applied sciences

MDPI

Article

Cattle Weight Estimation Using Fully and Weakly Supervised Segmentation from 2D Images

Chang-bok Lee [1], Han-sung Lee [1] and Hyun-chong Cho [1,2,*]

[1] Department Graduate Program for BIT Medical Convergence, Kangwon National University, Chuncheon 24341, Republic of Korea
[2] Department of Electronics Engineering, Kangwon National University, Chuncheon 24341, Republic of Korea
* Correspondence: hyuncho@kangwon.ac.kr

Abstract: Weight information is important in cattle breeding because it can measure animal growth and be used to calculate the appropriate amount of daily feed. To estimate the weight, we developed an image-based method that does not stress cattle and requires no manual labor. From a 2D image, a mask was obtained by segmenting the animal and background, and weights were estimated using a deep neural network with residual connections by extracting weight-related features from the segmentation mask. Two image segmentation methods, fully and weakly supervised segmentation, were compared. The fully supervised segmentation method uses a Mask R-CNN model that learns the ground truth mask generated by labeling as the correct answer. The weakly supervised segmentation method uses an activation visualization map that is proposed in this study. The first method creates a more precise mask, but the second method does not require ground truth segmentation labeling. The body weight was estimated using statistical features of the segmented region. In experiments, the following performance results were obtained: a mean average error of 17.31 kg and mean absolute percentage error of 5.52% for fully supervised segmentation, and a mean average error of 35.91 kg and mean absolute percentage error of 10.1% for the weakly supervised segmentation.

Keywords: convolutional neural network; deep learning; fully supervised segmentation; weakly supervised segmentation; cattle weight estimation

check for updates

Citation: Lee, C.-b.; Lee, H.-s.; Cho, H.-c. Cattle Weight Estimation Using Fully and Weakly Supervised Segmentation from 2D Images. *Appl. Sci.* **2023**, *13*, 2896. https://doi.org/10.3390/app13052896

Academic Editor: Jan Egger

Received: 30 January 2023
Revised: 17 February 2023
Accepted: 20 February 2023
Published: 23 February 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Periodic weight measurement can measure the growth rate of cattle, identify animals at target weight, and serve as a method for an indirect health checkup. However, because the weight of a cow after breeding is approximately 700–1100 kg, it is difficult for livestock farms to transport the cattle to a scale, and therefore, weight information cannot be easily gathered or actively used. According to the Korea Rural Economic Institute, the price of Korean beef has been rising steadily since 2015. Statistics Korea also explained that per capita beef consumption in Korea is increasing by an average of 4.3% per year due to an increased preference for meat. These increases in cattle, cattle prices, and beef consumption are the same worldwide. According to OECD Statistics, the production, consumption, and price of beef are all increasing [1]. Figure 1 shows the growth of the cattle livestock market in Korea and the world. Figure 1a shows the increasing number of livestock farms raising more than 100 cattle in Korea, Figure 1b shows the world's beef production and consumption, and Figure 1c shows the rise in global beef prices.

Related studies that investigated the correlation between body type and body weight of cattle demonstrate the potential of predicting body weight using the body parts of cattle; this was the motivation behind conducting the current study. Qiao et al. [2] used Mask R-CNN to separate cattle and background. They used a histogram difference-based approach to select repetitive and similar images in the video. As a result of the experiments, they obtained a mean pixel accuracy of 0.92, which showed better performance than

SharpMask (0.82) and DeepMask (0.53). Banos et al. [3] revealed that body height, chest width, and body depth had correlations of 0.65, 0.73, and 0.49, respectively, with body weight, and Koenen et al. [4] showed that body depth had a correlation of 0.49 with body weight. Vallimont et al. [5] investigated the correlation between body height and body weight and obtained a correlation of 0.94.

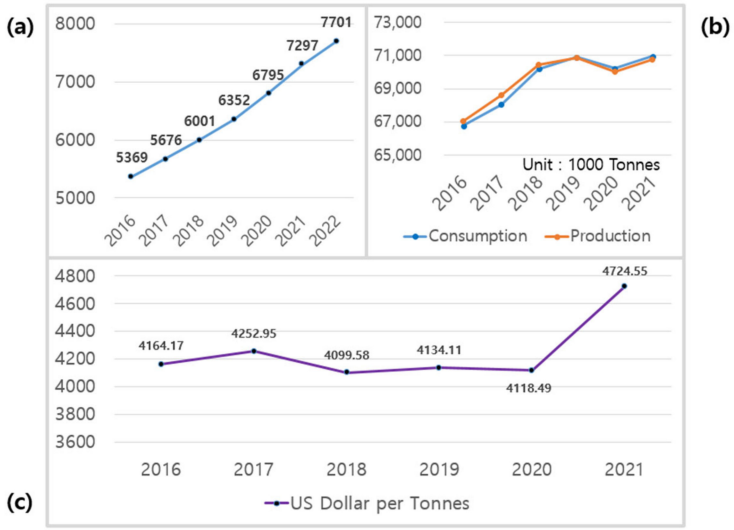


Figure 1. (**a**) Number of Korean beef farms producing more than a 100 head of cattle per year; (**b**) trends in world beef production and consumption; (**c**) trend of global beef prices.

