Real-Time Thermal Face Identification System for Low Memory Vision Applications Using CNN

Rami Reddy Devaram[1], Alessandro Ortis[1(✉)], Sebastiano Battiato[1], Arcangelo R. Bruna[2], and Valeria Tomaselli[2]

[1] Department of Mathematics and Computer Science, University of Catania, Viale A. Doria, 6, 95125 Catania, Italy
ramireddy.devaram@phd.unict.it, {alessandro.ortis,sebastiano.battiato}@unict.it

[2] STMicroelectronics, Catania 95121, Italy
{arcangelo.bruna,valeria.tomaselli}@st.com

Abstract. Image based face identification systems have attained optimal performance. However, the design of such systems often involves some issues related to extreme light conditions and privacy protection, among others. Since several years, Face Identification (FI) based on thermal images using deep neural networks (DNN) has received significant attention. Yet, the majority of the FI systems developed through DNN's need huge computational power; those systems are not suitable for the devices with memory limitations. In this paper, we proposed a new CNN framework based on depthwise separable convolutions for real-time face identification for low memory vision applications. The lack of publicly available thermal datasets makes very hard the research and developing of new techniques. In this work, we further present a new large-scale thermal face database called "ST_UNICT_Thermal_Face". As per our analysis, the evaluation of the learnt model using the data obtained in the single-day (without temporal variations), it might not stable over time. One of the main reasons behind the development of this database for the real-time evaluation of the proposed model depends on the fact that most thermal face identification systems are not stable over time and climate due to insufficient time data. The evaluation results exhibit that the proposed framework is suitable for the devices having limited memory and which is stable over time and different indoor environmental conditions.

Keywords: Thermal face identification · Thermal images · Convolutional neural networks

1 Introduction

Over the last two decades, Face Identification or Recognition, based on surveillance and traditional images obtained in the visible spectrum, have reached a

A. Del Bimbo et al. (Eds.): ICPR 2020 Workshops, LNCS 12665, pp. 529–543, 2021.
https://doi.org/10.1007/978-3-030-68821-9_44

significant level of advancement among various disciplines [5]. It has applied to many real world applications such as security and transportation. Although a face identification system based on visual images performs under controlled illumination conditions and there is a possibility to recognize faces of corpses, such system is not entirely suitable for the security applications. Nevertheless, in literature, we have a reliable biometric technology based on human physical or behavioural feature and can be used to identify human-beings such as finger-print effect. There are various biometric systems based on fingerprints such as iris patterns, signatures, palm-prints, and voice. However, such systems require the involvement of the subject [18]. Among these, Thermal Face Identification based biometric technology has been attracting significant attention due to its various advantages, such as ease of handling with low-resolution images and robust to ambience illumination, economic friendly hardware and high recognition accuracy and privacy preserving. To address the limitations of the above-mentioned biometrics systems, we were motivated to develop a face identification system based on thermal Vision. In the processes to overcome the aforementioned matters, there are huge works for many years to learn some usual innovations that can be transferred across various recognition tasks. Accompanying this direction, Deep Convolutional Neural Networks (DCNN) have attained enormous progress in grabbing various tasks of computer vision problems [10,17]. However, since such algorithms require high computational and memory resources, proper strategies must be employed [1]. In this paper, we focused on a novel CNN model for mobile or embedded vision applications, the one of the goal is the learned model must fit into the low-memory portable device like STM32F7 Family Microcontrollers, which includes limited memory, in terms of RAM and Flash Memory.

The convolutional neural network itself having a dense computational model, a large number of parameters, heavy computing load, and excess memory access leads to large power consumption, which makes the task more challenging to embed the learnt model into mobile or other portable devices with limited hardware resources (fixed memory). The motivation behind the development of the proposed work that previous works reveal that the compression of large networks according to the tasks attains the significant results, such as SqueezeNet [8] which is based on fire module and ShuffleNet [27] modified from the residual structure with group pointwise convolution and channel shuffle operation.

