Digital Signal Processing 135 (2023) 103967

Contents lists available at ScienceDirect

Digital Signal Processing

journal homepage: www.elsevier.com/locate/dsp

Augmenting the training database with the method of gradual similarity ratios in the face recognition systems

Mehmet Ali Kutlugün *, Yahya Şirin

Istanbul Sabahattin Zaim University, Computer Science and Engineering, Istanbul, Turkey

ARTICLE INFO

Article history:
Available online 10 February 2023

Dataset link: http://vis-www.cs.umass.edu/lfw

Dataset link: https://youtube.com

Keywords:
Classification
Data augmentation
Deep metric learning
Face recognition
Feature extraction
Image processing

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. Deep metric learning algorithms called representative learning are frequently preferred in this field. However, in addition to the model's success in feature extraction, factors such as the distribution of samples in this database and appropriate classifier preferences also affect the overall performance of the face recognition system. This study it is aimed to create integrity in the database of a pre-trained deep neural network model by obtaining augmented data for classes with a limited number of samples. Thanks to this method called Graded Similarity Rates (GSR), augmented data that could disrupt class integrity has been removed from the database. This way, classes with limited examples are kept integrity, and classifier behavior is used more effectively. The model proposed in the experimental study reached 99.38% accuracy values compared to traditional data augmentation models. Experimental results have shown that the database has an acceptable level of success even at smaller vector sizes and is more organized.

1. Introduction

The basis of facial recognition systems is to extract facial features and make comparisons using these features. Deep Learning (DL) models offer an approach that has successfully extracted the features of the analyzed data [1–4]. Advanced face recognition technologies are needed in many applications, such as obtaining face information, human-computer interface applications, multimedia communication, creating artificial faces, and managing content-based face databases [5]. Proximity values of each comparison process are calculated by comparing the distinctive features of the image questioned in the recognition process with the other feature sets in the face database. Then, the result is obtained by considering all these calculated proximity values. 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 in the face recognition processes [6–10]. The most critical point of face matching is determining the representations of the features of two different faces that will ensure correct matching [8]. Although excellent results are obtained when suitable conditions are provided, small changes in environmental conditions adversely affect face recognition processes [5].

For many computer vision tasks, increasing the data size is a crucial way to expand the training dataset, allowing the model to learn different conditions and preventing "overfitting" situations defined as memorization or overfitting [11]. This approach is the easiest for classification. In this way, the diversity of data is increased by artificial methods and significantly affects learning performance. Pictures are subject to some distortions with data augmentation processes. Operations such as angular rotation, perspective-changing, panning, and zooming are performed on the picture to increase the variety of pictures [3,12].

This study aims to create integrity in the database of a pre-trained deep neural network model by obtaining augmented data for classes with a limited number of samples. Thanks to this method called Graded Similarity Rates (GSR), augmented data that could disrupt class integrity has been removed from the database. This way, classes with limited examples are kept in integrity, and classifier behavior is used more effectively. In summary, the contributions of the study can be presented as items such as:

- optimally augmenting classes with limited data
- limiting excessive data augmentation
- more efficient use of the classifier
- high performance even with small-size vectors

* Corresponding author.
E-mail addresses: mehmet.kutlugun@std.izu.edu.tr (M.A. Kutlugün), yahya.sirin@izu.edu.tr (Y. Şirin).

https://doi.org/10.1016/j.dsp.2023.103967

2. Related studies

2.1. Data augmentation

Data augmentation methods are widely used in many fields and data sets, significantly contributing to success. Şahin O. In his study, he increased the number of images he used from 1540 to 2640 using data enlargement techniques. He calculated the consistency of the ResNet50 model on the training data as 100% by applying the 90-degree angular rotation, pan, and zoom filters to different models to obtain their different positions in the banknote data set [12]. Duman and Akın increased the data set they used in their study by 1000, 3000, and 20000 rates and increased the test success from 4% to 57% according to training [13]. Krishevsky et al., in the AlexNet model they designed, used two techniques called data augmentation and dropout to eliminate the overfitting problem and stated that the depth of the neural networks and the amount of data used during the training phase are significant for the classification performance [14]. Arpacı and Varlı studied the effect of the mixup method on retinal vessel segmentation performed with the U-Net model in IOSTAR dataset images. In this direction, they tested different variations in which the mixup method was applied in addition to traditional data augmentation processes such as horizontal mirroring, clipping, and flipping. According to this study, while low performance was obtained in the tests applied only to Mixup, the test results in which traditional and mixup methods were used together had a higher success rate [15].

