ESTIMATION OF BROILER CHICKEN MASS USING COMPUTER VISION WITH CONVOLUTIONAL NEURAL NETWORK

Igor de Souza Azevedo [1,2]; Leonardo Vian Erthal [1]; Rafael da Silveira Castro [3]; Aurélio Tergolina Salton [1];

[1] Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil

[2] Center for Embedded Devices and Research in Digital Agriculture (CEDRA), São Leopoldo, Brazil

[3] Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, Brazil

A paper from the Proceedings of the

17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture

13-15 July 2026

Porto Alegre, Brazil

Abstract.

In poultry farming, monitoring bird mass during rearing is crucial, as it enables farmers to adjust parameters such as feed supply and lighting to better control weight gain. However, the methods currently used in poultry houses, i.e. manual weighing or poultry scales, present drawbacks, including the inability to weigh a representative number of birds or the frequent maintenance required to keep the equipment clean. The present work aims to validate an alternative method for estimating the mass of Cobb500 broiler chickens using computer vision techniques based on convolutional neural networks, functioning as a visual scale. To this end, three datasets corresponding to three distinct flocks were collected in a poultry house, with the average mass of the chickens under a ceiling-mounted camera manually recorded up to twice per day. Using these datasets, two types of neural networks were trained through transfer learning, starting from the Xception architecture pretrained on the ImageNet dataset. First, a classification network selects suitable images, which are then processed by a regression network responsible for predicting the mean weight of the birds captured by the camera. Experiments demonstrated a root mean square error (RMSE) of 109.8 grams and a mean absolute error (MAE) of 82 grams when evaluated by the regression network trained on two flocks and tested on a third. To reduce fluctuations in the regression output, averages of daily measurements were evaluated, yielding a daily average MAE value of 48.5 grams. When tested using a camera positioned elsewhere in the poultry house, with images not represented in the training data, the model achieved RMSE, MAE and daily average MAE values of 135.5, 107.6 and 56.4 grams, respectively. Once fully implemented, the system enables daily monitoring of

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2026). Title of paper. In Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture and Brazilian Congress on Precision and Digital Agriculture.

broiler chicken's growth, providing insights into flock health and overall farming efficiency. It can also provide support for decision-making during the rearing process and by scheduling the flock removal based on the target market weight.

Keywords.
Mass Estimation, Precision Agriculture, Computer Vision, Convolutional Neural Network.

Introduction

In the field of animal production, one of the most important metrics is the animals' body mass and its variation over time, as this serves as an indicator of expected productivity or health issues. However, there are difficulties and complications associated with the manual collection of such data.

Specifically in the case of broiler chickens, the problems include: many birds in a single poultry house, making complete measurement unfeasible; difficulty in catching the birds; stress induced in the animals during the procedure; human errors during measurement, among others. There are solutions aimed at addressing these problems, primarily using floor-based scales so that the animals walk over them several times a day, but these solutions are also affected by external factors (Nyalala, et al., 2021).

In this context, research has emerged using computer vision to estimate body mass. The studies found regarding Convolutional Neural Networks (CNNs) with Deep Learning approach the problem by obtaining individual measurements of the animals. Some studies extract features such as perimeter and area with the aid of image segmentation (Szabo & Alexy, 2022), while others directly estimate mass from the animal's image using a regression model (Hewe, Mondia, & Empas, 2024).

Concepts from CNN applications in other fields are also examined in this research, such as in (Zhen, et al., 2015), where it is suggested that the direct use of a neural network is superior to segmentation for determining cardiac volume. In (Zhang, 2021), several networks for transfer learning were compared, with Xception (Chollet, 2017) performing best in estimating worm mass.

The objective of this study is to estimate the average body mass of Cobb500 broiler chickens present in a single image, as opposed to individually estimating mass. This goal is pursued by training two CNNs using transfer learning, based on images and measurements taken from a poultry house located in Feliz, RS, Brazil. The first CNN selects images that are under suitable conditions, while the second CNN estimates the average mass value.

