Article

Real-Time Identification of Rice Weeds by UAV Low-Altitude Remote Sensing Based on Improved Semantic Segmentation Model

Yubin Lan ^{1,2,†}, Kanghua Huang ^{1,2,†}, Chang Yang ^{2,3}, Luocheng Lei ^{2,3}, Jiahang Ye ^{2,3}, Jianling Zhang ^{2,3}, Wen Zeng ^{2,3}, Yali Zhang ^{2,3} and Jizhong Deng ^{2,3,*}

¹ College of Electronic Engineering, South China Agricultural University, Guangzhou 510642, China; ylan@scau.edu.cn (Y.L.); kh_huang@stu.scau.edu.cn (K.H.)

² National Center for International Collaboration Research on Precision Agricultural Aviation Pesticides Spraying Technology, Guangzhou 510642, China; changyang@stu.scau.edu.cn (C.Y.); llc@stu.scau.edu.cn (L.L.); yjh@stu.scau.edu.cn (J.Y.); jlingzhang@scau.edu.cn (J.Z.); zengwen@scau.edu.cn (W.Z.); ylzhang@scau.edu.cn (Y.Z.)

³ College of Engineering, South China Agricultural University, Guangzhou 510642, China

* Correspondence: jz-deng@scau.edu.cn

† These authors contributed equally to this work and should be considered as co-first authors.

Citation: Lan, Y.; Huang, K.; Yang, C.; Lei, L.; Ye, J.; Zhang, J.; Zeng, W.; Zhang, Y.; Deng, J. Real-Time Identification of Rice Weeds by UAV Low-Altitude Remote Sensing Based on Improved Semantic Segmentation Model. *Remote Sens.* **2021**, *13*, 4370. <https://doi.org/10.3390/rs13214370>

Academic Editors:

Michael Schirrmann and
Mario Cunha

Received: 23 August 2021

Accepted: 25 October 2021

Published: 30 October 2021

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

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Real-time analysis of UAV low-altitude remote sensing images at airborne terminals facilitates the timely monitoring of weeds in the farmland. Aiming at the real-time identification of rice weeds by UAV low-altitude remote sensing, two improved identification models, MobileNetV2-UNet and FFB-BiSeNetV2, were proposed based on the semantic segmentation models U-Net and BiSeNetV2, respectively. The MobileNetV2-UNet model focuses on reducing the amount of calculation of the original model parameters, and the FFB-BiSeNetV2 model focuses on improving the segmentation accuracy of the original model. In this study, we first tested and compared the segmentation accuracy and operating efficiency of the models before and after the improvement on the computer platform, and then transplanted the improved models to the embedded hardware platform Jetson AGX Xavier, and used TensorRT to optimize the model structure to improve the inference speed. Finally, the real-time segmentation effect of the two improved models on rice weeds was further verified through the collected low-altitude remote sensing video data. The results show that on the computer platform, the MobileNetV2-UNet model reduced the amount of network parameters, model size, and floating point calculations by 89.12%, 86.16%, and 92.6%, and the inference speed also increased by 2.77 times, when compared with the U-Net model. The FFB-BiSeNetV2 model improved the segmentation accuracy compared with the BiSeNetV2 model and achieved the highest pixel accuracy and mean Intersection over Union ratio of 93.09% and 80.28%. On the embedded hardware platform, the optimized MobileNetV2-UNet model and FFB-BiSeNetV2 model inferred 45.05 FPS and 40.16 FPS for a single image under the weight accuracy of FP16, respectively, both meeting the performance requirements of real-time identification. The two methods proposed in this study realize the real-time identification of rice weeds under low-altitude remote sensing by UAV, which provide a reference for the subsequent integrated operation of plant protection drones in real-time rice weed identification and precision spraying.

Keywords: low-altitude remote sensing; semantic segmentation model; real-time; rice weeds; target spraying

1. Introduction

Monitoring farmland weeds and taking timely control measures are conducive to the growth of crops and the increase in yield. In recent years, aerial remote sensing with Unmanned Aerial Vehicle (UAV) as a carrier has been widely used in farmland weeds monitoring due to its high spatial resolution, safe and convenient operation, and

low economic cost [1–3]. At the same time, image processing methods based on deep learning have made good progress in image classification, object detection, and instance segmentation recently [4–6]. The feature extraction method of the deep learning method is different from the previous machine learning method of artificially selecting features [7–9]. The deep learning method is to automatically complete feature selection and extraction from the massive labeled data through the learning algorithm, which enables it to obtain better generalization in complex agricultural analysis tasks [10–13]. Therefore, deep learning technology is increasingly applied to the recognition of agricultural scenes by UAVs [14–16], especially in the weeds recognition of offline UAV low-altitude remote sensing images, which has achieved remarkable results. Torres-Sánchez J. et al. [17] used UAV to collect RGB and multi-spectral remote sensing data of sunflower fields, and analyzed the different performance in Normalized Difference Vegetation Index (NDVI) between weeds, crops, and bare soil at different flight altitudes. The results proved that the difference in NDVI of each category is the most obvious at a flight altitude of 30 m. Huang HS. et al. [18–20] proposed to apply the Full Convolutional Neural (FCN) network to the research of rice weed recognition in UAV remote sensing images, and proved that this method can generate the weed coverage map of UAV remote sensing images. In order to improve the recognition accuracy of the rice weeds recognition model for UAV low-altitude remote sensing, Huang HS. et al. improved the classic Convolutional Neural Network (CNN) by combining the methods such as the Atrous Convolution and the fully connected Conditional Random Field (CRF). Further, the semantic annotation method is used to generate accurate weed distribution maps of high-resolution UAV remote sensing images. Zhu S. et al. [21] used UAV to obtain low-altitude remote sensing images of rice fields, and compared the classification effects on rice weeds of the three machine learning algorithms, Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and AdaBoost and the CNN. The results proved that the CNN has the best effect among these algorithms, with a classification accuracy of 92.41%.

In order to achieve high accuracy, deep learning image recognition requires significant computational and memory resources [22–24]. Running on an airborne UAV, the deep learning inference takes seconds to minutes. Given UAV have limited battery capacity that constrains flight time, these delays limit the number of missions that can be completed using real-time processing. While prior work has developed computer systems that execute these workloads efficiently, it is important to devise deep learning approaches that use as few computational resources as possible [25–28].

With the improvement of embedded hardware computing performance and the emergence of lightweight neural network architecture, it is possible to deploy deep learning models on embedded hardware to realize real-time recognition of farmland weeds. At present, lightweight models mainly include SqueezeNet [29], MobileNet [30] series and ShuffleNet [31] series. These lightweight neural networks reduce the amount of network parameters by designing more efficient convolution units without losing network performance, which greatly improves the calculation speed of the network. These network models and their improved structures have achieved good results in the fields of image classification in recent years [32–34], which can be used as a basic backbone network in the task of deep learning semantic segmentation.

