

Counting Livestock with Image Segmentation Neural Network^{*}

Petr Dolezel¹[0000-0002-7359-0764], Dominik Stursa¹[0000-0002-2324-162X],
Daniel Honc¹[0000-0001-8440-5272], Jan Merta¹[0000-0002-1521-5050], Veronika
Rozsivalova¹[⁻], Ladislav Beran¹[⁻], and Ivo Hora¹[⁻]

University of Pardubice, Studentska 95,
53210 Pardubice, Czech Republic
petr.dolezel@upce.cz
<http://www.upce.cz/fei>

Abstract. Livestock farming industries, as well as almost any industry, want more and more data about the operation of their business and activities in order to make the right decisions. However, especially when considering very large animal farms, the precise and up-to-date information about the position and numbers of the animals is rather difficult to obtain. In this contribution, a novel engineering approach to livestock positioning and counting, based on image processing, is proposed. The approach is composed of two parts. Namely, a fully convolutional neural network for input image transformation, and a locator for animal positioning. The transformation process is designed in order to transform the original RGB image into a gray-scale image, where animal positions are highlighted as gradient circles. The locator then detects the positions of the circles in order to provide the positions of animals. The presented approach provides a precision rate of 0.9842 and a recall rate of 0.9911 with the testing set, which is, in combination with a rather suitable computational complexity, a good premise for the future implementation under real conditions.

Keywords: Livestock Counting · Fully Convolutional Neural Network · U-Net · Precision Agriculture.

1 Introduction

Livestock farming industries, as well as almost any industry, want more and more data about the operation of their business and activities, in order to make the right decisions, at the right location, at the right time, and at the right intensity. These days, with the development of precision agriculture [3], farmers can acquire more data than ever before, including soil moisture and acidity, ground and air temperature, individual stock or crop increments, etc.

^{*} The work has been supported by SGS grant at Faculty of Electrical Engineering and Informatics, University of Pardubice, Czech Republic. This support is very gratefully acknowledged.

Nevertheless, especially when considering very large farms, the precise and up-to-date information about the position and numbers of the animals is still difficult to obtain. Counting livestock is often performed once in a time period, and animals have to be led through a drafting race or a narrow choke point, while being counted manually or using some sort of smart collar [5]. This approach does not provide the information continuously and it can also be uncomfortable for the animals. Hence, continuous, or at least more frequent, livestock counting directly in the pastureland is a desired task, which can be then applied for monitoring of animal numbers, animal growth, animal distress, distribution of the herds, etc.

In order to solve the task, several challenges, such as species characteristic, diversity of background, variable light conditions, overlapping of animals, animal reaction to monitoring and the mosaicking process appear [1], [12]. Hence, various approaches are proposed to deal with those challenges. Some of them are based on classical statistical techniques [11]. Others use more recent methods, such as K-means clustering [10], histogram of oriented gradients and local binary pattern [6], power spectral based methods [13], support vector machines with various sound processing approaches [4], etc.

Considering image or video as the input signal for livestock counting, deep learning techniques become one of the major approaches for implementation. Very good results were provided especially by the methods based on convolutional neural networks [7], [8], [16]. However, the cited approaches were tested on images, where animals occupy a substantial part of the image and each animal is depicted in high resolution. Contrary to these approaches, Farah Sarwar and Anthony Griffin published an approach to deal with images having hundreds of small animal silhouettes per image [15]. Their testing experiments provided a precision rate of 95.6 % and recall rate of 99.5 %. However, the dataset used in the cited work did not contain spacial clusters of animals and the diversity of the background was rather low.

Therefore, we focus on counting livestock animals from an Unmanned Aerial Vehicle (UAV) video, considering high angle take from an altitude bigger than 50 meters. We consider various types of background, different size of animals, big variance of animal numbers including crowded animals stocks in the source signal.

The rest of the contribution is structured as follows. In the next chapter, the aim of the paper is formulated and the proposed solution is described. Then, the implementation part follows and the results are presented. The paper is finished with some conclusions.

2 Methodology

In this section, the aim of this paper is defined and a method, which is based on image segmentation neural network, is presented to solve the problem.

