2020 16th IEEE International Conference on Automation Science and Engineering (CASE)
August 20-21, 2020, Online Zoom Meeting

Data Augmentation for Deep Learning based Cattle Segmentation in Precision Livestock Farming

Yongliang Qiao, Daobilige Su, He Kong, Salah Sukkarieh, Sabrina Lomax and Cameron Clark

Abstract— **Accurate segmentation of cattle is a prerequisite for feature extraction and estimation. Convolutional neural networks (CNN) based approaches that train models on the large-scale labeled datasets have achieved high levels of segmentation performance. However, pixel-wise manual labeling of a cattle image is challenging and time consuming due to the irregularity of the cattle contour. In this regard, data augmentation for deep learning based cattle segmentation is required. Our proposed data augmentation approach uses random image cropping and patching to expand the number of training images and their corresponding labels, then, a state-of-the-art deep neural net is trained to segment cattle images. Here we apply these techniques to images of cattle in a feedlot environment. Our data augmentation-based approach segmented cattle from a complex background with 99.5% mean Accuracy (*mAcc*) and 97.3% mean Intersection of Unions (*mIoU*), improving current techniques including a combination of random flipping, rotation and color jitter.**

I. Introduction

Varying deep learning based-networks have been proposed to segment objects from their background [1]–[3]. These networks achieve high segmentation accuracy after training thousands of images [4]. This automatic segmentation of cattle profile from the background is a prerequisite for vision based monitoring and tracking [5], [6]. However, for deep learning-based cattle segmentation, producing pixelwise labels for images acquired from a feedlot or paddock is challenging, with polygon based labeling for each image taking tens of minutes as cattle have an irregular body contour, and their body profile changes with movement. These challenges limit the amount of data available for training.

Data augmentation (DA) increases the variety of training samples and prevents over-fitting [7]–[10]. Random horizontal flipping and cropping [11] techniques have been used to increase the number of images, then color variations (e.g. color jitter technique, color translation) used to detect or segment colorful objects [12], [13]. The dropout technique was also used for data augmentation, which dropped certain pixels to increase the robustness of the trained model to noisy images [14]. Hinton et al. (2012) [14] added disturbances to the original images by dropping certain pixels, and Kang et al. (2017) [10] applied a unique kernel filter that randomly swapped the pixel values in a square sliding window. In addition, DeVries et al. (2017) [15] and Zhong et al. (2017) [16] masked a region of an image to force the model to learn different parts of objects as opposed to the overall image. Emmanuel et al. (2017) [17] randomly rotated images and introduced a natural background in the newly created image to enhance image information, which can be used to improve recognition accuracy of aerial images of cows. However, these DA approaches rely on complex computation and filters to generate new images.

More recently, Takahashi et al. (2018) [18] proposed the Random Image Cropping and Patching (RICAP) method to expand training datasets for the purpose of improving image classification accuracy. The RICAP method randomly crops four images and patches them to generate a new training image, producing one patched image together with mixed labels for each image, thereby improving image classification accuracy by increasing the variety of training images and preventing over-fitting. However, the original proposed RICAP is used in the field of image classification, which cannot be directly applied to image segmentation tasks. Here we cropped and patched both cattle images and segmentation masks simultaneously to generate new images and labels to expand the training dataset for cattle segmentation for training through CNNs.

Based on the expanded training dataset, the state-of-the-art encode-decoder architecture Bonnet is trained to segment cattle images.

II. Related work

Due to overfitting, models trained with small datasets often do not generalize well in validation and testing [19]–[21]. Data augmentation can be an effective way to reduce over-fitting by increasing the variety of training samples. Various data augmentation techniques have been proposed. A brief review of existing data augmentation techniques for improving the performance of various deep neural nets is provided in this section.

The standard data augmentation method usually uses random cropping and horizontal flipping to expand a dataset. One typical example is the work in AlexNet [12], where the images from CIFAR dataset [22] were cropped and flipped. Here random cropping can change the apparent features of the image, whilst horizontal flipping doubles the variation with specific orientations (e.g. a side-view of an airplane). The AlexNet [12] also used a principal component analysis to change intensity values of all image channels. Such color variations can increase the variations of images and help to detect or segment colorful objects such as flowers. A similar

Yongliang Qiao, Daobilige Su, He Kong, and Salah Sukkarieh are with the Australian Centre for Field Robotics, University of Sydney, Australia. E-mails: {yongliang.qiao, daobilige.su, he.kong, salah.sukkarieh}@sydney.edu.au.

