Article

Segmentation of Wheat Lodging Areas from UAV Imagery Using an Ultra-Lightweight Network

Guoqing Feng [1,2,3], Cheng Wang [1,2,3], Aichen Wang [1,*], Yuanyuan Gao [1], Yanan Zhou [2,3], Shuo Huang [2,3] and Bin Luo [1,2,3,*]

1 School of Agricultural Engineering, Jiangsu University, Zhenjiang 212013, China; grigorii_f@yeah.net (G.F.); wangc@nercita.org.cn (C.W.); gaoyy0910@ujs.edu.cn (Y.G.)
2 Research Center of Intelligent Equipment, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100089, China; zhouyn@nercita.org.cn (Y.Z.); huangs@nercita.org.cn (S.H.)
3 National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China
* Correspondence: acwang@ujs.edu.cn (A.W.); luob@nercita.org.cn (B.L.)

Abstract: Crop lodging is an important cause of direct economic losses and secondary disease transmission in agricultural production. Most existing methods for segmenting wheat lodging areas use a large-volume network, which poses great difficulties for annotation and crop monitoring in real time. Therefore, an ultra-lightweight model, Lodging-U2NetP (L-U2NetP), based on a novel annotation strategy which crops the images before annotating them (Crop-annotation), was proposed and applied to RGB images of wheat captured with an unmanned aerial vehicle (UAV) at a height of 30 m during the maturity stage. In the L-U2NetP, the Dual Cross-Attention (DCA) module was firstly introduced into each small U-structure effectively to address semantic gaps. Then, Crisscross Attention (CCA) was used to replace several bulky modules for a stronger feature extraction ability. Finally, the model was compared with several classic networks. The results showed that the L-U2NetP yielded an accuracy, F1 score, and IoU (Intersection over Union) for segmenting of 95.45%, 93.11%, 89.15% and 89.72%, 79.95%, 70.24% on the simple and difficult sub-sets of the dataset (CA set) obtained using the Crop-annotation strategy, respectively. Additionally, the L-U2NetP also demonstrated strong robustness in the real-time detection simulations and the dataset (AC set) obtained using the mainstream annotation strategy, which annotates images before cropping (Annotation-crop). The results indicated that L-U2NetP could effectively extract wheat lodging and the Crop-annotation strategy provided a reliable performance which is comparable with that of the mainstream one.

Keywords: UAV; wheat lodging; lightweight; deep learning; improved U2NetP

1. Introduction

Wheat is one of the most widely cultivated crops in the world [1], and meets the main food demand of over 3 billion people [2], providing protein and energy to nearly 20% of the world's population. The stable production of wheat is of great importance to food security. The lodging caused by various natural disasters and improper planting poses a serious threat to food production. Lodging in the middle and late stages of wheat growth even seriously affects the transportation of water [3] and nutrients, resulting in reduced wheat yield [4] and quality [5,6], and even more indirectly affects the genetic advancement of wheat [7]. However, quickly and accurately locating the location of wheat lodging after disasters, extracting information of the lodging area, can provide data and technical support for relevant departments and insurance claims [8]. Therefore, the efficient detection of wheat lodging is of great significance for stabilizing global food security.

Remote sensing technology provides a promising tool to grasp timely, synoptic, and repetitive information about the status of agricultural crops [9]. In numerous previous studies, various types of remote sensing images have provided high-quality lodging information such as RGB data [10], thermal infrared data [11], and spectral data [12]. Traditional

remote sensing platforms for crop lodging detection mainly include near-ground, airborne, and satellite remote sensing [13,14]. Hufkens et al. [15] used an inexpensive smartphone to take a stream of images of individual small farmlands to quantify the important phenological stages of crops and predicted lodging events. Based on manned aircraft, Gerten and Wiese [16] analyzed the lodging-induced winter wheat yield reduction. Yang et al. [9] proposed a polarization index method for crop lodging detection by utilizing the sensitivity of the polarization characteristics of synthetic aperture radar (SAR) data to wheat lodging. These studies have promoted the application of remote sensing for monitoring crop lodging. However, the limited detection range of near-surface equipment, expensive data acquisition costs for manned aircraft, and low temporal resolution of satellite images make it difficult to promote these methods for real-time lodging detection [17].

In recent years, the development and application of Unmanned Aerial Vehicles (UAVs) in smart agriculture have garnered extensive attention. UAVs have proven to be particularly effective in weed mapping and management [18], vegetation monitoring [19], disease detection [20], and various other agricultural technology applications. Compared with traditional remote sensing platforms, UAVs offer a more conservative data source for research on the innovation and application of various methods in precision agriculture, being more flexible, controllable, and cost effective. Vélez et al. [21] utilized UAV images to obtain a novel method for measuring the planar area and ground shadows of pistachio trees, and the obtained rapid and low-cost canopy characterization can be used for accurate information in management decision-making. Based on the UAV imagery, Matese and Di Gennaro [22] obtained a novel index based on the geometric structure of the grape canopy, which goes beyond traditional NDVI indicators, and based on this, obtained a flexible global yield assessment model. Li et al. [23] obtained a clear corn canopy structure based on drone oblique images, breaking through the complex and expensive traditional mapping methods. Additionally, research on lodging detection has also made significant progress with the aid of UAV images. Tian et al. [24] obtained multispectral and RGB images from UAV platforms, extracted various features of rice crop, and established a rice lodging detection model. Chu et al. [25] accurately extracted the corn lodging from UAV RGB images based on optimal texture features. With the development of deep learning, more deep-learning-based methods have been explored and applied for crop lodging detection. At present, the data for crop lodging detection based on deep learning have gradually developed towards multiple heights [26], growth stages [27] and image sizes [28]. However, this trend means more data should be processed and analyzed. The models for lodging detection are generally segmentation models, which require burdensome annotations and often have a huge model volume with massive parameters. Therefore, the annotation requirements and model size should be reduced to alleviate the research burden, making it feasible for models to be deployed on edge devices for real-time lodging extraction.

Semantic annotations are often directly applied to complete images to obtain a label of the lodging area, non-lodging area, and background area in mainstream research on lodging area segmentation [29]. The complete images are often images containing a complete plot of a wheat field that are concatenated from multiple images captured by a UAV camera. They often have a high resolution and contain rich information, which greatly reduce the loading and processing speed of annotation software and cause visual fatigue and errors for annotators. Shen et al. [30] also pointed out that, especially for high-resolution remote sensing images, the memory limitations of processing equipment necessitate downsampling or cropping after image annotation. However, downsampling may result in image distortion, and cropping can lead to misjudgment in shallow networks due to a lack of global information. Therefore, it is necessary to explore a new annotation strategy to alleviate these two issues. UNet and its various variants have been widely used in segmentation tasks in various complex scenes. Its ultra-lightweight variant U2NetP is also being applied in multiple fields such as electricity meter detection [31] and ship draft detection [32]. The network has achieved good results for the abovementioned tasks, but

there is still room for improvement in crop lodging segmentation where more complex features and unclear boundaries should be handled.

Therefore, the main objectives of this study were to (1) propose an ultra-lightweight segmentation method for a wheat lodging area that combines an annotation strategy and U2NetP based on UAV imagery, and to compare the performance with several classic segmentation models; (2) simulate real-time detection tasks on the proposed model and verify its robustness, and (3) verify the annotation quality of the Crop-annotation strategy in datasets obtained using the Annotation-crop strategy.

