Enhancement Techniques for Improving Facial Recognition Performance in Convolutional Neural Networks

Mhd Saeed Sharif
Computer Science and DT, ACE, UEL
University Way, London, UK
S.Sharif@uel.ac.uk

Matthew Olusegun AFOLABI
Computer Science and DT, ACE, UEL
University Way, London, UK
u2161761@uel.ac.uk

Abin Zorto
Computer Science and DT, ACE, UEL
University Way, London, UK
u2091940@uel.ac.uk

Wael Elmedany
College of Information Technology
University of Bahrain, Kingdom of Bahrain
welmedany@uob.edu.bh

***Abstract*—The advent of convolutional neural networks (CNNs) to the development of face recognition system has been a game changer in the field of computer vision and pattern recognition. This research work uses a pre trained MobileNet-V1 model to develop an effective CNN model capable of high performance. We also tackle several common facial recognition challenges which include occlusions, illumination variations, make-ups, pose variation and ageing through the use of several improvement techniques. The techniques include adopting a less computationally costly approach, transfer learning and hyper-parameter fine-tuning. The Top-1 accuracy 70.6% and Top-5 accuracy 89.5% of the base MobileNet-V1 model has been improved using these techniques to achieve training accuracy of 95% and accuracies of 96.4%, 98.0% and 99.1% on the Pins face recognition data-set, FaceScrub data-set and LFW data-set, respectively. The work done so far illustrates the need for further research into improvement techniques for convolutional neural networks.**

***Index Terms*—Receptive Field, Occlusion, Pose variation, Face detection, Face Landmark, Loss Function**

I. Introduction

This research study focuses on how to improve convolutional neural network models for face recognition. Face recognition is a popular security biometric because facial traits are distinctive and easily collected [1]. Due to its non-contact nature and easy installation, it is commonly used for identity verification and is a prominent biometric technology used in healthcare, banking/financial services, law enforcement, travel/tourism, education, and security, among other sectors.

Martinez [2] defines face recognition as the study of how computers replicate biological systems that recognise faces using visual sensors such as eyes. Faces are matched with name/details in a database. Face verification is 1:1, while face identification is 1:N (where N is the number of faces).

Researchers have made various attempts to make facial recognition as good as or better than humans. In "The Tichborne Claimant" case from 1871, two photos were compared to identify a person. No face recognition system existed to quickly resolve the case [3]. A century later, American researchers attempted semi-automatic face recognition [4]. O'Toole *et al*. [5] and Tang and Wang [6] claimed computer algorithms outperformed humans when it came to images and doodles. While initially there was some skepticism amongst researchers on the algorithms performance and whether facial recognition technologies increased or decreased security the use of algorithms have now become standard practice.

Face identification in computer vision and pattern recognition has undergone various eras of innovation, from the classical technique to recent methods driven by Convolutional neural networks (CNNs). Face recognition research has advanced, but there are opportunities for improvement. for example, a face recognition system, with the purpose of identifying/capturing offenders must be perfect, or the wrong person could be charged. CNN-based models have improved facial recognition systems, yet there are still limitations and challenges with CNNs. These include: high computational cost, facial expression variations, use of make-ups, spoofing problems, cross-age, occlusion problem, wrinkles/ageing, low resolution, variation in image poses, noisy-labelled data-set, illumination variations, masked face (as seen in the pandemic), loss function and choice of activation function problems [7]. Face recognition systems deployed in controlled environments such as financial services, train stations, and workplaces may have some of the identified problems minimised, since their subjects are usually cooperative. However, in many uncontrolled conditions where face recognition systems are vital, the subjects may not be as cooperative.

II. Literature Review

CNNs' effectiveness in image identification helped popularise deep learning [8]. CNN's research led to breakthroughs in computer vision and pattern recognition.

