Semi-Supervised Active Learning to Efficiently Train a Classifier on a High Volume of Unlabelled Images

Jette Korthals Altes
jette@korthalsaltes.com
University of Amsterdam

ABSTRACT

Training a deep convolutional network to be able to classify images requires a large volume of labelled images, which can be expensive to get by. Semi-supervised learning is a way to train a network on less labelled images. Active learning is a way to select the images whose labels will provide the most information to the training of the network. Semi-supervised active learning is applied to train a classifier on the Fashion MNIST dataset, and its performance is compared to a classic supervised learning approach, by training a classifier with the same architecture on the same number of labelled images. Semi-supervised active learning outperforms supervised learning in terms of accuracy obtained on the test set, especially on a lower number of manually labelled images.

1 INTRODUCTION

An abundance of images is widely available. This includes a great number of images, which could be used to complete computer vision tasks such as classification through deep learning. Unfortunately, training a classifier requires a large amount of labelled images. The more labelled images per class during training, the better the performance of the classifier [6]. However, these high volumes of available images do not automatically come perfectly labelled into the classes relevant for the classification problem of interest. This problem is brought forward by Chen & Lin [2] in their survey on deep learning for Big Data when discussing the challenges of deep learning from high volumes of Big Data. This labelling can be time-consuming and labour-intensive, or require experts on the contents of the images. This is the case in for instance the medical field, where laypeople can simply be unable to recognise the contents of an image [5]. This burden to label a large amount of images can be relieved by using semi-supervised learning instead of supervised learning [2]. With semi-supervised learning, the classifier learns the underlying data distribution from the unlabelled images, and outperforms a fully supervised classifier trained on the same amount of labelled images [6].

However, randomly labelling a number of images and training a semi-supervised classifier on them may not be the most efficient approach. By chance, only a few classes could be represented in the labelled subset. Active learning could prevent this. Active learning is a version of semi-supervised learning, where the learning algorithm selects which images require labelling [8]. This way, the most important images will be labelled, keeping the amount of labelling that is needed to a minimum.

2 RELATED WORK

Several different approaches are possible to implement active and semi-supervised learning. Gal et al. [4] use a Bayesian deep learning framework to quantify the uncertainty of their convolutional neural network in order to select which images to label.

Wang et al. [9] apply a pseudo-labelling approach when using active learning to train a classifier on images. After some initial labelling, the classifier starts to classify the images. The images with the lowest certainty are presented to a human annotator to give labels to. The images with the highest certainty get appointed a pseudo-label, and those pseudo-labels are used for further training. This leads to more images with labels, but the possibility for errors in the pseudo-labelled part of the data.

Several different approaches can be made to select the initial datapoints that are labelled. Randomly selecting these may induce a bias. Therefore, Dasgupta & Hsu [3] propose using the underlying structure of the data by clustering the data first. Clustering is a form of unsupervised learning, where data is grouped by similarity without the necessity of labels. By clustering first, the initial labelled samples follow from the data, and would not introduce a sampling bias. This clustering approach has not been applied to the classification of images.

3 RESEARCH QUESTION

This paper will compare the performance of a semi-supervised active learning approach to a supervised learning approach in classifying an image dataset to determine which approach is better to deal with a large volume of unlabelled images. For the active learning approach, the clustering approach of Dasgupta & Hsu [3] is adapted to work on an image dataset is and combined with the pseudo-labelling approach of Wang et al. [9]. For the supervised learning approach, a classifier with the same architecture is trained on the same number of labelled images as is labelled in the active learning approach.

4 IMPLEMENTATION

The Fashion MNIST dataset [10] is used to compare semi-supervised active learning to supervised learning. This dataset consists of 70,000 grayscale images, equally divided over ten classes depicting clothing items. These classes are T-Shirt, Top, Trouser, Pullover, Dress, Coat, Sandals, Shirt, Sneaker, Bag, and Ankle Boots. The train set consists of 60,000 images and the test set consists of 10,000 images. The images are 28 by 28 pixels.

No validation set is used during the training of the networks. Usually, a validation set is used to tweak the hyperparameters of a convolutional neural network. However, the use case for which active and semi-supervised learning would be used is one with an unlabelled dataset. Therefore, validation would likely not be executed.

