

RESEARCH ARTICLE

MS-Net: Multi-Segmentation Network for the Iris Region Using Deep Learning in an Unconstrained Environment

JASEM RAHMAN MALGHEET, NORIDAYU BT MANSHOR^{ib}, LILLY SURIANI AFFENDEY^{ib},
AND ALFIAN BIN ABDUL HALIN^{ib}, (Member, IEEE)

Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, (UPM), Serdang, Selangor Darul Ehsan 43400, Malaysia

Corresponding author: Noridayu Bt Manshor (ayu@upm.edu.my)

This work was supported by Putra Grant funded by Universiti Putra Malaysia (UPM) under Grant 9693600.

ABSTRACT Iris segmentation is a significant phase in the iris recognition process because segmentation errors cascade into all subsequent phases. Therefore, it is important that errors in iris segmentation are minimised. The U-Net architecture that uses a deep learning approach was previously adopted for this task, but its performance was affected by the deformation of iris images caused by various noise factors in unconstrained (non-ideal) environments. Scratches, blurriness, dirt, specular reflections and other noise factors are some of the challenges faced in unconstrained environments when eyeglasses are present in the original images. Additionally, the performance of the iris segmentation was degraded due to problems of exploding gradient or vanishing gradient and the loss of information. This paper proposes a multi-segmentation network called MS-Net, based on a deep learning approach, that aims to capture high-level semantic features while maintaining spatial information to improve the accuracy of iris segmentation. MS-Net consists of three principal segments: a feature encoder network, a multi-scale context feature extractor network (MSCFE-Net) and a feature decoder network. MSCFE-Net a multi-scale context feature extractor network is constructed from a dilated residual multi-convolutional network module and a pyramid pooling residual model based on an attention convolutional module. In addition, the proposed MS-Net contains dense connections within the feature decoder network to decrease training difficulty, by using only a few training samples. The accuracy of MS-Net was evaluated on the CASIA-Iris.V4-1000 and UBIRIS.V2 databases. The performance of our proposed MS-Net method on the CASIA-Iris.V4-1000 and UBIRIS.V2 databases achieved an overall accuracy of 97.11% and 96.128%, respectively. Experiment results show that MS-Net is able to achieve better results compared to earlier methods used for the same purpose.

INDEX TERMS Iris recognition, iris segmentation, traditional techniques, U-Net architecture, convolutional neural network (CNN) techniques, dilated convolution (DC), residual network (ResNet), deep learning (DL) techniques.

I. INTRODUCTION

Traditionally, authentication made use of passwords, smart cards, pin codes and patterns. These techniques are being replaced because they were misused and people were easily deceived [1]. Biometric recognition technologies are being adopted in many important tasks, such as to unlock

The associate editor coordinating the review of this manuscript and approving it for publication was Siddhartha Bhattacharyya^{ib}.

a smartphone, to withdraw cash from ATMs, to identify travellers at airports, for border control, to record attendance [2], [3] and to make purchases at retail outlets [4]. The need for reliable, quick and secure methods has led to the emergence of physiological and behavioural models in biometric recognition systems [5]. Physiological biometrics involves recognising the geometry of the iris [6], fingerprint [7], face [8], retina or hand. The behavioural model authenticates by recognising signature, voice or gait [9], [10]. Among all these,

recognising the iris has proven to be capable of achieving high distinctiveness, reliability and good performance [11] in terms of the false rejection rate (FRR) and false acceptance rate (FAR) [12], [13]. The most recent methods either have deep-learning features accompanied by a traditional pipeline [14], [15] or may simulate an existing model that has been trained in iris recognition [16].

Security systems need to be efficient, especially at public entry points such as airports, where people are put through numerous time-consuming screenings and stressful processes. Hence, it is crucial that a biometric authentication system be implemented to provide the shortest process time, with minimal user involvement. Delays can be shortened and the economic burden lightened if swift and secure systems were available [17].

Recent developments in iris recognition were focused on overcoming challenges in unconstrained environments to improve recognition rates. However, real-life situations present more complex scenarios such as low resolutions, off-angles, scaling, rotations, occlusions by eyelash, distortions, cropped images, blurred or off-focus images and noise from eyeglasses [18], [19], [20], [21], [22], [23], [24], [25]. Presently, hand-held mobile devices employ iris recognition, replacing traditional techniques as a more reliable method of authenticating in selected areas [26]. The iris segmentation (IS) phase is the most critical phase of an iris recognition system (IRS) because its accuracy is fundamental for the accurate performance of all subsequent phases, hence the entire system. The IS phase has been shown to be the most computationally challenging step in an IRS [27], [28], [29], [30], [31]. This phase has traditionally been tackled by utilising handcrafted and custom-designed segmentation methods [15], [32] that have achieved significant results in several databases of various qualities.

Considering the success of DL approaches for overcoming many other challenges in producing accurate iris images, researchers are performing in-depth studies of a technique called the convolutional neural network (CNN) to further improve the accuracy of current IS techniques [16], [33], [34], [35]. Past experiments on iris images have shown that errors in the segmentation phase impact the overall accuracy of an IRS [36]. Segmentation errors cause further errors in all subsequent phases, leading to even more recognition errors. The challenges of IS are due to noise from eyeglasses [1], [3], [25], [37] IS techniques can also fail because this noise is sometimes misconstrued as limbus boundaries or pupillary boundaries [38]. Thus, the IS phase of an iris recognition system for images with eyeglasses still faces challenges in unconstrained environments [25], [31], [39]. These challenges in fact enhance the development of an IRS because improving IS techniques significantly increases the accuracy of biometric recognition. DL techniques are better than traditional techniques at identifying poor quality samples of iris images [23]. Therefore, the significance of this study is in improving the accuracy of IS in the presence of noise such as

scratches, blurriness, dirt, specular reflections and other noise factors that are caused by eyeglasses.

To tackle the challenges encountered in IS due to eyeglasses, the multi-segmentation network (MS-Net) based on Deep CNNs (DCNNs) is proposed to segment the iris region. The authors propose the MS-Net approach to extract the iris region, which would allow the detection of distinguishing features, which were produced as a result of the simultaneous processing of information from multiple sources. MS-Net improves the accuracy of interconnected tasks because it is equipped with multi-task learning abilities and multi-loss functions that capture the differences and similarities between multiple samples. The architecture of the proposed MS-Net method comprises three main networks: the feature encoder network, the multi-scale context extractor network (MSCFE-Net) and the feature decoder network. Within the MSCFE-Net, the authors propose the Dilated Residual Multi-Convolutional Network (DRMC-Net) model to encode high-level feature maps and obtain deeper features using four convolution branches with multi-scale dilated convolutions. Additionally, the authors also introduce a Pyramid Pooling Residual Model based on the Attention Convolutional (PPR-AC) module, which was modelled on a pyramid pooling network module [40]. PPR-AC improves feature extraction from the multi-scale receptive field in high-level feature maps. The aim of this strategy is to avoid a degradation in performance when faced with problematic situations of noticeable scale differences. In this network, residual connectivity was used to prevent the gradient from vanishing. For the feature decoder network, a dense block was inserted directly after the convolution layer of each block to quicken training and minimise the number of parameters required to enhance performance.

The fundamental contributions of this paper are:

- a) Introduces an MS-Net architecture based on DL to address different types of noise caused by eyeglasses. MS-Net learns the representative features of an iris image from a small amount of training data.
- b) Proposes the DRMC-Net and PPR-AC modules to enhance the receptive field and capture multi-scale features from high-level semantic feature maps by combining them with the encoder-decoder network for improving the segmentation of iris images.
- c) The MS-Net architecture uses the dense block in the decoder network structure to improve segmentation accuracy. MS-Net has fewer parameters, requires less training and segments faster, thus it is suitable for practical applications.
- d) Evaluates the accuracy of the proposed MS-Net architecture on two available public databases, using five evaluation metrics.

The remaining sections of this paper are organised as follows: Section II presents an overview of related works on IS, comprising traditional and DL techniques; Section III gives details of the MS-Net method proposed in this paper;

TABLE 1. Comparison between previous methods and the proposed MS-Net method for iris segmentation.

Type	Method	Strengths	Weaknesses
Using traditional techniques	CHT method for Iris segmentation [43].	High performance in detecting the iris boundary in ideal environments with high user cooperation.	Low accuracy in non-ideal environments where images are accompanied with the noise of visible light.
	Delogne–Kása circlefitting based on OMT [44].	OMT provides gradient direction of a circle within a particular radial range.	Binarization may not succeed in noisy iris images, which might decrease accuracy.
	Coarse-to-fine method based on adaptive filtering and dynamic thresholding [47].	A step-by-step technique for determining the true edge that is based on basic image processing methods.	Specular reflection processing and pre-processing are required.
	The fusion of iterative shrinking and expanding of the AC [48].	Shrinking and expanding of the AC to detect the correct iris boundary.	The iterative process requires a long processing time.
	Iris segmentation using rough entropy with CSA [49].	An effective computational method to locate the visible region of the iris.	Pupil segmentation is impacted by noise due to the threshold value.
	Morphological operator and regional properties [29].	This technique is ideal for quick implementations on an FPGA, which is best suited for real-time systems.	This technique is unable to identify the actual pupillary border when there are thick eyebrows and eyelashes present.
Using deep learning techniques	CHT based on Canny’s edge detection method for iris segmentation [24].	High accuracy in constrained environments because the CHT detects edges of the iris region by estimating the coordinates of the centre and the radius.	Low accuracy in non-ideal environments with noise due to eyeglasses. In addition, it assumes that the border of the iris is circular.
	Four network models that combine DC and U-Net for iris segmentation [61].	This method used DC rather than the original convolution layer of the U-Net method to extract more global features.	The accuracy of this method in detecting the iris boundary is low on the iris dataset with VIS.
	An ATT-UNet that has a structure similar to the VGG16 without FC layers [2].	Uses an attention mask guided to extract distinctive features for classifying pixels of the iris and the non-iris regions.	This model fails to segment extreme samples like very thin hair, and noise due to eyeglasses.
	A U-Net for iris segmentation [15].	This method has the ability to learn relatively accurately from a small database.	This method is insufficient to distinguish between the non-iris and the iris pixels around the boundary region.
	IrisDenseNet: DenseNet with SegNet based on the encoder-decoder network [3].	Dense connectivity improves features by directly connecting the features from previous layers, which results in better accuracy.	A dense connection uses high memory; thus, a smaller mini-batch size should be employed.
	Fully residual encoder-decoder network (FRED-Net) [1].	Higher accuracy is achieved without increasing network depth by using residual skip connections between convolutional layers for both encoder and decoder networks.	The semantic segmentation performance relies on the number of training samples. In addition, there are many false-positive and false-negative errors.
	Faster Region-based Convolutional Neural Network (R-CNN) with only six layers [13].	To enhance the technique of limbal boundary localization, the technique of locating the pupillary region and detecting the points of the limbic edge is used.	The performance of this method can be heavily influenced by noisy pixels around the edge of the iris region and noise caused by eyeglasses.
	Two-stage CHT- and VGG-face-based deep learning for iris segmentation [16].	This method simply locates the rough edge of the iris and uses CNN only within the ROI determined based on the rough edge.	A huge amount of data is required for CNN training with the size of 21 × 21 px.
MS-Net (Proposed method)	The proposed MS-Net aims to capture high-level semantic features to improve the accuracy of IS in challenging situations due to eyeglasses. The MS-Net is effective in improving the accuracy of interconnected tasks, leading to multi-task learning by capturing the differences and similarities between multiple samples.	In cases where the image of the iris contains very difficult challenges due to eyeglasses, it can produce false positives and false negatives errors.	

