

Research Article

Robotic 3D Vision-Guided System for Half-Sheep Cutting Robot

Shuai Mu,^{1,2} Haibo Qin,^{1,2} Jia Wei,^{1,2} Qingkang Wen,^{1,2} Sihan Liu,^{1,2} Shucui Wang,^{1,2}
and Shengyong Xu ^{1,2}

¹College of Engineering, Huazhong Agricultural University, Wuhan 430070, China

²Key Laboratory of Agricultural Equipment in the Middle and Lower Reaches of Yanze River, Ministry of Agriculture, Wuhan 430070, China

Correspondence should be addressed to Shengyong Xu; xsy@mail.hzau.edu.cn

Received 30 July 2020; Revised 2 September 2020; Accepted 16 September 2020; Published 5 October 2020

Academic Editor: Jun-Jun Jiang

Copyright © 2020 Shuai Mu 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.

Sheep body segmentation robot can improve production hygiene, product quality, and cutting accuracy, which is a huge change for traditional manual segmentation. With reference to the New Zealand sheep body segmentation specification, a vision system for Cartesian coordinate robot cutting half-sheep was developed and tested. The workflow of the vision system was designed and the image acquisition device with an Azure Kinect sensor was developed. Furthermore, a LabVIEW software with the image processing algorithm was then integrated with the RGBD image acquisition device in order to construct an automatic vision system. Based on Deeplab v3+ networks, an image processing system for locating ribs and spine was employed. Taking advantage of the location characteristics of ribs and spine in the split half-sheep, a calculation method of cutting line based on the key points is designed to determine five cutting curves. The seven key points are located by convex points of ribs and spine and the root of hind leg. Using the conversion relation between depth image and the space coordinates, the 3D coordinates of the curves were computed. Finally, the kinematics equation of the rectangular coordinate robot arm is established, and the 3D coordinates of the curves are converted into the corresponding motion parameters of the robot arm. The experimental results indicated that the automatic vision system had a success rate of 98.4% in the cutting curves location, 4.2 s time consumption per half-sheep, and approximately 1.3 mm location error. The positioning accuracy and speed of the vision system can meet the requirements of the sheep cutting production line. The vision system shows that there is potential to automate even the most challenging processing operations currently carried out manually by human operators.

1. Introduction

Lamb carcass segmentation is the basis of carcass deep processing products. Good carcass grading standards and assessment technology can not only establish a good communication platform between consumers and manufacturers but also help improve carcass quality and enhance market competitiveness [1]. However, the traditional manual process has high requirements on the cutting experience of the workers. Sheep carcass segmentation is a labor-intensive technique, with a risk of infection and consequent costs. These aspects have enabled the development of automated machines to increase productivity and rooting success rate while reducing costs. Meat products are highly diverse and the production environment is complex,

making the automation of the meat production process a big challenge. In a few countries such as Australia, intelligent processing equipment for sheep bodies is used [2]. Not only can the fast and effective lamb carcass segmentation technology improve the quality of lamb and affect its taste, but also it can help to combine the processing procedure with intelligent equipment to improve production efficiency and bring more significant economic benefits [3].

The vision system is an indispensable functional module in the intelligent processing of livestock and poultry. It is used to identify and locate operation targets and provide operation parameters for processing equipment. The current vision system for livestock and poultry processing equipment can be divided into machine vision and computer vision. (1) Machine vision is a technology widely studied and

applied in intelligent equipment. Bonder applied machine vision to the sardine slaughter production line, reducing the number of operators from 4 to 1 [4]. Liu et al. used machine vision methods to extract pig belly contour curves, combined machine vision with spline fitting, and improved genetic algorithm to obtain segmentation trajectories [5]. Azarmde et al. developed an image processing algorithm to extract the appropriate head and abdomen cuts based on the size of the trout. The algorithm can detect the direction and position of the pectoral fins, anus, pelvis, and caudal fin of the fish [6]. Wang et al. used the largest inscribed circle method to locate the poultry muzzle to obtain the center coordinates of the muzzle and designed a poultry slaughter evisceration system based on machine vision positioning [7]. Feng et al. designed an online beef image acquisition platform to segment the image in the red chromaticity space and then use the area labeling method and the small area elimination method to extract the eye muscle area [8]. (2) Compared with machine vision, technologies such as three-dimensional reconstruction and deep learning models in computer vision can perform smarter and higher-precision processing. For example, Bondo et al. used 3D machine vision in the salmon slaughter production line to achieve automatic salmon slaughter, which reduced the demand for manual labor of the salmon slaughter plant [4]. Dasiewicz et al. used computer image analysis to analyze the linear and area measurement data of sheep carcass to estimate the weight of sheep cuts [9].

Compared with the automatic cutting assembly line, the intelligent robot has better flexibility and operation accuracy and has more research prospects. For example, Singh et al. developed an industrial manipulator based on 6 degrees of freedom based on sensors and PLC technology, using laser ranging methods to detect the distance between the carcass of the sheep and the end of the robot and realized the segmentation of the hanging cattle and sheep [10]. These automated systems have not done research on trajectory planning and can only segment them according to fixed trajectories. In order to adapt to the diversity of livestock breeds and body types, the information feedback of sensors must be used to ensure the quality and accuracy of robot operations. Kinect sensor is a classic RGB-D camera, which can simultaneously obtain the color, infrared, and depth information of the target. It has the characteristics and advantages of low cost and small size, making the sensor popular and applied in the field of modern agricultural information [11]. Misimi and others presented the GRIBBOT, a novel 3D vision-guided robotic concept for front half chicken harvesting. A computer vision algorithm was developed to process images from a Kinect v2 and locate the grasping point in 3D for harvesting operation. GRIBBOT shows that there is potential to automate even the most challenging processing operations currently carried out manually by human operators [12]. Image segmentation is one of the most important links in the visual system. Image segmentation technology based on convolutional neural networks has gradually emerged in the agricultural field [13]. For example, Deng et al. proposed a semantic segmentation optimization method based on RGB-D beef cattle image full

convolutional network. According to the mapping relationship between depth images and color images on pixel content, the semantic segmentation accuracy of beef cattle images under complex backgrounds is improved [14]. Multi-view or single-view 3D reconstruction technology has also been widely studied because of its low cost and fast calculation speed [15]. Cong et al. used binocular cameras to take images of pig carcasses, acquired depth information of the images, determined the movement trajectory by identifying the center line and path points, and guided the robot to complete the pig belly cutting operation [16]. Deep neural networks bring more and better opportunities to robot vision systems [17]. Yang et al. used VGG16 as the basic network and adopted an 8-fold jump structure on the feature map to achieve accurate and fast segmentation of lactating sows in the pig house scene [18].