The main objective of this study is to research the real-time identification of rice weeds by UAV low-altitude remote sensing. Combining the deep learning semantic segmentation technology with the high-performance embedded hardware, we proposed two improved rice weeds recognition models MobileNetV2-UNet and FFB-BiSeNetV2. Firstly, we introduced the lightweight MobileNetV2 feature extraction basic network to improve the U-Net [35] model which has a complex network structure to explore the operating efficiency of the model, and built the MobileNetV2-UNet model. In addition, we improved the real-time semantic segmentation model BiSeNetV2 [36] by introducing a feature fusion branch and the depthwise separable convolution to explore the segmentation accuracy of the model, and built the FFB-BiSeNetV2 model. Finally, we transplanted the improved models to the embedded device Jetson AGX Xavier to verify the real-time

performance of rice weeds recognition for exploring the feasibility of real-time recognition on the airborne end of UAV.

2. Materials and Methods

2.1. The Technical Route of the Study

In this study, the technical route of the research on real-time identification of rice weeds based on UAV low-altitude remote sensing is shown in Figure 1. The main process is described as follows:

1. The UAV equipped with an RGB camera was used to collect low-altitude remote sensing visible image data of rice fields, and then the data preprocessing and enhancement methods in Section 2.2.2 was used to process the remote sensing data. Further, the training set and validation set were divided for the training of the semantic segmentation models.
2. The deep learning semantic segmentation networks U-Net and BiSeNetV2 and their improved structures MobileNetV2-UNet and FFB-BiSeNetV2 were used to identify and analyze rice weeds in UAV remote sensing images. The training set and validation set data was used to train the semantic segmentation models. Further, through the evaluating indicators analysis, the original models and the improved models were compared the segmentation accuracy, operation efficiency, and segmentation effect in the UAV low-altitude remote sensing rice weeds images, and the optimal rice weeds segmentation model that meets the actual identification accuracy and real-time performance requirements was selected for further verification.
3. A low-altitude remote sensing rice weed recognition system was developed to verify the practicality of the optimal rice weeds segmentation model.
4. The improved semantic segmentation models were transplanted to the high-performance embedded device, the models structure was optimized through model quantization and the TensorTR [37] acceleration tool to improve inference speed of the models.
5. The collected rice field video data were used to verify the real-time recognition performance and compare the segmentation effect of the improved models, which will provide a technical basis for further airborne real-time identification of rice weeds.

Figure 1. The technical route for real-time identification of weed in rice fields.

2.2. Image Acquisition and Preprocessing

2.2.1. Image Acquisition

This research was carried out at the paddy field test base, Lubu Town, Gaoyao District, Zhaoqing City, Guangdong Province, in China (112.184548N, 23.134322E). In the experiment, a DJI P4 Multispectral UAV was used to collect RGB images of the rice fields in the test area. The RGB camera of the UAV has a resolution of 1300×1600 pixels, the UAV's flying height is 12 m, and the flying speed is 3.5 mph, the ground resolution is 0.6 cm/px, and the UAV route planning of the rice field image collection is completed in DJI GS PRO software and the image overlap rate was set to 50% forward overlap and 40% side overlap. According to field investigations, the seeding of rice in the experimental area was directly broadcasted by drones, the rice was growing in the tillering stage, and the herbicide was not sprayed in the rice fields. The weeds in the fields mainly include barnyard grass, *leptochloa chinensis*, etc. The weeds and rice have distinct color characteristics. The collection area of UAV remote sensing images is shown in Figure 2, the UAV took aerial photographs of seven test plots numbered ①~⑦, the size of each field was $30 \text{ m} \times 70 \text{ m}$, and a total of 660 RGB UAV remote sensing images were collected, including 84 UAV remote sensing images of test field numbered ③ as test set samples, and 576 images of other test fields as training set samples which were expanded by means of data enhancement. Some experimental UAV remote sensing images are shown in Figure 3.

Figure 2. UAV remote sensing image collection area.

Figure 3. Part of UAV remote sensing images.

2.2.2. Image Data Preprocessing

An open source annotation tool-LabelMe [38] was used to semantically label the collected UAV remote sensing images and establish a standard semantic label data set. The generated label images corresponds to the original images one-to-one, which provides a reliable data basis for further research. In terms of data processing, in order to keep the pixel resolution of local image features unchanged, each original UAV remote sensing image and a label image with a resolution of 1300×1600 (shown in the left of Figure 4a) was cut into four images and four label images with a resolution of 650×800 (shown in the right of Figure 4a) to increase the number of samples in the data set and to reduce the size of input images for deep learning models. At the same time, data enhancement methods including image flipping, rotation, and image brightness conversion were used to expand the data set and increase the noise data to improve the generalization ability and enhance the robustness of the models. The data enhancement effect is shown in Figure 4b. The expanded samples set have a total of 3854 images, which were randomly divided into training set and validation set for model training at a ratio of 9:1. The number of valid label categories is three, which are weed, rice, and others.

Figure 4. Image clipping and data enhancement. (a) Image clipping; (b) Data enhancement.

2.3. Construction of Rice Weeds Identification Model

2.3.1. Depth-wise Separable Convolution

Traditional convolutional neural networks use large-size convolution kernels to obtain larger receptive fields, however, this method also greatly increases the amount of model parameters. In 2017, Google proposed MobileNet [30], a lightweight model dedicated to embedded mobile devices, which is characterized by the use of a large number of depthwise separable convolutions instead of standard convolutions, as shown in Figure 5. The depthwise separable convolution integrates the traditional convolution into a two-step operation: a 3×3 depthwise convolution (DWC) and a 1×1 pointwise convolution (PWC) in two steps.

Assuming that the size of the input feature maps is $D_h \times D_w$, the number of channels of the feature maps is M , the size of the convolution kernel is D_k , and there are a total of N convolution kernels, so the ratio of the amount of computation between the depth separable convolution and the traditional convolution is:

$$\frac{D_h \cdot D_w \cdot M \cdot D_k \cdot D_k + D_h \cdot D_w \cdot M \cdot N}{D_h \cdot D_w \cdot M \cdot N \cdot D_k \cdot D_k} = \frac{1}{N} + \frac{1}{D_k^2} \quad (1)$$

In practice, the depthwise separable convolution generally uses a 3×3 convolution kernel. Through Equation (1), the calculation amount of depthwise separable convolution is about 1/9 of traditional convolution. It can be seen that using depthwise separable con-

volution to replace standard convolution can significantly reduce the calculation amount of the model, which also provides a theoretical basis for the realization of lightweight structure of the deep learning semantic segmentation networks.

Figure 5. Standard convolution and depthwise separable convolution. (a) Standard convolution; (b) Depthwise separable convolution.

