Combining Deep Neural Network with Traditional Classifier to Recognize Facial Expressions

Zixiang Fei[1], Erfu Yang[*1], David Li[2], Stephen Butler[3], Winifred Ijomah[1], Huiyu Zhou[4]

1 Department of Design, Manufacture and Engineering Management
University of Strathclyde, Glasgow G1 1XJ, UK
{zixiang.fei, erfu.yang, w.l.ijomah}@strath.ac.uk
2 Strathclyde Institute of Pharmacy & Biomedical Sciences
University of Strathclyde, Glasgow G4 0RE, UK
david.li@strath.ac.uk
3 School of Psychological Sciences and Health
University of Strathclyde, Glasgow G1 1QE, UK
stephen.butler@strath.ac.uk
4 Department of Informatics
University of Leicester, LE1 7RH, Leicester
hz143@leicester.ac.uk
* Correspondence: erfu.yang@strath.ac.uk; Tel.: +44-141-574-5279

Abstract— Facial expressions are important in people's daily communications. Recognising facial expressions also has many important applications in the areas such as healthcare and e-learning. Existing facial expression recognition systems have problems such as background interference. Furthermore, systems using traditional approaches like SVM (Support Vector Machine) have weakness in dealing with unseen images. Systems using deep neural network have problems such as requirement for GPU, longer training time and requirement for large memory. To overcome the shortcomings of pure deep neural network and traditional facial recognition approaches, this paper presents a new facial expression recognition approach which has image pre-processing techniques to remove unnecessary background information and combines deep neural network ResNet50 and a traditional classifier-- the multiclass model for Support Vector Machine to recognise facial expressions. The proposed approach has better recognition accuracy than traditional approaches like Support Vector Machine and doesn't need GPU. We have compared 3 proposed frameworks with a traditional SVM approach against the Karolinska Directed Emotional Faces (KDEF) Database, the Japanese Female Facial Expression (JAFFE) Database and the extended Cohn-Kanade dataset (CK+), respectively. The experiment results show that the features extracted from the layer 49Relu have the best performance for these three datasets.

Keywords- Facial Expression Recognition; Deep Convolution Network; Support Vector Machine

I. INTRODUCTION

Facial expressions play an important role in people's daily communication and interaction [1], [2]. Machine vision and machine learning as hot topics [3], [4], [5], [6] are widely used to develop facial expression recognition systems for many important applications such as healthcare and e-learning system, etc.

In a typical facial expression recognition system, there are mainly three parts: face detection, facial feature extraction and facial feature classification. The first step is face localization which helps removing non-facial information and may improve the recognition accuracy as well[7]. Next, the facial feature extraction part is used to extract the features related to the facial expressions which is a quite important part as it will determine the performance of facial expression recognition. Finally, there is a facial feature classification part to recognise the facial expression. In the proposed work, we have the image pre-processing part to localise the position of the face and remove unnecessary information. Also, we use the ResNet50 a deep residual neural network to extract facial features and use Support Vector Machine (SVM) for facial expressions classification. Meanwhile, the facial expression recognition system can be divided into two groups: using static images and using dynamic image sequences.

In general, facial expression recognition system can also be divided into two types: using traditional machine learning approaches and using deep learning approaches like deep neural networks AlexNet and ResNet. To begin with, the machine learning approaches are such as SVM, Linear Discriminant Analysis (LDA). Many researchers have used machine learning based approaches for facial expression recognition. For instance, Makhmudkhujaev et al. proposed a novel descriptor called Histograms of Prominent Edge Directions (HPED) to recognize facial expressions [8]. This descriptor can avoid sampling error and needs fewer samples by using less code-bins to describe the spatial regions. In a relevant work, Anh et al. proposed a facial expression recognition system using salient facial regions, which improved facial expression recognition accuracy [9]. They used Pyramid of Local Phase Quantization descriptor (PLPQ) to extract facial

features and the SVM was used in facial features classification. However, the disadvantages of traditional machine learning based approaches have been noted by researchers. For example, these approaches don't have good performance for unseen images.

