processes

MDPI

Article

Improved UNet-Based Detection of 3D Cotton Cup Indentations and Analysis of Automatic Cutting Accuracy

Lin Liu [1,2], Xizhao Li [3], Hongze Lv [1], Jianhuang Wang [4], Fucai Lai [5], Fangwei Zhao [1,*] and Xibing Li [1,*]

[1] College of Mechanical and Electrical Engineering, Fujian Agriculture and Forestry University, Fuzhou 350100, China
[2] School of Management, Fujian University of Technology, Fuzhou 350118, China
[3] Department of Intelligent Manufacturing, Shangdong Labor Vocational and Technical College, Jinan 250022, China
[4] Fujian Henge Build Co., Ltd., Longyan 364300, China
[5] Fujian Quanrun Construction Engineering Co., Ltd., Longyan 364200, China
* Correspondence: 000q812049@fafu.edu.cn (F.Z.); lixibing@fafu.edu.cn (X.L.)

Abstract

With the advancement of intelligent technology and the rise in labor costs, manual identification and cutting of 3D cotton cup indentations can no longer meet modern demands. The increasing variety and shape of 3D cotton cups due to personalized requirements make the use of fixed molds for cutting inefficient, leading to a large number of molds and high costs. Therefore, this paper proposes a UNet-based indentation segmentation algorithm to automatically extract 3D cotton cup indentation data. By incorporating the VGG16 network and Leaky-ReLU activation function into the UNet model, the method improves the model's generalization capability, convergence speed, detection speed, and reduces the risk of overfitting. Additionally, attention mechanisms and an Atrous Spatial Pyramid Pooling (ASPP) module are introduced to enhance feature extraction, improving the network's spatial feature extraction ability. Experiments conducted on a self-made 3D cotton cup dataset demonstrate a precision of 99.53%, a recall of 99.69%, a mIoU of 99.18%, and an mPA of 99.73%, meeting practical application requirements. The extracted 3D cotton cup indentation contour data is automatically input into an intelligent CNC cutting machine to cut 3D cotton cup. The cutting results of 400 data points show an 0.20 mm $\pm$ 0.42 mm error, meeting the cutting accuracy requirements for flexible material 3D cotton cups. This study may serve as a reference for machine vision, image segmentation, improvements to deep learning architectures, and automated cutting machinery for flexible materials such as fabrics.

Keywords: 3D cotton cup; UNet model; Atrous Spatial Pyramid Pooling; indentation contour; intelligent cutting

check for updates

Academic Editors: Xiong Luo and Dan Zhang

Received: 24 July 2025
Revised: 18 September 2025
Accepted: 25 September 2025
Published: 30 September 2025

Citation: Liu, L.; Li, X.; Lv, H.; Wang, J.; Lai, F.; Zhao, F.; Li, X. Improved UNet-Based Detection of 3D Cotton Cup Indentations and Analysis of Automatic Cutting Accuracy. *Processes* **2025**, *13*, 3144. https://doi.org/10.3390/pr13103144

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the improvement of living standards, the demand for personalized clothing and apparel products has significantly increased. The deep integration of information technology and industrialization, along with the enhancement of digital capabilities, has greatly improved the industry's resilience and competitiveness, making personalized demands a reality [1]. Therefore, improving labor productivity has become a key focus of digital transformation in the cotton spinning industry [2]. The textile industry is also seeking development through digital transformation, enhancing independent innovation

capabilities, and optimizing industrial structures. Currently, research on the improvement of the textile industry through artificial intelligence is relatively limited. However, this has also brought challenges such as increased production demands and labor shortages [3]. Some textiles, such as cotton cup indentation, with complex a curved machining process, still require manual segmentation.

Some researchers have attempted to optimize cutting paths [4] and cutting tool models [5–7] to reduce cutting errors, but the ability to improve cutting precision is limited. Achieving precise cutting of cotton cup indentations relies fundamentally on the fast and accurate identification of indentation trajectories. But textile materials such as cotton are highly flexible, and the production process of 3D cotton cups involves the coupling of rigid needle-type, knife-type, curved-type, and other mechanical parts. It is necessary to achieve its 3D segmentation and processing without destroying the softness. The increasing variety and shape of 3D cotton cups due to personalized requirements make the use of fixed molds for cutting inefficient, leading to a large number of molds and high costs. So, what we do in this study is develop a machine to segment accurately and automatically, based on a deep learning model.

The most commonly used deep learning image segmentation algorithms include UNet [8], PSPNet [9], and DeepLabv3+ [10], which have shown excellent performance in image segmentation tasks. UNet, a variant of the fully convolutional network, is the simplest and most efficient end-to-end segmentation algorithm. It mainly consists of a primary feature extraction part and an enhanced feature extraction part. Zhao et al. [11] proposed a region-based comprehensive matching method that extracts color information from different regions of clothing materials and retrieves matching areas through weighted region matching. Jiang et al. [12] proposed a deep learning-based method that combines dilated convolution features with traditional convolution features to achieve pixel-level material segmentation across the entire image. Zhong et al. [13] developed the FMNet algorithm with a multi-directional attention mechanism for clothing image segmentation. This algorithm learns semantic information from three perspectives: space, channel, and category, enabling accurate and fast segmentation of clothing images. Deng et al. [14] improved the detection of unclear cracks by introducing attention modules and enhanced residual connections. By using modified residual connections and adding attention modules, they improved the detection of unclear cracks.

There are also many recent studies on improving deep learning through PSP structure, ECA, ASPP, Leaky-ReLU, etc. For example, Ge et al. [15] improved the classification task (of GNNs) through PSP technique, achieved the performance improvement of PSP framework in the sample-less scenario, and pointed out that the PSP is an effective path to solve the problem of prototype vector construction with few samples. Shu et al. [16] improved the U-Net network by ECA to ensure accurate segmentation while significantly reducing the model parameters during fetal ultrasound cerebellar segmentation, and pointed out that ECA is an effective solution to the problem of "balancing high accuracy with low parameter redundancy" in the segmentation task. Ding et al. [17] improved the structure of deep learning models and efficient semantic segmentation in complex scenes in real time by ASPP, and also pointed out that ASPP can significantly improve the segmentation performance of DeepLabv3+, which is a good adaptation choice for segmentation tasks. Ma et al. [18] optimized the performance enhancement of neural network image classification, target detection, and semantic segmentation tasks through customized activation function techniques such as Leak-ReLU and meta-ACON, and pointed out that the activation mechanism can significantly improve the performance of models of different scales in multi-class computer vision tasks, which is an effective direction for the improvement of traditional activation functions. However, these studies demonstrate the significant

usefulness of deep learning in 3D cotton cup image segmentation [19]. However, existing research has discussed how to accurately segment the target under regularized conditions, but accurate cutting of soft materials such as cotton cups with rigid cutting control has rarely been seen [20], and in particular, deep learning techniques for automated cutting of flexible garments with CNC (computer numerical control) have not yet been discussed.

