

# Reducing false positive rate with the help of scene change indicator in deep learning based real-time face recognition systems

Mehmet Ali Kutlugün<sup>1</sup> · Yahya Şirin<sup>1</sup>

Received: 24 April 2022 / Revised: 25 April 2023 / Accepted: 2 May 2023 / Published online: 13 May 2023 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

# Abstract

In face recognition systems, light direction, reflection, and emotional and physical changes on the face are some of the main factors that make recognition difficult. Researchers continue to work on deep learning-based algorithms to overcome these difficulties. It is essential to develop models that will work with high accuracy and reduce the computational cost, especially in real-time face recognition systems. Deep metric learning algorithms called representative learning are frequently preferred in this field. However, in addition to the extraction of outstanding representative features, the appropriate classification of these feature vectors is also an essential factor affecting the performance. The Scene Change Indicator (SCI) in this study is proposed to reduce or eliminate false recognition rates in sliding windows with a deep metric learning model. This model detects the blocks where the scene does not change and tries to identify the comparison threshold value used in the classifier stage with a new value more precisely. Increasing the sensitivity ratio across the unchanging scene blocks allows for fewer comparisons among the samples in the database. The model proposed in the experimental study reached 99.25% accuracy and 99.28% F-1 score values compared to the original deep metric learning model. Experimental results show that even if there are differences in facial images of the same person in unchanging scenes, misrecognition can be minimized because the sample area being compared is narrowed.

**Keywords** Classification  $\cdot$  Deep metric learning  $\cdot$  Illumination and pose changes  $\cdot$  Image processing  $\cdot$  Real-time face recognition

 Mehmet Ali Kutlugün mehmet.kutlugun@std.izu.edu.tr
 Yahya Şirin yahya.sirin@izu.edu.tr

<sup>&</sup>lt;sup>1</sup> Istanbul Sabahattin Zaim University, Computer Science and Engineering, Istanbul, Turkey

# 1 Introduction

For years, face detection and facial recognition systems have attracted attention from scientific research groups and commercial application development teams. In particular, developing these systems' real-time applications is considered a very challenging issue [6, 35]. Advanced face recognition technologies are needed in many applications such as facial information retrieval, human-computer interface applications, multimedia communication, creation of artificial faces, and management of content-based face databases [7, 8, 23, 30].

Although excellent results are obtained when suitable conditions are provided, small changes in environmental conditions affect face detection and recognition problems negatively [11]. With the latest technological developments, many systems, including the locks of mobile phones, are integrated with the automatic face recognition system. A critical point is how to distinguish different human face data of different people in these applications or detect the similarity of two different facial images belonging to the same person [5, 37].

When speaking of representation for an image, one can think of features such as a vector of per-pixel intensity values or sets of edges. Some of these features may represent the data better, while others may represent it worse [26]. Deep Learning (DL) methods provide an advantage by using efficient hierarchical feature extraction algorithms that best represent the data instead of manually extracted features [36, 38].

In the recognition process, finding that the image queried matches the face in the database turns the face recognition process into a proximity calculation problem [34]. By comparing the distinctive features of the questioned picture with other feature sets in the face database. A decision is made considering all the calculated proximity values [11]. The quality of face recognition applications is directly proportional to the creation of the face recognition database, and the quality of the face poses obtained during the face recognition processes [8].

The most critical point of face matching is determining the representations of the features of two different faces that will ensure the correct matching [17, 25]. Therefore, a face recognition system is a recognition, verification, or classification system that creates a feature set to identify the person from the biometric features of the person and compares this feature set with another previously created data set [5, 32].

Image frames that move in real-time may undergo scene changes and environment changes from time to time. This does not mean that the general integrity is broken. However, some distortions or misidentifications may occur while the appropriate flow continues within the image set. In face recognition systems, light direction, reflection, and emotional and physical changes in facial expression can be the main factors that make recognition difficult [10, 29, 33, 41]. In this case, the system may tend to memorize or overfit as a different person instead of correctly.

Although the model generally defines the person correctly, it tends to resemble the person most similar from time to time in case of pose differences [15]. In other words, it does not know that the person actually has not changed because the vector with the most similarity in the dataset is returned as a result. In reality, only the facial expression of the same person has changed. An example is presented in Fig. 1 below. In Fig. 1, although there is no change of environment and person in the video image, the change in exposure angle in some frames causes misidentification.

In this study, a new model and perspective have been brought forward using the scene change indicator, proposed to reduce or eliminate the false recognition that may occur



Fig. 1 Example of misrecognition due to exposure angle differences

in sliding frames, as in Fig. 2. The application is thought to help reduce misidentification problems by dynamically changing the threshold value used in the face recognition phase, as shown in the figure below.