Sabrina Lomax and Cameron Clark are with Livestock Production and Welfare Group, School of Life and Environmental Sciences, University of Sydney, Australia. E-mails:{sabrina.lomax, cameron.clark}@sydney.edu.au

color translation technique was also proposed in the reimplementation of ResNet [23] by Facebook AI Research in [24]. In addition, Takahashi et al. (2019) proposed RICAP, which randomly cropped four images and patched them to create a new training image [25], and validated its applicability in image classification tasks.

Unlike standard data augmentation methods which generate natural images, data disrupting is a general data augmentation technique that produces unnatural images by disrupting image features. Dropout is a popular data disrupting method that disturbs and masks the original information by dropping pixels. Pixel dropping functions inject noise into an image [26], which makes the CNN robust to noisy images and contributes to generalization rather than enriching the dataset. As an extension of dropout, Cutout can mask the entire main section of an object in the image and makes CNNs robust to noisy images [27]. Similarity, Zhong et al. (2017) proposed a random erasing method, which also masks a certain area of an image with randomly generated size and aspect ratio of the masked region [16].

Pixel dropping functions as an injection of noise into an image [26], makes the CNN robust to noisy images and contributes to generalization rather than enriching the dataset. As an extension of dropout, Cutout can mask the entire main part of an object in the image and makes CNNs robust to noisy images [27]. Similarity, a random erasing method, which also masks a certain area of an image with randomly generated size and aspect ratio of the masked region, has been proposed in Zhong et al. (2017) [16].

With recent deep neural nets becoming increasingly advanced and the number of parameters of the model growing ever larger [28], [29], new data augmentation methods have emerged. AutoAugment [30] is a framework exploring the best hyperparameters of existing data augmentations via reinforcement learning [31], which achieves high performance in CIFAR-10 classification. In addition, Zhang et al. (2017) introduced a DA technique that alpha-blends two images into one new training image [32]. Combination of pairs of images and labels is used to train deep neural nets, which forces neural nets to form linear behavior between these pairs of samples.

More recently, Generative Adversarial Nets (GANs) have been proposed to perform unsupervised generation of new images for training [29]. Furthermore, GANs have been shown to be effective even with relatively small sets of data by performing transfer learning techniques [33]. Additionally, they can augment data sets, increasing the resolution of input images [34]. Tran et al. (2017) used a Bayesian approach to generate data based on the distribution learned from the training set and demonstrated its implementation via GAN [35]. However, one should note that the complexity of GANs is high, leading to a long computation time.

Although DA and other similar concepts have been proposed, existing methods have their respective limitations in agricultural applications. For example, Di Cicco et al. (2017) proposed a DA method to train a deep segmentation neural net, which was used to segment crops and weeds synthetically [36]. However, the method requires information about the distribution of leaves in all crops and weed species, and high definition of textures on leaves and illumination. Bah et al. (2018) proposed an unsupervised labeling method, which exploited line structures of crops on aerial images [37]. However, this method cannot tackle images captured by ground field robots due to the restriction of camera views.

III. The Proposed method

A. Data augmentation

Following the RICAP method in [18], we randomly cropped a given image I into M slices along the horizontal direction, and N slices along the vertical direction. Thus there were $M \times N$ number of independent regions. Each region was extracted from one randomly selected image, then these regions were patched to a new image. The whole process was formulated as follows,

$$\begin{aligned} m_i &= Round(\frac{m_{i'} M_I}{\sum m_{i'}}) \\ n_j &= Round(\frac{n_{j'} N_I}{\sum n_{i'}}) \end{aligned} \tag{1}$$

where, m_i and n_j are the width and height of the crop region; i and j are horizontal and vertical indices of each image crop region; Round $(\cdot)$ is the rounding to integer operation, and M_I and N_I are width and height of the raw images in dataset; $m_{i'}$ and $n_{j'}$ are crop ratio along the horizontal and vertical directions, and they follow a simple uniform distribution.