Section IV describes the experiment setup, covering databases, implementation details and evaluation metrics; experiment results are explained in Section V; and section VI ends this paper with our conclusions.

II. RELATED WORKS

This section presents a brief review of the important works on IS using traditional and DL techniques.

A. TRADITIONAL TECHNIQUES

Most techniques of executing IS use traditional image processing such as Hough Transform (HT), morphological operator, Histograms of Oriented Gradients (HOG), thresholding, Daugman’s method, Canny edge detector (CED) and active contours. The basic IS techniques were introduced by Daugman [41] using the Circular Hough Transform (CHT) and Integro-Differential Operator (IDO) methods [42]. Sahmoud & Abuhaiba [43] proposed a 2-step method for IS

by employing the k-means clustering technique to first determine the expected iris region on a vertical CED output of the iris image, followed by the CHT method to then determine the iris radius and centre. An algorithm was developed to detect and separate the upper eyelid and to remove the non-iris regions. To shorten the CHT processing time for iris edge detection, the k-means clustering algorithm was used. However, this method had difficulty localising the upper and lower eyelids, a step that was necessary in order to determine the natural edge of the iris. Chung et al. [44] had used CHT to roughly locate the pupils by determining the outer and inner edges of the iris, using the Orientation Matching Transform (OMT) method. Then, to obtain a better iris region, the Delogne-Kåsa circle fitting (DKCFS) technique was employed to remove outlier details of coarse inner and outer iris edges. Accuracy is affected when there is a combination of very noisy eye images and the possible failure of binarization. Improving the accuracy of the IS phase faces such challenges as off-angles, reflection from eyeglasses, occlusion by eyelids and eyelashes, rotation and different scanners. Shang et al. [24] developed CHT based on Canny's edge detection technique [45] for segmenting the iris region. CHT identifies boundaries by taking the duplicity between points on a boundary and the coordinate parameters of the centre and the radius of that boundary [46]. The weight of the matrix established by the radius' threshold was used to determine these parameters.

Gangwar et al. [47] proposed an adaptive filtering Iris-Seg method to utilise a coarse-to-fine approach of segmenting the pupil and iris edges. The method exploits dynamic thresholding assisted by local features to segment the pupil. The iris edge is located using adaptive filters in the polar space and then the cartesian space was improved. The performance of this technique may be significantly poorer in images within the visible wavelength spectrum (VIS) because they are more complicated than near-infrared (NIR) images. Abdullah et al. [48] developed a method for IS using iris and pupil segmentation. The pupillary boundary was identified using the thresholding method, where detection of the iris boundary was performed by a fusion-based shrinking and expanding of an iterative active contour (AC). This method was more robust for closed eye detection, using the Hue-Saturation-Intensity (HSI) method and separating the eyelids in the iris images. The challenge in this method is the longer processing time of the iteration process. Taking into account the unconstrained (non-ideal) environment, Sardar et al. [49] came up with a soft-computing method for IS, based on the rough entropy for localising the pupillary edge by reducing spatial ambiguity and the grey level, which was achieved by employing circular sector analysis (CSA) to determine the limbus boundary. The weakness of this method is the noise that affects the segmented pupil due to the threshold value. Khan et al. [29] devised an IS method to attain the best real-time results, using a Field Programmable Gate Array (FPGA). This technique uses the resulting image obtained

after removing the background, followed by morphological operations to locate the pupil region. The proposed non-iterative technique was implemented via an FPGA to achieve high speed and high accuracy, using less memory. However, it failed to determine the pupillary boundary in the presence of thick eyebrows and eyelashes, which worsened the accuracy of the visual spectrum images due to the poor contrast between the pupil and the iris in VIS environments. Although all these techniques achieved good speed and accuracy, their overall performance was still influenced by various environmental factors like varying illumination, off-angles, eyelids, eyelashes, reflections, blurring, various types of noise from eyeglasses and the ghost effect. Therefore, adopting a DL approach in this research could result in IS improvement.

B. DEEP LEARNING TECHNIQUES

This section explains the main contributions of DL techniques in iris segmentation. To overcome the challenges of traditional techniques and improve accuracy, IS has been done using DL techniques. Arsalan et al. [16] introduced a 2-phase IS system based on the CNN approach to detect the natural boundary of the iris in unconstrained environments. The first phase takes a rough iris boundary and uses modified CHT to define the region of interest (RoI) by slightly increasing the radius of the iris region. In the second phase, CNN is used by a fine-tuned VGG-face model on the RoI. The output of the CNN layers is two types of features by which the iris and non-iris pixels are classified to identify the correct iris edges. The difficulties are the unavailability of large public databases to investigate and train the DL approach (a large number of 21×21 iris images are needed to train CNN techniques). Despite the impressive results achieved previously, IS still faces challenges in adapting existing methods to low-quality iris images in unconstrained environments. These environments include blurriness, occlusions, off-angles, eyeglasses, rotations, low resolutions, reflections and other issues. Bazrafkan et al. [10] designed the Fully Convolutional Deep Neural Network (FCDNN) based on semi-parallel deep neural networks (SPDNN), which integrates several deep structure networks to allow the usage of parameters similar to SegNet-Basic that was introduced by Badrinarayanan et al. [50].

Bazrafkan & Corcoran [51] proposed a U-shaped system with 13 layers based on CNN to segment the iris region. The method does not have pooling layers to avoid unwanted artifact generated by the various networks. The terminal prediction of the system can utilise the traditional binary cross-entropy loss function. These techniques utilise DL approaches to achieve high IS performances. However, under non-ideal (unconstrained) conditions, such as with eyeglasses, off-angles, blurring, reflections and varying degrees of illumination, there is still potential for better accuracy in detecting the boundaries of the actual iris. Arsalan et al. [3] introduced a deep learning, densely

connected, fully convolutional network called IrisDenseNet to improve IS performance in unconstrained environments. IrisDenseNet comprises two main methods: the densely convolutional network (DenseNet) method [52] and the SegNet method as introduced in [50]. There are challenges to this method. The first is to keep the mini-batch sizes small because training takes longer due to its dense connectivity. The second is to consider high false-positive and false-negative errors. The latter is mainly due to pixels of eyelashes or pupils whose values are roughly similar to the pixel value of the iris region. In addition, there is reflection noise caused by eyeglasses or a dark iris region, which is still prevalent. To improve the accuracy of the IS, Arsalan et al. [1] suggested a deep learning approach based on the Fully Residual Encoder-Decoder Network (FRED-Net), also known as a semantic segmentation network. FRED-Net is based on the architecture of SegNet as proposed in [50], setting the convolutional layers at 8 rather than 13. FRED-Net uses the residual skip connections between convolutional layers for both the encoder and the decoder networks. Some limitations of this technique are that the accuracy of the semantic segmentation technique relies on the number of training iris images and that there are false-positive and false-negative errors. Hence, to improve the performance of the technique proposed in [3], Li et al. [13] introduced an IS method that utilises a redesigned structure of the CNN network to become a faster R-CNN network for segmentation of the iris region. A faster R-CNN was designed by Zeiler and Fergus [53]. Their method involves the VGG-16 model suggested by Simonyan and Zisserman [54] as presented in Othman et al. [55]. The procedure was constructed using the Gaussian mixture model (GMM), taking into account the expectation-maximisation (EM) method proposed by Bilmes [56], Figueiredo and Jain [57]. The constructed algorithm was applied to locate the pupillary region and detect the limbus edge points to enhance the limbus boundary localisation technique [58]. Lastly, the limbus and pupillary boundaries were used to locate the iris region. To further enhance the suggested method [13], merged heterogeneous techniques that consider the strength of the CNN approach and the quickness of the traditional approach can be applied. Accuracy can also be improved by using a semantic segmentation technique and the proposed algorithm.

Lozej et al. [15] introduced an end-to-end deep-learning method for IS, based on the famous U-Net architecture. There were still challenges to this technique caused by several factors, including occlusion due to eyelashes and/or eyelids, noisy pixels around the edge of the iris region, rotation, off-angles, over/poor illumination, blurring and noise caused by eyeglasses. To improve the IS performance of the U-Net architecture, Lian et al. [2] came up with the Attention U-Net (ATT-UNet) technique. This is a technique originally suggested by Ronneberger et al. [59] in connection with an attention mask guided by more distinctive features for classifying pixels of the iris and non-iris regions. ATT-UNet first selects

a bounding box of the possible iris region and produces an attention mask. The attention mask is then combined with the distinguishing feature maps as a weighted function. ATT-UNet has a similar architecture to VGG16 but lacks fully connected layers and uses the pre-trained ImageNet template to initialise the weights as mentioned by Krizhevsky et al. [60]. Meanwhile, challenges such as noise due to eyeglasses, occlusion caused by hair and eyelashes, blurring, reflection, bad illumination and off-angles still exist. Zhang et al. [61] designed four network architectures combining the dilated convolution (DC) proposed by Yu and Koltun [62] along with the U-Net architecture in Ronneberger et al. [59]. The DC technique extracts more information from the eye images and accordingly, enhances the efficiency of IS. In the first three architectures of the proposed network, the partial DC was merged with the U-Net architecture (PD-UNet), while in the fourth, the full DC was merged with the U-Net architecture (FD-UNet). The FD-UNet method extracts features using DC rather than the original convolution, allowing for better processing of image information. The accuracy of IS in network performance was enhanced by adjusting the parameters of the network.

The current techniques for semantic segmentation based on a deep learning approach show favourable performance accuracies for iris segmentation, but still face challenges in unconstrained environments, such as in the presence of various types of noises due to eyeglasses, illumination variations, occlusion due to eyelids and eyelashes, blurring, off-angles, ghost effects and reflections. Thus, we proposed a multi-segmentation network called MS-Net, based on a deep learning approach, to improve the accuracy of iris segmentation in challenging situations due to the presence of eyeglasses. Table 1 illustrates the comparison between a number of previous methods and the proposed MS-Net method for iris segmentation.

III. ARCHITECTURE OF THE PROPOSED MS-NET MODEL

This section provides details of the architecture of our proposed MS-Net model. It was developed based on the original U-Net model, aimed at enhancing U-Net in addressing the IS task. The purpose of this enhancement is to enable the modified model to better identify distinguishing features for separating the iris and non-iris pixels in images with eyeglass challenges.

The proposed MS-Net architecture (Figure 1) consists of three main elements: the feature encoder network (marked in navy blue), the MSCFE-Net (marked in straw yellow) and the feature decoder network (marked in saffron yellow). Another essential part of this architecture that is newly-introduced is the skip connection (for concatenation). It concatenates the encoded feature map with the corresponding decoded map.