A. From Traditional to Modern Facial Recognition

In 1964, American researchers imagined a semi-automatic approach and had computer operators enter 20 measures such as mouth and eye size. Triguero *et al*. [9] cites Kelly and Kanade's 1970s PhD theses as the first formal face recognition research. Traditional algorithms include geometry-based, holistic, and local appearance features [10]. Despite issues such as high training data volume and extended training time, modern algorithms include artificial neural network-based or deep learning-based approaches.

The position of a face in an image is determined and the corresponding coordinates of a bounding box for such an image is returned. Face alignment uses a reference point at a fixed position in an image to scale and crop the face image. Important features that distinguish individual faces are learned and the key features extracted are later used for face matching. Features extracted are used either as a 1:1 matching as in the case of face verification or as 1:N matching for identification against faces in the database [10].

B. CNNs-Based Face Recognition Models

One of the earliest successful CNN based face recognition systems was presented in 1997. While it performed relatively well it was not immune to challenges posed by low-resolution, spoofing, variation in pose, make-up issue and cross-age problem [13]. In 2018 Liu *et al*. [14] proposed SphereFace with the use of angular Softmax loss. That was also called deep hypersphere embedded approach; and it achieved an accuracy of 99.42% on the LFW face database. A study by Deng *et al*. [15] proposed the ArcFace by using an additive angular margin loss function to improve the discriminate ability of embedded feature learned when training CNNs. It was claimed that, it outperformed the state of the art face recognition algorithm and had an accuracy of 99.83% on the most used face database (labelled faces in the wild). Hassan and Abdulazeez [16] lauded the achievement pointing at the substantial improvements regarding response to occlusion, illumination challenges, variations in pose and expression, cost due to high level GPU usage and network depth as persistent problem.

C. Applications of Face Recognition Models

Face recognition is being used in many ways in travel, hospitality, and tourism and border control. Some transport companies have employed face recognition system to verify drivers for the passengers' peace of mind [17]. Also, it is now being implemented in hotels to reduce check-in duration and so that customers do not need to lock themselves out of their rooms [18]. The importance of face recognition for access control purposes has been discussed. It can serve to prevent an unauthorised individual from entering sensitive or restricted areas such as control rooms, laboratories, bank vaults, and lecture halls [19]. The Chinese Government has used face recognition techniques with monitoring equipment to slow the spread of covid-19 by identifying contacted suspects with a view to controlling their movements [20].

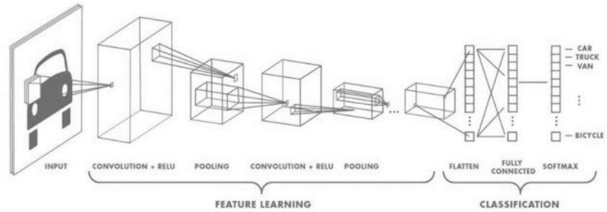

Fig. 1. The Basic Architecture of Convolutional Neural Networks [11].

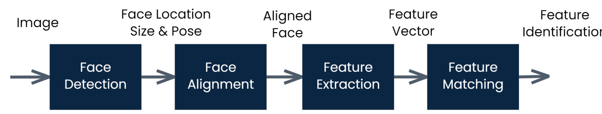

Fig. 2. Typical Face Recognition Processes [10].

III. Methodology

A. Convolutional Neural Network

1) The Convolutional Neural Networks Architecture: CNNs are feed-forward neural networks that consist of input layers, output layers, and multiple hidden layers. They have been successful in image classification and pattern recognition especially face recognition tasks, among others. Fig. 1 illustrates the CNNs architecture as identified by Kulkarni and Shivananda [11]. According to Imaoka *et al*. [12], The face recognition process flow consists of face detection, face alignment, feature extraction, and feature matching.

2) Improvement Methods: Transfer learning and hyper-parameter fine tuning are seen as effective ways of improving machine learning models. Transfer learning involves training a pre-trained network on a different data-set it was not trained on before. Fine-tuning is to tweak or work on the parameters of a model of interest with a view to get the best result out of it. In addition, injecting noise during training of an algorithm in machine learning has been viewed as means of data augmentation as well as regularization that improves the generalization of algorithms.