For each cropped region $I_{ij}(i = 1, 2, \cdots, M; j = 1, 2, \cdots, N)$, its top left and bottom right pixel indexes were formulated as follows,

$$\begin{aligned} x_{ij}^{tl} = \sum_{k=1}^{i-1} m_k + 1, y_{ij}^{tl} = \sum_{k=1}^{j-1} n_k + 1 \\ x_{ij}^{br} = \sum_{k=1}^{i} m_k, y_{ij}^{br} = \sum_{k=1}^{j} n_k \end{aligned} \tag{2}$$

where, x_{ij}^{tl}, y_{ij}^{tl} ,x_{ij}^{br}, y_{ij}^{br} are x, y coordinates along image width and height directions of the top left and bottom right corners of the crop region corresponding to I_{ij}.

As illustrated in Fig. 1, four training images and their corresponding labels were sliced into 2 parts horizontally and 2 parts vertically. Here the size of horizontal and vertical crops was completely random for each image to be generated. These selected images and labels were cropped and patched together to generate a new image and a new label. By doing this, the limited labeled images were augmented, which were used for training the segmentation model.

B. Cattle segmentation network

In this work, a state-of-the-art deep neural net, namely, Bonnet [1], was used for image segmentation. Bonnet is an open-source framework with Python training pipeline and C++ deployment library, which allows for fast multi-GPU training. The architecture used for Bonnet was ERFNet, and no pre-trained weights were used. The input image size was

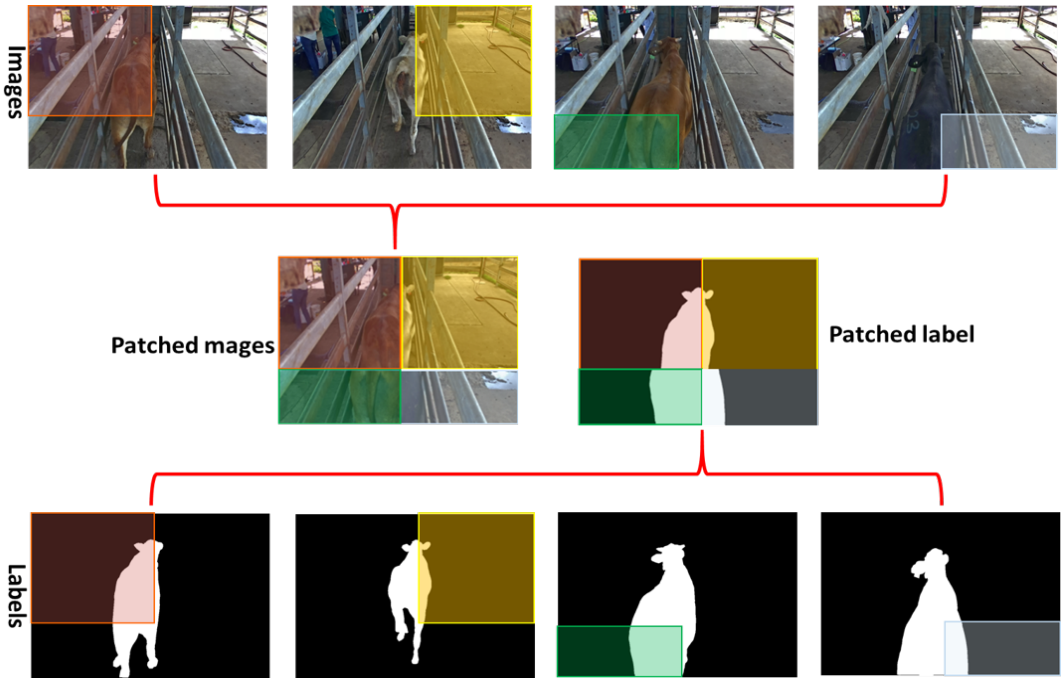


Fig. 1: Illustration of the data augmentation for deep learning based segmentation. Each image and its corresponding labels are randomly divided into 2 parts horizontally and 2 parts vertically. Then the corresponding parts from 4 images and labels are patched to generate a new image and a new label.

512×384. The architecture of Bonnet based cattle segmentation can be seen in Fig. 2. There are 3 downsampling and upscaling layers to encode and decode information. In addition, in order to speedup forward pass time and avoid overfitting, the variable receptive field non-bottleneck convolution layers were also used [38]. The last linear classification layer determined each pixel of the images belonging to either background or cattle.