2.3.2. The MobileNetV2-UNet Semantic Segmentation Model Structure

The classic semantic segmentation network U-Net model uses an encoding-decoding framework, and uses the VGG-Net [39] network as its basic feature extraction network, which can extract effective feature information. However, the model runs a large amount of calculation due to the complex network structure and excessive weight parameters, which is difficult to meet the real-time identification requirements of embedded devices. MobileNetV2, proposed by Mark Sandler [40], is a lightweight convolutional neural network developed for mobile terminals. MobileNetV2 is improved on the basis of the MobileNet network. Its main feature is that the inverse residual structure reduces the loss of low-dimensional spatial information, while also combined with depthwise separable convolution, which reduces the network parameters, thereby speeding up the calculation of the overall network, making it easier to meet the real-time requirements of the embedded platform. This article combines the MobileNetV2 network to improve the U-Net semantic segmentation model, and builds the MobileNetV2-UNet model. The model framework is shown in Figure 6. The coding part of the framework (the left side of Figure 6) is used as the basic network for feature extraction, retaining the first and second convolutional layers of the original U-Net (as shown in Figure 6(C1~C2)) to obtain more shallow features, and using the 17-layer inverted residual block of MobileNetV2 (as shown in Figure 6(B1~B17)) to replace the traditional VGG-Net convolutional network as the feature extraction backbone network of the model. The purpose is to reduce the amount of model parameter calculation, improve the feature information extraction efficiency of the model, and reduce the loss of information caused by image compression; the decoding part (right side of Figure 6) is a feature recovery process, of which the feature recovery layers are combined with the extracted features from the coding part through multiple concatenate operations, as shown in Figure 6(CN1~CN5), and through the upsampling operation, the compressed feature map is gradually restored to the size of original input images, and finally it is input to the Softmax layer for the output of the prediction results.

Figure 6. The structure of the MobileNetV2-UNet semantic segmentation model proposed in this study.

2.3.3. The FFB-BiSeNetV2 Semantic Segmentation Model Structure

Current real-time semantic segmentation algorithms mostly obtain faster processing speed by sacrificing shallow detail information, which often leads to a serious decrease in segmentation accuracy. The bilateral network model BiSeNetV2 proposed by Yu et al. [36] in 2020 can effectively balance the accuracy and the speed of image semantic segmentation. Its architecture is mainly composed of two different feature extraction branches: Detail Branch and Semantic Branch. The detail branch is used to capture shallow spatial detail information, and the semantic branch captures high-level semantic context information through rapid down sampling. Finally, the output features of the detail branch and the semantic branch are integrated and restored through the bilateral guided aggregation layer (Bilateral Guided Aggregation Layer) to output the prediction results; in addition, without increasing the complexity of network inference, the architecture adopts a boost training strategy to improve the accuracy of network training. Although BiSeNetV2 has greatly improved the inference speed and recognition accuracy, since the size of the feature maps output in the aggregation layer is only 1/8 of the original image size, when outputting the prediction results, it need to restore the feature maps to the original image size through an eight-times bilinear interpolation up-sampling method. In this process, the compressed feature maps do not make full use of the semantic context information and spatial detail information of different scales to restore, which to a certain extent causes the segmentation results of the predicted images roughness and lack of segmentation boundary information.

The FFB-BiSeNetV2 model structure based on multi-feature fusion is improved on the basis of the original BiSeNetV2 model. The improved framework structure is shown in Figure 7. The main improvement methods are as follows.

- Introduce a module with depthwise separable convolution in the detail branch to realize the lightweight of the branch structure.
- Introduce the Feature Fusion Branch, and integrate the output results from the aggregation layer with the feature information from different scales of semantic branch and detail branch through concatenate operations.
- Add the convolution output of the feature fusion branch to the boost training strategy to optimize the calculation of the loss function, so that make the overall segmentation result of the network more accurate.

Figure 7. Improved semantic segmentation model structure of FFB-BiSeNetV2.

The specific structure and improvement process of the framework are described as follows.

1. Improvements in the detail branch

The detail branch retains rich image spatial detail information through a shallow and wide-channel structure. The structure is shown in the blue dashed box in Figure 7. This branch is divided into three stages for feature extraction. The original BiSeNetV2 mainly uses the idea of VGG-Net for stacking layer output. In the original model structure, each stage performs two layers of 3×3 standard convolution operations. Each convolution layer in the stages is connected with Batch Normalization (BN) and a Relu activation function. The stride of the first convolution layer in each state is two, so the size of the feature maps output by this branch is $1/8$ of the original images. In this study, we combine the advantages of depth separable convolution to improve the detail branch on the premise that the receptive field of the feature map remains unchanged, which aims to reduce the parameter calculation amount of the detail branch. The main improvement method is to design a bottleneck module similar to the MobileNetV2 inverse residual structure as the feature extraction layer at each stage of the detail branch. The first layer of the bottleneck module maps the input feature map to the high-latitude space through a 1×1 standard convolution, and then connect two layers of depthwise convolutions to perform channel-by-channel convolution on the output feature maps of the first 1×1 convolution. The first depthwise convolution is set to a convolutional stride of 2, which is used to reduce the size of the output feature maps and expand the range of the receptive field. Finally, point-by-point convolution is performed on the feature map through a 1×1 convolution layer and the convolution result is output.

2. Introduction of the feature fusion branch

In the feature restoration stage, in order to combine more scales of feature information of semantic context and spatial detail, we design a feature fusion branch structure (shown in the purple dashed box in Figure 7), which receives the output from the bilateral guided aggregation layer (BGA Layer, as shown in the orange dashed box in Figure 7) for feature restoration. The main function of the BGA Layer is to achieve effective communication

between the semantic branch and the detailed branch through the two-way multi-scale guidance fusion of the two branches feature information. The feature fusion branch mainly have three stages. At each stage, the output feature maps of the semantic branch and the detail branch that have the same scale as the input feature maps are merged through a concatenation operation, and then it uses a 3×3 convolution to extract the fused features. Finally, the size of the feature maps is restored through the deconvolution upsampling operation. In the entire feature restoration process, the convolution output of each stage in the feature fusion branch participates in the booster training strategy (as shown in the yellow dashed box in Figure 7), that is, through the upsampling operation of the Seg Head module, the feature maps with the same dimension as the prediction results are output to supervise the overall training process of the network, thereby increasing the feature representation in the training phase. Since the Seg Head module only participates in the training stage of the network and can be completely discarded in the inference stage, it has less impact on the inference and calculation speed of the network.

3. The loss function

The multi-feature fusion BiSeNetV2 proposed in this study uses the Softmax function as the output layer activation function, which transforms the K-dimensional real vector z output by the network to another K-dimensional real vector $\varphi(z)$, so that each element of the neural network output is mapped to (0, 1) interval, its function expression is:

$$\varphi(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad (2)$$

where $\varphi(z)_j$ represents the j-dimensional real vector $\varphi(z)$, $j = 1, \dots, K$; z_j represents the j-dimensional real vector z , $j = 1, \dots, K$. The cross-entropy function is used as the multi-class loss function, and its mathematical expression is as follows:

$$L_s = -\frac{1}{N} \sum_{n=1}^N \sum_{i=1}^M y_i \ln(\varphi(z)_i) \quad (3)$$

In the formula, N means the number of samples in a batch, M means the total number of prediction categories; y_i means if the current category prediction corresponds to the i -th category label, it is recorded as y_i , otherwise it is equal to 0, a is the predicted probability of the network output by the Softmax function; in addition, due to the booster training strategy used in this article, a series of Seg Head modules are generated to participate in the calculation of the loss value to supervise the overall training and learning process of the network. The total loss function expression is as follows:

$$Loss_{Total} = L_s + \sum_{i=1}^T L_i(x_i) \quad (4)$$

where L_s is the loss function output by the last layer of the network, L_i is the loss function output by the i -th Seg Head module, $i = 1, \dots, T$; x_i is the corresponding input feature value.