B. Data-set Description

Pins face recognition data-set was used for used for training, validation and testing. The data-set comprises of 17,534 images of 105 celebrities with their identities revealed. Transfer learning and fine-tuning of few pre-trained models (ResNet-50, Inception V3 and Xception) that have been trained on thousands of images (ImageNet) was experimented.

The Pins face data-set [21] used for this research was heavily imbalanced with the least number of face images having 86 images while some have as high as 200. The researcher manually reduced the number of images in classes (celebrities) with more than 86 to equal the minimum class starting with those images that were like duplicates, and few images cropped with a little part of another person. This is not advisable in a real world scenario as it would have amount

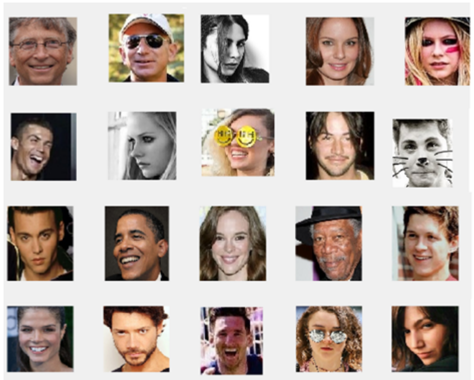

Fig. 3. Random Samples of Images from Pins Face Data-set [21].

to high loss of data, but it was done to reduce the high computational demand on the researcher.

C. Pre Processing/Augmentation

The balanced data-set was divided into training, validation and test sets in proportion of 75%, 20% and 5%. Images were reshaped into 224, 224, 3 corresponding to the image's height, width and channels respectively. Geometric transformation technique was adopted by randomly flipping training images horizontally and by rotating them slightly to present to the CNN model as if they are different leveraging on the invariance property that the CNN models exhibit.

D. Model Design

Two CNN models' architecture for transfer learning were proposed for a start but only one made it to the end of the research, see section V on critical analysis of the results and discussion for more on this. Xception, MobileNet-V1, ResNet50 and InceptionV3 models with ImageNet weights have been used for this research with the top layer not included and 1 Average Pooling 2D layer, 1 Dropout layer, 1 classification layer and 1 dense layer for models with architecture in Fig. 4 (a) while the models using the architecture in Fig. 4 (b) were with 3 dense layers. 297 neurons were activated for models with 3 dense layers and 105 for models with 1 dense layer. The base model was built using the functional API in Keras library because of the flexibility it provides.

IV. Results

The model was evaluated in terms of accuracy, precision, recall, and Area under the curve (AUC) metrics . The results of the experiments are visualized using tables, figures, and plots where appropriate.

A. Unbalanced and Balanced data-sets

Fig. 5 shows the result of data balancing, while the unbalanced data-set features celebrities having unequal number of images, the balanced data-set features celebrities having equal number of images (86 images) per subject/identity.

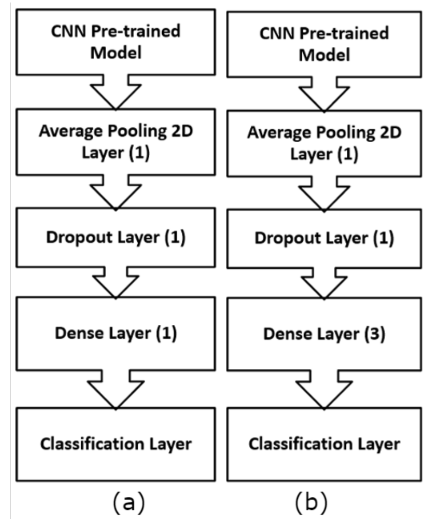

Fig. 4. The Proposed CNN Models Architectures.

