A Survey of CNN and Facial Recognition Methods in the Age of COVID-19

CNN and Facial Recognition Survey

ADINMA CHIDUMIJE

California State University East Bay, Department of Computer Science

achidumije@horizon.csueastbay.edu

FATIMA GOWHER

California State University East Bay, Department of Computer Science

fgowher@horizon.csueastbay.edu

EHSAN KAMALINEJAD

California State University East Bay, Department of Mathematics

ehsan.kamalinejad@csueastbay.edu

JUSTIN MERCADO

California State University East Bay, Department of Computer Science

jmercado17@horizon.csueastbay.edu

JIWANJOT SONI

California State University East Bay, Department of Computer Science

jsoni@horizon.csueastbay.edu

JIAOFEI ZHONG

California State University East Bay, Department of Computer Science

jiaofei.zhong@csueastbay.edu

The rising popularity of facial recognition technology has prompted a lot of questions about its application, reliability, safety, and legality. The ability of a machine to identify an individual and their emotions through an image with near perfect accuracy is a testament to how far Artificial intelligence (AI) models have come. This study rigorously analyzes and consolidates several reputable materials with the purposes of answering the following questions: What is facial recognition? How is data acquired? What is the machine learning process? How does the Convolution Neural Network (CNN) work? It also explores the potential obstructions such as face masks that affect the machine's accuracy, security vulnerabilities, reliability, and legal concerns of the technology.

CCS CONCEPTS • Computing methodologies • Machine learning • Machine learning approaches • Neural networks

Additional Keywords and Phrases: Facial recognition, CNN, Face detection, Expression, Mood recognition

1 INTRODUCTION

Facial Recognition has been a major focus of research for the past decades. As new technologies are arising, security and data acquisition has become a major component of facial recognition. From law enforcement to commercial tasks, the demand for better, sophisticated facial recognition tasks has increased exponentially. As the demand grows, so do the complications and struggles in terms of obtaining data, processing data, and applying the technology. Many researchers are figuring out a way to make facial recognition as efficient as possible. Changes in style, posture, stance, rotation, background, lighting conditions, and camera resolutions are some examples of the issues that need to be addressed for improved facial recognition accuracy.

Using facial recognition technology in security settings could see potential growth as it is less intrusive and difficult to commit forgery. In the current heated political climate and increasing crime rates, facial recognition could make catching suspects easier for law enforcement. Facial recognition could also be used for multimedia information processing. Many algorithms attempt to perfectly identify the individual or emotion, but none so far has garnered results with 100% accuracy.

In this paper, we study the facial recognition technology, explore the best ways to gather data, and discuss the applications of facial recognition, such as taking attendance and receiving student comprehension feedback in school settings. We also explain why machine learning and data collection are at the heart of facial recognition, and further discuss the accuracy of facial recognition with obstructions such as facial coverings during *Coronavirus Disease 2019* (COVID-19) pandemic, and how the technology could be improved to navigate around barriers as such. Finally, this paper goes into the legalities and potential security risks the technology poses.

2 BACKGROUND

Within just a decades, facial recognition has gone from being a highly advanced security system in movies, to existing all around us. Even in the palms of our hands, the technology has been implemented in several ways in our society. This section introduces the background of facial recognition, and answers the following questions: What is facial recognition and why do we need it? How is data acquired and what is the machine learning process?

2.1 Facial Recognition

Facial recognition captures several images through camera or video and then parses the image, attempting to record the data of certain points on a human face to create a unique record for that person which then could be stored in a database [1]. Specifically, facial recognition uses convolutional neural networks to recognize the face by understanding accurate classification of the coefficients calculated by the eigenface algorithm. The network is first coached on the pictures from the face database, and then it is used to match the face pictures given to it. When you think of a human face, you probably think of a basic set of features, eyes, nose, and mouth. However, there's more to a face than just these features. In fact, to get a good idea of how complex this problem can get, you can draw multiple different yet simple faces. You'll notice that the faces differ in a lot of factors such as the width of the nose, the distance between the eyes, the shape and the size of the mouth, etc. Facial recognition technologies look at up to 80 factors on the face to help identify unique features and ultimately match these features to identify a person.

