FewFaceNet: A Lightweight Few-Shot Learning-based Incremental Face Authentication for Edge Cameras

Abu Sufian
University of Gour Banga
English Bazar, India
sufian@ieee.org

Anirudha Ghosh
Visva-Bharati
Santiniketan, India
ghoshanirudha141@gmail.com

Debaditya Barman
Visva-Bharati
Santiniketan, India
debadityabarman@gmail.com

Marco Leo
CNR-ISASI
73100 Lecce, Italy
marco.leo@cnr.it

Cosimo Distante
CNR-ISASI
73100 Lecce, Italy
Cosimo.distante@cnr.it

Baihua Li
Loughborough University
Loughborough, UK
B.Li@lboro.ac.uk

Abstract

Face authentication is a widely used technique for verifying identity, but current approaches encounter limitations due to their reliance on extensive computing resources, large datasets, and well-lit environments. Additionally, these approaches often lack adaptability to accommodate new individuals and continuously improve performance. These constraints make them impractical for various edge applications such as smart home security, bio-metric, surveillance system, etc. To address these challenges, this paper introduces a novel technique called FewFaceNet, which leverages a very lightweight few-shot learning-based incremental face authentication. Unlike existing methods, FewFaceNet employs a shallow lightweight backbone model that can start work with just one face image and also can handle infrared images in dark environments. These features make it highly suitable for deployment on small-edge cameras like door security cameras. We curated a diverse dataset from various reliable sources, including our own infrared camera to train and evaluate the model. Through extensive experimentation, we assessed the performance of FewFaceNet with different backbone ablation studies across one-shot to five-shot scenarios. The experimental results convincingly demonstrate the effectiveness of FewFaceNet in overcoming the limitations of existing approaches. The code and data available at: https://github.com/Sufianlab/FewFaceNet.

1. Introduction

Due to the rapid transformation of lifestyles, the trend of getting services at home or apartment is becoming increasingly common[1]. Different service personnel such as delivery personnel, electricians, plumbers, technicians, housekeepers frequently visits consumer's home or apartment. Consequently, there is a growing need for security checks and enhanced authentication at the doors of houses or apartments [13]. Automated face authentication at doors shall be a solution, and this could be done in collaboration with respective service providers such as online shopping platforms, food delivery applications, home appliance maintenance agencies, etc., who send their service personnel to consumers' homes as a scenario depicted in Figure 1.

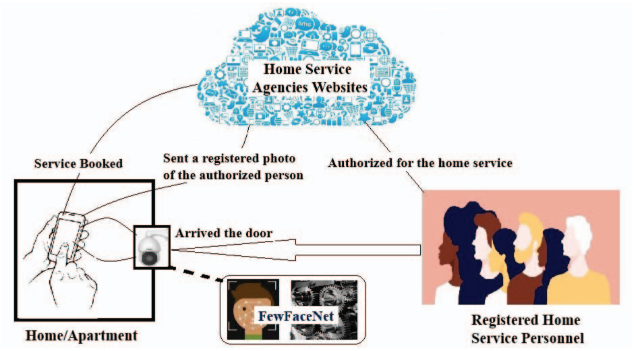

Figure 1. A working scenario of proposed FewFaceNet.

While numerous automated face authentication methods have been proposed [26, 4, 10], most of them primarily rely on large databases, a predefined number of classes, high computing resources, and well-illuminated environments, also scopes of continuous performance improvement were not considered. Moreover, as the field of security check-

[1] https://www.grandviewresearch.com/industry-analysis/online-on-demand-home-services-market-report.

ing advances, particularly in relation to face authentication mechanisms at doors, several important factors need to be considered. Firstly, it is not practical the creation large databases containing multiple images of home service personnel. Secondly, since the number of on-demand home services as well as the number of personnel associated with these businesses are increasing day by day, the number of classes in the system cannot be fixed. Thirdly, maintaining a well-lit environment at doors is neither practical nor convenient. Fourthly, it is essential to continuously enhance the performance of the face authentication model through feedback mechanisms. Last but not least, it is important to acknowledge that not every household has access to high-end computing devices or cloud services, making affordability a key consideration.

Developing a useful face authentication system at doors should address the above challenges. Therefore, we propose FewFaceNet, a technique to deal with such challenges. Figure 1 illustrates an application scenario of FewFaceNet through a door security camera. When a consumer book a home service, they will receive an image of the authorized service personnel. The FewFaceNet algorithm-based system will authenticate the in-person visit of the authorized service personnel with that image and door camera.

We aim to develop a practical solution to the challenges encountered in face authentication systems used in home security solutions. FewFaceNet is designed using few-shot learning, operates in low-light environments, and its suitability for small edge cameras makes it a promising advancement in face authentication techniques. FewFaceNet has been extensively studied including ablation studies with the following key contributions:

1. **Novel Lightweight Backbone Model**: The FewFaceNet model integrates a novel backbone network, enabling resource-friendly execution. The model consists three parallel branches of five layers. The proposed backbone model contains only 1.3 million parameters (Only 12% compare to classical ResNet 18 [14]). This feature is particularly beneficial for small-edge devices such as door security cameras.

2. **Incremental few-shot approach**: The proposed FewFaceNet model is capable of working with a single image, leveraging the state-of-the-art Siamese network [7, 29]. Our experiment considered one to five shots to asses the incremental performance improvement but compromising the response time of the authentication.

3. **Work with dynamic datasets**: Initially, the support set contains a single image of the incoming service person sourced from the agency's online platform. As the person arrives, additional images can be included in the support set through the door security camera once that query image is successfully authenticated. For new service personnel, a new class will be created.

4. **New query dataset**: We develop two different types of working query datasets. One is by taking RGB images from several reliable sources. Second, we created a dataset containing images captured by an infrared camera with the help of a group of volunteers to assess its suitability to work in low-light environments.

The rest of the paper is organized as follows: Section 2 presents a brief literature review. We present the proposed methodology in Section 3. Experimental results can be found in Section 4. Section 5 presents the discussion as well as the future scopes of our experiment. Finally, we conclude in Section 6.