A. THE FEATURE ENCODER NETWORK

The two architectural branches of MS-Net function similarly to the microscopic optic nerves of the human eye. It extracts

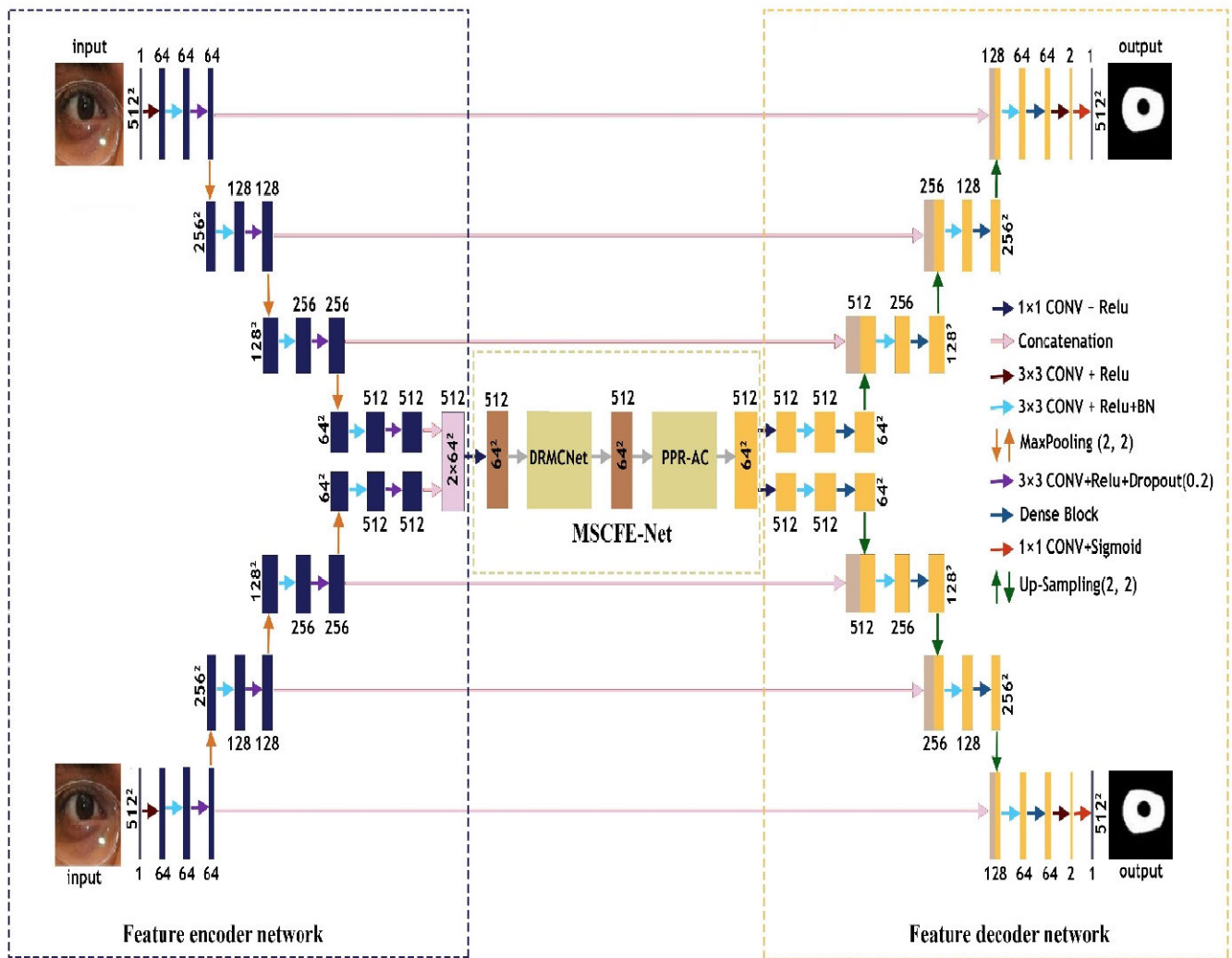


FIGURE 1. The architecture of the proposed MS-Net method for iris segmentation.

features using two similar frameworks and produces a single feature map from a single input image for each branch [63]. In the encoder, each block in the down-sampling path contains two 3×3 convolutions and dropout layers followed by a 2×2 max-pooling layer. The layers in both branches use batch normalisation, ReLU, zero-padding and the encoding process, represented by the navy-blue bands in Figure 1. This dual-input structure is capable of extracting prominent features, thanks to its simultaneous processing of information from multiple images of the eye. The encoder halves the image resolution at each down-sampling step and doubles the number of feature channels.

The encoder is designed as a Siamese [64], [65] structure to allow the model to pick out more distinguishing features and details by simultaneously processing information from two sources (similar to seeing with both eyes at the same time). Sharing the same weightage means using the same method to extract features from both eye images. In addition, neighbouring inputs from the two eye images provide

temporal contiguity [66], [67]. The weight-sharing technique also reduces the number of network parameters to avoid overfitting due to the small number of training samples.

B. MULTI-SCALE CONTEXT FEATURE EXTRACTOR NETWORK (MSCFE-NET)

MSCFE-Net consists of the DRMC-Net and the PPR-AC models, as indicated by the straw-coloured bands in Figure 1. The proposed module extracts multi-scale semantic information of context features and produces high-level maps by first encoding that information, then suppressing noise pixels and finally, enhancing the receptive field. The PPR-AC model consists of a residual pyramid pooling model and an attention convolutional network.

1) DILATED CONVOLUTIONAL (DC) MODEL

Deep convolutional neural networks (DCNNs) are adopted in object detection and semantic segmentation tasks to extract regions with distinguishing features of interest from

images [68], [69], [70]. However, the repeated sequence of pooling layers in these networks negatively affects the spatial resolution of the feature maps by a factor of 32 across each direction in modern DCNNs [54], [60], [71]. Deconvolutional layers [1], [50], [59], [69], [72], [73] are applied to be a partial solution, as utilised in [69], but this requires additional time and memory [74]. To address this, the dilated convolution (DC) technique (also known as “atrous” convolution) was used for semantic segmentation [74], as explained later in [75], [76], and [77]. The atrous rates were modified to obtain long-range information [78], dense up-sampling convolution and hybrid dilated convolution (DUC-HDC) [79]. To increase the accuracy of the images, this segmentation technique [70], [80], [81] uses captions on the images [82], includes depth information [83] and adds in video animation [84].

In this study, DC was applied to efficiently calculate the undecimated wavelet transforms in the technique mentioned in [85] and as applied previously in the context of a DCNN by [68], [77], and [86]. This technology permits us to calculate the responses of any layer with any required accuracy. It can be seamlessly combined with the training phase of the network but can also be implemented after a network is trained. Our proposed MS-Net model applies four DC rates, as shown in Figure 2. DC is a robust tool that controls the resolution of features calculated by DCNNs, modifies the fields in the filter view to capture multi-scale information and simplifies the traditional convolution process. Mathematically, DC is applied to the two-dimensional signal output $F(x)$ at any location (x) , with input signal $M(x)$ and convolution filter $N(K)$, whose length is K and dilation rate is (r) , as illustrated in Equation (1):

$$F(x) = \sum_{x=1}^K M[x+r.K] * N(K) \tag{1}$$

Note that conventional convolution is a state in which the dilation rate $r = 1$. We refer interested researchers to [74] for more details about the DC model.

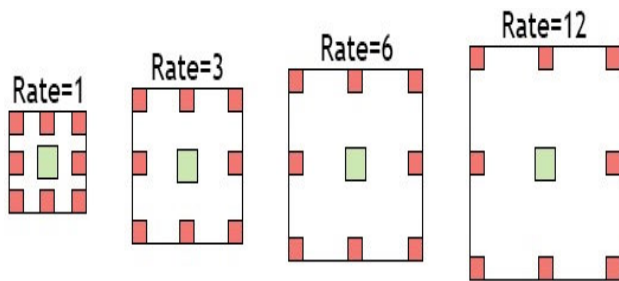


FIGURE 2. The proposed dilated (atrous) convolution rates.

2) RESIDUAL STRUCTURE (RS) MODE

DCNNs typically integrate the features of input objects at various levels. The deeper the network, the richer the extracted features and the stronger the ability to learn. Therefore, a network design with a more profound architecture should be able to extract richer features. However, increasing the number

of layers will not enhance performance but instead might cause information loss, vanish the gradient and degrade network performance when a deep network is trained [71], [87]. To prevent degradation, He et al. [71] and Zhang et al. [87] applied the idea of residual learning, which was introduced in the Residual Network (ResNet) technique. This means that the network is no longer learning total outputs. Residual learning allows the number of layers in the deep network to be raised in order to generate a map of the region of interest while avoiding degradation problems. This method effectively simplifies the learning objective and reduces the computational complexity, $F(x)$, by implementing a feature map function, $Y(x)$, where x is the network input as described in Equation (2):

$$F(x) = Y(x) - x \tag{2}$$

The residual block provides several advantages. It supports information flow and facilitates parameter optimisation by permitting the gradient to readily back-propagate over the associations. Furthermore, the residual learning structure has no extra parameters, which implies that enhancements are achieved without increasing the model’s computational complexity.

Figure 3(a) illustrates the architecture of the original ResNet network [71], which consists of multiple blocks of residual learning. Our proposed method avoids gradient vanishing and information loss without increasing the deep network’s parameters and computational complexity, as illustrated in Figure 3(b). In addition, the skip connection step that we introduced efficiently protects the original features during sending and prevents information loss from the network.

3) DILATED RESIDUAL MULTI-CONVOLUTIONAL NETWORK MODULE (DRMC-NET)

Residual connections [71] and inception [88] are two classical architectures based on DL. Inception [88] introduces several designs with different receptive fields to extend the convolutional network architecture, which enhances the performance of the networks at lower computational costs. ResNet was trained for profound networks [71]. It uses shortcut connections to avoid degradation problems such as vanishing and exploding gradients. Szegedy et al. [41] suggested an inception-ResNet architecture that combines inception [88] and ResNet [71], thereby inheriting the benefits of both methods.

Using DC [85] and Inception-ResNet [86], [89], the DRMC-Net module is proposed to encode high-level feature maps. DRMC-Net employs four parallel dilated convolutions of 3×3 kernels with different dilation rates (r) as illustrated in Figure 4. In the proposed model, DRMC-Net provides four r s, namely 1, 3, 6 and 12. The branch with the highest rate of 12 contains a 1×3 convolution and a 3×1 convolution. The technique is akin to inception architectures that use several different receptive fields. In addition, the current proposal employs a 1×1 convolution at the end of each branch

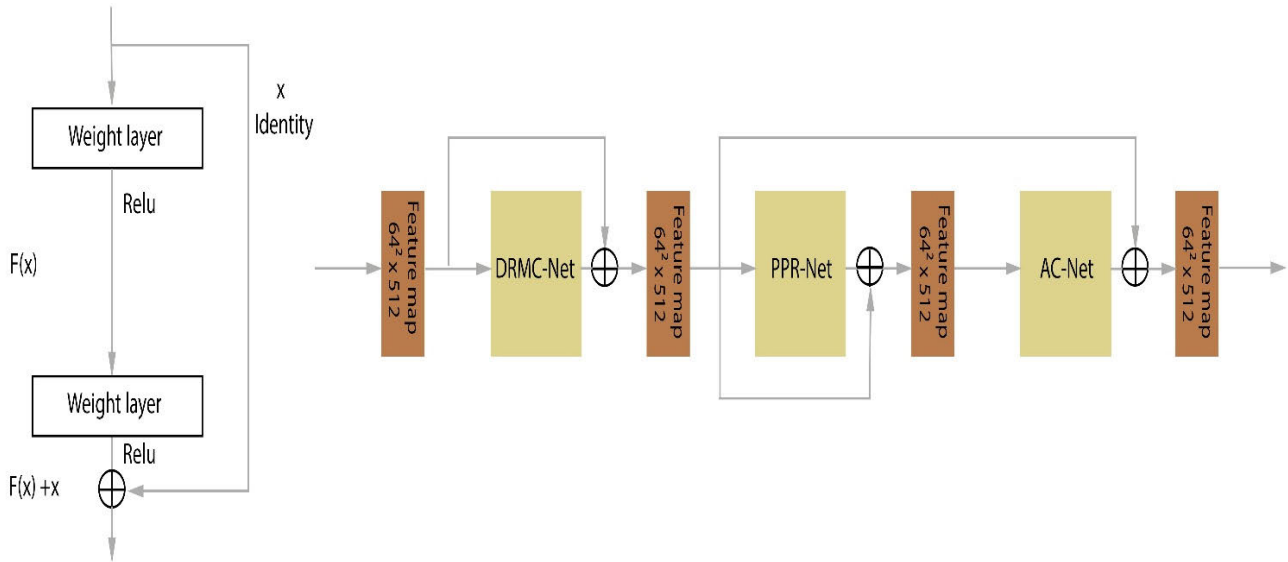


FIGURE 3. Block of residual learning: (a) the original block of residual learning, (b) our proposed multi-scale block of residual learning.

to improve linear activation and computational efficiency. The final design is provided with a shortcut mechanism (such as ResNet) to add the resulting features to the original ones. By merging the DCs at various dilation rates, the DRMC-Net model can extract semantic features from objects of various sizes and enhance the receptive field while maintaining computational cost.