2.2 Research Motivation

The motivation behind facial recognition technology is security and safety. With a database of past customers, a store owner who is a victim of theft could identify the suspect with greater ease thanks to facial recognition technology. Present day security systems are mostly known for their security alarms and live monitoring features. Yet no security system attempts to modernize and develop their camera system beyond mere recording functionality. Security breaches may occur sometimes, where hackers access sensitive user data such as personal addresses, bank accounts, and social security numbers. Thieves may use this data to their own personal gain, or attempt to sell it to other thieves on the black market. If banks would implement facial recognition technology, it would be difficult for thieves to use stolen information, as it would provide another layer of security that would be near impossible to forge. Nowadays smart phones use facial recognition to grant access, and some governments like China and the United States are using facial recognition on databases like driver's licenses for a variety of reasons.

2.3 Data Acquisition

Firstly we seek to explain how facial recognition technology works. When a camera takes a picture, it is saved to flash storage and each picture contains coordinates and pixels. Each coordinate can tell us what is part of the face and what is part of the background. In CNN facial recognition, there is an AI that is trained with algorithms to process the image in a way where it gets all the coordinates as accurately as possible and calculates who the person is. This AI not only checks the coordinates but also checks other factors such as facial tracking, position, and other parts of the face that the AI finds unique. One of the algorithms that is used to recognize images in the CNN model is called One-Shot-Learning [2]. It takes multiple similar images and tries to connect them and compare them with the coordinates to unlock the Iphone. Everytime it gets the face it learns and improves.

In one research paper [3], the researchers attempt to find the best setup that is cost effective, scalable and provides high quality facial recordings for facial recognition systems to parse. Firstly one must choose a recording device. Cameras mounted on tripods provide high quality recordings as they have a high frame rate and high resolution, but they are rather expensive and not very scalable. The Pan-Tilt-Zoom (PTZ) cameras are the worst option [3] as it requires a dedicated room to be set up with expensive recording equipment and restricts the subject from moving or sitting in specific spots in the room. Webcams have high scalability because they are common in today's devices, so the cost of this recording device trumps all others. However, webcams fail to produce a parsable video for the facial recognition algorithm when the subject partly turns their face away from the front of the camera. Additionally, some webcams have low frame rate and low resolution which may render the recorded footage unusable. The best option found by the research article are head mounted cameras. Head mounted cameras use inexpensive GoPros that can be attached to a rig that is worn by the subject. Since GoPros are inexpensive, the setup is well scalable just beat by webcams. However, this solution of head mounted cameras addresses the issues webcams faced with subject head rotations. No matter which way the subject moved their head or body, the GoPro was able to record in high resolution and high frame rate, a video that is usable for facial recognition programs. The only issue is synchronizing the audio and video that is recorded by the GoPro, but that was addressed through the advent of FaceSync, a open source framework that syncs the two together. Another recent work [4] utilized the Augmented Reality (AR) devices namely the Microsoft Hololens for face recognition application, which could be applied to other similar AR devices as well.

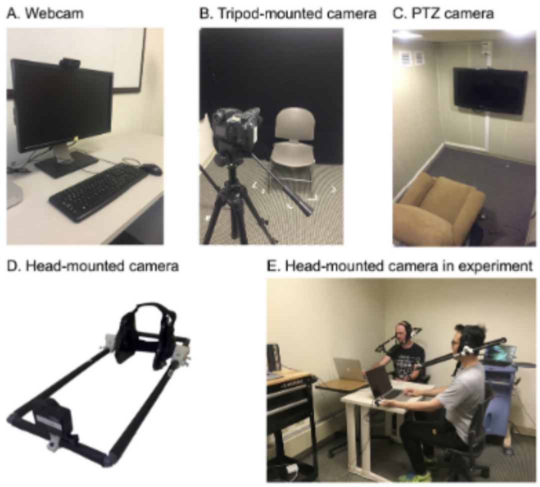

Figure 1: Different Recording Setups for Facial Recognition Recording [3]