The key techniques in vision-based control include vision information acquisition strategies, target recognition algorithms, and eye-hand coordination methods. Combined with the actual working conditions of a self-made robot system, this research focuses mainly on recognizing the key part and calculating the spatial coordination of cutting curves for the end-effector. The method for acquiring spatial information from half-sheep is presented using an Azure Kinect sensor. This method will determine the 3D coordinates of the cutting curves on the half-sheep body for the robot. This way, the robot can smoothly separate a half-sheep to several parts. Research on robot vision system is of great significance to the automation and intelligence of sheep cutting.

The detailed objectives of the current study can be summarized as follows: Section 2.1 describes the experimental materials and test platform of image processing algorithm used in the research. Section 2.2 introduces the hardware structure and workflow of the vision module in the self-made robot. Section 2.3 introduces the software of vision module system developed on LabVIEW platform. Sections 2.4 shows the algorithms of image recognition. Section 2.5 shows the location method of cutting curves. Section 2.6 introduces the method of transforming the space coordinates of the curves acquired by the vision system into the motion parameters of the three-axis Cartesian robot. Section 3 shows experiments and discussion. The last section summarizes the merits and demerits of the vision system.

2. Materials and Methods

2.1. Testing Materials and Equipment. The experimental images in this article are divided into two parts, color images and RGBD images. Color images taken with smartphones are used for the development and testing of image processing algorithms. These images are taken at the slaughterhouse of XIYANGYANG Food Co., Ltd. in Bayannaoer, Inner Mongolia, in September 2018. A total of 8,400 color images of the half-splitting part of BAMEI sheep, LAKE sheep, and DORPER sheep were taken under natural light conditions. In order to expand the diversity of images, the phone models, image resolution, and shooting angles are not uniform. RGBD images taken by an Azure Kinect are used

for testing of vision systems. Azure Kinect is the latest RGBD sensor launched by the Microsoft company in 2019. Compared with the previous generation Kinect v2, the performance of the Azure Kinect was greatly improved in the resolution and the noise. In order to eliminate the system noise and fluctuations of the Kinect sensor, 10 frames of depth images were taken continuously and then averaged. With the use of the image acquisition platform of the self-made robot, 10 sheep of each of the three species were photographed, and a total of 60 RGBD images of half-sheep were obtained. The camera environment has an LED light source, which simultaneously acquires a color image with a resolution of 4 K and a depth image of 1 million pixels. The shooting distance is between 1.0 m and 1.2 m.

The vision system is developed and tested with a general-purpose computer. The configuration is Intel Core i5-6500/32G/1TB/GTX1070-8G. The operating system is Windows 10 Professional. The required software is Python3.8, OpenCV 4.0, MATLAB 2018, and Anaconda 3-5.2.0. The deep learning framework is Caffe using PyCaffe as the interface, and using cuDNNv7.5+CUDA10.1 for GPU acceleration.

2.2. Vision System Architecture and Work Procedures.

Figure 1 is a structural diagram of a self-made Cartesian robot. The half-splitting sheep body is fixed on the cutting table by the sheep body fixing device. Only one RGBD image of half-sheep on the initial position was taken and processed in one cutting operation. The right-angle mechanical arm drives the end cutting saw to perform cutting operations. After the cutting is completed, the electric push rod pushes the cut pieces of sheep from the worktable. The vision system is mainly composed of a Kinect sensor, LED light source, a computer, a control circuit, and so on. After the sheep body is manually fixed on the cutting table, the vision system is started from the man-machine interface software, the system will automatically turn on the Kinect camera to take RGBD images and automatically call the image processing algorithm to calculate the motion control parameters required by the control circuit. The end cutting saw starts to cut the sheep body according to the motion parameters until the task is completed.

The algorithm flow chart is shown in Figure 1, which can be divided into three parts. (1) Semantic segmentation: color images of half-sheep body were used to make a training set for FCN and Deeplab v3+ models. The trained model was adopted to automatically recognize ribs and spine. (2) Cutting curves calculation: using segmented ribs and spine images, the contours and their minimum bounding rectangle were calculated. Based on some rules, 7 cutting key points were detected. Then, 5 dividing lines were calculated. (3) Coordinate conversion: the pixel in the color image can be mapped to the pixel coordinates in the depth map and space coordinates of the five cutting curves can be converted by Kinect image rules. With the hand-eye calibration procedures and the motion equation, the space coordinates will be converted to the robot motion parameters.

2.3. *LabVIEW Software of the Vision System.* A kind of human-computer interaction software was developed with LabVIEW2018, as shown in Figure 2. The software provides two modes of debugging and working. The debugging mode is convenient for the development of the vision system. The working mode is used for actual production. The man-machine software mainly integrates the image acquisition and processing module, the Python script node, and the MCU communication module. (1) With the use of the MakerHub toolkit to develop an image acquisition function module in LabVIEW, the language is used to connect and control the Kinect sensor to capture color images and depth images and store them in a designated location on the PC according to the shooting date. (2) The Python script node program can call the Python image processing code to analyze the sheep carcass image and return the result. The result is a character string merged by the space coordinates of the cutting curves, and the microcontroller accepts this character string as the motion control parameter of the robotic arm. (3) The MCU communication module uses VISA driver to realize the communication between LabVIEW and the MCU. The robot control circuit parses the received character string, controls the movement of the cutting saw, and realizes the automatic cutting of the sheep body.

2.4. *Image Recognition of Ribs and Spine.* Among the key parts of the sheep carcass, the ribs and spine are the most prominent. Ribs have fan-shaped distribution, red and white texture characteristics. Spine has a bamboo-like texture structure. Regardless of the sheep's breed, size, fatness, and light change, the special texture and structure make the rib and spine segmentation the most feasible. Research on targeted image segmentation algorithms to effectively segment ribs and spine regions will lay the foundation for the calculation of segmentation lines. The image segmentation of ribs and spine includes four steps. Firstly, the experimental images were expanded using data augmentation such as rotation, zoom, Gaussian blur, etc. Secondly, the color image is preprocessed to make a training network model data set. Thirdly, the full convolutional neural network (FCN) [19] and Deeplab v3+ [20] are used to train the image segmentation model. Finally, the best model was selected to achieve automatic image segmentation of ribs and spine.

