Smart Agricultural Technology 3 (2023) 100133

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Deep learning–based automated measurements of the scrotal circumference of Norwegian Red bulls from 3D images

Joanna Bremer [a,*], Michał Maj [b], Øyvind Nordbø [c], Elisabeth Kommisrud [a]

[a] Department of Applied Ecology, Agricultural Sciences and Biotechnology, Inland Norway University of Applied Sciences, Holsetgata 31, 2318, Hamar, Norway
[b] Climatica, Warsaw, Poland
[c] Norsvin SA, Storhamargata 44, 2317 Hamar, Norway

ARTICLE INFO

Keywords:
Convolutional Neural Networks
Semantic segmentation
U–Net
3D imaging
Scrotal circumference
Bull

ABSTRACT

The main aim of this study was to create an automated method for the measurement of the scrotal circumference (SC) of Norwegian Red bulls using 3D images of the scrotum based on convolutional neural networks. The study population was bull calves recruited for performance testing before the selection of bulls for semen production in the breeding program. Bulls were measured at four different time points: upon arrival in quarantine (Q) and thereafter at approximately 6, 9 and 12 months of age. Both 3D images and manual SC measurements were performed at all time points. In our approach, SC could be calculated without direct contact with the bull, using only 3D images and a simple, user–friendly application into which mentioned images are uploaded. The results show that SC measurements obtained using semantic segmentation are comparable with manual measurements. The mean prediction error was significantly different between age groups Q, 6, 9 and 12, and it was -3.07 cm, -3.02 cm, -1.79 cm and -1.11 cm, respectively. The results show a significant difference in the measurement error of the SC based on the quality of the images. Images were categorised into three quality groups. For good prediction accuracy, we recommend capturing 3D images of quality 2 – full circle from individuals older than 6 months.

1. Introduction

In dairy farms, automated herd control systems are used for multiple physiological and behavioural traits measurements such as estrus detection, calving time, lameness or pH of the rumen [1,2]. Scrotal circumference (SC) is an essential part of the breeding soundness evaluation of bulls due to its high repeatability and moderate to high heritability (from 0.36 to 0.69) [3]. Automation of SC measurement and implementation into feeding stations would be a valuable tool for performance testing stations and bovine semen collection centre. SC of bulls follows the sigmoidal growth pattern starting with an increase before six months of age and rapid growth during the peripubertal phase. It is broadly agreed that larger SC is associated with early puberty onset, increased sperm output and better fertility outcomes [3–7]. SC is influenced by age, body weight, nutrition (especially before 6 months of age) and breed with individual differences [5–7]. Traditional measurements of scrotum circumference are based on manual handling, using a measuring scrotal tape. This activity is associated with HSE (Health, Safety and Environment) issues for the technician and a stressful situation for the animal. Using 3D cameras combined with automatized image analysis might possibly constitute a good alternative.

Artificial Neural Networks (ANNs) are a type of mathematical model inspired by biological neural networks designed to mimic learning processes in the human brain. An ANN architecture is based on neurons (also called perceptrons) grouped in layers connected with each other using weights. The information is transferred from the input layer onto the output layer. ANN weights are fitted in a backpropagation process using a stochastic gradient descent algorithm to minimise a loss function, which corresponds to "learning" how to perform a specific task (regression, classification) by an ANN. Convolutional Neural Networks (CNNs) are a subclass of ANNs designed to perform complicated tasks on visual imagery (e.g. images, videos, spectrograms, holograms). CNNs are widely used for image classification, image segmentation, object detection, optical character recognition, etc. In contrast to classical ANNs, CNNs learn local filters that can be applied to the image data to extract interesting features. In CNNs those filters are calculated automatically in the backpropagation process, similar to those in ANNs [8–10].

* Corresponding author.
E-mail address: joanna.szlendak@inn.no (J. Bremer).

https://doi.org/10.1016/j.atech.2022.100133
Received 14 June 2022; Received in revised form 6 October 2022; Accepted 27 October 2022
Available online 1 November 2022

3D camera scanning combined with Convolutional Neural Networks (CNNs) or other machine learning algorithms was already studied with success in the field of animal production for body condition scoring of dairy and beef cattle [11–13], pigs [14] and horses [15]. Recently the 3D imaging combined with CNNs was used for supernumerary teat classification of Norwegian Red cattle udder [16]. Yang *et al.* [17] created a portable non–contact 3D measurement system for dairy cow body using smartphones. The point cloud hole completion method they used worked regardless of the posture of the animal, with relative errors for different traits from 2 to 6%. They showed that a low–cost method can be introduced for accurate non–contact measurement of livestock.

Deep learning methods such as CNNs were used in the field of reproduction to classify the human spermatozoa into WHO [18] shape–based morphology categories [19]. Butola *et al.* [20] combined a partially spatially coherent digital holographic microscope (PSC–DHM) for quantitative phase imaging (QPI) with deep neural networks (DNN) to differentiate with high accuracy normal human spermatozoa from abnormal. Another research used U–net image segmentation to automate the identification of the different stages of spermatogenesis in rats applied on stained testis tissue [21]. As seen in the presented examples, computer vision methods based on deep learning are increasingly used in fields of reproduction and other branches of biology, genetics, medicine and agriculture.

Our research's main aim was to explore the potential of using convolutional neural networks for automated SC measurements. Which would be of great interest to breeding companies and increase the accuracy of the SC measurements during breeding soundness evaluation. Our second objective was to investigate differences in prediction accuracy between age groups and image quality. Further, our practical goal was to create a framework of what is a good quality image to be directly uploaded into our user–friendly app and provide a quick and easy way of SC measurement in bulls without direct contact with the animals.

2. Materials and methods

2.1. Animals

Geno (*Geno*), the breeding organisation for Norwegian Red (NR), each year buys approximately 150 NR bull calves for their performance testing program [22]. Individuals aged 3-5 months arrive in 5-6 groups per year at the testing station and are quarantined for two weeks upon arrival. After isolation, they are housed in groups of 10 and consequently kept in the same group for the whole duration of the performance testing period. Bulls are subject to temperament, conformation, and andrology testing at the station. Around 12 months of age, they are approved or rejected to the bovine semen collection centre. Bulls are fed concentrate according to the age and grass silage ad libitum. This study was performed during a period for 1.5 years and included bulls enrolled in the performance testing program during this period.