2.4 Convolution Neural Network

In the machine learning process, the first thing you typically do is to calculate a couple of features out of faces. For example, first you can calculate the distance of the eyes, then the distance between your eyes and your nose, then the distance between the nose and the mouth and so on and so forth. Then we can put this all into a train machine learning model table which helps us predict whose face it actually is. This table is incredibly accurate as it recognizes people with nearly 100% accuracy and is even able to detect the minor differences in identical twins as the technology has improved and become more widespread the possibilities for the future.

To understand what a Convolution Neural Network (CNN) is, the Artificial Neural Networks (ANNs) should be introduced first. An ANN is the umbrella term for various Neural Networks, and is consisted of 3 main parts: an input layer, a hidden layer(s), and an output layer. The input layer contains that data that we would like to classify and is usually in the form of a matrix or vector. The computer will then read in this data and put it through the next layer, which is the hidden layer. This layer applies what is called an “activation function” to the previous layer, so that the program is enabled to learn nonlinear relationships. Each node in the layer has a weight associated with it, which can be determined by using gradient descent and find the local minima. Due to the large number of parameters that are used in neural networks, finding these gradients are computationally expensive, so we use a process called back-propagation, which allows us to use the chain rule without performing repeated calculations. Following this layer is our output layer, which gives a vector whose entries correspond to a different class. The output layer allows us to interpret the results as probabilities and we can then observe how well the network was trained on the data.

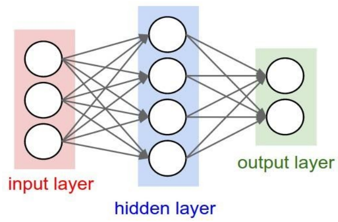

Figure 2: Three Layers of an Artificial Neural Network

A CNN is a much more sophisticated neural network, where the main operation performed is the convolution. Krizhevsky et. al. [5] gave way to the rise of CNNs, as their research showed that their neural network outperformed previous attempts by others. By adding a convolutional layer to neural network, it enables the learning of features in an image. Doing so, it is possible to create image filters by applying the convolution operator to the input matrix and the filter matrix, also known as a kernel. A kernel can be used as an edge kernel, which detects the horizontal, vertical, or both types of edges on an image. During convolution the filter (or kernel) is moved across the width and height of the input image. The dot product of the receptive field and the filter is computed at each step and the result is stored an activation map. There will be many features and many activation maps. This type of CNN paves the way towards a more generalized form of classification, which will lead us into faster and smarter image recognition.

3 FACIAL RECOGNITION METHODS

In this section, we analyze a number of recent research works of facial recognition methods and their applications. We seek to explain what drives the development of facial recognition technology, and what algorithmic and patterns are required to solve the ongoing problem.

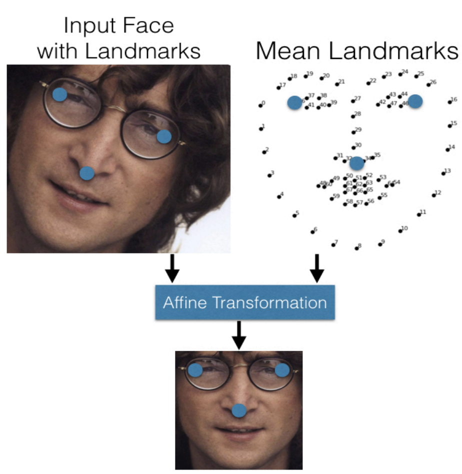

Figure 3: OpenFace's affine transformation [6]

3.1 OpenFace: a Face Recognition Library

When it comes to face recognition, OpenFace [6] is one of the face recognition systems using neural networks, based on the technology of Facebook's DeepFace [7] and Google's FaceNet [8] systems. In the preprocessing stage, OpenFace uses affine transformation to normalize faces. During the training stage, the preprocessed image is mapped to the low-dimensional face representation. OpenFace provides near-human accuracy on the LFW benchmark [9], and presents a new classification benchmark for mobile scenarios, where a real-time face recognition system adapts depending on context, meanwhile maintaining low training and prediction times.