2.4.1. *Image Preprocessing.* Image preprocessing includes three aspects: data enhancement, normalization processing, and data set creation:

- (1) Data enhancement: in order to prevent overfitting, the data set is expanded and the image data enhancement operation is performed. The original 8400 images are translated, rotated, and zoomed, respectively, to obtain more images. 75% of the images are classified as the training set and 25% of the images as the test set.
- (2) Normalization processing: the resolution of the experimental image is inconsistent, the file name is not

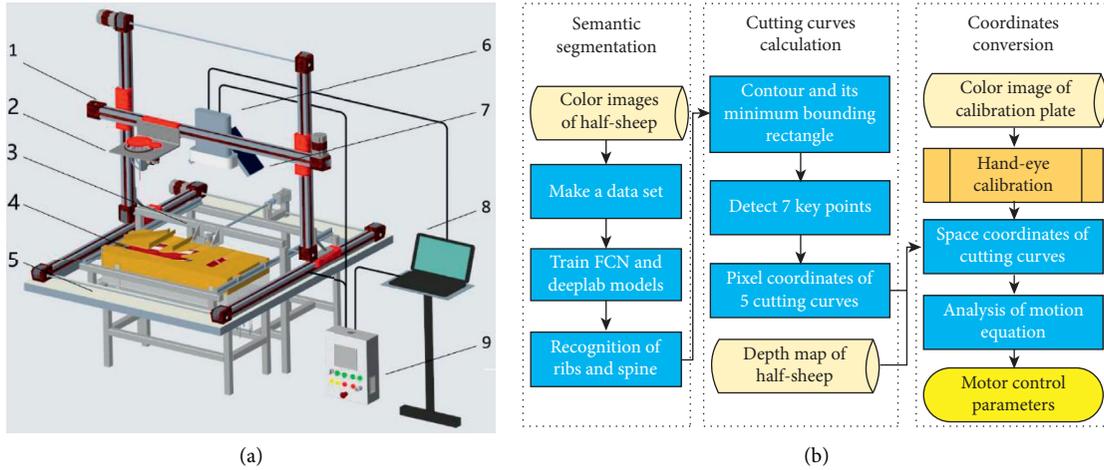


FIGURE 1: Half-sheep automatic cutting: (a) workbench structure and (b) procedures. (1) Cartesian coordinate robot arm; (2) end cutting saw; (3) electric push rod; (4) sheep body fixing device; (5) cutting table; (6) Azure Kinect; (7) light source; (8) computer; (9) electric control box.

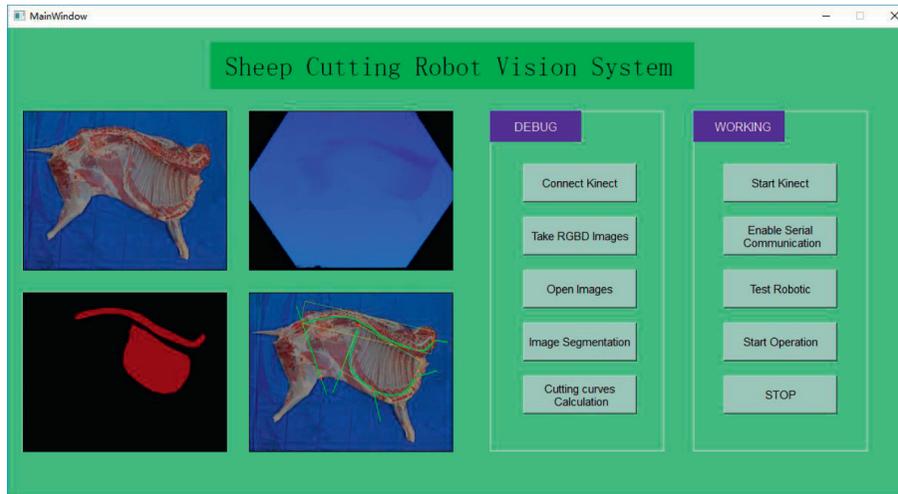


FIGURE 2: LabVIEW man-machine interface software.

uniform, and it cannot be directly input into the network. Therefore, the original images collected are renamed in the numerical order, and each image is cropped to a size of 256×256 (pixels).

- (3) Making a data set: MATLAB image labeler toolbox is used to manually label each sheep ketone body image, and the sheep carcass ribs and spine are used as separate connected domains to obtain a corresponding .json file. Then, the batch processing algorithm is used to convert the .json file into a label image containing label .png to complete the data set production.

2.4.2. Train FCN Model. Migration learning can speed up the training, greatly shorten the time required for traditional training, and make the results converge quickly. The FCN structure diagram and segmentation results are shown in Figure 3. The specific details are as follows.

(1) FCN Network. Based on the classic network VGG-Net, the FCN segmentation network changes the last two fully connected layers $fc6$ and $fc7$ to convolutional layers. First, $fc6$ is replaced with a convolutional layer with a convolution kernel size of $4096 \times 512 \times 7 \times 7$, and then $fc7$ is replaced with a corresponding convolution kernel with a size of $4096 \times 4096 \times 1 \times 1$. Therefore, the FCN network can be divided into two parts. The first part is to retain the Encoder layer of the first 5 convolution modules of the original VGG-Net, and the other part is to replace the Decoder layer after the fully connected layer by the convolution layer. Each of the first five convolution modules of the Encoder layer contains a pooling layer. The size of the feature map obtained after each layer of pooling becomes $1/2$ of the size of the input feature map. After five times of pooling, a size of input 1 is obtained, $1/32$ the size of the feature map, and finally the Decoder layer deconvolves the resulting feature map and restores it to the size of the original image, thereby achieving pixel-level prediction and classification.

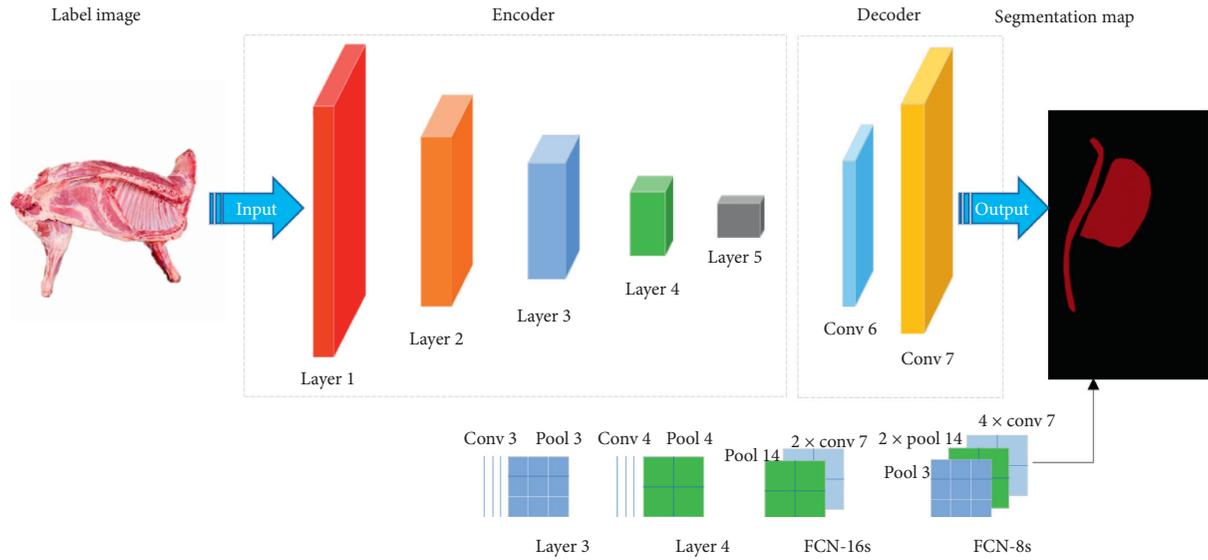


FIGURE 3: FCN image segmentation model.

