

Research Article

A Real-Time Semantic Segmentation Method of Sheep Carcass Images Based on ICNet

Shida Zhao,^{1,2} Guangzhao Hao,^{1,2} Yichi Zhang,^{1,2} and Shucai Wang^{1,2} 

¹Departments of Engineering, University of Huazhong Agricultural, Wuhan, Hubei 430070, China

²Key Laboratory of Agricultural Equipment in the Mid-lower Yanze River, Ministry of Agriculture and Rural Affairs, Wuhan, Hubei 430070, China

Correspondence should be addressed to Shucai Wang; zsd@webmail.hzau.edu.cn

Received 2 September 2020; Revised 17 March 2021; Accepted 9 April 2021; Published 20 April 2021

Academic Editor: Ramadoni Syahputra

Copyright © 2021 Shida Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to realize the accurate recognition of 3 parts of sheep carcass is the key to the research of mutton cutting robots. The characteristics of each part of the sheep carcass are connected to each other and have similar features, which make it difficult to identify and detect, but with the development of image semantic segmentation technology based on deep learning, it is possible to explore this technology for real-time recognition of the 3 parts of the sheep carcass. Based on the ICNet, we propose a real-time semantic segmentation method for sheep carcass images. We first acquire images of the sheep carcass and use augmentation technology to expand the image data, after normalization, using LabelMe to annotate the image and build the sheep carcass image dataset. After that, we establish the ICNet model and train it with transfer learning. The segmentation accuracy, MIoU, and the average processing time of single image are then obtained and used as the evaluation standard of the segmentation effect. In addition, we verify the generalization ability of the ICNet for the sheep carcass image dataset by setting different brightness image segmentation experiments. Finally, the U-Net, DeepLabv3, PSPNet, and Fast-SCNN are introduced for comparative experiments to further verify the segmentation performance of the ICNet. The experimental results show that for the sheep carcass image datasets, the segmentation accuracy and MIoU of our method are 97.68% and 88.47%, respectively. The single image processing time is 83 ms. Besides, the MIoU of U-Net and DeepLabv3 is 0.22% and 0.03% higher than the ICNet, but the processing time of a single image is longer by 186 ms and 430 ms. Besides, compared with the PSPNet and Fast-SCNN, the MIoU of the ICNet model is increased by 1.25% and 4.49%, respectively. However, the processing time of a single image is shorter by 469 ms and expands by 7 ms, respectively.

1. Introduction

Mutton is the fourth largest meat consumer product in the world, and its demand is increasing with the adjustment of people's dietary structure. According to the statistics, in 2017, the global mutton production was 15.1232 million tons, with an average growth rate of more than 1.6% compared with 2013. It has been estimated that the global mutton production will reach 17.5805 million tons in 2025. China has the largest mutton production and consumption in the world, with mutton output of 4.751 million tons in 2018 and a yearly increase of 0.8% [1–3]. Despite those impressive figures, most of the mutton slaughtering and processing enterprises in China still adopt a semiautomatic

production technology where the cutting process of sheep is completed manually, with high labor intensity and low cutting efficiency, i.e., consuming a lot of manpower and financial resources. At the same time, the harsh production environment is not conducive for the health of operators, and it may pose a risk to food safety [4, 5]. It is thus of great practical significance to design intelligent livestock segmentation robots to replace manual. To this aim, an accurate acquisition of the location of each part of the sheep body in the image is a prerequisite.

In the recent years, most of the studies about sheep body detection focus on measuring the sheep body size and assessing the mutton freshness. These methods are mainly focused on image processing, to combine spectral

technology with the machine learning algorithm, e.g., to extract the characteristics of specific sheep body and mutton (such as contour, color, texture, chemical substances, and so on) from the image or from spectrum at different waveband, and then, according to the extracted feature vector, to design machine learning classifiers to achieve accurate object recognition and detection. Li'na et al. developed a measurement system for sheep's morphological parameters without stress response [6]. Lin et al. have used a portable platform Linux real-time operating system, based on OpenCV, to perform morphological processing on the sheep body image to obtain the sheep body length, height, and other parameters [7]. Meng et al. have put forward an accurate identification method for lamb spine, front leg, and rear leg based on color contrast detection combined with deep learning, the support vector machine, and the back-propagation neural network [8]. Chandraratne has exploited image processing technology to analyze the different combinations features in sheep carcass images which, after ANN and DFA classification, has been used as input to MLP to achieve a classification accuracy of sheep carcass as high as 96.9% [9]. Fan et al. have extracted data from hyperspectral image of mutton in the range 460–1000 nm and have developed a BP/AdaBoost BP neural network to assess the quality of mutton freshness with accuracy about 94.44% [10]. Mohammed et al. have detected three kinds of sheep muscles (semitendon, dorsal longus, and psoas major) in the Charolais sheep using hyperspectral near-infrared technology with 100% accuracy [11]. Although the above research on sheep body detection and recognition has achieved significant results, two main challenges are still limiting the accuracy and application of automatic segmentation methods. First, the key ingredient of image recognition is the extraction of image features. In the above methods, the extraction of objects features depends on the manual work, which is affected by human subjectivity and involves a heavy and tedious workload. At the same time, the extracted features have poor recognition and are far from the expression of natural features of the object. On the other hand, the classifiers designed by machine learning are not robust, have poor generalization ability, and can only be used for specific target objects.

Current, the CNN shows excellent performance in object detection [12–15], image recognition [16, 17], and natural language processing [18] due to the automatic extraction of deep and shallow layers features of the image. At the same time, it has high precision and strong robustness. These features make the image semantic segmentation method based on the CNN a natural candidate to solve the above-mentioned problems of livestock detection. Shen et al. have proposed a real-time detection method, based on the deep convolution neural network, to detect piglets in a complex environment (e.g., the pig house) [19]. The resulting detection speed is 53.19 f/s for 500×375 pixels image with an accuracy rate equal to 95.76%. Yun et al. have designed a PigNet detection network with 44 convolution layers, which may effectively implement segmentation of bonded pigs in herd images [20]. Yang et al. have developed an image segmentation method based on the full convolution neural

network to solve the segmentation problem of sow and piglet. The scheme is based on taking 8 feature maps as sampling and combining deep abstract features and shallow detail features of the image [21]. Noor et al. used a VGG16 network to extract features from sheep facial expression images and then predict sheep expression labels based on the transfer learning training model. Test results show that the accuracy rate may reach 98.4% [22]. Our team has carried out related research on the classification of lamb and on the semantic segmentation of lamb rib images. We have realized the accurate recognition of 6 types of lamb splits based on the Fast-RCNN object detection model, and the mAP of the model reached 84.5% accuracy. Moreover, our team has also implemented the end-to-end pixel level semantic segmentation of the lamb rib region using the U-Net fully convolutional neural network with a segmentation accuracy PA of 92.34% [23, 24].