2. Literature Review

Face recognition and authentication have recently emerged as prominent applications of artificial intelligence, particularly within computer vision and image processing domains. The literature in this field is extensive and constantly evolving. In this section, we highlighted some recent works that are relevant to the proposed model.

One popular approach is the eigenfaces-based method, which utilizes Principal Component Analysis (PCA) to extract significant facial variations from datasets. However, recent eigenfaces-based methods [49, 32, 39] are yet to overcome the challenges of handling variations in lighting conditions. Another popular technique is Fisherfaces which employs Linear Discriminant Analysis (LDA) to find a low-dimensional subspace that maximizes the ratio of between-class to within-class scatter. However, like previous methods, recent Fisherfaces-based methods [3, 36] are also sensitive to lighting variations. Texture-based approaches such as Local Binary Patterns (LBP) have also been widely used in face recognition methods [11, 45, 20]. LBP encodes local pixel comparisons to represent facial patterns but has limited capability to capture global spatial information. Gabor wavelets-based methods [25, 24, 1] leverage Gabor filters and these methods are suitable for faces with different orientations and scales. However, the computational complexity of these methods is high.

Deep Learning (DL) methods gained popularity due to their capabilities and data-driven facilities. Convolutional Neural Network (CNN)-based methods [17, 50, 10] and Generative Adversarial Network (GAN)-based methods [51, 19] have been extensively used for face recognition and authentication. Several specialized DL approaches for face recognition like FaceNet [42], ArcFace [9], and DeepFace [37] have also been proposed [47, 35, 12]. Although deep learning-based methods perform very well, these methods necessitate substantial amounts of data and significant com-

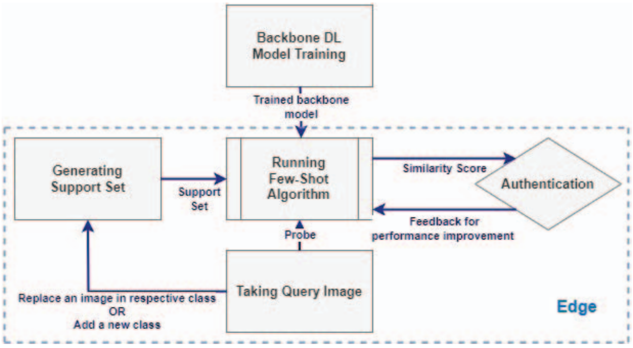
Backbone DL Model Training
Trained backbone model
Generating Support Set
Support Set
Running Few-Shot Algorithm
Similarity Score
Authentication
Feedback for performance improvement
Probe
Taking Query Image
Replace an image in respective class OR Add a new class
Edge

Figure 2. The working procedure of FewFaceNet.

puting power, making their deployment on edge devices challenging.

Researchers also explored various contemporary and hybrid approaches. For instance, methods based on 3D data [27, 15], infrared data [23, 2], multimodal fusion [46, 6], etc. Transformer models [30, 44] have also been utilized for face recognition. Recent contemporary methods include elastic margin loss-based deep face recognition [5], spherical confidence learning [28], universal representation and quality assessment [34], and quality adaptive margin [21]. Some methods combine traditional and DL approaches for improved recognition [45], while ensemble techniques have been used in others [8, 31] to get combined strength of different models. Recently, meta-learning and few-shot learning methods based such as Siamese Networks have been studied for face recognition and authentication [16, 43, 48].

The number of research works based on door-based face authentication or recognition is limited, with only a few related works proposed in [40, 33, 52, 41, 38]. However, these works primarily concentrate on traditional bio-metric applications, intelligent door lock systems, etc.

To the best of our knowledge and on existing literature review, no prior research has introduced an incremental few-shot learning technique specifically designed for face authentication or recognition that effectively operates with just a single image as input and which is suitable for edges.

3. Proposed Methodology

The operation of FewFaceNet, as illustrated in Figure 4, begins with the design and training of our proposed ensemble lightweight DL-based backbone network. This backbone network is trained using our curated dataset of facial images. We experimented with different backbone models including ablation studies towards the proposed novel lightweight backbone model. Subsequently, the trained model is utilized to develop the proposed FewFaceNet, which leverages few-shot learning techniques. During the authentication process, FewFaceNet compares the similarity between a query image and the support images.

3.1. Dataset

3.1.1 Data Collection

Training Dataset:
We compile the training dataset for the proposed FewFaceNet model by taking data from three different sources. Each of them is briefly described below. First of all, to optimize the training process for meta-learning, we refine the dataset by selecting a maximum of twenty images for each class and discarding classes with fewer than three images.
1) **Labelled Faces in the Wild (LFW) Dataset**: This dataset[2] was created and maintained by a group of researchers at the University of Massachusetts, Amherst. It contains 13,233 facial images of 5,749 individuals. The images were collected online and processed using the Viola-Jones face detector. Notably, 1,680 individuals have multiple distinct photos in the dataset.
2) **Pins Face Recognition**: This dataset[3] contains facial images collected from a social media platform Pinterest[4] and cropped for face recognition purposes. It includes 105 celebrities and a total of 17,534 faces.
3) **The ORL Database of Faces**: The ORL Database of Faces[5] was utilized in a face recognition project conducted in collaboration with the speech, vision, and robotics Group of the Cambridge University Engineering Department. This database contains ten different images of each of the 40 distinct subjects. The images were captured under varying conditions, such as different lighting, facial expressions (open or closed eyes, smiling or not smiling), and facial details (with or without glasses). All images were taken against a dark homogeneous background, with subjects positioned upright and frontal, allowing for slight side movement.
Test Dataset:
We use two test datasets to evaluate the proposed FewFaceNet technique in our experiment.
Test dataset 1: This dataset is constructed from the same distributions as the training set, but the splitting is performed prior to training. For our experiment, we consider 10 classes ranging from one shot to five shots. In each class, we utilize 20 images to obtain average authentication scores.
Test dataset 2: Our infrared dataset:
In addition to the previously mentioned sources, we created a test dataset consisting of infrared facial images. We selected a small group of cohorts consisting of 10 volunteers participating in the data collection process. Continu-

[2] https://www.kaggle.com/datasets/jessicali9530/lfw-dataset
[3] https://www.kaggle.com/datasets/hereisburak/pins-face-recognition
[4] pinterest.com
[5] https://www.kaggle.com/datasets/tavarez/the-orl-database-for-training-and-testing

ous video recordings were captured using an infrared night vision-based home security camera in a dark environment. Subsequently, we utilized OpenCV to extract facial images from the video feeds, resulting in the construction of this test dataset. We did not use these images for training the model to understand its generalizability without retraining.