2.2. Data augmentation for face recognition

Many data augmentation studies are developed in facial recognition studies, especially for real-time systems. In general, these studies can be listed as recognizing pose angle changes and recognizing despite the factors that make it difficult to recognize in the face region. Jiang-Jing Lv et al. in their study, they compared the performance between data augmentation-based augmentation and manual dataset expansion on five different methods by increasing the CASIA-WebFace dataset to an almost twice as large dataset containing 25,580 subjects, 2,545,659 face images. Finally, combining all these models achieved 99.33% training and 94.08% test success [16]. Kutlugün et al., in their study, wanted to determine which different data augmentation method was more effective in the face recognition system they proposed. They increased the number of images from 150 to 3300, 6300, and 9450, respectively, by filtering 22, 42, and 63 times for each face image. They found that this data augmentation process increased the recognition performance by 26% more in the last case than in the first case [3].

Studies in this area generally aim to train models by extracting much better features from the data and thus recognizing different situations. However, causing excessive data growth with data augmentation can waste training time. In this study, data augmentation was used not to better train the model but to classify the dataset of a pre-trained model more effectively. In this respect, our study is unique from other studies we have investigated in the literature.

3. Method

3.1. Aim

It is aimed to preserve class integrity in models where high-quality 128-dimensional vector features obtained from deep metric learning-based face recognition systems, especially k-nn classifiers, are used. Therefore, excessive data growth away from the targeted class should be filtered out or excluded. GSR proposes that it can take advantage of the classifier threshold value by introducing a new way to determine this distance. This way, only truly distinctive features will remain in the class, and integrity will be created in the database. The targeted distribution is presented in Fig. 1 as a representation. The details of filtering with GSR are explained in the section titled GSR.

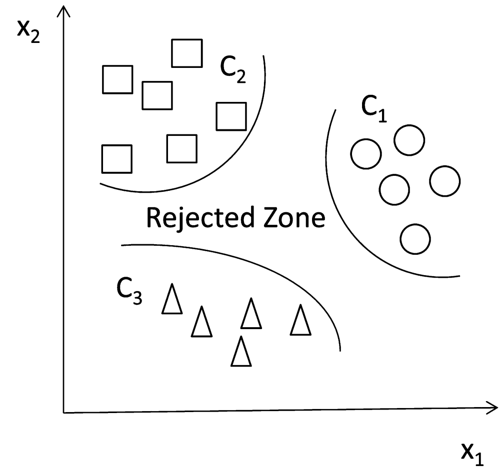

Fig. 1. Data distribution with ideal class distinction.

3.2. Model and architecture

In the study, as a model, the open-source Dlib face recognition library created by Davis King [17] and developed by Adam Geitgey [18] in 2017 was used. The k-NN algorithm was used as a classifier also. This library is built with state-of-the-art facial recognition trained with 3 million images. The default classifier tolerance value according to the dlib library is 0.6, and lower rates make face comparisons more stringent. This optimum threshold value has a 99.38% accuracy rate in the Labeled Faces in the Wild (LFW) dataset [16]. The model uses embedding, which can be represented by a 128-dimensional vector space for each face image. The Histogram of Oriented Gradient (HOG) algorithm, developed by Dalal and Triggs [19] in 2005, was preferred for face detection.

As shown in the flow diagram in Fig. 2, the original picture is augmented by applying some filters. Each augmented picture is converted into 128-dimensional vectors with the Convolutional Neural Networks (CNNs) model. In the classifier stage, only data that conforms to the accepted threshold value will remain, and other augmented artificial data will be discarded.

3.3. GSR

Let c_1 and c_2 be the elements of two different classes with a limited number of data. Let the first instance derived from class c_1 by data augmentation methods be named $c'_{1,1}$, the second $c'_{1,2}$, and the first instance derived from class c_2 as $c'_{2,1}$. Let's assume that as a result of data augmentation processes, data distribution is formed as in Fig. 3.

In this case, if each derived data were stored in the database, the second data derived from c_1 would be incorrectly classified as false positive because it was closer to the c_2 class. For $c'_{1,2}$, it can be said that it is subject to excessive distortion by excessive data augmentation. Thus, it caused an uneven distribution by moving away from its own class and getting closer to another class. With the proposed GSR method to prevent such problems, the classifier tries to keep the data augmentation under control by considering