Combined with the real-time requirements in the actual processing of sheep carcass, the recognition model should have the characteristics of small amount of computation. At present, the commonly used image semantic segmentation methods mainly focus on the DCNN, so that the large amount of floating-point calculations makes it difficult to guarantee real-time performance. Therefore, we choose the ICNet [25] with less calculation to accurately identify the three parts of sheep carcass and verify its feasibility through comparative experiments. Our method aims at images semantic segmentation of sheep carcass from which the head, legs, and abdomen have been removed, to accurately detect the neck, ribs, and spine in real-time. The first step consists in collecting a certain amount of sheep carcass images as experimental data and to expand the data using image augmentation technology. Then, by using the annotation tool on the sheep carcass images, we build the training, test, and validation sets. In the second step, we take into account the time constraint and look for a scheme with good real-time performance. To this aim, we have chosen the ICNet model for the real-time sheep carcass image segmentation model, and after training, we use accuracy, MIoU, and single image processing time as figures of merit to assess the model segmentation performance. As a further step in assessing our ICNet scheme, we consider the U-Net [26], DeepLabv3 [27], PSPNet [28], and Fast-SCNN [29] and perform comparative experiments. In particular, we consider accuracy, MIoU, computational time, and each model's segmentation accuracy and MIoU of the neck, ribs, and spine to assess the advantages and disadvantages of the ICNet in real-time semantic segmentation of sheep carcass images. In this study, a 2D image real-time semantic segmentation based on the ICNet model is utilized in a deep learning model for sheep carcass segmentation. The model's accuracy and real-time performance are realized mainly focusing on the sheep carcass's spine, ribs, and neck. Furthermore, the applicability and advantages of this method are verified through comparative experiments. This research provides theoretical basis and technical support for the development of vision systems of sheep carcass cutting robots.

This study is designed as follows. Section 2.1 describes the sheep carcass image acquisition and preprocessing in the

research. Section 2.2 introduces the algorithm principle and structure of this study. Section 2.4 describes the evaluation criteria of test results. Section 3.1 describes the model training of introducing transfer learning and nontransfer learning and the corresponding comparative experiment. Section 3.2 introduces segmentation result and analysis of sheep carcass images. Section 3.3 describes the ICNet generalization ability test. Section 3.4 introduces the comparative experiments of the method in this study and other methods. Section 4 presents the conclusions.

2. Materials and Methods

2.1. Images Acquisition and Preprocessing

2.1.1. Images Acquisition. The experimental samples are adult Boer goats, and its head, legs, and abdomen fur have been removed. Images have been acquired from the livestock segmentation production line of Inner Mongolia Meiyangyang Food Co., Ltd. The image acquisition consists in a HuaGu Power Technology WP-UC600 color camera with a Z4S-LE-SV-1214H Omron lens. The camera has been set to 1.4 meters from the ground and 0.8 meters from the hanging sheep carcass sample, without any specific background and light source. Figure 1 is the sketch of the image acquisition device. In order to expand the differences between the samples and to ensure the generalization ability of the model training, we have randomly collected ten batches of sheep carcass, each batch containing 350 sheep carcasses. We thus have 3500 mutton images in total, with image resolution 3024×4032 pixels.

2.1.2. Image Preprocessing. Deep learning is driven by large datasets, which improve the prediction accuracy of the model and suppress the overfitting phenomenon. On the other hand, since the number of sheep carcass image samples available for our analysis was limited, we resort to augmentation techniques for the images. In particular, we have used the MATLAB image processing toolbox to transform the sheep carcasses by rotation, flipping, and translation (the rotation angle is set to 30 degrees). In this way, we have simulated the sheep carcass in different postures and have collected images by cameras in different locations. Excessive image augmentation would lead to overfitting of the model, and thus, we set the image augmentation factor to 3. Finally, in order to speed up the convergence of the model, we have used the principle of proportional invariance to scale the image to 512×512 pixels. Examples from the image augmentation process are shown in Figures 2(b)–2(d).

2.2. Image Annotation and Generation of Datasets. Our semantic segmentation model training belongs to the class of supervised learning. However, the acquired sheep carcass images do not contain labels and semantics, and it is thus necessary to annotate them manually to meet the requirements of model training. We have made reference to the Cityscapes datasets format and used the LabelMe image annotation tool to annotate the sheep carcass images. After

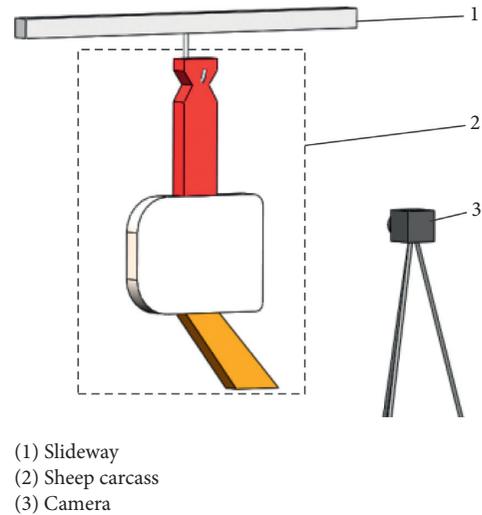


FIGURE 1: Illustration of the image acquisition device.

this stage, we have randomly split data into the training set, testing set, and validation set according to the principle of deep learning dataset division. We have used 8000 images, 2000 images, and 200 images, respectively. The labels of the spine, ribs, and neck of the sheep carcass are sheep 1, sheep 2, and sheep 3, as shown in Figure 3. The RGB 3 channel values of the three types of labels are given in Table 1.

2.3. Image Semantic Segmentation Model

2.3.1. Structure of the ICNet Model. There are several deep learning models currently available to realize image semantic segmentation, such as the DeepLabv3, PSPNet, U-Net, SegNet [30], and FCN [31]. All the models may achieve high accuracy in image segmentation for the ImageNet dataset, but the segmentation takes a long time, and they cannot be employed in real-time applications. The lightweight semantic segmentation model ENet [32] allows one to save time, however, at the price of a reduced accuracy in the segmentation. Zhao has proposed an image.

A cascade network real-time image semantic segmentation module, referred to as the ICNet, which may be used for real-time applications and guarantees a good accuracy rate at same time. The ICNet uses PSPNet's pyramid pooling module to fuse multiscale context information and split the network structure into three branches, i.e., low resolution, medium resolution, and high resolution, as shown in Figure 4. The low-resolution branch scales the feature map of the original from the image size of $1/16$ to the medium resolution output to $1/32$, uses dialed convolution to expand the receptive field, and outputs the feature map at the size of $1/32$ of the original image size, sharing the convolution parameters and weights with the medium resolution branch. The medium resolution branch takes the resolution of $1/2$ of the original image as input and obtains a feature map of the original image size of $1/16$ after the convolution layer, merging the output with the low-resolution output in the cascade feature fusion (CFF) unit to obtain the final output. The high-resolution branch takes the original image as the

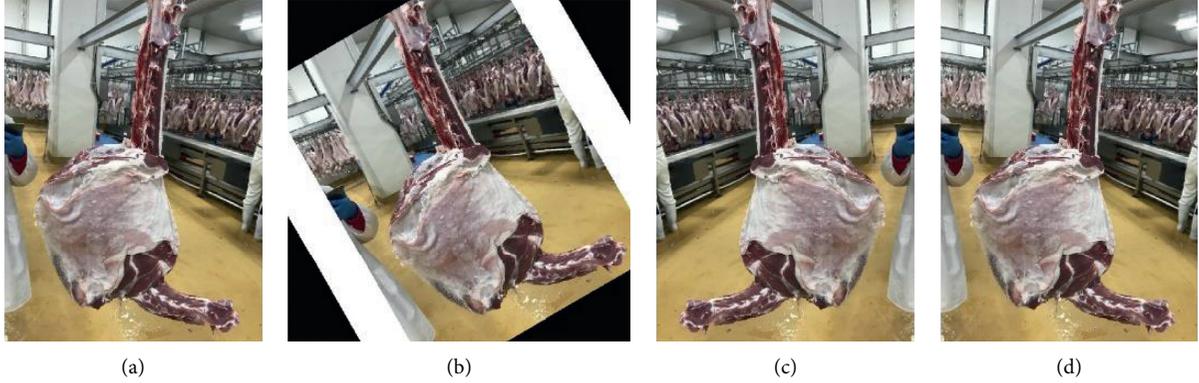


FIGURE 2: Examples of augmented images. (a) Original image. (b) Rotation image. (c) Flipping image. (d) Translation image.