First we performed attempts to retrain state of the art models such as ResNet50 and MobileNet on the publicly available dataset from scratch, in order to assess how these models perform on the thermal vision domain in terms of performance and the computational complexity. Although these models consumed huge memory and showing significant performance. However, to be adequately applied to real-time applications and low-memory portable devices like STM32F7 Family Microcontrollers having limited memory, we selected MobileNet architecture as the best alternative to meet the requirements of the task by compress and accelerate the network to reduce parameters, computation, ad the power consumption of the model. MobileNet is a lightweight network compared with

ResNet in terms of the number of parameters, and the performance is nearly similar with both models. Though, parameters and computational burden reduced by the implementation of depthwise separable convolutions.

2 Materials and Data

2.1 Thermal Camera

The proposed Thermal Image Database called "ST-UNICT-Thermal-Face" has been acquired using a PureThermal-2 FLIR Lepton Radiometric LWIR camera equipped with Smart I/O STM32F412CGU6 microcontroller and 50-degree field view module including stutter for automatic calibration. It is pre-configured to operate plug and play UVC 1.0 USB. The 'thermal image' acquired using a thermal camera can be handled as a Grayscale Image (2D-array @16bpp). Each value is related to the absolute temperature of the corresponding area. During acquisition, the camera can be configured to operate in a High gain state, which provides lower Noise Equivalent Differential Temperature and lower intra-scene range. It is using a focal-plane array of 160×120 active pixels. Integrated digital thermal image processing functions capable of filtering and compensation for environmental issues. The radiometric Lepton captures accurate, calibrated, non-contact temperature data in every pixel of each image with the thermal sensitivity of less then 50 millikelvin, Radiometric accuracy (35 °C blackbody) at High gain: ±5 °C @ 25 °C, at Low gain ±10 °C @ 25 °C and scene dynamic range at High gain is −10 °C to +140 °C and at Low gain is −10 °C to +400 °C. It scans 8 to 14 µm wavelength bands to generate uniform thermal images at up to 9 frame/s.

2.2 Microcontroller

With the help of STM32CubeMX extension AI conversion tool which is STM32Cube.AI, it is simple and effective interoperability with modern Deep Learning training tools broadly used by the AI developer can be directly imported into the STM32Cube.AI. In this research, to embed the pre-trained model, we utilized STM32F767 Microcontroller, it has a new set of AI solutions to map and run pre-trained ANN models on the board. Since, STM32F767 includes fixed high-speed embedded memories with a Flash memory up to 2 MB, 512 KB of SRAM. The objective of this research is the development of the smallest CNN model for vision applications, which adequately fit into such STM32F767 Microcontroller. Moreover, this device is incorporate high-performance Arm Cortex-M7 32 bit RISC core processor operating at up to 216 MHz, and it is enabled with Floating point unit (FPU) which supports double and single precision data processing instructions and data types.

2.3 Datasets

Development of a thermal face identification system requires input data to test and validate the performance of the system. Datasets are required to train deep

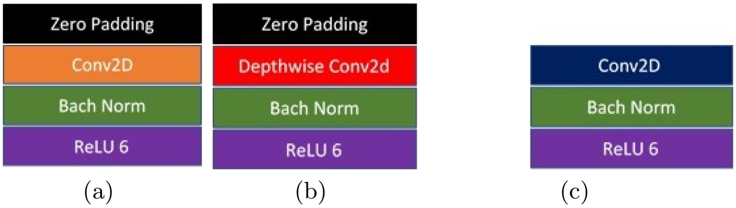

Fig. 1. CNN Blocks. (a). Standard Convolution Block, (b). Depthwise Convolution Block, (c). Pointwise Convolution Block

Table 1. Train, Validation and Test data distribution of Dataset-1

# Subjects	# Train	# Validation	# Test
12 Subjects	500	200	100
Total Samples	6000	2400	1200

learning models as well as to evaluate system performance. Proposed ST-UNICT-THERMAL-FACE DATABASE been obtained different states of ambience and the subject - a small Orientation of faces for flexible real-time system. It is very useful to obtain diverse data including various scenarios and to utilize the freely available sets to analyse the results of the experiments. Research Community frequently offered substantial datasets with respect to the various tasks. To evaluate the proposed model, we employed 5 Thermal Face datasets, including, public domain datasets. Since, unavailability of the subject, freely available data is not relevant to real-time evaluation. However, we proposed the following new benchmark datasets to evaluate the system in real-time. The description of all datasets we used in this research as follows.