IV. Experimental Setup

A. Data acquisition

The cattle image dataset was collected at a Southern Queensland commercial feedlot in 2018 at three different times during the feed period (induction, middle and end point) on 20 March, 30 April and 30 May, respectively. The data acquisition system is shown in Fig. 3. Our data was acquired when the cattle were walking along the race (path) from right to left in Fig. 3. In our data collection, only the left image of the rear view ZED camera was used, the image resolution is 1920×1080. A high frame acquisition rate (30fps) was adopted, to reduce the influence of motion blur during the herding process from the open-air pen to the restraint device (crush).

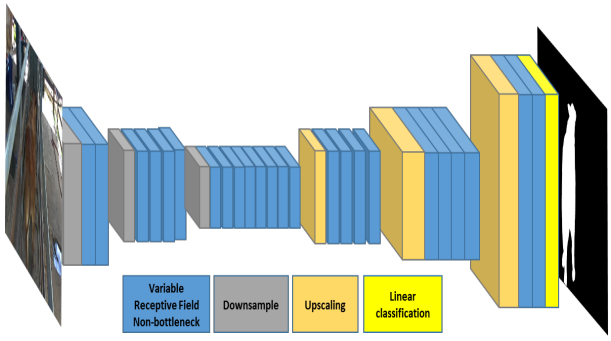


Fig. 2: Architecture of Bonnet based cattle segmentation.

Fig. 3: Area for data acquisition, the race leading to the crush at feedlot showing rear view camera clamped to overhead bar and the side view camera on tripod.

This cattle dataset was challenging for the image segmenter when considering complicated factors such as different illumination conditions, animal postural changes and complex backgrounds (including the cattle crush and ground). In some cases, the segmented animal nearly blends with the soil background. Only 400 images were labeled

among over 20K captured images due to the time-consuming labeling process.

B. Network training

In our work, all deep learning models were trained on a GTX1080Ti GPU with a batch size of 16 images. The training steps for all experiments were 20K iterations, and learning rate was set to 0.001 initially which decayed by 1:5 every 10 epochs. The optimiser used was Adam [39]. All other parameters were default, as used in Bonnet [1]. For the network training, the total labeled data (400 images) was split into training, validation and testing by 168, 32 and 200 respectively.

V. Experimental Results

The proposed data augmentation approach was compared to a baseline (conventional data augmentation). The latter was a combination of random flipping, rotation and color jitter operations together.

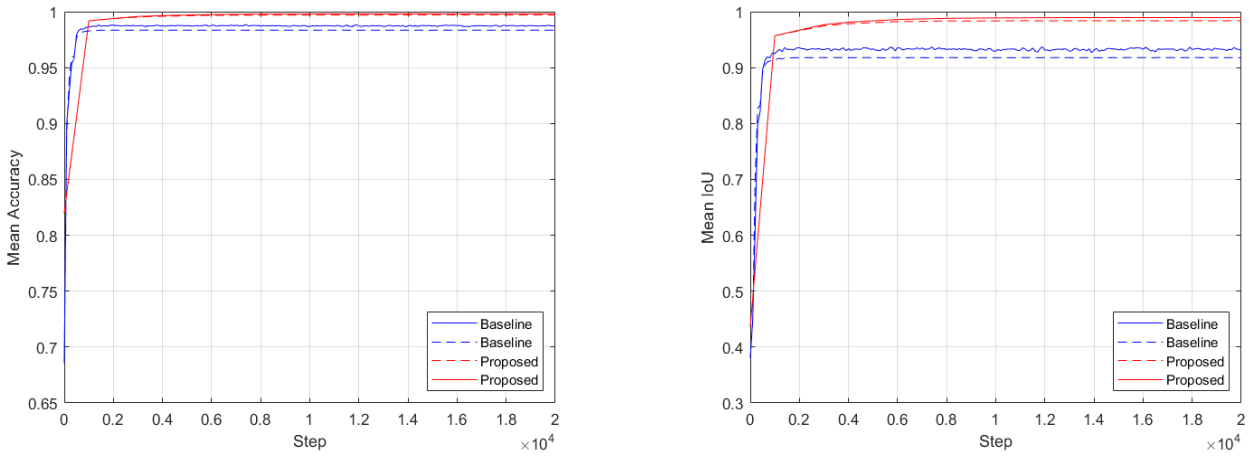
To compare the above two methods, we computed and showed the pixel-wise performance of the deep neural net tested in our acquired challenging cattle dataset. In particular, for both methods, we calculated the precision and recalls of all classes, i.e., background and cattle, as well as mean accuracy (*mAcc*) and mean intersection of unions (*mIoU*) of all pixels. The *mAcc* and *mIoU* are computed as follows:

$$mAcc = \frac{1}{C}\sum_{c=1}^{C}\frac{TP_c}{TP_c + FP_c}$$
$$mIoU = \frac{1}{C}\sum_{c=1}^{C}\frac{TP_c}{TP_c + FP_c + FN_c} \quad (3)$$

where, TP_c, FP_c, FN_c are true positively, false positively, and false negatively classified pixel numbers for cth class in the segmentation tasks. In our work, the value of C is 2 (we only consider cattle and background).

The performance comparisons over optimization steps are shown in Fig. 4. In Fig. 4, the red and blue solid lines in (a), (b) represent mean accuracies and mean *IoU* with our proposed baseline approach, while the red and blue dash lines in (a), (b) represent mean accuracies and mean *IoU* in validation data. It can be seen that the optimization has converged after more than 4000 steps. In addition, mean accuracies and mean *IoU* of the proposed approach are better than that of the Baseline.

The cattle image segmentation accuracy for the Bonnet trained with different data augmentation methods are presented in Table 1. It can be seen that the proposed data



(a) Mean accuracies over optimisation steps

(b) Mean IoU over optimisation steps

Fig. 4: Mean accuracies (a) and mean *IoU* (b) over optimization steps with different data augmentations. The solid lines represent the performance of training data. The dash lines represent the performance of validation data.

TABLE I: Accuracy of cattle identification

Augmentation	*mAcc*(%)	*mIoU*(%)	Precision(%)		Recall(%)	
			background	cattle	background	cattle
Baseline						
%Training	98.72	93.14	99.19	94.32	99.39	92.57
%Validation	98.34	91.78	98.78	95.09	99.44	89.33
%Testing	98.58	92.58	99.05	94.31	99.38	91.51
Proposed						
%Training	99.81	98.95	99.89	990.3	99.89	99.06
%Validation	99.70	99.38	99.83	98.53	99.83	98.53
%Testing	**99.50**	**97.31**	**99.70**	**97.73**	**99.75**	**97.32**