To address the aforementioned issues, the main contribution of this paper lies in the proposal of an improved UNet algorithm (UNet-IV) specifically optimized for the segmentation of 3D cotton cup indentations. Specifically, the innovations of this paper include: (1) The innovative combination of VGG16, ECA (Efficient Channel Attention) attention mechanism, ASPP (Atrous Spatial Pyramid Pooling) module, and Leaky-ReLU activation function significantly enhances the model's segmentation accuracy and robustness for cotton cup contours in complex backgrounds; (2) The algorithm has been successfully applied to industrial CNC cutting machines, achieving a fully automated process from image recognition to physical cutting; (3) Through error analysis of the cut products using a three-coordinate measuring machine, the high precision and feasibility of this method in industrial applications have been verified. The work in this paper focuses on algorithmic innovations and their application validation in specific industrial scenarios.

1.1. Image Collection

A Hikvision industrial camera (MV-CA060-11GM) (Hikvision Digital Technology Co., Ltd., Hangzhou, China) with an effective resolution of 6 megapixels was used for image acquisition. The camera was positioned 1.2 m away from the experimental shooting platform, resulting in the collection of 632 images, each with a resolution of 2048 × 2048 pixels. To enable the network to learn target features under various background noise conditions, the collected cotton cup images included different lighting conditions, different numbers of 3D cotton cups, and different levels of image clarity. Some of the collected images are shown in Figure 1.



Figure 1. Partial samples of image acquisition.

1.2. Image Labeling

Image labeling: From 632 images of cotton cups, 500 images of well-rendered 3D cotton cups were selected. The "LabelMe" software was used to label the semantic segmentation dataset; 450 of these 3D cotton cup images were labelled, with the remaining 50 serving as the test set, as shown in Figure 2. The dataset of cotton cup indentation annotations provides significant value for subsequent machine learning algorithms. Unlike data labeling in object detection tasks, where the target region in the image only needs to be enclosed in a rectangular box, in the semantic segmentation task, the image needs to be zoomed in,

and the contours of the target need to be labeled point by point for sub-pixel level cotton cup indentation annotation. The annotation is completed by enclosing all labeled points in a ring shape. Any areas outside the closed region are considered background. Throughout the labeling process, careful observation of each image is necessary to ensure that each cotton cup area is accurately labeled and to avoid mistakenly labeling the surrounding background as part of the cotton cup region. The JSON file recording the labeled results (Ground Truth mask and image name) was divided into training, validation, and test sets in a ratio of 7:2:1.

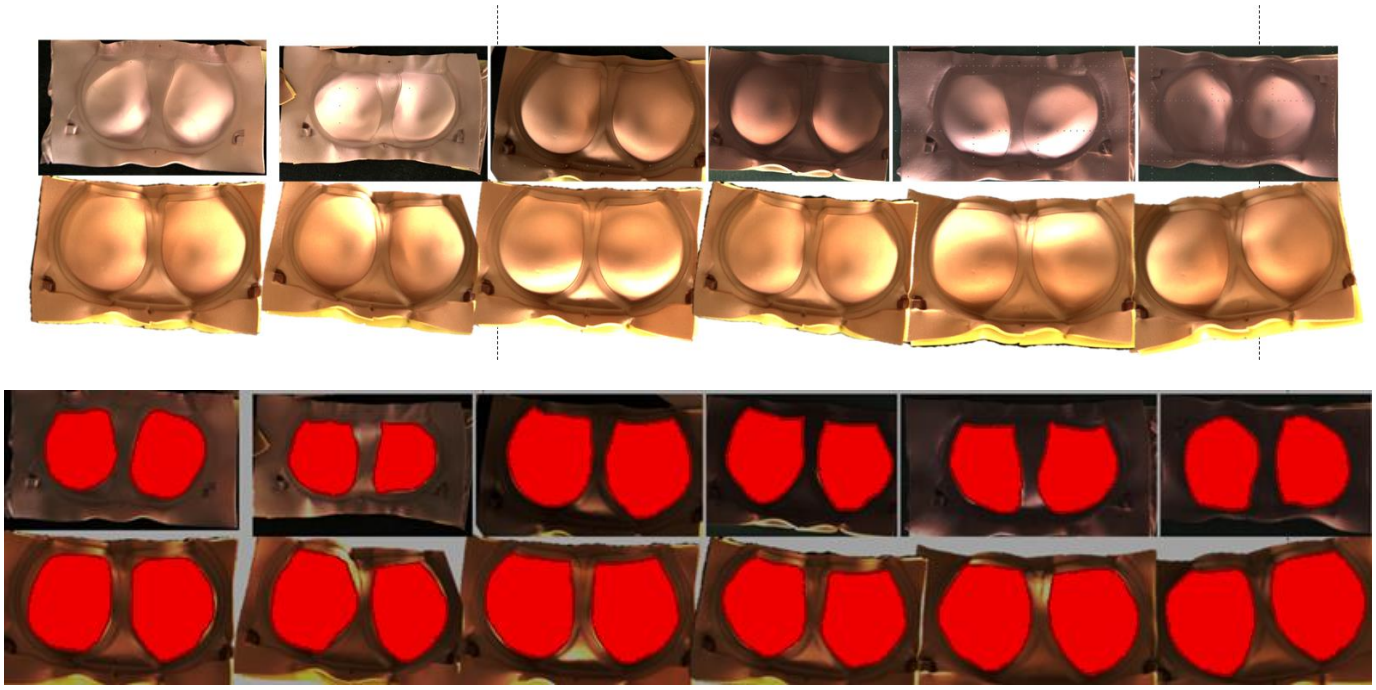


Figure 2. Some sample annotations of 3D cotton cups. The red regions mean the segmented areas and the edge for red regions means the cutting road for CNC.

2. Improvement of the UNet Model

2.1. UNet Model

UNet is a variant of a convolutional neural network (CNN) models. Its backbone feature extraction consists of convolution and down-sampling operations, allowing for deeper feature extraction. The UNet model employs 3×3 convolution layers for convolution operations and 2×2 max pooling layers for down-sampling. Throughout this process, the width and height of the feature maps gradually decrease while the number of feature channels increases, resulting in richer and more abstract features. This enhances the UNet model's ability to express segmentation targets. However, due to the continuous compression of feature maps in the UNet model, some detailed information required for accurate segmentation may be lost. To address this issue and improve the feature extraction capability of the UNet model, an improved structure combines up-sampling and skip connections to restore the original resolution of the feature maps. This approach integrates effective feature layers from the upper part of the model with those from the lower part. Finally, a 1×1 convolution layer is used to adjust the number of channels, resulting in the final segmentation output. In the UNet segmentation model, the main feature extraction part is used for encoding, while the enhanced feature extraction part is used for decoding. The overall structure of the UNet model is shown in Figure 3.