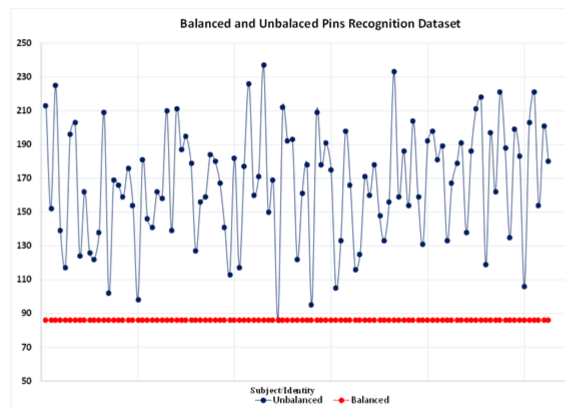

Fig. 5. Balanced and unbalanced data-sets compared.

B. Data Pre-processing/Augmentation Result

Fig. 6 shows a sample of the output of data augmentation performed on the training set with respect to horizontal flipping and slight rotation of the randomly selected face image.

C. Transfer Learning Results

The results presented in this section were obtained from the best performing runs of experiments. All results are presented in table II except for the plots comparing 1-dense layer MobileNet training using dropout only for regularization and 1-dense layer MobileNet training using both dropout and Gaussian noise for regularization.

D. Hyper-Parameter Fine-Tuning Results

Table II also illustrates the accuracies and losses as observed when training the models with 3 dense layers architecture, as well as with 1 dense layer architecture after 60 epochs of fine-tuning. In Fig. 11, training and validation accuracies and losses for MobileNet-V1 is presented while Fig. 12 reveals how MobileNet-V1 model responded to the addition of Gaussian noise to the augmented data during training.

Fig. 6. Training Data Augmentation Output.

TABLE I
MODELS PARAMETERS AS OBSERVED DURING TRANSFER LEARNING.

Model	3 Dense Layers		1 Dense Layer	
	Total	***Trainable***	***Total***	***Trainable***
Xception	21,007,761	146,281	21,076,625	215,145
MobileNet-V1	3,309,609	80,745	3,336,489	107,625
ResNet50	23,733,993	146,281	23,802,857	215,145
InceptionV3	21,949,065	146,281	22,017,929	215,145

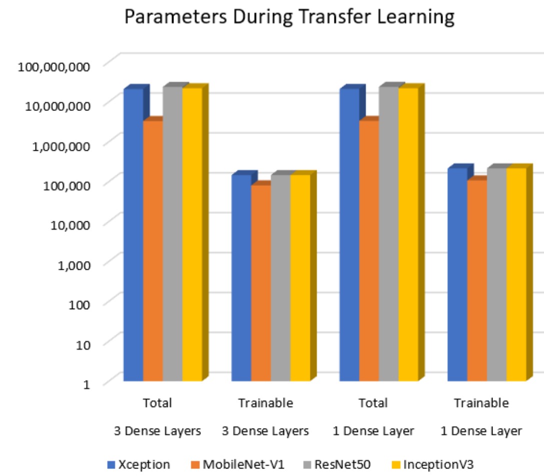

Fig. 7. Parameters as Observed During Transfer Learning

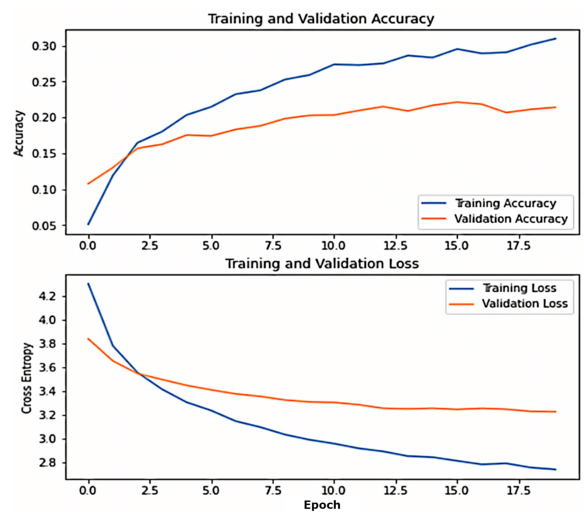

Fig. 8. Training and Loss Curves for 1-dense layer MobileNet-V1 trained with 0.6 dropout rate only.

