RESEARCH ARTICLE | DECEMBER 15 2023

Deep learning framework for smart patient attendant system using thermal images

Dinesh Dusa; Mallikarjuna Rao Gundavarapu

AIP Conf. Proc. 2901, 060022 (2023)
https://doi.org/10.1063/5.0187049

Deep Learning Framework for Smart Patient Attendant System Using Thermal Images

Dinesh Dusa [a)], Mallikarjuna Rao Gundavarapu[b)]

Department of Computer Science and Engineering,
Gokaraju Rangaraju Institute of Engineering and Technology(jntuh), Hyderabad, India

[a)] *Corresponding Author: dineshdusa28@gmail.*com
[b)] *gmr_333@yahoo.com*

Abstract. A terrible disease, Covid-19, has afflicted the world. As it spreads infectious and air droplet transmission disease, it is always mandatory to wear masks and temperature checks in offices and crowded environments. As a matter of fact, the temperature check has become more mandatory. The manual way of measuring temperature also may spread the disease. A process that is fully automatic and does not require human interaction has a considerable advantage in controlling the spread. A camera with an IR sensor equipped with artificial intelligence can help with this. In epidemics involving infectious disease pandemics such as Ebola and SARS, using infrared thermographs (IRTs) as a fever detection method has been implemented. As Infrared images show the unique heat-signature of the human face, which can be used for facial recognition also which provides an added advantage. IR images have inherent advantages over visible light images due to their characteristics, making them useful in various fields such as improving face recognition algorithms. It is evident that IR images remain constant even in extreme lighting conditions. As well as detecting facial temperature based on pixel intensity, this paper deals with the deep learning based facial recognition and temperature detection from facial thermal image dataset. We are proposing a CNN based densenet201 architecture for facial recognition and identification, and OpenCV methodologies for examining the temperature of the validated images to accomplish this experiment.

Keyword: CNN, OpenCV, IR, Covid-19, Facial Recognition, Neural Networks

INTRODUCTION

As infrared technology has become more affordable over the past 10 years, we have also seen an increase in their use. In reaction to the current COVID-19 pandemic, the number of people using these devices has significantly increased, and they are now being used for fever detection in public places such as shopping malls, auditoriums, airports, railway stations etc. In practical applications, these devices can precisely measure an individual's body temperature with high accuracy [1]. Many authors had already noted the potential for these instruments to be used in biomedical applications where accurate body temperature assessment is crucial. Through the use of infrared cameras, a variety of clinical conditions including burns, hemangiomas, extremity thrombosis, etc., can be studied. Local body temperature differences are well studied using such applications. Several companies offer infrared cameras aimed at the medical field already, including Spectron IR and Medithermal IR modules [1].

It is essential that imaging systems meet certain requirements in clinical scenarios: they must be precise, portable, easy to use, autonomous, and they must be easy to interpret. Medical specialties like dermatology rely heavily on images in the visible spectrum for additional information on color and texture. Image analysis also helps physicians identify patients and injuries. It is therefore imperative to acquire multispectral images of the surfaces in the visible as well as the infrared (IR) bands [1]. A system with extra peripherals and sensors can be easily adapted to a variety of medical situations because the addition of these components allows it to operate in a variety of modes.

Third International Conference on Advances in Physical Sciences and Materials: ICAPSM 2022
AIP Conf. Proc. 2901, 060022-1–060022-9; https://doi.org/10.1063/5.0187049
Published by AIP Publishing. 978-0-7354-4765-3/$30.00

Meanwhile, the development of Convolutional Neural Networks paved the way for the processing of digital images and videos and the recognition of various content categories. In the domain of deep learning, automatic recognition of human faces is one of the most intriguing problems [2]. Machines are becoming increasingly humanoid day-by-day, and the challenge of autonomously detecting facial structure and emotions makes it a very important procedure for researchers in this area. Because facial detection is a critical function for an intelligent machine, we are now able to make machines work on it as humans do [3].

Although most of the existing computer vision algorithms use visible light for their input, heat signature is utilized extensively for a number of different tasks in the animal kingdom. By utilizing Infra-red (IR) data for object detection and identification, particularly for detecting and recognizing faces, we have investigated its potential benefits. It has been observed that face recognition using IR is more invariant than CCD, particularly in different conditions such as varying head 3D orientation [4].