For the semi-supervised active learning approach, the initial images that receive labels are selected through clustering. Each image is transformed into a one dimensional array, and each pixel value is treated as a feature. The K-Means algorithm from the scikit learn library in Python is used to create ten clusters [7]. Ten initial centroids are initialized using the k-means++ initialisation scheme [1]. All remaining images are assigned to the centroid that is closest in the feature space, as measured by the Euclidean distance. New centroids are created by taking the mean value of the images in each cluster, and the images are reassigned accordingly. This process is repeated until the distance between the old and new centroids is smaller than 0.0001 or if the algorithm has run for 300 iterations. From these ten clusters, the five images closest to the cluster centroids are selected and manually labelled, resulting in 50 labelled images. Since the Fashion MNIST dataset is actually a labelled dataset, "manually labelling" does not refer to the author looking at the image and assigning it a label, but refers to a bit of code that looks up the corresponding label and assigning it.

The training process of the semi-supervised active learning approach is displayed in Figure 1. The ADADELTA optimizer is used, and the loss is defined as the categorical cross-entropy. A convolutional neural network will be trained for 15 epochs on these 50 labelled images. The architecture of this network is displayed in Figure 2. This network is then used to predict the labels of the 59,950 still unlabelled images. Of these 59,950 predictions, the 50 images with the least certain predictions are selected and labelled manually. Furthermore, the 50 images with the most certain predictions are selected and pseudo-labelled, meaning they receive the label that was predicted by the network. The network is trained further on the now 150 labelled images for another 15 epochs. This process is repeated until 10,000 images are manually labelled.

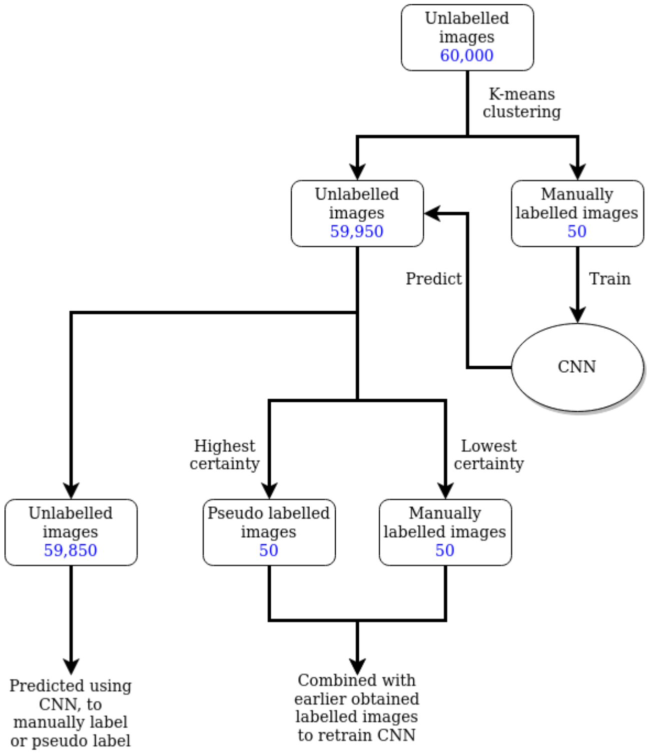

Unlabelled images 60,000
K-means clustering
Unlabelled images 59,950
Manually labelled images 50
Predict
Train
CNN
Highest certainty
Lowest certainty
Unlabelled images 59,850
Pseudo labelled images 50
Manually labelled images 50
Predicted using CNN, to manually label or pseudo label
Combined with earlier obtained labelled images to retrain CNN

Figure 1: Diagram of the training of the semi-supervised active learning approach. CNN = Convolutional Neural Network

For the supervised learning approach, a convolutional neural network with the same architecture will be trained on the same number of labelled images as is manually labelled in the active learning approach. A difference between this classic training approach and the semi-supervised active learning approach is that with the semi-supervised active learning approach, the network is retrained every iteration, whereas it is uncommon in a supervised learning setting to increase the number of training examples every few epochs. Therefore, two different versions will be executed of the supervised learning approach: the classic approach, where the weights of the network are reinitialised after every 15 epochs, and the retraining approach, where the weights are kept and more labelled images are added to the training set.

