

Article

Advanced Swine Management: Infrared Imaging for Precise Localization of Reproductive Organs in Livestock Monitoring

Iyad Almadani ^{1,*}, Brandon Ramos ², Mohammed Abuhussein ¹ and Aaron L. Robinson ¹

¹ Electrical and Computer Engineering, University of Memphis, Memphis, TN 38111, USA; mbhssein@memphis.edu (M.A.); alrobins@memphis.edu (A.L.R.)

² Sivananthan Laboratories, 586 Territorial Dr Suite A, Bolingbrook, IL 60440, USA; bramos@sivananthanlabs.us

* Correspondence: mlmadani@memphis.edu

Abstract: Traditional methods for predicting sow reproductive cycles are not only costly but also demand a larger workforce, exposing workers to respiratory toxins, repetitive stress injuries, and chronic pain. This occupational hazard can even lead to mental health issues due to repeated exposure to violence. Managing health and welfare issues becomes pivotal in group-housed animal settings, where individual care is challenging on large farms with limited staff. The necessity for computer vision systems to analyze sow behavior and detect deviations indicative of health problems is apparent. Beyond observing changes in behavior and physical traits, computer vision can accurately detect estrus based on vulva characteristics and analyze thermal imagery for temperature changes, which are crucial indicators of estrus. By automating estrus detection, farms can significantly enhance breeding efficiency, ensuring optimal timing for insemination. These systems work continuously, promptly alerting staff to anomalies for early intervention. In this research, we propose part of the solution by utilizing an image segmentation model to localize the vulva. We created our technique to identify vulvae on pig farms using infrared imagery. To accomplish this, we initially isolate the vulva region by enclosing it within a red rectangle and then generate vulva masks by applying a threshold to the red area. The system is trained using U-Net semantic segmentation, where the input for the system consists of grayscale images and their corresponding masks. We utilize U-Net semantic segmentation to find the vulva in the input image, making it lightweight, simple, and robust enough to be tested on many images. To evaluate the performance of our model, we employ the intersection over union (IOU) metric, which is a suitable indicator for determining the model's robustness. For the segmentation model, a prediction is generally considered 'good' when the intersection over union score surpasses 0.5. Our model achieved this criterion with a score of 0.58, surpassing the scores of alternative methods such as the SVM with Gabor (0.515) and YOLOv3 (0.52).

Keywords: deep learning; U-Net; reproductive organ; localization; image segmentation; computer vision; farming



Citation: Almadani, I.; Ramos, B.; Abuhussein, M.; Robinson, A.L. Advanced Swine Management: Infrared Imaging for Precise Localization of Reproductive Organs in Livestock Monitoring. *Digital* **2024**, *4*, 446–460. <https://doi.org/10.3390/digital4020022>

Academic Editors: Phivos Mylonas, Katia Lida Kermanidis and Manolis Maragoudakis

Received: 18 February 2024

Revised: 27 April 2024

Accepted: 27 April 2024

Published: 2 May 2024



Copyright: © 2024 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

In today's agricultural landscape, farmers have increasingly embraced animal breeding for diverse objectives such as boosting productivity, fortifying disease resistance, improving reproductive efficiency, and adapting to environmental challenges, notably heat and drought. Highly productive animals prioritize allocating a significant portion of their nutritional intake toward desired outputs like milk or meat production rather than solely maintaining their own bodily functions [1]. This emphasis on longevity and reproductive success leads to a more concentrated distribution of feed calories among the most productive animals, reducing the need for a large number of replacements within the herd [2]. On the other hand, farm workers have developed specialized techniques to discern subtle changes in sow characteristics, such as vulva size and activity, in order to track their reproductive cycles. However, when it comes to replicating this efficiency using autonomous

machines, the results have been largely disappointing. The main challenge lies in accurately identifying and capturing representative feature data, which has remained elusive despite technological advancements. This stark contrast between human expertise and machine capabilities underscores the intricate nature of farming practices and the ongoing challenges in automating certain aspects of animal husbandry. In this research, our primary objective is to detect the vulvae of sows using the U-Net segmentation model [3], which has been proven to be the most effective technique for breeding management on farms. The vulva serves as a crucial indicator of the sow's ovulation cycle, and by accurately detecting temperature fluctuations in this region, we can gain insights into the specific phase of the estrous cycle [4]. To achieve our goal, we employ the U-Net semantic segmentation method, which involves the extraction of masks through the color threshold. Our system is trained using grayscale images alongside their corresponding masks, allowing us to identify the vulva in input images effectively. Following vulva detection, the minimum area rectangle (MAR) algorithm is applied to reduce unwarranted background information, making the detection more precise and efficient. Our objective is to achieve the minimum possible dimensions for the bounding box since the size of the organ is important in future work. The next step is to track the size of the vulva. Therefore, any extra information in the bounding box will act as noise during the estrous detection phase, leading to false standing heat positives. The utilization of U-Net with the MAR technique not only ensures a light weight and simplicity but also enhances the robustness of our model, enabling its evaluation across a diverse range of images. This paper comprehensively explores the theoretical foundations underlying the employed methods. Section 4 offers a detailed description of our proposed system, elucidating its intricacies and highlighting the rationale behind our approach. Additionally, we present extensive experimental results in this section, showcasing the performance and efficacy of our methodology. Furthermore, Section 5 is dedicated to evaluating the outcomes of our research using the intersection over union (IOU) [5], while Section 6 provides conclusive remarks based on our findings and outlines potential avenues for future expansion and refinement of this work.

2. Background

The global agricultural landscape is undergoing a profound transformation driven by technological innovations. Among these innovations, machine learning and deep learning have emerged as game-changers, reshaping the way farming operations are conducted. Recently, there have been significant advancements in object detection and segmentation techniques [6,7], which have found widespread applications across various domains. These methodologies have been extensively researched and applied to diverse fields, such as autonomous driving [8], medical imaging [9], and industrial automation [10–12]. The development of deep learning algorithms, particularly convolutional neural networks (CNNs) [13], has revolutionized many fields [14–16], especially the field of computer vision, enabling more accurate and efficient detection and segmentation of objects within images [17]. With the advent of the intelligent pig breeding industry, precise pig management has emerged as a critical aspect, and the ability to individually recognize pigs has become the cornerstone of achieving precision breeding. However, despite its importance, research into the field of pig breeding remains limited. Previous studies have focused on various aspects such as pig face detection [18,19], posture analysis [20], interactive touch-based techniques [21], and machine learning approaches for sow identification [22]. Many recent research efforts used changes in posture as an indicator of heat in sows [23]. Others attempt to detect estrus by extracting features or indicators dependent upon the change in the size around the sow's vulva [24]. This research [25] builds upon the findings of a study that highlighted the correlation between temperature changes in the vulva and the sow's reproductive status. However, detecting and understanding the sow's reproductive cycle and its management within a farm setting requires further investigation to make generalization and widespread application of these efforts more feasible. Achieving widespread application of these efforts necessitates accurate detection of the sow's vulva. To address

this challenge, our study focuses on employing U-Net semantic segmentation, a method that stands out due to its unique design. U-Net utilizes a contracting path to capture the overall context and a symmetric expanding path for precise localization. This design enables the detection of intricate patterns and fine details in images, making it particularly valuable in fields such as medical image analysis and satellite imagery. The ability of U-Net to combine high-level context with detailed information distinguishes it among semantic segmentation algorithms, followed by using the minimum area rectangle (MAR) algorithm to allocate the exact possible area of the object, making our research noteworthy.

