Agtech 2030

Face recognition of horses using convolutional neural network

Ina Eriksson • Tommy Granström • Petra Gunnarsdotter

Evaluating the properties of dataset

Preface

The background of the subject chosen was that Petra Gunnarsdotter contacted us 16 September 2020 and asked for project ideas related to agriculture. Our answer was face recognition applied to horses.

We decided upon a cooperation between Agtech 2030 on the one hand and the team Gunnarsdotter on the other. Then we chose Hollstad farm outside Norrköping as test arena. After OK from the stable manager Caroline Sjödell we could start work.

A couple of horses were photographed and filmed both outdoor and inside the stable. This "ground truth" material became the empirical basis for the project team consisting of students at the Civil Engineering program Media technology at Linköping University, campus Norrköping.

It is with great pleasure that we now can present this report on face recognition of horses with the help of neural networks. The project team Ina Eriksson, Tommy Granström and Petra Gunnarsdotter did a splendid job.

Linköping 20 december 2020

Per Frankelius	*Matilda von Rosen*
Process leader	Coordinator

Agtech 2030
c/o Linköping University
Att. Per Frankelius
IEI
581 83 Linköping
www.agtech2030.com

Face recognition of horses using Convolutional neural network - Evaluating the properties of dataset

Ina Eriksson[1], Tommy Granström [2], Petra Gunnarsdotter[3]

Abstract
Identifying grazing animals is both time consuming and a matter of animal welfare. Being able to digitally monitor the animals, abnormal behavior can be detected which can prevent unnecessary suffering. In this thesis, the design of datasets has been evaluated for the deep learning method convolutional neural network. The model has been developed using data augmentation and drop-out both separately and combined to investigate how these two methods are affecting the accuracy of the model. This thesis also investigates different sizes of datasets and if cropped images increases the accuracy compared to not cropped images. The conclusion is that after evaluation of all the datasets, the best settings for CNN are the combination of both data augmentation and drop-out. The best size of the dataset was 500 images or more, for three classes using 10-15 epochs. Another important conclusion is that when using small datasets with 50 images, a cropped setup with 30 epochs results in an acceptable accuracy compared with a not cropped setup.
Source code: https://github.com/tommygranstrom/Horse-Facerec_TNM095.git

Authors
[1] *Master of Science in Communications,Transport and Infrastructure at Linköping University, inaer470@student.liu.se*
[2] *Master of Science in Communications,Transport and Infrastructure at Linköping University, tomgr497@student.liu.se*
[3] *Master of Science in Media Technology and Engineering at Linköping University, petgu692@student.liu.se*

Keywords: Convolutional Neural Network (CNN), face recognition of horses, horses identification, dataset.

Contents

1. Introduction

The interest of identifying the individuals of grazing animals has increased as the agricultural production systems expands. Agriculture plays an important role in the global economy and environment. Agri-technology and precision farming are new scientific fields with great potential. The Agri-technology could improve both the productivity in the industry and the effects on the environment. [1][2]

There are several reasons to identify each individual animal at a farm, for example a safety aspect. To be able to identify each animal, the daily supervision could be complemented with machine learning to identify behavior patterns. For example, abnormal behavior that could indicate illness. [3]

In order to create a face recognition algorithm, the machine learning model needs to be trained on images of horses. Sometime thousand of images are needed depending on the object and variation in the image. In this thesis images were collected by hand and used to create five different datasets. The datasets have different sizes and the performance of the machine learning model after each dataset are compared. The machine learning algorithm used in this thesis is Convolution Neural Network (CNN).

2. Theory

In machine learning models the data is often divided into three groups: training, validation and test. [4] The training data is used to fit the model and includes the weights and bias. The validation data is used to evaluate the models performance while tuning the parameters in the model. The validation data are used to improve the model during the development stage. The test data are used to evaluate the performance of the final model. The purpose with test data is to give a unbiased estimate of the generalized performance. A visualization of the three groups can be seen in Figure 1.