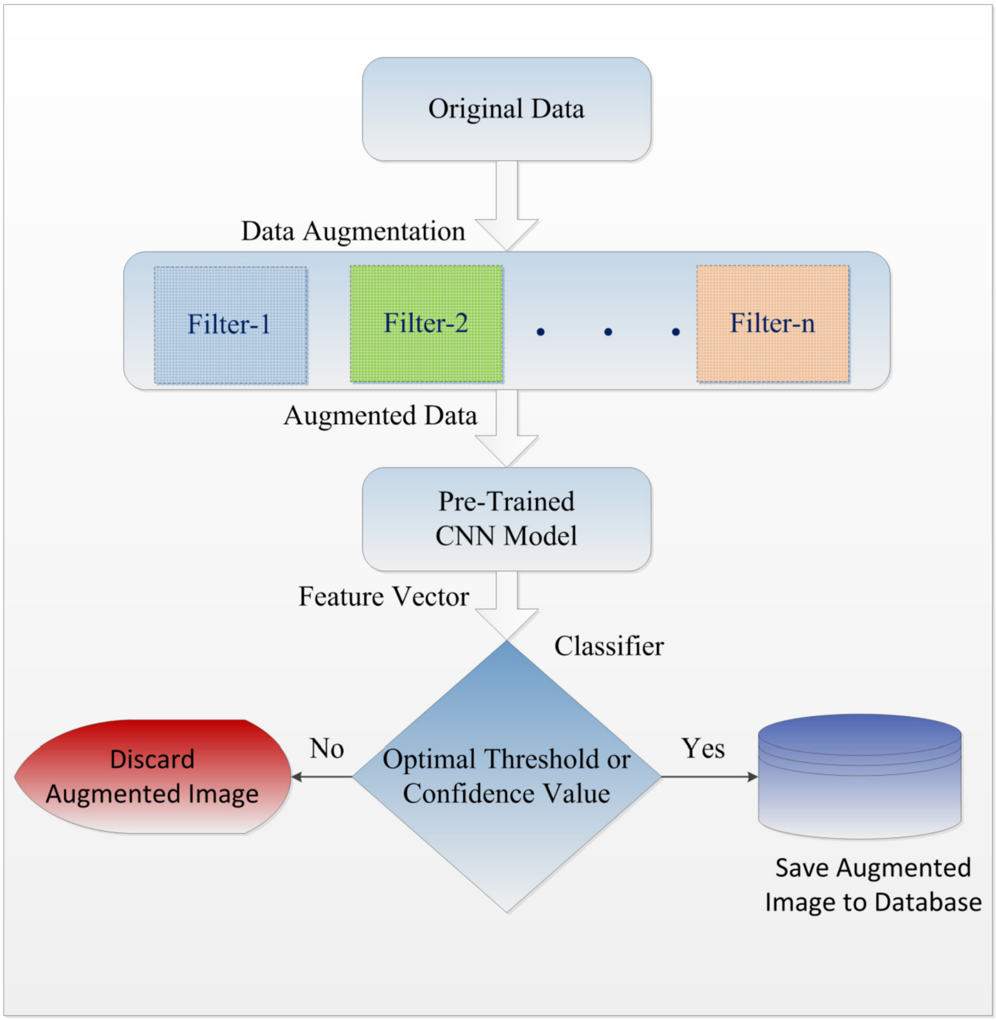

Fig. 2. Architecture.

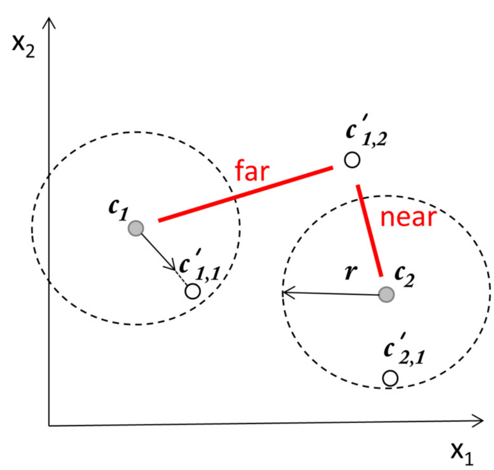

Fig. 3. Threshold relationship with GSR action.

the threshold value. Accordingly, when the two vectors are compared with the Euclidean distance formula [20,21] in equation (1), samples that exceed the classifier threshold value will be discarded according to the equation (2) rule. Instead, $c'_{1,1}$ and $c'_{2,1}$ samples that do not exceed the classifier threshold value will be located in the appropriate areas of the class.

$$d\left(x,x'\right)=\sqrt{\sum_{i-1}^{N}\left(x_i-x'_i\right)^2} \quad (1)$$

$$\text{GSR Action}=\begin{cases}0,\text{discard} & r<d\left(x,x'\right)\\ 1,\text{accept} & r\geq d\left(x,x'\right)\end{cases} \quad (2)$$

GSR will filter the fields corresponding to each different percentile separately for the data within the boundary. Suppose we want to get six synthetic data from an original image with a data augmentation filter. In that case, we can do it with GSR as in Fig. 4. However, it should be noted that dummy data corresponding to the 50% and 40% slices cannot be used as they are outside the classifier threshold. In other percentiles, the remaining four artificial data are limited to one data per percentile. In this way, a more homogeneous distribution of data is achieved.