3. Related Work

In this section, we thoroughly assess and discuss the methodologies and approaches we implemented to address the specific challenges presented by this problem. We delve into an in-depth examination of the strategies employed, providing a comprehensive analysis of their respective strengths, weaknesses, and overall effectiveness in tackling the intricacies inherent in the problem at hand.

3.1. Vulva Detection in Sows Using Gabor with the SVM Classifier

Figure 1 presents the full pipeline components of this method. The first stage involves collecting thermal RGB image datasets. Then, we extract their features using the Gabor filter bank, followed by feature selection via PCA. Consequently, our SVM classifier is trained using the resulting features. Finally, we elaborate upon the Gaussian pyramid with a sliding window to allocate the vulva locations through the trained SVM.

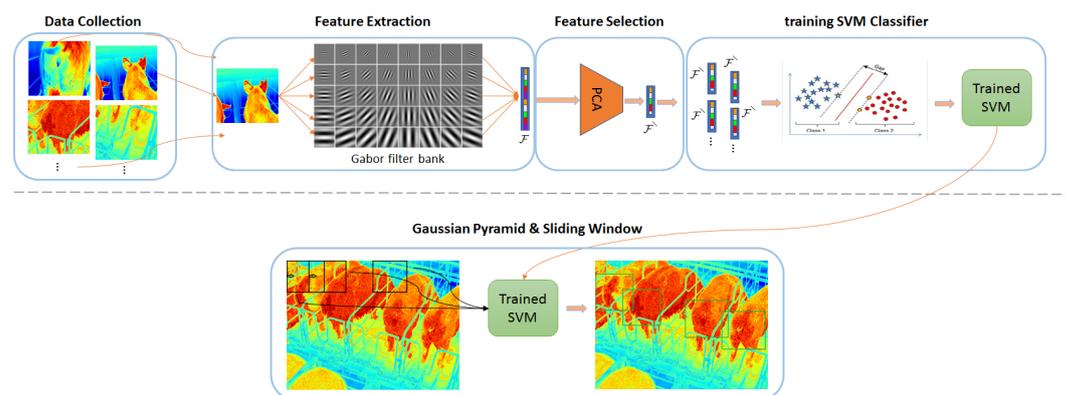


Figure 1. An overview of our Detection system construction.

This trained system's confusion matrix is shown in Figure 2. It demonstrates that this system could correctly classify the pig's vulva at a rate of 93%. It also shows that the percentages of false detection and missing detection were 4% and 7%, respectively. Since the confusion matrix is not sufficient to evaluate the object detection systems, other performance metrics and loss functions should be used, such as the IOU, and we will demonstrate the comparison between models using this metric in the following sections.

On the other hand, Figure 3 shows the qualitative results, illustrating that this system achieved success in a high-level range of image scenarios. For instance, Figure 3f,g presents the high detection speed of the system. When a farmer covered a sow's hindquarter with his hand in Figure 3f, the system was unable to detect the vulva in that instance, whereas the system became able to detect the sow's vulva right after the man removed his hand in Figure 3g. On the other hand, the figures represent the robustness of this system on different image scales. This can be seen in Figure 3a,e on one side and all other images on the other side. This demonstrates the importance of using the Gaussian pyramid technique, which derived the robustness of the system in a variety of scenarios and scales.

		True Class	
		Positive	Negative
Predicated Class	Positive	0.93	0.04
	Negative	0.07	0.96

Figure 2. The resultant confusion matrix of the trained system.

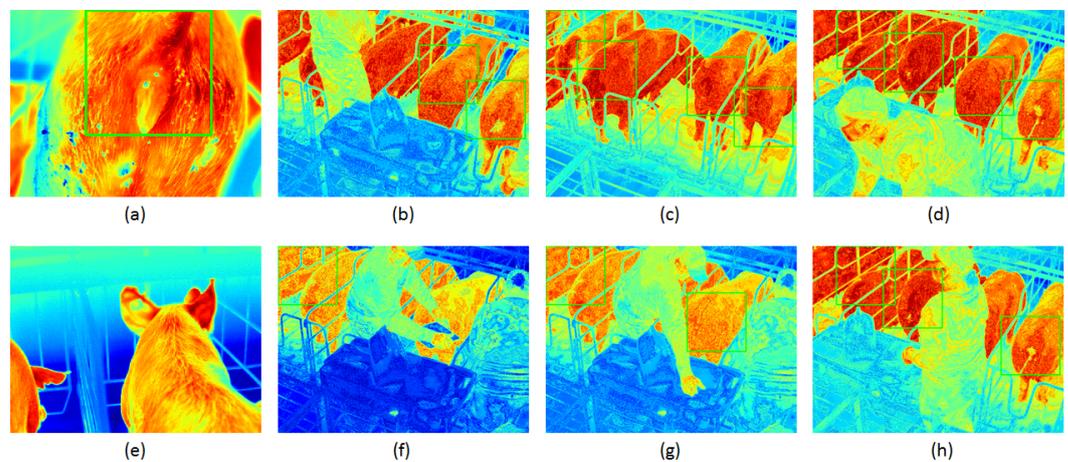


Figure 3. Results of the vulva detection system. The results cover a wide range of scenarios. (a) represents the robustness of our system on close-up images. (b–e) represent the robustness of our system from a distance. (f–h) present the high detection response of our system.

While this method is accurate and robust, it might not be the best fit for applications that require speedy decision making, like live monitoring systems.

3.2. Vulva Detection Using YOLOv3

The interpretation of a YOLO model's prediction results is just as nuanced as the model's implementation. A successful interpretation and accuracy rating are determined by several elements, including the box confidence score and class confidence score utilized when developing a YOLOv3 computer vision model. The results were evaluated using the mean average precision (mAP). The mAP computes a score by comparing the ground truth bounding box to the detected box. The higher the score, the more precise the model's detections. Figure 4 shows the results that were obtained utilizing our custom grayscale dataset. The images show the detection of the vulva and ear in the same model as well as the prediction accuracy.