Some studies attempted to estimate the weight of animals using image-based methods. Weber et al. [6]. attempted to estimate the weight of Girolando cattle using physical characteristics such as heart girth (HGP), abdominal circumference, and body length. Pearson's correlation was used to estimate the correlation between the traits and weight. HGP had the highest correlation coefficient at 0.88, followed by the circumference of the abdomen at 0.79, hip width at 0.65, and body length at 0.58. As a result, weight estimation using a linear regression model showed an RMSE of 42.52 kg. Na. [7] attempted to estimate the weight of a cow using a Bayesian Ridge algorithm applied to RGB-Depth camera images. A threshold was applied to the histograms of the depth image and the HIS image to obtain a mask of cattle from the background. The mask was used for feature extraction relevant to weight estimation. As a result, a mean square error (MSE) of 1046.0 was obtained. Weber et al. [8] obtained a mean absolute percentage error (MAPE) of 2.27% using a manually created segmentation mask to find the distances between points on the outline of the segmentation mask, and estimated the cattle weight using those values. Seo et al. [9] used cattle top-view and side-view data and achieved an error rate of 5% to 10.7% using multiple regression equations. Hansen et al. [10] obtained a weight estimation error rate of 6.1% for predicting the volume of cattle using a 3D depth camera. Wongsriworaphon et al. [11] predicted the weight of pigs with an error rate of 4% using a regression neural network and the area captured by the human eye in a top-view image of pigs. Using side-view data can further improve performance, but it is difficult to take this approach in livestock farms because of animal occlusions. In general, 3D cameras have a limited imaging distance so they capture a smaller area than 2D cameras, and are more expensive. We used 2D top-view images in this study.

We propose a method to estimate cattle weight using an image-based approach. Figure 2 shows a flow chart of the cattle weight estimation system. After extracting the cattle area from the image, features that are highly correlated with weight are extracted, and the weight is estimated using a deep neural network. Two methods were used to create the segmentation mask. The first is a precise but fully supervised segmentation method that requires a large amount of labor to generate the ground truth mask. The second

segmentation method is relatively less precise but is weakly supervised; its advantage is that it requires few human resources to generate the ground truth. The results obtained in this study show that the image-based method, which does not induce stress on animals, is effective, reduces the amount of repetitive manual effort needed to weigh cattle using scales, and can therefore improve the productivity of large livestock farms.

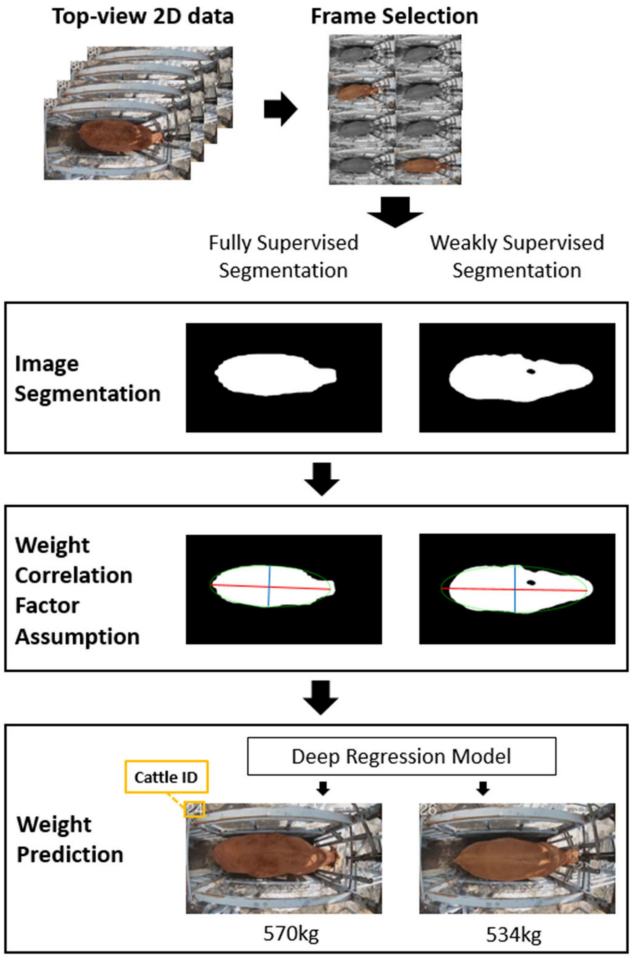


Figure 2. Flow chart of our cattle weight estimation system.

2. Materials and Methods

2.1. Database

Research data were obtained from the Gangwon-do Livestock Research Institute in Hoengseong, Gangwon-do, Republic of Korea. Because it is a research institution, this institute measures the body shape and weight of its cattle every three or six months. The guideways, equipment, and staff needed to measure body shape and weight were prepared for the experiment.

A camera was installed on the ceiling above the center of a scale to collect images during weight measurement. Examples of the collected images are shown in Figure 3a. We collected video data at 1280 × 720 (HD) resolution for a total of 43 cattle. Most were recorded between 10:00 h and 14:00 h, so the lighting was sufficiently bright. The actual weights ranged between 158 kg and 651 kg, with an average of 340.46 kg and a standard deviation of 148.73 kg. The weight distribution is shown in Figure 3b. Of the data for 43 cattle, the data for 23 cattle were used for training, and the remaining data were used for testing. To construct the datasets, all 43 cattle were first classified by weight into two groups: light and heavy. The average weight of the light and heavy groups was 260.33 kg and 421.23 kg, respectively. So that the average cattle weight of the training and testing data were similar, the training and testing data were randomly selected in the same ratio

from each group. The weight distribution can be biased if a dataset is randomly distributed without considering the weight. Since this cannot be seen as a generalized performance of the study, the data sets were distributed in groups according to weight in this study. The average weights of the training and testing data were 327.78 kg and 355.05 kg, respectively.