Otherwise, positioning four artificial data in the closest position to the original data is not suitable in terms of quality, although it is suitable in number. Because the main purpose of data augmentation is to learn different situations and classify them correctly by providing data diversity.

4. Experimental study

Many different data augmentation methods have been applied for experimental studies, especially with different data sets containing a limited number of samples. As a result of these applications, the effect of the proposed method was measured by examin-

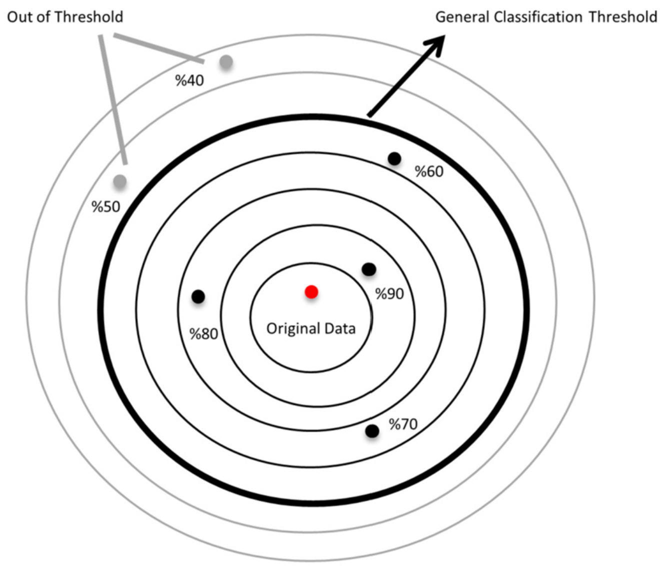

Fig. 4. Data distributed by gradual similarity ratios.

ing the GSR action according to the similarity rates obtained from the model.

The LFW dataset, frequently used in face recognition studies, contains more than 13000 face images. These contain two or more pictures of 1680 people. In the application, the facial images of 4069 people, excluded from these and only one sample are available, are used. Data enlargement filters were applied to the original data with this limited sample in groups of 150 people.

In the first stage, for each face image using data augmentation techniques, traditional methods such as angular rotation, panning, cropping, zooming, and lighting filters were applied. Augmented images in Fig. 5 are representatively shown with different similarity ratios and GSR action from the original image.

In the second phase, apart from traditional methods, the "state-of-the-art" MixUp [22] method (with alpha: 0.2) was applied to images between different classes. The new images created from these processes are shown in Fig. 6 with the GSR effect.

In the third stage, different data augmentation was applied to the original image belonging to the same class, and then new images were obtained with the MixUp method (with alpha: 0.2 value). The new images obtained from these processes are shown in Fig. 7 with the GSR effect.

In the fourth stage, the MixUp method was applied to datasets other than face databases. As a result of this application, the images obtained from the FashionMNIST data set [23] are as in Fig. 8 with the effect of GSR.

The CutMix [24] method was applied to two classes in the last stage. The new image created by the CutMix process and the GSR effect is shown in Fig. 9.

4.1. Experimental results

In this section, the results obtained from the applications performed with different methods are presented in Table 1. The Precision [25] (P) metric in equation (3) was used in this evaluation process.

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

In addition, a comparison with the test video detailed in Table 2 was made to measure the effect of GSR on the real-time face recognition system. First, random data augmentation was made by adding pictures of Aleksandra Wozniak and Maria Sharapova to the existing data set. The obtained 516 artificial images were compared without any filtering (without applying GSR) as in Fig. 10(a). Then, it was filtered with GSR up to 25 pieces each, and the results in Fig. 10(b) were obtained.

Table 1
3-Level performance chart of different data augmentation methods.

Method	Number of Data		
	Level-1	Level-2	Level-3
Traditional without GSR	3300	6300	9450
Performance Rate (%)	62.38%	66.56%	76.84%
Traditional + GSR	2400	5250	6150
Performance Rate (%)	93.74%	96.36%	**98.82**%
Traditional + Class based MixUP + GSR	2250	4650	5700
Performance Rate (%)	94.9%	97.86%	**99.38**%

Table 2
Features of the test video.

Label of the video clip	"Wozniak and Sharapova"
Duration (s)	324
Number of Frames	8103
FPS	25
Dimension	854x480
Category	Sports-Tennis
Characteristics of the video clip	It includes summary match images of different players taken with a motion camera in the quarter-finals. Original title in Youtube-8M dataset: "2015 Aegon International Eastbourne Quarterfinal WTA Highlights"

All 128 dimensional vectors stored in the database were reduced to two dimensions with the t-SNE [26] method and compared visually. Accordingly, a randomly augmented database distribution with traditional data augmentation methods will be as in Fig. 11. It is seen in the figure that the data are distributed with an accuracy of about 77%.

