

A Survey of Face Recognition Based on Convolutional Neural Network

R E Saragih^{*1}, Q H To²

¹Department of Informatics Engineering, Universitas Universal, Indonesia

²Department of Industrial Engineering and Management, Yuan Ze University Taiwan

E-mail: raymonde.saragih@gmail.com¹, qyunhuong.to@gmail.com²

Submitted: 7 Jan 2022, revised: 26 Feb 2022, accepted: 27 Feb 2022

Abstract. Face recognition is one of the interesting research topics in the field of computer vision. In recent years, deep learning methods, especially the Convolutional Neural Network, have progressed. One of the successes of CNN is in face recognition. Face recognition by computer is a technique done so that the computer can automatically recognize faces in an image. Various researchers have conducted related research on facial recognition. This survey presents research related to face recognition based on Convolutional Neural Network that has been conducted. The studies used are studies that have been published in the last five years. It was performed to determine the renewal that emerged in face recognition based on Convolutional Neural Network. The basic theory of the Convolutional Neural Network, face recognition, and description of the database used in various researches are also discussed. Hopefully, this survey can provide additional knowledge regarding face recognition based on the Convolutional Neural Network.

Keywords: convolutional neural network; deep learning; face recognition; survey

1. Introduction

The human face is one of the unique parts of a body. A person can recognize others through the face. In addition to voice recognition, retinal scanning, or fingerprint recognition, face recognition is included in biometric identification [1]. Face recognition has become one of the interesting research topics in the field of computer vision. There are various researches done with the aim of the computer being able to recognize a person's face [2]. Moreover, the emergence of interest in research on face recognition is driven by requests for the application of the technology.

Face recognition by computer is a technique that is done so that the computer can recognize faces in an image automatically [3]. In general, there are four stages in face recognition, and they are face detection, facial straightening, facial feature extraction, and face matching or classification [3], [2]. The deep learning method has progressed in recent years, specifically the Convolutional Neural Network (CNN) for computer vision. Some of the successes achieved through CNN for computer vision are object detection, object recognition, gesture recognition, and one of them is face recognition. The success is due

to many datasets available for training, the use of GPUs in parallel that can speed up the computing process, and the emergence of new and more effective CNN architectures [4].

Nowadays, the field of face recognition is becoming an important issue. However, there is much research that readers sometimes find difficulties to reach a general view of the field. The current study integrates many aspects of face recognition in a brief but usual way. The focus of this survey is to discuss face recognition based on Convolutional Neural Network (CNN). This survey presents theories about the Convolutional Neural Network, face recognition, related studies, and face databases used in various studies.

This survey's structure is as follows: Section 2 discusses the basic theory of the Convolutional Neural Network. Section 3 discusses face recognition. Section 4 discusses various studies on face recognition based on Convolutional Neural Network. Section 5 presents databases used in various researches and is followed by a conclusion in Section 6.

2. Convolutional Neural Network

Convolutional Neural Network or CNN was discovered by LeCun [5]. A Convolutional Neural Network consists of several layers. The layers are arranged in sequence. An example of the structure of a Convolutional Neural Network is shown in Fig. 1.

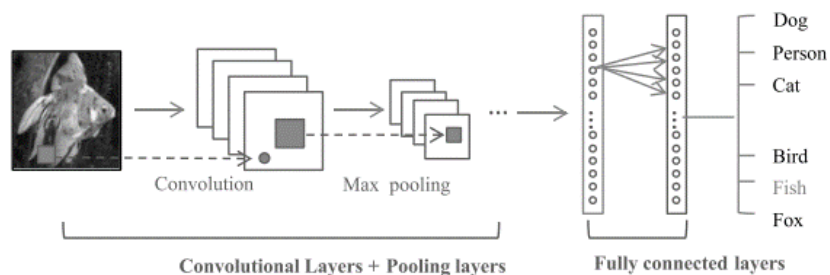


Figure 1. Structure of a Convolutional Neural Network [5]

2.1. Convolutional Layer

Convolutional Layer is the primary layer of a CNN, and the Convolutional Layer has several filters or can be called kernels. A filter is a matrix that contains numbers. The filter serves to produce a feature map from the input layer (for the first time) or a feature map from the previous layer. Filters are convoluted with the input image to produce a feature map. Convolution is a mathematical operation that multiplies two matrices and sums them. Filters are connected to the input layer through a small area called the receptive field. The extent of the filter determines the extent of the area. The values in the filter will change in the learning or training process. After the convolution calculation results are obtained, each will be processed with an activation function [6]. The size of the output of the convolutional layer is determined by several parameters, such as depth (number of filters), stride (the amount of shift of the filter), and zero paddings (adjust the spatial size of the output) [7]. Fig. 2 shows an image that is convoluted with a filter to produce a feature map.

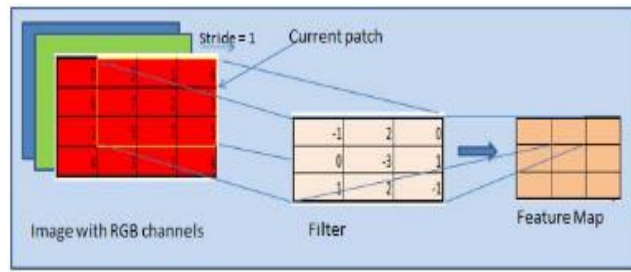


Figure 2. An image convoluted with a filter [7]

2.2. Pooling Layer

The pooling layer is placed after the convolutional layer. The pooling layer's function is to reduce the input size from the previous layer (feature map of the convolutional layer). The conventional method is by using the max-pooling or average pooling of several parts of the area in the feature map. An example of operation in the pooling layer is shown in Fig. 3.

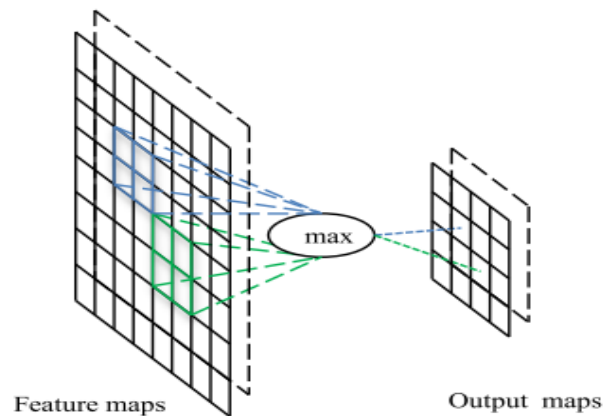


Figure 3. The operation in the pooling layer using max-pooling [8]

2.3. Fully Connected Layer

In the fully connected layer, each node is connected to all the feature map nodes from the result of the last pooling layer on CNN. An example of a fully connected layer is shown in Fig. 4.

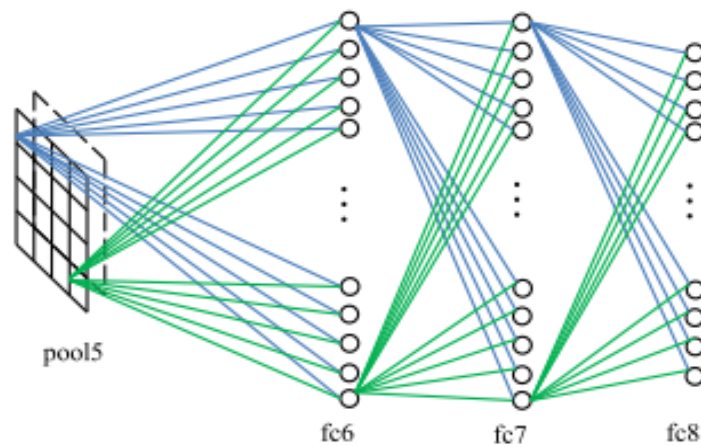


Figure 4. Fully connected layer [8]

Several fully connected layers change the feature map with the shape of 2D into a feature vector with the shape of 1D. A fully connected layer works similar to the neural network [8]. After the fully connected layer is the output layer. The amount of output depends on the goal to be achieved, for example, if CNN is used to recognize four classes' objects, then the number of outputs from the output layer is four. The relationship between the fully connected layer and the output layer is a softmax function. Softmax function changes the previous layer's features into the probability values of existing classes [5].