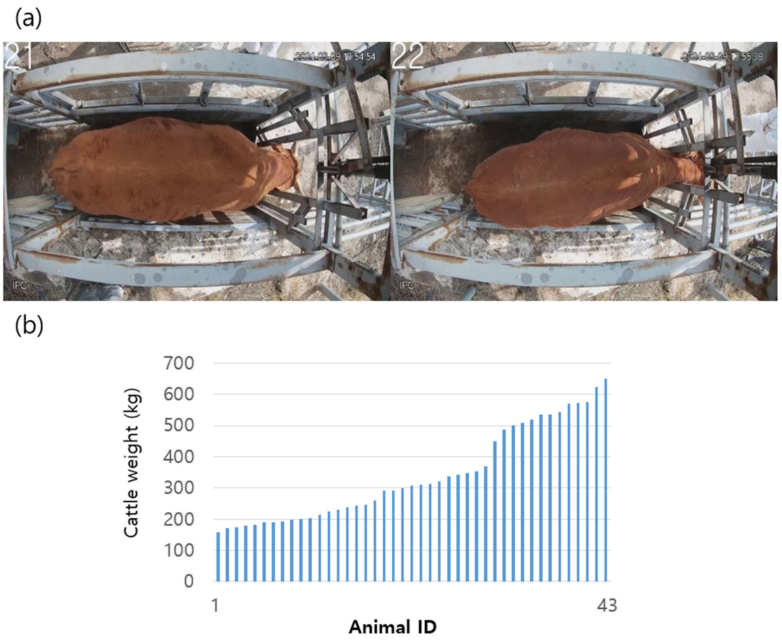


Figure 3. Collected research data: (**a**) examples of top-view image data; (**b**) weight distribution statistics.

2.2. Devices and Settings

The computer used in this study was equipped with an Intel Xeon processor (NVIDIA TITAN RTX 24 GB). The programming environment consisted of a Windows 10 operating system, CUDA 11.0, Keras 2.8.0, TensorFlow 2.5.0, PyTorch 1.10.0, and Torchvision 0.11.0. The camera used for data collection was a GB-CDX (GB-CDX04, GASI). The maximum resolution of the camera was 1920 × 1080 (FHD) at a speed of 60 FPS; however, we used a 1280 × 720 (HD) at a speed of 30 FPS. In the training of deep learning networks, large-resolution input requires many parameters and calculations. Max pooling, which is generally used in the pooling layer of convolutional neural networks (CNNs), leaves only large features and connects to the next layer. Therefore, if a resolution that is sufficient to clearly capture cattle is used, the analysis performance will not improve even if a larger resolution is used. Therefore, to improve the efficiency of the storage space, the camera was set to HD resolution.

2.3. Frame Selection for Effective Learning

In this study, the Mask R-CNN model, the Xception model, and the proposed regression artificial neural network model were used [12,13]. Supervised learning compares the ground truth with the model output and updates the weights according to the difference between the two. The weight update value is proportional to the input, and if there is a large amount of data, the characteristics of common features of the target class can be learned correctly. However, the repetition of similar data impedes learning and increases the risk of overfitting. The study data consisted of a video at 30 FPS, and the data for the continuous frames every 1/30th of a second were essentially the same because the difference was quite insignificant. Data were collected through a histogram difference-based approach to avoid a decrease in the efficiency of the learning process due to the similarly repeating images. The amount of data through histogram variation can be changed by the threshold value. In addition, each cow has a different degree of movement when measuring weight, so

it is necessary to select a threshold that can collect appropriate data for each cow. The threshold value was determined to collect approximately 100 images per cow from the training dataset. A total of 9359 images were extracted from about 125,550 images using this method. The process was as follows.

1. The image was converted into a grayscale image and the histograms were computed.
2. The sum of the histograms was calculated by adding all the values in each histogram.
3. Steps 1 and 2 were executed for successive frames, and the images were selected only when the difference between the histograms of the two images exceeded the threshold value of 50,000.

Because this method utilizes only the difference of the histogram summed values, its operation is simple. Examples of the selected and discarded data are shown in Figure 4a. The difference between the three images can be seen in Figure 4b. The composition of the dataset after frame selection is presented in Table 1. Validation was performed by randomly selecting 20% of the training dataset every epoch.

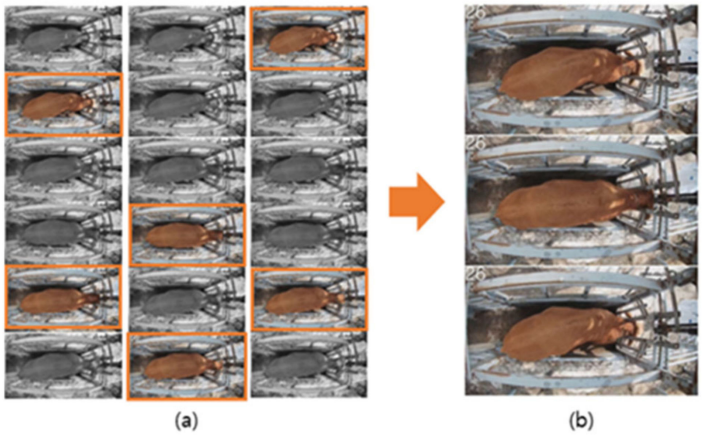


Figure 4. Image frame selection results: (**a**) original data; (**b**) frame selection results.

