agriculture

MDPI

*Article*

# Wheat-Seed Variety Recognition Based on the GC_DRNet Model

**Xue Xing [1], Chengzhong Liu [1,*], Junying Han [1], Quan Feng [2], Qinglin Lu [3] and Yongqiang Feng [1]**

[1] College of Information Sciences and Technology, Gansu Agricultural University, Lanzhou 730070, China; xingx@st.gsau.edu.cn (X.X.); hanjy@gsau.edu.cn (J.H.); fengyq@st.gsau.edu.cn (Y.F.)
[2] College of Mechanical and Electrical Engineering, Gansu Agricultural University, Lanzhou 730070, China; fengq@gsau.edu.cn
[3] Wheat Research Institute, Gansu Academy of Agricultural Sciences, Lanzhou 730070, China; gsnklql@gsagr.ac.cn
* Correspondence: liucz@gsau.edu.cn; Tel.: +86-138-9321-8815

**Abstract:** Wheat is a significant cereal for humans, with diverse varieties. The growth of the wheat industry and the protection of breeding rights can be promoted through the accurate identification of wheat varieties. To recognize wheat seeds quickly and accurately, this paper proposes a convolutional neural network-based image-recognition method for wheat seeds, namely GC_DRNet. The model is based on the ResNet18 network and incorporates the dense network idea by changing its residual module to a dense residual module and introducing a global contextual module, reducing the network model's parameters and improving the network's recognition accuracy. Experiments were conducted on the self-constructed wheat-seed dataset and the publicly available dataset CIFAR-100 by combining GC_DRNet with network models such as ResNet18, ResNet34, ResNet50, and DenseNet121. The GC_DRNet model achieved a recognition accuracy of 96.98% on the wheat-seed dataset, which was improved by 2.34%, 1.43%, 2.05%, and 1.77% compared to ResNet18, ResNet34, ResNet50, and DenseNet121, respectively. On the CIFAR-100 dataset, the recognition accuracy of the GC_DRNet model was 80.77%, which improved the accuracy of ResNet18, ResNet34, ResNet50, and DenseNet121 by 8.19%, 1.6%, 9.59%, and 16.29%, respectively. Analyzing the confusion-matrix results of the wheat-seed dataset, the average recognition precision of the test set was 97.02%, the recall rate was 96.99%, and the F1 value was 96.98%. The parameter size of the GC_DRNet model was smaller than that of the other three models, only 11.65MB. The experimental results show that the GC_DRNet has a high level of recognition accuracy and detection capability for images of wheat seeds and provides technical support for wheat-seed identification.

**Keywords:** convolutional neural network; wheat seeds; image recognition; ResNet18; wheat-seed recognition model

check for updates

**Citation:** Xing, X.; Liu, C.; Han, J.; Feng, Q.; Lu, Q.; Feng, Y. Wheat-Seed Variety Recognition Based on the GC_DRNet Model. *Agriculture* **2023**, *13*, 2056. https://doi.org/10.3390/agriculture13112056

Academic Editor: Jiangbo Li

Received: 8 September 2023
Revised: 20 October 2023
Accepted: 24 October 2023
Published: 27 October 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

Wheat is a grass plant and a cereal crop widely grown worldwide. The earliest regions where wheat was cultivated can be traced back to the Mesopotamian region, and China is one of the countries that started planting wheat relatively early. As one of the three significant grains, almost all wheat is used for consumption, with only about one-sixth used as feed. Wheat is a major cereal crop in China and one of the most significant staple crops in Gansu Province [1]. Gansu Province is characterized by complex and diverse climatic conditions and ecological environments, as well as rich types of wheat varieties; so, wheat grows all year round in Gansu Province [2]. There are many varieties of wheat, and, for specific reasons, different varieties of wheat seeds may be mixed, resulting in low seed purity. It is, therefore, of great importance and value to use scientific technology to identify wheat varieties quickly and accurately.

The advancement of agricultural information technology has led to the widespread implementation of technologies like image processing and machine vision in the agricultural sector, such as monitoring crop pests and diseases, detecting defects in agricultural products, grading the quality of agricultural products, and recognizing crop varieties [3–8]. Niu et al. [9] used a combination of transfer learning and the DenseNet network to create a tomato-leaf disease-recognition model that successfully classified these diseases, achieving a test accuracy of up to 97.76%, which is significantly higher compared to a model that combines transfer learning with AlexNet, VGG, and MobileNet. Zhang et al. [10] proposed a method for peanut-pod grade image recognition based on a transfer-learning approach for five grades of peanut-pod images and constructed a network model (Penut_AlexNet model, PA model) suitable for peanut-pod grade recognition by using the AlexNet model for transfer learning. Through a series of parameter tuning, the final PA model obtained 95.43% accuracy in peanut-pod grade classification recognition. Yang and other scholars [11] improved the VGG16 network and applied it to 16 peanut varieties and seven corn-seed varieties. The recognition accuracy of peanut pods is 96.7%, and the recognition accuracy of corn seeds is 90.1%, which outperforms other classical convolutional neural networks in terms of performance.

From the above literature, we can know that crop-variety classification using convolutional neural networks (CNNs) is feasible, and it is capable of achieving high classification accuracy. However, it is worth noting that these models suffer from the challenges of large parameter sizes, long training times, and models that are not easy to use. Lu et al. [12] improved the ResNet model by proposing A-ResNet50 and A-ResNet101, which are soft-attention mechanism-based methods for small-sample hops pest and disease recognition. These two methods achieved 93.27% and 93.11% accuracy on the test set, which provides a useful reference for solving the problem of high-precision recognition on small-sample datasets. Lingwal et al. [13] identified 15 varieties of wheat grains using CNNs with an accuracy of 97.53%. Vidyarthi et al. [14] proposed to combine the deep convolutional neural networks (DCNN) with a soft attention mechanism to identify the pests of hops. DCNN was combined with image-processing technology to grade cashew kernels. The results show that the accuracy of both Inception-V3 and ResNet50 neural networks can reach 98.4%.

Considering the limitations of storage and computing power when deployed on mobile and embedded devices, the complexity of the model needs to be compressed as much as possible under the premise of guaranteeing the accuracy. Wang et al. [15] proposed an improved lightweight disease-recognition model called Multiscale ResNet. The model achieved an accuracy of 93.05% on seven disease image data collected in a real environment. It is worth noting that, although the accuracy of Multiscale ResNet decreased by about 3.72% relative to the ResNet18 model, this reduction came at the cost of reducing the training parameters by about 93% and shrinking the overall model size by about 35%, which provides a useful direction for the deployment of vegetable disease-recognition systems at the edge. Zhang [16] used 56 peach varieties as the research object and improved ResNet18. The improved model has a Top-1 accuracy of 94.4% and a model size of 14.35 MB, which improves the accuracy by 20% compared with the traditional SVM algorithm and improves the accuracy by 1.2% and 1% compared with mainstream networks such as VGG and MobileNetv3 and loses a very small amount of accuracy compared with ResNet18 in exchange for a threefold model reduction. Li et al. [17] used RegNet, a lightweight transfer-learning network, for the identification of five apple-leaf diseases in the field environment and compared the proposed method with networks, such as ShuffleNet, EfficientNet-B0, MobileNetV3, and Vision Transformer, and concluded that the transfer-learning-based method can identify apple-leaf diseases quickly and accurately.

As the number of model layers increases, the problem of gradient vanishing in CNN training leads to degradation of classification performance. ResNet [18] prevents gradient vanishing and overfitting by introducing residual cells and short connections but underutilizes the output information from the convolutional layers inside the residual cells

as well as the information transfer across the cells. DenseNet [19] proposes a denser connectivity approach that allows convolutional layers to be connected to each other, but only considers connections between layers within a unit and does not take full advantage of the multiple layers of features extracted from different units. Zhang et al. [20] proposed ResDenNet, which combines the advantages of ResNet and DenseNet while considering the connection between different units and the fusion of multilayer features to achieve better results in image super-resolution reconstruction. Gao [21] proposed a local dense residual network (LDRNet) to address the problem of redundancy of weight parameters caused by the widening and deepening of convolutional neural networks in the image-classification tasks, experimented with it on four datasets of different types and sizes, i.e., cifar10, cifar100, STL-10, and Flowers-17, and compared it with some classical networks. Sets were experimented with and compared with some classical networks. The results show that LDRNet achieves higher test accuracies on these datasets, proving its effectiveness and superiority in image-classification tasks. Xu [22] proposed a modified residual dense block convolution neural network (MRDB-CNN) for two important fundus image-research tasks: fundus image quality classification and fundus diabetic retinopathy (DR) discrimination. The experimental results show that the accuracy of the MRDB-CNN network is higher than other network structures for both fundus image quality classification and fundus DR discrimination, reaching 99.90% and 94.90% accuracy, respectively. Wei et al. [23] designed a residual dense network model suitable for hyperspectral image classification and achieved 98.71%, 99.31%, and 97.91% classification accuracy on Indian Pines data, University of Pavia data, and Salinas data, respectively.