3.1.2 Pre-processing of data

As mentioned, we compile three datasets from three different sources to create our training dataset. During training, we follow the following pre-processing steps to normalize our training samples before feeding them to the model:

1) Transform the raw images into grayscale images. 2) Resize all images to a width and height of 100 pixels, ensuring uniform dimensions. Additionally, to enhance the model's prediction ability with night vision images from our infrared dataset, we applied Adaptive Histogram Equalization (AHE) during the testing phase. The utilization of AHE noticeably improved the model's performance. AHE redistributes the intensity values in the image, resulting in improved visibility of details that may be concealed in regions with low contrast. This technique brings out delicate structures, edges, and other crucial features that are vital for night vision applications.

To address the problem of data scarcity, we employ data augmentation techniques. These techniques generated modified versions of existing images, effectively increasing the size of the training dataset.

We utilize the following augmentation techniques: 1) Randomly rotate the image by a specified angle within the range of ± 15 degrees. 2) Randomly flip the image horizontally with a 50% probability, introducing diversity in the training data. 3) Randomly crop and resize the image, with a crop size of 100×100 pixels and a scale range from 80% to 100% of the original image size. 4) Apply random changes to the image's color, including brightness, contrast, saturation, and hue. The specified values determine the range of variation for each attribute.

3.2. Architecture of the Backbone Model

We encounter various challenges while using the classical Siamese network [22]. In addition to that, for a deep model, the cosine distance for positive and negative is becomes small, so, chances becomes very high to either overfit or collapsed. Thus we develop a novel architecture to overcome those issues. The key feature of our model is a shallow tree-like structure as shown in Figure 3 that offers several benefits, such as: 1). Very lightweight, **only 1.3 M parameters** which is highly suitable for edge camera. 2). Increased data modeling capacity enabling the network to learn more complex and nuanced relationships in input images. 3). Hierarchical feature extraction allows the network to capture fine-grained and high-level features, leading to more comprehensive representations of the inputs. 4). The hierarchical nature of this architecture also allows shared computations and parameter sharing at different levels, reducing redundancy, improving computational efficiency, and enhancing the network's discriminative power.

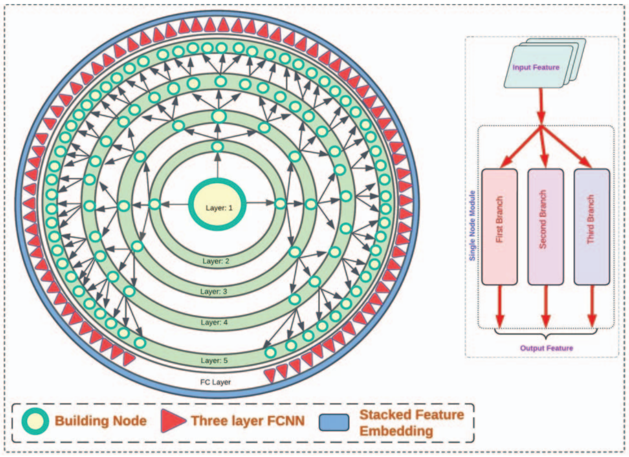

Figure 3. Architecture of the Backbone Model.

The tree-like parallel subnetworks allow multiple pathways for gradient flow. As gradients propagate through different subnetworks, they encounter different sets of parameters and operations. This path diversification helps to avoid **vanishing or exploding gradients** because even if one pathway suffers from these issues, other pathways may still carry meaningful gradients.

Our model's core building blocks consist of multiple layers of node modules, which collectively form a compact ensemble model. Each node contains three parallel branches, each inspired by different state-of-the-art architectures. Each layer in the model's architecture consists of 3^i node modules, where i represents the layer number. This allows us to create a model with the exponential growth of node modules with layers.

The first branch comprises a single 3×3 convolutional layer and a ReLU activation unit. The second branch draws inspiration from the ResNet architecture [14]. Here, we addresses the vanishing gradient problem through identity shortcuts. It begins with a 3×3 convolution, followed by batch normalization and a ReLU activation. The output is then convoluted sequentially through two lightweight residual blocks. Finally, the extracted feature maps of those blocks was further convoluted using a 3×3 convolution, batch normalized and passed through a ReLU activation. Each Lightweight ResNet block comprises a single 3×3 convolutional layer followed by batch normalization and a ReLU activation unit.

The architecture of the DenseNet [18] inspires us to de-

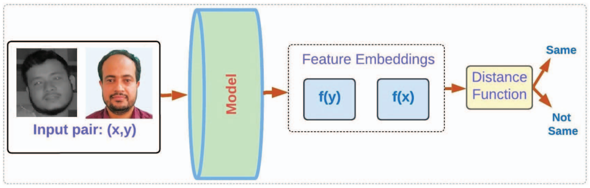

Figure 4. Block Diagram of Siamese Network based FewFaceNet.

sign of the last branch. Here, we utilizes dense connections, where each layer in a dense block receives feature maps from all preceding layers. This dense connectivity allows for direct information flow and feature reuse, enabling efficient learning from different scales and abstraction levels. This branch includes three lightweight dense blocks, each comprising one 1×1 convolution and one 3×3 convolution layer sequentially. Inside each block, the inputted feature maps are passed through batch normalization and a ReLU activation before passing to each convolution layer. Before passing the features into a dense block, we apply a 3×3 convolution to reduce the dimensions of the feature maps. The output of the dense block is then processed through another 3×3 convolution layer to reduce the number of channels.

Finally, the last layer of our model's architecture consists of a three-layer fully connected (FC) network. Each FC network computes similarity scores based on the features passed from each branch of the previous layer's Node modules.

