A Deep Learning Approach for Classification of Cleanliness in Restrooms

Lahiru Jayasinghe, Nipun Wijerathne, and Chau Yuen

Singapore University of Technology and Design, Singapore

Email: {*aruna_jayasinghe,gallage_wijerathne,yuenchau*}@*sutd.edu.sg*

***Abstract*—Facilities hire cleaning companies to maintain and manage cleaning operations on their restrooms by deploying cleaners who are responsible for performing frequent checks to ensure the cleanliness of restrooms. Nevertheless, the perception of quality and word, *clean* is very subjective to the observer. Hence, it is not an easy task to quantify the cleanliness. This paper presents a deep learning approach using deep convolutional neural networks (DCNN) to detect and classify the level of cleanliness in restrooms into three different categories; namely dirty, average, and clean. Our method sheds new lights on data augmentation, feature extraction and knowledge transfer between models. The proposed architecture achieved a precision of 0.98, 0.95, 0.80 and recall of 0.99, 0.83, 0.95 for dirty, average, and clean categories respectively utilizing a dataset collected from an active restroom facility.**

***Index Terms*—Deep Learning, cleanliness classification, Deep convolutional neural networks, principle component analysis**

I. Introduction

It's common knowledge that restrooms could be a harmful place when it is dirty. Therefore, people are profoundly concerned about the health hazards and cleanliness of shared and communal restrooms because, these places are usually not very clean.The toilet survey study in [1] conducted by Singapore government has shown, the people are unhappy about the cleanliness of public restrooms and Australian government persuades people to be more concern about diseases, which could get spread from these poorly maintained restrooms [2, 3]

Because of these reasons, people even tend to avoid restaurants with dirty restrooms [4]. Recently, countries trying very hard to address this pressing issue, the government of China has installed 50000 new restrooms nationwide [5], National Environment Agency of Singapore has launched a program called Happy Toilet [6], to give recognition to restroom owners, operators and cleaners for maintaining restrooms up to a standard expectation [7].

Since all these facts are directly pointing towards the restrooms cleanliness and maintenance, hence, cleaning service providers and facility managers need to change their traditional methodologies and they must attain new technologies and solutions to provide the expected quality of service from them.

A research group has proposed a solution for smart cleaning called Restroom Visitilizer System [8], which utilize people counters and odour sensors to determine the ammonia level present in restrooms and currently its been used at many facilities[9]. Two other research groups have proposed similar solutions, which utilize people counters and Ammonia Sensors to detect the dirtiness in restrooms and they have provided an end-to-end solution for the facility managers to monitor the cleanliness in their restrooms [10, 11].

To this end, the previous work focused on the odour based sensor systems to detect the dirtiness but there could be other odourless dirtiness, which can occur in a restroom. For instance, tissue papers often spread all over the restroom floor and sink mirrors, taps, knobs and other hand touching places get dirty very easily. Moreover, odour based sensor systems need considerable number of sensors to cover the whole restroom, which is sometimes can be inconvenient. Limited interpretability of existing odour sensors and complex correlations among readings from various sensors deployed in different facilities could be a challenge [10].

Recently, there has been much research conducted on deep learning architectures to address problems in medical image analysis, speech and audio processing, signal processing, time series prediction, visual recognition, games etc [12–14]. Visual recognition can be considered as one of the most popular fields among deep learning researchers, a vast amount of research studies have been published about image classification and detection, because most of the real world problems can be solved by analyzing specific images [12, 15, 16]. All these novel architectures were possible due to the use of Convolutional Neural Networks (CNN) [17]. It has the ability to learn meaningful features from a given dataset by iterative learning, then after, a classifier could be utilizes to classifying data. Based on CNN, many deep architectures have been introduced for image classification and detection using ImageNet's Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) dataset [18–21].

We employed this concept and implemented a new approach to address odourless dirtiness using Deep Convolutional Neural Network (DCNN), which is a different aspect from previous work [8, 10, 11]. To the best of authors knowledge, deep convolutional neural networks haven't been yet applied before in the context of smart cleaning for classifying orderless dirtiness in restrooms. The aim of our work is to implement an adaptive model that can be used in any restroom facility with less amount of fine-tuning. Our dataset comprises images which are captured by fixed cameras in restrooms. The data labeling need to be done according to the facilities owner or cleaning service provider standards and they must decide which images require to be considered as clean, average or

(a) Category:Dirty

(b) Category:Average

(c) Category:Clean

Fig. 1: Example data from each category in dataset s

dirty.