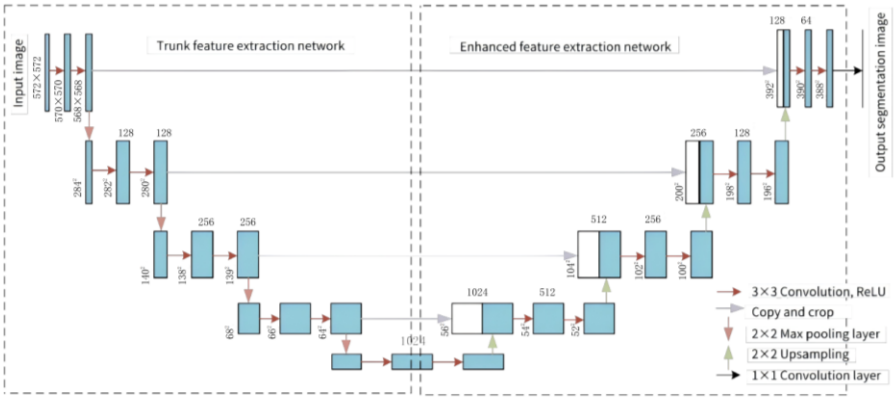


Figure 3. Structure diagram of the UNet network model.

2.2. Improvement of the UNet Architecture

2.2.1. Transfer Learning with the VGG16 Network

The VGG16 network is a classical convolutional neural network structure, known for its excellent feature extraction capabilities and model performance, and is often used as a primary feature extraction network [17]. The VGG16 network adopts a different convolution kernel design approach compared to Alex Net, using consecutive 3 × 3 small kernels instead of large ones, with the number of kernels set to 64, 128, 256, and 512. Compared to larger kernels, small kernels improve the nonlinear representation capability and increase the depth of the network while maintaining the same receptive field. Additionally, small kernels reduce the number of model parameters and the computational complexity of the model, making the network easier to train and optimize. This also enhances the model's nonlinear expression ability, improving its ability to extract and learn detailed features of segmentation targets, thereby boosting the learning capacity for indentations and positional features of three-dimensional cotton cups. To improve the model's generalization ability and reduce dependency on the training samples, a pre-trained VGG16 model is used for transfer learning, which lowers the training cost of the network. The five effective feature layers obtained from the VGG16 network are shown in Figure 4.

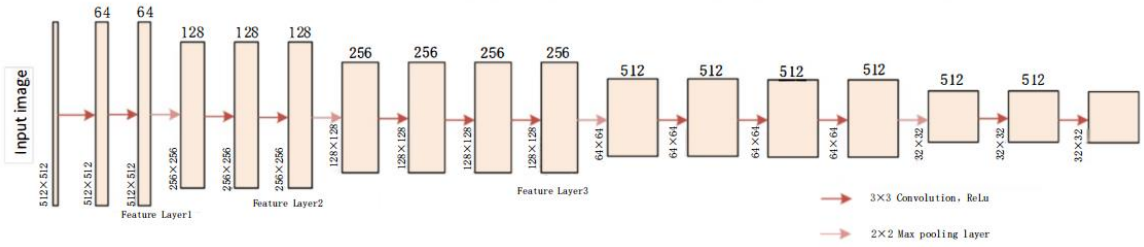


Figure 4. Structure diagram of the VGG16 network.

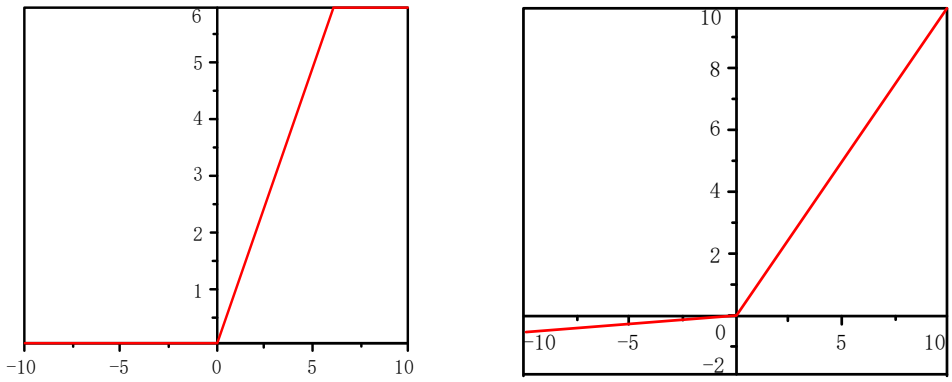
2.2.2. Optimizing the Activation Function in the VGG16 Network

The activation function is a key element in neural networks for handling nonlinear problems [18]. The VGG16 network uses the ReLU activation function, as illustrated in Figure 5a. The ReLU activation function has several advantages: when the input is positive, the derivative is non-zero, allowing for gradient-based learning. Both the function and its derivative involve no complex mathematical operations, leading to fast computation. However, its downside is that when the input is negative, the gradient is zero, and the weights cannot be updated, which may cause learning to become very slow or even lead to neuron failure, resulting in neurons staying inactive (referred to as the "dying ReLU" problem). To address this issue, the Leaky-ReLU activation function is introduced into the original network. As shown in Figure 5b, the Leaky-ReLU activation function is an

improved version of the ReLU function and solves the problem of neuron failure in the negative value range. When the input of the Leaky-ReLU activation function is negative, the output is not completely zero, but instead has a small slope, avoiding the gradient being zero. This reduces the occurrence of inactive neurons, allowing gradient-based learning to continue. Additionally, Leaky-ReLU maintains smoothness within the gradient range of the activation function, making backpropagation and gradient updates easier. This accelerates model convergence, reduces overfitting, and improves overall model performance. The Leaky-ReLU activation function is expressed as shown in Equation (1).

$$f(x) = \begin{cases} x & x \geq 0 \\ \alpha x & x < 0 \end{cases} \quad (1)$$

In Equation (1), "α" is the offset, with a default value of 0.01.



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

Figure 5. Activation function graph: (**a**) ReLU activation function; (**b**) Leaky-ReLU activation function.

2.2.3. Adding the ASPP Module

The ASPP module is added to the end of the enhanced feature extraction part of the UNet model [19]. This module uses atrous (dilated) convolutions to increase the receptive field of the convolutional kernels and constructs a pyramid structure with convolutional layers of different dilation rates. This allows for learning multi-scale information from the target features in three-dimensional cotton cup images, making feature extraction more precise [20]. The ASPP module is commonly used in semantic segmentation tasks as it effectively expands the receptive field to capture image information at different scales. The specific structure of the ASPP module is shown in Figure 6. It includes one 1×1 convolution, three 3×3 atrous convolutions (with dilation rates of 6, 12, and 18, respectively), and a global average pooling operation. Batch normalization is applied after each parallel convolutional layer, as expressed in Equation (2).

$$y(i,j) = \sum_{u=0}^{H} \sum_{v=0}^{W} \times (i + ar \times u, j + ar \times v) \times W(u,v) \quad (2)$$

In this context, the width and height of the input image or the feature map from the previous layer are represented as H and W, respectively, and the value of each pixel is denoted as "x(i,j)". The dilation rate of the dilated convolution is ar, and the output after the dilated convolution is represented as "y(i,j)". Therefore, "y(i,j)" is the feature map obtained after applying dilated convolution to the input image or the feature map from the previous layer. The coordinates of the center point of the convolution kernel are denoted as "(u,v)", where "$(0 \leq u < H, 0 \leq v < W)$"; "W(u,v)" represents the weight at the corresponding

coordinates of the convolution kernel. The output of global average pooling is defined as shown in Equation (3).

