Deep Face Verification Based Convolutional Neural Network

Hana Ben Fredj[†], Safa Bouguezzi[†], and Chokri Souani[††]
ben.fredj.hanaa@gmail.com safa_bouguezzi@outlook.com chokri.souani@gmail.com
[†]Université de Monastir, Faculté des Sciences de Monastir, Laboratoire de Micro-électronique et Instrumentation, Av. de l'Environnement 5000 Monastir, Tunisia
[††] Université de Sousse, Institut Supérieur des Sciences Appliquées et de Technologie de Sousse, 4003, Sousse, Tunisia

Abstract

The Convolutional Neural Network (CNN) has recently made potential improvements in face verification applications. In fact, different models based on the CNN have attained commendable progress in the classification rate using a massive amount of data in an uncontrolled environment. However, the enormous computation costs and the considerable use of storage causes a noticeable problem during training. To address these challenges, we focus on relevant data trained within the CNN model by integrating a lifting method for a better tradeoff between the data size and the computational efficiency. Our approach is characterized by the advantage that it does not need any additional space to store the features. Indeed, it makes the model much faster during the training and classification steps. The experimental results on Labeled Faces in the Wild and YouTube Faces datasets confirm that the proposed CNN framework improves performance in terms of precision. Obviously, our model deliberately designs to achieve significant speedup and reduce computational complexity in deep CNNs without any accuracy loss. Compared to the existing architectures, the proposed model achieves competitive results in face recognition tasks

Key words:
Deep learning, Face recognition, Lifting scheme, CNN

1. Introduction

For decades, face recognition has been a hot topic in pattern recognition. Face recognition has also increased in importance with the rising demand for deep learning. Recently, Convolutional Neural Networks (CNNs) have been considered as the first solution to address classification problems. Depending on the capabilities of the discriminative CNN processing, various work has succeeded in carrying out the main stages of face recognition. This effectiveness has been extensively demonstrated on the dataset of the Labeled Faces in the Wild (LFW) [1, 2]. Indeed, CNNs attain a significant improvement in facial representations within unconstrained environments such as severe occlusion, or using a large amount of data in the training step.

In order to achieve excellent accuracy, neural deep networks have become more profound with many parameters. Thus, directly training a deep network requires a massive amount of labelled face images. This generates a problem at the computational level [3, 4]. Furthermore, most work has dealt only with accuracy without finding any solutions for the requirement of calculation and storage during training. Moreover, the large size input face images has a great impact on the difficulty of computation operations in the approaches based on deep learning. Additionally, memory consumption increases when processing large amounts of images [5].

However, training with limited data will lead to overfitting [6]. Indeed, Deep convolutional networks have been shown to be very powerful in analysing classification tasks with large-scale datasets. Then, it is necessary to improve the CNN model performance to be much faster and more efficient without reducing the number of labelled face images and without losing the precision of recognition performance.

Among the most treated solutions in some algorithms, we find the decomposition of high-resolution images into patches. Thereafter, the evaluation is done independently on each patch [7, 8, 9]. However, an evident decrease in recognition rates can be recorded, since the patches may not contain all the information that we need in training. To tackle the previously mentioned challenges, we put forward an efficient approach for face recognition based on deep CNNs and the lifting method. In order to boost the performance of this proposed network, the deep CNN model is characterized by multi-scale convolution features from different layers.

Deep learning requires large computational resources because the basic CNN receives raw images as input redundant data or pixels. Hence, we utilize the lifting transform method in order to convert each input high-resolution face image to a small size. This strategy allows considerably minimizing the computational resources and the convolution time in deep CNNs without any downgrade

Manuscript received May 5, 2021
Manuscript revised May 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.5.35

in performance. We only consider the Low-Low (LL) part where the information is located, which contains sufficient feature details.

The advantage of the approach is that it does not require any additional space to store the facial features. In fact, it makes the model much faster during the training and classification phases. Moreover, this method contributes to better accuracy. In this paper, the model is deliberately designed to reduce the running time and to reach important performances. The experiments on the LFW and YouTube Faces (YTF) databases demonstrate that the suggested system improves the face verification performance compared with the state-of-the art methods. An overview of the rest of the paper is organized as follows. Section 2 analyzes some face recognition work. Section 3 presents the lifting method and section 4 shows the network architecture. In section 5, the experimental results are described. The conclusions are given in section 6.

2. Related work

Face recognition has been a dynamic topic within the field of pattern recognition, which is fundamental for biometric and security applications. Recently, it has progressed with the development of CNNs [6]. Besides, based on its great capabilities to present and learn the discriminative features, CNNs have become the ideal solutions for a variety of face recognition applications. Obviously, extensive research has made great progress in face recognition based on CNNs, and various techniques have shown surprising improvements in the LFW dataset; e.g., accuracy increased to 99.83% in [10].

Some face recognition systems, under excellently-controlled conditions, have already shown interesting results. However, face recognition remains a challenging problem in real-world scenarios due to dramatic facial variations caused by different expressions, poses, lighting conditions, occlusion and so on. In fact, the performance success in face recognition models have been largely due to the recent progress in deep CNN networks, and we find a large number of accessible models, such as, VGG [12], SphereFace [16], GoogLeNet [13], Facenet [14], ResNet [15], CosFace [17], AlexNet [11] and ArcFace [10]. These architectures vary with the presence of several ways, like the arrangement and number of layers, the size of filters and the number of filters. Moreover, flexible Graphical Processing Unit (GPU) tools and extensive amounts of data during training are two essential elements that have a greater role in the efficacy of these models.

In the real world and within uncontrolled environments, several methods have been proposed, which are capable of improving the performance of facial verification. Indeed, these methods have very important results. The authors in [18] put forward a new approach which expressed the changes in the face of each person. This method, which used the Bayesian model, was based on using intra-personal variations and identity-specific components. The results obtained on the LFW dataset demonstrate the effectiveness of this approach in face verification. In a cloud environment, the authors in [19] suggested a deeper automatic face recognition model. The results from the experiments showed that the proposed method achieved substantial precision through evaluation in various accessible databases. Mansoor Iqbal et al. [20] proposed hybrid angularly discriminatory features by using an additive cosine margin and a multiplicative angular. The authors in [21] created a deep network focusing on a generative adversarial network on real-world face datasets. The main contribution of this work was to optimize a reconstruction loss. This model gave significant performance for face recognition by high rank-1 recognition rates. The authors in [22] presented an improved framework, named Split-Net, which exploited global and local information by splitting the selective intermediate feature maps into several branches. The experimental results proved that this method effectively kept high accuracy of face recognition. In [23], the authors introduced an approach based on two self-adapting patch strategies, which were obtained by utilizing the integral projection technique. The extensive experiments on different datasets showed that the new methods surpassed some related patch-based methods. Mei Wang et al. [24] proposed an advanced clustering established domain adaptation approach created for face recognition. Comprehensive experiments on widely-used databases certainly highlighted the efficiency of the newly suggested approach. Ze Lu et al. [25] introduced a deep network model named as the Deep Coupled ResNet model, which was designed for the task of low resolution face recognition. The experiments showed that the proposed model attained significant performances.

The existing models take advantage of very deep CNN for face recognition and can attain excellent recognition rates. However, the models currently require a very massive size of data needed for training. This always makes the practical utilization of CNNs difficult. The problem here lies in the computational requirements of the CNNs. They often avoid the importance of CNNs on integrated platforms. Thereby, these methods cannot achieve a satisfactory speed on normal computing devices. Following the progression, we learn face features by using a CNN model with the lifting method to reduce the computational resources and the convolution time in deep CNNs without any performance decrease.

3. Lifting scheme

Most of the recognition frameworks based on CNNs require a huge size of training data. Therefore, the reduction in the execution time and the computational complexity is necessary, without losing any recognition performance. Consequently, to examine and improve our face representation, we integrate the lifting in our architecture.

In computer vision applications, the lifting transform is generally used as a mathematical method which is responsible for extracting information from various kinds of data. In this context, we implement the lifting wavelet scheme, to reduce the data treated within the CNN. Indeed, the lifting scheme is better than the convolution-based Discrete wavelet transform schemes in power consumption and computational complexity. The lifting scheme also allows obtaining low latency processing and a high throughput for high-resolution image signals.

The lifting scheme [26] was a very adaptable method to construct new and nonlinear wavelets from existing ones. This method made images into subframes and offered important insights into the frequency and spatial characteristics. It contained a wavelet transform, based on an update step and a prediction one.

The classical 1-D lifting scheme (lifting wavelet transform) consists of three basic steps: Firstly, the original signal S: $Z^d \rightarrow$ R is separated into an approximation signal x and a detail signal y by a precise wavelet transform.

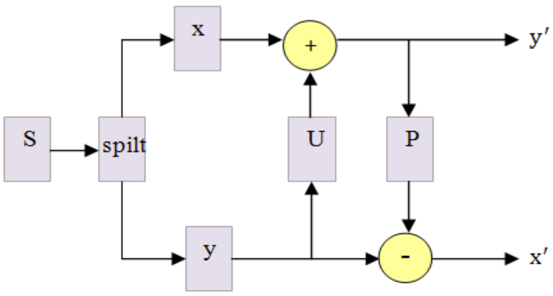
The update map U delegated on y is used to change x, resulting in a novelty approximation signal x' , i.e.

$$x' = x + U(y) \text{ (Eq.1).}$$

Afterwards, the prediction map P acting on x' is utilized to change y, resulting a new detail signal y', i.e.

$$y' = y + P\,(x') \text{ (Eq.2).}$$

The general 1-D lifting scheme is given in Fig 1.



U : Update
P : Predict

Fig.1 General 1-D lifting scheme

The lifting scheme is a separable transform; it can be computed by applying a 1-D lifting wavelet transform along the rows and columns of the input image of each level during the horizontal and vertical filtering stages. Every time the 1-D lifting is applied, it decomposes that image in two sets of coefficients: low-frequency (L) and high-frequency (H) components. After performing one level 2-D lifting wavelet decomposition, we can acquire the coefficients of sub-band LL (Low-Low), HL (High-Low), LH (Low-High) and HH (High-High).

Most energy is concentrated into an LL sub-band. However, the HL and LH subbands also contain edge and contour details of face images. Thus, the high frequency HH sub-band mainly comprises noise with negligible functionality details. Fig.2 represents the 2-D lifting scheme by applying the 1-D transform in the row direction and then in the column direction.

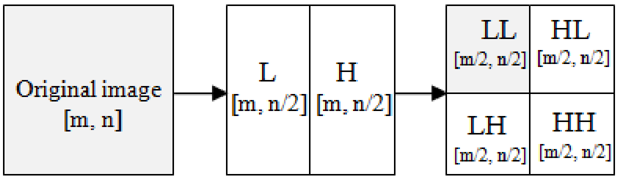


Fig. 2 Procedure of 2-D lifting Scheme. The resolution level have 4 bands: LL, HL, LH, and HH.

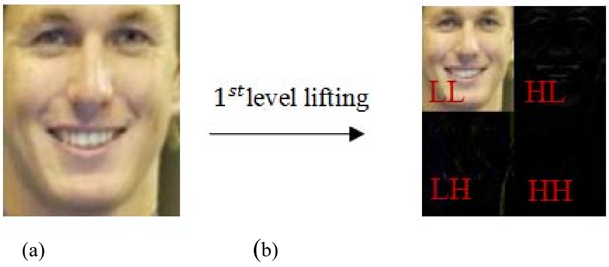


Fig. 3 (a) Original image, (b) 1-level 2-D lifting wavelet decomposition with an image from LFW dataset.

As illustrated shown in Fig 2, in the lifting decomposition of the image, the process is done line by line, and afterwards column by column. This figure describes the methodology for an n x m image size. The first decomposition according to the lines allows acquiring two n x (m/2) sizes of images. After making transformation according to the columns, four (m/2) x (n/2) sizes of images are obtained.

Fig.3 depicts a notation for a 1-level 2-D separable lifting decomposition of an image from the LFW dataset. The LL part presents an approximation of the original image.

One of the advantages of the lifting transform concerning data compression is that it tends to compact the energy of the input signal into a relatively small number of wavelet coefficients. Lower resolution band coefficients have high energy.

In our work, we concentrate on the LL part where the information is located, which contains sufficient features details. Subsequently, the LL sub-band represents an approximation of the original image. Thus, we consider the first level decomposed image for deep feature extraction. The integration of lifting analysis causes a suitable visual representation that contributes to the interpret ability of the network.

The CNN framework learns feature representations from large-scale data with massive labels. This generates large consumption in terms of memory and execution time. Therefore, to ensure that the model can extract a face representation faster than the traditional frameworks, we use the lifting method. In addition, we take advantage of this method by using the LL (m/2, n/2) part, which presents a smaller and smoothed version of the original image. Indeed, using just the compressed coefficients of the image, our model can learn enormous data.

4. Deep Convolutional Neural Network

Deep learning approaches in the area of digital imaging have made significant progress. In particular, deep CNNs have been extensively applied to image classification. This section discusses the suggested framework in our work (Fig.4).

The deep CNN arranges alternative convolutions, fully connected and max-pooling layers that can, using low level representations, determine high-level details. ResNet-50 is one of the most deep neural network models for benchmarking large-scale distributed deep learning. This Deep CNN model represents the residual network architecture [27]. Among these advantages, it is characterized by its adaptable design in such a way that it forces the convergence of networks and facilitates training by raising the depth.

The ResNet-50 architecture is represented in Fig.4 (b) with residual units, as well as sizes of filters. The fully connected layer is denoted by FC 1000 with 1000 neurons. The number on the top of the convolutional layer block describes the repetition of each unit. The NC represents the number of output classes. Different types of kernels with multiple sizes (7 × 7, 5 × 5 and 3 × 3) are used. This model has a great capability of learning features and properties of accessible training. As a consequence, we adopt ResNet-50 as a network to train the models for different data reduction methods throughout our experiments. Therefore, in order to train the network, we utilize the center loss combining with the Softmax loss to develop discriminative deep features discriminative.

For the facial recognition task, we have to separate and discriminate deeply learned features. The discriminating power characterizes the features in separable inter-class differences and in compact intra-class variations. As a result, we should use a highly efficient loss function to learn discriminant CNN features. To achieve these objectives, we consider a combination of Softmax and center losses.

Indeed, Softmax loss trained deep CNNs can achieve good results in several close-set recognition problems. The Softmax loss forces in an intuitive manner the deep characteristics of various classes to remain separate. However, such a Softmax loss does not encourage in an explicit way the intra-class feature compactness [37].

The center loss learns in a simultaneous way a center for each class. Added to that, it penalizes the distances between face images'deep features and their corresponding class centers [38]. In fact, a loss center pulls the same class'deep features towards their centers in an effective manner. In the same vein, training with the center loss makes it easier for CNNs to extract deep features with two desirable characteristics: intra-class compactness and inter-class separability. As a consequence, using joint supervision, the differences in interclass characteristics are widened and the variations in intra-class ones are reduced. Therefore, the discriminating power of deeply learned characteristics are greatly improved.

We follow a strategy of combining end-to-end training with multiresolution analysis via the lifting scheme that permits adeptly capturing the essential information from face image inputs. The network extracts features and captures information that can be trained in face classification. Also, the loss function utilized for training this network ensures that the captured information relates to the classification task.

The large amount of data which required by deep CNNs, leads to an enormous complexity in computation. Nevertheless, in order to extract and select distinctive features, face recognition tasks use input images with big number and size needed for careful processing. Our strategy is to reduce the computing burden by advancing the sparsity while using the LL sub-band, where most of the energy is concentrated. In fact, we consider the LL part for each image, where the information is located, which contains sufficient feature details.

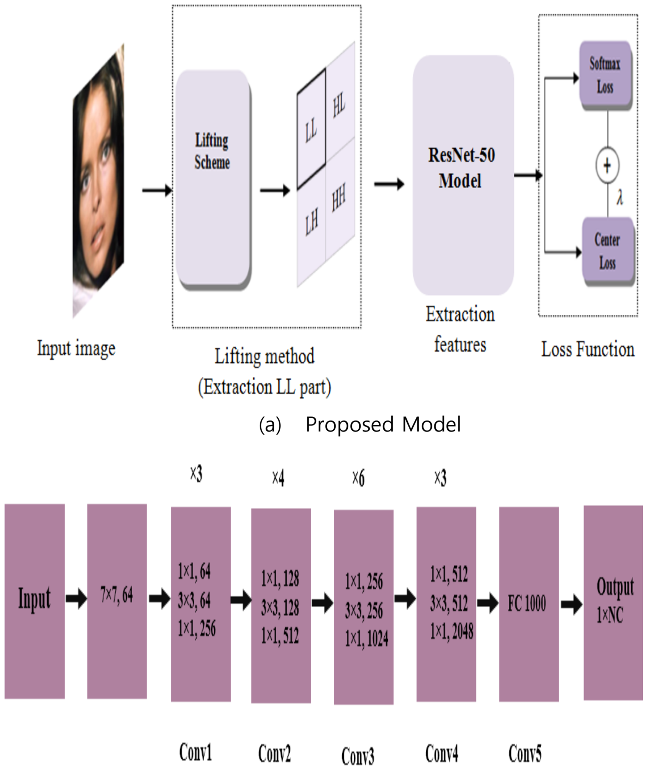
Obviously, the CNN layers are characterized by very strong learning capacities, but its time consumption is very important. While keeping the same necessary and useful information, we have to reduce the high dimensional data.

This approach helps minimize the space required to store data, and it also reduces the time needed during training. In fact, the lifting scheme is applied to the first convolutional layer output of the model to guarantee these objectives.

5. Experiments

We describe in this section the implementation details of our deep CNN in both training and testing steps. We perform our experiments using the LFW dataset to analyse the verification scenario. We also discuss the effectiveness of the lifting method. Added to that, our method is compared with the state-of-art models on the LFW dataset.

We train our model on GPU GeForce GTX 1080 Ti (with 11G memory size) using one publicly available deep learning framework, called Tensor Flow.



(a) Proposed Model

(b) CNN architecture (Resnet-50 Network); Conv is the convolution layer , NC is the Number of output classes, and FC 1000 denotes the fully connected layer

Fig. 4 Overview of our model

5.1 Implementation details

Preprocessing. For face alignement, we use the MTCNN method [29]. It is a very effective algorithm, especially for images presented in an uncontrolled environment. In fact, the reference points, two mouth corners, two eyes and nose points are occupied with the aim of making similarity transformation for face cropping.

Training. We train our model with the public deep learning framework TensorFlow. We use a large training dataset, VGGFace2, to improve the recognition face in an unconstrained environment. VGGFace2 is currently among the challenging databases, which allows using millions of images to assess face recognition algorithms. In fact, we download these images from Google Image Search. As a result, they are largely varied in illumination, age, pose, profession (e.g. politicians, actors) and ethnicity. This kind of wide benchmarks which are available makes it possible to further improve the current results for very deep CNNs. Training on VGGFace2 allows us to improve the classification performance over unconstrained environments. Fig.5 presents some samples from the VGGFace2 dataset.

As previously illustrated, we consider for each image the LL part where the information is located, which contains sufficient feature details. In this study, we use many techniques to strengthen the CNN model generalization.

The reduced data features (LL_details) are then integrated and trained to predict the face by using the CNN architecture. We train the framework after a 1-level lifting, and decomposed images for 100 epochs are applied with a root mean square propagation optimizer. Furthermore, to control the loss function regulation, we utilize the weight decay parameter, which can be fixed to $5e^{-4}$ for both fully-connected and convolutional layers. We apply also the momentum parameter to 0.9 and we adapt the model using a dropout function. Moreover, we adopt random cropping and horizontal flipping for data augmentation.

Testing. For testing, we determine the similarity score with the cosine distance. The results are performed on the LFW dataset according to the standard protocol of restricted outside-labeled data. LFW is a public benchmark for face verification. In fact, it contains 13,233 images of different persons, including various face images within an uncontrolled environment having different illuminations, poses, occlusions and expressions. To determine the efficiency of our proposed work to classify the facial images, we produce considerable experimental results. In addition, for testing, we use the YTF dataset of face videos which is produced for considering the issue of unconstrained video face recognition. It is composed of 3,425 videos of 1,595 different people.

5.2 Loss function

In order to develop and validate the evaluation parameters to be more appropriate and better correlated with intra-class compactness scoring and inter-class separability, we consider the joint between the Softmax and center losses.

For fair comparison, and to check the efficiency of the loss function used during training and composed of the center and Softmax losses after applying the lifting scheme for each image, we train our proposed model in two cases:

Indeed, after having used lifting and preserved the LL part for each image of the database, we train two kind of models using the Softmax loss (model (1)) and both Softmax and center losses (model (2)) on the augmented VGGFace2 dataset.

For detailed testing settings, the models are evaluated up on the LFW dataset. Table 1 shows the results concerning the recognition rate of each model. As shown, model (2) (supervised by both Softmax and center losses) will actually beat the baseline one (model (1) which will be only supervised by the softmax loss) through the use of a significant margin, while improving in particular the performance from (94.37% for LFW) to (99.4% for LFW).

Consequently, it is shown that the joint supervision is remarkably able to ameliorate the deeply learned features'discriminative power, hence showing the effectiveness of the center loss compared with training with Softmax only.

These results indicate that this face verification system, based on deep CNNs to reduce data using the lifting scheme, can achieve good performance with combined Softmax and enter losses. The accuracy on LFW is boosted by 5.03%, in a comparison to the model that is basically only trained with the Softmax loss.

Table 1: Classification accuracy (%) on LFW dataset: Model (1) (Softmax loss), Model (2) (Softmax+Center Losses)

Method	Accuracy (%)
Model (1)	94.37±0.013
Model (2)	99.4±0.04

5.3 Time analysis

Fig. 5 Example images for different subjects from VGGFace2 dataset.

Data training within a deep neural network presents an excessive execution time, which makes the recognition face system inefficient. Thus, to achieve an outstanding result for the face recognition framework, it is important to solve this problem. The lifting scheme is used in our work for reducing the input data of deep learning, in order to keep the maximal amount of information located in different images. We select the LL part where the information is located, which contains sufficient features to train our designed CNN architecture. Essentially, the necessary intact details are kept in approximate coefficients. This allows eventually decreasing the maximal data without losing the information required for the classification step.

We perform some experiments to examine the lifting technique effect on the performance of the designed model in terms of execution time. To better understand the optimization computational gain, we calculate the elapsed time with the lifting method and without it.

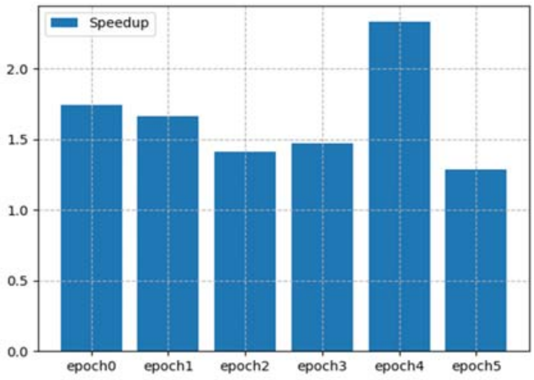


Fig. 6 Impact of lifting method for first 6 epochs during the training step

Firstly, we calculate the speedup training for the first six epochs during the training step. The speedup describes the ratio of the runtime with the CNN and the runtime with the lifting-CNN. The results are illustrated in Fig.6, and the performance is clearly improved. The obtained results justify that the approach achieves a better performance than the model without the lifting step in terms of execution time with an important speedup in each epoch. Hence, the lifting-CNN increases the processing speed and enhances the performance. Secondly, during the classification step, we study the execution time varying the number of images, in order to determine the impact of the lifting scheme. The results are described in Fig.7. As illustrated, the time increases progressively with the number of images. The execution time of the optimized method is significantly lower than the standard method, which decreases the consumption of the classification time in different cases. Indeed, the classification speed is accelerated using the lifting scheme. ResNet is a more compact model, and accelerating this network is even more difficult. However, our method can still achieve a significant speedup.

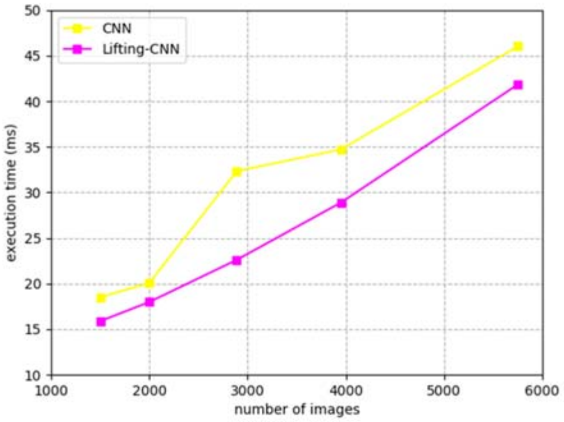


Fig. 7 Influence of lifting method on execution time during classification step

5.4 Memory analysis

During the training step of deep learning models, memory is considered one of the essential elements. In fact, Resenet-50 is a large model, which consumes lots of memory resources for training. Moreover, several critical variables are responsible for the increase in memory consumption, such as parameter (bias and weight), factors, input data and intermediate data. These latter appeared as the most consumers during training. Hence, the training time increases appreciably within the neural network, while the availability of GPU memory is low, in agreement with the requirement of training in deep learning. Usually, we want the GPU to spend most of the time on computing instead of fetching data from its memory. Therefore, it is necessary to reduce the time during training, while the GPU memory is reading or writing. One way to lower memory usage is to decrease the feature data, so data fetching becomes more efficient.

Thereafter, it is important to follow a data compression strategy in order to properly use the memory resources and to improve the baseline memory manager used by the current machine learning framework. This will reduce the time during training. Indeed, compression is much faster than transfer, and it also facilitates calculation.

As mentioned earlier, the memory requirement efficiency is also a critical aspect in the comprehensive evaluation of face recognition systems. To verify the reduction data within the CNN efficiency in terms of memory consumption, we compare the CNN model with the proposed model based on the lifting method for data reduction. Consequently, we analyze the performance of the percentage of memory usage as a function of time with and without the lifting method for a predefined period of training. Fig.8 shows the percentage of memory usage versus time for each method during training by randomly selecting a specific number of image samples.

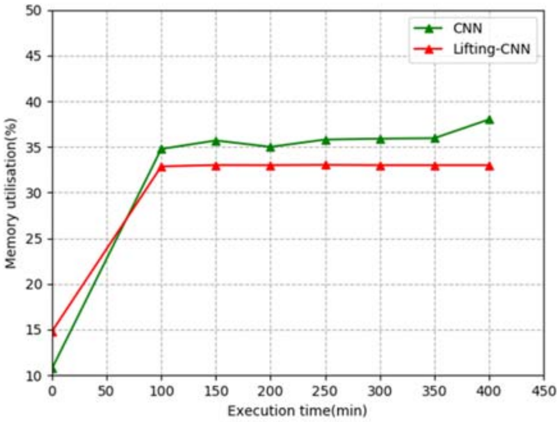


Fig. 8 Influence of lifting method on memory consumption

In terms of memory consumption, as shown in Fig.8, the model with the lifting method achieves comparative performances as the one without the lifting method. It is clear that the percentage decreases using the data reduction (lifting) method (red curve). However, the CNN without data reduction always consumes a high percentage (green curve).

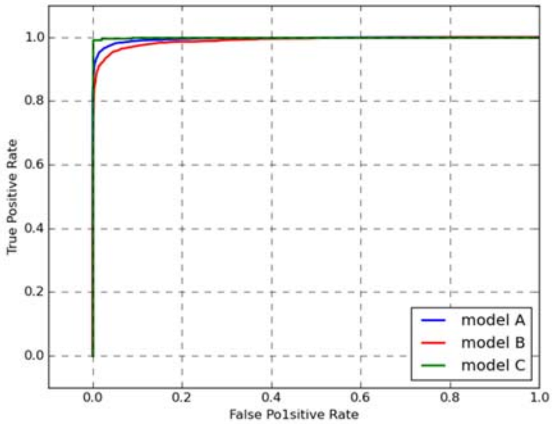


Fig. 9 Comparison with other data reduction techniques on LFW dataset: Model A: SA+CNN, Model B: Downsampling + CNN, Model C: Lifting+CNN

We note that the reduction data method based on lifting can improve the performance of the recognition model, which reduces the percentage of memory compared to other methods. This is because the amount of data has decreased, since we work with a small feature size. The lifting-CNN is potentially more suitable and practical, which consumes less memory. As a result, we work with a small feature size such that it needs no storage space for face features.

5.5 Performance with other data reduction methods

It is true that data reduction within CNNs helps us to decrease material consumption and is effective in terms of time and memory, but image quality and precision always remain priorities to achieve an efficient face recognition system.

In order to validate the importance of the lifting method decomposition compared with other data reduction methods, the images are compressed further with the Spiral Architecture (SA) [30] and on simple downsampling techniques [31].

We apply the data reduction methods and train our model with different configurations as follows: We work on SA-based image compression and we focus on the properties of the hexagonal pixel labeling scheme. Indeed, we generate an image with non-rectangular pixels to decrease the amount of data by changing the pixel shape of each image using the SA. Then, we implement simple downsampling, which is typically used to reduce the storage and/or transmission requirements of images.

Model A: SA+CNN

Model B: Downsampling+CNN

Model C: Lifting+CNN

Table 2 : Performance (%) of comparison with other data reduction techniques on LFW dataset.

Method	Accuracy (%)
SA+CNN (Model A)	97.0±0.006
Downsampling+CNN (Model B)	95.2±0.009
Lifting+CNN (Model C)	99.4±0.006

Table 2 comprises the detailed performance of different trained models with lifting, SA and simple downsampling methods. As shown, the model A reaches 97% on LFW. Model B attain only 95.2%. Evidently, our proposed method (Model C) has the most significant performance with 99.4%. The results evaluated with lifting decomposed images are mostly enhanced in terms of recognition rate.

Also, among the evaluation methods for face recognition systems, we have the ROC curves (receiver operating characteristic) which illustrate the ability of the classifier. It is created by plotting the true positive rate (TPR) against false positive rate (FPR). Fig.9 provides ROC curves for each model (Model A, Model B and Model C) on the LFW dataset.

As illustrated, the model implemented with the sampling method during training does not achieve good results (red curve). Although, the model training with the lifting method reaches the best result which presents the highest level (green curve).

By making the features more discriminating, our proposed model outperforms the rest of the models. Consequently, the obtained results exhibit the importance of the method compared to other methods of data reduction in terms of recognition rate using the LFW database. Table 2 and Fig.9 shows the influence of reduction data on the lifting-CNN (Model C) which clearly achieves the best performance in the verification task.

Although by applying SA during preprocessing, it gives acceptable results in terms of rate recognition but our model remains the most efficient.

Indeed, this model retains the amount of images more than the other two models, which generates a higher recognition rate. This is very useful for reducing the storage size of images while preserving as much of their information as possible. As it is evident the lifting scheme decomposition not only reduces the computational complexity of the CNN model but also enhances the recognition performance significantly.

Table 3: Performance (%) of comparison with the state-of-the-art on LFW dataset.

Method	Accuracy(%)
DeepFace[32]	97.15±0.27
FaceNet[14]	99.63±9
VGG[12]	98.37
WebFace[33]	97.73±0.31
DeepID2[34]	98.97±0.25
DeepID[31]	97.45±0.26
Joint-Alex[23]	98.0.3±0.23
LightCNN[22]	98.13
ArcFace[10]	99.83
PFEs[35]	99.82
Our Model	99.40±0.009

Table 4: Performance (%) ofcomparison with the state-of-the-art on YTF dataset.

Method	Acccuracy(%)
VGG[12]	97.30
DeepFace[32]	91.40±1.1
WebFace[33]	92.24±1.28
Joint-Alex[23]	92.32±0.40
Joint-Res[23]	93.12±0.43
ArcFace[10]	98.02
SeqFace[36]	98.12
Our Model	97.66

5.6 Comparison with the state-of the-art

In our practical application of face recognition, we take into consideration the accuracy as well as the prediction speed to achieve high performances. The objective is to find a trade-off between prediction speed and accuracy, important accuracy and high prediction speed. To verify the efficiency of our CNN framework in terms of recognition rate, we compare our model with widely used models. As shown in table 3, our proposed approach attains state-of-the-art results of 99.40% accuracy on LFW, which is best among all the methods, such as DeepID [31], DeepID2[34], DeepFace [32], WebFace [33], Joint-Alex[23], Light CNN 9[22] and VGG [12]. However, FaceNet [14] attains 99.63%. It is worth noting that the FaceNet has a distinctly important capability by utilizing the triplet loss function. PFEs [35] achieve 99.82%, and our training set is inferior to theirs. Moreover, the ArcFace method [10] achieves 99.83% accuracy. To attain high performances, the authors in [10] used eight GPU cards (four NVIDIA Tesla P40 (24GB) GPUs) on data training, which would generate high GPU memory consumption and a massive computational cost.

Besides, the computational performance is also a critical aspect in comprehensive evaluation of face recognition systems. We achieve an important performance on the LFW benchmark with only single CNN model in unconstrained environment by applying lifting technique to resolve the computational and memory requirements problems. As a consequence, it is shown that the lifting operation is suitable for our general CNN architecture, when the accuracy is significantly imperative without any precision degradation. We also study our model with the YTF dataset and we present the results in Table 4. The accuracy of our method achieves an important performance (97.66 %) on the YTF dataset.

The comparison with common recorded methods in Table 3 and table 4 indicates that our approach significantly improves the performance. Nevertheless, most recognition face applications have not addressed and resolved the problem of the increased usage of computational and memory resources yet. For example, despite having the very high recognition rate for the ArcFace method, many difficulties remain due to the huge consumption GPU memory and computational cost in the case that there are massive quantities of data training. The solution in our work to resolve this problem is to use the lifting method which allows reducing data during training as well during classification in the CNN framework. Particularly, we only consider the LL part where the information is located, which contains sufficient feature details using the lifting method.

Face recognition systems require a lot of data during learning in order to achieve excellent results, which makes learning very long. In our work, we use the large VGGFace2 dataset during training to improve face recognition in an unconstrained environment. Additionally, during training, we augment data by random cropping and horizontal flipping for data augmentation to make the learning environment more difficult. Therefore, these tasks still require a reduction in the input data while maintaining precision in finer details to reduce the number of parameters to be learned by the network and for a better inference speed. Our proposed model balances between the competing objectives. On the one hand, our model can reduce the input data to make the system faster. On the other hand, it reaches important accuracy in order to protect the role of the classification system. In fact, by using the lifting method, we only keep the LL sub-band; and the obtained results show the efficiency and impact of this method proposed on deep learning. Indeed, it improves both precision and robustness of the classification while improving the discriminating power of deeply learned features.

6. Conclusion

In this paper, a novel face verification system based on deep CNN with data reduction has been suggested to enhance robustness and achieve important performances on unconstrained environments. It is well-known that the CNNs are time-consuming with a massive memory cost in the training step. Hence, we have used the lifting method, which vastly overcome major computations and make the model more efficient and faster. Thus, it is an effective solution for the computational complexity of the CNN and the memory consumption when no extra storage space for face features is needed. The Experiments have been implemented on the publicly available LFW and YTF face databases. The proposed model has achieved high performances (99.4% on LFW dataset and 97.66% on YTF dataset) using unconstrained image faces with only one single CNN. Therefore, our model is designed in a way that the running time and memory consumption are reduced, which also improve the performance of the face representation.

7 References

[1] K.Simonyan, and A.Zisserman, "Very deep convolutional networks for large-scale image recognition,"arXiv preprint arXiv:1409.1556 , 2014.

[2] S. R Arashloo, andJ.Kittler, "Fast pose invariant face recognition using super coupled multiresolution Markov Random Fields on a GPU,"Pattern Recognition Letters, vol. 48,pp. 49-59, 2014.

[3] Y. Gong, L. Liu, M. Yang, andL.Bourdev, "Compressing deep convolutional networks using vector quantization,"arXiv preprint arXiv:1412.6115, 2014.

[4] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, andR. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors,"arXiv preprint arXiv:1207.0580, 2012.

[5] K. He, andJ.Sun, "Convolutional neural networks at constrainedtime cost," IEEE Conference on Computer Vision and PatternRecognition (CVPR),pp.5353–60, 2015.

[6] Q.Li, J.Yu, T.Kurihara, H.Zhang, andS. Zhan, "Deep convolutional neural network with optical flow for facial micro-expression recognition," Journal of Circuits, Systems and Computers, vol. 29, pp. 2050006, 2020.

[7] Y. Sun,X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," IEEE Conf. on Computer Vision and Pattern Recognition, Columbus, pp.1891–1898, 2014.

[8] Y.Sun, X. Wang, and X.Tang, "Deep learning face representation by joint identification-verification," Advances in Neural Information Processing Systems, Montreal, Canada, pp.1988–1996, 2014.

[9] G. Hu, Y.Yang, D. Yi *et al,*"When face recognition meets with deep learning, an evaluation of convolutional neural networks for face recognition,"IEEE Int. Conf. on Computer Vision Workshops, Santiago, pp.142–1502015.

[10] J. Deng, J.Guo, N.Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4690-4699, 2019.

[11] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," In Advances in neural information processing systems,pp. 1097-1105, 2012.

[12] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, " Return of the devil in the details, Delving deep into convolutional nets,"arXiv preprint arXiv:1405.3531, 2014.

[13] C. Szegedy, W.Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, and A. Rabinovich, "Going deeper with convolutions,"In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9, 2015.

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering,"In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 815-823, 2015.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

[16] W. Liu, Y. Wen, Z. Yu, M. Li, B.Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 212-220, 2017.

[17] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, and W. Liu, "Cosface: Large margin cosine loss for deep face recognition," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognitionpp. 265-5274, 2018.

[18] S. Munasinghe, C. Fookes, and S.Sridharan, "Human-level face verification with intra-personal factor analysis and deep face representation," IET Biometrics, vol. 7, pp. 467-473, 2018.

[19] M. Masud, G. Muhammad, H. Alhumyani, S. Alshamrani, O. Cheikhrouhou, S. Ibrahim, and M. S. Hossain, "Deep learning- based intelligent face recognition in IoT-cloud

environment," Computer Communications, vol. 152, pp. 215-222, 2020.

[20] M. Iqbal, M. S. I. Sameem, N. Naqvi, S. Kanwal, and Z.A. Ye, "deep learning approach for face recognition based on angularly discriminative features,"Pattern Recognition Letters, vol. 128, pp. 414-419, 2019.

[21] S. Banerjee, and S. Das, "LR-GAN for degraded Face Recognition," Pattern Recognition Letters, vol. 116, pp. 246-253, 2018.

[22] X. Wu, R. He, Z. Sun, and T.Tan, "A light cnn for deep face representation with noisy labels," IEEE Transactions on Information Forensics and Security,vol.13,pp. 2884-2896,2018.

[23] Li, Z. M., Li, W. J., and Wang, J. "Self-Adapting Patch Strategies for Face Recognition,"International Journal of Pattern Recognition and Artificial Intelligence, vol.34,pp. 2056002, 2020.

[24] Wang, M., andDeng, W. "Deep face recognition with clustering based domain adaptation," Neurocomputing, 2020.

[25] Lu, Z., Jiang, X., andKot, A. "Deep coupled resnet for low-resolution face recognition," IEEE Signal Processing Letters, vol.25, pp. 526-530, 2018.

[26] W. Ding, F. Wu, X. Wu, S. Li, and H. Li, "Adaptive directional lifting-based wavelet transform for image coding," IEEE Transactions on Image Processing, vol.16,pp. 416-427, 2017.

[27] K. Zhang, X. Ren, S. and Sun, J, "Deep residual learning for image recognition," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 770–778, 2016.

[28] J. Xiang, and G. Zhu,. "Joint Face Detection and Facial Expression Recognition with MTCNN," In 2017 4th International Conference on Information Science and Control Engineering (ICISCE), pp. 424-427, 2017.

[29] H. Wang, X. He, T. Hintz, and Q. Wu, "Fractal image compression on hexagonal structure," Journal of Algorithms & Computational Technology, vol.2,pp. 79-98, 2008.

[30] A. Youssef, "Image downsampling and upsampling methods," National Institute of Standards and Technology, 1999.

[31] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," In Advances in neural information processing systems,pp. 1988-1996,2014.

[32]M. Kirby, and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," IEEE Transactions on Pattern analysis and Machine intelligence, vol.12,pp. 103-108,1990.

[33] D. Yi, Z. Lei, S. Liao, and S.Z. Li,"Learning face representation from scratch," arXiv preprint arXiv:1411.7923,2014.

[34] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," In: IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 2892–2900,2015.

[35] Shi, Y., and Jain, A. K. "Probabilistic face embeddings," In Proceedings of the IEEE International Conference on Computer Vision, pp. 6902-6911,2019.

[36] Hu, W., Huang, Y., Zhang, F., Li, R., Li, W., andYuan, G. "SeqFace: make full use of sequence information for face recognition," arXiv preprint arXiv:1803.06524,2018.

[37] Liu, W., Wen, Y., Yu, Z., andYang, M. "Large-margin softmax loss for convolutional neural networks," In ICML, pp. 507–516,2016.

[38] Y. Wen, K.Zhang, Z. Li, andY. Qiao, "A comprehensive study on center loss for deep face recognition," International Journal of Computer Vision, vol.127,pp. 668-683,2019.

Hana ben fredj got her fundamental license and MS degree in Microelectronics from the Higher Institute of Informatics and Mathematics of Monastir, Tunisia, in 2011 and 2014, respectively. Currently, she is preparing her PhD degree in Electronics and Microelectronics in the Faculty of Sciences of Monastir. Her main research includes processing image, recognition pattern, parallel architecture and graphics processor.

Safa Bouguezzi got her fundamental license and MS degree in Microelectronics from the Higher Institute of Informatics and Mathematics of Monastir, Tunisia, in 2013. She got her MS degree in Microelectronics from Higher Institute of Applied Sciences and Technology Sousse, Tunisia. Currently, she is preparing her PhD degree in Electronics and Microelectronics in the Faculty of Sciences of Monastir. Her main research includes Embedded System, processing image, recognition pattern, parallel architecture.

Chokri Souani is Professor in Electronics and Microelectronics, at Higher Institute of Applied Sciences and Technology Sousse, Tunisia. He is currently team leader in the Microelectronics and Instrumentation Laboratory µEI (LR13ES12). His research interests include Software Defined System, SDR, SD-SoCSoC, MPSoC, Embedded System, Computer Vision, Big Data, IoT, Smart City, Communicant Vehicle & ITS, Small Satellite & Applications.