Classification of comparison vectors with appropriate values is an essential factor affecting False Positive (FP) rates and causing performance change [13, 25]. This model detects blocks where the scene has not changed and tries to more precisely determine the comparison threshold value used in the classifier stage with a new value. In summary, in this study, it is hypothesized that increasing the threshold sensitivity ratio between unchanging scene blocks will allow fewer comparisons between samples in the database, the classifier will avoid the tendency to return more similar results, and thus reduce the FP rates.



Fig. 2 False positive recognition in sliding frames

# 2 Related studies

Many studies have been conducted in the literature on the accuracy and performance of face recognition systems. In 1997, Lawrence et al. [27] first presented an Artificial Intelligence (AI) technique that uses Convolutional Neural Networks (CNN) to recognize faces. Since the 2010s, significant gains have been made in the performance and accuracy of DL-based systems.

In 2014, Y. Taigman et al. [42] Facebook presented a facial recognition system called DeepFace, developed in its research laboratory. This system is based on deep architecture for face recognition. This architecture used nine neural network layers with more than 120 million linked weights. They ran the CNN network with the most extensive dataset available, consisting of 4.4 million tagged images (4000 different people) generated from the internet. This network architecture was used to create DeepFace. This network works by applying the same CNN network to a pair of faces to obtain identifiers compared using Euclidean distance. The training aims to minimize the Euclidean distance between the same pair of faces and maximize the distance between 2 different faces. This is called metric learning [4, 40].

In 2015, the AI research group at Google presented FaceNet [16], a facial recognition system that could be implemented at a massive scale. This model is very similar to DeepFace. However, FaceNet uses a "triplet-based" loss function that minimizes the Euclidean distance between two faces of the same person and maximizes the Euclidean distance between a third face and the first two faces of a different person. This method achieved the best performance among all state-of-the-art methods, with an accuracy rate of 99.63% in the Labeled Faces in the Wild (LFW) dataset [4].

In 2016, B. Amos et al. [3] designed the OpenFace model called the face recognition library. This model combines Google's FaceNet architecture and pre-trained open source libraries. The system was tested on the LFW dataset, with an accuracy rate of 92.92%. Despite the smaller training dataset, this model shows competitive accuracy and performance results in the LFW validation comparison.

In 2016, the Dlib model, an open source facial recognition library created by Davis King [14, 22] and developed by Adam Geitgey [18] in 2017, was released. This model uses a version of a ResNet-34 network with 29 convolution layers with several layers removed and the number of filters per layer reduced by half. This library is built with state-of-the-art facial recognition trained with 3 million images. It has 99.38% accuracy on the LFW dataset. It uses embedding, which can be represented by a 128-dimensional vector space for each face image. Some generally accepted studies based on metric learning and their prominent features are presented in Table 1.

A study close to ours was published in 2022 by Chu et al. [9], which uses a deep metrics learning-based facial recognition system with the help of the scene changes module. In their study, they proposed the CFSM model, which detects scene change states in video images to achieve the target of face recognition and skips the frames that do not need to be re-recognized for the face.

According to the model, they said they would significantly reduce the workload of the real-time face recognition system in the embedded system. The application has a mechanism that uses MTCNN and FaceNet neural network models for the face detection and face extraction stages. They tested the model with ten video clips with square pixel dimensions of  $1280 \times 720$  from the Youtube-8 M dataset and reached an average accuracy of 0.896 [9].

Table 1 Comparison s	summary table of di	ifferent models	
Arch.	Year	Acc.(%)	Prominent Features
DeepFace	2014	97.35	It introduces the concept of metric learning for the first time using the CNN architecture created with a vast data set.
FaceNet	2015	99.63	It gave the best performance among the latest technology methods.
OpenFace	2016	92.92	It gives reasonable accuracy and performance results despite the smaller training dataset.
Dlib	2017	99.38	The computational cost has been reduced by reducing the number of filters and layers from the deep net- work model.

Ē
different
of
e
tab
summary
Comparison
_
ω.
0

Apart from these, some compelling studies are used for purposes other than face recognition systems, such as detecting scene changes [19, 28]. Detecting such abnormal situations is also crucial in face recognition systems since the light direction, image quality, and reflection affect the recognition process [10, 15, 26, 41]. The anomaly detection method recommended by Zhang et al. [43] can help overcome these issues by filtering out false positives.

Similarly, detecting face fraud has emerged as a new field. Zhou et al. [44] proposed a new algorithm and created a large-scale dataset to overcome this task. Zhou et al. [45] also achieved promising results by proposing a cascade parsing network (CP-HOI) to address detecting and recognizing human-object interactions (HOI) in images. Also, the salience map of facial features can help improve face detection accuracy. Jia et al. [21] proposed a deep CNN architecture and saliency map method to evaluate image quality and obtained improved results on different datasets.

