Deep Learning Techniques for Beef Cattle Body Weight Prediction

Conference Paper · July 2020

DOI: 10.1109/IJCNN48605.2020.9207624

7 authors, including:

Vanessa Aparecida de Moraes Weber
Universidade Católica Dom Bosco (UCDB)

Rodrigo da Costa Gomes
Brazilian Agricultural Research Corporation (EMBRAPA)

Hemerson Pistori
Universidade Católica Dom Bosco (UCDB)

Marco Alvarez
National University of Trujillo

Deep Learning Techniques for Beef Cattle Body Weight Prediction

Mikel Gjergji[1], Vanessa de Moraes Weber[2,3], Luiz Otávio Campos Silva[4], Rodrigo da Costa Gomes[4], Thiago Luís Alves Campos de Araújo[5], Hemerson Pistori[2,6], Marco Alvarez[1]

[1]*Departament of Computer Science and Statistics - University of Rhode Island - Kingston - USA*
[2]*Dom Bosco Catholic University (UCDB) - Campo Grande – MS – Brazil*
[3]*Universidade Estadual do Mato Grosso do Sul (UEMS) - Campo Grande– MS – Brazil*
[4]*Brazilian Agricultural Research Corporation - Embrapa Beef Cattle - Campo Grande - MS - Brazil*
[5]*Federal University of Ceará (UFC) - Fortaleza – CE – Brazil*
[6]*College of Computing, Federal University of Mato Grosso do Sul (UFMS) - Campo Grande – MS – Brazil*

***Abstract*—Following the weight of beef cattle is of great importance to the producer. The activities of nutrition, management, genetics, health and environment can benefit from the weight control of these animals. We explore different deep learning models performance in the regression task of predicting cattle weight. This is a hard problem since moving from 3d space to 2d images presents a loss of information in object shape, making weight prediction more difficult. A model that produces good results in this problem could potentially be applied more abstractly to similar problem spaces. We analyze convolutional neural networks, RNN/CNN networks, Recurrent Attention Models, and Recurrent Attention Models with Convolutional Neural Networks. In this paper we show the performance and speed comparisons of these networks for the problem of beef cattle weight prediction, and find that convolutional neural networks are most performant. Our top model has a Mean Average Error of 21642.963 Grams (21.64KG). This is nearly half the error as previous top linear regression models which reached an error of 38.46KG.**

***Index Terms*—deep learning, weight, cattle, attention based models, convolutional neural networks, recurrent neural networks**

I. Introduction

Monitoring and maintaining the weight history of cattle allows for timely intervention of cattle diet, cattle health, and for greater efficiency in genetic selection. Another great advantage of tracking weight gain is to identify the best time to market animals because animals that have already reached the point of slaughter represent burden for feedlot.

Removing animals from paddocks and leading them to scales is a costly and stressful activity for both the animal and the herdsman. This process can cause injuries or even weight loss [1] [2]. With this in mind, some companies have been working on solutions to track the weight of feedlot cattle and have tools such as GrowSafe, Intergado and the Bosch Precision Livestock Platform. These solutions consist of weighing cell equipment that must be installed in passageways or in front of feeders and drinkers. However these devices need constant maintenance that may encumber the cost of production. Still on weight measurement, some researchers propose research where they relate measurements of body parts of animals with their weight [3] [4] [5].

In addition, researchers have developed livestock-based applications based on image analysis through the computing area known as Computer Vision [6]. These applications allow automation of some farm work in key areas such as animal behavior, health and welfare including nutrition management [7], locomotion [8], identification [9], body conditions [10] [9], diseases [11], and cattle weighing [12] [13] [14]. To this end, equipment is described for acquisition of these images and can be divided into two large groups, the first for 2D images such as RGB cameras and thermal cameras, and the second for 3D images such as depth and Kinect sensors, stereo vision and stereo photogrammetry according to the review by Nasirahmadi et al. [15].