On the other hand, deep learning based approaches such as AlexNet are quite popular in facial expression recognition [10]–[12]. For instance, Ziyang et al. proposed a facial expression recognition system which used facial alignment technology to adjust inclined faces and used a sequential Convolution Neural Network (CNN) to extract facial features [12]. As huge amount of memory may be used to process the whole images, facial landmark detection was used to extract facial features and CNN was used as classifier [10]. However, these deep learning based systems have problems such as requirement for GPU, longer training time and requirement for large memory. On the other hand, the proposed method uses deep neural network to extract deep features and uses SVM to recognise facial expressions, which only require a single pass through the data. As a result, the GPU is not essential [13]. In spite of the higher computing requirement, the deep neural network based approaches have many good applications in machine vision [14], [15]. There are some good survey papers on facial expression recognition [16]–[20]. For example, Byoung et al. reviewed many aspects in facial expression recognition systems such as some famous facial expression datasets and performance evaluation for the facial expression recognition [21], etc.

This paper proposes a new facial expression recognition approach which combines deep neural network ResNet50 and a traditional classifier -- the multiclass model for Support Vector Machine. The proposed system has better recognition accuracy than traditional approaches like Support Vector Machine and doesn't need GPU. The approach has been tested on the three benchmarking facial expression datasets , i.e., the Karolinska Directed Emotional Faces (KDEF) Database [22], the Japanese Female Facial Expression (JAFFE) Database [23] and the extended Cohn-Kanade dataset (CK+) [24], [25]. This paper is organized as follows. Section 1 provides a general overview of the paper. Next, the proposed deep learning-based approach is detailed in Section 2. Section 3 presents the experiment results. The discussion is given in Section 4. Section 5 provides the conclusion.
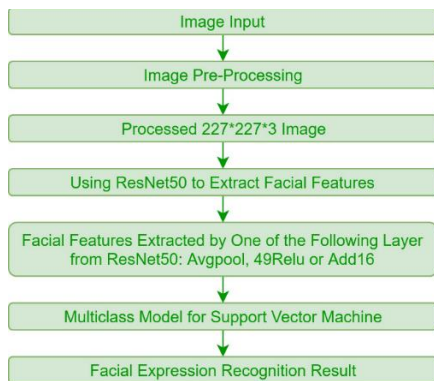

Figure 1Flowchart of the Proposed Approach

II. PROPOSED APPROACH

A. Overview of the Approach

The flowchart of the proposed approach is shown in Fig. 1. It includes several parts such as: image input, image pre-processing, facial expression analysis and system output. To begin with, the images of facial expressions are taken as the input. Next, the image pre-processing techniques are applied to the input images. In this work, the Viola-Jones algorithm is used[26], [27] to locate the position of the face and crop the face part from the images. By applying this image pre-processing technique, the irrelevant background image part is removed and the interference from the background is reduced. As the input of the ResNet50 need to be 224*224*3, the cropped image of the face will be resized. Then, the proposed facial expression approach which combines deep neural network ResNet50 and a traditional classifier-the multiclass model for Support Vector Machine is used to recognise the facial expression of images. Also, the features are extracted from the following layers of the ResNet50: avg-pool (avgpool), activation_49_relu (49Relu) and add_16 (add16). Finally, the images with the predicted facial expression labels are the output of the system.

B. Resnet50

A residual neural network (ResNet) is an artificial neural network [28]. The ResNet uses the skip connection to jump over some layers. The advantage of skipping over layers is to avoid the problem from vanishing gradients and to simplify the network in the training stages. In the experiment, we use the ResNet50 which contains 50 layers and the system runs on Matlab. The ResNet is originally trained by more than one million images from the ImageNet database and it can classify images into 1000 categories [29]. Also, the input size of the images is 224 by 224.