Dataset-1. This dataset developed for the competition of Benchmark/Test Datasets by IEEE OTCBVS Workshop series [3]. The images have been captured using a long wave Raytheon L-3 Thermal-Eye 2000AS thermographic camera. The camera produces a resolution of 320×240 pixels of infrared images and this has been the final resolution chosen to evaluate the proposed model. It contains 20 subjects, and due to unbalanced classes distribution, we selected 12 subjects of 9600 images Table 1. The data acquired in different acquisition modalities, such as face orientation, ambience and occlusions. The images captured with Front, Right and Left orientations of a video with 20 frames in each acquisition. Moreover, the data obtained in indoor/outdoor conditions and various face occlusions including with and without Eye Glasses and Hat. The whole dataset is pre-processed and stored in 8-bit Grayscale JPEG format.

Table 2. Train, Validation and Test data distribution of Dataset-2

# Subjects	# Train	# Validation	# Test
29 Subjects	100	≈50	≈50
Total Samples	2900	1303	1305

Dataset-2. The images were captured using a PureThermal-2 Radiometric camera Sect. 2.1. The dataset developed to evaluate the proposed model in real-time. It contains 29 male and female subjects/people of 6405 images Table 2 with the resolution of 160×120 pixels. Each subject has been recorded in a single day as both snapshots and continues images (Videos) for further analysis. The whole dataset was captured in indoor (room) environment with different viewpoints Fig. 2 and accessories (Eyeglass, Cap, free and tied hair for women) Fig. 4 under uniform lighting and climate conditions. Each subject was asked to look at a fixed view-point continuously to 9 equidistant positions Fig. 3 making an estimated semi-circle around the camera with each accessory. The average time required for the whole acquisition process per individual subject has been 3 to 5 min. Moreover, the distance between the face of the subject and the camera is approximately between 120 cm to 150 cm to minimize the inherent parallax error. Finally, for the later advances, the entire data was saved in RAW-16bit TIFF format and further pre-processed to 8-bit JPEG visible grayscale image format to train and test the CNN model.

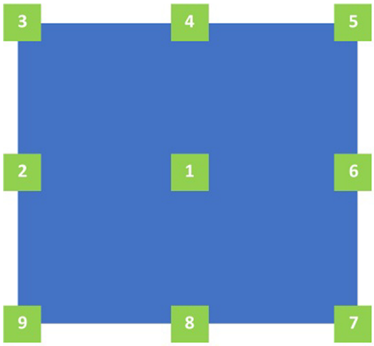

Fig. 2. Structure of head postures

Dataset-3. The whole dataset has been obtained with the same device and modalities used to as the Dataset-2 [2.3]. However, after the one-hold evaluation of the proposed model on the test set of the Dataset-2 [2.3], the system obtained optimal performance. Yet, due to temporal thermal variations between data used for training and real-time testing, the system performance is not optimal.

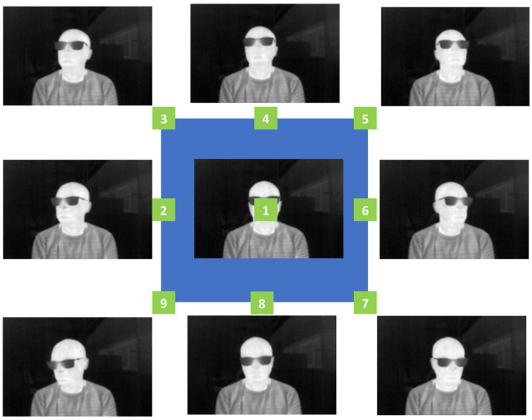

Fig. 3. Head Postures

Fig. 4. Samples from the dataset

However, we acquired new data over multiple sessions and with a diverse environments (background) due to develop the system robust to background and its artifacts.

The dataset eventually developed with 11 subjects, a total of 14740 images acquired in 6 distinctive days between 10 to 20 days period, including the data from dataset-2 [2.3]. The data distribution for the train, Validation and testing of this dataset presented in Table 3.

Note: The Final Dataset available in public domain soon.