If a difference is found between the two approaches, it would be interesting to determine whether this difference is due to the active learning or due to the semi-supervised learning. Therefore a training approach with only active learning is implemented. The same course of training is followed as displayed in Figure 1, except the pseudo-labelling is not carried out. A training approach with only the semi-supervised learning method is also implemented. In this approach, the initial images that receive labels are not selected through

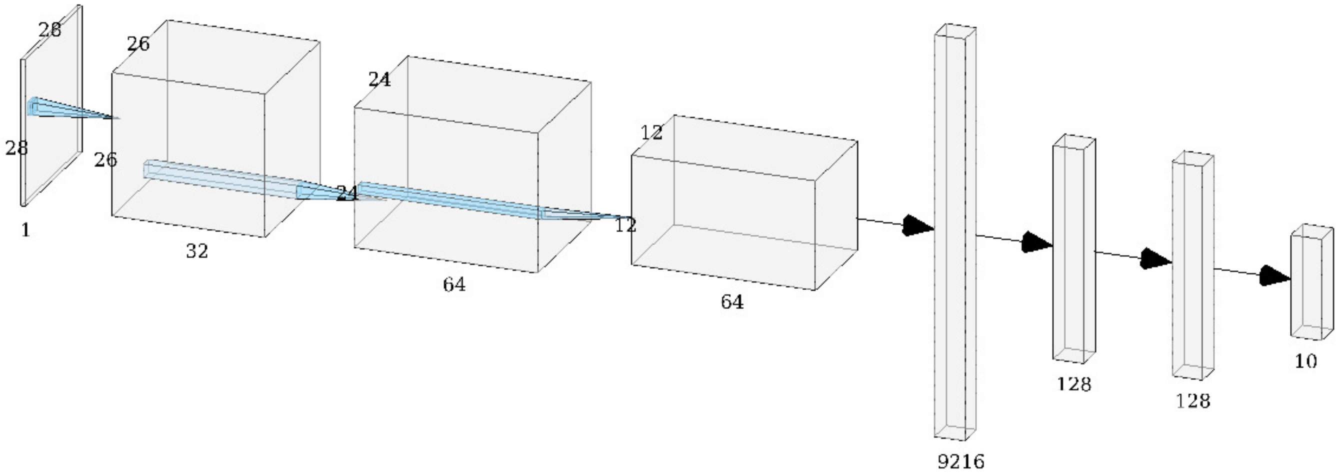


Figure 2: Architecture of the convolutional neural network

clustering, but through random selection. After training the network for 15 epochs and predicting the labels of the remaining unlabelled images, the 50 images with the most certain predictions are selected and pseudo-labelled with the predicted labels. So with this approach, in total only 50 images are manually labelled.

To determine the effectiveness of the clustering approach, another variation of the active learning approach is implemented as well. One in which the first 50 images that receive labels are selected at random, instead of obtained through clustering. This variation will be implemented in both the semi-supervised active learning approach as in the active learning approach.

To summarise, seven networks are trained:

- Semi-supervised active learning with clustering
- Semi-supervised active learning with random initialisation
- Active learning with clustering
- Active learning with random initialisation
- Semi-supervised learning
- Supervised learning with retraining
- Supervised learning without retraining

The accuracy of all networks on the test set of the Fashion MNIST dataset is compared to each other to determine whether a combination of active learning and semi-supervised learning outperforms supervised learning.

The implementation is available at https://github.com/JetteKA/BigData.

5 EVALUATION

Prior to training, the images were clustered in order to select the initial images that receive labels. The five images closest to the centroid of each of the ten clusters are displayed in Fig. 3. In six out of the ten clusters, the five centre images belong to the same class. Of these centre images per cluster, one consists fully of trousers, one of sneakers, two of ankle boots, and two of bags. The four classes corresponding to clothing of the upper body are spread over the four remaining clusters. There is a cluster with a mix of shirt and t-shirt/top, one with pullover and coat, and two with pullover and shirt. Images corresponding to the labels dress and sandal are not present in the centre images of the clusters.

The accuracy obtained on the test set is plotted against the number of manually labelled images the networks have been trained on, except for the semi-supervised learning network (see Figure 4a). This network is only trained on 50 manually labelled images, the initial 50, and the rest of the images it is trained on is pseudo-labelled. Therefore, the accuracy obtained over the test set is plotted separately against the number of pseudo labelled images (see Figure 4b).

In Figure 4c the difference in performance between the semi-supervised active learning with clustering approach and the supervised learning without retraining can be seen. After manually labelling 10,000 images, the semi-supervised active learning approach achieves an accuracy of 92% while the supervised learning approach reaches 90%. This 92% is already reaches after manually labelling 5,500 images, at which point the supervised learning approach has 87%. What can be observed from this graph is that the supervised learning approach appears to still improve in accuracy, while the semi-supervised active learning approach appears to have stopped improving.