2.4. Activation Function

The activation function used in CNN varies. Some commonly used functions are the sigmoid function, tanh function, Rectified Linear Units (ReLU), leaky-ReLU, exponential linear units (ELU), and scaled-ELU (SELU). ReLU is the most commonly used activation function [2], [5].

3. Face Recognition

Face recognition is one of biometric identification, apart from voice recognition, retinal scanning, and fingerprint recognition [1]. Face recognition is one of the challenges in the field of pattern recognition and computer vision, and one of the many studies conducted in recent years [9], [10], [5]. The trend is due to the increasing demand for face recognition technology to be applied in law enforcement, commercial use [10], human-computer interaction, access control, surveillance [11], video conference [1], authentication on mobile devices, payment transactions, autonomous car [9], automatic attendance system, and digital entertainment [12]. Face recognition become famous for its nature, which does not require physical contact with the scanner device, and tends to be relatively inexpensive [1].

Face recognition by computer is a computational technique to recognize someone in an image that is done automatically [3]. A person can be identified through certain features that make it different from other people's faces [13]. Based on its application, face recognition can be categorized into face identification and face verification. Face verification is when a computer is given a face pair, and the computer must be able to determine the faces in both images belong to the same person or not [14]. Face identification is the process carried out by a computer to find a face's identity in an image of a collection of faces whose identity has been stored in a database [10].

3.1. Stages of Face Recognition

Face recognition is carried out through four stages: the face detection stage, the face alignment stage, the face feature extraction stage, and the face matching/classification stage [3], [2]. The Diagram of the face recognition stages is shown in Fig. 5.

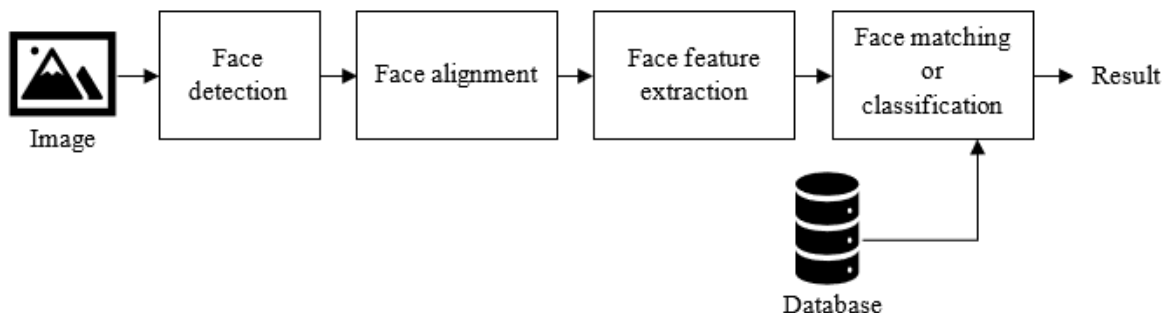


Figure 5. Diagram of stages in face recognition

3.1.1. Face Detection

Face detection is the primary step used for analyzing faces, such as aligning face, modeling a face, recognizing a face, recognizing facial expression, tracking facial poses, and recognizing a person's gender or age based on their face [15]. The purpose of the face detection algorithm is if an image is given to the computer, the face location will be displayed with a box marking the faces found in the image [16]. Fig. 6 shows an example of face detection in an image. The faces of people contained in the image were successfully detected, with the display of a colored box surrounding the face and the level of confidence. The faces are successfully detected even though the positions are different and not in a straight position to the camera. A face detector must have the ability to detect faces when the face is not in a straight position, varied lighting conditions, different expressions, different skin colors, obstacles, different face sizes in the images, low image resolution, and there are various other objects in an image [15], [16].

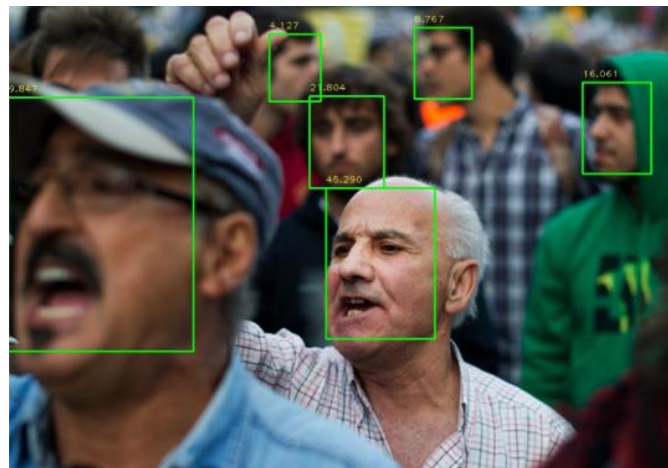


Figure 6. Example of face detection in an image [17]

Until today, there are several methods used to detect faces in an image [18]. Face detection in an image begins with the work done by Viola-Jones [18]. Viola-Jones applied the Haar-like features combined with multilevel classifiers that were trained using the AdaBoost learning algorithm to detect faces in an image quickly and accurately [19]. However, the weakness of the Viola-Jones method is the difficulty in detecting faces with a different point of view, blurred images, or partially covered faces [17], [19], [20]. Another method is the Deformable Part Model (DPM) proposed by Felzenszwalb et al. DPM model information contained between parts of the face [17]. The weakness of DPM is that it requires extensive computing resources [19].

Recent research shows that deep learning, using the Convolutional Neural Network (CNN), achieved success in computer vision. One of the successes is to detect faces. The reason for CNN's superiority compared to the previous methods is that CNN can automatically learn features that represent complex visual variations from a large amount of training data [21]. Several studies used CNN to detect the face. Li et al. used CNN cascade, which uses more than one CNN [22]. The use of six CNNs makes the computer able to detect faces more accurately. However, the study's weakness is that the training process is very complicated and heavy because they must train the six CNN separately [20]. Qin et al. proposed a joint training method to optimize the CNN cascade training process with the same goal, which is to detect faces [23]. Sawat and Hegadi proposed a method for detecting face by combining CNN and Cubic Support Vector Machine [24].

Face Alignment

The stage after detecting the face is to align the face. Face alignment is also known as detecting face landmarks [25]. Detecting face landmark is needed to make align to the front. It can increase the level of accuracy of face recognition. The main points of face landmark include the eyes, nose, mouth, and so on.

The shape or location gives a unique pattern to each face [26]. Fig. 7 shows the detection of face landmarks points. The face on the left shows 20 points of face landmarks, while on the right shows 68 points of face landmarks.

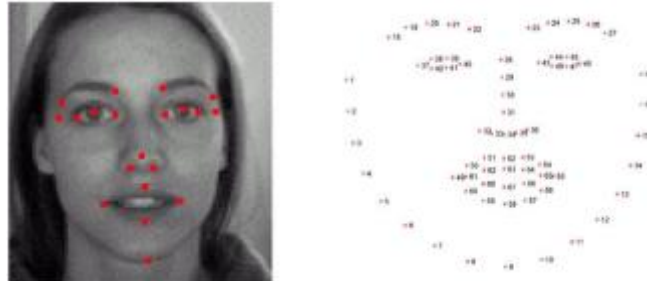


Figure 7. Detection of face landmarks [26]

3.1.2. Face Feature Extraction

The feature extraction stage is essential in face recognition [27]. As the name suggests, the primary purpose of this stage is to extract features of the face. Face feature extraction is taking and storing the most important information of the face. The information is in the form of a geometric distribution and the shape of the mouth, nose, eyes, or other features that make a face unique. Face features are represented in vectors that will be used for the next stage [10], [28]. There are several methods proposed to extract features of the face, including using Principal Component Analysis (PCA), Independent Component Analysis (ICA), Local Binary Pattern (LBP), Histogram, and the latest and better is using CNN [10].