3.3. Design of FewFaceNet

FewFaceNet utilizes few-shot learning (FSL), an ML technique enabling models to learn new tasks from a few examples. FSL addresses data scarcity and enhances generalization, making it ideal for real-world applications with limited data. Moreover, FSL can be implemented as incremental learning by updating a pre-trained model with new concepts. The model is fine-tuned using a few samples for the new task while retaining knowledge from previous tasks. This enables continuous learning and adaptation without retraining.

Incremental FSL is well-suited for face authentication applications as it allows the model to continuously learn and recognize new faces without retraining on the entire dataset. By updating the model with one image of each new face, it can incrementally expand its face recognition capabilities while keeping knowledge of previously learned faces.

We utilize the Siamese network, an FSL approach in this work, which comprises two identical subnetworks with matching weights and architecture. These subnetworks are commonly referred to as *twin networks* or *siamese twins*. The main objective of a Siamese network is to determine the similarity or dissimilarity between two input samples. Algorithm for training a Siamese network is below:

1. Initialization: $\theta_1 = \theta_2$ (Initialize the Siamese network with shared weights parameters: $w_1, w_2, \ldots, w_n$ and Bias parameters: $b_1, b_2, \ldots, b_m$)

2. Distance Metric: $D(x_1, x_2) = \|f(x_1) - f(x_2)\|$ (Compute the distance metric using the outputs of the subnetworks, $f(x_1)$ and $f(x_2)$ represents feature embeddings of x_1 and x_2 respectively.)

3. Training Dataset: $\{(x_1, x_2, y)\}$ (Training dataset consisting of input pairs (x_1, x_2) and their labels y)

4. Shuffle Dataset: Shuffle the training dataset during training

5. Training Loop:

   (a) For each training example (x_1, x_2, y):

      i. Forward Pass: $f_1 = f(x_1), f_2 = f(x_2)$ (Compute the output feature embeddings of the subnetworks)

      ii. Distance Calculation: $D(x_1, x_2) = \|f_1 - f_2\|$ (Compute the distance metric)

      iii. Loss Calculation: $L(x_1, x_2, y)$ (Compute the loss based on the distance and the ground truth)

      iv. Backpropagation: $\theta = \theta - \alpha \nabla L(x_1, x_2, y)$ (Update the weights and parameters using gradient descent rule)

6. Repeat the iterations for sufficient epochs or until the model convergence.

The network takes a pair of input images $\{(x_1, x_2)\}$ and processes them through the twin subnetworks, which share the same architecture and weights θ. The outputs of these shared subnetworks are then concatenated or combined to form a single feature vector $f(x_1), f(x_2)$, representing the embedded representation of the input pair. These feature vectors are subsequently used to calculate a similarity or dissimilarity metric $L(x_1, x_2, y)$ between the input images. Common distance metrics include Euclidean distance, cosine similarity, or contrastive loss are used.

3.4. Training Details

Our experiment utilizes a meta-learning approach to recognize faces with minimal support images. During training, the network employs these distance metrics to calculate the loss against known similarity or dissimilarity labels. The loss is then back-propagated through the network to update the shared weights and optimize its performance. Meta-learning with Siamese networks typically involves two phases: the meta-training phase and the meta-testing phase.

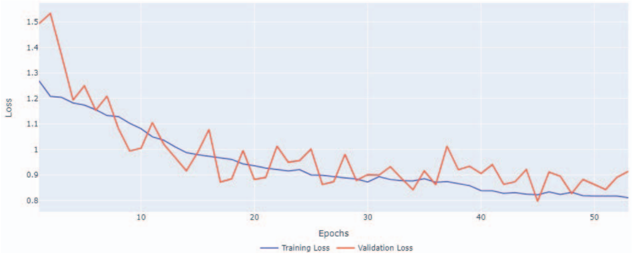

Figure 5. Number of epoch vs. training and validation loss.

In the meta-training phase, our model is trained to learn a generalizable representation that can quickly adapt to new tasks or data. For this phase, we utilize a training dataset consisting of 1036 classes of samples. The inputs are in the form of pairs of two images, and the corresponding label represents the dissimilarity score between the pair (0 if they belong to the same class, otherwise 1). Additionally, we incorporate 10% dropout for regularization, employ the Adam optimizer with an initial learning rate of 0.009, and utilize cross-entropy loss to adjust model weights and optimize the similarity/dissimilarity predictions based on the true labels. The cross-entropy loss equation for a pair of inputs in a Siamese network as below:

$$L(\hat{y}, y) = -(ylog(\hat{y}) + (1 - y)\log(1 - \hat{y}))$$

where $\hat{y}$ represents the similarity/dissimilarity score, and y denotes the true label for the pair. The $\hat{y}$ value is calculated by finding the Euclidean distance between the embedding of the input pairs. How the training and validation loss decreases with an increasing number of epochs is depicted in Figure 5.

While training, we use the early stopping technique as the convergence criteria in our experiment; for this, we use a separate validation set to monitor the model's performance metrics with the subsequent training epochs. The validation set is created by collecting 236 separate classes from the same domain as the training dataset. Figure 5 shows the training and validation loss against the training epochs. There we can notice that the decreasing rate of validation loss was very high during the initial epochs, and it gradually became very low with subsequent epochs. We also noticed that the model began to overfit after approximately 50 epochs. Therefore, we decided to early stop the training process at that point. At that point training loss was around 0.82 and validation loss was around 0.85.

4. Model Evaluation and Results

In the testing phase, the trained model is evaluated on new tasks to assess its ability to quickly adapt and generalize from a small amount of labeled data. To evaluate our model's performance, we employ two different datasets for testing. The first dataset consists of 10 classes collected from the same domain as the training dataset, and the second dataset comprises 10 classes of infrared facial images taken in a dark environment using our infrared camera. Detailed discussion could be found in the Section 3.1 namely as Test datast-1 and Test dataset-2.

