

Face Recognition Using Convolutional Neural Network

Alla Abdulkareem Saber¹, Asist. Prof. Dr. Gullanar M Hadi¹

¹Department of Software and Informatics Engineering (SIE), College of Engineering, Salahaddin University- Erbil, Kurdistan Region, Iraq

Abstract—One of the important areas of computer vision is the face recognition (FR) issue. FR is used to detect faces seen on dispersed cameras throughout a network. There are two categories of face recognition problems. The first category, called the face identification issue involves recognizing faces based on samples, per person. One method of face recognition known as Single Sample Per Person (SSPP) is the type commonly referred to as "current face recognition approaches." There is research that supports both methods and deep learning techniques. This research paper introduces a CNN model that aims to expedite network convergence and minimize the duration of training. As a step we incorporate a normalization layer, into the CNN framework. To accommodate the normalized data, we include an activation function, within the existing activation function allowing for modifications. Finally, we utilize max pooling to ensure that the feature information remains stable and intact while still preserving its details. The experimental findings demonstrate that our approach outperforms the leading technique and addresses the issue of results, between the training and test datasets resulting in a lower rate of recognition.

Keywords—layered activation function, fine-tuning, probabilistic max-pooling, face recognition, and convolutional neural network.

1. INTRODUCTION

Face recognition (FR) is a challenge, in the field of computer vision. Many security cameras use facial recognition algorithms to identify people, in videos (Dewan et al., 2016). Video monitoring and law enforcement serve functions. Play multiple roles (Matta and Dugelay, 2009). One major issue, with recognition is the lack of citations (Abdelmaksoud et al., 2020). In present years, there have been significant developments in face recognition technology, which is now extensively used in applications related to identity verification, security, and surveillance. Deep convolutional neural networks (CNNs) have shown great excellence in face recognition applications; nevertheless, training deep CNNs

may be laborious and necessitates a substantial quantity of labeled data. To address these issues, the method of "transfer learning," It has been suggested that information be moved from a source domain to a target domain. (Ikromovich 2023).

Biometric recognition technology is a form of identifying or confirmation based on physiological or behavioral features of persons. Iris, fingerprint, palm print, visage, ear kind, voiceprint, as well as gait are all biometric characteristics that are employed in the processes of authentication. Face recognition is a primary trend in biometric recognition technology, which involves face-based identification. Face recognition technology offers distinct benefits over other biometric technologies. The acquisition method is non-contact, and user interaction is not necessary. It conceals itself well in the meantime. According to Li et al. (2018), the advantages make it especially appropriate for applications including airport anti-terrorism and security surveillance. Numerous methods focus on facial recognition.

The first is a conventional method that uses characteristics or landmarks extracted from the picture to identify faces. For instance, an algorithm will look into the proportions, forms, and placement of the nose, cheekbones, and eyes. It will then use these features to look for other images that have similar features using techniques like principal component analysis (PCA) as well as support vector machines (SVM) (Bashbaghi et al., 2017). (Zhao et al., 1998). The second method is called dimensional recognition; it makes use of 3D sensors to gather a great deal of data on the geometry of the face and then uses this data to identify unique features on the face, like the shape of the nose. This information enhances the effectiveness of face recognition (Cook, Chandran, and Fookes, no date). Skin texture analysis is the third method. This innovative design makes use of the skin's aesthetic characteristics. Skin texture study improved the face recognition method's performance (Lal, Srilakshmi, and Venkateswarlu, 2022). Thermal cameras make up the fourth method; there is a difference in the format of the data that is being used for face identification, this method is the only way the cameras will identify head form. It also stays away from a person's accessories, such as hats,

spectacles, or cosmetics. When using thermal images for face identification, the primary issue is that face recognition datasets are limited (Socolinsky, Selinger, and Neuheisel, 2003; Abdelmaksoud et al., 2020).

2. RELATED WORK

CNN model algorithm: The algorithm's broad description is as follows: Separate the dataset into validation and training sets in **step 1**. Fully linked layers in **Step 2**. **Step 3:** Utilize the training set to train the CNN model, and then optimize it with methods like gradient descent as well as back propagation. Create a CNN architecture that works for face recognition. This design usually consists of many pooling layers, convolution layers, and **Step 4:** Adjust the model's hyper parameters (understanding rate, batch size, etc.) to improve performances. **Step 5:** Use the validation set to assess the model and adjust the architecture or hyper characteristics as needed. The aforementioned algorithm offers a high-level overview; more modification would be necessary depending on the particular needs, resources available, and technical limitations of the university's attendance and surveillance system (Gudikandhula and Kondapalli, 2023).