In addition to classical feature extraction techniques detailed in the related work section, deep learning for computer vision has stood out in the last decade mainly for the frequency with which they reach the state of the art. Deep learning architectures are known to have often outperformed even humans in classification problems [16] [17] [15] [18]. These same architectures supported by regression techniques can solve problems aimed at predicting continuous values [19], such as estimating head positions and detecting facial expressions [20] [21].

In this paper we analyze convolutional neural networks, recurrent convolutional neural networks, recurrent attention models [22], and recurrent attention models with convolutional neural networks. We find that convolutional neural networks outperform all of the other tested models. We were able to produce a model that nearly halved the error of a previous regression model based on more traditional computer vision techniques [**?**]. Our top model had a MAE of 21.64KG while the previous paper produced a top model who's MAE was 38.46KG. This demonstrates that convolutional neural networks vastly outperform models trained on hand picked features for this task.

In our Related Work section we review work done on weight calculation of animals, advancements and applications of convolutional neural networks, and applications of Recurrent Attention Models [22] with convolutional neural networks. In our materials and methods section we describe how we collected data and split it into train, validation, and test sets. In our models section we give an overview of different models tested, and provide a table of results for those models. In our analysis section we analyze the results of each model. In our future work section we talk about where to move forward with this problem based on problem areas in our models. We summarize our results in our conclusion.

II. Related Work

A. Weight Calculation of Animals

The use of computer vision techniques to predict weight of cattle has also been described for both 2D and 3D images [15]. Morphological characteristics such as croup height, croup width, body size, rib height and contours [23] among others were automatically measured and subsequently subjected to regression algorithms or Fuzzy logic, according to Stajnko et al. [24].

From 3D images Cominotte et al. [14] selected, segmented and extracted features from 234 images of Nellore beef cattle for regression algorithms and Artificial Neural Networks (ANN) to estimate the live weight of cattle. The images were collected at various stages of the animal's life, namely: Weaning, Stocker, Beginning of Feedlot, and End of Feedlot phase.

Although some authors describe that biometric measurements extracted from 3D images are highly correlated with animal weight [14] [25], ease of access and costs. Lower 2D images from ordinary cameras, such as security imaging equipment for example, should be considered.

The segmentation of animals in their natural environment has been the subject of research and constitutes a challenging task for automated animal body mass prediction systems [25] [26]. However, as it is from the contour of the animal that can be extracted several measurements with possible high correlations with the weight of the animal, systems and or automatic solutions for mass prediction of cattle are mostly hostage of this task. Like segmentation, extracting frames with the best positioning of the cattle is also a challenging task. Therefore, deep learning techniques that can predict the weight of cattle through 2D image processing without the specific task of segmentation and frame extraction seem promising.

B. Convolutional Neural Networks

Weight calculation in 2d space is a difficult task because there is a loss of information when migrating from 3d to 2d space. We are trying to estimate the amount of mass in a 3d object using only information in 2d space. Convolutional neural networks have been applied to tasks facing similar problems with success. Convolutional neural networks were applied to the problem of 3d pose estimation using 2d data [27] with success. Convolutional neural networks have also been used to create 3d point clouds from 2d images [28], an important example for our use case since it provides a concrete example of a convolutional neural network learning a mapping from $2d \mapsto 3d$ space. These examples demonstrate an ability of convolutional neural networks to work effectively on 2d images when the problem space lies more in 3d space, and make them a good candidate for weight prediction.

C. Combination RNN/CNN Attention Networks

Attention based RNN/CNN combination networks have seen success in several application domains. Some examples of this include image captioning [29], and object detection [30]. Attention based RNN/CNN combination networks have also seen use in a variety of different regression based tasks. Zhou et al. use a combination RNN/CNN network with attention to predict the price of precious metals [31]. Zhao et al. use a combination RNN/CNN network with attention for fine-grained visual emotion regression [32]. These successes, particularly those of regression tasks, make these more advanced models a good candidate for testing in weight prediction.

III. Materials and methods

A. Data collection

The images to form this data set were collected from October 8 to October 20, 2018 at the Embrapa Gado de Corte in Campo Grande MS (geographical location -20º 43 '64.915 "latitude South and - 54 ° 55 '65.227 "west longitude). As can be seen in the aerial image in Figure 1, where are distributed in two paddocks 20 male Nellore and Angus.