The research shows that the threshold value determined at the beginning is made with a constant comparison rate throughout the entire video processing process [3, 9, 14, 16, 18, 22]. Although it is known that changing the classification threshold has severe effects on performance, we could not find any other study in the literature [13] that compares dynamically changing threshold values. In this respect, our work is unique.

#### 3 Experimental method

#### 3.1 Aim

The most suitable model for the snapshot flow is expected to work with high performance in real-time face recognition systems. Over time, losses may occur in frames that change instantaneously, leading to erroneous recognition. This study proposes an SCI to support the face recognition system and reduce or eliminate false recognition in sliding windows. According to this model, a new perspective is brought to the behavior of the face recognition system with the help of a new parameter obtained from scene changes. It aims to ensure that the recognition behavior does not change unless there is a significant change on the scene.

#### 3.2 Model and architecture

In the proposed method, it has been suggested that scene change movements may affect recognition sensitivity. The method works based on dynamically changing the proximity value without deviating from the focus in real-time video images with a new parameter. In order to obtain the mentioned parameter, a block called scene change indicator (SCI) is used. Once scene changes are detected, the SCI module decides the sensitivity ratio of the current frame's classification threshold across whole unchanged scenes. In other cases, the default value is taken into account. The SCI module is described in the next section. The diagram showing the methodology of the proposed model is presented in Fig. 3.

The Histogram of Oriented Gradients (HOG) algorithm, developed by Dalal and Triggs [2, 12, 46], was preferred for face detection. Compared to the Viola-Jones [39] algorithm, this algorithm works faster and more accurately under many different conditions such as lighting, scaling, and shading.



Fig. 3 Architecture of the proposed model

Based on deep metric learning, the open source face recognition library Dlib [14, 18, 22] was used to obtain facial features. K-NN algorithm was used as a classifier also [24].

The algorithm of the proposed method is presented in Algorithm 1.

In the algorithm,  $D^{c,r}$  is used to determine the number of different blocks between the current frame and the previous frame.  $Tf_{imm}$  is used for comparison to detect whether a scene change has occurred.  $Td_{block}$  is used to determine the number of different blocks between the current frame and the referenced frame. If  $D^{c,r}$  is greater than the predefined  $Tf_{imm}$  the state of the scene has changed. Otherwise, it means the scene has not changed.

#### 3.3 SCI Block

This study used a pixel-based method to find scene transitions between video scenes. By thresholding, the absolute change of pixel values with a particular value, pixels that show too much change and therefore do not have scene transition features are distinguished as follows [19].

$$D_{i,j}^{p} = |I_{t}(x, y) - I_{t-1}(x, y)|$$
(1)

```
Input: InputVideoFrame(V)
Output: Face Recognition
count_of_face: 0
procedure SCI(V)
         Input Features(V)
         Initial threshold: Tdblock, Tfimm, default threshold, compare new threshold
         D^{c,r} = SCD(current frame, previus frame, Td_{block})
// Td<sub>block</sub> is used to determine the number of different blocks between the current frame and
// reference frame. D^{c,r} is the pixel-level difference between the (i,j)th previous and current blocks.
         if(D^{c,r} < Tf_{imm}) then
                   FaceRecognition(current frame, compare new threshold)
         else if(D^{c,r} > Tf_{imm}) then
                   FaceRecognition(current frame, default classifier threshold )
         end if
end procedure
//T_{f_{imm}} is used to determine if a sudden scene change has occurred. When the threshold value is not
// exceeded, there is no abrupt scene change.
function SCD(current, reference, Tdblock)
          \{ B_{0,0}, B_{0,1}, B_{0,2}, \dots, B_{N-1}, B_{M-1} \} = reference to NxM block
          for n=0 to N-1
                   for m=0 to M-1
                            \Delta^{c_{n,r}}_{n,m} = abs(B^{c_{n,m}} - B^{r_{n,m}})
                             Increase D^{c,r}, if \Delta^{c,n,r} > Td_{block}
                   end for
          end for
         Return Dc,r
end function
                   // This function counts the number of distinct blocks between adjacent frames.
function FaceRecognition(current frame)
         [count of face, bounding box] = FaceDetection(current frame)
          face encoding = FeatureExtraction(bounding box)
         FaceComparing(face_encoding)
         reset count of face
end function
                   // In this function, detecting and recognizing faces are performed.
```

Algorithm 1 The algorithm of SCI

$$M_t(x, y) = \begin{cases} 1, \ D_{i,j}^p < th_m \\ 0, \ other \end{cases}$$
(2)

 $D_{i,j}^{p}$  is the pixel-level difference between the (i,j)-th previous and current blocks.  $I_{t}$  and  $I_{t-1}$  represent consecutive image frames at moments t and (t-1) in the video sequence, respectively. The pixels' information to detect the scene transitions in the created  $M_{t}$  mask is kept [19]. The  $th_{m}$  used here represents the threshold value. Instead of pixel-based scene change detection, a histogram-based [28] or different detector can also be used for this module.