C. Deep feature extraction and classification

In the experiment, we use the ResNet50 as the deep feature extractor to extract the deep features from the images of facial expression. We will then use these deep features to train a traditional classifier, namely the Support Vector Machine. The advantages of this combination are such as improvement in speed and low requirement to experimental equipment. For instance, as feature extraction only requires a single pass through the data, we don't need a GPU to do this experiment [30]. Furthermore, in order to use the ResNet50 to do this work, a layer needs to be selected to extract the deep features. To achieve the best performance, we have done experiments to explore the best layer to extract the deep features. In the experiment, we compare the experiment performance for the following layer to extract the features: avg-pool (avgpool), activation_49_relu (49Relu) and add_16 (add16). Finally, we use a traditional classifier to train this deep features. As the SVM has good performance in object classification and face detection applications, we use the SVM to train these deep features in the experiment [31].

III. EXPERIMENT RESULT

A. JAFFE Dataset

In the experiment, the first facial expression dataset we used was the JAFFE dataset [23]. We selected 202 images

from this dataset and all the images were applied with image pre-processing techniques. This dataset contained 7 kinds of facial expressions: angry, happy, neutral, surprise, sad, afraid, disgust. In the experiment, 80% of the images were selected as the training images randomly and the rest of the images worked as the testing images. Table 1 compares the recognition accuracy for the facial expressions using the JAFFE Dataset with different methods. The experiments were repeated 5 times. The first column in the table shows the method used for facial expressions recognition. The second column shows the minimum recognition accuracy after 5 iterations, whilst the third column gives the average recognition accuracy over 5 iterations. Finally, the fourth column shows the maximum recognition accuracy after 5 iterations.

TABLE 1. COMPARISON OF RECOGNITION ACCURACY OF DIFFERENT METHODS FOR 5 ITERATIONS ON THE JAFFE DATASET

Method	MIN	Average	MAX
SVM	69.2%	74.4%	79.5%
RES + Avgpool + SVM	71.8%	79.0%	84.6%
RES + 49Relu + SVM	76.9%	83.1%	87.2%
RES + Add16 + SVM	74.4%	80.0%	84.6%

The result for different methods for the JAFFE Dataset are also shown in Fig. 2. As the figure shows, the features extracted from the layer 49Relu reaches the highest average recognition accuracy 83.1%. As the number of the images in the JAFFE Database is quite small, the difference between the minimum and maximum recognition accuracy for each method are quite large.
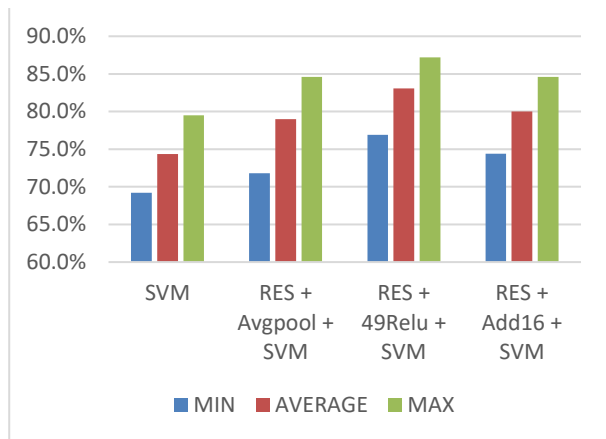

Figure 2. Comparison of Recognition Accuracy of Different Methods for 5 iterations on the JAFFE Dataset

B. KDEF Dataset and CK+ Dataset

In the experiment, the second facial expression dataset we used was the KDEF dataset [22]. We selected 980 front-view images from this dataset and all the images was applied with image pre-processing techniques. This dataset contained 7 kinds of facial expressions: angry, happy, neutral, surprise, sad, afraid, disgust. In the experiment, 80% of the images were selected as the training images randomly and the rest of the images worked as the testing images. Table 2 compares the recognition accuracy for facial expressions from the KDEF Dataset using each different methods. The first column in the table shows the method used for facial expressions recognition. The second column shows the recognition accuracy.