Fig. 1. Overhead view of Embrapa beef cattle feedlot.

For the collection of images, the experiment included the installation of a DVR set: MD-1004NS MD-DVR41 of MIDI brand, cameras with AHD 720p image quality and a HD with 1Tb recording capacity. Two cameras (b) of the equipment were installed so as to be fixed in the structure of the drinker (a) of the equipment known as Intergado ®, on an adapted rod, so that each camera collects the image of the dorsal area of the animal when taking water in a of the possible entries. Two other cameras (c) were installed in the trough cover structure to acquire the profile images of the animals moving to the trough. The collected videos were stored in the DVR and later transferred to computers for the purpose of preprocessing and extracting the frames that contained images of the animals, as shown in Figure 2.

Fig. 2. Embrapa Beef Cattle image acquisition system, Intergado drinking fountain (a), cameras installed on the drinking trough (b), cameras installed on the feeding troughs to capture profile images (c), aerial view of Embrapa Beef Cattle deeflot with the collection system (d).

This drinking system Figure 2 (a) is part of the Intergado equipment and allows individual identification of the animal through an RFID antenna, and every time the cattle travels to the drinking fountain it is positioned on a platform coupled to a weighing scale. Additionally, time and weight data of the animals are transmitted via transmission antenna to the company's software. By relating the weighing time indicated by the software with the video of the corresponding drinker entrance it is possible to identify the animal that is in the drinker and thus extract tables containing the image of the cattle as can be seen in Figure 3 a and b.

Fig. 3. Image of the dorsal area of bovine collected by the collection structure in the Embrapa Beef Cattle feedlot (a). Sequence of frames extracted from the bovine video in the trough (b).

After the image collection was completed, they were validated to compose the ESTMASSABOV400 image database.

B. Separation of Training, Validation, and Test Set

Since our data are video frames of an individual cow, moving from frame x to frame $x+1$ will be highly similar images. We want to avoid training on data points that are nearly identical with data points in our validation or test set in order to get a better gauge of how our model performs on unique and unseen data. To ensure fair tests, we first create a dictionary where each key is a unique cow identifier that indexes a list of tuples containing the video frame (jpeg), frame number, and cow weight. We convert this dictionary to a list, then separate 60% of the unique cows for training, 20% for validation, and 20% for our final test set once best hyper parameters for validation set have been found. All weight labels were scaled by using the formula

$$x = \frac{x}{x_{max}}$$

where x is the weight of a cow and x_{max} is the largest cow weight in our dataset. We do this in order to squash the range of labels to [0-1] and avoid extremely large gradients in training. For our MAE calculations, we average the loss of all our batches on a given set of data.

IV. Models

In this section, we detail the different types of networks that were tested and training procedures for each of them. We also provide results for these models in Tables I & II.

A. Convolutional Neural Networks

We trained 3 different convolutional neural networks using the Adam optimizer along with a learning rate of .0005. We used two EfficientNet models [33], EfficientNet-B1 and EfficientNet-B7 as well as the ResNet18 [34] CNN. Our reasoning for choosing EfficientNet is that it provides a smaller yet high performing model when trained on ImageNet, and better performing models on ImageNet seem to correlate to better accuracy when transferred to other problems [35]. Our reasoning for choosing the ResNet18 model is that it is a highly tested model and easily implementable in PyTorch. All of our convolutional neural networks were trained for 10 epochs with L1 Loss. L2 Loss models were trained for 5 epochs since MAE was much higher.

TABLE I

A table of each of our tested convolutional neural network models along with their lowest MAE acheived on the test set.