4) PYRAMID POOLING RESIDUAL MODEL BASED ON ATTENTION CONVOLUTIONAL (PPR-AC) MODULE

The IS is a complicated process when eyeglasses are present because it causes large differences in the intensity of the pixels. The eyeglasses result in blurriness, specular reflections, occlusions, scratches and shadows. Also, it is difficult to extract features from multi-scale receptive fields in high-level feature maps, which degrades performance in problematic situations of significant scale differences. The Pyramid Pooling Module (PPM) [40] is an effective technique for extracting multi-scale features from a single eye image. A representation of the distinguishing features in the region of interest can be obtained using the pyramid pooling residual model based on the attention convolutional (PPR-AC) module to improve multi-scale feature extraction. The general architecture of the suggested PPR-AC is illustrated in Figure 5. PPR-AC consists of a Pyramid Pooling Residual Network model (PPR-Net) with skip connections and an Attention Convolutional Network (AC-Net). The idea of extracting multi-scale features using PPM is to enhance the capturing of features from the various receptive fields of the object. To avoid vanishing of the gradient and information loss, a skip connection step for both phases of PPR-AC was added. The AC-Net Module was used to improve accuracy at a lower level of computational complexity to selectively emphasise beneficial features. The feature map that was obtained from the DRMC-Net module is

fed into the PPR-AC module as input. Figure 5(a) illustrates PPR-Net, which uses four different sizes of pooling, namely 16×16 , 8×8 , 4×4 and 2×2 . To get different representations of a subregion, a 1×1 convolution is used to decrease the dimensions of the original features, followed by the up-sampling and concatenation of the layers to generate a representation of the feature map. Finally, this representation is fed into AC-Net to extract the final distinguishing features, which contain both local and global information of the context, as shown in Figure 5(b).

- PYRAMID POOLING RESIDUAL NETWORK MODEL (PPR-NET) To avoid information loss of the context and improve the extraction of multi-scale features from high-level semantic feature maps, PPR-Net is proposed. The size of the receptive field determines the information of the required context. It is advantageous to extract deeper semantic features for better map predictions. The proposed PPR-Net model encodes global information of the context under four different-sized pyramid scales. As shown in Figure 5(a), PPR-Net contains four levels with fixed-sized average pooling: 16×16 , 8×8 , 4×4 and 2×2 . A 1×1 convolution layer and up-sampling are also used. The global feature weightage, restricted to the pyramid pooling level, can employ a 1×1 convolution to decrease the feature dimension to $1/N$ of the original dimension, where N represents the pyramid level. The next step is to directly apply sample operations on the low-dimensional feature maps from the 1×1 convolution layer to obtain a feature map with the same dimensions as the original by bilinear interpolation. Finally, different levels of features are linked with the original features and the residual idea in [71] is applied to produce the final feature.

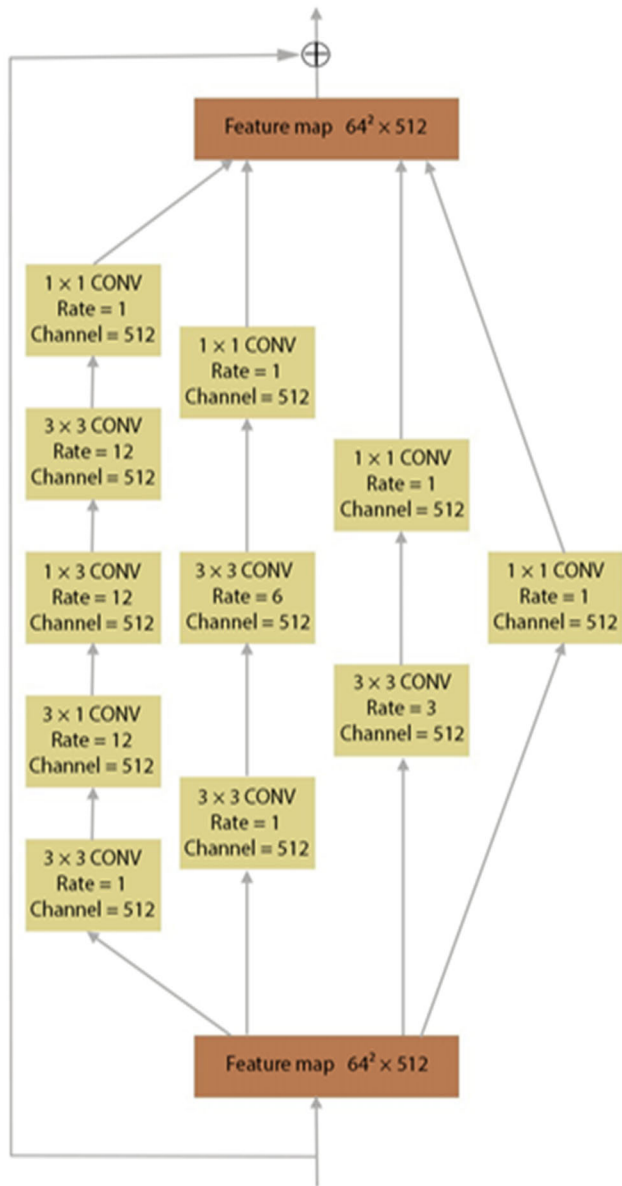


FIGURE 4. The proposed dilated residual multi-convolutional network module (DRMC-Net) contains four dilation rates, which are 1, 3, 6 and 12. There are multiple parallel convolution layers, namely 3 × 3, 3 × 1 and 1 × 3 kernels, within a residual learning structure. Therefore, DRMC-Net can extract semantic feature maps of different scales and enhance the receptive field.

- **ATTENTION CONVOLUTIONAL NETWORK MODEL (AC-NET)** Excessive noise in eye images due to eyeglasses causes the iris information to be unclear, which results in difficult detection and segmentation of the iris region. The Attention Convolutional Network (AC-Net) model is designed to improve accuracy of our proposed network whilst requiring less computational complexity to extract the distinguishing features before it is decoded. The main aim of AC-Net is to pick out important information related to the objective of the task among the various types of information from the

region of interest. Our experiments showed that network performance was enhanced by including AC-Net in the architecture of the suggested network. The AC-Net used in this research includes two convolutional layers, one average pooling layer, one up-sampling layer and a sigmoid layer. All layers use zero-padding and ReLUs. The architecture and main parameters of the AC-Net model are illustrated in Figure 5(b).

C. THE FEATURE DECODER NETWORK

The network’s decoding process is performed by MS-Net, comprising two branches with the same configurations, as shown in the saffron-coloured part of Figure 1. The decoding process consists of the convolution operation, dense connection and up-sampling. It decodes the high-level features extracted by the encoder and MSCFE-Net. The dense connections block was employed to increase training speed and reduce the required number of parameters in order to enhance performance. The skip connection was included to connect the encoded and decoded elements for the purpose of addressing the low resolution of high-level semantic features and information loss across the consecutive convolutional layers and during repeated up-sampling operations. Finally, we applied a 1 × 1 convolution layer to minimise the feature channels and sigmoid function to produce a pair of output maps. The effectiveness of MS-Net is more apparent in images with bad illumination and complex noise from eyeglasses such as blurriness, reflections, shadows, scratches, dirt and other factors. It is for this reason that we included the MS-Net structure in the decoder module, as it can further enhance segmentation.

D. LOSS FUNCTIONS

During the training of the proposed MS-Net framework for IS, binary cross entropy was proposed to depict loss function segmentation, which is described as follows:

$$LF_{bce} = \sum_{i=1}^J [(M_i \log N_i) + (1 - M) \log(1 - N_i)] \quad (3)$$

where $N_i \in [0, 1]$ indicates the output of the last layer of the proposed framework, $M_i \in [0, 1]$ indicates the ground truth labels of the iris image and J indicates the number of pixels in each iris image.

IV. EXPERIMENT SETUP

The following three sections describe the databases, implementation details and evaluation metrics.

A. DATABASES

In this study, the CASIA-Iris.V4-1000⁽¹⁾ and UBIRIS.V2⁽²⁾ databases were used to evaluate the accuracy of the proposed techniques. Both of these databases are utilised to evaluate the performance of the IS task and the IRS. Details of these two databases are as follows:

¹[http://www.cbsr.ia.ac.cn/china/Iris% 20Databases% 20CH.asp](http://www.cbsr.ia.ac.cn/china/Iris%20Databases%20CH.asp)
²<http://iris.di.ubi.pt/ubiris2.html>

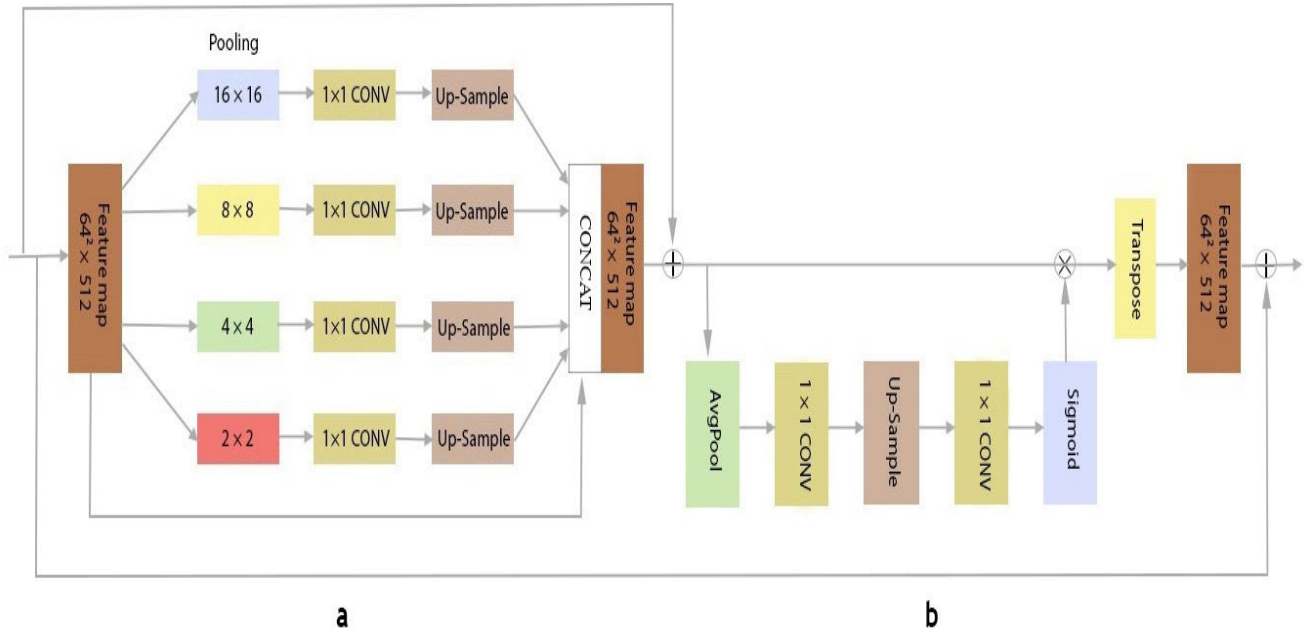


FIGURE 5. An illustration of the proposed pyramid pooling residual model based on the attention convolutional network (PPR-AC). (a) Pyramid Pooling Residual Network Model (PPR-Net), (b) Attention Convolutional Network Model (AC-Net).