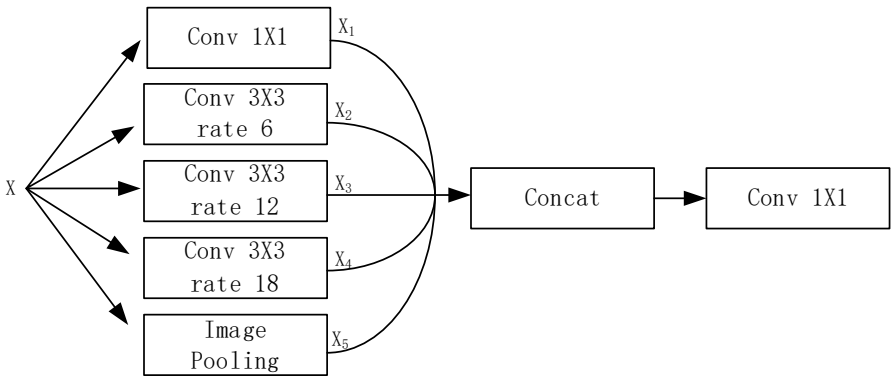


Figure 6. ASPP module.

$$X_5'(c,1,1) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X(c,i,j) \quad (3)$$

Here, X (c,i,j) represents the pixel value at the corresponding coordinates in the c-th channel of the input feature map X; "(c,1,1)" represents the output feature map.

As shown in Figure 6, the ASPP module provides five paths, each with a different receptive field size. Convolution layers with larger dilation rates provide the network with more global contextual information, while those with smaller dilation rates supplement the network with detailed information. The feature maps processed at different receptive fields represent multi-scale information of the same feature level. The first four convolution operations do not alter the spatial dimensions of the feature maps; they only modify the number of channels to produce feature maps such as X_1, X_2, X_3, and X_4. The feature map obtained through global average pooling is up-sampled using bilinear interpolation so that the spatial dimensions of X_5 match the input feature map. Afterward, the processed feature maps from all branches are concatenated along the channel dimension. To reduce computational complexity, the feature maps from each branch are reduced in dimensionality before concatenation, ensuring that all feature maps have the same number of channels. Finally, the concatenated feature maps are passed through a convolution layer with a 1×1 kernel, yielding the final semantic segmentation result. This design effectively improves the model's accuracy and efficiency. When compressing the feature maps to the specified dimensions, multi-scale information within the same feature level interacts along the channel dimension, further extracting deep, high-level features of the three-dimensional cotton cup indentations. This provides more abstract and advanced features to the feature maps in the decoding path.

2.2.4. Adding an Attention Mechanism

To improve the processing of relevant information within the perceptual range of the image, the system employs convolutional neural network deep learning algorithms to further enhance the ability to capture useful feature information. Traditional convolution operations typically fuse all channels of the input feature map, preventing the network from focusing on important feature channels. By introducing the ECA module [21], after a series of convolution operations, the feature map predicts a set of weight values, which are then used to weight the various channel features. This allows the model to focus more on important target information in the feature map, increasing attention to target position information during feature extraction, removing some unimportant information, and reducing the impact of noise on the target feature extraction process. This improves

the accuracy of target detection. A diagram illustrating the ECA attention module process is shown in Figure 7.

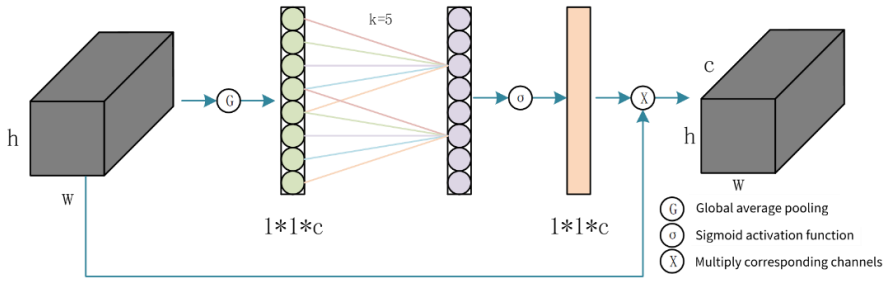


Figure 7. ECA Attention Module.

The ECA attention module applies global average pooling to the input feature map, converting the feature map from a [h, w, c] matrix into a [1, 1, c] vector. After global average pooling, a "1 × 1" convolution layer is used to replace the fully connected layer. This module avoids dimensionality reduction and effectively captures cross-channel interactions, achieving good results with very few parameters and reducing the complexity of the network. Channel-wise interaction is implemented using 1D convolution, and an adaptive 1D convolution kernel size k is used based on the number of channels in the feature map, as shown in Equation (4).

$$k = \left|\frac{\log_2(c)}{\gamma} + \frac{b}{\gamma}\right| \tag{4}$$

In Equation (4), c is the number of channels in the feature map, with "$\gamma = 2$", "$b = 1$".

The convolution kernel size obtained through calculation is used for 1D convolution operations to compute the weight values of each channel in the feature map. These weight values are normalized after being processed by the Sigmoid activation function, allowing layers with a larger number of channels to interact more across channels. Finally, the normalized weights are multiplied by the original input feature map across channels, resulting in a weighted feature map. This module design helps the network better learn the importance of each channel, thereby improving the model's accuracy and robustness. By incorporating the ECA module into the enhanced feature extraction part of the original network, the improved model is obtained, as shown in Figure 8.

In conclusion, the improved architecture proposed in this study is designed to address the specific challenges of the 3D cotton cup segmentation task. The introduction of VGG16 utilizes pre-trained weights to accelerate convergence and enhance generalization ability. Leaky-ReLU addresses the potential issue of neuronal inactivation, which is crucial for detailed indentation edge learning. The ASPP module effectively deals with cotton cups of different sizes and shapes through multi-scale feature fusion. The ECA attention mechanism enables the network to focus more on the indentation area and suppress the interference of background noise. This combination and synergy have significantly enhanced the model's accuracy and robustness.

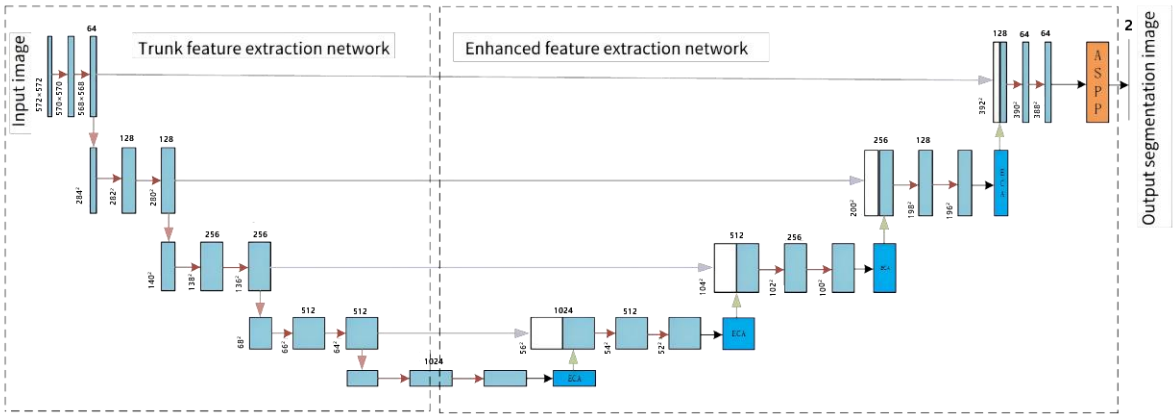


Figure 8. Improved UNet Network Model.