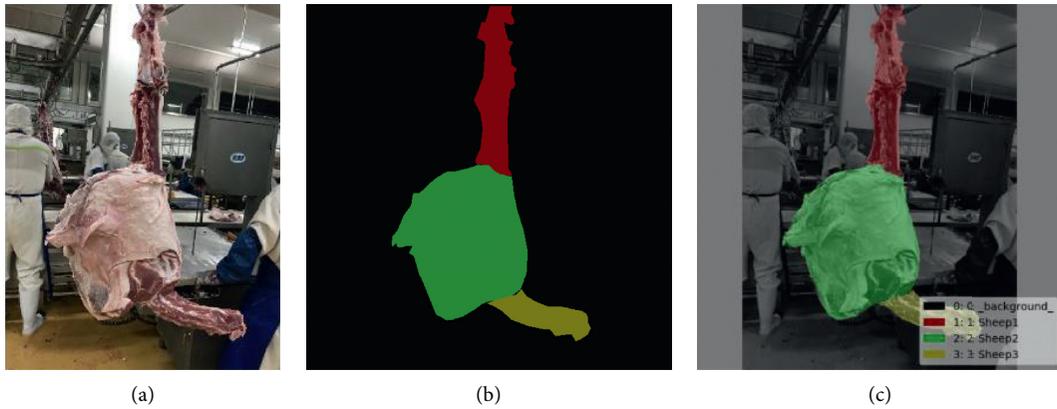


FIGURE 3: Examples of annotated images. (a) Original image. (b) Annotated image. (c) Index image.

TABLE 1: Marked color of three parts in sheep carcass.

Classification	Label	R	G	B
Background	0	0	0	0
Spine	Sheep 1	255	0	0
Ribs	Sheep 2	107	142	35
Neck	Sheep 3	255	245	225

input, obtains the feature map with the size of $1/8$ of the original image after the convolution layer, and then combines the output feature map with the medium resolution through CFF unit, expanding the feature map to the original size after multiple sampling. The ICNet uses low resolution images to complete semantic segmentation and high-resolution refinement for the segmentation results to improve segmentation accuracy. Furthermore, the ICNet uses cascading tags to guide the training of each branch, speeding up the convergence and prediction, and improving the real-time performance. Overall, the ICNet is able to suitably match the conflicting requirements of accuracy and real-time operation.

2.3.2. Cascade Feature Fusion Unit. ICNet uses CFF unit to achieve feature fusion. The CFF unit includes three parts: a low-resolution branch feature map F1, a high-resolution

feature map F2, and a cascade tag. At first, F1 uses 2 times upsampling to restore its size to F2 size and then uses 3×3 dilated convs to input F2 branch after batch normalization. F2 uses 1×1 dilated convs to ensure that it has the same output size as F1, fuses the output features from the F1 branch through the batch normalization layer, and finally obtains F2' with the same resolution as F2 through the ReLU nonlinear activation function, leaving it ready for the next level. The effect of cascaded tags is to strengthen the learning of F1, to optimize the softmax cross-entropy, and obtain a new loss value to update the model weight. The structure of CFF unit is shown in Figure 5 [25].

2.3.3. Loss Function. ICNet adds loss weight to each branch training and optimizes the weighted softmax cross-entropy. The loss function can be expressed as in equation (1).

$$\mathcal{L} = - \sum_{t=1}^{\mathcal{T}} \lambda_t \frac{1}{\mathcal{Y}_t \mathcal{X}_t} \sum_{y=1}^{\mathcal{Y}_t} \sum_{x=1}^{\mathcal{X}_t} \log \frac{e^{\mathcal{F}_{n,y,x}^t}}{\sum_{n=1}^{\mathcal{N}} e^{\mathcal{F}_{n,y,x}^t}}. \quad (1)$$

In the formula, \mathcal{T} is the number of resolution branches, i.e., 3. \mathcal{N} denotes the number of categories, \mathcal{F}^t is the predicted feature map in the branch t , $\mathcal{Y}_t \mathcal{X}_t$ denotes the space size of the feature map \mathcal{F}^t . $\mathcal{F}_{n,y,x}^t$ is the value of position (n, y, x) ; the true label of 2D position (y, x) is \hat{n} . If

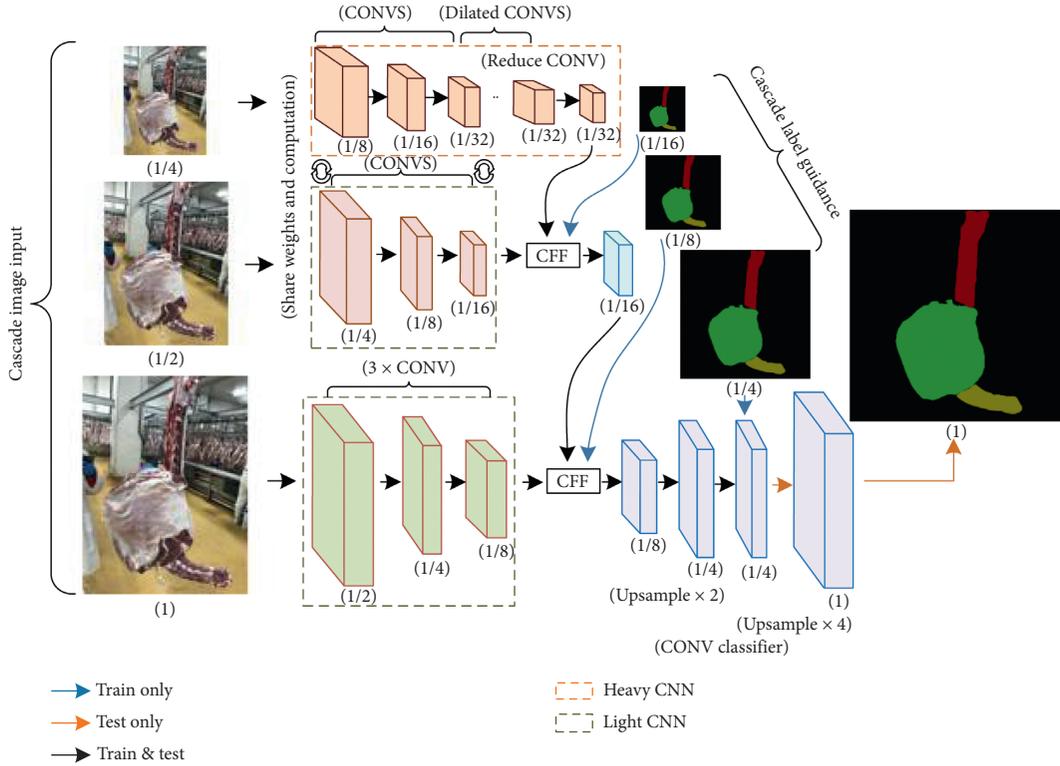


FIGURE 4: Architecture of the image cascade network.