(2) *Jump Structure*. By directly replacing the fully connected layer in the CNN network, the FCN-32s is obtained. Since the FCN-32s network model is obtained by directly upsampling the convolutional features of the last layer 32 times, it uses fewer convolutional features and is difficult to capture images, so the segmentation error is large. In order to solve this problem, on the basis of FCN-32s, jump structure optimization was added to obtain FCN-16s and FCN-8s network models. FCN-16s is based on the deconvolution of the last layer of conv7, fused with the features of the pool4 layer, and then restored to the original image size after 16 times of upsampling. FCN-8s is based on FCN-16s, fused with the features of pool3 layer, after 8 times of upsampling, the feature map is enlarged by 8 times, and finally the crop layer is cut to restore the same size as the original image. After being tested, two segmentation models were trained using FCN-VGG16 network, ribs segmentation model Ribs-fcn16s, and spine segmentation model Spine-fcn8s.

(3) *Transfer Learning*. Since training a deep neural network requires a large amount of data and high-performance hardware, it is difficult to implement the actual process. In order to solve this problem, this paper adopts the transfer learning method. By using the pretrained model on the VOC2012 data set as the initialization parameter of the Encoder layer, the training speed can be accelerated. This article resets the main parameters of network training: learning rate is 0.01, epoch is 100, max_iter = 50000, and momentum = 0.99.

2.4.3. *Train Deeplab v3+ Model*. Due to the many advantages of FCN, many semantic segmentation models adopt or learn from this model. The use of convolutional neural networks for semantic segmentation has another problem in addition to the fully connected layer; that is, the pooling layer loses its location information due to the aggregation of context and the expansion of the receptive field. In 2015,

Google open-sourced the first-generation Deeplab model. After that, after continuous optimization and improvement of convolutional neural network target scale modeling technology, contextual information processing, feature extractor, and model training process, the Deeplab model was upgraded to Deeplab-v3. Google upgraded to Deeplab-v3+ by adding an effective and simple decoder module to improve the segmentation effect and further applied deep separable convolution to the spatial porous pyramid pooling (ASPP) and decoder modules and obtained a more powerful and faster semantic segmentation model.

In order to solve the problems of low accuracy and inaccurate edges in image segmentation, a Deeplab-v3+ model of two types of subdivision labels of ribs and spine is established. A Deeplab v3+ network was created from the pretrained ResNet-18 network. Use GPU for training and use CUDA environment to accelerate. Figure 4 shows the model structure. The Deeplab-v3+ model applies the ASPP module to the encoder-decoder network to perform semantic segmentation tasks. First, during encoding, the original image features are extracted by using the ResNet network, and then through the ASPP module, the input features of the original image are extracted through convolution with filters on multiple ratios and multiple effective fields of view, and the pooling operation is performed to encode multi-scale dense features. In the decoder stage, the low-level features and the ASPP features are concatenated and then convolved, and finally upsampling is performed to gradually restore the spatial information to the more refined target boundary of the target image.

2.5. *Automatic Method of Cutting Curves in Half-Sheep Images*

2.5.1. *Scheme of the Cutting Curves Calculation*. According to the New Zealand sheep carcass grading standard, the sheep body will be divided into 7 parts: waist-hip, belly, ribs, spine, hind nates/leg, neck, and

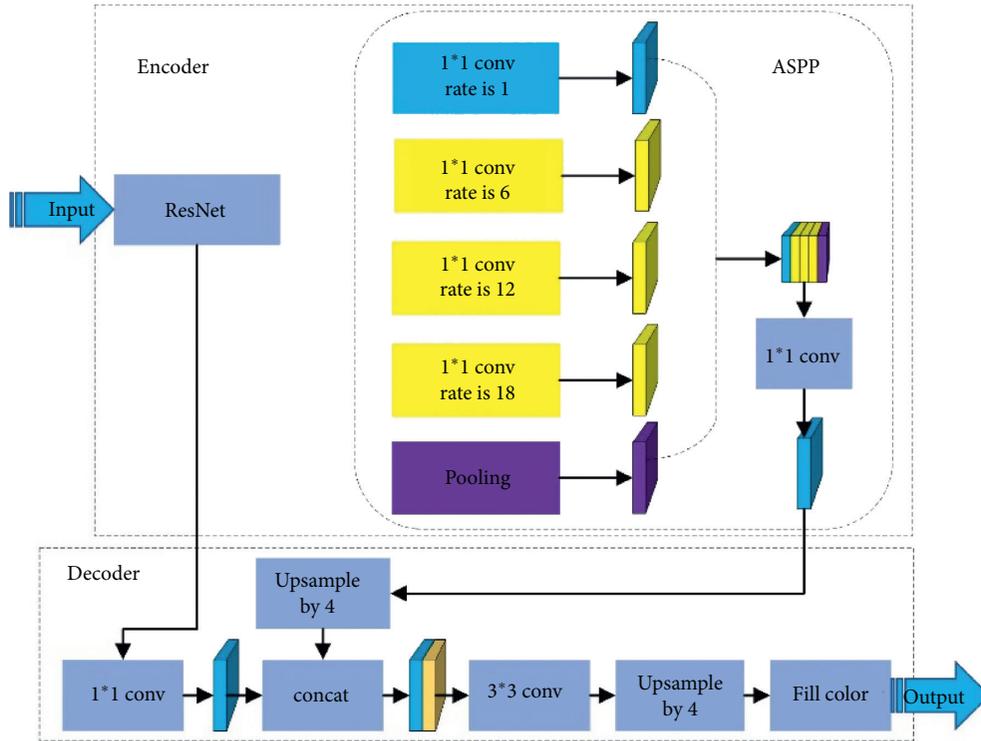


FIGURE 4: Deeplab v3+ semantic segmentation model.

foreleg [21]. The ribs and spine of sheep occupy a significant and important position in the sheep. The upper edge of the ribs is adjacent to the waist and abdomen, the lower edge is adjacent to the spine, the left edge is the cutting curves between the buttocks/loin and the waist and abdomen, and the right edge and the left edge of the spine together determine the location of the neck of the lamb. Therefore, the recognition of these two key areas can prompt other key parts to be segmented one after another. According to the relative spatial relationship between each part of the sheep body, ribs, and spine, 5 cutting curves $\Gamma_1 \sim \Gamma_5$ are designed, as shown in Figure 5. Sheep body segmentation is carried out in the order of the number of the segmentation line, and the hind nates/leg, spine, waist-hip, belly, foreleg, and ribs are segmented in sequence.