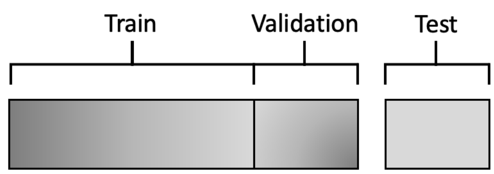

Figure 1. Visualization of the three groups of data.

The amount of data in each group depends on the amount of available data and the actual model that is trained.[4]

The performance of the model can be evaluated by analyzing the training accuracy versus the validation accuracy and the training loss versus the validation loss. The training accuracy is a estimate of potential model performance and validation accuracy is the actual model performance. The training loss is a estimate of the potential number of incorrect predictions and the validation loss is the actual incorrect number of incorrect predictions.

Tuneing the parameters of the model considering the validation set could lead to overfitting, especially models with many hyperparameters. Overfitting is when the model preforms well on the training data but poorly on the unseen data or test data as it is called in this thesis. The validation data could be classified as unseen data since the model never sees the true label of the data. However, the developer tunes the model after model performance and the parameters are configured to give good results. Therefore, a test set of images is used to ensure a reliable model.

It is not desirable with a overfit model. An overfit model would result in a model that can only recognize very similar images. Nor is it desirable with a underfit model. Underfitting is when the model can not learn the relationship between the data point. An underfit model would believe it could recognize almost anything.

2.1 CNN

CNN [5] is an artificial neural network that has a type of specialization for being able to detect patterns and are therefore popular in image processing. CNN takes advantage of the spatial context and when dealing with images one pixel is very dependent on the neighboring pixels. [5]

The network in CNN is built up on convolutional layers. The convolutional layers transform the input information and send the output information to the next layer. For each convolution layer, the number of filters needs to be specified. A filter in this context is a matrix with numbers representing the convolutional matrix. More filters at a convolution layer means more matrices. Initially the matrix is set with random numbers and the filter is slid over all input, this is called convolving. When convolving the dot product between the input and the filter is calculated and stored which constitutes the convolutional layer. The result from this convolutional layer is the output that is sent to the next layer. [5]

One drawback with CNN is that it often needs a large training dataset compared to other machine learning algorithms. Large datasets bring long computation time and the dataset is often split up to batches. These batches contain a smaller number of images and are all run through the network. When building a CNN, the number of epochs is set. The epochs are the number of times that all batches will be run through the network, each time in randomized order. Too few epochs will result in underfitting and too many epochs will result in overfitting. [6]

One tool for implementing this machine learning algorithms is TensorFlow [7]. TensorFlow is an open source platform which provides relevant CNN tools and libraries. TensorFlow is available in many programming languages.

2.1.1 Max-pooling

Max-pooling [8] is an operation often added to the CNN algorithm. Max-pooling is added after a layer and lowers the dimension of the images by reducing the number of pixels. The number of pixels is reduced by replacing a given area, called pool, of pixels with the largest pixel value. In max-pooling, a filter size and a stride is set. The filter size tells the size of the pool and the stride tells how the filter is slid over the image. Max-pooling could be done to help overfitting since it provides an abstract form of the image. [8].

2.1.2 Data augmentation

Small datasets that contains similar images could be favored using data augmentation. Data augmentation creates more variety in the dataset by duplicating the images and transforms the images by zooming, rotating or shearing. [9]

2.1.3 Drop-out

Drop-out [10] is a type of regularization where layers in the neural network are randomly ignored during training . In the CNN algorithm the probability for the drop-out is set. The probability of rejection could be between 0 to 0.5 but are often close to 0. Drop-out is used to prevent overfitting.

When training a network and not using drop-out, the neurons can become co-depending on each other. This could also be avoided using more data, but if the amount of data is limited, a good alternative is to use drop-out. [10]

2.2 Relevant work

There are several research projects that are similar to the subject of this thesis. In the article "Towards on-farm pig face recognition using convolutional neural networks" by Hansen, Smith, et al. [11], three face recognition methods have been applied on images of pigs. The three methods are Fisherfaces, transfer learning using the pre-trained VGG-Face model and CNN.