# 4 Experimental study

### 4.1 Dataset

In the application phase of the model, images of Adile Naşit, Ata Demirer, Mehmet Öz, Mesut Özil, and Naim Süleymanoğlu from the Turkish celebrities dataset and Aleksandra

Wozniak, Maria Sharapova, Martin Palermo, Matt Smith and Morgan Freeman from the Youtube-8 M dataset [1] were used. A data set of approximately 419 people was created by further addition to this dataset of our images, other images: we collected from our close circle and some more people from the LFW [20] dataset. Each of these datasets contained an average of ten pictures. An example of the dataset is shown in Fig. 4.

The database containing 128-dimensional face vectors (embeddings) from this trained dataset was tested with 12 video sets that the system had not seen before, containing video images of the celebrities mentioned above for real-time testing. The attributes of these video clips are shown in Table 2. In these test videos, preprocessing was done. Firstly, determining manually the scene transition places and the places where the scene changed very little. Secondly, performance comparisons were made with the proposed model, especially for the frames where the scene did not change.

#### 4.2 Implementation

The application has been carried out to evaluate the face recognition performances of the classifier according to different sensitivity ratios by using test videos in the dataset. The comparison ratio is considered low if the same person is in the flowing image and the scene does not change more than the determined threshold value. In short, the classifier threshold value is changed under appropriate conditions by utilizing the scene change information. The default tolerance value is 0.6 according to the Dlib library, and lower numbers make face comparisons more strict [3]. The indicator moves with the default value unless there is a background change. In each change, the sensitivity rate is reset to the default value. The hardware/operating system information used for the experiments in this study is presented in Table 3. Experimental application settings are as follows:

 $Tf_{imm} = \%30$ , default\_classifier\_threshold = 0.6, compare\_new\_threshold = 0.55 (high sensitivity value), compare\_low\_sensitivity\_threshold = 0.65.

#### 4.3 Experimental Results

We present the results obtained from the experimental study in some graphs and tables below. First, we shared the findings from comparisons with different threshold values.

The scene switching module is not used in the examples shown in Fig. 5a. Although there was no change in the scene, misidentification occurred in some image frames by testing with



Fig. 4 Example of Dataset

Table 2	The attributes of the dataset						
Video Id	Name of the video clip	Duration (s)	Num- ber of frames	FPS	Dimension	Category	Characteristics of the video clip
_	"Adile Na.it"	58	892	25	$854 \times 468$	Cinema / Theater	It includes a part from the movie "Dream Family" starring the art- ist. In this episode, different people are shown in different scenes.
7	"Aleksandra Wozniak"	134	4028	30	1280×720	Sports - Tennis	The player makes statements to the fixed camera in an interview. Original title in Youtube-8 M dataset: "WTA Tennis Star and USANA Brand Partner Aleksandra Wozniak talks about her life."
ŝ	"Ata Demirer"	113	2843	25	$640 \times 360$	Cinema / Comedy	It includes images taken with different camera angles in the stand- up show of the artist named "One Man Giant Staff."
4	"Maria Sharapova"	246	6172	25	854×480	Sports- Tennis	It includes a part of a match where the player plays with his oppo- nent. Original title in Youtube-8 M dataset:" 2015 Internazionali BNL d'Italia Final WTA Highlights   Maria Sharapova vs Carla Suarez Navarro"
Ś	"Martin Palermo"	127	3175	25	640×480	Sports- Football	It includes the player's farewell speech captured with a motion cam- era at the stadium after a match. Original name in Youtube-8 M dataset: "Palermo en la Bombonera, la despedida"
9	"Matt Smith"	140	3504	25	$1280 \times 720$	Cinema / Theater	It includes the actor's scenes Matt Smith's from BBC's Doctor Who. Original name in Youtube-8 M dataset: "Matt Smith Trib- ute Doctor Who HD."
7	'Mehmet Oz"-1	297	7444	25	$640 \times 360$	Health	Dr. Öz gives information about the Coronavirus protection plan to a fixed camera with different scenes.
8	"Mehmet Oz"-2	182	4555	25	$640 \times 360$	Health	Dr. Öz is interviewed with a moving camera outdoors.
6	"Mesut Ozil"	94	2352	25	$640 \times 360$	Sports- Football	The player is interviewing a live camera in front of a post-match advertising booth.
10	"Morgan Freeman"	161	3876	24	$1280 \times 720$	Cinema	It contains footage from a movie clip shot with a motion camera containing different scenes played by the actor. Original title in Youtube-8 M dataset: "Gone Baby Gone (6/10) Movie CLIP - To Lose a Child (2007) HD."

