Semantic Segmentation of Germinated Oil Palm Seeds Based on Deep Convolutional Neural Networks with a Novel Channel Attention Mechanism

Sihang Gao
University of Nottingham
Xin Duan
Beijing Institute of Technology
Iman Yi Liao (Iman.Liao@nottingham.edu.my)
University of Nottingham Malaysia
Mohammad Fakhry Jelani
Applied Agricultural Resources Sdn. Bhd
Zi Yan Chen
Advanced Agriecological Research Sdn. Bhd
Choo Kien Wong
Advanced Agriecological Research Sdn. Bhd
Wei Chee Wong
Advanced Agriecological Research Sdn. Bhd

Research Article

Keywords: Oil palm seeds, Deep learning, Semantic segmentation, Attention mechanism

Posted Date: June 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3050864/v1

Abstract

Oil palm is a high value crop with an estimated 5% yearly replanting rate. To dispatch high quality seeds, stringent culling upon the germinated seeds is necessary. The shape of germinated part of an oil palm seed is a key manual criterion for distinguishing good seeds from bad ones. Accurate segmentation of the germinated part would serve as an important preprocessing step for automatic phenotypic analysis and quality classification of germinated oil palm seeds. In this paper, we pioneer the study of semantic segmentation of germinated oil palm seeds by convolutional neural networks (CNNs). Leveraging the state-of-the-art ‘SE-ResNext + U-Net’ architecture for image segmentation, we propose two modifications to address the difficulty of accurately segmenting the germinated part of a seed since it is much smaller compared to the seed body. Firstly, we design local spatial channel attention (LS-SE) to replace the Squeez-Excitation (SE) module to retain local information of a feature channel. Then we pass the features generated by the encoder to the same level of the decoder part twice along the decoding direction (DC-UNet) to retain the original features. This helped address the problem where the edge segmentation details of oil palm seeds require higher-resolution detail information to improve the segmentation accuracy. In addition, the number of parameters of the proposed DC-UNet is much smaller than that of other state-of-the art U-Net variants such as Unet++, significantly reducing the training time. Our proposed DC-UNet with LS-SE obtained an MIOU that is 1.2% higher than U-Net, with a clearly better visual segmentation around the boundaries of the germinated parts.

1 Introduction

Oil palms, which are extensively grown near the equator, are a major source of edible oil. Their economic lifespan is only about 25 years, and it is estimated that 5% of the planted areas will need to be replanted annually to maintain sustainable production of palm oil. To dispatch high quality seeds, stringent culling upon the germinated seeds is necessary. The current classification of palm seeds relies mainly on manual inspection by employees to determine the quality of the germinated part of the seeds. This has high labour costs, requires a large number of workers, and is also time-consuming (Liao et al. 2021). Therefore, automated seed quality classification is desirable to save labour costs and time. In the process of classification, the shape of germinated part is one of the key factors to distinguish the good seeds from the bad ones. Therefore, accurately segmenting the germinated part of a seed may be an important preprocessing step to achieve accurate classification of seed quality.

In recent years, image segmentation has been used increasingly to analyse the morphology and numbers of plant seeds, e.g., semantic segmentation of maize seeds, soybean sprouts. Khoenkaw (2016) used HSV colour space to extract seeds, roots, and coleoptiles in order to count grains of rice and also judge the germination of the crop. Xu et al. (2021) used computer vision and machine learning techniques for the cultivar classification of corn seeds. A classification accuracy of 89%-99% was obtained under different machine learning methods. Kurtulmuş et al. (2016) sampled pepper seeds with an office scanner and then extracted image features representing colour, texture, and shape. Finally, these features were used to classify pepper seeds with a classification accuracy of 84.94%. Lin and Lin (2022) applied a soybean image segmentation method based on multi-scale retinex colour restoration (MSRCR) to segment soybeans which obtained a segmentation accuracy of 98% in terms of the problem of seed overlap and adhesion in the actual quality inspection of soybeans. However, the seed shape and germinated part of these studies are very different from that of oil palm seeds and cannot be directly used for the segmentation of oil palm seeds. As for oil palm seeds, the colour of the germinated part of bad seed is similar to the colour of the seed body, hence it is difficulty to segment the germinated part from the bad seed exactly. In addition, to better classify seeds, higher precision of germinated part segmentation is needed. However, the current segmentation method is not suitable for accurately delineating the germinated part from the seed body.

In recent years, deep learning has greatly developed within the field of image segmentation, which is mainly due to the proposal and development of convolutional neural networks (CNNs) (Albawi et al. 2017; Liu et al. 2019). Image segmentation techniques are fundamental for computer vision tasks. Image segmentation technology is now widely used in transportation, medical, service, and other industries, with applications such as face segmentation, medical image processing, and perception applications (Li, Hao, & Lei 2016). In the agricultural field, segmentation technology is mainly used in the shape analysis of crops, including seed identification (Chitra, Suguna & Sujatha 2016), fruit detection (Mizushima & Lu 2013), weed detection (Manh et al. 2001), and identifying crop diseases (Singh & Misra 2017; Sriwastwa et al 2018). These applications show that image segmentation technology is valuable for agricultural research.

Among deep learning research, attention mechanism has risen as one of the most important research directions. Attention mechanism is a concept of letting neural network to learn where to focus on a given image for a specific task. Chen et al. (2016) proposed an attention mechanism that can soft-weigh different features at different locations in the feature map. Huang et al. (2017) proposed a reverse attention network (RAN) architecture that applies a reverse attention mechanism. Li et al. (2018) proposed a pyramid attention network that utilises global context information for semantic segmentation. To avoid complex dilated convolution and decoder networks, they combined the spatial pyramid and attention mechanism to accomplish accurate dense feature extraction from pixel labels. The SE block suggested by Hu, J. et al. (2018) is a lightweight gating mechanism that is designed to model channel relationships, thereby enhancing the representation of basic modules across the entire network.

The current mainstream attention mechanisms in computer vision include channel attention, spatial attention, and self-attention. Many of the latest attention mechanism methods are also improved based on these three attention mechanisms. In particular, the channel attention, which has a simple structure with remarkable effects, can adaptively adjust the feature response between channels through feature recalibration. It is now widely used in image classification, and image segmentation among other tasks. In addition to the SE block in 2019, Gao et al. (2019) recommended a global second-order pooling (GSoP) block that models high-order statistics while collecting global information, thereby improving the throughput ability of the SE block to gather global information. Gated channel transformation (GCT) as introduced by Yang et al. (2020), is a method that reduces the amount of computation and facilitates the addition of more layers of attention mechanisms to the backbone. Following Qin et al. (2021) submitted the new multi-spectral channel attention, which had a more powerful representation ability.

As the germinated part of an oil palm seed is much smaller than the seed body and the colour of the germinated part of a bad seed is similar to the colour of the seed body, it is, therefore, beneficial to design a segmentation model that is able to retain information at a fine scale both in terms of spatial resolution and colour resolution. We achieve this goal by proposing two modifications to a state-of-the-art image segmentation model, i.e., a U-Net model with ResNext backbone and Squeez-Excitation (SE) attention mechanism (Hu et al. 2018). Firstly, we design local spatial channel attention (LS-SE) to replace the Squeez-Excitation (SE) module to retain local information of a feature channel. Then we pass the features generated by the encoder to the same level of the decoder part twice along the decoding direction (DC-UNet) to retain the original features. This helped address the problem where the edge segmentation details of oil palm seeds require higher-resolution detailed information to improve the segmentation accuracy.

2 Methodology

2.1 Model Structure

The sample images of the oil palm seed dataset in this experiment are unambiguous and have distinct edges, which is convenient when they are manually segmented into different categories. Supervised learning is generally more accurate than unsupervised learning, so the method of supervised learning was chosen to semantically segment the oil palm seeds.