2. Materials and Methods

2.1. Study Site

Runguo Farm (32°11′ N, 119°40′ E, WGS84) in Zhenjiang New Area, Zhenjiang City, Jiangsu Province, China, was the study area (Figure 1). The area has a northern subtropical monsoon climate, with warm winters and hot summers, an annual precipitation of 700–2000 mm, and an average annual temperature of 13–21 °C. Conditions are eminently suitable for wheat cultivation and production. The soil in this area is mainly yellow-brown soil and paddy soil, and the planting wheat varieties are mainly the "Zhenmai" series. The 14,000 acres of wheat in this region are crucial for the basic food supply and economic source of farmers in the area.
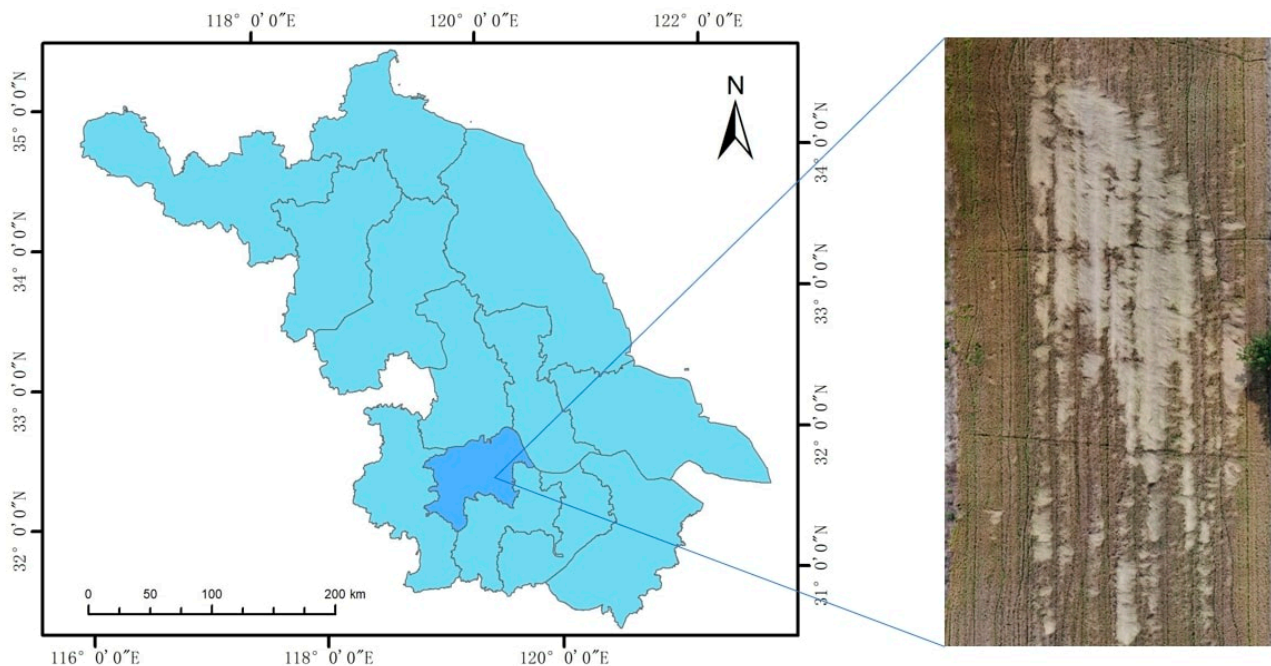

Figure 1. The left side of image is geographical location of Runguo Farm, Zhenjiang New Area, Zhenjiang City, Jiangsu Province, China, based on WGS84, and the right side of image is the field area of this study.

2.2. Data Collection and Preprocessing

This study collected UAV images of the wheat lodging caused by dense planting and strong winds (Beaufort Wind Scale 4–5 in the daytime) at the maturity stage (11:00–14:00 on 8 May 2022). The image is captured using DJI MAVIC AIR (DJI Innovation Technology, Shenzhen, China) with flight altitude of 30 m, which is appropriate for capturing clear images and has been widely used in prior similar studies [24,28,29] and speed of 1.5 m/s, respectively. The spatial resolution of the altitude is 243.2 mm^2/pixel. Forward and side overlapping of the images were set at 85% and 75%, respectively. The shooting mode is set to take photos at equal time intervals. All the adjustable parameters mentioned above have been adjusted on site to minimize the impact of lighting and height on the obtained

images, providing clearer and more usable images captured by drones for subsequent models. During the noon of image capture, the weather was sunny and windless, which met the requirements for UAV flight and image capture.

For image preprocessing, Agisoft Metashape 1.4.1 (Russia) software was used to concatenate UAV images and add control points in the middle to complete geometric correction. WGS84 (ESPG::4326) and extrapolation were chosen as geographic and interpolation type, respectively, when we built the digital elevation model (DEM). Also, we chose mosaic as the blending mode to enable hole filling in image preprocessing. After cropping out the experimental area, RGB images of 18 plots covering approximately 1.0 ha were obtained. As shown on the right side of Figure 1, the image includes interference from trees, field boundaries, and agricultural machinery operation traces.

2.3. Dataset Construction and Annotation

For dataset construction, two strategies were employed: the Crop-annotation strategy and the Annotation-crop strategy. The procedure of the Crop-annotation strategy is illustrated in Figure 2a. Firstly, the complete image of the wheat field obtained through concatenation was cropped by a sliding window to the size of 512×512 pixels, which is moderate and has been successfully applied in many models in previous studies. Then, we annotated lodging in LabelMe 5.2.1 [33] for preparation of network training. The red color represents the lodging area, and the black color represents the non-lodging area. Finally, after deleting some samples that were fully lodging or non-lodging, and then data augmentation, the Crop-annotation strategy resulted in a total of 10,460 images, constituting the CA set. The data augmentation consists of random horizontal flipping, vertical flipping, and viewing angles.
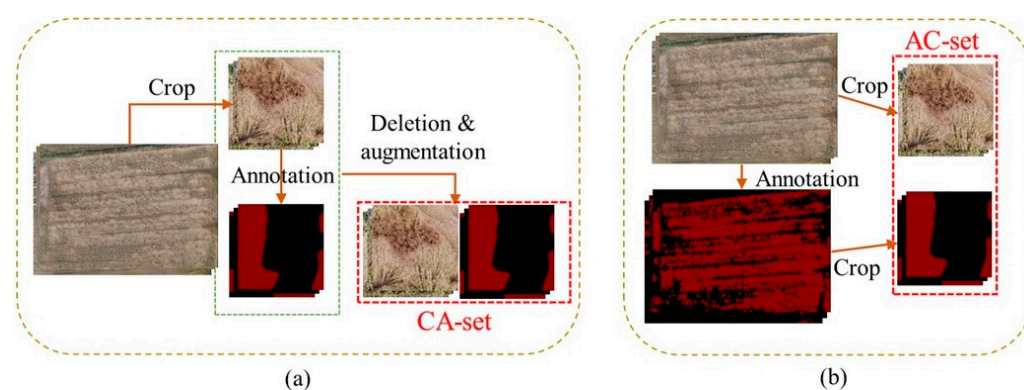

Figure 2. The (**a**,**b**) sides represent the procedure of producing CA set by Crop-annotation strategy and AC set by Annotation-drop, respectively.

The procedure of the Annotation-crop strategy is shown in Figure 2b. In contrast to the Crop-annotation strategy, the Annotation-crop strategy firstly annotated the mosaic images, then cropped the RGB and annotation images into small ones with a size of 512×512 pixels. The Crop-annotation strategy yielded a total of 3000 images, constituting the AC set, which was only used for the test.

Table 1 describes the different datasets used in this research. In establishing the training set, this study was inspired by [34] to improve model performance by adjusting the proportion of positive and negative samples. In this article, we define images with more than half and less than half of the lodging information as positive and negative samples, respectively. To ensure an adequate representation of valuable lodging and interference information in the training set, we manually adjusted the ratio of images with more than half of the lodging area to those with less than half of the lodging area to approximately 5:3 before applying data enhancement. When preparing the test sets, this article proposed a set of main design principles for simple and difficult datasets and divided multiple test sets into simple and difficult sets accordingly. As shown in Figure 3A–D, the difficult set

mainly consisted of several types of samples: (A) the area of lodging or non-lodging areas in the image is small or scattered at the boundaries, (B) the degree of lodging area varies, (C) each lodging area is compressed to form suspected strip-shaped lodging areas, (D) the boundaries of the lodging areas are unclear due to factors such as similar colors to the surrounding areas. Simple set shown in figures E–H in Figure 3 mainly consist of several types of samples: (E,F) the image is clear and either lodging or non-lodging, (G,H) the boundary between lodging and non-lodging is clear and the texture is distinct. On one hand, the division verified the basic performance of the model in simple samples. On the other hand, it aimed to ensure that the model can still guarantee efficiency in difficult samples. Moreover, the principles can avoid having too many simple samples in the test set, which could lead to an artificially high performance and affect researchers' judgment of model improvement.