In recent years, attention mechanisms have been the focus of extensive research, being introduced into CNNs to enhance their learning capabilities. In 2019, Cao et al. [24] proposed GCNet (global context network), whose core component is the global context block (GCBlock), which helps the network to better understand the relationships between objects and the overall context of an image by capturing global information in the image in an efficient way. Yan [25] improved the YOLOv3 target-detection model by introducing the GCNet attention module and the custom feature fusion module as a way to detect tea pests and diseases, and the experimental results show that the improved model can detect tea pests and diseases very well. Zhang [26] added GCBlock to ResNet and introduced Gram's Corner field, which fully extracted the time-associated features and long-distance dependence and achieved a 99.37% classification accuracy for arrhythmia signals. Using the SMS augmentation algorithm and two-way feature fusion, Zeng et al. [27] proposed a citrus huanglong disease-detection method for natural contexts, which also introduces a global-context module to establish effective long-range dependencies. The experimental results show that the method achieves an average accuracy of 84.8% and outperforms other target-detection algorithms such as SSD, RetinaNet, YOLO v3, YOLO v5s, Faster RCNN, and Cascade RCNN in terms of performance.

The above findings show that convolutional neural networks present excellent performance in image-classification tasks, with the ability to automatically learn and extract complex features from images. However, CNNs also have some shortcomings, including the problems of cumbersome network structure, a large number of parameters, and the inability to completely guarantee classification accuracy. It is also worth noting that residual dense networks have been widely used in hyperspectral image classification but are relatively less used in RGB image-classification tasks. Meanwhile, GCBlock is usually used in target-detection tasks, while its use is more limited in image-classification tasks, especially in the field of wheat-seed image recognition. Therefore, in view of the aforementioned problems, the aim of this paper is to find out the wheat-variety recognition model suitable for smart terminal deployment with few parameters, low computing power, and high classification accuracy. The main contributions of this paper are as follows:

1. We collected images of 29 mainstream wheat seeds in Gansu Province and prepared the data by screening and different image-preprocessing techniques to ensure the

quality and availability of the data and to provide sufficient training data for the model to learn effectively;

2. Since the residual dense network and global context module are less applied in the field of wheat-seed image recognition, we improve the network based on the ResNet18 network by combining the idea of a dense network in the residual module and introducing the global context module and propose the GC_DRNet network model for wheat-seed recognition;
3. Comparative experiments and analyses of the proposed GC_DRNet model with other classical networks on the self-constructed wheat-seed dataset and the public dataset CIFAR-100 have fully demonstrated the fast accuracy of the GC_DRNet model, which can successfully identify the varieties of wheat seeds.

In conclusion, the research objectives of this study are to develop an efficient model for wheat-seed image classification to address the challenges of convolutional neural networks in this task, to provide a solution with fewer parameters and suitable for smart terminal applications, to provide a research idea for wheat-variety recognition, and to provide technical support for practical wheat-seed recognition applications.

The remainder of this paper is structured as follows. The experimental wheat-image dataset, the methodology used in this paper, and the constructed network-recognition model are described in Section 2. The experimental setup and training of the model, as well as the analysis and evaluation of the experimental results, are described in Section 3. This is summarized in Section 4.

## 2. Materials and Methods

### *2.1. Data Analysis*

#### 2.1.1. Self-Built Dataset

The study described in this paper made use of a wheat-seed dataset that was gathered at the Tianshui Integrated Experiment Station of the Gansu Provincial Technology System for Wheat. The dataset was captured with a Nikon COOLPIX B700 (Nikon Corporation, Tokyo, Japan) digital camera in JPG format, with multiangle and multiscale images taken under natural outdoor lighting conditions. The filming took place between 15 July and 20 July 2021 and covered a wide range of weather conditions. All seeds were stored in the same environment after harvesting by natural air drying, packing, and packaging. Seed samples were removed without any physical or chemical treatment until the experiment. The moisture content ranged from 7.5% to 10% for all seeds during the emergence period.

A total of 29 seed images of the main winter-wheat varieties in Gansu Province were collected for the dataset [2]. Because of the natural morphological differences between seed singletons, no less than 900 seeds per variety were used in this study to construct the germplasm external morphological population phenotype dataset. Under natural outdoor-lighting conditions, blue cardboard was used as the carrier and background, the spacing between seeds was greater than twice the length of the seed, and 30 wheat seeds were collected in groups of five viewpoint photographs each. The dataset was based on 29 varieties $\times$ 900 kernels/variety $\div$ 30 kernels/group $\times$ 5 viewpoints, and a total of 4385 seed images were retained after filtering based on the quality of the photographs taken. For the experiment, 29 varieties of wheat were used; the dataset contains from 145 to 160 images of each variety. A wheat-seed image has a storage space of about 4 MB and a size of 5184 $\times$ 3888 pixels.

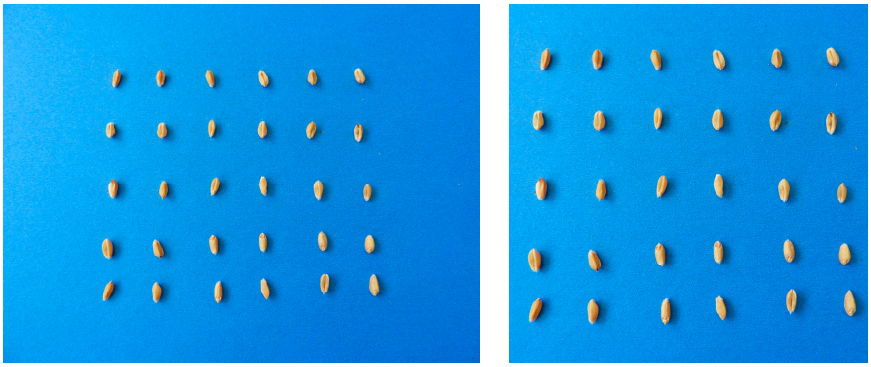
#### 2.1.2. Open Dataset CIFAR-100

The CIFAR-100 dataset comprises 60,000 images, which are categorized into 100 groups, each group comprising 600 images. The training set comprises 500 images per category, whereas the test set comprises 100 images per category. All images have a size of pixels. Each image in the CIFAR-100 dataset is labeled with a corresponding class label, making it suitable for classification tasks. The CIFAR-100 dataset is versatile and can be utilized for

various tasks, including feature extraction, object detection, and image segmentation, in addition to model training and evaluation.

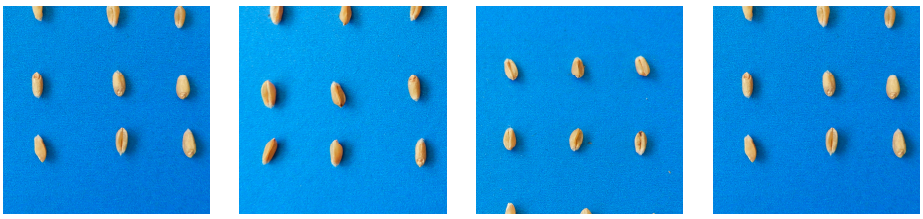
#### 2.1.3. Data Processing

In the actual training process, the original wheat-seed images are segmented in this article to prevent overfitting due to limited image quantity and improve the classification accuracy and robustness of the model. First, the image size is cropped to 3300 × 3300 and then divided into four parts. Figure 1 shows the original image and the cropped image, and Figure 2 shows the segmented image. After image segmentation, the number of images is increased by four times, totaling 17,540. To reduce model-training time, improve recognition efficiency, and meet the network's input requirements for image pixels, the image size is adjusted to the desired 224 × 224 sizes during training.



(**a**) (**b**)

**Figure 1.** The wheat-seed image: (**a**) original image; (**b**) cropped image.



(**a**) (**b**) (**c**) (**d**)

**Figure 2.** The segmented image: (**a**) the upper left portion of the image; (**b**) the lower left portion of the image; (**c**) the upper right portion of the image; and (**d**) the lower right portion of the image.

All the images are disordered and then divided into three sets in the ratio of 8:1:1, which are the training set, the validation set, and the test set. Among them, the training set and validation set are bravely used for model training and parameter tuning, and the test set is employed to evaluate the performance of each model.