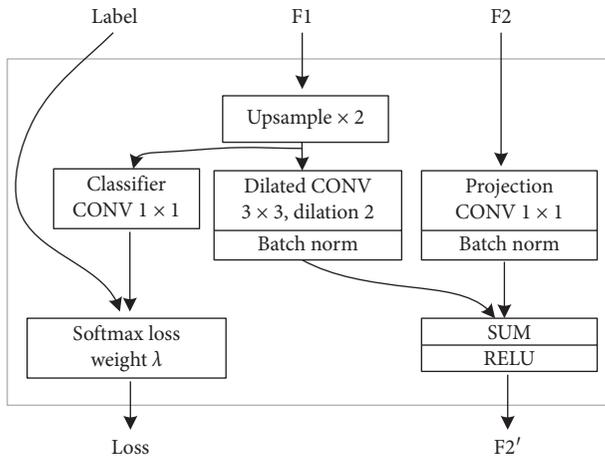


FIGURE 5: Cascade feature fusion unit.

the high-resolution branch weight λ_3 is set to 1, the usual values of the weights λ_2 and λ_1 of the medium-resolution and low-resolution branches are 0.4 and 0.16, respectively.

2.4. Evaluating Criteria. In order to quantitatively analyze the performance of the ICNet model in semantic segmentation of the sheep carcass image dataset, we introduce the PA (pixel accuracy) and MIoU (mean intersection over union). PA and MIoU can be expressed as in equations (2) and (3).

$$PA = \frac{\sum_{i=0}^{N-1} n_{ii}}{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} n_{ij}}, \quad (2)$$

where N is the number of semantic categories; in this study, $N = 4$; n_{ii} denotes the number of real pixels in i class semantics, that is, TP (true positives); n_{ij} is the number of pixels in which the semantics of class i are wrongly identified as class j , that is, FP (false positives).

$$MIoU = \frac{1}{N} \sum_{i=0}^{N-1} \frac{n_{ii}}{\sum_{j=0}^{N-1} n_{ij} + \sum_{j=0}^{N-1} (n_{ji} - n_{ii})}, \quad (3)$$

where n_{ji} is the number of pixels whose class j semantics are recognized as class i , namely, FN (false negatives). The MIoU is positively correlated with the segmentation effect of the model, and because of its simplicity and representativeness, it is often used as the main figure of merit to evaluate performance of image semantic segmentation models.

3. Real-Time Semantic Segmentation of Sheep Carcass Image

Our real-time semantic segmentation of sheep carcass image based on the ICNet includes the following three steps:

- A Building a dataset of images of sheep carcass
- B Building and training the ICNet model and generalization ability experiment

C Introducing the U-Net, DeepLabv3, PSPNet, and Fast-SCNN to make a comparative study with the ICNet

The procedure is illustrated in the flow chart of Figure 6.

3.1. Real-Time Semantic Segmentation of Sheep Carcass Images Based on the ICNet

3.1.1. Test Platform. Based on the PyTorch deep learning framework, we built and trained the ICNet model with the Python language in a parallel computational framework that used Cuda 10.1. The hardware platform is a Dell t5820 image processing workstation with 64 g operating memory, Xeon w-2145 3.70 GHz processor, and P4000-8 g GPU.

3.1.2. The ICNet Model Training Based on Transfer Learning. Transfer learning ensures that the model has the ability to suppress overfitting while processing small sample datasets and is capable to accelerate model convergence and improve generalization ability. Before training, we adopt ICNet's pretraining weights for the Cityscapes dataset, load the trained weights and parameters into the convolutional layer while keeping the structure of all convolutional layers unchanged, and conduct all-layer transfer learning training of the ICNet. ICNet training is then performed with stochastic gradient descent using the ADAM optimizer, and the hyperparameters of transfer learning and nontransfer learning are set to be the same, as given in Table 2. The model is designed to automatically save the optimal model during training and use this as the final input model for semantic segmentation of sheep carcass images. During the model training process, trend variations for the loss values of transfer learning and nontransfer learning methods are evident as shown in Figure 7.

From Figure 7, rapid decrease of the loss values for both methods can be observed at the beginning of training. The values then slightly decrease with the increased number of iterations and finally converge to 0.032 and 0.318 for transfer learning and nontransfer learning, respectively. This indicates that the introduction of transfer learning into the model can not only reduce the loss value of the model but also improve the training effect of the model. Therefore, pretraining weights are loaded to all models used in this study.

3.2. Segmentation Result and Analysis of Sheep Carcass Image. We randomly select 200 images of sheep carcass as additional test datasets and conduct tests based on the optimal model to obtain the MIoU and accuracy of the three parts of the sheep, as well as the overall MIoU and total segmentation accuracy of the model. Furthermore, in order to assess whether the model has good real-time performance, the time needed to process a single image is recorded, as well as its average value. Examples of the segmentation process are shown in Figure 8.

The accuracy, MIoU, and the average processing time of single image of the ICNet for the semantic segmentation of

the three parts of the spine, ribs, and neck in the sheep carcass image are given in Table 3.

From Figure 8 and Table 3, we conclude that the segmentation of sheep carcass image in the spine, rib, and neck is accurately performed by our method; each part is clearly distinguished, with oversegmentation and undersegmentation to reduce to a nonsignificant value. In particular, the sheep spine edges with complex features are clearly identified, which is probably due to the deeper low-resolution branch convolution layer of the ICNet and to the multilayer convolution operation, which ensures the extraction of detail abstract features. Besides, ICNet's multiple upsampling feature fusion improves the recognition accuracy. For the additional test set, the overall accuracy and MIoU of the ICNet model are 97.68% and 88.47%, respectively, and the average processing time of the single image is 83 ms, which indicates that the ICNet achieves accurate semantic segmentation of sheep carcass images, with good real-time performance.

3.3. Generalization Ability Test of the ICNet Model. In order to verify the generalization ability of the ICNet model for semantic segmentation of sheep carcass images, we selected 600 sheep skeleton images with the same light source randomly and converted its RGB color space to HSV. In addition, we set 1.5 times and 0.8 times image brightness to simulate different light intensity and establish different brightness image datasets of sheep carcass, including 300 "bright" images and 300 "dark" images. Finally, it is manually labeled and input into the ICNet model for testing. Some segmentation results are shown in Figure 9.

From the image segmentation results in Figure 9, it can be seen that the ICNet can still achieve accurate segmentation of three parts of sheep carcass at the two levels of "light" and "dark," and the edges of spine-rib and rib-neck adhesion areas are obviously distinguished, the area is complete, the contour is clear, and the over segmentation and under segmentation is not significant. Finally, the segmentation accuracy and MIoU of the ICNet for different brightness sheep skeleton images reached 95.23% and 81.67%, which indicates that for the sheep carcass image dataset, the ICNet has strong generalization ability and can meet the segmentation requirements of different brightness scenes.

3.4. Comparative Study on Different Segmentation Methods for Sheep Carcass Images. In order to further test the advantages and disadvantages of our ICNet real-time semantic segmentation method of sheep carcass image, we here compare its performance with that of the U-Net, DeepLabv3, PSPNet, and Fast-SCNN, which are commonly used methods in image semantic segmentation tasks.

