma, W.H.; Wang, K.; Li, J.W.; Yang, S.X.; Li, J.F.; Song, L.P.; Li, Q.F. Infrared and Visible Image Fusion Technology and Application: A Review. *Sensors* **2023**, *23*, 599. [[CrossRef](#)]
40. Jin, X.; Jiang, Q.; Yao, S.W.; Zhou, D.M.; Nie, R.C.; Hai, J.J.; He, K.J. A survey of infrared and visual image fusion methods. *Infrared Phys. Technol.* **2017**, *85*, 478–501. [[CrossRef](#)]

41. Helin, R.; Indahl, U.G.; Tomic, O.; Liland, K.H. On the possible benefits of deep learning for spectral preprocessing. *J. Chemom.* **2022**, *36*, e3374. [[CrossRef](#)]
42. Lu, J.; Behbood, V.; Hao, P.; Zuo, H.; Xue, S.; Zhang, G.Q. Transfer learning using computational intelligence: A survey. *Knowl.-Based Syst.* **2015**, *80*, 14–23. [[CrossRef](#)]
43. Ji, J.; Zhu, X.; Ma, H.; Wang, H.; Jin, X.; Zhao, K. Apple Fruit Recognition Based on a Deep Learning Algorithm Using an Improved Lightweight Network. *Appl. Eng. Agric.* **2021**, *37*, 123–134. [[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.