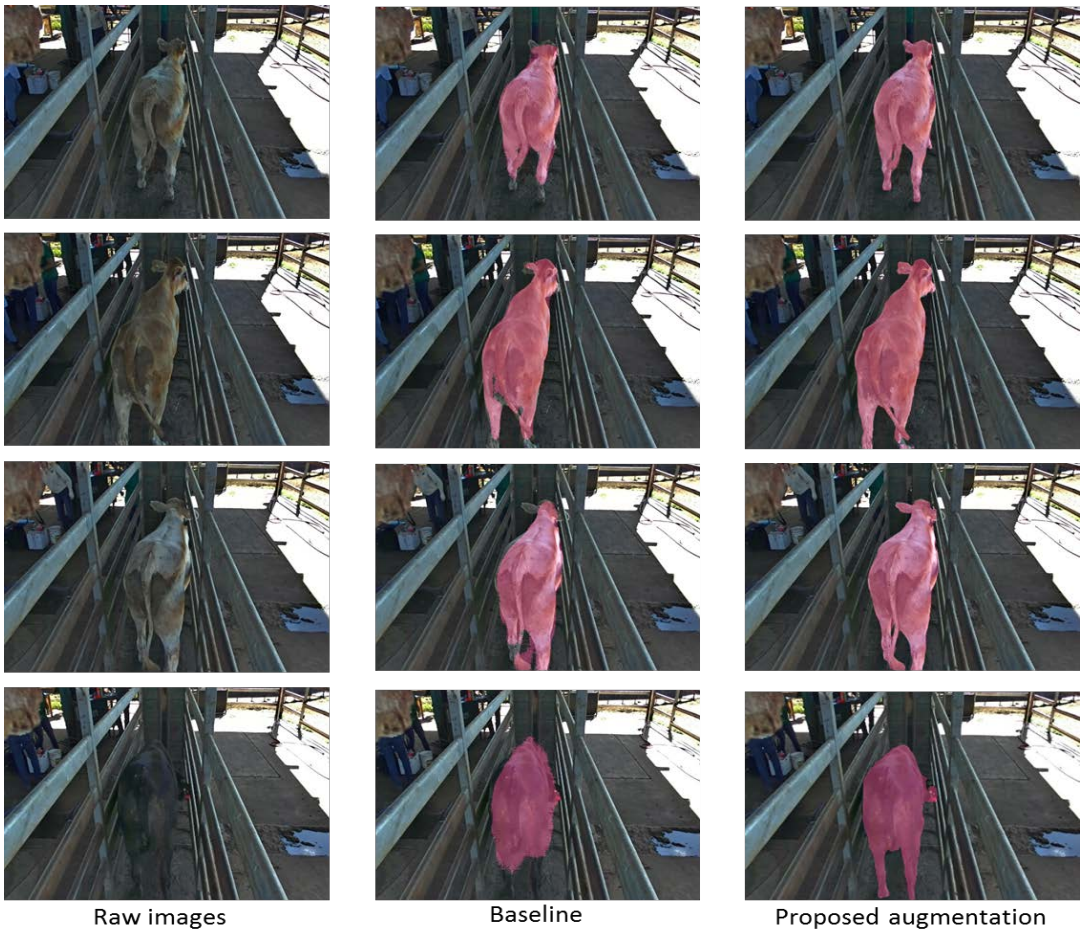


Fig. 5: Segmentation results for the network trained with different data augmentation methods. The first column is raw images, the second column is the results of baseline, while the third column is the results with the proposed data augmentation.

augmentation based approach segmented cattle from the complex background with 99.50% *mAcc* and 97.31% *mIoU*, which is better than that of baseline (98.58% *mAcc* and 92.58% *mIoU*). In addition, during the training and validation phases, the segmentation accuracy of the proposed approach was also higher than those of baseline.

To better understand the effect of the proposed data augmentation approach on performance, the qualitative comparisons of cattle segmentation results with different data augmentation methods are also displayed in Fig. 5. It can be seen that some pixels nearby the animal's head and rump are mis-segmented as background (second column in Fig. 5) in the baseline approach, while the animal's leg and head regions are correctly segmented (third column in Fig. 5) when the proposed data augmentation was used for network training. This is due to the generated images having more variability when being cropped and patched randomly. When the augmented dataset with more variability was used for network training, it prevented the model from over-fitting and improved the segmentation accuracy in the test data.

The illustrated results in Table I and Fig. 5 show that the proposed method can effectively improve the performance of a deep neural net for image segmentation by bringing a greater variety of information for training images from a relatively small numbers of labeled images. The proposed approach is a solution for segmentation tasks with limited label datasets in the field of agriculture and livestock production.

VI. Conclusion and Future Work

In this paper, a data augmentation and deep learning based framework for cattle image segmentation is proposed. The proposed framework randomly crops images into different regions and then patches them into a new image. By doing this, the proposed data augmentation produces new, more informative images together with the corresponding labels for cattle segmentation network training. According to the experimental results from our acquired cattle dataset, the proposed data augmentation based approach segmented cattle from its complex background with 99.5% *mAcc* and 97.3% *mIoU*. These results highlight the ability to improve the segmentation accuracy of the deep neural net, which offers a solution to solve tasks with limited label datasets in agriculture or livestock faming. Further work will determine the impact of image patch sizes and numbers on segmentation accuracy.

ACKNOWLEDGMENTS

The authors acknowledge the support of the Meat & Livestock Australia Donor Company through the project: Objective, robust, real-time animal welfare measures for the Australian red meat industry (P.PSH.0819). The authors also express their gratitude to Khalid Rafique, Javier Martinez, Amanda Doughty, Ashraful Islam, and Mike Reynolds for their help in experiment organization and data collection.

REFERENCES

[1] A. Milioto and C. Stachniss, "Bonnet: An open-source training and deployment framework for semantic segmentation in robotics using cnns," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 7094–7100.

[2] 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.

[3] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2980–2988.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[5] T. Van Hertem, L. Rooijakkers, D. Berckmans, A. P. Fernández, T. Norton, and E. Vranken, "Appropriate data visualisation is key to precision livestock farming acceptance," *Computers and electronics in agriculture*, vol. 138, pp. 1–10, 2017.