Table 1. Configuration of the dataset.

Training Data	Validation Data	Test Data	Total
4251	1063	4765	9359

2.4. Segmentation Methods

We employed two methods to obtain the segmentation masks of the cattle.

The first is fully supervised segmentation (FSS). FSS uses the Mask R-CNN model, which requires a polygon-shaped ground truth mask to train a deep learning model. The masks in the FSS model are precise; however, creating a ground truth mask requires a human to outline the animal by selecting image points. A total of 9359 images were used in the study, and three livestock experts worked over two months to generate the ground truth masks for all images. The opinions of three experts were cross-validated to determine the final decision by majority vote in cases of disagreement.

The time spent generating the ground truth masks is not the same for all image segmentation tasks. This is labor intensive, as the ground truth must be obtained each time additional data are acquired. These shortcomings of FSS motivated the development of a second method.

The second method is a weakly supervised segmentation (WSS). It is a deep learning model developed using the Xception model as a backbone with Grad-CAM and Puzzle-CAM. The Xception network is trained to classify binary classes of 'Cattle' and 'Background', and the training dataset consisted of 5100 cattle images and 1554 images of the background. This model can perform image segmentation with the same level of ground

truth as a general classification model without the need to use the ground truth in the form of a mask. WSS is less accurate than FSS, but is easier to use when additional data are collected. The two methods have different advantages and disadvantages. The difference in the ground truth generation process is visualized in Figure 5.

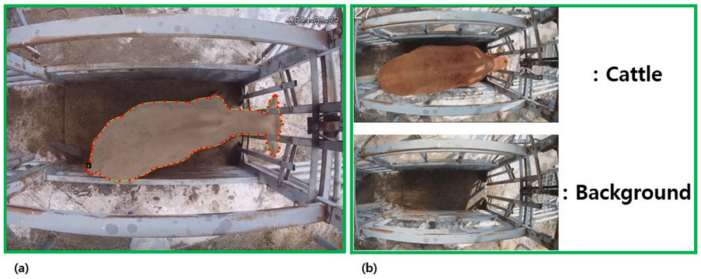


Figure 5. Difference between two ground truths; (**a**) FSS ground truth; (**b**) WSS ground truth.

2.4.1. FSS

Our FSS approach employs a Mask R-CNN, which combines a Faster R-CNN as the object detection model with a fully convolutional network (FCN) to form the image segmentation model [14,15]. Combining these algorithms and using them together improves the image segmentation performance. The FCN model is entirely composed of convolutional layers and is different from a general CNN in that it can output feature maps. The feature map output from the last convolutional layer of this model is upsampled to the same size as that of the input image, and the weights of the entire model are updated using the difference between the labelled ground truth mask and the output. The trained FCN divides the object from the background. Because the Faster R-CNN model performs the function of classifying objects along with detection, the FCN model eliminates the need to classify pixels into multiple classes. Therefore, instead of classifying pixels into multiple classes using softmax as an activation function, we employed the sigmoid function, which classifies a pixel as a mask pixel or non-mask pixel (i.e., 1 or 0) to yield clear predictions. Figure 6 shows the Mask R-CNN structure. The results of the cattle image segmentation using a Mask R-CNN are shown in Figure 7. The image segmentation task uses intersection over union (IoU) as an evaluation index, which is calculated by dividing the overlapping area by the combined area of the two masks to indicate the amount by which the ground truth mask and the model output mask overlap. The mean IoU of the cattle image segmentation using FSS was 0.92. As demonstrated by this result and related research, Mask R-CNN can effectively separate cows and the background with high performance, and its performance has been validated to the extent that it can be used as the backbone for subsequent segmentation models. Therefore, it was used as the segmentation model in this study.

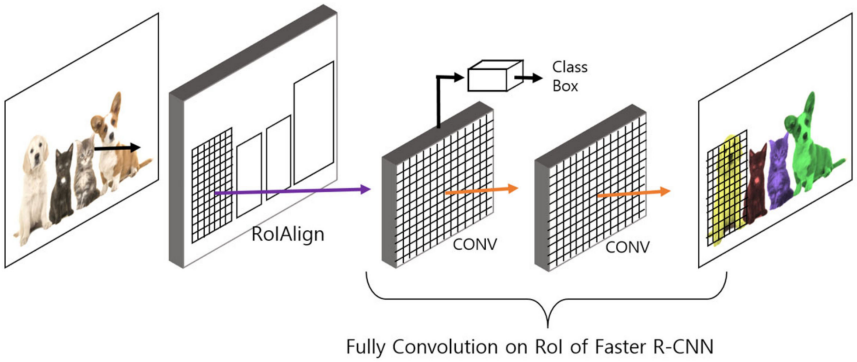


Figure 6. Mask R-CNN structure.