The data distribution obtained from discarding the data filtered by the GSR method is as in Fig. 12. It is seen that the data are distributed with approximately 99% accuracy.

In Fig. 10(a), the artificial data obtained from the random data augmentation methods show that the distance between classes is quite close. Therefore, elements of the two classes can be mixed, and losses may occur at different face exposure angles. In this test video, approximately 80% accuracy value was obtained with the traditional method.

In Fig. 10(b), more stable results can be obtained since artificial data is filtered with GSR. In this test video, approximately 88% accuracy has been reached with the GSR method.

The performance distribution graph obtained as a result of the comparison of the different methods obtained from our proposed method with the studies (Method-1 [3], Method-2 [15]) conducted in other fields and different data sets similar to our study is as in Fig. 13.

5. Conclusion and evaluation

Deep learning methods are proven methods with effective feature extraction, especially in structures where the number of data is relatively high. Many different models are generally accepted in this field today. However, there may be difficulties that each model may face. Especially for systems with limited sample data, additional methods such as data augmentation are needed. Indiscriminate data augmentation will result in unnecessary data storage and misclassification. A much more complex structure will emerge in structures where the data set grows.

The method proposed in this study is provided to shape the database in a systematic order. Data augmentation steps are ob-

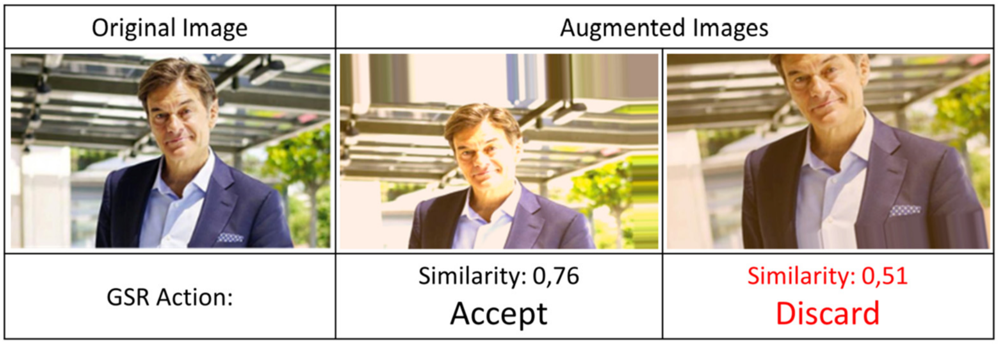

Fig. 5. Augmented images by applying traditional methods.

Fig. 6. Images obtained with the MixUp method applied between different classes.

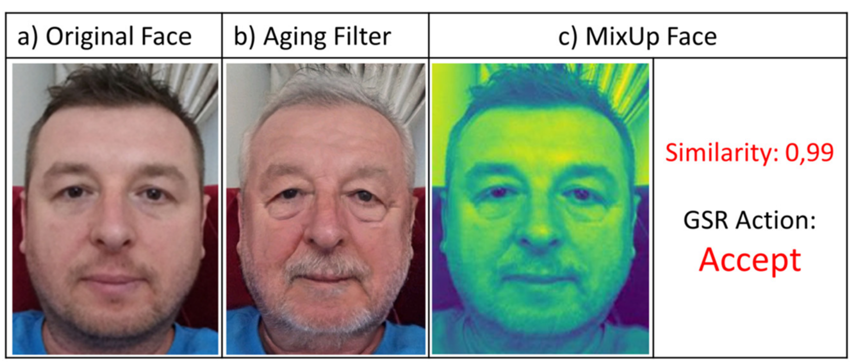

Fig. 7. MixUp images applied to images belonging to the same class.

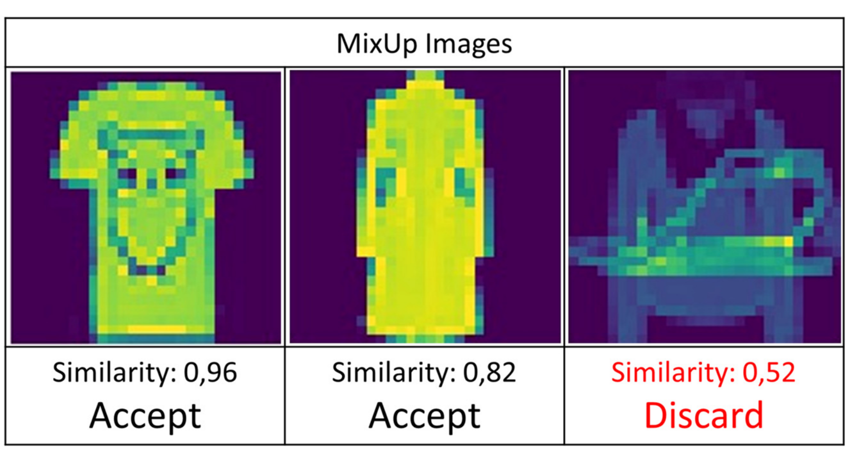

Fig. 8. MixUp example for the FashionMNIST dataset.