In order to generate a useful CNN model, we employ many layers that collaborate to extract significant characteristics from the provided input picture. Among these levels are:

A. Layer of Convolution: The Convolutional Layer extracts edges and other fine details from the input image. It finds gradients and patterns by applying filters. The network may comprehend pictures more thoroughly by adapting to more complicated characteristics when later layers pick up on low-level properties like edges. This starts with the first convolutional layer.

B. Batch Normalization Layer: Because of the sensitivity of starting weights and learning algorithm design, deep neural network training with several layers can be challenging. This problem is solved by incorporating batch normalization, which decreases the number of training cycles required for networks stabilizes the learning process and standardizes inputs, for each group of data.

C. Max Pooling Layer: The Max Pooling Layer helps to speed up computations by reducing the size of the features in space through a decrease, in dimensions. This technique aids, in identifying characteristics that remain unchanged when the model is rotated or moved thereby enhancing the effectiveness of model training. Max pooling serves a purpose by not reducing dimensionality but also eliminating disruptive activations.

D. Dropout Layer: Dropout serves as a technique, for regularization, in networks preventing overfitting by introducing some level of imperfection. Dropout works by turning off neurons at random during training. This helps the

network build stronger features that work with a wider range of neuron combos. It lessens neuronal interdependency and encourages improved generalization.

E. Fully Connected Layer: The Fully Connected Layer combines high-level characteristics acquired by preceding layers in non-linear ways. It flattens the input picture and applies a non-linear modification to enable the network to categorize it using methods such as SoftMax. According to Jagtap et al. (2023), this layer gives the model the ability to identify prominent traits and generate precise predictions.

In summary one of the primary trends in biometric recognition is face recognition, or face-based identifying technology. Face recognition technology offers distinct benefits over other biometric technologies. The acquisition method is non-contact, and user interaction is not necessary. It conceals itself well in the meantime. According to Li et al. (2018), the advantages make it especially appropriate for applications including airport anti-terrorism and security surveillance. Using facial recognition to track attendance involves many steps. The student's facial image must first be entered into the system as a dataset. The camera capture application may be used to collect the student's facial image, which can then be uploaded into the database (Goyal et al., 2021).

Using a face recognition method that is selected from the database's data, the second stage involves training the dataset (Shrestha and Panday, 2020). A face recognition system is utilized to record and identify student profiles through the use of a classroom-installed camera in the third phase. The captured faces are then compared to facial photos stored in the database. Students whose presence should be reported as absent will be those who are expected to be there but are not found (Chowdhury et al., 2020). The university's attendance system will automatically record the attendance data (Budiman et al., 2022). Two categories exist for face recognition problems: the first, known as the classic face identification issue, deals with recognition from several samples per subject. The identification of faces utilizing a single specimen per person (SSPP) is the second. Even though deep learning algorithms and other contemporary face recognition approaches attained excellent accuracy, they performed poorly in situations with little data. The reason is that deep learning methods require a lot of data in order to achieve high accuracy (Abdelmaksoud et al., 2020).

Using deep learning techniques is an additional way to utilize face recognition. Deep learning optimizes face recognition speed by utilizing the CPU and GPU power of the device. CNN is one use of deep learning that could produce useful outcomes. The benefit of the CNN method is that it contains many layers that work together to extract the most useful information from a picture using machine learning, saving us the trouble of manually entering these features. CNN also has the benefit of increasing accuracy since the algorithm learns more from the usage of additional datasets, which increases accuracy (Kakadiaris et al., 2017). Videos of various people facing various directions are used to construct the dataset. A

comparison of each of the deep learning facial recognition algorithms that were used in the study is presented in Table 1 (Damale and Pathak, 2018).