A difference can also be observed at the start of training. After the networks are trained on the 50 initially labelled images, the semi-supervised active learning approach achieves 45% accuracy, whereas the supervised learning approach

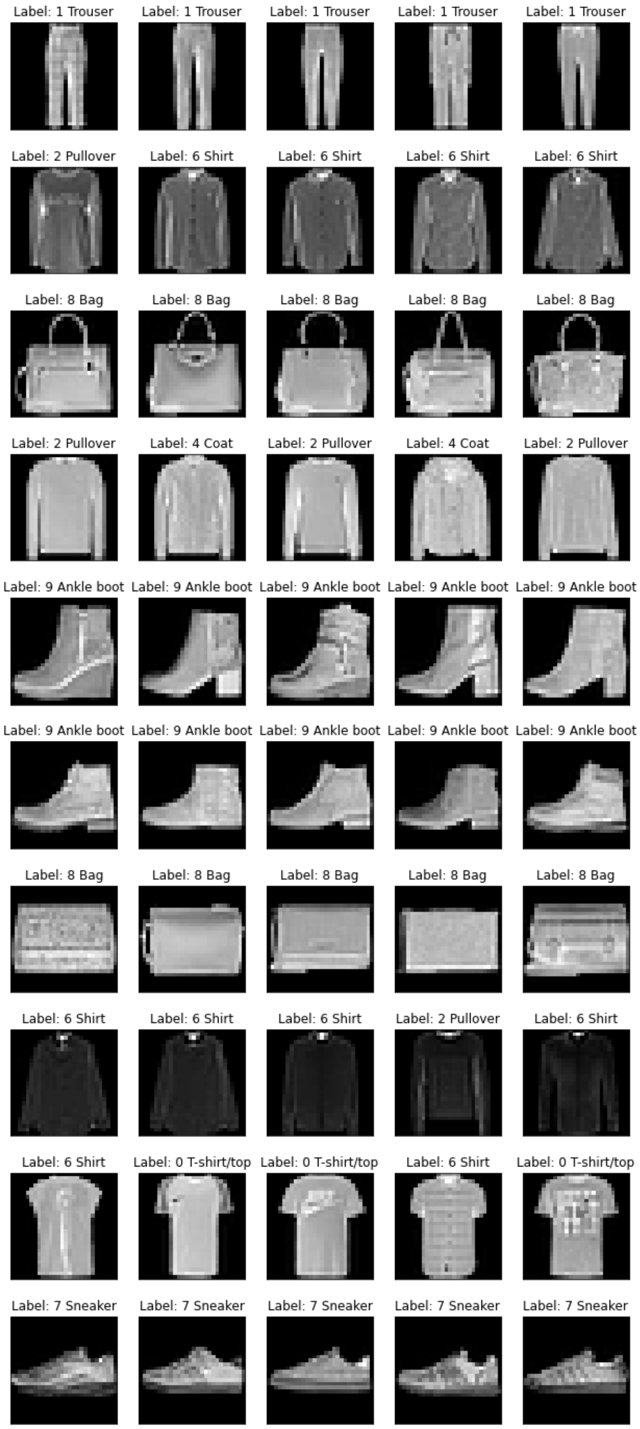


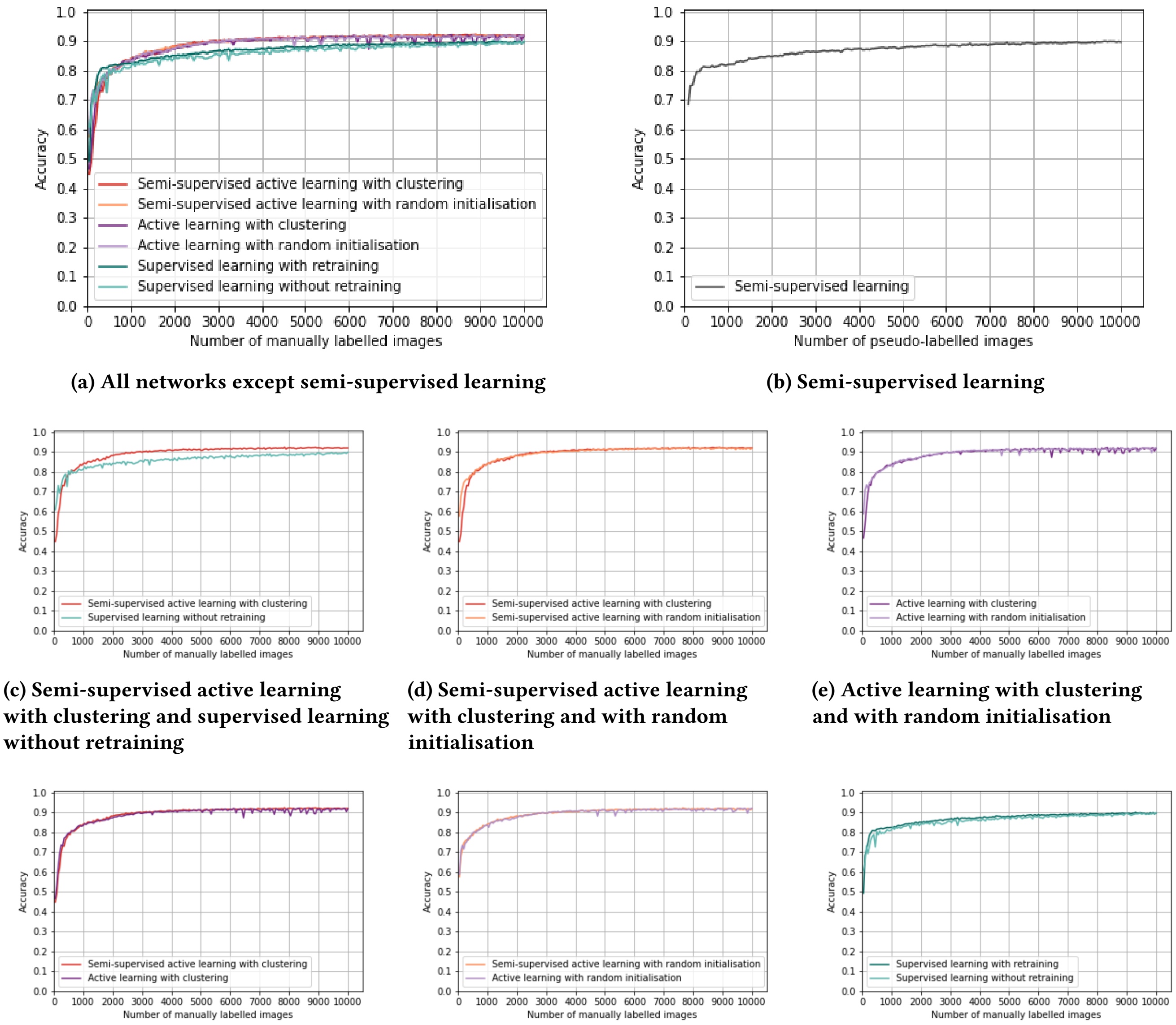
Figure 3: Each row represents one cluster. The five images closest to the centroid of each cluster are displayed, with their correct label. These 50 images are used to start training the network in the semi-supervised active learning approach with clustering.

achieves 61% accuracy. The difference between these two approaches at this point in training is the clustering versus the random selection of the initial images that are humanly labelled. This difference can also be observed when comparing semi supervised active learning with clustering to semi-supervised active learning with random initialisation (Figure 4d) and comparing active learning with clustering to active learning with random initialisation (Figure 4e). The approaches using clustering perform worse after the first round of training than the approaches using clustering to select the images. This difference cannot be observed after 500 images are manually labelled.

Figure 4h depicts the accuracy of the supervised learning approaches with and without retraining of the networks. Learning with retraining appears to achieve slightly higher accuracy than without retraining, this can no longer be observed after manually labelling 9,000 images. Furthermore, supervised learning with retraining appears more stable than without retraining. There is less variation in the accuracy obtained. This can be explained by the random initialisation of the weights of the network. Without retraining, each new iteration, the weights of the network are randomly initialised. By chance, these could be closer or further away from where they would need to be to accurately predict labels.

Is this difference between supervised learning and semi-supervised active learning due to the active learning, the semi-supervised learning, or the combination of the two? Comparing semi-supervised active learning to active learning, both appear to perform equally well (see Figure 4f for with clustering and Figure 4g for with random initialisation). In both cases, the obtained accuracies are pretty much equal, which could suggest that semi-supervised learning does not add much to the active learning approach. All achieve 92% accuracy after 5,000 manually labelled images. However, the semi-supervised active learning approach does appear to be more stable than the active learning approach. The semi-supervised learning approach is only trained on 50 manually labelled images, the rest of the images it is trained on are all pseudo labelled, making the comparison not as straightforward. After training for the same amount of epochs as the (semi-supervised) active learning approach, on 50 manually labelled and 9,950 pseudo-labelled images, an accuracy of 90% is achieved. Comparing this to the performances of the (semi-supervised) active learning approaches considering the amount of manually labelled images, the semi-supervised learning approach performs a lot better. Comparing this to the performance at the same amount of epochs, the (semi-supervised) active learning approach performs better.