3.1.3. Face Matching / Classification

This stage is the stage of comparing face features that have been extracted in the previous stage with face features contained in the database. Face recognition has two types of applications, namely, face identification and face verification [28]. Face verification is a one-to-one matching process. A test image will be compared with an image from the database to determine whether they are the same. Face identification is a one-to-many matching process. A test image is compared to a set of faces in the database to find the most likely match [10].

4. Face recognition based on Convolutional Neural Network

Various studies on face recognition have been carried out. This section will discuss several studies that focus on face recognition using the Convolutional Neural Network. The selected studies are studies published from 2015 to 2020.

Zafar et al. focused on making Bayesian DCNN for face recognition [2]. The researchers formed three CNN models with different depths and a Bayesian-DCNN. The use of different models is done to test the accuracy of each model. The dataset used in the study is the AT&T Face Database and the EURECOM Kinect Face Database. The accuracy obtained from the ordinary CNN models reached 94.2% when tested with the EUROCOM Kinect Face Database, and 97.5% with the AT&T Face Database. The test on the B-DCNN model shows that the accuracy achieved was 98.1% with the EUROCOM Kinect Face Database and 100% with the AT&T Face Database.

Peng et al. introduced a CNN called NIRFaceNet [5]. NIRFaceNet is a modification of GoogLeNet, used for face recognition whose input is in the form of Near-Infrared (NIR) images. The researchers chose to do face recognition with a NIR image because it has advantages over lighting changes. The dataset used is the CASIA NIR database. The advantage of this research is that the researchers added image variations to the dataset. These variations are motion blur, Gaussian blur, salt-and-pepper, and Gaussian noise. The addition of variations is done so that the model can recognize faces in unclear image conditions. The model

used is based on GoogLeNet, but only consists of 8 layers, different from the original, 27 layers. The reduction in layers provides the advantage of faster time needed for training. The test results show that NIRFaceNet obtained an accuracy rate of 100% to recognize faces without expression and normal position. Subsequent testing, with facial expressions and different facial positions, the accuracy level obtained, was 98.28%. Then, the testing with blurred and noise images achieved accuracy ranging from 96.02% to 98.48%.

Ben Fredj et al. trained a CNN to recognize faces in an uncontrolled environment, in the sense that face images have noise, or partially covered faces [29]. The researchers used the data augmentation method of flipping, histograms, noise, blur, differences in lighting, partially covered faces, and parts of cut-off faces. Data augmentation increases the number of images in the dataset and adds variation. The research in [29] has little in common with research in [5]. Softmax loss and center loss are used. The dataset used in the study is CASIA-WebFace. The accuracy obtained is 99.2% when tested with the LFW dataset, and 96.83% when tested with the YTF dataset.

Pei et al. made a student attendance system using face recognition based on deep learning [30]. Researchers revealed the problem faced was the difficulty in getting a large amount of training data. Data augmentation was used as a solution to increase the number of images in the dataset that can be used. The dataset images are modified through geometric transformation (enlarging the image, translation, rotation), lighting, mean filter, median filter, Gaussian filter, and bilateral filter. The dataset used is privately owned, with 3538 student face images for training, and 372 for testing. Researchers used the CNN VGG-16 architecture. The accuracy obtained from this study was 86.3%. Researchers then increased the amount of training data through face capture via video, and the level of accuracy increased to 98.1%.

Moon et al. trained CNN to recognize a person's face at different distances [31]. The used dataset is a dataset formed by the researchers themselves, with 12 individuals, and 270 images for each individual. Everyone's faces in the dataset are taken from 1-meter to 9 meters, with 30 images taken at each distance. The average level of accuracy obtained from this study was 88.9%.

Zheng et al. made a face recognition using the Deep Convolutional Neural Network and the Vector of Locally Aggregated Descriptor (VLAD) feature encoding [32]. The CNN was trained using the CASIA-WebFace dataset and tested using the IJB-A and JANUS CS2 datasets. Data augmentation was used to add to the dataset by turning the image in the dataset horizontally. The highest accuracy achieved in face verification testing is 97.90% using the IJB-A dataset, and 96.66% using the JANUS CS2 dataset. The highest level of accuracy achieved in face identification testing was 96.4% using the IJB-A dataset, and 96.90% using the JANUS CS2 dataset.

Hu et al. used 3 Convolutional Neural Networks arranged in parallel and proposed a Diversity Combination method for face recognition [33]. Inception-ResNet-v1 was used as the basic architecture of CNN. The three CNNs were used for feature extraction, and diversity combination is a strategy used to adaptively adjust the weight value in each CNN and make joint classification decisions. VGG2-Face, MS-Celeb-1M, and CAISA WebFace were used to train the CNN. Face matching testing was performed using CASIA NIR-VIS 2.0 and Oulu-CASIA NIR-VIS dataset. The accuracy level obtained was 98.9% using the CASIA NIR-VIS 2.0 and 99.8% using the Oulu-CASIA NIR-VIS dataset.

Binti Mat Kasim et al. used CNN to recognize faces [34]. The research focuses on celebrity face recognition. The author used three types of CNN architecture, which are ordinary CNN, AlexNet, and GoogLeNet. The purpose of the three architectures is to find out and compare the accuracy obtained from each architecture. The dataset used for training is the CelebFaces dataset. When tested, CNN could get 99.72% accuracy, while AlexNet and GoogLeNet managed to achieve an accuracy level of 100%.

Bendjillali et al., in their research, compared three types of architecture VGG16, ResNet50, and Inception-v3 [35]. The Viola-Jones algorithm is used to detect the face. The authors increased the contrast of the training image to determine the impact given to the accuracy of face recognition. The contrast

enhancement method used is Modified Contrast Limited Adaptive Histogram Equalization (M-CLAHE). Testing is done using Extended Yale B and CMU PIE datasets. The results obtained in testing using the Extended Yale B are the accuracy of face recognition with VGG16 is 97.28%, ResNet50 is 98.35%, and Inception-v3 is 99.44%. In testing using the CMU PIE, the accuracy of face recognition with VGG16 is 97.41%, ResNet50 is 99.53%, and Inception-v3 is 99.89%.

Nam et al. proposed a Pyramid-Based Scale-Invariant Convolutional Neural Network (PSI-CNN) model [36]. The method is used to overcome the problem of input image scale, for CNN to recognize faces on low-resolution images, therefore increasing the performance of CNN. Training is carried out using the CASIA WebFace database. Evaluation is done using the LFW database, which is a standard dataset for face recognition evaluation. Besides, the authors tested it with a private CCTV dataset. The results obtained are that the model has a face matching accuracy level of 98.87%.

Chandran et al. used CNN and Multi-Class Support Vector Machine (SVM) to recognize children's faces [37]. The architecture created is based on VGG-Face. The SVM Multiclass is used as a substitute for Softmax. The authors made the face database used for training and testing. The database contains 846 faces of children with 43 individuals. The test results show that the level of accuracy achieved is 99.41%.

Khan et al. used a Convolutional Neural Network to detect and recognize faces [38]. Region Proposal Network which is part of the R-CNN that is used for object detection, is used by researchers to detect faces in images. The dataset used for training is LFW, and the authors used the data augmentation method by reversing each image in the dataset. The accuracy achieved was 97.9%.

Hu et al. proposed a method for increasing the number of images in the dataset through image synthesis [4]. Two CNN architectures are used to test the resulting accuracy. The first architecture is named CNN-S, and the second architecture is named CNN-L. CNN-L has more layers than CNN-S. The dataset used for training is LFW and CASIA NIR-VIS 2.0. The test results show the highest accuracy obtained using CNN-L, which is 95.77% using the LFW dataset and 85.05% using the CASIA NIR-VIS 2.0 dataset.

Ding proposed a CNN model called Trunk-Branch Ensemble CNN (TBE -CNN) for video-based face recognition [39]. The model was designed to be able to extract additional information from the face holistically, and facial parts took around facial components. The TBE-CNN model is based on GoogLeNet. The researcher used the CASIA-WebFace database for training. Data augmentation such as the horizontal reversal and adding Gaussian noise was used. The test was carried out using the PaSC, COX Face, and YouTube Faces databases. Up to 98% accuracy was achieved using the PaSC database, 94.96% accuracy was achieved using the YouTube Faces database, and 99.33% accuracy was achieved using the COX Face.