2.4. The Computer Platform and the Embedded Hardware Platform of Experiments

The experiment of models comparison and analysis before and after improvement in this study is firstly carried out on the computer platform with a 64-bit Windows 10 system, which is configured with an Intel® Core™ i5-4460 CPU, a GTX1080Ti graphics card and a 16 G running memory. The system is installed with the CUDA10.0 parallel computing framework and the CUDNN7.6 deep learning acceleration library. The semantic segmentation models are trained and quantified on the Tensorflow deep learning framework of the computer, and finally the trained improved model networks are transplanted to the embedded hardware platform NVIDIA Jetson AGX Xavier for models inference and prediction experiments. Xavier uses a 64-bit CPU with an eight-core ARM architecture, a GPU with 512 CUDA cores NVIDIA Volta architecture and 64 Tensor cores, a running memory

of 16 GB. Further, it supports FP32/FP16/INT8 operation precisions, among which the calculation capacity under FP16 precision is 11 TFLOPS, and under INT8 precision it can reach 22 TOPS, which can provide a high-performance hardware computing platform for the deployment of deep learning edge computing. The embedded hardware platform used in this experiment is configured with the same deep learning running environment as the computer platform, on which the embedded transplantation and structure optimization of the rice weeds recognition models are carried out to analyze the real-time recognition performance on the UAV low-altitude remote sensing data.

2.5. Models Training Strategy

In this study, we used the training set remote sensing data processed by the method in Section 2.2.2 to train the models on the PC platform, and the relevant training hyperparameters need to be set before training. To prevent memory overflow and to adjust the input image to fit the size requirements of the network input, the input images were resized to 352×480 pixels. The Adam optimization algorithm with momentum was used to optimize the gradient descent during model training. The exponential decay rate β_1 of the first moment estimation is 0.9, the exponential decay rate β_2 of the second moment estimation is 0.99, and the value of the epsilon ϵ is 10^{-8} . The initial learning rate is set to 0.001, and the batch size is six. The L2 regularization method was used to prevent the model from over fitting. The L2 regularization coefficient decay is 10^{-4} . In this study, the experiment set up a total of 50 epochs for training iterations, each epoch is trained 300 times, and the model training ends after 15,000 times.

2.6. Experimental Evaluation

This experiment evaluates the model performance from three aspects of segmentation accuracy, operation efficiency and structural complexity. The model segmentation accuracy mainly uses the pixel accuracy, the Intersection over Union ratio (IoU) of each category, and the mean Intersection over Union ratio (MIoU) to evaluate the models, the mathematical expressions of each evaluation index are as follows:

$$PA = \frac{\sum_{i=0}^k P_{ii}}{\sum_{i=0}^k \sum_{j=0}^k P_{ij}} \quad (5)$$

$$IoU = \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (6)$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (7)$$

In the above formula, the prediction categories have a total of $k+1$ categories, including an empty category or background category, P_{ii} means the proportion of pixels that are correctly predicted, and P_{ij} means the probability that the pixels of category i are predicted to be category j , P_{ji} means that the probability that the pixels of category j are predicted to be category i .

In terms of the model operation efficiency, the real-time segmentation performance of the model is verified by the evaluating indicator of Frames Per Second (FPS) to calculate the number of frames of predicted images inferred by the model per second and the evaluating indicator of Floating Points of Operations (FLOPs) to calculate number of floating-point calculations of the model to verify the real-time segmentation performance.. The structural complexity is measured by the parameters of the training model and the size of the memory space occupied by the model. Finally, a suitable rice weed recognition model is selected for the next step of embedded transplantation and real-time verification after weighing the evaluation indicators in these three aspects.

2.7. Models Training Parameters

Parameter quantity and inference speed are important indicators to measure the complexity and operating efficiency of deep learning models [13]. When testing the reasoning ability of semantic segmentation models or applying them in actual offline detection situations, the models only need to calculate the prediction result through the forward propagation algorithm, and nodes such as variable initialization, back propagation, auxiliary training, etc. are not required [41]. Saving the training parameters of the deep learning model as a closed serialized Protocol Buffer (PB) file is a process of turning the variable parameters of the model into fixed constants and discarding irrelevant network nodes in the inference stage, thereby the model structure is simplified through this method, which greatly reduces the memory space occupied by the model, and makes the deep learning model more suitable for deployment on embedded platforms. In this study, firstly, the training parameters of the U-Net, BiSeNetV2 and their improved models MobileNetV2-UNet and FFB-BiSeNetV2 are persistently frozen to generate the PB files on the computer platform. Then, the size of the parameters and the efficiency of inference operation on the test set of each model before and after parameter persistent freezing were compared and analyzed.

2.8. Model Optimization and Inference Acceleration

To study the effect of deep learning inference frameworks on the energy efficiency of the improved semantic segmentation models inference, we used the TensorRT tool to optimize model structure and accelerate model inference. NVIDIA TensorRT is a high-performance inference engine of neural networks for deploying deep learning applications in production environments. The operation of the TensorRT optimization model inference process is mainly divided into two parts: reducing the operation precision of floating point type and optimizing the network structure [42]. In the training process, the convolutional neural network needs to continuously update the weight parameters of the convolutional layer through the iterative calculation of the backpropagation algorithm. The calculation precision of this process generally adopts FP32, which is a 32-bit floating point type. In the inference recognition stage, the network model outputs the prediction results through forward propagation, and has no need to update the weights through back propagation. Therefore, using lower-precision weights for inference will not have much impact on the accuracy of the inference results. On the contrary, it can obtain higher reasoning speed and higher throughput. In terms of network optimization, TensorRT uses the vertical integration, horizontal integration, and layer elimination operations to simplify the network structure to improve the utilization of hardware computing resources, as shown in Figure 8.

Figure 8. Optimization of convolutional neural network structure by TensorRT.