 $\underline{\textcircled{O}}$  Springer

Characteristics of the video clip	It contains scenes from the player's weightlifting competition in Cardiff for the European Championships.	It includes summary match images of different players taken with a motion camera in the quarter-finals. Original title in Youtube-8 M dataset: "2015 Aegon International Eastbourne Quarterfinal WTA Highlights"
Category	Sports- Weightlifting	Sports- Tennis
Dimension	$640 \times 360$	854×480
FPS	30	25
Num- ber of frames	1955	8103
Duration (s)	65	324
Name of the video clip	"Naim Stileymanoglu"	"Wozniak and Sharapova"
Video Id	11	12

Table 3Hardware and OSinformation table	Hardware:		
	Manufacturer and Model	:	HP-EliteBook-840-G4
	System Type	:	×64-based PC
	Processor	:	Intel(R)Core(TM) i5-7300U CPU @ 2.60GHz
	Graphics	:	Intel® HD Graphics 620
	Physical Memory	:	8.00 GB (RAM)
	Hard Disk	:	250 GB
	OS Type	:	Ubuntu 16.04 LTS 64-bit

the default comparison threshold or sensitivity. By applying the proposed model to unchanging scenes, we achieved correct identification in the same frames as seen in Fig. 5b.

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{3}$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{4}$$

$$F_1 = 2x \frac{PxR}{P+R} \tag{5}$$

N

Frame:		2306	2311	2316	2321	
Name		Mesut Özil	Mesut Özil	Mali	Mesut Özil	Mesut Özil
Thr:0.6		ТР	ТР	FP	ТР	TP
	·· ·· ·· ·· ·· ·· ·· ·· ·· ·· ·· ·· ··					
				(a)		
Name		Mesut Özil				
Thr:0.55		ТР	ТР	ТР	TP	TP
	4 N					

(b)

Fig. 5 Comparison of different threshold values in real-time video images. (a) with threshold value:0.6 (b) with threshold value:0.55

$$S = \frac{N_{TN}}{N_{TN} + N_{FP}} \tag{6}$$

In Figs. 6 and 7, we graphically show all TP and FP rates occurring throughout the video according to two different comparison rates. In these sections, we compared the two methods by choosing the most misidentified people as FP1 and FP2. To make the changes in the graphs more understandable, we have shown the FP ratios in the negative region and the TP ratios in the positive region.

Second, the application results are given in Table 4 to show how the sensitivity ratio changes at different threshold values. In this table, we evaluated scene change detection accuracy using the most popular evaluation criteria: Precision (P), Recall (R), Specificity (S) and F1 score ( $F_1$ ) [28, 31]. We achieved an average specificity of approximately 87.21% in the proposed model. These evaluation metrics are defined as:

The default threshold value of 0.6 in Fig. 6a was dynamically converted to 0.55 in the unchanged scene segments in Fig. 6b. Accordingly, we have shown the TP and FP ratios distribution in the video in the graphics above.

In another test video made with the SCI method, we saw that the FP ratios shown in the negative region in Fig. 7a completely disappeared in Fig. 7b.

Third, we presented the comparison results with two different approaches similar to our model Tables 5, 6 and 7.

We also presented the graph of FP change according to different models in all test videos in Fig. 8.

The average performance graph of different models is as in Fig. 9.

### 5 Conclusions and Discussion

This study proposes a new approach to reducing false identifications in real-time face recognition systems during snapshot changes. The primary purpose is to teach the system that the person in the scene has not changed by discovering additional measures to prevent false face recognition. According to this approach, allowing for scene changes in successive images is included in the real-time face recognition system. For this purpose, experiments were conducted with different threshold values used in the classifier stage and how the FP ratios changed. A more sensitive classifier is obtained by taking this threshold value from 0.6 in the default model to 0.55 in the unchanged scene sections. In the model proposed, we reached 99.25% accuracy and 99.28% F-1 score values compared to the original deep metric learning model. Experimental results show that even if there are differences in facial images of the same person in unchanged scenes, false recognition can be minimized as the compared sample area is narrowed. When the performance table is examined, it shows that the proposed model works as well as the original dlib model and even gives better results. Therefore, it can be said that the method is compatible.

**Limitations and future research directions** Some limitations of the proposed system are explained in the following:

The absence of equivalent numbers of samples belonging to different classes in the database is a factor that directly affects the k-nn classifier. Uneven data distribution will limit the performance of the model.