Fisherfaces is a method that uses a combination of Principal Component Analysis and Fisher's Linear Discriminant. The CNN was trained with self-captured images, as in this thesis. [11] brings out the challenges of lightning and contamination from dirt. One important difference with [11] and this thesis is that the images are taken in the same environment whereas this thesis has focused on varying the environment as the equipment and lightning. In this thesis the data collection was done manually by filming the horses which could give more control over the image collection, compared to [11] where the collection was done by a mounted webcam with less control over the image collection. The images in the dataset was selected using structural-similarity index measure (SSIM) in [11], whereas this thesis has selected the images manually. The SSIM resulted in that 70 % of the raw images was discarded and the final dataset was a lot smaller than the dataset used in this thesis.

The result in [11] shows that the best performing algorithm is CNN with the highest classification accuracy. It is explained that human face recognition tends to report higher accuracy for the Fisherfaces algorithm than in [11] which is probably because the training and test images are similar than face recognition for pigs.

3. Method

The method describes how the face recognition model and the dataset is created and evaluated.

3.1 Dataset

The five datasets used in this thesis was created from films and images of three horses. In Table 1 the number of images in each dataset can be seen. One dataset contained images that were cropped after the face of the horse.

Table 1. Summary of created datasets.

Size of dataset	Cropped	Number of horses
10	no	3
50	no	3
500	no	3
1000	no	3
50	yes	3

The datasets were created from 56 images and 67 short films of the horses. Table 2 shows the total number of images from each camera. The images and films were taken with a Canon 700D camera, a Nikon D700 and an Iphone 8 camera.

Table 2. Quantity of images collected.

Camera	Extracted images	Images
Canon 700D	13560	41
Nikon D700	0	8
Iphone 8	2310	7
Total	15870	56

From all films, 12 images per second were extracted. The images were then sorted by horse. The blurred images and images that not the whole head was involved in were removed. The datasets were then created by manually selecting images to the datasets according to Table 1.

The images and films were collected from two different occasions where the majority was collected under the second occasion. The goal of the data collection was to create datasets with a great variety between the images. This is illustrated with examples from the dataset in Figure 1 to Figure 5.

In Figure 2, an example of a horse from three different directions can be seen.

Figure 2. Example images from the dataset - from different directions. Left: right side, middle: front side, right: left side.

The directions in Figure 2 show a wide perspective. Small changes of the head alignment were also captured from the films as well as different facial expressions. The horses constantly moves their ears and mule creating different facial expressions.

Not only the different directions were captured, in Figure 3 an example of images with different weathers can be seen.

Figure 3. Example images from the dataset - different weathers. Left: cloudy, right: sunny.

In Figure 3, the most important difference between the images is that the images taken on a sunny day has more shadows. Images taken towards the sun can also contribute with characteristics in the dataset.

The data was also collected with three different cameras, cameras that had different settings. Using different equipment, the quality was varied as was settings of each camera. Figure 4 shows an example of two images taken at the same occasion using two different cameras.

Figure 4. Example images from the dataset - different cameras. Left: iphone 8, right: Canon 700D.

In Figure 4, two images of the same horse can be seen. One difference between the images is that the image to the left has warmer colors than the image to the right.

Figure 5 shows example of two images in two different environment.

Figure 5. Example images from the dataset - different environment. Left: outside, right: inside.

To the left in Figure 5 is a images of the horse outside and to the right is an image of the horse inside. Changing the environment gave different lightning and shadowing but also different noises in the background. Different noises in the images was captured. The noises captured in the background was cars, houses, other horses or humans. Example of one noise can be seen to the right in Figure 6. To the left in Figure 6, the noise has been removed by cropping the image.

A dataset to compare images with and without noise was created by cropping each image by the horse's head.

3.2 CNN algorithm

The face recognition consists of the deep learning algorithm CNN. The CNN algorithm was implemented in Python mainly using the libraries Tensorflow and Numpy.

