agriculture

MDPI

Article

Dead Broiler Detection and Segmentation Using Transformer-Based Dual Stream Network

Gyu-Sung Ham [1] and Kanghan Oh [2,*]

[1] AI Convergence Research Institute, Wonkwang University, Iksan 54538, Republic of Korea; ham1231@wku.ac.kr
[2] Department of Computer and Software Engineering, Wonkwang University, Iksan 54538, Republic of Korea
* Correspondence: khoh888@wku.ac.kr; Tel.: +82-63-850-6897

Abstract: Improving productivity in industrial farming is crucial for precision agriculture, particularly in the broiler breeding sector, where swift identification of dead broilers is vital for preventing disease outbreaks and minimizing financial losses. Traditionally, the detection process relies on manual identification by farmers, which is both labor-intensive and inefficient. Recent advances in computer vision and deep learning have resulted in promising automatic dead broiler detection systems. In this study, we present an automatic detection and segmentation system for dead broilers that uses transformer-based dual-stream networks. The proposed dual-stream method comprises two streams that reflect the segmentation and detection networks. In our approach, the detection network supplies location-based features of dead broilers to the segmentation network, aiding in the prevention of live broiler mis-segmentation. This integration allows for more accurate identification and segmentation of dead broilers within the farm environment. Additionally, we utilized the self-attention mechanism of the transformer to uncover high-level relationships among the features, thereby enhancing the overall accuracy and robustness. Experiments indicated that the proposed approach achieved an average IoU of 88% on the test set, indicating its strong detection capabilities and precise segmentation of dead broilers.

Keywords: dead broiler segmentation; CNN; deep learning; precision agriculture

Citation: Ham, G.-S.; Oh, K. Dead Broiler Detection and Segmentation Using Transformer-Based Dual Stream Network. *Agriculture* **2024**, *14*, 2082. https://doi.org/10.3390/agriculture14112082

Academic Editor: Hai Lin

Received: 15 October 2024
Revised: 4 November 2024
Accepted: 16 November 2024
Published: 19 November 2024

1. Introduction

Precision agriculture has emerged as a crucial framework for industrial farming and is crucial in boosting productivity. This field is increasingly embracing automation technologies to reduce labor requirements and leverage advanced computer vision methods to achieve high efficiency in diverse applications [1,2].

The poultry industry has adopted precision agriculture tools, such as animal wearables, computer vision, and other sensing technologies, for the real-time tracking of animal conditions to boost efficiency and reduce costs. In particular, computer vision is considered a game changer, offering solutions for various needs, such as poultry house management [3,4], early disease detection [5–7], weight monitoring [8–10], carcass evaluation [11,12], and egg quality analysis [13]. These technologies ensure better monitoring of health and welfare [14,15] and drastically reduce the manual labor required for routine operations.

Recent research has increasingly focused on broiler chickens as a vital food source [5,16,17]. In broiler farms, rapid detection and removal of dead chickens is crucial for maintaining animal welfare, minimizing the spread of disease, and optimizing production efficiency. Traditional manual methods for detecting dead broilers are labor-intensive, time-consuming, and prone to human errors. This often leads to delays in identifying and removing dead chickens, increasing the risk of contamination and the spread of infection among flocks [6]. An automated dead broiler detection and segmentation system can significantly reduce these risks by providing real-time monitoring and alerts, ensuring a healthier environment

for birds and better productivity for the farm. Implementing such systems can minimize costs by reducing labor requirements and minimizing potential losses owing to disease outbreaks. Specifically, the automated segmentation results for dead broilers provided more detailed information. These detailed data are invaluable for further precise analyses, offering insights that can be used for future improvements and management strategies on broiler farms. Overall, automated detection systems are essential for modern broiler farm management to promote economic and ethical benefits.

Recent advances in computer vision have significantly affected the agricultural sector by offering new solutions to traditional challenges. This technology is valued for its ability to replicate human visual capabilities and affordability. In particular, convolutional neural networks (CNNs) are crucial in achieving remarkable accuracy in image recognition [18], surpassing human performance in certain cases. These networks excel in image classification, object detection [19], and image segmentation [20,21]. Therefore, CNNs are highly valuable because of their ability to handle complex visual tasks, which has led to their widespread use across academic and commercial sectors.

Recently, computer vision and CNN have been widely used to study broiler behavior [1,6,22–24]. A summary of these studies is provided in Table 1. The YOLO v4-based automated system [1] was proposed to detect and remove dead chickens in poultry houses. It operates in both remote-controlled and fully automatic modes, effectively reducing human–poultry contact while maintaining the efficient removal of deceased chickens. van der Eijk et al. [22] explored the use of computer vision algorithms, specifically, Mask R-CNN [25], to automatically detect individual broilers and monitor their resource use in both experimental and commercial settings. The developed models demonstrated high accuracy in identifying broilers and tracking their interactions with resources, such as feeders, bales, and perches. Yang et al. [23] addressed the challenge of detecting dead hens in caged environments by improving the YOLOv7 model. These enhancements include using a convolutional block attention module (CBAM) for improved feature extraction. The system integrates edge devices and inspection robots to achieve high detection accuracy in real-world farms. Bao et al. [24] proposed an AI-based sensor-detection method for identifying dead and sick chickens on farms, addressing the limitations of manual inspection. This method uses foot rings to measure chicken movement and calculates the three-dimensional variance in activity intensity. Machine learning algorithms analyze data to determine the health status of chickens. Okinda et al. [17] developed a machine vision system for the early detection and prediction of diseases in broiler chickens using a non-intrusive depth camera. It monitors posture and mobility changes and extracts features such as shape descriptors and walking speed to identify the health status. Hao et al. [6] developed a detection system for identifying dead broilers in stacked cage environments using an autonomous wheeled vehicle. This vehicle navigates through the cement aisles of a broiler house equipped with camera sensors that capture images of the broilers within the cages. A YOLOv3-based model was then applied to analyze the side-view images and accurately detect dead broilers. Li et al. [26] presented a method for detecting sick laying hens by using infrared thermal imaging combined with deep learning. It utilizes CNN to identify areas of interest (head, body, and legs) from thermal images and extracts temperature data. Massari et al. [27] explored the use of computer vision-based indices to monitor the broiler chicken responses to different rearing environments, particularly under thermal comfort and heat stress conditions. It introduces two indices, cluster and unrest, to analyze movement patterns and behavior in enriched versus non-enriched environments. This system utilizes video analysis to assess how environmental factors, such as enrichment, affect bird activity and well-being.

