*Avrupa Bilim ve Teknoloji Dergisi*
*Özel Sayı 36, S. 191-196, Mayıs 2022*

***Araştırma Makalesi***

www.ejosat.com ISSN:2148-2683

*European Journal of Science and Technology*
*Special Issue 36, pp. 191-196, May 2022*

***Research Article***

# Face Detection by Measuring Thermal Value to Avoid Covid-19

Kubilay Tuna [1*], Bayram Akdemir [2]

[1*] Konya Technical University, Facutly of Engineering and Natural Sciences, Department of Electrical Electronics Engineering, Konya, Turkey, (ORCID: 0000-0002-6448-074X), kubilaytuna26@hotmail.com

[2] Konya Technical University, Facutly of Engineering and Natural Sciences, Department of Electrical Electronics Engineering, Konya, Turkey, (ORCID: 0000-0002-0565-2345), bakdemir@ktun.edu.tr

(1st International Conference on Engineering and Applied Natural Sciences ICEANS 2022, May 10-13, 2022)

(**DOI:** 10.31590/ejosat.1113302)

**ATIF/REFERENCE:** Tuna, K. & Akdemir, B. (2022). Face Detection by Measuring Thermal Value to Avoid Covid-19. *European Journal of Science and Technology*, (36), 191-196.

**Abstract**

In this study, custom Single Shot Detection (SSD) was used to detect infected people faces because of Covid-19. It has been suggested to determine around the eyes area where the body temperature of the person most accurate by determining the facial landmarks with Ensemble of Regression Trees (ERT) model on these detected faces. Finally, the thermal value was measured from around the eyes area in a non-contact way using sensor fusion. As a result of the analyzes made, it was observed that the proposed system gave results close to the different measurement methods.

**Keywords:** SSD, Facial Landmark, ERT, Non-contact, Sensor Fusion, Covid-19.

# Covid-19'u Önlemek İçin Termal Değeri Ölçerek Yüz Tespiti

**Öz**

Bu çalışmada, Covid-19 nedeniyle enfekte olmuş kişilerin yüzlerini tespit etmek için özel Single Shot Detection (SSD) modeli kullanıldı. Tespit edilen bu yüzler üzerinde Ensemble of Regresyon Trees (ERT) modeliyle yüz işaret noktaları belirlenerek kişinin vücut sıcaklığının en doğru olduğu göz çevresinin tespit edilmesi önerilmiştir. Son olarak, termal değer, sensör füzyonu kullanılarak temassız bir şekilde göz çevresinden ölçülmüştür. Yapılan analizler sonucunda önerilen sistemin farklı ölçüm yöntemlerine yakın sonuçlar verdiği gözlemlenmiştir.

**Anahtar Kelimeler:** SSD, Yüz İşaret Noktaları, ERT, temassız, Sensör Füzyonu, Covid-19.

[*] Corresponding Author: kubilaytuna26@hotmail.com

# 1. Introduction

In these periods of rapidly growing societies, human beings and wildlife are increasingly intertwined with each passing day. In addition to, Pandemics, which have many examples in history, are one of the biggest problems of humanity. The most important step in the fight against such contagious diseases is to detect the diseased people as soon as possible and to prevent the transmission of the disease to more people (Hays, 2005).

The common symptom of almost all pandemic diseases is high fever. Thus, in the simplest way, the detection of people with a fever of 38 °C and above, which is described as abnormal body temperature, will eliminate the risk of transmission, and ensure that the epidemic is brought under control. However, it is difficult to control these people in environments where hundreds of people enter and exit. This situation requires labor and cost, causing time loss. Therefore, in the advanced technology world we live in, it is necessary to automate this process by eliminating human influence (Cai et al., 2020).

When the studies in the literature are examined, various approaches are presented regarding the detection of faces and facial landmarks.