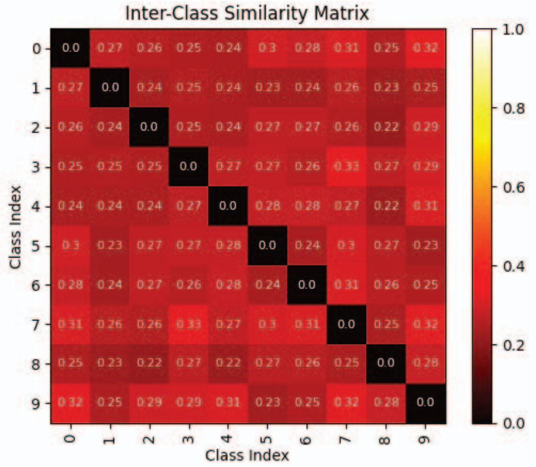

Figure 6. Representing the inter-class similarity of test dataset 1.

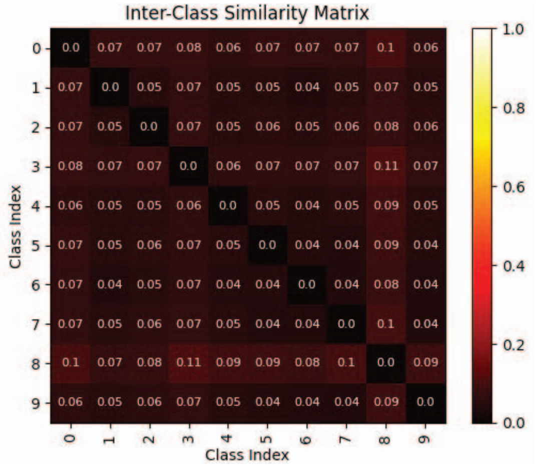

Figure 7. Representing the inter-class similarity of test dataset 2.

The first test dataset is designed to have inter-class similarity following a normal distribution, representing a typical scenario. In contrast, the second test dataset is specifically created to maintain a high inter-class similarity, representing a more challenging scenario. By experimenting with these two datasets, we aim to comprehensively assess the model's performance across different levels of inter-class similarity. Figure 6 illustrates the inter-class similarity of the first test dataset whereas Figure 7 illustrates the inter-class similarity of the second test dataset. While calculating similarity scores, we use the cosine distance between two feature vectors.

The performance metrics, including the true authentication rate (TAR), false authentication rate (FAR), false un-

		One-shot	Two-shot	Three-shot	four-shot	Five-shot
Ablation 1	TAR	48.50 ± 0.00	70.68 ± 0.00	51.93 ± 0.00	62.69 ± 0.00	53.98 ± 0.00
	TUR	83.50 ± 0.80	70.05 ± 1.80	81.32 ± 1.00	77.67 ± 1.40	83.31 ± 1.60
	FUR	51.50 ± 0.00	29.31 ± 0.00	48.06 ± 0.00	37.20 ± 0.00	46.01 ± 0.00
	FAR	16.50 ± 0.80	29.94 ± 1.80	18.67 ± 1.00	22.32 ± 1.40	16.68 ± 1.60
Ablation 2	TAR	72.00 ± 0.00	79.58 ± 0.00	74.58 ± 0.00	81.97 ± 0.00	69.32 ± 0.00
	TUR	33.10 ± 0.70	27.53 ± 0.02	36.90 ± 1.70	28.95 ± 1.30	40.00 ± 1.12
	FUR	28.00 ± 0.00	20.41 ± 0.00	25.41 ± 0.00	18.02 ± 0.00	30.67 ± 0.00
	FAR	66.89 ± 0.70	72.46 ± 0.20	63.09 ± 1.70	71.04 ± 1.30	60.00 ± 1.12
Ablation 3	TAR	37.00 ± 0.00	59.68 ± 0.00	39.77 ± 0.00	53.48 ± 0.00	38.65 ± 0.00
	TUR	84.40 ± 0.70	69.52 ± 0.60	88.17 ± 1.40	81.51 ± 1.50	89.69 ± 2.10
	FUR	63.00 ± 0.00	40.31 ± 0.00	60.22 ± 0.00	46.51 ± 0.00	61.34 ± 0.00
	FAR	15.60 ± 0.70	30.47 ± 0.60	11.82 ± 1.40	18.48 ± 1.50	10.30 ± 2.10
FewFaceNet	TAR	70.50 ± 0.00	80.10 ± 0.00	75.69 ± 0.00	78.48 ± 0.00	74.84 ± 0.00
	TUR	66.70 ± 2.70	63.87 ± 0.50	71.71 ± 1.20	71.97 ± 0.85	79.01 ± 1.50
	FUR	29.50 ± 0.00	19.89 ± 0.00	24.30 ± 0.00	21.51 ± 0.00	25.15 ± 0.00
	FAR	33.29 ± 2.76	36.12 ± 0.50	28.28 ± 1.27	28.02 ± 0.85	20.98 ± 1.50

Table 1. Mean Values with standard deviations of TUR, TAR, FUR, and FAR on test dataset-1 across different shot scenarios.

		One-shot	Two-shot	Three-shot	four-shot	Five-shot
FewFaceNet	TAR	55.00 ±0.00	63.68 ±0.00	56.35 ±0.00	59.06 ±0.00	51.85 ±0.00
	TUR	74.60 ±0.70	66.52 ±0.70	75.80 ±1.40	75.20 ±0.87	84.69 ±0.70
	FUR	45.00 ±0.00	36.31 ±0.00	43.64 ±0.00	40.9 ±0.00	48.14 ±0.00
	FAR	25.40 ±0.70	33.47 ±0.70	24.19 ±1.40	24.79 ±0.87	15.30 ±0.70

Table 2. Mean Values with standard deviations of TUR, TAR, FUR, and FAR on test dataset-2 across different shot scenarios.

authentication rate (FUR), and true un-authentication rate (TUR) are computed using the following formulations:

$$TAR = TN/(FP + TN)$$

$$FUR = FN/(FN + TP)$$

$$FAR = FP/(FP + TN), \text{ and}$$

$$TUR = TP/(TP + FN)$$

These metrics are observed for each one-shot, two-shot, three-shot, four-shot, and five-shot scenarios. Table 1 displays the values for the mentioned performance metrics, illustrating the results obtained by deploying three ablation models and the proposed FewFaceNet model on test dataset 1 under various shot scenarios. Here, Ablation 1 represents two branches of FewFaceNet except the traditional CNN branch. Ablation 2 represents two branches of FewFaceNet except the ResNet-based branch. Ablation 3 represents two branches of FewFaceNet except the DenseNet-based branch. In Table 2, the results obtained using the proposed model, FewFaceNet, exclusively on test dataset 2 are presented for different shot scenarios.