2.5.2. Detection of the Cutting Key Points. In order to calculate the cutting curves, reference points are set in the image of the split half-sheep body, which are called the cutting key points. As shown in Figure 6, 7 key points are set in the half-sheep body image, which are marked with $P_{k1}(u_1, v_1) \sim P_{k7}(u_7, v_7)$, respectively. $P_{k1} \sim P_{k3}$ are the three intersection points of the rib profile Ω_R and its smallest bounding rectangle. $P_{k4} \sim P_{k6}$ are the three intersection points between the outline Ω_S of the spine and its smallest bounding rectangle. P_{k7} is the junction between the hind leg and the waist.

Taking the calculation of the key points corresponding to the ribs as an example, firstly, the ribs and spine regions obtained by the aforementioned deep learning model segmentation are binarized using the OTSU algorithm, and

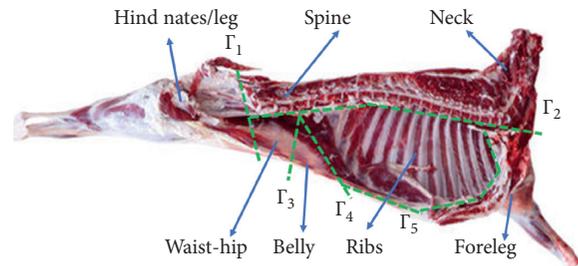


FIGURE 5: Half-sheep division standard and the cutting curves.

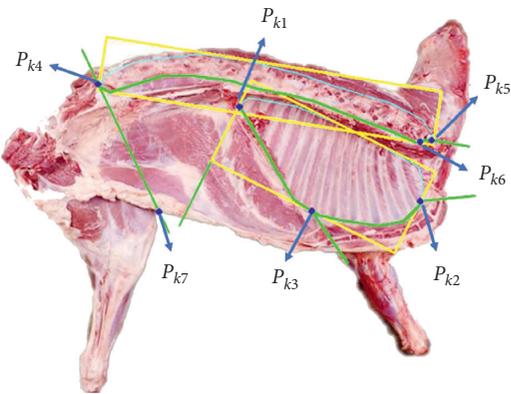


FIGURE 6: Automatic detection result of the key point and cutting curves.

then the outline of the binary image and its minimum bounding rectangle are calculated.

In the detection of $P_{k1} \sim P_{k6}$, assume that $P_{R0}(u_0, v_0)$ is the pixel coordinate of the centroid. Both the angle between the line formed by P_{k1} and P_{k2} and the U axis of the pixel coordinate system are calculated. The calculation formula is as follows:

$$\alpha = \arctan \frac{u_2 - u_1}{v_2 - v_1}, \quad (1)$$

with P_{R0} as the center; the rib area is rotated clockwise by an angle. Extreme points in the rotated rib profile are searched for. The minimum point and maximum point of the U axis and the maximum point of the V axis, respectively, correspond to $P_{k1} \sim P_{k3}$. The centroid P_{R0} is taken as the center and the extreme point coordinates are rotated counter-clockwise by α degree to get the key point coordinates. The calculation method of $P_{k4} \sim P_{k6}$ is similar to $P_{k1} \sim P_{k3}$.

The key point P_{k7} is located in the depression between the hind leg tendon and waist. This feature is taken advantage of to design the calculation method of the key point: First, the overall outline of the sheep is calculated. The second step is to calculate the width of the 1/4~1/2 part of the left side of the contour. The third step is to search for an interval with continuously decreasing width, and the contour point corresponding to the end of the interval is P_{k7} .

2.5.3. Calculating the Image Cutting Curves Equation. In view of these 7 key points, ribs and spine contours, the calculation rules of 5 cutting curves are designed, which are introduced as follows. (u_i, v_i) in the following formula represents the pixel coordinates of the key point. The detection result is as shown in Figure 6:

- (1) Cutting curve Γ_1 : it consists of two parts. Two key points P_{k4} and P_{k7} are connected to form a straight line, and the straight line equation is as in formula (2). The straight line from P_{k7} to P_{k4} is the first part of Γ_1 . The upward extension of this straight line has an intersection with the outline of the sheep body. The line segment formed by P_{k4} and the intersection is the second part of Γ_1 :

$$\frac{(y - v_7)}{(x - u_7)} = \frac{(y - v_4)}{(x - u_4)}. \quad (2)$$

- (2) Cutting curve Γ_2 : the lower contour of the spine region constitutes the first part. The connecting line of two key points P_{k4} and P_{k6} forms a straight line, and the form of the straight line equation is similar to formula (2). The extension of this straight line to the right has an intersection with the outline of the sheep body. The straight line segment formed by P_{k6} and the intersection is the second part of Γ_2 .

- (3) Cutting curve Γ_3 : it is a straight line. The slope of the left side of the rib rectangle Ω_R and the key point P_{k1} constitute the straight line equation of the cutting curves Γ_3 , as shown in

$$y - v_1 = \tan \beta (x - u_1). \quad (3)$$

Here, β is the tilt angle parameter obtained by fitting the smallest bounding rectangle of Ω_R .

- (4) Cutting curve Γ_4 : it consists of two parts. The section of rib profile curve from P_{k1} to P_{k4} constitutes the first part. P_{k1} and P_{k3} form a straight line, and the straight line equation is similar to formula (2). The extension of this straight line has an intersection with the outline of the sheep body. The straight line segment from P_{k3} to the intersection constitutes the second part of the cutting curves Γ_4 .
- (5) Cutting curve Γ_5 : it is composed of two parts. The section of rib profile curve from P_{k4} to P_{k2} constitutes the first part. A section of the rib profile curve between P_{k2} and P_{k3} constitutes the second part.

2.6. From Curves Image to Robot Motion Parameters

2.6.1. Hand-to-Eye Calibration. A visual system model of the robot was established through hand-to-eye model. It meant that the camera was installed in a fixed position and separated from the manipulator. The mathematical relation between the vision system and the robot system needed to be established, which meant that the hand-eye calibration was required. The hand-eye calibration schematic is shown in Figure 7. Let the arm base coordinate system be OBASE-xyz and its origin is located at a point on the planet. Let the arm end coordinate system be OTOOL-xyz, the camera coordinate system be OCAM-xyz, and the calibration plate coordinate system be OPLATE-xyz.