Zangeneh et al. used two CNNs to map face images with a high and low resolution into 4096 common space dimensions using nonlinear transformations [40]. The model used is based on the VGGnet model. The two models are FECNN and SRFECNN. Model training uses the PubFig, UMIST, YALE B, AT&T, and FERET datasets. Testing is done using LFW, MBGC, and FERET face datasets. The accuracy obtained using the FERET dataset reached 92.1%; using LFW the accuracy reached 76.3%, and using the MBGC the accuracy reached 68.64%. The level of accuracy is based on the use of low-resolution images.

Yang et al. proposed a face matching method named SR-CNN [41]. The study combines Rotation-Invariant Texture Features (RITF), Scale-Invariant Feature Transform (SIFT), and Convolutional Neural Network (CNN). Testing and training are done using the LFW dataset. The results obtained reached 98.98%. The combination of RITF, SIFT, and CNN provides an increase in the level of accuracy of face matching.

Singh and Om used CNN to recognize the face of a newborn baby [42]. The CNN architecture used is very simple, consisting of 2 convolutional layers, two pooling layers, and a fully connected layer. The CNN was trained using the IIT (BHU) database. The database contains the faces of newborn babies with

different poses, such as normal facial conditions, crying, and falling asleep. The highest level of accuracy achieved was 91.03% in normal facial conditions.

Ríos-Sánchez et al. compared four CNN models, FaceNet, OpenFace, gb2s_Model1, and gb2s_Model2 [43]. All models are based on GoogLeNet, with the last two models being made by researchers. The four models are used to recognize faces with only one sample for each person. The goal is to determine the accuracy of face recognition when the amount of available data is very small. The LFW database is used to train the gb2s_Model1 and gb2s_Model2. The test was carried out using Extended Yale B, ORL, BioID, EUCFI, PrintAttack, gb2sMOD_Face_Dataset, gb2sTablet, gb2s_Selfies, and gb2s_IDCards database. The last three databases are private. The result shows the highest False Match Rate (FMR) and False Non-Match Rate (FNMR) is using OpenFace.

Zhou et al. used ResNet-face18, which is a CNN model for face recognition, and modification of ResNet [44]. The research introduces a Softmax function named double additive margin Softmax loss (DAM-Softmax), and they use CASIA-WebFace to train the model. CFP-FP, CALFW, and CPLFW datasets were used to test the model. Researchers compared three types of Softmax functions, namely Softmax, AM-Softmax, and DAM-Softmax. The highest accuracy level is obtained using DAM-Softmax. The accuracy achieved is 90.17%, 82.08%, and 93.26%, using CALFW, CPLFW, CFP-FP, respectively.

Another research used three different CNN models, namely MTCNN, self-designed CNN, and IFaceNet [45]. MTCNN is used to detect faces in an image. The self-designed CNN is used to determine whether the face in the image is fake or not. IFaceNet is used to recognize faces that are otherwise not fake. The IFaceNet is a modification of FaceNet. CNN was trained using a private dataset called NenuLD. CNN has an accuracy of 99.8% in determining whether a face is fake or not. The accuracy of face recognition obtained is also high, which is 99.7%. The researcher states that the total accuracy of the proposed system is 99.5%.

Son et al. used CNN in making face recognition systems for attendance via CCTV [46]. The models used are MTCNN, FaceNet, and ArcFace. MTCNN is used to detect faces. The researcher wants to compare the use of FaceNet and Arc Face to extract features. For the face classification process, the researchers tested four methods, namely Linear SVM, RBF-SVM, Gaussian Naïve Bayes, and Weighted-KNN. The use of W-KNN produces the highest level of accuracy compared to the three. The dataset used is the face of students taken manually by the author. The highest level of accuracy is achieved by combining ArcFace with W-KNN, which is 91.3% in the test dataset.

Li et al. in their research used Multi-CNN and Bayes probability to recognize faces [47]. Detection of facial components (eyes, eyebrows, nose, mouth) is done using the Active Shape Model (ASM). CNN is then used to extract features from each facial component. Bayes probability is then used to match faces. The dataset used for training and testing is private. The test results showed that the proposed method obtained an Equal Error Rate (EER) of 0.51, a False Positive Rate of 0.08, and a True Positive Rate of 0.92.

Song et al., in their research, proposed a CNN model for extracting features on the face [48]. The model is named IE-CNN. IE-CNN can extract unique internal features (eye, nose, and mouth) and external features (location of internal features on the head, chin, and ears) from the face. The training was carried out using the CASIA-WebFace database. LFW and CASIA-WebFace are used as test database. The results of Top1 and Top5 accuracy obtained in LFW are 98.86% and 99.20%, respectively. Furthermore, Top1 and Top5 accuracy obtained at CASIA-WebFace were 84.92% and 93.31%, respectively.

Ma et al. proposed a Local Binary Pattern (LBP) Guiding Pooling (G-RLBP) mechanism [49]. The mechanism can reduce the feature map's size and reduce the impact of noise in the image. The use of G-RLBP has advantages compared to the pooling layer that is generally used. The CNN architecture used is based on Alexnet, ZF-5net, and GoogLeNet. G-RLBP is used instead of the pooling layer while CASIA-WebFace is used as a training database. ORL and AR databases are used as test databases. The Nearest

Neighbor classifier is used to classify faces. The highest level of accuracy obtained by using Alexnet using ORL databases reached 92.74%, and ZF-5net reached 87.68%. The highest level of accuracy obtained in the AR database using Alexnet reaches 69.44%, and ZF-5net reaches 72.52%. GoogLeNet produces an accuracy rate of up to 93.54% using the ORL database, and 76.17% using the AR database.

Nimbarte and Bhoyar discussed one of the problems in face recognition, which is increasing one's age [50]. If a person's face stored in the database changes due to aging, it is worried that the face recognition system cannot recognize that person's face. Researchers proposed a CNN architecture to overcome this problem. CNN is used as a feature extractor, and SVM is used to classify faces. FGNET and MORPH (Album II) datasets are used to train and test the model. The accuracy level obtained using the FGNET dataset is 76.5%, and 92.5% using the MORPH (Album II) dataset.

Kamencay et al. compared the performance of CNN with Principal Component Analysis (PCA), Local Binary Patterns Histograms (LBPH), and K-Nearest Neighbor (K-NN) in face recognition [51]. The database used in the study is the ORL database. The highest face recognition accuracy was obtained using CNN, which is up to 98.3%. While the level of accuracy obtained using PCA, LBPH, and K-NN is 85.6%, 88.9%, and 81.4%, respectively. By using CNN, the level of accuracy obtained is very high.

Simón et al. proposed a method to improve face recognition [52]. The study introduces multimodal facial recognition using CNN. CNN is combined with Local Binary Patterns (LBP), Histograms of Gabor Ordinal Measures, Haar features, and Histograms of Oriented Gradients. The RGB-D-T database is used in training and testing. Data augmentation is used to reproduce training data. The results obtained are that by combining the old-fashioned features, CNN can improve facial recognition performance using the RGB-D-T database.

Zeng et al. proposed a method called the traditional and deep learning method (TDL) to recognize faces that only use a sample per individual in training (single sample per person or SSPP) [53]. In the research, the authors used a Convolutional Neural Network model for face recognition. The CNN model used is based on lightened CNN. Lightened CNN has been trained using the CASIA-WebFace database. Four databases were used in this study, namely AR face, Extend Yale B, FERET, and LFW face database. The accuracy level reached 100% using AR face, 88.3% using the Extend Yale B, 93.9% using the FERET database, and 74% using the LFW database.

Chen et al. [54] tested two CNN models, namely DCNNs and DCNNL. The DCNNL model is based on the AlexNet architecture, while the DCNNs is based on the architecture in the study by Chen et al. [55]. DCNN-based face detection was also used in this study. The model is the Deep Pyramid Deformable Parts Model for Face Detection (DP2MFD). The study used IJB-A and JANUS CS2 databases. The accuracy rate in the study reached 98.8% using the IJB-A dataset, and 98.6% using the JANUS CS2 dataset.