Figure 8 and Figure 9 present the authentication capability of FewFaceNet for different samples from test dataset-1 and test dataset-2 respectively with varying shot accuracy (i.e. one, two, three, four, and five-shot). These metrics give a comprehensive insight into different aspects of the model's performance, such as its ability to identify negative and positive cases correctly and its tendency to make false positive and false negative errors.

In the one-shot authentication, when each class consists of one support image and 20 query images; we generate 20 positive pairs (by combining the support image with each query image) and an equal number of negative pairs (by combining the support image with negative images) for testing. The negative images are generated by selecting random images from the dataset, except the support class's images. This experimental setup provides 20 positive and 20 negative simulated results. Then we transfer one previously successfully authenticated image from the query set to the support set for the two-shot authentication. This process will continue for two-shot to three-shot authentication, three-shot to four-shot authentication, and four-shot to five-shot authentication.

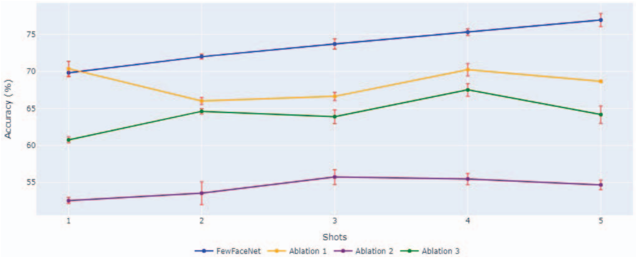

Figure 8. Mean accuracy of FewFaceNet with ablation study in different shot scenarios on test dataset-1.

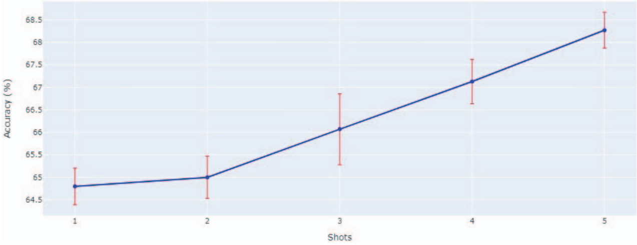

Figure 9. Mean accurcy of FewFaceNet in different shot scenarios on test dataset-2.

During our experiment, threshold values of 0.82 and 0.92 are used for test dataset-1 and test dataset-2, respectively. Since inter-class similarity for test dataset-2 is relatively high than that for test dataset-1, we set higher threshold value for test dataset-2.

When the similarity score between the query and support set sample is in unison with at least the threshold value, we classified the query image as belonging to the class of the supported image. For two, three, four, and five-shot classifications, if the number of support images with a similarity score equal to or higher than the threshold was greater than or equal to $\lfloor(shot + 1)/2\rfloor$, we classified the query image to the corresponding support class.

5. Discussion and Future Scopes

The performance of the proposed FewFaceNet is thoroughly evaluated through extensive experimentation, and the results unequivocally demonstrate its effectiveness. Several metrics are used to assess the model's performance.

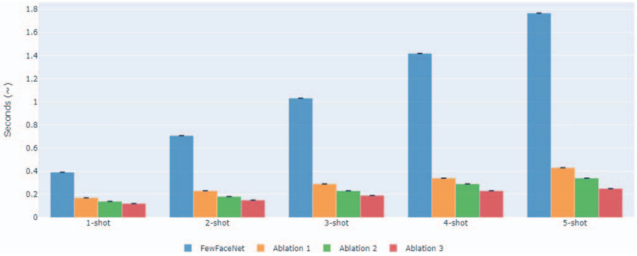

Figure 10. Response time vs. Number shot in different scenarios.

The accuracy on the normal light test dataset and low light infrared dataset are depicted in Figure 8 and Figure 9, respectively. Figure 8 also present the results of ablation study of the proposed backbone model. Based on ablation studies, the results suggest that the proposed FewFaceNet is stable. However, when the ResNet-based branch is removed, performance is worse, as indicated by the purple line in the graph. On the other hand, combining ResNet and DenseNet leads to improved performance, as shown by the orange line in the graph.

However, the graph of these two figures clearly shows improvement from one-shot to five-shot scenarios, albeit with a slight increase in time cost as depicted in Figure 10. Response time increases each and every possible model. Time taken by proposed model is little high compared to it's ablation models but it is more stable and give higher performances. Therefore, if the authentication are not required real time then number of shot could be increases as incremental learning. This allows the model to progressively improve its performance and generalize better to new classes or categories with minimal additional training data. Additionally, the means and standard deviations as presented in Table 4, are observed to be minimal which also indicating the stability of the proposed backbone model.

The performance of FewFaceNet is also assessed on infrared images, specifically examining its capability to operate effectively in low-light environments. The inter-class similarity matrix (shown in Figure 7 of this dataset shows high similarity between classes, which lead performance degradation. But the proposed FewFaceNet demonstrates satisfactory performance, albeit slightly lower compared to images in normal lighting conditions. However, the authentication rate of the model remains reliable, as depicted in Figure 9.

Based on the findings of this investigation, three immediate future research directions have been identified:

1) **Evaluation with a broader range of classes using a more diverse dataset**: To enhance the generalizability of the proposed FewFaceNet algorithm, it is crucial to assess its performance with an expanded set of classes and a diverse datasets. This assessment will provide insights into the algorithm's adaptability and its ability to handle more complex practical scenarios.

2) **Implementation with a real edge camera setup for authentication evaluation**: In order to validate the practical applicability of the FewFaceNet algorithm, it is very important to implement it in a real-world edge camera environment. This setup will allow for authenticating actual users and assessing the algorithm's performance under realistic conditions. By conducting such experiments, we can measure the algorithm's accuracy, reliability, and potential limitations in real-world scenarios.

3) **Experimented with end-users-in-the-loop**: While we have experimented with an infrared low-light dataset collected from a cohort, but actual studies with end users have not been conducted yet. We plan to address this in the future as part of the deployment study.