H was a 4×4 homogeneous transformation matrix of two coordinate systems, including a 3×3 rotation matrix R and a 3×1 translation matrix. There are 4 matrices in the transformation. Equation (4) shows the conversion relation of each matrix:

$$H_{cam}^{base} = H_{tool}^{base} \cdot H_{plate}^{tool} \cdot H_{cam}^{plate}, \quad (4)$$

where H_{cam}^{plate} represents the transformation matrix of pixels in the plate image relative to its space coordinates. H_{tool}^{base} represents the homogeneous transformation matrix of OTOOL-xyz relative to OBASE-xyz, which could be obtained through robot kinematics. H_{plate}^{tool} represents the homogeneous transformation matrix of OPLATE-xyz relative to OTOOL-xyz. H_{cam}^{base} represents the homogeneous transformation matrix of OCAM-xyz relative to OBASE-xyz, namely, the matrix which needs to be calibrated.

In this paper, the camera did not move with the robot arm because of the hand-to-eye model, so H_{cam}^{base} was fixed. The matrix conversion relation at two different positions of the simultaneous calibration plate yielded

$$H_{tool}^{base} \cdot H_{plate}^{tool} \cdot H_{cam}^{plate} = H_{tool}^{base} \cdot H_{plate}^{tool} \cdot H_{cam}^{plate}. \quad (5)$$

After transformation, equation (6) was obtained as follows:

$$\left[H_{tool}^{base} \right]^{-1} \cdot H_{plate}^{base} \cdot H_{plate}^{tool} = H_{plate}^{tool} \cdot \left(H_{cam}^{plate} \cdot \left[H_{tool}^{base} \right]^{-1} \right). \quad (6)$$

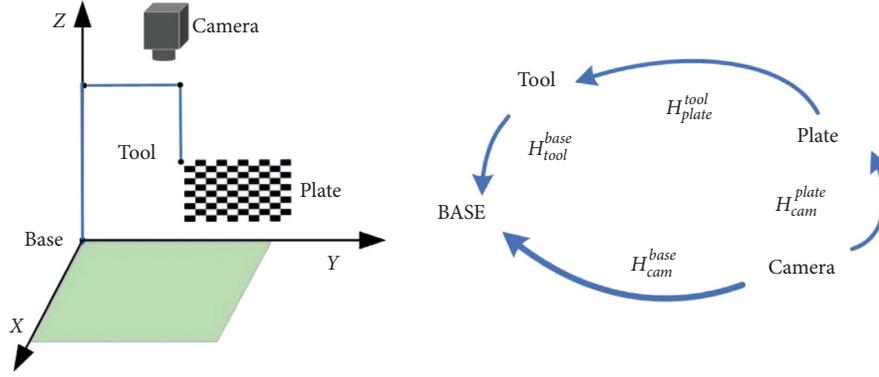


FIGURE 7: Hand-eye relationship diagram.

Therefore, H_{plate}^{tool} could be obtained by solving equation (6). And H_{cam}^{base} could be calculated through equation (4). H_{cam}^{plate} represents the transformation matrix of pixels in the plate image relative to its space coordinates. Kinect can obtain both color and depth images at the same time. There is a fixed mapping relationship between color information and depth information. Microsoft provides the Kinect SDK mapping function, which can be conveniently used to map the pixels in the color image to the pixels in the depth map. And then, according to the imaging principle of Kinect, the pixels in the depth map can be converted into space coordinates [22].

2.6.2. From Space Coordinates to Robot Motion Parameter. Each axis is driven by an asynchronous motor, and the rotation of the motor drives the slider on the axis to move linearly. The movement of the slider on the cartesian coordinate robot system can be regarded as a mass point. Since each axis is a linear axis, that is, it can only do translational motion but not rotary motion, the Cartesian coordinate robot motion parameter model can be equivalent to

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{d_x}{2\pi} & 0 & 0 \\ 0 & \frac{d_y}{2\pi} & 0 \\ 0 & 0 & \frac{d_z}{2\pi} \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_z \end{bmatrix}. \quad (7)$$

Here, x, y, z is the moving distance of each axis, and $[q_x, q_y, q_z]^T$ is the angle vector of the robot joint.

3. Experimental Results and Analysis

3.1. Image Segmentation Performance Test. In order to evaluate the performance of the image processing algorithm in this paper, an experiment was designed to test it. From the 8400 color images of sheep introduced in Section 2.1, 20% of the images are test images for the models development. For simulating a real environment in sheep cutting, 50 color images of each of the three breeds were selected to test the present algorithm. These images are shoot from an approximate vertical angle. The ribs and spine regions in a test

image are manually annotated as the reference image. This image is automatically segmented using the two algorithms as the resulting image. IoU (intersection over union) is used to evaluate the overall segmentation effect.

Firstly, the segmentation effect of FCN-8s, FCN-16s, FCN-32s, and Deeplab-v3+ on ribs and spine is evaluated. Among them, FCN-8s and FCN-16s are obtained by combining the pool3 and pool4 layers on the basis of the FCN-32s network model. Table 1 shows the histogram of segmentation error. For ribs, the IoU of the FCN-16S network is the largest in three FCN models. Deeplab-v3+ model is better than FCN-16s. For the spine, the results of Deeplab-v3+ are better than the three FCN models. So, Deeplab-v3+ is more suitable for spine image segmentation. Figure 8 shows a set of pictures showing the difference in spine segmentation. For the spine image, the FCN-32s segmentation result is poorer, with some broken parts.

The test images are divided into three categories according to sheep breeds (BAME MUTTON sheep, HU sheep, and DORPER sheep), each with 50 images, which are used to evaluate the adaptability of the algorithm to sheep breeds. FCN-16s network is used to segment ribs and FCN-8s network is used to segment spine. Deeplab-v3+ is used to segment both ribs and spine. Table 2 shows the average time of manual segmentation and algorithm segmentation, the total area of the segmentation result, and the IoU. The experimental results show that the rate of algorithm's automatic segmentation is 4 times higher than that of manual segmentation. If the hardware configuration of the computer is upgraded, the gap will be even greater. For the segmentation of ribs, FCN-16s has the best effect on BAME sheep, with an IoU as high as 98.3% and smallest IoU of 93.1% for DOBER sheep. For the segmentation of the spine, the IoU of FCN-8s for the three types of sheep is relatively close, ranging from 93.1% to 95.1%. Deeplab-v3+ is still better than FCN model. The time consumption of those two methods are approximately same. On the whole, the segmentation accuracy of the spine is greater than that of the ribs. There are two main reasons. One reason is that the image features of the ribs are more obvious than that of the spine. The other reason is that the ribs will not be damaged when the sheep's body is split in half, but the spine is often split not smoothly, causing damage of image texture, leading to greater segmentation errors.