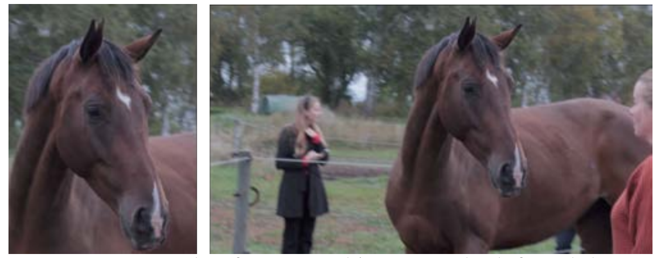

Figure 6. Example of a cropped image to the left and the original image to the right.

The first step in the algorithm was to load the dataset and randomly divide it into 80 % training data and 20 % validation data, as recommended by Tensorflow [7]. The training data was used to train the model and the validation data was used to evaluate the model.

All images in the loaded dataset were resized to the size 180x180. This is because the images in the original dataset had different sizes. RGB values of the dataset images, 0-255, is not ideal for neural network and all values were therefore normalized to the range of 0-1. After the preprocessing of the dataset the model was trained.

The model was trained using three convolutional layers. The first layer had 16 filters, the second layer had 32 and the third had 64 filters. All three layers had the kernel size 3x3. Max-pooling was applied for each convolutional layer, using filter size 2x2 and the stride 2. Lastly the input was flattened. The model was thereafter trained with different numbers of epochs.

Data augmentation and drop-out were implemented when training the model with the purpose to increase the accuracy of the model. The model was trained with data augmentation and drop-out both separately and combined. When adding drop-out to the CNN model, the probability 20 % was used, which was the overall best drop-out after trying different values.

3.3 Evaluation of dataset

All datasets were compared mainly using 5, 10, 15 epochs but also two cases using 30 epochs in the CNN algorithm. The CNN models were compared using data augmentation and drop-out, both separately and combined.

The datasets was first evaluated by comparing the accuracy and the loss between the training data and validation data.

The test images were manually selected for all three horses from images that were not included in neither of training nor validation datasets.

4. Result

In this section, the result of using different CNN methods and datasets is presented.

4.1 Evaluating the CNN method

Evaluating the CNN method, data augmentation and drop-out the dataset with 500 images was used. For comparability, the results are shown with 15 epochs.

Figure 7 shows the CNN performance without data augmentation and drop-out. The x-axis is the number of epochs and the y-axis shows the amount of correctly classified images for the left graph. The left graph in Figure 7 shows the accuracy of the CNN model. The blue curve is the training accuracy that represent a predicted accuracy based on the training images. The orange curve is the validation accuracy and represent the accuracy of how many images of the validation data that was classified correctly.

The right graph in Figure 7 shows the loss of the CNN model and is the errors of the predictions. The blue curve is the training loss and the orange curve is the validation loss. The y-axis in the right graph shows the sum of the errors that occurred when the algorithm classified images incorrect.

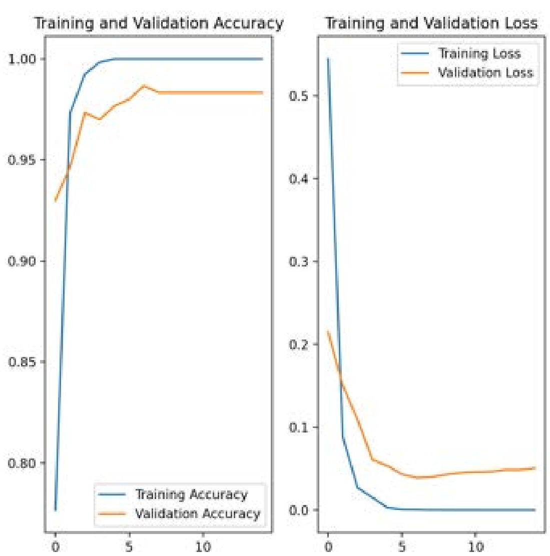

Figure 7. Model accuracy and loss using 500 images with three classes.

In Figure 7, the gap between the training accuracy and validation accuracy indicates overfitting. It can be seen that the validation accuracy stabilizes after six epochs. Considering the validation loss, the elbow shape of the orange curve indicates a maximum of six epochs.

Figure 8 shows the CNN performance with drop-out.