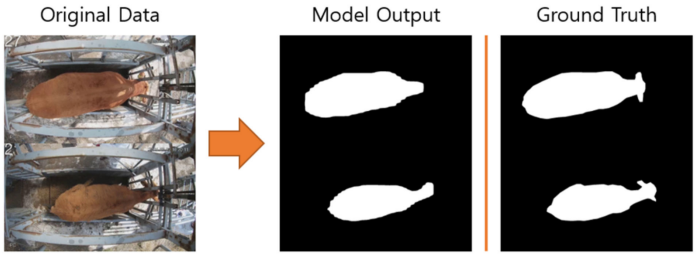


Figure 7. Image segmentation result of the FSS method.

2.4.2. WSS

For WSS, to obtain an activation map for the segmentation mask, the Xception model was used with two classes of data, an image with an animal, and an image with only a background; an activation map was obtained using the Grad-CAM method [16]. Because the desired mask quality could not be obtained by using Grad-CAM alone, it was obtained using the Puzzle-CAM method [17]. Grad-CAM is a learning visualization technique for interpreting processes in CNNs and has been studied to identify the features that models learn [18]. Grad-CAM uses the feature map A^k of the trained model and gradient map α_k^c, which is a gradient value for updating A^k then, α_k^c and A^k are matched element-wise. Map α_k^c can be expressed as follows:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y_c}{\partial A_{i,j}^k}. \tag{1}$$

where i and j are the width and height, respectively, and Z is the number of pixels in the map. Moreover, y_c is the predicted value of the model. Therefore, Equation (1) represents the global average pooling processed values of the gradient map that occurs when the classification model classifies class c and the weights are updated. Next, $L_{Grad-CAM}^c$ is obtained by multiplying Equation (1) by A^k, and then passing it through the ReLU function to discard values less than zero. This process can be expressed as follows:

$$L_{Grad-CAM}^c = \text{ReLU}(\sum_k \alpha_k^c A^k). \tag{2}$$

Because Grad-CAM is based on a classification model, it is unsuitable for segmentation tasks. In Figure 8, which shows the Grad-CAM results for our data, it can be seen that the activation map is generated only for a specific part of the cattle. This is because the classification model was able to distinguish the presence or absence of cattle using only the head and buttocks. This result cannot be expected in the case of segmentation, which is the purpose of WSS. To solve this problem, three types of Puzzle-CAM losses were used.

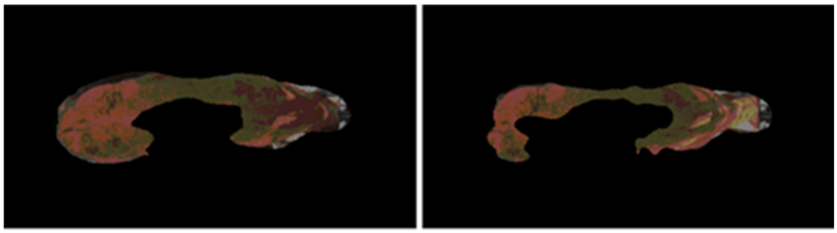


Figure 8. Grad-CAM results.

Puzzle-CAM defines three types of loss, as shown in Figure 9. The first is the classification loss of the original input image, the second is the classification loss of four image patches, and the third is the difference between the CAM of the original input image and the CAM of the four-patch image. Because the difference between the CAM of the original input image and the CAM of the four-patch image is used as a loss, it is possible to prevent only a specific part of the object from being visualized. To leverage the advantages of

Puzzle-CAM and Grad-CAM simultaneously, an activation map is generated using the Grad-CAM method, and the new loss of Puzzle-CAM (i.e., loss1 + loss2 + loss3) is used to update the Xception model. When training the Xception model, a Grad-CAM activation map is generated every time the weights are updated, and a new loss is calculated so that the Xception model learns by considering both the classification ability and the activation map generation ability. The results are depicted in Figure 10. As can be seen, the results of Grad-CAM, which only used a part of the animal as a region of interest, are improved.

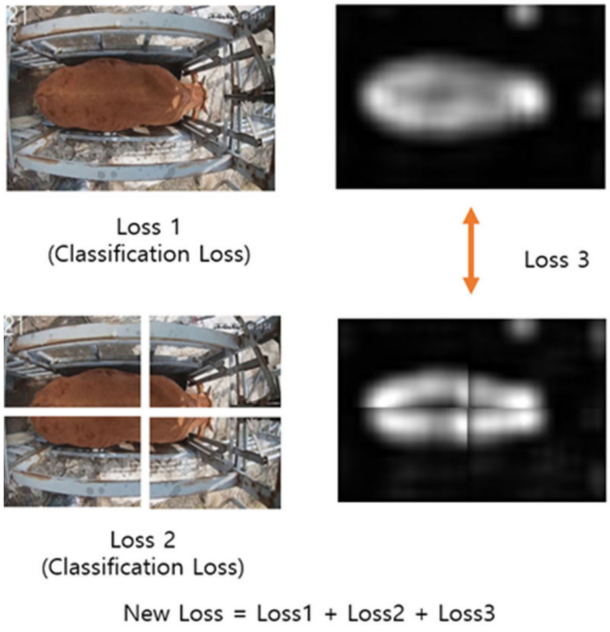


Figure 9. Puzzle-CAM loss.