By adopting these future scopes, we committed to further refine and validate the FewFaceNet algorithm, enabling its potential deployment as a practical authentication system through edge cameras.

6. Conclusion

This paper introduced FewFaceNet, a lightweight incremental few-shot learning technique designed for face authentication suitable for deployment on edge camera, especially door security cameras. Extensive experimentation across different shot scenarios, ranging from one shot to five shots along with ablation studies, showcases the effectiveness and robustness of FewFaceNet, even in challenging low-light environments. FewFaceNet offers low resource consumption, efficiency, adaptability, and quick authentication, addressing the requirements of current automated face authentication systems.

Future research includes optimizing FewFaceNet, exploring additional datasets, evaluating performance under diverse conditions, implementing it with edge cameras and experimentation with end users.

In conclusion, this study establishes FewFaceNet as a promising lightweight incremental few-shot learning for face authentication suitable at edge cameras. Its performance and potentiality for real-world applications such as authentication at doors make it valuable for advancing face authentication technology.

Acknowledgment

This research was partially supported by Future Artificial Intelligence Research–FAIR CUP B53C220036 30006 grant number PE0000013.

References

[1] Sulayman Ahmed, Mondher Frikha, Taha Darwassh Hanawy Hussein, and Javad Rahebi. Optimum feature selection with

particle swarm optimization to face recognition system using gabor wavelet transform and deep learning. *BioMed Research International*, 2021:1–13, 2021. 2

[2] Muhammad Eka Setio Aji, Rocky Alfanz, and Esa Prakasa. Infrared image analysis for human face recognition. In *2022 5th International Conference on Computing and Informatics (ICCI)*, pages 157–162. IEEE, 2022. 3

[3] Mustamin Anggo and La Arapu. Face recognition using fisherface method. In *Journal of Physics: Conference Series*, volume 1028, page 012119. IOP Publishing, 2018. 2

[4] Shahina Anwarul and Susheela Dahiya. A comprehensive review on face recognition methods and factors affecting facial recognition accuracy. *Proceedings of ICRIC 2019: Recent Innovations in Computing*, pages 495–514, 2020. 1

[5] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1578–1587, 2022. 3

[6] Usman Cheema and Seungbin Moon. Sejong face database: A multi-modal disguise face database. *Computer Vision and Image Understanding*, 208:103218, 2021. 3

[7] Davide Chicco. Siamese neural networks: An overview. *Artificial neural networks*, pages 73–94, 2021. 2

[8] Jae Young Choi and Bumshik Lee. Ensemble of deep convolutional neural networks with gabor face representations for face recognition. *IEEE Transactions on Image Processing*, 29:3270–3281, 2019. 3

[9] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019. 2

[10] Hang Du, Hailin Shi, Dan Zeng, Xiao-Ping Zhang, and Tao Mei. The elements of end-to-end deep face recognition: A survey of recent advances. *ACM Computing Surveys (CSUR)*, 54(10s):1–42, 2022. 1, 2

[11] Shamsul J Elias, Shahirah Mohamed Hatim, Nur Anisah Hassan, Lily Marlia Abd Latif, R Badlishah Ahmad, Mohamad Yusof Darus, and Ahmad Zambri Shahuddin. Face recognition attendance system using local binary pattern (lbp). *Bulletin of Electrical Engineering and Informatics*, 8(1):239–245, 2019. 2

[12] Andrian Firmansyah, Tien Fabrianti Kusumasari, and Ekky Novriza Alam. Comparison of face recognition accuracy of arcface, facenet and facenet512 models on deepface framework. In *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, pages 535–539. IEEE, 2023. 2

[13] Badis Hammi, Sherali Zeadally, Rida Khatoun, and Jamel Nebhen. Survey on smart homes: Vulnerabilities, risks, and countermeasures. *Computers & Security*, 117:102677, 2022. 1

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2, 4

[15] Mingjie He, Jie Zhang, Shiguang Shan, and Xilin Chen. Enhancing face recognition with self-supervised 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4062–4071, 2022. 3

[16] Ashwamegha Holkar, Rahee Walambe, and Ketan Kotecha. Few-shot learning for face recognition in the presence of image discrepancies for limited multi-class datasets. *Image and Vision Computing*, 120:104420, 2022. 3

[17] Guosheng Hu, Yongxin Yang, Dong Yi, Josef Kittler, William Christmas, Stan Z Li, and Timothy Hospedales. When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 142–150, 2015. 2

[18] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 4

[19] Seyed Mehdi Iranmanesh, Benjamin Riggan, Shuowen Hu, and Nasser M Nasrabadi. Coupled generative adversarial network for heterogeneous face recognition. *Image and Vision Computing*, 94:103861, 2020. 2

[20] Shekhar Karanwal and Manoj Diwakar. Od-lbp: Orthogonal difference-local binary pattern for face recognition. *Digital Signal Processing*, 110:102948, 2021. 2

[21] Minchul Kim, Anil K Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18750–18759, 2022. 3

[22] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2. Lille, 2015. 4

[23] Marcin Kopaczka, Jan Nestler, and Dorit Merhof. Face detection in thermal infrared images: A comparison of algorithm-and machine-learning-based approaches. In *Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium, September 18-21, 2017, Proceedings 18*, pages 518–529. Springer, 2017. 3

[24] Chaorong Li, Yuanyuan Huang, Wei Huang, and Fengqing Qin. Learning features from covariance matrix of gabor wavelet for face recognition under adverse conditions. *Pattern Recognition*, 119:108085, 2021. 2

[25] Chaorong Li, Yuanyuan Huang, and Yu Xue. Dependence structure of gabor wavelets based on copula for face recognition. *Expert Systems with Applications*, 137:453–470, 2019. 2

[26] Lixiang Li, Xiaohui Mu, Siying Li, and Haipeng Peng. A review of face recognition technology. *IEEE access*, 8:139110–139120, 2020. 1

[27] Menghan Li, Bin Huang, and Guohui Tian. A comprehensive survey on 3d face recognition methods. *Engineering Applications of Artificial Intelligence*, 110:104669, 2022. 3