**Basic Model** 

Fig.6 Comparison of different model thresholds for video id-3. (a) with threshold value: 0.6 (b) with threshold value: 0.55

Since reference regions taken from the background are used in detecting scene changes, the possibilities of this module may not be used in videos with a fixed background.

Applying an over-sensitivity rate may result in unknown recognition rather than the person expected to be recognized. In this study, experiments were carried out with three



Fig.7 Comparison of different model thresholds for video id-7. (a) with threshold value: 0.6 (b) with threshold value: 0.55

different values. Other studies may be required to determine the ideal ratio according to the state of the database.

In future studies, it is planned to obtain new indicators showing scene changes with models different from the method used.

#### 47531

Video Id	Precision			Recall			Accuracy			Fl-Score		
	Thr. (0.65)	Thr. (0.6)	Thr. (0.55)	Thr. (0.65)	Thr. (0.6)	Thr. (0.55)	Thr. (0.65)	Thr. (0.6)	Thr. (0.55)	Thr. (0.65)	Thr. (0.6)	Thr. (0.55)
1	0,9929	0,9929	0,9971	0,9957	0,9957	10,000	0,9913	0,9913	0,9972	0,9943	0,9943	0,9986
2	0,9997	10,000	10,000	10,000	10,000	10,000	0,9997	10,000	10,000	0,9999	10,000	10,000
3	0.9834	0,9885	0,9919	0,9970	0,9978	10,000	0,9814	0,9867	0,9919	0,9902	0,9931	0,9959
4	0,9314	0,9279	0,9518	0,9794	0,9717	10,000	0,9694	0,9572	0,9669	0.9548	0,9493	0,9753
5	0,8746	0,9314	0,94,154	0,9684	0,9855	0,9954	0,9688	0,9733	0,9873	0,9191	0,9576	0,9698
9	0,9867	0,9808	0,9949	0,9952	0,9980	0,9990	0,9852	0,9897	0,9940	6066,0	0,9894	0.9970
7	0,9996	0,9998	10,000	10,000	10,000	10,000	0,9996	0,9998	10,000	0,9998	6666'0	10,000
8	0,9884	0,9927	10,000	0,9992	10,000	10,000	0,9879	0,9927	10,000	0,9938	0,9963	10,000
6	0,9413	0,9730	0,9877	0,9784	0,9930	0,9954	0,9617	0.9769	0,9974	0,9595	0,9829	0,9916
10	0,9925	0,9924	0,9950	0,9979	0,9979	0,9979	0,9943	0,9928	0,9948	0,9952	0,9951	0,9965
11	0,9437	0,9889	0,9974	0,9888	10,000	10,000	0,9575	0,9910	779977	0,9657	0,9944	0,9987
12	0,9685	0,9796	0,9898	0,9915	0,9902	0,9912	0,9700	0,9754	0,9823	0,9798	0,9849	0,9905
Average	0,9669	0,9790	0,9876	0,9910	0,9942	0,9982	0,9806	0,9856	0,9925	0,9786	0,9864	0,9928

 Table 4
 Performance table

Video Id	Video Name	CFSM Accuracy
1	"Changeling"	0,92
2	"Before Sunset"	0,96
3	"The International"	0,97
4	"Patriot Games"	0,95
5	"Say Anything"	0,96
6	"Instant Family"	0,86
7	"Funny People"	0,88
8	"City Slickers"	0,74
9	"Uncle drew"	0,85
10	"Animal house"	0,87

# **Table 6**Performance table ofdefault Dlib with K-nn model

 Table 5
 Performance table of

CFSM model

Video Id	Video Name	Dlib+K-nn
		Accuracy
1	"Adile Naşit"	0,99
2	"Aleksandra Wozniak"	1,00
3	"Ata Demirer"	0,99
4	"Maria Sharapova"	0,96
5	"Martin Palermo"	0,97
6	"Matt Smith"	0,99
7	"Mehmet Öz"-1	1,00
8	"Mehmet Öz"-2	0,99
9	"Mesut Özil"	0,98
10	"Morgan Freeman"	0,99
11	"Naim Süleymanoğlu"	0,99
12	"Wozniak and Sharapova"	0,98

Video Id	Video Name	SCI
		Accuracy
1	"Adile Naşit"	1,00
2	"Aleksandra Wozniak"	1,00
3	"Ata Demirer"	0,99
4	"Maria Sharapova"	0,97
5	"Martin Palermo"	0,99
6	"Matt Smith"	0,99
7	"Mehmet Öz"-1	1,00
8	"Mehmet Öz"-2	1,00
9	"Mesut Özil"	1,00
10	"Morgan Freeman"	0,99

"Naim Süleymanoğlu"

"Wozniak and Sharapova"