Table 1. Comparative overview of studies on automated detection and monitoring in poultry farming.

Study	Year	Proposed Architecture	Dataset Used	Result
Liu et al. [1]	2021	YOLO v4-based automated chicken removal system	Chicken Mortality Detection Dataset	Accurately identifies and removes dead chickens, reducing human contact and improving biosecurity.
J. A. J. et al. [22]	2022	Mask R-CNN with zone-based classifiers for resource monitoring	Custom dataset from research facility and commercial farm	Achieved accurate broiler detection and effective monitoring of resource use in various zones (feeders, bales, perches, etc.).
Yang et al. [23]	2024	Enhanced YOLOv7 with CBAM	Caged hen farm images	Achieved high accuracy for dead hen detection, optimized for mobile deployment
Bao et al. [24]	2021	Machine learning-based sensor network	Custom sensor data from chicken farms	Effectively detects dead and sick chickens, enhancing automation in large-scale farms.
Okinda et al. [17]	2019	Machine vision with RBF-SVM classifier	Video and depth camera data	Enables automated, early disease detection in broilers.
Hao et al. [6]	2022	Improved YOLOv3 with SPP and CIoU loss	Custom broiler farm dataset	Achieved effective dead broiler detection in stacked cages, enhancing farm inspection automation.
Li et al. [26]	2021	CNN with infrared thermal imaging	Infrared images of laying hens	Detects sick hens by analyzing temperature patterns, enabling early identification.
Massari et al. [27]	2022	Computer vision with cluster and unrest indices	Video data from controlled environment	Demonstrated effectiveness in monitoring broiler movement and detecting heat stress, highlighting environmental impact on behavior.

Inspired by previous studies, we propose a machine-learning-based approach for detecting and segmenting dead broilers. The proposed network was built on a dual-stream framework consisting of separate detection and segmentation modules. A dual-stream network is a type of deep learning architecture that uses two parallel pathways to simultaneously process different aspects of input data [28,29]. Each stream was tailored to handle a specific task or feature extraction process, allowing the model to learn complementary information. For instance, in a detection and segmentation task, one stream focuses on detecting the presence and location of objects, whereas the other refines the segmentation of object regions. This design enables the network to efficiently combine detection accuracy with precise boundary delineation for more robust performance. The proposed method aims to segment regions containing dead broilers. Although the initial detection results of dead broilers are necessary for the segmentation task, this method was designed to perform both detection and segmentation in parallel. The detection network identified the locations of dead broilers and provided a segmentation network with information to exclude live birds. We employed heatmap regression techniques [30] to highlight the regions of dead broiler chickens. Heatmap regression is a computer vision method that generates a spatial heat map representing the probability of a target's presence at each pixel location. In this approach, Gaussian distributions were overlaid on heat maps at the locations of dead broilers, creating smooth peaks that correspond to their positions. This allows the network to identify and focus on the most likely areas where dead broilers are located, thereby enhancing detection accuracy. In addition, we successfully integrated the self-attention mechanism into the final part of the encoder. The self-attention mechanism allows the model to weigh the importance of different spatial features and capture long-range dependencies and interactions [31]. This is particularly beneficial for broiler detection networks

because it enables the model to better differentiate between dead and live broilers by focusing on subtle and contextually significant cues in crowded or complex environments.

The structure of this paper is organized as follows: Section 2 describes the materials and the proposed method in detail. Section 3 presents the experimental results and an in-depth discussion of the results. Finally, Section 4 summarizes the conclusions drawn from this study and suggests directions for future research.

2. Materials and Methods

2.1. Dataset Collection and Description

We used the publicly available dead broiler detection and segmentation datasets provided in the study [32]. This dataset contained 86 instances of dead broilers and included regions of interest (ROIs) for their locations. For the segmentation task, we manually generated binary segmentation masks with three different individuals and labeled each sample. The dataset was split into training, validation, and test sets, with approximately 71.4%, 17.1%, and 11.4% of the data, respectively. The final mask is selected by averaging the annotations. For the detection task, we placed a Gaussian mask at the center of each dead broiler. This Gaussian mask acted as a heat map, highlighting the most likely location of the broiler. The intensity of the Gaussian distribution peaks at the center of the target and gradually decreases toward the edges, effectively guiding the detection network toward the most relevant regions. Figure 1 provides an overview of the dataset used for training and evaluating the dead-broiler detection models. The first row shows raw images collected under different environmental conditions, illustrating the variability in the appearance of dead broilers. The second row contains binary masks that accurately outline the regions where dead broilers were present, serving as the ground truth for the segmentation tasks. The third row shows Gaussian heat maps centered on the locations of dead broilers, which were used to guide the models during training to better localize the target areas. Owing to the limited dataset of 86 samples, we conducted a 5-fold cross-validation to evaluate the performance of the model and used the average of the results as the final performance metric.

Figure 1. Examples of dead broiler dataset. From top to bottom: images of dead broilers, GT masks, and Gaussian heatmaps centered around the locations of the dead broilers.