Convolutional neural networks are currently used in supervised learning. Besides, the oil palm seed dataset consists of around 2000 training samples. For deep learning, using a large model with a small sample will likely lead to overfitting. Therefore, a model, with a complex network structure and a large number of parameters would not be suitable for this dataset. The original U-Net has a parameter size of 28M, which is a very lightweight network. Furthermore, even if the amount of data is insufficient, data enhancement methods could be applied, which generally have good adaptability. In summary, the U-Net network model was chosen to complete the task of oil palm seed segmentation. The encoder of the original U-Net network is replaced with a SE-ResNeXt50 encoder. SE-ResNeXt50 absorbs the advantages of VGG/ResNet's stacked repeated network layers and the split-transform-merge strategy of the Inception family. The SE-ResNeXt50 module contains a series of transformations, each of which is based on a part of the feature map. Finally, the outputs of these transformations are fused by addition.

2.2 Local Spatial Channel Squeeze and Excitation

In order to improve the unclear segmentation results brought by U-Net, we introduce the SE-Net network structure, and then analyse the seed mask of U-Net + SE network model segmentation. We found that some edge details segmented by the seeds still lacked clarity.

The U-Net + SE network model inserts a channel attention module after each convolution of each layer in the decoder part. Initially, the Squeeze operation was executed on the feature map to compress the features of the spatial dimension making each two-dimensional feature map 1*1 in size. The Excitation operation was then implemented to generate weights for each feature channel through the parameters. Finally, the weight of each feature channel was attained and applied to each original feature channel to determine the importance of different channels. However, the disadvantage of this model is that the process of being globally average pooled must be applied to the feature map at the beginning. Unfortunately, global average pooling causes the feature map to losing information, resulting in some image details not being well segmented.

To mitigate this, local average pooling was enacted on the feature map, then the Excitation operation was performed, and finally, was re-spliced (LS-SE). The insertion position of this module in the neural network is shown in Fig. 1. The blue arrows in the figure signify the convolution and activation functions, the black arrows are the attention module, and the grey arrows show replication and show the skip connection which is used for feature fusion. The green arrows display the upsampling and convolution; the orange arrows represent downsampling; the yellow arrows signify convolution operations with a kernel of 1×1. The overall structure is comparable to the U-Net network, which is also an Encoder-Decoder structure. The only difference is that the attention module which was redesigned, was inserted in the decoder part. As shown in Fig. 2, the feature map was first divided into four parts from the centre by length and width, and the feature map changed from the original [C, H, W] size to four [C, H/2, W/2] sizes up. It uses the global average pooling method to use the global spatial feature of each channel as the representation of the channel to form as a channel descriptor, and the size of the map was also changed from the original [C, H/2, W/2] to [C, 1, 1]. Then, the corresponding weight of each channel was obtained, and a new feature map was produced according to weight. That is, two 1×1×1 convolutions were used to process information and generate a C-dimensional vector. Then the sigmoid function was used for normalisation to get the corresponding weight for each channel. According to the significance of different channels, the weight was applied to the feature channels of the original feature map. Through these steps, the information-calibrated feature map was thus established. The above operation was performed on the previous four feature maps so that they had been calibrated by the information. The four feature maps were re-spliced together according to their original positions to form a new feature map for further processing. The advantage of this is that from the original global average pooling only the ‘number of channels’ features was retained. It can now retain several features of the four channels, increasing the features and making the segmentation more refined.

2.3 Double Concat U-Net

In addition to the improvements made to the channel attention mechanism, an analysis of the U-Net network structure itself was conducted. The U-Net network, like most segmentation networks, uses the first half (left) for feature extraction, and the second half (right) for upsampling. The structure can also be called an Encoder-Decoder structure. U-Net performs four maximum pooling downsampling and after each sampling, convolution is used to extract information for the feature map. Four rounds of upsampling are performed to restore the input pixel size. However, the most critical and characteristic part of the U-Net is its skip connection. It was found that using skip connection can provide a great level of feature information for later image segmentation, including the details of segmentation edges, thus improving the overall segmentation effect.

Therefore, a new U-Net variant was designed, namely a double concat U-Net (DC-UNet) based on the previous U-Net network. As shown in the figure above, the blue arrows in the figure represent convolution and activation functions, and the grey arrows represent replication (skip connection), which is used for feature fusion. The green arrows in the diagram are the upsampling and convolution, with the orange arrows indicating downsampling. Finally, the yellow arrows represent the convolution operations with a kernel of 1x1. In order to maintain the original U-Net network structure, this network combines the feature map that was obtained from the last convolution where each layer was coded. The feature map was derived from the first convolution result of the decoding of each layer on the channel. Skip connection was applied, and then a second convolution was performed to generate the final feature map for each layer. The information of feature maps is particularly relevant to segmentation tasks. In the encoder part of the network, the pooling layer reduces the resolution of the feature maps. The lack of resolution is not conducive to mask generation.

A U-Net with a double-layer skip connection structure can introduce shallower convolutional layer features via two skip connections. Those features have higher resolution and shallow layers, which will contain relatively rich convolutional layers. The feature information of the encoder is appropriate for generating segmentation masks. During the decoding process, the network can retain high-resolution detailed information by fusing more feature information with the feature map. This is better than simply using the U-Net framework as it facilitates great image accuracy in segmentation. Correctly defining the edges in images of oil palm seeds when they are segmented requires high-resolution and detailed information. In addition, the number of parameters for DC-UNet is smaller than that of other U-Net variants such as Unet++ (Zhou et al. 2018), which can reduce training time.

2.4 DC-Unet + LS-SE

The diversity of shapes found in palm seeds coupled with the irregularity of their germination poses new challenges for palm seed segmentation. Inspired by pioneering research on U-Net and the attention mechanism, we proposed that this DC-UNet network and LS-SE attention mechanism address the above problems. Finally, we integrate the two methods we proposed to obtain the DC-UNet + LS-SE network structure (as shown in Fig. 4), so that the model can better identify the difference between the seed body and the germination part, as well as the different shapes they produce. This is because they can better extract global and local context to circumvent the problem of shape diversity.

3 Implementation and Result

3.1 Experiment Preparation

5.1.1 Dataset

In this paper, in terms of data set production, the 'labelme' software was used first of all to divide the oil palm seeds into two parts, namely, the seed body and the parts of it that had germinated. In total there are 1752 training set images, including 901 positive samples containing oil palm seed regions and 851 negative samples containing oil palm seed regions. The test set consists of 401 images, of which 201 are positive samples that contain oil palm seed regions and 200 are negative samples therein. Each image is 256×256 in size and is divided into three parts: background, seed body, and germinated part. The germinated parts of some of the positive sample images are relatively short, however, others exceed the original size of the picture, and some have two or even three germinated parts that need to be marked. Among the negative sample images, there are some that do not exhibit signs of germination, and in others, the germinated part has become multiple branches. The training set was randomly divided into five parts, 80% of the training set pictures were used for training, 20% were used for verification, and five-fold cross-validation was completed. Positive and negative samples were randomly assigned. Five results were obtained for training and testing respectively for comparison. In addition, it is also necessary to test the seed sample images under different lights to check the generalisation of the model.

The images segmented by the 'labelme' software were saved in .json files, which were then converted to .png files through the code, to generate a corresponding mask for each original image. Some examples of original images and their corresponding masks are shown below:

To verify the validity of the scheme proposed in this chapter, two types of experiments were to analyse the enhancement effect of the model from the perspective of quantitative and qualitative analysis. This chapter mainly analyses the visual effects of images after segmentation in terms of qualitative analysis. From a quantitative point of view, five sets of training sets will be put through cross-validation and consequently taken out. Then, each group of models with the highest verification results after training will be tested. Finally, the MIOU values of the test results will be compared to verify how effective the new method is.