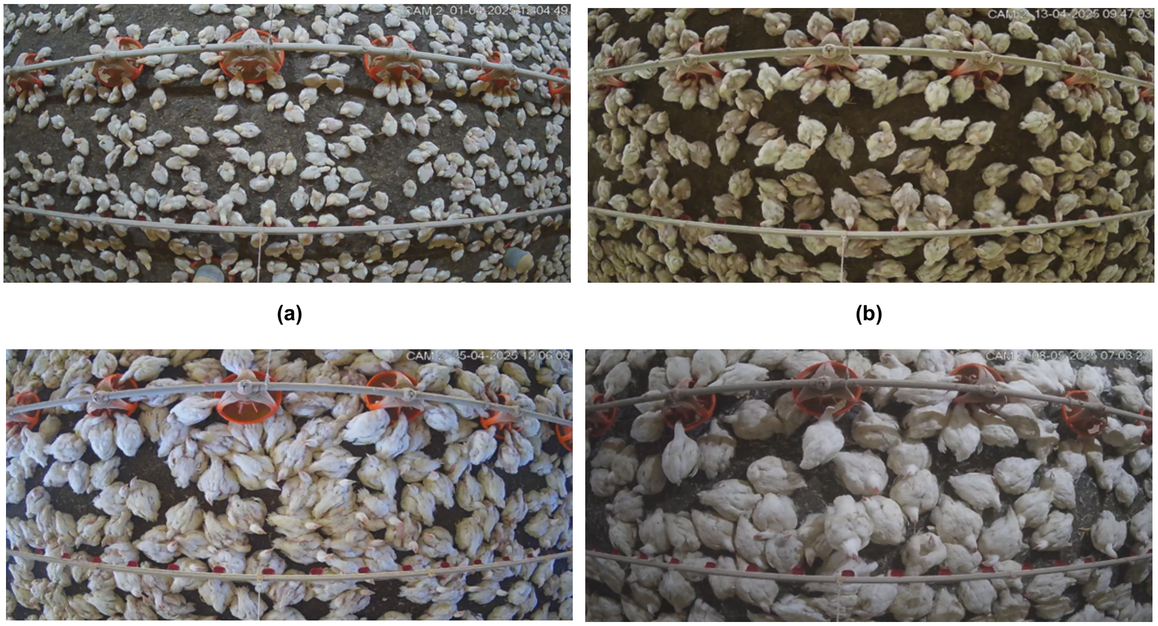
Dataset

To train the proposed CNNs, datasets were collected in Feliz, RS, Brazil inside a poultry house where Cobb500 broiler chicken are reared. This process involved gathering and selecting images, gathering measurements and creating labels associated with each image. All these steps are detailed in the subsequent sections.

Image gathering

Using Intelbras IP 1220D model cameras fixed to the ceiling of the poultry house pointing toward the floor, videos were recorded continuously during the illuminated periods (natural or artificial) of the day, when the animals tend to move and feed. Two cameras were arranged along a poultry house, at its beginning and middle, and referred to as A and B, respectively. A total of three flocks were monitored, with images from Camera A being collected for all of them, while Camera B had images captured only during the third flock.

The videos obtained from the cameras were recorded in Full HD resolution, at a frame rate of 10 frames per second, with a bit rate of 3000 kilobytes per second, and compressed using the H.264 encoding standard. From these videos, samples were extracted every minute, which was a sufficient period for the animals to move around and the scene to change. As an example, the format of the images obtained from Camera A in the first flock can be seen in Figure 1.



(a) (b) (c) (d)

Fig 1. Good images from Camera A of first flock: (a) with 7 days; (b) 14 days; (c) 21 days; and (d) 28 days.

Before being input into the neural networks, the images are cropped laterally into a 1:1 aspect ratio, to suppress the parts of the image most affected by the camera lens distortion. Therefore, only the middle section remains visible to the network.

Image selection

Since not all collected images presented a homogeneous condition in the distribution of chickens, an Xception network was first trained to classify images as good, exemplified in Figure 1, or bad (separated by the reason for rejection), shown in Figure 2. Good images were defined as those in which there is a large presence of birds in the frame, assuming that uniformity in space filling during flock development would be beneficial for training the neural network, allowing it to analyze a larger number of birds to estimate their average weight.

Firstly, using images captured in the first flock, a total of 568 good and 568 bad images were manually selected as the dataset used to train a classification CNN, which will choose images deemed adequate for the whole dataset collected from three flocks. The bad images have been distributed as follows: 92 with insufficient or excessive lighting, 128 with no or few birds, 64 with interference from people, and 284 with a gap in the center of the image.

Label gathering

To perform the regression of the average weight, data were obtained from the poultry house by the grower, using the methodology previously adopted by them to obtain this value, which allows them to monitor and report animal growth. The animals were weighed using a digital handheld scale to obtain the total weight of groups of up to 20 birds, as shown in Figure 3. This process was performed daily, obtaining measurements under each of the cameras.