2.1 Problem formulation

The aim of the paper is to design a monitoring system for livestock positioning and counting, using images acquired by the UAV. The monitoring system should be robust enough to handle various light conditions and background types, size of animals and both crowded and blank situations. The examples of these variants are shown in Fig. 1.



Fig. 1. Examples of livestock considered to be processed by the proposed system.

The proposed monitoring system is supposed to work as shown in Fig. 2. The UAV continuously takes the images of the area. The images are then properly combined together through mosaicking. During mosaicking, the images are also geometrically distorted to fit to the detailed map of the scanned area. Subsequently, the resulting complex image is processed to provide positions of each individual animal. Such information can be then used for various analyses required by farmers.

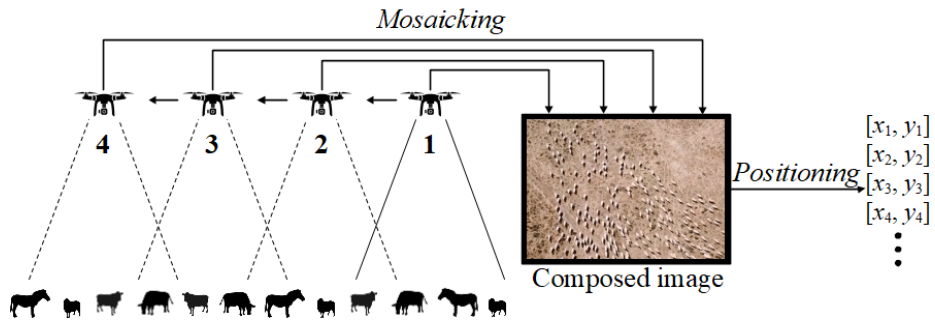


Fig. 2. Proposed functionality of the monitoring system.

In this particular paper, the process of mosaicking is not discussed in detail. It is clearly a non-trivial process and has the potential to produce various kinds of errors [1]. Hence, it will be comprehensively dealt with in authors' future work. Here, the positioning of the animals in the combined image is dealt with. In other words, we propose a novel approach to detect, to locate and to count livestock animals in visual data.

2.2 Proposed solution

Recent advance in convolutional neural networks (CNNs) has resulted in them outperforming traditional computer vision techniques in object detection as well as image classification tasks across various benchmark datasets [17], [18]. The proposed approach uses CNN for the original image transformation into a gray-scale image. The reason for this transformation is to create a segmented image, where animals are highlighted as radial gradients and the rest of the image remains black. The approach is depicted in Fig. 3.

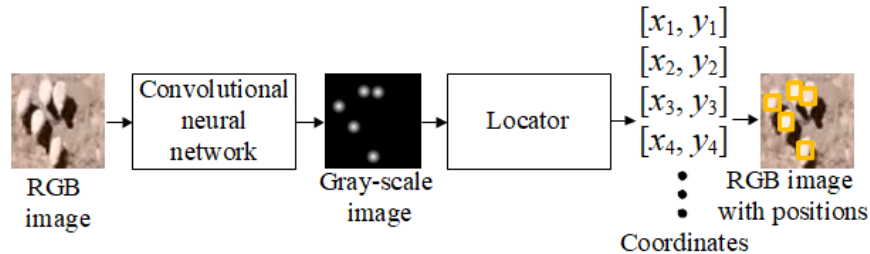


Fig. 3. Proposed positioning approach.

It is obvious that the function of the Locator is a routine process of finding local maxima in a gray-scale image. The transformation of the RGB image is, on the other hand, a tricky issue. One of the possible solutions is the implementation of a representative from the family of image segmentation neural networks. These neural networks are often based on deep encoders and decoders. It involves encoding the input image into low dimensions, and then recovering it with orientation invariance capabilities in the decoder. If the training set is prepared correctly, the approach should generate a segmented image at the decoder end. Although the image segmentation process is mostly supposed to produce images covered with monotonous areas of colors, the authors believe, that a similar approach is able to code gradient circles, as required in the presented method. Therefore, the image segmentation neural network, which transforms an RGB image into a gray-scale image where animals are highlighted as radial gradients, is defined in the next section.