### *2.2. Model Architecture*

#### 2.2.1. Convolutional Neural Networks

A convolutional neural network (CNN) is a kind of deep-learning neural network widely used in image processing, computer vision, and other fields [28]. Compared with traditional neural networks, CNN uses convolutional layers and pooling layers to abstract the local features of the image through convolution operations and impair the dimensionality of the feature maps through pooling operations. Meanwhile, CNN employs multiple layers of abstract representations to gradually extract high-level semantic information from images, thereby achieving image classification and recognition tasks. CNNs are essentially multilayer perceptrons with forward propagation, consisting of an input layer, convolutional layers, pooling layers, fully connected layers, and an output layer,

as shown in Figure 3. The input layer takes unprocessed data, such as images, text, or other formats. The convolutional layer uses multiple convolutional kernels to process the input data and extract the local features of the data. The pooling layer downsamples the output of the convolutional layer to decrease the feature map's dimensionality while preserving the map's primary information, thus reducing the computational workload of the model. The fully connected layer takes the output from the pooling layer and transforms it into a single, one-dimensional vector. This process involves flattening the feature map into a linear array of values that can be used as the input for the fully connected layer. The output layer is responsible for taking the output generated by the fully connected layer and producing a prediction based on that output. This prediction could take various forms depending on the specific task being performed, such as a classification or regression result. Usually, convolutional neural networks use multiple layers of convolutional and pooling layers to extract multilevel features, resulting in more efficient and accurate models [29].

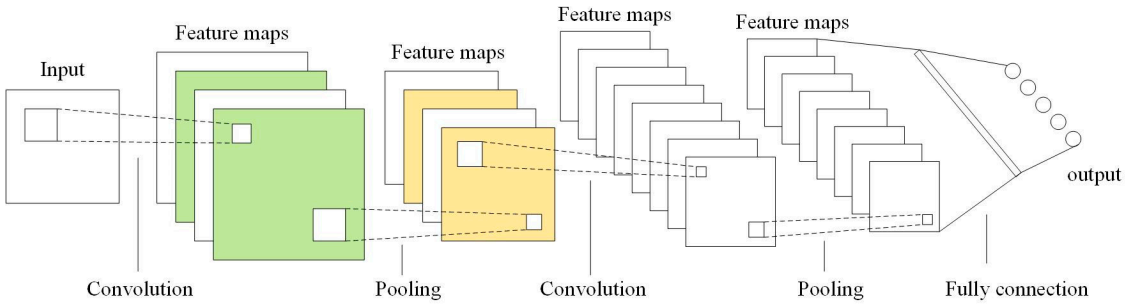



**Figure 3.** Convolution Neural Network architecture.

#### 2.2.2. ResNet18

Deep convolutional neural networks have made significant breakthroughs in computer vision. With an increase in the number of layers within a convolutional network, the extracted features become more prosperous and the training results become better. However, as deeper networks begin to converge, a problem of network degradation arises, where the training loss gradually decreases and then saturates, and when the depth of the network is increased, the training loss increases instead. In order to address this issue, He et al. [18] advanced the residual structure, which incorporates skip connections between layers to incorporate low-level features. The fundamental concept behind shortcut connections is to selectively combine the output of later layers with the output of earlier layers by adding them together so that the network can retain the previous layer's information during the training process, enabling better gradient propagation and solving the problem of gradient vanishing. The residual structure is shown in Figure 4. In residual structures, residual blocks are commonly used as basic units. Within each residual block, the first convolutional layer is employed to capture relevant features from the input data, the second convolutional layer maps the extracted features to the target dimension, and the shortcut connection directly adds the input to the output, forming a residual block. Stacking multiple residual blocks can create a more profound and powerful residual network. The input fed into the block, denoted by x, is added to the output of the block via the shortcut connection. The output of the block is H(x) = F(x) + x, where F(x) is the residual learned by the block. When F(x) is zero, H(x) reduces to x, an identity mapping. Therefore, the residual block learns the residual H(x) – x. Subsequent layers in the network approximate the residual to zero, enabling the model to improve its performance as the depth increases.

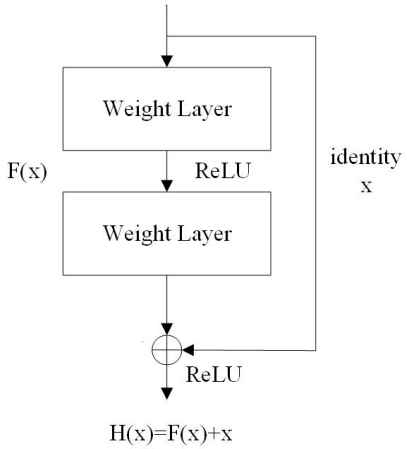



**Figure 4.** The residual structure.

This study opted to use ResNet18 as the underlying network architecture to strike a balance between achieving high accuracy and enabling real-time performance for wheat-variety recognition. As shown in Figure 5, the ResNet18 model consists mainly of a series of residual blocks, including one $7 \times 7$ convolutional layer, eight residual modules, and a fully connected layer. The input of ResNet18 is a $224 \times 224$ image, which is processed by convolutional layers to extract image features, and residual structures are used to preserve and pass on information from previous layers. Additionally, the model uses pooling layers to reduce feature map dimensionality without sacrificing essential features. At the end of the model, global average pooling compresses the feature map into a single-dimensional vector and maps it to the model output via fully connected layers. For tasks involving image classification, the neural network output layer often employs the softmax function to distribute output vectors over different classes, allowing the model to predict the wheat variety. Table 1 shows the parameters of the ResNet18 network structure. Although ResNet18 solves the problem of gradient vanishing, some issues remain, such as single feature extraction size and a high number of trainable parameters. Therefore, the model undergoes a series of improvements.

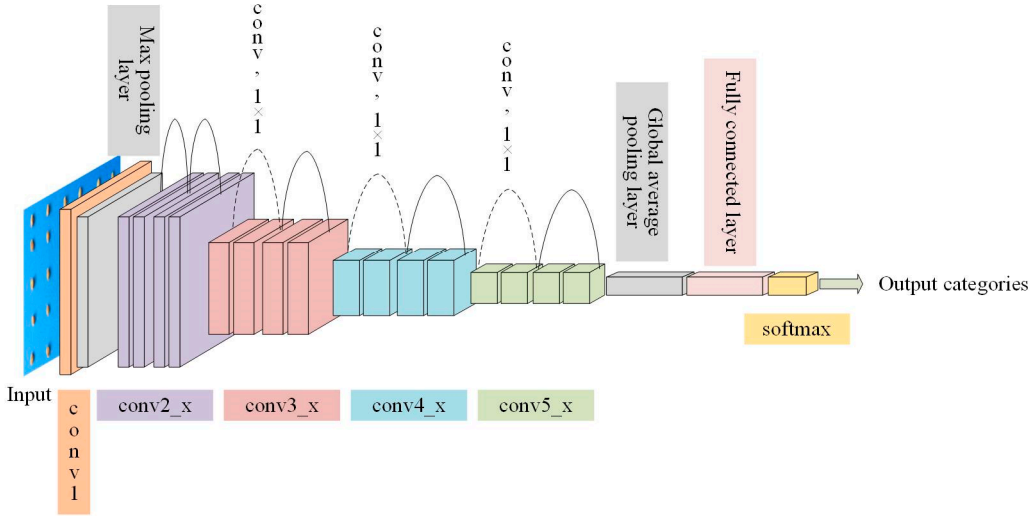



**Figure 5.** ResNet18 network architecture.

**Table 1.** ResNet18 network parameters.

| Layer Name | ResNet18 | Output Size |
|---|---|---|
| conv1 | $7 \times 7$, 64, stride 2 | $112 \times 112$ |
| conv2_x | $3 \times 3$ max pool stride 2<br>$\begin{pmatrix} 3\times3, 64 \\ 3\times3, 64 \end{pmatrix} \times 2$ | $56 \times 56$ |
| conv3_x | $\begin{pmatrix} 3\times3, 128 \\ 3\times3, 128 \end{pmatrix} \times 2$ | $28 \times 28$ |
| conv4_x | $\begin{pmatrix} 3\times3, 256 \\ 3\times3, 256 \end{pmatrix} \times 2$ | $14 \times 14$ |
| conv5_x | $\begin{pmatrix} 3\times3, 512 \\ 3\times3, 512 \end{pmatrix} \times 2$ | $7 \times 7$ |
| | Average pool, fc, softmax | $1 \times 1$ |

#### 2.2.3. Dense Residual Block

DenseNet is an improvement on the residual network proposed by Huang et al. [19], which addresses some issues in the residual network. It proposes a new type of connection called the concatenation connection. In this connection, the features of each layer are concatenated with the features of all the preceding layers and passed to the next layer. Figure 6 shows the structure of the dense block, which optimizes information flow between different network layers. Assuming that the DenseNet has n layers, "[]" represents the concatenate, and the output expression of the nth layer is:

$$x_{\mathrm{n}} = F([x_0, x_1, \cdots, x_{n-1}]) \tag{1}$$

In Equation (1), $x_n$ denotes the output feature vector of the nth layer, and $F(\cdot)$ denotes the function mapping.