(a) (b)

(c) (d)

Fig 2. Bad images classified as: (a) with a gap in the center; (b) with few chickens; (c) with excessive light; and (d) with people.

Fig 3. Broiler chicken weighing at the poultry house.

Given the discrete nature of the obtained measurements and the continuous nature of chicken growth, linear interpolation between daily averages was performed, providing an approximation of the average bird weight in the periods between measurements. These values, exemplified by the data from Camera A of flock 1 in Figure 4, were used as Ground Truth during the training of the regression CNNs.

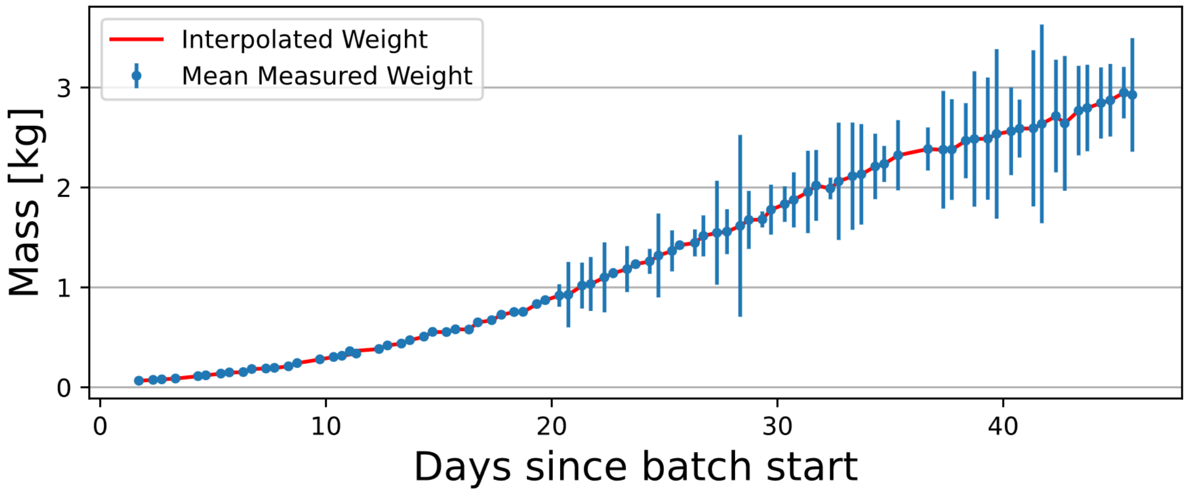


Fig 4. Average mass measurements from Camera A of first flock.

Classification Network

As mentioned previously, a classification network was used to select the adequate images. An Xception network pre-trained on the ImageNet dataset (Deng, et al., 2009) was trained using the Keras deep learning API. This network had its classification head modified with the layers defined in Table 1. The desired output from this network is a classification into ‘good’ or ‘bad’ images, hence the dense layer with 2 neurons at its end.

Table 1. Layers of the classification head.

Global Average Pooling
Dense with 2 SoftMax activation neurons

For network optimization during training, the Adam optimizer was used with an initial learning rate of 0.001 and Binary Cross-entropy as the loss metric. The learning rate value is reduced by a factor of 20% during training if it reaches a plateau after 50 epochs without improvement so that slower weight learning would allow continued progress in the metrics.

The training proceeded with no Xception network layers frozen, allowing the network to fully specialize in the dataset used in this work. The maximum number of epochs was set to 200 epochs, with Early Stopping applied in case of stagnation in validation loss for more than 100 epochs.

From the 1136 images of the dataset used in this training, they were split as follows: 60% images went into training, 20% into validation and 20% into test. The datasets used in training underwent Just-in-Time style data augmentations, in which the augmentations are reapplied at each training epoch, allowing greater variability in the dataset to promote better generalization of the final network. These augmentations are described in Table 2.

Table 2. Data augmentation preprocessing layers.

Augmentation	Value
Flip	Horizontal
Rotation	Up to 5º
Contrast	Up to 10%

With these characteristics defined, the network was trained, and its training curves are shown in Figure 5, achieving a validation Binary Cross-entropy of 0.2117, validation Precision, Recall and F1 Score of 0.9450 at the 24th epoch.

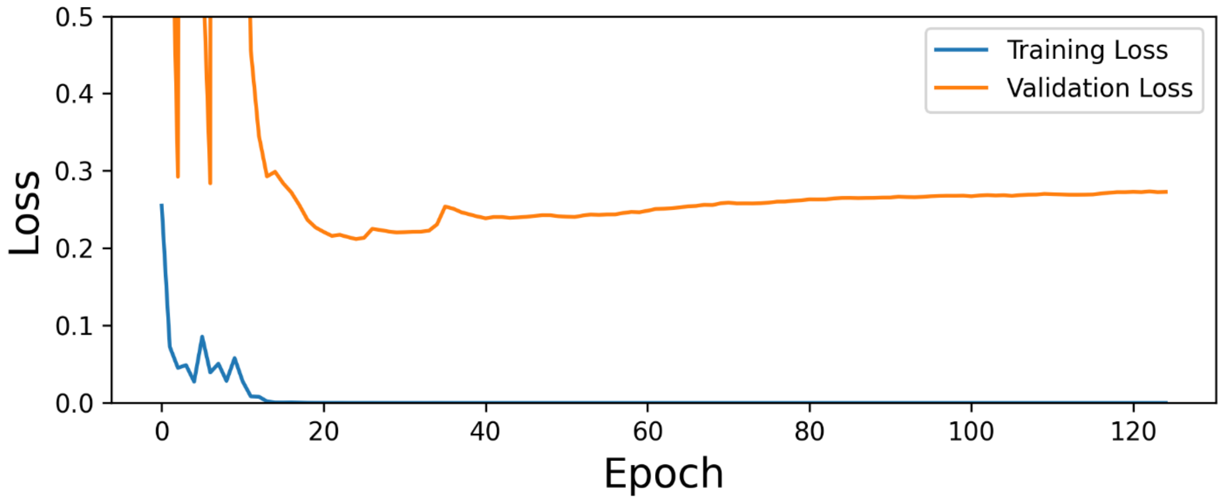


Fig 5. Classification network's training loss history

Taking the best weights, images from the test dataset were evaluated and the confusion matrix for the predictions by the network can be seen in Figure 6, showing a good performance by the network in correctly separating good and bad images. The model achieved a test Precision of 0.9230, test Recall of 1.0 and an F1 Score of 0.9599.

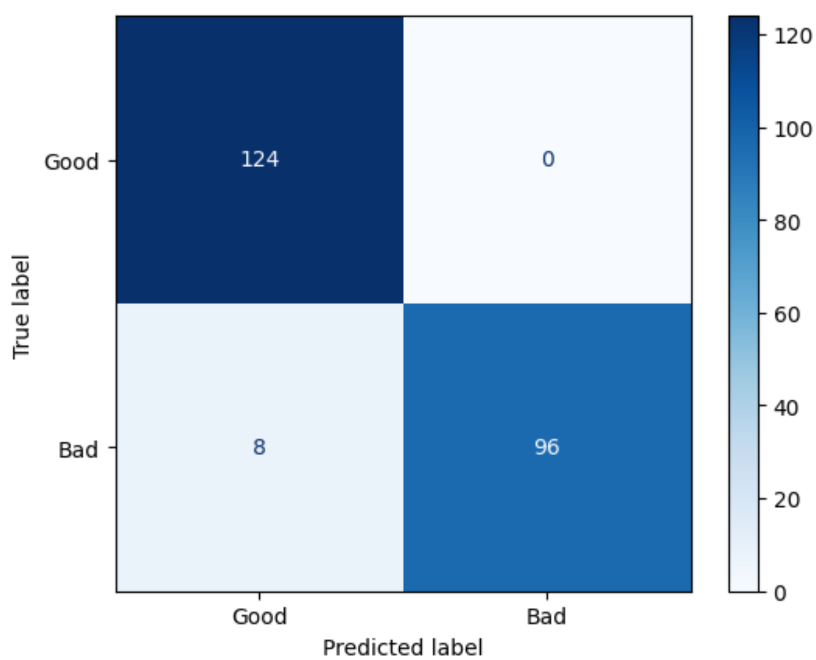


Fig 6. Confusion matrix for the test dataset evaluated by the classification network.

Since the model achieved a good performance, every image from the three flocks was evaluated. Taking Camera A of the first flock as an example, out of a total of 30168 images, there are 16457 images considered good and 13711 bad images, which are divided into 5.88% with insufficient or excessive lighting, 29.27% with no or few birds, 2.03% with interference from people, and 62.82% with a gap in the center of the image. Finally, 50 good images per day were chosen to compose the datasets, totaling 1800 images from days 9 to 44.

Given the images captured from other flocks and with cameras placed elsewhere, Table 3 briefly presents the number of good and bad images for each of the collected datasets. The dataset related to Camera B refers to a reduced number of days, due to an interval between the start of the flock and the time when the animals reach the point in the poultry house where this camera is located.

Table 3. Datasets summary.

Flock	Camera	Total	Good (%)	Bad (%)	Dataset Total
1	A	30168	54.55	45.45	1800
2	A	49358	39.85	60.15	1800
3	A	50186	45.20	54.80	1800
	B	23800	44.13	55.87	1100

Regression Network

Finally, the datasets from flocks 1A and 2A were used as the training and validation set for the regression network, with a training-validation split of 80–20%, while the datasets from the third flock, 3A and 3B, were used entirely as the test set. It is important to note that this means the network was not trained with any images from the test set camera from dataset 3B, allowing an evaluation of the generalization of the results.

The regression network follows the steps taken with the classification network, with the Xception network pre-trained on the ImageNet dataset as the base network. The main difference relies on the change from its classification head to a regression head with the layers defined in Table 4. Since the current task involves estimating a real value, the final layer chosen was a Dense layer with only 1 neuron and linear activation.

Table 4. Layers of the regression head.

Global Average Pooling 50% Dropout Dense with 512 ReLu activation neurons
Batch Normalization 40% Dropout Dense with 256 ReLu activation neurons
Batch Normalization 30% Dropout Dense with 1 linear activation neuron

The remaining differences between the training methods are as follows: the metric chosen as the training loss was the Mean Squared Error (MSE), the addition of a Standard Scaler to the output of the network and the training was split into two stages.

The Standard Scaler scales the value of labels to fit a normal distribution, affecting the range in which the neural network outputs its estimations. This can help the network convergence by removing the bias from the labels and keeping weight values contained to smaller values.

The two stages can be differentiated as: the first was performed with the base network layers frozen, thus training only the weights of the regression head; in a second fine-tuning stage, the first 20% of the base network layers remained frozen, aiming to keep part of the feature extraction intact while specializing the network to the dataset used in this work. The maximum number of epochs for each stage was set to 600 and 300 epochs, respectively, with Early Stopping applied in case of stagnation in validation loss for more than 100 and 50 epochs, respectively.

Due to having 3 flocks with the same Camera A, multiple networks were trained with different combinations of two flocks as training datasets and a third as the test dataset. These findings can be found on Table 5, where using flocks 1 and 3 to train the networks yielded the best results, testing the model on the second flock. Its training curves are shown in Figure 7, achieving at epoch 352 a validation Root Mean Squared Error (RMSE) of 92.1 grams and a validation Mean Average Error (MAE) of 66.3 grams after rescaling the output values. This model will be referred to as the first regression network for the remainder of this work.

Table 5. Regression networks trained in two flocks with Camera A.

Trained in flocks:	Test RMSE (g)	Test MAE (g)	Test Average daily MAE (g)
1 and 2	178.0	141.9	120.0
2 and 3	215.6	187.7	183.0
1 and 3	**109.8**	**82.0**	**48.5**

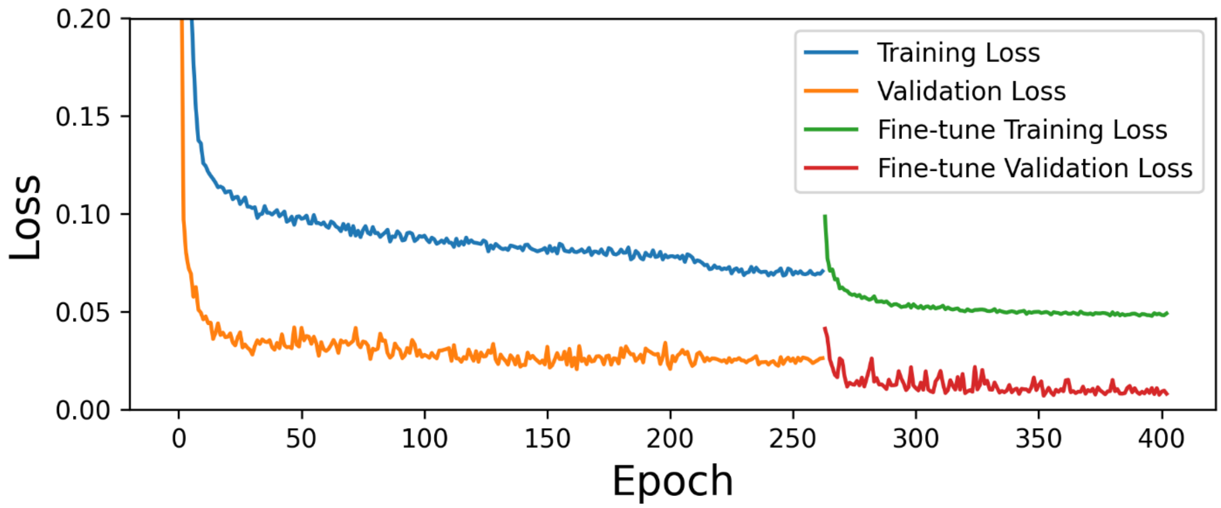


Fig 7. First regression network's training loss history

Since the variable of interest in this study does not need to be obtained instantaneously but rather evaluated daily, we sought to filter the uncertainties inherent to the neural network by using a daily average. From these daily averages, an overall average is obtained, which will be referred to as the average daily MAE metric. In addition to this metric, RMSE and MAE were also calculated to obtain the instantaneous error of the estimates. For dataset 2A, whose estimates are shown in Figure 8, these errors were 109.8 (RMSE), 82.0 (MAE), and 48.5 grams (average daily MAE).

Looking at the tail end of the graph, the average daily MAE starts to diverge from the real measurements, probably due to a lack of representation from the last days of the rearing process and the changes in scenery from the removal of feeders and waterers.

Applying these same principles to a new training, a neural network was trained utilizing all datasets form Camera A (flocks 1 through 3), hereby called the second regression network. This network achieved a validation RMSE and MAE of 68.2 and 50.8 grams, respectively, after a total of 770 epochs. The training history is presented in Figure 9.

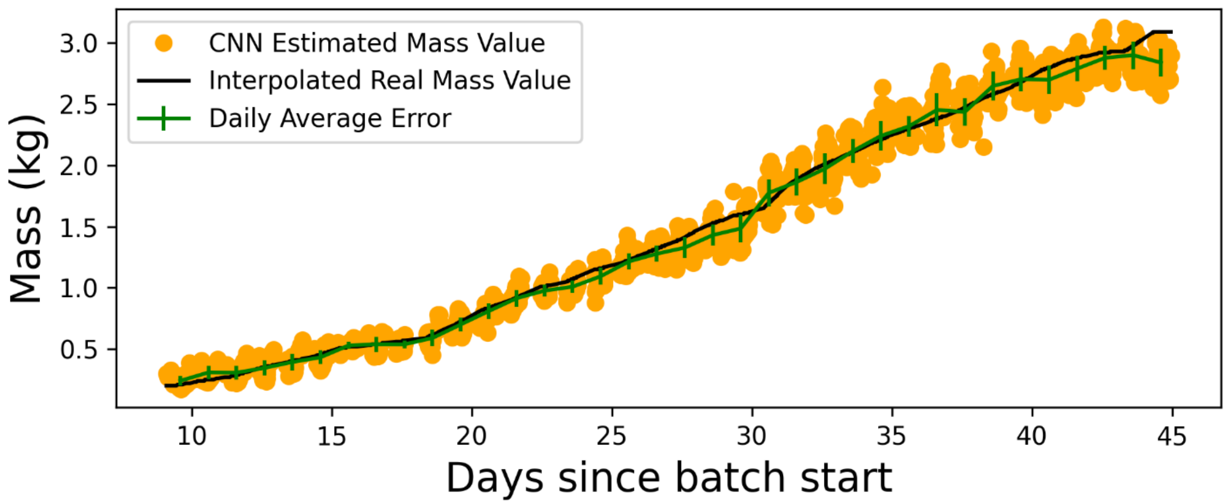


Fig 8. Mass predictions on Camera A from the second flock, using the first regression network.

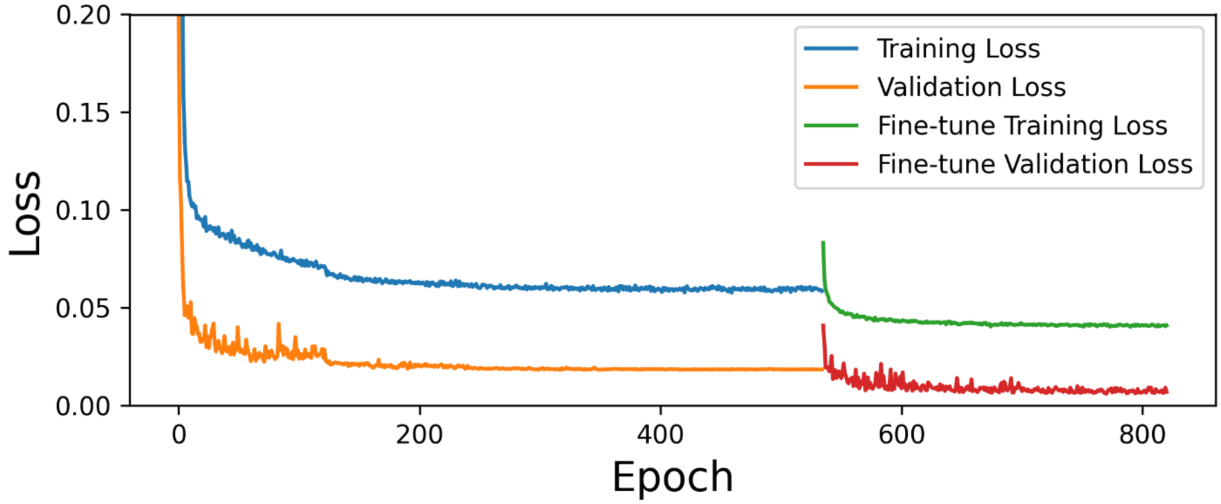


Fig 9. Second regression network's training loss history

Results

Having two trained networks, the First and Second Regression Networks, a comparison can be made. Firstly, by analyzing their validation metrics, summarized in Table 6, the latter achieved smaller error values, indicating that the increased number of images from an additional flock helped with its generalization of the problem. Secondly, when a dataset from a camera positioned elsewhere in the poultry house, Camera B, is evaluated by both networks, a direct comparison can be drawn.

Table 6. Comparison between first and second regression networks' validation metrics.

Trained in flocks:	Validation RMSE (g)	Validation MAE (g)
1 and 3	109.8	82.0
1, 2 and 3	**68.2**	**50.8**

Figures 10 and 11 present the mass estimation on Camera B from the third flock, evaluated by the First and Second networks, respectively. The latter has a visibly better performance, indicating the generalization of the training by adding newer information to the training dataset.

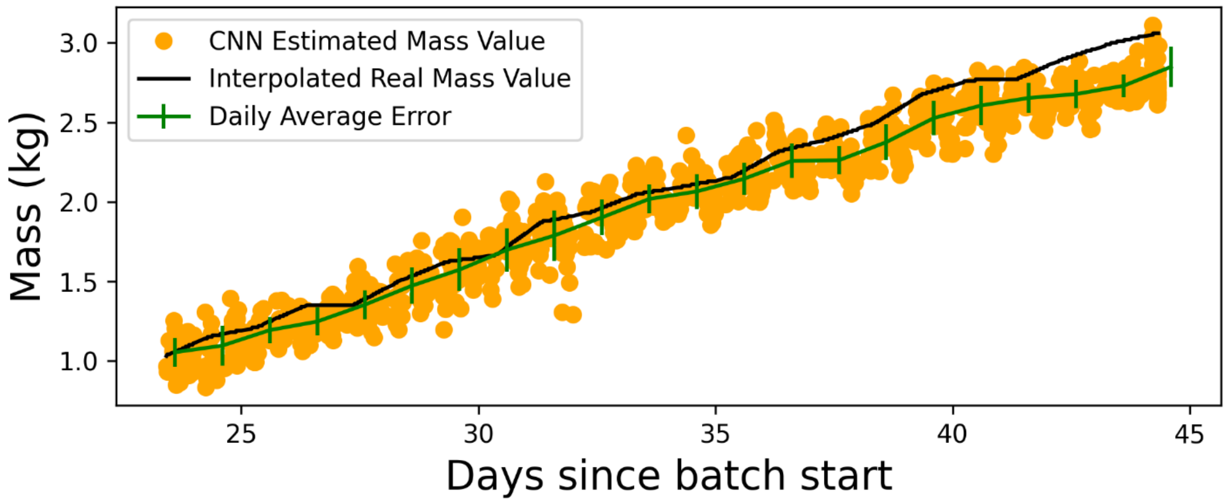


Fig 10. Mass predictions on Camera B from the third flock, using the first regression network.

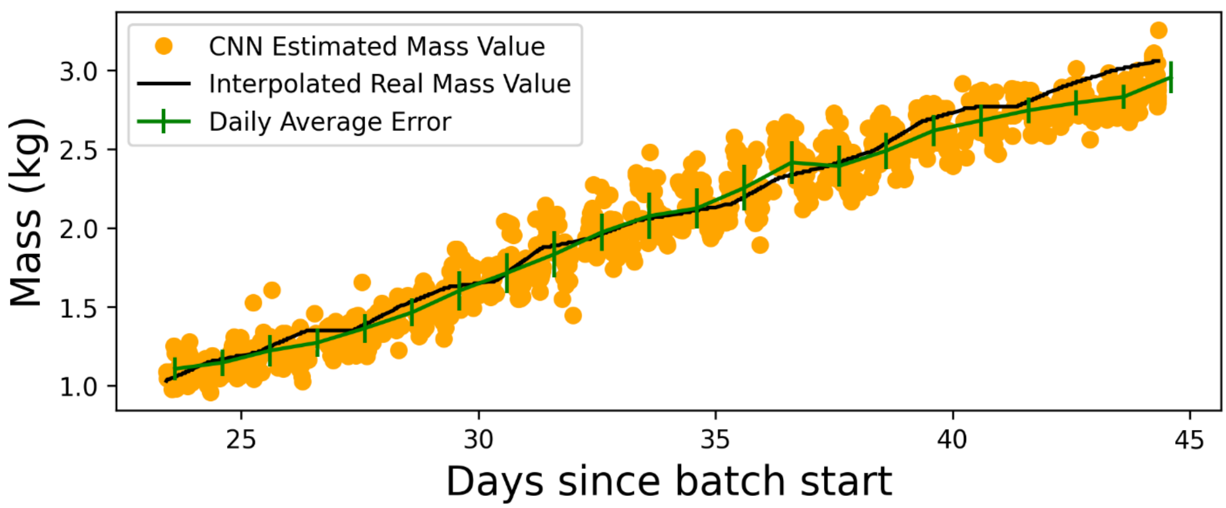


Fig 11. Mass predictions on Camera B from the third flock, using the second regression network.

This improvement can also be measured, and is shown in Table 7, comparing metrics previously used, as well as the Coefficient of Determination (R^2), which measures how well the model fits the data it measures.

Table 6. Comparison between trained regression networks tested with Camera B from third flock.

Trained in flocks:	Test RMSE (g)	Test MAE (g)	Test Average Daily MAE (g)	R^2
1 and 3	173.2	139.3	110.0	0.9248
1, 2 and 3	**135.5**	**107.5**	**56.4**	**0.9527**

Conclusion

This work presented a different approach to measuring broiler chickens in a poultry house, by estimating their mean mass, due to constraints in the data gathering methods available. This method proved capable of approximating the mass with a simple data collection method, considering the growers previously established methodologies and installing a couple cameras in the poultry house.

A fully functioning system with this technology could provide daily reports to growers, giving insights into their flocks and helping with decision making, such as the number of light hours should be given to the birds.

Future work might investigate improving the measurements collected, with floor scales providing reliable values for each chicken mass, unlocking different techniques to be used such as Object Detection or Segmentation networks to extract features from the birds.

Acknowledgments

This work was partly translated to English with help from the DeepSeek-V4 Flash Large Language Model. The text in its original language was written entirely by the authors, and the translated sections were thoroughly reviewed.

This study was financed in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Brazil, Grant/Award Number 409376/2022-1 and supported by the Center for Embedded Devices and Research in Digital Agriculture (CEDRA) - SENAI-RS.

References

Chollet, F. (2017). 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). *Xception: Deep Learning with Depthwise Separable Convolutions*, (pp. 1800-1807). Honolulu, HI, USA.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 248-255). doi:10.1109/CVPR.2009.5206848

Hewe, T. R., Mondia, K. M., & Empas, P. E. (2024). Advancing Poultry Farming Efficiency Through YOLOv5 and Image Regression-Based Broiler Weight Estimation. *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*, (pp. 7-12).

Nyalala, I., Okinda, C., Kunjie, C., Korohou, T., Nyalala, L., & Chao, Q. (2021). Weight and volume estimation of poultry and products based on computer vision systems: a review. *Poultry Science*, p. 101072.

Szabo, S., & Alexy, M. (2022). Practical Aspects of Weight Measurement Using Image Processing Methods in Waterfowl Production. *Agriculture, 12*. doi:10.3390/agriculture12111869

Zhang, J. a. (2021). Pig Weight and Body Size Estimation Using a Multiple Output Regression Convolutional Neural Network: A Fast and Fully Automatic Method. *Sensors*.

Zhen, X., Wang, Z., Islam, A., Bhaduri, M., Chan, I., & Li, S. (2015). Direct volume estimation without segmentation. *Medical Imaging 2015: Image Processing*, (pp. 671-676).