Looking at the pseudo-labels assigned in the semi-supervised learning approach, 97% is correctly labelled. When using the semi-supervised active learning approach, 99% of pseudo-labels are correct.



(a) All networks except semi-supervised learning

(b) Semi-supervised learning

(c) Semi-supervised active learning with clustering and supervised learning without retraining

(d) Semi-supervised active learning with clustering and with random initialisation

(e) Active learning with clustering and with random initialisation

(f) Semi-supervised active learning and active learning, both with clustering

(g) Semi-supervised active learning and active learning, both with random initialisation

(h) Supervised learning, with and without retraining

Figure 4: Accuracy obtained by the different networks on the test set of Fashion MNIST after training on a number manually labelled images, or pseudo-labelled images for the semi-supervised learning approach

6 DISCUSSION

Semi-supervised active learning outperformed the supervised learning approach as hypothesised. In terms of amount of human effort necessary, semi-supervised learning outperforms both active learning and semi-supervised active learning, since a better accuracy is obtained with less manual labelling. However, the pseudo-labels assigned in the semi-supervised learning approach are more often false than they are in the semi-supervised active learning approach. Therefore, it would depend on the situation for which the classifier would be deployed. If a certain error rate is acceptable, semi-supervised learning greatly reduces the amount of human effort, while with the combination of active learning and semi supervised learning, a large number of images would still need to be manually labelled, but the error rate of the labels during training would be less.

The clustering was not a great success. The images closest to the centroids are clustered pretty well, except the four classes of upper body clothing. Semantically, it makes sense that there would be overlap between these classes, since they are similar, but this shows that the clustering does not divide the classes perfectly. Because of this, not all classes are equally represented in the initial labelled set. The dataset was clustered by applying k-means to the flattened images and treating the pixel values as features. A more accurate result could have been obtained by extracting features using a convolutional neural network and clustering those features. This approach was not taken in this project for the sake of simplicity. In this case, the number of classes was known beforehand. There are cases imaginable where this is not the case. The amount of classes could be determined during the clustering process.

A problem in this approach in practice is the test set. In this project, a dataset was used for which labels were available. A large test set was also available. But if such a large test set were available for the dataset a network is trained on, there is little need for using a semi-supervised or active learning approach. In reality, such a large test set would not be available, making it more difficult to assess the performance of the trained network.

This was also the reason for not using a validation set in this project. Therefore, the hyper-parameters were set kind of randomly. This includes the number of images to label each iteration, as well as the number of epochs to train for.

In conclusion, semi-supervised and active learning could be very useful when resources are scarce to label a large dataset.

REFERENCES

[1] David Arthur and Sergei Vassilvitskii. 2006. *k-means++: The advantages of careful seeding.* Technical Report. Stanford.

[2] Xue-Wen Chen and Xiaotong Lin. 2014. Big data deep learning: challenges and perspectives. *IEEE access* 2 (2014), 514–525.

[3] Sanjoy Dasgupta and Daniel Hsu. 2008. Hierarchical sampling for active learning. In *Proceedings of the 25th international conference on Machine learning*. 208–215.

[4] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. 2017. Deep bayesian active learning with image data. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 1183–1192.

[5] Ali Madani, Mehdi Moradi, Alexandros Karargyris, and Tanveer Syeda-Mahmood. 2018. Semi-supervised learning with generative adversarial networks for chest x-ray classification with ability of data domain adaptation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 1038–1042.

[6] Augustus Odena. 2016. Semi-supervised learning with generative adversarial networks. *arXiv preprint arXiv:1606.01583* (2016).

[7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[8] Burr Settles. 2009. *Active learning literature survey.* Technical Report. University of Wisconsin-Madison Department of Computer Sciences.

[9] Keze Wang, Dongyu Zhang, Ya Li, Ruimao Zhang, and Liang Lin. 2016. Cost-effective active learning for deep image classification. *IEEE Transactions on Circuits and Systems for Video Technology* 27, 12 (2016), 2591–2600.

[10] Han Xiao, Kashif Rasul, and Roland Vollgraf. 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747* (2017).