2.2. Manual Scrotal Circumference measurements

The SC of NR bulls were measured manually at four–time points: upon arrival at the performance testing station (3-5 months) and later at approximately 6, 9 and 12 months of age. The bulls were restrained during the procedure. Three qualified veterinarians under the supervision of centre veterinarian measured the SC manually by scrotal tape for all bulls and time points.

2.3. 3D Scrotal Circumference measurements

After manual SC measurements, each bull's scrotum was photographed using a handheld device consisting of an Intel Real Sense d415 camera connected with a tablet by a stick [23]. The camera was carefully placed on the floor between the bull's legs. Qualified personnel assisted the procedure by keeping the tail of the restricted animal. One image per bull was captured at each time point. In total, four images per bull were taken. For five individuals, multiple images on the same day were captured and used for quality control of the method. Depth images were saved in the OneDrive cloud and used for further analysis. The number of animals per age group per measurement is shown in Table 1. Due to technical issues, the schedule of the project, and losses in animals, a few observations in each age group were lost. The proportion of the training, validation and test data were 60%, 20%, and 20%, respectively.

Table 1
Number of animals per age group (Q[1],6,9,12 months) per method of measurement used in our study.

Age group	n SC [2]	n 3D [3]
Q	96	76
6	111	99
9	137	131
12	123	116

[1] Quarantine - age 3-5 months
[2] Number of bulls from which scrotum circumference (SC) was manually measured
[3] Number of bulls from which 3D pictures of scrotum were collected

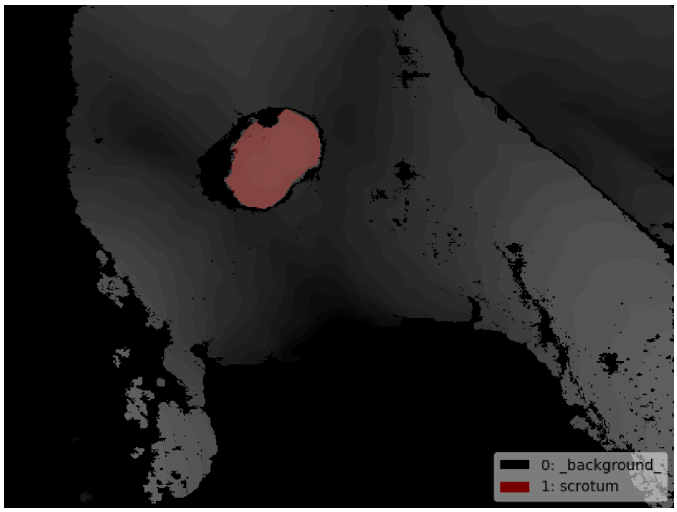

Fig. 1. Image segmentation visualisation. An example of a true scrotum segmentation mask (1) overlapped with the original 3D image of bull.

2.4. Artificial Neural Network (ANN)

In our approach, we decided to use CNNs, specifically the U–Net architecture; it is already successfully used in biomedicine biology and genetics for image semantic segmentation [24–26]. In an image classification task, we are interested in predicting the correct label for the whole image (e.g. cat or dog). In the semantic segmentation task, this approach was extended for each pixel of the image. This created a ground truth segmentation mask, showing the localisation of pixels (objects) belonging to the same class. This predicted segmentation mask was compared with the ground truth segmentation mask to check if a prediction was correct or not (Fig. 1). This task's most common compatibility measures are the Dice coefficient, the IoU coefficient and the Tversky coefficient which represents coverage between ground truth and predicted segmentation masks. In our approach, we created ground truth segmentation masks of the bull scrotum using the 'labelme' tool [27]. For each image the scrotum boundary was outlined using 30 to 40 unique points. Pixels inside the boundary were marked as 'scrotum' and outside the boundary as 'background'. Those segmentation masks were used to train, validate and test the U–Net model.