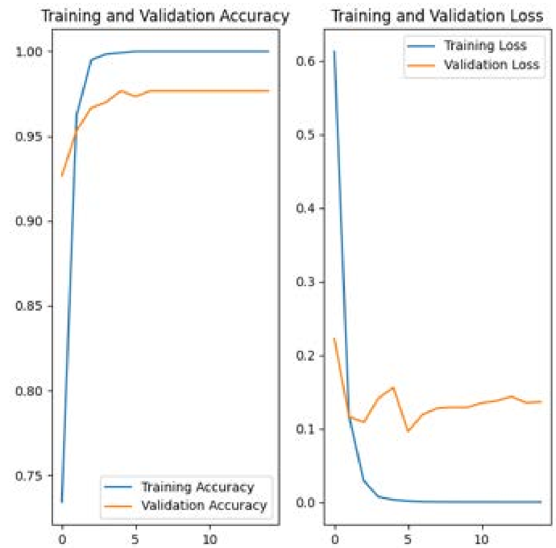

Figure 8. Model accuracy and loss with drop-out using 500 images in three classes.

The accuracy in Figure 8 performs similarly to the accuracy in Figure 7. The accuracy of CNN with drop-out stabilizes with fewer epochs than drop-out. Comparing the validation loss, CNN with drop-out testify worse results.

Figure 9 shows the CNN performance with data augmentation.

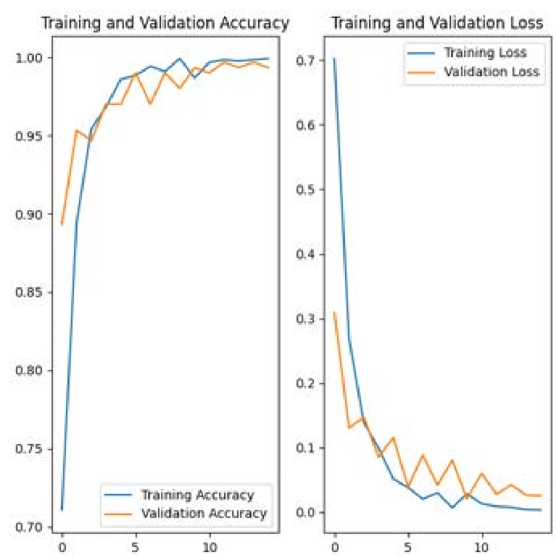

Figure 9. Model accuracy and loss with data augmentation using 500 images in three classes.

The CNN method with data augmentation has validation accuracy and validation loss with a closer fit to the training curve. However, the accuracy and loss curves in Figure 9 are more unstable than in both Figure 7 and Figure 8. It can be seen that data augmentation requires more epochs than the original CNN and CNN with drop-out. The training accuracy and the training loss stabilize after 10 epochs.

Figure 10 shows the CNN performance with both data augmentation and drop-out.

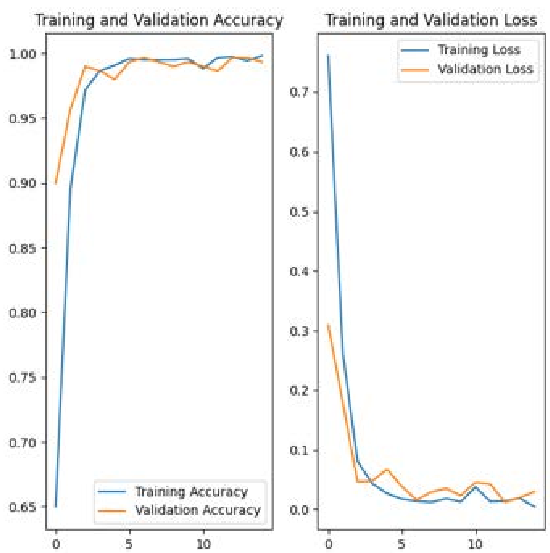

Figure 10. Model accuracy and loss with both data augmentation and drop-out using 500 images in three classes.

The closest fit between the training accuracy and validation accuracy and training loss and validation loss can be seen when both data augmentation and drop-out are used. The same pattern is identified for the other datasets.

The following results is therefore shown with the best CNN method using both data augmentation and drop-out.