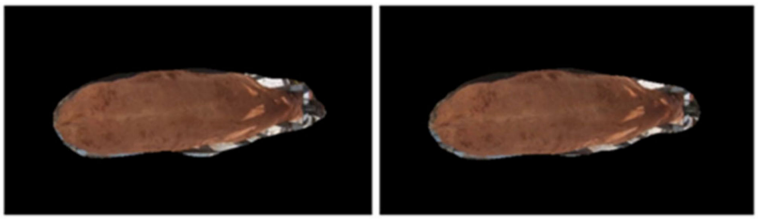


Figure 10. Segmentation results of the WSS method.

2.5. Weight Features for FSS and WSS

To design our weight estimation system, we extracted features based on the automatic segmentation. The area, body length, and width are good features for weight estimation. The area can be easily obtained from the segmentation result, but the body length and width cannot be calculated from the horizontal and vertical lengths because the direction changes as the animal moves. As a result, the optimal ellipse was used to measure a constant body length and width regardless of the direction the animal was facing unless it took a position parallel to the camera. After drawing an optimal ellipse for the segmentation mask, the major and minor axes of the optimal ellipse were measured and used as the body length and body width of the animal. In this way, the same value can be obtained for the same animal even when its posture changes. Figure 11a shows an example of an optimal ellipse fitted to a segmentation mask. The additional features used in this study are grid lengths, perimeter, minimum area rectangle, convex hull area, convex hull-to-contour ratio (solidity), circumscribed circle, eccentricity, and aspect ratio. Grid lengths were extracted through the following process. After obtaining the major axis and the minor axis of the optimal ellipse, the angle between the major axis and the horizontal axis was calculated. Then, the image was adjusted so that the major axis was parallel to the horizontal axis so that the animal was

horizontal with the camera. The segmentation result was vertically divided into 10 equal parts, and the length was measured. The same process was performed horizontally on the segmentation result, but the image was divided into five sections instead of 10. As a result, 15 lengths corresponding to 10 vertical lines and five horizontal lines were obtained. Figure 11b shows the obtained grid length and Figure 12 shows the perimeter, convex hull area, minimum area rectangle, and circumscribed circle. The correlation between each variable and the weight was analyzed using Pearson correlation coefficients. Only the train dataset was used to calculate the Pearson correlation coefficient to ensure analytical reliability. The correlation coefficients for each feature of the FSS mask are shown in Table 2, and those of the WSS mask are presented in Table 3.

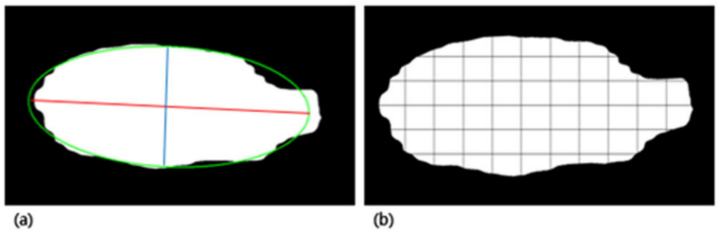


Figure 11. Calculating weight features: (**a**) optimal ellipse of the segmentation mask; (**b**) grid lengths.

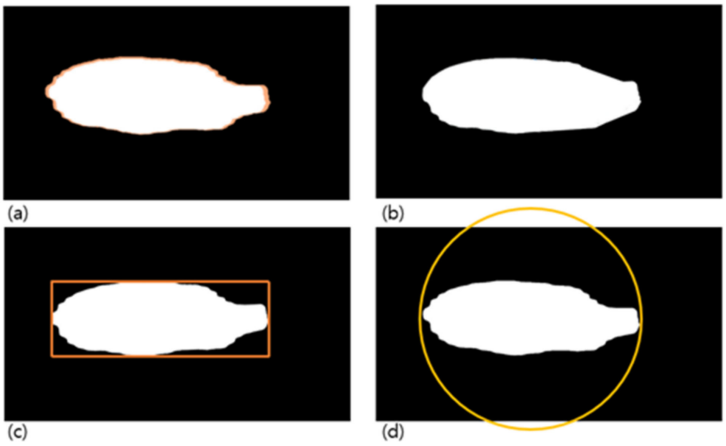


Figure 12. Features: (**a**) perimeter; (**b**) convex hull area; (**c**) minimum area rectangle; (**d**) circumscribed circle.

Table 2. Features selected for the FSS method and their correlation coefficients.

Feature	Pearson Correlation Coefficient
Area	0.953
Minimum Area Rectangle	0.941
Convex Hull Area	0.936
Minor Axis Length	0.935
Grid Length (Vertical)	0.852
Circumscribed Circle	0.847
Perimeter	0.836
Major Axis Length	0.785
Grid Length (Horizontal)	0.776
Solidity	0.547
Aspect Ratio	0.127
Eccentricity	−0.029

Table 3. Features selected for the WSS method and their correlation coefficients.

