PeerJ Computer Science

SeedSortNet: a rapid and highly effificient lightweight CNN based on visual attention for seed sorting

Chunlei Li[1], Huanyu Li[1], Zhoufeng Liu[1], Bicao Li[1] and Yun Huang[2]

[1] School of Electrical and Information Engineering, Zhongyuan University of Technology, Zhengzhou, Henan, China

[2] Xiamen Vision+ Technology Co. Ltd, Xiamen, Fujian, China

ABSTRACT

Seed purity directly affects the quality of seed breeding and subsequent processing products. Seed sorting based on machine vision provides an effective solution to this problem. The deep learning technology, particularly convolutional neural networks (CNNs), have exhibited impressive performance in image recognition and classification, and have been proven applicable in seed sorting. However the huge computational complexity and massive storage requirements make it a great challenge to deploy them in real-time applications, especially on devices with limited resources. In this study, a rapid and highly efficient lightweight CNN based on visual attention, namely SeedSortNet, is proposed for seed sorting. First, a dual-branch lightweight feature extraction module Shield-block is elaborately designed by performing identity mapping, spatial transformation at higher dimensions and different receptive field modeling, and thus it can alleviate information loss and effectively characterize the multi-scale feature while utilizing fewer parameters and lower computational complexity. In the down-sampling layer, the traditional MaxPool is replaced as MaxBlurPool to improve the shift-invariant of the network. Also, an extremely lightweight sub-feature space attention module (SFSAM) is presented to selectively emphasize fine-grained features and suppress the interference of complex backgrounds. Experimental results show that SeedSortNet achieves the accuracy rates of 97.33% and 99.56% on the maize seed dataset and sunflower seed dataset, respectively, and outperforms the mainstream lightweight networks (MobileNetv2, ShuffleNetv2, etc.) at similar computational costs, with only 0.400M parameters (vs. 4.06M, 5.40M).

Subjects Artificial Intelligence, Computer Vision, Embedded Computing, Real-Time and Embedded Systems

Keywords Seed sorting, Computer vision, Lightweight CNN, Attention mechanism

Submitted 9 April 2021
Accepted 21 June 2021
Published 5 August 2021

Corresponding author
Chunlei Li, lichunlei1979@zut.edu.cn

Academic editor
Pengcheng Liu

Additional Information and Declarations can be found on page 17

DOI **10.7717/peerj-cs.639**

INTRODUCTION

Seed purity directly affects the quality of seed breeding and subsequent processing products. For example, in the process of seed harvest and storage, the impurities or hybrids may be mixed in the normal seed, which results in the economic losses to agricultural production and processing. Therefore, it is crucial to sort impurities and hybrids to ensure that the seed purity meet the market criteria. However, the traditional manual sorting methods are laborious and time-consuming, and hence less efficient. With the evolution of the

technology, there has been a tremendous development in the field of machine vision (*Rehman et al., 2019*; *Wu et al., 2020*) and robot control technology (*Liu, Yu & Cang, 2019*; *Liu et al., 2020*; *Han et al., 2020b*). The automatic sorting methods (*Li et al., 2019*) based on the above technologies provide a promising solution.

Traditional automatic seed sorting methods adopt hand-crafted features for image characterization, such as color, shape, texture, and wavelet features or their combinations (*Liu et al., 2015*; *HemaChitra & Suguna, 2018*; *Li et al., 2019*). Then, the effective classifiers are employed to realize seed recognition such as linear discriminant analysis (LDA) (*Choudhary, Paliwal & Jayas, 2008*), support vector machine (SVM) (*Altuntas et al., 2018*), decision tree (DT) (*Kayacan, Sofu & Cetisli, 2016*), least square (LS) (*Mebatsion, Paliwal & Jayas, 2013*) and artificial neural network (ANN) (*Liu et al., 2015*). However, these methods are designed for a specific kind of seed and lack self-adaptivity. In the last three years, mainly due to the advances of deep learning, more concretely convolutional neural networks (CNNs), the quality of image classification (*Krizhevsky, Sutskever & Hinton, 2012*; *Han et al., 2018*), object detection (*Ren et al., 2015*; *Sun et al., 2019*; *Bochkovskiy, Wang & Liao, 2020*) and semantic segmentation (*Chen et al., 2014*) has been progressing at a dramatic pace. Recently, some researchers also adopted deep learning technology in crop identification tasks and achieved good performance (*Ni et al., 2019*; *Kurtulmus, 2021*).

The crop recognition and classification methods, especially for seed sorting, should be deployed on a fast and stable embedded system due to the requirement of higher processing speed. However, the performance of these methods often depends on a deeper, wider network structure, thus it suffers from huge computational complexity and massive storage requirements (*Han et al., 2018*). Therefore, the deep CNN model should be compressed and streamlined while maintaining high recognition accuracy. Recently, some lightweight and efficient CNN models have been designed, such as MobileNet (*Howard et al., 2017*), MobileNetv2 (*Sandler et al., 2018*), ShuffleNet (*Zhang et al., 2018*) and ShuffleNetv2 (*Ma et al., 2018*) for the real-time detection and recognition tasks. However, due to the lower discrimination of different types of seed, the feature extraction ability of these models is insufficient, thus leads to low recognition accuracy.

In this paper, a lightweight CNN based on visual attention for seed sorting is proposed. A dual-branch lightweight feature extraction module (i.e., Shield-block) is designed to improve the feature characterization ability while utilizing fewer parameters and lower computational complexity, and the traditional MaxPool is replaced as MaxBlurPool (*Zhang, 2019*) to improve the shift-invariant of the network in the down-sampling layer. In addition, an extremely lightweight sub-feature space attention module (SFSAM) is proposed as the basic unit of the built CNN model to selectively emphasize fine-grained features and suppress the interference from complex backgrounds. Overall, ours contributions are three-fold as follows:

- We designed a dual-branch lightweight feature extraction module (i.e, Shield-block) to alleviate information loss and effectively characterize the multi-scale feature while utilizing fewer parameters and lower computational complexity.

- We proposed an extremely lightweight sub-feature space attention module, which divides the feature maps into different subspaces and infers different attention maps for each subspace. To selectively emphasize fine-grained features, and suppress the interference of complex backgrounds.
- Experiments are conducted on the maize seed dataset and sunflower seed dataset, and the results show that SeedSortNet achieves higher accuracy compared with the mainstream lightweight networks (MobileNetv2, ShuffleNetv2, etc.) at the similar computational cost, even outperforms the deeper and wider networks, such as VGG (*Simonyan & Zisserman, 2014*), GoogleNet (*Szegedy et al., 2015*), and ResNet (*He et al., 2016*).

The remainder of the paper is organized as follows. In 'Related Work', we summarize some related work on seed sorting and lightweight model design. 'Proposed Method' introduces the technical details of the proposed method and network architecture. In 'Experiments', we carry out a series of comparative experiments on maize and sunflower seed datasets and the experimental results are analyzed. Finally, we conclude in 'Discussion'.

RELATED WORK

In the following, we review the existing crop identification methods and related technologies, such as CNN model compression and lightweight model design.

Crop identification

Agricultural product assessment and recognition based on machine vision technology have been a research focus in agricultural applications, which is widely used in the detection and sorting of agricultural products such as wheat, corn, fruits, and the identification of plant diseases and insect pests.