Table 3. Train, Validation and Test data distribution of Dataset-3

# Subjects	# Train	# Validation	# Test
11 Subjects	1000	200	140
Total Samples	11000	2200	1540

Dataset - 4. The Carl Database is one of the benchmark datasets obtained in three different scenarios such as visible, NIR and Thermal domain under various illumination conditions. In this paper, we utilise one the dataset from Carl-database, which is the dataset obtained in the thermal domain by using thermographic camera TESTO 880-3, it contains 41 subjects/people and 60 images/subject with the resolution of 320 $\times$ 240 pixels. In [4] provides the complete description of the dataset. However, the Carl thermal images dataset divided into two datasets. Such as the dataset contains the images with and without background (segmented). We evaluate the proposed model on both datasets to investigate is the background matters. In the following sections, we named dataset-4.1 and Dataset-4.2 for the images with background and without background respectively Table 4.

Table 4. Train, Validation and Test data distribution of Dataset-4A and Dataset-4B

# Subjects	# Train	# Validation	# Test
41 Subjects	40	10	10
Total Samples	1640	410	410

3 Related Works

Face Identification (FI) technologies are used in security related applications, it needs to be analysed well before going to deploy, such analysis well described in [6]. However, most of the technologies are utilized images obtained in the visible spectrum. Since the huge availability of thermal imagery technology, FI system based on thermal imagery seems more secured and robust compare with the FI system based on visible images. However, the features of skin colour and texture, often exploited to make inferences on face images [23], are not present in the thermal images, which are mainly determined by the absolute temperature values of the face. There are well traditional feature extraction or appearance based methods works for images acquired in the thermal domain, yet in [24] analysed and explored the conflicts by utilizing those approaches. Moreover, to extract thermal features in [15] performed Fast Independent Component Analysis and Gabor Wavelet Transformation, but it does not improve the recognition rate. Certainly, face identification in the thermal domain has been relatively limited compare with the visual domain. However, [11] explains the comparisons, benefits

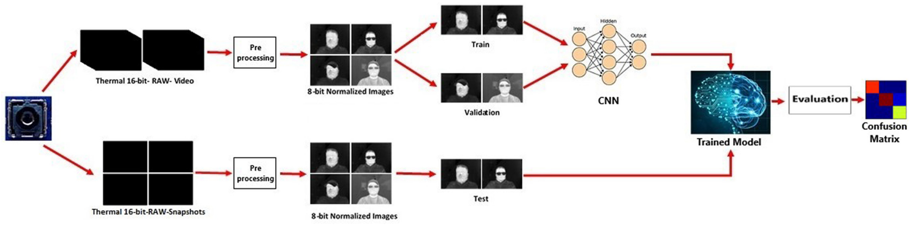

Fig. 5. Training and Testing Pipeline

and weaknesses of various thermal face recognition methods. [21,22] explores a relative study on the performance of various approaches for the FI system using both thermal and visual images and in authors in [9] provides an opinion on most of the research implemented on the face recognition approaches in both IR and Visible spectrum. Deep Learning methods might be an alternative to robust and real-time face identification system. In literature, many experiments showing great interest in the thermal domain. In [19] employed a standard pre-trained CNN based model to identify the person in thermal images. [13] works to develop a face recognition system by fusion of both CNN and SVM, but this approach needs huge computation power. In [25] proposed a multilayer CNN architecture, since the evaluation results are optimal, since those obtained on the data taken the same day as data used for training. In [16] proposed TIRFaceNet based on CNN, it takes extracted features from both visible and thermal images to train the network. Compare with all the previous works proposed approach is more convenient in terms of computational cost and temporal variations.

To improve the performance of the model with respect to their task, authors of [2,12,14,20,26] modified the state of the art models in terms of hyperparameters, size of the kernels and the architecture.

4 Proposed CNN Architecture

In this paper, we introduced a small and robust model based on CNN which is described in Fig. 6 for thermal face identification system, with advantages of the Depthwise Separable Convolution (DWSC) and modality of the MobileNet [7]. The architecture, presents an efficient tradeoff between complexity and performance of the system. It is typically designed for the devices with a shortage of computational power such as mobile or embedded based vision applications. The main goal of the proposed model is to significantly reduce the number of parameters by reducing the multiplication of the total number of floating-points. Proposed modernised architecture Fig. 6, employed two types layers such as standard convolution Fig. 1(a) and depthwise separable convolution. Depthwise separable convolution is divided into depthwise convolutions Fig. 1(b) followed by pointwise convolution Fig. 1(c). However, standard convolution built only on the top of the network, and the whole network remains work with only depthwise separable convolutions. Proposed topology having one standard convolution, eight depthwise convolutional layers followed by pointwise convolution layer each, one fully connected and a softmax classifier at the end of the network respectively.