Selected Feature	Pearson Correlation Coefficient
Area	0.763
Convex Hull Area	0.721
Circumscribed Circle	0.651
Minimum Area Rectangle	0.644
Major Axis Length	0.619
Perimeter	0.576
Minor Axis Length	0.539
Grid Length (Horizontal)	0.462
Grid Length (Vertical)	0.450
Aspect Ratio	0.316
Eccentricity	0.141
Solidity	0.057

3. Results and Discussion

Twelve kinds of features were extracted from the segmentation mask; the nine features with correlation coefficients greater than 0.5 were used as input to the weight estimation model, and the actual weight was used as the ground truth. The proposed model has a fully connected layer consisting of an input layer, nine hidden layers, and an output layer. The residual connection of ResNet is used in every two layers. Generally, neural networks can solve multi-dimensional problems as the network deepens; however, this causes the gradient vanishing phenomenon to increase. A residual connection is used in the weight estimation model structure because it solves the gradient vanishing problem using only an addition operation, enabling the network to be deeper. The mean squared error was used as a loss function to handle large errors. Every layer has 32 nodes. The batch size, depth of the model, and number of nodes were set to the values that yielded the best performance in the experiments. Overfitting was prevented by stopping learning when no improvement in the validation results was observed over 20 epochs. For evaluation, we used the mean average error (MAE), which is a commonly used error metric. Of the 43 cattle, 23 (4251 images) were used for learning, and 20 (4765 images) were used for testing. In addition, the results of the proposed method were compared with those obtained using support vector regression (SVR) based on the radial basis function (RBF) kernel and polynomial kernel. Grid search was used to tune the optimal parameters for SVR. The results are shown in Table 4. Our proposed weight estimation system achieved the highest performance for both FSS and WSS. For the FSS mask, the RBF and polynomial kernel SVRs achieved MAPE scores of 7.7% and 11.6%, respectively, and the proposed method achieved a MAPE of 5.5%. For the WSS mask, the RBF and polynomial kernel SVRs achieved MAPE scores of 11.3% and 14.4%, respectively, and our method achieved 10.1%.

Table 4. Weight estimation performance of FSS and WSS.

	Method	MAE (kg)	MAPE (%)
FSS	Ours	17.31	5.5
	SVR (RBF)	27.57	7.7
	SVR (Polynomial)	41.29	11.6
WSS	Ours	35.91	10.1
	SVR (RBF)	40.12	11.3
	SVR (Polynomial)	51.45	14.4

In SVR, the kernel maps data to higher dimensions, allowing more complex problems to be solved. The Gaussian and polynomial kernels can be respectively expressed as follows.

$$\mathrm{K}(\mathrm{x}, \mathrm{y}) = \exp\left(-\gamma \|x - y\|^2\right) \quad (3)$$

$$K(x, y) = \left(x^T y + c\right)^d. \quad (4)$$

In Equation (3), x-y is the Euclidean distance, and γ is a parameter that controls the area over which the kernel filters. In Equation (4) d determines the order of the polynomial and c adjusts the coefficients of the polynomial.

We performed additional experiments to determine the optimal features that correlate with cattle weight. The experiments were conducted based on the Pearson correlation coefficients of the features. The results obtained by our model are presented in Table 5 and the results obtained by SVR are presented in Table 6. As can be seen in Table 2, for the FSS mask, two features with correlation coefficients less than 0.5 have very small correlation coefficients, −0.029 and 0.127. Therefore, from 0.5 to 0.9, the applied value was increased by 0.1. When the WSS mask was used, since the largest features had a correlation coefficient of 0.763, the applied value was increased by 0.1 from 0 to 0.7.

Table 5. Performance of the proposed method according to correlation coefficient intervals.

Mask	Correlation Coefficient	Number of Features	MAE (kg)	MAPE (%)
FSS	0	12	19.14	5.3
	0.5	10	17.31	5.1
	0.6	9	18.37	5.1
	0.7	7	23.20	6.5
	0.8	6	23.21	6.5
	0.9	4	23.77	6.6
WSS	0	12	35.19	9.9
	0.1	11	35.53	10.0
	0.2 = 0.3	10	35.01	9.8
	0.4	9	36.69	10.3
	0.5	7	35.91	10.1
	0.6	5	37.65	10.6
	0.7	2	40.49	11.4

Table 6. Performance of RBF kernel-based SVR according to the correlation coefficient interval.

Mask	Correlation Coefficient	Number of Features	MAE (kg)	MAPE (%)
FSS	0	12	28.23	7.9
	0.5	10	27.57	7.7
	0.6	9	28.28	7.9
	0.7	7	30.92	8.7
	0.8	6	30.11	8.4
	0.9	4	28.56	8.4
WSS	0	12	40.12	11.3
	0.1	11	40.76	11.4
	0.2 = 0.3	10	40.90	11.5
	0.4	9	41.27	11.6
	0.5	7	44.06	12.4
	0.6	5	50.36	14.1
	0.7	2	61.62	17.3

Our model outperformed the SVR model, improving the MAPE by 2.2% and 1.2% when FSS and WSS were used, respectively. In this experiment, we were able to obtain an MAE of 35.01 kg and MAPE of 9.8% using WSS. The FSS mask features yielded the best performance when the correlation coefficient was 0.5. When the value was 0, the performance was slightly degraded. Moreover, as the number of features decreased, the performance tended to decrease. This trend was also true for the WSS mask. The best performance was obtained when the value was 0.2 and 0.3, and the performance decreased as the number of features decreased.

If the number of features is large, even those features with small correlation coefficients are included in the input of the model, and a large amount of information can be conveyed to the model. If the number of features is small, only features with large correlation coefficients should be used for analysis. Because the machine learning model can determine the optimal weight, it can be concluded that including a feature with a small correlation coefficient in the model input can help the model perform better than entirely excluding the coefficient.