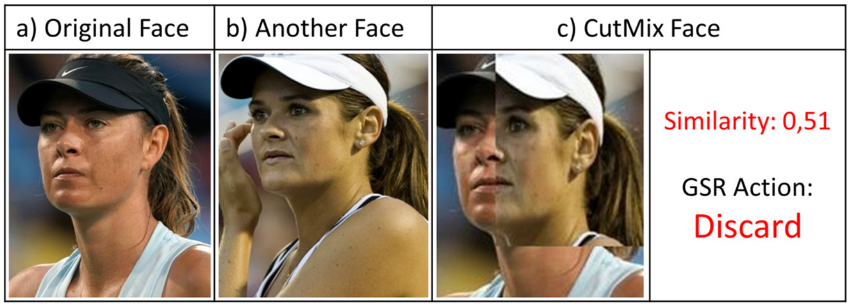

Fig. 9. Images obtained with the CutMix method applied between different classes.

Fig. 10. Comparison of face recognition performance in test video of data augmentation with GSR.

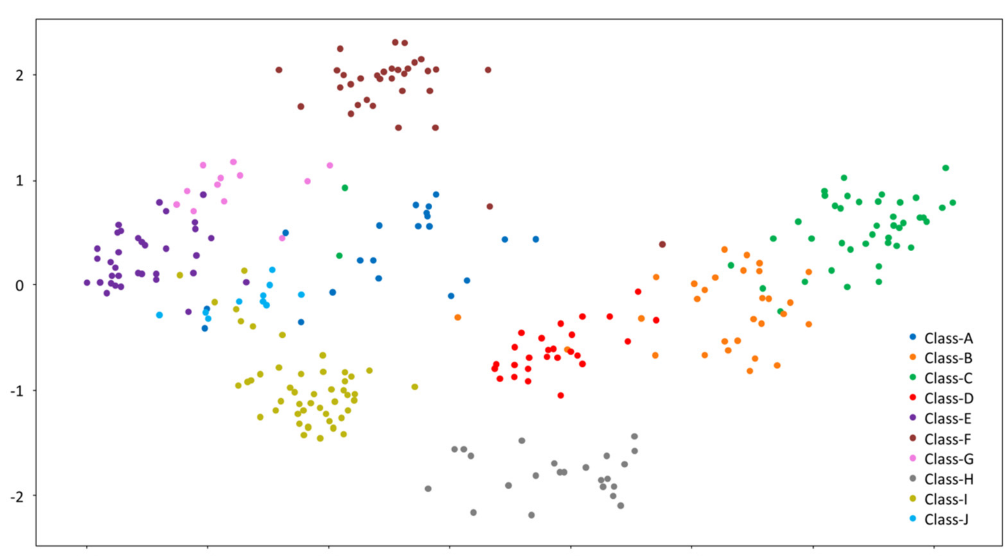

Fig. 11. Random augmented training dataset distribution.

