

Article

Research on Maize Seed Classification and Recognition Based on Machine Vision and Deep Learning

Peng Xu ¹, Qian Tan ², Yunpeng Zhang ², Xiantao Zha ², Songmei Yang ² and Ranbing Yang ^{2,*}

¹ College of Information and Communication Engineering, Hainan University, Haikou 570228, China; xupeng@hainanu.edu.cn

² College of Mechanical and Electrical Engineering, Hainan University, Haikou 570228, China; tanqian@hainanu.edu.cn (Q.T.); zhangyunpeng@hainanu.edu.cn (Y.Z.); zhaxt@hainanu.edu.cn (X.Z.); 995093@hainanu.edu.cn (S.Y.)

* Correspondence: yangranbing@hainanu.edu.cn

Abstract: Maize is one of the essential crops for food supply. Accurate sorting of seeds is critical for cultivation and marketing purposes, while the traditional methods of variety identification are time-consuming, inefficient, and easily damaged. This study proposes a rapid classification method for maize seeds using a combination of machine vision and deep learning. 8080 maize seeds of five varieties were collected, and then the sample images were classified into training and validation sets in the proportion of 8:2, and the data were enhanced. The proposed improved network architecture, namely P-ResNet, was fine-tuned for transfer learning to recognize and categorize maize seeds, and then it compares the performance of the models. The results show that the overall classification accuracy was determined as 97.91, 96.44, 99.70, 97.84, 98.58, 97.13, 96.59, and 98.28% for AlexNet, VGGNet, P-ResNet, GoogLeNet, MobileNet, DenseNet, ShuffleNet, and EfficientNet, respectively. The highest classification accuracy result was obtained with P-ResNet, and the model loss remained at around 0.01. This model obtained the accuracy of classifications for BaoQiu, ShanCu, XinNuo, LiaoGe, and KouXian varieties, which reached 99.74, 99.68, 99.68, 99.61, and 99.80%, respectively. The experimental results demonstrated that the convolutional neural network model proposed enables the effective classification of maize seeds. It can provide a reference for identifying seeds of other crops and be applied to consumer use and the food industry.

Keywords: machine vision; maize seeds; classification; deep learning; convolutional neural network



Citation: Xu, P.; Tan, Q.; Zhang, Y.; Zha, X.; Yang, S.; Yang, R. Research on Maize Seed Classification and Recognition Based on Machine Vision and Deep Learning. *Agriculture* **2022**, *12*, 232. <https://doi.org/10.3390/agriculture12020232>

Academic Editor: Dimitre Dimitrov

Received: 9 December 2021

Accepted: 3 February 2022

Published: 6 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Maize (*Zea mays* L.) is a significant fundamental agricultural product for the economies and markets of countries. With the development of society, the widespread use of biotechnology has improved maize breeding technologies and accelerated the renewal and iteration of varieties. However, the increasing number of varieties of maize seed and their color characteristics overlap to make it more challenging to classify seeds after harvest [1]. In addition, the phenomenon of seeds being mixed may occur during production activities such as planting, harvesting, transportation, and storage [2]. Therefore, variety identification plays a crucial role in the production, processing, and marketing of seeds. It will provide markets and consumers with pure seeds that will ensure yields and stabilize their market value.

Traditionally, there are many methods for variety identification [3]. Morphological identification is limited by the range of morphological characteristics, the interference of human and environmental factors, and the impact of testing period or cost, which will decrease the accuracy of identification. Biochemical identification enables the recognition of seeds with different genetic characteristics, but it is difficult to identify closely related varieties. Molecular identification through DNA markers has the advantage of genetic stability and is independent of environmental conditions. But the cost of primer design is

high, and the identification process can damage the sample [4]. In summary, these detection methods are difficult to adapt to be online detection in the seed processing industry [5] and cannot complete the sorting of samples during processing. Therefore, it is necessary to develop non-destructive, rapid, and efficient methods for the variety identification and classification of maize seeds.

Machine vision is the method of image processing adapted to multi-classification, which has been successfully applied in several fields. As for seed classification, the non-destructive nature hereof is undoubtedly a better choice than traditional detection methods. This method extracts color, texture, and shape features from seed images for classification. In [6], 12 color features were extracted to distinguish between the different types of damage in maize, with an accuracy of 74.76% for classifying normal and six damaged maize. In [7], 16 morphological features were extracted to classify dry beans, and the overall correct classification rate of SVM was 93.13%. In [8], developed a machine that automatically extracts shape, color, and texture feature data of cabbage seeds and uses them to classify the quality of seeds. The research of maize seeds has focused on bioactivity screening and quality inspection. However, an additional issue that demands consideration is the classification of maize seeds of different varieties [9].

Deep learning techniques have developed rapidly. The convolutional neural network (CNN) is a part of them, which has strong self-learning ability, adaptability, and generalization [2,10,11]. It has achieved considerable success in image classification, object detection, and face recognition [12]. CNN is a deep feedforward network inspired by the receptive field mechanism, which has the properties of local-connectivity, weight sharing, and aggregation in structure [13,14]. The network was composed of an input layer, convolution layers, pooling layers, fully connected layers, and an output layer [15]. CNN models have emerged since 2012, such as AlexNet [16], GoogLeNet [17], VGGNet [18], ResNet [19], DenseNet [20], MobileNet [21], ShuffleNet [22], EfficientNet [23], and more.

Machine vision has also been combined with deep learning to classify seeds [2,24]. In [15], used a CNN to automatically identify haploid and diploid maize seeds through a transfer learning approach. The experiment showed that the CNN model achieved good results, significantly outperforming machine learning-based methods and traditional manual selection. In [25], a wheat recognition system was developed based on VGG16, and the classification accuracy was 98.19%, which could adequately distinguish between different types of wheat grains. In [26], they used their self-designed CNN and ResNet models to identify seven cotton seed varieties, and it achieved good results, with 80% accuracy of the model identification. Reference [27] determined HSI images of 10 representative high-quality rice varieties in China and established a rice variety determination model using the PCANet, with a classification accuracy of 98.66%. In [24], used the CNN-ANN model to classify maize seeds, completing a test of 2250 instances in 26.8 s, with a classification accuracy of 98.1%.

Many studies have combined deep learning with machine vision because of its high accuracy, speed, and reliability. However, the increasing number of seed varieties and consumption are placing new demands on these studies, and the applicability of previous research methods has diminished. Therefore, inspired by the successful classification of agricultural products by deep CNNs, this paper studied the classification of maize seeds of different varieties. It is an in-depth exploration from another perspective based on the reference [9]. Specifically, this research used maize seed images from [9] and increased the number of samples by data augmentation. The proposed CNN network and transfer learning were used to study this classification task to obtain the best classification performance. This study not only extends [9], but its distinction lies in the attempt to automatically obtain deeper features from the data to achieve end-to-end problem-solving.

In summary, the objective of this study was to propose a non-destructive method for the automatic identification and classification of different varieties of maize seeds from images, thus overcoming the time-consuming and inefficient problems of traditional identification methods. We would pursue this study objective by: (1) implementing

machine vision combined with deep learning by applying a CNN network with P-ResNet architecture for varietal detection; (2) establishing a seed dataset and dividing it into training and validation sets in the ratio of 8:2 for experiments; as well as (3) evaluating and comparing the classification performance of the models and using visualization to validate the results. In addition, we address the following specific hypotheses: (1) transfer learning can acquire knowledge learned in other Settings and be used to complete similar tasks in deep learning, thus helping to save the training time of the model; and (2) compared with manual feature extraction methods, the CNN model can be used to automatically extract more depth features from images, thus improving the classification performance.

2. Materials and Methods

2.1. Sample Preparation

In this study, 8080 maize seeds of five common varieties in China were used as a dataset to train the deep learning model for image classification. These maize seeds were provided by the National Seed Breeding Base in Hainan (Longitude 109.17° E, Latitude 18.35° N). These seeds were selected and certified by experts and have manually been cleaned for impurities and dust [9]. The selected for the experiment were of excellent quality, without noticeable defects or damage. The image in the dataset included 1710 BaoQiu, 1800 KouXian, 570 LiaoGe, 2000 ShanCu, and 2000 XinNuo. Figure 1 shows RGB images of five varieties of maize seeds.