TABLE 2. COMPARISON OF RECOGNITION ACCURACY OF DIFFERENT METHODS ON THE KDEF DATASET

Method	KDEF
SVM	69.2%
RES + Avgpool + SVM	76.8%
RES + 49Relu + SVM	83.9%
RES + Add16 + SVM	82.3%

As the table 2 shows, the features extracted from the layer 49Relu reaches the highest average recognition accuracy 83.9%. It was noticed that the traditional classifiers SVM did not have such a good performance, for such a large dataset. In the experiment, the third facial expression dataset we used was the CK+ dataset [24], [25]. We selected 693 images from this dataset and all the images was applied with image pre-processing techniques. This dataset contained 7 kinds of facial expressions: angry, happy, neutral, surprise, sad, afraid, disgust. In the experiment, 80% of the images were selected as the training images randomly and the rest of the images worked as the testing images. Table 3 compares the recognition accuracy for facial expressions from the CK+ Dataset using each different methods. The first column in the table shows the method used for facial expressions recognition. The second column shows the recognition accuracy.

TABLE 3. COMPARISON OF RECOGNITION ACCURACY OF DIFFERENT METHODS ON THE CK+ DATASET

Method	CK+
SVM	87.7%
RES + Avgpool + SVM	92.8%
RES + 49Relu + SVM	93.5%
RES + Add16 + SVM	93.5%

As the table 3 shows, the features extracted from the layer 49Relu and the layer Add16 reaches the highest recognition accuracy 93.5%. It was noticed that the traditional classifiers SVM had a lower recognition accuracy.

IV. DISCUSSION

Facial expression is an important way to show emotion and attitude in people's daily communication. There are many important applications for facial expression recognition. However, existing algorithms may have some problems such as interference from the background. The proposed system have the image pre-processing part to crop the head part from the image and remove the irrelevant background image in order that the interference from the background is reduced.

TABLE 4. COMPARISON OF RECOGNITION ACCURACY OF DIFFERENT METHODS ON THE JAFFE, KDEF AND CK+ DATASET

Method	JAF	KDEF	CK+
SVM	74.4%	69.2%	87.7%
RES + Avgpool + SVM	79.0%	76.8%	92.8%
RES + 49Relu + SVM	83.1%	83.9%	93.5%
RES + Add16 + SVM	80.0%	82.3%	93.5%

In the research, we did experiments of facial expression recognition using traditional classifier SVM and three kinds of combinations of the ResNet and SVM on JAFFE, KDEF and CK+ facial expression datasets. We compared performance for the features extracted from 3 different layers from the ResNet: layer Avgpool, layer 49Relu and

layer Add16. The general performance can be shown in Table 3 which shows average recognition accuracy for each method for the 3 datasets. The first column in the table shows the 4 different methods used for facial expressions recognition. The columns from the second to the fourth show the recognition accuracy for each method for JAFFE, KDEF and CK+ dataset respectively. The comparison result for each method on these datasets are also shown in Fig. 3, which shows that the method using the features extracted from the layer 49Relu have the best performance on these datasets.
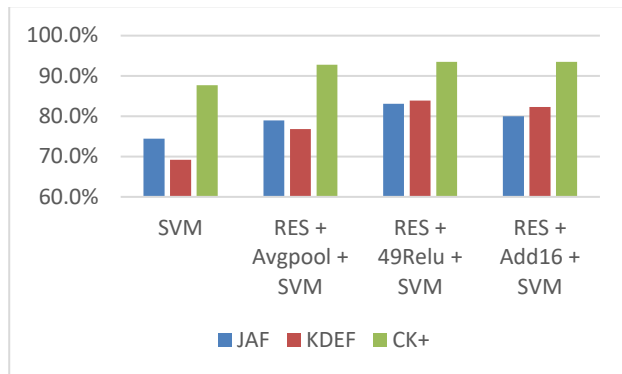
Figure 3. Comparison of Recognition Accuracy of Different Methods on the JAFFE, KDEF and CK+ Dataset

Furthermore, we have found out the recognition performance for each emotion. Table 5 shows the average recognition accuracy for each emotion for the three datasets. The first column in the table shows the 7 different facial expressions. The columns from the second to the fourth show the average recognition accuracy for each emotion using 4 methods on the three datasets respectively. It is noted that all the methods don't have a good facial expressions recognition performance for the emotion of sad in both three datasets. On the other hand, the emotion with the highest recognition accuracy is different for the three datasets.