3. Results

3.1. Analysis of Training Process of the Semantic Models

To verify the real-time performance of the improved model, the BiSeNetV2 and U-Net models and their improved structures were trained with the same training hyperparameter settings in Section 2.5 and data sets. When the model was trained on the training set, the model training data was saved once every epoch and 300 iterations, and the pixel accuracy, loss value, and the MIoU value of the corresponding verification set were recorded at the same time. According to the recorded training log, the change curve of model training was generated as shown in Figure 9. From the loss value change curve (shown in Figure 9b), the MobileNetV2-UNet model and the U-Net model have little difference in the change of loss value during the training process. The loss value of FFB-BiSeNetV2 is higher than that of BiSeNetV2. This is because FFB-BiSeNetV2 adds the convolution output of the introduced feature fusion branch to the original enhanced training strategy of BiSeNetV2. Therefore, the composition of the loss function is changed, which also enhances the model's ability to learn the data characteristics of the training set to a certain extent. From the pixel accuracy change curve (Figure 9a) and the MIoU change curve (Figure 9c), it can be seen that during the training process, the pixel accuracy and the MIoU of FFB-BiSeNetV2 are slightly higher than BiSeNetV2, which shows that the segmentation and recognition accuracy of FFB-BiSeNetV2 on the verification set is better than that of the original model; the pixel accuracy and the MIoU of MobileNetV2-UNet on the verification set are significantly higher than that of U-Net, which shows that the ability of MobileNetV2-UNet to segment and recognize rice weeds in the UAV images is better than the original model.

Figure 9. Changes of pixel accuracy, loss value and MIoU in the validation set. (a) Changes of the pixel accuracy; (b) Changes of the loss value; (c) Changes of the MIoU in the validation set.

3.2. Recognition Results and Analysis of Rice Weeds

3.2.1. Comparison of Segmentation Accuracy of Different Models

In order to evaluate the segmentation accuracy of the models, we compared U-Net, BiSeNetV2 and their improved models MobileNetV2-UNet and FFB-BiSeNetV2, and the test set images were used to test the evaluation indicator scores of each model in terms of segmentation accuracy. From the results in Table 1, it can be seen that the MobileNetV2-UNet and FFB-BiSeNetV2 improved the pixel accuracy, the mean Intersection over Union ratio and the Intersection over Union ratio of each category compared to the original models U-Net and BiSeNetV2. From the results, the MobileNetV2-UNet model improved the PA by 0.62% and the MIoU was 2.94% higher than the original model U-Net; the PA and MIoU of FFB-BiSeNetV2 reached 93.09% and 80.28%, which were 0.88% and 3.03% higher than the original model BiSeNetV2, and the FFB-BiSeNetV2 model achieved the best recognition effect in the comparative test method of this article.

Table 1. Comparison of segmentation precision of different models.

Methods	Pixel Accuracy/%	Intersection over Union/%			
		Mean Intersection over Union Ratio/%	Weed	Rice	Others
U-Net	92.01	76.52	87.19	87.41	54.94
MobileNetV2-UNet	92.58	78.77	88.02	88.08	60.19
BiSeNetV2	92.28	77.92	87.26	87.78	58.73
FFB-BiSeNetV2	93.09	80.28	88.02	88.46	64.38

3.2.2. Comparison of Model Parameters and Operating Efficiency

In this section, we used the methods in Section 2.7 to freeze the training parameters of the models and compared the performance of the model before and after the parameters freezing. From the results in Table 2, it can be seen that after the model parameters freezing, the MobileNetV2-UNet model and the FFB-BiSeNetV2 model have reduced the model size by 62.75% and 78.67%, and the inference speed has increased by 14% and 25.22% compared with that before parameters freezing. It shows that the persistent freezing operation on the model parameters is beneficial to improve the real-time recognition efficiency of the model. After the parameters are frozen, in the comparison and analysis of the model performance before and after the improvement, the number of parameters, the model size, and the floating points of operations of the MobileNetV2-UNet were greatly reduced compared with the original model U-Net, which were reduced by 89.12%, 86.16%, and 92.6%, the inference speed has also been greatly improved, which was 2.77 times that of the original model; the FFB-BiSeNetV2 model has reduced the model size and the floating points of operations by 10.37% and 3.6% compared with the original model BiSeNetV2, however, the inference speed has dropped. The main reasons for this result are: (1) Synthesizing the method described in Section 2.3.3, due to the use of the Bottleneck structure with depthwise convolution to replace the standard convolutional layer in the detail branch of the original model BiSeNetV2, which reduced the calculation amount of the improved model, thereby the size and the floating points of operations of the model were reduced; (2) The introduction of depthwise separable convolution increased the usage of GPU memory to a certain extent, resulting in a decrease in inference speed [43]. In summary, the MobileNetV2-UNet model has a significant improvement in the operating efficiency and structure complexity of the model due to the introduction of MobileNetV2 as the feature extraction backbone network. Compared with the original model BiSeNetV2, the FFB-BiSeNetV2 model not only improved the segmentation accuracy, but also optimized the model structure. Although the operating efficiency has decreased, the inference speed of the FFB-BiSeNetV2 model has reached 104.17 FPS, which is 16.68% higher than the MobileNetV2-UNet model, with a good real-time recognition performance.

Table 2. Comparison of model parameters and operation efficiency.

Methods	Number of Parameters	Before Freezing			After Freezing		
		Size of Model/MB	Floating Points of Operations/GFLOPs	Speed of Inference/FPS	Size of Model/MB	Floating Points of Operations/GFLOPs	Speed of Inference/FPS
U-Net	93,140,046	357	248.68	31.5	118	248.53	32.26
MobileNetV2-UNet	10,129,614	49.4	18.41	78.31	13	18.4	89.28
BiSeNetV2	16,065,682	98.8	33.84	100.2	9.64	11.1	149.25
FFB-BiSeNetV2	17,394,450	111	40.51	83.19	8.64	10.7	104.17

3.2.3. Segmentation Results of UAV Remote Sensing Image of Rice Weeds

To analyze the segmentation effect of the improved semantic segmentation model proposed in this article, we selected four pictures randomly from the test set samples as the

example pictures, and obtained the segmentation results of rice weeds of different models on the example pictures. As shown in Figure 10, the first column is the original input images, the second column is the semantic label images, and the third to sixth columns are the predicted segmentation results of different models. It can be seen from the figure that the overall segmentation results and the restoration of local details of the MobileNetV2-UNet model were not much different from the original model U-Net. The improved model not only improves the recognition efficiency and the network structure, but also maintains the segmentation and recognition performance of the original model. Since the FFB-BiSeNetV2 model introduces a feature fusion branch in the up-sampling feature restoration stage, the recognition accuracy has been significantly improved. The semantic segmentation results are more abundant in contour details than the original model BiSeNetV2, which are closer to the labeled image. Based on the above experimental results and discussion, measuring the performance of each model in segmentation accuracy, operating efficiency, and model complexity, the FFB-BiSeNetV2 model with the best segmentation effect is selected for further experimental verification in the case of meeting the requirements of real-time recognition performance.

Figure 10. Segmentation results of weeds in rice fields based on low-altitude remote sensing.