3 Image segmentation neural network

In recent years, many types of deep neural networks, including CNNs, fully convolutional networks, ensemble learning systems, and SegNet neural networks, were introduced [2]. Based on the authors' previous experience, from the wide set of possibilities, the U-Net is selected for the initial testing. It is robust, reasonably fast, and it is supposed to need only a small training set to be sufficiently trained.

The U-Net is a fully convolutional neural network developed originally for biomedical image segmentation. It consists of a contracting path (left side) and an expansive path (right side). The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer, a 1x1 convolution is used to map the output. In total, the network has 23 convolutional layers [14].

3.1 Dataset for training and validation

The dataset for training and validation is prepared in order to fulfill the conditions described in Section 1. Therefore, several videos taken by the UAV were processed, and images, which satisfied the conditions, were extracted. Altogether, 440 images [288 x 288]px, which cover livestock (sheep and cows), taken from the height 50-100 m, were selected. These images were then divided into a training and testing set. The overall information is summarized in Table 1.

Table 1. Parameters of dataset

-	Training set	Testing set	Together
Images	330	110	440
Animals	10621	3707	14328

Then, the target images for the training and validation needed to be prepared. This process was performed manually by a custom tagging application. For each input image, a gray-scale target image, where animal positions were highlighted by a gradient circle, was prepared. The examples of input-target pairs are shown in Fig. 4.



Fig. 4. Examples of input-target pairs for training set. The dimensions are 288 x 288 px.

3.2 U-Net training

Consequently, the training of the U-Net architecture was performed. The ADAM algorithm is implemented as an optimizer based on its generally acceptable performance [9]. Initial weights were set randomly with normal distribution (mean = 0, standard deviation = 0.05). The experiments are performed twenty times due to a stochastic character of training. All the parameters are shown in Table 2.

Table 2. Parameters of the training

Input shape	288 x 288 x 3
Training algorithm	ADAM algorithm
Initialization	Normal distribution (mean = 0, std = 0.05)
Maximum epochs	50
Stopping criterion	Maximum epochs reached
Learning rate α	0.001
Exponential decay rate 1 β_1	0.9
Exponential decay rate 2 β_2	0.999

3.3 Results

In this section, the performance of the best U-Net, trained according to the previous paragraph, is introduced. A good practice for the evaluation is to determine the accuracy over the testing set. However, two additional metrics, precision and recall, are added. The metrics are described by the following equations.

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (3)$$

where TP (true positive) is the number of correctly classified animals, FN (false negative) is the number of misclassified animals, FP (false positive) is the number of incorrectly detected animals. All these quality indicators are expressed as percentage values in the rest of the article.

The results obtained by the best U-Net for image transformation, and the local maxima detector for livestock positioning, are summarized in Table 3, where the testing set is divided into three subsets according the height of the camera.

Table 3. Results

Dataset	Accuracy	Precision	Recall
Testing	97.56 %	98.42 %	99.11 %
Training	100.00 %	100.00 %	100.00 %
Testing (50-66 m)	97.55 %	98.58 %	98.93 %
Testing (66-83 m)	96.37 %	97.22 %	99.10 %
Testing (83-100 m)	98.34 %	99.18 %	99.13 %

According to the results, the proposed method obviously works perfectly on the training images, and shows only slight degradation on the test images. In addition, there is no obvious relevance between the accuracy (and other metrics) and the height of the UAV with the camera.

4 Conclusion

A novel engineering approach to livestock positioning and counting is proposed in this contribution. The approach is composed of two parts. Firstly, a fully convolutional neural network for input image transformation, and secondly, a locator for animal positioning. The transformation process is designed in order to transform the original RGB image into a gray-scale image, where animal positions are highlighted as gradient circles. After a set of experiments, the U-Net was selected for the transformation. In combination with the local maxima function for positioning, the U-Net provides a precision rate of 0.9842 and a recall rate of 0.9911 with the testing set.

The presented contribution should be understood as a first step in the development of a robust livestock counting device. Work in the near future will include convolutional neural network architecture optimizing, and computational complexity testing in order to prepare the approach for implementation.