Table 1. Dataset description.

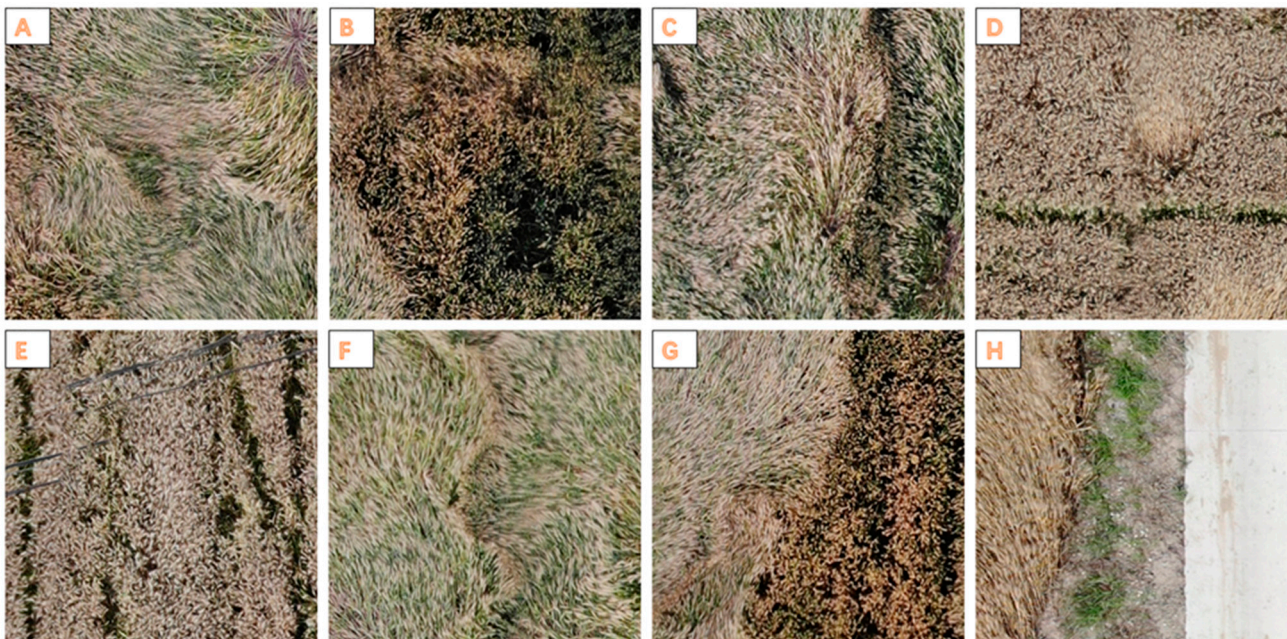| Dataset	Total	Train Set	Test Set	
CA set	10,460	8000	Normal set1 880 (Difficult) 760 (Simple)	Disturb set 660 (Difficult) 570 (Simple)
AC set	3000	/	Normal set2 3000	Normal set3 500 (Difficult) 500 (Simple)

Figure 3. Corresponding images of the main design principles of the difficult and simple set. (**A**–**D**) are examples of the four design principles corresponding to the difficult set, and (**E**,**F**) and (**G**,**H**) are examples of the two design principles for the simple set, respectively.

The following is the detailed description of every test set. The original test set in the CA set consisted of 220 difficult samples and 190 simple samples. Normal set1 was obtained by performing the same data augmentation operation as the training set on the original test set. The Disturb set was obtained by performing three data augmentation methods on the original test set: light changes, construction of airborne foreign objects, and motion blur.

In order to evaluate the annotation quality of Crop-annotation strategy and the ability of the ultra-lightweight network, L-U2NetP, to extract fine-grained feature information, this paper established a new test set using the Annotation-crop strategy. By extracting

annotation points from Json files generated by annotation results, the number of annotation points obtained using the Annotation-crop strategy in the same original complete image was approximately 1.5~2.5 times of that obtained using the Crop-annotation strategy. However, this fine-grained annotation greatly increases the training burden of the network, and all networks mentioned in the article could not obtain usable models under the same computing resources. Therefore, this dataset was only for testing purposes. Normal set2 contained all data of the AC set, while Normal set3 consisted of an equal number of difficult and simple samples selected from it.

3. Methodology

3.1. Technical Workflow

Figure 4 shows the flow chart of the research method for the extraction of the wheat lodging area. Firstly, UAV RGB images of the wheat lodging areas at the maturity stage were collected. Secondly, the CA set and AC set were produced. Then, the L-U2NetP network and other models were trained by the training set in the CA set. Finally, the classic segmentation network SegNet [35] and UNet [36], U2Net, and U2NetP [37], which are typical representatives of different depths and cascade modes in the UNet family of networks, were selected as comparative models.
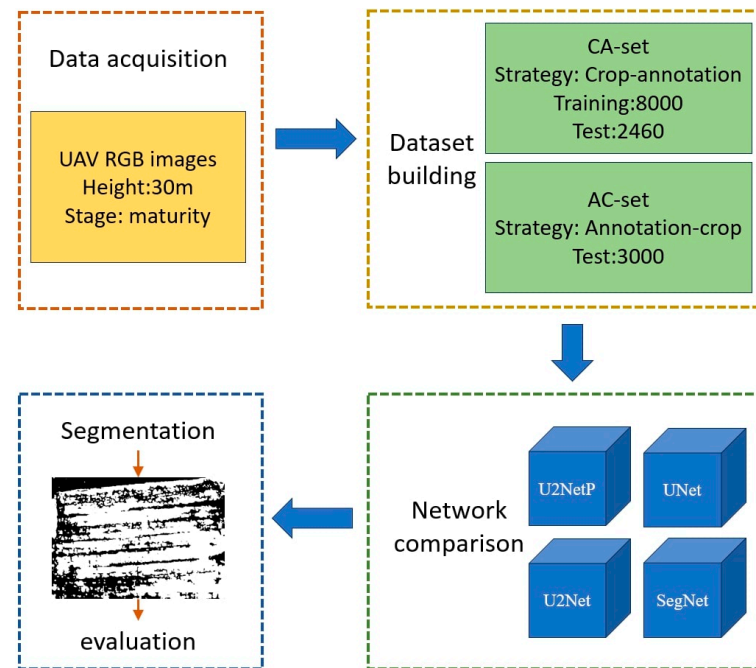

Figure 4. Research flowchart, including data acquisition, dataset building, network comparison, and result evaluation.