Instead of treating the detection of facial landmarks as a single and independent problem, Zhang et al. (2014) suggested that detection robustness should be improved with multi-task learning. In particular, facial mark detection has been optimized with heterogeneous but finely correlated tasks. As a result of the practice, it has been observed that the proposed task-constrained learning outperforms existing methods, especially in dealing with faces with severe occlusion and pose change, and greatly reduces model complexity when compared to the state-of-the-art method based on progressive deep modeling. Li H. et al. (2015) proposed a CNN cascade model that can accurately distinguish real-world faces from the background by isolating them from large visual differences such as lighting. The detector evaluates the input image at low resolution to quickly reject non-face areas and carefully process difficult areas at higher resolution for accurate detection. Calibration nets have been gradually added to speed detection and improve bounding box quality. The proposed detector is very fast, reaching 14 FPS for typical VGA images on the CPU and can be accelerated up to 100 FPS on the GPU. Liu et al. (2015) proposed a new deep learning framework for facial feature estimation in the wild. The network consists of two different cascade CNN models. The LNET and ANET rungs are pre-trained differently. LNET is trained for face location estimation, while ANET is trained with large face IDs for feature estimation. The model is resistant to background confusion with designed pre-training strategies. Fast forward feed algorithm is used to save unnecessary computation. Yang et al. (2015) presented the WIDER FACE dataset, which is 10 times larger than existing datasets. The dataset contains rich descriptions, including gags, poses, event categories, and face bounding boxes. Faces in the proposed dataset are extremely difficult due to the large differences in scale, exposure, and occlusion. These factors are ubiquitous in many real-world applications. Therefore, the models trained with this dataset perform quite well in the real world. Ranjan et al. (2017) proposed an algorithm for simultaneous face detection, localization of landmarks, pose estimation and gender recognition using deep convolutional neural networks (CNN). The proposed method, called HyperFace, combines the interlayers of a deep CNN using a separate CNN and then follows a multi-task learning algorithm that works on the combined features. Extensive experiments show that the proposed models can capture both global and local information in hundreds, significantly outperforming many competing algorithms for each of these four tasks. Jiang and Learnend-Miller have recently proposed to apply Faster RCNN to face detection, which has shown impressive results on various object detection criteria. They trained a faster R-CNN model on the large-scale WIDER face dataset and compared it with the latest results in the WIDER test set. As a result of the study, it was concluded that although Faster R-CNN was designed for general object detection, it showed impressive face detection performance when retrained on an appropriate face detection training set (Jiang & Learned-Miller, 2017). Sun et al. proposed the application of a combined visible and thermal image processing approach that uses an IRT-equipped CMOS camera to screen patients with infectious diseases. An IRT system producing visible and thermal images was used to acquire the image. Subjects respiratory rates were measured by monitoring temperature changes around the nasal areas on thermal images; Facial skin temperatures were measured at the same time. As a result, the proposed system efficiently detected patients with suspected infectious diseases (Sun et al., 2017). Li et al. (2018) presented a new framework for real-time thermal comfort interpretation using infrared thermography. The main contribution of this work is the proposed data collection and analysis framework continuously and automatically acquire, retrieve, and analyze facial skin temperature data and interpret thermal comfort conditions for each building user in real operational environments. The proposed framework uses interdisciplinary techniques, including thermoregulation theory, computer vision, and machine learning. The results show that facial skin temperature collected from non-intrusive low-cost infrared thermal cameras can help to obtain a robust estimate of thermal comfort in real time, offering the possibility of synchronous control of indoor environments with minimal disruption to building works.

In this study proposes fast, efficient, and easy detection of risky people with a deep learning-based face detection system. Thus, the pandemic, which negatively affects the flow of daily life, can be reduced and disruptions can be eliminated. Single Shot Detector (SSD) and Ensemble of Regression Trees (ERT) were used as deep learning methods. The study consists of five parts in total. Section 1 is called the introduction and the literature studies are mentioned. The second part is the material and method part. Section 3 contains the findings and discussion. Section 4 contains the outcomes.

# 2. Material and Method

## 2.1. Dataset Description

In this study, SSD network is trained on the dataset, which is a combination of various open-source datasets and pictures, which included WIDER Face (Yang et al., 2015) and MAFA (Ge et al., 2017) datasets. All images were labelled and checked in accordance with the VOC format with a series of developed algorithms while preparing the dataset. The data set used for training was reviewed one by one and checked before reaching its final form.

*Fig. 1 Combined dataset*

## 2.2. Deep Learning

The Deep learning method is a popular machine learning approach that emerged with the deepening of ANN (Artificial Neural Networks) methods. In this method, the computational structure is based on ANN calculations and better success is achieved compared to classical ANN approaches.

In previous years, it was difficult to train ANNs due to limited hardware possibilities and data. Therefore, the layers in these models were less numerous. With the introduction of machines with high processing power into our lives, it became possible to train deeper models. Especially with the ability of the GPU to perform matrix operations quickly, models have started to be trained on GPUs, and the increasing number of data has made deep learning popular (Mesnil et al., 2014). Figure 2 shows the performances of deep learning and classical machine learning.