References

1. Arnal Barbedo, J.G., Koenigkan, L.V.: Perspectives on the use of unmanned aerial systems to monitor cattle. *OUTLOOK ON AGRICULTURE* **47**(3), 214–222 (Sep 2018). <https://doi.org/10.1177/0030727018781876>

2. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(12), 2481–2495 (Dec 2017). <https://doi.org/10.1109/TPAMI.2016.2644615>
3. Bhakta, I., Phadikar, S., Majumder, K.: State-of-the-art technologies in precision agriculture: a systematic review. *Journal of the Science of Food and Agriculture* **99**(11), 4878–4888 (2019). <https://doi.org/10.1002/jsfa.9693>
4. Bishop, J.C., Falzon, G., Trotter, M., Kwan, P., Meek, P.D.: Livestock vocalisation classification in farm soundscapes. *COMPUTERS AND ELECTRONICS IN AGRICULTURE* **162**, 531–542 (Jul 2019). <https://doi.org/10.1016/j.compag.2019.04.020>
5. Cowlar: Streamline your dairy business! (2020), <https://www.cowlar.com/>
6. Dhulekar, P.A., Gandhe, S.T., Bagad, G.R., Dwivedi, S.S.: Vision based technique for animal detection. In: 2018 International Conference On Advances in Communication and Computing Technology (ICACCT). pp. 344–348 (Feb 2018). <https://doi.org/10.1109/ICACCT.2018.8529660>
7. Ismail, Z.H., Chun, A.K.K., Shapiai Razak, M.I.: Efficient herd – outlier detection in livestock monitoring system based on density – based spatial clustering. *IEEE Access* **7**, 175062–175070 (2019). <https://doi.org/10.1109/ACCESS.2019.2952912>
8. Kellenberger, B., Marcos, D., Courty, N., Tuia, D.: Detecting animals in repeated uav image acquisitions by matching cnn activations with optimal transport. In: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. pp. 3643–3646 (Jul 2018). <https://doi.org/10.1109/IGARSS.2018.8519012>
9. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *CoRR abs/1412.6980* (2014), <http://arxiv.org/abs/1412.6980>
10. Li, X., Xing, L.: Use of Unmanned Aerial Vehicles for Livestock Monitoring based on Streaming K-Means Clustering. *IFAC PAPERSONLINE* **52**(30), 324–329 (2019). <https://doi.org/10.1016/j.ifacol.2019.12.560>
11. McKinlay, J., Southwell, C., Trebilco, R.: Integrating Count Effort By Seasonally Correcting Animal Population Estimates (Icescape): A Method For Estimating Abundance And Its Uncertainty From Count Data Using Adelie Penguins As A Case Study. *CCAMLR SCIENCE* **17**, 213–227 (2010)
12. Nyamuryekung’e, S., Cibils, A.F., Estell, R.E., Gonzalez, A.L.: Use of an Unmanned Aerial Vehicle-Mounted Video Camera to Assess Feeding Behavior of Raramuri Criollo Cows. *RANGELAND ECOLOGY & MANAGEMENT* **69**(5), 386–389 (Sep 2016). <https://doi.org/10.1016/j.rama.2016.04.005>
13. Parikh, M., Patel, M., Bhatt, D.: Animal detection using template matching algorithm. *International Journal of Research in Modern Engineering and Emerging Technology* **1**(3), 26–32 (Apr 2013)
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. *CoRR abs/1505.04597* (2015), <http://arxiv.org/abs/1505.04597>
15. Sarwar, F., Griffin, A., Periasamy, P., Portas, K., Law, J.: Detecting and counting sheep with a convolutional neural network. In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). pp. 1–6 (Nov 2018). <https://doi.org/10.1109/AVSS.2018.8639306>
16. Seo, J., Sa, J., Choi, Y., Chung, Y., Park, D., Kim, H.: A yolo-based separation of touching-pigs for smart pig farm applications. In: 2019 21st International Conference on Advanced Communication Technology (ICACT). pp. 395–401 (Feb 2019). <https://doi.org/10.23919/ICACT.2019.8701968>

17. Sharma, P., Singh, A.: Era of deep neural networks: A review. In: 8th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2017 (2017). <https://doi.org/10.1109/ICCCNT.2017.8203938>
18. Xu, Y., Zhou, X., Chen, S., Li, F.: Deep learning for multiple object tracking: a survey. *IET Computer Vision* **13**(4), 355–368 (2019). <https://doi.org/10.1049/iet-cvi.2018.5598>