3.4.1. Model Training. The training parameters of the above four models are consistent with the ICNet, and they are all trained automatically saving the optimal model. The

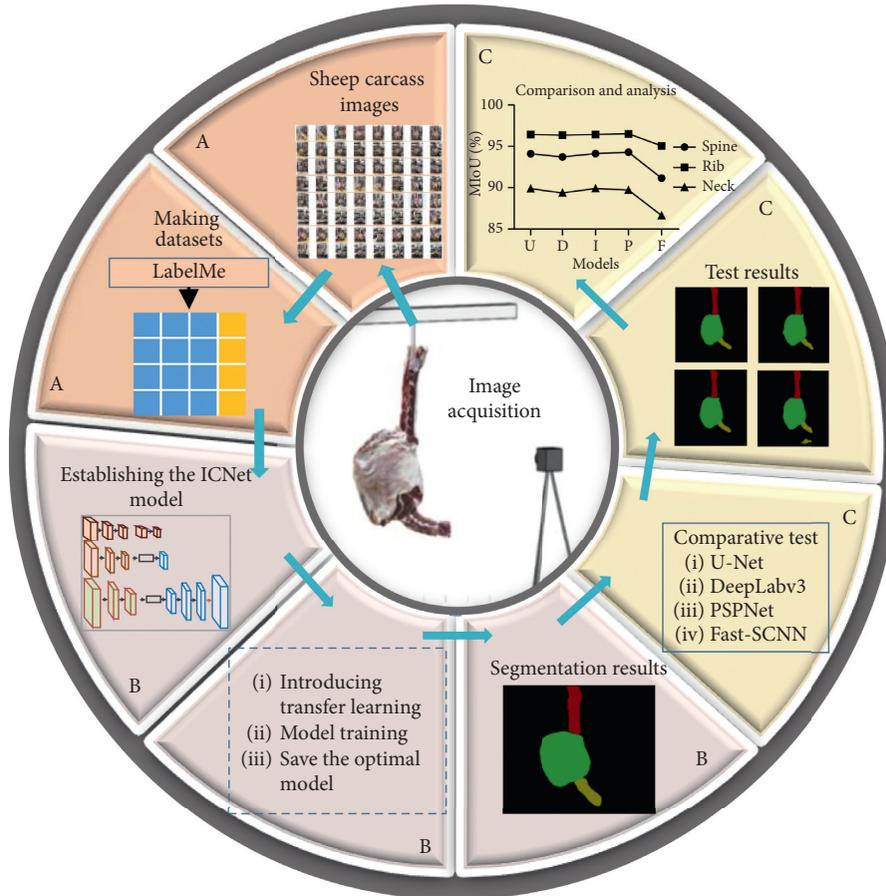


FIGURE 6: Technical flowchart for real-time semantic segmentation of sheep carcass images.

TABLE 2: Training parameters of transfer learning and nontransfer learning methods.

Training parameters	Transfer learning	Nontransfer learning
Initial learning rate	0.001	0.001
Momentum factor	0.9	0.9
Batch size	4	4
Gamma factor	0.1	0.1
Iterations	20,000	20,000

behavior of the loss function with the number of iterations during the training process for each model is shown in Figure 10.

As shown in Figure 10, the loss value of DeepLabv3 and U-Net decreases rapidly at the beginning of training and then shows a plateau around 0.05, finally converging to 0.033 and 0.034, respectively, for a number of iterations around 19,000. The behavior of the loss function for the PSPNet and Fast-SCNN is similar; it decreases rapidly at the beginning of training and then slowly decreases. After 17,000 iterations, the loss value converges to 0.036 and 0.057, respectively.

Moreover, the model's parameter size is also significant toward its deployment on the robot processing system. Ensuring model's accuracy and real-time performance, smaller model parameters mean higher practicality. In this study, the memory size of the model is used to measure its

parameter scale. According to statistics, the minimum memories are 13.6 MB, 100 MB, 224 MB, 679 MB, and 764 MB for the Fast-SCNN, ICNet, U-Net, DeepLabv3, and PSPNet models, respectively.

3.4.2. Semantic Segmentation Results of Sheep Carcass Images. We have used the best model from the training process of the four segmentation models to test the additional test datasets and obtain the final segmentation results of each model, using them to compare results with those of the ICNet. The segmentation results of some sheep carcass images are shown in Figure 11.

From the segmentation results in Figure 11, we see that the U-Net, DeepLabv3, ICNet, and PSPNet may achieve accurate segmentation of the spine, ribs, and neck in 4 sheep carcass sample images, with smooth edges of each part, i.e., they meet the requirements of production for cutting accuracy. However, there are oversegmentation and undersegmentation in the image processing of sample 1 and sample 3 in the Fast-SCNN, which mainly shows that the background and ribs are wrongly segmented into the neck region. The reason for this may be the shallow depth of the Fast-SCNN network and the shallow learning downsampling module that is used to extract multibranch low-level features; when the scale of sheep carcass image

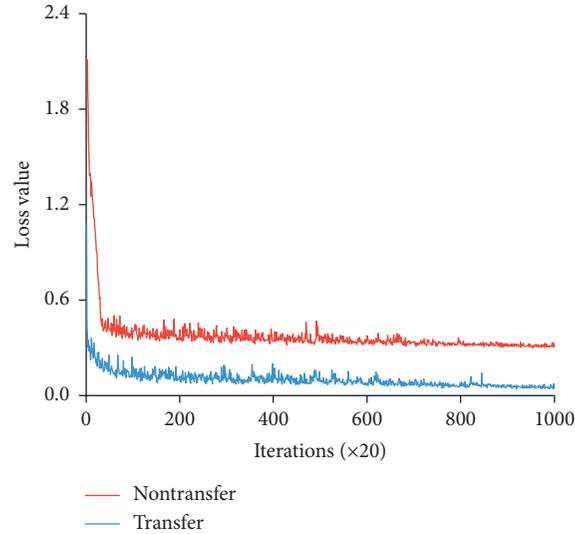


FIGURE 7: Loss value of nontransfer learning and transfer learning on the ICNet model.

TABLE 3: Segmentation results of sheep carcass on the ICNet model.

Types	Accuracy (%)	MIoU (%)	Time (ms)
Spine	94.13	84.61	83
Ribs	96.42	90.88	
Neck	89.90	80.33	
Additional test datasets	97.68	88.47	

data is limited, it is difficult to extract deep abstract features in the image for network learning, making it difficult for the localization of features. On the contrary, the U-Net, DeepLabv3, and PSPNet are deeper than the Fast-SCNN, and they all adopt encoder-decoder structure, so that the models are able to extract more abundant semantic features and better recover the edge information of objects. In addition, the ASPP structure of the DeepLabv3 and the pyramid pooling module of the ICNet and PSPNet make the models able to accurately obtain context information and multiscale features. The overall accuracy, MIoU, and average processing time of single image of the U-Net, DeepLabv3, PSPNet, and Fast-SCNN are given in Table 4.

According to the numbers of Table 3 and Table 4, we conclude that the semantic segmentation model of sheep carcass image based on the U-Net has the highest segmentation accuracy and MIoU, reaching 97.72% and 88.69%, which is 0.03%, 0.19%, 0.04%, and 0.22%, and 0.25%, 1.47%, 0.87%, and 4.71% higher than the DeepLabv3, ICNet, PSPNet, and Fast-SCNN, respectively. The above results also show that the segmentation performance of the five segmentation models is only slightly different.

According to the visual segmentation results in Figure 11, the Fast-SCNN is affected by undersegmentation and oversegmentation, and thus, only the U-Net, DeepLabv3, ICNet, and PSPNet models are suitable to practical segmentation of sheep carcass image. In terms of real-time,

the U-Net, DeepLabv3, ICNet, and PSPNet take 269 ms, 513 ms, 83 ms, and 552 ms respectively, and the ICNet takes the shortest time, which is 69.14%, 83.82%, and 84.96% shorter than the U-Net, DeepLabv3, and PSPNet, respectively. Those results show that ICNet achieves high segmentation accuracy and good real-time performance, thus meeting the actual needs of sheep carcass cutting production line.

Due to the differences in the characteristics of the three parts of the neck, ribs, and spine of the carcass, it is equally important to judge the segmentation effect of the model for each part. We thus have obtained the segmentation accuracy and MIoU of the five models for the three parts. They are as shown in Figures 12 and 13.