3.2 A Web Service for Facial Recognition Using Convolutional Networks

There is one study [10] about how to use convolutional networks built in web development for facial recognition. This system is designed to use facial recognition with the web service called Vision, which is developed with Flask and Tensorflow. These programs give a user access to a deep learning framework simultaneously which helps them generate the data. Furthermore, the data helps the development team figure out authentication and user memorization, so not only do the users benefit from this service. This is all achieved through using facial recognition in web development.

3.3 Image Net Classification With Deep Convolutional Neural Networks

The research work discussed in one paper [5] studies improvements in object recognition through machine learning methods over relatively small datasets. Object recognition was significantly improved by collecting larger datasets, learning more powerful training models with multinomial logistic regression objective, and using better techniques for preventing overfitting via data augmentation and dropouts.

3.4 Deep Convolutional Neural Network Used In Single Sample Per Person Face Recognition

The study that is discussed in one paper [11] talks about how they have trained a facial recognition system using per person recognition. This means that it uses a picture recognition system where it puts filters in each picture it takes to get the best accuracy. These pictures are then filtered and the background is erased and every time it recognizes the face. The AI is trained every time the person is on the camera, meanwhile the AI is able to improves itself, by means of different ways to expand the knowledge of their AI. Every data is obtained and stored into a storage which grows and improves over time.

3.5 Convolution Neural Network Cascade for Face Detection

One paper [12] proposes a cascade architecture which enhances facial detection despite difficulties arising from a large search space of possible face positions and sizes, as well as large visual variations resulting from factors such as pose changes, exaggerated expressions and extreme illuminations. With the use of calibrations nets, this cascade architecture is a much faster and a more effective approach to facial detection.

3.6 Facial Expression and Mood Recognition

With the use of facial recognition, one system in the research paper [13] studies how an integrated committee neural network system accurately and reliably classified the emotions of subjects from facial expressions and related changes in facial patterns. Two methods use eight real-valued parameters and seven binary parameters from each facial image on a database containing the facial images of 97 subjects. The study demonstrates how the system is able to classify facial images and distinguish between seven distinct emotions: neutral, angry, disgust, fear, sad, surprised or happy. Researchers sought to explain how their system correctly identifies a person's emotion with 90.43% accuracy. They did this by breaking down the human face into two types of parameters, real valued and binary. Real valued parameters include the following criterions: Eyebrow raise distance, Upper eyelid to eyebrow distance, Inter-eyebrow distance, Upper eyelid-lower eyelid distance, Top lip thickness, Lower lip thickness, Mouth width, and Mouth opening. Binary parameters are broken up into the following criterions: Upper teeth visible, Lower teeth visible, Forehead lines, Eyebrow lines, Nose lines, Chin lines, and Nasolabial lines.

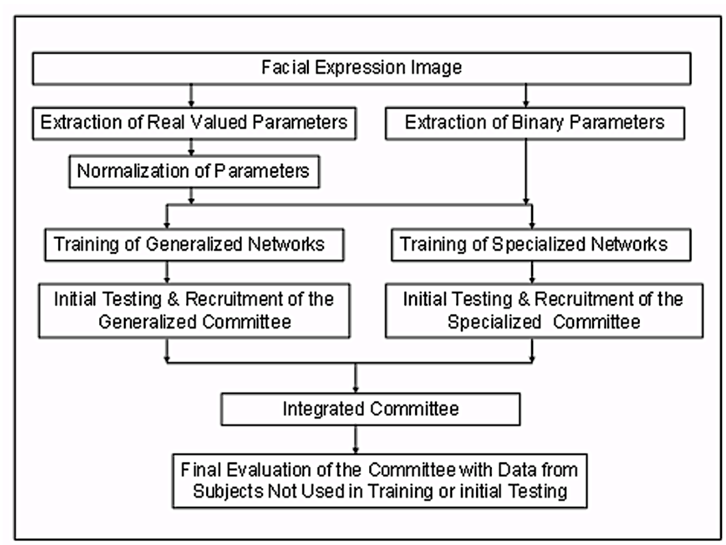

Figure 4: An Overall Block Diagram of Facial Expression Mood Recognition [13]