While this method operates in real time, providing swift processing of data, it did not exhibit improvement in the intersection over union (IOU) when compared with the Gabor with the SVM classifier approach. Despite its real-time capabilities, the system's IOU performance remained comparable to or possibly less favorable than that achieved by the Gabor with the SVM classifier method. This indicates that while the system excelled in delivering speedy results, further enhancements may be explored to optimize the IOU metrics for more accurate and robust outcomes.

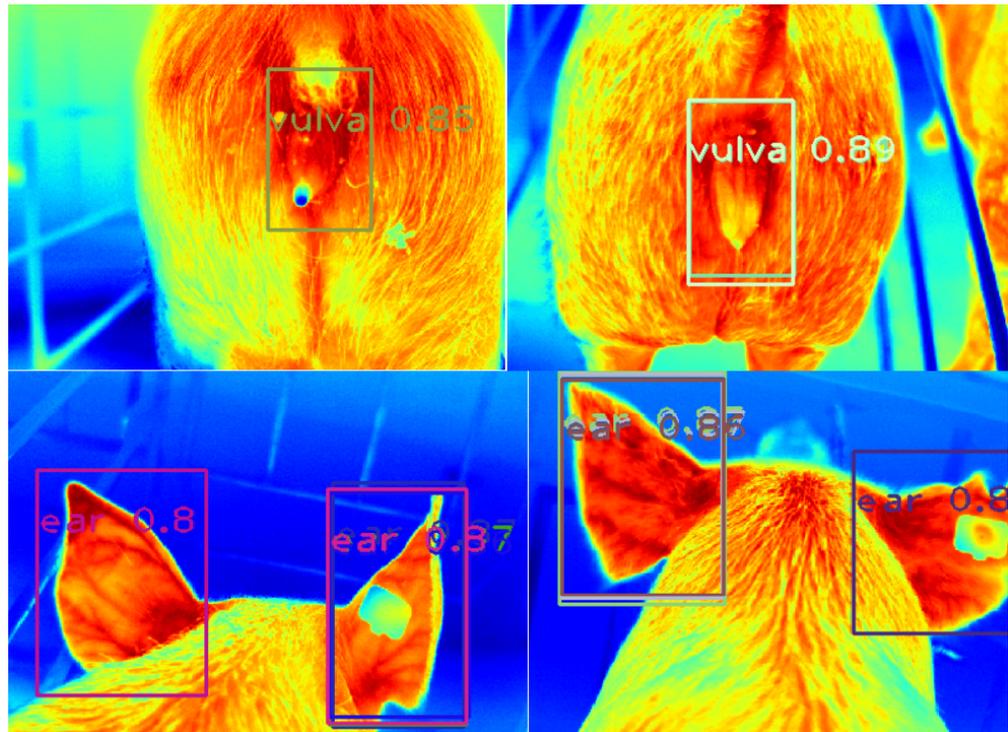


Figure 4. YOLOv3 results.

4. Method Overview

In this overview, we will explain the steps we took to find things accurately in our research. First, we extracted masks of the region of interest using OpenCV. This step involved precisely defining the areas we wanted to focus on and laying the foundation for our in-depth analysis. Then, we used grayscale images along with the extracted masks to train a U-Net semantic segmentation model. This integration was vital as it allowed us to harness the advanced capabilities of the U-Net model, enabling it to recognize and understand the specific areas defined by the masks. Finally, the model used a minimum area rectangle algorithm to optimize the predicted mask for better intersection over union (IOU) values. By fine-tuning the results using this technique, we ensured the accuracy and reliability of our findings, enhancing the overall quality of our research outcomes. Figure 5 presents our full pipeline components that are covered in detail in the following subsections.

4.1. Data Collection and Preparation

To train our models, a dataset was needed which had images that only contained vulva (positive dataset) and images that represented anything other than vulva (negative dataset). To accomplish this, we used a long-wave infrared (LWIR) camera to capture 4755 images of sows in the barn and then cropped the parts that contained only the sows' vulvae. In contrast, we created images with other sow parts, such as ears and heads, as well as other unspecified scenes, as shown in Figure 6. LWIR is one of the three wavelength bands in which infrared thermal imaging operates. The other two are medium-wavelength infrared (MWIR) and very long-wavelength infrared (VLIR). LWIR can cover wavelengths that range from 8000 nm to 14,000 nm (from 8 μm to 14 μm) [26]. An LWIR camera is capable of detecting thermal emissions of animals, vehicles, as well as people as they stand out when the temperature of the environment is different. Since it measures heat rather than light, LWIR imaging is frequently used as a thermal imaging solution. This makes it easier to distinguish living objects in images, which is why we used it in our research.