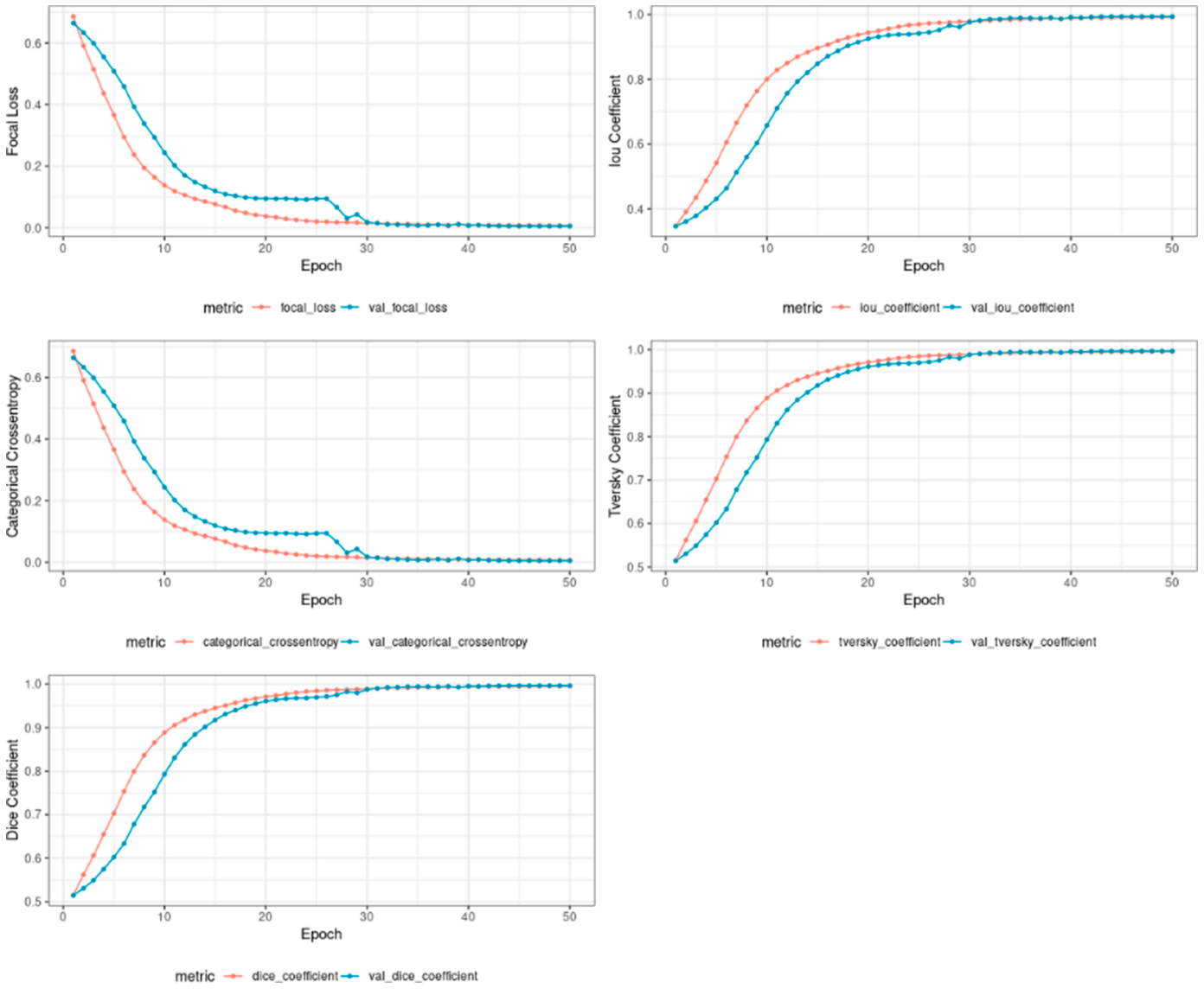

Fig. 2. Learning curves for Focal loss, IoU coefficient, Categorical Crossentropy, Tversky coefficient and Dice coefficient, respectively for training (red) and validation (blue) sets.

2.5. Model architecture

In the U–Net model used in our study for scrotum segmentation, the input image was a tensor of shape 256 × 512 × 1, rescaled from the original image 640 × 480 × 1. Our model was composed of 5 double–convolutional downscaling blocks – each block included two iterations of 2D convolution with the same padding (the activation map had the same shape as input) and batch normalisation layers with ReLU activation followed by 2D max–pooling and dropout layers – and 5 deconvolutional upscaling blocks – each block included 2D deconvolution and concatenation layers followed by two iterations of 2D convolution with same padding (the activation map had the same shape as input) and batch normalisation layers with ReLU activation. Downscaling and upscaling blocks were connected with a bridge composed of two iterations of 2D convolution with same padding (the activation map had the same shape as input) and batch normalisation layers with ReLU activation. The model was fitted using Adam optimizer with a Focal loss function (Fig. 2). The model was build, trained and validated using 'pyplatypus' software (https://github.com/maju116/pyplatypus).

Fig. 3. Multiple objects as a result of segmentation. Example of predicted scrotum segmentation mask containing 2 objects (scrotum and incorrect artefact).

2.6. Connected–component labeling (CCL) algorithm

For most of the images, the predicted segmentation mask contained one solid object, which was expected and desirable. Some artefacts that created a second smaller object for the remaining images were found in the predicted segmentation mask (Fig. 3). To solve this problem, we used a connected–component labeling (CCL) algorithm (also known as blob extraction or region labeling) to count the number of solid objects in a predicted segmentation mask [28]. In the case of finding more than

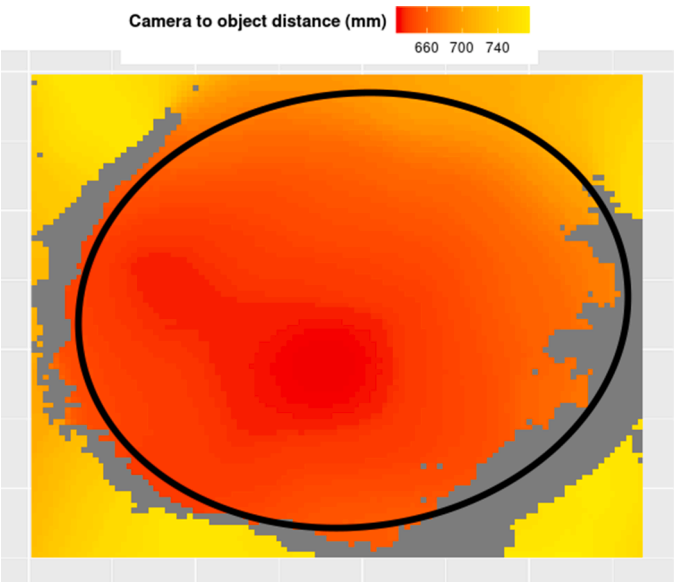

Fig. 4. Direct Linear Least Squares fitting of an ellipse. An example of an ellipse fitted onto the predicted scrotum segmentation mask using the Direct Linear Least Squares method is plotted onto the original depth image. The colour scale represents the camera to object distance in mm. The grey represents NA.

one object, only the object with the highest area would stay on the segmentation mask.

2.7. Direct Linear Least Squares fitting of an ellipse

After segmentation and cleaning faze, the Direct Linear Least Squares algorithm was used to fit an ellipse onto the boundary of a segmented mask (Fig. 4) [29,30].

From this fit, semi–major and semi–minor axes of the ellipse were calculated using angles per pixel from the median camera to object distance. The angles used in the angles per pixel calculation 66.5° x 40.5° were selected from the metafile. The ellipse perimeter was calculated using the Pade approximation [31].

2.8. Validation of capturing the 3D SC images and statistical analysis

To validate the method, multiple images from the same individuals were analysed and subsequently compared with their predicted 3D SC. The results were repeatable with a difference of +/-1cm. If the difference was higher, the quality of the image was low. Metafiles from the 3D camera contained angle range DPFOV: 65°±2° x 40°±1°, identical for every image. The angles are required to calculate the angle per pixel, which were further used for the 3D SC calculations. To choose one pair of angles, we predicted the 3D SC for a combination of angles and compared the results, which showed a difference of +/-1cm. We chose the following pair of angles based on this validation: 66.5° x 40.5°.