3.7 Class Attendance Management

We can use facial recognition to take attendance in classrooms. Taking attendance creates wasted learning time for both students and professors. Facial recognition could give back this wasted time. There is a research work in paper [14] studying how to meet the growing needs of efficient and automatic techniques of taking attendance in the classroom with the use of facial recognition. It shows the limitations of already existing methods to take attendance, such as pen and paper, biometrics, fingerprint and Radio Frequency Identification tags. It then explains why and how facial recognition is a more efficient, effective, and safer approach. Through the use of facial recognition, professors could easily and confidently identify who is attending class and who is not. This system of facial recognition is highly secure because it is near impossible to pose as someone else, an issue that plagues many other attendance taking methods. Pen and paper are subject to false signatures and proxies.

3.8 Face-sync for Class Attendance with Open Source Framework

One paper [3] discusses how facial recognition through machine learning can not only can identify people but also identify their emotions, feelings and thoughts. However these advancements in facial recognition come with limitations and difficulties such cost, reliability, adaptability, and flexibility. The paper talks about an alternative that addresses these limitations. A mobile head-mounted camera that can be easily constructed, is cheaper, more flexible, and more reliable than conventional video recording setups such as webcams, tripod-mounted cameras, or pan-tilt-zoom cameras.

3.9 An Adaptive Pig Face Recognition Approach Using Convolution Neural Networks

Another work [15] introduced how to use facial recognition on animals. It talks about how as of right now they are using *Radio-Frequency Identification* (RFID) tags to recognize the animals and this is working efficiently. The only problem is that these tags take so much time and labor as they have to put it in each animal individually. To fix this problem, they have made an AI that uses CNN Model to recognize each animal. In this research they used pigs as the subjects and they have found about 96.7% success rate. The research work goes into the details of how they achieved this.

3.10 Fast R-CNN

One paper [16] proposes a Fast Region-based Convolutional Neural Network method (Fast R-CNN) as an improvement over already existing technology such as R-CNN and SPPnet, for object detection. This is achieved through the use of a single-stage training algorithm that jointly learns to classify object proposals and refine their spatial locations, Fast R-CNN is significantly faster than R-CNN and SPPnet.

3.11 Mask R-CNN

Another paper [17] introduces a simple, flexible, and fast system for instance segmentation called Mask R-CNN. By adopting and extending the two step procedure in Faster R-CNN, Mask R-CNN outperforms previous state of the art instance segmentation methods. Like Farter R-CNN, It applies the principles of bounding-box classification and regression in parallel. However because Faster R-CNN was not designed for pixel to pixel alignment, it often has misalignments and this is where Mask R-CNN outshines Faster R-CNN. With the integration of RolPool to Mask R-CNN, the misalignment problem is fixed and mask accuracy is improved by 10% to 50% relatively.

3.12 Facial Recognition Application using Augmented Reality Devices

Kim et. al. [4] implemented a facial recognition application on the augmented reality device, i.e. Microsoft Hololens. They use a four-step process to detect faces in the image, transform the detected faces as input to a CNN, pass each face through a CNN to produce a feature vector, and finally compare the feature vector to the feature vectors of known faces and pick the closest one as the correct face. The implement of this process on a non-GPU based server can result in a frame rate of only approximately one frame per second to detect and identify faces in real time.

4 DISCUSSION

With COVID-19 pandemic, students and workers alike switched from working in person to working online. With the surge of online users, many unknowledgeable of the dangers on the Internet, there is an increased risk of data breaches targeting new computer users. Facial recognition provides an extra layer of security that everyone would benefit from. This section delves deep into the key areas of facial recognition and answer the following pressing questions: What are the effects of face masks and other facial coverings to the accuracy of the technology? What are the threats to the security, reliability, and legal concerns of the technology?