2.2. Methodology

2.2.1. Overall Architecture

An overview of the proposed method is presented in Figure 2. This figure illustrates the architecture of the proposed dual-task network for dead-broiler detection and segmentation. The input image was first passed through an encoder consisting of multiple convolutional layers with 3 × 3 kernels, instance normalization (IN), ReLU activation, and 2 × 2 Max Pooling operations. The encoded features are then fed into a transformer module that captures long-range dependencies and relationships between the features. Following the transformer, the network is split into two branches: one for the segmentation task and the other for heatmap regression. The segmentation branch uses a series of convolutional and upsampling layers to generate a binary mask for the dead broilers, optimized using Dice Loss. The heatmap regression branch produced a Gaussian heatmap centered on the detected location of dead broilers, optimized using L2 Loss. In addition, the features from the heatmap regression decoder were merged with the segmentation layers to provide an additional spatial context that helped refine the segmentation results. Heatmap regression is intended mainly as a supportive feature rather than a core output. Thus, we adopted a one-way feature-sharing approach, where information flows from the heatmap regression branch to the segmentation branch only and not in the reverse direction. This design choice ensures that the segmentation network receives priority, allowing it to isolate dead broiler areas more accurately. Feature concatenation and skip connections were employed to preserve spatial information throughout the decoding process, thereby contributing to improved accuracy in both tasks.

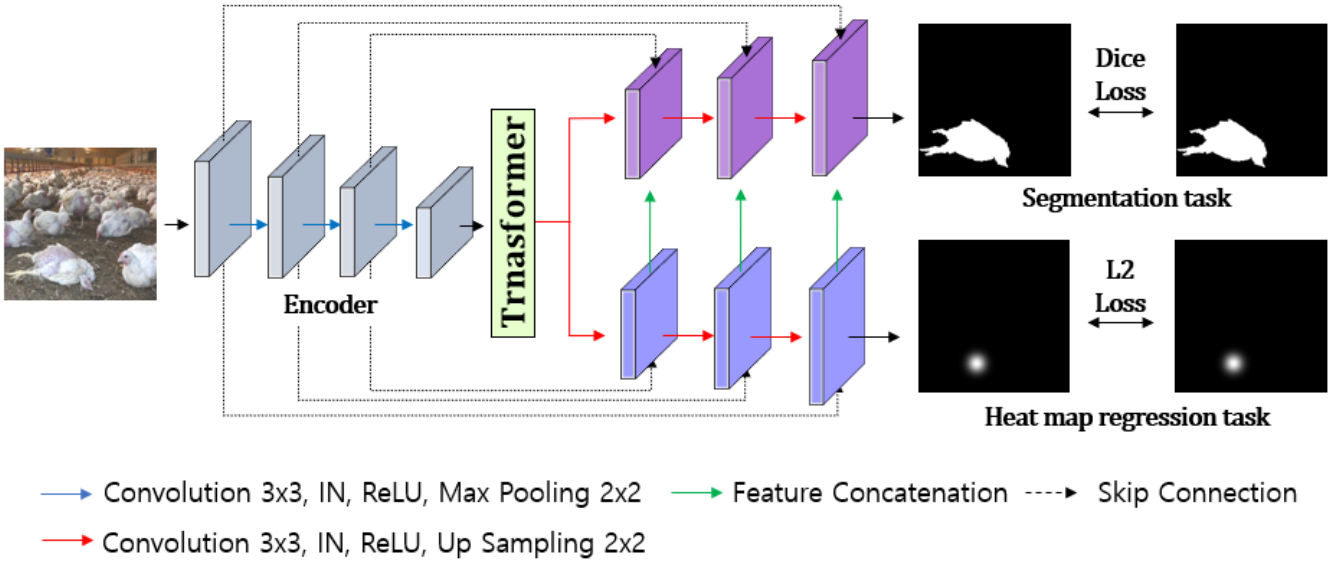

Figure 2. Proposed dual-stream network for the detection and segmentation of dead broiler.

During the encoding phase, the network progressively increased the number of feature maps from 64 to 512, intensifying its ability to capture and process more complex features from the input images. Conversely, in the decoding phase, the feature maps were gradually reduced from 512 to 64, allowing the network to reconstruct finer details while retaining essential information from the encoded features. The segmentation decoder then receives additional feature information from the detection decoder. Specifically, at each corresponding layer of the segmentation decoder, feature maps from the detection decoder are concatenated to provide spatial and contextual cues that assist in refining the segmentation process. For instance, if the detection decoder outputs a feature map with 128 channels in a particular layer, and the segmentation decoder has 128 channels in the corresponding layer, the concatenation results in a feature map with 256 channels. The additional feature channels are calculated as follows: if the detection decoder outputs a

feature with 64, 128, 256, and 512 channels at various stages, the segmentation decoder at each corresponding stage will double its channels owing to concatenation. Thus, the input channels for the segmentation decoder were 128, 256, 512, and 1024.

2.2.2. Transformer Module

In this section, we employ a transformer block to learn the contextual relationships between the tillage boundary points extracted from the soft Argmax function. Transformers [31] are a type of neural network architecture that has become foundational in various fields of machine learning, including natural language processing and computer vision. Transformers utilize a multi-layered structure that incorporates Self-Attention mechanisms, allowing each input element to interact with others in the sequence. This enables the model to capture contextual relationships among features more effectively, making it especially useful for computer vision tasks that involve complex spatial patterns. In our study, we leverage the Transformer's ability to accurately predict the locations of dead broilers, incorporating these predictions into the segmentation process.

Figure 3 illustrates the architecture of a transformer block utilized to recalibrate the encoded feature maps, which is crucial in learning the relationships among different spatial regions of an image. The input to the transformer block is an encoded feature map of dimensions $c \times h \times w$, where c represents the number of channels and h and w correspond to the height and width, respectively. The encoded features were first reshaped into a sequence of length $hw \times c$ and normalized using layer normalization [33] (LN):

$$\mathrm{LN}(x) = \frac{x - \mu}{\sigma} \times \gamma + \beta \tag{1}$$

where μ denotes the mean of the input vector x, and σ is the standard deviation. The learnable parameters γ and β control the scaling and shifting of the normalized output, respectively.