The predicted 3D SC values and their corresponding manual measurements were used for the calculation of mean prediction error (MPE), mean squared prediction error (MSPE) and mean percentage prediction error (MPPE) adjusted by the mean of the group for different age groups (Q,6,9,12 months) and different image quality categories (0,1,2). The data were tested for normality, and the nonparametric Scheirer–Ray–Hare test was performed to evaluate significant differences in MPE $p < 0.01$ between age groups and the image quality categories in each age group.

3. Results

The values of Dice, IoU and Tversky coefficients for the training, validation and test sets are represented in the Table 2. Mean prediction errors for each age group Q, 6, 9, 12 were -3.07 cm, -3.02 cm, -1.79 cm and -1.11 cm respectively. Fig. 7 shows the distribution of prediction error for each age group. Mean and percentage prediction errors were significantly different ($p < 0.05$) between the age groups (Table 3). Age group 12 showed significantly lowest MPE and MPPE. While capturing 3D images, we observed that natural or artificial light influenced the quality of the images significantly. Fig. 5 shows the 3D images and their matching RBG images of one individual in two different light conditions.

To improve future prediction accuracy, we conducted a test of image quality control. Each 3D image was rated based on the quality of the image. The scale was the following: 2 – full circle, 1– partial circle / "hanging testicle", 0 – not enough information/scrotum not well pronounced (Fig. 6).

The results show that the quality of image category 2 shows

Table 2
Values of Dice, IoU and Tversky coefficients for the training, validation and test sets.

Metric	Training set	Validation Set	Test Set
Dice	99,8%	99,5%	99,3%
IoU	99,3%	99,4%	99,2%
Tversky	99,4%	99,3%	99,2%

Table 3
Mean prediction error (MPE), mean squared prediction error (MSPE) and mean percentage prediction error (MPPE) adjusted by mean of the group for different age groups (Q[1],6,9,12 months). Based on the nonparametric Scheirer–Ray–Hare test, significant differences in MPE $p < 0.05$ between the age groups are marked by letters [a-d].

Age group	n images [2]	MPE [3]	MSPE[4]	MPPE [5]	mean SC (cm) [6]
Q [a]	57	-3.07 cm [cd]	16.6	-19.9 %	15 cm
6 [b]	88	-3.02 cm [cd]	13.9	-14.0 %	22 cm
9 [c]	124	-1.79 cm [abd]	7.89	-6.16 %	29 cm
12 [d]	101	-1.11 cm [abc]	7.46	-3.28 %	34 cm

[1] Quarantine – age 3-5 months
[2] Number of analysed images
[3] MPE – Mean of (SC-prediction)
[4] MSPE – Mean of (SC-prediction)^2
[5] MPPE – Mean of [(SC-prediction)/group_mean * 100%]
[6] Mean manually measured SC in cm

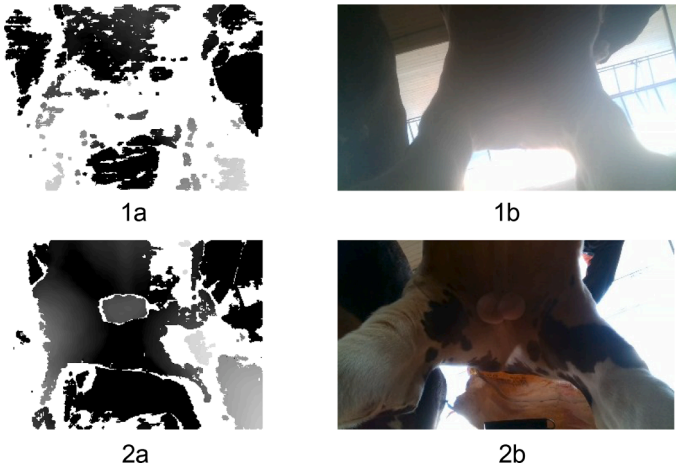

Fig. 5. Influence of natural light on 3D image quality for the same individual. 1a: 3D image with high exposure to natural light. 1b: RGB image matching 3D image, scrotum not visible. 2a: 3D image with the covered back of the bull to block natural light in the area of interest. 2b: RGB image matching 3D image, scrotum visible.

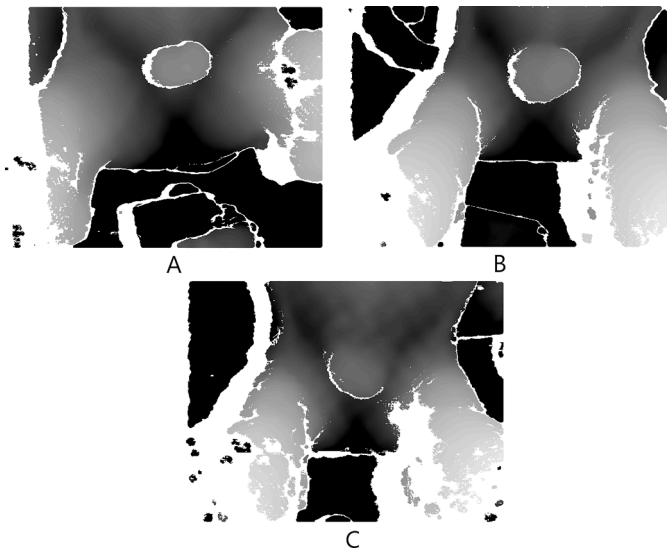

Fig. 6. Image quality categories. Examples of the 3D images from the category 2 – full circle (A), 1 – partial circle/" hanging testicle" (B), 0 – not enough information/scrotum not well pronounced (C).

significantly lower mean prediction error and mean percentage prediction error in age groups 6, 9 and 12 compared to image quality 0 and 1 as well as the mix of all images (Tables 3 and 4). For group Q, we can observe the lowest MPE and MPPE for image quality 0.