All the layers are followed by Batch Normalization and ReLU non-linearity except final softmax layer. To improve the performance by preserving the spatial information at the boarders after convolution operation we used Zeropadding in both standard and DWSC layers. One dropout layer is implanted at the end of the fully connected layer. To reduce the spatial resolution to one a final global average pooling used before the fully connected layer. Initially, the top layer of the network which is standard convolution layer gets the input. The input is a thermal image of dimensions $H \times W \times C$. The Conv2d layer holds all the pixel values and computes the output of the neuron and passes them to the following layer. The parameter of the layer is composed of a set of learnable kernels. Kernels or filters are convolved across dimensions of the input and enlarging with its depth, with the implementation of the dot product between input values and kernels. It outputs a two-dimensional activation map of the kernel. The network learns the kernels thought the features from the input spatial position. The ReLU perform an elementwise activation function, it activates and only grows linearly with positive values, but there is no impact on the size of the volume. Then it passes the output values to the following depthwise separable convolution layer.

Depthwise convolution is a spatial convolution based on the number of channels, each of the filters only computes on a single input channel to do convolution. In thermal images are having a single channel, in such case 1×1 spatial convolutions and integrate depthwise convolution outputs linearly with the help of 1×1 pointwise convolution. The computational operations using depthwise separable convolutions are much lesser than the standard convolutions. As like MobileNet, we have two hyperparameters such as width multiplier α and depth multiplier σ, both parameters are chosen as 1. In this work, we did not use any pooling operation except before the fully connected layer. In proposed topology we have designed a unique structure. Our goal is to reduce the complexity of the network, since, the complexity of the network depends on the size of the output feature maps of each layer. The size of the feature map influenced by three important parameters, such as size of the kernel, stride and zero padding.

To find the key features from the input data, the size of the kernel plays an important role. Large kernels may overlook at the features it leads to loss essential features, since small size kernel may find more information it may lead to over fitting. However, it is very hard task to find the optimal size of the kernel, it may depends on the task. In this paper, we proposed, instead of finding optimal size of the kernel, we would fix the size of the kernel to the entire network, we increase the number of kernels with small difference in each of the following layer, it prevents the confusion with more information. However, Strides also actually influences on the number of steps that moves the kernel on the input image, and also the size and volume of the output feature maps.

The entire training processes implemented in Keras following TensorFlow backend, for fast computation we employed NVIDIA Graphic Processing Unit having 12 GB of RAM. All the training process implemented from scratch.

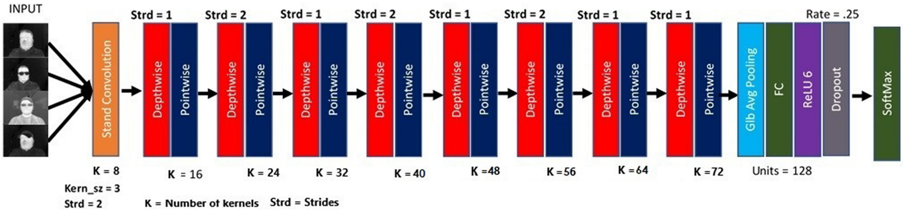

Fig. 6. Proposed CNN Architecture

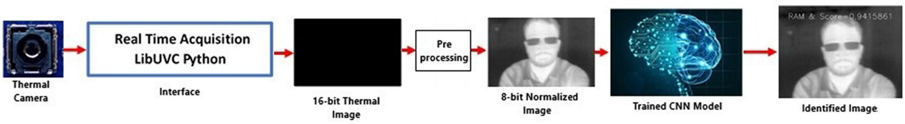

Fig. 7. Real-Time Testing Phase Pipeline

The proposed model having total 29,798 parameters and it consumes 335.17 KB and 117.14 KB of RAM and FLASH memory respectively and which obtained optimal performance on five thermal face datasets, three of them are publicly available dataset-1, dataset-4&5 and datasets-2 and dataset-3 are developed by ourselves to test the learnt model in real-time.