3.2. DCA Module

Dual Cross-Attention (DCA) is a simple and effective attention module that can enhance skip connections in U-structures for segmentation [38]. The DCA module consists of three main sections, as shown in Figure 5. Firstly, it consists of patch embedding modules of different scales. Then, DCA is implemented on the token obtained in the previous step combining the Channel Cross Attention and Spatial Cross Attention modules with the residual connection to capture the long-distance dependencies. Finally, these tokens are serialized and upsampled using normalization layers and GeLU, connecting them to the decoder counterparts. The DCA module can effectively extract the channel and spatial dependencies between multi-scale encoder features to make up the semantic gap.
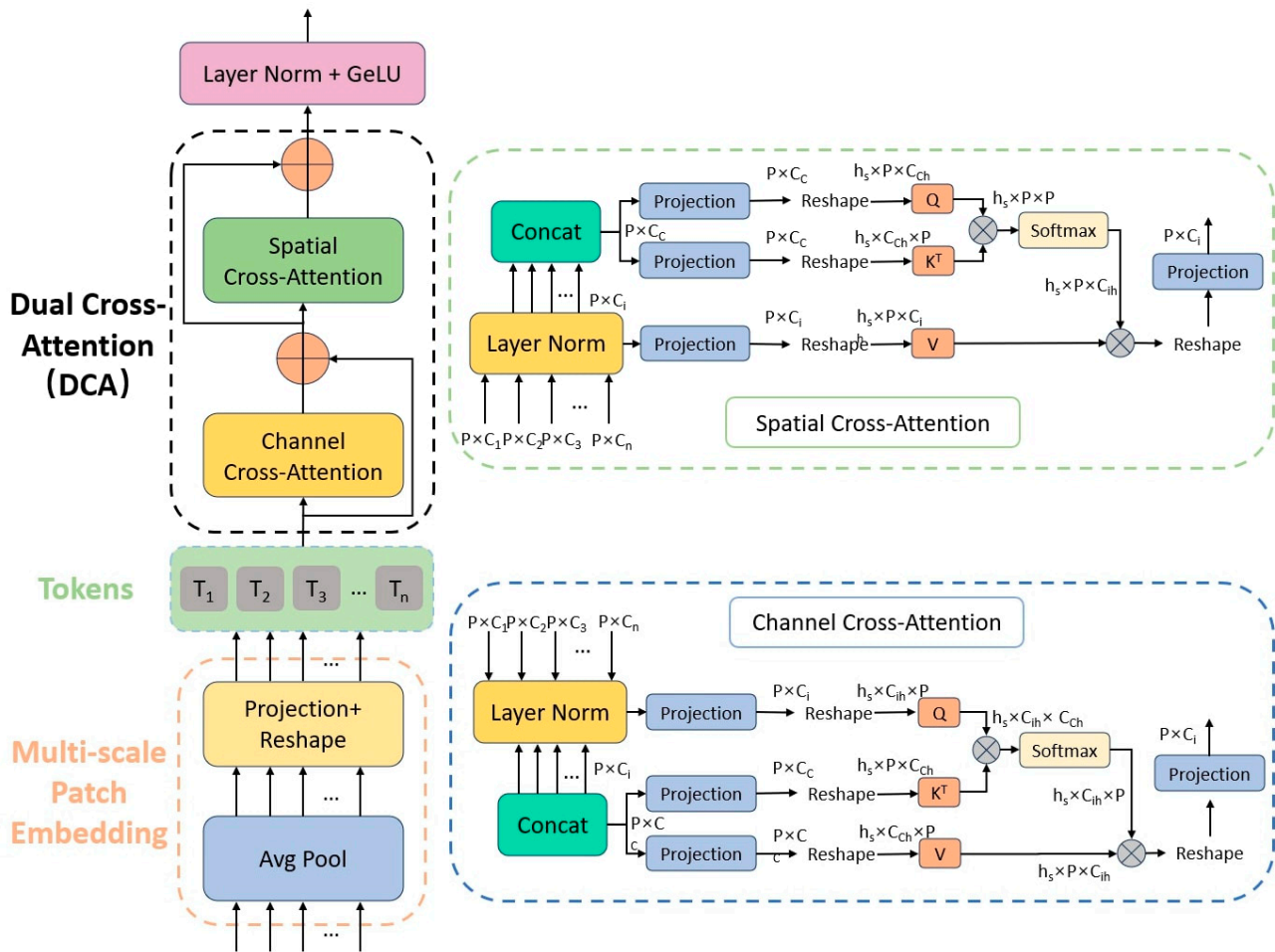
Figure 5. The specific implementation steps and structural details of the DCA module.

3.3. Structure of L-U2NetP

U2Net was proposed for salient object detection (SOD). U2NetP is an improved version of U2 Net, which further improves the segmentation performance and reduces inference costs by reducing the parameters and computational complexity of the latter without changing its network structure. As an improved version, the parameter counts and model size of U2NetP are about 1/40 and 1/67 of U2Net, respectively. The smaller model size reduces the requirement for storage and computing resources, allowing for faster loading and adaptation to different platforms and device. Hence, this paper used U2NetP as the main architecture. The following improvements are made, as shown in Figure 6:

1. CCA [39] was used to replace the large dilation rate convolution blocks and bottom-most blocks of U2NetP. At the cost of ultra-lightweight computation and memory, the CCA module modeled the correlations between the neighborhood features of the full image. Compared with the original image, the CCA-enhanced feature map could selectively aggregate context features through the attention feature map while obtaining a larger context receptive field.

2. As shown in the dark gray module in Figure 6, the DCA module was added to the connection channels of each small-scale U-structure's encoder and decoder, which could make it simple and effective to enhance the skip connection in the structure.

3. All activation functions, ReLU, in U2Net were replaced by LeakyReLU [40] in L-U2NetP. The ReLU is often used as an activation function in order to obtain a sparse network, but with an ReLU, the dead ReLU problem was observed when the input data were normalized [41]. The Leaky ReLU function can adjust the dead ReLU problem of

negative values by assigning very small linear components of the input vector to negative inputs.
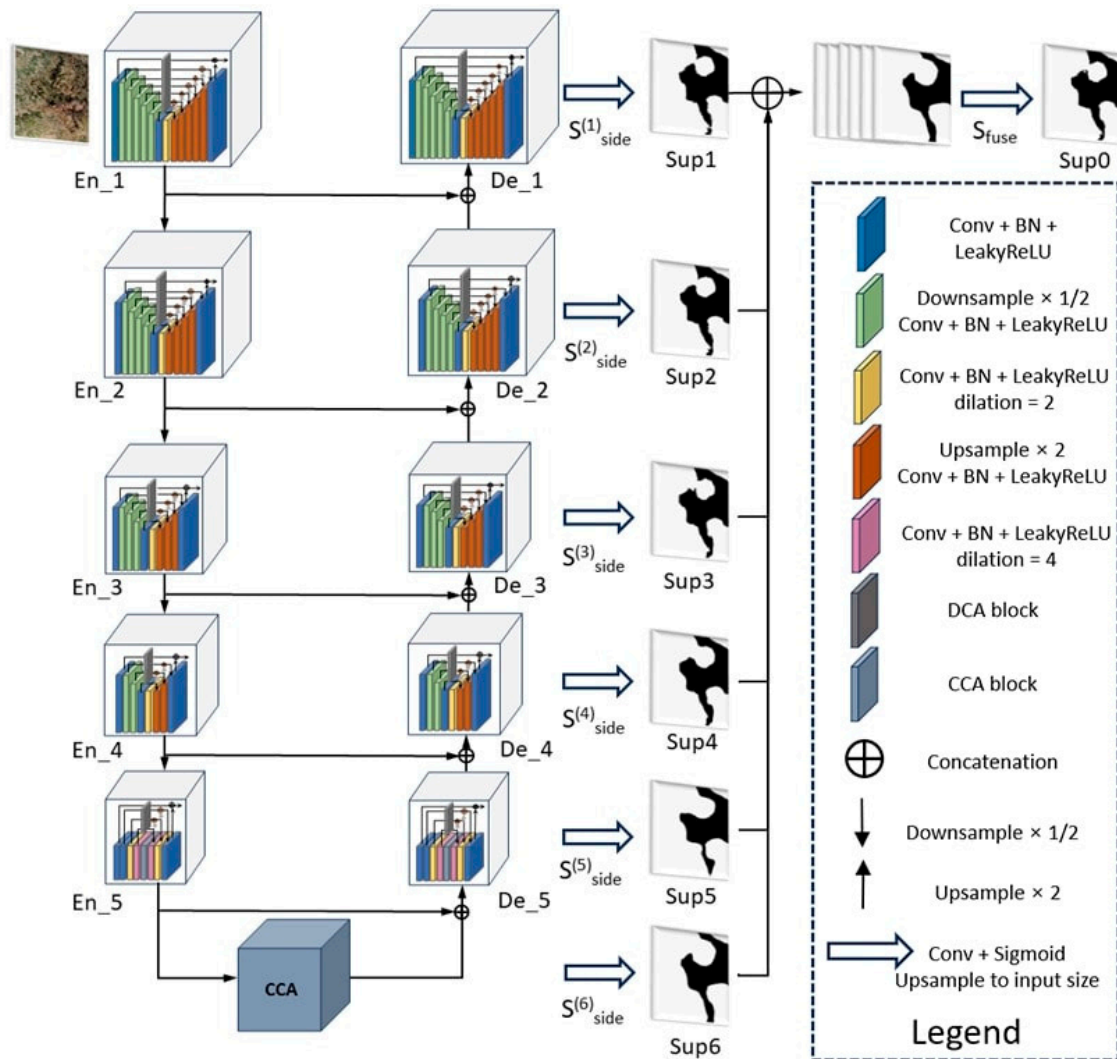

Figure 6. L-U2NetP network structure and output processing diagram.