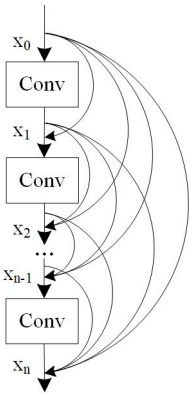



**Figure 6.** The dense block structure.

This paper proposes an improvement to the residual module in ResNet18 by incorporating the idea of DenseNet, resulting in a dense residual block (DRB). Figure 7 illustrates the architecture of the DRB, where the input features are first convolved and then concatenated together, and the weights for each channel are adjusted using $1 \times 1$ convolution. Eventually, identity connection is used to add the input and output features. The DRB can extract image features more comprehensively and partially alleviate the problem of gradient disappearance during training. By connecting features from different layers using concatenation, the DRB promotes feature reuse and reduces network parameters and computational costs achieving good wheat-seed image-classification performance.

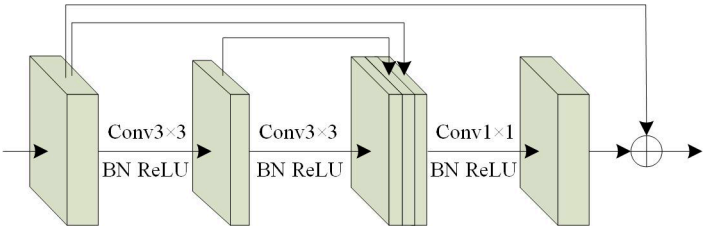



**Figure 7.** The DRB structure.

#### 2.2.4. The Global Context Module

Capturing long-range dependencies has been proven to be beneficial in computer vision tasks. The purpose of capturing long-range relationships or correlations is to comprehensively comprehend the visual context. In convolutional neural networks, the convolutional layers establish contextual relationships in local regions, so it is necessary to continuously stack convolutional layers to obtain global contextual information. However, this approach increases computation and makes it challenging to optimize the network. The solution to this problem, the nonlocal network, was present [30], which can seize long-range dependencies and model them through attention mechanisms. Although the nonlocal network avoids stacking convolutional layers, it has a high computational cost. The global context module [24] combines a streamlined variation of the nonlocal network and SENet [31] (squeeze-excitation network), which has fewer computations and can integrate global information well. Figure 8 illustrates the nonlocal network and SENet structures.

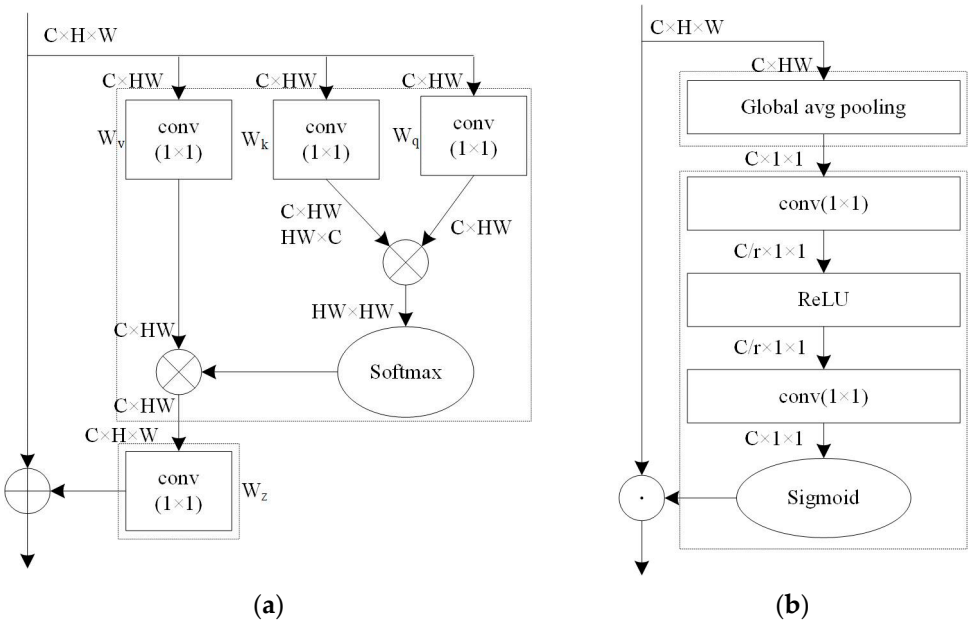



**Figure 8.** (**a**) NLNet structure; (**b**) SENet structure.

The global context block consists of three operations: (1) global attention—a global attention mechanism is used to model the context information. This is achieved by computing self-attention weights using a $1 \times 1$ convolutional layer $W_k$ and a softmax function. The global contextual features are then obtained by attention pooling from the attention weights. (2) Feature transformation—the obtained global contextual features are transformed to capture channel dependencies using a $1 \times 1$ convolutional layer. (3) Feature fusion:—the transformed global contextual features are merged with the input features by element-wise addition, producing the final output of the GCBlock. The overall goal of the GC module is the capture of global context information and the enhancement of the representational capacity of input features for better performance in various computer vision tasks. The GCBlock can be seen in Figure 9.

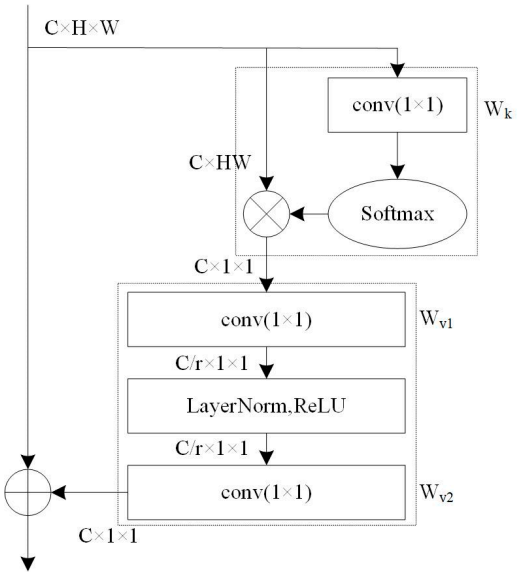



**Figure 9.** The GCBlock structure.

The GCBlock is a lightweight module that can be flexibly inserted into any neural network structure. In this paper, the GCBlock is incorporated into the improved dense residual block to effectively extract global context information, obtain richer shallow and deep features, and improve the network's generalization ability during training. The dense residual block with the introduced global context module (GC_DRBlock) is demonstrated in Figure 10.

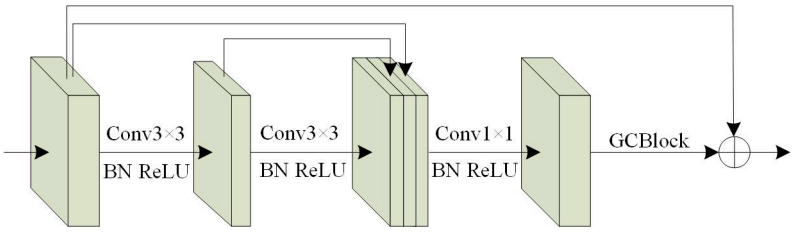



**Figure 10.** The GC_DRBlock structure.

#### 2.2.5. GC_DRNet Model Architecture

The whole structure of the designed wheat-seed classification model, GC_DRNet, is depicted in Figure 11, which consists of input, output, dense residual blocks, transition layers, and a softmax classifier. First, the network receives the input image, and shallow features of the wheat-seed image are extracted using a $7 \times 7$ convolutional layer. Then, the wheat-seed image features obtained through max pooling are input to the GC_DRBlock, which extracts deep image features using dense and skip connections. The next step is to construct a transition layer. This is done by $1 \times 1$ convolution and average pooling to connect two adjacent GC_DR blocks. The $1 \times 1$ convolutional operation adjusts the feature map's channels. In contrast, to downsample the feature map and reduce computational complexity, the average pooling operation is used. Ultimately, the features that have been extracted are subjected to global average pooling. This is followed by a fully connected layer and the softmax classifier, in order to produce the final classification output. The GC_DRNet network-structure parameters are listed in Table 2.