4.2 Evaluation of different datasets

The different sizes of datasets were evaluated when both data augmentation and drop-out was added to the model. In Figure 11, the accuracy and loss of the model using the smallest dataset can be seen.

Using a dataset with 50 images, there are signs of both overfitting and underfitting. Both the validation accuracy and validation loss have a zig-zag shape and moves around the training curve. The oscillations of the training accuracy and training show that the model has not stabilized. Because the validation curves are unstable it is not possible to read the preferred number of epochs.

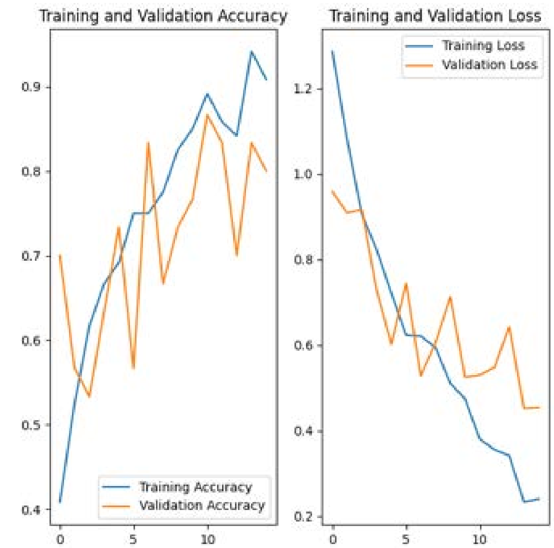

Figure 11. Model accuracy and loss with both data augmentation and drop-out using 50 images in three classes.

The second largest dataset is seen in Figure 10. Looking at Figure 10, using the dataset of 500 images, both training and validation accuracy is quite good, the lines are following each other quite accurately. From this graph six epochs can be considered enough although there are still some oscillations.

Figure 12 shows the accuracy of the model using the dataset with 1000 images.

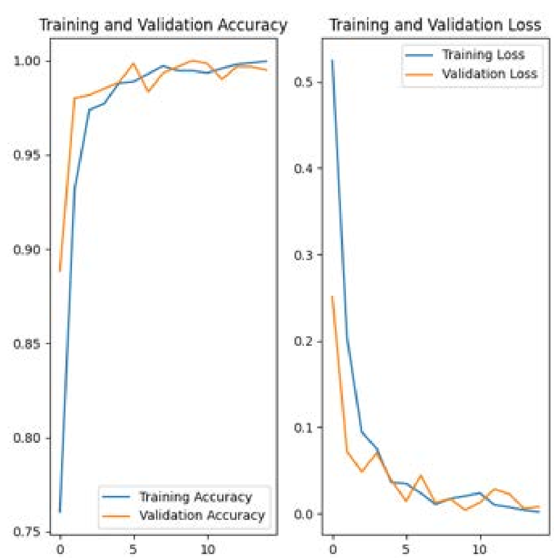

Figure 12. Model accuracy and loss with both data augmentation and drop-out using 1000 images in three classes.

Using the dataset of 1000 images there are some oscillations, but the accuracy of the model is quite good, since both training and validation accuracy are following each other closely.

Comparing the two large datasets, the accuracy and loss has not improved with more images, see Figure 10 and Figure 12. The dataset with 500 images and the dataset with 1000 images both has unstable curves with oscillations. The training accuracy are close to 100 % in both models with a validation accuracy following closely.

Table 3 presents the numbers of incorrect predictions of the 15 test images of each horse. The results are generated with a model with both data augmentation and drop-out and 15 epochs.

Table 3. Incorrect classifications of 15 test images of each horse.

Horse	10 Images	50 Images	500 Images	1000 Images
A	4	1	2	1
B	3	1	0	1
C	1	5	3	1
Total	8	7	5	3

In Table 3, the most incorrect classifications can be seen for the smaller datasets. When the dataset with 10 images is used a total of nine images are incorrect classified.

Table 3 show that 10 and 50 images is not enough for a dataset with this CNN model. From this result it can be seen that the largest dataset of 1000 images are better than the dataset with 500 images, compared from the result in Figure 10 and Figure 12.