3.4. Model Training

The model training was based on the Windows Server 2022 operating system. The Python 3.10 programming language was adopted and PyTorch was used as the back end. It runs under the Intel Xeon Gold 6130T processor, 128 GB video memory, NVIDIA GeForce RTX4070 video card, and GPU (CUDA 12.1) environment. The models were trained to obtain parameters for segmentation after initializing their parameters. Their training parameters in the paper are shown in Table 2.

Table 2. Network model training parameter configuration.

Parameter	Value
Batch size	8
Learning rate	0.001
Weight decay	0.0001
Input shape	512×512
Num classes	2
Epoch	80

The Loss function used in training consists of two parts: Multi BCE Loss [37] and Focal Loss [42]. Multi BCE Loss is often used in U2Net series networks which use Sigmoid to classify pixels. The focal loss can dynamically adjust the cross-entropy loss based on confidence, making the model training loss more focused on difficult cases, thereby alleviating the problem of imbalanced positive and negative samples and differences. This paper set the corresponding weight of each layer loss to 1 by observing the output of the feature maps. Therefore, the formula of the Multi Focal Loss L_{MF} is as follows:

$$L_{MF} = \sum_{n=1}^{N} \xi_n + \xi_{fuse} \tag{1}$$

$$\xi = -\alpha_t(1 - p_t)^r lg p_t \tag{2}$$

where N represents the number of nested layers of L-U2NetP, and n represents the present nested layer, ξ_n and ξ_{fuse} represent the m-th layer loss value and the loss value of the fusion feature map, respectively; every focal loss ξ is calculated by Formula (2), where α_t represents the weighted values of the positive and negative samples, p_t represents the classification probability for each category, r represents the weighted value of the sample difficulty degree.

3.5. Evaluation Metrics

The evaluation metrics of the wheat lodging area extraction used here are the *Accuracy*, *F1*, and *IoU*. The extraction results of the wheat lodging can be divided into *TP*, which is a lodging feature detected correctly; *TN*, which is a non-lodging feature detected correctly; *FP*, which is a non-lodging feature falsely detected as a lodging feature; and *FN*, which is a lodging feature of wheat that is falsely detected as a non-lodging feature.

Accuracy is used to evaluate the global accuracy of the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Before calculating the other four evaluation metrics, in this article, we make improvements to the calculation premise of them based on the task characteristics. Unlike other segmentation tasks, the test set of the lodging segmentation task has a large number of images without targets (lodging areas). *TP* is 0 when these images are predicted completely correctly, which means that all metrics with *TP* as the product factor will also be 0. This will greatly reduce the final average value of the metrics and affect the researcher's judgment on the performance of the model. Therefore, in this this paper, we propose that the other four evaluation metrics should be included in the calculation of the final average value only when the accuracy is 1 and the target image is not fully non-lodging, or the accuracy is not 1.

Precision represents the proportion of the number of pixels correctly extracted as the wheat lodging area in the image to be tested.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

Recall is the proportion of all the wheat pixels in the test set that are correctly recognized as wheat lodging area pixels.

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

The *F1* score is a harmonized average of precision and recall that is used to take into account the performance of a classifier.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

IoU calculates the intersection ratio between the real lodging of the tag and the lodging predicted by the system.

$$IoU = \frac{TP}{TP + FP + FN} \tag{7}$$

The *IoU* reflects the coincidence degree between the predicted and the real values. The closer the ratio is, the higher the degree of coincidence is and the higher the quality of the semantic segmentation is.

4. Results

4.1. Performance of Different Models on CA Set

In order to accurately extract the lodging area of wheat, the L-U2NetP, U2NetP, U2Net, UNet, and SegNet models were used to segment the UAV images of the wheat lodging. In this part of the comparison, the above five models all used the training and test set in CA set.

It can be seen from Table 3 that except for UNet, the accuracy of the other four models on the simple and difficult set are over 94% and 87%. Compared with U2NetP, the Acc, F1, and IoU values of L-U2NetP are 95.45%, 93.11%, and 89.15% higher on the simple set, and 95.45%, 93.11%, and 89.15% higher on the difficult set. Moreover, the network L-U2NetP, which had the least parameters, outperformed multiple deep networks in all the evaluation metrics, with both an advantage on the simple and difficult set.

Table 3. Segmentation performance of different models in Normal set1.

Model	Simple Set			Difficult Set			Parameters (Million)
	Acc	F1	IoU	Acc	F1	IoU	
L-U2NetP	95.45%	93.11%	89.15%	89.72%	79.95%	70.24%	1.10
U2NetP	94.80%	91.97%	87.89%	87.83%	78.90%	69.23%	1.13
UNet	75.95%	73.01%	63.57%	66.18%	49.02%	36.41%	29.44
U2Net	94.61%	91.69%	87.62%	88.24%	78.06%	68.07%	44.01
SegNet	94.17%	91.16%	86.33%	88.57%	76.27%	65.65%	31.03

4.2. Comparison of Robustness of Different Models in Real-Time Detection Simulation

To simulate and assess the influence of factors such as changes in lighting, obstructions caused by aerial foreign objects, and the motion blur of UAVs on the model during real-time detection, the models were used to segment the dataset with designated data augmentation. The Light set, Occlusion set, and Motion_blur set were created through adjustments in the contrast and brightness changes, random occlusion, and mild blur, respectively.

As shown in Table 4, L-U2NetP, U2Net, and SegNet achieve an accuracy of more than 85% in the simple set despite changes in lighting. On the difficult set, L-U2NetP exhibits an increase of more than 11% in all the evaluation metrics compared to U2NetP. As can be seen from Table 5, the influence of random occlusion on the difficult set is more pronounced than on the simple set. Meanwhile, L-U2NetP still has certain accuracy advantages compared with the other networks. As can be seen from Table 6, motion blur has a significant influence on the lodging segmentation models. L-U2NetP achieves nearly 90% accuracy on the simple set, and performs significantly better on the difficult set than other deeper and wider networks. The results verified the excellent robustness of L-U2NetP in simulating the real-time extraction of wheat lodging.

Table 4. Segmentation performance of different models in Light set.

Model	Simple Set			Difficult Set		
	Acc	F1	IoU	Acc	F1	IoU
L-U2NetP	85.96%	83.02%	77.97%	83.58%	69.42%	59.44%
U2NetP	73.84%	69.78%	64.36%	71.96%	55.70%	46.91%
UNet	73.56%	71.70%	62.49%	63.08%	49.58%	37.02%
U2Net	87.55%	86.30%	81.50%	77.89%	64.72%	55.27%
SegNet	85.49%	82.38%	76.03%	81.92%	69.21%	58.69%

Table 5. Segmentation performance of different models in Occlusion set.

Model	Simple Set			Difficult Set		
	Acc	F1	IoU	Acc	F1	IoU
L-U2NetP	94.13%	89.61%	84.55%	85.82%	72.93%	61.60%
U2NetP	93.42%	88.48%	83.46%	84.74%	73.82%	62.71%
UNet	72.48%	71.77%	60.53%	64.40%	49.63%	36.05%
U2Net	93.31%	89.70%	84.71%	85.55%	74.33%	63.27%
SegNet	91.16%	87.70%	81.63%	83.56%	72.26%	60.54%

Table 6. Segmentation performance of different models in Motion_blur set.