3.3. Semantic Segmentation Model Validation

In order to verify the practicability of the rice weeds identification models constructed in this study and simulate the effect of UAV real-time identification and offline identification of rice weeds, a rice weed identification system based on UAV low-altitude remote sensing was developed. The development environment include Pycharm2020, PyQt5 and Python3.6.5. The software includes two parts: the front-end control interface and the back-end development. The front-end control interface uses Qt for visual display, and uses the component tools such as Tab, Button, QLabel, TextEdit, and Label for layout. The specific process of the back-end development is shown in Figure 11a. Two main functions of real-time UAV detection and offline detection by users as shown in Figure 11b, were implemented through a Python program. The real-time detection function calls the camera through OpenCV, and sets the button to take pictures instead of the UAV way-point photography operation. When the captured images are detected, the recognition program automatically outputs the corresponding recognition results, which are saved

by the system at the same time. The offline detection function is mainly designed for the situation when the embedded hardware computing resources are insufficient, and the user can only perform offline identification through the ground station server. The user can open the collected UAV remote sensing image library and select the picture that needs to be recognized, click the detection button to call the recognition program of the deep learning algorithm, and realize the recognition of the images one by one. At the same time, through the console output, the users also can view the running results of the system, including the running information of the program, the density of rice weeds, etc.

Figure 11. Development process and the running results in the offline detection function of rice weeds detection system. (a) Development process and the running results in the offline detection function of rice weeds detection system; (b) Running results in the offline detection function of rice weeds detection system.

Combined with the conclusion in Section 3.2.3, we used the offline recognition function of the rice weeds detection system based on the FFB-BiSeNetV2 model with the best segmentation effect to identify 84 UAV remote sensing images collected by the UAV in the test field 3. The prediction results of the FFB-BiSeNetV2 model for each test image in each category are showed in Figure 12. By counting the percentage of pixels in each category in the prediction results, the density proportions of rice, weeds, and other categories can be obtained, which are combined with the waypoints set by the UAV planning route to generate a weed density forecast distribution map. As shown in Figure 13a, it is the actual operation route planning of the UAV in the test field 3. The red point corresponds to the UAV's waypoint. When the UAV arrives at the waypoint, the UAV performs a photographing task and collects low-altitude remote sensing images of the rice field. As shown in Figure 13b, it is the generated heat map of the predicted distribution of rice weed density. The color scale on the right of Figure 13b indicates that the range of weed density is 0–100%, 0 indicates that there are no weeds in the waypoint area, which is shown green, and 100 indicates that the area is all weeds, which is shown red. Through the generated weed density distribution map, it is possible to accurately obtain the position information of the distribution of weeds in the rice field, which can also provide guidance for the subsequent precise spraying operation of plant protection UAV.

Figure 12. Proportion of pixels of different categories in the prediction results.

Figure 13. UAV actual operation route planning and distribution map of weed density prediction. (a) UAV actual operation route planning; (b) Distribution map of weed density prediction.

3.4. Embedded Transplantation and Real-Time Verification Analysis

In this experiment, the MobileNetV2-UNet model and the FFB-BiSeNetV2 model were transplanted to the embedded hardware Jetson AGX Xavier, and the TensorRT acceleration tool in Section 2.8 was used to improve the model's ability of inference. We used the test set data on the embedded hardware platform to compare the inference speed for a single picture of the two improved models before and after TensorRT optimization. The test results are shown in Table 3. After TensorRT optimization, MobileNetV2-UNet's single image inference speed under FP32 precision was 75.37% faster than before optimization, and the inference speed under FP16 precision reached 45.05 FPS, which was 3.14 times that before optimization. The optimized FFB-BiSeNetV2 had a 31.7% increase in inference speed under FP32 precision, and the inference speed under FP16 precision reached 40.16 FPS, which was 2.07 times that before optimization. From the analysis of the results, after TensorRT optimization, MobileNetV2-UNet and FFB-BiSeNetV2 both meet the basic performance requirements of embedded real-time recognition in inference speed, which can be well applied to the real-time and accurate identification of rice weeds at the airborne terminal, and further guide the precise target application task of plant protection UAV.

Table 3. Real-time comparison of the improved semantic segmentation model before and after optimization.

Methods	Before TRT Optimization/FPS	After TRT Optimization//FPS	
		FP32	FP16
MobileNetV2-UNet	14.33	25.13	45.05
FFB-BiSeNetV2	19.42	25.58	40.16

We also collected a piece of low-altitude remote sensing video data of the rice fields as the test data to verify the real-time recognition and segmentation effect of the models proposed in this study in a complex farmland environment. Firstly, the video data are input into the semantic segmentation models transplanted on the embedded hardware for recognition. At the same time, the density proportion information corresponding to the predicted image result of each frame is output, and the video data and segmented image results are saved every five frames. Three frames of video pictures and the corresponding predicted segmentation results are randomly extracted for comparing the segmentation effects of the two improved semantic segmentation models. As shown in Figure 14, compared to the MobileNetV2-UNet model, the FFB-BiSeNetV2 model has smoother segmentation boundaries for each category on the video data, and the contour of the segmentation results are closer to the corresponding position of the rice weeds distribution in the original image. The MobileNetV2-UNet model has many misrecognition areas. In the video frame 1 and video frame 3 shown in Figure 14, the MobileNetV2-UNet model recognizes the part of the rice area on the left side of the picture as a weed area. In the video frame 2, the MobileNetV2-UNet model and the FFB-BiSeNetV2 model have incorrect prediction, and the remote part of the rice area in the image was recognized as other categories. The reason for this result may be: due to the different shooting angles and shooting heights of the training data and video data of the model, each category in the rice fields presents similar characteristics under different light intensities and ground resolutions, which leads to the incorrect prediction results. Therefore, it is necessary to further increase and enhance the data set to improve the generalization ability of the rice weed recognition models, so that the models can better adapt to complex farmland real-time detection tasks.

Figure 14. Real-time segmentation result of the improved semantic segmentation model.

4. Discussion

Previous research works [18–20] has proved that the semantic segmentation method based on the full convolutional neural network can be used to generate accurate rice weed distribution maps to achieve full-pixel end-to-end prediction and recognition of rice weed images. However, due to the low inference efficiency of the CNN network, it is limited to offline analysis of rice weeds remote sensing images. Based on the previous research works, in this study, we improved on the semantic segmentation model U-Net and BiSeNetV2, and proposed two rice weeds recognition models MobileNetV2-UNet and FFB- BiSeNetV2 suitable for UAV low-altitude remote sensing. We transplanted the improved models to the embedded hardware platform, and analyzed the real-time performance of the model, aiming to study whether the improved model proposed in this paper is feasible for real-time recognition of drones on the airborne end, which can provide theoretical guidance for the future airborne terminal real-time weeds identification test.