Fig. 8 presents the training and validation curves for training the MobileNet-V1 with a dropout rate of 0.6 at a learning rate of 0.0001. That was obtained after experimenting with fine-tuning by varying the learning rate. While Fig. 9 shows the result of adding Gaussian noise with a standard deviation of 0.3, Fig. 10 was the outcome of increasing the Gaussian noise 0.7 standard deviation with dropout rate and the learning rate kept constant.

E. Model Performance Evaluation Results

The model finished with a training accuracy of 95% and 80% accuracy on the validation set. The test set was not used during the training as validation set was used to monitor how the model was performing during training and that helped during hyper-parameter fine-tuning.

V. CRITICAL ANALYSIS OF RESULTS AND DISCUSSION

A. Analysis of Results

The model was trained on a balanced data-set, to avoid the tendency that it would favour one celebrity over the other when predicting an image. It can be inferred that MobileNet V1 has shown less susceptibility to overfitting on small data-sets. Whether this view holds for other transfer learning domains may be a function of the uniqueness of that domain. Varying the standard deviation of the Gaussian noise coupled with dropout rate benefitted the model. The resultant effect of fine-tuning appeared to have yielded improvement as compared to the result obtained from training the top layer of the new model. MobileNet has a Top-1 accuracy and Top-5 accuracy of 70.6% and 89.5% respectively [23]; and that achievement of 96% in this research can be said to be a significant improvement. On the pins face recognition dataset, Ali and Kumar [24] achieved high accuracies of 93.08%, 94.04% using Inception model with SVM and achieved 93.3% using SqueezNet while Saib and Pudaruth [25] achieved the highest accuracy of 95.31%. The leading algorithm for the 2018 MegaFace challenge was trained with over 4 million of face images and was tested on the FaceScrub data-set, on which it achieved 99.939% accuracy while this model trained on 6,720 images achieved an accuracy of 98.02% [26]. That implies the MobileNet-V1 in this research compare favourably with the mentioned models.

B. Affirming or Dispelling Hypothesis

1) Transfer learning does not enhance CNN models positively: The generic knowledge gained by MobileNet-V1 model has been transferred from ImageNet domain, to be useful for a different computer vision domain entirely or to solve a different computer vision problem. The resulting CNNs models due to the added knowledge have also acquired the ImageNet weights from the pre-trained MobileNet-1, which is a positive enhancement.

TABLE II
MODEL RESULTS

Epoch	**20**								**40**							
Dense Layer	**1**				**3**				**1**				**3**			
Model	**Accuracy**		**Loss**		**Accuracy**		**Loss**		**Accuracy**		**Loss**		**Accuracy**		**Loss**	
	Training	***Validation***	***Training***	***Validation***	***Training***	***Validation***	***Training***	***Validation***	***Training***	***Validation***	***Training***	***Validation***	***Training***	***Validation***	***Training***	***Validation***
Xception	0.2698	0.2040	2.7382	3.2235	0.1746	0.1742	3.3152	3.3354	0.5942	0.5308	1.8322	2.1891	0.4523	0.4042	2.3196	2.1324
MobileNet-V1	0.3272	0.3289	2.6856	2.2276	0.2288	0.2297	3.2889	2.8836	0.7269	0.6234	0.9876	1.8276	0.5173	0.5197	1.9899	2.7632
ResNet50	0.30071	0.2800	2.2929	2.7841	0.1550	0.1406	3.0302	3.0988	0.5826	0.5142	1.8245	1.9232	0.4850	0.3806	2.0301	2.3412
InceptionV3	0.2531	0.2314	3.1844	3.1852	0.1539	0.1630	3.4371	3.3666	0.5741	0.0516	1.4354	2.4921	0.4139	0.3530	2.6854	2.4653