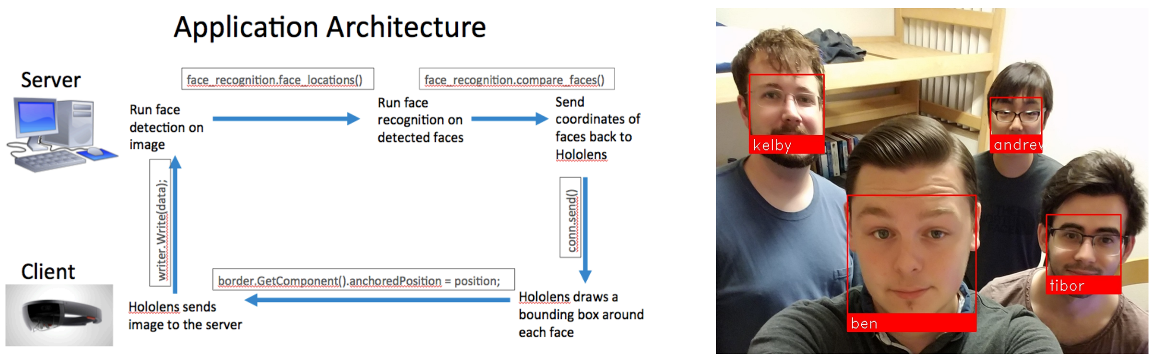

Figure 5: Face Recognition using Augmented Reality Devices [4]

4.1 Effects of Face Masks and Other Facial Coverings During COVID-19

In the age of COVID-19 facial recognition is even more essential as it allows for increases in classroom efficiency, security, and public safety just to name a few use cases. As facial recognition technology uses biometric data from your eyes, nose, and mouth to securely unlock devices, facial coverings make it nearly impossible to implement the technology to its fullest potential during the age of COVID-19. One suggested solution to this is to use biometric iris recognition scanners to illuminate the iris with invisible infrared light to pick up unique patterns. Although this does fully fall into the realm of fictional technology, it is still a solution to be considered.

4.2 Threats to the Security, Reliability, and Legal Concerns

One of the issues with facial recognition is its intersection with surveillance possibilities. Consider the number of surveillance cameras in both your private and public life. If you were to connect those into a network, unsolicited data would flow in a continuous manner, and when combined with facial recognition, that data would be processed automatically. Essentially, you could track everyone, all of the time.

One paper [18] reveals the adverse consequences that facial recognition has on children. It argues that not only does surveillance undermine a free society and threaten democracy, it also discourages mind-wandering and creativity, works against children becoming autonomous, fully functional citizenry, promotes racial inequality and weakens the bond of trust between children and adults.

5 CONCLUSION AND FUTURE WORK

Facial Recognition is at the forefront of innovation in today's society. There are a variety of ways one can collect data for facial recognition parsing. As mentioned earlier, webcams are not the ideal recording device for facial recognition data collection, however, the projected benefits from utilizing facial recognition outweighs the time it will take to get the system running. People could use a combination of webcam, GoPro, and AR devices, etc. for facial recognition applications. With complex algorithms, we have a system that is ready to process the facial data.

One of the ways that facial recognition could be utilized is through the implementation in classrooms, such as taking attendance in classrooms. These applications although sounding flawless have much room to be improved. One of the struggles we would face if using a facial recognition algorithm in a physical classroom is the presence of masks obstructing the face, especially during COVID-19. Either one would have to code a facial recognition method that could focus on the uncovered parts of the face, or we would have to implement a system in person that each person can display their face to the camera by taking off their mask momentarily. Another problem would be implementing the technology in online classrooms, as some student's webcams may not be high enough resolution or poor conditions to make the algorithm not work.