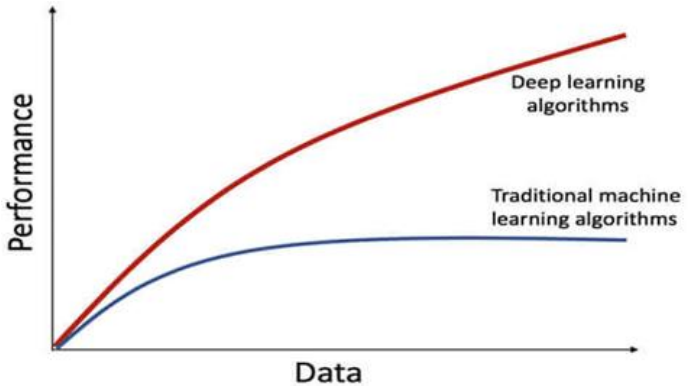


*Fig. 2 Comparison of deep learning and machine learning*

## 2.3. Convolutional Neural Network

CNN basically consists of two parts. The first part is the feature extraction that consists of one or more convolution and pooling layers. The second part is the part where the classification process consisting of fully connected layers is performed. Figure 3 shows the CNN stages.


*Fig. 3 CNN stages*

These networks are mostly used in the field of image classification. Estimates are made by extracting certain features, such as the human vision mechanism, as the operation of the network. Feature extraction from image pixels is difficult in image classification problems. CNN automatically runs this difficult process in the first part to obtain feature maps. Afterwards, the obtained features are transmitted to the fully connected layers, and certain number values are obtained thanks to the multi-layer detectors, and predictions are made about which class the image belongs to.

In the feature extraction phase, one or more convolution, pooling and Relu operations are applied sequentially.

The purpose of convolution layers is to extract feature maps by using pixel matrices of the image. In this layer, filters (weights) of different sizes such as 3x3, 5x5 are applied to the pixel matrices of the image, starting from the top left and descending to the bottom right, by shifting.

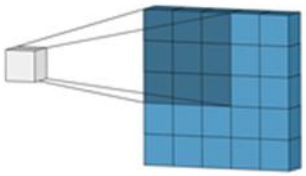

*Fig. 4 Convolution process*

The biggest problems encountered in the CNN method are overfitting and linearity. Overfitting means that the proposed deep learning model has high success on training data and low on test data, while linearity problem means that the obtained data cannot be separated linearly.

Pooling and dropout layers are used to prevent overfitting. The pooling process is applied in the next stage after the filtering process in the convolution layer during the feature extraction stage. The purpose of this process is to reduce the dimensions of the obtained feature maps. On the other hand, dropout deletes some of the neurons on the network during training. In this way, nerve cells are prevented from representing the same features, and each nerve cell is ensured to represent different features.

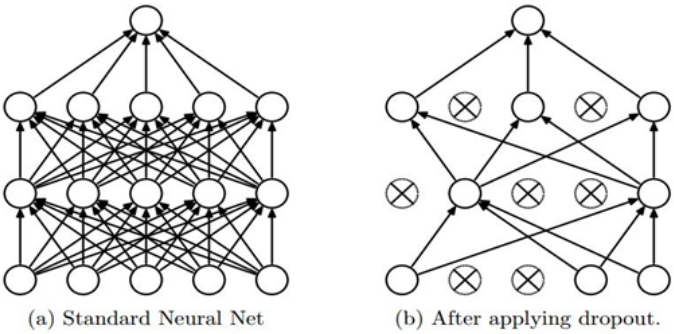


*Fig. 5 ANN model and dropout applied*

The Relu is active only when the input value is above a certain amount, thus avoiding linearity problem (Öztürk & Akdemir, 2019).

In the classification problem, the output of the network is equal to the number of classes used in training. Here, probability values are used to indicate which class the image belongs to from the feature map and which class it is more related to.

At this stage, softmax is used as a probabilistic class estimator. The softmax function is shown in Figure 6.

$$s\left(x_i\right) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$

*Fig. 6 Softmax function*

## 2.4. Single Shot Detector (SSD)

The SSD model, which appeared at the end of November 2016, has set new records in terms of performance and sensitivity for object detection tasks. It also showed 74% mAP (average precision) at 59 frames per second (FPS) in standard datasets such as Pascal VOC and COCO (Liu et al., 2016). As the name suggests, SSD works by considering three basic methods. These,

• Single shot: This means that object localization and classification tasks are done in a single forward pass of the network.

• Multibox: It is the name of the bounding box regression technique developed by Szegedy et al. (2016).

• Detector: Network is an object detector that also classifies detected objects.