As it is apparent from the plot, the segmentation accuracy and MIoU value of the ICNet for three parts of sheep carcass are larger than the PSPNet and Fast-SCNN, and there is no significant difference with U-Net and DeepLabv3 results. The ICNet's segmentation accuracies of the neck, ribs, and spine are 94.13%, 96.42%, and 89.90% and MIoU is 84.61%, 90.88%, and 80.33%. This set of figures shows that ICNet results meet the requirements of production lines for the segmentation accuracy of each part of sheep body. The test results indicate that ICNet's ability to segment the sheep's neck in the image is the weaker. The reason may be that there is adhesion between the neck and the rib, and the characteristics of the adhesion region are very similar to those of the neck region, despite the adhesion region belongs to the ribs.

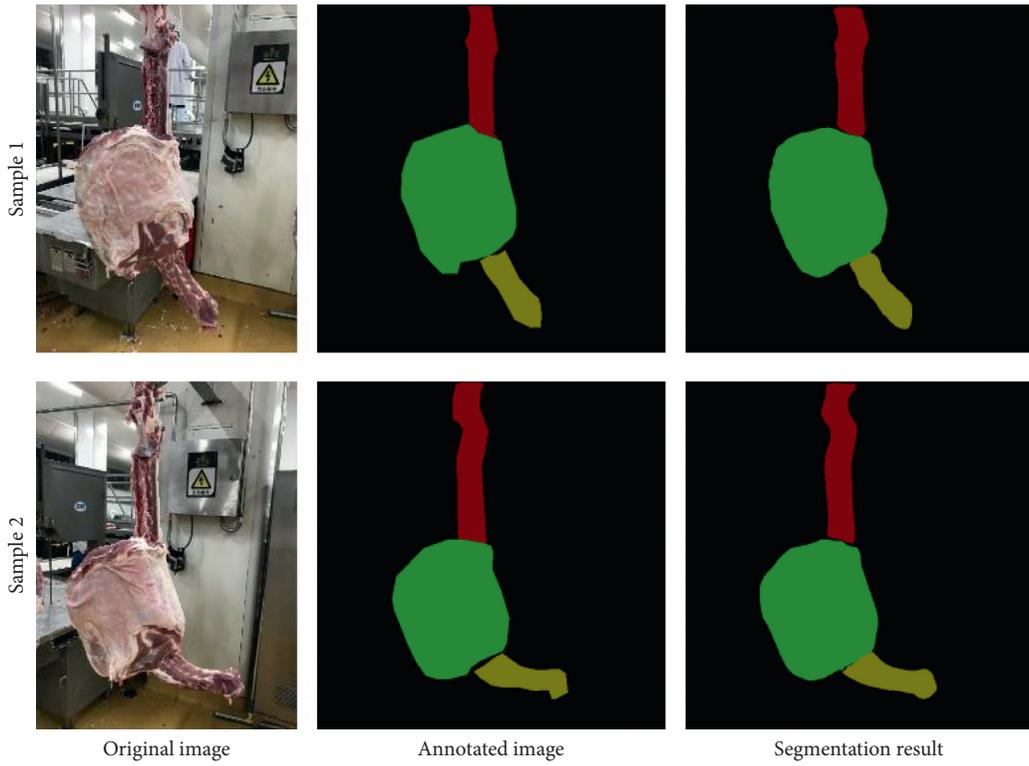


FIGURE 8: Examples of segmentation results on verifying images.

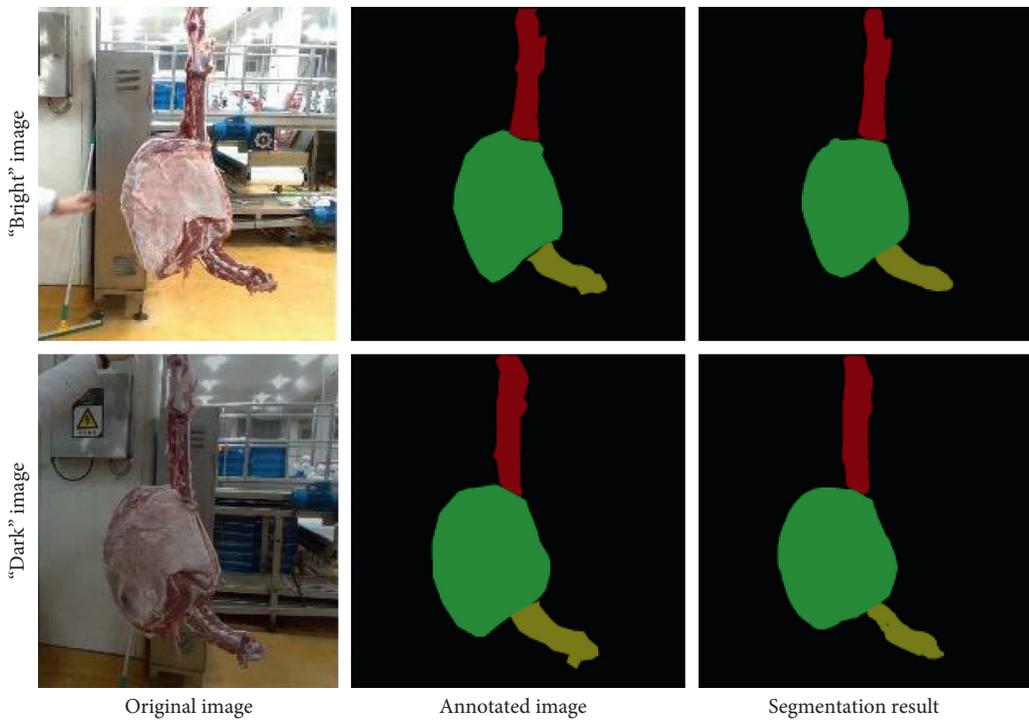


FIGURE 9: Examples of generalization test results.