- 1) CASIA-IRIS.V4-1000 is a subset of the CASIA-Iris.V4 [38] database provided by the Center for Biometrics and Security Research (CBSR). The dataset contains 20,000 NIR images taken from 1,000 subjects, composed of the left and right eyes with 1 to 10 samples each. The resolution for all images is 640×480 pixels and captured with an IKEMB-100 camera. All images are in (.jpeg) format. The database includes samples with eyeglasses and specular reflections. Additionally, for subjects who wear eyeglasses, it includes images with eyeglasses and without eyeglasses as a standard for measuring performance. A subset of 700 images for the training phase and 400 for the testing phase was used.
- 2) UBIRIS.V2 [90] is an iris database captured in an unconstrained (non-ideal) environment using visible light illumination. It contains 2 sets of images for 261 subjects (left and right eyes) totalling 522 or (2×261). The total number of samples in this database is 11,102 VIS images captured using a Canon EOS DSLR camera with a focal range of 400 mm, an exposure time of $1/200$ s and an ISO of 1600. All images are in (.tiff) format. The resolution of each image is 800×600 pixels. A subset of 700 images for the training phase and 500 for the testing phase was used. The database includes samples with eyeglasses and specular reflections.

B. IMPLEMENTATION DETAILS

The simulation platform for conducting our experiment is PyCharm, using Keras with TensorFlow backend.

The experiment was implemented using a laptop with Intel (R) Core (TM) i7-9750H processor, Nvidia GeForce GTX 1660Ti GPU(6GB-GDDR6) Graphics Processing Unit, 32GB-DDR4 RAM and 512GBNVMe SSD, running on a 64-bit Windows10 Home OS. The proposed networks were trained on the CASIA-Iris.V4-1000 and UBIRIS.V2 databases. The number of epochs was set at 500, with 10 steps per epoch for both U-Net and our proposed network. During the training phase, augmentation was not used. The Adam optimisation process was employed with a learning rate of 10^{-4} and zero decay. Each network was then tested on two iris image settings, with and without eyeglasses. The images with eyeglasses posed several challenges, because of blurriness, reflections, dirt, scratches, shadows and other issues.

C. EVALUATION METRICS

The performance evaluation of the proposed IS method and the comparisons were carried out using the following five evaluation metrics: Accuracy (ACC) whose value lies between 0 and 1 as defined by Equation (4), the Equal Error Rate (EER) whose value ranges from 0 to 1 as defined by Equation (5), Precision (Pr) as defined by Equation (6), Recall (Rec) as defined by Equation (7) and F-Measure (FM) as defined by Equation (8). All definitions are based on the true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values.

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

$$EER = \frac{FP + FN}{TP + FP + TN + FN} \quad (5)$$

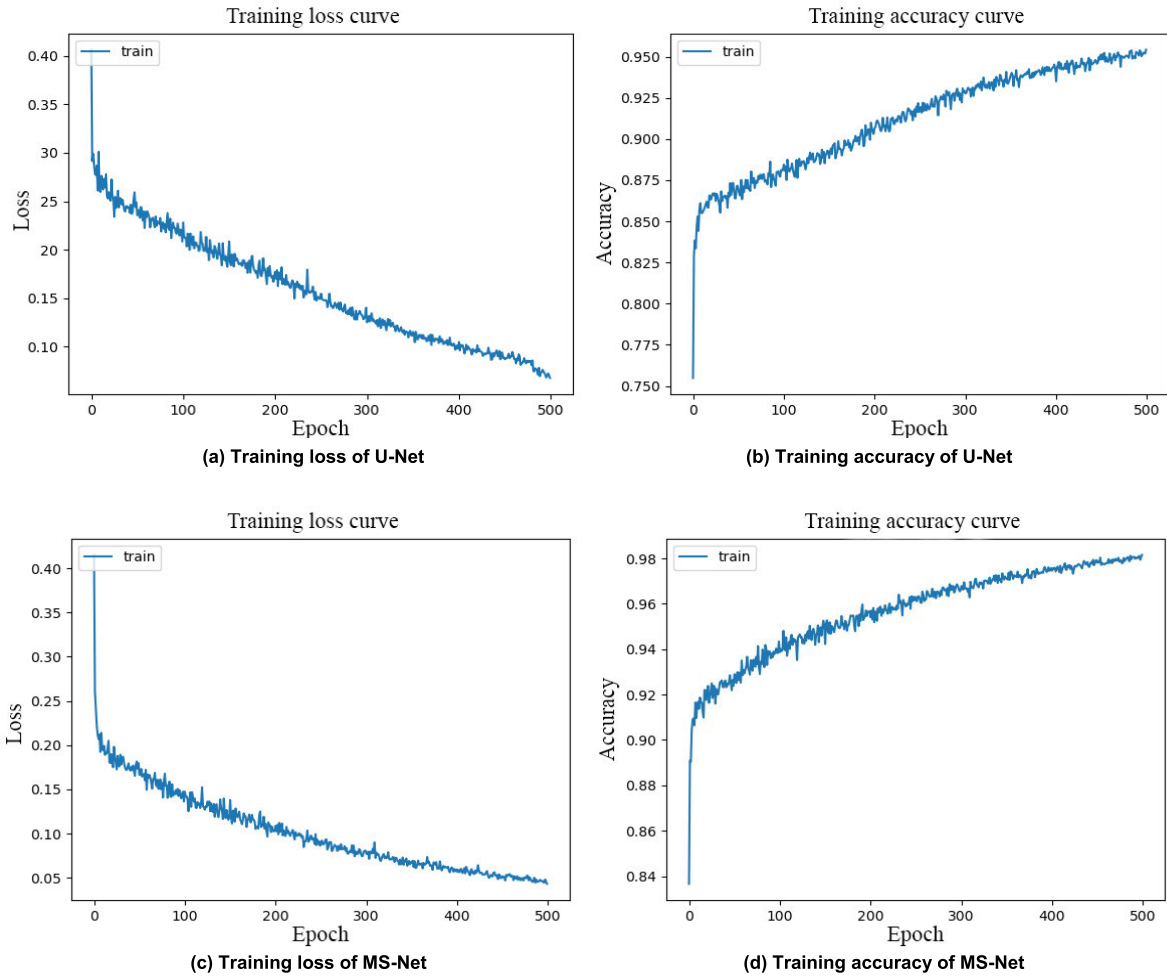


FIGURE 6. The training accuracy (right) and training loss (left) curves for the original U-Net method (a and b) and for our proposed MS-Net method (c and d).

$$Pr = \frac{TP}{TP + FP} \tag{6}$$

$$Rec = \frac{TP}{TP + FN} \tag{7}$$

$$FM = \frac{2 * Pr * Rec}{Pr + Rec} \tag{8}$$

V. EXPERIMENT RESULTS

This part evaluates the segmentation result, measured using five metrics of the proposed MS-Net and the original U-Net on two databases. The three objectives of this evaluation are to measure the accuracy of MS-Net, to compare the accuracy of MS-Net with that of U-Net [15] and to compare MS-Net with a number of state-of-the-art methods.

A. EVALUATION OF THE PROPOSED MS-NET METHOD

The segmentation performance was evaluated against three main criteria: with eyeglasses, without eyeglasses and both with and without eyeglasses (overall). The proposed MS-Net method was compared with the original U-Net method [15]. The original U-Net is an open-source algorithm and was

tested using the same CASIA-Iris.V4-1000 and UBIRIS.V2 databases.

Figure 6 shows the changes in training loss (a and c) and training accuracy (b and d) in the original U-Net (a and b) and the proposed MS-Net (c and d). The *x*-axis represents the number of epochs while the *y*-axis represents the loss and the accuracy of the training phase per batch of epochs between 0 to 500. A comparison between the original U-Net algorithm and the suggested MS-Net method shows that the accuracy of MS-Net is higher and the loss lower. This suggests that the MS-Net has improved the convergence rate and accuracy of the training phase, besides achieving significant optimisation.

Figure 7 presents a comparison of the IS without eyeglasses obtained from the CASIA-Iris.V4-1000 and UBIRIS.V2 databases. It shows the original eye images (a and g) followed by the three resulting images obtained, which are the ground truth (b and h), segmentation by U-Net (c and i), the false positive and negative pixels of the U-Net segmentation results are presented in green and blue colours, respectively (d and j), segmentation by the proposed MS-Net

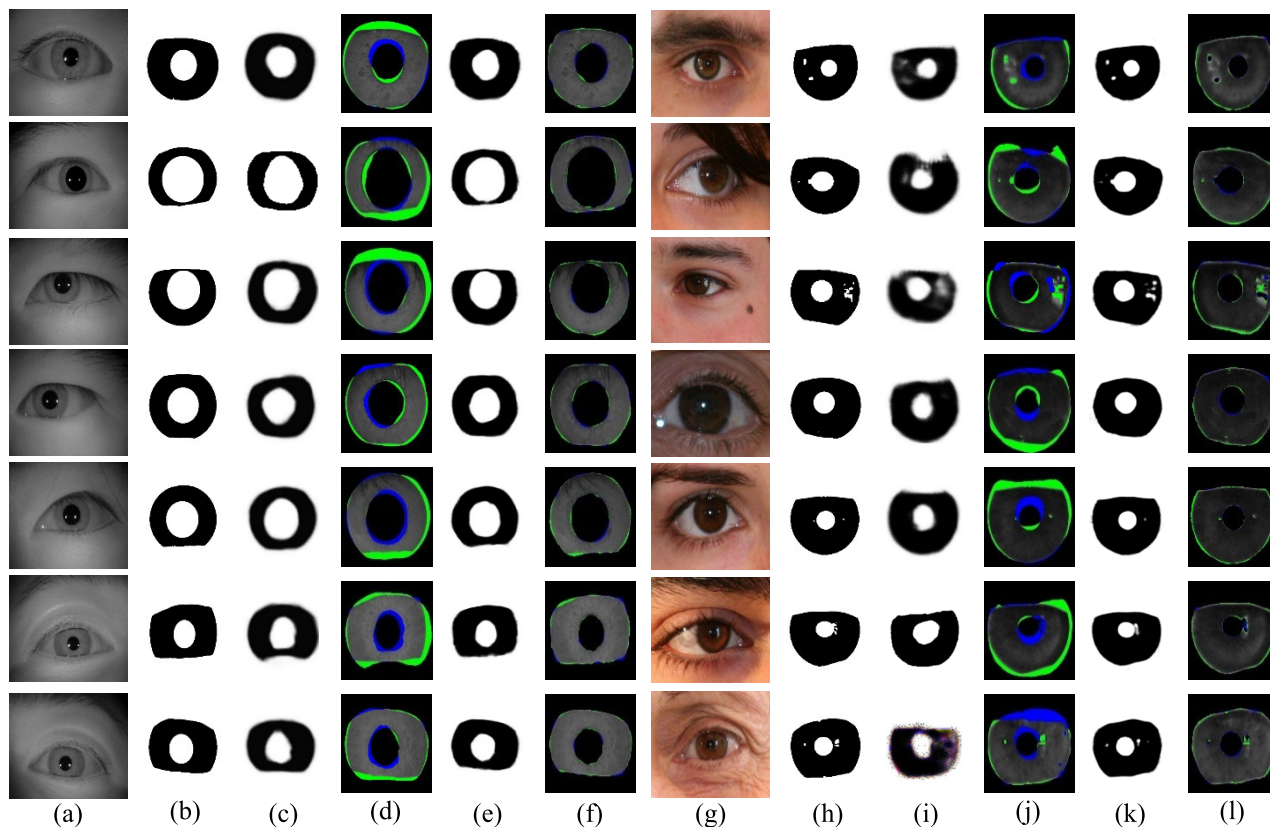


FIGURE 7. Visualisation of the iris segmentation without eyeglasses: Original eye images from the CASIA-Iris.V4-1000 and UBIRIS.V2 databases (a and g respectively), ground truth images (b and h), segmentation by U-Net (c and i), the false positive and negative pixels of the U-Net segmentation results are presented in green and blue colours, respectively (d and j), segmentation by the proposed MS-Net method (e and k), and the false positive and negative pixels of the MS-Net segmentation results are presented in green and blue colours, respectively (f and l).