4. Conclusions

Artificial intelligence techniques can enable analyses that were not previously possible, and can be applied to the livestock field for great benefits. In the livestock industry, the weight of cattle is directly related to the competitiveness of the farm.

In this study, an image-based cattle weight estimation system was proposed. To estimate the body weight, a segmentation mask was obtained from the top-view images of 43 cattle, 12 kinds of features were extracted from the segmentation, and regression analysis was performed. The MAPE of the proposed system was 5.06% for FSS and 9.8% for WSS. The results of the FSS method demonstrated that it can be effectively used in livestock farms. Fully supervised segmentation is generally a method that can achieve high accuracy compared to weakly supervised methods. However, preparing a sufficient dataset for the deep learning model to train requires a significant investment of human resources and time. On the other hand, this study has demonstrated that weakly supervised segmentation can achieve a similar performance while dramatically reducing the required human resources and time. In addition, the performance of the deep learning model improves as the amount of training data increases; therefore, if enough data are available, the performance of the WSS method can be improved. In future studies, we plan to find more effective features in the WSS mask and improve the performance of the WSS method by collecting more research data.

Author Contributions: Conceptualization, H.-c.C.; methodology, C.-b.L., H.-s.L. and H.-c.C.; software, C.-b.L.; validation, C.-b.L. and H.-s.L.; formal analysis, C.-b.L.; investigation, C.-b.L.; resources, H.-s.L.; data curation, C.-b.L. and H.-s.L.; writing—original draft preparation, C.-b.L.; writing—review and editing, H.-c.C.; visualization, C.-b.L. and H.-s.L.; supervision, H.-c.C.; project administration, H.-c.C.; funding acquisition, H.-c.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2022R1I1A3053872) and was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2022RIS-005).

Institutional Review Board Statement: The animal study was approved by the Gangwon-do Livestock Research Institute in Republic of Korea.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available upon request from the corresponding author.

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

References

1. OECD Statistics. OECD-FAO Agricultural Outlook 2022–2031. Available online: https://stats.oecd.org/# (accessed on 24 August 2022).
2. Qiao, Y.; Truman, M.; Sukkarieh, S. Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming. *Comput. Electron. Agric.* **2019**, *165*, 104958. [CrossRef]
3. Banos, G.; Coffey, M.P. Prediction of liveweight from linear conformation traits in dairy cattle. *J. Dairy Sci.* **2012**, *95*, 2170–2175. [CrossRef] [PubMed]
4. Koenen, E.P.C.; Groen, A.F. Genetic evaluation of body weight of lactating Holstein heifers using body measurements and conformation traits. *J. Dairy Sci.* **1998**, *81*, 1709–1713. [CrossRef] [PubMed]

5. Vallimont, J.E.; Dechow, C.D.; Daubert, J.M.; Dekleva, M.W.; Blum, J.W.; Barlieb, C.M.; Baumrucker, C.R. Genetic parameters of feed intake, production, body weight, body condition score, and selected type traits of Holstein cows in commercial tie-stall barns. *J. Dairy Sci.* **2010**, *93*, 4892–4901. [CrossRef] [PubMed]
6. Weber, V.A.M.; Weber, F.D.L.; Gomes, R.D.C.; Oliveira Junior, A.D.S.; Menezes, G.V.; Abreu, U.G.P.D.; Pistori, H. Prediction of Girolando cattle weight by means of body measurements extracted from images. *Rev. Bras. Zootec.* **2020**, *49*, e20190110. [CrossRef]
7. Na, M.H.; Cho, W.H.; Kim, S.K.; Na, I.S. Automatic weight prediction system for Korean cattle using Bayesian ridge algorithm on RGB-D image. *Electronics* **2022**, *11*, 1663. [CrossRef]
8. Weber, V.A.M.; de Lima Weber, F.; da Silva Oliveira, A.; Astolfi, G.; Menezes, G.V.; de Andrade Porto, J.V.; Pistori, H. Cattle weight estimation using active contour models and regression trees Bagging. *Comput. Electron. Agric.* **2020**, *179*, 105804. [CrossRef]
9. Seo, K.W.; Kim, H.T.; Lee, D.W.; Yoon, Y.C.; Choi, D.Y. Image processing algorithm for weight estimation of dairy cattle. *J. Biosyst. Eng.* **2011**, *36*, 48–57. [CrossRef]
10. Hansen, M.F.; Smith, M.L.; Smith, L.N.; Jabbar, K.A.; Forbes, D. Automated monitoring of dairy cow body condition, mobility and weight using a single 3D video capture device. *Comput. Ind.* **2018**, *98*, 14–22. [CrossRef] [PubMed]
11. Wongsriworaphon, A.; Arnonkijpanich, B.; Pathumnaku, S. An approach based on digital image analysis to estimate the live weights of pigs in farm environments. *Comput. Electron. Agric.* **2015**, *115*, 26–33. [CrossRef]
12. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
13. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
14. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
15. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
16. Jo, S.; Yu, I.-J. Puzzle-CAM: Improved localization via matching partial and full features. In Proceedings of the 2021 IEEE International Conference on Image Processing, Anchorage, AK, USA, 19–22 September 2021; pp. 639–643.
17. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
18. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep feature for discriminative localization. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.

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