TABLE 5. AVERAGE RECOGNITION ACCURACY FOR EACH EMOTION FOR THE JAFFE, KDEF AND CK+ DATASET

Expression	JAF (%)	KDEF (%)	CK+(%)
Angry	87.5	76.8	90.0
Disgust	75.0	79.5	95.7
Fear	66.7	61.6	90.6
Happy	83.3	94.6	91.4
Neutral	95.8	87.5	95.5
Sad	80.0	79.5	62.5
Surprise	80.0	79.5	95.0

Figure 4. Confusion Matrix for SVM on the JAFFE and KDEF

Fig. 4 and Fig.5 show the confusion matrixes for the SVM and the proposed approach that uses the features extracted from the layer 49Relu for the JAFFE and KDEF Datasets. It shows that the SVM only has a good performance in the neutral and happy emotion on the KDEF dataset. Also, it is noted that the proposed approach has good performance for all the expressions for the JAFFE dataset. However, the proposed approach does not have a good performance for the expression of fear for the KDEF dataset. It is noticed that this algorithm may recognize the fear emotion as sad wrongly.

V. CONCLUSION

This paper proposed a facial expression recognition approach which combines a deep neural network ResNet50 and a traditional classifier - the multiclass model for Support Vector Machine. The approach has been tested on three benchmarking facial expression datasets including the Karolinska Directed Emotional Faces (KDEF) Database, the Japanese Female Facial Expression (JAFFE) Database and the extended Cohn-Kanade dataset (CK+).

The experiments have discovered that the proposed approach which extracted the features from the layer 49Relu have the best performance on KDEF, JAFFE and

CK+ datasets. In addition, the proposed approach may solve problems such as interference from background, requirement for GPU, longer training time and requirement for large memory.

Figure 5. Confusion Matrix for Res + 49Relu + SVM on the JAFFE and KDEF Dataset

For future work, we will have a further research to modify the network such as the convolution layers and pooling layers in order that the recognition performance for the expression of fear can be improved. In addition, because of some equipment limitation, we didn't test the algorithm using a database with large amounts of images and we didn't do the experiment using ResNet. We will do these in the future work.

REFERENCE

[1] Z. Kasiran and S. Yahya, "Facial Expression as an Implicit Customers' Feedback and the Challenges," in Computer Graphics, Imaging and Visualisation (CGIV 2007), 2007, pp. 377–381.

[2] S. J. Wilson, L. M. Martire, F. J. Keefe, J. A. Mogle, M. A. P. Stephens, and R. Schulz, "Daily verbal and nonverbal expression of osteoarthritis pain and spouse responses," Pain, vol. 154, no. 10, pp. 2045–2053, Oct. 2013.

[3] F. Gao, F. Ma, J. Wang, J. Sun, E. Yang, and H. Zhou, "Visual Saliency Modeling for River Detection in High-Resolution SAR Imagery," IEEE Access, vol. 6, pp. 1000–1014, 2017.

[4] F. Gao et al., "Biologically Inspired Progressive Enhancement Target Detection from Heavy Cluttered SAR Images," Cognit. Comput., vol. 8, no. 5, pp. 955–966, 2016.

[5] T. U. Rehman, M. S. Mahmud, Y. K. Chang, J. Jin, and J. Shin, "Current and future applications of statistical machine learning algorithms for agricultural machine vision systems," Comput. Electron. Agric., vol. 156, pp. 585–605, Jan. 2019.

[6] Y. Lu, T. Q. Chen, J. Chen, J. Zhang, and A. Tisler, "<title>Machine vision systems using machine learning for industrial product inspection</title>," 2002, pp. 161–170.