method (e and k) and the false positive and negative pixels of the MS-Net segmentation results are presented in green and blue colours, respectively (f and l). The images produced by MS-Net were shown to be more accurate than those obtained by other techniques.

Figure 8 shows the same layout as Figure 7 but with eyeglasses. These segmentation results were also obtained using samples taken from the CASIA-Iris.V4-1000 and UBIRIS.V2 databases. The results show that the proposed MS-Net method is more reliable in performing IS for samples with eyeglasses as compared with U-Net method.

Generally, eyeglasses are a significant source of noise in non-cooperative and unconstrained environments of iris identification. The following effects were considered to be noise in iris images with eyeglasses [38], [39]: specular reflections, occlusions, shadows, dirt, scratches and blurriness. In addition, the edges/frames of eyeglasses may be misidentified as limbus/pupillary boundaries by public IS methods such as the open-source OSIRIS [55]. As a result, these problems frequently lead to erroneous segmentation of the iris region, which further degenerates the performance accuracy of the IRS. For this reason, the authors propose the MS-Net architecture to improve the accuracy of IS for iris images with eyeglasses. Figure 9 presents the results of IS obtained

by the U-Net algorithm compared to the proposed MS-Net method using challenging samples due to eyeglasses, taken from the CASIA-Iris.V4-1000 and UBIRIS.V2 databases. As can be observed, certain iris images were very difficult to process due to noise from eyeglasses. When plotting these results, such challenging examples of the IS from these two databases had to be further processed. Results showed that the images underwent inaccurate segmentation of the iris region using the original U-Net [15] algorithm. In contrast, the proposed MS-Net method successfully segmented the iris in those challenging samples and produced higher accuracy IS, as shown in Figure 9. MS-Net detects the iris region with high accuracy against challenges with eyeglasses caused by blurriness, specular reflections, dirt, scratches and other factors.

There were samples with eyeglasses that were poorly segmented by both the original U-Net algorithm and the proposed MS-Net method, as shown in Figure 10. The noise generated by the eyeglasses, such as specular reflection or blurring, poses a significant challenge to segmenting the iris region. Nevertheless, the IS results obtained by the suggested MS-Net are better than those of the original U-Net. Iris images in the UBIRIS.V2 database are visible light (VI) images, which means that challenges could be caused by low

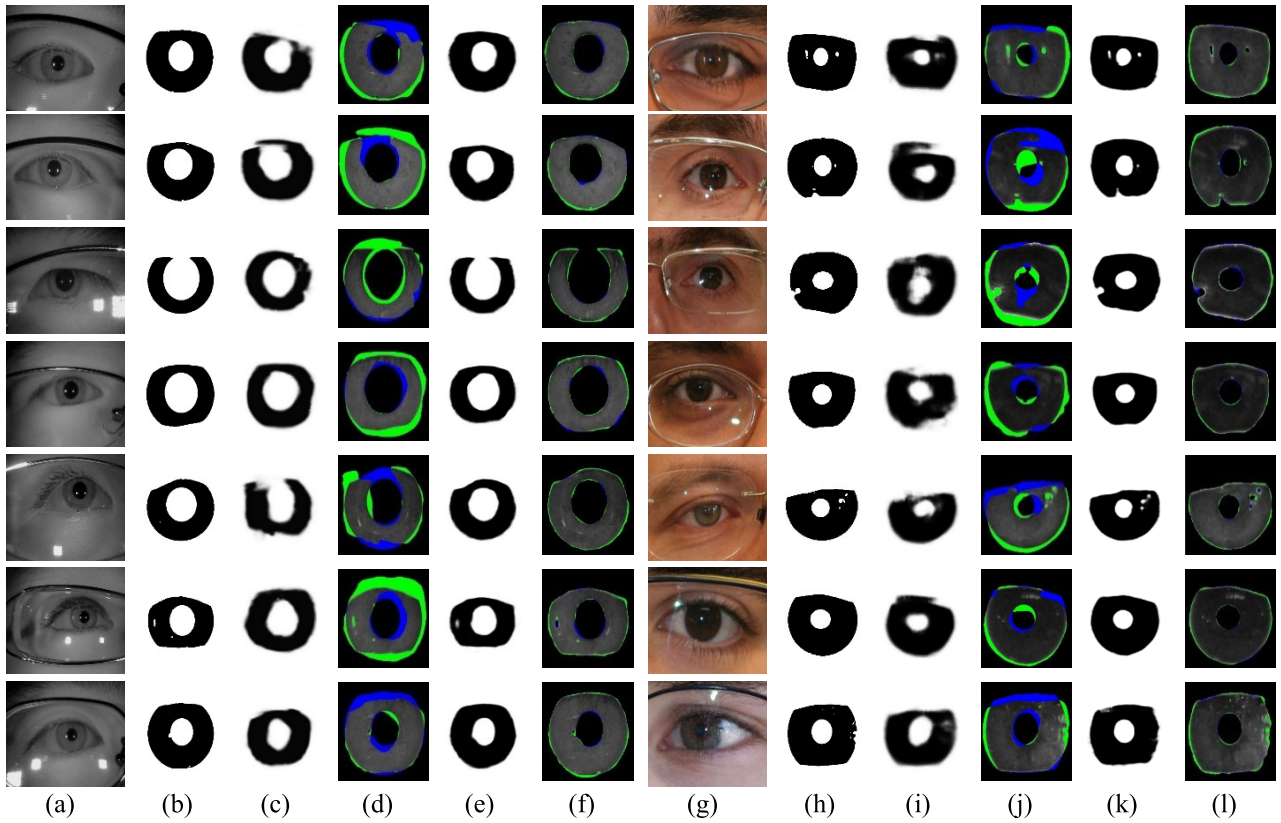


FIGURE 8. Visualisation of the iris segmentation with eyeglasses: Original eye images from the CASIA-Iris.V4-1000 and UBIRIS.V2 databases (a and g respectively), ground truth images (b and h), segmentation by U-Net (c and i), the false positive and negative pixels of the U-Net segmentation results are presented in green and blue colours, respectively (d and j), segmentation by the proposed MS-Net method (e and k), and the false positive and negative pixels of the MS-Net segmentation results are presented in green and blue colours, respectively (f and l).

TABLE 2. A comparison of the performance of iris segmentation with/without eyeglasses by u-net against the proposed ms-net on the casia-iris.v4-1000 and ubiris.v2 databases.

Database	Method	Pr	Rec	FM	EER	ACC		
CASIA-Iris.V4-1000	Without eyeglasses	U-Net	0.9112	0.9053	0.9081	0.09459	0.90541	
		Proposed MS-Net	0.9765	0.97631	0.9764	0.02361	0.97639	
	With eyeglasses	U-Net	0.85883	0.85321	0.85641	0.14668	0.85332	
		Proposed MS-Net	0.96652	0.96483	0.96612	0.03422	0.96578	
UBIRIS.v2	Without + with eyeglasses (overall)		U-Net	0.88504	0.87925	0.88225	0.12065	0.87935
	Without eyeglasses	Proposed MS-Net	0.97156	0.97057	0.97129	0.02891	0.97109	
		U-Net	0.90262	0.89756	0.89965	0.10224	0.89776	
	With eyeglasses	Proposed MS-Net	0.97086	0.97052	0.97061	0.02946	0.97054	
		U-Net	0.84314	0.83852	0.84065	0.16118	0.83882	
	Without + with eyeglasses (overall)		Proposed MS-Net	0.9527	0.95184	0.95253	0.04798	0.95202
		U-Net	0.87288	0.86804	0.87015	0.13171	0.86829	
		Proposed MS-Net	0.96178	0.96118	0.96157	0.03872	0.96128	

contrast. Even if we relabelled the iris images in the database with the human eye, some are still suspect and require further investigation.

As shown in Fig. 7, Fig. 8, Fig. 9 and Fig10, there are two types of errors called the false positive (FP) and the false negative (FN). When a non-iris pixel is predicted by both

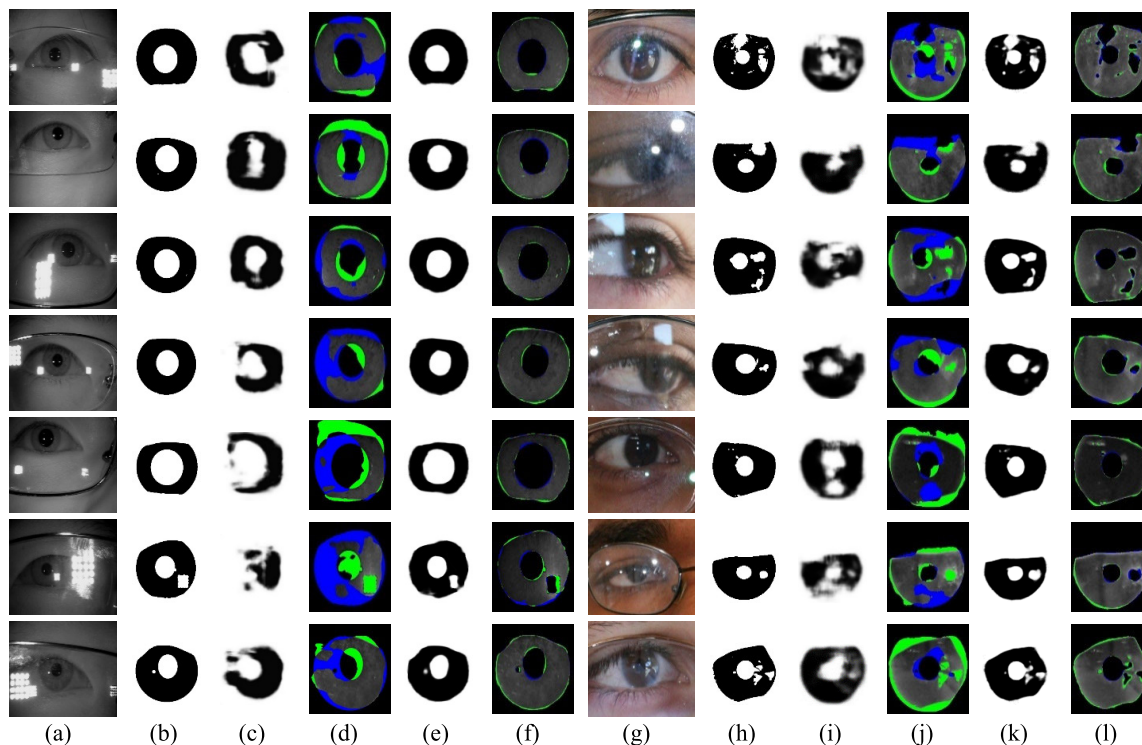


FIGURE 9. Poor segmentation by the original U-Net method for some challenging samples: Original eye images from the CASIA-Iris.V4-1000 and UBIRIS.V2 databases (a and g respectively), ground truth images (b and h), segmentation by U-Net (c and i), the false positive and negative pixels of the U-Net segmentation results are presented in green and blue colours, respectively (d and j), segmentation by the proposed MS-Net method (e and k), and the false positive and negative pixels of the MS-Net segmentation results are presented in green and blue colours, respectively (f and l).

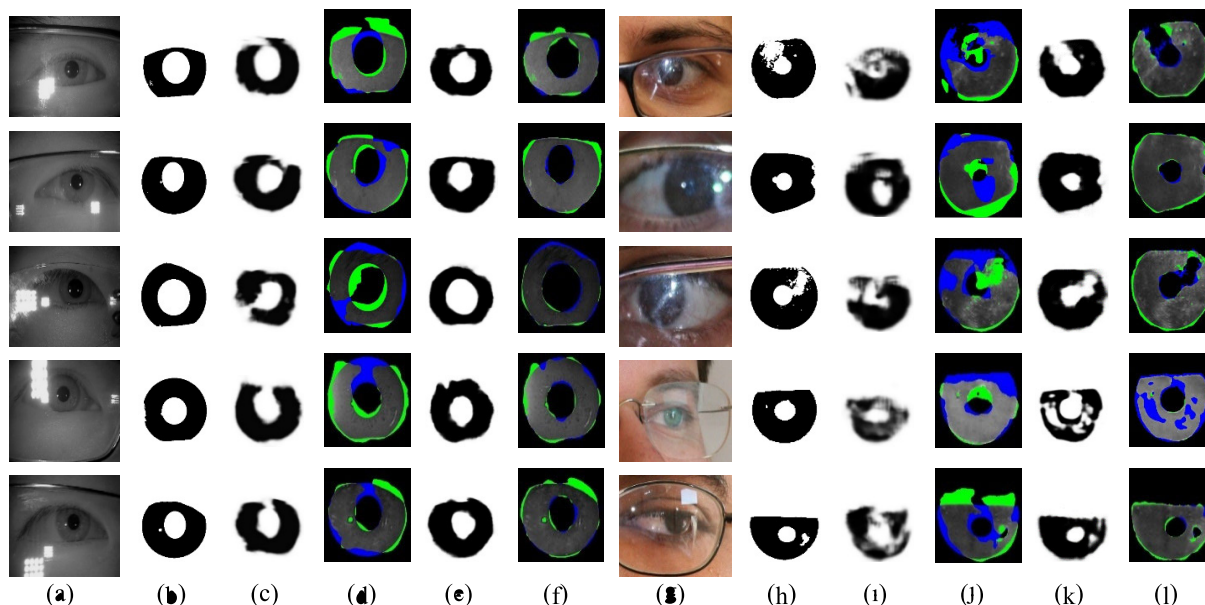


FIGURE 10. Poor segmentation by both the original U-Net method and the proposed MS-Net method for some challenging samples: Original eye images from the CASIA-Iris.V4-1000 and UBIRIS.V2 databases (a and g respectively), ground truth images (b and h), segmentation by U-Net (c and i), the false positive and negative pixels of the U-Net segmentation results are presented in green and blue colours, respectively (d and j), segmentation by the proposed MS-Net method (e and k), and the false positive and negative pixels of the MS-Net segmentation results are presented in green and blue colours, respectively (f and l).

the original U-Net method and the proposed MS-Net method to be an iris pixel, the false positive (FP) error (shown by the green colour) happens, whereas the false negative (FN)

error (shown by the blue colour) happens when an iris pixel is predicted by both the original U-Net method and the proposed MS-Net method to be a non-iris pixel. The FP error is caused

by scratches and blurring noises due to eyeglasses or similar non-iris region pixels that have a pixel value closer to the pixel value of the iris region, while the FN error is caused by reflection noises due to eyeglasses or the dark region.

B. COMPARISON OF THE PROPOSED MS-NET METHOD WITH THE ORIGINAL U-NET NETWORK

The proposed MS-Net method was evaluated and compared with the original U-Net method [15] in terms of F-Measure (FM), Accuracy (ACC), Precision (Pr), Recall (Rec) and Equal Error Rate (EER) as defined by Equations (4) through (8). Table 2 lists the comparison between the original U-Net and the suggested MS-Net on the CASIA-Iris.V4-1000 and UBIRIS.V2 databases. The MS-Net method achieved better results compared with the original U-Net method in both databases with an error rate of 0.011 and 0.033 on CASIA-Iris.V4-1000 and UBIRIS.v2, respectively. It is observed that the suggested MS-Net method always achieved excellent IS results when dealing with iris images taken in both constrained (ideal) and unconstrained (non-ideal) environments, demonstrating its exceptional generalisation capacity to various databases.

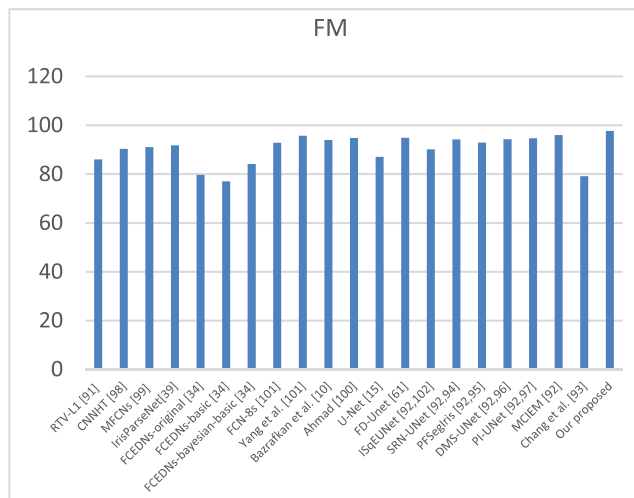


FIGURE 11. Comparison of FM results (%) between our proposed MS-Net method and other earlier iris segmentation methods on the UBIRIS.v2 database.

C. COMPARISON OF THE PROPOSED MS-NET WITH EARLIER STATE-OF-THE-ART METHODS

Figure 11 is a graphical representation of the performance of several IS techniques. Our proposed MS-Net is compared with previous state-of-the-art methods that were based on the traditional approach [91] and CNN methods for IS [10], [15], [34], [39], [61], [92], [93], [94], [95], [96], [97], [98], [99], [100], [101], [102]. FM testing was conducted to compare the performance of MS-Net with earlier methods, utilising the same open database, the UBIRIS.v2. It was observed that the MS-Net method achieves better FM results compared to the other iris segmentation methods tested.

VI. CONCLUSION

The accuracy of an IS technique is essential in determining the accuracy of an IRS. Our proposed multi-segmentation network (MS-Net) framework based on a DL approach enhances the IS of images with eyeglasses. This technique has been tested, evaluated and compared with other techniques. MS-Net was effective in improving the accuracy of inter-connected tasks, leading to multi-task learning by capturing the differences and similarities between multiple samples. In terms of architecture, MS-Net consists of a feature encoder network, a feature extractor by a multi-scale context network (MSCFE-Net) and a feature decoder network. In MSCFE-Net, the DRMC-Net model was adopted to encode high-level feature maps and obtain deeper features using four convolution branches with multi-scale dilated convolution. PPR-AC was also introduced to improve extraction features of the multi-scale receptive field in high-level feature maps and to avoid drops in performance when faced with problems of significant scale differences. The proposed method contains dense connections within the feature decoder network in each branch. This dense connection reduces training difficulty as only a few training samples of eye images are used. The results of experiments conducted on the proposed MS-Net method showed that it is able to increase the accuracy of IS, particularly with cases that involve eyeglasses.

In future work, the researchers would be designing more efficient techniques by exploiting spatial associations between the iris mask, the outer iris edge, and the inner iris edge to enhance the iris segmentation accuracy. Also, the researchers would further improve the speed of our proposed MS-Net method by designing heterogeneous networks that combine the speed of traditional methods and the power of the deep learning approach with different forms of ResNet to make it more efficient for iris segmentation tasks. Additionally, our method would be used for different segmentation tasks, such as medical image segmentation (brain anatomy, melanoma, vessel segmentation, cancer segmentation [skin, liver, breast, and lung cancer segmentations], fractures, etc), dental image segmentation, market segmentation, and facial segmentation.

REFERENCES

- [1] M. Arsalan, D. S. Kim, M. B. Lee, M. Owais, and K. R. Park, "FRED-Net: Fully residual encoder-decoder network for accurate iris segmentation," *Expert Syst. Appl.*, vol. 122, pp. 217–241, May 2019, doi: 10.1016/j.eswa.2019.01.010.
- [2] S. Lian, Z. Luo, Z. Zhong, X. Lin, S. Su, and S. Li, "Attention guided U-Net for accurate iris segmentation," *J. Vis. Commun. Image Represent.*, vol. 56, pp. 296–304, Oct. 2018.
- [3] M. Arsalan, R. Naqvi, D. Kim, P. Nguyen, M. Owais, and K. Park, "IrisDenseNet: Robust iris segmentation using densely connected fully convolutional networks in the images by visible light and near-infrared light camera sensors," *Sensors*, vol. 18, no. 5, p. 1501, May 2018.
- [4] A. Pando, "Beyond security: Biometrics integration into everyday life," Forbes Community Council, 2017.
- [5] K. W. Bowyer, K. P. Hollingsworth, and P. J. Flynn, "A survey of iris biometrics research: 2008–2010," in *Handbook of Iris Recognition*. Cham, Switzerland: Springer, 2013, pp. 15–54.