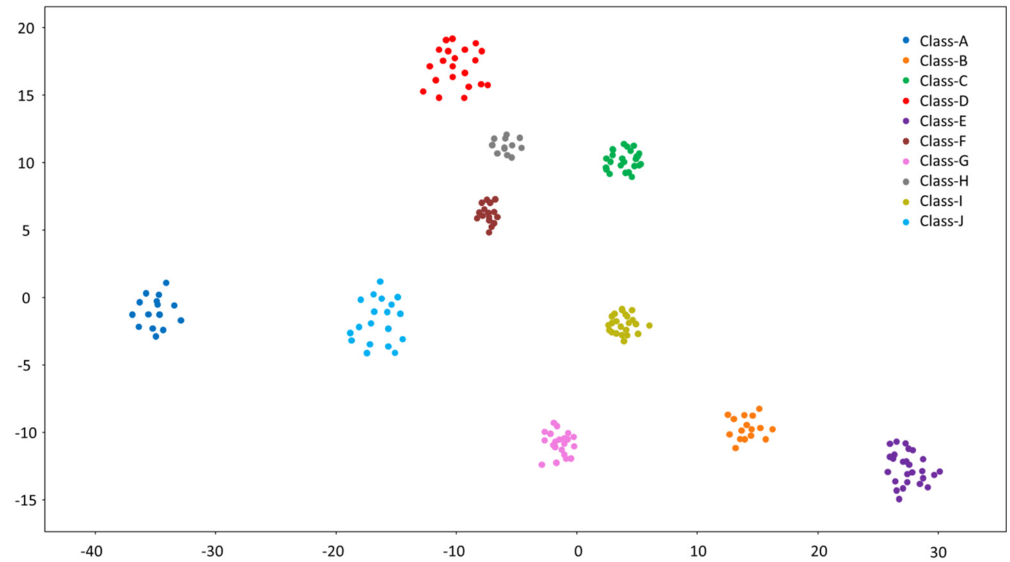

Fig. 12. Dataset distribution after GSR.

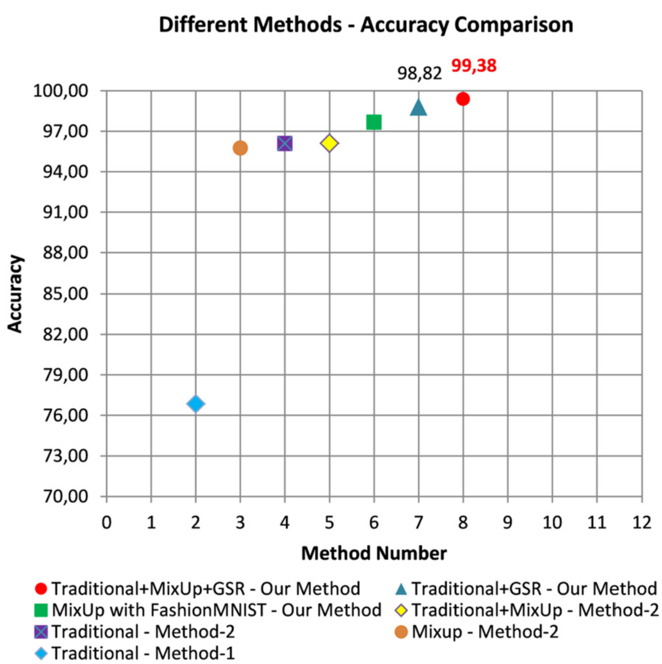

Fig. 13. Dataset distribution after gradual augmentation.

tained by using the threshold value to determine the target image's class. This stratification will ensure effective data samples with appropriate criteria instead of having many irregular data in the dataset. For example, in the case of augmentation with many pictures consisting of precisely the same images belonging to the same class, data distribution that almost overlaps or is very close to each other in that class is obtained. This situation will cause the system to become more rigid in the classifier stage. However, the fact that it contains different examples from each level within acceptable limits will lead the system to behave much more consistently. This way, classes with limited examples are kept in integrity, and classifier behavior is used more effectively. The model proposed in the experimental study reached 99.38% accuracy values compared to traditional data augmentation models. Experimental results have shown that the database has an acceptable level of success even at smaller vector sizes and is more organized.

Discussion and limitations

According to this study, it has been suggested that the classification reference threshold should be considered in the data augmentation method, which is often preferred in applications with a limited number of datasets, as it has the effect of deciding whether a person is real or fake.

However, the proposed method determines the boundaries of the classes more sharply and can be preferred to create a stable face recognition system. Because this method, which aims to reduce the FP and FN ratios, will increase the "Unknown" person identification in cases where the threshold value is not determined appropriately. In addition to the distribution of the dataset, the data augmentation capacity (i.e., the number of data in each class to be found in the database) is also essential, especially in models using k-nn classifiers.

The proposed model can be preferred in the applications given below, where it is necessary to monitor up-to-date data constantly and require more stringent classification.

- Employee tracking systems,
- Airport passenger control systems,
- Passport, visa, and driver's license control systems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The following information was supplied regarding data availability: The images are available at LFW [16], FashionMNIST [23] and YouTube Databases: http://vis-www.cs.umass.edu/lfw, https://youtube.com.

References

[1] J. Brownlee, Deep Learning for Natural Language Processing: Develop Deep Learning Models for Your Natural Language Problems, Machine Learning Mastery, 2017.
[2] Ian Goodfellow, et al., Deep Learning, MIT Press, 2016.