Model	Simple Set			Difficult Set		
	Acc	F1	IoU	Acc	F1	IoU
L-U2NetP	89.64%	88.38%	83.72%	75.16%	68.89%	58.24%
U2NetP	91.73%	89.63%	84.67%	82.57%	73.05%	61.99%
UNet	76.70%	76.37%	68.47%	60.41%	50.92%	38.83%
U2Net	85.49%	85.02%	79.04%	67.97%	63.42%	52.28%
SegNet	78.69%	74.75%	65.99%	75.87%	56.48%	44.68%

4.3. Comparison of Generalization Ability of Different Models on AC Set

The performance of the different models in the AC set is shown in Table 7 and Figure 7. The accuracy of L-U2NetP and SegNet is greater than 90% on every set. However, the model parameters of the L-U2NetP model are only 1/28 of the SegNet model, which showed a significant lightweight advantage. Compared with U2NetP, L-U2NetP showed an accuracy advantage of approximately 7% to 12% on the three datasets with a slight reduction in the parameters. Overall, L-U2NetP outperforms the other models in the test set with a finer annotation strategy than the training set.

Table 7. Accuracy of different models trained by CA set in AC set.

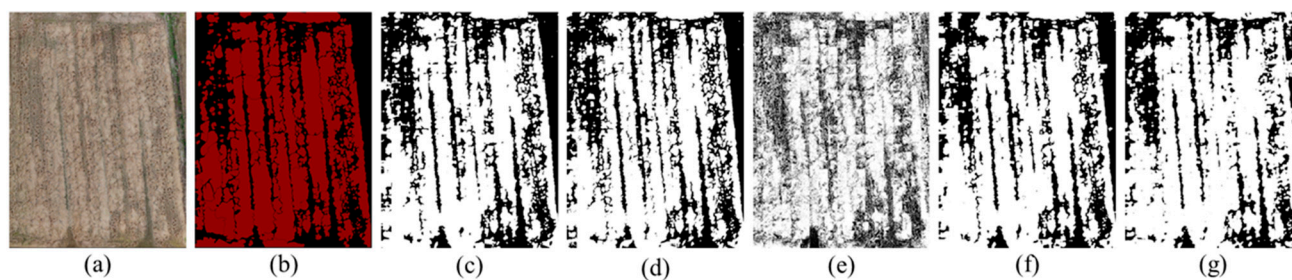| Models	All	Simple	Difficult
L-U2NetP	91.67%	97.30%	90.63%
U2NetP	84.36%	85.53%	81.35%
UNet	61.89%	50.68%	63.49%
U2Net	87.93%	94.49%	85.89%
SegNet	90.98%	97.34%	90.53%

Figure 7. Comparison of performance of different network models on AC set. From left to right are (**a**) the original image, (**b**) annotated image, and the segmentation results obtained from (**c**) L-U2NetP, (**d**) U2NetP, (**e**) UNet, (**f**) U2Net, and (**g**) SegNet.

5. Discussion

5.1. Comparison of the Proposed Method with Previous Crop Lodging Segmentation Methods Based on RGB Images

Currently, more research on crop lodging extraction based on UAV RGB images adopts deep learning methods based on the Annotation-crop strategy for lodging area segmentation. Based on RGB images extracted from UAV multispectral images, SegNet was used to predict sunflower lodging information, achieving an accuracy of 83.98% and 88.53% on two plots, respectively [43]. UNet was used to extract the rice lodging area based on UAV RGB images, and a dice coefficient of 0.9442 was obtained [44]. In addition to utilizing large-scale neural networks, there remain some studies that have achieved promising results using a lightweight model based on the Annotation-crop strategy. Based on UAV RGB images, a lightweight PSPNet was used to segment the wheat lodging areas during the filling and ripening stages, with an accuracy of 89.32% [45]. The Mobile UNet with only 1/3 of the original network parameters and fused data were used to segment the wheat lodging area, achieving an accuracy of 88.99% [29]. However, its parameter quantity still reached 9.49 million. The accuracy of the method proposed in this study is comparable to previous methods and has shown an accuracy of over 90% on multiple test sets. The ultra-lightweight model in this work only used the 1.10 million parameter, which is far smaller than some existing segmentation models that can be used for real-time detection [46,47], and demonstrated good robustness and generalization ability in various test sets. However, we only used images of a single growth stage and height, while many previous research methods have used deep learning with more dimensional data. Therefore, the training samples used in this study are slightly smaller than those in previous studies. Similarly, obvious ridge information should also be identified [29] or avoided [28] in the detection of lodging areas. However, due to the dense planting of the experimental plot, this article only obtained information composed of wheat and slight agricultural machinery operation traces, as shown in Figure 1. So, this question has not been addressed.

5.2. Necessity of Model Robustness Testing in Real-Time Detection Simulation

This study used an ultra-lightweight network, L-U2NetP, to accurately segment wheat lodging areas based on UAV imagery. One of the ultimate goals of network lightweighting is to conveniently deploy it on various edge devices to complete various real-time tasks [17]. In fact, many previous studies have mentioned the importance of the real-time monitoring of wheat [48] and the cost-effectiveness of RGB images [49] in this task. However, we have found that many studies have not conducted real-time performance verification experiments or simulations on segmentation models due to their large volume. However, the model proposed in this article has made significant progress in terms of being lightweight while ensuring a certain performance. Therefore, testing the possible influence of the real-time detection status can effectively evaluate the problems that may arise in reality. Light changes [50], obstructions caused by aerial foreign objects [51], and motion blur [52] are three common and significant issues that can affect UAV imagery. It can be inferred from the performance of the models in the Disturb set that these three issues have varying

degrees of impact on the segmentation accuracy. However, the ultra-lightweight model still shows reliable robustness while achieving a deeper networks' ability to extract and learn crop lodging information.

5.3. Necessity of Using Crop-Annotation Strategy for Crop Lodging Extraction

According to the research of [28], it can be found that the appropriate size of the input images will cause a certain improvement in the model performance. This means that obtaining more information from the input image cannot be achieved solely by enlarging the image size. Moreover, previous studies have found the importance of boundary [53] and texture [54] features in lodging images. Compared with the target object in other segmentation tasks, the structural and positional information of the lodging area is less important. Therefore, directly annotating cropped images with a fixed size can make annotators focus on local information which consist of boundary and texture information that better meets the requirements of this task. Meanwhile, this strategy further alleviates the problem of lagging annotation software caused by a large image size and high resolution under certain hardware constraints [29], making the annotation process smoother. And it can also alleviate the problem of image distortion caused by downsampling the input [30]. The results showed that the new annotation strategy has a reliable quality and the ultra-lightweight model based on it is capable of extracting fine-grained information and achieving high-quality segmentation. However, although the task requires less global features, the proposed annotation strategy results in annotators not being able to grasp global information well during the annotation stage. This may require annotators to compare cropped images with the original image in some blurry areas when using this annotation strategy. Hence, this annotation strategy should also be combined with more semi-automatic annotation strategies [55,56] to release more pressure on annotators.

5.4. Future Challenges

To cope with more complex application scenarios, it is necessary to obtain more UAV imagery data of different-height cultivation techniques to further validate the method the study proposed. Meanwhile, the temporal and spatial generalization abilities of models for lodging detection is equally important, and more reproductive stages, areas of study, or field conditions should also be considered [57]. In the future, more sophisticated image processing techniques can be used to simulate the many problems that models may encounter practically, in order to more realistically verify the model performance. Utilizing these simulation methods, the model can not only benefit from augmented data, but can better evaluate its feasibility in more complex application scenarios.

6. Conclusions

This paper presented an ultra-lightweight network, L-U2NetP, and a novel annotation strategy for wheat lodging area extraction based on UAV imagery. And the model was compared with the results from the U2NetP, UNet, U2Net, and SegNet models. The model with only 1.10 M parameters, which is approximately 1/40 of U2NetP, 1/27 of UNet, and 1/28 of SegNet, exhibited an excellent performance in the following aspects.