3. Experiment and Result Analysis

3.1. Experimental Environment and Evaluation Metrics

The hardware configuration used in the experiment includes the following: GPU: NVIDIA GeForce RTX 3060; Processor: Intel Core i7-11800H; Video Memory: 6 GB; Software environment: 64-bit Windows 10 operating system; Python 3.9.12, Torch 1.11.0.

For the results obtained in the experiment, commonly used image segmentation performance metrics are selected for evaluation, including Precision, Recall, Mean Intersection over Union (MIoU), and Mean Pixel Accuracy (mPA) [22–24]. The formulas for these metrics are as follows:

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

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{MIoU} = \frac{1}{\text{k} + 1} \times \frac{\text{TP}}{\text{FN} + \text{TP} + \text{FP}} \tag{7}$$

$$\text{mPA} = \frac{1}{\text{N}} \sum_{\text{i}=1}^{\text{N}} \text{P}_\text{i} \tag{8}$$

Here, TP represents the number of pixels correctly predicted as the 3D cotton cup surface, i.e., the number of pixels correctly classified as belonging to the 3D cotton cup surface. FP represents the number of pixels incorrectly predicted as the 3D cotton cup surface, i.e., the number of background pixels misclassified as the 3D cotton cup surface. FN represents the number of pixels incorrectly predicted as background, i.e., the number of 3D cotton cup surface pixels misclassified as background. In Equation (7), k + 1 refers to k classes plus one background class. Equation (8) represents the average precision across all target classes, which is used to measure the overall performance of the network.

3.2. Ablation Experiment

The improved UNet-IV uses the VGG16 network as the backbone for feature extraction in transfer learning, incorporating the ECA attention mechanism and ASPP module, and utilizing the Leaky-ReLU activation function. To compare the impact of the improved modules on the overall performance of the network and to analyze the effect of different improvement modules on image segmentation [25,26], an ablation experiment is conducted. To ensure the accuracy of the ablation experiment, the same dataset and training parameters are used for each group of experiments. In the experiment, the total number of training batches for the model is set to 100, with the first 50 batches employing frozen training. At this stage, the backbone of the model is frozen, allocating more resources to train the

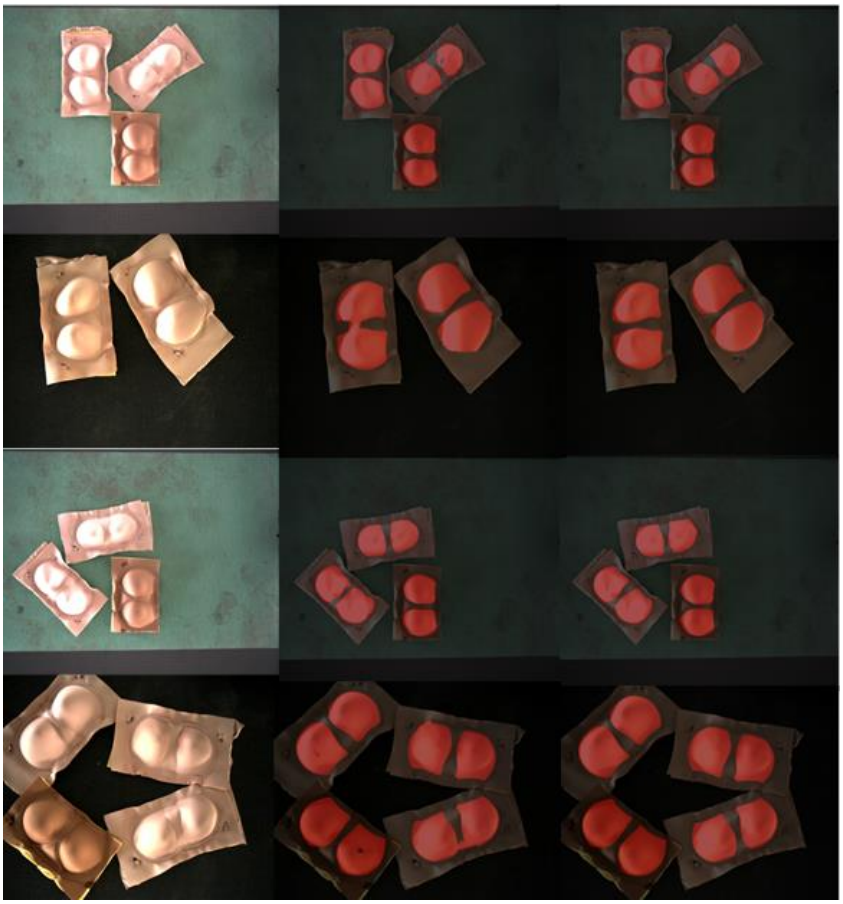
parameters of the subsequent network layers, aiming to reduce the complexity of the model and improve its generalization capability. The remaining 50 batches employ unfrozen training. The batch size is set to 2 throughout the experiment. The results of the experiment are shown in Table 1.

Table 1. Ablation experiment results.

Network Model	VGG16	ECA	ASPP	Leaky-ReLU	Precision/%	Recall/%	MIoU/%	mPA/%
UNet	√	×	×	×	99.43	99.42	98.92	99.49
UNet-I	√	√	×	×	99.50	99.62	99.02	99.55
UNet-II	√	×	√	×	99.46	99.48	98.85	99.65
UNet-III	√	×	×	√	99.41	99.43	98.95	99.50
UNet-IV	√	√	√	√	99.53	99.69	99.18	99.73

In Table 1, checkmarks (√) and crosses (×) indicate whether such models are adopted or not. Although the improvements in the accuracy indicators of each model in Table 1 seem minor (for example, UNet-IV has improved mIoU by 0.26% compared to the baseline UNet), such differences are of great significance in large-scale and high-demand industrial production. As shown in Figure 9, these numerical improvements are directly reflected in the quality of the segmentation results: the UNet-IV model (Figure 9c) can generate smoother and more complete edge contours, effectively avoiding the edge sawtooth, breakage, and false detection problems that occur in the baseline UNet model (Figure 9b). In the automated cutting process, these defects can lead to an increase in the scrap rate and material waste. Therefore, even seemingly minor improvements in precision are crucial for ensuring the quality of the final product and controlling production costs.