There were some challenges we had to address in order to achieve an acceptable result. The limited dataset for classification and the dataset contains variety of inconsistencies mainly on the camera angel and lighting conditions, recognizing a qualified deep architecture for our dataset as the feature extractor and implementing a classifier that could classify the extracted features. These can be consider as challenges we encountered in this work.

Since the size of dataset is limited for each category and this data contains some inconsistencies, data augmentation techniques are applied on the dataset. We have proposed a novel method of finding the best model and the transfer layer to apply transfer-leaning. After employing this method, we selected InceptionV3[18] as the DCNN to extract features from training data and extracted features are classified by using multilayer perceptron neural network (MLP-NN) into three categories; namely *dirty*, *average* and *clean*. They are the solutions we proposed for our challenges and also could be consider as contribution of our work.

Although we have covered a different aspect compared to previous work [8, 10, 11], it should be noted that since cameras encompass a wide range, the number of cameras required, significantly less than the number of sensors required in the previous methods. However, sometimes cameras could not cover every details of environment, cameras could have blind spots. Therefore, by combining our solution and previous work, a sophisticated and intact system can be implement.

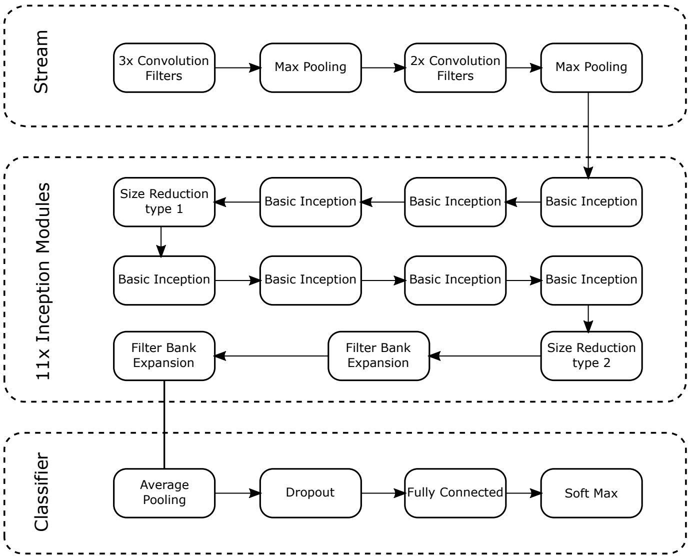


Fig. 2: Inception V3 architecture without the auxiliary classifiers

Another advantage of our proposed architecture is, after applying fewer numbers of fine-tuning, the system can be easily deployed in a different facility. These can be considered as our system adaptabilities for different environment. The paper validates the proposed scheme experimentally using urinal bowl images.

The paper is organized as follows. The proposed architecture is presented in Section II. Section III discuss the results of each phase described in Section II thoroughly. Finally, Section IV concludes the paper.

II. Proposed Architecture

As shown in Fig 3, the proposed solution consists of five major phases, dataset preparation, data augmentation, feature extraction, proposed classifier and model training-testing . The collected dataset is limited and unbalance, hence, the dataset needs to be artificially enlarged by using data augmentation techniques (Reason explained in subsection II-A). The data augmentation increases the size of a given dataset by five times and collectively the training data set becomes six times larger than the initially balanced dataset see Fig. 3. Then, pre-trained Inception V3 [18] DCNN model was selected to extract features from the data. The selected feature extractor is a very deep architecture consists of, image stream, eleven inception modules, and a fully connected neural network classifier see Fig. 2. The input image stream consists of five convolutional layers along with two max pooling layers. The most crucial part of feature extraction in Inception V3 model is the stack of eleven inception modules. Inception stack is made of four types of different inception modules, namely, basic inception module, size reduction module and two types of filter bank expansion modules. The basic inception module designed to approximate the optimal local sparse structure of features. Size reduction modules are responsible for reducing model dimension otherwise the computational requirements would be too heavy. The last inception modules are the filter bank