Additionally, with all the data, facial recognition algorithms would be able to tell the instructor truthfully how well the students are understanding the material in class. Unlike traditional teaching methods that require active student feedback, facial recognition allows the professor to adjust the speed and teaching style of the course from passive student feedback through their facial expressions in class. Futhermore, facial recognition could also assist in monitoring online tests.

ACKNOWLEDGMENTS

We would like to thank the California State University Fresno, the California State University East Bay, and the National Science Foundation for their support (Grant No.: NSF-DMS-1460151), the FURST program (Grant No.: NSF-DMS-1620268), and CSUEB FURST program (Grant No.: NSF-DMS-1620500).

REFERENCES

[1] Jain, A.K. and Li, S.Z., 2011. Handbook of face recognition, Vol. 1, New York: springer.

[2] Bengio, Y., Goodfellow, I. and Courville, A., 2017. Deep learning (Vol. 1). Massachusetts, USA: MIT press.

[3] Cheong, J.H., Brooks, S. and Chang, L.J., 2019. FaceSync: Open source framework for recording facial expressions with head-mounted cameras. F1000Research, 8.

[4] Kim, A., Kamalinejad, E., Madal-Hellmuth, K. and Zhong, F., 2020, March. Deep Learning Based Face Recognition Application with Augmented Reality Devices. In Future of Information and Communication Conference (pp. 836-841). Springer, Cham.

[5] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), pp.84-90.

[6] Amos, B., Ludwiczuk, B. and Satyanarayanan, M., 2016. Openface: A general-purpose face recognition library with mobile applications. CMU School of Computer Science, 6(2).

[7] Taigman, Y., Yang, M., Ranzato, M.A. and Wolf, L., 2014. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1701-1708.

[8] Schroff, F., Kalenichenko, D. and Philbin, J., 2015. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 815-823.

[9] Huang, G.B., Mattar, M., Berg, T. and Learned-Miller, E., 2008, October. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Erik Learned-Miller and Andras Ferencz and Frédéric Jurie, Oct 2008, Marseille, France.

[10] Archilles, A. and Wicaksana, A., 2020. Vision: a web service for face recognition using convolutional network. Telkomnika, 18(3).

[11] Zeng, J., Zhao, X., Gan, J., Mai, C., Zhai, Y. and Wang, F., 2018. Deep convolutional neural network used in single sample per person

face recognition. Computational intelligence and neuroscience, 2018.

[12] Khan, S., Javed, M.H., Ahmed, E., Shah, S.A. and Ali, S.U., 2019, March. Facial recognition using convolutional neural networks and implementation on smart glasses. In 2019 International Conference on Information Science and Communication Technology (ICISCT), pp. 1-6, IEEE.

[13] Kulkarni, S.S., Reddy, N.P. and Hariharan, S.I., 2009. Facial expression (mood) recognition from facial images using committee neural networks. Biomedical engineering online, 8(1), pp.1-12.

[14] Salim, O.A.R., Olanrewaju, R.F. and Balogun, W.A., 2018, September. Class attendance management system using face recognition. In 2018 7th International Conference on Computer and Communication Engineering (ICCCE), pp. 93-98, IEEE.

[15] Marsot, M., Mei, J., Shan, X., Ye, L., Feng, P., Yan, X., Li, C. and Zhao, Y., 2020. An adaptive pig face recognition approach using Convolutional Neural Networks. Computers and Electronics in Agriculture, 173, p.105386.

[16] Girshick, R., 2015. Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, pp. 1440-1448.

[17] He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision, pp. 2961-2969.

[18] Bala N., 2020. The Danger of Facial Recognition in Our Children's Classrooms. Duke Law & Technology Review. 18(1), pp.249-267.

Authors' background

Your Name	**Title***	**Research Field**	**Personal website**
Adinma Chidumije, Fatima Gowher, Justin Mercado, Jiwanjot Soni,	Undergrad students	Computer Science	
Jiaofei Zhong, Ehsan Kamalinejad	Associate Professor	Computer Science, Mathematics	

***This form helps us to understand your paper better, the form itself will not be published.**

***Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor**