Journal of Advanced Engineering Trends

ISSN : 2682 - 2091

Vol. 44, No. 1, January 2025

Developing Deep Learning Based Facial Recognition Technique

Hossam M. Elian[1], Gamal M. Dousoky [1,*], Ali Hafez[2]

[1] *Electrical Engineering Dep., Faculty of Engineering, Minia University, Minia, Egypt*

[2] *National Research Institute Of Astronomy and Geophysics, Giza, Egypt*

* *Corresponding author(s) E-mail: dousoky@mu.edu.eg*

A R T I C L E I N F O

Article history:
Received: 22 September 2024
Accepted: 16 December 2024
Online: 25 February 2025

Keywords:
Face Recognition
Deep Learning
Auto-Encoder
Capsule Network

A B S T R A C T

Identity verification is becoming more and more crucial in a variety of real-world applications, including identity checks at airports, apartment door locks, and cell phones. This process requires a methodology that is fast, precise, scalable to accommodate additional users, and adaptable to variations in face angle, brightness, and other variables. In order to address the aforementioned difficulties, we provide four facial recognition techniques in this work. First, the CNN architecture was presented. Then, We developed the CNN Decoder for face encoding to improve model accuracy and overcome the difficulty of retraining the model when adding new people. Next, we presented two capsule network topologies to address face angle-related problems. The COMSATS Face Dataset is the dataset that we used in our study for testing, training, and assessment. According to an experiment, CNN recognizes faces with 93% accuracy, the decoder with 99% accuracy, the CapsNet with CNN 81% accuracy, and the CapsNet with VGG-19 with 99% accuracy. The latter is thought to yield the greatest results when it comes to distinguishing faces from various viewing angles.

1. Introduction

Facial recognition technology has witnessed remarkable advancements in recent years, driven primarily by breakthroughs in deep learning. This technology has found widespread applications across various domains, including security, surveillance, and personal identification. The ability to accurately and efficiently identify individuals based on their facial features has become increasingly crucial in today's world. Convolutional Neural Networks (CNNs) have emerged as a cornerstone of modern facial recognition systems. Their proficiency in extracting meaningful features from images has significantly improved the accuracy and reliability of these systems. By leveraging the capabilities of CNNs, researchers have been able to develop models that can effectively handle variations in lighting, facial expressions, and poses. While CNNs have achieved significant success, Capsule Neural Networks (CapsNets) represent a promising alternative approach. CapsNets introduce a novel concept of routing by agreement, which allows for a more nuanced understanding of spatial relationships within images. This capability can potentially enhance the performance of facial recognition systems, particularly in handling diverse facial expressions and viewpoints. To effectively implement facial recognition using deep learning, a comprehensive understanding of neural network architectures, data preparation, training methodologies, and hyper-parameter optimization is essential. CNNs, with their ability to capture spatial hierarchies, are well-suited for facial recognition tasks. Additionally, auto-encoders can be employed to perform feature extraction and dimensionality reduction, aiding in the preprocessing of facial data. This paper will delve into the intricacies of deep learning techniques for facial recognition, focusing on CNNs and CapsNets. We will explore the key components of these architectures, discuss data preparation strategies, and provide insights into training methodologies and hyper-parameter optimization. By understanding these fundamental aspects, researchers and practitioners can develop more robust and effective facial recognition systems.

2. Literature Review

Deep learning has revolutionized face recognition technology, enabling highly accurate and efficient identification systems. Convolutional Neural Networks (CNNs) form the backbone of modern face recognition algorithms, in addition Auto-encoder and Capsule network also used in learning intricate facial features from large datasets. These deep learning models can overcome challenges like varying lighting conditions, facial expressions, and partial occlusions, making them invaluable in security, authentication, and social media applications. In [1] they used binary classification with Capsule-Net and achieve 99.92% in classification. In [2] the authors propose a multi-layer capsule network with a joint dynamic routing algorithm and a new loss function. This enhanced model outperforms traditional CNNs and current state-of-the-art methods in fire recognition, In [3] the author developed model recognizing people while they are wearing masks using DeepMaskNet model, In [4] This study evaluates four neural network architectures, including two types of convolutional neural networks (CNNs) as a baseline, and CapsNet , the CapsNet demonstrates enhanced efficiency and speed in facial recognition tasks, particularly in detecting a driver's head position within the bus cabin's wide view, In [5] the study proposes an advanced approach that refines the AdaBoost method and the skin color method, aiming to reduce false detection rates, The study also evaluates the performance of the AdaBoost method, skin color method, and a combined skin color + AdaBoost approach. In [6] and [7] and [8] the study focusses on recognize hand written digit using CNN and CapsNet. In [9] they develop a deep Convolutional Neural Network (CNN) with four layers achieve an accuracy of 92.2% on the test set. In [10]This paper summarizes the Masked Face Recognition Competitions (MFR) held during the 2021 International Joint Conference on Biometrics (IJCB 2021), which featured 10 participating teams from diverse academic and industry backgrounds across nine countries. A total of 18 valid solutions were submitted, focusing on improving face recognition accuracy for masked individuals while also considering model compactness for deploy ability. The evaluation utilized a private dataset that simulated real masked capture scenarios. Notably, 10 of the submitted solutions outperformed one of the leading academic face recognition models in masked face verification accuracy. In [11] This Paper descript A capsule where A capsule is a group of neurons whose activity vector represents the instantiation parameters of a specific type of entity such as an object or an object part. We use the length of the activity vector to represent the probability that the entity exists and its orientation to represent the instantiation parameters. Active capsules at one level make predictions, via transformation matrices, for the instantiation parameters of higher-level capsules. When multiple predictions agree, a higher-level capsule becomes active. We show that a discriminatively trained, multi-layer capsule system achieves state-of-the-art performance on MNIST and is considerably better than a convolutional net at recognizing highly overlapping digits. To achieve these results we use an iterative routing-by-agreement mechanism: A lower-level capsule prefers to send its output to higher level capsules whose activity vectors have a big scalar product with the prediction coming from the lower-level capsule. In [12] this paper discusses the face inversion effect (FIE), which highlights the greater difficulty in recognizing faces compared to other objects when they are presented upside-down. It examines the neural mechanisms involved in face processing, particularly focusing on the fusiform face area (FFA) and the N170 component, and how these are affected by face orientation. Additionally, it explores the role of expertise in face recognition and its implications for understanding developmental differences, such as those seen in individuals with autism. In [13] this paper explores the application of CNNs to face recognition using a small dataset, comparing their performance to traditional methods like Principal Component Analysis. It also investigates the effects of data augmentation, a technique used to enhance dataset size and mitigate under fitting or overfitting. Results show that while data augmentation improves accuracy, it can lead to an unstable learning curve. Ultimately, the CNN achieved a 94% accuracy on the augmented dataset with just two convolutional layers. In [14] and [15] offers a comprehensive overview of Convolutional Neural Networks (CNNs) in face recognition, analyzing 150 research papers to explore CNN models, primary datasets, accuracy levels, research focuses, and future prospects. It highlights common CNN architectures and techniques used in facial recognition. In [16] this work examines the impact of convolutional network depth on accuracy in large-scale image recognition. The main contribution is a comprehensive evaluation of increasingly deeper networks, revealing that significant improvements over previous configurations can be achieved by Utilizing 16-19 weight layers. These findings formed the foundation for our submission to the ImageNet Challenge 2014, where our team achieved first and second places in the localization and classification tracks, respectively. In [17] the ace recognition is achieved using the LSCO method, integrating Light Spectrum Optimizer and Chimp Optimization, resulting in an accuracy of 93.5%, True Positive Rate (TPR) of 90.7%, True Negative Rate (TNR) of 87.5%, and an execution time of 0.236 seconds. In [18] this paper present a facial recognition technique leveraging capsule networks, which model facial features hierarchically for efficient recognition. It begins by contrasting capsule networks with convolutional neural networks, focusing on their operational principles and structures. The research delves into dynamic routing algorithms and the internal workings of capsules. Experiments on face datasets, optimized with the Adam algorithm alongside boundary and reconstruction loss, enable the capsule network to learn more robust feature representations. Results indicate that this capsule network-based system achieves a 93.5% accuracy on the WebFace dataset, demonstrating its feasibility and effectiveness in facial recognition. In [19] this research focuses on developing a facial recognition system using capsule networks, which are designed to model facial features hierarchically for efficient recognition. It compares the advantages of capsule networks over traditional convolutional neural networks (CNNs) and highlights their ability to capture spatial relationships in images. The study demonstrates the system's effectiveness, achieving a 93.5% accuracy rate on the WebFace dataset. In [20] they used Capsules networks to identify animal from its faces for different bat families and same cow family but different members. In [21] this research investigates the effectiveness of a deep learning model (VGG16-CBAM) for classifying seven species of horseshoe bats, which exhibit high interspecific similarity and intraspecific variation. It utilizes a dataset of 879 images collected over nine years to achieve a classification accuracy of 92.15%. The study aims to enhance species identification methods in taxonomy through advanced image-based techniques. In [22] the author proposed the Residual Capsule Network, a novel framework that combines the strengths of Capsule Networks and Residual Networks. By replacing conventional convolutional layers in Capsule Networks with skip connections from Residual Networks, the author aimed to reduce complexity and improve training efficiency. The model was evaluated on the MNIST and CIFAR-10 datasets, demonstrating a significant decrease in the number of parameters compared to baseline models. In [23] the author try to find best face recognition technique by develop 10 methods for Face Recognition. In [24] the author introduced capsule networks (CapsNets) as an improvement over traditional convolutional neural networks (CNNs) for computer vision tasks. They highlighted the limitations of CNNs, particularly in situations where the relevance of detected objects changes, and demonstrated that CapsNets can achieve higher detection rates in challenging environments. This work aims to enhance the robustness and effectiveness of object detection in deep learning applications. In [25] he author proposed a new angular margin loss function called X2-Softmax to improve face recognition by enhancing the reparability of facial features extracted by neural networks. Recognizing the limitations of fixed margins used loss functions like CosFace and ArcFace, which can hinder model convergence and discriminative ability due to uneven sample distributions, the author introduced adaptive angular margins. These margins increase with the angles between different classes, providing greater flexibility. In [26] the author developed a novel neural network approach for face mask recognition using capsule networks, The proposed system consists of two phases: the first phase employs VGG16 and VGG19 as pre-training modules for deep feature extraction, while the second phase utilizes the capsule network for recognizing face masks. Achieving high accuracies of 99.87%, 99.90%, and 99.94% for CapsNet, VGG16, and VGG19 on RMFD, and 99.94% accuracy for CapsNet with VGG19 on SMFD. In [27] the author outlines the evolution of limitations of traditional symbolic AI in addressing complex tasks like image recognition and object detection. Starting BY (CNNs), while also highlighting CNNs' shortcomings, such as their dependence on large datasets and difficulties in recognizing object pose and deformation. To address these challenges, the author presents Capsule Networks as a promising advancement in deep learning, noting their superior performance. The paper aims to provide a comprehensive review of current architectures, tools, and methodologies related to Capsule Networks serving as a resource for researchers and industry professionals. In [28] the paper discusses development of a deep learning framework for the early detection of Diabetic Retinopathy (DR) using Efficient Channel Attention (ECA) mechanisms. It presents various models, including ECA-Resnet101, ECA-VGG19, and ECA-InceptionV4, which are designed to improve the accuracy of DR classification from retinal fundus photographs. The study emphasizes the importance of automated diagnosis in managing DR effectively, given the increasing prevalence of diabetes and the need for regular screening. In [29] the paper discusses the design and implementation of a real-time digitizer for an earthquake monitoring system, which converts analog seismic signals into digital format. It details the components of the digitizer, including a deep learning module (CapsPhase) for detecting earthquake events and accurately picking the first arrival times of seismic waves. The study aims to enhance seismic data acquisition and monitoring capabilities using advanced technology.

3. The Proposed Technique

3.1. Dataset

The dataset used for training and evaluate the model are COMSATS Face Dataset. The dataset includes 850 images of 50 individuals under 17 different poses (0°, 5°, 10°, 15°, 20°, 25°, 30°, 35°, 55°, -5°, -10°, -15°, -20°, -25°, -30°, -35°, -55°). These images were captured closed to real-world conditions in the time span of five months in COMSATS University, Abbottabad Campus. Face images included in this dataset can reveal the efficiency and robustness of future face detection and face recognition algorithms. This dataset are available in Kaggle website.

3.2. Face Detection

we detect the face using frontal face Haar cascade algorithm detect the face only we tried with another algorithm like MTCNN to but the frontal face cascade algorithm are very fast and efficient when it comes to face detection according to the way its work, where its only filter multiply the gray image contains the faces as a result for this the position of the faces in the image become more dark than other image so we can identify the value we need to crop from image then crop at this areas.

3.3. Convolutional Neural Network

One of the most effective algorithms in deep learning is the Convolutional Neural Network (CNN), a layered architecture designed for data pattern recognition used in classification, detection, and segmentation tasks. In our proposed CNN model architecture, we utilize the frontal face Haar cascade algorithm to detect and extract faces from images. These cropped faces are then fed into the CNN model, which consists of several layers depicted in Figure 1: CNN Model Architecture Conv2D layers apply filters (kernels) to capture local patterns and features from the input data. Following this, MaxPool2D layers down-sample the feature maps, reducing spatial dimensions. This process is repeated to emphasize the most significant patterns in the image. Subsequently, a Flatten layer prepares the data for input into fully connected layers, which learn and culminate in an output layer that assigns values to represent individuals from the training dataset, employing a softmax activation function. The model was trained on a dataset collected with volunteer consent, using categorical cross-entropy loss to measure and optimize training performance, achieving an accuracy of 93%. However, we encounter two significant issues during live testing. Firstly, our deep learning model requires retraining from scratch whenever we add a new person. This necessitates editing the number of output nodes to match the number of individuals in the dataset. This is a major issue due to the time-consuming nature of training and the need to repeat this process every time a new person is added, making this approach unsuitable for the production phase Secondly, The model struggles to detect faces that are not in a frontal pose or when a person is not looking directly at the camera. In such cases, the model returns zero results, rendering the identification process incomplete until the person faces the camera directly.

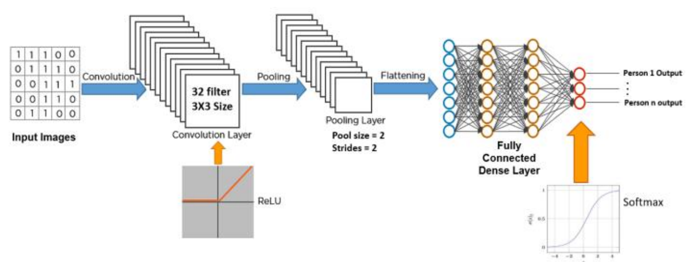

Figure 1: CNN Model Architecture

3.4. Auto-Encoder

An auto-encoder is a specialized neural network architecture designed to efficiently compress input data into its essential features (encode) and reconstruct the original input (decode), as illustrated in Figure 2: Auto-Encoder Model ***Architecture***. Depending on its purpose, an auto-encoder can be trained to extract facial landmarks and convert them into a set number of dimensions that represent a face as a numerical list. This representation can then be stored in a database rather than saving the entire face in model weights. To identify a known user, the auto-encoder decodes the user's face into a numerical list of dimensions (128 dimensions) then compares it with stored dimensions in database using the Euclidean distance, as shown in the equation. (**Error! Reference source not found.**) [30] The Euclidean distance are a number between 0, 1

$$d = \sqrt{(P1_{D1} - P2_{D2})^2 + \cdots + (P1_{D128} - P2_{D128})^2} \quad (1)$$

where in this equation P represent person face and D represent dimension. Finally the result will be number between zero and one zero means this is the identical object but one is complete different object. Then we compare between calculated distance from the equation (**Error! Reference source not found.**) and the threshold we adjust by experimental test which is (0.25) if the distance less than threshold this is the person else it's another person. In the model we train the encoder to generate 128 dimensions to represent the face in the photo. As a result for this model we achieved 99\% accuracy for frontal face recognition. As in CNN we detect the face using frontal face Haar cascade algorithm to crop the face from image then feed faces to our proposed model to identify the person in the image according to process mentioned earlier.

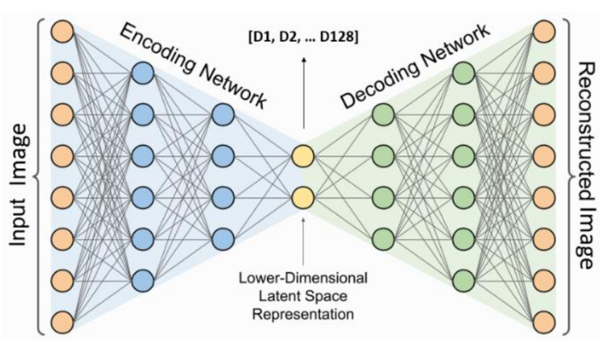

Figure 2: Auto-Encoder Model Architecture

3.5. Capsule Network

Capsule network become one of the most powerful technology because of its ability to detect the patterns exist in a photo even when the pattern pose changed. The capsule are able to collect more features about the photos more than normal CNN and auto-encoder models. The Capsule neural network is computer vision algorithm this algorithm are a unit that learns to detect an implicitly defined entity over a limited domain of viewing conditions. It outputs both the probability that entity is present in a photo and set of "instantiation parameters" [31] that reflect the features of the entity such as pose information. The presence probability is viewpoint invariant Capsule. Where the basic idea of capsule network is to encrypt the relationship between various entities (scales, location, pose, and orientation). We will start by build the capsule itself then build the model, then training the model to reach final point which is evaluate the result.

3.5.1. Capsule Structure

Capsules: These are groups of neurons that represent entities or parts of entities within an image to enable multiple entities identification within the single image. Capsules are designed to capture the pose, size, and other properties of the entities they represent. Where the Capsules represent entities (e.g., objects, parts of objects) rather than just individual features like in CNN model.

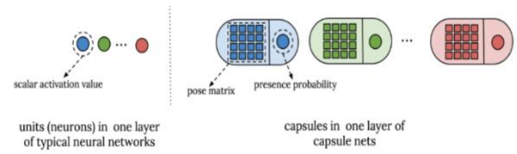

Figure 3: Capsule Structure

3.5.2. Input layer

The typical input for a capsule network is a color image represented as three-channel pixel data.

3.5.3. Feature extraction layer

These layers consist of convolutional and pooling operations that transform raw pixel values into higher-level features like edges, textures, and patterns.

3.5.4. Primary Capsule

The primary capsules represent the most basic multi-dimensional entities. From an inverse graphics standpoint, activating these primary capsules involves reversing the rendering process. This form of computation contrasts sharply with the assembly of instantiated parts to form recognizable wholes, which is the intended strength of capsules.

3.5.5. Secondary Capsule

Secondary capsule are the layer where every capsule have to represent entity, the capsule in primary capsule layer try to get the parent for each primary capsule.

3.5.6. Routing by Agreement

The dynamic routing mechanism ensures that the output of each capsule is directed to an appropriate parent capsule in the layer above. Initially, the capsule's output is distributed to all potential parents, but this distribution is adjusted using coupling coefficients that sum to 1. For each potential parent capsule, the capsule computes a "prediction vector" [32] by multiplying its own output with a weight matrix. If this prediction vector shows a significant scalar product with the output of a particular parent capsule, top-down feedback increases the coupling coefficient for that parent while decreasing it for others. This process enhances the contribution that the capsule makes to the chosen parent, thereby increasing the scalar product between the capsule's prediction and the parent's output.

$$\widehat{U_{j|i}} = W_{ij}\, U_i \tag{2}$$

$$C_{ij} = \frac{\exp(b_{ij})}{\exp(b_{ik})} = Softmax(b_{ij}) \tag{3}$$

$$S_j = \sum_i C_{ij}\ \widehat{U_{j|i}} \tag{4}$$

$$V_j = \frac{||S_j||}{1 + ||S_j||} * \frac{S_j}{||S_j||} \tag{5}$$

Procedure 1: Routing algorithm

1: **procedure** ROUTING ($\hat{\mathbf{u}}_{j|i}$, r, l)

2: for all capsule i in layer l and capsule j in layer (l + b_{ij}) ← 0

3: for r iterations do

4: for all capsule i in layer l: $\mathbf{c}_i$ ← e.q. (3)

5: for all capsule j in layer (l + 1): $\mathbf{S}_j$ ← e.q. (4)

6: for all capsule j in layer (l + 1): $\mathbf{V}_j$ ← e.q. (5)

7: for all capsule i in layer l and capsule j in layer (l + 1): $b_{ij} \leftarrow b_{ij} + \hat{\mathbf{u}}_{j|i}.\mathbf{v}_j$
Return $\mathbf{V}_j$

In the above Equations [32] we start by multiply the input with width as vectors (2) then in (3) we set the probability for this input to classified as entity in parent layer "Known faces". In first iteration the probability are divided between parent layer entities in equal. Then we multiple vectors (4) of input image entities with vectors represent parent entities. The result of this multiplication are the similarity between entities in parent layer and input image. We squash this result to value between 0, 1 to represent how much entity in parent layer are similar to parent in input in percentage. Repeating this process 3 times is more than enough to recognize input face.

3.6. Margin loss for object existence

We are using the length of the instantiation vector to represent the probability that a capsule's entity exists. We would like the top-level capsule for object class k to have a long instantiation vector if and only if that object is present in the image. To allow for multiple object, we use a separate margin loss, Lk for each object capsule [31]:

$$L_k = T_k\ \max(0, m^+ - ||V_k||)^2 + \lambda\ (1 - T_k) \max(0, m^- - ||V_k||)^2 \tag{6}$$

3.7. CapsNet with CNN layer

The basic capsule network consists of two convolutional network layers designed for feature extraction. The initial RGB face, sized [224 × 224 × 3], first undergoes the Conv1 layer (kernel size: [9 × 9 × 64], stride = 2, padding = 1), these maps then pass through the Conv2 layer (kernel size: [53 × 53 × 128], stride = 2, padding = 1) as illustrated in Figure 4: CapsNet with CNN ***Architecture***. Importantly, the absence of a pooling layer between Conv1 and Conv2 ensures that crucial spatial features of the target are retained.

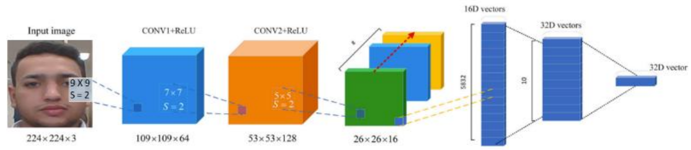

Figure 4: CapsNet with CNN Architecture

3.8. CapsNet with VGG16

VGG16 is a popular deep convolutional neural network architecture used for image classification and object detection. It was introduced by Simonyan and Zisserman [16] and achieved state-of-the-art performance on the ImageNet classification dataset. VGG16 is characterized by its simple and uniform architecture, consisting of multiple convolutional layers followed by pooling layers, and finally fully connected layers. With 16 convolutional layers, which contributes to its strong performance. The structure of layers utilize 3×3 filters, designed with a small receptive field to capture spatial relationships such as left/right, up/down, and center The small convolutional filters allows for greater depth without increasing the computational cost. Additionally, some configurations include 1×1 convolution filters, which act as linear transformations of the input channels followed by non-linear activation. The convolution operation maintains the spatial resolution of the input with a fixed stride of 1 pixel and utilizes 1-pixel padding for the 3×3 conv. layers to preserve spatial dimensions after convolution. The network incorporates five max-pooling layers for spatial pooling, applied after certain conv. layers. Each max-pooling operation reduces the spatial dimensions by a factor of 2, using a 2×2 pixel window with a stride of 2. Following the stack of convolutional layers, three Fully-Connected (FC) layers are employed. The first two FC layers consist of 4096 channels each, while the third layer is tailored for 1000-way ILSVRC classification, comprising 1000 channels, each representing a distinct class in the dataset. The final layer employs a soft-max function for classification. The configuration of these fully connected layers remains consistent across different network architectures. Throughout the network, rectified linear units (ReLU) serve as the non-linear activation function for all hidden layers. VGG16 and CapsNet can be combined into a hybrid architecture, where the output of VGG16 is used as input to a CapsNet layer. This can leverage the strengths of both architectures, potentially leading to best result In our model VGG16 are used to extract robust features from face images to, which then be fed into a Capsule Network for further processing and finally train on those features to maximize the training process.

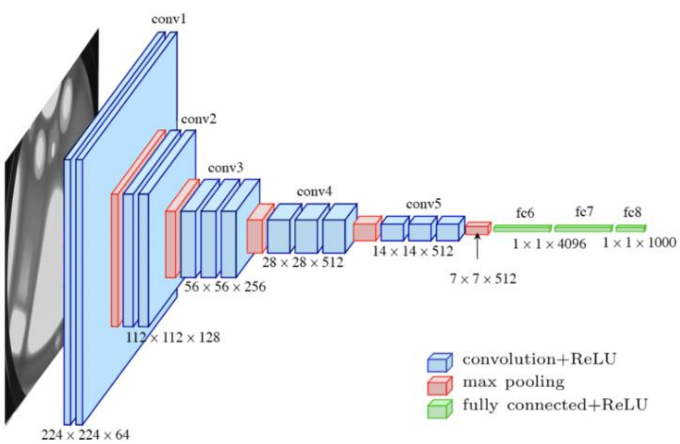

Figure 5: CapsNet with VGG-16 Model Architecture

3.9. CapsNet Vs CNN

In convolutional neural networks every neuron in model work as single unit containing its single representation of data, on other hand the CapsNet model contains capsule network which is group of neurons combined together as treated as a single unit representing object in image having the existence of this object as value between zero and one and angle to represent variation in object characteristics such as image brightness, image color thickness,..etc. All this representation of data support face recognition process and increase the probability of distinguish between people similar face landmark and increasing the CapNet model efficiency over the CNN Model.

4. Result

The result represent that the models has been listed are able to identify the individuals according to Table 1.

5. Conclusion

This study proposes four face-recognition methodologies to address these challenges effectively. Initially, we introduced the CNN Model architecture consisting of a Convolutional layer followed by a max pooling layer for feature extraction; in addition, to layer flatten fully connected layer generating with soft-max activation function for output layer, but with the need to retrain the model to add new person which affects model deploy ability we develop new decoder model which decodes the face to 128 dimensions enabling saving those 128 dimensions per person and comparing again with Appling threshold to distinguish between faces all this without the need for retraining when new individuals are added, thereby improving overall accuracy. We introduced two capsule network architectures to overcome issues related to face angles. The first one used a normal CNN layer for feature extraction. This counter-measure angle poses limitations but leads to poor performance issues due to the poor feature extraction layer. In this result, we increase the feature extraction layer in the last model using VGG-19, which increases the model's efficiency. Our experiments, conducted using COMSATS Face Dataset, demonstrated significant performance: CNN achieved 93% accuracy “Frontal only”, the decoder achieved 99 % “Frontal only”, CapsNet with CNN achieved 81\%, and the CapsNet with VGG-19 achieved the highest accuracy of 99\%, particularly notable for its ability to recognize faces from diverse viewing angles.

Table 1: Comparison

Compare point	CNN	Auto-Encoder	CapsNet & CNN	CapsNet &VGG16
Accuracy (Frontal)	**93.5%**	**99%**	**81%**	**99%**
Accuracy (non Frontal)	**77**	**NA**	**79**	**98**
Training Image Required	**300 ↑**	**1**	**1**	**1**
Retrain Required to add person?	**Yes**	**No**	**Yes**	**Yes**
Translation Invariant	**No**	**No**	**Yes**	**Yes**
Reliability	**Low**	**High**	**Low**	**High**
Training Speed	**Low**	**Highest**	**Low**	**Low**

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Holguin-Garcia, S. Alejandro, and Guevara-Navarro, Ernesto, and Daza-Chica, A. Eduardo, and Patin-Claro, M. Alejandra, Arteaga-Arteaga and H. B. , "A comparative study of CNN-capsule-net, CNN-transformer encoder, and Traditional machine learning algorithms to classify epileptic seizure," BMC Medical Informatics and Decision Making, p. 60, 2024.

[2] Y. W. Xie, a. L. Cen, a. S. Kan and a. Yongfang, "Multi-layer capsule network with joint dynamic routing for fire recognition," Image and Vision Computing, vol. 139, p. 104825, 2023.

[3] Bourouis, N. Ullah, a. A. Javed, a. Mustansar, A. Ghazanfar, a. A. Alsufyani and a. Sami, "A novel DeepMaskNet model for face mask detection and masked facial recognition," Journal of King Saud University - Computer and Information Sciences, vol. 34, pp. 9905-9914, 2022.

[4] Hollósi, János and Ballagi, Áron and Kovács, Gábor and Fischer, Szabolcs and Nagy and Viktor, "Bus Driver Head Position Detection Using Capsule Networks under Dynamic Driving Conditions," Computers, vol. 13, no. https://www.mdpi.com/2073-431X/13/3/66, p. 13, 2024.

[5] Y. Sun, Z. and Ren and W. and Zheng, "Research on Face Recognition Algorithm Based on Image Processing," Computational Intelligence and Neuroscience, 2022.

[6] S. Nalluri, S. and Thota, A. and Thota, V. and v and R. and Ac, "International Journal of Innovative Research in Computer and Communication Engineering," Handwritten Digit Recognition Using CNN, no. 10.15680/IJIRCCE.2023.11010012, p. 8849, 2023.

[7] K. SOOD and J. and Fiaidhi, "Capsule Networks: An Alternative Approach to Image Classification Using Convolutional Neural Networks," 2020.

[8] R. Dixit, R. and Kushwah and S. and Pashine, "Handwritten Digit Recognition using Machine and Deep Learning Algorithms," International Journal of Computer Applications, 2020.

[9] Q. Kuang, "Face Image Feature Extraction based on Deep Learning Algorithm," Journal of Physics: Conference Series, vol. 1852, p. 032040, 2021.

[10] F. B. Nieto, a. N. Damer, a. J. N. Kolf, a. K. Raja, a. F. Kirchbuchner, a. R. Ramachandra, a. A. Kuijper, a. P. Fang, a. C. Zhang, a. F. Wang and a. D. Montero, "MFR 2021: Masked Face Recognition Competition," https://arxiv.org/abs/2106.15288, 2021.

[11] G. E. Hinton, Z. Ghahramani, Teh and Y. Whye, "Learning to Parse Images," MIT Press, vol. 12.

[12] Gauthier and B. Rossion, ""How Does the Brain Process Upright and," Inverted Faces?, Behavioral and Cognitive Neuroscience Reviews.

[13] Chen and S. , "CNN combined with data augmentation for face recognition on small dataset," Journal of Physics: Conference Series, no. https://dx.doi.org/10.1088/1742-6596/2634/1/012040, 2023.

[14] E. Barcic, P. and Grd and I. and Tomicic, "Convolutional Neural Networks for Face Recognition: A Systematic Literature Review," 2023.

[15] Q. Wang, Z. and Yu and Y. and Zhang, "Face Recognition Using Convolutional Neural Networks," 2021.

[16] K. Simonyan and A. and Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," no. http://dblp.uni-trier.de/db/conf/iclr/iclr2015.html#SimonyanZ14a, 2015.

[17] S. M. Ramesh, a. D. Thamaraiselvi, a. S. Dhariwal, a. D. Aggarwal and a. J. V. Naga, "LSCO: Light spectrum chimp optimization based spinalnet for live face detection and recognition," Expert Systems with Applications, vol. 250, no. 0957-4174, p. 123585, 2024.

[18] N. Bourdakos, "Capsule Networks Are Shaking up AI–Heres How to use them," Hacker Noon, 2017..

[19] J. Shi and L. and Zhao, "Face Recognition System Based on Capsule Networks," International Journal of Advanced Network, Monitoring and Controls, vol. 9, pp. 22-31, 2024.

[20] F. Xu, J. and Gao and X. and Pan, "Cow Face Recognition for a Small Sample Based on Siamese DB Capsule Network," IEEE Access, vol. 10, 2022.

[21] Cao Z, Wang K, Wen J, Li C, Wu Y, Wang X and Y. W., "Fine-grained image classification on bats using VGG16-CBAM: a practical example with 7 horseshoe bats taxa (CHIROPTERA: Rhinolophidae: Rhinolophus) from Southern China. Front Zool," Frontiers in Zoology, vol. 21, 2024.

[22] S. B. Shruthi Bhamidi and M. and El-Sharkawy, "Residual Capsule Network," IEEE, pp. 0557-0560, 2019.

[23] M. Jawad, and Zaeer Dhaam, Haidar and F. A. , "An overview of face recognition methods," BIO Web of Conferences, 2024.

[24] Han, P. ZHang, a. P. Wei and a. SHuhuan, "CapsNets algorithm," Journal of Physics: Conference Series, 2020.

[25] Gong, Xu, Jiamu Liu, a. Xiaoxiang, a. X. Zhang, a. Y.-W. Si, a. X. Li, a. Z. Shi, a. K. Wang and a. Xueyuan, "X2-Softmax: Margin Adaptive Loss Function for Face Recognition," Elsevier Ltd., 2023.

[26] Abou El-Magd, Lobna, and Goda, Essam, and Darwish, Ashraf, and Hassnien and Aboul, "Face Mask Recognition System using Adapted Capsule Neural Networks for Medical Institutions," 2023.

[27] Edward, Mensah, K. Patrick, a. Adebayo, F. Adekoya, a. Ayidzoe, A. Mighty and a. B. Y., "Capsule Networks – A survey," Journal of King Saud University - Computer and Information Sciences, vol. 34, pp. 1295-1310, 2022.

[28] D. S. Salem, S. M. Sami, A. A. Donkol and a. G. M. Dousoky, "Classification of Diabetic Retinopathy (DR) using ECA Attention Mechanism Deep learning Networks," Journal of Advanced Engineering Trends (JAET), vol. 43, no. 2682-2091, p. 425–431, 2024.

[29] O. M. Saad, E. B. Helal, A. G. Hafez and a. G. M. D. , "Design and Implementation of a Trigger Digitizer for Earthquake Monitoring System," Journal of Advanced Engineering Trends (JAET), vol. 43, no. 2682-2091, pp. 419-424, 2024.

[30] L. Liberti, C. and Lavor, N. and Maculan and A. and Mucherino, "Euclidean Distance Geometry and Applications," SIAM Review, vol. 56, no. 10.1137/120875909.

[31] S. Sakib, G. and Hinton, S. and Sabour, N. and Google and B. and Toronto, "MATRIX CAPSULES WITH EM ROUTING," conference paper at ICLR 2018 MATRIX CAPSULES WITH EM ROUTING, no. 10.13140/RG.2.2.27416.44800.

[32] G. Hinton, S. S. and a. N. Frosst, "Dynamic Routing Between Capsules," 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA., 2017.

[33] Haq, M. Ul, M. and Sethi, R. and Ullah, A. and Shazhad, L. and Hasan and M. and Karami, "COMSATS Face: A Dataset of Face Images with Pose Variations, Its Design, and Aspects," Mathematical Problems in Engineering, vol. 11, no. 10.1155/2022/4589057, pp. 1-11, 2022.

Abbreviation and symbols

CNNs	**Convolution Neural Networks**
MTCNN	**Multi Task Convolution Neural Networks**
CapsNet	**Capsule Network**
VGG16	**Visual Geometry Group**