Model	Lowest MAE Test Set (grams)
ResNet18 w/ L1 Loss	24510.853
EfficientNetB1 w/ L1 Loss	21642.963
EfficientNetB7 w/ L1 Loss	23482.815
ResNet18 w/ L2 Loss	40008.518
EfficientNetB1 w/ L2 Loss	44064.731

B. Recurrent Attention Model without Convolutional Neural Network

We trained a RAM model using the same hyper parameters from our top performing Combination RNN/CNN with attention model. This network follows the same architecture as the Combination RNN/CNN with attention model but does not process glimpses through a convolutional neural network, instead concatenating glimpse scales together and processing it through a fully connected layer. Our model replicates the RAM model specified in Recurrent Models of Visual Attention [22] with a few exceptions.

We utilize three fully connected layers in our glimpse network by adding an additional fully connected layer processing the 'what' of our network. This is because when adding a convolutional neural network the output size was dependent on the number of scales, so adding a fully connected layer who's input size is $(h_g * s)$ where h_g is our hidden layer size for the what of our network and s is the number of scales to perform on the patch location and who's output size is h_g accounts for this.

Our action network must reflect that of the regression task at hand, so our action network is a fully connected layer who's input is h_t and output is one neuron producing a continuous value. Another modification that was made to accommodate our change in task was to the baseline network. Rather than using a rectifier activation, we omit this from our baseline network. This is due to how our reward is defined for the reinforcement loss in our hybrid loss. Reward is defined by the following formula:

$$R = -1 * |p - y|$$

where R is our reward, p are our predictions, and y are our labels. This is simply the absolute difference between our predicted versus real labels. A smaller difference produces a greater reward, with an exact match giving the highest reward of 0. Scaling our labels caused the model to fail to learn, so for the training of this model weights were not processed through the function

$$x = \frac{x}{x_{max}}$$

For this model and all sections following, we trained for up to 100 epochs and stopped training if validation accuracy did not increase after 10 epochs.

C. Combination RNN/CNN without attention model

In this model, we remove the attention portion of the RAM with CNN and instead feed fixed locations to the model. This is accomplished by removing the location and baseline network from the Recurrent Attention Model. Since we do not need to train a location network on this model, our loss function for this network is simply L1 or L2 loss rather than the hybrid loss used with attention networks.

D. Combination RNN/CNN with attention

Our Combination RNN/CNN with attention is achieved by attaching a convolutional neural network to the glimpse network to better extract features from each glimpse. We implemented this by feeding EfficientNet-B1 each glimpse and outputting an encoding of size h_g. Since a hyper-parameter of the model is how many scales to perform on each glimpse location, we process each scale through the network individually, and then concatenate the encodings together into a feature vector of size $(h_g * s)$. This is why we needed to add an additional fully connected layer as talked about in the Recurrent Attention Model without Convolutional Neural Network section. The rest of the model is identical to the model specified in the Recurrent Attention Model without Convolutional Neural Network.

After performing a random search for which hyper-parameters performed best on our validation set, we found the highest performing was 6 glimpses, a patch size of $96x96$ pixels, and no additional scales. We took these hyper-parameters and retrained a model using both the training and validation data, and tested on our held out test data. We evaluated the final error using both L1 and L2 loss metrics. A box plot versus true value can be seen in Figures 7 and 8.

V. Analysis

Our original intuition behind testing different variations of recurrent models was inspired by a recent paper that shows that CNNs trained on ImageNet prioritize textures versus object shapes [36]. Since mass is much more closely related to the object shape of the cow rather than the textures of the cow, we thought that a network that incorporates a location policy will be forced to give greater attention to object shape.

We tried different variations that attempted to exploit this idea, the first being a close implementation of the original recurrent models of visual attention [22]. This model did not converge to a low enough MAE, and actually reached a MAE close to that of the model trained on hand picked features. This is most likely caused by a lack of ability to learn high level features that a convolutional neural network knows to exploit. We then attached the EfficientNet-B1 network to our glimpse network to create a combination RNN/CNN with attention and found that this reduced error a great deal, but still did not beat the standalone EfficientNet-B1 network. Finally, we tested a model that removed the attention portion of the model, and found slightly degraded results over our attention module.

One possible explanation for these results could be that the convolutional neural network always has full view of the image from input to output, while the recurrent models are selecting subsections of the image. This is a loss of information and can negatively affect performance of the models. An interesting datapoint that supports this explanation can be seen in our box plot versus actual weight graphs for our models. There is a huge failure of each model of a cow weighing 392500 grams, where each model massively overestimates the weight of the cow. In the frames of the datapoints for this label, other cows are entering and leaving the frame, an example of which is shown in figure 6. The models could be learning to segment all areas of the picture that have the texture of a cow, and are including the additional cow area contributed by

Fig. 4. A random batch of seven cows from the test set used in our combination rnn/cnn with attention model and the glimpses taken using using L1 Loss.

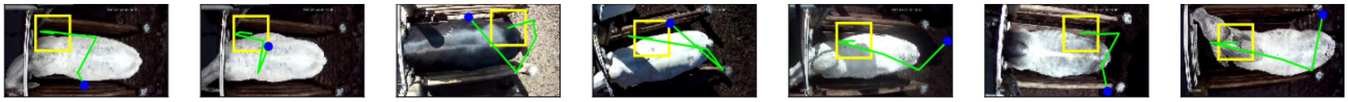

Fig. 5. A random batch of seven cows from the test set used in our combination rnn/cnn with attention model and the glimpses taken using using L2 Loss.

TABLE II
A TABLE OF EACH OF OUR TESTED RECURRENT MODELS ALONG WITH THEIR LOWEST MAE ACHEIVED ON THE VALIDATION/TEST SET. ALL RECURRENT MODELS USED A GLIMPSE SIZE OF 96X96 AND TOOK A TOTAL OF 6 GLIMPSES FOR A GIVEN IMAGE.

Model	Lowest MAE Validation Set	Lowest MAE Test Set
Combination RNN/CNN without attention (L1 Loss)	27197.365	27197.365
Combination RNN/CNN without attention (L2 Loss)	27622.287	27622.287
Recurrent Attention Model without CNN (L1 Loss)	38314.773	38061.492
Combination RNN/CNN with attention (L1 Loss)	25031.570	28338.379
Combination RNN/CNN with attention (L2 Loss)	25285.855	26476.070

these straying cows that do not belong. The only model that makes some predictions near the actual weight of the cow is our combination RNN/CNN network using L1 loss. Since it does not observe the entire image, it may have only taken glimpses in the area containing the cow we are attempting to predict on, leading to better results for some predictions. This can be seen in figure 7, where there are a number of outlier predictions near the true label for that cow.

Fig. 6. A frame where another cow is seen in the upper right corner, potentially being picked up by the networks and causing large error rates.

VI. Future Work

Following the bad data example demonstrated in the analysis section and shown in figure 6, future work can be done on developing a model that avoids outlier areas that it should not be looking at. Perhaps a more advanced attention module can be implemented to accomodate for this, since ours was able to predict fairly accurately on some occasions.

VII. Conclusion

Our experiments appear to show that convolutional neural networks are high performing on the task of weight calculation in 2d images. However, they are highly prone to bad data as shown in figure 6. While we reached an error rate much lower than that of the models trained on hand-picked features, there is still work to do to eliminate the large errors that can occur from these bad datapoints as they are likely to occur when a model is implemented in practical use.

References

[1] E. D. G. Santos, M. F. Paulino, R. d. P. Lana, S. d. C. Valadares Filho, and D. S. Queiroz, "Influência da suplementação com concentrados nas características de carcaça de bovinos f1 limousin-nelore, não-castrados, durante a seca, em pastagens de brachiaria decumbens," *Revista Brasileira de Zootecnia*, vol. 31, no. 4, pp. 1823–1832, 2002.

[2] S. Ozkaya and Y. Bozkurt, "The relationship of parameters of body measures and body weight by using digital image analysis in pre-slaughter cattle," *Archiv fur Tierzucht*, vol. 51, no. 2, p. 120, 2008.

[3] M. d. O. Franco, M. I. Marcondes, J. M. d. S. Campos, D. R. d. Freitas, E. Detmann, and S. d. C. Valadares Filho, "Evaluation of body weight prediction equations in growing heifers," *Acta Scientiarum. Animal Sciences*, vol. 39, no. 2, pp. 201–206, 2017.

[4] G. L. Reis, F. H. M. A. R. Albuquerque, B. D. Valente, G. A. Martins, R. L. Teodoro, M. B. D. Ferreira, J. B. N. Monteiro, M. d. A. Silva, and F. E. Madalena, "Predição do peso vivo a partir de medidas corporais em animais mestiços holandês/gir," *Ciência Rural*, vol. 38, no. 3, pp. 778–783, 2008.

[5] G. Bretschneider, A. Cuatrin, D. Arias, and D. Vottero, "Estimation of body weight by an indirect measurement method in developing replacement holstein heifers raised on pasture," *Archivos de Medicina Veterinaria*, vol. 46, no. 3, pp. 439–443, 2014.

[6] J. G. A. Barbedo and L. V. Koenigkan, "Perspectives on the use of unmanned aerial systems to monitor cattle," *Outlook on Agriculture*, vol. 47, no. 3, pp. 214–222, 2018.

[7] S. Nyamuryekung'e, A. F. Cibils, R. E. Estell, and A. L. Gonzalez, "Use of an unmanned aerial vehicle- mounted video camera to assess feeding behavior of raramuri criollo cows," *Rangeland ecology & management*, vol. 69, no. 5, pp. 386–389, 2016.

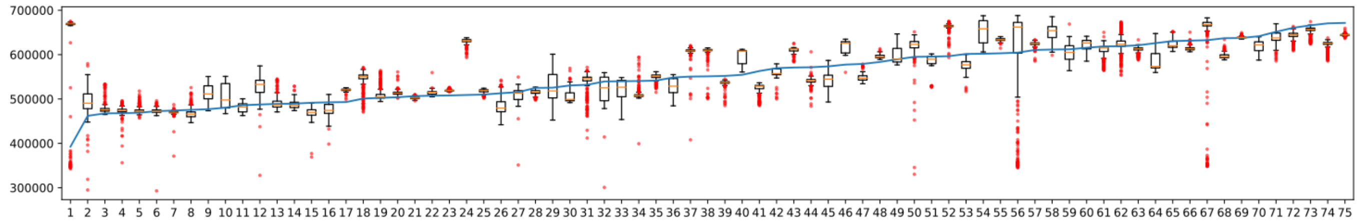

Fig. 7. Model pictures is combination rnn/cnn network with attention using L1 loss. Actual weight values are plotted in sorted order in blue. Box plots of predictions for that given weight are shown for that label, with outliers plotted in red.

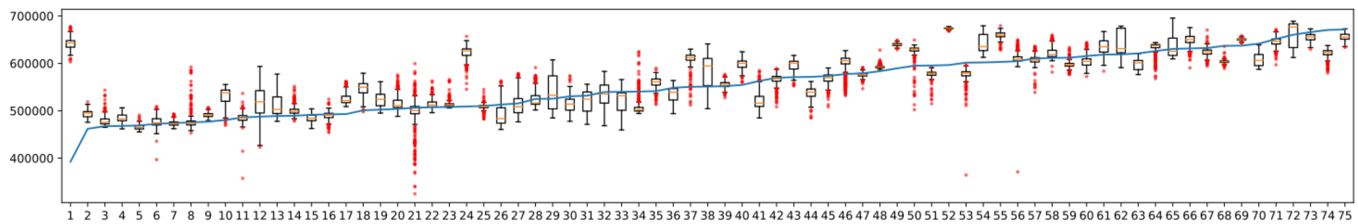

Fig. 8. Model pictures is combination rnn/cnn network with attention using L2 loss. Actual weight values are plotted in sorted order in blue. Box plots of predictions for that given weight are shown for that label, with outliers plotted in red.