In the experiment comparing segmentation accuracy of competing models, MobileNetV2-UNet and FFB-BiSeNetV2 have both improved the segmentation accuracy PA and MIoU compared to their original model U-Net and BiSeNetV2. The results show that the FFB-BiSeNetV2 achieved the best segmentation accuracy among the four models in this paper. By introducing the feature fusion branch and the depthwise separable convolution to improve the lightweight model BiSeNetV2, the feature extraction and recognition capabilities of the model are improved to a certain extent, and it is more suitable for the identification and segmentation task of rice weeds than the original model BiSeNetV2. In the experiment of comparison of model parameters and operating efficiency, the MobileNetV2-UNet has significant improvements in parameters, model size, and floating point calculations, and the inference speed has also been significantly improved. It indicates that the introduction of lightweight networks MobileNetV2 improves the original complex feature extraction network of the U-Net model, significantly improves the structure of the model and reduces the demand for computing resources of the model, and improves the model real-time recognition performance. On the embedded hardware platform, the inference speeds for a single image of the two improved models proposed in this paper at FP16 weight accuracy are both greater than the 25 frames/s required for real-time inference speed [36], indicating MobileNetV2-UNet and FFB-BiSeNetV2 also has good real-time recognition performance on embedded devices with limited computing resources. In summary, the two improved models proposed in this paper can be applied to task of the real-time identification of rice weeds on the airborne end of UAVs.

In the follow-up research work, the real-time detection system of rice weeds and the spraying system of the plant protection UAV can be combined to realize the integration of real-time identification and precise spraying operations on the airborne end. By adjusting different flight operation parameters, the reliability of real-time semantic segmentation method in identifying rice weeds in low-altitude remote sensing and guiding plant protection drones to accurately apply pesticides will be further verified.

5. Conclusions

In this study, we proposed two lightweight models suitable for low-altitude remote sensing rice weed recognition combining with deep learning semantic segmentation technology and embedded hardware to realize the real-time recognition of rice weeds on the airborne end. Firstly, the MobileNetV2-UNet model reduced by 89.12%, 86.16%, and 92.6% the parameter amount, model size, and floating points of operations compared with those of the original model U-Net, and there was a significant improvement in operating efficiency and model structure. In addition, the FFB-BiSeNetV2 model maintained real-time recognition performance of the original model BiSeNetV2 while improving the segmentation accuracy, and the PA and the MIoU ratio reached the highest accuracy among the comparison methods in this study, which were 93.09% and 80.28%, respectively.

Then, we developed a set of rice weeds detection system based on UAV low-altitude remote sensing based on the FFB-BiSeNetV2 model to identify rice weeds of the 84 low-

altitude remote sensing images collected in experimental plot 3. Further, a heat map of the predicted distribution of rice weeds in the corresponding waypoint area was generated base on the identification results, which can provide guidance for the follow-up precise spraying operations of plant protection UAVs.

Finally, to verify the real-time recognition performance of the model, the improved semantic segmentation models were transplanted to the embedded device Jetson AGX Xavier, optimized and accelerated through TensorRT to further improved the real-time inference efficiency. The test results showed that after TensorRT optimization, the MobileNetV2-UNet model and the FFB-BiSeNetV2 model had inference speeds of 45.05 FPS and 40.16 FPS under FP16 weight precision, which can meet the performance requirements of real-time detection. Through the verification of the real-time segmentation effect of the two improved models on the video data collected in the actual rice field, the FFB-BiSeNetV2 model is better than the MobileNetV2-UNet model in the real-time segmentation of rice weeds.

The improved semantic segmentation models proposed in this study balances the three indicators of segmentation accuracy, operation efficiency, and model structure complexity, which have good segmentation effects and real-time detection performance. The models can realize the real-time accurate recognition of rice weeds distribution area under UAV low-altitude remote sensing, which provides a strong theoretical basis and technical guidance for subsequent plant protection UAV to achieve real-time and precise target spraying operation.

Author Contributions: Conceptualization, Y.L. and K.H.; Methodology, J.D. and K.H.; Software, K.H.; Data curation, K.H. and C.Y.; writing—original draft preparation, K.H.; writing—review and editing, J.D., Y.Z. and Y.L.; investigation, K.H., C.Y., L.L., J.Y., J.Z. and W.Z.; visualization, K.H. supervision, J.D. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Guangdong Provincial Innovation Team for General Key Technologies in Modern Agricultural Industry, Grant No. 2020KJ133, Guangdong Science and Technology Plan Project, grant number 2018A050506073, National Key Research and Development Program, grant number 2018YFD0200304, and the 111 Project, grant number D18019.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful to the plant protection experts Maoxin Zhang and Bing Ling from South China Agricultural University for guiding us to conduct field investigation of the study site and for providing professional instruction in semantic annotation of UAV image data.

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

References

1. Lan, Y.B.; Chen, S.D.; Fritz, B.K. Current status and future trends of precision agricultural aviation technologies. *Int. J. Agric. Biol. Eng.* **2017**, *10*, 1–17.
2. Stroppiana, D.; Villa, P.; Sona, G.; Ronchetti, G.; Candiani, G.; Pepe, M.; Busetto, L.; Migliazzi, M.; Boschetti, M. Early season weed mapping in rice crops using multi-spectral UAV data. *Int. J. Remote Sens.* **2018**, *39*, 5432–5452. [[CrossRef](#)]
3. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A Review on UAV-Based Applications for Precision Agriculture. *Information* **2019**, *10*, 349. [[CrossRef](#)]
4. Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M.S. Deep learning for visual understanding: A review. *Neurocomputing* **2016**, *187*, 27–48. [[CrossRef](#)]
5. Xiao, Y.; Tian, Z.; Yu, J.; Zhang, Y.; Liu, S.; Du, S.; Lan, X. A review of object detection based on deep learning. *Multimed. Tools Appl.* **2020**, *79*, 23729–23791. [[CrossRef](#)]
6. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A survey on deep learning techniques for image and video semantic segmentation. *Appl. Soft Comput.* **2018**, *70*, 41–65. [[CrossRef](#)]
7. Mao, H.P.; Hu, B.; Zhang, Y.C.; Qian, D.; Chen, S.R. Optimization of color index and threshold segmentation in weed recognition. *Trans. Chin. Soc. Aric. Eng.* **2007**, *23*, 154–158. (In Chinese)
8. Kazmi, W.; Garcia-Ruiz, F.; Nielsen, J.; Rasmussen, J.; Andersen, H.J. Exploiting affine invariant regions and leaf edge shapes for weed detection. *Comput. Electron. Agric.* **2015**, *118*, 290–299. [[CrossRef](#)]