4. Discussion

Our results show that the quality of the image and the age of the bull are important variables affecting the accuracy of the prediction of SC based on 3D images. Three–month–old bull calves have small testicles that are not very well pronounced in the picture compared to older bulls. Furthermore, the intensity and quantity of natural and artificial light seem to influence the images' quality, as shown in Fig. 5. The Intel RealSense d415 depth sensor is mainly based on stereoscopy. Therefore, it is differently dependent on light conditions than sensors using structured light (like e.g. Kinect v1) or time of flight (like e.g. Kinect v2). In our case, strong backlight (Figs. 5, 1a and 1b) gave poorer image quality because it generates too large light contrasts in the scene, with a very light background compared with the relative dark scrotum area. On the other hand, outdoor light should improve the quality of depth images taken by the Intel RealSense d4XX cameras [32], but this probably requires that the object of interest is properly illuminated. Similarly, based on other types of depth camera technology, previous research has observed that strong illumination affects the quality of the depth images [33–35]. Azzari, Goulden and Rusu [33] showed that the quality of images increased with decreasing light exposure. Azzari, Goulden and Rusu, [33] observed that sunlight and infrared radiation influenced the contrast of the camera's laser pattern, which caused the lower quality of the images.

Dice, IoU and Tversky coefficients used in our study as a performance measure for semantic segmentation were chosen already by others [36–41]. A high Dice, IoU and Tversky coefficients of our model can be explained by the fact that a scrotum is a big object easily detectable by the human eye, which could imply that it should be just as easy for an advanced computer vision algorithm as a convolutional neural network.

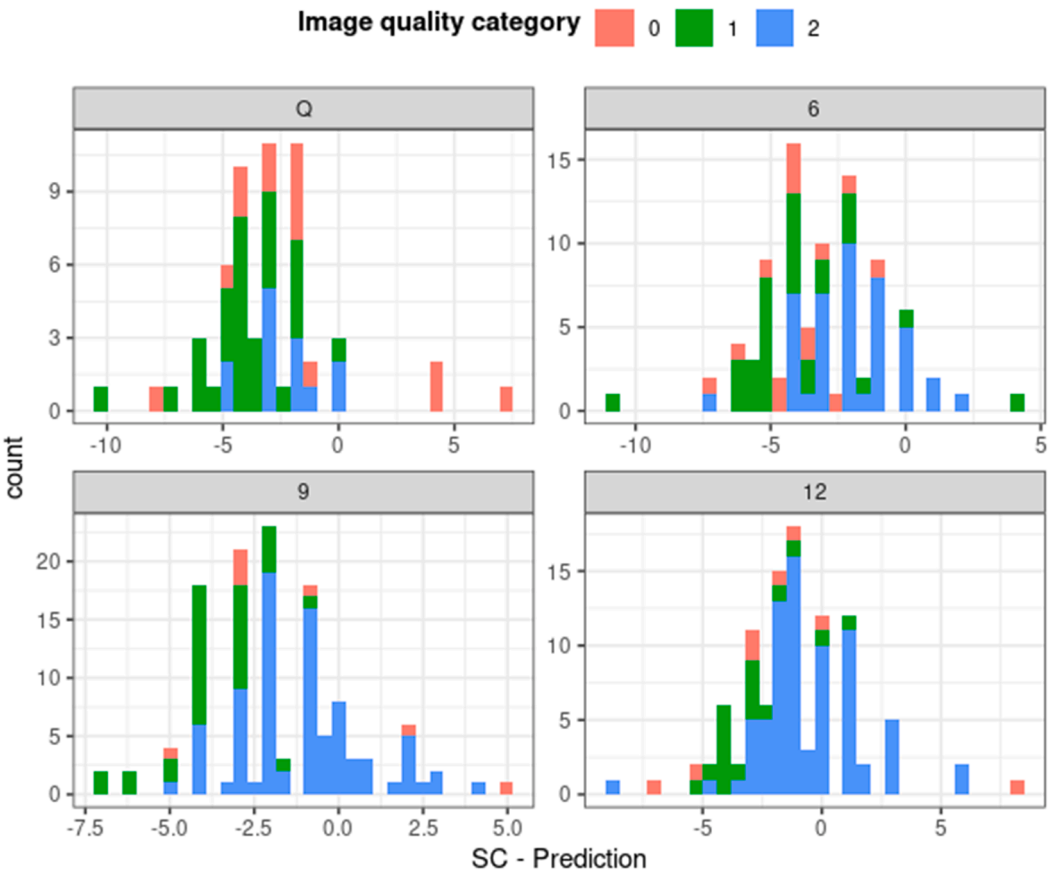

Fig. 7. Distribution of prediction errors. Histograms of prediction errors (SC – prediction) for different bull age groups (Q, 6, 9, 12) with the distinction of image quality (0, 1, 2).

Table 4
Mean prediction error (MPE)), mean squared prediction error (MSPE) and mean percentage prediction error (MPPE) adjusted by mean of the group for different age groups (Q[1],6,9,12 months) and different image quality categories (0[2],1[3],2[4]). Based on the nonparametric Scheirer–Ray–Hare test, significant differences in MPE $p < 0.01$ between the image quality categories in each age group are marked by letters [a-c].

Age group	Image quality category	n images [5]	MPE [6]	MSPE[7]	MPPE [8]	mean SC (cm) [9]
Q	0 [a]	14	-1.5 cm [bc]	16.9	-10 %	15 cm
Q	1 [b]	30	-4.1 cm [ac]	20.1	-26.3 %	15 cm
Q	2 [c]	13	-2.5 cm [ab]	8.31	-15.6 %	15 cm
6	0 [a]	14	-3.9 cm [bc]	17.4	-18.9 %	22 cm
6	1 [b]	31	-4 cm [ac]	22.2	-19.3 %	22 cm
6	2 [c]	43	-2 cm [ab]	6.80	-8.9 %	22 cm
9	0 [a]	7	-1.1 cm [bc]	11.7	-3.8 %	29 cm
9	1 [b]	33	-3.7 cm [ac]	15.5	-12.8 %	29 cm
9	2 [c]	84	-1.1 cm [ab]	4.58	-3.8 %	29 cm
12	0 [a]	8	-1.63 cm [bc]	20.1	-4.8 %	34 cm
12	1 [b]	18	-2.97 cm [ac]	11.1	-9.1 %	34 cm
12	2 [c]	75	-0.6 cm [ab]	5.21	-1.8 %	34 cm