3.1.2 Experimental Configuration and Parameter Settings

The seed segmentation network was implemented through the PyTorch framework, and all network models were trained from scratch. Batch_size = 4 for training set and batch_size = 2 for validation set were used. All experiments were performed on NVIDIA 2060 GPU; the processor is an AMD Ryzen 7 4800H; its memory is 16G and OS is 64-bit Windows 10.

The parameter settings in the ‘DICE LOSS’ loss function were experimentally acquired to determine the paramount configuration combination. Additionally, the Adam optimiser is a suitable optimiser for use in convolutional neural networks as it can handle a wide variety of problems. It has a capacity for fine-tuning which means that it can achieve more than satisfactory results swiftly. For these reasons the

Adam optimiser was used during the training process. The initial learning rate was 1∗10 − 4, step_size = 1, gamma = 0.9, that is, the learning rate of each epoch was multiplied by 0.9, and finally the validation set was taken after 40 iterations. The group of models with the highest IOU value will be used as the final test model.

To verify the effectiveness of the five-stage curriculum training strategy proposed in this chapter, the proposed method was contrasted with other related learning modalities under the same seed segmentation network and within the same experimental setup. The methods of these comparisons include: 1) Use some basic network models for training and compare the test results to find the most suitable model. 2) In the second stage of learning, the cSE channel attention mechanism was added after the convolution of each layer during the decoding element of U-Net for learning. The results of this were then noted. 3) During the third stage of learning the new attention module (LS-SE) was added to the U-Net network in the same location as in the second stage to verify the effectiveness of the new attention module. 4) Only the Unet + + network was used for the fourth stage of learning to garner results. 5) The new network structure (DC-UNet) was applied in the fifth stage of learning, and the test results were used to verify the effectiveness of the new network structure.

3.1.3 Evaluation Indicators

When thinking about evaluation indicators, Mean Intersection over Union (MIOU) is a standard measure in evaluation metrics in semantic segmentation. Due to its popularity, MIOU was used as our evaluation metric.

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}}$$

1

Among them, p_{ij}predicts i as j, which is false negative (FN); p_{ji} predicts j as i, which is false positive (FP); p_{ii}predicts i as i, which is true (TP).

3.6 Experimental Model Selection

For experimental model selection, as for the choice of experimental model, there is no relevant research on the semantic segmentation direction of germinated palm seeds. Therefore, in terms of model selection, we tried a number of currently relatively common models for training and testing. We applied U-Net network structure (Ronneberger et al. 2015), PSP-NET network structure (Zhao et al. 2017), MA-NET network structure, DeepLab v3 network structure (Chen et al. 2017) with SE-ResNeXt as backbone. From Table 1, we can find that the training speed of the PSP-NET network structure is faster, but the final accuracy rate is lower than that of the U-NET network. The MIOU obtained by the MA-NET and DeepLab v3 networks is similar to that obtained by the U-NET network, but the training time is much longer than that of the U-NET network. Therefore, after comparing the training time and the final MIOS, we selected the U-NET network as our final experimental network.

Table 1
MIOU and time achieved by U-Net, PSP-Net, MA-Net, and DeepLab v3 models respectively when tested on the 'Palm Seeds' dataset.

	U-Net	**PSP-Net**	**MA-Net**	**DeepLab v3**
MIOU (%)	93.067	92.418	93.117	92.831
Time (s)	4536	2700	7300	21600

3.7 U-Net + LS-SE Experiment of Oil Palm Seeds Data Set

The segmentation MIOU is shown in Table 2. The new attention mechanism method proposed in this paper was quantitively analysed on the average index: the proposed U-Net + LS-SE model segmented the germinated part and seed body of oil palm seeds. Compared with the traditional U-Net network structure, the average result of MIOU is improved from 93.31–92.99%. The model was used again in the same manner and compared with the U-Net + SE model, the average result of the secondary MIOU improved from 93.31–92.89%.

Figure 6 shows the semantic segmentation results of the three methods mentioned above in the bad oil palm seed dataset. Through this, the efficacy of the proposed U-Net + LS-SE model can be scrutinised. The first column in the figure is the result of the segmentation of oil palm seed germinated parts by the traditional U-Net network, and the second column is the result of the segmentation by the U-Net + SE model. The third column applies to the suggested U-Net + LS-SE model. It can be seen from the figure that all three models can approximately segment the germinated parts of seeds. However, the traditional U-Net network segmented the edges relatively roughly, and could not fit the edges sufficiently well. The colour of the germinated parts of bad seeds is similar to the colour of the seed itself, and the traditional U-Net network demonstrated that it could only identify parts with large colour differences. After adding the cSE channel attention mechanism, the segmentation effect in terms of some of the bad seeds that had germinated was better than that of the traditional U-Net network.

Nonetheless, there was still some deviations from the actual shape. On the other hand, the proposed U-Net + LS-SE model, as shown in the graph, demonstrated a better fitting effect on the edges of the germinated seeds. Moreover, the model can identify the geminated parts of bad seeds that are similar in colour to the seed body efficiently, and perform more accurate segmentation. The non segmentation part of Fig. 7 has been cut out to provide a clearer comparison of the quality of the segmentation results of different models.

Table 2
MIOU achieved by U-Net, U-Net + SE, and U-Net + LS-SE models respectively when tested on the 'Palm Seeds' dataset.

		First Group (%)	**Second Group (%)**	**Third Group (%)**	**Fourth Group (%)**	**Fifth Group (%)**	**Average (%)**
Seed Body	U-Net	97.043	**97.119**	96.991	96.742	97.072	96.993
	U-Net + cSE	**97.075**	96.785	96.998	**97.053**	96.984	96.975
	U-Net + LS-cSE (Ours)	97.057	97.088	**97.125**	97.035	**97.089**	**97.079**
Germinated Part	U-Net	89.029	89.016	89.473	88.737	88.698	88.991
	U-Net + cSE	88.029	88.677	88.440	89.335	89.529	88.802
	U-Net + LS-cSE (Ours)	**89.374**	**89.508**	**89.759**	**89.335**	**89.697**	**89.535**
Overall	U-Net	93.036	93.067	93.232	92.740	92.885	92.992
	U-Net + cSE	92.552	92.731	92.719	**93.185**	93.256	92.889
	U-Net + LS-cSE (Ours)	**93.216**	**93.298**	**93.442**	**93.185**	**93.393**	**93.307**

3.8 DC-UNet Experiment of Oil Palm Seeds Data Set

Next, the new network structure (DC-UNet) proposed in this paper was quantitively analysed. The proposed DC-UNet network structure segmented the germinated part and seed body of palm seeds in five rounds. Compared with the U-Net network structure, the average MIOU result improved from 92.99–93.42%. Compared with the Unet + + network structure, the results improved from 93.27% to 93,42. In addition, the training time required to train the germinated parts and seed body of palm seeds with the proposed DC-UNet network structure is about 4950 seconds, while taking 10600 seconds for the Unet + + network structure. Therefore, the training time required for the training using the proposed DC-UNet network structure was much less than that required by the U-Net + + network. In this way, the proposed DC-UNet network structure greatly reduces the training time, improves the efficiency, and refines the MIOU.

Figure 8 shows the semantic segmentation results of the U-Net, U-Net++, and DC-UNet network structures with the bad oil palm seed dataset. The first column in the figure is the result of the segmentation of the germinated parts of the seeds by the traditional U-Net network. The second column is the result of the U-Net + + network's segmentation, and the third column shows the results of the proposed DC-UNet network. It can be seen from the figure that all three models can roughly segment the germinated parts of the seed. The segmentation results of the U-Net + + network on the germinated parts of some bad seeds were better than those of the traditional U-Net network, as shown in the segmentation results. However, the U-Net + + network requires a long training time and has relatively low efficiency. The segmentation results for the proposed DC-UNet network in terms of the edges of germinated oil palm seeds are similar to the segmentation results of the U-Net + + network, with both having better fitting effects than the traditional U-Net network segmentation. In the segmentation results of the germinated parts of some bad seeds, it is also found that the proposed DC-UNet network can better identify the edges of germinated parts that have a similar colour to the seed body. To better visualise the segmentation effect of different models, the segmentation map of the germinated parts was enlarged in Fig. 9 for clearer comparison of the segmentation quality of the different models. 3.5 DC-UNet + LS-SE Experiment of Oil Palm Seeds Data Set

Table 3
MIOU obtained by U-Net, Unet++, and DC-UNet models tested on the 'Palm Seeds' dataset.