[7] A. Kartali, M. Roglic, M. Barjaktarovic, M. Duric-Jovicic, and M. M. Jankovic, "Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches," in 2018 14th Symposium on Neural Networks and Applications (NEUREL), 2018, pp. 1–4.

[8] "Person-Independent Facial Expression Recognition with Histograms of Prominent Edge Directions," KSII Trans. Internet Inf. Syst., vol. 12, no. 12, Dec. 2018.

[9] A. Vo and B. T. Nguyen, "Facial Expression Recognition Based on Salient Regions," in 2018 4th International Conference on Green Technology and Sustainable Development (GTSD), 2018, pp. 739–743.

[10] N. P. Gopalan, S. Bellamkonda, and V. Saran Chaitanya, "Facial Expression Recognition Using Geometric Landmark Points and Convolutional Neural Networks," in 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 2018, pp. 1149–1153.

[11] M. Ma, Z. Gao, J. Wu, Y. Chen, and X. Zheng, "A Smile Detection Method Based on Improved LeNet-5 and Support Vector Machine," in 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2018, pp. 446–451.

[12] Z. Han, H. Huang, and J. Wang, "Convolutional Neural Network Based Expression Classification with Face Alignment," in 2018 5th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS), 2018, pp. 408–412.

[13] "Extract Image Features Using Pretrained Network - MATLAB & Simulink - MathWorks United Kingdom." [Online]. Available: https://uk.mathworks.com/help/deeplearning/examples/extract-image-features-using-pretrained-network.html. [Accessed: 16-Jul-2019].

[14] Z. Yue et al., "A Novel Semi-Supervised Convolutional Neural Network Method for Synthetic Aperture Radar Image Recognition," Cognit. Comput., Mar. 2019.

[15] F. Gao, T. Huang, J. Sun, J. Wang, A. Hussain, and E. Yang, "A New Algorithm for SAR Image Target Recognition Based on an Improved Deep Convolutional Neural Network," Cognit. Comput., Jun. 2018.

[16] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," Pattern Recognit., vol. 25, no. 1, 1992.

[17] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," Pattern Recognit., vol. 36, no. 1, 2003.

[18] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3D facial expression recognition: A comprehensive survey," Image Vis. Comput., vol. 30, no. 10, 2012.

[19] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, 2009.

[20] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 6, 2015.

[21] B. Ko, Ko, and B. Chul, "A Brief Review of Facial Emotion Recognition Based on Visual Information," Sensors, vol. 18, no. 2, p. 401, Jan. 2018.

[22] D. Lundqvist, A. Flykt, and A. Ohman, "The Karolinska directed emotional faces (KDEF)," CD ROM from Dep. Clin. Neurosci. Psychol. Sect. Karolinska Institutet, pp. 91–630, 1998.

[23] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in Proceedings - 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998, 1998, pp. 200–205.

[24] T. Kanade, J. F. Cohn, and Yingli Tian, "Comprehensive database for facial expression analysis," in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), pp. 46–53.

[25] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2010, pp. 94–101.

[26] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1, p. I-511-I-518.

[27] "Detect objects using the Viola-Jones algorithm - MATLAB - MathWorks United Kingdom." [Online]. Available: https://uk.mathworks.com/help/vision/ref/vision.cascadeob jectdetector-system-object.html. [Accessed: 29-Jan-2019].

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015.

[29] "Pretrained ResNet-50 convolutional neural network - MATLAB resnet50 - MathWorks United Kingdom." [Online]. Available: https://uk.mathworks.com/help/deeplearning/ref/resnet50.h tml. [Accessed: 26-Apr-2019].

[30] "Feature Extraction Using AlexNet - MATLAB & Simulink - MathWorks United Kingdom." [Online]. Available: https://uk.mathworks.com/help/deeplearning/examples/feat ure-extraction-using-alexnet.html;jsessionid=a9dd0dd508fd2b96854cd6f11d5c. [Accessed: 23-Jan-2019].

[31] P. McAllister, H. Zheng, R. Bond, and A. Moorhead, "Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets," Comput. Biol. Med., vol. 95, pp. 217–233, Apr. 2018.