[28] Shen Li, Jianqing Xu, Xiaqing Xu, Pengcheng Shen, Shaoxin Li, and Bryan Hooi. Spherical confidence learning for face recognition. In *Proceedings of the IEEE/CVF Con-*

ference on Computer Vision and Pattern Recognition, pages 15629–15637, 2021. 3

[29] Yikai Li, CL Philip Chen, and Tong Zhang. A survey on siamese network: Methodologies, applications, and opportunities. *IEEE Transactions on Artificial Intelligence*, 3(6):994–1014, 2022. 2

[30] Mandi Luo, Haoxue Wu, Huaibo Huang, Weizan He, and Ran He. Memory-modulated transformer network for heterogeneous face recognition. *IEEE Transactions on Information Forensics and Security*, 17:2095–2109, 2022. 3

[31] Zhuo Ma, Yang Liu, Ximeng Liu, Jianfeng Ma, and Kui Ren. Lightweight privacy-preserving ensemble classification for face recognition. *IEEE Internet of Things Journal*, 6(3):5778–5790, 2019. 3

[32] Alina L Machidon, Octavian M Machidon, and Petre L Ogrutan. Face recognition using eigenfaces, geometrical pca approximation and neural networks. In *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, pages 80–83. IEEE, 2019. 2

[33] Lukas Mecke, Ken Pfeuffer, Sarah Prange, and Florian Alt. Open sesame! user perception of physical, biometric, and behavioural authentication concepts to open doors. In *Proceedings of the 17th international conference on mobile and ubiquitous multimedia*, pages 153–159, 2018. 3

[34] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14225–14234, 2021. 3

[35] David Montero, Marcos Nieto, Peter Leskovsky, and Naiara Aginako. Boosting masked face recognition with multi-task arcface. In *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 184–189. IEEE, 2022. 2

[36] Bauyrzhan Omarov, Batyrkhan Omarov, Shirinkyz Shekerbekova, Farida Gusmanova, Nurzhamal Oshanova, Alua Sarbasova, Zhanna Yessengaliyeva, Agyn Bedelbayev, Akmarzhan Maikhanova, Nurzhan Omarov, et al. Applying face recognition in video surveillance security systems. In *Software Technology: Methods and Tools: 51st International Conference, TOOLS 2019, Innopolis, Russia, October 15–17, 2019, Proceedings 51*, pages 271–280. Springer, 2019. 2

[37] Omkar Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In *BMVC 2015-Proceedings of the British Machine Vision Conference 2015*. British Machine Vision Association, 2015. 2

[38] Manik Rakhra, Dalwinder Singh, Arun Singh, Kamal Deep Garg, and Deepa Gupta. Face recognition with smart security system. In *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, pages 1–6. IEEE, 2022. 3

[39] Rika Rosnelly, Mutiara S Simanjuntak, Ade Clinton Sitepu, Mulkan Azhari, Sandy Kosasi, et al. Face recognition using eigenface algorithm on laptop camera. In *2020 8th International Conference on Cyber and IT Service Management (CITSM)*, pages 1–4. IEEE, 2020. 2

[40] Mrutyunjaya Sahani, Chiranjiv Nanda, Abhijeet Kumar Sahu, and Biswajeet Pattnaik. Web-based online embedded door access control and home security system based on face recognition. In *2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]*, pages 1–6. IEEE, 2015. 3

[41] Aditya Saroha, Anant Gupta, Aditya Bhargava, Anup Kumar Mandpura, and Himanshu Singh. Biometric authentication based automated, secure, and smart iot door lock system. In *2022 IEEE India Council International Subsections Conference (INDISCON)*, pages 1–5. IEEE, 2022. 3

[42] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 2

[43] Cunli Song and Shouyong Ji. Face recognition method based on siamese networks under non-restricted conditions. *IEEE Access*, 10:40432–40444, 2022. 3

[44] Yaozhe Song, Hongying Tang, Fangzhou Meng, Chaoyi Wang, Mengmeng Wu, Ziting Shu, and Guanjun Tong. A transformer-based low-resolution face recognition method via on-and-offline knowledge distillation. *Neurocomputing*, 509:193–205, 2022. 3

[45] Jialin Tang, Qinglang Su, Binghua Su, Simon Fong, Wei Cao, and Xueyuan Gong. Parallel ensemble learning of convolutional neural networks and local binary patterns for face recognition. *Computer Methods and Programs in Biomedicine*, 197:105622, 2020. 2, 3

[46] Leslie Ching Ow Tiong, Seong Tae Kim, and Yong Man Ro. Multimodal facial biometrics recognition: Dual-stream convolutional neural networks with multi-feature fusion layers. *Image and Vision Computing*, 102:103977, 2020. 3

[47] Ivan William, Eko Hari Rachmawanto, Heru Agus Santoso, Christy Atika Sari, et al. Face recognition using facenet (survey, performance test, and comparison). In *2019 fourth international conference on informatics and computing (ICIC)*, pages 1–6. IEEE, 2019. 2

[48] Cuican Yu, Zihui Zhang, Huibin Li, Jian Sun, and Zongben Xu. Meta-learning-based adversarial training for deep 3d face recognition on point clouds. *Pattern Recognition*, 134:109065, 2023. 3

[49] G Md Zafaruddin and HS Fadewar. Face recognition using eigenfaces. In *Computing, Communication and Signal Processing: Proceedings of ICCASP 2018*, pages 855–864. Springer, 2019. 2

[50] Erfan Zangeneh, Mohammad Rahmati, and Yalda Mohsenzadeh. Low resolution face recognition using a two-branch deep convolutional neural network architecture. *Expert Systems with Applications*, 139:112854, 2020. 2

[51] Teng Zhang, Arnold Wiliem, Siqi Yang, and Brian Lovell. Tv-gan: Generative adversarial network based thermal to visible face recognition. In *2018 international conference on biometrics (ICB)*, pages 174–181. IEEE, 2018. 2

[52] Zhiguo Zhu and Yao Cheng. Application of attitude tracking algorithm for face recognition based on opencv in the intelligent door lock. *Computer Communications*, 154:390–397, 2020. 3