9. Pflanz, M.; Nordmeyer, H.; Schirrmann, M. Weed Mapping with UAS Imagery and a Bag of Visual Words Based Image Classifier. *Remote Sens.* **2018**, *10*, 1530. [CrossRef]
10. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
11. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
12. Wang, S.S.; Wang, S.; Zhang, H.; Wen, C.J. Soybean field weed recognition based on light sum-product networks and UAV remote sensing images. *Trans. Chin. Soc. Agric. Eng.* **2019**, *35*, 81–89. (In Chinese)
13. Sun, J.; Tan, W.J.; Wu, X.H.; Shen, J.F.; Lu, B.; Dai, C.X. Real-time recognition of sugar beet and weeds in complex backgrounds using multi-channel depth-wise separable convolution model. *Trans. Chin. Soc. Agric. Eng.* **2019**, *35*, 184–190. (In Chinese)
14. Yang, M.-D.; Tseng, H.-H.; Hsu, Y.-C.; Yang, C.-Y.; Lai, M.-H.; Wu, D.-H. A UAV Open Dataset of Rice Paddies for Deep Learning Practice. *Remote Sens.* **2021**, *13*, 1358. [CrossRef]
15. Yang, M.-D.; Boubin, J.G.; Tsai, H.P.; Tseng, H.-H.; Hsu, Y.-C.; Stewart, C.C. Adaptive autonomous UAV scouting for rice lodging assessment using edge computing with deep learning EDANet. *Comput. Electron. Agric.* **2020**, *179*, 105817. [CrossRef]
16. Qiao, M.; He, X.; Cheng, X.; Li, P.; Luo, H.; Zhang, L.; Tian, Z. Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102436. [CrossRef]
17. Torres-Sánchez, J.; Lopez-Granados, F.; De Castro, A.I.; Peña-Barragan, J.M. Configuration and Specifications of an Unmanned Aerial Vehicle (UAV) for Early Site Specific Weed Management. *PLoS ONE* **2013**, *8*, e58210. [CrossRef]
18. Huang, H.; Deng, J.; Lan, Y.; Yang, A.; Deng, X.; Zhang, L. A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery. *PLoS ONE* **2018**, *13*, e0196302. [CrossRef]
19. Huang, H.; Deng, J.; Lan, Y.; Yang, A.; Deng, X.; Wen, S.; Zhang, H.; Zhang, Y. Accurate Weed Mapping and Prescription Map Generation Based on Fully Convolutional Networks Using UAV Imagery. *Sensors* **2018**, *18*, 3299. [CrossRef]
20. Huang, H.; Lan, Y.; Deng, J.; Yang, A.; Deng, X.; Zhang, L.; Wen, S. A Semantic Labeling Approach for Accurate Weed Mapping of High Resolution UAV Imagery. *Sensors* **2018**, *18*, 2113. [CrossRef]
21. Zhu, S.; Deng, J.Z.; Zhang, Y.L.; Yang, C.; Yan, Z.W.; Xie, Y.Q. Study on distribution map of weeds in rice field based on UAV remote sensing. *J. South China Agric. Univ.* **2020**, *41*, 67–74. (In Chinese)
22. Chen, J.S.; Ran, X.K. Deep Learning With Edge Computing: A Review. *Proc. IEEE* **2019**, *107*, 1655–1674. [CrossRef]
23. Boubin, J.G.; Chumley, J.; Stewart, C.; Khanal, S. Autonomic computing challenges in fully autonomous precision agriculture. In Proceedings of the 2019 IEEE International Conference on Autonomic Computing (ICAC), Umea, Sweden, 16–20 June 2019; pp. 11–17. [CrossRef]
24. AlKameli, A.; Hammad, M. Automatic Learning in Agriculture: A Survey. *Int. J. Comput. Digit. Syst.* **2021**, in press. Available online: <http://journal.uob.edu.bh/handle/123456789/4427> (accessed on 1 June 2021).
25. Boubin, J.G.; Babu, N.T.R.; Stewart, C.; Chumley, J.; Zhang, S.Q. Managing edge resources for fully autonomous aerial systems. In Proceedings of the 4th ACM/IEEE Symposium on Edge Computing, Arlington, VA, USA, 7–9 November 2019; pp. 74–87. [CrossRef]
26. Guillén, M.A.; Llanes, A.; Imbernón, B.; Martínez-España, R.; Bueno-Crespo, A.; Cano, J.-C.; Cecilia, J.M. Performance evaluation of edge-computing platforms for the prediction of low temperatures in agriculture using deep learning. *J. Supercomput.* **2021**, *77*, 818–840. [CrossRef]
27. Hu, J.; Bruno, A.; Ritchken, B.; Jackson, B.; Espinoza, M.; Shah, A.; Delimitrou, C. HiveMind: A Scalable and Serverless Coordination Control Platform for UAV Swarms. *arXiv* **2020**, arXiv:2002.01419.
28. Hadidi, R.; Asgari, B.; Jijina, S.; Amyette, A.; Shoghi, N.; Kim, H. Quantifying the design-space tradeoffs in autonomous drones. In Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Online, 19–23 April 2021; pp. 661–673. [CrossRef]
29. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *Computer Vision and Pattern Recognition*. Available online: <https://arxiv.org/abs/1602.07360> (accessed on 1 June 2021).
30. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weiyand, T.; Andreetto, M.; Hartwing, A. MobileNets: Efficient convolutional neural networks for mobile vision applications. *Computer Vision and Pattern Recognition*. Available online: <https://arxiv.org/abs/1704.04861> (accessed on 1 June 2021).
31. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
32. Shin, J.; Chang, Y.K.; Heung, B.; Nguyen-Quang, T.; Price, G.W.; Al-Mallahi, A. A deep learning approach for RGB image-based powdery mildew disease detection on strawberry leaves. *Comput. Electron. Agric.* **2021**, *183*, 106042. [CrossRef]
33. Rançon, F.; Bombrun, L.; Keresztes, B.; Germain, C. Comparison of SIFT Encoded and Deep Learning Features for the Classification and Detection of Esca Disease in Bordeaux Vineyards. *Remote Sens.* **2019**, *11*, 1. [CrossRef]
34. Tang, Z.; Yang, J.; Li, Z.; Qi, F. Grape disease image classification based on lightweight convolution neural networks and channelwise attention. *Comput. Electron. Agric.* **2020**, *178*, 105735. [CrossRef]

35. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015 18th International Conference, Munich, Germany, 5–9 October 2015; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9351. Available online: https://doi.org/10.1007/978-3-319-24574-4_28 (accessed on 1 June 2021).
36. Yu, C.; Gao, C.; Wang, J.; Yu, G.; Shen, C.; Sang, N. BiSeNet V2: Bilateral Network with Guided Aggregation for Real-time Semantic Segmentation. *arXiv* **2020**, arXiv:2004.02147.
37. NVIDIA. TensorRT. Available online: <https://developer.nvidia.com/tensorrt> (accessed on 1 June 2021).
38. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [[CrossRef](#)]
39. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
40. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNet V2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
41. Jang, J.; Cho, H.; Kim, J.; Lee, J.; Yang, S. Deep neural networks with a set of node-wise varying activation functions. *Neural Networks* **2020**, *126*, 118–131. [[CrossRef](#)] [[PubMed](#)]
42. Yao, C.; Liu, W.; Tang, W.; Guo, J.; Hu, S.; Lu, Y.; Jiang, W. Evaluating and analyzing the energy efficiency of CNN inference on high-performance GPU. *Concurr. Comput. Pract. Exp.* **2021**, *33*, e6064. [[CrossRef](#)]
43. Ma, N.; Zhang, X.; Zheng, H.-T.; Sun, J. Shufflenet v2: Practical guidelines for efficient CNN architecture design. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.