Table 1: Deep Learning Facial Recognition Algorithm Accuracy (Budiman *et al.*, 2022)

Algorithm	Accuracy
SVM	88%
MLP	87%
CNN	98%

Convolutional neural networks (CNNs) are primarily powerful because of their deep architecture, that makes it possible to extract a collection of feature representations that discriminate across several levels of abstraction. Due to its excellent performance, CNN has been extensively used in face recognition in recent years (Lu, Wang, and Zhou, 2017). (Hu *et al.*, 2018). CNN's success can be attributed to its capacity to recognize complex picture characteristics. But learning millions of network parameters is necessary for training a deep CNN network, and a various of labeled datasets are needed for pre-training (Oquab *et al.*, 2014). According to Abate *et al.* (2007), face recognition may be broadly classified into three technological categories. First, there is picture-based 2D face recognition. 2D face recognition has been around for a while and has been very successful. The primary obstacle to the further advancement of 2D face recognition is the brightness picture, which is invariably affected by variables like lighting. 2D face recognition detected a face using this image.

Based on 3D geometric data, 3D face recognition falls into second categories. A growing amount of focus has been on 3D facial recognition in recent years. Factors like lighting, posture, and occlusion may readily alter two-dimensional face photos, which will reduce the face recognition system's capacity to recognize faces. These factors, however, hardly have an impact on a 3D face recognition system. Solving facial difficulties with different lighting and posture has greater possibilities. This is so that variations in posture won't cause information loss and the three-dimensional data could accurately depict the surface geometric aspects of the human face. Furthermore, three-dimensional data will not be impacted by changes in light if the acquisition procedure is ignored since three-dimensional data lacks brightness information.

The 2D+3D dual-mode fusion system (Soltanpour, Wu, and Anvaripour, 2015; Wang *et al.*, 2013) is the third kind of face recognition technology that has demonstrated high performance. 2D+3D is a method that effectively combines the more advanced 2D image recognition technology with the three-dimensional shape aspects of the face in order to enhance the performance of face identification. Advances in 3D-2D face recognition have been made recently (Wang *et al.*, 2013; Kakadiaris *et al.*, 2017). Our suggestion is C2D-CNN, a deep face recognition technique with cheap computing costs and good accuracy that integrates the two characteristics at the decision-making stage.

We look at a fresh CNN model, Li *et al.* (2018) The network has been carefully designed, incorporating three components (1) To tackle problems associated with saturation and gradient diffusion a new algorithm called activation has been implemented. The main objective of this algorithm is to improve the nonlinearity of the activation function (2) To speed up the networks convergence and reduce the training time needed normalization is applied (3) Max pooling is used to keep the features stable while also ensuring their consistency.

3. METHODOLOGY

Color 2-Dimensional Principal Component Analysis-Convolutional Neural Network (C2D-CNN) for Face Recognition Model

- 1. Summary of the Suggested Approach** The process of recognizing faces using decision level fusion and a C2D CNN model comprises three components; CNN feature extraction, color based 2DPCA feature extraction and decision level fusion. You can see the flow of this algorithm, in Figure 1.

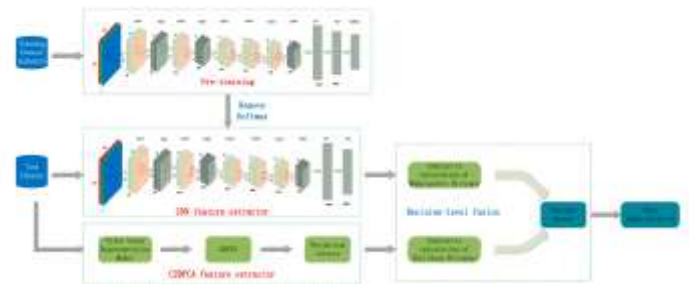


Figure 1: The diagram demonstrates how face identification is achieved by combining two principal component analysis, with convolution network (CD2 CNN). (Li *et al.*, 2018).

The overall process of the recommended C2D CNN approach can be seen in Figure 1. First, the input picture in the suggested approach is scale-normalized. Next, we separate our algorithm into three components: decision-level fusion, color 2DPCA feature extraction, as well as CNN feature extraction. Additionally, the normalized pictures are subjected to CNN as well as color 2DPCA applications, roughly, for feature extraction. To acquire the depth characteristics between them, the CNN model is trained using the FRGC (Face Recognition Grand Challenge) v2.0 dataset. Handcrafted characteristics with more detailed color and spatial information are extracted using Color 2DPCA. The Mahalanobis and Euclidean distances are used to find out how similar the feature vectors are so that they can be put together, Li and colleagues (2018)