11

12

# 

1,00

0,98



Fig. 8 The average performance graph of different models



Fig. 9 The average performance graph of different models

Abbreviations Acc: Accuracy; Arch: Architecture; AI: Artificial Intelligence; CFSM: Continuous Frames Skipping Mechanism; CNN: Convolutional Neural Networks; DL: Deep Learning; FP: False Positive; HOG: Histogram of Oriented Gradient; K-NN: K-Nearest Neighbors; LFW: Labeled Faces in the Wild; SCI: Scene Change Indicator; Thr: Threshold; TP: True Positive; TN: True Negative

Data availability The following information was supplied regarding data availability: The images and videos are available at LFW and YouTube-8 M Databases: http://vis-www.cs.umass.edu/lfw/, https://research.google.com/youtube8m/explore.html **Conflict of interest** It is to state that "No Competing interests are at stake and there is No Conflict of Interest" with other people or organizations that could inappropriately influence or bias the paper's content. None of the authors of this paper has a financial or personal relationship with other people or organizations that could inappropriately influence or bias the paper's content.

# References

- Abu-El-Haija S, et al. (2016) Youtube-8m: A large-scale video classification benchmark. arXiv preprint arXiv:1609.08675
- Al Machot F, Ali M, Haj Mosa A, Schwarzlmüller C, Gutmann M, Kyamakya K (2019) Real-time raindrop detection based on cellular neural networks for ADAS. J Real-Time Image Proc 16(4):931–943
- Amos B, Ludwiczuk B, Satyanarayanan M (2016) Openface: A general-purpose face recognition library with mobile applications, CMU School of Computer Science, Tech. Rep., CMU-CS-16-118.
- 4. Ananthakumar A (2018) Efficient face and gesture recognition for time sensitive application. In 2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), IEEE, pp.117-120.
- Atik M, Duran Z (2020) 3D face recognition and accuracy analysis using local feature-based methods. J Fac Eng Archit Gazi Univ 36(1):359–372
- Bhattarai S, Madanayake A, Cintra RJ, Duffner S, Garcia C (2017) Digital architecture for real-time CNN-based face detection for video processing. In 2017 Cognitive Communications for Aerospace Applications Workshop (CCAA), IEEE, pp1–6.
- Bilgiç A, Kurban OC, Yıldırım T (2017) Face recognition classifier based on dimension reduction in deep learning properties. In: 25th Signal Processing and Communications Applications Conference (SIU), IEEE, Antalya, Turkey, pp 1–4. https://doi.org/10.1109/SIU.2017.7960368
- Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA (2020) Albumentations: fast and flexible image augmentations. Information 11(2):125
- 9. Chu SL, Chen CF, Zheng YC (2022) CFSM: a novel frame analyzing mechanism for real-time face recognition system on the embedded system. Multimed Tools Appl 81(2):1867–1891
- Cuculo V, D'Amelio A, Grossi G, Lanzarotti R, Lin J (2019) Robust single-sample face recognition by sparsity-driven sub-dictionary learning using deep features. Sensors 19(1):146
- Dabhade RG, Waghmare LM (2017) Optimal Neural Network Based Face Recognition System for Various Pose and Occluded Images. Int J Appl Eng Res 12(22):12625–12636
- 12. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection, inProc. IEEE Int. Conference on Computer Vision Pattern Recognition, pp. 886–893
- Datta AK, Datta M, Banerjee PK (2015) Face detection and recognition: theory and practice. CRC Press Taylor & Francis Group, New York
- Davis EK (2017) High Quality Face Recognition with Deep Metric Learning. Blog.dlib.net. http:// blog.dlib.net/2017/02. Accessed: 2 Feb. 2020.
- Elaggoune H, Belahcene M, Bourennane S (2022) Hybrid descriptor and optimized CNN with transfer learning for face recognition. Multimed Tools Appl 81(7):9403–9427
- Florian S, Dmitry K, James P (2015) Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 815–823.
- Gao Y, Lee HJ (2019) Pose-invariant features and personalized correspondence learning for face recognition. Neural Comput Applic 31(1):607–616
- Geitgey A (2017) Ageitgey Face Recognition. Github. https://github.com/ageitgey/face\_recognition. Accessed: 2 Feb. 2020.
- Güllü MK, Urhan O, Ertürk S (2004) Dissolve detection for archive films. Proceedings of the IEEE 12th Signal Processing and Communications Applications Conference, IEEE, Kusadasi, Turkey, pp. 510-513. https://doi.org/10.1109/SIU.2004.1338577.
- Huang GB, et al. (2008) Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In: Workshop on faces in'Real-Life'Images: detection, alignment, and recognition
- Jia S, Zhang Y (2018) Saliency-based deep convolutional neural network for no-reference image quality assessment. Multimed Tools Appl 77(12):14859–14872
- 22. King DE (2009) Dlib-ml: A machine learning toolkit. J Mach Learn Res 10(3):1755-1758
- Koestinger M, Wohlhart P, Roth PM, Bischof H (2011) Annotated facial landmarks in the wild: A largescale, real-world database for facial landmark localization. In 2011 IEEE international conference on computer vision workshops (ICCV workshops), IEEE, pp.2144-2151.