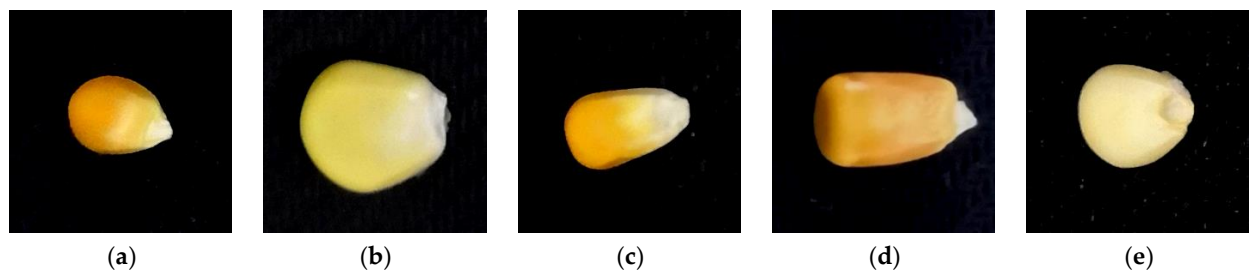


Figure 1. RGB images of maize seed grains. (a) BaoQiu. (b) KouXian. (c) LiaoGe. (d) ShanCu. (e) XinNuo.

These seeds were placed individually on a black background for image acquisition. The influence of the seed storage situation in the National Seed Breeding Base in Hainan results in different quantities of seeds for each variety. However, the situation does not complicate the assessment of the accuracy of the different varieties, as there was no reuse of maize seeds. In addition, there were different shapes and sizes of the five maize seeds in the images, which provided some assistance in the classification of this study. In this work, all of the seeds were randomly divided into a training set (80%) and a validation set (20%), then stored in their respective subdirectories. Finally, the training and validation sets contained 6464 and 1616 maize seeds, respectively. To establish the classification model, BaoQiu, KouXian, LiaoGe, ShanCu, and XinNuo samples were collected in 2020, as shown in Table 1.

Table 1. Data for training and validation.

No.	Cultivar Name	Seeds		
		Training Set	Validation Set	Number
1	BaoQiu	1368	342	1710
2	KouXian	1440	360	1800
3	LiaoGe	456	114	570
4	ShanCu	1600	400	2000
5	XinNuo	1600	400	2000
	Total	6464	1616	8080

2.2. Image Acquisition and Segmentation

In the machine vision part, an image acquisition system was built for capturing maize seeds. The system has cameras mounted on top and light sources on either side to provide illumination. It is impractical to capture a single seed in a stretch during image acquisition. Therefore, hundreds of seeds were photographed in an area of $12\text{ cm} \times 12\text{ cm}$, with them not touching each other. All of the photos were taken in the same environment, with a camera distance of 16 cm. The resolution of the acquired image was 3384×2708 pixels, which contained multiple single seeds that cannot be directly included as input in the CNN model. Therefore, the image was segmented into 350×350 pixels size and saved in PNG format for use.

2.3. Image Preprocessing and Data Augmentation

Large amounts of training data can avoid over-fitting and improve the accuracy of CNN, so data augmentation operation is often used to extend the dataset [28]. The training images were randomly rotated [29], flipped horizontally and vertically [30], and normalized, considering the uncertainty of the state of the detected seeds in the actual situations. The enhanced image was trained together with the sample image to improve the classification precision and robustness of the model and further improve its applicability. Detailed information is shown in Figure 2.

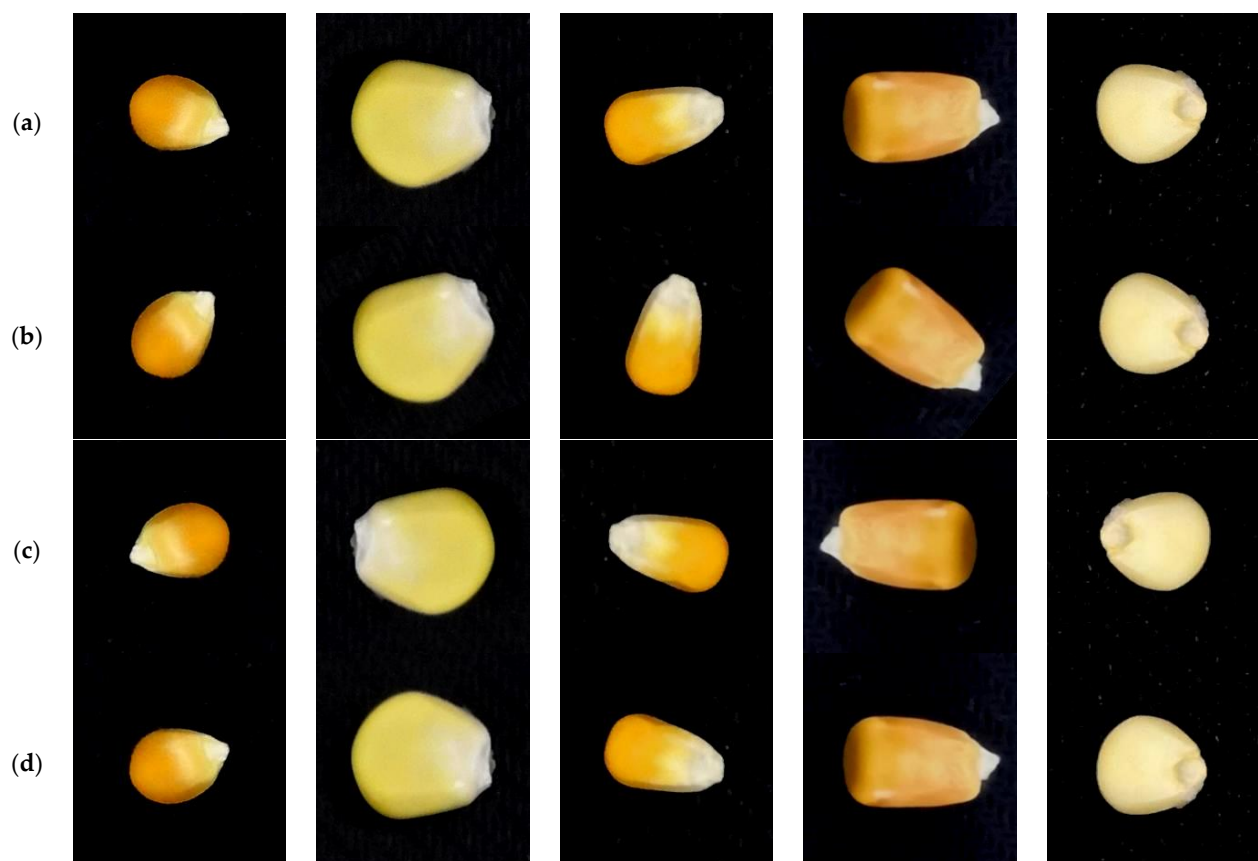


Figure 2. Data enhancement: (a) Original images; (b) Randomly rotated; (c) Flipped horizontally; (d) Flipped vertically.

2.4. Convolutional Neural Network

Deep learning is an emerging algorithm in machine learning, which has attracted extensive attention from researchers because of its remarkable effect on learning image features. Deep learning extracts higher-dimensional and abstract features by autonomic learning from training samples through neural networks [10]. This research proposed a

new model (P-ResNet) based on an improvement of ResNet, which provides a method to classify maize seeds. The network architecture of P-ResNet consists of six parts, five of which are the convolution layer, and the last one is a fully connected layer. The convolution operation is followed by batch normalization, and then ReLU is applied as the activation function to complete the output of the convolution layer. In addition, to avoid over-fitting and reduce the number of parameters and computation in the network, which adopted a strategy of max pooling and average pooling. The input image was resized to $224 \times 224 \times 3$. According to the prepared dataset, the output of the fully connected layer was fed into softmax to generate a probability distribution to predict the varieties of 5 maize seeds. Table 2 provides a detailed description of the P-ResNet network.