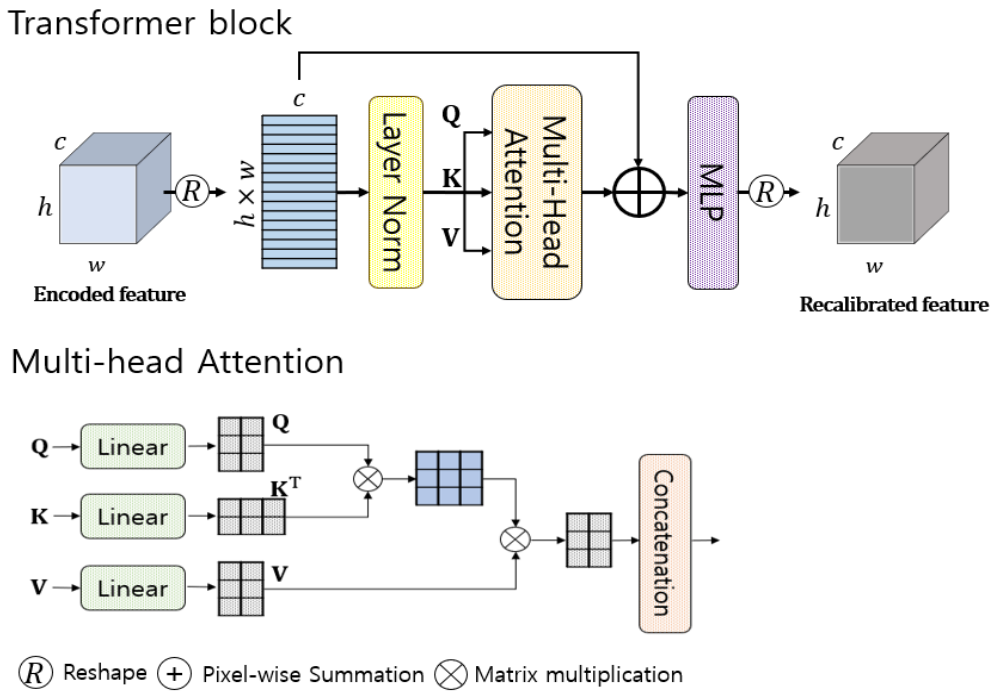

Figure 3. Overview of the transformer block with multi-head attention. The figure illustrates the process of recalibrating encoded features using a transformer block, which includes layer normalization, multi-head attention, and an MLP (multi-layer perceptron).

The transformer block uses multi-head attention to capture dependencies among different spatial regions within the feature map. Queries (Q), keys (K), and values (V) are derived from the normalized features using linear transformations [31]. The multi-head attention mechanism calculates the attention scores by performing a scaled dot product, as defined by the following:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

where $\sqrt{d_k}$ is a scaling factor that stabilizes the gradients. The formulas $Q = XW^Q$, $K = XW^K$, and $V = XW^V$ define the transformations applied to input feature X using the query, key, and value weight matrices, respectively. Each head i computes its own set of queries, keys, and values, as follows:

$$head_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{3}$$

In multi-head attention, the outputs from all heads are concatenated and projected using a weight matrix W^O:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_\text{n})W^O \tag{4}$$

This multi-head approach enables the model to focus on different aspects of the input simultaneously, capturing richer feature representations and improving its ability to understand complex relationships within the data. This is particularly useful for distinguishing subtle differences between live and dead broiler chickens. The ability to focus on the distant parts of an image allows the model to recognize the presence of dead broilers, even when they are partially hidden or surrounded by live broilers. A residual connection integrates the attention output with the original input features, followed by a multi-layer perceptron (MLP) that refines the recalibrated feature map. The output was reshaped into dimensions $c \times h \times w$, producing a feature map that encoded both local and global contextual information, thereby enhancing the ability of the model.

2.2.3. Joint Loss Function

To optimize the model for both the segmentation and heatmap regression tasks, we employed a joint loss function that combined the Dice loss and weighted L2 loss:

$$\mathcal{L}_{\text{Joint}} = \mathcal{L}_{\text{Dice}} + \gamma\mathcal{L}_{\text{L2}} \tag{5}$$

where $\mathcal{L}_{\text{Dice}}$ represents the Dice loss, $\mathcal{L}_{\text{L2}}$ is the L2 loss used for heatmap regression and γ is a weight parameter that controls the influence of the L2 loss on the overall optimization process. In our experiments, we set $\gamma = 0.1$, ensuring that the primary focus remains on the segmentation task while still incorporating valuable spatial information from the heatmap regression. This balance leads to an improved overall performance in detecting and segmenting the regions of dead broilers.

2.2.4. Implementation Detail

Experiments were performed using a system equipped with an Intel Core i9-10900X CPU at 3.70 GHz, 48 GB RAM, and an Nvidia GeForce RTX 3090 GPU. We used PyTorch version 1.8.0) to train and test the network. The input images were resized to 224 × 224 pixels to streamline processing and reduce computational overhead. The training process utilized the Adam optimizer, starting with an initial learning rate of 0.0001, which was multiplied by 0.1 every 200 epochs over a total of 600 epochs. To enhance model robustness and generalization, we applied data augmentation techniques such as brightness adjustments, rotations ranging from −25 to 25 degrees, and horizontal and vertical shifts. In the transformer block, we set the number of attention heads to four, with each head having an

embedding vector with a size of 64. In the heat map regression task, the Gaussian scale score was set to 15. A Gaussian heatmap was used to indicate the location of dead broilers without incorporating additional information about size or shape. With this approach, our goal is not to generate highly precise Gaussians but to produce a general representation of the target locations.

2.3. Evaluation Metrics

The proposed method was evaluated using four metrics: the intersection over union (IoU), precision, recall, and F-measure. As previously mentioned, our primary goal was to evaluate the segmentation performance of the model. The IoU measures the overlap between the predicted segmentation *Pred* and ground truth *GT*. It is defined as the ratio of the intersection area to the union area of *Pred* and *GT*:

$$Pred = \frac{|Pred \cap GT|}{|Pred \cup GT|} \tag{6}$$

Precision measures the accuracy of positive predictions and is calculated as the ratio of true-positive pixels to all pixels predicted to be positive.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{7}$$

where *TP* (True Positives TPs) refer to correctly predicted pixels, while *FP* (False Positives FPs) refer to pixels incorrectly identified as positive. The recall evaluates the proportion of actual positives correctly identified by the model.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{8}$$

False Negatives (FNs) represent the actual positive pixels that the model failed to predict. The F-measure is a balanced metric that combines precision and recall into a single value:

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

The F-measure ranges from 0 to 1, with values closer to 1 indicating a better balance between precision and recall rates. To evaluate the performance, we applied a threshold of 0.5 to the model's output and converted the segmentation results into binary masks.

3. Results and Discussion

3.1. Performance Comparison

In previous studies, detection methods such as YOLO primarily focused on predicting bounding boxes to identify dead broilers, making direct comparisons with segmentation-focused approaches challenging. Since our method outputs segmentation maps, we selected representative segmentation models—U-Net [20], FCN [34], LinkNet [35], and DeepLabV3 [21]—to benchmark our performance. These models are well-established for segmentation tasks, providing a more relevant comparison. All models were trained and tested using a 5-fold cross-validation process, and the results reflect the average performance across the folds. Table 2 presents the segmentation performance of the proposed method alongside four other methods widely used models, using key metrics such as IoU, precision, recall, and F-measure. Figure 4 shows the training and validation loss of our model over epochs.

Table 2. Comparison of segmentation performance between the proposed method and existing methods.

Method	Metrics (Standard Deviation)			
	IOU	Precision	Recall	F-Measure
U-Net	83.86 (0.69)	87.76 (0.67)	89.13 (0.68)	88.17 (0.99)
FCN	82.94 (1.46)	82.50 (0.54)	91.98 (0.59)	86.15 (0.69)
LinkNet	82.78 (1.52)	82.73 (0.49)	89.43 (1.36)	85.64 (0.78)
DeepLabV3	84.53 (0.88)	91.39 (0.78)	89.61 (0.87)	89.97 (0.87)
Proposed method	85.58 (1.41)	90.60 (1.06)	93.76 (0.53)	92.05 (1.14)

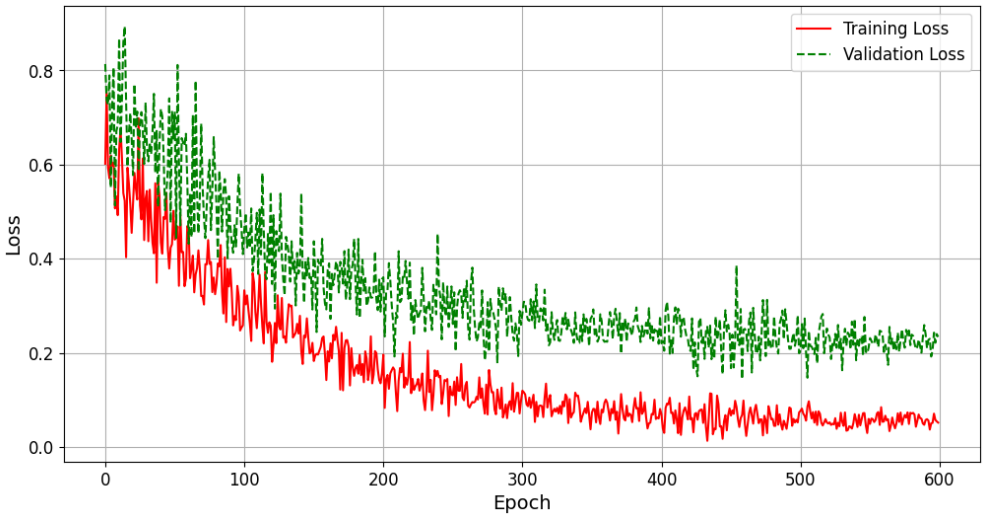

Figure 4. Training and validation loss graph of our model.

The proposed method demonstrates superior performance in terms of IoU, achieving 85.58%, which surpasses DeepLabV3's 84.53% and shows a significant improvement over U-Net, FCN, and LinkNet. In terms of precision, the proposed method reaches 90.60%, which is slightly lower than the highest precision achieved by DeepLabV3 at 91.39%. However, the proposed method excels in recall, obtaining a score of 93.76%, notably higher than all other methods. The balance between precision and recall is summarized by the F-measure, where the proposed method outperforms the existing methods. These results highlight the capability of the proposed approach to achieve more consistent and reliable segmentation across various conditions. In Figure 5, the box plots illustrate the distribution of segmentation performance metrics (IoU, precision, recall, and F-measure) for each method, including U-Net, FCN, LinkNet, DeepLabV3, and the proposed method. The proposed method demonstrates a higher median and tighter distribution across most metrics, indicating more consistent and superior performance. Table 3 presents the performance comparison of the proposed model across different numbers of attention heads in the transformer. As the number of heads increases, there is a general improvement in IoU. Starting from 84.82 with one head, the IoU improves to 85.58 with four heads. The F-measure, which balances precision and recall, also shows improvement as the number of heads increases. The value starts at 90.33 with one head, increases gradually, and reaches 92.05 with four heads. Overall, the results suggest that increasing the number of attention heads in the transformer architecture enhances the ability of the model to understand complex spatial relationships, leading to better segmentation performance. Table 4 presents the performance comparison between different configurations of the proposed method, including variations without heatmap regression and feature sharing. The results indicate that the full version of the proposed method, which incorporates both heatmap regression and feature sharing, achieves the highest overall performance, with an IoU of 85.58% and an F-measure of 92.05%. The absence of the heatmap regression layer results in a slight

decrease in IoU and Recall, suggesting that the spatial guidance provided by the heatmap aids in better segmentation of target regions. When feature sharing is removed, the performance drops slightly. Overall, the results emphasize the beneficial role of the heatmap regression layer in improving the segmentation performance of the proposed method.