TABLE 1: Comparison of image segmentation of the four models.

Network type	Target	Number of images	IoU (%)
FCN-32s	Ribs	50	92.2
	Spine	50	88.7
FCN-16s	Ribs	50	95.5
	Spine	50	92.3
FCN-8s	Ribs	50	94.8
	Spine	50	94.2
Deeplab-v3+	Ribs	50	96.3
	Spine	50	95.3

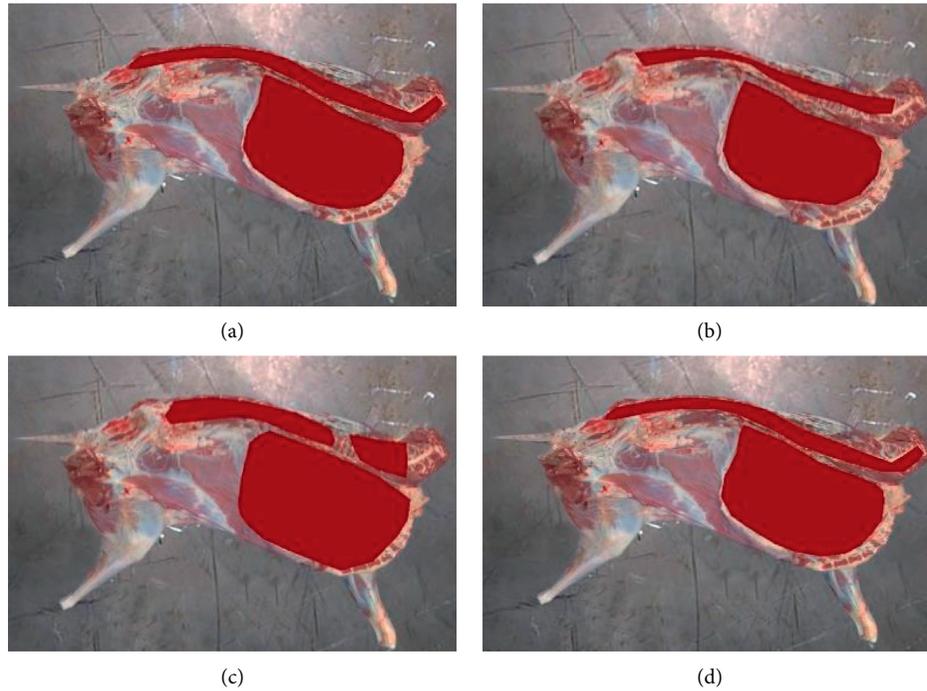


FIGURE 8: Segmentation result using FCN-8S, FCN-16S, FCN-32S, and Deeplab-v3+.

TABLE 2: Comparison between manual and automatic methods in image segmentation.

Sheep type	Target	Manual labeling		FCN Deeplab v3+		
		Time/s	Time (s)	IoU (%)	Time (s)	IoU (%)
BAMEI	Ribs	11.12	3.89	98.3	7.76	98.8
	Spine	18.07	3.65	95.1		96.2
LAKE	Ribs	12.05	3.91	95.4	8.21	96.2
	Spine	19.37	3.68	94.8		95.3
DOBER	Ribs	11.72	3.85	94.1	8.11	94.7
	Spine	17.54	3.62	93.1		93.6

3.2. Accuracy Test of Cutting Curves Recognition. An experiment was designed to evaluate the recognition performance of the cutting curves calculation algorithm. The 60 RGB images taken by KINECT described in Section 2.1 were used for this test. According to the cutting standard, a cutting curve is manually drawn in an image as a reference. The cutting curves automatically recognized by the algorithm are drawn in another image. IoU (intersection over union) has been identified in

Section 3.1. Table 3 shows the difference between the manual method and the automatic recognition method. The time consumption of the recognition algorithm is only 12% that of the manual method, and the overall positioning accuracy is 95.1%. There are three main sources of errors. The first is that the error of image segmentation leads to the error of contour and key points, which directly reduces the accuracy of segmentation line positioning. The second reason is that the cutting curves are

TABLE 3: Comparison of the manual model and the automatic model in recognition of cutting curves.

Segmentation type	Number of images	Time (s)	IoU (%)
Manual	50	17.7	—
FCN	50	4.2	95.1
Deeplab v3+	50	4.4	96.2

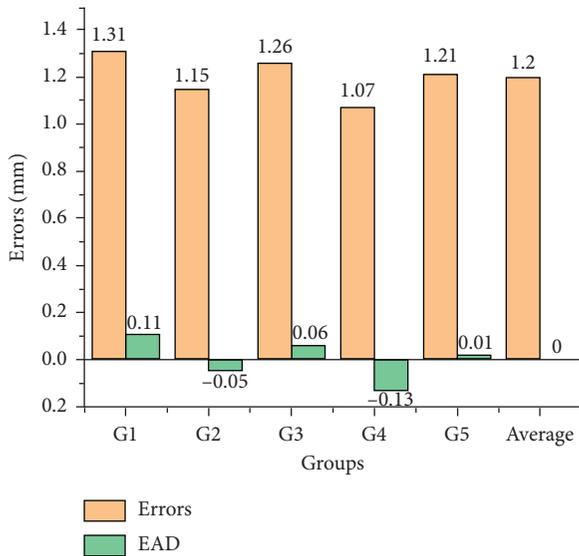


FIGURE 9: Positioning deviation.

composed of straight segments and curves, and the manual dividing method is flexible and includes many curves. This calculation method causes inherent errors. The third factor is that there are large differences in the internal structure of the carcass of sheep of different growth stages and breeds, which makes it difficult to use the algorithm.

3.3. Cutting Curves Positioning Accuracy. The spatial coordinates of the actual sheep body are difficult to mark, which is not conducive to verifying the positioning accuracy of the vision system. In order to evaluate the accuracy of the visual system, a calibration board with different horizontal planes is used to replace the sheep body. A checkerboard with a side length of 100 mm × 100 mm is used as the calibration board. The calibration board consisting of 120 checkerboards was put on the cutting platform. With a height of 800 mm, one vertex of the calibration board on the cutting table is the origin of the world coordinate system, and the corner points on the calibration board can be assigned spatial coordinates. These space coordinates are used as position reference values. Then, the vertical axis of the robotic arm is adjusted, and the height of the Kinect sensor stays in the range of 1000 mm~1200 mm, respectively, and takes pictures every 40 mm. The corner points in the image of the calibration board were detected, and then the RGB-D mapping relationship and hand-eye calibration rules were adopted to calculate the space coordinates of the corner points. Repeat 6 sets of the tests to compare the millimeter difference between the measured value and the true Euclidean distance, and the error analysis diagram is shown in Figure 9. EAD is the difference between each group error and the total error. The

results indicate that the average pixel error for the visual vision positioning was 1.3 mm. The errors in different horizontal planes were stable. The accuracy and stability of the system is considerable and it could be applied in practical engineering.