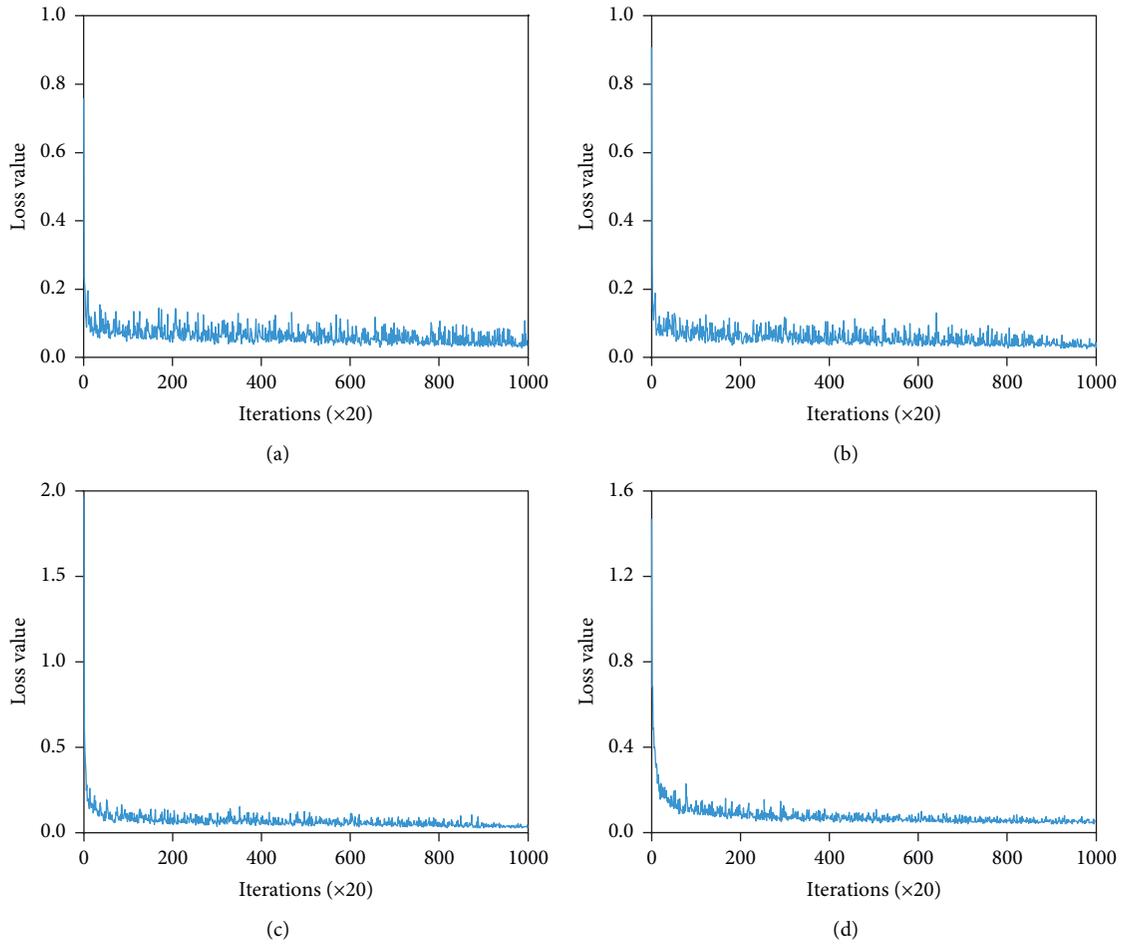


FIGURE 10: The loss value in the training process of four methods. (a) U-Net loss. (b) DeepLabv3 loss. (c) PSPNet loss. (d) Fast-SCNN.

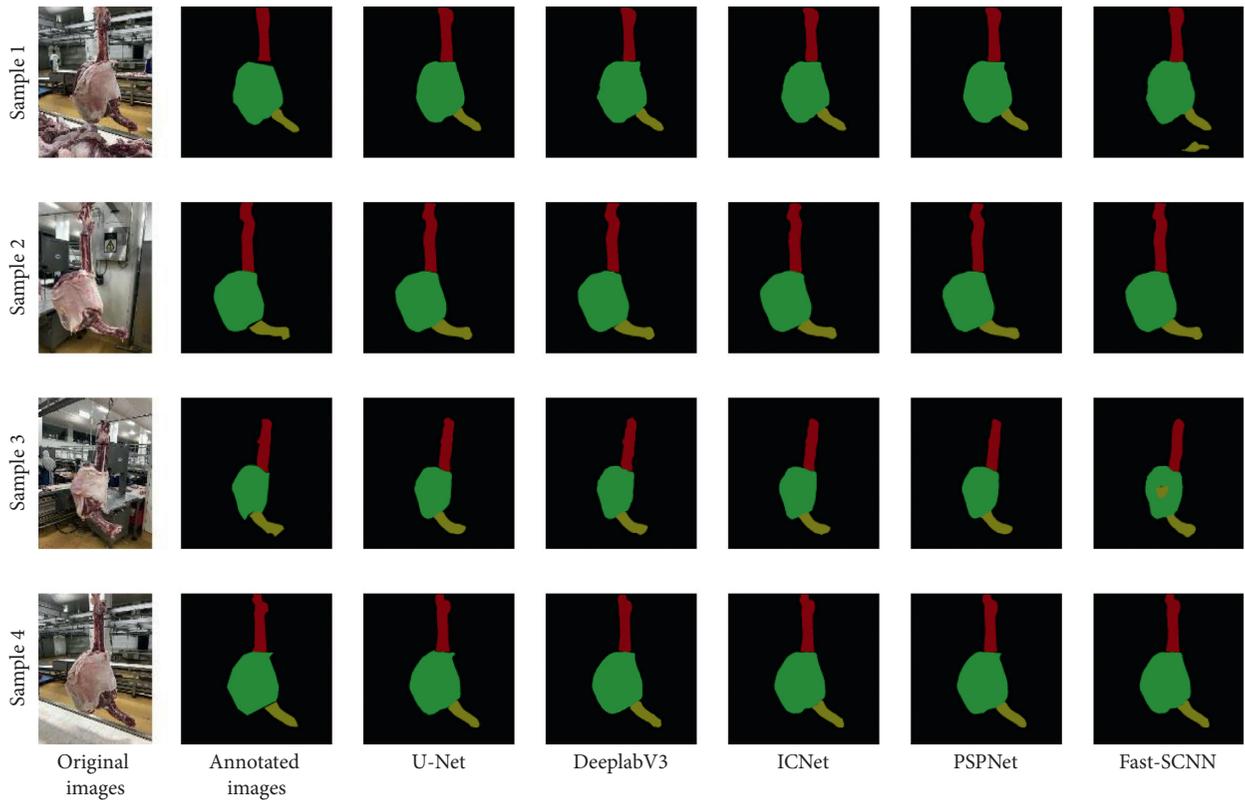


FIGURE 11: Examples of segmentation by different methods on four samples.

TABLE 4: Segmentation results of sheep carcass on the ICNet model.

Models	Accuracy (%)	MIoU (%)	Time (ms)
U-Net	97.72	88.69	269
DeepLabv3	97.69	88.50	513
PSPNet	97.47	87.22	552
Fast-SCNN	96.85	83.98	76

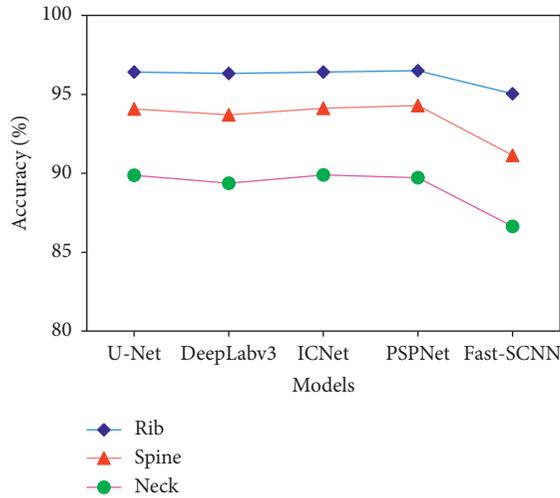


FIGURE 12: The accuracy of the five segmentation models for the three parts of sheep body.

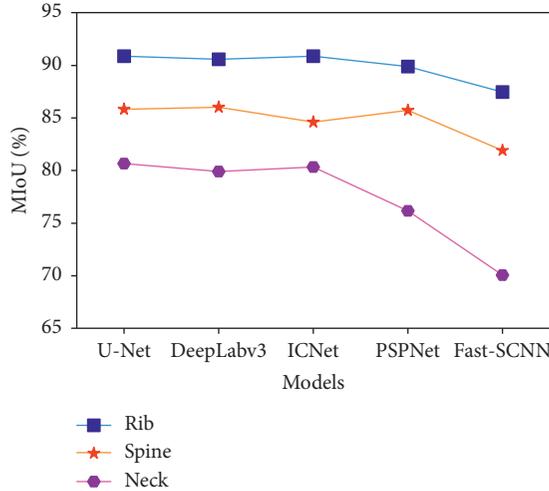


FIGURE 13: The MIoU of the five segmentation models for the three parts of sheep body.