In this article, a deep learning method is proposed that uses face information in the frontal position to analyze hot spots [5]. To achieve hot spot and temperature measurement, the proposed method utilizes DenseNet201 as an architecture for face detection and OpenCV techniques for image processing.

The structure of the paper is further divides as follows: section-II focuses on the literature review, section-III focuses on the techniques implemented in mentioned works, section-IV describes the methodology to process a thermal image to detect faces using transfer learning upon that DenseNet201 architecture, section-V discuss about the results and analysis and finally section-VI discusses about the conclusion of the whole experimentation and ends with References section

LITERATURE REVIEW ON RELATED WORKS

The authors in [6] employed infrared thermal images to alter the functions of building heating, ventilation, and air conditioning (HVAC) systems. In this study, the central facial regions' skin temperature was measured using a Face identification method with a Haar-Cascade algorithm, More precisely, in a facial orientation that is frontal. As part of the study mentioned, twelve subjects were tested for infrared comfort using the proposed method. 85% accuracy is shown by results that indicate infrared thermal comfort is best assessed by the ears, nose, and cheeks. One of the downsides of using the Haar Cascade algorithm is that it demands a significant amount of images and necessitates the adjustment of the bounding box size, which may lead to potential false detections. Therefore, it is not a very attractive option [6].

In 2017, [7] conducted a study that used infrared thermal images to localize the eyes of cattle, employing techniques such as ellipse detection and image processing. Detection of ellipses was carried out by using a randomized Hough Transform algorithm. To identify the eye region, this study utilized ellipse detection and thresholding techniques to eliminate non-eye regions. According to the results, the proposed method exhibits good performance in localizing the eyes in different orientations and locations. Using this method, it would be inappropriate to change the shape of an ellipse in response to the direction of the animal face. Moreover, it is difficult to apply to detect the eye region when several animals' faces are overlapping at the same time due to the fact that they are often seen as the same [7].

Marco Dell et al., developed contactless temperature detection which uses the matrix of the sensors for an accurate measurement of the body temperature [8]. The result is a temperature calculation that is 89% accurate. Using contactless temperature measurement technology, Marco Dell is able to measure ambient temperature up to 12 feet away by relating the maximum amount of energy available in the environment to the body temperature. This is called the Stefan-Boltzmann theory.

The research by Irving et al. was based on the integrated use of a smart thermal system that is capable of detecting emotions as they shift over time [9]. Thermal cameras were used to record the subjects as they walked through an environment, while an expert monitored them from a distance. In order to facilitate the emergence of emotions in the subjects, videos were handpicked with the aid of Psychology experts to create an environment that encourages emotional responses. This is the technique for detecting emotions that takes the first thermogram from the subject at the beginning and analyzes the ROIs following the previous examples that have been discussed. The ROI measures are determined by evaluating a fuzzy logic system with four inputs: cheeks, forehead, maxilla and nose. Then, a thermogram is taken once more when the emotion is forming and another analysis of ROIs are calculated. Apart from that, four inputs are transmitted to calibration with new temperature levels. During the calibration step, ROIs are divided into 3 groups each with a different phase; a low phase indicates that temperature has decreased, a normal phase implies temperature has stayed the same, and a high phase indicates that temperature has risen. Each ROI is outputted based on an increase or decrease in temperatures. Moreover, a top-down hierarchical classification approach was utilized by this classifier to examine temperature variances in accordance with established emotion rules and identify one of the five emotions (i.e., happy, disgust, fear, angry, sad). The results revealed that 89.9% of the males and 89.5% of the females were accurately diagnosed.

BACKGROUND

Dataset

Datasets from the OTCBVS Benchmark Dataset Collection are publicly available benchmarks designed for testing and evaluating computer vision algorithms that are state-of-the-art. The benchmark, which comprises videos and images captured both inside and outside of the visible spectrum, is accessible free of charge to all researchers in the field of computer vision. Furthermore, it will provide a broad audience of vision conference and workshop attendees from IEEE and SPIE with the chance to examine the applications of non-visible spectra in actual situations, and make a contribution to the OTCBVS workshop series aimed at promoting advancements in this research domain. Riad Hammoud launched this initiative in 2004. The Ohio State University hosted the project until 2013 under the direction of Dr. James W. David. Guoliang Fan, a researcher at Oklahoma State University, is currently part of the project's management team [10]. The data base consists of 18 people facial IR images.