5 Algorithm Pipeline

The algorithm pipeline to train the network with the proposed datasets is described in Fig. 5. We have acquired 16-bit RAW thermal Videos and Images (snapshots). Further, we extracted all the frames from the videos, then we preprocessed those 16-bit RAW non-visible IR Images into 8-bit visible grayscale images by the implementation of Min-Max Normalization (Eq. 1) using Python-OpenCV. Those Normalized images are randomly shuffled and split into train and validation sets. To train the proposed CNN, those train and validation datasets are given as an input to the model. However, we obtain 16-bit RAW thermal images as snapshots rather than the videos to test the learnt model. We have implemented the same preprocessing approach to test set shown in Fig. 5.

The preprocessing, such conversion of 16-bit image to 8-bit image by the following Min-Max Normalization equation.

$$\widehat{I}(x,y) = round(\frac{I(x,y) - min(I)}{max(I) - min(I)}) * 255) \quad (1)$$

Where I is the input image, min(I), max(I) are, respectively, the minimum and maximum value of the image I pixel values. Where $\widehat{I}$ is the input to the CNN.

Table 5. Evaluation results of proposed model

Dataset	# Persons	# Training	# Validation	# Testing	Avg. validation accuracy (%)	Avg. test accuracy(%)
Dataset- 1	12	6000	2400	1200	100	100
Dataset- 2	29	2885	1303	1046	99.61	97.57
Dataset- 3	**11**	**11000**	**2200**	**1755**	**99.95**	**99.96**
Dataset- 4A	41	4140	410	410	93.90	95.12
Dataset- 4B	41	4140	410	410	100	100

Table 6. Complexity of the proposed architecture compare with state of the art models

Models	Number of parameters	RAM memory consumption	FLASH memory consumption
ResNet50	23,794,333	4.30 MB	94.96 MB
MobileNet-V1	3,342,341	1.32 MB	13.32 MB
Proposed Model	29,798	335.17 KB	117.14 KB

6 Experimental Analysis and Results

In this section we describe all the experimental analysis has been implemented to obtain the optimal performance of the real-time thermal face identification system. However, all the obtained results are described in the Table 5 and complexity of the proposed model compared with the state of the models are described in the Table 6.

Experiment - 1
Initially, we have analysed several state of the art CNN topologies with publicly available dataset-1, such as ResNet50, MobileNet to find the suitable networks in terms of low memory consumption. However, ResNet50 and MobileNet are consuming huge computation power. ResNet50 has 23,794,33 parameter and it need minimum of 5 MB and 100 MB of RAM and FLASH memory respectively. Compare with ResNet50, MobileNet-V1 consumes very less memory Table 6.

However, as we decided by utilizing the advantages of MobileNet, we designed the small proposed CNN topology especially for the devices having low computational power, which described in Fig. 6 and confusion matrix of evaluation presented in Fig. 8.

Experiment - 2
To development of Face Identification system, we trained the proposed model with 12 subjects having total of 9600 images selected from the dataset-1. We split the dataset as train, validation and test of 500, 200 and 100 images per subject respectively; the distribution of the data is described in Table 1. We evaluate learnt model on both validation and test sets, we obtain the maximum results, presents in the Table 5.

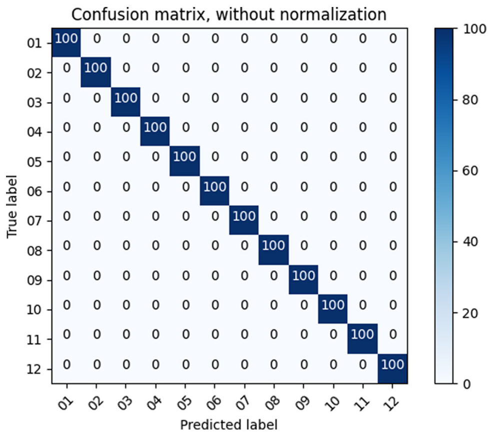

Fig. 8. Confusion Matrix of Test Set of Dataset-1

Experiment - 3

Due to unavailability of subject, we are not able evaluate learnt model in real-time. However, we developed a new dataset described in 'Sect. 2.3' and trained the network from scratch. The distribution of the data to train the network described in the Table 2. The proposed model shows optimal performance on both validation and test sets presented in Table 5 and evaluation confusion matrix shown in Fig. 10. However, the performance of real-time testing is poor. Thermal images contain absolute temperature values of the face surface, yet temperature may changes with emotions, ambience and/or time.