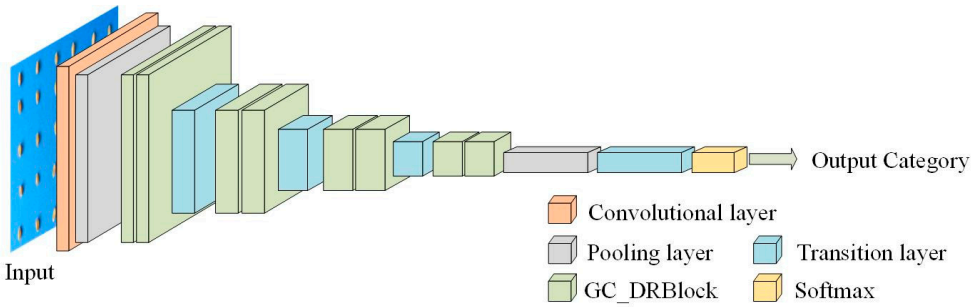



**Figure 11.** The GC_DRNet structure.

**Table 2.** GC_DRNet network parameters.

| Layer Name | GC_DRNet | Output Size |
|---|---|---|
| convolution | $7 \times 7$, 64, stride 2 | $112 \times 112$ |
| pooling | $3 \times 3$ max pool, stride 2 | $56 \times 56$ |
| GC_DRBlock (1) | $\begin{pmatrix} 3 \times 3 \text{ conv} \\ 3 \times 3 \text{ conv} \\ 1 \times 1 \text{ conv} \\ \text{GCBlock} \end{pmatrix} \times 2$ | $56 \times 56$ |
| Transition Layer (1) | $1 \times 1$ conv<br>$2 \times 2$ avgpool, stride 2 | $28 \times 28$ |
| GC_DRBlock (2) | $\begin{pmatrix} 3 \times 3 \text{ conv} \\ 3 \times 3 \text{ conv} \\ 1 \times 1 \text{ conv} \\ \text{GCBlock} \end{pmatrix} \times 2$ | $28 \times 28$ |
| Transition Layer (2) | $1 \times 1$ conv<br>$2 \times 2$ avgpool, stride 2 | $14 \times 14$ |
| GC_DRBlock (3) | $\begin{pmatrix} 3 \times 3 \text{ conv} \\ 3 \times 3 \text{ conv} \\ 1 \times 1 \text{ conv} \\ \text{GCBlock} \end{pmatrix} \times 2$ | $14 \times 14$ |
| Transition Layer (3) | $1 \times 1$ conv<br>$2 \times 2$ avgpool, stride 2 | $7 \times 7$ |
| GC_DRBlock (4) | $\begin{pmatrix} 3 \times 3 \text{ conv} \\ 3 \times 3 \text{ conv} \\ 1 \times 1 \text{ conv} \\ \text{GCBlock} \end{pmatrix} \times 2$ | $7 \times 7$ |
| | Average pool, fc, softmax | $1 \times 1$ |

## 3. Experimental Design and Results Analysis

### *3.1. Experimental Environment*

The computer environment used in this experiment is the Windows 11 system, with an AMD Ryzen 7 5800H processor and an NVIDIA GeForce RTX 3060 GPU. Table 3 gives the specific configuration of the experimental environment.

**Table 3.** Experimental environment configuration.

| Experimental Environment | Configuration Parameters |
|---|---|
| Operating system | Windows 11 |
| Processor | AMD Ryzen 7 5800H |
| GPU | NVIDIA GeForce RTX 3060 |
| Code-management software | Pycharm 2020.1.3 |
| Programming language | Python 3.8 |
| Deep-learning framework | Pytorch 1.12.1 |
| GPU-acceleration library | CUDA 11.6.134 |

### *3.2. Parameter Setting*

During the training of the CNN model, this paper used the stochastic gradient descent (SGD) algorithm with momentum (0.9) and a weight decay coefficient of $1 \times 10^{-4}$ to optimize the model. Cross entropy loss was used to calculate the loss value. The batch size was set to 32, the number of epochs was set to 100, and the initial learning rate was set to 0.001. The cosine annealing decay strategy was used to update the learning rate.

### *3.3. Evaluation Metrics*

Accuracy, precision, recall, F1 score, and confusion matrix are the main metrics used in this article to evaluate the model. Accuracy is the percentage of samples that were correctly classified out of the entire test set. Precision is defined as the percentage of true positive predictions out of the total number of positive predictions. Recall is the predicted true positives as a percentage of actual positive specimens. The F1 score is a metric that balances precision and recall.

The formulas for each index are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{3}$$

$$Recall = \frac{TP}{TP + TN} \times 100\% \tag{4}$$

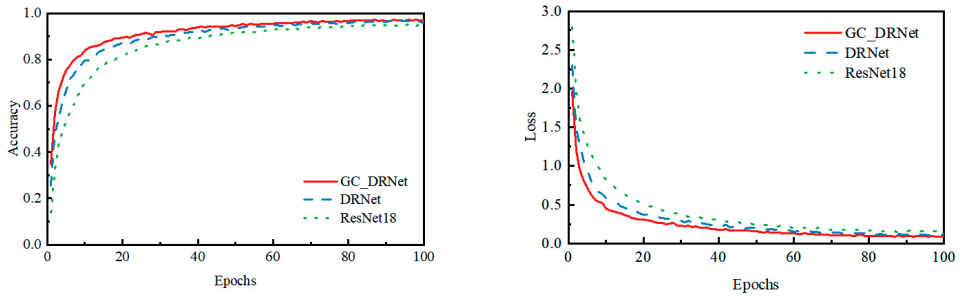
$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{5}$$

where $TP$ stands for true positive, $TN$ stands for true negative, $FP$ stands for false positive, and $FN$ stands for false negative.

### *3.4. Results and Analysis*

#### 3.4.1. Analysis of Wheat-Seed Dataset Ablation Experiment Results

We perform comparison experiments using GC_DRNet, DRNet, and ResNet18 networks on the wheat dataset to check the recognition capability of the optimized model GC_DRNet. Figure 12 depicts the loss and accuracy curves of the three models on the training set. The training loss curve shows that GC_DRNet has the fastest decay of the loss value at the beginning of training, followed by DRNet, and ResNet18 has the slowest. The accuracy curve shows that GC_DRNet achieves the highest recognition accuracy of wheat-seed images, followed by DRNet, and ResNet18 has the lowest accuracy. All models converge almost entirely around 70 epochs.




(**a**) (**b**)

**Figure 12.** Model training loss and accuracy curve graph: (**a**) Accuracy of models; (**b**) Loss of models.

Table 4 shows the results of ablation experiments based on ResNet18. The highest accuracy of GC_DRNet, DRNet, and ResNet18 on the training set was 97.07%, 96.64%, and

94.97%, as seen from the table, respectively. When applied to the test set, the wheat-seed image-recognition accuracies of GC_DRNet, DRNet, and ResNet18 were 96.98%, 96.52%, and 94.64%, respectively. Compared with the original ResNet18 model, GC_DRNet showed a 2.34% improvement in recognition accuracy on the test set, and DRNet showed a 1.88% improvement. The ResNet18 network model was 44.59 MB in parameters and $1.82 \times 10^9$ in float-point operation. The GC_DRNet model, which replaced the residual modules of ResNet18 with dense residual modules, explicitly distinguished the network's information and the preserved information, eliminating the need to relearn redundant feature maps and requiring each layer to learn very few features. This significantly reduces the model's parameter size and computational cost, reducing the computational and memory demands of the network model.

**Table 4.** Results of ablation experiments based on ResNet18 network.

| Model | Training Set Highest Accuracy/% | Accuracy of Test Set/% | Parameter/MB | Flops/$10^9$ |
|---|---|---|---|---|
| ResNet18 | 94.97 | 94.64 | 44.59 | 1.82 |
| DRNet | 96.64 | 96.52 | 10.30 | 0.87 |
| GC_DRNet | 97.07 | 96.98 | 11.65 | 0.87 |

The experiments demonstrate that replacing the residual modules in the ResNet18 network with dense residual modules can increase the diversity and richness of feature information, enhance the network model's wheat-seed image-identification accuracy, and effectively alleviate the issue of vanishing gradients during the training process. At the same time, introducing the global context module into the improved dense residual blocks can effectively extract global contextual information and raise the network model's ability to make accurate predictions on unseen data.