FIGURE 1. OTCBVS Terravic Facial IR Database

Transfer Learning

Deep learning is a method that reuses an existing model. As shown in Figure 3, the basic concept of transfer learning can be seen. This study focuses on using an already-trained, relatively complex and successful model built on significant amounts of data, such as image dataset from ImageNet, in order to apply the learned knowledge to a relatively simple task with a limited amount of data. To facilitate the transfer, three attributes are there [11].

- The user does not have to tune hyper-parameters because the pretrained model is successful.
- This method is also known as pretraining, where the first layers of the model are used to extract all forms of low-level properties, such as shades, edges, blobs, tints, and textures at low resolution.
- In our view, the majority of the identification task is carried out by the final layers of the pretrained model. Hence, the target model may only require retraining of these final layers to accomplish the complicated identification task.

DenseNet201 Architecture

DenseNet is a term that is used to describe residual networks that utilize the concept of reused features. DenseNet relies on direct connections between early and later layers in a feedforward fashion, enabling the later layers to utilize the features from the previous layers.

The dense interconnections are the result of this process. The concept of pre-activations that underpins Pre-activation of ResNet is also implemented in DenseNet in order to achieve better accuracy than ResNets while reducing the number of parameters. DenseNet refers to this number as the growth rate at each layer. As a result, a particular layer's feature map grows as its feature maps of previous layers are concatenated. A DenseNet is an order of increasing size, such as DenseNet-121 and DenseNet-201, which exhibit a growth rate of 32 [12].

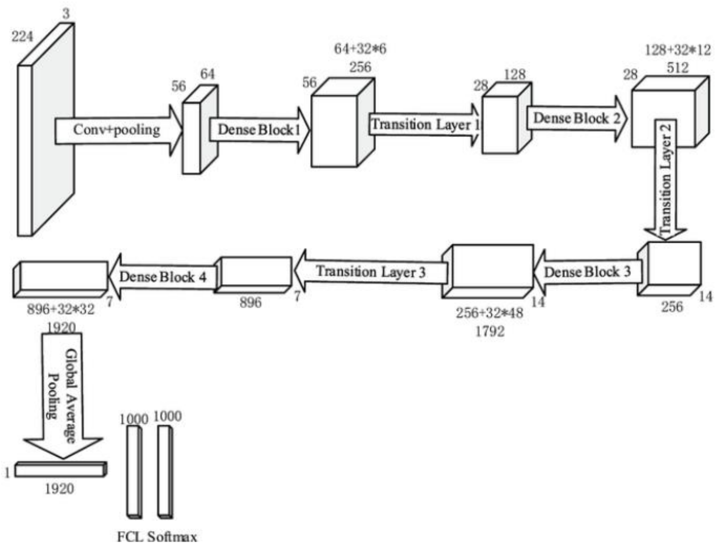

FIGURE 2. DenseNet201 Architecture

Temperature Measurement

To evaluate the proposed system, a dataset called Infra-red Thermography was generated by converting a series of grayscale images and infrared matrices. Using equation (1) we created the infrared matrix.

$$T_{min} + \left(\frac{T_{gray}}{T_{high\_gray}}\,(T_{max} - T_{min})\right) \quad (1)$$

A thermal image of an infrared object provides a calculation of temperature according to its thermal value. A thermal camera's maximum and minimum temperatures are measured by T_{min} and T_{max} respectively. For grayscale images, T_{gray} is the pixel intensity. High intensity gray images, in most cases 255, are referred to as T_{high_gray}. In this way, a temperature range of 94 to 108 Fahrenheit can be calculated on the basis of an infrared matrix [13].

ImageNet

The dataset used in our research is the ILSVRC 2012 classification dataset [14], which includes 1.2 million images for training, 50,000 images for validation, and 1,000 classes. To train our model, we employed the data augmentation methods outlined in [14], using either a 1-crop or 10-crop of 224x224 size during test time . Our validation set is analyzed in accordance with [14] and we report classification errors.

METHODOLOGY

The below pipeline in Fig1 illustrates how temperature and facial recognition can be extracted from thermal images in the proposed system. Fig 3 illustrates how the proposed technique loads an infrared image and a thermal matrix in grayscale. Grayscale images are pre-processed by removing noise before passing the Dense Net architecture for face detection and classification. The Region of Interest (ROI) is established by applying face identification to the input image. After detecting the faces, the face region is cropped and binarized to identify crucial points. A blob refinement stage preserves only hot spots in the face by calculating the temperature in each blob. Temperature measurement is the final stage. The following sections will explain every step of the proposed methodology in detailed way.