2. CNN for feature extraction

The CNN model that is recommended in this work is made up of a fully connected layer, a stacked activation function layer, a probabilistic max-pooling layer, as well as a convolution layer. After being trained on a lot of data (without the Softmax

classifier), this method uses CNN as a feature extractor to pull out face features. Color shots of faces that have been scale-normalized are fed into convolutional neural networks. Several convolution kernels are first used to make the original image more complicated so that the picture's features can be changed and extracted. The filter factors are set automatically throughout the the process of training based on the features of the training pictures that are fed in. The normalization layer then adds a normalizing layer to the convolutional operation's results. This keeps the network from fitting too well and makes sure that gradient descent works well. That data is then sent to the network's stacked activation function layer to make it better at non-linear expression. Ultimately, the outcome of the activation function is summed up. This keeps the most important features and makes the model less likely to become distorted. Once there are a lot of convolutional and pooling layers, the fully linked layer is where the important features are run. This produces outputs for CNN feature representation. CNN's pre-training is mostly made up of two parts: forward learning and back transmission of network characteristics (Li et al., 2018).

3. Fine-Tuning Information:

There is usually a big variation among the pre-trained dataset and the dataset of the goal job, no matter how many groups or picture patterns there are. When using a pre-trained CNN model to pull out an image's visual features, it can be hard to get better results in the target dataset retrieval job. Because of this, we use the target dataset picture to tweak the pre-trained CNN model settings so that they work better for extracting features from the target dataset. The process of retraining the CNN model with the FRGC training set is called fine-tuning training. During training, Gaussian random values are only used to set the output layer weights. The matched pre-training model weights are used to set the other CNN model weights. The first seven layers of the network should have a modest learning rate set, which is different from the pre-training, to prevent the parameters acquired from the pre-training CNN model from being lost throughout the fine-tuning procedure. Setting a greater learning rate for the last layer guarantees that the network swiftly converges to the new optimum point on the target dataset. According to Li et al. (2018), the CNN-0 network's learning rate (lr) is lowered to 1/10 for every 20 epochs in order to guarantee the network's normal convergence.

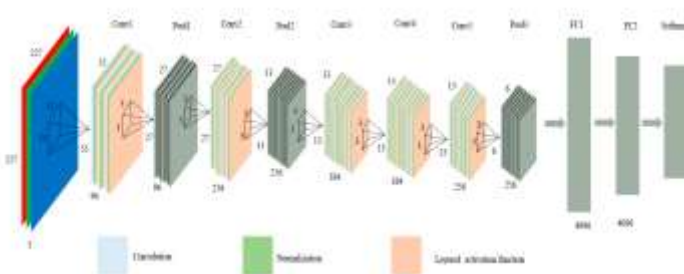


Figure 2: Facial recognition is the foundation upon which the CNN network is built.(Li et al., 2018)

The face features in this technique are extracted using a feature extractor that is a pre-trained CNN model. CNN uses the 4096-dimensional output of the final fully linked layer to derive the eigenvector. Figure 3 depicts the completely linked layer's structure.

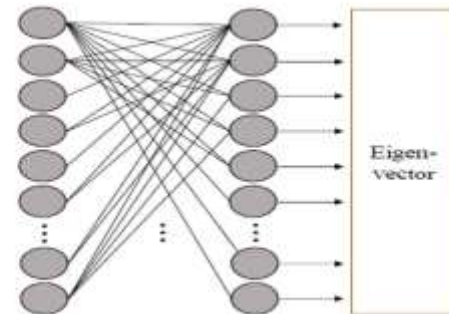


Figure 3: This is the structure of the layer that is entirely connected.(Li et al., 2018)

Comparing the Efficaciousness of Decision-Level Fusion

With small changes in this work, we looked at how well decision-level fusion as well as fine-tuning improved recognition accuracy and how well the C2D-CNN model recognized things. We only checked how well it worked on the FRGC database, so we chose to use FRGC Experiment 4 since there is no label data to utilize for improving the pre-trained model. This has four sets of trials: CNN-0, CNN-0 + fine-tuning, C2D-CNN, and C2D-CNN + fine-tuning. The closeness number is always set at 0.6. Figure 4 shows the outcomes of the experiment (Li et al., 2018).