[1] Quarantine – age 3-5 months
[2] Full circle
[3] Partial circle/"hanging testicle"
[4] Not enough information/scrotum not well pronounced
[5] Number of analysed images
[6] MPE – Mean of (SC-prediction)
[7] MSPE – Mean of (SC-prediction)^2
[8] MPPE – Mean of [(SC-prediction)/group_mean * 100%]
[9] Mean manually measured SC in cm

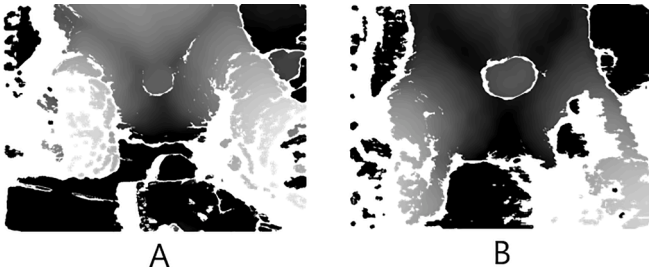

Fig. 8. Age differences and 3D image quality. Examples of 3D pictures of a three–month–old NR bull calf (A) and twelve–month–old NR bull (B).

Images used in the study are standardised since the scrotum, in most cases, can be found in the centre of the image. Similar methods to the ones used in our study were applied to the field of horticulture for the growth rate, size and yield measurement of the fruits and vegetables [33, 42–46]. The successful easy–to–use prediction of the in–field mango fruit size was performed with high accuracy by Wang, Walsh and Verma [46]. As a scrotum, the mango has an approximately elliptical shape. The authors used this to predict the mango's length and width, concluding that measurement error is caused by the not entirely elliptical shape of the fruits, which agrees with our findings. In our study, during manual measurement, bulls were moving from side to side, and as a reaction to the stressful situation, some bulls pulled the scrotum closer to the body using cremaster muscle, which might have influenced the measurements. Above mentioned variables assess our prediction error complex since we do not know the "real true" value. We believe that the best way of reducing prediction error is to standardise image quality. As our results demonstrate, the lowest MPE and MPPA were achieved for the image quality 2 in age groups 6, 9 and 12, making it the best candidate for the reference image quality for future analysis. The quality category 0 exhibits the lowest error measurements in age groups Q and 9, which we consider false due to the significant lack of information on the images of quality 0 (Fig. 8). This result could speak for the excellent performance of our segmentation model (dice coefficient – 99, 8%, 99,3%), which succeeded in classifying the scrotum of the incomplete image correctly. It is of importance to point out a very low number of images analysed in image quality category 0 in all age groups, which can explain this result. Future research should be devoted to the validation of the results described in this paper by capturing 3D SC images of only image quality 2 and performing manual SC measurements of the same individuals. Our results indicate that the mean prediction error and mean percentage prediction error in all age groups will not increase if the quality of the image, including light conditions, is taken into consideration. On top of a collection of new samples, data augmentation methods could be beneficial to model performance. Another interesting development opportunity would be machine learning model ensembling. We noticed that applying even a simple linear model (Fig. 9) on top of the previously described modelling methodology could potentially decrease prediction even more. This hypothesis, however, has to be confirmed.

5. Conclusion

To our knowledge, this is the first time SC measurements of bulls were automated with the use of convolutional neural networks and 3D images. This innovative approach, combined with a user–friendly application, allows a fast integration into breeding soundness evaluation of Norwegian Red bulls at the performance testing and bovine semen collection centres. To keep a high prediction accuracy, we recommend analysing individuals older than 6 months, paying attention to light conditions and capturing 3D images of quality 2 only.

Ethical statement

Ethical approval was not required in this study. We worked at the performance testing station under the supervision of a qualified veterinarian employed at the breeding company Geno. The performance testing station fulfils Norwegian legislation for the housing of bulls.

Role of the funding source

This work was funded by the internal scholarship of Inland Norway University of Applied Sciences sponsored by Sparebankstiftelsen Hedmark.

CRediT authorship contribution statement

Joanna Bremer: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Visualization, Project administration. **Michał Maj:** Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Writing – original draft, Visualization. **Øyvind Nordbø:** Conceptualization, Methodology, Resources, Writing – original draft. **Elisabeth Kommisrud:** Conceptualization, Methodology, Writing – original draft, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