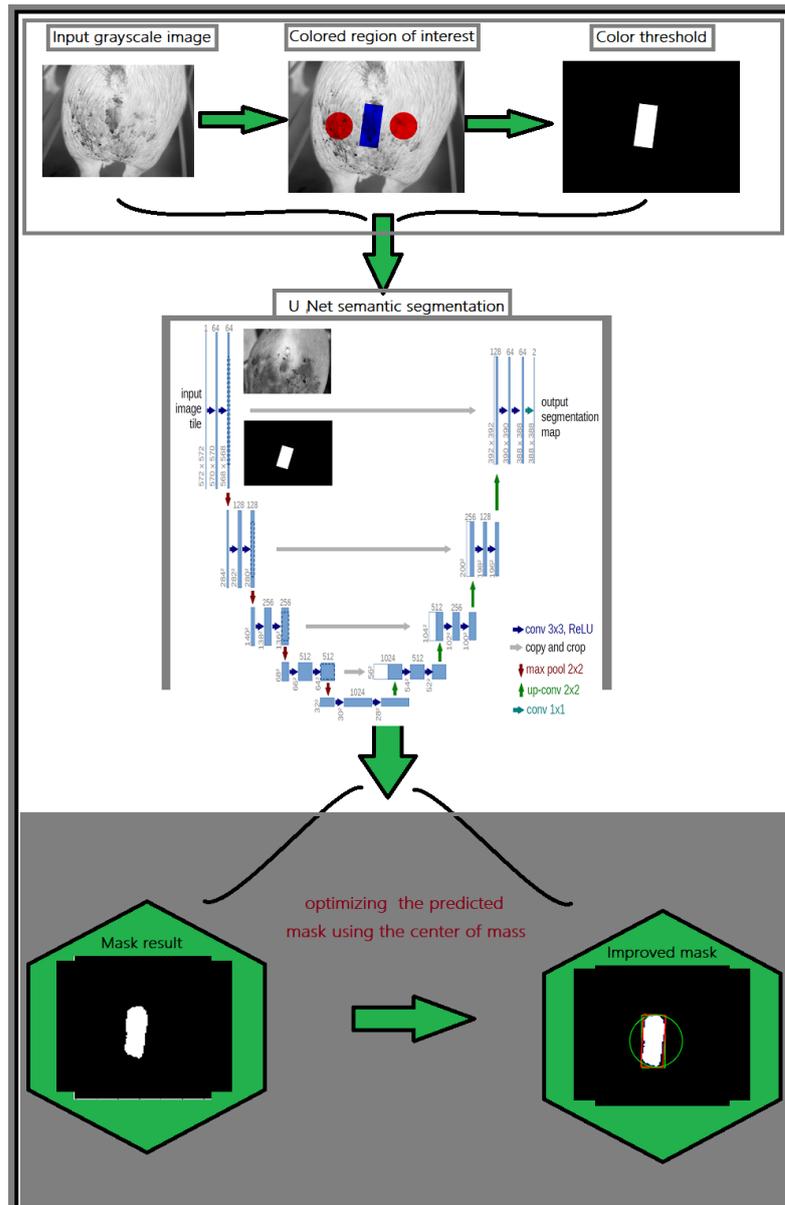


Figure 5. An overview of the segmentation model used in this research.

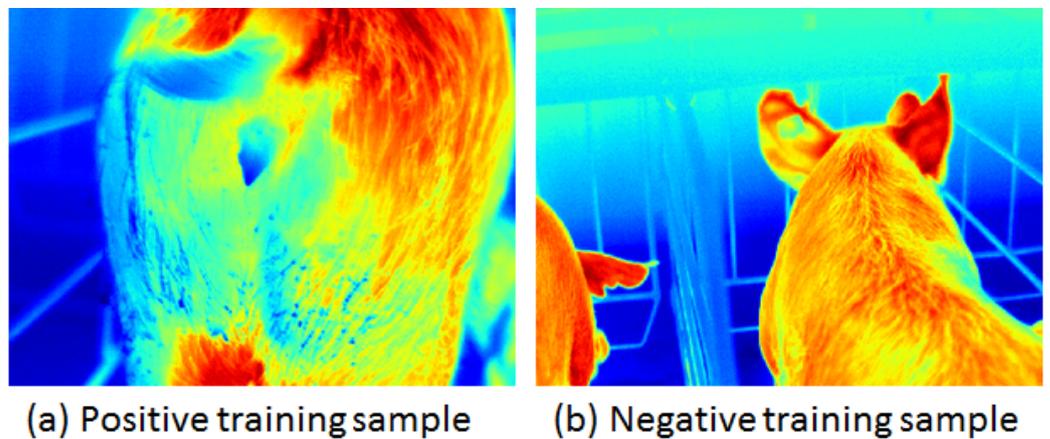


Figure 6. (a) presents a sample training image containing only sows' vulvae, representing our positive training dataset. The image in (b) shows a sample image of our negative training dataset.

4.2. Color Threshold

Color thresholding is a technique used to separate objects or regions of interest from the background in an image based on their color or intensity characteristics. It is a form of image segmentation that divides an image into distinct parts, facilitating further analysis or manipulation. Color thresholding operates on the principle that objects or regions of interest in an image often exhibit unique color characteristics that distinguish them from the surrounding background. In image processing, thresholding involves comparing pixel values to a predetermined threshold. Pixels meeting the specified condition are assigned new values, leading to the creation of a binary mask. This mask highlights areas in the image where the condition is met; pixels equal to or greater than the threshold are often set to 255 (representing white), and pixels below the threshold are set to 0 (representing black). In our research, color thresholding played a pivotal role. It enabled us to identify and extract specific regions of interest, such as those segmented in red. This targeted segmentation was crucial for our analysis. We obtained a binary representation by thresholding the image, providing a clear delineation of the region we were investigating. Figure 7 shows a sample of a segmented image and its mask.

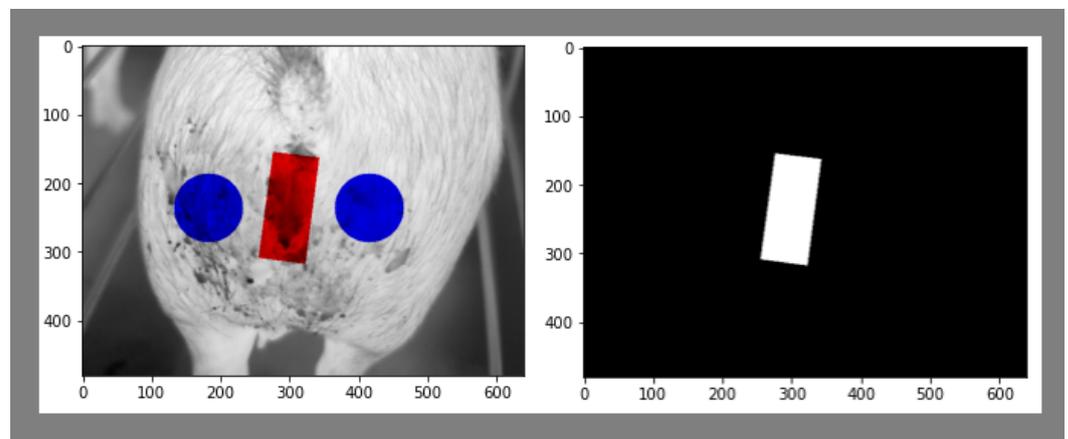


Figure 7. Segmented image sample and its mask.

4.3. U-Net Semantic Segmentation Architecture

Before delving into the specifics of U-Net, it is essential to grasp the fundamental concept of semantic segmentation. Unlike image classification, where the entire image is assigned a single label, semantic segmentation aims to categorize each pixel within an image, providing a detailed delineation of the object boundaries and classes [27]. This pixel-level understanding has a myriad of applications, including object detection, scene understanding, and medical image analysis. Semantic segmentation algorithms traditionally grapple with several challenges. One of the most pressing challenges is the need to capture both local and global contexts within an image while preserving high-resolution spatial information. The U-Net architecture was conceived as an answer to these challenges, offering an innovative solution that struck a balance between feature extraction and precise localization. U-Net is a deep learning architecture specifically designed for semantic segmentation tasks (Table 1). The U-Net architecture derives its name from its U-shaped structure (Figure 8), which resembles an encoder-decoder network. The model consists of two main components: the contracting path and the expansive path. The contracting path serves as the encoder and is responsible for capturing the context and extracting high-level features from the input image. It consists of several convolutional and max pooling layers that progressively reduce the spatial dimensions while increasing the number of feature channels. The contracting path in this network follows a common pattern used in convolutional networks. It repeatedly applies two 3×3 convolutions followed by a ReLU activation and then a 2×2 max pooling operation to reduce the size. Each time it reduces the image, it also doubles the number of features it looks for. The expansive path performs

the opposite functions. It enlarges the feature map and reduces the number of features. It accomplishes this by upsampling the features and then using a 2×2 convolution to decrease the number of features. It then combines this with the cropped feature map from the contracting path and applies more convolutions with ReLU activation. Finally, a 1×1 convolutional layer is used at the end to convert the features into the desired output classes. The entire network has 23 layers in total (Figure 8). The choice of U-Net as the segmentation model stemmed from its well-documented effectiveness in biomedical image segmentation tasks, particularly in scenarios where there is a need to preserve spatial information and capture intricate details in low-rank images, while other larger models may provide a slight performance improvement but are more costly during training and inference times.

