Leveraging Active and Continual Learning for Improving Deep Face Recognition in-the-Wild

Pavlos Tosidis, Nikolaos Passalis and Anastasios Tefas
Computational Intelligence and Deep Learning (CIDL) group, AIIA lab.
School of Informatics, Aristotle University of Thessaloniki
Thessaloniki, Greece
{ptosidis, passalis, tefas}@csd.auth.gr

***Abstract*—Face recognition systems play a vital role in various applications by providing identification and verification based on facial features. However, these systems face challenges in large-scale in-the-wild applications, where the current static pipelines are usually unable to cope with the high velocity, variety, and volume of the data, negatively affecting their accuracy and reliability. To overcome these challenges, in this paper, we propose departing from traditional static face recognition pipelines and moving towards dynamic and adaptable approaches. To this end, we propose a combination of an active and continual learning approach that can automatically augment the model with informative samples gathered during the inference, provided that the model is already confident enough, significantly improving its recognition accuracy. Furthermore, the proposed pipeline also natively incorporates active learning, allowing for using human feedback, when available, to further improve its performance. The effectiveness of the proposed method is validated on a challenging setup using two large-scale face recognition datasets.**

***Index Terms*—Face Recognition, Active Learning, Continual Learning**

I. Introduction

Face recognition systems have become increasingly important in various applications, ranging from security and surveillance to social media and online authentication. These systems enable the identification and verification of individuals based on their facial features, providing a convenient and efficient means of access control and person recognition. Since the seventies, face recognition is one of the most researched topics in computer vision and biometrics, with older methods being based on hand-crafted features and traditional machine learning techniques [1]. However, during the past decade, deep learning (DL) methods have been developed [2], that show incredible accuracy in face recognition. However, even state-of-the-art systems face challenges in large-scale in-the-wild applications, e.g., large-scale media monitoring [3], where the current static pipelines are usually unable to cope with the high velocity, variety, and volume of the data, negatively affecting their accuracy and reliability.

Indeed, recent face recognition systems rely on fixed databases of known faces, without the ability to re-train as additional samples are gathered, which can limit their accuracy in real-world scenarios. As a result, they struggle to accurately recognize new faces or faces that have undergone changes in appearance over time, such as aging or cosmetic alterations. Therefore, most pipelines lack the necessary adaptability and fail to exploit new information that could be readily available without undergoing a tedious annotation and re-training process. Additionally, despite the enormous representation capabilities of the recently developed facial feature extractors [4], most systems ignore faces that are not currently matched towards their reference database, losing a potentially useful source of information and limiting their active learning and human-in-the-loop capabilities [5].

To address the limitations mentioned above, this paper aims to investigate whether it is possible to incorporate two well-known learning paradigms that can address these challenges, i.e., continual learning [6] and active learning [7], into the existing face recognition pipelines. Continual learning enables machine learning models to incrementally update their knowledge over time, allowing them to adapt to new faces and changes in appearance. However, continual learning faces challenges in preserving previously learned knowledge while incorporating new information. Active learning, on the other hand, can enable human-in-the-loop approaches, increasing the effectiveness of annotations provided by users and minimizing their involvement. However, selecting the most informative samples for active learning can be challenging due to the vast and dynamic nature of face data.

In this paper, we propose a novel approach that combines the strengths of continual learning and active learning to enhance the performance of DL-based face recognition systems in large-scale in-the-wild applications. To avoid challenges arising from updating the actual face recognition model, we proposed manipulating the reference representations stored for the identities to be recognized. In this way, the proposed method can enable efficient self-supervised continual learning, allowing the system to be more robust to changes, as well as to easily incorporate human feedback in the most efficient way. The proposed method is evaluated on a challenging setup, simulating real in-the-wild applications, demonstrating significant improvements in face recognition accuracy.

The rest of the paper is structured as followed. First, the related work is briefly introduced in Section II. Then, the proposed method is introduced in Section III. The experimental evaluation is provided in Section IV, while conclusions are drawn in Section V.

II. Related Work

Face recognition has been extensively studied over the years, and a large number of approaches have been proposed to address various challenges in this field. In recent years, deep learning-based methods have shown remarkable results in face recognition tasks [8], [9]. The proposed method can be incorporated in any recent face recognition system with ease, to further improve their efficiency.

Furthermore, continual learning has been proposed as a method to address the issue of catastrophic forgetting in deep neural networks [10], [11]. Continual learning approaches enable models to learn continuously from new data, without forgetting previously learned information. The proposed method incorporates continual learning and continuously updates the existing feature representations in the reference database faces, in order to be more consistent with potentially facial feature changes. In this way, it can readily overcome issues associated with catastrophic forgetting, since the model used for feature extraction, which is the most vulnerable to such phenomena, is kept fixed.

Several approaches have also been proposed for incremental face recognition, where the goal is to recognize new individuals while retaining the ability to recognize previously learned individuals [12]. These approaches typically rely on training a model on a fixed set of identities, and then updating the model with new identities over time. Our approach, omits the need of a new trained model, since the continual learning method used, can also add new identities to the reference database, allowing the face recognition system to easily incorporate additional knowledge and accurately recognize new identities in the future.

Active learning has also been extensively studied in the field of face recognition, where it has been used to reduce the amount of labeled data required for training [13]. In the context of face recognition, active learning has been used to identify the most informative samples for labeling, thereby reducing the need for large amounts of labeled data. In this work, an active learning approach is used to provide a means for a human, to enhance, or even locate problematic entries in the reference database.

Therefore, the proposed method combines the strengths of continual and active learning to enable face recognition systems to both continuously adapt to new individuals, as well as, improve their accuracy over time. The proposed method is designed to identify the most informative samples for updating the reference database, while also retaining the ability to recognize individuals from preciously acquired images.

III. Proposed Method

Face recognition systems employ deep learning models to extract discriminative features from face images. The process involves detecting and aligning the face in the image, followed by encoding it into a feature vector that represents the unique characteristics of the face. Let's denote an input image containing a face as $\mathbf{x} \in \mathbb{R}^{W \times H \times C}$, where W, H, and C represent the width, height, and number of channels of the image, respectively.

To extract facial features, a preprocessing pipeline is typically applied to the image. This pipeline includes face detection and alignment, ensuring that only the face region is considered for further processing. Let $f_c(\cdot)$ denote this preprocessing pipeline, and $\mathbf{x}_c = f_c(\mathbf{x}) \in \mathbb{R}^{W_c \times H_c \times C_c}$ represent the cropped and aligned face image. Here, W_c, H_c, and C_c denote the dimensions of the cropped image.

Deep face recognition models, such as convolutional neural networks (CNNs), are then utilized to extract feature embeddings from the preprocessed face image. These models aim to learn a mapping function $f_r(\cdot)$ that transforms the input face image into a discriminative feature vector $\mathbf{y} = f_r(\mathbf{x}_c) \in \mathbb{R}^D$, where D represents the dimensionality of the embedding space. The learned features are expected to capture the essential facial characteristics necessary for accurate identification.

During the recognition process, a probe image $\mathbf{x}_p$ is compared with the feature vectors of known identities stored in a reference database $\mathcal{X}_d$. The similarity between the probe image's feature vector and the reference database entries is evaluated using the Euclidean distance metric d. Specifically, the identity l of a person in the probe image is determined by finding the feature vector in the reference database that is closest to the feature vector of the probe image:

$$k = \arg\min_i ||f_r(\mathbf{x}_i) - f_r(\mathbf{x}_p)||_2 \quad \forall (\mathbf{x}_i, l_i) \in \mathcal{X}_d, \quad (1)$$

where k represents the index of the closest feature vector in the reference database, and l_k denotes the corresponding identity label. The Euclidean distance $||f_r(\mathbf{x}_i) - f_r(\mathbf{x}_p)||_2$ measures the similarity between the feature vectors, and the minimum distance indicates the closest match.

To determine the confidence of the face recognition system in identifying a person, a confidence score c is calculated based on the Euclidean distance and a predefined threshold a. The confidence score is computed as:

$$c = 1 - \frac{||f_r(\mathbf{x}_i) - f_r(\mathbf{x}_p)||_2}{a}, \quad (2)$$

where c is bounded below 1, reflecting the confidence level of the identification. A higher confidence score indicates a stronger match between the probe image and the reference database. In summary, deep face recognition models extract discriminative features from face images using deep learning techniques. The extracted features are then compared to the feature vectors of known identities in a reference database using the Euclidean distance. The similarity measure and confidence score facilitate the identification of individuals based on their facial characteristics.

One of the challenges of face recognition systems, is the introduction of new identities that the system has not encountered before. The proposed method uses a continual learning approach to address this challenge. When a probe image is compared to the reference database, if the Euclidean distance

between the feature vectors exceeds a threshold a, indicating a significant dissimilarity, the proposed continual learning module stores the new identity in the reference database. This is similar to the way humans handle person recognition, i.e., a new identity can be formed as a concrete entity, even when part of information (e.g., exact name) is missing. By doing so, the system becomes capable of recognizing these new identities accurately in future appearances, even though the exact identity should be provided at a later stage by a human annotator, thereby expanding its capability to handle a broader range of individuals.

Another challenge of recent face recognition systems, lies in cases where the system lacks sufficient confidence in identifying a known person. This often occurs when the confidence score falls between a predefined lower bound c_l and a higher bound c_h. In such instances, the system is not confident enough in its identification and requires adaptation to improve accuracy. The proposed continual learning module addresses this challenge by updating the feature vector of the known identity in the reference database with the feature vector extracted from the new image. This adaptation process enables the system to refine its representation of the known identity, potentially capturing changes in appearance, pose, or other factors that affect recognition performance.

The adaptation of feature vectors is achieved by averaging the existing feature vector $\mathbf{y}_l = f(\mathbf{x}_l)$ in the reference database with the feature vector $f(\mathbf{x}_p)$ extracted from the new image as:

$$\mathbf{y'}_l = \beta \mathbf{y}_l + (1 - \beta) f(\mathbf{x}_p), \tag{3}$$

where $\mathbf{y'}_l$ denotes the updated feature vector for the l-th identity and β is a parameter that controls the updating process (typically set to $\beta = 0$ for performing averaging). The averaging operation aims to incorporate the updated facial information while preserving the identity characteristics learned from previous encounters. The resulting feature vector becomes an enhanced representation of the known identity, which can contribute to improved recognition accuracy in subsequent appearances.

In addition to the continual learning module, the proposed method incorporates an active learning component to further enhance the performance of the face recognition system. One of the challenges faced by face recognition systems is the presence of incorrect or problematic face images in the reference database, which can negatively impact recognition accuracy. The active learning module addresses this challenge by storing face images of identities that are recognized with a confidence score lower than the predefined lower bound c_l.

These stored face images are then presented to a human annotator for further examination. The proposed approach suggests these face images to the human annotator, who can determine whether they should be used to enhance the existing feature vector in the reference database or be discarded. This approach helps prevent the inclusion of incorrect or problematic face images in the database, such as images with low resolution, motion blur, or occlusions, which could lead to incorrect identifications.

The involvement of a human annotator in the active learning process allows for subjective judgment and domain expertise to be applied. The human annotator can carefully evaluate the stored face images and make informed decisions regarding their inclusion in the reference database. This human oversight ensures the overall quality and integrity of the database, helping to maintain high recognition accuracy.

The enhanced reference database resulting from the active learning process contributes to improved performance on future probe images. With a more accurate and comprehensive database, the face recognition system can achieve better identification and verification outcomes. This advancement is particularly valuable in real-world scenarios where the system encounters various environmental conditions, diverse facial appearances, and potential image quality challenges.

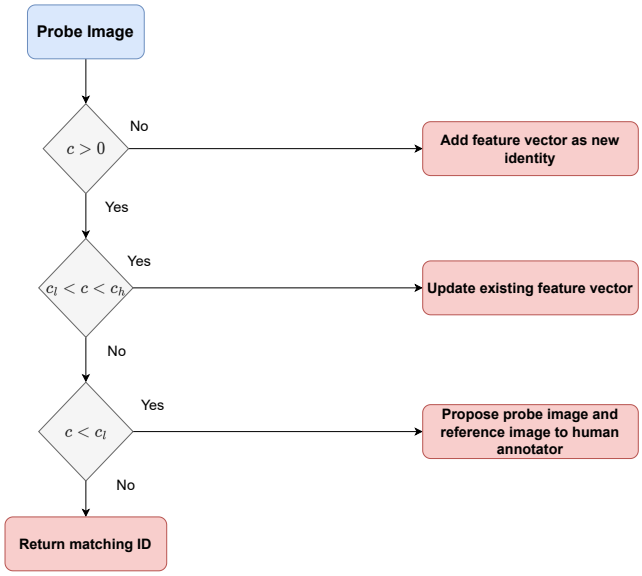

Fig. 1. Overview of the proposed method

An overview of the proposed method can be seen in Fig. 1. Given a new face image, the system measures its confidence on recognizing the person depicted in the image. If the confidence is 0, a new entry is created in the reference database. If the confidence lies between two thresholds, c_l and c_h, the existing matching entry is updated, based on the facial features extracted from the new face image. If the confidence is lower than c_l, but greater than 0, the image, coupled with the corresponding image of the matching entry is proposed to a human annotator for further action. Finally, if the confidence is higher than c_h, the system returns the matching identity.

IV. Experiments

We conducted an evaluation of our proposed method using the YouTube Faces DB dataset [14], specifically the provided aligned set, as well as a subset of MSCeleb [15]. Preprocessing of the images involved cropping and realigning them around

the face using OpenDR's [16] face recognition align method, which utilizes Multi-task Cascaded Convolutional Networks (MTCNN) [17] for both face detection and alignment. This resulted in a dataset comprising 620,852 images of 1,595 different identities from the YouTube Faces DB dataset, and 100,000 images of 1,932 different identities from the MSCeleb subset. Example images depicting each step of the preprocessing procedure can be seen in Fig. 2.

Subsequently, we constructed a reference database of known identities by randomly selecting one image from each identity in our preprocessed dataset. The feature vectors, for each of the selected face images, were produced by a deep residual network with inverted residual blocks [18], that was trained using ArcFace [4]. We evaluated the proposed method on an empty reference database (Table I), a reference database containing only half the known identities (Table II) and a reference database containing all known identities (Table III). We compare the proposed method with a modern face recognition system, first by incorporating only the continual learning approach to add the missing identities for the first two scenarios, as well as update the feature vector for each identity when needed. Then, we added the active learning approach as well, where we engaged a human annotator. The annotator had the opportunity to select which feature vectors to update in the reference database for each identity. Note that the proposed active learning pipeline relies on the ability of the system to perform continual learning. Subsequently, we repeated the evaluation on all three scenarios with the enhanced reference database.

To evaluate the performance of the methods, we measured the precision of the recognition results. Precision was calculated as the ratio of the number of correctly recognized faces to the total number of probe faces. For each probe image, we compared the identity of the closest embedding returned by the baseline method to the ground truth identity of the person in the probe image. A recognition was deemed correct if the identities matched. This process was repeated for all probe images in our dataset.

In Table I, we provide the results, evaluating the proposed method on an empty reference database. As shown in Table I, the face recognition system used fails to identify any faces, since the reference database is empty, resulting to a 0% accuracy. When the continual learning approach is used, we achieve an accuracy of more than 90%, up to 96%, for both datasets, which we further improve by almost 1% when the active learning approach, further incorporating human feedback, is added to the pipeline. Note that new identities are added as concrete entries, yet with an unknown id, in the database. Then these entries can be annotated by a human annotator. However, since these entries have already receive a unique id, we can still calculate the precision assuming that the correct ground truth information will be provided at a later time.

We then provide the evaluation results of the proposed method on a reference database containing only half of the known identities in Table II. As shown here, the face recognition system, fails to recognize faces that do not exist in the reference database, achieving an accuracy of only 45%. Our approach improves the face recognition system, which can reach an accuracy of 97.32% when the continual learning approach is used, and an accuracy of 97.88% when the active learning approach is also added. Finally, we provide the results of the proposed method on a reference database containing all known identities in Table III. Even with a fully built reference database, the use of both the continual and the active learning approach, can improve the face recognition system by 1% and 1.5% respectively, further highlighting the flexibility of the proposed method.

Fig. 2. Face alignment and cropping examples

TABLE I
FACE RECOGNITION PRECISION WHEN AN EMPTY REFERENCE DATABASE IS USED

Dataset	Baseline	Proposed (Cont. Learning)	Proposed (Cont.+Active Learning)
YouTubeFaces	0%	90.56%	**90.97%**
MSCeleb	0%	96.26%	**97.02%**

TABLE II
FACE RECOGNITION PRECISION WHEN A REFERENCE DATABASE CONTAINING ONLY HALF THE KNOWN IDENTITIES IS USED

Dataset	Baseline	Proposed (Cont. Learning)	Proposed (Cont.+Active Learning)
YouTubeFaces	40.14%	90.78%	**91.52%**
MSCeleb	45.78%	97.32%	**97.88%**

V. CONCLUSION

In this paper, we introduced a novel approach that can enhance face recognition systems through the integration of

TABLE III
FACE RECOGNITION PRECISION WHEN A REFERENCE DATABASE CONTAINING ALL KNOWN IDENTITIES IS USED

Dataset	**Baseline**	**Proposed (Cont. Learning)**	**Proposed (Cont.+Active Learning)**
YouTubeFaces	90.73%	91.56%	**92.56%**
MSCeleb	96.92%	98.14%	**98.84%**

continual learning and active learning methodologies. By addressing the challenges associated with adapting to new faces, updating reference databases, and ensuring data quality, the proposed method significantly improved the adaptability, accuracy, and reliability of the system. Experimental results on benchmark datasets demonstrated the superior performance of our approach compared to a baseline method in recognizing individuals across diverse scenarios.

Looking ahead, there are promising directions for future research in the field of face recognition. One important aspect to explore is the improvement of the continual learning module's feature vector adaptation process. While our method currently employs averaging to update the feature vectors of known identities, further investigation can be conducted to explore more sophisticated adaptation strategies. For instance, instead of averaging, each value in the feature vector could be individually adapted based on the new information, potentially leading to better preservation of important discriminative features, while multiple feature vectors could also be stored to better model the underlying feature distribution, at the expense of slightly higher retrieval complexity. Finally, spatial and temporal constraints could also be incorporated, e.g., using information and tracking faces from previous frames where a face has been correctly identified, going beyond the confidence of the model, further improving the accuracy of updating the database even in cases where all identities are already known.

ACKNOWLEDGMENT

This research was funded by the project "SEMANTIC ANNOTATION AND METADATA ENRICHMENT OF OPEN VIDEO STREAMS USING DEEP LEARNING" (Project code: KMP6-0079092) that was implemented under the framework of the Action "Investment Plans of Innovation" of the Operational Program "Central Macedonia 2014 2020", that is co-funded by the European Regional Development Fund and Greece.

REFERENCES

[1] Wenyi Zhao, Rama Chellappa, P Jonathon Phillips, and Azriel Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[2] Muhtahir O Oloyede, Gerhard P Hancke, and Hermanus C Myburgh, "A review on face recognition systems: recent approaches and challenges," *Multimedia Tools and Applications*, vol. 79, pp. 27891–27922, 2020.

[3] N Passalis, M Tzelepi, P Charitidis, S Doropoulos, S Vologiannidis, and A Tefas, "Deep video stream information analysis and retrieval: Challenges and opportunities," in *Proceedings of the IEEE International Conference on Multimedia Information Processing and Retrieval*, 2022, pp. 336–341.

[4] Jiankang Deng, Jia Guo, Jing Yang, Niannan Xue, Irene Kotsia, and Stefanos Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 5962–5979, 2022.

[5] Xingjiao Wu, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, and Liang He, "A survey of human-in-the-loop for machine learning," *Future Generation Computer Systems*, 2022.

[6] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars, "A continual learning survey: Defying forgetting in classification tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3366–3385, 2021.

[7] Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B Gupta, Xiaojiang Chen, and Xin Wang, "A survey of deep active learning," *ACM Computing Surveys*, vol. 54, no. 9, pp. 1–40, 2021.

[8] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman, "Deep face recognition," in *Proceedings of the British Machine Vision Conference*, September 2015, pp. 41.1–41.12.

[9] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proceedings oft the IEEE Conference on Computer Vision and Pattern Recognition*, jun 2015.

[10] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, mar 2017.

[11] Zhizhong Li and Derek Hoiem, "Learning without forgetting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 2935–2947, 2017.

[12] Suresh Madhavan and Nitin Kumar, "Incremental methods in face recognition: A survey," *Artif. Intell. Rev.*, vol. 54, no. 1, pp. 253–303, 2021.

[13] Robin Hewitt and Serge Belongie, "Active learning in face recognition: Using tracking to build a face model," in *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop*, 2006, pp. 157–157.

[14] Lior Wolf, Tal Hassner, and Itay Maoz, "Face recognition in unconstrained videos with matched background similarity," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2011, pp. 529–534.

[15] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *Proceedings of the European Conference on Computer Vision*, 2016, pp. 87–102.

[16] Nikolaos Passalis, Stefania Pedrazzi, Robert Babuska, Wolfram Burgard, Daniel Dias, Francesco Ferro, Moncef Gabbouj, Ole Green, Alexandros Iosifidis, Erdal Kayacan, Jens Kober, Olivier Michel, Nikos Nikolaidis, Paraskevi Nousi, Roel Pieters, Maria Tzelepi, Abhinav Valada, and Anastasios Tefas, "Opendr: An open toolkit for enabling high performance, low footprint deep learning for robotics," in *Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (to appear)*, 2022.

[17] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503.

[18] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.