Table 2. Network architecture for P-ResNet.

Layer Name	Output Shape	The Network Layer	Stride
		Input (224×224 RGB image)	
Convolution layer 1	112×112	$7 \times 7, 64$	2
Max pooling		$3 \times 3, 64$	2
Convolution layer 2	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	1, 1, 1
Convolution layer 3	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 3$	2, 1, 1
Convolution layer 4	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 3$	2, 1, 1
Convolution layer 5	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	2, 1, 1
Classification	1×1	average pooling, 5-d fully-connected, softmax	1

As can be seen from Table 2, the convolutional layer 1 of the P-ResNet network goes through a 7×7 convolution. The receptive field is large enough to be used for the feature extraction of images in this database. In order to classify maize seeds more accurately, more subtle features need to be extracted. Furthermore, a suitable network depth was required to be designed and to reduce the size of the presented model. Therefore, the convolution layer of layers 2–5 was improved in the architecture of the network to make it more suitable for the model classification task. The design of this study used twenty-four 3×3 stacked convolution layers for learning, with more nonlinear activation functions to make the decision function more accurate; on the other hand, it can effectively decrease the number of parameters in calculation. Furthermore, in online inspection in the seed processing industry, the objective region occupies a small area of the whole image, and the proportion of information obtained is weak. In order to avoid redundant and useless information, this study adds a pooling layer to integrate spatial information before the convolution kernel of the residual module does down-sampling.

2.5. Transfer Learning

The RGB images with labeled data were input into the improved network P-ResNet. In the experiment, 6464 images were utilized for training and 1616 images for validation. Transfer learning [31] was performed for 5 varieties: BaoQiu, ShanCu, XinNuo, LiaoGe, and KouXian. After training the model, its performance was evaluated and compared through the training and validation sets. The models were developed using the open-source software framework of PyTorch 1.9.0, the programming language of Python 3.8.10, and the Integrated Development Environment of PyCharm 1.3. The classification model was

trained on a server equipped with one NVIDIA GeForce GTX 1660 SUPER GPU and 16 GB GDDR4 on-board memory.

In this study, as shown in Table 3, some classical CNN models have been used to compare with P-ResNet. The acquired data were fed into a pre-trained network, storing the activation values of each layer as features. The cross-dataset fine-tuning method was used for training. According to the new task, the weights of the presented model were updated and back-propagated through the network. This approach can transfer weights from the pre-trained model to the one we want to train. Details of the hyper-parameters applied during the fine-tuning procedure are listed in Table 4. Using these enables transferring the knowledge gained from the large dataset to the classification problem of maize seeds. For the purpose of this study, the convolution layer was used as a fixed feature extractor. Then a fully connected layer with merely five neurons was constructed. Finally, the categorization results were obtained with the prediction layer.

Table 3. Compare the properties of CNN models.

Network Name	Depth	Image Input Size	Parameters (Millions)	FLOPs (G)	Total Memory (MB)
AlexNet	8	224-by-224	16.63	0.31	2.77
VGGNet	16	224-by-224	138.36	15.50	109.29
P-ResNet	26	224-by-224	17.96	2.75	32.83
GoogLeNet	22	224-by-224	6.99	1.59	30.03
MobileNet	19	224-by-224	3.50	0.32	74.26
DenseNet	121	224-by-224	7.98	2.88	147.10
ShuffleNet	19	224-by-224	1.37	0.04	11.24
EfficientNet	18	224-by-224	21.46	2.87	144.98

Table 4. Hyper-parameters were applied to the fine-tuning procedure.

Parameter	Value
Epochs	30
Batch size	32
Learn rate	0.001
Momentum	0.9
Learn rate weight coefficient	15
Learn rate bias coefficient	15
Learn rate schedule	Exponential
Weight decay	0.005
Decay period	10

The whole network has 17,960,232 parameters. The proposed CNN model uses Adam [32] as the optimizer to train with an initial learning rate value of 0.001, and the loss function of the network was declined by updating the weight parameters. Batch Normalization was used between each convolution and ReLU layer during network training, instead of the traditional dropout to improve training and reduce over-fitting. Epoch is the complete training cycle of the entire dataset with maize seeds, and its maximum value corresponds to the limit value of the minimum loss function. The maximum training epoch was set to 30, and the minimum batch size was set to 32. These parameters achieved better results in the optimizer. The process of transfer learning and classification of maize seeds in the network involved in the experiment was given in Figure 3.

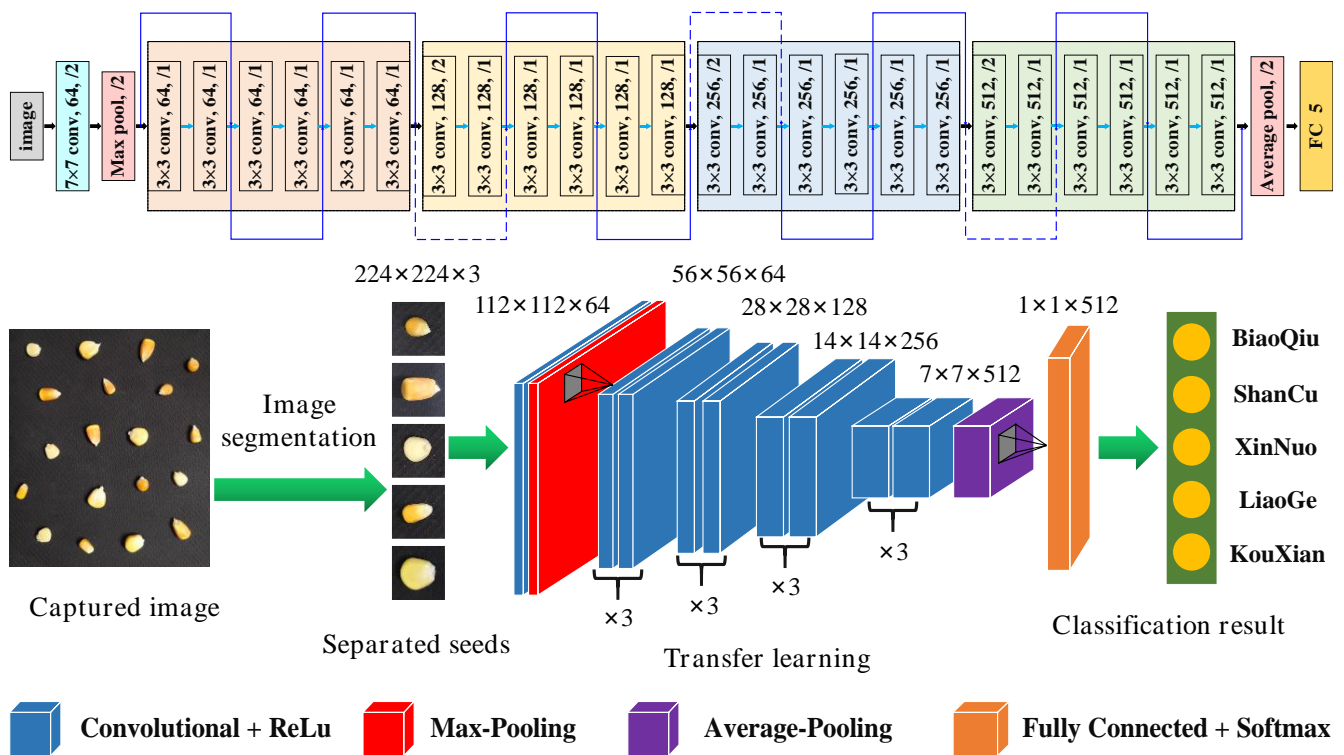


Figure 3. Process of transfer learning and classification of maize seeds.