Table 1. The network structure.

Contraction Path (Encoder)			
Block	Layer	No. of Filters	Activation Function
Block 1	Conv2D (3,3)	16	relu
	Conv2D (3,3)	16	relu
Block 2	Conv2D (3,3)	32	relu
	Conv2D (3,3)	32	relu
Block 3	Conv2D (3,3)	64	relu
	Conv2D (3,3)	64	relu
Block 4	Conv2D (3,3)	128	relu
	Conv2D (3,3)	128	relu
Block 5	Conv2D (3,3)	256	relu
	Conv2D (3,3)	256	relu
Expansive Path (Decoder)			
Block	Layer	No. of Filters	Activation Function
Block 6	Conv2DTranspose (2,2)	128
	Concatenate (c4,Transpose(2,2))	128
	Conv2D (3,3)	128	relu
Block 7	Conv2DTranspose (2,2)	64
	Concatenate (c3,Transpose(2,2))	64
	Conv2D (3,3)	64	relu
Block 8	Conv2DTranspose (2,2)	32
	Concatenate (c2,Transpose(2,2))	32
	Conv2D (3,3)	32	relu
Block 9	Conv2DTranspose (2,2)	16
	Concatenate (c1,Transpose(2,2))	16
	Conv2D (3,3)	16	relu

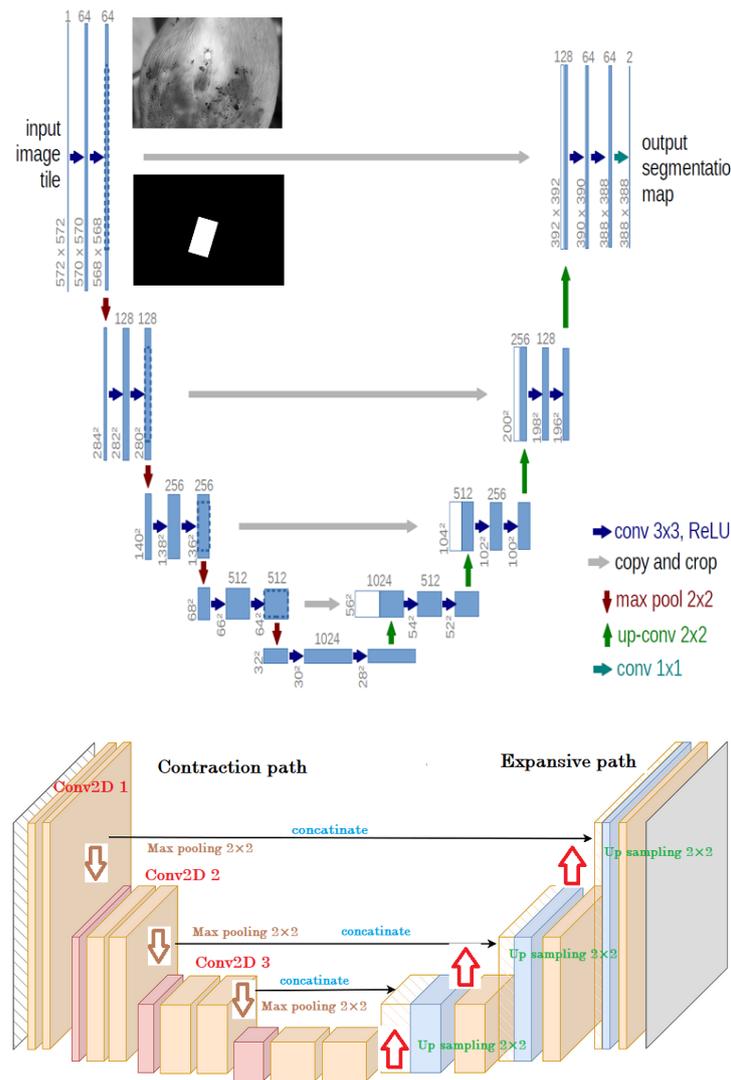


Figure 8. U-Net architecture.

4.4. Training Masks

In U-Net semantic segmentation, the training process typically involves using input masks and grayscale images to train the model to accurately segment objects of interest in an image. Input masks play a crucial role in the training of U-Net. An input mask is a binary image that has the same size as the original input image, where each pixel is assigned a value of either 0 or 1. The purpose of the input mask is to indicate the regions of interest or the ground truth segmentation annotations for the corresponding input image. During training, the U-Net model takes both the original grayscale image and the corresponding input mask as inputs. The grayscale image provides the visual information, while the input mask guides the model to learn the correct segmentation boundaries and object classes. The grayscale image is typically used as the input to the U-Net model. It represents the original image in a single channel, where each pixel value corresponds to the intensity or brightness of the pixel. The grayscale image is preprocessed and normalized before being fed into the model.

The input mask, on the other hand, is a binary image where the pixels belonging to the segmented object are assigned a value of 1, while the remaining pixels are labeled as background pixels and are assigned a value of 0. The input mask is usually created through manual annotation. However, automatic techniques can be used, depending on the availability of labeled data. The input mask provides the ground truth labels for the

training process, guiding the model to learn the correct segmentation boundaries and object classes. During the U-Net training phase, the model learns to map the input grayscale image to a predicted segmentation map, which is then compared to the ground truth input mask using a suitable loss function such as the dice coefficient or cross-entropy loss [28]. The model is optimized to minimize the discrepancy between the predicted segmentation map and the ground truth input mask. By utilizing input masks and grayscale images, U-Net learns to segment objects in the grayscale image accurately. The model learns to capture both low-level and high-level features, enabling it to detect and classify objects in the image based on the guidance provided by the input mask.

The network is trained using stochastic gradient descent implementation, where pairs of input images and their respective segmentation maps are used for training. The energy function is calculated through a softmax operation applied pixel by pixel on the final feature map, and this result is combined with the cross-entropy loss function, where the softmax operation is defined as

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}}$$

where z_i is the i th element of the input vector z and N is the total number of elements in the vector z . In addition, the cross-entropy function is defined as

$$\text{Loss} = - \sum_{k=1}^K y_k \cdot \log(\hat{y}_k)$$

where K represents the number of classes or categories, y_k is the true label for each pixel corresponding to class k , and \hat{y}_k is the predicted probability (or score) for class k .