In Table 1, all network models use VGG16 as the backbone for feature extraction. UNet-I adds the ECA attention mechanism, resulting in a 0.07% and 0.2% improvement in precision and recall, respectively, a 0.1% increase in the mean mIoU, and a 0.06% increase in the mPA. UNet-II strengthens feature extraction by adding the ASPP module to the UNet, leading to a 0.03% and 0.06% increase in precision and recall, respectively, while the mIoU decreased by 0.07%, and the mPA increased by 0.16% compared to the original network. UNet-III introduces the Leaky-ReLU activation function, resulting in a 0.02% decrease in precision, a 0.01% increase in recall, and a 0.03% and 0.01% increase in mIoU and mPA, respectively, compared to the original network. UNet-IV represents the final improved network model, with a 0.1% and 0.27% increase in precision and recall, respectively, a 0.26% improvement in mIoU, and a 0.24% increase in mPA compared to the original network. Through the improvement of different network modules, the final improved network model shows varying degrees of performance enhancement compared to the original UNet. The precision reached 99.53%, the recall reached 99.69%, mIoU reached 99.18%, and mPA reached 99.73%, meeting the performance requirements for detecting indentations in 3D cotton cups. The feasibility of these proposed methods was mutually verified, and it was found that UNet-IV had the best performance.



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

Figure 9. Original image and detection result images from UNet and UNet-IV: (**a**) Original image; (**b**) UNet model detects images; (**c**) UNet-IV model detects images.

3.3. Comparison of Different Network Models

To further evaluate the performance of the improved algorithm, a comparative analysis of the mean mIoU and mPA was conducted between UNet-IV and other methods such as PSPNet, UNet++, and DeepLabV3+ on the dataset. The analysis results are shown in Table 2. UNet-IV improved by 0.35% and 0.22% in mIoU and mPA, respectively, compared to PSPNet; by 0.21% and 0.17% compared to UNet++; and by 2.35% and 2.41% compared to DeepLabV3+. By comparing the data from these experiments, it is evident that the improved UNet-IV outperforms PSPNet, UNet++, and DeepLabV3+ in overall performance. This demonstrates the significance of the final improvements, providing a solid foundation for the practical application of 3D cotton cup indentation detection.

Table 2. Performance comparison of different models.

Network Model	MIoU/%	mPA/%	Recall/%	F1 Score	Parameter Count/M	FLOPs/G	Training Time/min	Inference Time/ms
UNet-IV	99.18	99.73	99.20	99.11	41	720	550	280
PSPNet	98.83	99.51	98.73	98.50	43	800	600	320
UNet++	98.97	99.56	98.51	98.85	37	640	480	240
DeepLabV3+	96.83	97.32	95.96	96.25	31	560	380	180

In terms of the four parameters MIoU/mPA/Recall/F1, the accuracy of different deep learning models is ranked as follows: UNet-IV (F1 score: 99.11%) > UNet++ (98.85%) > PSPNet (98.50%) > DeepLabV3++ (96.25%), and the other three parameters have a similar

order of magnitude. It can be seen that UNet-IV performs optimally in terms of accuracy and is more in line with the flexibility of the cotton cup, complex object segmentation; it can also be demonstrated that better results are obtained under the dataset and training conditions of this paper [27]. In addition, from the perspective of four parameters: parameter count/FLOPs/training time/inference time, the performance ranking in terms of efficiency is as follows: DeepLabV3++ (the shortest inference time: 180 ms, with a performance of 1/0.18 HZ) > UNet++ (240 ms) > UNet-IV (280 ms) > PSPNet (320 ms). It can be seen that DeepLabV3 has the best speed performance with poor accuracy performance, UNet++ has balanced efficiency, UNet-IV has lower efficiency but high accuracy, and PSPNet has the worst efficiency and average accuracy, and this result parameter can be analyzed similarly from a similar number of parameters/FLOPs/training time. UNet-IV is the best in terms of accuracy and faster in terms of speed performance, although there is still a processing speed difference of about 100 ms compared to DeepLabV3++. But in fact, for the later CNC startup and working process, the impact of 100 ms is not significant; and for the accuracy, higher accuracy has higher cutting precision and lower bad product rate, which is very important for the cost control of the industrial Winfield.

3.4. Analysis of Test Results

To further demonstrate the effectiveness of the UNet-IV model, a more in-depth analysis was conducted. By comparing the performance of the UNet and UNet-IV models across different metrics, it was found that the UNet-IV model outperforms the original UNet model in terms of detection accuracy and robustness. Additionally, application experiments were conducted with the UNet-IV model to detect indentations in 3D cotton cups across different scenarios, and the detection results were evaluated and analyzed. The detection results are shown in Figure 10. Figure 9a represents the original image, Figure 9b shows the detection results of the UNet model, and Figure 9c shows the detection results of the UNet-IV model. As observed in Figure 10, the UNet model's detection results show issues such as incomplete detection of the 3D cotton cup surface, unsmooth edges, and incorrect detection of the surface, leading to significant errors in the extracted 3D cotton cup indentations. In contrast, the detection results of the UNet-IV model are complete and accurate, with smooth edges and no isolated points, successfully extracting complete 3D cotton cup indentations for subsequent cutting processes. This demonstrates the strong practical value of the UNet-IV model.

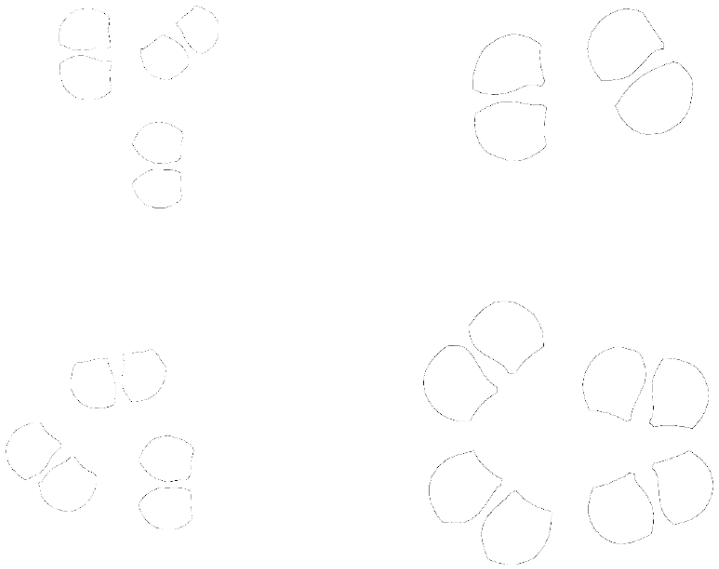


Figure 10. Extracted 3D cotton cup impression image.

Future research will enrich the 3D cotton cup dataset with samples under extreme conditions (e.g., strong reflection, uneven texture) and explore dynamic adaptive segmentation techniques to enhance the model's stability in variable industrial environments. Efforts will be made to lightweight the UNet-IV model (e.g., via depth-wise separable convolution or knowledge distillation) to meet the real-time response requirements of high-speed production lines and enable deployment on edge computing devices. Multi-sensor data fusion (integrating 3D point cloud, force, and temperature data) and extension to multi-category flexible materials (e.g., sponge pads, non-woven fabrics) will be explored to broaden the method's application scope and improve comprehensive quality control.