- 24. Kutlugün MA (2017) Gözetimli Makine Öğrenmesi Yoluyla Türe göre Metinden Ses Sentezleme. Yüksek Lisans Tezi, İstanbul Sabahattin Zaim Üniversitesi
- Kutlugün MA, Şirin Y (2023) Augmenting the training database with the method of gradual similarity ratios in the face recognition systems. Digital Signal Processing, https://doi.org/10.1016/j.dsp.2023. 103967.
- Kutlugün MA, Şirin Y, Karakaya MA (2019) The Effects of Augmented Training Dataset on Performance of Convolutional Neural Networks in Face Recognition System. Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE, Leipzig, Germany, pp 929–932. https://doi.org/10.15439/2019F181
- Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: A convolutional neural-network approach. IEEE Trans Neural Netw 8(1):98–113
- 28. Lee HS, Cho SI (2022) Luminance Level of Histogram-Based Scene-Change Detection for Frame Rate Up-Conversion. IEEE Access,
- Li Y, Guo K, Lu Y, Liu L (2021) Cropping and attention based approach for masked face recognition. Appl Intell 51(5):3012–3025
- Moghaddam B, Pentland A (1999) Bayesian image retrieval in biometric databases. Int Conf Multimed Comput Syst, IEEE 2:610–615
- Monaghan TF et al (2021) Foundational statistical principles in medical research: sensitivity, specificity, positive predictive value, and negative predictive value. Medicina 57(5):503
- Ozkaya N, Sagıroglu Ş (2008) Face recognition from fingerprints. J Fac Eng Archit Gazi Univ 23(4):785–793
- Patil H, Kothari A, Bhurchandi K (2016) Expression invariant face recognition using semidecimated DWT, Patch-LDSMT, feature and score level fusion. Appl Intell 44(4):913–930
- 34. Qi X, Liu C, Schuckers S (2018) CNN based key frame extraction for face in video recognition. In 2018 IEEE 4th International Conference on Identity Security and Behavior Analysis (ISBA), IEEE, pp.1-8
- Saypadith S, Aramvith S (2018) Real-time multiple face recognition using deep learning on embedded GPU system. In: 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE, pp1318–1324
- Şeker A, Diri B, Balık HH (2017) A review of deep learning methods and applications. Gazi J Eng Sci 3(3):47–64
- Soltanpour S, Boufama B, Wu QJ (2017) A survey of local feature methods for 3D face recognition. Pattern Recogn 72:391–406
- 38. Song HA, Lee SY (2013) Hierarchical Representation Using NMF. International Conference on Neural Information Processing, pp. 466–473
- 39. Viola P, Jones M (2004) Robust real-time object detection. Int J Comput Vis 57:137-154
- 40. Vu HN, Nguyen MH, Pham C (2021) Masked face recognition with convolutional neural networks and local binary patterns. Appl Intell 1-16
- 41. Wang K, Chen Z, Wu QM, Liu C (2019) Face recognition using AMVP and WSRC under variable illumination and pose. Neural Comput Applic 31(8):3805–3818
- Yaniv T, Ming Y, Marc'Aurelio R, Lior W (2014) DeepFace: Closing the gap to human-level performance in face verification, In Computer Vision and Pattern Recognition (CVPR), IEEE, 1701–1708.
- Zhang Y, Hutchinson P, Lieven NA, Nunez-Yanez J (2019) Adaptive event-triggered anomaly detection in compressed vibration data. Mech Syst Signal Process 122:480–501
- Zhou T, Wang W, Liang Z, Shen J (2021) Face forensics in the wild. In Proceedings of the IEEE/ CVF conference on computer vision and pattern recognition (CVPR), pp. 5778-5788. https://doi.org/ 10.1109/CVPR46437.2021.00572
- Zhou T, Qi S, Wang W, Shen J, Zhu SC (2021) Cascaded parsing of human-object interaction recognition. IEEE Trans Pattern Anal Mach Intell 44(6):2827–2840. https://doi.org/10.1109/TPAMI.2021. 3049156
- Zhu X, Meng Q, Gu L (2018) Real-time image recognition using weighted spatial pyramid networks. J Real-Time Image Proc 15:617–629. https://doi.org/10.1007/s11554-017-0743-y

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.