This loss function is often used in scenarios where the task is to assign each pixel to one of K classes. The log term penalizes the model more if the predicted probability diverges from the true label. Minimizing this loss during training helps the model learn to produce predicted probabilities that align with the true labels.

4.5. Minimum Area Rectangle

The minimum area rectangle (MAR) algorithm [29] is a fascinating solution to a fundamental geometric problem in computational geometry. This algorithm, also known as the minimum bounding rectangle (MBR) algorithm, tackles the challenge of finding the smallest possible rectangle that can enclose a group of points on a flat surface. It is crucial to note that this rectangle must have sides that align with the coordinate axes, meaning they are either horizontal or vertical. To grasp the essence of the problem, imagine having a collection of points scattered across a two-dimensional plane and wanting to discover the smallest rectangle that can cover all these points. This rectangle should be aligned with the x and y axes, making it an axis-aligned rectangle. The MAR algorithm employs mathematical concepts like the convex hull, principal component analysis (PCA), and angle calculations to solve this problem efficiently. Firstly, it starts by finding the convex hull of the input points. The convex hull is the smallest convex polygon that encompasses all the given points. Think of this as a rubber band stretched around the outermost points. Secondly, the algorithm utilizes PCA, a mathematical technique often used in statistics and machine learning, to determine the principal axes of the points. These principal axes correspond to the directions with the maximum variability in the point distribution. In the context of the MAR algorithm, these axes are vital for figuring out the orientation of the minimum area rectangle. Lastly, the algorithm calculates the angle between the principal axis and the horizontal axis (x axis). This angle information helps determine how the minimum area rectangle should be oriented. Additionally, the information will be useful for future estrus detection research and will serve to improve the results obtained from the U-Net model. The masks are returned to the original rectangle shape by drawing a minimum rectangle around the predicted masks while considering the rotation angle. Figure 9 shows two rectangles: the green one is the regular bounding rectangle, while the blue one is the minimum area rectangle.

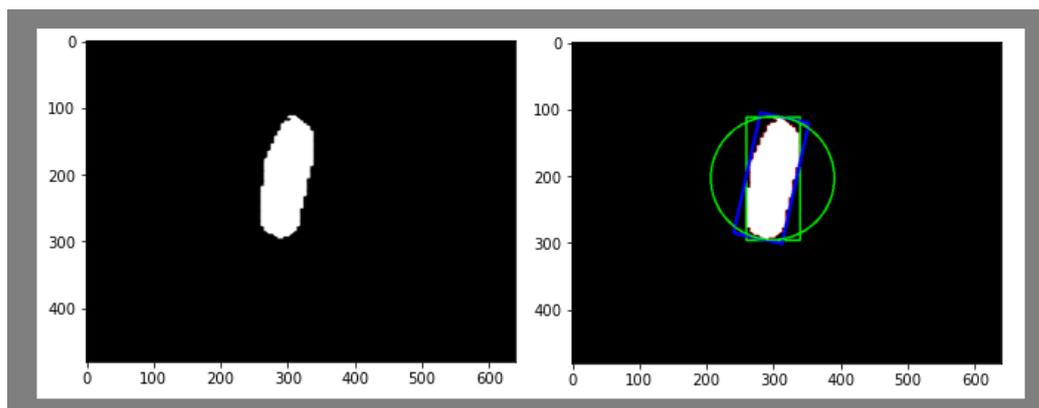


Figure 9. Figure on the (left) shows the predicted mask, while the figure on the (right) shows 1. Green cycle which represents the minimum enclosing circle drawn around each detected contour. 2. Green Rectangle: This represents the bounding box drawn around each detected contour. 3. Blue Rectangle: This represents the minimum area rectangle drawn around each detected contour.

5. Results and Discussion

The intersection over union (IOU) refers to the amount of overlap between two bounding boxes as shown in Figure 10. In the case of U-Net, it initially exhibited an IOU score of 0.50. However, after implementing a minimum area rectangle (MAR), the IOU score was further improved to 0.58. This method was compared with previous methods, including YOLOv3 and machine learning using Gabor with SVM classifier filters. In the machine learning algorithm, Gabor filters and a histogram of oriented gradients (HOGs) were used for feature extraction, followed by support vector machines (SVMs) for classification. However, the results revealed shortcomings in the model's performance, prompting the introduction of a second stage. In this stage, the dataset was trained using YOLOv3, a deep-learning object detection framework. The segmentation model developed was found to be the optimal choice for assisting in vulva localization. The outcomes of the ultimate approach demonstrate that the suggested system is capable of accurately segmenting the vulva region of a sow, even in images of poor quality, and exhibits outstanding performance efficiency.

Table 2 displays the decision-making process that led us to choose the U-Net model over the YOLOv3 model and other classical models. Based on the table, we observed that the performance of the SVM was enhanced by augmenting the dataset. Furthermore, when we trained the SVM with Gabor features, we noticed an improved intersection over union (IOU) score. We also observed better performance by increasing the number of Gabor filters from 12 to 96 feature maps. However, we did not make any changes to YOLOv3 since we switched to U-Net for better performance and a faster runtime. We achieved improved results with U-Net by adding the minimum average recall (MAR) as a post-processing step. we found that U-Net outperformed YOLOv3 significantly when both models were trained on the same number of labeled images. Furthermore, compared with YOLOv3, quantized U-Net runs at a faster speed, making it a more suitable choice for our project and setting the benchmark for future research improvements. In real-world applications, the model will be quantized and deployed on an edge device with a limited bandwidth, making speed a crucial factor in our selection. We referred to [30] for a comprehensive benchmarking of all YOLO versions and their speeds on multiple platforms. We were able to run the quantized U-Net model on our platform faster and with minimal loss of performance, confirming our decision.