Chen et al., in their study, proposed a CNN with deep transformation learning [56]. The method increases the robustness and degree of discrimination of the extracted features. In the study, face detection and face marker detection were performed using MTCNN and Dlib. The datasets used for training are FaceScrub, cad2000, and CASIA-Web face. Tests were carried out by researchers using the LFW and IJB-A datasets. The test results show that the accuracy rate obtained using LFW is 99.16%, and identification using IJB-A is 93.1%.

Ling et al. proposed a method called Attention-Based Convolutional Neural Network (ACNN) that can distinguish facial features embedding [57]. The proposed method aims to increase the ability to distinguish possessed by Convolutional Neural Network such as ResNet-50 and ResNet-101. The research used ResNet as a basic architecture, specifically ResNet-50 and ResNet-101. Databases used for training include CASIA-WebFace and MS-Celeb-1M. Then the researchers used the Labeled Faces in the Wild (LFW) database, Age Database (AgeDB), and Celebrities in Frontal Profile (CFP) for validation. MegaFace database is used for testing. The accuracy of using a validation database is 99.83% using LFW,

98.57% using AgeDB, and 95.85% using CFP. The identification accuracy level results using MegaFace are 88.74% using the ACNN-Res50 and 98.35% using the ACNN-Res-101. Then the accuracy of verification is 91.12% using the ACNN-Res50 and 98.42% using the ACNN-Res101.

Khan et al. discussed making a student attendance recording system using face recognition [58]. The proposed system is expected to replace the manual and biometric attendance system, which takes a long time to record attendance. Researchers use YOLO v3 to detect faces and the Microsoft Azure face API to recognize detected faces. Researchers chose to use the YOLO algorithm because the processing is faster than R-CNN. A database containing students' faces is used to recognize faces. Twenty photos are taken from each student. The level of accuracy is obtained when the system is tested up to 100%. The result indicates that the proposed system can be used as a substitute for the manual recording system.

The work of Nakajima, Moshnyaga, and Hashimoto compared the performance of two facial recognition approaches [59]. The two approaches are CNN and Local-Binary Pattern Histograms. Both approaches were experimented on Raspberry-Pi and were trained using a private dataset consisting of 12 classes. Each class consists of 50 facial images. The number of images used for training is 540 images, and 60 images are used for verification. The CNN achieved the highest accuracy of 100% and average accuracy of 96%. In contrast to CNN, the highest accuracy achieved by the LBP is 76%, while the average accuracy is 64%. Therefore, the study concludes that the CNN achieved more robust recognition than the LBP.

Hussain et al. proposed an authentication system for the medical and healthcare area using face recognition [60]. The proposed system is comprised of face detection, extraction of facial features, and classification. The face detection was done using the Haar cascade technique. Three methods were compared in the facial features extraction: pre-trained ResNet-50, VGG-16, and the Linear Binary Pattern Histogram (LBPH). Lastly, the Support Vector Machine (SVM) was utilized for classification. A total of 8422 face images of 100 individuals were used, of which 70% were used for training, 15% for testing, and 15% for validation. The performance of each method was compared. The ResNet-50 + SVM could achieve an accuracy of 99.56%, VGG-16 + SVM achieved 98.49%, while the LBPH achieved an accuracy of 98.47%. Hence, the ResNet + SVM achieved the highest accuracy.

Farhi, Abbasi, and Rehman proposed a face recognition-based identity management system for office and academic environments [61]. The proposed system comprises face detection, facial features extraction, and classification. The work utilized MTCNN for face detection purposes and the well-known FaceNet to extract the facial features. Like the work of [60], the classification process was done using the Support Vector Machine (SVM). The authors experimented with different angles, distances, and illumination in their work. The proposed system could achieve 97.1% - 98.8% accuracy with face positioning from -15 to +15-degree angle. In normal light conditions and a distance of 4-5 meters, the proposed system achieved 98% to 99% accuracy. However, in low light conditions, the accuracy achieved is lower, 96.47%.

Xing, Wang, and Zheng proposed a VGG-16 with an improved pooling method for face recognition [62]. The work utilizes image size normalization and de-averaging operation to preprocess the image data. Furthermore, Gabor Wavelet Transform-based image enhancement was used to reduce noise in the image, while histogram homogenization was used to reduce the effect of light shadow in the image. The improved pooling method proposed in their work is the improved stochastic pooling, which enhances the generalization of the network and abstraction process. Face detection was carried out using the combination of Harr and Adaboost and the deep learning-based Faster R-CNN. The proposed method was experimented with using the self-built and LFW dataset. The resulting accuracy reached 97.2%.

Akter et al. proposed a framework to detect autistic children through facial recognition [63]. The facial recognition was done through improved transfer learning using pre-trained CNN models. The pre-trained CNNs used in their work are DenseNet121, ResNet50, VGG16, VGG19, MobileNet-V1, and MobileNet-V2. Besides using the pre-trained CNN, the work experimented with several machine learning

classifiers, such as Adaboost, k-Nearest Neighbor (kNN), Decision Tree, Logistic Regression, Gradient Boosting, Naïve Bayes, Support Vector Machine, Multi-layer Perception, and Random Forest. The pre-trained models were modified by adding three Batch Normalization layers and two fully connected layers before the output layer. The dataset used is facial images of normal and autistic children with 2936 images. The highest accuracy achieved was 91% on the test set using the improved MobileNet-V1. The authors used k-means clustering for classifying binary sub-types of autism and achieved 92.10% using the improved MobileNet-V1.

The work of Gwyn, Roy, and Atay compared the performance of popular deep learning network architectures for face recognition [64]. The architectures used in their work are AlexNet, Xception, Inception V2 and V3, ResNet50 and ResNet101, VGG16, and VGG19. Training those models was done using the LFW image dataset, with 4788 images of 423 individuals. Several dataset splitting strategies for training and testing were carried out, which are 80/20 split, 70/30 split, and 60/40 split. VGG-16 achieved the highest accuracy among other models, 84% on an 80/20 split, while AlexNet, Xception, Inception v2, Inception v3, ResNet50, ResNet101, and VGG19 achieved 61%, 52%, 68%, 67%, 71%, 72%, and 80%, respectively.

In the Covid-19 pandemic, the challenge to facial recognition emerges because people must wear face masks. Several works have tried to address this problem. Talahua et al. proposed a facial recognition system that could recognize a person even when using a face mask [65]. The system utilized OpenCv for facial detection, MobileNetV2 handles the face mask-wearing recognition. FaceNet handled the face recognition task as the facial features extractor and feedforward multilayer perceptron used for classification. A total of 13,359 images were used for training face recognition. The highest accuracy achieved is 99.52% for facial recognition with a mask and 99.96% without a mask.

Deng et al. proposed a facial recognition algorithm to recognize a masked face based on large margin cosine loss called MFCosface [66]. In their work, MTCNN was utilized for face detection, and the base architecture used is the Inception-ResNet-v1, and the proposed large margin cosine loss was used to train the model. VGGFace2_m was used to train the model, while CASIA-FaceV5_m, LFW_m, RMFD, and MFR2 were used for testing. Some faces from the datasets are generated with a mask. The accuracy achieved on LFW_m, CF_m, MFR2, RMFD is 99.33%, 97.03%, 98.50%, 92.15%, respectively.

Ullah et al. proposed a unified framework for mask detection and masked facial recognition [67]. The authors made a custom CNN called DeepMaskNet, which comprises 17 layers, including the input and classification layers. A largescale called masked detection and masked facial recognition (MDMFR) was used for training and testing the model. The DeepMaskNet achieved 100% accuracy in face mask detection and 93.33% in face recognition, outperforming other state-of-the-art models.