2.6. Performance Evaluation

In this paper, the confusion matrix was used to visualize the performance of the CNN model. This data on the confusion matrix represents the actual class in the samples and the class predicted by the CNN classifier. The four metrics typically included true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) [24]. In this work, TP and TN correspond to the correct identification of maize seeds, while FP and FN correspond to false identification of it. The performance of models was evaluated based on some statistical parameters of the confusion matrix, such as accuracy, sensitivity, specificity, precision, and F1-score, which can be obtained from them [33]. The performance evaluation was performed using images from the validation set and their respective labels, which were not used for training. Table 5 represents the formulae for performance evaluation and their evaluation focus.

Table 5. Performance evaluation to measure the performance of the CNN models.

Metrics	Formula	Evaluation Focus
Accuracy	$\frac{TP + TN}{TP + FP + FN + TN}$	It is the sum of correct predictions divided by all the predictions.
Specificity	$\frac{TN}{TN + FP}$	It reflects the ability of the classifier to exclude misclassification images.
Sensitivity	$\frac{TP}{TP + FN}$	It reflects the ability of the model to detect instances of certain classes.
Precision	$\frac{TP}{TP + FP}$	Its high value indicates the low number of false positives hence better classification.
F1-score	$\frac{2 * TP}{2 * TP + FP + FN}$	Its high value means the model classifies well.

3. Results

The parameters given in Table 4 were selected for transfer learning. The prepared dataset was trained using AlexNet, VGGNet, P-ResNet, GoogLeNet, MobileNet, DenseNet, ShuffleNet, and EfficientNet. The optimal parameters in Table 4 were used to prevent

over-fitting during training and avoid spending more time. All networks have been trained for 30 epochs. The accuracy and loss of the training and validation data for each epoch are shown in Figure 4. In the initial phase (1–10 epochs), the loss values declined sharply, but the accuracy improved dramatically. Finally, the CNN models reached an accuracy of over 92% in the training phase, and the loss of models was steady below 0.15, indicating that these are very robust and dependable. Also, the model achieved the convergence procedure in approximately 15 epochs. As can be depicted in Figure 4, after this period, the validation accuracy and loss curves smoothed out, and the difference between the accuracy and loss values of the validation and training data decreased. There, for the fact, was some fluctuation with accuracy and loss for GoogLeNet. This condition suggests that the model is not stable until the 25 epochs, possibly because some of the varieties were easily confounded. There are gaps in GoogLeNet’s handling of the dataset for this study compared to other models. Nevertheless, even in the worst case, the metrics were above 90% or below 0.2. This result indicates that the classifier’s performance is satisfactory and did not prevent it from achieving its final classification purpose.

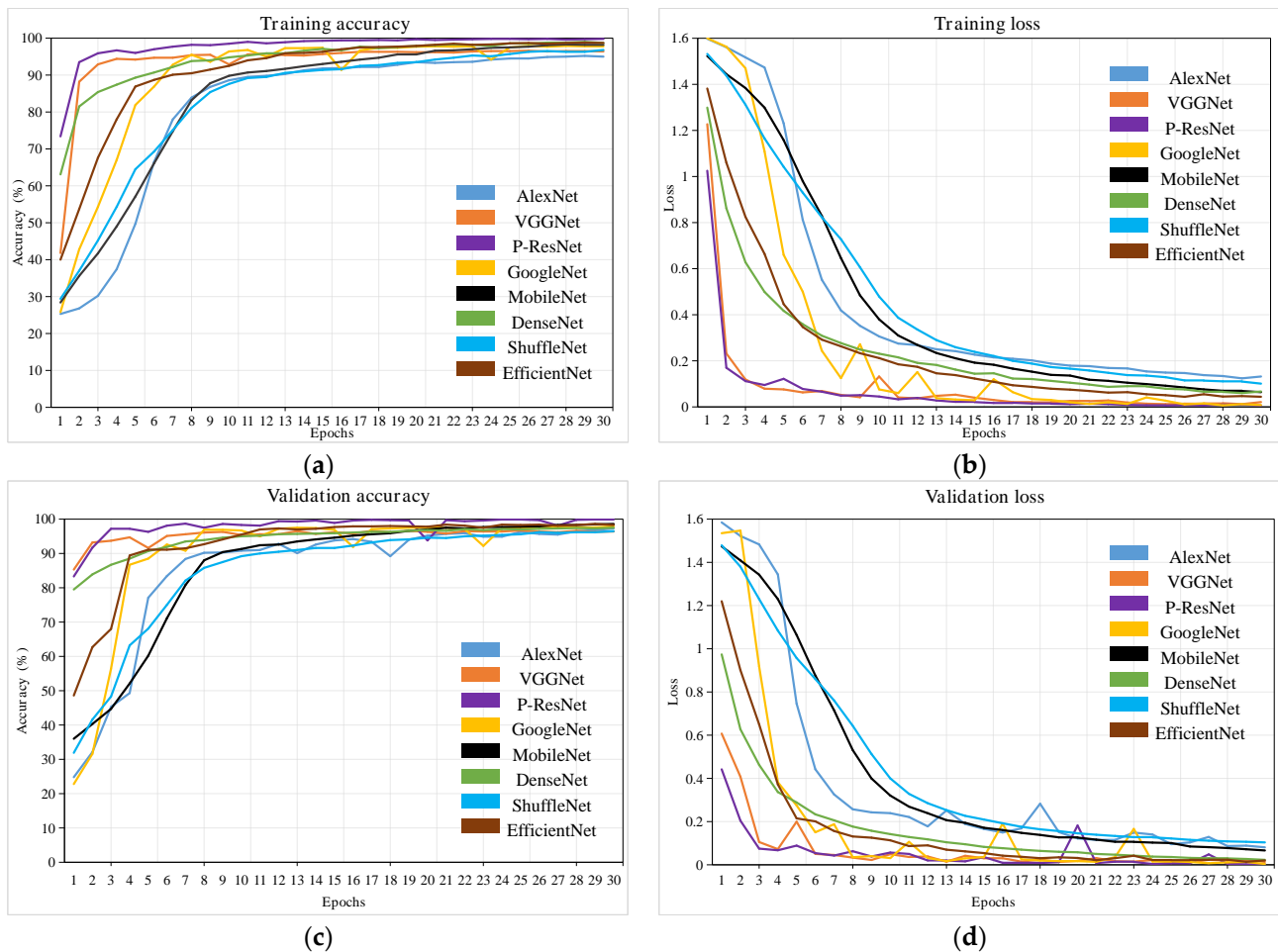


Figure 4. Each epoch of the CNN model: (a) Training accuracy, (b) Training loss, (c) Validation accuracy, (d) Validation loss.

After the training, a confusion matrix was created for each classification algorithm, and performance evaluation was visualized using the values on the confusion matrix (TP, TN, FP, FN). The confusion matrices for the validation set of the CNN model are depicted in Figure 5. In addition, the performance metrics depicted from the confusion matrix are presented in Table 6, including their mean values for precision, specificity, sensitivity, accuracy, and f1 scoring. In this experiment, all CNN models can identify five classes of

maize seeds, and all had an accuracy rate of over 92%. The highest accuracy was obtained by P-ResNet (99.70%), followed by AlexNet (97.91%), VGGNet (96.44%), GoogLeNet (97.84%), MobileNet (98.58%), DenseNet (97.13%), ShuffleNet (96.59%), and EfficientNet (98.28%). Even though these models were trained with images from self-made datasets, fine-tuning these models can achieve similar results to using the end-to-end models in datasets with limited samples. This situation will make image acquisition more convenient and fast, will save effort and time, and will thus improve efficiency. Besides, the confusion matrix and classification results also prove that P-ResNet has excellent performance. This result also illustrates that the presented network can catch the detailed information of the samples. These can provide relatively high accuracy classification under complex datasets, which is beneficial for transferring it to similar classification tasks. The experimental results also demonstrate that enhancing the data used for training has a positive impact on the performance of the presented model on datasets with a small number of samples. In particular, these include datasets with low sample sizes. At the same time, the deep learning-based feature extraction method can effectively preserve information about the maize seeds, reduce the loss of information due to manual feature extraction.