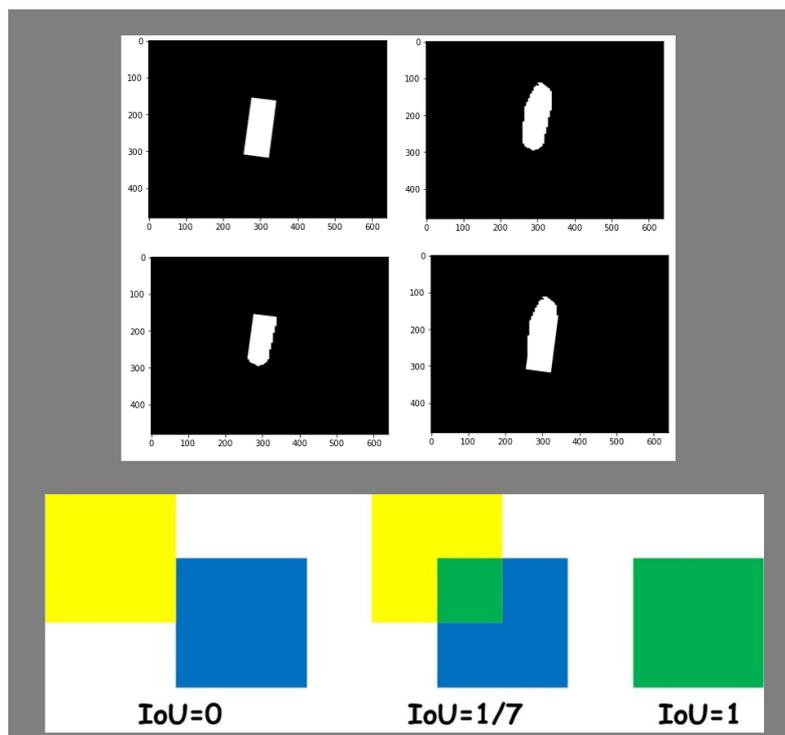


Figure 10. Intersection over union results. The (first left) image is the ground truth mask. The (first right) image is the prediction. The (second left) image shows the intersection between the ground truth and the predicted mask, while the (second right) image represents the union.

Table 2. Presentation of the performance of each model.

Comparison between Models			
Approach	Mean IOU	Improved IOU	FPS
SVM	0.400	0.450	35
SVM with Gabor	0.490	0.515	25
YOLOv3	0.520	0.520	40
U-Net	0.500	0.586	45

The U-Net architecture is well-known for its ability to perform image segmentation tasks with high accuracy, as shown in Figure 11. However, it has a limitation in that it only considers spatial information, which can result in suboptimal performance in certain scenarios. In our research, we discovered that when using U-Net to analyze vulva temperature data, some of the temperature information was lost during the normalization step of the preprocessing stage. Since the vulva temperature is a crucial indicator of estrous, this loss of information can lead to false negatives and impact the accuracy of estrous prediction. Therefore, it is essential to address this limitation of the U-Net architecture and devise more effective strategies to preserve the temperature information while performing image segmentation. Moreover, this will make it hard to generalize over other species. The overall pipeline will work, but it will require training from scratch instead of fine-tuning the existing model. In our upcoming research, we will address these limitations.

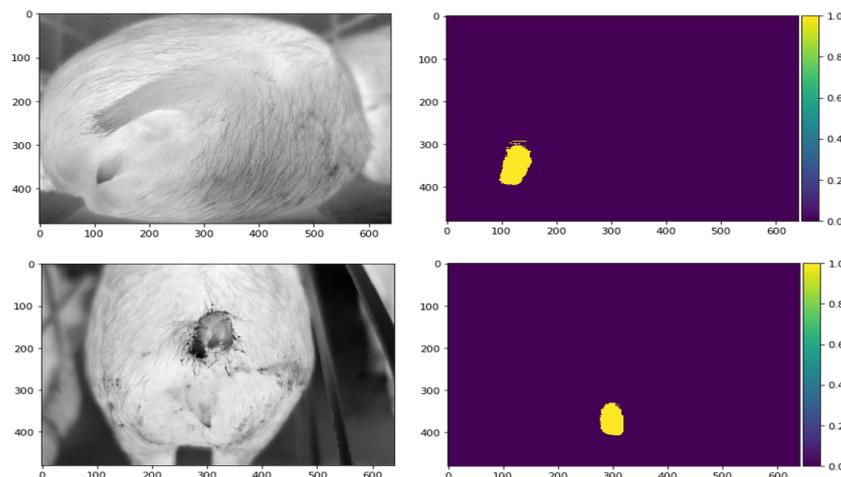


Figure 11. U-Net segmentation results.

6. Conclusions

Traditionally, the task of vulva localization has largely relied on manual labor, with farm workers employing specialized techniques to identify and track the reproductive cycles of sows. While this approach can yield valuable insights, it is labor-intensive, prone to human error, and challenging to scale efficiently in large farms. Moreover, farm workers are exposed to various occupational hazards, including toxins in the air, repetitive stress injuries, and mental health issues, highlighting the urgent need for automation and innovative solutions in the industry. Our research represents a paradigm shift by introducing the U-Net semantic segmentation model as a robust and reliable alternative for vulva localization. By harnessing the capabilities of deep learning, we demonstrated the potential to significantly reduce the reliance on manual labor while enhancing accuracy and efficiency. This transition from human-based methods to computer vision and machine learning is pivotal for the well-being of farm workers and the overall productivity of the livestock industry. In conclusion, this paper presents a novel approach for vulva localization using the U-Net semantic segmentation model. The proposed method was compared with alternative techniques, including YOLOv3 and machine learning using SVMs with Gabor filters. Through rigorous experimentation and evaluation, the results of our experiments unequivocally demonstrate the advantages of the U-Net model. It consistently outperformed both YOLOv3 and the SVM-Gabor filter combination across multiple metrics. While YOLOv3 exhibited competitive performance, it fell short in terms of localization accuracy and was sensitive to variations in pig posture and lighting conditions. On the other hand, the SVM-Gabor approach, while demonstrating some merit, lacked the adaptability and generalization capabilities offered by deep learning models like U-Net.

Further plans include employing the output of this method as the input for the next step in this research. The overall pipeline improves the automation of estrous detection based on tracking the size of the vulva and the changes associated with it over time. However, achieving precise and reliable results relies heavily on maintaining a consistent camera distance during image capture. Camera distance variations can lead to erroneous estrus estimations, potentially resulting in missed breeding opportunities or false positives. Future research avenues will make this approach robust at varying imaging distances by incorporating depth information. Later, we plan on using the minimized bounding box we achieved to track the size of the vulva.

Author Contributions: Conceptualization, I.A. and M.A.; methodology, I.A.; software, I.A.; validation, I.A., M.A. and A.L.R.; formal analysis, I.A.; investigation, I.A.; resources, A.L.R.; data curation, B.R.; writing—original draft preparation, I.A.; writing—review and editing, I.A.; visualization, I.A.; supervision, A.L.R.; project administration, A.L.R.; funding acquisition, A.L.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Sivananthan Laboratories.

Data Availability Statement: Data is unavailable due to privacy or ethical restrictions.

Acknowledgments: The authors express gratitude to Sivananthan Laboratories for their generous funding and provision of the dataset essential for this study.

Conflicts of Interest: The authors declare no conflicts of interest

References

1. Koketsu, Y.; Tani, S.; Iida, R. Factors for improving reproductive performance of sows and herd productivity in commercial breeding herds. *Porc. Health Manag.* **2017**, *3*, 1. [[CrossRef](#)] [[PubMed](#)]
2. Fry, J.P.; Mailloux, N.A.; Love, D.C.; Milli, M.C.; Cao, L. Feed conversion efficiency in aquaculture: Do we measure it correctly? *Environ. Res. Lett.* **2018**, *13*, 024017. [[CrossRef](#)]
3. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
4. Labrecque, J.; Rivest, J. A real-time sow behavior analysis system to predict an optimal timing for insemination. In Proceedings of the 10th International Livestock Environment Symposium (ILES X), Omaha, NE, USA, 25–27 September 2018; American Society of Agricultural and Biological Engineers: St. Joseph, MI, USA, 2018; p. 1.
5. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
6. Barbole, D.K.; Jadhav, P.M.; Patil, S. A review on fruit detection and segmentation techniques in agricultural field. In Proceedings of the Second International Conference on Image Processing and Capsule Networks: ICIPCN 2021, Bangkok, Thailand, 27–28 May 2021; Springer: Berlin/Heidelberg, Germany, 2022; pp. 269–288.
7. 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](#)]
8. Feng, D.; Harakeh, A.; Waslander, S.L.; Dietmayer, K. A review and comparative study on probabilistic object detection in autonomous driving. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 9961–9980. [[CrossRef](#)]
9. Latif, J.; Xiao, C.; Imran, A.; Tu, S. Medical imaging using machine learning and deep learning algorithms: A review. In Proceedings of the 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 30–31 January 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5.
10. Almadani, M.; Elhayek, A.; Malik, J.; Stricker, D. Graph-Based Hand-Object Meshes and Poses Reconstruction With Multi-Modal Input. *IEEE Access* **2021**, *9*, 136438–136447. [[CrossRef](#)]
11. Almadani, M.; Waheed, U.b.; Masood, M.; Chen, Y. Dictionary Learning With Convolutional Structure for Seismic Data Denoising and Interpolation. *Geophysics* **2021**, *86*, 1–102. [[CrossRef](#)]
12. Abuhussein, M.; Robinson, A. Obscurant Segmentation in Long Wave Infrared Images Using GLCM Textures. *J. Imaging* **2022**, *8*, 266. [[CrossRef](#)] [[PubMed](#)]
13. Chauhan, R.; Ghanshala, K.K.; Joshi, R. Convolutional neural network (CNN) for image detection and recognition. In Proceedings of the 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 15–17 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 278–282.
14. Moinuddin, K.A.; Havugimana, F.; Al-Fahad, R.; Bidelman, G.M.; Yeasin, M. Unraveling Spatial-Spectral Dynamics of Speech Categorization Speed Using Convolutional Neural Networks. *Brain Sci.* **2022**, *13*, 75. [[CrossRef](#)] [[PubMed](#)]
15. Havugimana, F.; Moinuddin, K.A.; Yeasin, M. Deep Learning Framework for Modeling Cognitive Load from Small and Noisy EEG data. *IEEE Trans. Cogn. Dev. Syst.* **2022**. [[CrossRef](#)]
16. Havugimana, F.; Muhammad, M.B.; Moinudin, K.A.; Yeasin, M. Predicting Cognitive Load using Parameter-optimized CNN from Spatial-Spectral Representation of EEG Recordings. In Proceedings of the 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), Pasadena, CA, USA, 13–16 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 710–715.
17. Jogin, M.; Mohana; Madhulika, M.S.; Divya, G.D.; Meghana, R.K.; Apoorva, S. Feature extraction using convolution neural networks (CNN) and deep learning. In Proceedings of the 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 18–19 May 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 2319–2323.
18. Li, G.; Jiao, J.; Shi, G.; Ma, H.; Gu, L.; Tao, L. Fast Recognition of Pig Faces Based on Improved Yolov3. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2022; Volume 2171, p. 012005.
19. Mao, Q.C.; Sun, H.M.; Liu, Y.B.; Jia, R.S. Mini-YOLOv3: Real-time object detector for embedded applications. *IEEE Access* **2019**, *7*, 133529–133538. [[CrossRef](#)]
20. Sivamani, S.; Choi, S.H.; Lee, D.H.; Park, J.; Chon, S. Automatic posture detection of pigs on real-time using Yolo framework. *Int. J. Res. Trends Innov.* **2020**, *5*, 81–88.

21. Ju, M.; Choi, Y.; Seo, J.; Sa, J.; Lee, S.; Chung, Y.; Park, D. A Kinect-based segmentation of touching-pigs for real-time monitoring. *Sensors* **2018**, *18*, 1746. [[CrossRef](#)] [[PubMed](#)]
22. Almadani, I.; Abuhussein, M.; Robinson, A.L. Sow Localization in Thermal Images Using Gabor Filters. In *Advances in Information and Communication, Proceedings of the FICC2022, San Francisco, CA, USA, 3–4 March 2022*; Springer International Publishing: Cham, Switzerland, 2022.
23. Xue, H.; Chen, J.; Ding, Q.; Sun, Y.; Shen, M.; Liu, L.; Chen, X.; Zhou, J. Automatic detection of sow posture and estrus based on convolutional neural network. *Front. Phys.* **2022**, *10*, 1037129. [[CrossRef](#)]
24. Xu, Z.; Sullivan, R.; Zhou, J.; Bromfield, C.; Lim, T.T.; Safranski, T.J.; Yan, Z. Detecting sow vulva size change around estrus using machine vision technology. *Smart Agric. Technol.* **2023**, *3*, 100090. [[CrossRef](#)]
25. Simões, V.G.; Lyazrhi, F.; Picard-Hagen, N.; Gayrard, V.; Martineau, G.P.; Waret-Szkuta, A. Variations in the vulvar temperature of sows during proestrus and estrus as determined by infrared thermography and its relation to ovulation. *Theriogenology* **2014**, *82*, 1080–1085. [[CrossRef](#)]
26. Cabib, D.; Lavi, M.; Gil, A.; Milman, U. Long wave infrared (8 to 14 microns) hyperspectral imager based on an uncooled thermal camera and the traditional CI block interferometer (SI-LWIR-UC). In *Proceedings of the Infrared Technology and Applications. International Society for Optics and Photonics, Orlando, FL, USA, 25–29 April 2011*; Volume 8012, p. 80123H.
27. Shen, F.; Zeng, G. Semantic image segmentation via guidance of image classification. *Neurocomputing* **2019**, *330*, 259–266. [[CrossRef](#)]
28. Jadon, S. A survey of loss functions for semantic segmentation. In *Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Via del Mar, Chile, 27–29 October 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 1–7.
29. Freeman, H.; Shapira, R. Determining the minimum-area encasing rectangle for an arbitrary closed curve. *Commun. ACM* **1975**, *18*, 409–413. [[CrossRef](#)]
30. GitHub-Qengineering/YoloV8-ncnn-Raspberry-Pi-4: YoloV8 for a Bare Raspberry Pi 4—github.com. Available online: <https://github.com/Qengineering/YoloV8-ncnn-Raspberry-Pi-4> (accessed on 26 April 2024).

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.