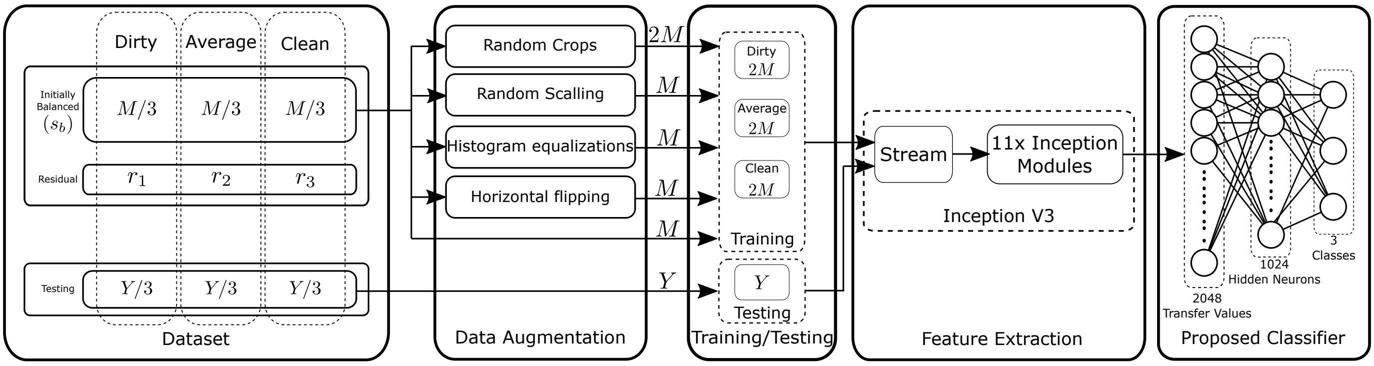


Fig. 3: The proposed architecture for the cleanliness classification

expansion modules, which used to increase the dimensional representations to the classifier. Finally, two layered (including average pooling layer) neural network classifier, as the output of Inception V3 model. Size of the fully connected layer in Inception V3 classifier represents the number of classes, which it's designed for classification.

The proposed method for selecting this deep architecture as our feature extractor is explained in subsection II-C. After that, according to Fig. 3, the extracted features are classified by using a three-layered MLP-Neural network classifier into given categories. This proposed classifier explained in subsection II-D and the training and testing methods for the classifier described in subsection II-E.

A. *Dataset preparation*

The dataset consists of 4032×3024 pixels sized RGB urinal bowl images taken from an active restroom facility. According to the definitions provided by the cleaning company, the image dataset labeled with one of the three classes, dirty, average, and clean see Fig. 1. Since urinal bowls not frequently getting dirty, the number of images relevant to dirty and average categories, significantly less than the images in the Clean category. In [22], the authors investigated, the algorithms trained with balanced datasets usually surpass those trained with imbalanced datasets, hence, prepared an initially balanced dataset s_b from the collected data s, and maintained the remaining data as residuals to accommodate in accuracy calculations in training process (explains in subsection II-E). For the testing phase, we acquired a new unbiased dataset from the active restroom facility environment to ensure that there are no overlaps between training and testing data. However, given the limited number of labeled data, the dataset needs to be artificially enlarged to reduce the over-fit in classification phase.

B. *Data augmentation*

Though there are many techniques to generate data, a mere and plausible way could be random crops. Since the dataset consists of rather high-resolution images, the cropped frame needs to incorporate a considerable part of the urinal bowl. Because of that, we included 50% and 75% crops from the original image, which leads to 2016×1512 and 3024×2268 pixel resolutions respectively.

According to the Fig. 1, it can be clearly see that, collected dataset is not uniform, they accommodate a variety of inconsistencies between data due to the variations of camera angles and lighting conditions. For instance, images show various scales of urinal bowl sizes, some images are flipped compared to others and lighting conditions are fairly inconsistent between data.