Song et al. proposed the Spartan Face Mask Detection and Facial Recognition system to address the challenge of mask detection, mask type classification, mask position classification, and identity recognition that emerged during the Covid-19 pandemic [68]. The proposed system utilized the MTCNN to detect faces in an image, FaceNet to extract embedded facial features, and Support Vector Machine and XGBoost as the classifiers for the facial recognition scope. The training and testing were done using a total of 2000 images. The accuracy achieved using the FaceNet + SVM is 97% on the test set and 100% on the training set, while the FaceNet + XGBoost is 88% on the test set and 100% on the training set.

Table 1. Summary of CNN-based Face Recognition Research

Research	CNN Model	Face Database	Accuracy
Zafar et al. [2]	B-DCNN	EUROCOM Kinect, AT&T	98.1% - 100%
Peng et al. [5]	GoogLeNet	CASIA NIR	96.02% - 100%
Ben Fredj et al. [29]	CNN	CASIA-WebFace, LFW, YTF	96.83% - 99.2%
Pei et al. [30]	VGG-16	Private	86.3% - 98.1%
Moon et al. [31]	CNN	Private	88.9%
Zheng et al. [32]	DCNN + VLAD	CASIA-WebFace, IJB-A,	96.4% - 97.90%

Research	CNN Model	Face Database	Accuracy
		JANUS CS2	
Hu et al. [33]	Inception-ResNet-v1 + Diversity Combination	VGG2-Face, MS-Celeb-1M, CAISA WebFace, CASIA NIR-VIS 2.0, Oulu-CASIA NIR-VIS	98.9% - 99.8%
Binti Mat Kasim et al. [34]	CNN, AlexNet, GoogLeNet	CelebFaces	99.72% - 100%
Bendjillali et al. [35]	VGG16, ResNet50, Inception-v3	Extended Yale B, CMU PIE	97.28% - 99.89%
Nam et al. [36]	PSI-CNN	CASIA WebFace, LFW, CCTV (private)	98.87%
Chandran et al. [37]	VGG-Face	Private	99.41%
Khan et al. [38]	CNN	LFW	97.9%
Hu et al. [4]	CNN	LFW, CASIA NIR-VIS 2.0	85.05% - 95.77%
Ding [39]	GoogLeNet	CASIA-WebFace, PaSC, COX Face, YTF	94.96% - 99.33%
Zangeneh et al. [40]	VGGnet	PubFig, UMIST, YALE B, AT&T, FERET, LFW, MBGC	68.64% - 92.1%
Yang et al. [41]	CNN	LFW	98.98%
Singh and Om [42]	CNN	IIT(BHU)	91.03%
Ríos-Sánchez et al. [43]	FaceNet, GoogLeNet	OpenFace, Extended Yale B, ORL, BioID, EUCFI, PrintAttack, gb2sµMOD_Face_Dataset, gb2sTablet, gb2s_Selfies, gb2s_IDCards	-
Zhou et al. [44]	ResNet-face18	CASIA-WebFace, CFP-FP, CALFW, CPLFW	82.08% - 93.26%
Liu et al. [45]	FaceNet	NenuLD (Private)	99.5%
Son et al. [46]	FaceNet, Arc Face	Private	91.3%
Li et al. [47]	Multi-CNN	Private	-
Song et al. [48]	IE-CNN	LFW, CASIA-WebFace	84.92% - 99.20%
Ma et al. [49]	Alexnet, ZF-5net, GoogLeNet	CASIA-WebFace, ORL, AR face	69.44% - 92.74%
Nimbarte and Bhoyar [50]	CNN	FGNET, MORPH (Album II)	76.5% - 92.5%
Kamencay et al. [51]	CNN	ORL	98,3%
Simón et al. [52]	CNN	RGB-D-T	-
Zeng et al. [53]	CNN	AR face, Extend Yale B, FERET, LFW	74% - 100%
Chen et al. [54]	CNN, AlexNet	IJB-A, JANUS CS2	98.6% - 98.8%
Chen et al [56]	CNN	FaceScrub, cad2000, CASIA-Webface, LFW, IJB-A	93.1% - 99.16%
Ling et al. [57]	ResNet	CASIA-WebFace, MS-Celeb-1M, LFW, AgeDB, CFP, MegaFace	88.74% - 98.35%
Khan et al. [58]	YOLO v3, Microsoft Azure Face API	Private	100%
Nakajima, Moshnyaga, and Hashimoto [59]	CNN and LBP	Private	CNN: 100% (avg. 96%) LBP: 76% (avg. 64%)
Hussain et al. [60]	ResNet-50 + SVM, VGG-16 + SVM, LBPH	Private	ResNet-50 + SVM: 99.56% VGG-16 + SVM: 98.49% LBPH: 98.47%
Farhi, Abbasi, and Rehman [61]	FaceNet + SVM	Private	96.47% - 99%
Xing, Wang, and Zheng [62]	VGG-16 with improved stochastic pooling	LFW and Self-Built	97.2%
Akter et al. [63]	Improved MobileNet-V1	Facial images of normal and	91%

Research	CNN Model	Face Database	Accuracy
		autistic children	
Gwyn, Roy, and Atay [64]	AlexNet, Xception, Inception v2, Inception v3, ResNet50, ResNet101, VGG16, VGG19	LFW	52% - 84%
Talahua et al. [65]	MobileNetV2 + feedforward MLP	Private	99.52% - 99.96%
Deng et al. [66]	Inception-ResNet-v1 with large margin cosine loss	VGGFace2_m, CASIA-FaceV5_m, LFW_m, RMFD, and MFR2	92.15% - 99.33%
Ullah et al. [67]	Deepmasknet (CNN)	MDMFR	93.33%
Song et al. [68]	FaceNet + SVM, FaceNet + XGBoost	Private	FaceNet + XGBoost: 88%, FaceNet + SVM: 97%

5. Face Recognition Database

Face database is important in face recognition. The face database contains face images that can be used to train the Convolutional Neural Network. There are a variety of face databases that are available online and can be downloaded by researchers or developers who need them. Table 2 lists the databases used in several studies to train and test facial recognition, along with their information.

Table 2. Face Recognition Databases

Name	Description
Labeled Faces in the Wild (LFW) [29]	13,233 images from 5,749 individuals
VGGFace2 [69]	3.31 million images from 9,131 individuals
MS-Celeb-1M [70]	10 million images from 100 thousand individuals
FaceScrub [71]	141,130 images from 695 individuals
FERET [72]	14,126 images from 1199 individuals
CASIA-WebFace [73]	494,414 images from 10,575 individuals
CMU Multi-PIE [72]	41,368 images from 68 individuals
YouTube Faces (YTF) [29]	3,425 videos from 1,595 individuals
ORL (AT&T) [2]	400 gray images with 40 individuals
EURECOM Kinect (KinectFaceDB) [74]	936 RGB images with 52 individuals
CASIA NIR	3,940 images from 197 individuals
IJB-A & JANUS CS2 [32]	5,712 images and 2,085 videos from 500 individuals
Oulu-CASIA NIR-VIS [33]	480 images from 80 individuals
CelebFaces [34]	200,000 images from 40 individuals
Extended Yale B [43]	16,128 images from 28 individuals
PaSC [39]	2,802 videos from 265 people
COX Face [39]	1,000 images and 3,000 videos from 1,000 individuals
PubFig [40]	58,797 images of 200 individuals
UMIST [40]	564 images from 20 individuals
YALE B [40]	5,760 images from 10 individuals
FERET [40]	14,126 images from 1,199 individuals
MBGC [40]	147 images from 147 individuals
IIT(BHU) [42]	2,100 images from 210 individuals
BioID [43]	1,521 images from 23 individuals
EUFCI [43]	7,900 images from 395 individuals
PrintAttack [43]	1,400 images from 38 individuals
AR Face [53]	4,000 images from 126 individuals
FGNET [50]	1,002 images from 82 individuals
MORPH Album II [50]	55,000 images from 13,000 individuals
Cad2000 [75]	160,000 images from 2,000 individuals
Celebrities in Frontal Profile (CFP) [57]	7,000 images from 500 individuals
AgeDB [57]	16,488 images from 568 individuals
MegaFace [57]	More than 1 million images from 690,000 individuals

The table above shows that the images used for training and testing are very large. The use of many images is done so that CNN can learn more about variations in human faces. Through learning about human faces' variations, CNN will understand the various patterns. Training by using many face images makes CNN have a high degree of accuracy.