1. L-U2NetP achieved a segmentation accuracy of 95.45% and 89.72% on the simple and difficult set with the same enhancement as the training set, which outperformed all the other comparative networks.
2. In the real-time detection simulation test, L-U2NetP demonstrated accuracy rates of 85.96% and 83.58% on the simple and difficult sets, respectively, under changing light conditions. Similarly, under aerial foreign object obstruction, the accuracy rates were 94.13% and 85.82% on the simple and difficult sets, respectively. When motion blur was present, the accuracy rates stood at 89.64% and 75.16% on the simple and difficult sets, respectively. L-U2NetP outperformed the other models in the majority of the evaluation metrics and showed strong robustness on the above simulation.

3. Compared with the U2NetP network, the accuracy improved by 7.31% using L-U2NetP on the test set obtained by the Annotation-crop strategy. The results indicated that the L-U2NetP can effectively extract lodging areas and the novel strategy was reliable.

In this study, the proposed method outperforms multiple models previously used in the field of wheat lodging detection in terms of their light weight. The accuracy and robustness of the model have also been fully validated. In addition, the proposed novel annotation strategy also provides new annotation ideas for this work or similar works. Above all, the method we proposed has good potential for use in the detection of wheat lodging.

Author Contributions: Conceptualization, G.F.; methodology, G.F.; software, Y.Z.; validation, G.F.; formal analysis, Y.G.; investigation, G.F.; resources, A.W.; data curation, S.H.; writing—original draft preparation, G.F.; writing—review and editing, C.W., A.W. and B.L.; supervision, B.L.; funding acquisition, A.W. and B.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data are contained within https://github.com/FGQ24/Lodging-area-segmentation.git, accessed on 1 January 2024.

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

References

1. Brune, P.F.; Baumgarten, A.; McKay, S.J.; Technow, F.; Podhiny, J. A biomechanical model for maize root lodging. *Plant Soil* **2018**, *422*, 397–408. [CrossRef]
2. Bing, L.; Jingang, L.; Yupan, Z.; Yugi, W.; Zhen, J. Epidemiological Analysis and Management Strategies of Fusarium Head Blight of Wheat. *Curr. Biotechnol.* **2021**, *11*, 647.
3. Martínez-Peña, R.; Vergara-Díaz, O.; Schlereth, A.; Höhne, M.; Morcuende, R.; Nieto-Taladriz, M.T.; Araus, J.L.; Aparicio, N.; Vicente, R. Analysis of durum wheat photosynthetic organs during grain filling reveals the ear as a water stress-tolerant organ and the peduncle as the largest pool of primary metabolites. *Planta* **2023**, *257*, 81. [CrossRef]
4. Islam, M.S.; Peng, S.; Visperas, R.M.; Ereful, N.; Bhuiya, M.S.U.; Julfiquar, A. Lodging-related morphological traits of hybrid rice in a tropical irrigated ecosystem. *Field Crops Res.* **2007**, *101*, 240–248. [CrossRef]
5. Dong, H.; Luo, Y.; Li, W.; Wang, Y.; Zhang, Q.; Chen, J.; Jin, M.; Li, Y.; Wang, Z. Effects of Diferent Spring Nitrogen Topdressing Modes on Lodging Resistance and Lignin Accumulation of Winter Wheat. *Sci. Agric. Sin.* **2020**, *53*, 4399–4414.
6. Wang, D.; Ding, W.H.; Feng, S.W.; Hu, T.Z.; Li, G.; Li, X.H.; Yang, Y.Y.; Ru, Z.G. Stem characteristics of different wheat varieties and its relationship with lodqing-resistance. *Chin. J. Appl. Ecol.* **2015**, *27*, 1496–1502.
7. Del Pozo, A.; Matus, I.; Ruf, K.; Castillo, D.; Méndez-Espinoza, A.M.; Serret, M.D. Genetic advance of durum wheat under high yielding conditions: The case of Chile. *Agronomy* **2019**, *9*, 454. [CrossRef]
8. Zhu, W.; Feng, Z.; Dai, S.; Zhang, P.; Ji, W.; Wang, A.; Wei, X.-H. Multi-Feature Fusion Detection of Wheat Lodqing lnformation Based on UAV Multispectral lmages. *Spectrosc. Spectr. Anal.* **2022**, *44*, 197–206.
9. Yang, H.; Chen, E.; Li, Z.; Zhao, C.; Yang, G.; Pignatti, S.; Casa, R.; Zhao, L. Wheat lodging monitoring using polarimetric index from RADARSAT-2 data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 157–166. [CrossRef]
10. Zhang, G.; Yan, H.; Zhang, D.; Zhang, H.; Cheng, T.; Hu, G.; Shen, S.; Xu, H. Enhancing model performance in detecting lodging areas in wheat fields using UAV RGB Imagery: Considering spatial and temporal variations. *Comput. Electron. Agric.* **2023**, *214*, 108297. [CrossRef]
11. Liu, T.; Li, R.; Zhong, X.; Jiang, M.; Jin, X.; Zhou, P.; Liu, S.; Sun, C.; Guo, W. Estimates of rice lodging using indices derived from UAV visible and thermal infrared images. *Agric. For. Meteorol.* **2018**, *252*, 144–154. [CrossRef]
12. Mardanisamani, S.; Maleki, F.; Hosseinzadeh Kassani, S.; Rajapaksa, S.; Duddu, H.; Wang, M.; Shirtliffe, S.; Ryu, S.; Josuttes, A.; Zhang, T. Crop lodging prediction from UAV-acquired images of wheat and canola using a DCNN augmented with handcrafted texture features. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
13. Chauhan, S.; Darvishzadeh, R.; van Delden, S.H.; Boschetti, M.; Nelson, A. Mapping of wheat lodging susceptibility with synthetic aperture radar data. *Remote Sens. Environ.* **2021**, *259*, 112427. [CrossRef]
14. Chauhan, S.; Darvishzadeh, R.; Lu, Y.; Boschetti, M.; Nelson, A. Understanding wheat lodging using multi-temporal Sentinel-1 and Sentinel-2 data. *Remote Sens. Environ.* **2020**, *243*, 111804. [CrossRef]

15. Hufkens, K.; Melaas, E.K.; Mann, M.L.; Foster, T.; Ceballos, F.; Robles, M.; Kramer, B. Monitoring crop phenology using a smartphone based near-surface remote sensing approach. *Agric. For. Meteorol.* **2019**, *265*, 327–337. [CrossRef]

16. Gerten, D.; Wiese, M. Microcomputer-assisted video image analysis of lodging in winter wheat. *Photogramm. Eng. Remote Sens.* **1987**, *53*, 83–88.

17. Chauhan, S.; Darvishzadeh, R.; Boschetti, M.; Pepe, M.; Nelson, A. Remote sensing-based crop lodging assessment: Current status and perspectives. *ISPRS J. Photogramm. Remote Sens.* **2019**, *151*, 124–140. [CrossRef]

18. Bah, M.D.; Hafiane, A.; Canals, R. Weeds detection in UAV imagery using SLIC and the hough transform. In Proceedings of the 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), Montreal, QC, Canada, 28 November–1 December 2017; pp. 1–6.