4.3 Evaluations of small datasets

Previous result concluded that using the dataset with 50 images resulted in a several incorrect classifications and an unstable model.

In this section the noise in the background of the image has been removed by cropping the images in the dataset. Figure 13 shows that the model is still not stable, but more accurate compared to the not cropped setup seen in Figure 11.

Both the accuracy and the loss in Figure 13 have zig-zag shape where signs of both underfitting and overfitting can be seen. None of the models in Figure 11 or Figure 13 are stabilized.

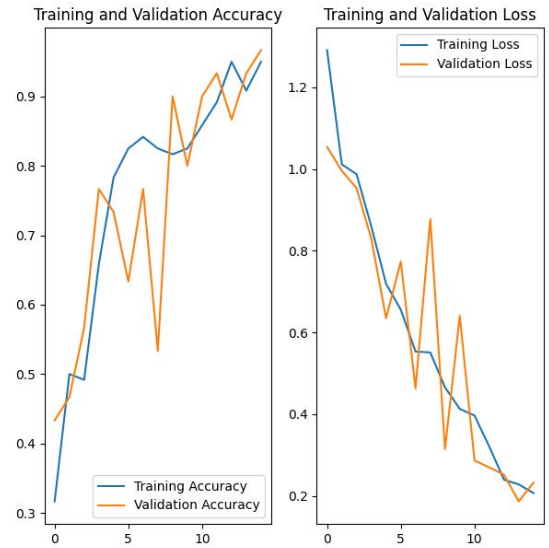

Figure 13. Model accuracy and loss with both data augmentation and drop-out using 50 cropped images in three classes.

As an attempt to find the number of epochs of where the model stabilizes, the model with both data augmentation and drop-out is generated with 30 epochs. Figure 14 shows the accuracy of the model using cropped images and 30 epochs with both data augmentation and drop-out.

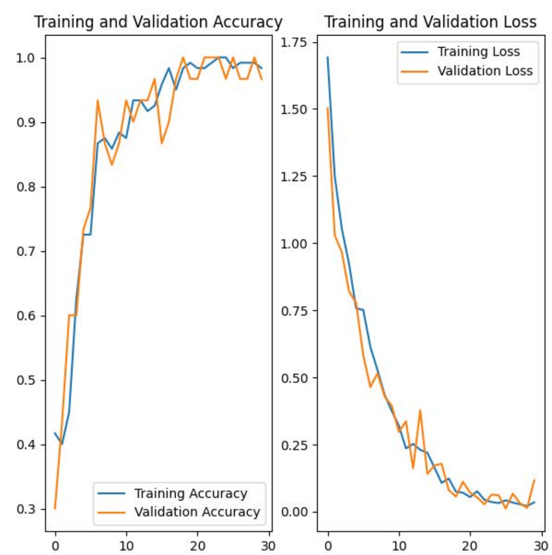

Figure 14. Model accuracy and loss with both data augmentation and drop-out using 50 cropped images in three classes.

In Figure 14, the training accuracy are closer to 100 % and the validation accuracy seem to follow more closely. Using a large number of epochs, the accuracy is not fully stabilized but the oscillations are not as large as in Figure 13.

The results from the test images using the cropped dataset with 30 epochs was five incorrect predictions for horse A, seven for Horse B and five for Horse C which in total are 17 wrong predictions. In comparison with the test results in Table 3 there is a big increase of the number incorrect predictions. However, it is important to note that the test images are the same throughout this thesis and the test images are not cropped.

5. Discussion

The CNN performance is known to be dependent on the size of the dataset, whereas large dataset gives better performance. This could be seen in the result of this thesis as well. The advantage of large datasets is that it includes varied images and the noisy images does not have a large effect on the learning of the algorithm. The problem is that it is time consuming to create large datasets.

The algorithm is designed as a tool to help the farmers in their daily work, daily supervision of horses, behavior analyses and authorization in stables. Some serious problem can accrue if the algorithm is used incorrectly. The chance for miss classification needs to be considered, as this could result in catastrophically outcome if the algorithm is blindly trusted. The algorithm should therefore be used as an tool in combination with human supervising to reduce errors.

