Journal of Ambient Intelligence and Humanized Computing (2024) 15:373–387
https://doi.org/10.1007/s12652-022-03897-8

ORIGINAL RESEARCH

Deep learning-based face detection and recognition on drones

Mohsen Rostami[1] · Amirhamzeh Farajollahi[1] · Hashem Parvin[2]

Received: 15 January 2022 / Accepted: 3 May 2022 / Published online: 29 May 2022

Abstract

Unmanned aerial vehicles as known as drones, are aircraft that can comfortably search locations which are excessively dangerous or difficult for humans and take data from bird's-eye view. Enabling unmanned aerial vehicles to detect and recognize humans on the ground is essential for various applications, such as remote monitoring, people search, and surveillance. The current face detection and recognition models are able to detect or recognize faces on unmanned aerial vehicles using various limits in height, angle and distance, mainly where drones take images from high altitude or long distance. In the present paper, we proposed a novel face detection and recognition model on drones for improving the performance of face recognition when query images are taken from high altitudes or long distances that do not show much facial information of the humans. Moreover, we aim to employ deep neural network to perform these tasks and reach an enhanced top performance. Experimental evaluation of the proposed framework compared to state-of-the-art models over the DroneFace dataset demonstrates that our method can attain competitive accuracy on both the recognition and detection protocols.

Keywords Unmanned aerial vehicles · Face detection · Face recognition · Deep learning · Drones

1 Introduction

Drone or unmanned aerial vehicle (UAV) are flying machines without pilots which able to fly remotely (Mishra et al. 2021; Wang and Siddique 2020). Recently, drones have been utilized to detect and recognize humans and track them on the ground which are broadly applied in remote sensing, surveillance and photogrammetry (Wang and Siddique 2020). Face recognition is undoubtedly a capability for UAVs to recognize special humans within a crowd. To employ UAVs to seek missing children or elderlies, the UAV requires to realize who the targets are, and seek can proceed. So, face detection on UAV scan be an essential part; how well face detection or recognition accomplish on UAV is a worth research issue to be considered (Deeb et al. 2020; Yang et al. 2021). Face detection using UAV, due to the small size drone can fly in compact building blocks and easy to fly in the different climate and it has the good stability. The distance from UAV and their goals immediately influence the size of the face images in pixels. Because, UAV takes an image from the air, height and distance of UAVs keep them remote from their goals on the ground. In addition, speed and angles influence the quality of the face images and reduce the accuracy of face detection and recognition (Kalra et al. 2019; Bonetto et al. 2015; Cao et al. 2021; Lv et al. 2021).

Many face recognition and detection algorithms have been presented in the past decades (Kumar et al. 2019; Bae and Kim 2005). For instance, Yang et al. (2002) introduced several face recognition and detection models that utilize direct data of facial images which are speedy. However, they are usually less accurate concerning background interference and huge size face variance. Li et al. (2004)introduced a novel model to address multiple views face detection problems, e.g., self-shadowing, self-occlusion, rotation in-depth or nonlinear variation, and frontal view. In Yuan (2020), to reduce detection and recognition (DAR) performance due to face occlusion and increase the precision of face recognition, an attention algorithm supervision framework is presented which indicates the visible region of the occluded face. Recently, face detection has advanced highly due to its beginning, though tiny model has been proposed on how this technology would employ in UAVs. Nowadays, deep

✉ Amirhamzeh Farajollahi
a.farajollahi@sharif.edu

[1] Department of Engineering, University of Imam Ali, Tehran, Iran

[2] Department of Artificial Intelligence, Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran

learning (DL) has been utilized broadly in machine learning (ML) scenarios, and many models to make face detection or recognition protocols exist, for example, the Eigenfaces algorithm and neural networks models (Bhattacharyya 2011; Hjelmås and Low 2001). Inspired by the Bonetto et al. (2015), some face DAR models (Hsu and Chen 2015; Sarath et al. 2019; Jurevičius et al. 2019; Herrera and Imamura 2019) were introduced to evaluate their performances in attaining favorable functionality on UAVs. The evaluations of this method demonstrated that the frameworks could detect or identify faces where taken in ideal conditions at high performance. Besides, if the requirements of the photographs are not perfect, the accuracy of DAR drastically falls. Therefore, when UAVs with face detection functionality is utilized in surveillance, they will be needed to serve in poor weather conditions, in much higher heights, and with large distance available to them. Because, UAV scan flies outdoor or in-door over the various environment or illumination conditions and may take images with any combination of the angle of depression, altitude, and distance. The effects caused by heights and distances from UAV to the subjects are considered to systematically analyze the limits of the proposed face DAR models when employed on UAVs.

In the present work, we aim to tackle the drawbacks of the current face DAR models while they are employed on UAVs, and design appropriate architecture to utilize face DAR in UAV-based applications. We propose an efficient DL-based framework inspired by the single-shot multi-box detector (SSD) model (Liu et al. 2016), two architectures based on the CNN for image face detection and recognition are proposed that enhance the accuracy and speed of the search. Our model is built utilizing two stages. The first stage is an SDD-based face localization process so that the face images are extracted, and the second stage is DL-based face classification which is able to increase the accuracy of face recognition. The primary highlights of this study are as follows:

- We present a DL model for face detection in drone images. Different from the previous method (Davis et al. 2013; Wang and Siddique 2020; Deeb et al. 2020). Our face detection model is regression-based face detection which is more efficient for multi-scale face detection on UAV.
- In our detection network, Mobile-Net is employed as the backbone model to decrease the number of parameters and computation, which makes the model more capable. Moreover, seven feature maps and two feature fusion units are utilized in our framework to improve detection of small faces, multi-scale faces and enhance the detection accuracy.
- We propose a DL architecture for face recognition in drone images by employing the pre-trained CNN to extract features and softmax layer for classification. Moreover, CNN networks have been utilized for extracting abstract image features and employing them in the recognition step. Unlike previous models (Hsu and Chen 2015; Saha et al. 2018; Davis et al. 2013; Luo et al. 2020; Li et al. 2021), which have many false detections, our model is more efficient under different heights and distance conditions on drones.
- It should be noted that our model utilizes the appropriate optimization algorithm and loss function in its process, which leads to a decrease in the number of iterations needed to obtain stable results. As a result, this improvement indicates the capability of our model to satisfy industrial requirements.
- Experimental evaluations indicate that compared with the state-of-the-art drone-based methods, our work demonstrates higher accuracy and solves low DAR rate over the DroneFace dataset, compared to previous leading performance.

2 Related works

2.1 Pattern recognition

Pattern recognition (PR) is the task of recognizing patterns by utilizing a ML method. PR can be defined as the clustering or classification of data based on knowledge generated from patterns or/and their representation.PR has a variety of applications, including speech recognition, aerial photo interpretation, image processing, and medical imaging. Some modern models to PR include utilize of ML for clustering or classification of data. For example, Li et al. (2018a, b), presented a constrained spectral clustering utilizing flexible embedding model. Moreover, a flexible probabilistic neighborhood technique is used to create the similarity matrix of a target graph. Li et al. (2018a, b), proposed a novel model to recover a robust similarity graph utilizing multiple features to generate optimal weights for each feature. Furthermore, they designed a novel model to determine a set of optimal affinity matrixes, one for each dimension that determines the lower-dimensional space. Li et al. (2019a, b) introduced a novel zero-shot event detection model which employs the semantic similarity between concepts and events. This model focuses on the most related concepts for the zero-shot event detection and learns the semantic similarity from the vocabulary. In Luo et al. (2018), a new semi-supervised feature selection framework is introduced that the samples with similar classes have a high similarity of being neighbors. The authors Zhang et al. (2020), introduced two DL-based models with new spatio-temporal preserving representations to accurately recognize human intentions. This model contains both RNN and CNN architectures effectively employing the preserved temporal and spatial data for human intention recognition. In Chen et al. (2020), a semi-supervised DL method is proposed for imbalanced activity recognition using multi-modal sensory data. They aim to consider the challenges of multi-modal sensor data and limited labeled data along with the imbalance problems jointly.

In Yan et al. (2020), a new self-weighted robust model with *l21*-norm based pairwise between-class distance criterion is developed for multi-label classification mainly with edge labels. Moreover, anew re-weighted method is used to determine the global optimum of the challenging *l21*-norm maximization issue. In Chang et al. (2016), a new rank projection framework for bi-linear analysis is proposed which reduces the computation complexity. This method uses multiple rank projection methods to provide a larger search space in which the optimal point can be found. Most related works use the ML strategy which needs labeled training data which reduces their scalability to real-world scenarios where major unlabeled data are available. To address this problem, in Zhou et al. (2020), a multi-feature fusion with similarity learning method is proposed which aims to communicate general assessment on the graph structure by using special data of feature descriptors.

2.2 Traditional face DAR

Face DAR is a technology that is used to identify a person from a video or image source (Zhu and Jiang 2020; Suri et al. 2021). The scenario of face detection and recognition has an excellent scientific basis that the main idea focus on face DAR dates back to the 1990s. After that, these systems are being optimized and improved continually and these technologies become broadly utilized in people's daily life (Liu and Chen 2021; Cheng et al. 2019). It has been used increasingly for forensics, military professionals and mobile security. Face detection and recognition include solutions to some complex applications, such as training support vector machine (SVM) for face recognition, face DAR in a complex background, and convolutional neural network (CNN)-based face DAR model (Yang et al. 2017). A systematic framework to understanding the models for face DAR is proposed in Hjelmås and Low (2001). The overview of previous works indicates an understandable perspective on the proposed models, such as neural networks methods, statistical models, feature-based methods and linear subspace approaches (Wang et al. 2021; Iqbal et al. 2019). The face DAR model can be applied in stationary cameras; UAV provides an advantage of flying at low heights at face level and mobility. Face DAR is accurate when cameras take the face images from zero degrees of height, direct angle, and a low distance, while the pictures that would be taken using UAV would often be from extremely various conditions (Hsu and Chen 2015). Therefore, the potential of using drones as a tool for face detection and recognition in surveillance to improve security measures is evident. Some factors reduce the accuracy of face DAR models on drones. Previousf ace recognition models in UAV scan classify faces, but with a number of drawbacks in distance, heights and with a large depression angle.

In early development, Gao and Lu (2008) presented a novel model to speed up the Haar-classifier-based face detection method. With parallel structure, this model attained real-time face detection accuracy which has suitable performance and resources. Korshunov and Ooi (2011) considered the critical image modality face DAR which reduces the limits of face recognition under strict environment on UAVs that it applied in rescue missions, and robot competition. Matai et al. (2011)designed face detection using the AdaBoost model that Haar features have been employed in it (Davis et al. 2013). Developed local binary pattern (LBP) model to utilize face DAR onto a commercial drone for security application. Their system was economical and can be broadly used; however, they do not evaluate the efficacy rate and limits of the proposed method. Kumar et al. (2014) presented the design and validation of a face detector model utilizing techniques of discrete cosine transform and principal component analysis. Bold et al. (2016) proposed a model that is able to recognize faces utilizing accelerated robust feature extraction based on the Eigen-face recognizer on AR drone 2.0. This model uses various models, such as SURF, Haar-Cascade (HC) classifier, modified Eigen-face for face classification. The evaluation of this model was performed in an indoor environment to test its accuracy in different conditions of angle, distance, and height. HC is also employed to make a smart parking application (Meduri and Telles 2018). Saha et al. (2018) presented a novel model about advancements in UAVs to recognize persons using drone cameras. This model is designed using high-tech specification which is more stable and a noise reduction feature is employed. It is appropriate for military operation and surveillance ideas. Sarath et al. (2019) provided an additional feature, to the global authorities to seek for people that are blacklisted which can be utilized for military and local surveillance; the model is well stabled in its application. However, this method has specific limits with future scope to make effective improvements. Wang and Siddique (2020) proposed a face recognition model utilizing the LBP face recognizer framework installed on UAV technology. Drone-based strategy in this model can be used in a surveillance drone that can cover more area, unlike the stationary model. Atmaja et al. (2021) implemented HC-based technique for face DAR on drones that the face DAR process is performed in real-time. This method is used indoors at a distance of 50 m, while the angle of the face is highly effective. Atmaja et al. (2021) employed a spy UAV with a face localization algorithm using Eigen-face information in testing and training facial images. The overall structure of this model involves feature extraction, face detection and classification. The focus of the model is to realize the performance of the Eigen-faces algorithm in face localization.

2.3 DL-based face DAR

Face DAR model can be developed by utilizing DL algorithms (Wang et al. 2019; Zhu and Jiang 2020; Yang et al. 2002). A face DAR algorithm built-in DL employed in drones can be improved to detect criminals and raise security (Bhattacharyya 2011) DL aims to make the high-level information abstraction by utilizing neural network architecture constructed of multiple non-linear/linear transformations, especially the CNN, which indicates significant advantages. For instance, Lin et al. (1997) proposed a face recognition model by using the probabilistic decision-based deep network. Nair and Cavallaro (2009) presented a robust and accurate method for segmenting and detecting faces, detecting landmarks, and attaining appropriate registration of face patches using the fitting of face information. This method used a 3D point distribution algorithm without depending on pose, texture or orientation data. Yang et al. (2017) proposed a DL-based framework for face detection exploiting on face attribute-based supervision that component detectors in CNN learned to recognize attributes from facial images, without explicit guidance. Kim et al. (2019) presented a model of people localization by creating an environment similar to a UAV utilizing a large angle camera in various places to develop training images for individual detection. Tiny Yolo (Fang et al. 2019), which is a kind of object detection framework, was employed as a recognition algorithm. Almabdy and Elrefaei (2019) applied pre-trained CNNframeworksto improve face DAR performance by analyzing accuracy utilizing the pre-trained Alex-Net to extract features, followed by SVM, and then utilizing transfer learning based on CNN for both classification and feature extraction. Li et al. (2019a, b)proposed an attention-based feature agglomeration model to extract a feature pyramid by semantic data for multi-scaling face localization. Moreover, high-level representations are immediately combined into low-level features using skip connection. Deeb et al. (2020) attempted to use DL frameworks to perform its tasks and achieve enhanced top performance in face detection. They compared various DL methods; when evaluating them in some appropriate face recognition tasks, utilizing the DroneFace dataset to illustrate that they can be employed for attaining excellent UAV face recognition performance. Specifically, they evaluated how the height at that the images were taken, utilizing the UAV, influenced the method's performance at localizing faces. In summary, our approach belongs to the DL models; unlike the techniques introduced above, our method can learn the relationship between labels and images by using face DAR networks to recognize facial images collaboratively.

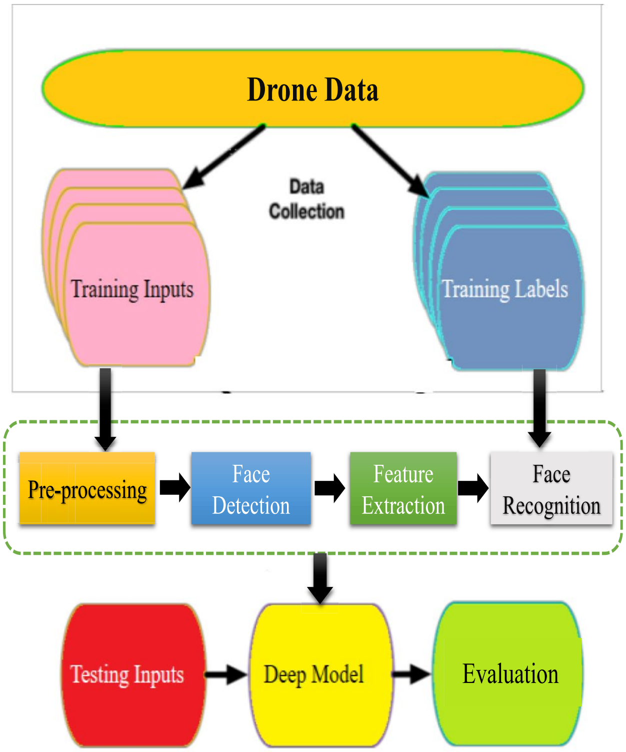
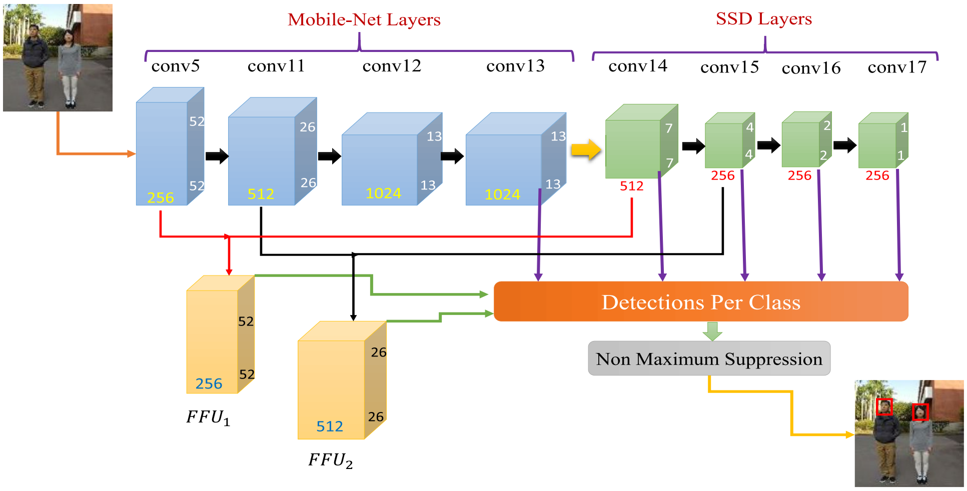


Fig. 1 The flow graph of our framework

3 Proposed model

In this section, we present a face DAR framework using DL architecture. Our framework contains two parts which are the detection and recognition parts. Essentially, the detection part detects the faces in an image and extracts the faces as an image by using our detection network. Then, the recognition part recognizes the person based on our recognition network. Then the training modules train the system utilizing the DL model. As shown in Fig. 1, the structure of our method is broken down into four stages include pre-processing, face detection, feature extraction and face recognition. The input for the model is of an image and the output is the recognized face of the person. The system starts image pre-processing. After that, the images are processed, we resize each image to an appropriate size for the CNN models. In face detection, we localize faces by employing a deep CNN model and use them in face recognition. In the features extraction step, deep features are extracted from the generated face images to employ in the next step. Finally, the process of classifying

Fig. 2 The overall structure of the face detection framework



faces occurred with CNN which involves recognizing a face from the facial images.