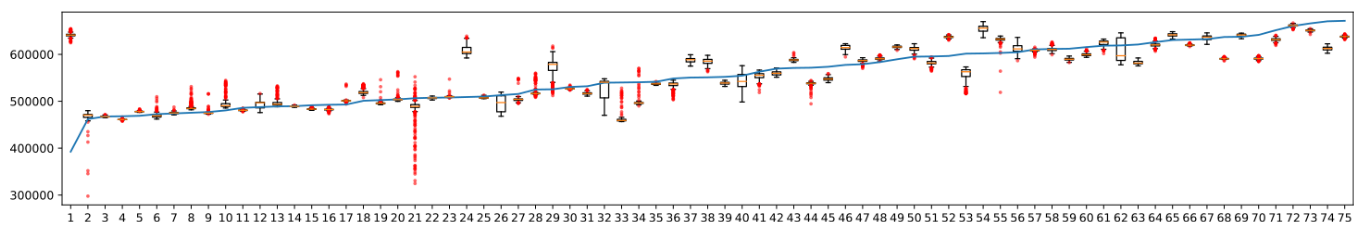

Fig. 9. Model pictured is EfficientNet-B1 using L1 loss. Actual weight values are plotted in sorted order in blue. Box plots of predictions for that given weight are shown for that label, with outliers plotted in red.

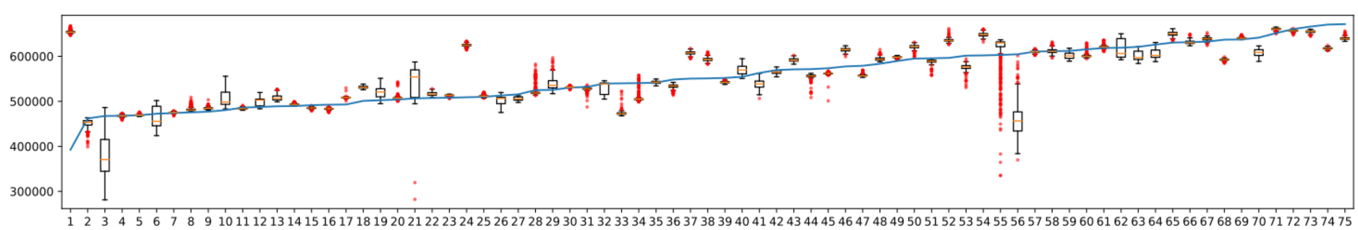

Fig. 10. Model pictured is EfficientNet-B1 using L1 loss. Actual weight values are plotted in sorted order in blue. Box plots of predictions for that given weight are shown for that label, with outliers plotted in red.

[8] F. Lao, T. Brown-Brandl, J. Stinn, K. Liu, G. Teng, and H. Xin, "Automatic recognition of lactating sow behaviors through depth image processing," *Computers and Electronics in Agriculture*, vol. 125, pp. 56–62, 2016.

[9] A. Bercovich, Y. Edan, V. Alchanatis, U. Moallem, Y. Parmet, H. Honig, E. Maltz, A. Antler, and I. Halachmi, "Development of an automatic cow body condition scoring using body shape signature and fourier descriptors," *Journal of dairy science*, vol. 96, no. 12, pp. 8047–8059, 2013.

[10] S. Jung and K. B. Ariyur, "Strategic cattle roundup using multiple quadrotor uavs," *International Journal of Aeronautical and Space Sciences*, vol. 18, no. 2, pp. 315–326, 2017.

[11] S. G. Matthews, A. L. Miller, T. PlÖtz, and I. Kyriazakis, "Automated tracking to measure behavioural changes in pigs for health and welfare monitoring," *Scientific reports*, vol. 7, no. 1, pp. 1–12, 2017.

[12] M. Tscharke and T. Banhazi, "Review of methods to determine weight and size of livestock from images," *Australian Journal of Multi-disciplinary Engineering*, vol. 10, no. 1, pp. 1–17, 2013.

[13] S. Tasdemir, A. Urkmez, and S. Inal, "Determination of body measurements on the holstein cows using digital image analysis and estimation of live weight with regression analysis," *Computers and electronics in agriculture*, vol. 76, no. 2, pp. 189–197, 2011.