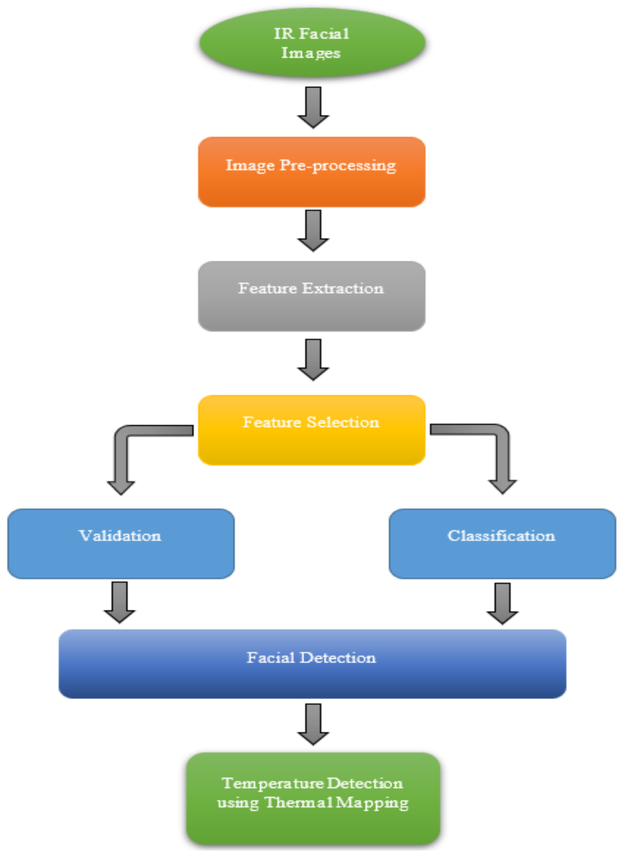

FIGURE 3. FlowChart

The backbone of our system is a version of DenseNet architecture. This is initialized by the pretrained weights obtained by training the DenseNet network on ImageNet pretrained weights.

We focus on fine-tuning the DenseNet201 network pretrained with a OTCBVS Terravic Facial IR Database in this experiment. The concept of fine-tuning involves setting parameters in the CNN model for the target task based on parameters previously pretrained for another task that relates to the target task. The stochastic gradient descent (SGD) was utilized to train all networks using the Adam algorithm. For our face-IR model, we chose 64 for batch sizes of 1088 images and 20 epochs, respectively, on the OTCBVS Terravic Facial IR Dataset. During the training process, the learning rate is reduced by a factor of 10 and set to 0.0001 at 50% and 75% of the total epochs. After successful training and validation, we saved the model weights. Further the model weights are used to load the trained network model for facial recognition or detection using the validation dataset. The images are converted from BGR to gray image scale to pass on through the pipeline to measurement the temperature using the equation-1. Instead of measuring the temperature on whole image we coordinated the pixel on center of the face to measure the exact pixel on the face. The measured pixel was highlighted as red dot in color.

```
Epoch 1/20
17/17 [==============================] - ETA: 0s - loss: 2.6875 - accuracy: 0.2217
Epoch 1: val_accuracy improved from -inf to 0.50531, saving model to .\best_weights.hdf5
17/17 [==============================] - 253s 13s/step - loss: 2.6875 - accuracy: 0.2217 - val_loss: 1.4297 - val_accuracy:
0.5053 - lr: 1.0000e-04
Epoch 2/20
17/17 [==============================] - ETA: 0s - loss: 1.4857 - accuracy: 0.5198
Epoch 2: val_accuracy improved from 0.50531 to 0.77282, saving model to .\best_weights.hdf5
17/17 [==============================] - 219s 13s/step - loss: 1.4857 - accuracy: 0.5198 - val_loss: 0.8789 - val_accuracy:
0.7728 - lr: 1.0000e-04
Epoch 3/20
17/17 [==============================] - ETA: 0s - loss: 0.9519 - accuracy: 0.6771
Epoch 3: val_accuracy improved from 0.77282 to 0.80255, saving model to .\best_weights.hdf5
17/17 [==============================] - 232s 14s/step - loss: 0.9519 - accuracy: 0.6771 - val_loss: 0.5449 - val_accuracy:
0.8025 - lr: 1.0000e-04
Epoch 4/20
17/17 [==============================] - ETA: 0s - loss: 0.6515 - accuracy: 0.7893
Epoch 4: val_accuracy improved from 0.80255 to 0.85563, saving model to .\best_weights.hdf5
17/17 [==============================] - 232s 14s/step - loss: 0.6515 - accuracy: 0.7893 - val_loss: 0.3571 - val_accuracy:
0.8556 - lr: 1.0000e-04
Epoch 5/20
17/17 [==============================] - ETA: 0s - loss: 0.4879 - accuracy: 0.8381
Epoch 5: val_accuracy improved from 0.85563 to 0.92569, saving model to .\best_weights.hdf5
17/17 [==============================] - 240s 14s/step - loss: 0.4879 - accuracy: 0.8381 - val_loss: 0.2452 - val_accuracy:
0.9257 - lr: 1.0000e-04
Epoch 6/20
17/17 [==============================] - ETA: 0s - loss: 0.3779 - accuracy: 0.8813
Epoch 6: val_accuracy improved from 0.92569 to 0.95754, saving model to .\best_weights.hdf5
17/17 [==============================] - 230s 14s/step - loss: 0.3779 - accuracy: 0.8813 - val_loss: 0.1413 - val_accuracy:
0.9575 - lr: 1.0000e-04
Epoch 7/20
...
17/17 [==============================] - ETA: 0s - loss: 0.0717 - accuracy: 0.9788
Epoch 13: val_accuracy did not improve from 1.00000
17/17 [==============================] - 231s 14s/step - loss: 0.0717 - accuracy: 0.9788 - val_loss: 4.0070e-04 - val_accuracy:
1.0000 - lr: 1.0000e-04
Epoch 13: early stopping
```