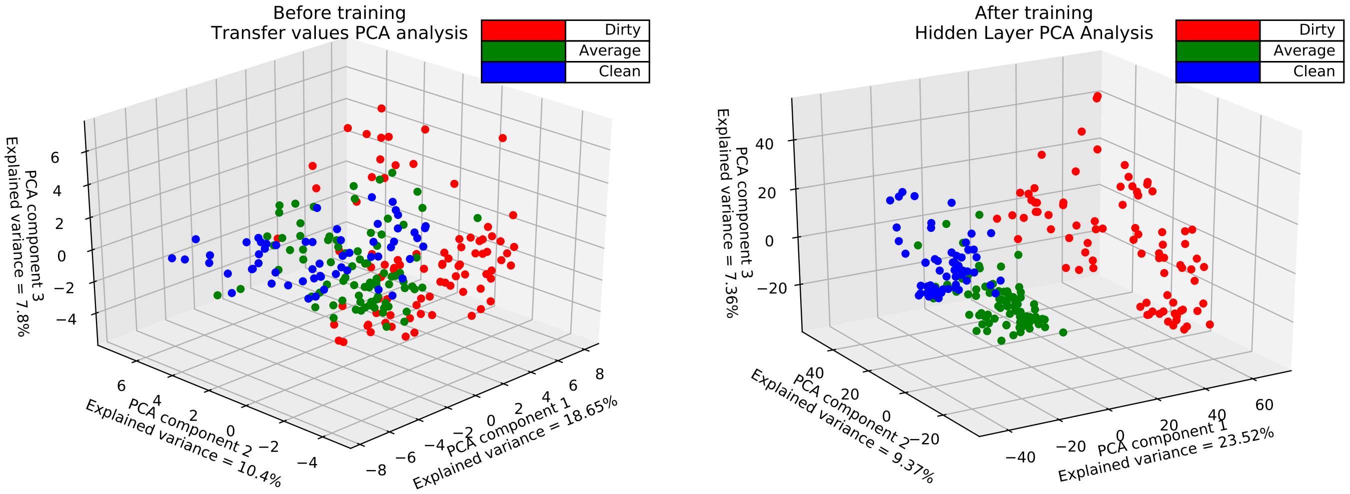
Furthermore, these disparities could influence the decision-making process. Subsequently, the classifier will learn these differences as the features for classification rather than learning the dirtiness feature included in the data, this phenomena is know as over-fit in deep learning. Thus, the solution would be to eliminate these dissimilarities by introducing additional data with the same types of disparities, then the classifier will be further robust to variants in test data. Therefore, we apply random scaling, horizontal flipping and histogram equalizations. See Fig 3. Implementations have been done using the libraries opencv-python-3.4 and TensorFlow TFLearn.

A method of balancing an unbalanced dataset without wasting labeled data, and a solution for the problem of limited sized datasets has been proposed in this sections that can be considered as a contribution in our work. Next subsection is focused on a method of selecting a suitable DCNN model to extract features.

C. *Feature extraction*

In recent history, many deep architectures for image classification came to light, and all these models are proven to give an acceptable range of image classification accuracies. Even though these deep architectures perform well in practice, training such an architecture from scratch requires millions of images and significant computational power. Since our dataset is limited, a suitable pre-trained model needed to be selected in order to fine-tune on training dataset. Then arise a question, how to select a pre-trained model suitable for a given dataset. This section outlines a solution to this problem, which can be considered as another contribution to our work.

Most of the deep architectures designed and trained to tackle ImageNet's ILSVRC2012 dataset [21]. It contains 1.2 million images for 1000 categories, but our categories are not one of them. Therefore, these deep architectures call into question that the capability of understanding the required features of other application specific image datasets like ours. We have



(a) Distribution of extracted features from Inception V3

(b) Distribution of extracted features from proposed classifier

Fig. 4: Analysis of clustering ability using extracted features

addressed this problem by exploiting principal component analysis (PCA) on extracted features from unique models, and different hidden layers of those models. We utilized pre-trained weights of VGG-16, ResNet V1-50 and Inception V3 models using TensorFlow TF-Slim library [18–20].

Using our training dataset as inputs and pre-trained weights as model parameters, we calculated values for several hidden layers, then performs the PCA per layer using calculated layer values and repeat the process for each deep model. In VGG-16, RestNet, and Inception V3 we used latter three dense layers, final flattened layer, and average pooling layer respectively for the PCA calculations. Since PCA can be considered as a clustering method [24], results from a compatible feature extractor (model) need to demonstrate considerable clusters in PCA space. If PCA results are not showing any clusters or PCA results are completely random, then feature extractor has failed in extracting features of the given dataset and there is no use of utilizing it as a feature extractor, because it only adds a randomness which can also achieve by a random weight initializer. After utilizing this procedure, we chosen the Inception V3 as the feature extractor. Fig. 4a depicts the PCA results of feature values (transfer values) extracted from Inception V3 average pooling layer using a part of the training dataset. According to the Fig. 4a, it can be clear that the distribution is not random, there is a visible cluster form from dirty category, but the average and clean categories seem to merge together with some part of the dirty category. Then after selecting the average pooling layer of the Inception V3 model as the transfer layer, we left with the input stream and the eleven inception modules from Inception V3 model Fig. 2, as the feature extractor for the proposed architecture Fig. 3.