In conclusion, the improved UNet-IV image segmentation model has a significant advantage in detection accuracy compared to the original UNet model. This model is capable of clearly extracting the track of the three-dimensional cotton cup impression, providing a reliable basis for subsequent segmentation work. The extracted three-dimensional cotton cup impression, as shown in Figure 10, displays a clear impression trajectory, indicating that the improved UNet-IV model effectively addresses the shortcomings of the original model. This improves the accuracy and efficiency of image segmentation, making it highly significant in practical applications.

4. Segmentation and Error Analysis of the 3D Cotton Cup

After collecting the impression trajectory data of the three-dimensional cotton cup using the improved UNet-IV image segmentation model, it was sent in real time to the high-frequency vibration CNC cutting machine (RZCRT5-2516EF, Guangdong Ruizhou Technology Co., Ltd., Foshan, Guangdong, China) for cutting, as shown in Figure 11. A cut 3D cotton cup result with the CNC cutting machine is shown in Figure 12. As shown in Figure 12b, a non-contact coordinate measuring machine (model: DuraMax HTG, Zhejiang Langtong Precision Instrument Co., Ltd., Hangzhou, China), an error analysis is conducted on the three-dimensional cotton cups after cutting.



Figure 11. High-frequency vibrating CNC cutting machine.



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

Figure 12. The CNC cutting result of a 3D cotton cup example. (**a**) A 3D cotton cup example before cutting with the CNC; (**b**) The 3D cotton cup result cut with the CNC; (**c**) Scraps of the 3D cotton cup example.

This device collects 400 uniformly distributed data points on the cutting edge and compares them with the original theoretical design model (such as a CAD digital model). Error is defined as the deviation value of each measurement point from the theoretical model in the direction of its contour normal. Figure 13 shows the error distribution of these 400 points." The average error is 0.20 mm, and the σ (standard deviation) is 0.14 mm. According to the 3σ principle, the cutting error is within the range of 0.20 mm ± 0.42 mm, meeting the cutting accuracy requirements of 3D cotton cups in flexible clothing materials. Then, Figure 13 shows the boxplot of the cutting result, and its *X*-axis represents different cup types, while the *Y*-axis indicates the cutting errors of the CNC. It can be observed that the margin of error varies slightly across different cup types: smaller cup types exhibit a slightly larger margin of error, whereas larger cup types demonstrate a smaller margin of error (e.g., A cup versus D cup). It can be seen from the quartile types of the four cup types, the "range of fluctuation" in error is essentially the same (about 0.14 mm) regardless of cup types. Although the error of the A cup has an overall high bias, the magnitude of error fluctuation of a single measurement is similar to the magnitude of the error fluctuation of the D cup. This suggests that the influence of cup type on the error is mainly reflected in the "overall offset" (mean), rather than "stability" (dispersion). This also may suggest that the small curved edges of the smaller (A cup) are difficult capture accurately and are prone to measurement bias; the large curved edges of the large (D cup) are easier to be measured by the measurement tools (e.g., laser distance measurement, visual inspection), and the deviation is smaller. This may also be related to the fabric's softness under stress; under identical fabric conditions, smaller cup sizes may exhibit differing degrees of post-cutting elongation. This paper focuses on reducing manual labor involvement through camera detection and automated cutting, with an average error margin of 0.2 mm that largely meets the requirements outlined herein. We further infer that subsequent refinements could enhance accuracy by leveraging material microstructural characteristics and CNC control methodologies.

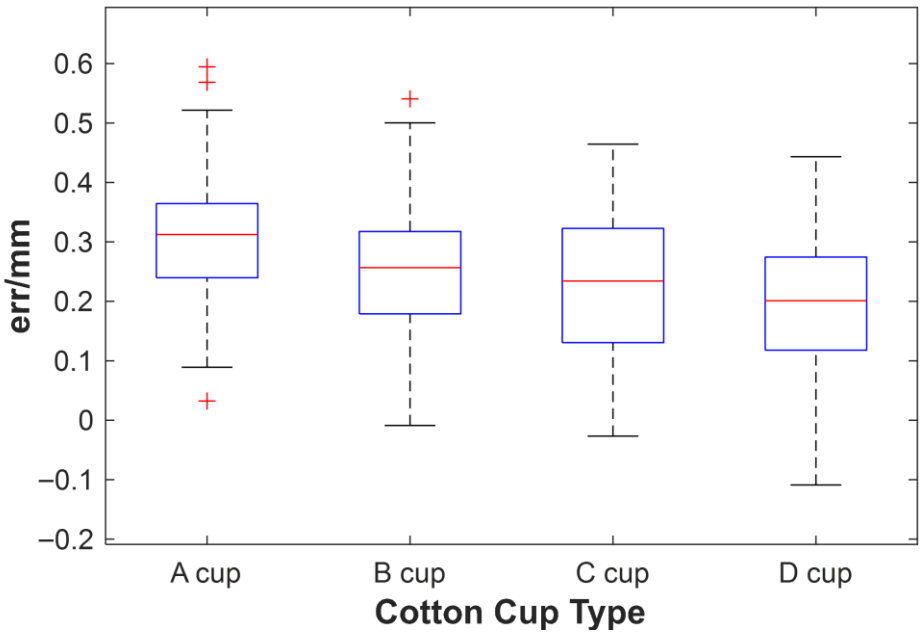


Figure 13. Cutting error of the 3D cotton cup for different cup styles.

5. Conclusions

This paper proposes a 3D cotton cup impression segmentation network model based on an improved UNet algorithm. The model not only saves labor costs and improves

productivity but also meets the cutting accuracy requirements for 3D cotton cup impressions. The VGG16 network is used as the backbone feature extraction network, reducing the number of training parameters while enhancing the model's nonlinear representation capabilities. The introduction of the Leaky-ReLU activation function accelerates the model's convergence and suppresses overfitting. The addition of the ASPP module improves the feature extraction capability, and the inclusion of the ECA attention mechanism enhances the model's ability to learn important features. Experimental results show that the mPA and recall results of UNet-IV reach 99.53% and 99.69%, while the mIoU and mPA reach 99.18% and 99.73%, demonstrating that the model can clearly extract the impression features of the 3D cotton cup. The overall performance of the improved network has been significantly enhanced. The subsequent cutting of the 3D cotton cup resulted in errors within the range of 0.20 mm $\pm$ 0.42 mm, and different cup styes show simple error, meeting the cutting accuracy requirements for flexible garment materials.