FIGURE 4. Training Model and Accuracies

Further the model is evaluated using the various parameters. To assess the model's performance, we constructed a confusion matrix for the validation set and computed four evaluation metrics: recall, precision, accuracy, and F1-score. These metrics were determined using the quantities of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), these five measures can be summarized as follows:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$TPR = \frac{TP}{TP+FN} \tag{4}$$

$$FPR = \frac{FP}{FP+TN} \tag{5}$$

A confusion matrix, which is also referred to as a C matrix, is a square matrix of dimensions NxN. The value of N represents the number of target classes, and the matrix is utilized to assess the performance of a classification model. This matrix is a comparison between the actual target values and the ones predicted by the machine learning model [15]. Our classification model is being evaluated comprehensively in this way, and we can see how well it is performing and how often it makes errors.

RESULTS

The objective of this thesis is to present research discoveries pertaining to thermal image-based facial recognition and temperature measurement. In order to recognize facial identification, we proposed the DenseNet201 architecture. We trained this model by separating 12 classes from the 18 classes OTCBVS Terravic Facial IR Dataset and deleted some images and applied data augmentation technique to overcome the overfitting issue. We fed 1088 images to the architecture for training purpose and 471 images for validation purpose and it achieved 97.88% of training accuracy with a loss of 0.0717% and 99.98% validation accuracy achieved. The comparison graph of trained and validated accuracies, as well as trained and validated losses, is shown below fig.6 & 7. Also, we showed the training model in the fig.4.