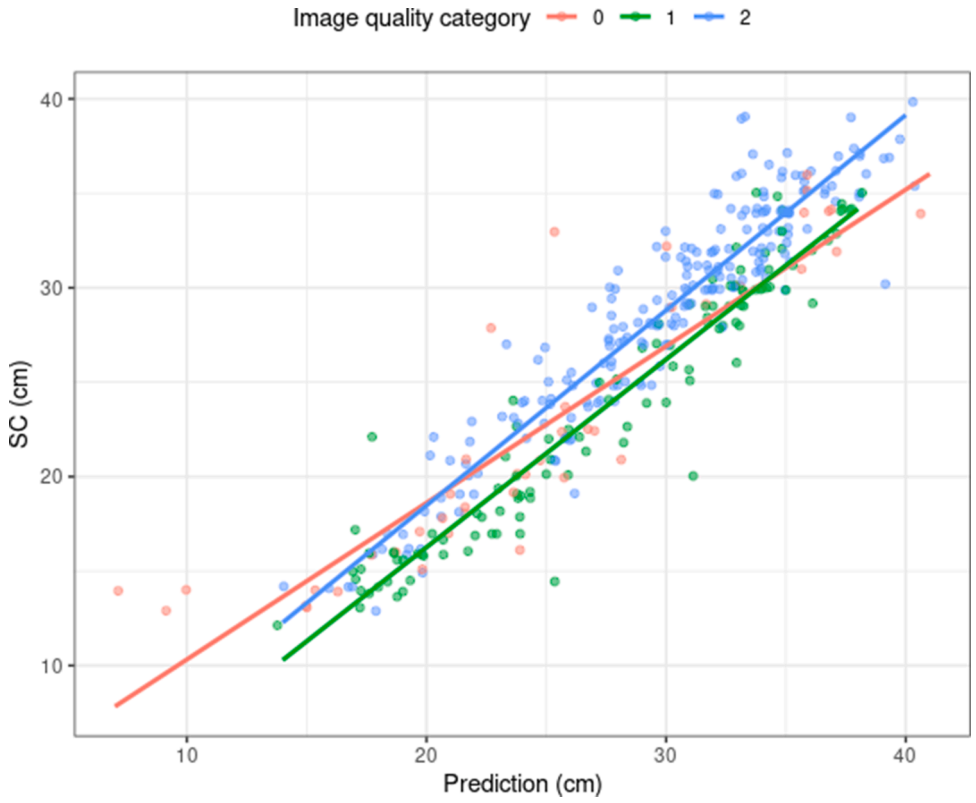

Fig. 9. Fitted linear regression models for the scrotum circumference (cm) based on the prediction (cm) value obtained from the fitted ellipse for different image quality (0,1, 2) groups.

Data availability

The authors do not have permission to share data.

Acknowledgements

This study used a 3D camera from the project supported by the Research Council of Norway under the BIONÆR program, project number 282252, "New traits in pigs and cattle based on 3D imaging technology." We would like to thank Hallstein Holen and his team at the Geno performance testing station: Jan Tore Rosingholm, Erik Skogli, Sigmund Høibakken and Stein Marius Brumoen for their time and indispensable help with data collection. The authors thank Prof. Jane Morrell and Prof. Bjørg Heringstad for their valuable ideas and discussions.

References

[1] G. Grodkowski, et al., Comparison of different applications of automatic herd control systems on dairy farms – a review, J. Sci. Food Agric. 98 (14) (2018) 5181–5188, https://doi.org/10.1002/jsfa.9194.
[2] X. Kang, et al., Dimension-reduced spatiotemporal network for lameness detection in dairy cows, in: Computers and Electronics in Agriculture, 197, Elsevier B.V., 2022, 106922, https://doi.org/10.1016/j.compag.2022.106922.
[3] J.M. Penitente-Filho, et al., Can scrotal circumference-based selection discard bulls with good productive and reproductive potential? PLoS One 13 (3) (2018) 1–14, https://doi.org/10.1371/journal.pone.0193103.
[4] H. Bollwein, F. Janett, M. Kaske, Impact of nutritional programming on the growth, health, and sexual development of bull calves, in: Domestic Animal Endocrinology, 56, Elsevier, 2016, pp. S180–S190, https://doi.org/10.1016/j.domaniend.2016.02.006.
[5] J.P. Kastelic, Understanding and evaluating bovine testes, in: Theriogenology, 81, Elsevier Inc, 2014, pp. 18–23, https://doi.org/10.1016/j.theriogenology.2013.09.001.
[6] N. Rawlings, et al., Sexual Maturation in the Bull, Reprod. Domest. Anim. 43 (SUPPL.2) (2008) 295–301, https://doi.org/10.1111/j.1439-0531.2008.01177.x.
[7] R.K. Waite, et al., Scrotal circumference, bodyweight and semen characteristics in growing dairy-breed natural-service bulls in Tasmania, Australia, N. Z. Vet. J. 67 (3) (2019) 109–116, https://doi.org/10.1080/00480169.2018.1563512.
[8] S. Albawi, T.A.M. Mohammed, S. Alzawi, Understanding of a Convolutional Neural Network, in: International Conference on Engineering and Technology (ICET), 2017, pp. 1–6, https://doi.org/10.1109/ICEngTechnol.2017.8308186.
[9] K. He, et al., Deep residual learning for image recognition, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem, 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.
[10] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015, pp. 1–14.
[11] T. KOJIMA, et al., Estimation of beef cow body condition score: a machine learning approach using three-dimensional image data and a simple approach with heart girth measurements, in: Livestock Science, 256, Elsevier B.V., 2022, 104816, https://doi.org/10.1016/j.livsci.2021.104816.
[12] M. Shigeta, et al., Automatic measurement and determination of body condition score of cows based on 3D images using CNN, J. Robot. Mechatronics 30 (2) (2018) 206–213, https://doi.org/10.20965/jrm.2018.p0206.
[13] Y. Tao, F. Li, Y. Sun, Development and implementation of a training dataset to ensure clear boundary value of body condition score classification of dairy cows in automatic system, in: Livestock Science, 259, Elsevier B.V., 2022, 104901, https://doi.org/10.1016/j.livsci.2022.104901.
[14] S. Shuai, et al., Research on 3D surface reconstruction and body size measurement of pigs based on multi-view RGB-D cameras, in: Computers and Electronics in Agriculture, 175, Elsevier, 2020, 105543, https://doi.org/10.1016/j.compag.2020.105543.
[15] F. Pallottino, et al., Comparison between manual and stereovision body traits measurements of Lipizzan horses, in: Computers and Electronics in Agriculture, 118, Elsevier B.V., 2015, pp. 408–413, https://doi.org/10.1016/j.compag.2015.09.019.
[16] H. Afridi, M. Ullah, Ø. Nordbø, F.A. Cheikh, et al., Optimized Deep-Learning-Based Method for Cattle Udder Traits Classification, Mathematics 10 (17) (2022) 3097, https://doi.org/10.3390/math10173097.
[17] G. Yang, et al., Automated measurement of dairy cows body size via 3D point cloud data analysis, in: Computers and Electronics in Agriculture, 200, Elsevier B.V., 2022, 107218, https://doi.org/10.1016/j.compag.2022.107218.
[18] World Health Organization, WHO laboratory manual for the examination and processing of human semen Sixth Edition, World Health Organization, 2021. Available at: http://whqlibdoc.who.int/publications/2010/9789241547789_eng.pdf.
[19] Riordon, J., Mccallum, C. and Sinton, D. (2019) 'Deep learning for the classification of human sperm', 111(June).