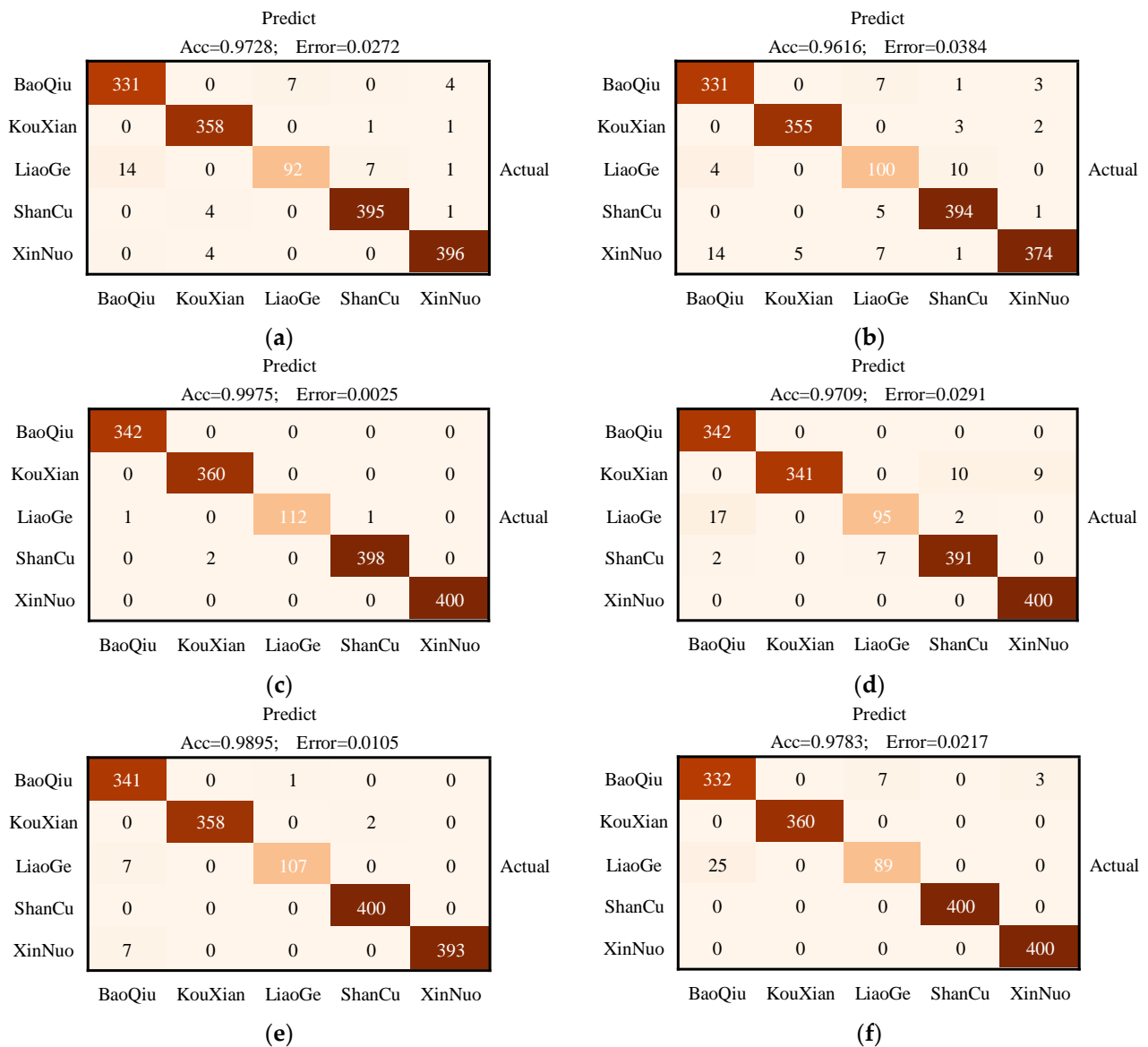


Figure 5. Cont.

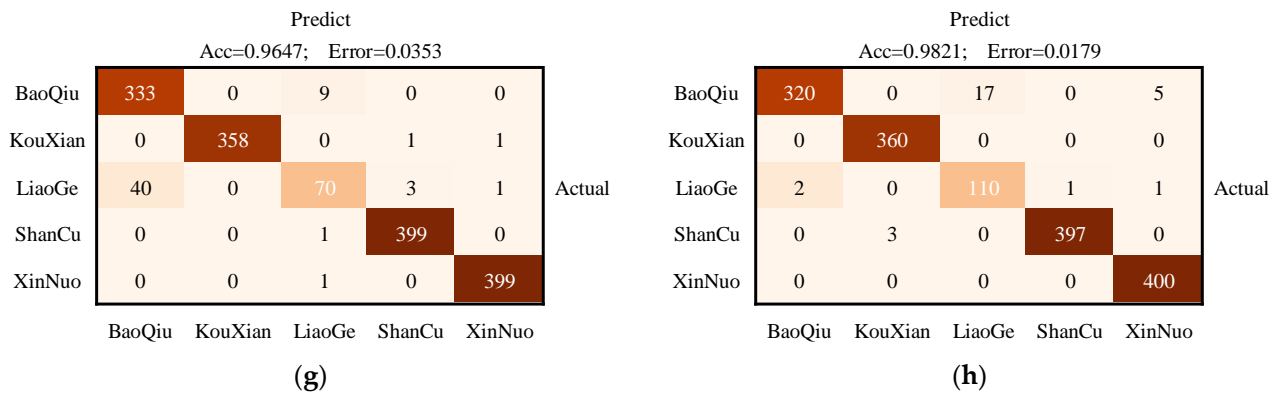


Figure 5. Confusion Matrix: (a) AlexNet; (b) VGGNet; (c) P-ResNet; (d) GoogLeNet; (e) MobileNet; (f) DenseNet; (g) ShuffleNet; and (h) EfficientNet.

The experimental analysis showed that the deep learning architecture with updated weights and fine-tuning had good generalization capability in the maize seed dataset. Compared with the networks in the literature, the proposed P-ResNet has relatively better performance and higher accuracy. It also found that the value of the maximum difference in classification accuracy between all models was no more than 3%. Although there were differences between them, they performed similarly for multi-classification. Therefore, an improved ResNet-based network has been used for transfer learning in the study. Due to its better classification results, it confirms that the idea of balancing its depth and width when designing the network is feasible. It would also increase the complexity of the model and consume more computation time. As can be observed in Table 3, the P-ResNet proposed generates a relatively small number of parameters (17.96 Million) and memory (32.83 MB). VGGNet has 7.6 times as many parameters as it does, while the memory footprint is close to GoogLeNet. The FLOPs value (2.75 G) is approximately the same as that of DenseNet and EfficientNet, indicating the low complexity of the model. It also demonstrates the potential for the network to be lightweight and mobile. With the continuous improvement of the model, a version more suitable for mobile and embedded devices can be achieved.

Table 6. Classification results of CNN models.

Name	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)	F1-Score (%)
AlexNet	97.91	98.31	93.94	95.59	94.80
VGGNet	96.44	97.04	92.98	92.45	92.83
P-ResNet	99.70	99.94	99.55	99.78	99.71
GoogLeNet	97.84	98.26	94.16	95.54	94.99
MobileNet	98.58	98.73	97.25	97.93	97.57
DenseNet	97.13	97.46	93.03	94.99	93.88
ShuffleNet	96.59	97.13	91.54	94.84	92.76
EfficientNet	98.28	99.58	97.86	96.69	97.28

Figure 6 shows a comparison of the performance of the CNN models tested, from which it can be seen that the training time of the proposed network is comparable to that of the lightweight networks (MobileNet and ShuffleNet). AlexNet had the shortest training time (14 min) and VGGNet the longest (62 min). It is important to note that the time required for training the network depends on the hardware resources. The use of advanced GPU can reduce the training time of CNN. However, when time and model complexity were considered, P-ResNet’s training time was only 4 min slower than AlexNet, and the values for Parameters, FLOPs, and total memory were relatively better. This situation is because of the use of Adam optimizer in the proposed model, which minimizes error loss, and transforms the training and validation data for each epoch. In this work, when the

region of interest of seeds was extracted from the image and applied to the deep learning model, the processing time of training can be reduced.