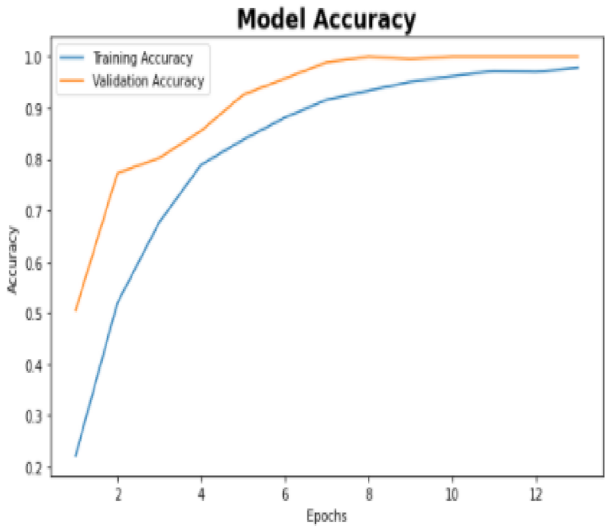

FIGURE 5. Train and validation Accuracy

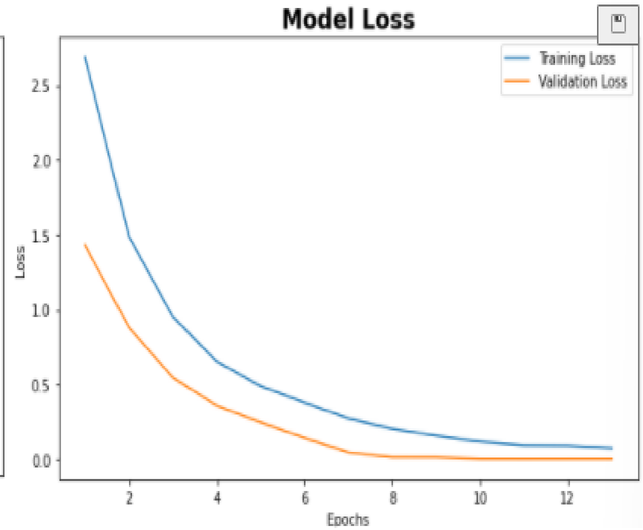

FIGURE 6. Train and Validation Loss

Further we evaluated our model to know the quality of a classifiers output we used Precision-Recall metrics. These metrics are useful to measure the Predictive success when classes are very imbalanced and we achieved 100% accuracy. Information retrieval measures precision by how relevant results are, while recall measures how many results are truly relevant.

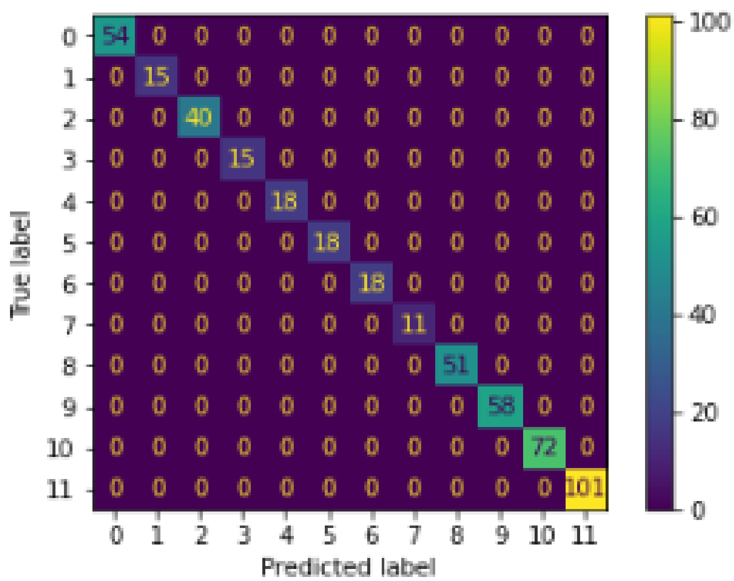

FIGURE 7. Confusion matrix of the validation dataset

To analyze the performance of the classification algorithm, we utilized a confusion matrix to visualize and summarize the data. A confusion matrix is shown in fig 5, where how many images classified correctly and how many misclassified. As discussed earlier 12 classes are available in both validation and training datasets. We used the validation dataset to detect the images and all the images classified correctly.

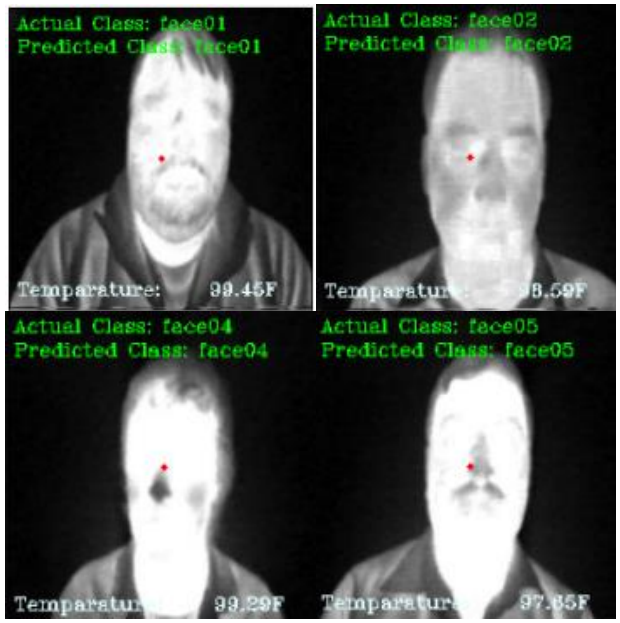

FIGURE 8. Facial Recognition and temperature finding

After achieved good accuracies in facial classification, we saved the model weights for temperature detection and further evaluation purposes. We used OpenCV techniques to detect the temperature on the facial recognition images shown in the above fig. 8.

CONCLUSION

In this paper, we introduce a new method for facial classification using DenseNet and temperature measurement. The evaluation results of our approach are summarized in the following section, along with a step-by-step explanation of the method. Our experimental findings demonstrate that our proposed approach, DenseNet-201, outperforms current state-of-the-art methods.

Our work consists of developing a system for facial recognition and temperature measurement using the infrared images. This architecture has been designed to implement the IR based temperature and facial recognition systems in a crowd areas like malls, offices, parks, exhibitions, etc., to detect and restrict entering. This approach achieved highest classification accuracy of 97.88%.

REFERENCES

1. J. A. Leñero-Bardallo, R. De la Rosa-Vidal, R. Padial-Allué, J. Ceballos-Cáceres, Á. Rodríguez-Vázquez and J. Bernabéu-Wittel, "A Customizable Thermographic Imaging System for Medical Image Acquisition and Processing," in IEEE Sensors Journal, doi: 10.110".
2. G. Mallikarjuna Rao, Praveen Kumar, Ankush Mittal, G. Vijaya Kumari, G. R. Babu. Exploration of Active Pixels in Intelligent Face Recognition for Constrained Mobile Systems. In Bhanu Prasad, Pawan Lingras, Ramakant Nevatia, editors, Proceedings of the 5th".
3. B. Ilikci, L. Chen, H. Cho and Q. Liu, "Heat-Map Based Emotion and Face Recognition from Thermal Images," 2019 Computing, Communications and IoT Applications (ComComAp), 2019, pp. 449-453, doi: 10.1109/ComComAp46287.2019.9018786.".
4. Friedrich, Gil and Yehezkel Yeshurun. "Seeing People in the Dark: Face Recognition in Infrared Images." Biologically Motivated Computer Vision (2002).".
5. Rao, G. & Kumar, Praveen & Kumari, G. & Pande, Amit & Babu, G.. (2011). An Investigation into the Use of Partial Face in the Mobile Environment. 526-535. 10.1007/978-3-642-24031-7_53.".

6. Li, Da, Carol C. Menassa, and Vineet R. Kamat. "Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography." Energy and Buildings 176 (2018): 246-261.".
7. "Jaddoa, Mohammed Ahmed, et al. "Automatic eyes localization in thermal images for temperature measurement in cattle." 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). IEEE, 2017.".
8. Dell'Isola, Giovanni Battista, et al. "Noncontact body temperature measurement: uncertainty evaluation and screening decision rule to prevent the spread of COVID-19." Sensors 21.2 (2021): 346.".
9. Cruz-Albarran, Irving A., et al. "Human emotions detection based on a smart-thermal system of thermographic images." Infrared Physics & Technology 81 (2017): 250-261.".
10. R. Roland Miezianko, "OTCBVS Benchmark Dataset Collection," [Online]. Available: http://vcipl-okstate.org/pbvs/bench/.
11. Zhuang, Fuzhen, et al. "A comprehensive survey on transfer learning." Proceedings of the IEEE 109.1 (2020): 43-76.".
12. Wang, Shui-Hua, and Yu-Dong Zhang. "DenseNet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16.2s (2".
13. Jaddoa, Mohammed Ahmed, et al. "Automatic temperature measurement for hot spots in face region of cattle using infrared thermography." ICINCO 2019-Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics. 2019.".
14. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012).".
15. Rao, G. & Babu, G.R. & Kumari, G. & Chaitanya, N.Krishna. (2009). Methodological Approach for Machine based Expression and Gender Classification. 2009 IEEE International Advance Computing Conference, IACC 2009. 1369 - 1374. 10.1109/IADCC.2009.4809216.".