3.1 Pre-processing

Image Pre-processing (IP)is a crucial step and most of the models spend a reasonable amount of time in image pre-processing before building the model. IP performs the operations on the image at the lowest level of abstraction. The aim of IP is an enhancement of the image that decreases undesired distortion and improves many image features for other processing. Our model starts IP and resizes each image to an appropriate size for our CNN model and employssome operations, such as histogram equalization, super-resolution and pixel brightness transformations over the input images by using existing methods (Bhattacharyya 2011).

3.2 Face detection

The aim of this stage is to localize the individual faces in a query image. The individual face consists of many unique features that can be utilized by a CNN method to detect facial images. We use the DL algorithm that uses the CNN structure to achieve high performance in face detection.

3.2.1 Model architecture

Inspired by the in-depth analysis of the challenges and characteristics of face detection in drone data, based on the SSD framework, we proposed a novel framework shown in Fig. 2. Our face detection model is regression-based face detection which is more efficient for multi-scale face localization in drone images. Different from the SSD model, our backbone structure utilizes the Mobile-Net (Sinha and El-Sharkawy 2019) instead of VGG-Net. Our light weight Mobile-Net utilizes depth-wise separable convolution (DWSC) to effectually decrease the number of parameters and computations of the network, which is useful to face localization in embedded scenarios, mainly in a UAV application environment. By decreasing the framework's computational complexity and parameters, our model framework was more efficient in face detection. Due to layer-by-layer pooling, the high-level features miss details, and it is usually appropriate for multi-scale face localization. The important features were converted to the lower layer features with richer details and higher resolution which combines to enhance the effect of small face detection. Hence, in the detection network, the low-level feature vectorConv5 (52 × 52 × 256) was attached, and the Conv 14 and Conv 15 were respectively up sampled and combined with Conv 5 and Conv 11 layers to increase the accuracy of small face detection. Figure 2 illustrates the structure of our model. The Fusion layer1, Fusion layer 2, Conv17, Conv16, Conv15, Conv14, and Conv13 were employed to predict both confidences and locations. In this framework, the dimension of the input images and the set of bounding boxes for each class of face changes the performance of the localization. Inspired by the SSD network, we regulated the bounding boxes and resized the input images from 300 × 300 to 416 × 416, to enhance the detection accuracy of our model. Besides, the non-maximum suppression (NMS) method is used to extract the redundant face outputs. These face detection models consist of several essential computation layers, including the convolutional, pooling, fully connected (FC), activation function, element-wise sum, normalization, de convolutional (DCL), and softmax layers. Computation of each layer in the DL architecture and optimization are present in the next section.

3.2.2 Feature fusion unit (FFU)

There are two FFU in the detection network as illustrated in Fig. 2 and its architecture is shown in Fig. 3. The high-level

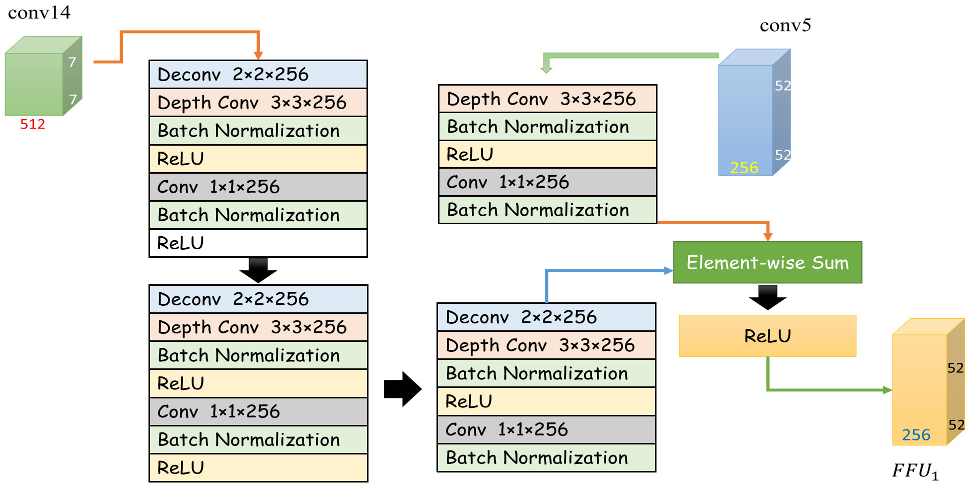


Fig. 3 The architecture of FFU

feature vectors are learned by DCL to the same dimension as the low-level feature vectors. These features are combined using an element-wise sum. Taking the FFU 1 as an instance, for fusing the feature vectors of the Conv 5 and Conv14, it is required to up sample the dimension of the Conv14 layer by eight times. Thus, for Conv14, we created three DCLs with stride 2 to attain up-sampling. Because the feature vectors in the FFU are mathematically costly, we employed depth-DWSCs to decrease computational complexity and parameters. The DCL was followed by DWSC. The DWSC contains 3×3 batch normalization (BN), 1×1 point-wise convolutional layer, depth-wise convolutional layer, and rectified linear unit (ReLU) layer. After the BN layer, we combined them using element-wise sum, and finally passed the ReLU to perform the fusion. The FFU 2 utilized the same computation strategy, and only the channels were regulated. Only little changes were needed for methods with various input dimensions.

3.2.3 Convolutional and deconvolutional layers

The convolutional layer contains several trainable parameters. Each kernel (filter) is small in height and width that they expanded using the entire depth of the input data. After an image is forwarded into a network, each kernel is convolved over the height and width of the input, and the dot operation is calculated between the entries of the image and the entries of the kernel. In other words, the convolutional layer contains several filters, which are utilized to generate a set of features from the image feature vectors. The dimension of input feature vectors F_i is expressed as $W \times H \times C$; the fitters are defined as $K_x \times K_y \times C \times Y$ (Y is the set of filters, which is equal to the set of output feature vectors). The output neuron N at position (i, j) of output feature map F_o is calculated as:

$$N^{F_i,F_o}_{i,j} = \sum_{z=0}^{C-1} \sum_{y=0}^{K_y-1} \sum_{x=0}^{K_x-1} W^{F_i,F_o}_{x,y} * N^{F_i,F_o}_{i*S_i+x,j*S_j+y} + b^{F_i,F_o} \qquad (1)$$

where b and W denote the bias parameters and filters between F_i and F_o respectively, and S_i and S_j represent the sliding steps where the input is convoluted in the x and y directions. Besides, DCL in face localization usually refers to dilated convolution or transposed convolution, which is employed to up sample the output of the convolution layer back to the dimension of the input image. The computations of DCL are similar to the convolution layer that its basic computation is also composed of addition and multiplication.

3.2.4 Pooling layer

This layer is called the down-sampling layer, which decreases the size of input onto feature vectors utilizing averaging or maximizing the output in each pooling window. The computation of pooling not only keeps the main features, but also reduces the computation of the model which effectually decreases the risk of over-fitting of DL networks. Average pooling and max-pooling are two popular utilized pooling techniques. In the pooling computation of the two-dim input feature vector, the pooling window is represented as $P_x \times P_y$, and the F_i and F_o represent one-to-one correspondence. The max-pooling equation for the output neuron N at position (i, j) is:

$$N^{F_o}_{i,j} = \max_{0 \le x \le P_x - 1, 0 \le y \le P_y - 1} N^{F_i}_{i+x,j+y} \qquad (2)$$

in which equation (2) is done by consecutively comparing the maximum outputs of neurons in the pooling window.

3.2.5 Non-linearity activation layer

The nonlinear activation layer is broadly employed in DL, which makes the CNN have nonlinear learning to improve the ability of the model to learn the high-level features. The nonlinear layer accomplishes a nonlinear transformation of the input to learn complex relations. There are various nonlinear activation functions, such as Tanh, Relu, Sigmoid, and Softmax. The ReLu is the most utilized because it can to reduce over-fitting and is less sensitive to gradient loss effectually. Its computation equation is:

$$Relu(D) = \begin{cases} 0, x \leq 0 \\ D, x > 0 \end{cases} \tag{3}$$

where D is the input of the function, the ReLu equation is nonlinear for negative inputs as it outputs all negative values as zero and is linear for all positive inputs.

3.2.6 Normalization layer

The normalization layer in the DL is to overcome the issue that the data distribution in the middle layer changes during the learning iterations to control the vanishing and exploding gradient and speed up the training (Du et al. 2015). The BN proposed by Ioffe and Szegedy (2015), is broadly employed in the DL network, which effectually enhances the speed of convergence and the training. The BN is computed as a separate layer in the DL process that its equation for the neuron N at position (i,j) is as:

$$N_{i,j}^{F_o} = \frac{\left(N_{i,j}^{F_i} - \frac{ME_{i,j}^{F_i}}{SF}\right)}{\sqrt{\frac{Var_{i,j}^{F_i}}{SF} + \varepsilon}} \tag{4}$$

in which SF is the scaling factor and Var and ME are the variance and mean of the F_i respectively. These parameters are trained using training data. The ε is a constant which is taken as 0.0001.

3.2.7 Element-wise sum layer

The computation of this layer is accomplished when FFU combines feature vectors of the same size generated on various paths. The major computation of this layer is the adding of the neuron at the corresponding position of F_i.

3.2.8 Full connection layer

This layer is utilized to fuse the extracted features from previous layers, which is at the end of the network and employed as a classifier. However, this layer is utilized in current facial localization models, in order not to reduce generality. The F_i is a vector of $1 \times 1 \times C$, and the F_i is a vector of $1 \times 1 \times Y$. The output neuron N at position (i,j) of F_o is calculated as:

$$N_{i,j}^{F_o} = \sum_{F_i=0}^{C-1} W_{i,j}^{F_i,F_o} * N_{i,j}^{F_i} + b^{F_i,F_o} \tag{5}$$

where b and W denote the bias parameters and filters between F_i and F_o respectively. The computation of this layer similar to the convolution layer is composed of addition and multiplication.

3.2.9 Softmax layer

This layer is utilized for the multi-class classification in CNN, which connecting the A-dim vector B into an A-dim vector S inrange (0, 1). The computation of this layer is:

$$S_i = \frac{e^{B_i}}{\sum_{j=1}^{A} e^{B_j}} (i = 1, 2, 3, \ldots, A) \tag{6}$$

where B and S are both A-dimensional vectors.

3.2.10 Training

During training, we utilized a similar method as SSD that several default boxes were compared to the Ground-Truth (GT) boxes. For each GT box, we matched it to the default box using the Jaccard overlap higher than a threshold (e.g., 0.6). This was favorable to predict several bounding boxes for overlapped faces with high confidence. We chose the non-matched default boxes with top loss value as the negative instances when the ratio of positive and negative was 3:1. We have taken z to be an indicator to match the default box to the GT box, which is 0 or 1. The p, g, and c denotethe predicted box, GT box and confidences, respectively. The loss function for our detection network is combining the confidence loss L_{conf} and the detection loss L_{loc}, as:

$$L(z, c, p, g) = \frac{1}{U}\left(\beta * L_{loc}(z, p, g) + L_{conf}(z, c)\right) \tag{7}$$

where the weight term β is adjusted to 1 using cross-validation and U is the set of matched default boxes. The L_{loc} is a smooth L_1 loss between the GT box g and the predicted box p parameters. The L_{conf} is the softmax loss across several classes' confidences c. The L_{loc} and the L_{conf} are equal to SSD. Several feature vectors were utilized to predict both confidence and location in our framework. We employed the data augmentation technique to improve the robustness of the model in various input face shapes and sizes. These models included expansion augmentation, photometric

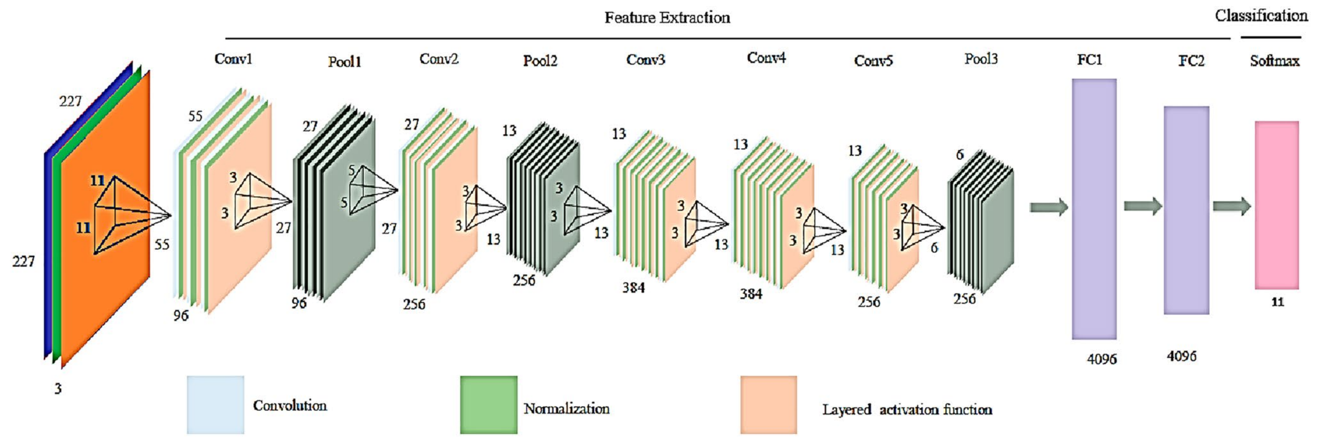


Fig. 4 The overall diagram of the recognition architecture

distortion, random cropping and flipping which are more appropriate to detect small faces.

3.3 Face recognition

After face detection, face recognition is the final step. In face recognition, the model seeks for the query facial image to tell who she/he is the extracted face batches are needed to achieve an automatic recognition. First, the faces are localized, then run the faces recognition algorithm using the proposed recognition framework shown in Fig. 4.

3.3.1 Feature extraction

In our recognition model presented in Fig. 4, the feature extraction part is first pretrained utilizing great data, and then the pre-trained network (softmax layer is dropped) is employed as a feature extractor for extracting face features. Due to pre-train the proposed model with large images, we utilized the ATT dataset (Samaria and Harter 1994) which was utilized alongside the DroneFace dataset, which was captured utilizing the ARDrone1.0. The 4096-dim output of the last FC part is employed as the Eigen-vector. The output of the FC part is determined by the following formula:

$$O^{l+1} = \varphi\left(W^{l+1}O^{l} + b^{l+1}\right) \tag{8}$$

where O^{l} and O^{l+1} are the output vector of both lth and $(l+1)$th layers respectively, W^{l+1} is a weight of the linear coefficients, b^{l+1} is the bias vector, and $\varphi(\cdot)$ is the nonlinear function.

3.3.2 Recognition architecture

The diagram of the recognition network is indicated in Fig. 4. The input of our model is facial images that have been scale normalized. The input image is convoluted using some convolution filters for extracting the image features. The kernel coefficients are acquired during the training which is related to the features of the training images. After that, the convolutional operation outputs are normalized by the BN layer, which reduces the over fitting of the network and provides efficient gradient descent. Due to an increase the nonlinear ability of our network, the normalized image is forward in the activation layer. Finally, the output of the activation layer is pooled, keeping the appropriate features and enhancing the distortion tolerance of the model. After some pooled layers and convolutional layers, the obtained useful features are employed by the FC layer to acquire CNN feature representation outputs. The pre-training of the network includes the forward and backpropagation of network parameters.

3.3.3 Training

To enhance the self-adaptability of our recognition model, backpropagation method was utilized to regulate the parameters in inverse. In other words, the parameters of the network upgrade very slowly by employing the variance loss function; the cross-entropy loss (Kline and Berardi 2005) was utilized that causes big errors can guide to quickly update the parameters of the network, while little errors can slowly update the parameters of the network. Therefore, for the dataset with the size of T, $e = [(I(1), q(1)), \ldots, (I(T), q(T))]$, the cross-entropy loss function is:

$$\Gamma(\theta) = -\frac{1}{T}\sum_{i=1}^{T}\left[q^{i}log\left(O^{i}\right) + \left(1 - q^{i}\right)\log\left(1 - O^{i}\right)\right] \tag{9}$$

where O^{i} is the actual output related to the image $I(i)$, and $q(i)$ denotes the label of the ith image, the backpropagation gradient of the convolutional parameters b and w were calculated as:

$$\frac{\partial}{\partial w^i}\Gamma(\theta) = \frac{1}{T}\sum_{i=1}^{T}\left[x^i\left(O^i - q^i\right)\right] \tag{10}$$

$$\frac{\partial}{\partial b^i}\Gamma(\theta) = \frac{1}{T}\sum_{i=1}^{T}\left[x^i\left(O^i - q^i\right)\right] \tag{11}$$

The equations of the learnable parameters w^i and b^i of the lth layer are as follows:

$$w_i^l = w_i^l - \eta\frac{\partial \mathrm{E}}{\partial w_i^l} \tag{12}$$

$$b_i^l = b_i^l - \eta\frac{\partial \mathrm{E}}{\partial b_i^l} \tag{13}$$

where η denotes the learning rate (LR), and E represents the error of training data for the current batch samples. Note that the calculations of the pooling layer, convolution layer, FC layer and softmax layer are similar to the detection network.

3.4 Algorithm design

The pseudocode of our framework is given in Algorithm 1. There are several parameters and matricesas inputs in this algorithm. These arguments are images (I), input labels (q_i), output labels (q_o), and parameters η and β. Then, initialize the parameters of our model (line 1). We simply have to loop over our data iterator, and feed the inputs to the networks and optimize. The training process needs that we determine an optimization method and a loss function. Training the proposed model requires enumerating the data loader for the training sets. First, a loop is used for the number of training epochs. Each update to the model (both Detection and Recognition networks) involves the same general pattern comprised of:

- A forward pass of the input through the model (lines 5 and 14).
- Calculating the loss for the model output (lines 6 and 15).
- Backpropagating the error through the model (lines 7 and 16).
- Update the model to reduce a loss (lines 8 and 17).

The detection network continues until all the images are given to the model and faces are extracted (lines 2–10). Moreover, the recognition network continues until all the extracted faces are recognized by the model (lines 11–19).

Algorithm 1. *Pseudocode of the training processin the proposed method*

```
Inputs: $I$, $q_i$
Outputs: $q_o$
begin:
 1:  Initilaize  $\eta, \beta,\ t = 0, n$ =maxiteration
Face Detection:
 2:  While  $t \leq n$ do
 3:  For  each image $i$ in  $I$
 4:      For  each BoundingBox $b$ in  $q_i$
 5:          Predicted_bbx = Forwrad($i, b$)to DetectionNetwork
 6:  Loss = Criterion (Predicted_bbx, GroundTruth_bbx) usingEq. (7)
 7:  Loss. Backward()
 8:  Optimizer. Step()
 9:      Enf for
10:  End while
Face Recognition:
11:  While  $t \leq n$ do
12:      For  each image $i$ in  $I$
13:          For  each label $c$ in  $q_i$
14:  Predicted labels = Forwrad($i, c$)to RecognitionNetwork
15:  Loss = Criterion (Predicted labels, GroundTruth) usingEq. (9)
16:  Loss. Backward()
17:  Optimizer. Step()
18:      Enf for
19:  End while
End
```

4 Experimental analysis

In this section, we indicate the outputs acquired by employing our model on the face DAR system. To measure the performance of our framework, we accomplish various analyses on the DroneFace dataset. Firstly, we discuss the experimental settings. Then, we compared our framework with state-of-the-art models on the face DAR scenario. In order to evaluate the precision of the face DAR algorithms, the DroneFace dataset is utilized as face DAR dataset. DroneFace[1]was created to enhance UAV capabilities and to resolve the present defects of face DAR which is very little work concerning it. The DroneFace dataset contains the following contents:

- 11 subjects including 4 females and 7 males.
- 2057 images including 1364 facial images, 620 raw images.
- The raw samples are in 3680 × 2760 dimension.

[1] https://hjhsu.github.io/DroneFace/.

- The dimensions of the face images are 384 × 384 and 23 × 31.
- The raw samples are taken from 5, 4, 3, and 1.5 m high.
- The raw samples are taken 2 to 17 m away from the subjects with 0.5 m distance.

4.1 Evaluation index

The evaluation of our face DAR model is accomplished utilizing standard evaluation indexes such as detection rate (DR) and recognition rate (RR). Thus, the standard metrics are chosen as accuracy metrics, which is the DR of each image.

$$DR = \frac{TP}{TP + FN} \tag{14}$$

where TP denotes positive faces localized correctly and FN is the set of false-negative instances. In other hands, the RR is the useful metric which is the whole number of correctly recognized face images divided by the whole number of face images as:

$$RR = \frac{TP}{TP + FN} \tag{15}$$

where TP denotes the set of correctly recognized faces, FP is the set of face samples incorrectly assigned to a category and FN are miss-recognized faces.

4.2 Experimental setup

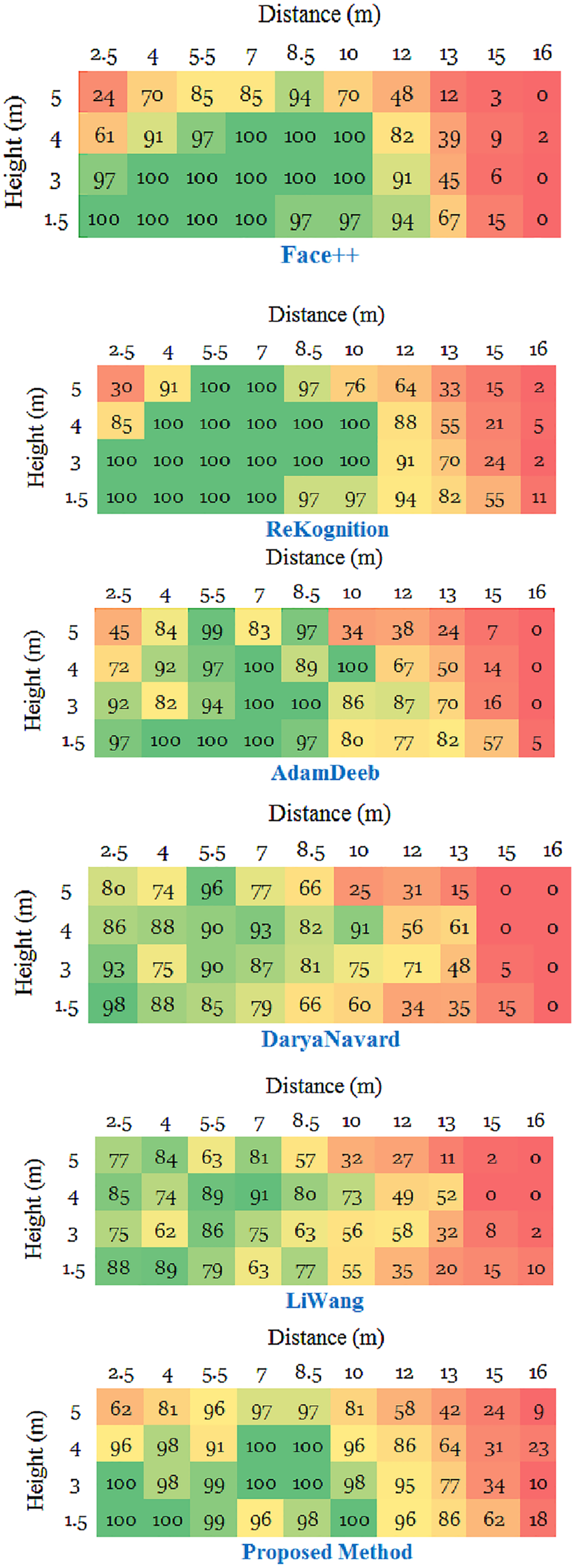
First, the DroneFace dataset is correctly adjusted for each evaluation. Thus, the samples split into test, train and validation sets. So, the validation and train data were made up of 11 sub-folders, each consist of the images of its corresponding labels for different distances and heights. The validation and train data included various images that were randomly split, 20% for the validation and 80% for the training. Our proposed detection network uses pre-trained Mobile-Net as the backbone network that the parameters of this network are defined as: maximum number of epochs is adjusted to 1000, mini-batch is set to 12;LR is set to 0.0001, depth = 30, growth-rate = 10, bottleneck = True, reduction = 0.4, and dropout is adjusted to 0.6; to enhance the optimization usefulness, we evaluated the method with a different number of iterations until it reached a stable accuracy. The highest accuracy was achieved with an epoch number 20. We utilize the Adam optimization method which is an extension of the gradient descent method, which can iteratively update the network weights using the training image; the initialization of LR is adjusting to 0.15; then, where training to the 30th iteration, the LR is changed to 0.023. In this paper, we employed the same loss function and categorical cross-entropy. The NMS method is utilized to extract the redundant face outputs.

4.3 Performance analysis for face detection

In this section, the accuracy of the drone-based face detection models is compared with different settings in distances and heights between UAVs and their targets. We performed several evaluations over the DroneFace database, and the outputs are shown in Fig. 5.Tocomprehend the performance of our DL network in face detection, Fig. 5 shows the heat map face detection rate. The y-axis denotes the heights and the x-axis is ground distances. This paper selects the drone-based face detection algorithm as Face++ (Hsu and Chen 2015), AdamDeeb (Deeb et al. 2020), ReKognition (Hsu and Chen 2015), Daryanavard (Daryanavard and Harifi 2018) and LiWang (Wang and Siddique 2020) to further verify the accuracy of our model. For the evaluation, we detect the faces when the camera is set up at 2–17 m in distances and 1:5 m in heights. Compared with the baseline model, our face detector has a significant improvement by using the deep network model, demonstrating that our detection network is able to focus on the various scale of faces effectually. Compared with Face++ and ReKognition methods, our method has a significant enhancement in DR over the DroneFace dataset, indicating that our model has high efficiency for face localization in multi-scale images. Through observation, it can be found that the DL network can lead the model to detect multi-scale face images. As a result, the heat map of the face detection result of our model compare to other models over different settings in height and distance indicates that our framework can attend higher accuracy on the face detection scenario.

4.4 Performance analysis for face recognition

In this section, we analyze how heights and distances of face recognizers change the RR of methods as shown in Fig. 6. To quantitatively and qualitatively analyze the accuracy of our model for scaled and occluded face recognition, we consider how distance between UAV and their target influence the accuracy of face recognition. We recognize, the faces acquired while the camera is set up at 2–17 m in distances and1:5 m in height. The x-axis indicates the ground distances, and the y-axis shows the heights. Fig. 6 demonstrates the heat map about how drone-based face recognition accomplishes in recognizing faces for different distances and heights. Both ReKognition and Face++ show stable and high RR measures. Therefore, the face recognition utilizing the DL network introduced in this work is excellent compared to other models; our model attains the best recognition accuracy in most cases. Thus, the faces with different scales, low contrast and deformation can be correctly recognized



Distance (m)

Height (m)	2.5	4	5.5	7	8.5	10	12	13	15	16
5	24	70	85	85	94	70	48	12	3	0
4	61	91	97	100	100	100	82	39	9	2
3	97	100	100	100	100	100	91	45	6	0
1.5	100	100	100	100	97	97	94	67	15	0

Face++

Distance (m)

Height (m)	2.5	4	5.5	7	8.5	10	12	13	15	16
5	30	91	100	100	97	76	64	33	15	2
4	85	100	100	100	100	100	88	55	21	5
3	100	100	100	100	100	100	91	70	24	2
1.5	100	100	100	100	97	97	94	82	55	11

ReKognition

Distance (m)

Height (m)	2.5	4	5.5	7	8.5	10	12	13	15	16
5	45	84	99	83	97	34	38	24	7	0
4	72	92	97	100	89	100	67	50	14	0
3	92	82	94	100	100	86	87	70	16	0
1.5	97	100	100	100	97	80	77	82	57	5

AdamDeeb

Distance (m)

Height (m)	2.5	4	5.5	7	8.5	10	12	13	15	16
5	80	74	96	77	66	25	31	15	0	0
4	86	88	90	93	82	91	56	61	0	0
3	93	75	90	87	81	75	71	48	5	0
1.5	98	88	85	79	66	60	34	35	15	0

DaryaNavard

Distance (m)

Height (m)	2.5	4	5.5	7	8.5	10	12	13	15	16
5	77	84	63	81	57	32	27	11	2	0
4	85	74	89	91	80	73	49	52	0	0
3	75	62	86	75	63	56	58	32	8	2
1.5	88	89	79	63	77	55	35	20	15	10

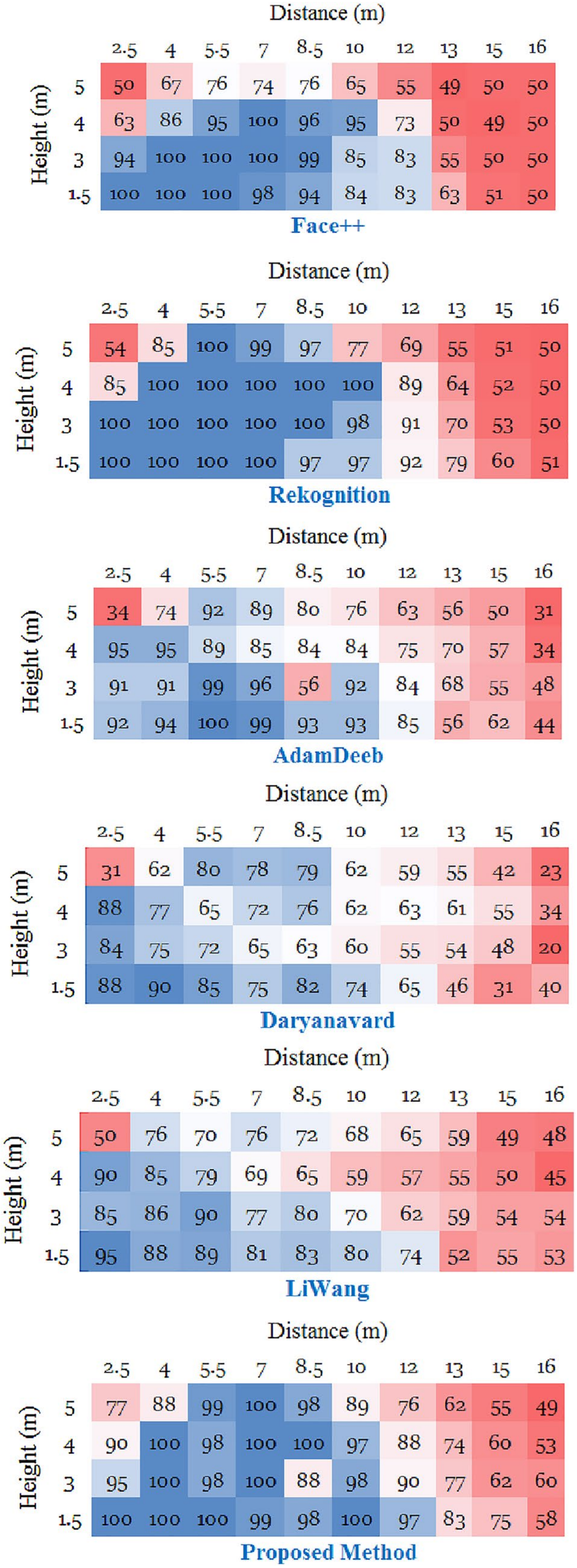
LiWang

Distance (m)

Height (m)	2.5	4	5.5	7	8.5	10	12	13	15	16
5	62	81	96	97	97	81	58	42	24	9
4	96	98	91	100	100	96	86	64	31	23
3	100	98	99	100	100	98	95	77	34	10
1.5	100	100	99	96	98	100	96	86	62	18

Proposed Method

Fig. 5 The face detection rate in correspondence to heights and distances



Distance (m)

Height (m)	2.5	4	5.5	7	8.5	10	12	13	15	16
5	50	67	76	74	76	65	55	49	50	50
4	63	86	95	100	96	95	73	50	49	50
3	94	100	100	100	99	85	83	55	50	50
1.5	100	100	100	98	94	84	83	63	51	50

Face++

Distance (m)

Height (m)	2.5	4	5.5	7	8.5	10	12	13	15	16
5	54	85	100	99	97	77	69	55	51	50
4	85	100	100	100	100	100	89	64	52	50
3	100	100	100	100	100	98	91	70	53	50
1.5	100	100	100	100	97	97	92	79	60	51

Rekognition

Distance (m)

Height (m)	2.5	4	5.5	7	8.5	10	12	13	15	16
5	34	74	92	89	80	76	63	56	50	31
4	95	95	89	85	84	84	75	70	57	34
3	91	91	99	96	56	92	84	68	55	48
1.5	92	94	100	99	93	93	85	56	62	44

AdamDeeb

Distance (m)

Height (m)	2.5	4	5.5	7	8.5	10	12	13	15	16
5	31	62	80	78	79	62	59	55	42	23
4	88	77	65	72	76	62	63	61	55	34
3	84	75	72	65	63	60	55	54	48	20
1.5	88	90	85	75	82	74	65	46	31	40

Daryanavard

Distance (m)

Height (m)	2.5	4	5.5	7	8.5	10	12	13	15	16
5	50	76	70	76	72	68	65	59	49	48
4	90	85	79	69	65	59	57	55	50	45
3	85	86	90	77	80	70	62	59	54	54
1.5	95	88	89	81	83	80	74	52	55	53

LiWang

Distance (m)

Height (m)	2.5	4	5.5	7	8.5	10	12	13	15	16
5	77	88	99	100	98	89	76	62	55	49
4	90	100	98	100	100	97	88	74	60	53
3	95	100	98	100	88	98	90	77	62	60
1.5	100	100	100	99	98	100	97	83	75	58

Proposed Method

Fig. 6 The face recognition in correspondence to heights and distances

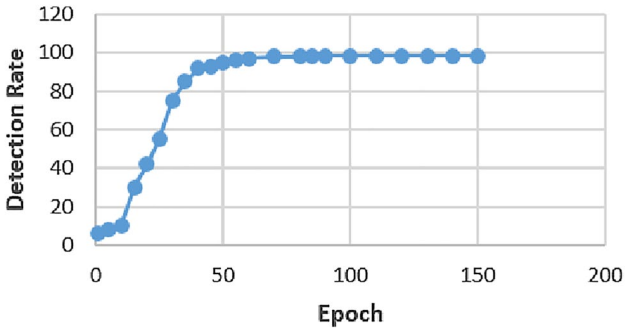


Fig. 7 Convergence analysis of detection network

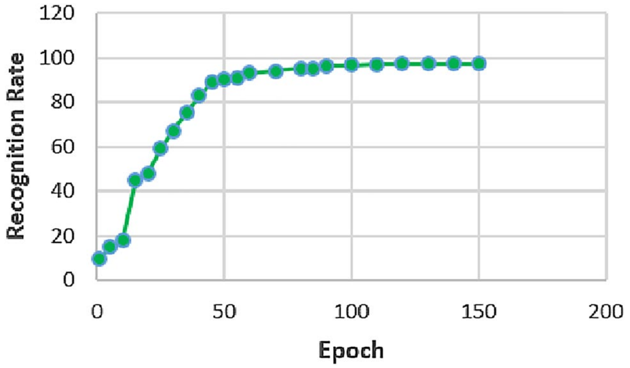


Fig. 8 Convergence analysis of recognition network

using the proposed method. Our proposed model uses the DL algorithm to recognize the scaled and occluded faces. Thus, our model can better adjust to the control of scaled and occluded samples in face recognition. As a result, our face recognition model is better than the other methods due to the multi-scale face recognition layers to take high-level semantic features.