As seen in Figure 7, the SSD architecture is based on the VGG-16 architecture. However, the fully connected layers in VGG have been discarded. The reason for using VGG-16 as the base mesh is its powerful performance in high-quality image classification tasks and helps to improve results with transfer learning. Instead of the fully connected layers in the original VGG, a few auxiliary convolution layers (as of conv6) have been added, allowing features to be extracted at multiple scales and gradually reduced the input size of each subsequent layer.

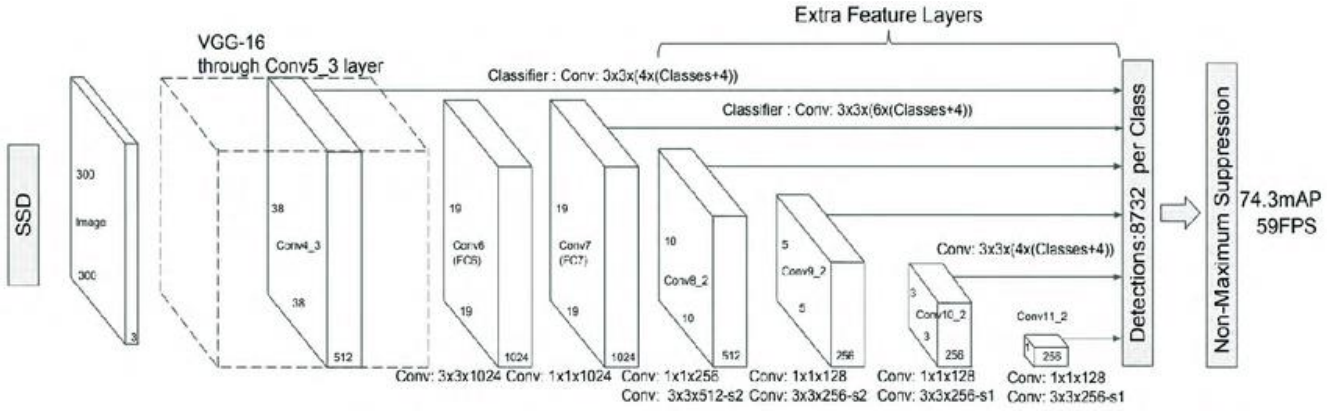


*Fig. 7 SSD architecture*

## 2.5. Facial Landmarks

Detection of facial markings is a method that is frequently used in almost every field recently. By identifying facial markings, face alignment, head pose estimation, face swapping, blink detection, etc. Many applications have been successfully implemented. Detection of facial markings is a subset of the shape prediction problem. Given an input image (and an ROI that normally indicates the object of interest), a shape predictor attempts to localize key points of interest throughout the shape. It was suggested to use the Ensemble of Regression Trees (ERT) model as the face point detector in this study (Kazemi & Sullivan, 2014). Figure 8 shows facial markings.

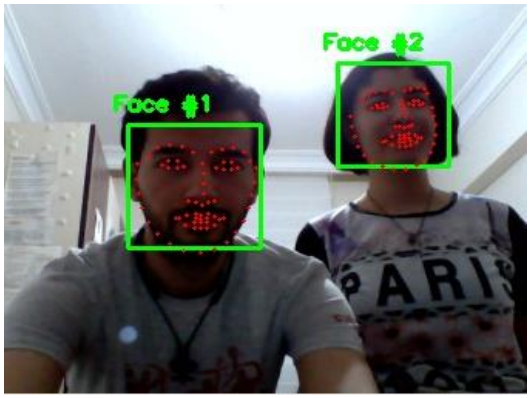


*Fig. 8 Facial landmarks*

## 2.6. Infrared Temperature Measurement Theory

Infrared (IR) energy is emitted by all materials above 0 °K. Infrared radiation is part of the Electromagnetic Spectrum and occupies frequencies between visible light and radio waves. The IR part of the spectrum covers wavelengths from 0.7 micrometers to 1000 micrometers (microns). In this waveband, only 0.7 micron to 20-micron frequencies are used for practical, everyday temperature measurement.

Although IR radiation is invisible to the human eye, it is useful to imagine it visible when dealing with measurement principles and evaluating applications because it behaves in many ways the same as visible light. IR energy travels in straight lines from the source and can be reflected and absorbed by material surfaces in its path. Radioactive heat exchange is shown in Figure 9.