Experiment - 4

To address the drawbacks in the previous experiments, such as the model expects having data with temporal variations. However, we acquired a new dataset in several distinct days, which is described in Sect. 3. The evaluation of the proposed model on the dataset-4 showing optimal performance on the both train and test sets presented in Table 5 and also showing significant performance on the real-time test. The pipeline of the real-time evaluation described in Fig. 7 and Fig. 9 is the real-time identified image.

Experiment - 5

In order to improve the analysis, we evaluate our model with another publicly available dataset described in Sect. 2.3. This dataset having two type of images, such as segmented (without background) and original (with background) images. The evaluation results Table 5 reveals that the recognition rate may improve by removing the background.

Fig. 9. Identified Image in Real-Time

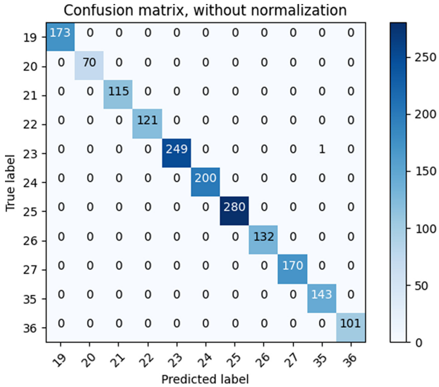

Fig. 10. Confusion Matrix of Test Set of Dataset-3

7 Conclusion

In this work we addressed the problem of the complexity of CNN for low memory portable devices and real-time evaluation. The proposed framework which includes data acquisition modalities and a new CNN topology based on depthwise separable convolution makes the model more efficient for real-time thermal face identification. Despite that already high performance achieved by various previous works, yet those systems evaluate the learnt model only with validation or test data. Our method performs well to identify the person from thermal images. Based on this research in future works, hopefully by the analysis of proper prepossessing techniques may improve the performance and robustness of the system.

References

1. Battiato, S., Conoci, S., Leotta, R., Ortis, A., Rundo, F., Trenta, F.: Benchmarking of computer vision algorithms for driver monitoring on automotive-grade devices. In: 2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE), pp. 1–5 (2020)
2. Chen, L., Peng, L., Yao, G., Liu, C., Zhang, X.: A modified inception-ResNet network with discriminant weighting loss for handwritten chinese character recognition. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, September 2019, https://doi.org/10.1109/icdar.2019.00197
3. Davis, J.W.: Otcbvs benchmark dataset collection, http://vcipl-okstate.org/pbvs/bench/
4. Espinosa-Duró, V., Faundez-Zanuy, M., Mekyska, J.: A new face database simultaneously acquired in visible, near-infrared and thermal spectrums. Cogn. Comput. **5**, 119–135 (2013). https://doi.org/10.1007/s12559-012-9163-2