4.5 Convergence behavior

This section aims to analyze the convergence property of our model for both face DAR scenarios. That is, the number of epochs required for our model to converge is evaluated. Figures 7 and 8 provide the convergence speed of our model for the DroneFace dataset; we can see that our model has appropriate convergence properties. For example, Figs. 7 and 8 show that the proposed deep network has a suitable convergence speed while it has high DR and RR values in the DroneFace dataset. For example, we can see from the convergence curves that the proposed method could escape from the local optima in later iterations. In addition, the results illustrate that our proposed model generally requires only 100 iterations to converge. This evidence demonstrates that our model has appropriate convergence behavior, and it can be used effectively in practical applications. The convergence behavior of our model indicates that it exhibits its capability of convergence. In addition, the proposed model converges slowly in Fig. 9 after 120 iterations. It should be noted that our model utilizes the appropriate optimization algorithm in its process, which leads to reduce the number of epochs needed for obtaining stable results.

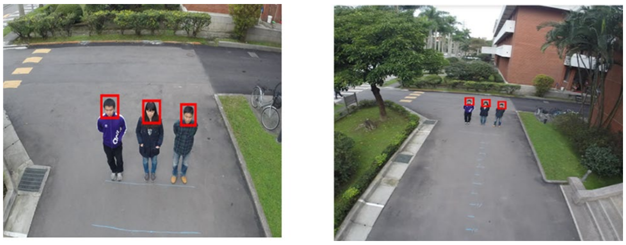


Fig. 9 Sample faces detected/recognized

4.6 Qualitative results

To more comprehend the efficiency of our framework over the face DAR, we perform qualitative results on the DroneFace dataset as shown in Fig. 9. That outputs are taken for the different distances and heights of the faces. As seen in the samples, two faces are in all images that the detection and recognition of these facial images in correspondence to various heights and distances is the optimal result for the problem. Our model can localize multiple faces by using the training set. The qualitative results in Fig. 9 show sample face images detected/recognize correctly by the proposed method. It can be found that our DL network can lead the model to detect and recognize the visible area of the scales face, while decreasing the influence of background on DAR accuracy.

4.7 Scalability analysis

To determine the scalability of our model, it is evaluated on varying parts of the dataset from 0.1 to 1 portions stepping by 0.1 and the outputs are demonstrated in Fig. 10. The results indicate that the training time raises linearly by increasing the size of the training data. This means that our model can be employed to large-scale datasets.

4.8 Discussion

The proposed model includes three main steps. The first stage is an SDD-based face localization process so that the face images are extracted, and the second stage is DL-based face classification which is able to improve the accuracy of face recognition. The main property of the original face detection/recognition methods is utilizing a DL method to

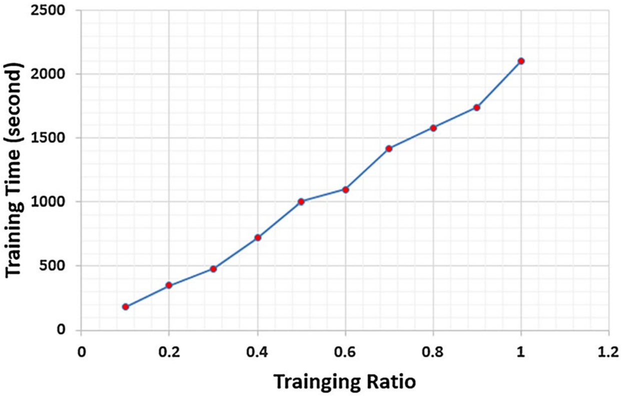


Fig. 10 The scalability of our method over the DroneFace dataset

search through data space to recognize the faces. Although our model uses this property, it uses a novel mechanism based on the Mobile-Netto improve the detection of small faces, multi-scale faces and enhance the detection accuracy. The proposed detection network is a regression-based model which is more efficient for multi-scale face detection on UAVs and thus it is more effective than the previous models. This property makes our model to accurately detect multi-scale faces with a small number of trainable parameters. Note that the previous algorithm has many false detections and false recognition under different heights and distance conditions on drones. To address this issue our model uses FFU for combining low-level and high-level features through the SSD-based architecture. In each iteration, the depth-DWSCs is used to decrease computational complexity and parameters. Besides, we combined features using element-wise sum. This refinement strategy results in improving the quality of detection results. Similar to any DL model, our framework requires an optimizer and loss function. Therefore, our model utilizes the appropriate optimization algorithm and loss function in its process, which leads to reduce in the number of iterations needed to obtain stable results. As a result, this improvement indicates the capability of our model to satisfy industrial requirements.

5 Conclusion and future works

The accuracy of face DAR is affected when UAVs take images in a long-distance and from high altitudes. To enhance the accuracy of face DAR, a DL guidance framework is presented in this work, which employs the CNN model for employing face DAR in drone-based applications. In this paper, we mainly investigated how heights and distances influence the performance of face recognition on UAVs. An efficient DL-based model based on the CNN for face DAR is proposed. Due to CNN's strong learning ability, it enhances the accuracy and speed of the network. Our method contains two phases. In the first phase, the SDD-based face detection process extracts faces image, and the second phase is DL-based face classification which can appreciably enhance the accuracy of face recognition. Several evaluations indicate that our framework is superior to other models for the performance of face DAR on the DroneFace dataset. Moreover, our model attains a higher balance between speed and accuracy, which can be utilized in the security surveillance scenario.

There are several future directions for improving our model. First, the current face recognition methods and our method are capable of recognizing faces on UAVs with some limits in angle, especially when UAVs take photos with a large angle of depression. To address this limitation, the idea of the grid can be employed in our model which focuses on features, such as eyes, the mouth or cheeks and geometric of face to model facial components. The second future direction is to utilize generative adversarial networks for large-pose and small-pose face recognition. Although our model has high stability with different poses, it also requires identifying faces under the dynamic environment. Third, another future direction is to reduce the computation complexity of our framework facing with large-scale datasets. A general way to this aim is to propose a new architecture by using the advantages of the MobileNet. Finally, one can design an incremental model for other severe environments. Furthermore, we would develop our model for the facial expression scenario.

References

Almabdy S, Elrefaei L (2019) Deep convolutional neural network-based approaches for face recognition. Appl Sci 9:4397

Atmaja AP, Setyawan SB, Setia LD, Yulianto SV, Winarno B, Lestariningsih T (2021) Face recognition system using micro unmanned aerial vehicle. J Phys Conf Ser 1845:012043

Bae H, Kim S (2005) Real-time face detection and recognition using hybrid-information extracted from face space and facial features. Image vis Comput 23:1181–1191

Bhattacharyya S (2011) A brief survey of color image preprocessing and segmentation techniques. J Pattern Recognit Res 1:120–129

Bold S, Batchimeg S, Seong RL (2016) Implementation of autonomous unmanned aerial vehicle with moving-object detection and face recognition. In: Information science and applications (ICISA). Springer

Bonetto M, Pavel K, Giovanni R, Touradj E (2015) Privacy in mini-drone based video surveillance. In: 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG). IEEE, pp 1–6

Cao B, Li M, Liu X, Zhao J, Cao W, Lv Z (2021) Many-objective deployment optimization for a drone-assisted camera network. IEEE Trans Netw Sci Eng 8:2756–2764

Chang X, Nie F, Wang S, Yang Y, Zhou X, Zhang C (2016) Compound rank-projections for bilinear analysis. IEEE Trans Neural Netw Learn Syst 27:1502–1513

Chen K, Yao L, Zhang D, Wang X, Chang X, Nie F (2020) A semi-supervised recurrent convolutional attention model for human activity recognition. IEEE Trans Neural Netw Learn Syst 31:1747–1756