[14] A. Cominotte, A. Fernandes, J. Dorea, G. Rosa, M. Ladeira, E. van Cleef, G. Pereira, W. Baldassini, and O. M. Neto, "Automated computer vision system to predict body weight and average daily gain in beef cattle during growing and finishing phases," *Livestock Science*, vol. 232, p. 103904, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1871141319310856

[15] A. Nasirahmadi, S. A. Edwards, and B. Sturm, "Implementation of machine vision for detecting behaviour of cattle and pigs," *Livestock Science*, vol. 202, pp. 25–38, 2017.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[17] A. dos Santos Ferreira, D. M. Freitas, G. G. da Silva, H. Pistori, and M. T. Folhes, "Weed detection in soybean crops using convnets," *Computers and Electronics in Agriculture*, vol. 143, pp. 314 – 324, 2017. [Online]. Available:

http://www.sciencedirect.com/science/article/pii/S0168169917301977
[18] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[19] S. Lathuilière, P. Mesejo, X. Alameda-Pineda, and R. Horaud, “A comprehensive analysis of deep regression,” *IEEE transactions on pattern analysis and machine intelligence*, 2019.
[20] G. Fanelli, J. Gall, and L. Van Gool, “Real time head pose estimation with random regression forests,” in *CVPR 2011*. IEEE, 2011, pp. 617–624.
[21] X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 2879–2886.
[22] V. Mnih, N. Heess, A. Graves *et al.*, “Recurrent models of visual attention,” in *Advances in neural information processing systems*, 2014, pp. 2204–2212.
[23] D. Anglart, “Automatic estimation of body weight and body condition score in dairy cows using 3d imaging technique,” 2014.
[24] D. Stajnko, M. Brus, and M. Hočevar, “Estimation of bull live weight through thermographically measured body dimensions,” *Computers and Electronics in Agriculture*, vol. 61, no. 2, pp. 233–240, 2008.
[25] A. F. Fernandes, J. R. Dórea, R. Fitzgerald, W. Herring, and G. J. Rosa, “A novel automated system to acquire biometric and morphological measurements and predict body weight of pigs via 3d computer vision,” *Journal of animal science*, vol. 97, no. 1, pp. 496–508, 2019.
[26] Y. Qiao, M. Truman, and S. Sukkarieh, “Cattle segmentation and contour extraction based on mask r-cnn for precision livestock farming,” *Computers and Electronics in Agriculture*, vol. 165, p. 104958, 2019.
[27] S. Li and A. B. Chan, “3d human pose estimation from monocular images with deep convolutional neural network,” in *Asian Conference on Computer Vision*. Springer, 2014, pp. 332–347.
[28] D. Crispell and M. Bazik, “Pix2face: Direct 3d face model estimation,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 2512–2518.
[29] K. Chen, J. Wang, L. Chen, H. Gao, W. Xu, and R. Nevatia, “ABC-CNN: an attention based convolutional neural network for visual question answering,” *CoRR*, vol. abs/1511.05960, 2015. [Online]. Available: http://arxiv.org/abs/1511.05960
[30] Y. Ji, H. Zhang, and Q. J. Wu, “Salient object detection via multi-scale attention cnn,” *Neurocomputing*, vol. 322, pp. 130 – 140, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231218311342
[31] J. Zhou, Z. He, Y. N. Song, H. Wang, X. Yang, W. Lian, and H.-N. Dai, “Precious metal price prediction based on deep regularization self-attention regression,” *IEEE Access*, 2019.
[32] S. Zhao, Z. Jia, H. Chen, L. Li, G. Ding, and K. Keutzer, “Pdanet: Polarity-consistent deep attention network for fine-grained visual emotion regression,” in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 192–201.
[33] M. Tan and Q. V. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” 2019.
[34] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385
[35] S. Kornblith, J. Shlens, and Q. V. Le, “Do better imagenet models transfer better?” *CoRR*, vol. abs/1805.08974, 2018. [Online]. Available: http://arxiv.org/abs/1805.08974
[36] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, “Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness,” 2018.