[3] M.A. Kutlugün, Y. Sirin, M.A. Karakaya, The effects of augmented training dataset on performance of convolutional neural networks in face recognition system, in: Federated Conference on Computer Science and Information Systems, FedCSIS, 2019, pp. 929–932.
[4] R.G. Dabhade, L.M. Waghmare, Optimal neural network based face recognition system for various pose and occluded images, Int. J. Appl. Eng. Res. 12 (22) (2017) 12625–12636.
[5] M. Koestinger, P. Wohlhart, P.M. Roth, H. Bischof, Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization, in: 2011 IEEE International Conference on Computer Vision Workshops, ICCV Workshops, IEEE, 2011, pp. 2144–2151.
[6] X. Qi, C. Liu, S. Schuckers, CNN based key frame extraction for face in video recognition, in: 2018 IEEE 4th International Conference on Identity Security and Behavior Analysis, ISBA, IEEE, 2018, pp. 1–8.
[7] E. Sütçüler, Real-time face localization and recognition system by using video sequences, Master Thesis, Yıldız Technical University Institute of Science, 2006.
[8] Y. Gao, H.J. Lee, Pose-invariant features and personalized correspondence learning for face recognition, Neural Comput. Appl. 31 (1) (2019) 607–616.
[9] M. Atik, Z. Duran, 3D face recognition and accuracy analysis using local feature-based methods, J. Fac. Eng. Archit. Gazi Univ. 36 (1) (2020) 359–372.
[10] N. Özkaya, Ş Sağıroğlu, Face recognition from fingerprints, J. Fac. Eng. Archit. Gazi Univ. 23 (4) (2008) 785–793.
[11] X. Zheng, T. Chalasani, K. Ghosal, S. Lutz, A. Smolic, STaDa: style transfer as data augmentation, arXiv preprint, arXiv:1909.01056, 2019.
[12] Ö. Şahin, TL recognition for visually impaired people on iOs platform, Master Thesis, Selçuk University Institute of Science and Technology, 2017.
[13] E. Duman, F. Akın, Evaluation of effects of Vgg16-Cnn parameters on model success, in: International Mardin Artuklu Multidisciplinary Studies Congress, 2019, pp. 26–34.
[14] A. Krizhevsky, et al., Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, vol. 25, Curran Associates, Inc., 2012, pp. 1097–1105.
[15] S.A. Arpaci, S. Varli, Retinal vessel segmentation with the mixup data augmentation method, J. Presidency Türk. Health Inst. 5 (1) (2022) 41–50.
[16] J.J. Lv, X.H. Shao, J.S. Huang, X.D. Zhou, X. Zhou, Data augmentation for face recognition, Neurocomputing 230 (2017) 184–196.
[17] E.K. Davis, High quality face recognition with deep metric learning [Online], Blog.dlib.net, Available at http://blog.dlib.net/2017/02/high-qualityface-recognition-with-deep.html, 2017. (Accessed 2 February 2020).
[18] G. Adam, Face recognition [Online], Available at https://github.com/ageitgey/face_recognition#face-recognition, 2017. (Accessed 2 February 2020).
[19] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proc. IEEE Int. Conference on Computer Vision Pattern Recognition, 2005, pp. 886–893.
[20] P.E. Danielsson, Euclidean distance mapping, Comput. Graph. Image Process. 14 (3) (1980) 227–248.
[21] M.A. Kutlugün, Text to speech synthesis using supervised machine learning according to types, Master Thesis, Istanbul Sabahattin Zaim University, 2017, https://hdl.handle.net/20.500.12436/569.
[22] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, mixup: beyond empirical risk minimization, arXiv preprint, arXiv:1710.09412, 2017.
[23] H. Xiao, K. Rasul, R. Vollgraf, Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms, arXiv preprint, arXiv:1708.07747, 2017.
[24] Y. Sangdoo, H. Dongyoon, J.O. Seong, C. Sanghyuk, C. Junsuk, Y. Youngjoon, CutMix: regularization strategy to train strong classifiers with localizable features, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV, October 2019.
[25] D.L. Olson, D. Delen, Advanced Data Mining Techniques, Springer Science & Business Media, 2008.
[26] Laurens Van der Maaten, Geoffrey Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).

Mehmet Ali Kutlugün was born in Istanbul in 1980. He graduated from Istanbul Sabahattin Zaim University with B.S. and M.S. degrees in Computer Engineering. He is continuing his Ph.D. at Istanbul Sabahattin Zaim University, Department of Computer Science. He has been working in the IT department of Turkish Airlines Inc. company for 12 years. Mehmet Ali Kutlugün is married and has two children.

Dr. Yahya Şirin was born in 1977 in Ankara. He graduated from Ankara University with B.S. and M.S. degrees in Electronics Engineering. He completed his Ph.D. at TOBB University of Economics and Technology, Department of Computer Engineering. He worked at institutions such as TOBB ETU, TUBITAK, Access Providers Association and Istanbul Sabahattin Zaim University. Yahya Şirin is married and has three children.