3.4.2. Comparison and Analysis of Different Convolutional Neural Network Models

We compared its training results on the wheat-seed image dataset with commonly used image-classification models, ResNet34, ResNet50, and DenseNet121. In order to ensure the experimental results are reliable, we trained the comparative models with the same network parameters as the GC_DRNet model. The accuracy curves of wheat-seed image recognition on the training set can be seen in Figure 13.

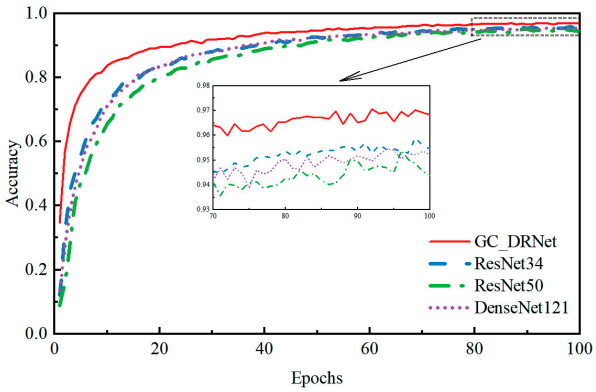



**Figure 13.** Different Network Models Training Accuracy Curve.

This figure shows that this improved the GC_DRNet model and achieved the best recognition results on the wheat-seed images. By replacing the residual module of the original ResNet18 model with a dense residual module and introducing a global context module in the dense residual module, the diversity of wheat-seed information was increased, and the fusion of global information was enhanced, thereby improving the effectiveness of the model.

Table 5 presents the training results of the four models, including their highest training-set accuracy, test-set accuracy, and model-training parameters.

**Table 5.** Comparison of experimental results among different network models.

| Model | Training Set Highest Accuracy/% | Accuracy of Test Set/% | Parameter/MB | Flops/$10^9$ |
|---|---|---|---|---|
| ResNet34 | 95.90 | 95.55 | 83.15 | 3.67 |
| ResNet50 | 95.37 | 94.93 | 97.49 | 4.12 |
| DenseNet121 | 95.44 | 95.21 | 30.44 | 2.88 |
| GC_DRNet | 97.07 | 96.98 | 11.65 | 0.87 |

The table presents all the network models that achieved a recognition accuracy of over 90% on the wheat-seed image dataset. The GC_DRNet network model achieved the highest test accuracy among the four models, at least 2.05% higher than the other three models. After 100 epochs of iterative training, the GC_DRNet network model achieved a training set with the highest accuracy of 97.07% and a test set accuracy of 96.98%, which is 1.17% and 1.43% higher than the ResNet34 model's training set highest accuracy and test-set accuracy, respectively. The ResNet50 network model performed with a 94.93% accuracy on the test set, which is 2.05% lower than the GC_DRNet model's test accuracy. The DenseNet121 network model had a recognition accuracy of 1.77% lower than the GC_DRNet model, with a recognition accuracy of 95.21%. Although ResNet34, ResNet50, and DenseNet121 have deeper network layers than GC_DRNet, the parameters and computational cost of GC_DRNet are smaller than these three models. The parameter size of the GC_DRNet model is 11.65 MB, and the floating-point operation is $8.67 \times 10^8$, indicating that this network model was less expensive to compute than the other three.

After comparing the performance of four different convolutional neural network models, including ResNet34, ResNet50, DenseNet121, and the proposed GC_DRNet, we can conclude that the GC_DRNet is superior to the other three models with respect to test accuracy and training efficiency. Moreover, the GC_DRNet achieved a higher accuracy while having fewer parameters and requiring fewer floating-point operations than the ResNet50 and DenseNet121 models. Therefore, the GC_DRNet is a more effective and efficient model for classifying wheat-seed images.

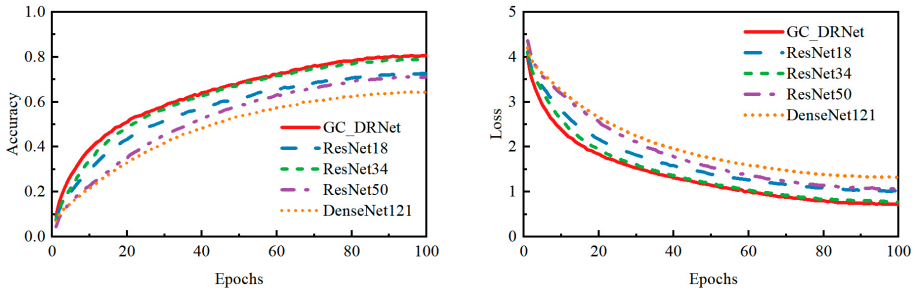
#### 3.4.3. Analysis of Experimental Results on Public Dataset CIFAR-100

In order to avoid the influence of irrelevant features, such as blue background and seed spacing, on the classification accuracy of the experiments, comparative experiments with ResNet18, ResNet34, ResNet50, and DenseNet121 on the CIFAR-100 public dataset were conducted to validate the classification performance of the GC_DRNet model proposed in this paper. The training accuracy and loss value curves obtained are shown in Figure 14. The figure illustrates that, after 100 epochs of training, the proposed GC_DRNet model achieved the best classification performance on the CIFAR-100 dataset, with a training-set accuracy of 80.77% and a loss value of 0.73. Compared with the other four classification-network models, the training-set accuracy was improved by 8.19%, 1.6%, 9.59%, and 16.29%, and the loss value was reduced by 0.29, 0.05, 0.34, and 0.59, respectively.

#### 3.4.4. Performance Evaluation of GC_DRNet Model

The confusion matrix intuitively reflects a neural network model's classification-recognition ability and is an essential indicator for evaluating network models [32]. In this paper, the improved model GC_DRNet was tested on a test set of 29 wheat-seed varieties, and the classification-confusion matrix obtained is shown in Figure 15. The correct category of wheat-seed varieties is shown on the horizontal axis, while the predicted category is shown on the vertical axis. The bar chart on the right side displays a heat map of the predicted quantity of wheat seeds. Confusion-matrix diagonals represent correctly classified example quantities, while off-diagonals represent incorrectly classified

example quantities. From the confusion matrix, it can be seen that the main diagonal is darker, which indicates that the GC_DRNet model has good recognition performance for 29 different varieties of wheat-seed images. The best recognition performance was achieved for Lantian15, Lantian42, and Lantian45, with all test images being correctly predicted. The recognition performance of Jimai22, Jimai44, Lantian35, Lantian43, Lantian48, and Zhoumai20 was poor, with Zhoumai20 having the lowest recognition accuracy, with only 53 of the 60 actual samples being correctly predicted. The next worst was Jimai44, with 5 of the 62 actual samples being incorrectly predicted. By analyzing the confusion matrix, Zhoumai20 is most easily misclassified as Zhoumai23, and Jimai44 is easily misclassified as Jimai47. The easy misclassification may be because Zhoumai20 and Zhoumai23, Jimai44, and Jimai47 belong to the same large category of varieties, and their appearance has certain similarities, which increases the difficulty of classification for the network model. Overall, this paper's improved GC_DRNet network model has good classification performance on the wheat-seed image dataset and can accurately identify wheat-seed varieties.




(**a**) (**b**)

**Figure 14.** The training accuracy and loss curves of each model on CIFAR-100: (**a**) Accuracy of models; (**b**) Loss of models.

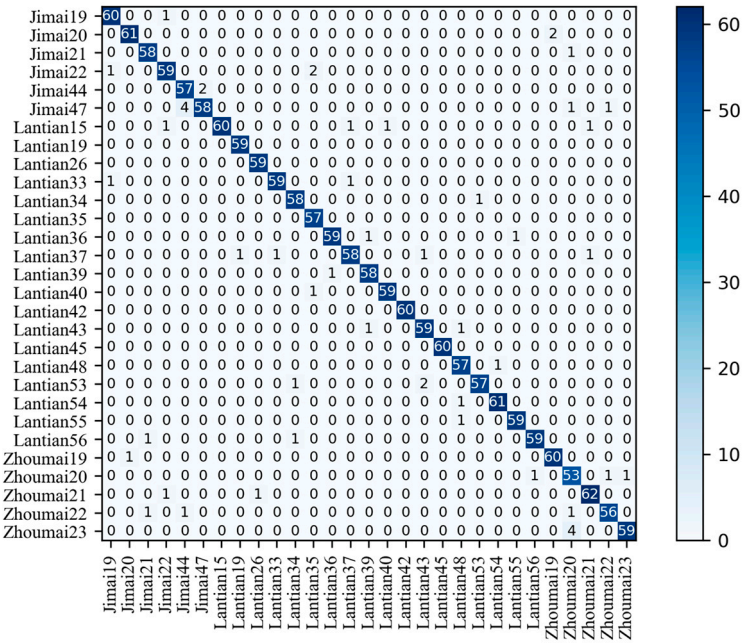



**Figure 15.** Confusion matrix of GC_DRNet model on test set.

The precision, recall, and F1 score of each wheat-seed variety can be calculated according to the confusion matrix. Table 6 gives the results of the calculations.