19. Jung, J.; Maeda, M.; Chang, A.; Landivar, J.; Yeom, J.; McGinty, J. Unmanned aerial system assisted framework for the selection of high yielding cotton genotypes. *Comput. Electron. Agric.* **2018**, *152*, 74–81. [CrossRef]

20. Zhang, X.; Han, L.; Dong, Y.; Shi, Y.; Huang, W.; Han, L.; González-Moreno, P.; Ma, H.; Ye, H.; Sobeih, T. A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral UAV images. *Remote Sens.* **2019**, *11*, 1554. [CrossRef]

21. Vélez, S.; Vacas, R.; Martín, H.; Ruano-Rosa, D.; Álvarez, S. A novel technique using planar area and ground shadows calculated from UAV RGB imagery to estimate pistachio tree (*Pistacia vera* L.) canopy volume. *Remote Sens.* **2022**, *14*, 6006. [CrossRef]

22. Matese, A.; Di Gennaro, S.F. Beyond the traditional NDVI index as a key factor to mainstream the use of UAV in precision viticulture. *Sci. Rep.* **2021**, *11*, 2721. [CrossRef]

23. Li, M.; Shamshiri, R.R.; Schirrmann, M.; Weltzien, C.; Shafian, S.; Laursen, M.S. UAV oblique imagery with an adaptive micro-terrain model for estimation of leaf area index and height of maize canopy from 3D point clouds. *Remote Sens.* **2022**, *14*, 585. [CrossRef]

24. Tian, M.; Ban, S.; Yuan, T.; Ji, Y.; Ma, C.; Li, L. Assessing rice lodging using UAV visible and multispectral image. *Int. J. Remote Sens.* **2021**, *42*, 8840–8857. [CrossRef]

25. Chu, T.; Starek, M.J.; Brewer, M.J.; Masiane, T.; Murray, S.C. UAS imaging for automated crop lodging detection: A case study over an experimental maize field. In *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II*; SPIE: Bellingham, WA, USA, 2017; pp. 88–94.

26. Yu, J.; Cheng, T.; Cai, N.; Lin, F.; Zhou, X.-G.; Du, S.; Zhang, D.; Zhang, G.; Liang, D. Wheat lodging extraction using Improved_Unet network. *Front. Plant Sci.* **2022**, *13*, 1009835. [CrossRef] [PubMed]

27. Zhang, D.; Ding, Y.; Chen, P.; Zhang, X.; Pan, Z.; Liang, D. Automatic extraction of wheat lodging area based on transfer learning method and deeplabv3+ network. *Comput. Electron. Agric.* **2020**, *179*, 105845. [CrossRef]

28. Yu, J.; Cheng, T.; Cai, N.; Zhou, X.-G.; Diao, Z.; Wang, T.; Du, S.; Liang, D.; Zhang, D. Wheat Lodging Segmentation Based on Lstm_PSPNet Deep Learning Network. *Drones* **2023**, *7*, 143. [CrossRef]

29. Yang, B.; Zhu, Y.; Zhou, S. Accurate wheat lodging extraction from multi-channel UAV images using a lightweight network model. *Sensors* **2021**, *21*, 6826. [CrossRef] [PubMed]

30. Shen, H.; Su, X.; Zhao, Q. Extraction of lodging area of wheat varieties by unmanned aerial vehicle remote sensing based on deep learning. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 252–260.

31. Yang, W.; Luo, W.; Mao, J.; Fang, Y.; Bei, J. Substation meter detection and recognition method based on lightweight deep learning model. In Proceedings of the International Symposium on Artificial Intelligence and Robotics 2022, Shanghai, China, 21–23 October 2022; pp. 199–207.

32. Li, W.; Zhan, W.; Han, T.; Wang, P.; Liu, H.; Xiong, M.; Hong, S. Research and Application of U 2-NetP Network Incorporating Coordinate Attention for Ship Draft Reading in Complex Situations. *J. Signal Process. Syst.* **2023**, *95*, 177–195. [CrossRef]

33. 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]

34. Tao, F.; Yao, H.; Hruska, Z.; Kincaid, R.; Rajasekaran, K. Near-infrared hyperspectral imaging for evaluation of aflatoxin contamination in corn kernels. *Biosyst. Eng.* **2022**, *221*, 181–194. [CrossRef]

35. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]

36. 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; Proceedings, Part III 18. Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

37. Qin, X.; Zhang, Z.; Huang, C.; Dehghan, M.; Zaiane, O.R.; Jagersand, M. U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognit.* **2020**, *106*, 107404. [CrossRef]

38. Ates, G.C.; Mohan, P.; Celik, E. Dual Cross-Attention for Medical Image Segmentation. *arXiv* **2023**, arXiv:17696. [CrossRef]

39. Huang, Z.; Wang, X.; Huang, L.; Huang, C.; Wei, Y.; Liu, W. Ccnet: Criss-cross attention for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 603–612.

40. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; p. 3.

41. Baloch, D.; Abdullah, S.; Qaiser, A.; Ahmed, S.; Nasim, F.; Kanwal, M. Speech Enhancement using Fully Convolutional UNET and Gated Convolutional Neural Network. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*, 831–836. [CrossRef]
42. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
43. Zhao, S.; Peng, Y.; Liu, J.; Wu, S. Tomato Leaf Disease Diagnosis Based on Improved Convolution Neural Network by Attention Module. *Agriculture* **2021**, *11*, 651. [CrossRef]
44. Zhao, X.; Yuan, Y.; Song, M.; Ding, Y.; Lin, F.; Liang, D.; Zhang, D. Use of unmanned aerial vehicle imagery and deep learning unet to extract rice lodging. *Sensors* **2019**, *19*, 3859. [CrossRef]
45. Zhao, J.; Li, Z.; Lei, Y.; Huang, L. Application of UAV RGB Images and Improved PSPNet Network to the Identification of Wheat Lodging Areas. *Agronomy* **2023**, *13*, 1309. [CrossRef]
46. He, Y.; Zhang, X.; Zhang, Z.; Fang, H.J.C. Automated detection of boundary line in paddy field using MobileV2-UNet and RANSAC. *Comput. Electron. Agric.* **2022**, *194*, 106697. [CrossRef]
47. Yoon, H.-S.; Park, S.-W.; Yoo, J.-H. Real-time hair segmentation using mobile-unet. *Electronics* **2021**, *10*, 99. [CrossRef]
48. Zhang, Z.; Flores, P.; Igathinathane, C.; Naik, D.L.; Kiran, R.; Ransom, J.K. Wheat lodging detection from UAS imagery using machine learning algorithms. *Remote Sens.* **2020**, *12*, 1838. [CrossRef]
49. Jianing, L.; Zhao, Z.; Xiaohang, L.; Yunxia, L.; Zhaoyu, R.; Jiangfan, Y.; Man, Z.; Paulo, F.; Zhexiong, H.; Can, H.; et al. Wheat Lodging Types Detection Based on UAV Image Using Improved EfficientNetV2. *Smart Agric.* **2023**, *5*, 62–74.
50. Burdziakowski, P.; Bobkowska, K. UAV Photogrammetry under Poor Lighting Conditions—Accuracy Considerations. *Sensors* **2021**, *21*, 3531. [CrossRef]
51. Chen, X.; Peng, D.; Gu, Y. Real-time object detection for UAV images based on improved YOLOv5s. *Opto-Electron. Eng.* **2022**, *49*, 210372-1–210372-13.
52. Nasrullah, A.R. Systematic Analysis of Unmanned Aerial Vehicle (UAV) Derived Product Quality. Master's Thesis, University of Twente, Enschede, The Netherlands, 2016.
53. Zhang, G.; He, F.; Yan, H.; Xu, H.; Pan, Z.; Yang, X.; Zhang, D.; Li, W. Methodology of wheat lodging annotation based on semi-automatic image segmentation algorithm. *Int. J. Precis. Agric. Aviat.* **2022**, *5*, 47–53.
54. Liu, H.Y.; Yang, G.J.; Zhu, H.C. The extraction of wheat lodging area in UAV's image used spectral and texture features. *Appl. Mech. Mater.* **2014**, *651*, 2390–2393. [CrossRef]
55. Yu, N.; Hua, K.A.; Cheng, H. A Multi-Directional Search technique for image annotation propagation. *J. Vis. Commun. Image Represent.* **2012**, *23*, 237–244. [CrossRef]
56. Bhagat, P.; Choudhary, P. Image annotation: Then and now. *Image Vis. Comput.* **2018**, *80*, 1–23. [CrossRef]
57. Xiao, Y.; Dong, Y.; Huang, W.; Liu, L.; Ma, H. Wheat Fusarium Head Blight Detection Using UAV-Based Spectral and Texture Features in Optimal Window Size. *Remote Sens.* **2021**, *13*, 2437. [CrossRef]