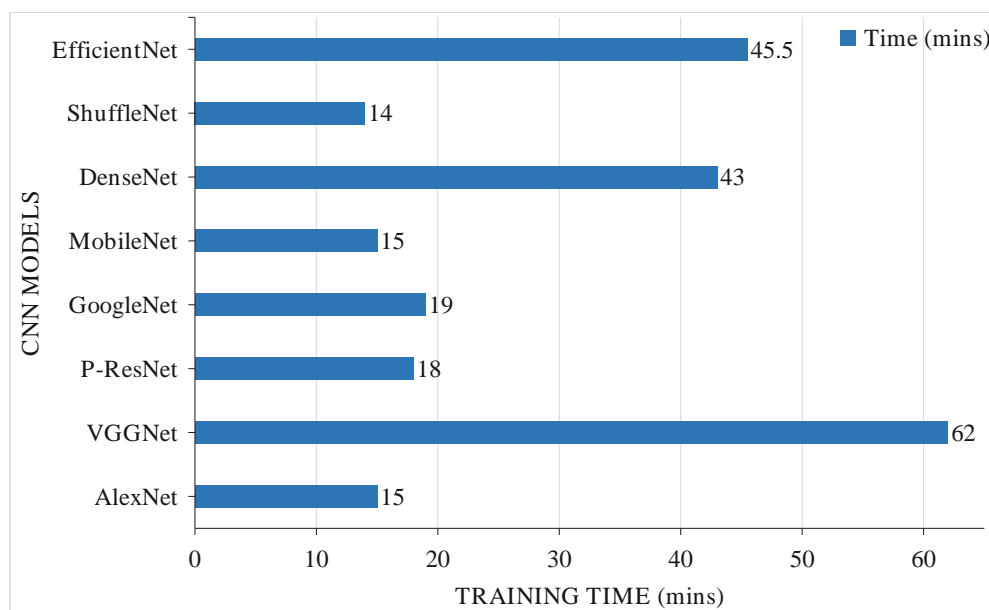


Figure 6. Consumed time for training.

The classification results for all varieties of maize seeds in the different models were shown in Table 7, and this statistic clearly shows how the model performance stays in general and as a whole. As can be seen from Figure 7, the classification accuracy of 8 different models for five maize seeds was over 90%. The P-ResNet network had the best classification performance, with 99.74, 99.68, 99.68, 99.61, and 99.80% accuracy for the BaoQiu, KouXian, LiaoGe, ShanCu, and XinNuo, respectively. However, BaoQiu and LiaoGe had lesser classification performance among all models, and the lowest values were 94.97 and 94.60%, respectively. These results indicated that VGGNet, DenseNet, and ShuffleNet models were not the best adapted for these two varieties. It also revealed that there probably is overlap in features between BaoQiu and LiaoGe and the other three varieties, resulting in poor distinction. In addition, the low number of LiaoGe in the dataset may also have contributed to this situation. Through, their classification results were still very encouraging.

Table 7. Statistics of classification results for all maize varieties.

Model	Classification Accuracy (%)				
	BaoQiu	KouXian	LiaoGe	ShanCu	XinNuo
AlexNet	97.45	98.38	97.21	98.20	98.32
VGGNet	96.21	97.38	95.96	96.70	95.96
P-ResNet	99.74	99.68	99.68	99.61	99.80
GoogleNet	97.82	97.82	97.39	97.70	98.44
MobileNet	98.07	98.88	98.50	98.88	98.57
DenseNet	95.83	98.00	96.02	98.00	97.81
ShuffleNet	94.97	97.88	94.60	97.69	97.81
EfficientNet	97.51	98.81	97.70	98.75	98.63

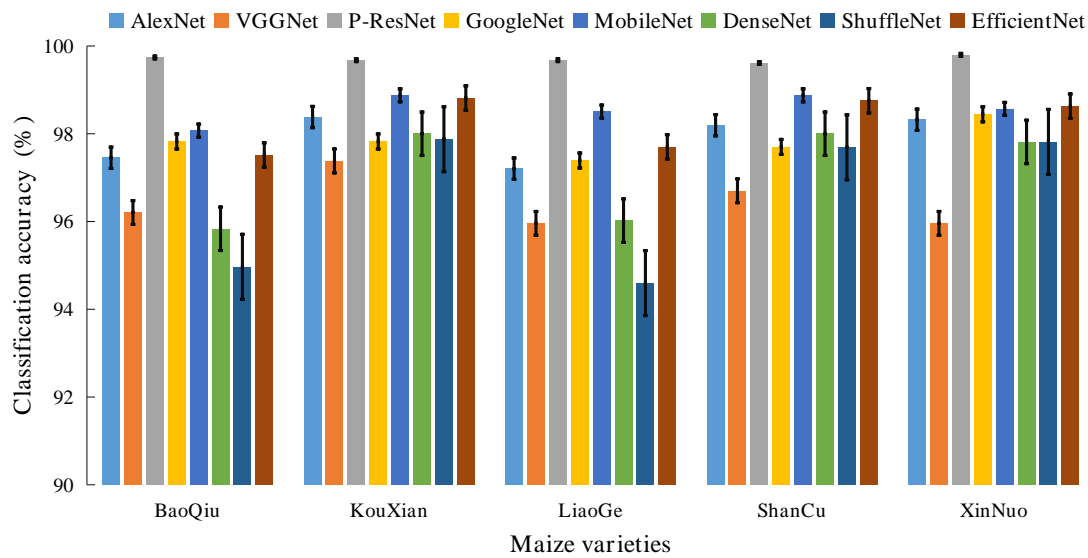


Figure 7. Classification results for maize seeds. The error bars are standard deviations of means.

In this experiment, given the different methods, datasets, and classification criteria employed, relevant studies cannot be compared in detail. Nevertheless, it compared some applications in agricultural classification tasks, and the results are shown in Table 8. These comparisons considered several criteria, such as dataset size, application, the method used, and accuracy. The results showed that the accuracy for the different classification tasks was above 95%, which indicated that the CNN model proposed in this paper and the pre-training method using transfer learning were feasible. It can provide a reference for the classification of agricultural products. The five types of maize seeds utilized in the research are relatively common in China, with a wide distribution of planting areas. In this situation, the credibility of this study has been enhanced. Although the P-ResNet model achieved good results, maize seeds may vary depending on storage time and cultivation conditions (soil or climate). These conditions lead to changes in the dataset, which may influence its accuracy in distinguishing the target varieties. Therefore, it will be necessary to update the algorithm in the future, and the aim is to retain the classification precision and robustness of the model.

Table 8. Comparison of the proposed model and related studies (maize seeds).

Imaging Method	Dataset Size	Application	Approach	Result	References
Hyperspectral imaging	1632	Variety identification	LDA	99.13%	[1]
Digital camera	700	Quality detection	Maximum likelihood	96.67%	[6]
Near-infrared spectroscopy	760	Variety identification	PLS-DA	99.19%	[3]
Digital camera	5400	Variety identification	ANN	98.10%	[24]
Near-infrared spectroscopy	2250	Variety identification	LSTM	95.22%	[34]
Digital camera	8080	Variety identification	SVM	96.46%	[9]
Digital camera	1600	Quality detection	VGG16	98.00%	[2]
Digital camera	8080	Variety classification	P-ResNet	99.70%	Our work

These results were obtained using the Gradient-weighted Class Activation Mapping (Grad-CAM) technique to visualize the used regions of a random input image to extract features for image classification prediction [35]. The gradient of any target feature through the last convolution layer produces a roughly local feature map, highlighting the regions of the image that are important for it. Figure 8 shows the achieved results of the implementation of this method on maize seed images. It can be shown that the image locations for seeds were accurately calculated, with the class activation heat map indicating the importance rank and similarity of the location relative to the particular variety.