Liu et al. (2015) proposed a novel soybean seed sorting based on neural network. Eight shape features, three-color features, and three texture features are extracted to characterize the soybean seed, and BP neural network is used as the classification model to recognize the different defects. Experiments are conducted on the collected image set which includes 857 images of soybean seeds with insect damage, mildew, and other defects, and the results achieve an average recognition accuracy of 97.25%. *Huang (2012)* proposed a neural network-based quality evaluation and classification method for areca nuts. The axis length, secondary axis length, axis number, area, perimeter, compactness, and the average gray level are used as the feature, and a back-propagation neural network classifier is employed to sort the quality of the areca nuts. *Aznan et al. (2016)* adopted machine vision methods to discriminate the variety of cultivated rice seed, namely M263. They firstly extracted different morphological features and then adopted a stepwise discriminant function analysis (DFA) to classify different types of rice. The classification accuracy for testing and training sets is 96% and 95.8%, respectively. *HemaChitra & Suguna (2018)* presented a novel sorting method of Indian pulse seeds based on image analysis techniques. In this method, they extracted the colors, shapes, and texture features, and adopted SVM for classification. The accuracy of their method can reach 98.9% accuracy. *Li et al. (2019)* designed a system to distinguish different damaged types of corn. An image database including normal corn

and six different damaged corns is constructed. The features such as color and shape are extracted, then the maximum likelihood classifier is leveraged to discriminate these corns. Experiment results show that the classification accuracy is above 74% for all the classes. However, these methods adopt the handcraft features designed for the specific crops and the traditional classifier for sorting, and suffer from poor adaptability and low accuracy.

Due to the excellent feature representation ability, the deep learning models represented by CNN have achieved good performance in image classification, object detection, and semantic segmentation, and have also been successfully applied in plant disease detection and crop type classification. *Sladojevic et al. (2016)* proposed a CNN-based system to identify 13 types of plant diseases out of healthy leaves. The performance of this approach exhibited a top-1 success of 96.3%. *Veeramani, Raymond & Chanda (2018)* studied the effect of the number of convolution kernels in the two layers CNN on the recognition performance of haploid and diploid corn seeds. *Veeramani, Raymond & Chanda (2018)* adopted VGG19 and GoogleNet to classify corn seed defects and analyzed the influence of the two networks with different depths on the recognition performance. *Dolata & Reiner (2018)* proposed a method for the classification of barley varieties based on CNN, which is based on two separate convolutional layers to analyze dorsal and ventral sides, respectively. The network is trained on a small sample set of 200-500 cases in 8 categories, and the classification accuracy reaches 97%. *Kurtulmuş (2021)* adopted AlexNet (*Krizhevsky, Sutskever & Hinton, 2012*), GoogleNet, and ResNet to identify sunflower seed varieties, and then they were also evaluated in terms of both accuracy and training time, GoogleNet obtained the highest classification accuracy (95%).

The CNN-based crop identification method can achieve the better recognition rate and has higher self-adaptivity. However, the performance of the deep learning method depends on the depth and width of the model, the researchers often boost the depth and width of the model to improve the performance of the detection and recognition system. But this strategy results in slow speed and difficult deployment in industrial applications.

Model lightweight

For the specific application of crop seed sorting, due to the extremely fast production speed, it is necessary to develop the lightweight CNN model while maintaining a higher recognition accuracy. To trade off the model size and performance for deep neural network architectures has been an active research area, the related technologies include model compression, lightweight network design, etc (*Liang et al., 2021*).

Model compression

Model compression aims at generating the small network models from the trained large network models while keeping the performance. The typical techniques include pruning, quantization, and knowledge distillation. Pruning technology is based on the assumption that many parameters in the deep neural network are redundant, then the weights (*Guo, Yao & Chen, 2016*; *Aghasi et al., 2016*; *Liu et al., 2018*) or filters (*Li et al., 2016*; *Liu et al., 2017*; *Lin et al., 2020*) with low correlation can be removed to make the network structure sparse. Quantization methods aim to deploy the CNN model on the terminal hardware and encode

the weights and activations using 8-bit integers (INT8) without incurring a significant loss in accuracy. Some other quantization methods even adopt INT4 or lower, such as binary quantization (*Courbariaux et al., 2016*) and ternary quantization (*Mellempudi et al., 2017*) to reduce the model size. Knowledge distillation is firstly proposed by *Bucilu, Caruana & Niculescu-Mizil (2006)* and generalized by *Hinton, Vinyals & Dean (2015)* and can generate a small student network by learning the behavior of a large teacher network. *Cho & Hariharan (2019)* empirically analyzed in detail the efficacy of knowledge distillation. However, compared with the original network, model compression is difficult to achieve better performance. The compression size is too large, which will lead to significant decrease of performance.

Lightweight network design

Lightweight network design refers to the redesign of the network structure based on the existing CNN model to reduce the parameters and the computational complexity. *Lin, Chen & Yan (2013)* proposed a Network-In-Network architecture, which used 1×1 convolution to increase network performance while maintaining a lower computational complexity. SqueezeNet (*Iandola et al., 2016*) is a lightweight network structure based on 1×1 convolution. The squeeze and expand module proposed by this model can effectively reduce the parameters while ensuring recognition accuracy. The recognition accuracy of the proposed method can be up to 57.55%, and it is similar with the AlexNet with the model size of $50\times$smaller. Google developed two efficient architectures denoted as MobileNet (*Howard et al., 2017*) and MobileNetv2 (*Sandler et al., 2018*) in 2017 and 2018, respectively. MobileNet proposed depthwise separable convolutions to reduce the computational complexity and achieved the state-of-art accuracy with low latency. Thereafter, the linear bottleneck with inverted residual structure is proposed in MobileNetV2 to construct a more efficient architecture. ShuffleNet (*Zhang et al., 2018*) proposed the pointwise group convolution and channel shuffle operations to improve the recognition accuracy while reducing latency. Combining the advantage of MobileNet and ShuffleNet, *Ma et al. (2018)* proposed ShuffleNetV2, which improves group convolution by channel split and used channel shuffle for the split channel as well. *Wang & Yu (2020)* proposed the Tied Block Convolution (TBC) which shares the same thinner filters over equal blocks of channels and produces multiple responses with a single filter, to design a lightweight model. GhostNet (*Han et al., 2020a*) applied a series of linear transformations to generate many Ghost feature maps, and it can characterize the required information from the original features at a small cost, which effectively reduces calculation and parameters. However, due to the low discrimination of crop seeds, the recognition accuracy will be significantly reduced when the existing lightweight models are directly applied to the seed sorting. Therefore, a rapid and highly efficient lightweight CNN model should be developed based on the characteristics of crop seeds while keeping the accuracy.