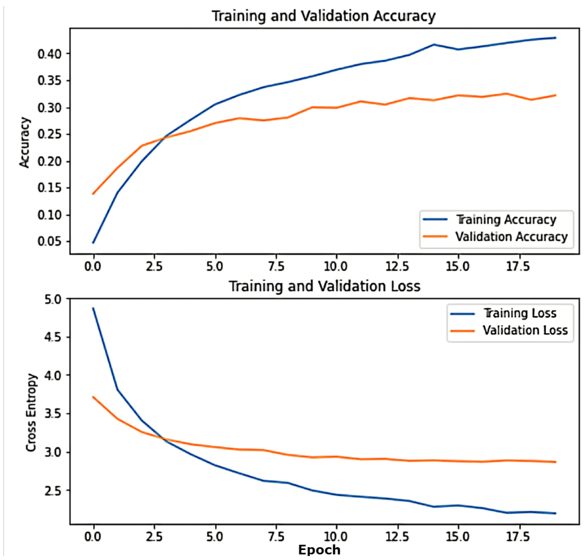

Fig. 9. Training and Loss Curves for 1-dense layer MobileNet-V1 trained with 0.6 dropout rate and and 0.3 standard deviation of Gaussian noise.

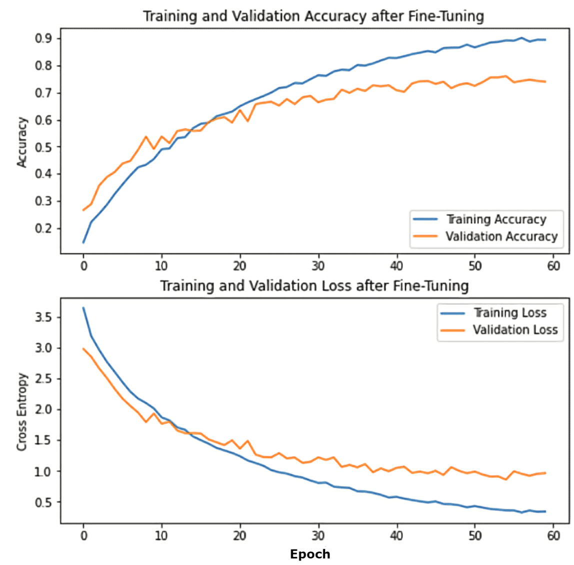

Fig. 11. Training and Loss Curves for 1-dense layer MobileNet-V1 trained with 0.6 dropout rate and and 0.3 standard deviation of Gaussian noise.

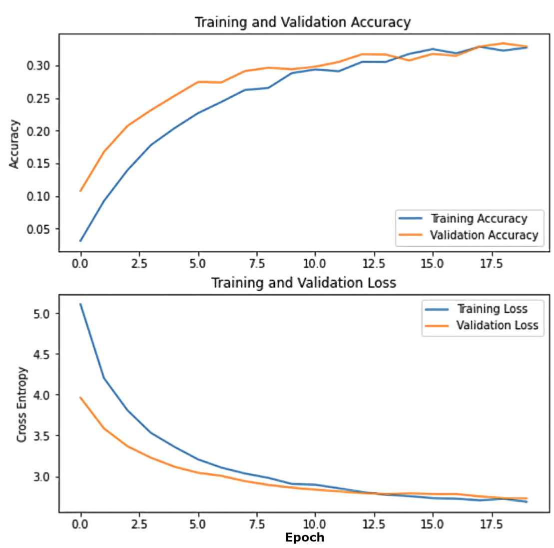

Fig. 10. Training and Loss Curves for 1-dense layer MobileNet-V1 trained with 0.6 dropout rate and 0.7 standard deviation of Gaussian noise.