6. Conclusion

This survey discusses face recognition based on the Convolutional Neural Network. Face recognition is one of the challenges in pattern recognition and computer vision and one of the many studies conducted in recent years. Several studies have been discussed trying to find renewal for face recognition. The renewed face recognition proposed various kinds of architecture used, modifying the images in the dataset or combining several methods. The main goal is to obtain a high level of accuracy so that the face recognition system has a high performance. This survey also discusses some face databases used in several studies. The face databases available have data from hundreds of images to millions. It is hoped that through this survey, readers can gain additional knowledge about face recognition based on Convolutional Neural Networks.

7. Acknowledgement

We would like to thank Universitas Universal for funding this research.

References

- [1] M. Chihaoui, A. Elkefi, W. Bellil, and C. Ben Amar, "A Survey of 2D Face Recognition Techniques," *Computers*, vol. 5, no. 4, p. 21, Sep. 2016.
- [2] U. Zafar et al., "Face recognition with Bayesian convolutional networks for robust surveillance systems," *EURASIP J. Image Video Process.*, vol. 2019, no. 1, p. 10, Dec. 2019.
- [3] Z. Lei and S. Z. Li, "Face Recognition Models: Computational Approaches," in *International Encyclopedia of the Social & Behavioral Sciences*, vol. 8, Elsevier, 2015, pp. 658–662.
- [4] G. Hu, X. Peng, Y. Yang, T. M. Hospedales, and J. Verbeek, "Frankenstein: Learning Deep Face Representations Using Small Data," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 293–303, 2018.
- [5] M. Peng, C. Wang, T. Chen, and G. Liu, "NIRFaceNet: A Convolutional Neural Network for Near-Infrared Face Identification," *Information*, vol. 7, no. 4, p. 61, Oct. 2016.
- [6] S. Pouyanfar et al., "A Survey on Deep Learning," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–36, Jan. 2019.
- [7] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks," *Proc. 2017 IEEE Int. Conf. Commun. Signal Process. ICCSP 2017*, vol. 2018-Janua, pp. 588–592, 2018.
- [8] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016.
- [9] R. Ranjan et al., "Deep Learning for Understanding Faces: Machines May Be Just as Good, or Better, than Humans," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 66–83, Jan. 2018.
- [10] S. Almabdy and L. Elrefaei, "Deep Convolutional Neural Network-Based Approaches for Face Recognition," *Appl. Sci.*, vol. 9, no. 20, p. 4397, Oct. 2019.
- [11] S. Zhou and S. Xiao, "3D face recognition: a survey," *Human-centric Comput. Inf. Sci.*, vol. 8, no. 1, 2018.
- [12] S. B. Ahmed, S. F. Ali, J. Ahmad, M. Adnan, and M. M. Fraz, "On the frontiers of pose invariant face recognition: a review," *Artif. Intell. Rev.*, Jul. 2019.
- [13] P. Kaur, K. Krishan, S. K. Sharma, and T. Kanchan, "Facial-recognition algorithms: A literature review," *Med. Sci. Law*, vol. 60, no. 2, pp. 131–139, Apr. 2020.
- [14] Q. Hua, C. Dong, and F. Zhang, "A Novel Approach to Face Verification Based on Second-Order Face-Pair Representation," *Complexity*, vol. 2018, pp. 1–10, Jun. 2018.
- [15] A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: a review," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 927–948, 2019.
- [16] R. Ranjan et al., "A Fast and Accurate System for Face Detection, Identification, and Verification," *IEEE Trans. Biometrics, Behav. Identity Sci.*, vol. 1, no. 2, pp. 82–96, Apr. 2019.

- [17] D. Triantafyllidou, P. Nousi, and A. Tefas, "Fast Deep Convolutional Face Detection in the Wild Exploiting Hard Sample Mining," *Big Data Res.*, vol. 11, pp. 65–76, Mar. 2018.
- [18] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: Past, present and future," *Comput. Vis. Image Underst.*, vol. 138, no. March, pp. 1–24, Sep. 2015.
- [19] D. Luo, G. Wen, D. Li, Y. Hu, and E. Huan, "Deep-learning-based face detection using iterative bounding-box regression," *Multimed. Tools Appl.*, vol. 77, no. 19, pp. 24663–24680, 2018.
- [20] W. Wu, Y. Yin, X. Wang, and D. Xu, "Face detection with different scales based on faster R-CNN," *IEEE Trans. Cybern.*, vol. 49, no. 11, pp. 4017–4028, 2019.
- [21] C. Peng, W. Bu, J. Xiao, K. Wong, and M. Yang, "An Improved Neural Network Cascade for Face Detection in Large Scene Surveillance," *Appl. Sci.*, vol. 8, no. 11, p. 2222, Nov. 2018.
- [22] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, vol. 07-12-June, pp. 5325–5334.
- [23] H. Qin, J. Yan, X. Li, and X. Hu, "Joint Training of Cascaded CNN for Face Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, vol. 2016-Decem, pp. 3456–3465.
- [24] D. D. Sawat and R. S. Hegadi, "Unconstrained face detection: a Deep learning and Machine learning combined approach," *CSI Trans. ICT*, vol. 5, no. 2, pp. 195–199, Jun. 2017.
- [25] X. Jin and X. Tan, "Face alignment in-the-wild: A Survey," *Comput. Vis. Image Underst.*, vol. 162, pp. 1–22, Sep. 2017.
- [26] Y. Wu and Q. Ji, "Facial Landmark Detection: A Literature Survey," *Int. J. Comput. Vis.*, vol. 127, no. 2, pp. 115–142, 2019.
- [27] H. Wang, J. Hu, and W. Deng, "Face Feature Extraction: A Complete Review," *IEEE Access*, vol. 6, no. c, pp. 6001–6039, 2018.
- [28] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face Recognition Systems: A Survey," *Sensors*, vol. 20, no. 2, p. 342, Jan. 2020.
- [29] H. Ben Fredj, S. Bouguezzi, and C. Souani, "Face recognition in unconstrained environment with CNN," *Vis. Comput.*, no. 0123456789, Jan. 2020.
- [30] Z. Pei, H. Xu, Y. Zhang, M. Guo, and Y.-H. Yang, "Face Recognition via Deep Learning Using Data Augmentation Based on Orthogonal Experiments," *Electronics*, vol. 8, no. 10, p. 1088, Sep. 2019.
- [31] H.-M. Moon, C. H. Seo, and S. B. Pan, "A face recognition system based on convolution neural network using multiple distance face," *Soft Comput.*, vol. 21, no. 17, pp. 4995–5002, Sep. 2017.
- [32] Z. Jingxiao, J. Chen, N. Bodla, V. M. Patel, and R. Chellappa, "VLAD encoded Deep Convolutional features for unconstrained face verification," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 4101–4106.
- [33] W. Hu, H. Hu, and X. Lu, "Heterogeneous Face Recognition Based on Multiple Deep Networks with Scatter Loss and Diversity Combination," *IEEE Access*, vol. 7, pp. 75305–75317, 2019.
- [34] N. A. Binti Mat Kasim, N. H. Binti Abd Rahman, Z. Ibrahim, and N. N. Abu Mangshor, "Celebrity Face Recognition using Deep Learning," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 2, p. 476, Nov. 2018.
- [35] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Illumination-robust face recognition based on deep convolutional neural networks architectures," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 2, p. 1015, May 2020.
- [36] G. P. Nam, H. Choi, J. Cho, and I. J. Kim, "PSI-CNN: A Pyramid-based scale-invariant cnn architecture for face recognition robust to various image resolutions," *Appl. Sci.*, vol. 8, no. 9, 2018.
- [37] P. S. Chandran, N. B. Byju, R. U. Deepak, K. N. Nishakumari, P. Devanand, and P. M. Sasi, "Missing child identification system using deep learning and multiclass SVM," *2018 IEEE Recent Adv. Intell. Comput. Syst. RAICS 2018*, pp. 113–116, 2019.
- [38] M. Z. Khan, S. Harous, S. U. Hassan, M. U. Ghani Khan, R. Iqbal, and S. Mumtaz, "Deep Unified Model for Face Recognition Based on Convolution Neural Network and Edge Computing," *IEEE Access*, vol. 7, pp. 72622–72633, 2019.
- [39] C. Ding and D. Tao, "Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 1002–1014, 2018.
- [40] E. Zangeneh, M. Rahmati, and Y. Mohsenzadeh, "Low resolution face recognition using a two-branch deep convolutional neural network architecture," *Expert Syst. Appl.*, vol. 139, p. 112854, 2020.