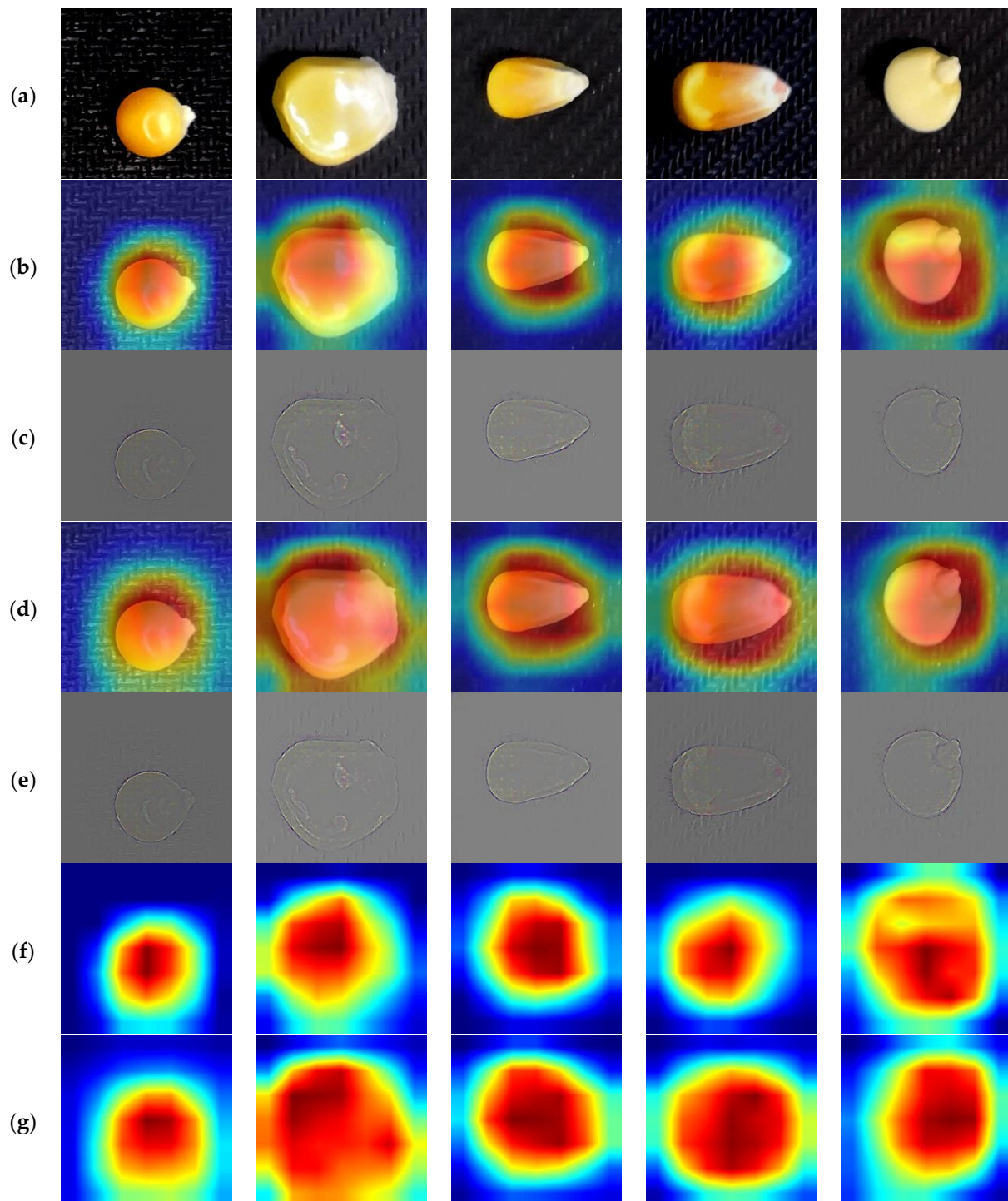


Figure 8. (a) Original images; (b) Grad-CAM visualization; (c) Guided Grad-CAM visualization; (d) Grad-CAM++ visualization; (e) Guided backpropagation visualization; (f) HeatMap visualization; (g) HeatMap++ visualization.

4. Discussion

There are two primary methods for training CNN models using sample data: (1) starting from zero; and (2) transfer learning. In practice, while training a CNN model from the ground up gives us the best active control concerning the network, it may not have enough data and time to train in some cases, or the data to create the markers may be difficult to obtain. Moreover, over-fitting and convergence states are also potential problems. In such

cases, transfer learning can be applied to gain knowledge gained in other settings. It is a convenient and effective method of knowledge adaptation [31], which is usually more efficient than training a new neural network since all parameter values are not required to start from zero. In higher-layers networks, some features are more applicable to a specific task. However, there are many similar features like color and texture for the lower layers of the network. These can be transferred to other tasks and are very helpful for performing similar tasks in deep learning.

P-ResNet was designed based on the principle of balancing the width and depth of the network according to the specific task, which has a better architecture than GoogLeNet, DenseNet, and EfficientNet. It can reduce parameters for computation and avoid gradient disappearance and gradient explosion during training. Meanwhile, it does not need to crop or scale the input image like AlexNet and VGGNet, which can maximally protect the information integrity. In addition, the underlying implementation of this network was simplified to make it more lightweight as possible as MobileNet and ShuffleNet. Since only images are required, which can be produced by low-cost digital cameras, this approach can be widely deployed and disseminated in intelligent agriculture. Machine vision can only obtain phenotypic information of seeds, while spectral information can reflect the internal quality of seeds. The combination of CNN-based machine vision and spectroscopic techniques for seed classification and detection was considered in the follow-up work.

The proposed method has been compared with related work based on the unification of the research objectives (classification or identification) and the object of study (maize seeds). In Table 7, it can be seen that automatic extraction of image features for recognition using CNN is better than manual extraction, and these results illustrate that deep learning is more effective than traditional machine learning methods in cultivar classification. However, the variety and number of samples collected in this study are limited and cannot represent all maize seeds within China. Therefore, the number of samples should be increased to improve its applicability to the model. Moreover, this experiment only considered the classification effects of seed samples from the same year, so the impact of different planting years, growing regions, and climatic conditions on the classification of seeds of the same variety can be compared in subsequent studies.

5. Conclusions

In this work, a combination of deep learning algorithms and machine vision has been used to automatically classify five varieties of maize seeds using a CNN model. In terms of classification, the model architecture developed can be applied to different regions and types of seeds to ensure the provision of high-quality seeds for agricultural production. Also, the method has application potential in identifying varieties of seeds, and the developed variety classification model can be applied to seed sorting machinery to provide an idea and reference for real-time industrial detection. This study proposes an improved model, P-ResNet, and compares it with AlexNet, VGGNet, GoogLeNet, MobileNet, DenseNet, ShuffleNet, and EfficientNet models. The results showed that the P-ResNet model achieved the best accuracy to classify maize seeds in a non-destructive, fast, and efficient manner. These results highlight the advantages of transfer learning and its potential to work with deep learning using a few quantities of training samples. In addition, the Grad-CAM has been used to visualize the regions of use of the input seed images, making this work more efficient and productive. This machine vision technology based on CNN with high accuracy and reliability can also be applied in other intelligent agricultural equipment to facilitate the analysis of seeds or other crops to save cost, labor, and time.

Based on the work presented in this paper, further studies on more varieties of maize seeds and the environment in which they are grown would be appropriate, and it is in order to optimize the stability of the proposed model. Considering that each seed in maize has its genetic characteristics, grouping seeds of selected varieties might avoid natural variability in seedlings. Given the real-time nature of the data, that helps to develop an integrated and

intelligent automated seed sorting system for the food industry and smartphone-based applications used by consumers.