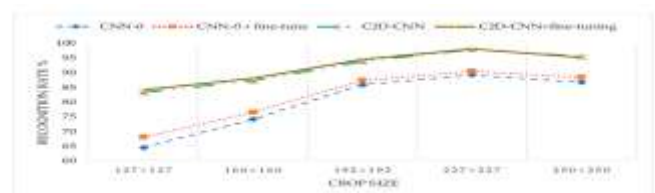


Figure 4: Conduct an analysis to see how the implementation of fine-tuning and decision-level fusion can improve the overall performance of CNN.(Li et al., 2018)

4. DISCUSSION

This part will compare the suggested C2D-CNN algorithm to other cutting-edge algorithms to show that ours is a reliable and useful one. First, we will assess the viability of the layered activation function, probabilistic max-pooling, and suggested and introduced normalizing layers, respectively. Second, we'll go back to the initial goal of the suggested innovative idea—a comparison between decision-level fusion and fine-tuning. Lastly, we will contrast the suggested algorithm with the most advanced approach. Figure 4 depicts the particular flow (Li et al., 2018).

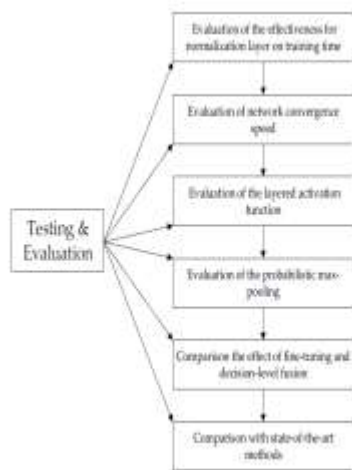


Figure 5: the flow chart of experiment(Li *et al.*, 2018)

5. CONCLUSIONS

In computer vision, face recognition (FR) is one of the major problems, especially appropriate for airport anti-terrorism applications and security surveillance. It finds applications, in domains, such, as the monitoring of video footage and supporting law enforcement efforts. The CNN model proposed in this study consists of components, including a fully connected layer, a stacked activation function layer, a probabilistic max pooling layer and a convolution layer. This technique employs a Convolutional Neural Network (CNN) to extract features by training it on extensive datasets containing various faces. Next we utilize the trained CNN model (excluding the Softmax classifier) to extract the desired features. There are ways to utilize detection. The initial approach involves utilizing features or distinctive points extracted from the image to recognize faces.

For example an algorithm examines the ratios, shapes and positioning of the nose, cheekbones and eyes. Then it will use these characteristics to search for images that share properties, by employing methods such, as support vector machines (SVM) and principal component analysis (PCA). The alternative approach is referred to as "recognition," which utilizes 3D sensors to gather data on the facial structure. This information is then utilized to identify characteristics, on the face such, as the shape of the nose. This understanding enhances the accuracy of recognition. Skin texture analysis serves as the third technique employed. This clever design utilizes the appeal of the skin. By incorporating analysis of skin texture the effectiveness of facial recognition techniques has been enhanced.