PROPOSED METHOD

Seed sorting based on deep learning is a promising method for seed breeding and subsequent processing products. In this paper, we proposed a rapid and efficient lightweight

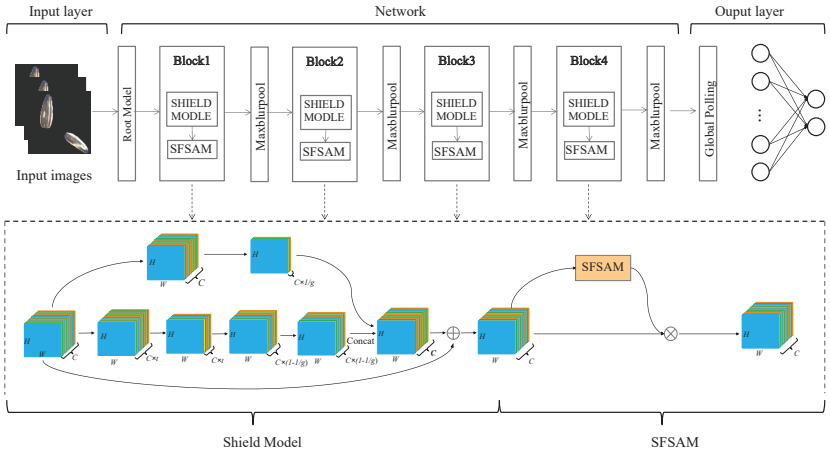

Figure 1 **The flowchart of the proposed SeedSortNet model.**
Full-size DOI: 10.7717/peerjcs.639/fig-1

CNN model with a dual-branch network structure based on visual attention for seed sorting, denoted as SeedSortNet. It is an efficient and lightweight end-to-end recognition framework, which is mainly composed of sequential cascade layers and basic blocks, and the overall structure of the model is shown in Fig. 1. First, a dual-branch lightweight feature extraction module, namely Shield-block, is designed for effective feature extraction. Then, the traditional convolution is replaced by depthwise convolution and pointwise convolution to achieve the trade-off between classification accuracy and efficiency. Moreover, in the down-sampling layer, MaxPool is substituted as MaxBlurPool to improve the shift-invariant of the network. Finally, we propose an extremely lightweight sub-feature space attention module to selectively emphasize fine-grained features and suppress the interference of complex backgrounds. And the proposed method is specifically described as follows.

Network construction

Due to the required higher processing speed and recognition accuracy, the representative deep neural network model (eg, ResNet, VGG, GoogLeNet, etc.) cannot efficiently tackle with the seed sorting task because of the lower efficiency and insufficient feature extraction ability. To address these issues, we construct a novel lightweight and efficient network which consists of Root-model, Shield-block, and a novel down-sampling module.

A.Root-model. To effectively improve the feature representation ability while reducing calculation, a dual-branch structure, namely Root-model (Fig. 2A), is designed as the first stage of SeedSortNet. First, the sixteen 3×3 filters are utilized to extract the shallow feature information (such as texture, shape, color, etc.) of the test image. Then, in one branch, the MaxBlurPool which is a non-overlapping 2×2 window is designed for reducing the aliasing effect and improving the shift-invariant of the network. In another branch, we firstly use 3×3 filters with the stride of 2 to convolute the input features and then adopt 1×1 filters

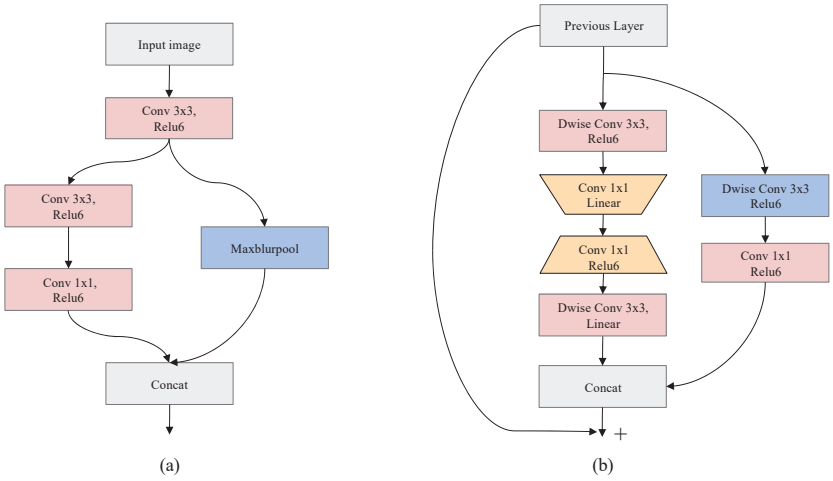

Figure 2 **The schema of Root-model (A) and Shield-block (B).**
Full-size DOI: 10.7717/peerjcs.639/fig-2

to reduce the output dimension of the branch. Finally, the features generated by the two branches are concatenated together as the input of the next layer.

B.Shield-block. The inverted residual block (*Sandler et al., 2018*) which shifts the identity mapping from high-dimensional representations to low-dimensional ones (i.e., the bottlenecks), has been successfully applied in the design of lightweight networks. However, the connection of identity mapping between thin bottlenecks would inevitably lead to information loss since the residual representations are compressed (*Daquan et al., 2020*). In addition, this connection would also weaken the propagation capability of gradients across layers due to gradient confusion arising from the narrowed feature dimensions, and hence affect the training convergence and model performance (*Sankararaman et al., 2020*). To address these issues, we propose a dual-branch feature extraction module by improving the inverted residual block in this article (shown in Fig. 2B).

In the main branch, two 3×3 depthwise convolution layers are utilized to encode richer spatial information to generate a more expressive representation. Then we adopt two pointwise convolutional layers between two 3×3 depthwise convolutional layers, the first point convolution layer reduces the feature channel dimension and the latter increases its dimension, to encode the cross-channel information of the feature maps and reduce the computational complexity. Also, the linear activation function is adopted after the first pointwise convolutional layer and the last depthwise convolutional layer, which can prevent the feature values from being zeroed and hence reduce information loss.

For the other branch, a 3×3 depthwise separable convolution is designed to acquire the spatial representation of different receptive fields, and thus improve the feature representation ability. In the end, the concatenation of the two branches and its shortcut connection with the input feature are combined as the final output.

In the following, we present the detailed data processing operator of Shield-block, and it is shown in Table 1, where H, W and C represents the height, width, and channel number of the feature map, and $1/t$ represents the reduction rate of channels. Moreover, to ensure

Table 1 Data processing in the Shield-block.

Input dimension	Operator	Output dimension
$\begin{bmatrix} H\times W\times C \\ H\times W\times C & H\times W\times C \\ H\times W\times C\times\frac{1}{t} & H\times W\times C \\ H\times W\times C\times\left(1-\frac{1}{r}\right) & \end{bmatrix}$	$\begin{bmatrix} 3\times 3 \text{ Dwise Conv, Re}lu6 & \\ 1\times 1 \text{ Conv, Linear} & 3\times 3 \text{ Dwise Conv, Relu6} \\ 1\times 1 \text{ Conv, Re } lu6 & 1\times 1 \text{ Conv , Re } lu6 \\ 3\times 3 \text{ Dwise Conv, Linear} & \\ & \text{Concat} \end{bmatrix}$	$\begin{bmatrix} H\times W\times C \\ H\times W\times C\times\frac{1}{t} & H\times W\times C \\ H\times W\times C\times\left(1-\frac{1}{r}\right) & H\times W\times C\times\frac{1}{r} \\ H\times W\times C\times\left(1-\frac{1}{r}\right) & \\ H\times W\times C & \end{bmatrix}$

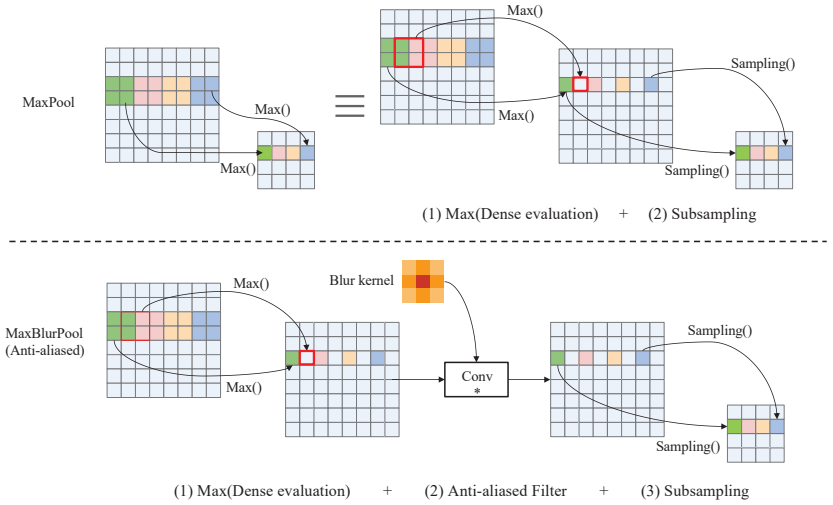

Figure 3 Operation details of MaxPool and anti-aliased MaxBlurPool.
Full-size DOI: 10.7717/peerjcs.639/fig-3

the input channel dimension is consistent with the output channel dimensional, and the hyperparameter $1/r$ (referring to the proportion of the input channel of the sub-branch output channel) is adopted in this paper, here we empirically set r to 6.

C.Down-sampling. Down-sampling operator can reduce the feature dimensionality while retaining the valid information. Traditional down-sampling methods (eg. MaxPool, Strided-Convolution, AvgPool, etc.) violate the shift-equivariance and results in small shifts in the input that can drastically change the output (*Azulay & Weiss, 2018*). And this phenomenon will become more obvious with the increase of the network depth. To solve this problem, a fuzzy sampling method, namely MaxBlurPool, proposed in (*Zhang, 2019*) is adopted in our method. The specific process of MaxPool and MaxBlurPool is shown in Fig. 3. From this figure, we can see that a blur kernel is inserted between max and subsampling to remove aliased in the MaxBlurPool method, thereby improving the shift-invariant and enhancing the robustness of the CNN model.

Lightweight sub-feature space attention module (SFSAM)

Due to the low discrimination of different types of seeds, the fine-grained spatial features are crucial for seed sorting. Therefore, an extremely lightweight sub-feature space attention module is proposed to selectively emphasize fine-grained features, and suppress the

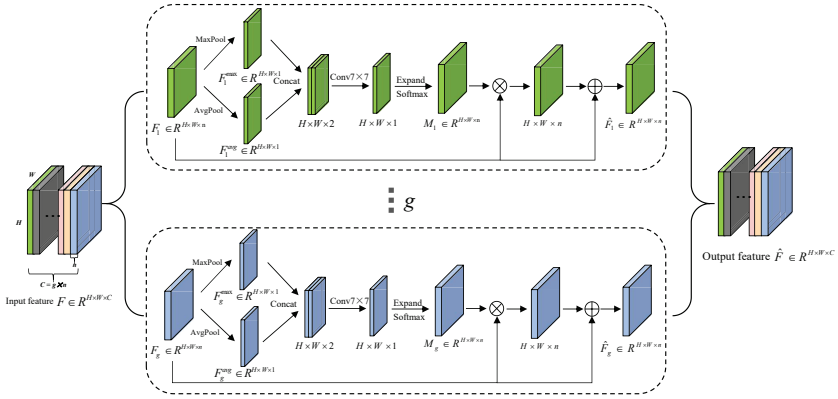

Figure 4 **The schematic diagram of SFSAM structure.**
Full-size DOI: 10.7717/peerjcs.639/fig-4

interference of complex backgrounds. It divides the feature maps into different subspaces and infers different attention maps for each subspace, thus can generate the multi-scale feature representation, it is shown in Fig. 4. The detailed process is described as follows.

The input feature map $F \in R^{H\times W\times C}$ is firstly divided into g mutually exclusive groups $[F_1, F_2, F_3, \ldots, F_i, \ldots, F_g]$ (i.e., sub-feature spaces),where each sub-feature space F_i contains n intermediate feature maps. *Zagoruyko & Komodakis (2016)* have proved that pooling operations along the channel axis are effective in highlighting informative regions. Therefore, AvgPool and MaxPool operations are applied to g sub-feature spaces along the channel axis to generate g groups of average-pooled features $F_i^{\max} \in R^{1\times H\times W}$ and max-pooled features $F_i^{avg} \in R^{1\times H\times W}$. Then, these features are concatenated separately to generate g efficient feature descriptors $[F_i^{\max}, F_i^{avi}]$. Thereafter, the g group's subspace attention maps are generated using Eq.(1).

$$\begin{aligned} M_i &= \mathrm{softmax}(f^{k\times k}([\mathrm{MaxPool}(F_i), \mathrm{AvgPool}(F_i)])) \\ &= \mathrm{softmax}\left(f^{k\times k}\left(\left[F_i^{\max}, F_i^{avi}\right]\right)\right) \end{aligned} \tag{1}$$

where $f^{k\times k}$ represents a convolution operation with a filter size of $k\times k$. In this paper, k is empirically set to 7. The attention map in each group (subspace) can capture the non-linear dependencies among the feature maps by learning to gather cross-channel information. Meantime, we employ a gating mechanism with a softmax activation to map the attention weighting tensor into $[0, 1]$.

Then, each group of feature maps gets the refined set of feature maps ($\hat{F}_i$) after the feature redistribution in Eq. (2).

$$\hat{F}_i = (M_i \otimes F_i) \oplus F_i \tag{2}$$

where $\otimes$ is element-wise multiplication and $\oplus$ is element-wise addition.

The final output $\hat{F}$ of SFSAM is obtained by concatenating the feature maps of each group, and it is described as Eq. (3).

$$\hat{F} = \mathrm{concat}\left(\hat{F}_1, \hat{F}_2, \hat{F}_3, \ldots, \hat{F}_i, \ldots, \hat{F}_g\right). \tag{3}$$

Table 2 Parameter configuration diagram of the SeedSortNet.

Stage	Input	Operator	*t*	C	R
1	224×224×3	Root-module	—	32	1
2	112×112×32	Shield-block	2 |6	64	2
3	112×112×64	SFSAM	—	64	1
4	112×112×64	MaxBlurPool	—	64	1
5	56×56×64	Shield-block	2 |6	128	4
6	56×56×128	SFSAM	—	128	1
7	56×56×128	MaxBlurPool	—	128	1
8	28×28×128	Shield-block	2 |6	128	5
9	28×28×128	SFSAM	—	192	1
10	28×28×192	MaxBlurPool	—	192	1
11	14×14×192	Shield-block	2 |6	256	4
12	14×14×256	SFSAM	—	256	1
13	14×14×256	MaxBlurPool	—	256	1
14	7×7×256	GlobalAvgpool	—	256	—
15	1×1×256	Dropout 2D-FC	—	2	—

Network topology

In this paper, a novel lightweight CNN model with a dual-branch network structure based on visual attention, denoted as SeedSortNet, is proposed for seed sorting with higher efficiency and recognition accuracy. The setting of the proposed SeedSortNet is outlined in Table 2. Each row denotes a sequence of building blocks, which is the repeated times of 'R'. The reduction ratio of channels is used in each Shield-block is denoted by ' $1/t$', and 'C' represents the number of channels in the output feature map.

We first use Root-module to generate 32 feature maps with the size of 112×112. Then, it is followed by the 15 Shield-blocks, four SFSAM attention modules, and four down-sampling layers (i.e., MaxBlurPool) spatial location distributions described in Table 2. At the first Shield-block of stages 2, 5, 8, and 11, the identity mappings do not need to be set because of the increasement of the feature map depth. Besides, we set 't' to 2 to avoid the information loss due to the low-dimensional input. Finally, the output of the fourth down-sampling layer is followed by a global average pooling layer, which can convert 2D feature maps into 1D feature vectors.

EXPERIMENTS

Experimental datasets

In this section, two datasets are selected for verifying the effectiveness of the proposed network architecture.

Maize seed dataset

The first dataset is a public haploid and diploid maize seed dataset of the maize research institute in Sakarya (Turkey), including 3000 RGB images of corn seeds (*Altunta, Cömert & Kocamaz, 2019*), and it includes 1230 haploid seeds images and 1770 diploid seed images. The dimensions of these images depend on the sizes of the seeds and vary between 300×289

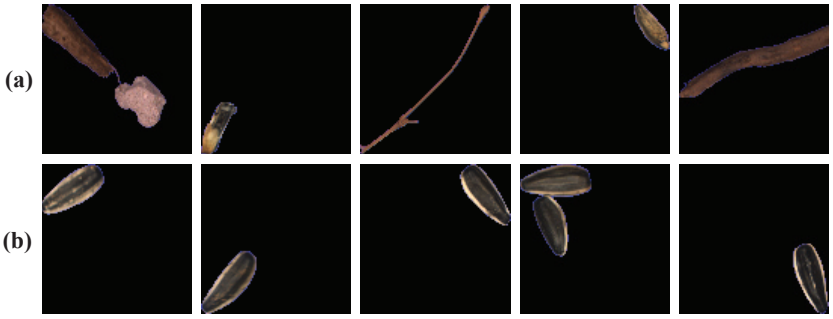

Figure 5 Representative samples of sunflower seed dataset. (A) Abnormal sunflower seeds; (B) normal sunflower seeds.

Full-size DOI: 10.7717/peerjcs.639/fig-5

Table 3 Category distribution and proportion of the training set on maize seed dataset and sunflower seed dataset.

Dataset	Category distribution	The proportion of the training set
Maize	1770 (Diploid) 1230 (Haploid)	75%
Sunflower	7837 (Normal) 7997 (Abnormal)	≈75%

pixels and 610×637 pixels. In the experiment, three-quarters of the dataset are used for training, and the remaining images are used for testing, as shown in Table 3.

The number of maize seed dataset is limited and may bring the overfitting for the proposed model. Therefore, the data augmentation methods, such as horizontal flip, vertical flip, and angle rotation, are adopted to augment the maize seed data set by a factor of 4. The experimental results prove that such a large dataset is enough to train a model with very strong generalization ability.

Sunflower seed dataset

To thoroughly evaluate the effectiveness of the proposed method, we constructed our sunflower seed dataset on an industrial production line for the experiments. The image acquisition device equipped with a color line scan camera is established to collect 15834 sunflower seed RGB images with the size of 100×100 pixels. And we divided them into two categories, as shown in Fig. 5. The top row is the abnormal seed images composed of leaves, stones, defective seeds, etc. The bottom row is the normal sunflower seed images. It is worth noting that when the picture contains several seeds and impurities or hybrids, we will classify them as abnormal to ensure a low false alarm. In our experiment, about three-quarters of the dataset are randomly selected as the training set, and the remaining images are used for testing, as shown in Table 3.

Implementation details and evaluation metric

Implementation details

All experiments were performed on a 64-bit Linux-based operation system, Ubuntu 18.04. The software is mainly based on the deep learning architecture of Pytorch and python development environment Spyder. The hardware is based on an Intel(R) Xeon(R)

Table 4 Calculation formulas and explanations of binary class metrics.

Measure	Formulation	Evaluation Focus
Accuracy (Acc)	$\frac{tp+tn}{tp+fp+tn+fn}$	The overall accuracy of a model.
Precision (p)	$\frac{tp}{tp+fp}$	The ratio of correctly classified positive samples to estimated total positive sample.
Recall (r)	$\frac{tp}{tp+fn}$	The proportion of positive values classified as true.
F1-score	$\frac{2*p*r}{p+r} = \frac{2*\frac{tp}{tp+fp}*\frac{tp}{tp+fn}}{\frac{tp}{tp+fp}+\frac{tp}{tp+fn}}$	The harmonic mean between precision and recall.

CPU E5-2650 v4 @2.20 GHz and two NVIDIA Quadro M5000 GPUs, with CUDA10.2 accelerating calculation.

And we train the network by mini-batch SGD, with an initial learning rate of 0.001 and a reducing factor of 0.1 after 30 epochs. The momentum parameter is set to 0.9 and the weight decay parameter is 0.0001. The number of iterations in training is 100, and the batch size is set to 16 and 64 on the maize and sunflower datasets, respectively. Besides, the input image size is resized to 224×224-pixel by the CenterCrop function, and the parameter g is set to 4 by analyzing experimental results.

Evaluation metric

To quantitively evaluate the effectiveness of the proposed method, four metrics, such as true positive (tp), true negative (tn), false positive (fp), and false negative (fn) are adopted in our method. tp is the true positive and represents correctly recognized haploid maize seeds or the normal sunflower seed. tn is the true negative and represents correctly recognized diploid maize seeds or the abnormal sunflower seed. fp is the false positive and represents the falsely recognized haploid maize seeds or the normal sunflower seed. fn is the false negative and represents falsely recognized diploid maize seeds or the abnormal sunflower seed. Based on these metrics, four evaluation metrics, accuracy (Acc), precision (p), recall (r) and F1-score, are calculated as Table 4.

It should be noted that the F1-score metric can better interpret the true performance when the number of samples is not balanced. Receiver operating characteristic (ROC) curves are also a useful tool for measuring a model performance without considering class distribution or error costs. Also, the number of parameters and required float points operations (denoted as FLOPs) are also employed to evaluate the model size and computational complexity, which are widely-used protocols.

Result analysis

Results on maize seed dataset

To assess the performance of our network (i.e., SeedSortNet) in the maize dataset. Six representative CNN models (i.e., AlexNet, VGG, ResNet, GoogleNet, DenseNet (*Huang et al., 2017*), and Resnext (*Xie et al., 2017*) are selected to conduct comparative experiments. The experimental results are shown in Table 5.

From the results in Table 5, we can observe that the adopted network can achieve good classification accuracy, and reach more than 90% under the same experimental environment. SeedSortNet has the best performance, with accuracy, precision, recall,

Table 5 Performance comparison of different network on maize seed dataset.

Model	Parameters	FLOPs	Acc	Precision	Recall	F1-score
AlexNet	57.01M	711.46M	93.33	93.39	92.80	93.07
VGG11	128.77M	7.63G	94.67	94.54	94.43	94.48
VGG13	128.96M	11.33G	94.50	94.51	94.10	94.29
ResNet18	11.18M	1.82G	95.33	95.18	95.18	95.18
ResNet50	23.51M	4.12G	96.00	95.73	96.05	95.88
DenseNet121	6.96M	2.88G	95.83	95.58	95.85	95.71
GoogleNet	5.60M	1.51G	96.67	96.50	96.62	96.56
ResNext101	86.75M	16.48G	96.00	95.77	95.99	95.88
SeedSortNet	0.40M	512.06M	97.33	97.30	97.18	97.24

Table 6 Performance comparison of maize seed dataset in lightweight CNNs.

Model	Parameters	FLOPs	Acc	F1-score
MobileNetv1	3.22M	587.94M	94.00	93.81
MobileNetv2 1.4×	4.06M	566.33M	96.00	95.89
ShuffleNetv1 2×(g=3)	3.53M	537.48M	96.67	96.55
ShuuffleNetv2 2×	5.35M	591.79M	96.00	95.90
GhostNet 2×	12.96M	529.89M	96.50	96.41
SeedSortNet	0.40M	512.06M	97.33	97.24
MobileNetv1 0.75×	1.83M	339.80M	91.83	91.65
MobileNetv2	2.23M	318.96M	95.83	95.71
ShuffleNetv1 1.5×(g=3)	2.00M	301.90M	96.17	96.05
ShuffleNetv2 1.5×	2.48M	302.65M	95.50	95.38
GhostNet 1.5×	7.79M	310.76M	95.83	95.72
SeedSortNet 0.75×	0.23M	338.64M	97.00	96.90

and F1-score of 97.33%, 97.30%, 97.18%, and 97.24%, respectively, with a relatively low computational complexity and model size. These results verify the effectiveness of the proposed method.

Meanwhile, we also compared the performance of mainstream lightweight CNN models (eg, MobileNetv1, MobileNetv2, ShuffleNetv1, ShuffleNetv2, GhostNet) under different calculation benchmarks. The experimental results on the maize dataset in terms of computational complexity, model parameters, classification accuracy, and F1-score are shown in Table 6. The models are typically grouped into two levels of computational complexity for embedded device applications, i.e., ∼300MFLOPs and 500 ∼600MFLOPs. From the results, we can see that the larger FLOPs lead to higher accuracy in these lightweight networks. SeedSortNet outperforms other competitors consistently in classification accuracy and F1-score at various computational complexity levels. Furthermore, the number of parameters has also greatly decreased for the proposed method.

In order to further demonstrate the effectiveness of the proposed method, ROC curve is adopted to measure the model performance. Figures 6A–6C shows the ROC curves and

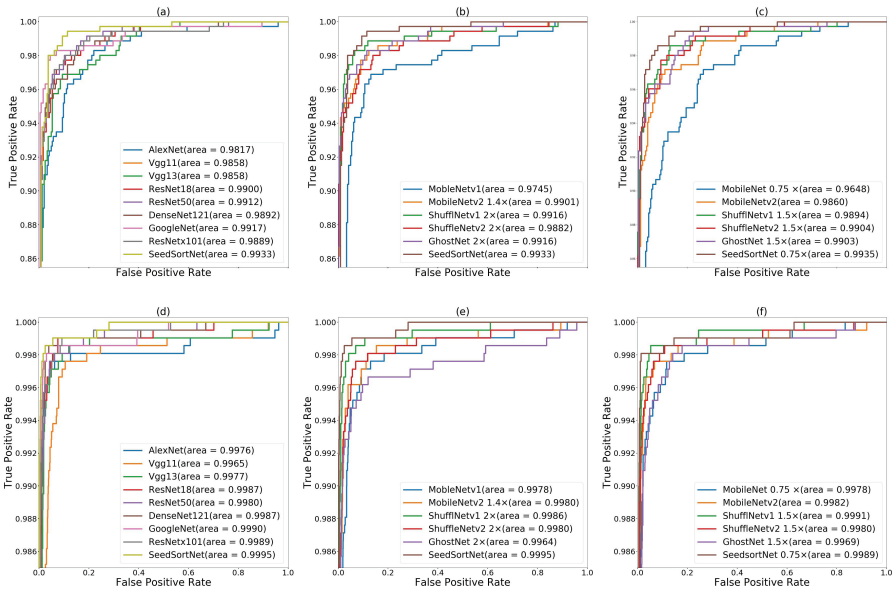

Figure 6 ROC curves of the CNN models on maize seed dataset (A, B, C) and sunflower seed dataset (D, E, F). (A & D) ROC curves of SeedSortNet and Six representative CNN models (i.e., AlexNet, VGG, ResNet, GoogleNet, DenseNet, and Resnext), (B & E) ROC curves of the lightweight network (500 ∼600MFLOPs), (C & F) ROC curves of the lightweight network (∼300MFLOPs).

Full-size DOI: 10.7717/peerjcs.639/fig-6

the calculated area under curve (AUC) scores for using the proposed method and other network models (i.e., the above comparison network) on the maize seed dataset. From AUC scores, it is observed that the method achieve the best result of 99.33% compared with other models, which is superior to the above representative representative CNN models and lightweight networks.

Results on sunflower seed dataset

Table 7 demonstrates the model size, computational complexity, accuracy, recall, specificity and F1-score of different network models on the sunflower seed dataset. From the table, we can see that the proposed method has the highest accuracy, precision, recall, and F1-score, while it has the lower FLOPs and parameters.

Similar to the maize seed sorting, we also conducted comparative experiments with the mainstream lightweight CNN models (eg, MobileNetv1, MobileNetv2, ShuffleNetv1, ShuffleNetv2, GhostNet) under different calculation benchmarks on the sunflower seed dataset. From Table 8, we can find that the classification accuracy and F1-score of the SeedSortNet are higher than other network models under a similar calculation cost. Meantime, we find that the test dataset is relatively balanced, thus its F1-score and accuracy are almost the same. Therefore, the proposed SeedSortNet is more suitable for deployment on edge devices and has the ideal sorting accuracy.

In Figs. 6D–6F, we find that the AUC score of the proposed method is closer to 1.0 (i.e., 0.9995) on the sunflower seed dataset compared with other CNN models, which demonstrates SeedSortNet has a good ability to prevent misclassification.

Table 7 Performance comparison of different network on sunflower seed dataset.

Model	Parameters	FLOPs	Acc	Precision	Recall	F1-score
AlexNet	57.01M	711.46M	99.00	99.01	98.99	99.00
VGG11	128.77M	7.63G	97.78	97.78	97.80	97.78
VGG13	128.96M	11.33G	98.44	98.43	98.45	98.44
ResNet18	11.18M	1.82G	99.05	99.04	99.05	99.05
ResNet50	23.51M	4.12G	99.22	99.22	99.22	99.22
DenseNet121	6.96M	2.88G	98.90	98.89	98.91	98.90
GoogleNet	5.60M	1.51G	99.34	99.34	99.34	99.34
ResNext101	86.75M	16.48G	99.22	99.22	99.22	99.22
SeedSortNet	0.40M	512.06M	99.56	99.56	99.56	99.56

Table 8 Performance comparison of sunflower seed dataset in lightweight CNNs.

Model	Parameters	FLOPs	Acc	F1-score
MobileNetv1	3.22M	587.94M	98.36	98.36
MobileNetv2 1.4×	4.06M	566.33M	98.83	98.83
ShuffleNetv1 2×(g=3)	3.53M	537.48M	99.19	99.19
ShuuffleNetv2 2×	5.35M	591.79M	99.00	99.00
GhostNet 2×	12.96M	529.89M	98.80	98.80
SeedSortNet	0.40M	512.06M	99.56	99.56
MobileNetv1 0.75×	1.83M	339.80M	98.32	98.31
MobileNetv2	2.23M	318.96M	98.90	98.90
ShuffleNetv1 1.5×(g=3)	2.00M	301.90M	99.12	99.12
ShuffleNetv2 1.5×	2.48M	302.65M	98.73	98.73
GhostNet 1.5×	7.79M	310.76M	98.44	98.44
SeedSortNet 0.75×	0.23M	338.64M	99.34	99.34

Ablation study

The ablation study is carried on SeedSortNet and the network without SFSAM attention mechanism. The experimental results in Table 9 show that F1-score of 96.33% and 99.37% are obtained without SFSAM on the maize and sunflower seed datasets, respectively, which proves that Root-model and Shield block have better information extraction abilities. Meanwhile, SeedSortNet can get 97.33% and 99.56% F1-score, respectively. These demonstrate that SFSAM can selectively emphasize information features and suppress the interference of complex backgrounds, thereby improving the performance. At the same time, it can also be observed from Table 9 that SFSAM does not introduce too many parameters and calculations.

Effects of g selection in SFSAM

As described in the SFSAM section, the feature maps are divided into g groups and generate g attention maps. Each attention map can capture cross-channel information from the feature maps in its respective group. When $g = 1$, the cross channel information for the whole feature volume is captured by a single attention map, which is not sufficient

Table 9 F1-score of SeedSortNet and SeedSortNet (without SFSAM) on maize seed dataset and sunflower seed dataset.

Model	Parameters (M)	FLOPs	F1-score(maize)	F1-score (sunflower)
SeedSortNet(without SFSAM)	0.399M	505.26M	96.22	99.39
SortSeedNet	0.400M	512.06M	97.24	99.56

Table 10 F1-score of SeedSortNet (with fewer parameters/FLOPs; $g = 1, 4, 8, 16$) on maize seed dataset and sunflower seed dataset.

Model	Parameters	FLOPs	F1-score(maize)	F1-score(sunflower)
SortSeedNet (g=1)	0.399M	506.56M	96.39	99.44
SortSeedNet (g=4)	0.400M	512.06M	97.24	99.56
SortSeedNet (g=8)	0.402M	518.85M	96.90	99.51
SortSeedNet (g=16)	0.405M	532.45M	96.57	99.46

to capture the complex relationships in the entire feature space and will result in lower predictive performance. When $1 < g < C$, the better exchange of cross-channel information can be obtained. Therefore, we conduct experiments on the different parameters assigned by g (such as $g = 1, 4, 8, 16$), and the results in Table 10 confirm the correctness of the above analysis. It can also be observed that the maize and sunflower seed datasets have achieved higher performance gains, and the FLOPs and parameters increase with the increase g. Based on the experimental results, we adopt $g = 4$ to conduct the above series of comparative experiments which provides a reasonable trade-off between preserving good performance and improving computational efficiency.

DISCUSSION

In this paper, we present a rapid and highly efficient lightweight CNN for seed sorting (i.e., SeedSortNet). We first design a novel dual-branch lightweight feature extraction module (i.e., Shield block) for building efficient neural network architectures. In the down-sampling layer, MaxBlurPool is employed instead of frequently-used MaxPool to improve the shift-invariant of the network. Then we proposed a lightweight sub-feature space attention module (SFSAM), which improves the representational power of the model by learning different attention feature maps. A wide range of experiments show the effectiveness of SeedSortNet, which achieves state-of-the-art identification performance on maize seed and sunflower seed datasets while utilizing fewer parameters and lower computational complexity. In future research, the number of seed varieties and images will be further increased to test the performance of these models, and we hope that these methods can be applied in the seed market.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by NSFC (U1804157, No. 61772576, No. 62072489), the Henan Science and Technology Innovation Team (CXTD2017091), IRTSTHN (21IRTSTHN013), the ZhongYuan Science and Technology Innovation Leading Talent Program (14200510013), and the Program for Interdisciplinary Direction Team in Zhongyuan University of Technology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
NSFC: U1804157, No. 61772576, No. 62072489.
Henan science and technology innovation team: CXTD2017091.
IRTSTHN: 21IRTSTHN013.
ZhongYuan Science and Technology Innovation Leading Talent Program: 14200510013.
Program for Interdisciplinary Direction Team in Zhongyuan University of Technology.

Competing Interests

Yun Huang is employed by Xiamen Vision+ Technology Co. Ltd.

Author Contributions

- Chunlei Li and Huanyu Li conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Zhoufeng Liu conceived and designed the experiments, performed the experiments, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Bicao Li conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Yun Huang conceived and designed the experiments, performed the computation work, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The maize seed dataset comes from Altuntaş et al. (2019): https://doi.org/10.1016/j.compag.2019.104874 and is available at: http://www.rovile.org/datasets/haploid-and-diploid-maize-seeds-dataset/.

The seedsortnet code and sunflower seed dataset are available at GitHub: https://github.com/Huanyu2019/Seedsortnet.

REFERENCES

Aghasi A, Abdi A, Nguyen N, Romberg J. 2016. Net-trim: convex pruning of deep neural networks with performance guarantee. ArXiv preprint. arXiv:1611.05162.

Altuntas Y, Kocamaz AF, Cengiz R, Esmeray M. 2018. Classification of haploid and diploid maize seeds by using image processing techniques and support vector machines. In: *2018 26th signal processing and communications applications conference (SIU)*.

Altunta Y, Cömert Z, Kocamaz AF. 2019. Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach. *Computers & Electronics in Agriculture* **163(2019)**:1–11.

Aznan A, Rukunudin I, Shakaff A, Ruslan R, Zakaria A, Saad F. 2016. The use of machine vision technique to classify cultivated rice seed variety and weedy rice seed variants for the seed industry. *International Food Research Journal* **23**:S31–S325.

Azulay A, Weiss Y. 2018. Why do deep convolutional networks generalize so poorly to small image transformations? ArXiv preprint. arXiv:1805.12177.

Bochkovskiy A, Wang C-Y, Liao H-YM. 2020. Yolov4: optimal speed and accuracy of object detection. ArXiv preprint. arXiv:2004.10934.

Bucilu C, Caruana R, Niculescu-Mizil A. 2006. Model compression. In: *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining*. New York: ACM, 535–541.

Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. 2014. Semantic image segmentation with deep convolutional nets and fully connected crfs. ArXiv preprint. arXiv:1412.7062.

Cho JH, Hariharan B. 2019. On the efficacy of knowledge distillation. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 4794–4802.

Choudhary R, Paliwal J, Jayas DS. 2008. Classification of cereal grains using wavelet, morphological, colour, and textural features of non-touching kernel images. *Biosystems Engineering* **99(3)**:330–337 DOI 10.1016/j.biosystemseng.2007.11.013.

Courbariaux M, Hubara I, Soudry D, El-Yaniv R, Bengio Y. 2016. Binarized neural networks: training deep neural networks with weights and activations constrained to+ 1 or-1. ArXiv preprint. arXiv:1602.02830.

Daquan Z, Hou Q, Chen Y, Feng J, Yan S. 2020. Rethinking bottleneck structure for efficient mobile network design. In: Vedaldi A, Bischof H, Brox T, Frahm JM, eds. *Computer Vision—ECCV 2020. ECCV 2020. Lecture Notes in Computer Science*, vol. 12348. Cham: Springer DOI 10.1007/978-3-030-58580-8_40.

Dolata P, Reiner J. 2018. Barley variety recognition with viewpoint-aware double-stream convolutional neural networks. In: *2018 federated conference on computer science and information systems (FedCSIS)*. 101–105.

Guo Y, Yao A, Chen Y. 2016. Dynamic network surgery for efficient dnns. ArXiv preprint. arXiv:1608.04493.

Han K, Guo J, Zhang C, Zhu M. 2018. Attribute-aware attention model for fine-grained representation learning. In: *Proceedings of the 26th ACM international conference on multimedia, MM '18*. New York, NY, USA: Association for Computing Machinery, 2040–2048.

Han K, Wang Y, Tian Q, Guo J, Xu C. 2020a. Ghostnet: more features from cheap operations. In: *2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. Piscataway: IEEE.

Han S-Y, Zhou J, Chen Y-H, Zhang Y-F, Tang G-Y, Wang L. 2020b. Active fault-tolerant control for discrete vehicle active suspension via reduced-order observer. In: *IEEE transactions on systems, man, and cybernetics: systems* DOI 10.1109/TSMC.2020.2964607.

He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE, 770–778.

HemaChitra HS, Suguna S. 2018. Optimized feature extraction and classification technique for indian pulse seed recognition. *International Journal of Computer Engineering & Applications* **12**:421–427.

Hinton G, Vinyals O, Dean J. 2015. Distilling the knowledge in a neural network. ArXiv preprint. arXiv:1503.02531.

Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. 2017. Mobilenets: efficient convolutional neural networks for mobile vision applications. ArXiv preprint. arXiv:1704.04861.

Huang G, Liu Z, Laurens V, Weinberger KQ. 2017. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE, 4700–4708.

Huang K-Y. 2012. Detection and classification of areca nuts with machine vision. *Computers & Mathematics with Applications* **64(5)**:739–746 Advanced Technologies in Computer, Consumer and Control DOI 10.1016/j.camwa.2011.11.041.

Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. 2016. Squeezenet: alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. ArXiv preprint. arXiv:1602.07360.

Kayacan MC, Sofu MM, Cetisli BO. 2016. Design of an automatic apple sorting system using machine vision. *Computers and Electronics in Agriculture* **127**:395–405.

Krizhevsky A, Sutskever I, Hinton GE. 2012. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* **25**:1097–1105.

Kurtulmuş F. 2021. Identification of sunflower seeds with deep convolutional neural networks. *Food Measure* **15**:1024–1033 DOI 10.1007/s11694-020-00707-7.

Li H, Kadav A, Durdanovic I, Samet H, Graf HP. 2016. Pruning filters for efficient convnets. ArXiv preprint. arXiv:1608.08710.

Li X, Dai B, Sun H, Li W. 2019. Corn classification system based on computer vision. *Symmetry* **11(4)**:591 DOI 10.3390/sym11040591.

Liang F, Tian Z, Dong M, Cheng S, Sun L, Li H, Chen Y, Zhang G. 2021. Efficient neural network using pointwise convolution kernels with linear phase constraint. *Neurocomputing* **423**:572–579 DOI 10.1016/j.neucom.2020.10.067.

Lin M, Chen Q, Yan S. 2013. Network in network. ArXiv preprint. arXiv:1312.4400.

Lin M, Ji R, Wang Y, Zhang Y, Zhang B, Tian Y, Shao L. 2020. Hrank: filter pruning using high-rank feature map. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Piscataway: IEEE, 1529–1538.

Liu D, Ning X, Li Z, Yang D, Li H, Gao L. 2015. Discriminating and elimination of damaged soybean seeds based on image characteristics. *Journal of Stored Products Research* **60**:67–74 DOI 10.1016/j.jspr.2014.10.001.

Liu P, Huda MN, Sun L, Yu H. 2020. A survey on underactuated robotic systems: bio-inspiration, trajectory planning and control. *Mechatronics* **72**:102443 DOI 10.1016/j.mechatronics.2020.102443.

Liu P, Yu H, Cang S. 2019. Adaptive neural network tracking control for underactuated systems with matched and mismatched disturbances. *Nonlinear Dynamics* **98(2)**:1447–1464 DOI 10.1007/s11071-019-05170-8.

Liu Z, Li J, Shen Z, Huang G, Yan S, Zhang C. 2017. Learning efficient convolutional networks through network slimming. In: *Proceedings of the IEEE international conference on computer vision*. 2736–2744.

Liu Z, Xu J, Peng X, Xiong R. 2018. Frequency-domain dynamic pruning for convolutional neural networks. In: *Proceedings of the 32nd international conference on neural information processing systems*. 1051–1061.

Ma N, Zhang X, Zheng H-T, Sun J. 2018. Shufflenet v2: practical guidelines for efficient cnn architecture design. In: *Proceedings of the European conference on computer vision (ECCV)*. 116–131.

Mebatsion HK, Paliwal J, Jayas DS. 2013. Automatic classification of non-touching cereal grains in digital images using limited morphological and color features. *Computers & Electronics in Agriculture* **90(90)**:99–105 DOI 10.1016/j.compag.2012.09.007.

Mellempudi N, Kundu A, Mudigere D, Das D, Kaul B, Dubey P. 2017. Ternary neural networks with fine-grained quantization. ArXiv preprint. arXiv:1705.01462.

Ni C, Wang D, Vinson R, Holmes M, Tao Y. 2019. Automatic inspection machine for maize kernels based on deep convolutional neural networks. *Biosystems Engineering* **178**:131–144 DOI 10.1016/j.biosystemseng.2018.11.010.

Rehman TU, Mahmud MS, Chang YK, Jin J, Shin J. 2019. Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Computers and Electronics in Agriculture* **156**:585–605 DOI 10.1016/j.compag.2018.12.006.

Ren S, He K, Girshick R, Sun J. 2015. Faster r-cnn: towards real-time object detection with region proposal networks. ArXiv preprint. arXiv:1506.01497.

Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. 2018. Mobilenetv2: inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE, 4510–4520.

Sankararaman KA, De S, Xu Z, Huang WR, Goldstein T. 2020. The impact of neural network overparameterization on gradient confusion and stochastic gradient descent. In: *International conference on machine learning*. PMLR, 8469–8479.

Simonyan K, Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. ArXiv preprint. arXiv:1409.1556.

Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D. 2016. Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience* **2016**:3289801 DOI 10.1155/2016/3289801.

Sun L, Zhao C, Yan Z, Liu P, Duckett T, Stolkin R. 2019. A novel weakly-supervised approach for rgb-d-based nuclear waste object detection. *IEEE Sensors Journal* **19**(**9**):3487–3500.

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. 2015. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE, 1–9.

Veeramani B, Raymond JW, Chanda P. 2018. Deepsort: deep convolutional networks for sorting haploid maize seeds. *BMC Bioinformatics* **19**(**9**):1–9 DOI 10.1186/s12859-017-2006-0.

Wang X, Yu SX. 2020. Tied block convolution: leaner and better cnns with shared thinner filters. ArXiv preprint. arXiv:2009.12021.

Wu S, Wang Z, Shen B, Wang J-H, Dongdong L. 2020. Human–computer interaction based on machine vision of a smart assembly workbench. *Assembly Automation*.

Xie S, Girshick R, Dollár P, Tu Z, He K. 2017. Aggregated residual transformations for deep neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE, 1492–1500.

Zagoruyko S, Komodakis N. 2016. *Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer*. ArXiv preprint. arXiv:1612.03928.

Zhang R. 2019. Making convolutional networks shift-invariant again. In: *International conference on machine learning*. PMLR, 7324–7334.

Zhang X, Zhou X, Lin M, Sun J. 2018. Shufflenet: an extremely efficient convolutional neural network for mobile devices. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE, 6848–6856.