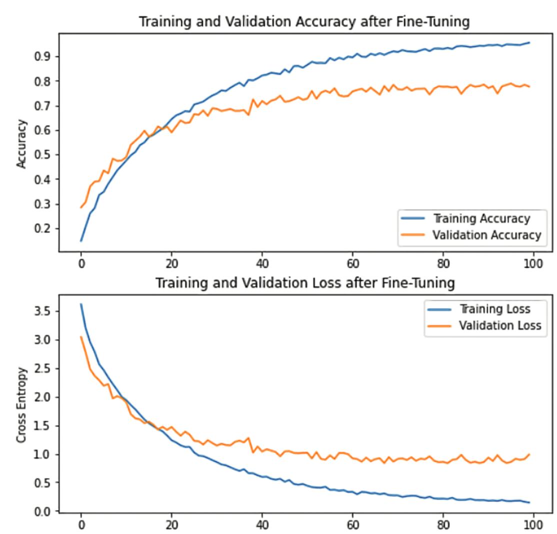

Fig. 12. MobileNet-V1 accuracy and loss curves after fine-tuning with 0.6 dropout rate only.

TABLE III
MobileNet-V1 Evaluation on Pins, LFW, and FaceScrub Datasets

Metrics / Dataset	Accuracy	AUC	Precision	Recall
Pins	0.9638	0.9744	0.9462	0.9229
FaceScrub	0.9802	0.9853	0.9760	0.9602
LFW	0.9910	0.9743	0.9862	0.9529

2) Noise has no effect on CNN models: As evidenced in the results and supporting literature of this paper, noise has effect on CNN models and considering the total outcome of this research, the model has been positively impacted by noise; therefore, the hypothesis does not hold in this regard and is there by dispelled.

3) Can the present state-of-the-art CNNs based face recognition be improved?: A look at this model predictions on faces with glasses, different illumination conditions with high confidence rate shows that the state-of-the-art CNNs based face recognition can be improved with the appropriate training dataset and continuous research and innovation.

VI. Conclusion

Considering MobileNet-V1 accuracy in TABLE II, training and loss curves in figures 7, 8 and 9 and the progressive improvement in accuracy and loss in Fig. 11 and 12 due to the effect of hyper-parameter fine-tuning, it can be concluded that hyper-parameter fine-tuning has a positive effect on the model. AI practitioners should be enlightened on the positive effects/benefits of training with noise when appropriate to do so. Researchers should do more to the existing works on noisy data driven approaches to machine learning. There is a need for more research to investigate the relationship between CNN models overfitting the data, the models' learn-able parameters and the depth or number of layers a pertained model has. A closer look at the behaviour of CNNs with the support of several enhancement techniques would benefit artificial intelligence developers and the community at large.

References

[1] C. Zhu, Y. Zheng, K. Luu, and M. Savvides, "CMS-RCNN: Contextual Multi-Scale Region-based CNN for Unconstrained Face Detection," arXiv [cs.CV], Jun. 17, 2016. [Online]. Available: http://arxiv.org/abs/1606.05413

[2] A. M. Martinez, "Face Recognition, Overview," in Encyclopedia of Biometrics, S. Z. Li and A. Jain, Eds. Boston, MA: Springer US, 2009, pp. 355–359.

[3] G. Porter and G. Doran, "An anatomical and photographic technique for forensic facial identification," Forensic Sci. Int., vol. 114, no. 2, pp. 97–105, Nov. 2000.

[4] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, present, and future of face recognition: A review," Preprints, Jul. 21, 2020. doi: 10.20944/preprints202007.0479.v1.