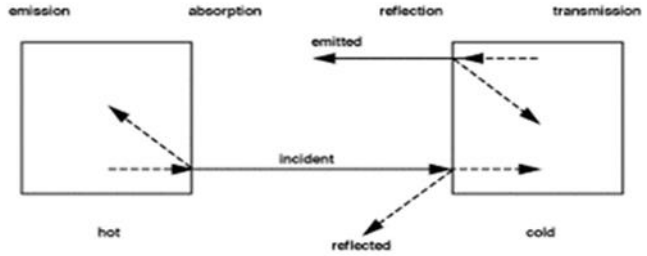


*Fig. 9 Radioactive heat exchange*

# 3. Results and Discussion

In this study, custom SSD model was trained using ResNet-10 like architecture as a backbone on the combined dataset. The dlib library, which includes the implementation of the ERT model, was used as a facial landmarks detector. In this way, temperature measurement was performed from around the eyes using the Optris CS LT infrared thermometer (sensor fusion). Finally, people with a measured temperature value of 38 °C and above were considered risky.


*Fig. 10 SSD architecture*

The proposed method has been implemented on Nvidia Jetson Nano and has been tested in real-time scenarios. The Nvidia Jetson Nano and dual-sensor camera (RGB+IR) are shown in Figure 11.

*Fig. 11 Nvidia Jetson Nano and dual-sensor camera*

As a result of these tests, it has been observed that the proposed method is robust, trustful, easy to use, being suitable for real-time applications (~15 FPS), saving time and resources.

The performance of the SSD model against occlusion is shown in Figure 12.


*Fig. 12 SSD model against occlusion*

The comparison of the proposed method with different measurement methods is shown in table

*Table 1. Temperature comparison between different measurement method*

| **Type of Thermometer** | **Body Side** | **Mean Temperature (˚C)** |
|---|---|---|
| Classic infrared thermometer | Forehead | 37.0 |
| Classic thermal thermometer | Around the eyes | 36.5 |
| **Sensor fusion (proposed method)** | **Around the eyes** | **36.0** |

## 4. Conclusions and Recommendations

In this study, a deep learning-based face detection system by measuring thermal value is proposed that can support human vision to avoid Covid-19.

When this system is implemented, it can be easily used for public health measurement and analysis by eliminating human influence in airports, hospitals, public buildings, shopping malls, educational institutions, justice palaces and penitentiary institutions, military institutions, universities, and any environment with a large human population.

## 5. Acknowledge

I would like to thank Asst. Prof. Bayram Akdemir for his valuable supports.

## References

Cai, Q., Huang, D., Ou, P., Yu, H., Zhu, Z., Xia, Z., ... & Chen, J. (2020). COVID-19 in a designated infectious diseases hospital outside Hubei Province, China. Allergy, 75(7), 1742-1752.

Ge, S., Li, J., Ye, Q., & Luo, Z. (2017). Detecting masked faces in the wild with lle-cnns. In Proceedings of the IEEE

conference on computer vision and pattern recognition (pp. 2682-2690).
Hays, J. N. (2005). Epidemics and pandemics: their impacts on human history. Abc-clio.
Jiang, H., & Learned-Miller, E. (2017, May). Face detection with the faster R-CNN. In 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017) (pp. 650-657). IEEE.
Kazemi, V., & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1867-1874).
Li, D., Menassa, C. C., & Kamat, V. R. (2018). Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography. Energy and Buildings, 176, 246-261.
Li, H., Lin, Z., Shen, X., Brandt, J., & Hua, G. (2015). A convolutional neural network cascade for face detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5325-5334).
Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In Proceedings of the IEEE international conference on computer vision (pp. 3730-3738).
Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector,‖ in European conference on computer vision (ECCV).
Mesnil, G., Dauphin, Y., Yao, K., Bengio, Y., Deng, L., Hakkani-Tur, D., ... & Zweig, G. (2014). Using recurrent neural networks for slot filling in spoken language understanding. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(3), 530-539.
Öztürk, Ş., ve Akdemir, B. (2019). Cell-type based semantic segmentation of histopathological images using deep convolutional neural networks. International Journal of Imaging Systems and Technology, 29(3), 234-246.
Ranjan, R., Patel, V. M., & Chellappa, R. (2017). Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE transactions on pattern analysis and machine intelligence, 41(1), 121-135.
Sun, G., Nakayama, Y., Dagdanpurev, S., Abe, S., Nishimura, H., Kirimoto, T., & Matsui, T. (2017). Remote sensing of multiple vital signs using a CMOS camera-equipped infrared thermography system and its clinical application in rapidly screening patients with suspected infectious diseases. International Journal of Infectious Diseases, 55, 113-117.
Yang, S., Luo, P., Loy, C. C., & Tang, X. (2015). From facial parts responses to face detection: A deep learning approach. In Proceedings of the IEEE international conference on computer vision (pp. 3676-3684).
Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2014, September). Facial landmark detection by deep multi-task learning. In European conference on computer vision (pp. 94-108). Springer, Cham.