4. Conclusions

This paper designed a vision system to carry out the robotic cutting of half-sheep. Cutting is a challenging operation and is predominantly carried out manually. The vision system combines a 3D vision algorithm for calculation and localization of the cutting curves using Azure Kinect camera, with a self-made robot. This system could automatically collect RGBD images of a half-sheep on the workbench. The ribs and spine in a half-sheep may be segmented using Deeplab v3+ model. Based on the identification of ribs and spine, the five cutting curves may be calculated and 3D-located. The paper includes a proof-of-concept demonstration showing that the entire robot-based cutting procedure for six parts was carried out in less than 22.6 seconds. The system is an example of research and technology development with potential for flexible and adaptive robot-based automation in livestock processing. It demonstrates that there is scope for automating even the most challenging food processing operations. Factors such as improper operation and deterioration of the knives during the lamb splitting operation will result in poor quality of the spine splitting, with burrs, cracks, and blurring. This will cause holes in the image segmentation of the spine, or even break into several sections. This situation will lead to incorrect dividing line calculation results. Adopting better deep learning models and adding damaged spine training images will reduce the impact of this situation.

Data Availability

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 Key Research and Development Program of China (2018YFD0700804).

References

- [1] L. Pannier, G. E. Gardner, R. A. O'Reilly, and D. W. Pethick, "Factors affecting lamb eating quality and the potential for their integration into an MSA sheepmeat grading model," *Meat Science*, vol. 144, pp. 43–52, 2018.
- [2] Y. Lu, J. Liu, H. Zhou, M. Pan, and H. Wu, "Status and development trend of intelligent equipment and technology

- on livestock and poultry processing,” *Agricultural Engineering*, vol. 9, pp. 42–55, 2019.
- [3] 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.
 - [4] M. Steen Bond, J. Reidar Mathiassen, P. Aaby Vebenstad et al., “An automated salmonid slaughter line using machine vision,” *Industrial Robot: An International Journal*, vol. 38, no. 4, pp. 399–405, 2011.
 - [5] Y. Liu, M. Cong, D. Liu, and F. Zhu, “Trajectory planning for porcine abdomen cutting based on an improved genetic algorithm and machine vision for industrial robot,” *Robot*, vol. 39, no. 3, pp. 377–384, 2017.
 - [6] H. Azarmdel, S. S. Mohtasebi, A. Jafari, and A. Rosado Muñoz, “Developing an orientation and cutting point determination algorithm for a trout fish processing system using machine vision,” *Computers and Electronics in Agriculture*, vol. 162, pp. 613–629, 2019.
 - [7] S. Wang, K. Tao, and L. I. Hang, “Design and experiment of poultry eviscerator system based on machine vision positioning,” *Transactions of the Chinese Society for Agricultural Machinery*, vol. 49, no. 1, pp. 335–343, 2018.
 - [8] B. Pang, X. Sun, D. Liu, and K. Chen, “On-line acquisition and real-time segmentation system of beef rib-eye image,” *Transactions of the Chinese Society for Agricultural Machinery*, vol. 44, no. 6, pp. 189–193, 2013.
 - [9] A. Dasiewicz, A. Radzik-rant, W. Rant, and M. Świątek, “Assessment of slaughter value of sheep on the basis of linear measurements made on digital carcass images,” *Annals of Warsaw University of Life Sciences—SGGW—Animal Science*, vol. 56, no. 1, pp. 53–63, 2017.
 - [10] J. Singh, J. Potgieter, and W. L. Xu, “Ovine automation: robotic brisket cutting,” *Industrial Robot: An International Journal*, vol. 39, no. 2, pp. 191–196, 2012.
 - [11] K. Guo, F. Xu, T. Yu, X. Liu, Q. Dai, and Y. Liu, “Real-time geometry, albedo, and motion reconstruction using a single RGB-D camera,” *ACM Transactions on Graphics*, vol. 36, no. 3, pp. 1–13, 2017.
 - [12] E. Misimi, E. R. Øye, A. Eilertsen et al., “GRIBBOT—robotic 3D vision-guided harvesting of chicken fillets,” *Computers and Electronics in Agriculture*, vol. 121, pp. 84–100, 2016.
 - [13] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, 2016.
 - [14] H. Deng, Y. Zhou, T. Xu, M. Teng, and J. Xu, “Optimization of cattle’s image semantics segmentation with fully convolutional networks based on RGB-D,” *Transactions of the Chinese Society of Agricultural Engineering*, vol. 35, no. 18, pp. 151–160, 2019.
 - [15] J. Ma, H. Zhang, P. Yi, and Z. Wang, “SCSCN: a separated channel-spatial convolution net with attention for single-view reconstruction,” *IEEE Transactions on Industrial Electronics*, vol. 67, no. 10, pp. 8649–8658, 2020.
 - [16] M. Cong, H. Wang, X. Ren, Yu Du, and D. Liu, “Design of porcine abdomen cutting robot system based on binocular vision,” in *Proceedings of the 2019 14th International Conference on Computer Science & Education (ICCSE)*, pp. 188–193, Toronto, Canada, August 2019.
 - [17] Yi Peng, Z. Wang, K. Jiang, Z. Shao, and J. Ma, “Multi-temporal ultra dense memory network for video super-resolution,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 8, pp. 2503–2516, 2020.
 - [18] A. Yang, Y. Xue, H. Huang, and N. Huang, “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.
 - [19] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, Boston, MA, USA, June 2015.
 - [20] D. Wu, X. Yin, B. Jiang, M. Jiang, Z. Li, and H. Song, “Detection of the respiratory rate of standing cows by combining the Deeplab V3+ semantic segmentation model with the phase-based video magnification algorithm,” *Biosystems Engineering*, vol. 192, pp. 72–89, 2020.
 - [21] M. R. Chandraratne, D. Kulasiri, C. Frampton, S. Samarasinghe, and R. Bickerstaffe, “Prediction of lamb carcass grades using features extracted from lamb chop images,” *Journal of Food Engineering*, vol. 74, no. 1, pp. 116–124, 2006.
 - [22] T. Chan, D. Lichti, A. Jahraus et al., “An egg volume measurement system based on the Microsoft Kinect,” *Sensors*, vol. 18, no. 8, p. 2454, 2018.