5. Farinella, G.M., Farioli, G., Battiato, S., Leonardi, S., Gallo, G.: Face re-identification for digital signage applications. In: Distante, C., Battiato, S., Cavallaro, A. (eds.) Video Analytics for Audience Measurement, pp. 40–52. Springer, Cham (2014)
6. Gondhi, N.K., Kour, E.N.: A comparative analysis on various face recognition techniques. In: 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE (2017). https://doi.org/10.1109/iccons.2017.8250626
7. Howard, A.G., et al.: Mobilenets: efficient convolutional neural networks for mobile vision applications (2017)
8. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: Squeezenet: alexnet-level accuracy with 50× fewer parameters and <0.5 mb model size (2016)
9. Kong, S.G., Heo, J., Abidi, B.R., Paik, J., Abidi, M.A.: Recent advances in visual and infrared face recognition—a review. Comput. Vis. Image Understanding **97**(1), 103–135 (2005). https://doi.org/10.1016/j.cviu.2004.04.001
10. krishna, M., Neelima, M., Mane, H., Matcha, V.: Image classification using deep learning. Int. J. Eng. Technol. **7**, 614 (2018). https://doi.org/10.14419/ijet.v7i2.7.10892
11. Kristo, M., Ivasic-Kos, M.: An overview of thermal face recognition methods. In: 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, May 2018. https://doi.org/10.23919/mipro.2018.8400200
12. Li, X., Li, W., Xu, X., Du, Q.: Cascadenet: Modified resnet with cascade blocks. In: 2018 24th International Conference on Pattern Recognition (ICPR), pp. 483–488 (2018)
13. Lin, S.D., Chen, K.: Thermal face recognition under disguised conditions. In: 2019 International Conference on Machine Learning and Cybernetics (ICMLC). IEEE, July 2019. https://doi.org/10.1109/icmlc48188.2019.8949194
14. Liu, S., Deng, W.: Very deep convolutional neural network based image classification using small training sample size. In: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). IEEE, November 2015. https://doi.org/10.1109/acpr.2015.7486599
15. Majumder, G., Bhowmik, M.K.: Gabor-fast ICA feature extraction for thermal face recognition using linear kernel support vector machine. In: 2015 International Conference on Computational Intelligence and Networks. IEEE, January 2015. https://doi.org/10.1109/cine.2015.14
16. Manssor, S.A.F., Sun, S.: TIRFaceNet: thermal IR facial recognition. In: 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, October 2019. https://doi.org/10.1109/cisp-bmei48845.2019.8966066
17. Mateen, M., Wen, J., Nasrullah, D., Song, S., Huang, Z.: Fundus image classification using VGG-19 architecture with PCA and SVD. Symmetry **11**, 1 (2018). https://doi.org/10.3390/sym11010001
18. Santarcangelo, V., Farinella, G.M., Battiato, S.: Gender recognition: methods, datasets and results. In: 2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW), pp. 1–6 (2015). https://doi.org/10.1109/ICMEW.2015.7169756
19. Sayed, M., Baker, F.: Thermal face authentication with convolutional neural network. J. Comput. Sci. **14**(12), 1627–1637 (2018). https://doi.org/10.3844/jcssp.2018.1627.1637

20. Sheng, T., Feng, C., Zhuo, S., Zhang, X., Shen, L., Aleksic, M.: A quantization-friendly separable convolution for MobileNets. In: 2018 1st Workshop on Energy Efficient Machine Learning and Cognitive Computing for Embedded Applications (EMC2). IEEE (2018). https://doi.org/10.1109/emc2.2018.00011
21. Socolinsky, D., Selinger, A.: A comparative analysis of face recognition performance with visible and thermal infrared imagery. In: Object recognition supported by user interaction for service robots. IEEE Comput. Soc. https://doi.org/10.1109/icpr.2002.1047436
22. Socolinsky, D.A., Selinger, A., Neuheisel, J.D.: Face recognition with visible and thermal infrared imagery. Computer Vision and Image Understanding **91**(1–2), 72–114 (2003). https://doi.org/10.1016/s1077-3142(03)00075-4
23. Trenta, F., Conoci, S., Rundo, F., Battiato, S.: Advanced motion-tracking system with multi-layers deep learning framework for innovative car-driver drowsiness monitoring. In: 2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019), pp. 1–5 (2019). https://doi.org/10.1109/FG.2019.8756566
24. Vigneau, G.H., Verdugo, J.L., Castro, G.F., Pizarro, F., Vera, E.: Thermal face recognition under temporal variation conditions. IEEE Access **5**, 9663–9672 (2017). https://doi.org/10.1109/access.2017.2704296
25. Wu, Z., Peng, M., Chen, T.: Thermal face recognition using convolutional neural network. In: 2016 International Conference on Optoelectronics and Image Processing (ICOIP). IEEE, June 2016. https://doi.org/10.1109/optip.2016.7528489
26. Yuan, A., Bai, G., Jiao, L., Liu, Y.: Offline handwritten english character recognition based on convolutional neural network. In: 2012 10th IAPR International Workshop on Document Analysis Systems. IEEE, March 2012. https://doi.org/10.1109/das.2012.61
27. Zhang, X., Zhou, X., Lin, M., Sun, J.: Shufflenet: An extremely efficient convolutional neural network for mobile devices (2017)