[20] A. Butola, et al., High spatially sensitive quantitative phase imaging assisted with deep neural network for classification of human spermatozoa under stressed condition, in: Scientific Reports, 10, Nature Publishing Group UK, 2020, pp. 1–12, https://doi.org/10.1038/s41598-020-69857-4.
[21] D.M. Creasy, et al., Deep learning-based spermatogenic staging assessment for hematoxylin and eosin-stained sections of rat testes, Toxicol. Pathol. 49 (4) (2021) 872–887, https://doi.org/10.1177/0192623320969678.
[22] *Startpage*, Geno, |, 2022. Available at: https://www.geno.no/en/. Accessed: 31 May 2022.
[23] H. Afridi, M. Ullah, Ø. Nordbø, F. Alaya Cheikh, Deep learning based udder classification for cattle traits analysis, in: IS&T International Symposium on Electronic Imaging 2022 7, 2022, https://doi.org/10.2352/EI.2022.34.10.IPAS-390, 390-1-390-6.
[24] M. Benazzouz, M.L. Benomar, Y. Moualek, Modified U-Net for cytological medical image segmentation, Int. J. Imaging Syst. Technol. (March) (2022) 1761–1773, https://doi.org/10.1002/ima.22732.
[25] H. Lin, et al., Variance-aware attention U-Net for multi-organ segmentation, Med. Phys. 48 (12) (2021) 7864–7876, https://doi.org/10.1002/mp.15322.
[26] J. Wu, et al., A medical assistant segmentation method for MRI images of osteosarcoma based on DecoupleSegNet, Int. J. Intell. Syst. (June) (2022), https://doi.org/10.1002/int.22949.
[27] Kentaro, W. (2021) 'Labelme: Image Polygonal Annotation with Python'. doi: 10.5281/zenodo.5711226.
[28] L. He, et al., The connected-component labeling problem: A review of state-of-the-art algorithms, in: Pattern Recognition, 70, Elsevier Ltd, 2017, pp. 25–43, https://doi.org/10.1016/j.patcog.2017.04.018.
[29] A. Fitzgibbon, M. Pilu, R.B. Fisher, Direct least square fitting of ellipses, IEEE Trans. Pattern Anal. Mach. Intell. 21 (5) (2000), https://doi.org/10.1109/34.765658.
[30] R. Halir, J. Flusser, Numerically stable direct least squares fitting of ellipses, in: Conference in Central Europe on Computer Graphics and Visualization WSCG, 1998, p. 98.
[31] Roberts, M. (2020) A Formula for the Perimeter of an Ellipse. Available at: http://extremelearning.com.au/a-formula-for-the-perimeter-of-an-ellipse/. (Accessed 17 March 2022).
[32] A. Grunnet-Jepsen, et al., Projectors for D400 Series Depth Cameras, 2020. Available at: https://dev.intelrealsense.com/docs/projectors. Accessed: 29 May 2022.
[33] G. Azzari, M.L. Goulden, R.B. Rusu, Rapid characterization of vegetation structure with a microsoft kinect sensor, Sensors (Switzerland) 13 (2) (2013) 2384–2398, https://doi.org/10.3390/s130202384.
[34] Y. Chéné, et al., On the use of depth camera for 3D phenotyping of entire plants, Comput. Electron. Agric. 82 (2012) 122–127, https://doi.org/10.1016/j.compag.2011.12.007.
[35] Y. Jiang, et al., Quantitative analysis of cotton canopy size in field conditions using a consumer-grade RGB-D camera, Front. Plant Sci. 8 (January) (2018) 1–20, https://doi.org/10.3389/fpls.2017.02233.
[36] K.O. Babalola, et al., Comparison and evaluation of segmentation techniques for subcortical structures in brain MRI, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5241 LNCS(PART 1) (2008) 409–416, https://doi.org/10.1007/978-3-540-85988-8_49.
[37] R. Cárdenes, R. de Luis-García, M. Bach-Cuadra, A multidimensional segmentation evaluation for medical image data, Comput. Methods Programs Biomed. 96 (2) (2009) 108–124, https://doi.org/10.1016/j.cmpb.2009.04.009.
[38] H. Khotanlou, et al., 3D brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models, Fuzzy Sets Syst. 160 (10) (2009) 1457–1473, https://doi.org/10.1016/j.fss.2008.11.016.
[39] S. Klein, et al., Segmentation of the prostate in MR images by atlas matching, in: *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro - Proceedings*, 2007, pp. 1300–1303, https://doi.org/10.1109/ISBI.2007.357098.
[40] K.H. Zou, S.K. Warfield, et al., Statistical validation of image segmentation quality based on a spatial overlap index, Acad. Radiol. 11 (2) (2004) 178–189, https://doi.org/10.1016/S1076-6332(03)00671-8.
[41] K.H. Zou, W.M. Wells, et al., Three validation metrics for automated probabilistic image segmentation of brain tumours, Stat. Med. 23 (8) (2004) 1259–1282, https://doi.org/10.1002/sim.1723.
[42] D. Andújar, et al., Using depth cameras to extract structural parameters to assess the growth state and yield of cauliflower crops, Comput. Electron. Agric. 122 (2016) 67–73, https://doi.org/10.1016/j.compag.2016.01.018.
[43] M. Fukuda, T. Okuno, S. Yuki, Central object segmentation by deep learning to continuously monitor fruit growth through rgb images, Sensors 21 (21) (2021) 1–18, https://doi.org/10.3390/s21216999.
[44] D. Wang, et al., Deep Learning Approach for Apple Edge Detection to Remotely Monitor Apple Growth in Orchards, IEEE Access 8 (2020) 26911–26925, https://doi.org/10.1109/ACCESS.2020.2971524.
[45] Z. Wang, et al., In field fruit sizing using a smart phone application, Sensors (Switzerland) 18 (10) (2018), https://doi.org/10.3390/S18103331.
[46] Z. Wang, K.B. Walsh, B. Verma, On-tree mango fruit size estimation using RGB-D images, Sensors (Switzerland) 17 (12) (2017) 1–15, https://doi.org/10.3390/s17122738.