Author Contributions: Conceptualization, L.L., F.Z. and X.L. (Xibing Li); methodology, H.L. and X.L. (Xizhao Li); software, F.L. and H.L.; validation, L.L., F.Z. and X.L. (Xibing Li); formal analysis, H.L. and J.W.; resources, F.Z., J.W. and X.L. (Xibing Li); data curation, L.L. and J.W.; writing—original draft preparation, L.L.; writing—review and editing, X.L. (Xibing Li) and F.Z.; visualization, L.L., F.Z. and X.L. (Xibing Li). All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Fujian Research Institute of China Engineering Science and Technology Development Strategy: Research on Intelligent Construction Level and Innovation Ecosystem in Fujian Province (No. 2021-DFZ-20-2), the Natural Science Foundation of Fujian (No. 2022J01609), and the Scientific and Technological Innovation of Fujian Agriculture and Forestry University (Grant No. KFB24039).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Acknowledgments: We thank the Fujian Research Institute of China Engineering Science and technology development strategy, the Fujian Provincial Natural Science Foundation and the Scientific and Technological Innovation of Fujian Agriculture and Forestry University for their support.

Conflicts of Interest: Author Jianhuang Wang was employed by the company Fujian Henge Build Co., Ltd. Author Fucai Lai was employed by the company Fujian Quanrun Construction Engineering Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Ma, W.J.; Gu, G.D.; Li, J.Q. Research on the impact of artificial intelligence on international specialization status in textile industry. *J. Silk* **2022**, *59*, 1–9.
2. Chen, L.; Zhang, X.W. Mechanism of Digital Transformation and Upgrading of Manufacturing. *Jinan J.* **2023**, *290*, 99–110.
3. Heo, S.; Han, S.; Shin, Y.; Na, S. Challenges of Data Refining Process during the Artificial Intelligence Development Projects in the Architecture, Engineering and Construction Industry. *Appl. Sci.* **2021**, *11*, 10919. [CrossRef]
4. Chang, C.Z.; Gao, W.L.; Yang, X.; Li, Z.C.; Huang, Y.F. Laser cutting path optimization using image processing and annealing algorithm with double chain gene. *Manuf. Technol. Mach. Tool* **2023**, *5*, 35–40. [CrossRef]
5. Zhao, M.Y.; Zhai, X.D.; Lin, M. Optimization and test of the working parameters of the cutting disc for the whole-skin peeling system of snakehead. *Trans. Chin. Soc. Agric. Eng. (Trans. CSAE)* **2023**, *39*, 249–255.
6. Feng, C.H.; Chen, X.; Zhang, J.Y.; Huang, Y.G.; Qu, Z.B. Minimizing the energy consumption of hole machining integrating the optimization of tool path and cutting parameters on CNC machines. *Int. J. Adv. Manuf. Technol.* **2022**, *121*, 215–218. [CrossRef]
7. Wojtkowiak, D.; Talaśka, K.; Wilczyński, D.; Górecki, J.; Wałęsa, K. A Coupled Eulerian-Lagrangian Simulation and Tool Optimization for Belt Punching Process with a Single Cutting Edge. *Materials* **2021**, *14*, 5406. [CrossRef] [PubMed]
8. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for Semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef]

9. Zhao, H.S.; Shi, J.P.; Qi, X.J.; Wang, X.G.; Jia, J.Y. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239. [CrossRef]
10. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*; European Conference on Computer Vision; Springer: Cham, Switzerland, 2018. [CrossRef]
11. Zhao, W.L.; Zhou, J.X.; Xu, T.W. National costume image retrieval based on integrated region matching. In Proceedings of the 2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP), Singapore, 4–6 August 2017; Volume 6, pp. 123–129. [CrossRef]
12. Jiang, X.Y.; Du, J.N.; Sun, B.H.; Feng, X.Y. Deep dilated convolutional network for material recognition. In Proceedings of the 8th International Conference on Image Processing Theory, Tools and Applications, Xi'an, China, 7–10 November 2018; IEEE Press: Piscataway, NJ, USA, 2018; pp. 1–6. [CrossRef]
13. Zhong, H.; Zhang, Z.L.; Peng, T.; He, R.H.; Hu, X.R.; Zhang, J. FMNet: Feature alignment based multi-directional attention mechanism clothing image segmentation network. *China Sci.* **2023**, *18*, 275–282.
14. Deng, Z.L.; Luo, R.Z.; Fei, Y.; Li, H.F. Crack detection of airport pavement based on FE-Unet. *J. Optoelectron. Laser* **2023**, *34*, 34–42. [CrossRef]
15. Ge, Q.; Zhao, Z.; Liu, Y.; Cheng, A.; Li, X.; Wang, S. PSP: Pre-training and Structure Prompt Tuning for Graph Neural Networks. In *Machine Learning and Knowledge Discovery in Databases*; Springer Nature: Cham, Switzerland, 2024; pp. 423–439. [CrossRef]
16. Shu, X.; Chang, F.; Zhang, X.; Shao, C.; Yang, X. ECAU-Net: Efficient channel attention U-Net for fetal ultrasound cerebellum segmentation. *Biomed. Signal Process. Control.* **2022**, *75*, 103528. [CrossRef]
17. Ding, P.; Qian, H.; Zhou, Y.; Yan, S.; Feng, S.; Yu, S. Real-time efficient semantic segmentation network based on improved ASPP and parallel fusion module in complex scenes. *J. Real-Time Image Process.* **2023**, *20*, 41. [CrossRef]
18. Ma, N.; Zhang, X.; Liu, M.; Sun, J. Activate or Not: Learning Customized Activation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021), Nashville, TN, USA, 20–25 June 2021. [CrossRef]
19. Yang, W.; Hao, C.; Liang, S. Using Deep Learning Algorithms to Improve Energy Resolution in the Semileptonic Decays. *Wuhan Univ. J. Nat. Sci.* **2022**, *27*, 161–168. [CrossRef]
20. Qi, K.; Pan, R.; Zhou, J. Self-Supervised Fabric Defect Segmentation Based on Optimal Discrete Codebook. *J. Shanghai Jiaotong Univ.* **2022**, *27*, 161–168. [CrossRef]
21. Tang, S.Y.; Liu, Y.R.; Yang, M. Detection of pulmonary nodules based on transfer learning and three dimensional convolutional neural network. *Chin. J. Med. Imaging Technol.* **2020**, *36*, 1882–1886. [CrossRef]
22. Pardede, J.; Sitohang, B.; Akbar, S.; Khodra, M.L. Implementation of Transfer Learning Using VGG16 on Fruit Ripeness Detection. *Int. J. Intell. Syst. Appl.* **2021**, *13*, 52–61. [CrossRef]
23. He, J.C.; Li, L.; Xu, J.C. ReLU deep neural networks from the hierarchical basis perspective. *Comput. Math. Appl.* **2022**, *120*, 105–114. [CrossRef]
24. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef]
25. Hou, Q.; Zhang, L.; Cheng, M.M.; Feng, J. Strip pooling: Rethinking spatial pooling for scene parsing. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 4002–4011. [CrossRef]
26. Jiang, L.X.; Zhou, Y. Efficient Channel Attention Feature Fusion for Lightweight Single Image Super Resolution. *J. Phys. Conf. Ser.* **2021**, *1828*, 012020. [CrossRef]
27. Gao, J.L.; Zhou, H.; Song, H.T.; Guo, J.; Zhang, H. Image Semantic Segmentation Based on Residual Attention and Pyramid Upsampling. *J. Xinyang Norm. Univ.* **2022**, *35*, 134–140. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.