Author Contributions: Conceptualization, P.X., Q.T., Y.Z. and R.Y.; methodology, P.X.; software, P.X.; validation, P.X. and R.Y.; formal analysis, P.X., Q.T., Y.Z., X.Z., S.Y. and R.Y.; investigation, P.X.; resources, R.Y.; data curation, P.X.; writing—original draft preparation, P.X., Q.T. and Y.Z.; writing—review and editing, P.X., X.Z., S.Y. and R.Y.; visualization, P.X.; supervision, X.Z., S.Y. and R.Y.; project administration, P.X.; funding acquisition, R.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Talent Foundation Project of China, grant number T2019136.

Institutional Review Board Statement: Not applicable.

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

Data Availability Statement: All data are presented in this article in the form of figures and tables.

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

References

- Xia, C.; Yang, S.; Huang, M.; Zhu, Q.; Guo, Y.; Qin, J. Maize Seed Classification Using Hyperspectral Image Coupled with Multi-Linear Discriminant Analysis. *Infrared Phys. Technol.* **2019**, *103*, 103077. [[CrossRef](#)]
- Tu, K.; Wen, S.; Cheng, Y.; Zhang, T.; Pan, T.; Wang, J.; Wang, J.; Sun, Q. A Non-Destructive and Highly Efficient Model for Detecting the Genuineness of Maize Variety 'JINGKE 968' Using Machine Vision Combined with Deep Learning. *Comput. Electron. Agric.* **2021**, *182*, 106002. [[CrossRef](#)]
- Qiu, G.; Lü, E.; Wang, N.; Lu, H.; Wang, F.; Zeng, F. Cultivar Classification of Single Sweet Corn Seed Using Fourier Transform Near-Infrared Spectroscopy Combined with Discriminant Analysis. *Appl. Sci.* **2019**, *9*, 1530. [[CrossRef](#)]
- Cui, Y.; Xu, L.; An, D.; Liu, Z.; Gu, J.; Li, S.; Zhang, X.; Zhu, D. Identification of Maize Seed Varieties Based on near Infrared Reflectance Spectroscopy and Chemometrics. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 177–183. [[CrossRef](#)]
- Xie, C.; He, Y. Modeling for Mung Bean Variety Classification Using Visible and Near-Infrared Hyperspectral Imaging. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 187–191. [[CrossRef](#)]
- Li, X.; Dai, B.; Sun, H.; Li, W. Corn Classification System Based on Computer Vision. *Symmetry* **2019**, *11*, 591. [[CrossRef](#)]
- Koklu, M.; Ozkan, I.A. Multiclass Classification of Dry Beans Using Computer Vision and Machine Learning Techniques. *Comput. Electron. Agric.* **2020**, *174*, 105507. [[CrossRef](#)]
- Huang, K.Y.; Cheng, J.F. A Novel Auto-Sorting System for Chinese Cabbage Seeds. *Sensors* **2017**, *17*, 886. [[CrossRef](#)]
- Xu, P.; Yang, R.; Zeng, T.; Zhang, J.; Zhang, Y.; Tan, Q. Varietal Classification of Maize Seeds Using Computer Vision and Machine Learning Techniques. *J. Food Process Eng.* **2021**, *44*, e13846. [[CrossRef](#)]
- Traore, B.B.; Kamsu-Foguem, B.; Tangara, F. Deep Convolution Neural Network for Image Recognition. *Ecol. Inform.* **2018**, *48*, 257–268. [[CrossRef](#)]
- Zhao, G.; Quan, L.; Li, H.; Feng, H.; Li, S.; Zhang, S.; Liu, R. Real-Time Recognition System of Soybean Seed Full-Surface Defects Based on Deep Learning. *Comput. Electron. Agric.* **2021**, *187*, 106230. [[CrossRef](#)]
- Nie, P.; Zhang, J.; Feng, X.; Yu, C.; He, Y. Classification of Hybrid Seeds Using Near-Infrared Hyperspectral Imaging Technology Combined with Deep Learning. *Sens. Actuators B Chem.* **2019**, *296*, 126630. [[CrossRef](#)]
- Qiu, Z.; Chen, J.; Zhao, Y.; Zhu, S.; He, Y.; Zhang, C. Variety Identification of Single Rice Seed Using Hyperspectral Imaging Combined with Convolutional Neural Network. *Appl. Sci.* **2018**, *8*, 212. [[CrossRef](#)]
- De Medeiros, A.D.; Bernardes, R.C.; da Silva, L.J.; de Freitas, B.A.L.; dos Dias, D.C.F.S.; da Silva, C.B. Deep Learning-Based Approach Using X-Ray Images for Classifying Crambe Abyssinica Seed Quality. *Ind. Crops Prod.* **2021**, *164*, 113378. [[CrossRef](#)]
- Altuntaş, Y.; Cömert, Z.; Kocamaz, A.F. Identification of Haploid and Diploid Maize Seeds Using Convolutional Neural Networks and a Transfer Learning Approach. *Comput. Electron. Agric.* **2019**, *163*, 104874. [[CrossRef](#)]
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
- 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 Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
- Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.

20. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [[CrossRef](#)]
21. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
22. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. Shufflenet V2: Practical Guidelines for Efficient Cnn Architecture Design. In Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 15th European Conference, Munich, Germany, 8–14 September 2018; Part XIV.
23. Tan, M.; Le, Q.v. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, CA, USA, 9–15 June 2019; Volume 2019.
24. Javanmardi, S.; Miraei Ashtiani, S.H.; Verbeek, F.J.; Martynenko, A. Computer-Vision Classification of Corn Seed Varieties Using Deep Convolutional Neural Network. *J. Stored Prod. Res.* **2021**, *92*, 101800. [[CrossRef](#)]
25. Özkan, K.; Işık, Ş.; Yavuz, B.T. Identification of Wheat Kernels by Fusion of RGB, SWIR, and VNIR Samples. *J. Sci. Food Agric.* **2019**, *99*, 4977–4984. [[CrossRef](#)]
26. Zhu, S.; Zhou, L.; Gao, P.; Bao, Y.; He, Y.; Feng, L. Near-Infrared Hyperspectral Imaging Combined with Deep Learning to Identify Cotton Seed Varieties. *Molecules* **2019**, *24*, 3268. [[CrossRef](#)]
27. Weng, S.; Tang, P.; Yuan, H.; Guo, B.; Yu, S.; Huang, L.; Xu, C. Hyperspectral Imaging for Accurate Determination of Rice Variety Using a Deep Learning Network with Multi-Feature Fusion. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *234*, 118237. [[CrossRef](#)]
28. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
29. Xie, S.; Tu, Z. Holistically-Nested Edge Detection. *Int. J. Comput. Vis.* **2017**, *125*. [[CrossRef](#)]
30. Kok, K.Y.; Rajendran, P. Validation of Harris Detector and Eigen Features Detector. In Proceedings of the IOP Conference Series: Materials Science and Engineering, International Conference on Aerospace and Mechanical Engineering (AeroMech17), Batu Ferringhi, Penang, Malaysia, 21–22 November 2017; Volume 370.
31. Salaken, S.M.; Khosravi, A.; Nguyen, T.; Nahavandi, S. Seeded Transfer Learning for Regression Problems with Deep Learning. *Expert Syst. Appl.* **2019**, *115*, 565–577. [[CrossRef](#)]
32. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015.
33. Ishengoma, F.S.; Rai, I.A.; Said, R.N. Identification of Maize Leaves Infected by Fall Armyworms Using UAV-Based Imagery and Convolutional Neural Networks. *Comput. Electron. Agric.* **2021**, *184*, 106124. [[CrossRef](#)]
34. Zhang, C.; Zhao, Y.; Yan, T.; Bai, X.; Xiao, Q.; Gao, P.; Li, M.; Huang, W.; Bao, Y.; He, Y.; et al. Application of Near-Infrared Hyperspectral Imaging for Variety Identification of Coated Maize Kernels with Deep Learning. *Infrared Phys. Technol.* **2020**, *111*, 103550. [[CrossRef](#)]
35. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359. [[CrossRef](#)]