5.1 Datasets

The images in the dataset was selected manually with the aim to create datasets with large variations. Although there are many aspects included in the dataset there are some missing. For example images with snow in the background. It would have been interesting to test the model with an image with snow in the background to see if the model is able to recognize the horse.

Selecting images and creating the dataset manually was time consuming and could have been done by an algorithm as SSIM as in [11]. The algorithm could also have assured more variations between the images, variations that are difficult to see with the eye. With this in mind, the model could have preformed differently.

The number of images in the datasets were 10, 50, 500 and 1000. There is a big difference between the graphs using a dataset with 50 images and a dataset with 500 images. It would have been interesting to add datasets of a size between 50 images and 500 images since it is a large range. Using a dataset of 100, 200 or 300 images, a tipping point could have been identified where the size of the dataset is minimized, and the performance of model is good enough. Instead, other methods had to be tested to work around the problems with small datasets.

5.2 Evaluation

The algorithm divides the images randomly into training and validation and each run gives small variations in the results. Although there are small variations in the result it is important to keep in mind during the comparison between results. The variations in the results are considered to have a small impact of the comparison but to increase the reliability 15 test images was used.

The test images were manually selected from the original dataset. Depending on the images that was selected to create the dataset, the model was prepared for certain images and some of the test images could be more difficult for the model to classify. Although some of the images could be more difficult to classify, all models have been evaluated with the same test images.

5.3 Future Work

As a part of future work, the CNN algorithm could be implemented with transfer learning which could improve the results. Transfer learning is often used with small datasets and is a development of CNN. In transfer learning an already trained model is used, which means that you do not need a large dataset to train the model.

6. Conclusion

The project has achieved its goals of finding an algorithm that works to identify horses. The conclusion that can be drawn is that the more images and epochs, the better result. CNN as a method itself does not work so well but need dropout and data augmentation to get the best result. Using a small dataset, the number of epochs needs to be increased to achieve a sufficient results. There are improvements and this project is designed according to CNN with three convolutional layers, which means that the result may differ if other datasets, properties or layers of the CNN method are used.

Acknowledgments

We got our idea for this thesis after a discussion with Per Frankelius and Matilda von Rosen. Thanks for supporting us and enlighten us about this subject. Special thanks to Per Frankelius who helped us collecting images for our dataset.

References

[1] Konstantinos G. Liakos Patrizia Busato Dimitrios Moshou Simon Pearson and Dionysis Bochtis. Machine Learning in Agriculture: A Review. *Sensors 2018*, 18(8), 2018.

[2] Mundy Jon. How artificial intelligence could save humanity's food supply. *World of tech*, 2016.

[3] Miller A.L. PlÖtz T. et al. Matthews, S.G. Automated tracking to measure behavioural changes in pigs for health and welfare monitoring. *Scientific Reports*, 7:17582, 2017.

[4] Tarang Shah. Measuring object detection models — map — what is mean average precision?, 2018.

[5] Jason Brownlee. How do convolutional layers work in deep learning neural networks?, 2019. Deep Learning for Computer Vision; Online; accessed 20 October 2020.

[6] Gareth James Daniela Witten Trevor Hastie and Robert Tibshirani. An introduction to statistical learning with applications in r. In *Springer New York Heidelberg Dordrecht London*, 2013.

[7] TensorFlow. An end-to-end open source machine learning platform. Online; accessed 16 October 2020.

[8] A. Giusti, D. C. Cireşan, J. Masci, L. M. Gambardella, and J. Schmidhuber. Fast image scanning with deep max-pooling convolutional neural networks. In *2013 IEEE International Conference on Image Processing*, pages 4034–4038, 2013.

[9] TensorFlow. Image classification, 2020. Online; accessed 15 October 2020.

[10] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[11] L. Salter M. Baxter E. Farish M. Grieve B. Elsevier B.V. Hansen, M. Smith M. Smith. Towards on-farm pig face recognition using convolutional neural networks. . *Computers in Industry*, 98:145–152, 2018.