- [41] Y. X. Yang, C. Wen, K. Xie, F. Q. Wen, G. Q. Sheng, and X. G. Tang, "Face recognition using the SR-CNN model," *Sensors (Switzerland)*, vol. 18, no. 12, 2018.
- [42] R. Singh and H. Om, "Newborn face recognition using deep convolutional neural network," *Multimed. Tools Appl.*, vol. 76, no. 18, pp. 19005–19015, 2017.
- [43] B. Ríos-Sánchez, D. Costa-da-Silva, N. Martín-Yuste, and C. Sánchez-Ávila, "Deep Learning for Facial Recognition on Single Sample per Person Scenarios with Varied Capturing Conditions," *Appl. Sci.*, vol. 9, no. 24, p. 5474, Dec. 2019.
- [44] S. Zhou, C. Chen, G. Han, and X. Hou, "Double Additive Margin Softmax Loss for Face Recognition," *Appl. Sci.*, vol. 10, no. 1, p. 60, Dec. 2019.
- [45] S. Liu, Y. Song, M. Zhang, J. Zhao, S. Yang, and K. Hou, "An Identity Authentication Method Combining Liveness Detection and Face Recognition," *Sensors*, vol. 19, no. 21, p. 4733, Oct. 2019.
- [46] N. T. Son et al., "Implementing CCTV-Based Attendance Taking Support System Using Deep Face Recognition: A Case Study at FPT Polytechnic College," *Symmetry (Basel)*, vol. 12, no. 2, p. 307, Feb. 2020.
- [47] P. Li, J. Xie, W. Yan, Z. Li, and G. Kuang, "Living Face Verification via Multi-CNNs," *Int. J. Comput. Intell. Syst.*, vol. 12, no. 1, p. 183, 2018.
- [48] A. P. Song, Q. Hu, X. H. Ding, X. Y. Di, and Z. H. Song, "Similar Face Recognition Using the IE-CNN Model," *IEEE Access*, vol. 8, pp. 45244–45253, 2020.
- [49] Z. Ma, Y. Ding, B. Li, and X. Yuan, "Deep CNNs with robust LBP guiding pooling for face recognition," *Sensors (Switzerland)*, vol. 18, no. 11, pp. 1–18, 2018.
- [50] M. Nimbarte and K. Bhoyar, "Age Invariant Face Recognition using Convolutional Neural Network," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 4, p. 2126, Aug. 2018.
- [51] P. Kamencay, M. Benco, T. Mizdos, and R. Radil, "A new method for face recognition using convolutional neural network," *Adv. Electr. Electron. Eng.*, vol. 15, no. 4 Special Issue, pp. 663–672, 2017.
- [52] M. O. Simón et al., "Improved RGB-D-T based face recognition," *IET Biometrics*, vol. 5, no. 4, pp. 297–303, Dec. 2016.
- [53] J. Zeng, X. Zhao, J. Gan, C. Mai, Y. Zhai, and F. Wang, "Deep Convolutional Neural Network Used in Single Sample per Person Face Recognition," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–11, Aug. 2018.
- [54] J.-C. Chen et al., "Unconstrained Still/Video-Based Face Verification with Deep Convolutional Neural Networks," *Int. J. Comput. Vis.*, vol. 126, no. 2–4, pp. 272–291, Apr. 2018.
- [55] J.-C. Chen, V. M. Patel, and R. Chellappa, "Unconstrained face verification using deep CNN features," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–9.
- [56] G. Chen, Y. Shao, C. Tang, Z. Jin, and J. Zhang, "Deep transformation learning for face recognition in the unconstrained scene," *Mach. Vis. Appl.*, vol. 29, no. 3, pp. 513–523, Apr. 2018.
- [57] H. Ling, J. Wu, J. Huang, J. Chen, and P. Li, "Attention-based convolutional neural network for deep face recognition," *Multimed. Tools Appl.*, vol. 79, no. 9–10, pp. 5595–5616, Mar. 2020.
- [58] S. Khan, A. Akram, and N. Usman, "Real Time Automatic Attendance System for Face Recognition Using Face API and OpenCV," *Wirel. Pers. Commun.*, vol. 113, no. 1, pp. 469–480, 2020.
- [59] K. Nakajima, V. Moshnyaga, and K. Hashimoto, "A comparative study of conventional and CNN-based implementations of facial recognition on Raspberry-Pi," *SAMI 2021 - IEEE 19th World Symp. Appl. Mach. Intell. Informatics, Proc.*, no. 205010, pp. 217–221, 2021.
- [60] T. Hussain et al., "Internet of Things with Deep Learning-Based Face Recognition Approach for Authentication in Control Medical Systems," vol. 2022, 2022.
- [61] L. Farhi, H. Abbasi, and R. Rehman, "Smart Identity Management System by Face Detection Using Multitasking Convolution Network," *Secur. Commun. Networks*, vol. 2021, 2021.
- [62] C. Xing, J. S. Wang, and B. W. Zheng, "Hybrid Face Recognition Method Based on Gabor Wavelet Transform and VGG Convolutional Neural Network with Improved Pooling Strategy," *IAENG Int. J. Comput. Sci.*, vol. 48, no. 2, pp. 1–14, 2021.
- [63] T. Akter et al., "Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage," *Brain Sci.*, vol. 11, no. 6, 2021.
- [64] T. Gwyn, K. Roy, and M. Atay, "Face recognition using popular deep net architectures: A brief comparative study," *Futur. Internet*, vol. 13, no. 7, pp. 1–15, 2021.

- [65] J. S. Talahua, J. Buele, P. Calvopina, and J. Varela-Aldas, "Facial recognition system for people with and without face mask in times of the covid-19 pandemic," *Sustain.*, vol. 13, no. 12, pp. 1–19, 2021.
- [66] H. Deng, Z. Feng, G. Qian, X. Lv, H. Li, and G. Li, "MFCosface: a masked-face recognition algorithm based on large margin cosine loss," *Appl. Sci.*, vol. 11, no. 16, 2021.
- [67] N. Ullah, A. Javed, M. Ali Ghazanfar, A. Alsufyani, and S. Bourouis, "A novel DeepMaskNet model for face mask detection and masked facial recognition," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2022.
- [68] Z. Song, K. Nguyen, T. Nguyen, C. Cho, and J. Gao, "Spartan Face Mask Detection and Facial Recognition System," *Healthc.*, vol. 10, no. 1, 2022.
- [69] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A Dataset for Recognising Faces across Pose and Age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 67–74.
- [70] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9907 LNCS, 2016, pp. 87–102.
- [71] H.-W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 343–347.
- [72] S. S. Gangonda, P. P. Patavardhan, and K. J. Karande, "A Comprehensive Survey of Face Databases for Constrained and Unconstrained Environments," *Proc. - 2018 IEEE Glob. Conf. Wirel. Comput. Networking, GCWCN 2018*, pp. 173–177, 2019.
- [73] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning Face Representation from Scratch," Nov. 2014.
- [74] R. Min, N. Kose, and J.-L. Dugelay, "KinectFaceDB: A Kinect Database for Face Recognition," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 44, no. 11, pp. 1534–1548, Nov. 2014.
- [75] B. C. Chen, C. S. Chen, and W. H. Hsu, "Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset," *IEEE Trans. Multimed.*, vol. 17, no. 6, pp. 804–815, 2015.