REFERENCES

- Budiman, A. *et al.* (2022) 'Student attendance with face recognition (LBPH or CNN): Systematic literature review', *Procedia Computer Science*, 216, pp. 31–38. doi: 10.1016/j.procs.2022.12.108.
- Li, J. *et al.* (2018) 'Robust face recognition using the deep C2D-CNN model based on decision-level fusion', *Sensors (Switzerland)*, 18(7), pp. 1–27. doi: 10.3390/s18072080.
- Abate, A. F. *et al.* (2007) '2D and 3D face recognition: A survey', *Pattern Recognition Letters*, 28(14), pp. 1885–1906. doi: 10.1016/j.patrec.2006.12.018.
- Abdelmaksoud, M. *et al.* (2020) 'A Novel Neural Network Method for Face Recognition with a Single Sample per Person', *IEEE Access*, 8, pp. 102212–102221. doi: 10.1109/ACCESS.2020.2999030.
- Bashbaghi, S. *et al.* (2017) 'Dynamic ensembles of exemplar-SVMs for still-to-video face recognition', *Pattern Recognition*, 69, pp. 61–81. doi: 10.1016/j.patcog.2017.04.014.
- Budiman, A. *et al.* (2022) 'Student attendance with face recognition (LBPH or CNN): Systematic literature review', *Procedia Computer Science*, 216, pp. 31–38. doi: 10.1016/j.procs.2022.12.108.
- Chowdhury, S. *et al.* (2020) 'Development of an Automatic Class Attendance System using CNN-based Face Recognition', *ETCCE 2020 - International Conference on Emerging Technology in Computing, Communication and Electronics*, pp. 2–6. doi: 10.1109/ETCCE51779.2020.9350904.
- Cook, J., Chandran, V. and Fookes, C. (no date) '1 Introduction 2 Gabor and Log-Gabor Filters', pp. 1–10.
- Damale, R. C. and Pathak, B. V. (2018) 'System Using Machine Learning Algorithms', *Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018)*, (Iciccs), pp. 414–419.
- Dewan, M. A. A. *et al.* (2016) 'Adaptive appearance model tracking for still-to-video face recognition', *Pattern Recognition*, 49, pp. 129–151. doi: 10.1016/j.patcog.2015.08.002.
- Goyal, A. *et al.* (2021) 'Online Attendance Management System Based on Face Recognition Using CNN', *SSRN Electronic Journal*, (Icicnis). doi: 10.2139/ssrn.3883841.
- Gudikandhula, N. R. A. O. and Kondapalli, B. (2023) 'Enhancing University Security: A Machine Learning and IoT Driven Face Recognition System for Surveillance and Attendance', pp. 1–11.
- Hu, G. *et al.* (2018) 'Frankenstein: Learning Deep Face Representations Using Small Data', *IEEE Transactions on Image Processing*, 27(1), pp. 293–303. doi: 10.1109/TIP.2017.2756450.
- Jagtap, V. *et al.* (2023) 'International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Biometric Face Recognition System using Deep Dream and CNN', *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE*, 2023(10s), pp. 145–154. Available at: www.ijisae.org.
- Kakadiaris, I. A. *et al.* (2017) '3D-2D face recognition with pose and illumination normalization', *Computer Vision and Image Understanding*, 154, pp. 137–151. doi: 10.1016/j.cviu.2016.04.012.
- Lal, P. V., Srilakshmi, U. and Venkateswarlu, D. (2022) 'MHA_VGG19: Multi-Head Attention with VGG19 Backbone Classifier-based Face Recognition for Real-Time Security Applications', *International Journal of Intelligent Systems and*

Applications in Engineering, 10(1s), pp. 34–44.

Li, J. *et al.* (2018) 'Robust face recognition using the deep C2D-CNN model based on decision-level fusion', *Sensors (Switzerland)*, 18(7), pp. 1–27. doi: 10.3390/s18072080.

Lu, J., Wang, G. and Zhou, J. (2017) 'Simultaneous Feature and Dictionary Learning for Image Set Based Face Recognition', *IEEE Transactions on Image Processing*, 26(8), pp. 4042–4054. doi: 10.1109/TIP.2017.2713940.

Matta, F. and Dugelay, J. L. (2009) 'Person recognition using facial video information: A state of the art', *Journal of Visual Languages and Computing*, 20(3), pp. 180–187. doi: 10.1016/j.jvlc.2009.01.002.

Oquab, M. *et al.* (2014) 'Learning and transferring mid-level image representations using convolutional neural networks', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1717–1724. doi: 10.1109/CVPR.2014.222.

Shrestha, R. and Panday, S. P. (2020) 'Face Recognition Based on Shallow Convolutional Neural Network Classifier', *ACM International Conference Proceeding Series*, pp. 25–32. doi: 10.1145/3388818.3388825.

Socolinsky, D. A., Selinger, A. and Neuheisel, J. D. (2003) 'Face recognition with visible and thermal infrared imagery', *Computer Vision and Image Understanding*, 91(1–2), pp. 72–114. doi: 10.1016/S1077-3142(03)00075-4.

Soltanpour, S., Wu, Q. M. J. and Anvaripour, M. (2015) 'Multimodal 2D-3D face recognition using structural context and pyramidal shape index', *IET Seminar Digest*, 2015(5). doi: 10.1049/ic.2015.0100.

Wang, X. *et al.* (2013) 'A new approach for 2D-3D heterogeneous face recognition', *Proceedings - 2013 IEEE International Symposium on Multimedia, ISM 2013*, pp. 301–304. doi: 10.1109/ISM.2013.58.

Zhao, W. *et al.* (1998) 'Discriminant Analysis of Principal Components for Face Recognition', *Face Recognition*, pp. 73–85. doi: 10.1007/978-3-642-72201-1_4.