		First Group (%)	Second Group (%)	Third Group (%)	Fourth Group (%)	Fifth Group (%)	Average (%)
Seed Body	U-Net	97.043	97.119	96.991	96.742	97.072	96.993
	Unet++	97.020	97.058	97.114	97.055	97.092	97.068
	DC-Unet (Ours)	**97.053**	**97.173**	**97.178**	**97.204**	**97.324**	**97.186**
Germinated Part	U-Net	89.029	89.016	89.473	88.737	88.698	88.991
	Unet++	88.524	89.529	**89.674**	90.193	**89.416**	89.467
	DC-Unet (Ours)	**89.044**	**90.285**	89.501	**90.294**	89.366	**89.698**
Overall	U-Net	93.036	93.067	93.232	92.740	92.885	92.992
	Unet++	92.772	93.293	**93.393**	93.624	**93.254**	93.267
	DC-Unet (Ours)	**93.048**	**93.729**	93.340	**93.749**	93.245	**93.422**

3.9 DC-UNet + LS-SE Experiment of Oil Palm Seeds Data Set

We combined the LS-SE module with the DC-UNeT, and the resulting DC-Unet + LS-SE network structure segmented the germinated part and seed body of palm seeds in five rounds. The germinated part is the most important part for judging whether palm seeds are good or bad. As shown in Table 4, compared with the U-Net network structure, The average result of MIOU for the germination part improved from 88.99–90.15%. Compared with the U-Net + LS-SE network structure, the results improved by 89.54–90.15%. Compared with the DC-UNet network structure, the results improved by 89.70–90.15%.

Figure 10 shows the semantic segmentation results of the U-Net + LS-SE, DC-UNet, and DC-Unet + LS-SE network structures with the bad oil palm seed dataset. The first column in the figure is the result of the segmentation of the germinated parts of the seeds by the traditional U-Net + LS-SE network. The second column is the result of the DC-UNet network's segmentation, and the third column shows the results of the proposed DC-Unet + LS-SE network. The three models that we proposed can more accurately segment palm tree seeds. However, we can also find that the segmentation effect of using the DC-UNet + LS-SE model on some test pictures is better than that of the DC-UNet and U-Net + LS-SE models. To better visualise the segmentation effect of different models, the segmentation map of the germinated parts was enlarged in Fig. 11 for clearer comparison of the segmentation quality of the different models.

Table 4
MIOU obtained by U-Net + LS-SE, DC-UNet, and DC-Unet + LS-SE models tested on the 'Palm Seeds' dataset.

		First Group (%)	Second Group (%)	Third Group (%)	Fourth Group (%)	Fifth Group (%)	Average (%)
Seed Body	U-Net + LS-SE	**97.057**	97.088	97.125	97.035	97.089	97.079
	DC-Unet	97.053	**97.173**	**97.178**	**97.204**	**97.324**	**97.186**
	DC-Unet + LS-SE	96.918	97.148	97.037	97.074	96.989	97.033
Germinated Part	U-Net + LS-SE	89.374	89.508	89.759	89.335	89.697	89.535
	DC-Unet	89.044	90.285	89.501	90.294	89.366	89.698
	DC-Unet + LS-SE	**89.761**	**90.315**	**89.902**	**90.418**	**90.358**	**90.151**
Overall	U-Net + LS-SE	93.216	93.298	93.442	93.185	93.393	93.307
	DC-Unet	93.048	93.729	93.340	**93.749**	93.245	93.422
	DC-Unet + LS-SE	**93.340**	**93.732**	**93.470**	93.746	**93.674**	**93.592**

3.10 Segmentation of Germinated Oil Palm Seeds Under Different Lighting Conditions

Finally, to discern whether these methods could be applied in practical situations, the segmentation of oil palm seeds under different lighting conditions was tested. The first column in the figure is the segmentation results for the traditional U-Net network. The second column is the segmentation for the U-Net + LS-SE model, the third is the results of the DC-UNet model, and the fourth column is the segmentation results of the DC-Unet + LS-SE network.

The below figure qualitatively analyses the segmentation of oil palm seeds under different lighting environments. When contrasting the different lines on the map with the shape of the 'seed body' and 'germinated part' in the original image, it is evident that the previously tested models can be used with oil palm seed images under different lighting conditions. Good segmentation of the seed body and germinated parts of the palm seed is demonstrated. At the same time, the images segmented by the U-Net + LS-SE and DC-UNet networks perform better in regards to the edge details of both the 'seed body' and 'germinated part' when compared to the image segmented by the U-Net network. The final DC-UNet + LS-SE network structure is the best. Therefore, based on the qualitative analysis of the experimental results, it can be concluded that using the proposed DC-UNet + LS-SE network for segmenting images of oil palm seeds under different lighting conditions yields better results than using the traditional U-Net network. Additionally, the application of a convolutional neural network for segmenting oil palms seeds is feasible in practical work.

4 Conclusion

In terms of innovation, this paper proposes a pioneering study on the semantic segmentation direction of germinated palm seeds. The deep learning method of a convolutional neural network was applied to the semantic segmentation of oil palm seeds, achieving respectable segmentation results. In this paper, we propose the LS-SE method based on the channel attention mechanism of SE-Net. Based on the current U-Net network, a new U-Net variant network was designed, namely the double concat U-Net (DC-UNet). The MIOU effect obtained by this network in tests was enhanced compared to that of the U-Net network. Additionally, the segmentation accuracy of the edge segmentation details of the oil palm seeds was higher. The number of parameters of the proposed DC-UNet model is much smaller than that of the U-Net variants such as U-Net++, which can greatly reduce the training time. Finally, we integrated the two methods we proposed to get the DC-UNet + LS-SE network structure and obtained the best segmentation performance.

Furthermore, this is a brand-new private dataset that has not been annotated before. As such, the entire dataset was manually annotated, providing a suitable dataset for future research on palm seeds. In the field of agriculture, distinguishing between good and bad seeds has been an important research direction. At present, to complete this task, workers have to manually inspect the germinated part of the seeds to judge their quality. If seed quality could be automatically detected by machine learning algorithms, it could greatly save on labour costs and improve the speed of detection. The size and shape of the seed and the length and shape of the germinated part are among the most important predictors of whether a seed will develop normally. No similar oil palm seed segmentation experiment has been conducted before. Through related papers on other seed segmentation in the agricultural field, the use of the convolutional neural network deep learning method to semantically segment the seed body and germinated part of oil palm seeds was proposed.

The fact that palm seeds are irregular in shape, the germinated parts of good seeds and bad seeds are very different in colour, and the germinated parts of bad seeds are similar in colour to the seed body collectively causes poor segmentation. Although the training time of the U-Net network is fast, its segmentation accuracy needs to be improved. The U-Net + + network training accuracy rate is relatively high, but its training time is long, and there are many parameters. Therefore, in response to the above problems, a new U-Net variant network, the double concat U-Net (DC-UNet) was designed based on the current U-Net network structure. The MIOU score obtained by this network test was better than that of the U-Net network, and the segmentation accuracy of the edge details of the oil palm seeds was higher. In addition, the proposed DC-UNet has fewer parameters than that of U-Net deformation structures such as U-Net++, which can greatly reduce the training time.

At present, the channel attention mechanism mentioned in the SE-net network has been widely used, but due to the global average pooling, it loses a lot of feature map information, meaning that some image details cannot be well preserved. This leads to a deviation in some of the results. To address this problem, the method of local average pooling, convolution, and final re-splicing of feature maps based on the SE-net channel attention mechanism (LS-SE) was introduced. The advantage of this is that only the 'channel number' is retained from the original global average pooling. Through this, it is possible to retain several features in four channels, so that the overall number of retained features is increased, and the details of the segmentation are more obvious. Compared with the original SE-net, the MIOU results obtained by the test showed improvement.

Finally, we combined both DC-UNet and LS-SE in a network structure and obtained a segmentation result map with more obvious segmentation details and better MIOU results, which laid a good foundation for the future palm tree seed classification task.

This research is not without its limitations, and some problems need to be further explored:

(1) Although two improved methods for semantic segmentation of the seed body and germinated part of palm seeds have been proposed, and the final DC-UNet + LS-SE network structure has been obtained by fusion, the model was unable to identify the shape of the germinated parts of some seeds that are similar in colour to the seed body, resulting in a certain bias within the segmentation results.

(2) When applying the segmentation model to the oil palm seed images under different lighting conditions, it was found that the shadows cast by the lighting caused the model to segment parts of the shadow with the seed body. Therefore, differentiation between the two is one of the research directions that needs to be considered.

Declarations

Acknowledgements

The research was partially supported by Kuala Lumpur Kepong Berhad and Boustead Plantations Berhad. Iman Yi Liao would like to acknowledge the support by the Malaysian Ministry of Education for the grant awarded under the Fundamental Research Grant Scheme (No. FRGS/1/2014/ICT07/UNIM/02/1).

References

1. Albawi, S., Mohammed, T.A., Al-Zawi, S.: ‘Understanding of a convolutional neural network’. In *2017 international conference on engineering and technology (ICET)* (pp. 1–6). Ieee. (2017), August
2. Brostow, G.J., Fauqueur, J., Cipolla, R.: Semantic object classes in video: A high-definition ground truth database. Pattern Recognit. Lett. **30**(2), 88–97 (2009)
3. Chen, L.C., Yang, Y., Wang, J., Xu, W., Yuille, A.L.: ‘Attention to scale: Scale-aware semantic image segmentation’. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3640–3649). (2016)
4. Chitra, H.S.H., Suguna, S., Sujatha, S.N.: A survey on image analysis techniques in agricultural product. Indian J. Sci. Technol. **9**(12), 1–13 (2016)
5. Gao, Z., Xie, J., Wang, Q., Li, P.: ‘Global second-order pooling convolutional networks’. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3024–3033). (2019)
6. Huang, Q., Xia, C., Wu, C., Li, S., Wang, Y., Song, Y., Kuo, C.C.J.: (2017). ‘Semantic segmentation with reverse attention’. arXiv preprint arXiv:1707.06426.
7. Khoenkaw, P. (2016). ‘An image-processing based algorithm for rice seed germination rate evaluation’. *In 2016 International Computer Science and Engineering Conference (ICSEC)* (pp. 1–5). IEEE.
8. Kurtulmuş, F., Alibaş, İ., & Kavdır, I. (2016). ‘Classification of pepper seeds using machine vision based on neural network’. *International Journal of Agricultural and Biological Engineering*, 9(1), 51–62.
9. LeCun, Y., Bengio, Y., & Hinton, G. (2015). ‘Deep learning’. *nature*, 521(7553), 436–444.
10. Li, H., Xiong, P., An, J., & Wang, L. (2018). ‘Pyramid attention network for semantic segmentation’. arXiv preprint arXiv:1805.10180.
11. Li, Y., Hao, Z., & Lei, H. (2016). ‘Survey of convolutional neural network’. *Journal of Computer Applications*, 36(9), 2508.
12. Liao, I. Y., Shen, B. S. K., Chen, Z. Y., Jelani, M. F., Wong, C. K., & Wong, W. C. A (2021). ‘A Preliminary Study on Germinated Oil Palm Seeds Quality Classification with Convolutional Neural Networks’. *In: 7th workshop on Computer Vision in Plant Phenotyping and Agriculture.*
13. Lin, W., & Lin, Y. (2022, June). ‘Soybean image segmentation based on multi-scale Retinex with color restoration’. *In Journal of Physics: Conference Series* (Vol. 2284, No. 1, p. 012010). IOP Publishing.
14. Liu, X., Deng, Z., & Yang, Y. (2019). ‘Recent progress in semantic image segmentation’. *Artificial Intelligence Review*, 52(2), 1089–1106.
15. Manh, A. G., Rabatel, G., Assemat, L., & Aldon, M. J. (2001). ‘AE—Automation and emerging technologies: Weed leaf image segmentation by deformable templates’. *Journal of agricultural engineering research*, 80(2), 139–146.
16. Mizushima, A., & Lu, R. (2013). ‘An image segmentation method for apple sorting and grading using support vector machine and Otsu’s method’. *Computers and electronics in agriculture*, 94, 29–37.

17. Qin, Z., Zhang, P., Wu, F., & Li, X. (2021). 'Fcanet: Frequency channel attention networks'. *In Proceedings of the IEEE/CVF international conference on computer vision* (pp. 783–792).
18. Singh, V., & Misra, A. K. (2017). 'Detection of plant leaf diseases using image segmentation and soft computing techniques'. *Information processing in Agriculture*, 4(1), 41–49.
19. Sriwastwa, A., Prakash, S., Swarit, S., Kumari, K., & Sahu, S. S. (2018). 'Detection of pests using color based image segmentation'. *In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)* (pp. 1393–1396). IEEE.
20. Xu, P., Yang, R., Zeng, T., Zhang, J., Zhang, Y., & Tan, Q. (2021). 'Varietal classification of maize seeds using computer vision and machine learning techniques'. *Journal of Food Process Engineering*, 44(11), e13846.
21. Yang, Z., Zhu, L., Wu, Y., & Yang, Y. (2020). 'Gated channel transformation for visual recognition'. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11794–11803).
22. Hu, J., Shen, L., & Sun, G. (2018). 'Squeeze-and-excitation networks'. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132–7141).
23. Ronneberger, O., Fischer, P., & Brox, T. (2015). 'U-net: Convolutional networks for biomedical image segmentation'. *In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18* (pp. 234–241). Springer International Publishing.
24. Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). 'Pyramid scene parsing network'. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2881–2890).
25. Fan, T., Wang, G., Li, Y., & Wang, H. (2020). 'Ma-net: A multi-scale attention network for liver and tumor segmentation'. *IEEE Access*, 8, 179656–179665.
26. Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). 'Rethinking atrous convolution for semantic image segmentation'. arXiv preprint arXiv:1706.05587.
27. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). 'Unet++: A nested u-net architecture for medical image segmentation'. *In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4* (pp. 3–11). Springer International Publishing.

Figures

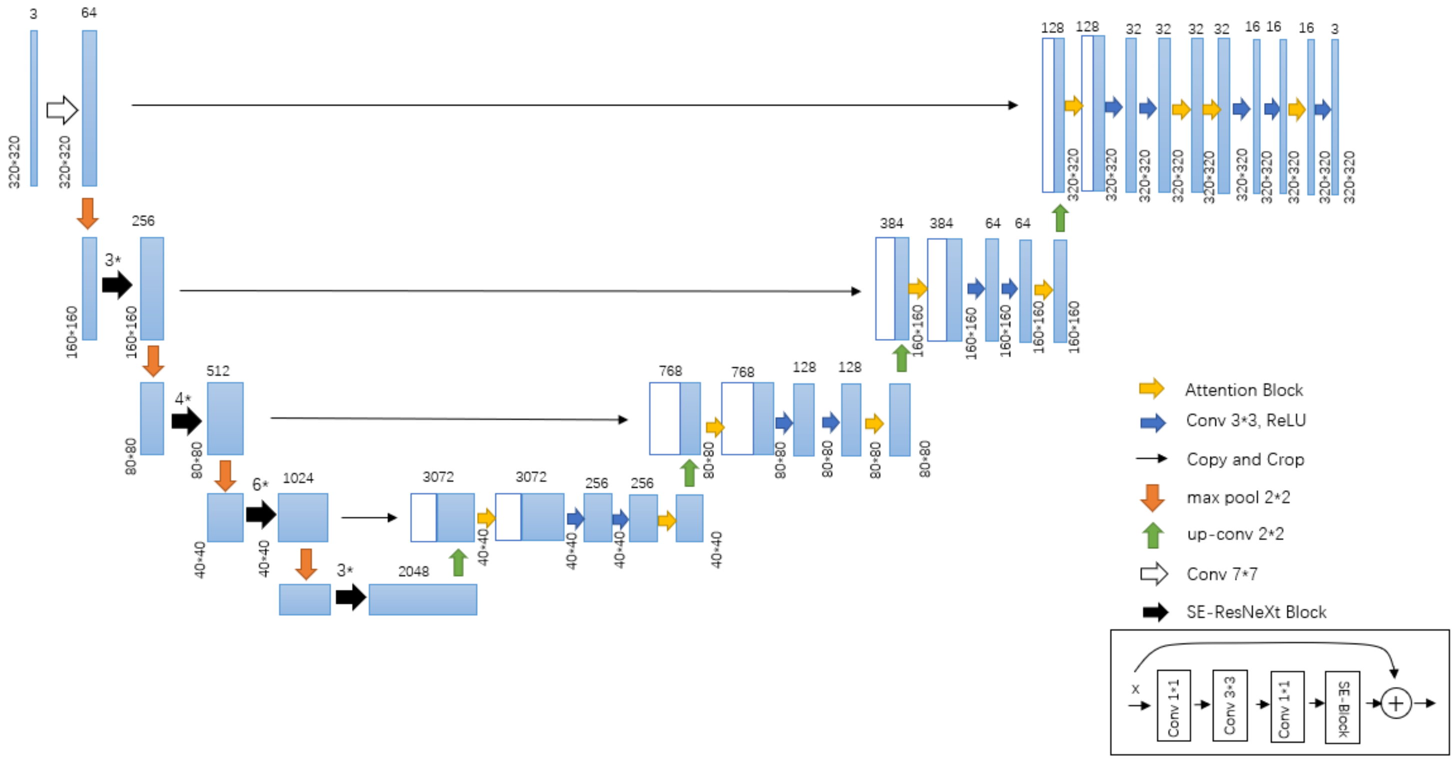

Figure 1

The proposed U-Net + LS-SE network structure

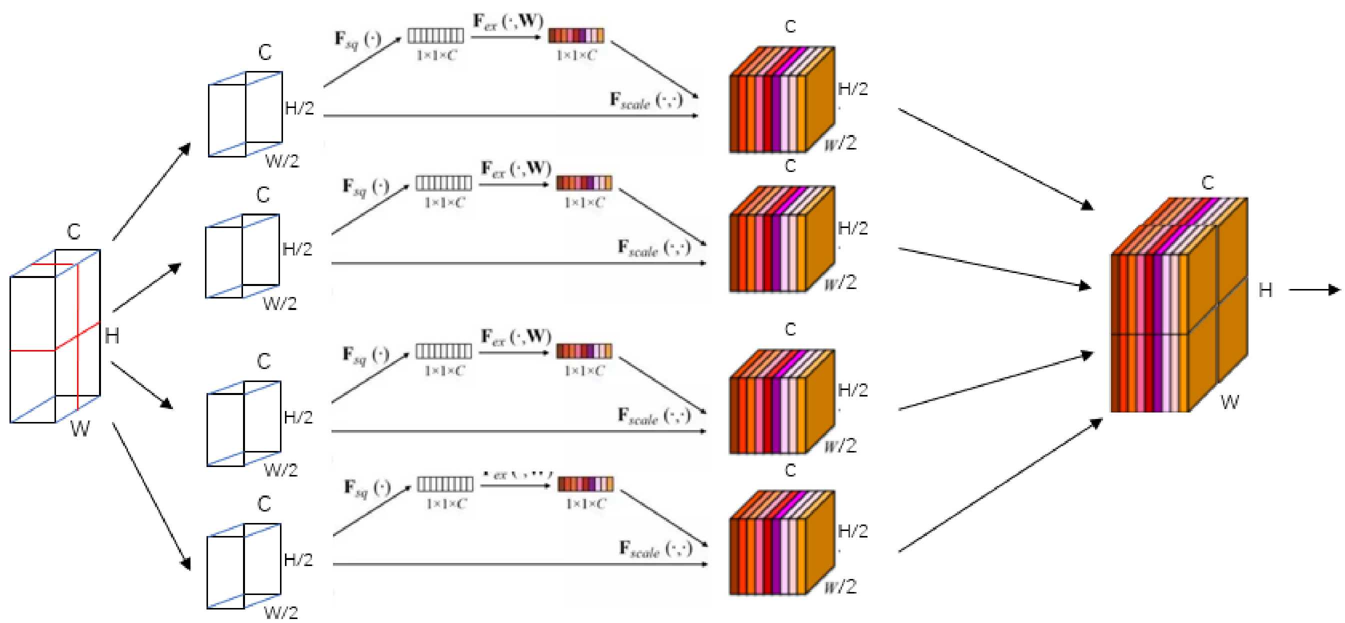

Figure 2

The proposed LS-SE structure

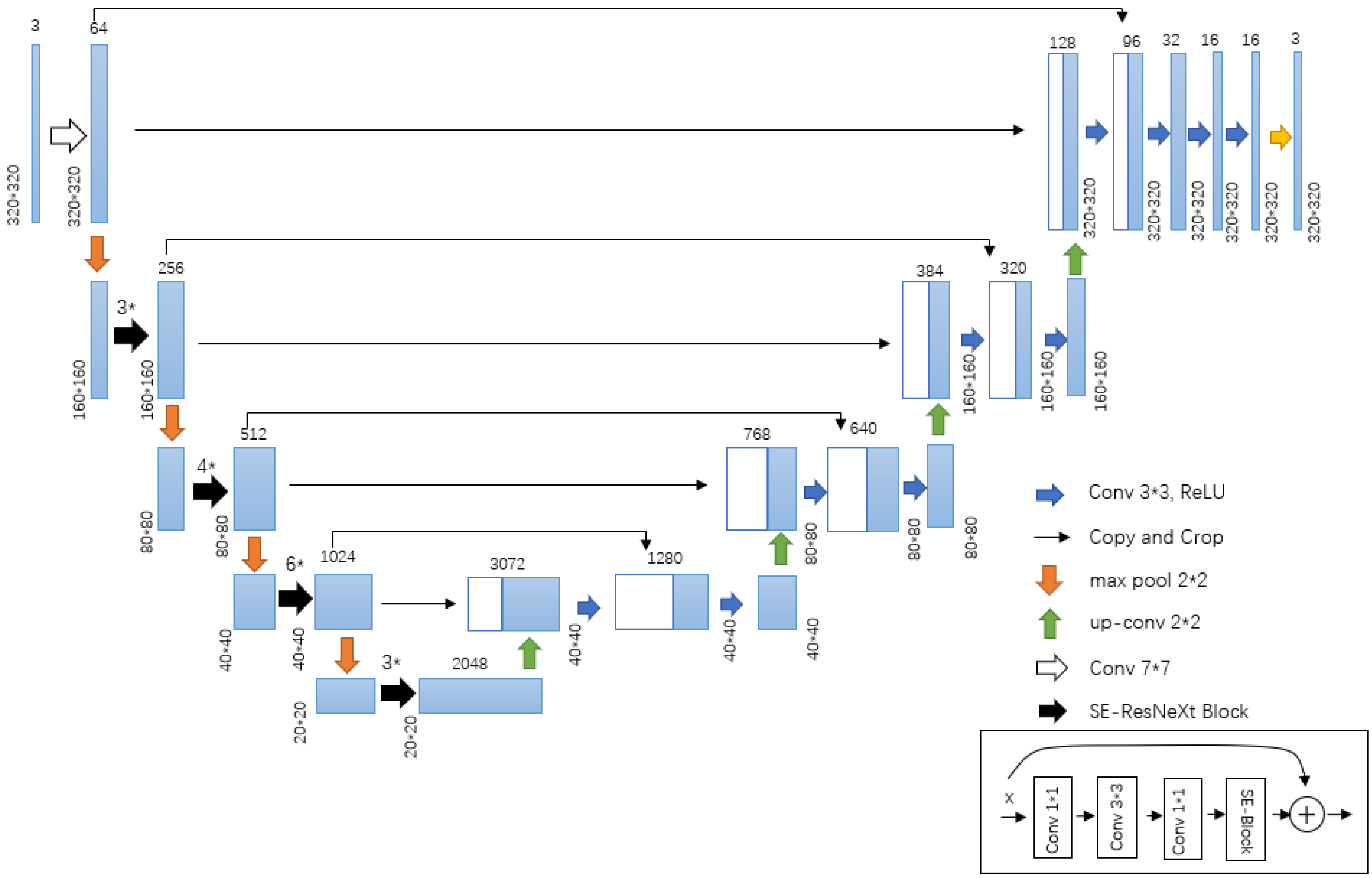

Figure 3

The proposed DC-UNet network structure.

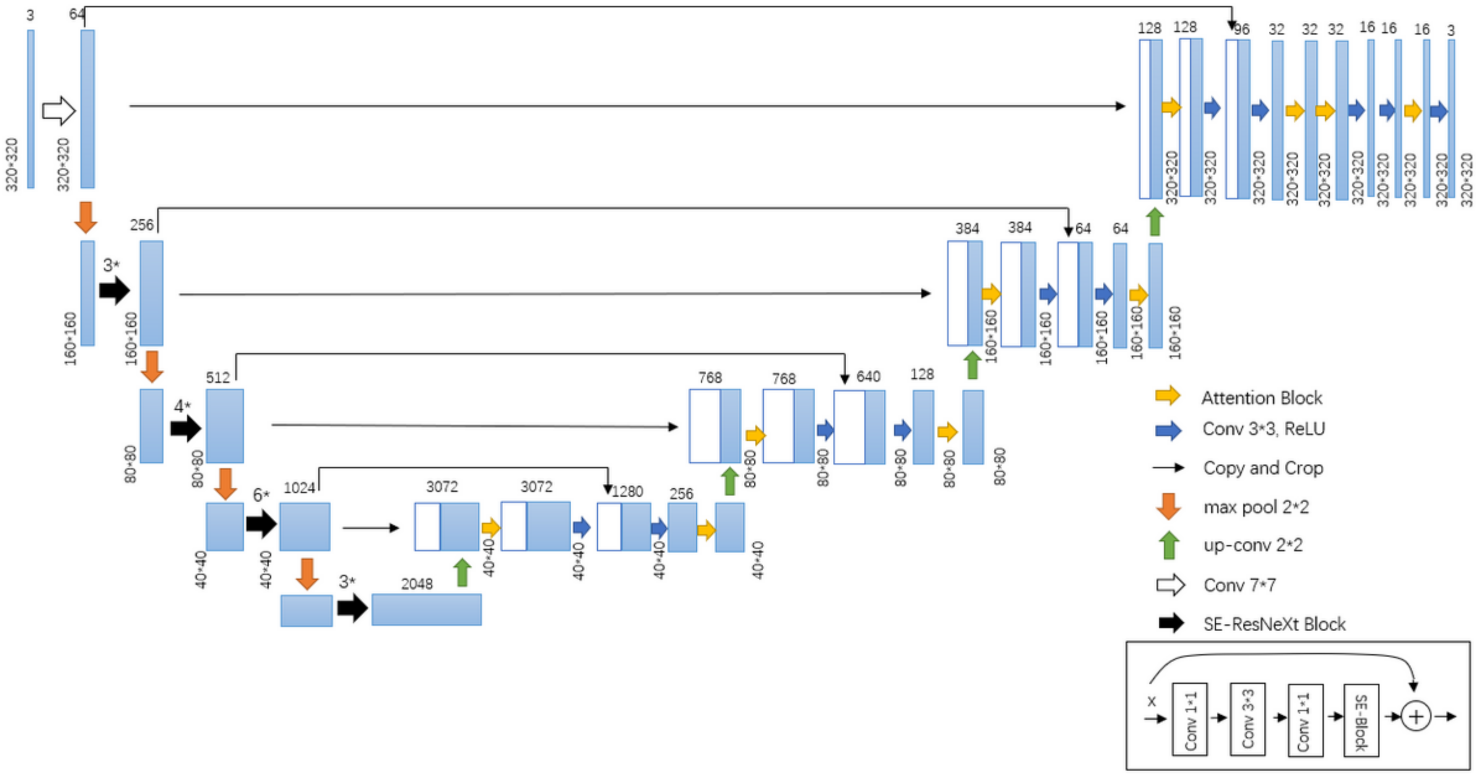

Figure 4

The proposed DC-UNet + LS-SE network structure.

Figure 5

The original image and the corresponding annotation image in the oil palm seed dataset.

Figure 6

Segmentation map of the three models tested on the 'Palm Seeds' dataset (red line is for ground truth and green line is for prediction).

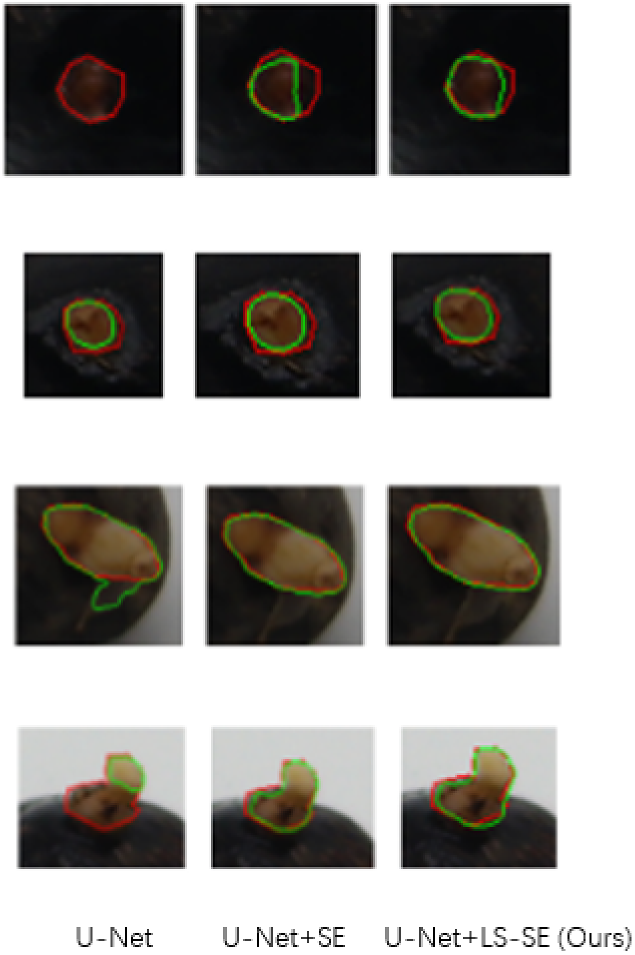

Figure 7

The enlarged version of the segmentation map of the segmentation parts obtained by the three model tests on the 'Palm Seeds' dataset (red line is for ground truth and green line is for prediction).

Figure 8

Segmentation map of the three models tested on the 'Palm Seeds' dataset (red line is for ground truth and green line is for prediction).

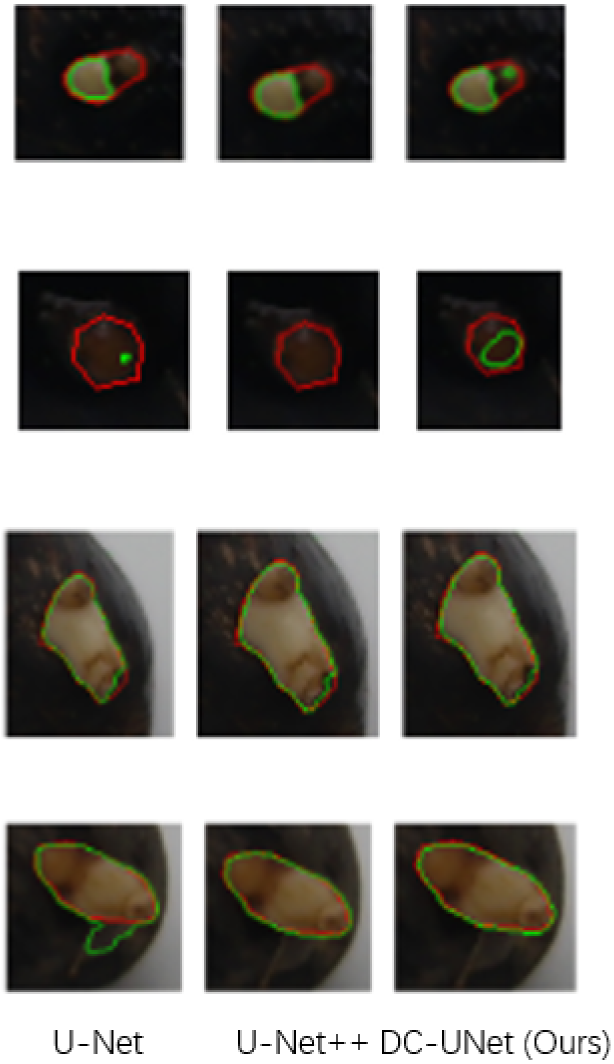

Figure 9

The enlarged version of the segmentation map of the segmentation parts obtained by the three model tests on the 'Palm Seeds' dataset (red line is for ground truth and green line is for prediction).

Figure 10

Segmentation map of the three models tested on the 'Palm Seeds' dataset (red line is for ground truth and green line is for prediction).

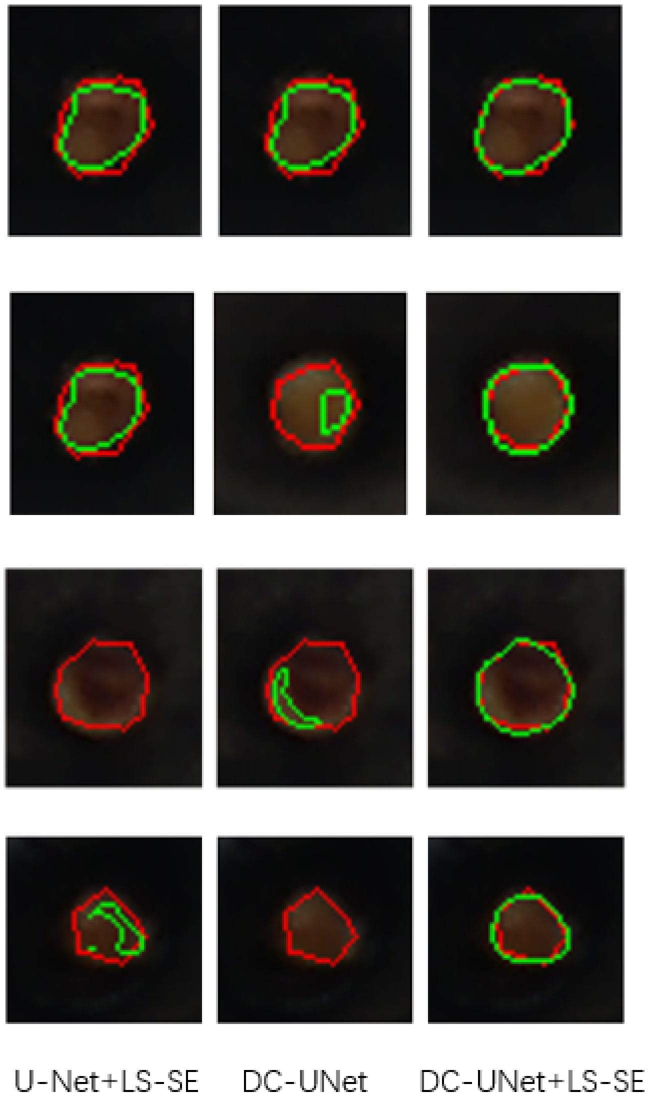

Figure 11

The enlarged version of the segmentation map of the segmentation parts obtained by the three model tests on the 'Palm Seeds' dataset (red line is for ground truth and green line is for prediction).

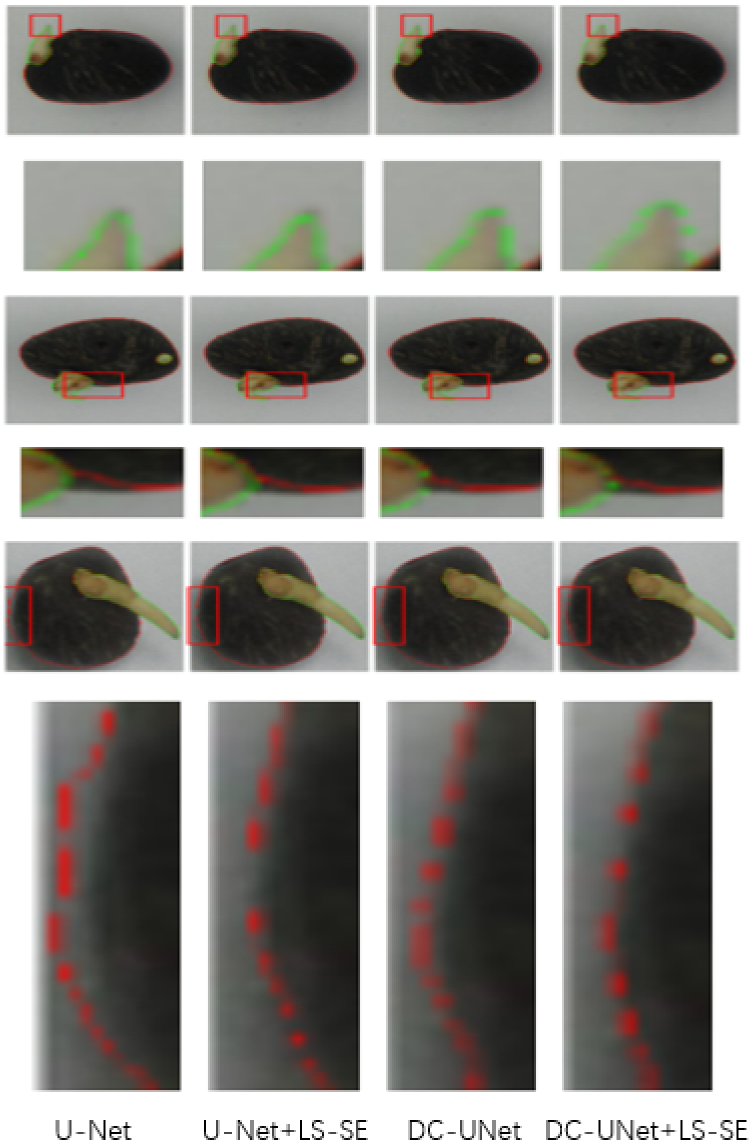

Figure 12

Segmentation of oil palm seeds under different lighting conditions (green line is for 'germinated part' and red line is for 'seed body').