However, the typical remedy now is to fine-tune the Inception V3 classifier using training dataset, but the fine-tuning approach did not demonstrate compelling performances on this special dataset. Therefore, we implemented a unique classifier to classify these extracted features that can be found in the next section in details.

D. Proposed Classifier

A three layered MLP-Neural network has been proposed as the feature classifier. The motivation is to transfer the extracted features for a given image, as numerical values to the classifier, in order to classify the image into the accurate class, hence numerical values of extracted features are called *transfer values*. As depicted in Fig. 4a, extracted transfer values are not yet properly clustered, hence, the model requires further feature extraction. Therefore, we used another hidden layer with dropout regularization before the soft-max classifier layer. After training the classifier using extracted features from Inception V3, PCA cluster analysis has been performed on the extracted features from the classifier hidden layer Fig. 4b. According to Fig. 4b, it's clear that the proposed classifier is capable of extracting unique features of training data and distinguish between them.

E. Training and Testing

The training dataset is created using the initially balanced dataset s_b and using the results from data augmentations, see Fig. 3. During the training stage, remaining residual data and 5% of total training data are selected before every epoch and kept away in order to calculate the accuracy after completing one full epoch. The focus of this arrangement is to enforce the model, not to learn noisy features like backgrounds, translations, etc. We used polynomial decaying learning rate for our experiment starting from 0.1 learning rate to 0.0001 learning rate within 100000. After model started to show compelling results, we stopped the training process and used test data Y to evaluate proposed architecture.

III. Results and Discussion

Our experiment setups are implemented by, using TensorFlow framework [25] and python 3.5 languages on an HP Z840 workstation with NVIDIA GeForce GTX 1080Ti graphics card. The size of the collected dataset $\boldsymbol{s}$ is $N = 323$ and the size of the initially balanced dataset $\boldsymbol{s_b}$ is $M = 273$. Even though an initial dataset is created, deep learning techniques could not be applied, because the dataset is significantly very small. However, proposed method introduced data augmentation techniques, therefore, we managed to increase the dataset to a considerable size according to

$$Dataset_i = Dataset_{i-1} + DataAugmentation(Dataset_{i-1}) \tag{1}$$

where $i \in \{1, \ldots, 4\}$, when $i = 0$ $Dataset_0$ represent the initial balanced dataset $\boldsymbol{s_b}$ and $DataAugmentation$ enlarge a given dataset by five times. Therefore, we iterate Eq.(1) to generate data and train the network until we achieve a compelling result.

The results of finding the most competent model for transfer-learning, using the proposed method describe in subsection II-C, is outlined in Fig.4. It's apparent in Fig. 4a, features relevant to dirty and clean classes in PCA space have formed some partial clusters, but features relevant to average class seems to mix with other clusters. Since there are partial clusters formed, the MLP-neural network classifier able to classify the data with an acceptable accuracy. Table I shows the classification performances of our proposed classifier. Confusion matrix (CM) calculated as

$$CM(i,j) = \frac{1}{|Y_{class_i}|} \sum\nolimits_{y \in Y_{class_i}} P(class(\hat{y}) = class_j | y) \tag{2}$$

where $CM(i,j)$ denotes the prediction of j^{th} class when an i^{th} class element given, it represents as i^{th},j^{th} element of Table I, y represent an element of i^{th} class and $class(\hat{y})$ denotes the predicted class label. For some data, the classifier yields accurate class, but with a low probability, therefore even the results are accurate, the decision uncertainty is high. Evaluation performances using Eq. 2 will reflect those uncertainties in the results. According to table I, it's clear that the average category has the lowest prediction accuracy. This result can be validated using Fig. 4b, the average category seems to spread into the clean category cluster as well as the dirty category boundary.

It is clear that our proposed architecture able to classify cleanliness in restrooms and above results verify our methods using real-life data. Since our model used only images of restrooms, the proposed architecture can be easily tailored to perform in different restroom facility after iterating very few amount of fine-tunes.