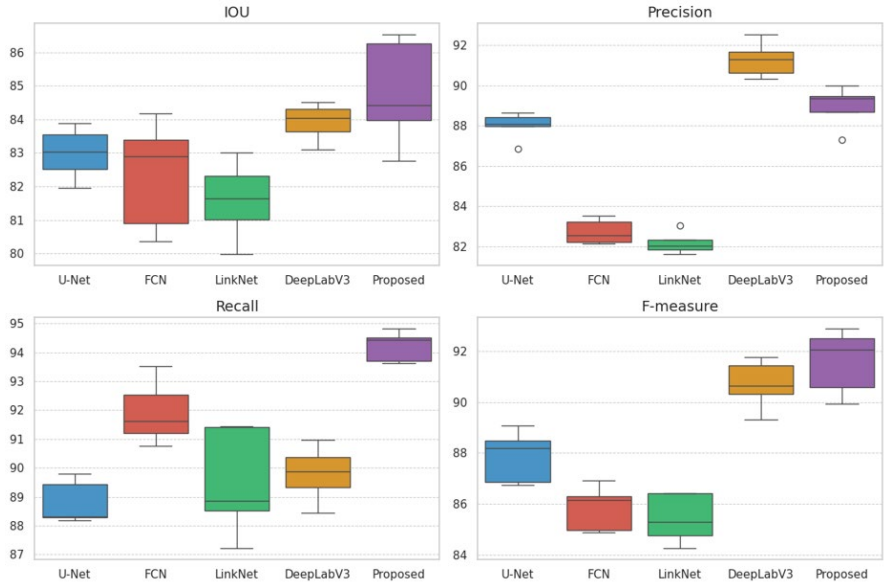

Figure 5. Box plots display the distribution of performance metrics (IoU, Precision, Recall, and F-measure) for each segmentation method (U-Net, FCN, LinkNet, DeepLabV3, and the Proposed method).

Table 3. Performance comparison according to the number of heads in the transformer.

Number of Heads	Metrics (Standard Deviation)			
	IOU	Precision	Recall	F-Measure
1	84.82 (0.55)	88.23 (0.77)	93.43 (1.06)	90.33 (0.81)
2	85.04 (1.32)	87.58 (0.39)	92.58 (1.06)	90.27 (0.81)
3	84.73 (0.42)	89.25 (1.32)	94.51 (1.19)	91.36 (0.52)
4	85.58 (1.41)	90.60 (1.06)	93.76 (0.53)	92.05 (1.14)

Table 4. Performance comparison of different variations of the proposed method, including the impact of heatmap regression and feature sharing.

Method	Metrics (Standard Deviation)			
	IOU	Precision	Recall	F-Measure
Proposed method without heatmap regression	84.88 (1.66)	91.11 (1.13)	89.44 (0.68)	91.33 (1.33)
Proposed method without feature sharing	85.13 (1.33)	89.85 (0.99)	94.45 (0.63)	91.53 (0.86)
Proposed method	85.58 (1.41)	90.60 (1.06)	93.76 (0.53)	92.05 (1.14)

3.2. Visualization Results

Figure 6 shows the segmentation results of the proposed method. The top row contains examples of single-dead broilers, demonstrating the ability of the model to accurately detect isolated instances. The bottom row, on the other hand, presents more complex scenarios where the dead broilers are surrounded by other broilers. Despite the complexity of these scenarios, the proposed method effectively distinguished between the live and dead broilers. The ability to separate live from deceased animals is critical for practical applications, ensuring accurate detection in crowded farm environments. Figure 7

presents the visualization results of the proposed method for segmenting dead broilers and generating heat maps. The first column shows the original input images of the broilers. The second column presents the ground truth segmentation masks (GT-seg), which serve as a reference for evaluating the segmentation accuracy. The third column displays the segmentation results produced by the model (Output-seg), which closely align with the ground-truth masks, pointing to an accurate segmentation. Although the output heat maps did not perfectly match the ground truth, they effectively highlighted relevant areas. This is because our approach focuses more on obtaining accurate segmentation results than on making the heat maps appear the same as the ground truth. The heat maps assist the model in focusing on dead broiler regions, thus supporting better segmentation performance, even if they do not exactly mirror the ground truth. Overall, the figure demonstrates the effectiveness of the proposed method in achieving precise segmentation and producing heat maps that support the accurate localization of dead broilers. Further emphasizing the balance between segmentation accuracy and practical guidance through heat maps.

Figure 6. Visualization results of the proposed method. Blue outlines indicate the predicted segmentation boundaries, while green outlines represent the ground truth (GT).

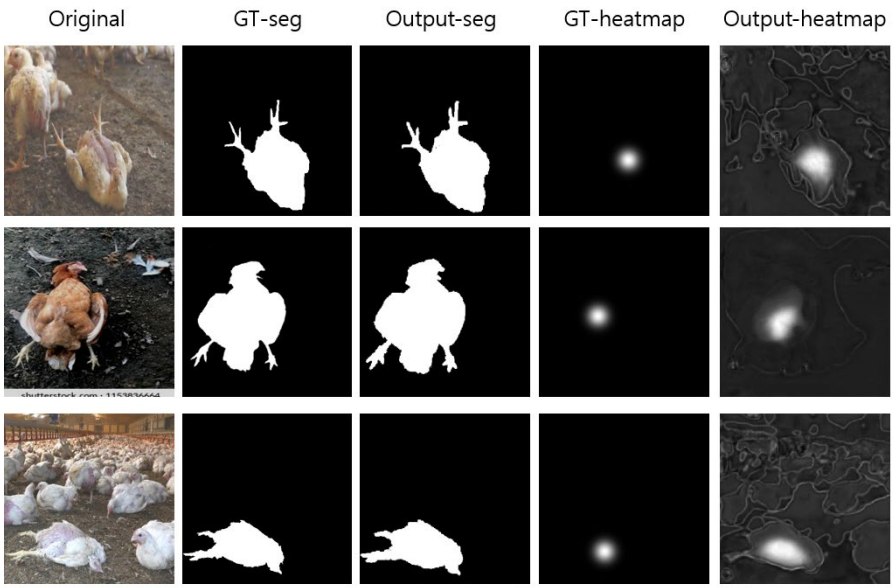

Figure 7. Comparison of segmentation and heatmap results. From left to right: original images, ground truth segmentation (GT-seg), output segmentation (Output-seg), ground truth heatmap (GT-heatmap), and output heatmap (Output-heatmap). Output heatmaps guide the model's focus, enhancing segmentation accuracy.

3.3. Discussion

In our current dataset, there were no occluded chickens, which could limit the model's performance in real-world applications where occlusion is frequent. However, if we include data with occluded chickens during training, we expect the model could better learn to predict occluded instances. Additionally, our methodology employs a dual-stream approach, where heatmaps provide spatial guidance by predicting the locations of dead chickens. This design could aid the model in detecting occluded regions, potentially improving segmentation accuracy even under challenging conditions. Future work could incorporate occluded chicken data to further validate and enhance the robustness of the proposed method.

While the proposed system generally performs well, certain configurations may lead to suboptimal results. For instance, when the influence of the heatmap is set too high, the model may become overly reliant on the heatmap guidance, causing false positives in regions with similar textures or patterns. Conversely, reducing the heatmap's influence can lead to missed detections, as the model may not be adequately guided to the target areas. These cases highlight the importance of carefully tuning the heatmap parameters to balance detection accuracy and robustness.

4. Conclusions

In this study, we introduce a novel automatic detection and segmentation system for identifying dead broilers using a transformer-based dual-stream network. The dual-stream design integrates detection and segmentation networks, where the detection network provides valuable location feature information to the segmentation network. This collaborative structure effectively prevents the mis-segmentation of live broilers, ensuring that dead broilers are accurately identified and segmented. Furthermore, the system leverages the self-attention mechanism of transformers to capture the complex relationships between the features. This feature allows the network to analyze high-level spatial relationships, making it particularly effective in distinguishing dead broilers from their surroundings. Experiments indicate that the proposed method showed a competitive performance with existing methods, proving its effectiveness in the precise segmentation of dead broilers. In future work, we plan to expand our experiments to larger datasets, evaluate the system under more challenging conditions, and enhance the precision of the segmentation results. The contributions of this study are as follows:

1. We proposed a novel dual-stream architecture that simultaneously handles the detection and segmentation of dead broilers. The detection module located potential areas containing dead broilers, and the segmentation module refined these identified areas, enabling precise boundary delineation.
2. To improve the ability of the model to capture long-range dependencies and contextual information, we integrated self-attention layers into the network architecture. This enabled a better understanding of the spatial relationships between the detected dead broilers and the overall scene context, enhancing the robustness of the model in complex environments.
3. This method utilizes heat map regression to map the probable locations of dead broilers by overlaying Gaussian distributions at these points. This approach provides a refined method for prioritizing critical areas during detection and segmentation.

Author Contributions: Conceptualization, G.-S.H. and K.O.; methodology, K.O.; software, G.-S.H. and K.O.; validation, G.-S.H. and K.O.; formal analysis, G.-S.H. and K.O.; investigation, G.-S.H. and K.O.; writing—original draft preparation, G.-S.H. and K.O.; writing—review and editing, K.O.; visualization, K.O.; supervision, K.O.; project administration, K.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) and Korea Smart Farm R&D Foundation (KoSFarm) through Smart Farm Innovation Technology Development Program, funded by Ministry of Agriculture,

Food and Rural Affairs (MAFRA), Ministry of Science and ICT (MSIT), and Rural Development Administration (RDA) (grant number: RS-2021-IP421024).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Wayne, B. Dead Chickens Dataset. Universe by Roboflow [32]. Available online: https://universe.roboflow.com/bruce-wayne-wja03/dead-chikens (accessed on 11 October 2024).

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

References

1. Liu, H.-W.; Chen, C.-H.; Tsai, Y.-C.; Hsieh, K.-W.; Lin, H.-T. Identifying Images of Dead Chickens with a Chicken Removal System Integrated with a Deep Learning Algorithm. *Sensors* **2021**, *21*, 3579. [CrossRef] [PubMed]
2. Abdalla, A.; Cen, H.; Wan, L.; Mehmood, K.; He, Y. Nutrient Status Diagnosis of Infield Oilseed Rape via Deep Learning-Enabled Dynamic Model. *IEEE Trans. Ind. Inform.* **2020**, *17*, 4379–4389. [CrossRef]
3. Abdanan Mehdizadeh, S.; Neves, D.P.; Tscharke, M.; Nääs, I.A.; Banhazi, T.M. Image Analysis Method to Evaluate Beak and Head Motion of Broiler Chickens During Feeding. *Comput. Electron. Agric.* **2015**, *114*, 88–95. [CrossRef]
4. Pereira, D.F.; Miyamoto, B.C.B.; Maia, G.D.N.; Sales, G.T.; Magalhães, M.M.; Gates, R.S. Machine Vision to Identify Broiler Breeder Behavior. *Comput. Electron. Agric.* **2013**, *99*, 194–199. [CrossRef]
5. Zhuang, X.; Zhang, T. Detection of Sick Broilers by Digital Image Processing and Deep Learning. *Biosyst. Eng.* **2019**, *179*, 106–116. [CrossRef]
6. Hao, H.; Fang, P.; Duan, E.; Yang, Z.; Wang, L.; Wang, H. A Dead Broiler Inspection System for Large-Scale Breeding Farms Based on Deep Learning. *Agriculture* **2022**, *12*, 1176. [CrossRef]
7. Bist, R.B.; Yang, X.; Subedi, S.; Chai, L. Mislaying Behavior Detection in Cage-Free Hens with Deep Learning Technologies. *Poult. Sci.* **2023**, *102*, 102729. [CrossRef]
8. Mollah, M.B.R.; Hasan, M.A.; Salam, M.A.; Ali, M.A. Digital Image Analysis to Estimate the Live Weight of Broiler. *Comput. Electron. Agric.* **2010**, *72*, 48–52. [CrossRef]
9. Mortensen, A.K.; Lisouski, P.; Ahrendt, P. Weight Prediction of Broiler Chickens Using 3D Computer Vision. *Comput. Electron. Agric.* **2016**, *123*, 319–326. [CrossRef]
10. Amraei, S.; Mehdizadeh, S.A.; Nääs, I.D.A. Development of a Transfer Function for Weight Prediction of Live Broiler Chicken Using Machine Vision. *Eng. Agric.* **2018**, *38*, 776–782. [CrossRef]
11. Ye, C.-W.; Yu, Z.-W.; Kang, R.; Yousaf, K.; Qi, C.; Chen, K.-J.; Huang, Y.-P. An Experimental Study of Stunned State Detection for Broiler Chickens Using an Improved Convolution Neural Network Algorithm. *Comput. Electron. Agric.* **2020**, *170*, 105284. [CrossRef]
12. Mansor, M.A.; Baki, S.R.M.S.; Tahir, N.M.; Rahman, R.A. An Approach of Halal Poultry Meat Comparison Based on Mean-Shift Segmentation. In Proceedings of the 2013 IEEE Conference on Systems, Process & Control (ICSPC), Kuala Lumpur, Malaysia, 13–15 December 2013; pp. 279–282. [CrossRef]
13. Alon, A.S.; Technological Institute of the Philippines-Manila. An Image Processing Approach of Multiple Eggs' Quality Inspection. *Int. J. Adv. Trends Comput. Sci. Eng.* **2019**, *8*, 2794–2799. [CrossRef]
14. Neethirajan, S. Recent Advances in Wearable Sensors for Animal Health Management. *Sens. Bio-Sens. Res.* **2017**, *12*, 15–29. [CrossRef]
15. Neethirajan, S.; Tuteja, S.K.; Huang, S.T.; Kelton, D. Recent Advancement in Biosensors Technology for Animal and Livestock Health Management. *Biosens. Bioelectron.* **2017**, *98*, 398–407. [CrossRef]
16. Syauqi, M.N.; Zaffrie, M.M.A.; Hasnul, H.I. Broiler Industry in Malaysia. Available online: http://ap.fftc.agnet.org/files/ap_policy/532/532_1.pdf (accessed on 25 January 2018).
17. Okinda, C.; Lu, M.; Liu, L.; Nyalala, I.; Muneri, C.; Wang, J.; Zhang, H.; Shen, M. A Machine Vision System for Early Detection and Prediction of Sick Birds: A Broiler Chicken Model. *Biosyst. Eng.* **2019**, *188*, 229–242. [CrossRef]
18. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
19. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929. [CrossRef]
20. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III*; Springer International Publishing: Cham, Switzerland, 2015; Volume 18, pp. 234–241. [CrossRef]
21. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 801–818. [CrossRef]
22. van der Eijk, J.A.J.; Guzhva, O.; Voss, A.; Möller, M.; Giersberg, M.F.; Jacobs, L.; de Jong, I.C. Seeing Is Caring—Automated Assessment of Resource Use of Broilers with Computer Vision Techniques. *Front. Anim. Sci.* **2022**, *3*, 945534. [CrossRef]

23. Yang, J.; Zhang, T.; Fang, C.; Zheng, H.; Ma, C.; Wu, Z. A Detection Method for Dead Caged Hens Based on Improved YOLOv7. *Comput. Electron. Agric.* **2024**, *226*, 109388. [CrossRef]
24. Bao, Y.; Lu, H.; Zhao, Q.; Yang, Z.; Xu, W. Detection System of Dead and Sick Chickens in Large Scale Farms Based on Artificial Intelligence. *Math. Biosci. Eng.* **2021**, *18*, 6117–6135. [CrossRef]
25. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2961–2969. [CrossRef]
26. Li, P.; Zhang, T.; Wang, X.; Liu, J.; Huang, Y. Detection of Sick Laying Hens by Infrared Thermal Imaging and Deep Learning. *J. Phys. Conf. Ser.* **2021**, *2025*, 012008. [CrossRef]
27. Massari, J.M.; de Moura, D.J.; de Alencar Nääs, I.; Pereira, D.F.; Branco, T. Computer-Vision-Based Indexes for Analyzing Broiler Response to Rearing Environment: A Proof of Concept. *Animals* **2022**, *12*, 846. [CrossRef] [PubMed]
28. Qiu, Z.; Hu, Y.; Chen, X.; Zeng, D.; Hu, Q.; Liu, J. Rethinking Dual-Stream Super-Resolution Semantic Learning in Medical Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 451–464. [CrossRef] [PubMed]
29. Liu, L.; Yang, Z.; Li, G.; Wang, K.; Chen, T.; Lin, L. Aerial Images Meet Crowdsourced Trajectories: A New Approach to Robust Road Extraction. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 3308–3322. [CrossRef] [PubMed]
30. Ham, G.S.; Oh, K. Learning Spatial Configuration Feature for Landmark Localization in Hand X-rays. *Electronics* **2023**, *12*, 4038. [CrossRef]
31. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010. [CrossRef]
32. Wayne, B. Dead Chickens Dataset. *Universe by Roboflow*. Available online: https://universe.roboflow.com/bruce-wayne-wja03/dead-chikens (accessed on 11 October 2024).
33. Ba, J.L. Layer normalization. *arXiv* **2016**, arXiv:1607.06450. [CrossRef]
34. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [CrossRef]
35. Chaurasia, A.; Culurciello, E. LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation. In Proceedings of the IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2018; pp. 1–4. [CrossRef]