[5] A. J. O'Toole, P. J. Phillips, F. Jiang, J. Ayyad, and H. Abdi, "Face recognition algorithms surpass humans matching faces in images that vary in illumination," J. Vis., vol. 6, no. 6, pp. 11–11, Jun. 2010.

[6] X. Tang and X. Wang, "Face sketch recognition," IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 1, pp. 50–57, Jan. 2004.

[7] M. T. H. Fuad *et al*.., "Recent Advances in Deep Learning Techniques for Face Recognition," arXiv [cs.CV], Mar. 18, 2021. [Online]. Available: http://arxiv.org/abs/2103.10492

[8] J. Patterson and A. Gibson, Deep Learning: A Practitioner's Approach, 1st ed. O'Reilly Media, Inc., 2017.

[9] D. S. Trigueros, L. Meng, and M. Hartnett, "Face Recognition: From Traditional to Deep Learning Methods," Oct. 31, 2018. Accessed: Jul. 21, 2022. [Online]. Available: http://dx.doi.org/

[10] M. Hassaballah and S. Aly, "Face Recognition: Challenges, Achievements, and Future Directions," IET Comput. Vision, vol. 9, no. 4, pp. 614–626, Aug. 2015.

[11] A. Kulkarni and A. Shivananda, Natural Language Processing Recipes. Apress.

[12] H. Imaoka *et al*.., "The future of biometrics technology: from face recognition to related applications," APSIPA Transactions on Signal and Information Processing, vol. 10, 2021, doi: 10.1017/ATSIP.2021.8.

[13] A. Atoum, A. Al-Zoubi, M. A. Jaber, M. Al-Dmour, and B. Hammad, "A New Approach for Delivering eLearning Courses in Jordanian Universities," Advances in Social Sciences Research Journal, vol. 4, no. 48, pp. 1–13, May 2017.

[14] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep Learning Face Attributes in the Wild," arXiv [cs.CV], Nov. 28, 2014. [Online]. Available: http://arxiv.org/abs/1411.7766

[15] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2019, pp. 4685–4694.

[16] R. J. Hassan and A. Abdulazeez, "Deep Learning Convolutional Neural Network for Face Recognition: A Review," International Journal of Science and Business, vol. 5, no. 2, pp. 114–127, 2021.

[17] L. William, A. Winda, D. Satrio, T. Sofyan, and M. I. Solihin, "Automotive Start-Stop Engine Based on Face Recognition System," E3S Web of Conferences, vol. 130, p. 01020, 2019.

[18] F. Z. Xu, Y. Zhang, T. Zhang, and J. Wang, "Facial recognition check-in services at hotels," Journal of Hospitality Marketing and Management, vol. 30, no. 3, pp. 373–393, Apr. 2021.

[19] L. Li, X. Mu, S. Li, and H. Peng, "A Review of Face Recognition Technology," IEEE Access, vol. 8, pp. 139110–139120, 2020.

[20] E. McClellan, "Facial Recognition Technology: Balancing the Benefits and Concerns," Journal of Business and Technology Law, vol. 15, no. 2, p. 363, 2020.

[21] Burak, "Pins Face Recognition." Mar. 07, 2020. Accessed: Jul. 21, 2022. [Online]. Available: https://www.kaggle.com/datasets/hereisburak/pins-face-recognition

[22] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[23] A. G. Howard *et al*.., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv [cs.CV], Apr. 17, 2017. [Online]. Available: http://arxiv.org/abs/1704.04861

[24] M. Ali and D. Kumar, "A Combination between Deep learning for feature extraction and Machine Learning for Recognition," in 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Jul. 2021, pp. 01–05.

[25] Y. M. Saib and S. Pudaruth, "Is Face Recognition with Masks Possible?," International Journal of Advanced Computer Science and Applications, vol. 12, no. 7, Jan. 2021, doi: 10.14569/IJACSA.2021.0120706.

[26] "MegaFace." https://megaface.cs.washington.edu/results/.html (accessed Jul. 21, 2022).