**Table 6.** Wheat-seed image-recognition results on the GC_DRNet model.

| Wheat-Seed Variety | TP | FP | FN | TN | Precision | Recall | F1 Score |
|---|---|---|---|---|---|---|---|
| Jimai19 | 60 | 1 | 2 | 1691 | 98.36 | 96.77 | 97.56 |
| Jimai20 | 61 | 2 | 1 | 1690 | 96.83 | 98.39 | 97.60 |
| Jimai21 | 58 | 1 | 2 | 1693 | 98.31 | 96.67 | 97.48 |
| Jimai22 | 59 | 3 | 3 | 1689 | 95.16 | 95.16 | 95.16 |
| Jimai44 | 57 | 2 | 5 | 1690 | 96.61 | 91.94 | 94.22 |
| Jimai47 | 58 | 6 | 2 | 1688 | 90.62 | 96.67 | 93.55 |
| Lantian15 | 60 | 4 | 0 | 1690 | 93.75 | 100.00 | 96.77 |
| Lantian19 | 59 | 0 | 1 | 1694 | 100.00 | 98.33 | 99.16 |
| Lantian26 | 59 | 0 | 1 | 1694 | 100.00 | 98.33 | 99.16 |
| Lantian33 | 59 | 2 | 1 | 1692 | 96.72 | 98.33 | 97.52 |
| Lantian34 | 58 | 1 | 2 | 1693 | 98.31 | 96.67 | 97.48 |
| Lantian35 | 57 | 0 | 3 | 1694 | 100.00 | 95.00 | 97.44 |
| Lantian36 | 59 | 2 | 1 | 1692 | 96.72 | 98.33 | 97.52 |
| Lantian37 | 58 | 4 | 2 | 1690 | 93.55 | 96.67 | 95.08 |
| Lantian39 | 58 | 1 | 2 | 1693 | 98.31 | 96.67 | 97.48 |
| Lantian40 | 59 | 1 | 1 | 1693 | 98.33 | 98.33 | 98.33 |
| Lantian42 | 60 | 0 | 0 | 1694 | 100.00 | 100.00 | 100.00 |
| Lantian43 | 59 | 2 | 3 | 1690 | 96.72 | 95.16 | 95.93 |
| Lantian45 | 60 | 0 | 0 | 1694 | 100.00 | 100.00 | 100.00 |
| Lantian48 | 57 | 1 | 3 | 1693 | 98.28 | 95.00 | 96.61 |
| Lantian53 | 57 | 3 | 1 | 1693 | 95.00 | 98.28 | 96.61 |
| Lantian54 | 61 | 1 | 1 | 1691 | 98.39 | 98.39 | 98.39 |
| Lantian55 | 59 | 1 | 1 | 1693 | 98.33 | 98.33 | 98.33 |
| Zhoumai56 | 59 | 2 | 1 | 1692 | 96.72 | 98.33 | 97.52 |
| Zhoumai19 | 60 | 1 | 2 | 1691 | 98.36 | 96.77 | 97.56 |
| Zhoumai20 | 53 | 3 | 7 | 1691 | 94.64 | 88.33 | 91.38 |
| Zhoumai21 | 62 | 2 | 2 | 1688 | 96.88 | 96.88 | 96.88 |
| Zhoumai22 | 56 | 3 | 2 | 1693 | 94.92 | 96.55 | 95.73 |
| Zhoumai23 | 59 | 4 | 1 | 1690 | 93.65 | 98.33 | 95.93 |

The table demonstrates that, when analyzing only precision, the performance of Jimai47 is the worst at 90.62%; when analyzing only recall, the performance of Zhoumai20 is the worst at 88.33%; when analyzing the F1 score, the performance of Lantian42 and Lantian45 is the best, while Zhoumai20 has the worst performance at 91.38%. The results of ranking the classification, the accuracy, and other performance metrics achieved by the models in this research based on the F1 value are Lantian45, Lantian42, Lantian26, Lantian19, Lantian54, Lantian55, Lantian40, Jimai20, Zhoumai19, Jimai19, Lantian56, Lantian36, Lantian33, Lantian39, Lantian34, Jimai21, Lantian35, Zhoumai21, Lantian15, Lantian53, Lantian48, Zhoumai23, Lantian43, Zhoumai22, Jimai22, Lantian37, Jimai44, Jimai47, and Zhoumai20. Overall, the recognition precision of the test set is 97.02%, the recall rate is 96.99%, and the F1 score is 96.98%. The results demonstrate that the proposed model in this paper achieves strong classification performance on the wheat-seed image dataset.

Some of the wheat seeds were recognized, as shown in Figure 16, and all the inference results were correct.

#### 3.4.5. Feature-Map Visualisation

Feature maps are core components in convolutional neural networks used to represent different layers of input data features. They play a crucial role in image processing and deep learning, enabling the network to effectively understand and process complex visual information. Figure 17 illustrates the feature maps of the first 10 layers of the GC_DRNet model. From the figure, it can be seen that, as the layers of the network increase, the feature map gradually captures higher level features, from simple edges to more abstract parts of the object. Typically, convolutional layers close to the input are more likely to capture image details and low-level features, such as edges, textures, colors, etc. The feature maps of

these layers usually have a higher spatial resolution and, therefore, provide more detailed information about the image. This detailed information is important for identifying local features in the image. As the depth of the convolutional layers increases, the network tends to capture more abstract, high-level features, such as parts of objects, patterns, overall shape, etc. The feature maps in these layers typically have lower spatial resolution, but they play a key role in image-classification tasks. The later layers focus more on the global structure and context of the image.

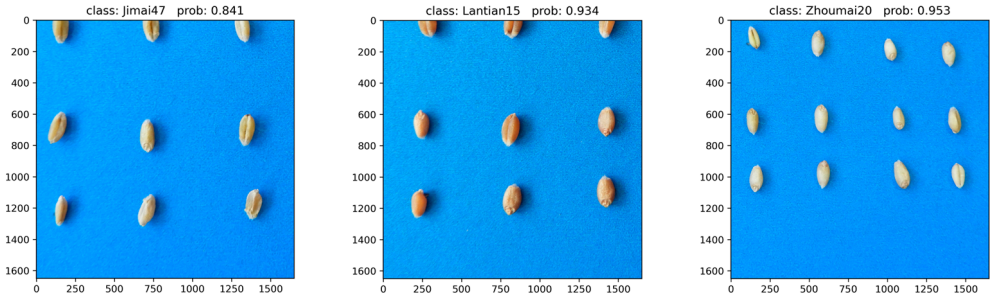



**Figure 16.** Selected wheat-seed classification results: (**a**) Jiami47; (**b**) Lantian15; (**c**) Zhoumai20.

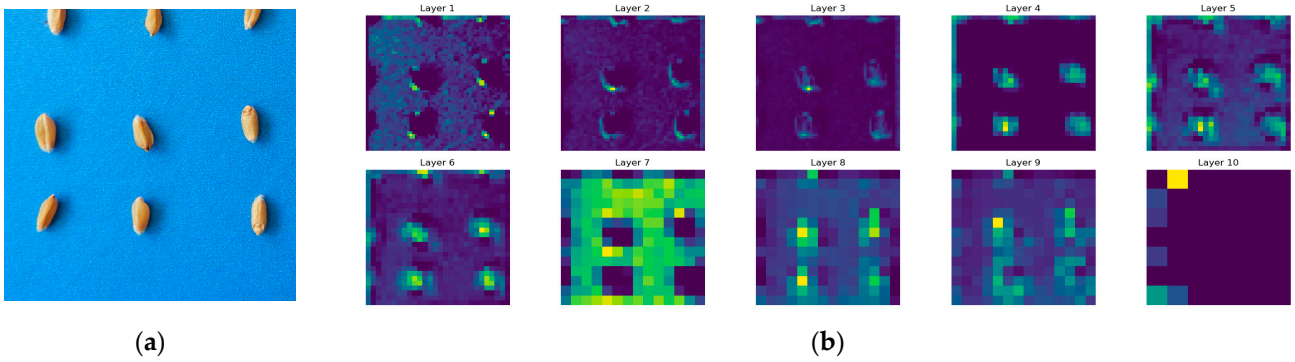



**Figure 17.** Feature maps of the first 10 layers of the GC_DRNet model: (**a**) original image; (**b**) feature maps.

## 4. Conclusions

In this paper, a deep learning variety-classification model GC_DRNet for 29 wheat-seed image datasets is established based on a convolutional neural network. Through a series of comparison experiments, the following conclusions are drawn:

(1) On the self-constructed wheat-seed dataset, the recognition accuracy of this model is improved by 2.34% and the amount of parameters is reduced by 73.87%, compared with the original ResNet18 network model. By comparing the ResNet34, ResNet50, and DenseNet121 networks, the GC_DRNet model has more significant advantages in terms of recognition accuracy and parameter quantity, with 1.43%, 2.05%, and 1.77% improvement in recognition accuracy and 85.98% and 88.05% reduction in parameter quantity, respectively, on the test set, 61.73%. It indicates that the introduction of the dense network idea in the residual block of the ResNet18 network model to achieve feature reuse can effectively reduce the network parameters and computational cost, and improve the recognition accuracy of the network model; the introduction of global context information in the improved dense residual module fully extracts the global information, increases the differences between different wheat-seed varieties, and can effectively increase the network model's wheat-seed recognition accuracy. In addition, the recognition accuracy of the model is 96.98%, the number of references is

only 11.65 MB, and the floating-point operation is $8.67 \times 10^8$, which indicates that the model in this paper can be embedded in smart terminals;

(2) On the publicly available dataset CIFAR-100, the recognition accuracy of the GC_DRNet network model is 80.77%, which is 8.19%, 1.6%, 9.59%, and 16.29% higher than that of the ResNet18, ResNet34, ResNet50, and DenseNet121 models, respectively. It indicates that features such as blue background and seed spacing have no effect on model robustness.

Going forward, our main areas of focus will be (1) expanding the scope of wheat varieties in the dataset studied to better reflect the market demand; (2) optimizing the proposed model further, we plan to investigate and incorporate other state-of-the-art neural network models, such as VoVNet [33], Swin Transformer [34], and HRNet [35], into the training process; and (3) considering deploying the model on mobile devices such as smartphones to achieve real-time wheat variety classification.

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

**Funding:** This work was supported by the National Natural Fund Project of China (Grant 32360437), the Innovation Fund Project of Colleges and Universities in Gansu of China (Grant 2021A-056), and the Industrial Support and Guidance Project of Universities in Gansu Province of China (Grant 2021CYZC-57).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We thank the editor and reviewers for their helpful suggestions to improve the quality of this manuscript.

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

## References

1. Liu, G.; Zhao, G.; Li, B.; You, Y.; Zhou, D. *Current Situation and Development Countermeasures of the Wheat Industry in Gansu Province*; Gansu Agricultural Science and Technology: Lanzhou, China, 2020; pp. 70–75.
2. Lu, Q.; Ma, Z.; Yang, W.; Zhang, K.; Zhang, L.; Cao, S.; Zhang, W.; Zhou, G. *Current Status and Countermeasures of Wheat Breeding in Gansu*; Gansu Agricultural Science and Technology: Lanzhou, China, 2022; Volume 53, pp. 1–5.
3. Yang, S.; Feng, Q.; Zhang, J.; Wang, G.; Zhang, P.; Yan, H. Non-destructive grading of potato external defects based on a lightweight convolutional network. *Food Sci.* **2020**, *42*, 284–289.
4. Liu, C.; Han, J.; Chen, B.; Mao, J.; Xue, Z.; Li, S. A Novel Identification Method for Apple (*Malus domestica* Borkh.) Cultivars Based on a Deep Convolutional Neural Network with Leaf Image Input. *Symmetry* **2020**, *12*, 217. [CrossRef]
5. Jeyaraj, P.R.; Asokan, S.P.; Samuel Nadar, E.R. Computer-Assisted Real-Time Rice Variety Learning Using Deep Learning Network. *Rice Sci.* **2022**, *29*, 489–498. [CrossRef]
6. Du, G.; Lu, M.; Ji, Z.; Liu, J. Research on Apple Defect Detection Method Based on Improved CNN. *Food Mach.* **2023**, *39*, 155–160.
7. Altuntas, Y.; Comert, 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]
8. Javanmardi, S.; Ashtiani, S.H.M.; 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]
9. Niu, X.; Gao, B.; Nan, X.; Shi, Y. Tomato Leaf Disease Detection Based on Improved DenseNet Convolutional Neural Network. *Jiangsu J. Agric. Sci.* **2022**, *38*, 129–134.
10. Zhang, R.; Li, Z.; Hao, J.; Sun, L.; Li, H.; Han, P. Convolutional Neural Network Peanut Pod Rank Image Recognition Based on Migration Learning. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 171–180.
11. Yang, H.; Ni, J.; Gao, J.; Han, Z.; Luan, T. A novel method for peanut variety identification and classification by Improved VGG16. *Sci. Rep.* **2021**, *11*, 15756. [CrossRef]
12. Lu, Y.; Chen, B. Identification of hops pests and diseases in small samples based on attentional mechanisms. *J. Chin. Agric. Mech.* **2020**, *42*, 189–196.

13. Lingwal, S.; Bhatia, K.K.; Tomer, M.S. Image-based wheat grain classification using convolutional neural network. *Multimed. Tools Appl.* **2021**, *80*, 35441–35465. [CrossRef]
14. Vidyarthi, S.K.; Singh, S.K.; Tiwari, R.; Xiao, H.W.; Rai, R. Classification of first quality fancy cashew kernels using four deep convolutional neural network models. *J. Food Process Eng.* **2020**, *43*, e13552. [CrossRef]
15. Wang, C.; Zhou, J.; Wu, H.; Teng, G.; Zhao, C.; Li, J. Identification of vegetable leaf diseases based on improved Multi-scale ResNet. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 209–217.
16. Zhang, F. *Research on Peach Variety Recognition Technology Based on Machine Vision*; Hebei Agricultural University: Baoding, China, 2021.
17. Li, L.; Zhang, S.; Wang, B. Apple Leaf Disease Identification with a Small and Imbalanced Dataset Based on Lightweight Convolutional Networks. *Sensors* **2021**, *22*, 173. [CrossRef] [PubMed]
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
19. Huang, G.; Liu, Z.; Maaten, L.v.d.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
20. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual Dense Network for Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.
21. Gao, X. *A Study on the Improvement of Convolutional Neural Networks and Their Application in Image Classification*; Yanshan University: Qinhuangdao, China, 2022.
22. Xu, X. *Quality Classification and Lesion Discrimination of DR Fundus Images Based on Residual-Dense Networks*; Tianjin University of Technology: Tianjin, China, 2021.
23. Wei, X.; Yu, X.; Tan, X.; Liu, B. Hyperspectral Image Classification Using Residual Dense Networks. *Laser Optoelectron. Prog.* **2019**, *56*, 95–103.
24. Cao, Y.; Xu, J.; Lin, S.; Wei, F.; Hu, H. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.
25. Yan, Z. *Research on Identification Technology of Tea Pests Based on Deep Learning*; Chongqing University of Technology: Chongqing, China, 2022.
26. Zhang, G. *Arrhythmia Classification System Based on ResGC-Net*; Jilin University: Changchun, China, 2022.
27. Zeng, W.; Chen, Y.; Hu, G.; Bao, W.; Liang, D. Detection of Citrus Huanglongbing in Natural Background by SMS and Two-way Feature Fusion. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 280–287.
28. Gai, L.; Cai, J.; Wang, S.; Cang, Y.; Chen, N. Research review on the application of convolutional neural network in image recognition. *J. Chin. Comput. Syst.* **2021**, *42*, 1980–1984.
29. Chen, L.; Li, S.; Bai, Q.; Yang, J.; Jiang, S.; Miao, Y. Review of Image Classification Algorithms Based on Convolutional Neural Networks. *Remote Sens.* **2021**, *13*, 4712. [CrossRef]
30. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
31. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
32. Chen, J.K.; Han, J.Y.; Liu, C.Z.; Wang, Y.F.; Shen, H.C.; Li, L. A Deep-Learning Method for the Classification of Apple Varieties via Leaf Images from Different Growth Periods in Natural Environment. *Symmetry* **2022**, *14*, 1671. [CrossRef]
33. Lee, Y.; Hwang, J.-W.; Lee, S.; Bae, Y.; Park, J. An energy and GPU-computation efficient backbone network for real-time object detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR), Long Beach, CA, USA, 15–20 June 2019.
34. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 10–17 October 2021; pp. 10012–10022.
35. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 3349–3364. [CrossRef] [PubMed]


**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.