	Dirty	Average	Clean
Dirty	.99	.00	.08
Average	.02	.83	.15
Clean	.00	.04	.95

TABLE I: Confusion matrix for the classifier results

IV. Conclusion

In summary, this paper has proposed a novel solution for real-time cleanliness classification in active restrooms facilities using deep learning architectures. We have devised methodologies for recognizing a qualified model for feature extraction, selecting a transfer layer for the transfer-learning and altogether a competent classifier along with a training procedure capable of classifying an application specific dataset up to an acceptable accuracy. Since there is an only limited number of studies which have addressed this issue of cleanliness in restrooms, the application of deep learning itself can be considered as a contribution with others mentioned above. Our study limitations highlight the difficulty of collecting data, sometimes cameras have blind spots and image quality will depend on lighting conditions and camera quality. Despite the limitations of this method, it demonstrated acceptable results in an active restroom environment. The future experiments would involve implementing a sophisticated data augmentation method and establish trained model in a different restroom facility.

References

[1] Restroom Association, "Toilet survey study 2011." [online]. Available: https://www.toilet.org.sg/docs/npsurvey.pdf, 2011. [Accessed on 19 February 2018].

[2] "Poor hygiene and disease." [online]. Available: http://www.health.gov.au/internet/publications/publishing.nsf/Content/ohp-enhealth-manual-atsi-cnt-l ohp-enhealth-manual-atsi-cnt-l-ch3 ohp-enhealth-manual-atsi-cnt-l-ch3.2, 2010. [Accessed on 19 February 2018].

[3] A. Das, "Infections that spread by using dirty toilet." [online]. Available: https://www.boldsky.com/health/wellness/2015/infections-that-spread-by-using-dirty-toilets-075057.html, 2015. [Accessed on 19 February 2018].

[4] S. Yi, "Dirty eatery toilets a no-go for diners." [online]. Available: http://www.straitstimes.com/singapore/dirty-eatery-toilets-a-no-go-for-diners, 2015. [Accessed on 19 February 2018].

[5] "China toilet revolution targets dirty lavatories." [online]. Available: http://www.todayonline.com/chinaindia/china/chinas-toilet-revolution-targets-dirty-lavatories, 2017. [Accessed on 19 February 2018].

[6] "Happy toilet programme." [online]. Available: http://www.nea.gov.sg/public-health/public-toilets. [Accessed on 19 February 2018].

[7] S. A. Assomull, "Loo awards show cleaners efforts cannot go to waste with low wages and little respect." [online]. Available: http://www.straitstimes.com/singapore/loo-awards-show-cleaners-efforts-cannot-go-to-waste-with-low-wages-and-little-respect, 2017. [Accessed on 19 February 2018].

[8] "Restroom visitlizer system." [online]. Available: http://smarttechnologies.sg/. [Accessed on 19 February 2018].

[9] A. Ten, "New gadget sniffs out when toilets need cleaning." [online]. Available: http://www.straitstimes.com/singapore/environment/new-gadget-sniffs-out-when-toilets-need-cleaning, 2016. [Accessed on 19 February 2018].

[10] "Cleaner, smarter toilets for your employees and visitors." [online]. Available: https://www.ncs.com.sg/documents/20184/73669/NCS+Smart+Toilet+Analytics.pdf/ab34c9ab-5e0f-4fe5-b794-fcd6546e4f24. [Accessed on 19 February 2018].

[11] SmartClean, "Virtual cleaning supervisor." [online]. Available: http://www.smartclean.sg/. [Accessed on 19 February 2018].

[12] Y. Bengio *et al.*, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[13] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," *arXiv preprint arXiv:1802.00752*, 2018.

[14] T. N. Sainath, B. Kingsbury, B. Ramabhadran, P. Fousek, P. Novak, and A.-r. Mohamed, "Making deep belief networks effective for large vocabulary continuous speech recognition," in *Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on*, pp. 30–35, IEEE, 2011.

[15] R. B. Girshick, "Fast R-CNN," *CoRR*, vol. abs/1504.08083, 2015.

[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*, pp. 21–37, Springer, 2016.

[17] Y. LeCun, L. Jackel, L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, P. Simard, *et al.*, "Learning algorithms for classification: A comparison on handwritten digit recognition," *Neural networks: the statistical mechanics perspective*, vol. 261, p. 276, 1995.

[18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826, 2016.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[21] O. Russakovsky and et al, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[22] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on knowledge and data engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[23] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[24] S. Raychaudhuri, J. M. Stuart, and R. B. Altman, "Principal components analysis to summarize microarray experiments: application to sporulation time series," in *Biocomputing 2000*, pp. 455–466, World Scientific, 1999.

[25] M. Abadi and et al, "TensorFlow: Large-scale machine learning on heterogeneous systems." Software available from tensorflow.org.