[6] M. Tscharke and T. M. Banhazi, "A brief review of the application of machine vision in livestock behaviour analysis," *AGRÁRINFORMATIKA/JOURNAL OF AGRICULTURAL INFORMATICS*, vol. 7, no. 1, pp. 23–42, 2016.

[7] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for sar target recognition," *IEEE Geoscience and remote sensing letters*, vol. 13, no. 3, pp. 364–368, 2016.

[8] Q. Xie, Z. Dai, E. Hovy, M.-T. Luong, and Q. V. Le, "Unsupervised data augmentation," *arXiv preprint arXiv:1904.12848*, 2019.

[9] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation strategies from data," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 113–123.

[10] G. Kang, X. Dong, L. Zheng, and Y. Yang, "Patchshuffle regularization," *arXiv preprint arXiv:1707.07103*, 2017.

[11] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., 2012, pp. 1097–1105.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[14] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.

[15] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv preprint arXiv:1708.04552*, 2017.

[16] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," *arXiv preprint arXiv:1708.04896*, 2017.

[17] E. Okafor, R. Smit, L. Schomaker, and M. Wiering, "Operational data augmentation in classifying single aerial images of animals," in *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*. IEEE, 2017, pp. 354–360.

[18] R. Takahashi, T. Matsubara, and K. Uehara, "Ricap: Random image cropping and patching data augmentation for deep cnns," in *Asian Conference on Machine Learning*, 2018, pp. 786–798.

[19] Y. Qiao, C. Cappelle, Y. Ruichek, and T. Yang, "Convnet and LSH-based visual localization using localized sequence matching," *Sensors*, vol. 19, no. 11, p. 2439, 2019.

[20] Y. Qiao, D. Su, H. Kong, S. Sukkarieh, S. Lomax, and C. Clark, "Individual cattle identification using a deep learning based framework," *IFAC-PapersOnLine*, vol. 52, no. 30, pp. 318–323, 2019.

[21] Y. Qiao, D. Su, H. Kong, S. Sukkarieh, S. Lomax *et al.*, "Bilstm-based individual cattle identification for automated precision livestock farming," in *16th IEEE International Conference on Automation Science and Engineering*. Accepted and to appear, IEEE, 2020.

[22] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*. Springer, 2016, pp. 630–645.

[24] F. AI, "Reimplementation of resnet by facebook ai," 2019. [Online]. Available: https://github.com/facebook/fb.resnet.torch. [Online;accessed19-March-2019].

[25] R. Takahashi, T. Matsubara, and K. Uehara, "Data augmentation using random image cropping and patching for deep cnns," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[26] J. Sietsma and R. J. Dow, "Creating artificial neural networks that generalize," *Neural networks*, vol. 4, no. 1, pp. 67–79, 1991.

[27] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv preprint arXiv:1708.04552*, 2017.

[28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[29] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[30] E. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. Le, "Autoaugment: learning augmentation policies from data. arxiv preprint," *arXiv preprint arXiv:1805.09501*, 2018.

[31] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," *arXiv preprint arXiv:1611.01578*, 2016.

[32] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[33] S. Gurumurthy, R. Kiran Sarvadevabhatla, and R. Venkatesh Babu, "Deligan: Generative adversarial networks for diverse and limited data," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 166–174.

[34] M. Marchesi, "Megapixel size image creation using generative adversarial networks," *arXiv preprint arXiv:1706.00082*, 2017.

[35] T. Tran, T. Pham, G. Carneiro, L. Palmer, and I. Reid, "A bayesian data augmentation approach for learning deep models," in *Advances in neural information processing systems*, 2017, pp. 2797–2806.

[36] M. Di Cicco, C. Potena, G. Grisetti, and A. Pretto, "Automatic model based dataset generation for fast and accurate crop and weeds detection," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 5188–5195.

[37] M. D. Bah, A. Hafiane, and R. Canals, "Deep learning with unsupervised data labeling for weed detection in line crops in uav images," *Remote sensing*, vol. 10, no. 11, p. 1690, 2018.

[38] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo, "Erfnet: Efficient residual factorized convnet for real-time semantic segmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 263–272, 2017.

[39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.