Cheng E-J, Chou K-P, Rajora S, Bo-Hao Jin M, Tanveer C-T, Young K-Y, Prasad M (2019) Deep sparse representation classifier for facial recognition and detection system. Pattern Recogn Lett 125:71–77

Daryanavard H, Harifi A (2018) Implementing face detection system on uav using raspberry pi platform. In: Iranian conference on electrical engineering (ICEE). IEEE, pp 1720–23

Davis N, Francesco P, Karen P (2013) Facial recognition using human visual system algorithms for robotic and UAV platforms. In: 2013 IEEE conference on technologies for practical robot applications (TePRA). IEEE, pp 1–5

Deeb A, Kaushik R, Kossi DE (2020) Drone-based face recognition using deep learning. In: International conference on advanced machine learning technologies and applications. Springer, pp 197–206

Du Z, Robert F, Tianshi C, Paolo I, Ling L, Tao L, Xiaobing F, Olivier T (2015) ShiDianNao: shifting vision processing closer to the sensor. In: Proceedings of the 42nd annual international symposium on computer architecture, pp 92–104

Fang W, Wang L, Ren P (2019) Tinier-YOLO: a real-time object detection method for constrained environments. IEEE Access 8:1935–1944

Gao C, Lu S-L (2008) Novel FPGA-based Haar classifier face detection algorithm acceleration. In: International conference on field programmable logic and applications. IEEE, pp 373–78

Herrera D, Imamura H (2019) Design of facial recognition system implemented in an unmanned aerial vehicle for citizen security in Latin America. In: ITM web of conferences, 04002. EDP Sciences

Hjelmås E, Low BK (2001) Face detection: a survey. Comput vis Image Underst 83:236–274

Hsu H-J, Chen K-T (2015) Face recognition on drones: issues and limitations. In: Proceedings of the first workshop on micro aerial vehicle networks, systems, and applications for civilian use, pp 39–44

Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. PMLR, pp 448–56

Iqbal MM, Sameem SI, Naqvi N, Kanwal S, Ye Z (2019) A deep learning approach for face recognition based on angularly discriminative features'. Pattern Recognit Lett 128:414–419

Jurevičius R, Goranin N, Janulevičius J, Nugaras J, Suzdalev I, Lapusinskij A (2019) Method for real time face recognition application in unmanned aerial vehicles. Aviation 23:65–70

Kalra I, Singh M, Nagpal S, Singh R, Vatsa M, Sujit PB (2019) Dronesurf: benchmark dataset for drone-based face recognition. In: 2019 14th IEEE international conference on automatic face & gesture recognition (FG 2019). IEEE, pp 1–7

Kim S, Kwon D, Ji Y (2019) CNN based human detection for unmanned aerial vehicle (poster). In: Proceedings of the 17th annual international conference on mobile systems, applications, and services, pp 626–27

Kline DM, Berardi VL (2005) Revisiting squared-error and cross-entropy functions for training neural network classifiers. Neural Comput Appl 14:310–318

Korshunov P, Ooi WT (2011) Video quality for face detection, recognition, and tracking. In: ACM transactions on multimedia computing, communications, and applications (TOMM), vol 7, pp 1–21

Kumar A, Kaur A, Kumar M (2019) Face detection techniques: a review. Artif Intell Rev 52:927–948

Kumar A, Suresh K, Kubakaddi S (2014) Multipiple face detection and tracking using adaboost and camshift algorithm

Li Y, Gong S, Sherrah J, Liddell H (2004) Support vector machine based multi-view face detection and recognition. Image vis Comput 22:413–427

Li Z, Nie F, Chang X, Nie L, Zhang H, Yang Y (2018a) Rank-constrained spectral clustering with flexible embedding. IEEE Trans Neural Netw Learn Syst 29:6073–6082

Li Z, Nie F, Chang X, Yang Y, Zhang C, Sebe N (2018b) Dynamic affinity graph construction for spectral clustering using multiple features. IEEE Trans Neural Netw Learn Syst 29:6323–6332

Li Z, Tang Xu, Xiang Wu, He R (2019a) Progressively refined face detection through semantics-enriched representation learning. IEEE Trans Inf Forensics Secur 15:1394–1406

Li Z, Yao L, Chang X, Zhan K, Sun J, Zhang H (2019b) Zero-shot event detection via event-adaptive concept relevance mining. Pattern Recogn 88:595–603

Li B, Yang J, Zhang Y (2021) Sign language/gesture recognition based on cumulative distribution density features using UWB radar. IEEE Trans Instrum Meas 70:1–13

Lin S-H, Kung S-Y, Lin L-J (1997) Face recognition/detection by probabilistic decision-based neural network. IEEE Trans Neural Netw 8:114–132

Liu Y, Chen J (2021) Unsupervised face Frontalization for pose-invariant face recognition. Image vis Comput 106:104093

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC (2016) Ssd: single shot multibox detector. In: European conference on computer vision. Springer, pp 21–37

Luo M, Chang X, Nie L, Yang Y, Hauptmann AG, Zheng Q (2018) An adaptive semisupervised feature analysis for video semantic recognition. IEEE Trans Cybern 48:648–660

Luo J, Liu J, Lin J, Wang Z (2020) A lightweight face detector by integrating the convolutional neural network with the image pyramid. Pattern Recogn Lett 133:180–187

Lv Z, Qiao L, Hossain MS, Choi BJ (2021) Analysis of using blockchain to protect the privacy of drone big data. IEEE Netw 35:44–49

Matai J, Irturk A, Kastner R (2011) Design and implementation of an fpga-based real-time face recognition system. In: 2011 IEEE 19th annual international symposium on field-programmable custom computing machines. IEEE, pp 97–100

Meduri P, Telles E (2018) A Haar-cascade classifier based smart parking system. In: Proceedings of the international conference on image processing, computer vision, and pattern recognition (IPCV). The Steering Committee of The World Congress in Computer Science, Computer, pp 66–70

Mishra NK, Dutta M, Singh SK (2021) Multiscale parallel deep CNN (mpdCNN) architecture for the real low-resolution face recognition for surveillance. Image vis Comput 115:104290

Nair P, Cavallaro A (2009) 3-D face detection, landmark localization, and registration using a point distribution model. IEEE Trans Multimed 11:611–623

Saha A, Kumar A, Sahu AK (2018) Face recognition drone. In: 2018 3rd international conference for convergence in technology (I2CT). IEEE, pp 1–5

Samaria FS, Harter AC (1994) Parameterisation of a stochastic model for human face identification. In: Proceedings of 1994 IEEE workshop on applications of computer vision, pp 138–42

Sarath RNS, Varghese JT, Pandya F (2019) Unmanned aerial vehicle for human tracking using face recognition system. In: 2019 advances in science and engineering technology international conferences (ASET). IEEE, pp 1–5

Sinha D, El-Sharkawy M (2019) Thin mobilenet: an enhanced mobilenet architecture. In: 2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON), 0280–85. IEEE

Suri S, Sankaran A, Vatsa M, Singh R (2021) Improving face recognition performance using TeCS2 dictionary. Pattern Recogn Lett 145:88–95

Wang Li, Siddique AA (2020) Facial recognition system using LBPH face recognizer for anti-theft and surveillance application based on drone technology. Meas Control 53:1070–1077

Wang L, Xiang Yu, Bourlai T, Metaxas DN (2019) A coupled encoder–decoder network for joint face detection and landmark localization. Image vis Comput 87:37–46

Wang P, Wang P, Fan En (2021) Violence detection and face recognition based on deep learning. Pattern Recogn Lett 142:20–24

Yan C, Chang X, Luo M, Zheng Q, Zhang X, Li Z, Nie F (2020) Self-weighted robust LDA for multiclass classification with edge classes. ACM Trans Intell Syst Technol (TIST) 12:1–19

Yang M-H, Kriegman DJ, Ahuja N (2002) Detecting faces in images: a survey. IEEE Trans Pattern Anal Mach Intell 24:34–58

Yang S, Luo P, Loy CC, Tang X (2017) Faceness-net: face detection through deep facial part responses. IEEE Trans Pattern Anal Mach Intell 40:1845–1859

Yang S, Wang J, Deng B, Azghadi MR, Linares-Barranco B (2021) Neuromorphic context-dependent learning framework with fault-tolerant spike routing. IEEE Trans Neural Netw Learn Syst 1–15

Yuan Z (2020) Face detection and recognition based on visual attention mechanism guidance model in unrestricted posture. Sci Program 2020

Zhang D, Yao L, Chen K, Chang X, Liu Y (2020) Making sense of spatio-temporal preserving representations EEG-based human intention recognition. IEEE Trans Cybern 50:3033–3044

Zhou R, Chang X, Shi L, Shen YD, Yang Y, Nie F (2020) Person reidentification via multi-feature fusion with adaptive graph learning. IEEE Trans Neural Netw Learn Syst 31:1592–1601

Zhu Y, Jiang Y (2020) Optimization of face recognition algorithm based on deep learning multi feature fusion driven by big data. Image vis Comput 104:104023

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.