4. Conclusion

For the spine, ribs, and neck in sheep carcass images, we have proposed a real-time semantic segmentation method of sheep carcass images based on the ICNet. The test results show that for the sheep carcass image datasets, the model has high segmentation accuracy and good real-time performance, which are suitable to overcome the noise effects caused by the complex production line background. Finally, the segmentation accuracy of the ICNet is 97.68%, MIoU

reaches 88.47%, and it takes 83 ms to process a single sheep carcass image. In addition, by different brightness image segmentation experiments verified that the ICNet has strong generalization ability for the sheep carcass image dataset. Moreover, we have taken the U-Net, DeepLabv3, PSPNet, and Fast-SCNN to conduct a semantic segmentation comparison test on the sheep carcass image datasets. The experiment has shown that the U-Net, DeepLabv3, and PSPNet can achieve accurate segmentation of sheep carcass images, whereas the Fast-SCNN shows undersegmentation and oversegmentation. In addition, the segmentation accuracy and MIoU of the U-Net and DeepLabv3 are 0.04%, 0.22%, and 0.01%, 0.03% higher than the ICNet, respectively, and the segmentation accuracy and MIoU of the ICNet are larger by 0.21%, 1.25%, and 0.83%, 4.49% compared to the PSPNet and Fast-SCNN, respectively. Finally, for a single sheep carcass image, the ICNet takes the shortest time, which is 69.14%, 83.82%, and 84.96% shorter than the U-Net, DeepLabv3, and PSPNet, respectively, which indicates that the ICNet has a significant real-time advantage. This method can provide technical reference for the development of the vision system for sheep carcass cutting robot.

In the future, we plan to optimize the method of this study with a more lightweight real-time image semantic segmentation model without decreasing the cutting accuracy.

Data Availability

The sheep carcass images data used to support the findings of this study were supplied by Huazhong Agricultural University and Chinese Academy of Agricultural Mechanization Sciences under license and cannot be made freely available. The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National “Thirteenth Five-Year” Plan for Science and Technology Support of China (2018YFD0700804) in Huazhong Agricultural University. The authors would also like to thank Prof. Shu-Cai Wang, Inner Mongolia Meiyangyang Food Co., Ltd., for their assistance in the experimental setup and image acquisition.

References

- [1] Azouz, Y. Liu, “Import and export situation: import and export situation: lamb imports continue to increase trade and the deficit has expanded significantly,” *China Animal Industry*, vol. 521, no. 2, p. 35, 2019.
- [2] National Bureau of Statistics, “National data,” 2019, <http://data.stats.gov.cn/search.htm?s=%E7%BE%8A%E8%82%89%E5%88%86%E4%BD%93>.

- [3] D. Zhang, T. Hui, and Z. Wang, "Current situation and future trend of meat processing technology in China," *Meat Research*, no. 1, pp. 1–8, 2020.
- [4] D. Cun-Zhe and Z. Rui-Ying, "Current status, issues and countermeasures of sheep slaughtering and processing industry in China," *Meat Research*, vol. 4, no. 3, pp. 31–35, 2014.
- [5] N. Guo, M. Pan, Z. Wang, and J. Jiang, "Development trend and current situation of key technology and equipment in sheep slaughtering and processing in China," *Meat Industry*, vol. 41, no. 8, pp. 27–31, 2020.
- [6] Z. Li'na, Y. Jianning, W. Pei, X. Chuazhong, H. Ding, and W. Jing, "Design and experiment of non-stress measuring system for sheep's conformation parameters," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 47, no. 11, pp. 307–315, 2016.
- [7] Z. Lin, Z. Wen, L. Qi, and L. Gang, "Measuring system of sheep body size based on embedded machine vision," *Computer Measurement & Control*, vol. 22, no. 8, pp. 2396–2398, Article ID 2408, 2014.
- [8] L. Meng, Z. Bai, M. Zheng, R. Zhu et al., "Discrimination of chilled lamb in different parts under different storage time based on mobile phone images," *Food Science*, vol. 41, no. 23, pp. 21–26, 2020.
- [9] M. R. Chandraratne, D. Kulasiri, and S. Samarasinghe, "Classification of lamb carcass using machine vision: comparison of statistical and neural network analyses," *Journal of Food Engineering*, vol. 82, no. 1, pp. 26–34, 2007.
- [10] Z. Fan, Zhu, F. Zhang, X. Yao, Y. Qiu, and Yan, "Qualitative analysis of mutton freshness by hyperspectral imaging based on BP and adaboost-BP neural network," *Xinjiang Agricultural Sciences*, vol. 55, no. 1, pp. 183–188, 2018.
- [11] K. Mohammed, G. Elmasry, D. W. Sun, and P. Allen, "Application of NIR hyperspectral imaging for discrimination of lamb muscles," *Journal of Food Engineering*, vol. 104, no. 3, pp. 332–340, 2011.
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, pp. 1097–1105, 2012.
- [13] S. Cui, Y. Zhou, Y. Wang, and L. Zhai, "Fish detection using deep learning," *Applied Computational Intelligence and Soft Computing*, vol. 2020, Article ID 3738108, 13 pages, 2020.
- [14] S. Mu, H. Qin, J. Wei et al., "Robotic 3D vision-guided system for half-sheep cutting robot," *Mathematical Problems in Engineering*, vol. 2020, Article ID 1520686, 11 pages, 2020.
- [15] F. Zeng, X. Cai, and S. S. Ge, "Low-shot wall defect detection for autonomous decoration robots using deep reinforcement learning," *Journal of Robotics*, vol. 2020, Article ID 8866406, 7 pages, 2020.
- [16] M. Long, C. Ouyang, H. Liu, and Q. Fu, "Image recognition of camellia oleifera diseases based on convolutional neural network & transfer learning," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 34, no. 18, pp. 194–201, 2018.
- [17] F. Longsheng, F. Yali, E. Tola, L. Zhihao, L. Rui, and C. Yongjie, "Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 34, no. 2, pp. 205–211, 2018.
- [18] L. Jin, Y. Fu, and Q. Dong, "The auto-question answering system based on convolution neural network," *Journal of East China Normal University: Natural Science*, vol. 2017, no. 5, pp. 66–79, 2017.
- [19] M. Shen, M. Tai, C. E. D. R. I. C. Okinda, L. Liu, Li, and Y. Sun, "Real-time detection method of newborn piglets based on deep convolution neural network," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 50, no. 8, pp. 270–279, 2019.
- [20] G. Yun, G. Jiliang, L. Xuan, L. Minggang, L. Jun, and T. Yu, "Instance-level segmentation method for group pig images based on deep learning," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 50, no. 4, pp. 179–187, 2019.
- [21] A. Yang, Y. Xue, H. Huang et al., "Lactating sow image segmentation based on fully convolutional networks," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 33, no. 23, pp. 219–225, 2017.
- [22] A. Noor, Y. Zhao, A. Koubaa et al., "Automated sheep facial expression classification using deep transfer learning," *Computers and Electronics in Agriculture*, vol. 175, Article ID 105528, 2020.
- [23] Z. Li, S. Wang, S. Zhao, and Y. Wang, "Automatic segmentation of sheep skeleton based on machine vision and machine learning," *Food & Machinery*, vol. 36, no. 6, pp. 125–132, 2020.
- [24] S. Zhao, S. Wang, Z. Li, and Y. Chen, "Segmentation of sheep rib image based on U-type convolution neural network," *Food & Machinery*, vol. 36, no. 9, pp. 116–121, 2020.
- [25] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "Icnet for real-time semantic segmentation on high-resolution images," 2017, <http://arxiv.org/abs/1704.08545>.
- [26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Lecture Notes in Computer Science, Munich, Germany, October 2015.
- [27] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July 2017.
- [28] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July 2017.
- [29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 91–99, Montreal, Canada, December 2015.
- [30] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [31] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2014.
- [32] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: a deep neural network architecture for real-time semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 2016.