- [6] S. Viriri and J. R. Tapamo, "Integrating iris and signature traits for personal authentication using user-Specific Weighting," *Sensors*, vol. 12, no. 4, pp. 4324–4338, Mar. 2012.
- [7] A. K. Jain, S. S. Arora, K. Cao, L. Best-Rowden, and A. Bhatnagar, "Fingerprint recognition of young children," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 7, pp. 1501–1514, Jul. 2017.
- [8] K. Bonnen, B. F. Klare, and A. K. Jain, "Component-based representation in automated face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 1, pp. 239–253, Jan. 2013.
- [9] R. Devi and P. Sujatha, "A study on biometric and multi-modal biometric system modules, applications, techniques and challenges," in *Proc. Conf. Emerg. Devices Smart Syst. (ICEDSS)*, Mar. 2017, pp. 267–271.
- [10] S. Bazrafkan, S. Thavalengal, and P. Corcoran, "An end to end deep neural network for iris segmentation in unconstrained scenarios," *Neural Netw.*, vol. 106, pp. 79–95, Oct. 2018, doi: [10.1016/j.neunet.2018.06.011](https://doi.org/10.1016/j.neunet.2018.06.011).
- [11] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric recognition: Security and privacy concerns," *IEEE Secur. Privacy*, vol. 1, no. 2, pp. 33–42, Mar. 2003.
- [12] J. Daugman, "Information theory and the IrisCode," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 2, pp. 400–409, Feb. 2016.
- [13] Y.-H. Li, P.-J. Huang, and Y. Juan, "An efficient and robust iris segmentation algorithm using deep learning," *Mobile Inf. Syst.*, vol. 2019, pp. 1–14, Jan. 2019, doi: [10.1155/2019/4568929](https://doi.org/10.1155/2019/4568929).
- [14] Z. Zhao and A. Kumar, "Towards more accurate iris recognition using deeply learned spatially corresponding features," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3829–3838, doi: [10.1109/ICCV.2017.411](https://doi.org/10.1109/ICCV.2017.411).
- [15] J. Lozej, B. Meden, V. Struc, and P. Peer, "End-to-end iris segmentation using U-Net," in *Proc. IEEE Int. Work Conf. Bioinspired Intell. (IWOB)*, Jul. 2018, pp. 1–6.
- [16] M. Arsalan, H. Hong, R. Naqvi, M. Lee, M. Kim, D. Kim, C. Kim, and K. Park, "Deep learning-based iris segmentation for iris recognition in visible light environment," *Symmetry*, vol. 9, no. 11, p. 263, Nov. 2017.
- [17] V. Nagaraju and L. Fiondella, "A survey of homeland security biometrics and forensics research," in *Proc. IEEE Symp. Technol. Homeland Secur.*, May 2016, pp. 1–7.
- [18] N. Zainal, A. Radman, and M. Ismail, "Iris segmentation for non-ideal images," *Jurnal Teknologi*, vol. 5, p. 74, Jan. 2015.
- [19] Y. Alvarez-Betancourt and M. Garcia-Silvente, "A keypoints-based feature extraction method for iris recognition under variable image quality conditions," *Knowl.-Based Syst.*, vol. 92, pp. 169–182, Jan. 2016, doi: [10.1016/j.knsys.2015.10.024](https://doi.org/10.1016/j.knsys.2015.10.024).
- [20] K. Nguyen, C. Fookes, R. Jillela, S. Sridharan, and A. Ross, "Long range iris recognition: A survey," *Pattern Recognit.*, vol. 72, pp. 123–143, Dec. 2017, doi: [10.1016/j.patcog.2017.05.021](https://doi.org/10.1016/j.patcog.2017.05.021).
- [21] Y. Jung, D. Kim, B. Son, and J. Kim, "An eye detection method robust to eyeglasses for mobile iris recognition," *Expert Syst. Appl.*, vol. 67, pp. 178–188, Jan. 2017.
- [22] Z. Wang, C. Li, H. Shao, and J. Sun, "Eye recognition with mixed convolutional and residual network (MiCoRe-Net)," *IEEE Access*, vol. 6, pp. 17905–17912, 2018.
- [23] J. J. Winston and D. J. Hemanth, "A comprehensive review on iris image-based biometric system," *Soft Comput.*, vol. 23, no. 19, pp. 9361–9384, Oct. 2019.
- [24] M. Liu, Z. Zhou, P. Shang, and D. Xu, "Fuzzified image enhancement for deep learning in iris recognition," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 1, pp. 92–99, Jan. 2020.
- [25] J. R. Malgheet, N. B. Manshor, and L. S. Affendey, "Iris recognition development techniques: A comprehensive review," *Complexity*, vol. 2021, pp. 1–32, Aug. 2021, doi: [10.1155/2021/6641247](https://doi.org/10.1155/2021/6641247).
- [26] D. Freire-Obregón, F. Narducci, S. Barra, and M. Castrillón-Santana, "Deep learning for source camera identification on mobile devices," *Pattern Recognit. Lett.*, vol. 126, pp. 86–91, Sep. 2019.
- [27] S. Dey and D. Samanta, "A novel approach to iris localization for iris biometric processing," *Int. J. Biomed. Biol. Eng.*, vol. 1, pp. 293–304, Jan. 2007.
- [28] T. M. Khan, M. Aurangzeb Khan, S. A. Malik, S. A. Khan, T. Bashir, and A. H. Dar, "Automatic localization of pupil using eccentricity and iris using gradient based method," *Opt. Lasers Eng.*, vol. 49, no. 2, pp. 177–187, Feb. 2011.
- [29] T. M. Khan, D. G. Bailey, M. A. U. Khan, and Y. Kong, "Real-time iris segmentation and its implementation on FPGA," *J. Real-Time Image Process.*, vol. 17, no. 5, pp. 1089–1102, Oct. 2020, doi: [10.1007/s11554-019-00859-w](https://doi.org/10.1007/s11554-019-00859-w).
- [30] M. López, J. Daugman, and E. Cantó, "Hardware–software co-design of an iris recognition algorithm," *IET Inf. Secur.*, vol. 5, no. 1, p. 60, 2011.
- [31] C. Wang, Y. Wang, K. Zhang, J. Muhammad, T. Lu, Q. Zhang, and Q. Tian, "NIR iris challenge evaluation in non-cooperative environments: Segmentation and localization," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Aug. 2021, pp. 1–10.
- [32] I. Nigam, M. Vatsa, and R. Singh, "Ocular biometrics: A survey of modalities and fusion approaches," *Inf. Fusion*, vol. 26, pp. 1–35, Nov. 2015.
- [33] E. Jalilian, A. Uhl, and R. Kwitt, "Domain adaptation for CNN based iris segmentation," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2017, pp. 1–6, doi: [10.23919/BIOSIG.2017.8053502](https://doi.org/10.23919/BIOSIG.2017.8053502).
- [34] E. Jalilian and A. Uhl, "Iris segmentation using fully convolutional encoder–decoder networks," in *Deep Learning for Biometrics*. Cham, Switzerland: Springer, 2017, pp. 55–133.
- [35] N. Susitha and R. Subban, "Reliable pupil detection and iris segmentation algorithm based on SPS," *Cognit. Syst. Res.*, vol. 57, pp. 78–84, Oct. 2019, doi: [10.1016/j.cogsys.2018.09.029](https://doi.org/10.1016/j.cogsys.2018.09.029).
- [36] H. Proença and L. A. Alexandre, "Iris recognition: Analysis of the error rates regarding the accuracy of the segmentation stage," *Image Vis. Comput.*, vol. 28, no. 1, pp. 202–206, Jan. 2010.
- [37] X. You, P. Zhao, X. Mu, K. Bai, and S. Lian, "Heterogeneous noise iris segmentation based on attention mechanism and dense multiscale features," *Laser Optoelectron. Prog.*, vol. 2021, pp. 1–19, Jan. 2021.
- [38] P. Drodowski, F. Struck, C. Rathgeb, and C. Busch, "Detection of glasses in near-infrared ocular images," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 202–208.
- [39] C. Wang, J. Muhammad, Y. Wang, Z. He, and Z. Sun, "Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2944–2959, 2020.
- [40] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.
- [41] J. Daugman, "How iris recognition works," in *The Essential Guide to Image Processing*. Amsterdam, The Netherlands: Elsevier, 2009, pp. 715–739.
- [42] R. P. Wildes, "Iris recognition: An emerging biometric technology," *Proc. IEEE*, vol. 85, no. 9, pp. 1348–1363, May 1997.
- [43] S. A. Sahmoud and I. S. Abuhaiba, "Efficient iris segmentation method in unconstrained environments," *Pattern Recognit.*, vol. 46, no. 12, pp. 3174–3185, Dec. 2013.
- [44] P. C. Chung, C. M. Lyu, J. Liu, and S. S. Yu, "An iris segmentation scheme using Delogne–Kåsa circle fitting based on orientation matching transform," in *Proc. Int. Symp. Comput., Consum. Control*, Jun. 2014, pp. 127–130.
- [45] J. F. Canny, "A computation approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 670–700, Apr. 1986.
- [46] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," in *Readings in Computer Vision*. Amsterdam, The Netherlands: Elsevier, 1987, pp. 714–724.
- [47] A. Gangwar, A. Joshi, A. Singh, F. Alonso-Fernandez, and J. Bigun, "IrisSeg: A fast and robust iris segmentation framework for non-ideal iris images," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2016, pp. 1–8, doi: [10.1109/ICB.2016.7550096](https://doi.org/10.1109/ICB.2016.7550096).
- [48] M. A. M. Abdullah, S. S. Dlay, W. L. Woo, and J. A. Chambers, "Robust iris segmentation method based on a new active contour force with a noncircular normalization," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 12, pp. 3128–3141, Dec. 2017, doi: [10.1109/TSMC.2016.2562500](https://doi.org/10.1109/TSMC.2016.2562500).
- [49] M. Sardar, S. Mitra, and B. Uma Shankar, "Iris localization using rough entropy and CSA: A soft computing approach," *Appl. Soft Comput.*, vol. 67, pp. 61–69, Jun. 2018.
- [50] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [51] S. Bazrafkan and P. Corcoran, "Enhancing iris authentication on hand-held devices using deep learning derived segmentation techniques," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2018, pp. 1–2.

- [52] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [53] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 818–833.
- [54] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [55] N. Othman, B. Dorizzi, and S. Garcia-Salicetti, "OSIRIS: An open source iris recognition software," *Pattern Recognit. Lett.*, vol. 82, pp. 124–131, Oct. 2016.
- [56] J. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," *Int. Comput. Sci. Inst.*, vol. 4, p. 126, Jan. 1998.
- [57] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.
- [58] Y.-H. Li and P.-J. Huang, "An accurate and efficient user authentication mechanism on smart glasses based on iris recognition," *Mobile Inf. Syst.*, vol. 2017, pp. 1–14, 2017.
- [59] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
- [60] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [61] W. Zhang, X. Lu, Y. Gu, Y. Liu, X. Meng, and J. Li, "A robust iris segmentation scheme based on improved U-Net," *IEEE Access*, vol. 7, pp. 85082–85089, 2019.
- [62] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.
- [63] J. Zhang, Y. Li, F. Chen, Z. Pan, X. Zhou, Y. Li, and S. Jiao, "X-Net: A binocular summation network for foreground segmentation," *IEEE Access*, vol. 7, pp. 71412–71422, 2019, doi: [10.1109/ACCESS.2019.2919802](https://doi.org/10.1109/ACCESS.2019.2919802).
- [64] A. Nandy and S. S. Mondal, "Kinship verification using deep Siamese convolutional neural network," in *Proc. 14th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Jun. 2019, pp. 5–9.
- [65] H. Spitzer, K. Kiwitz, K. Amunts, S. Harmeling, and T. Dickscheid, "Improving cytoarchitectonic segmentation of human brain areas with self-supervised Siamese networks," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2018, pp. 1–12.
- [66] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep Siamese convolutional network for optical aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, Oct. 2017.
- [67] H. Chen, C. Wu, B. Du, and L. Zhang, "Change detection in multi-temporal VHR images based on deep Siamese multi-scale convolutional networks," 2019, *arXiv:1906.11479*.
- [68] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," 2013, *arXiv:1312.6229*.
- [69] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
- [70] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.
- [71] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [72] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters—Improve semantic segmentation by global convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1743–1751.
- [73] M. D. Zeiler, G. W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2018–2025.
- [74] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
- [75] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," 2016, *arXiv:1605.06409*.
- [76] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, "Speed/Accuracy trade-offs for modern convolutional object detectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3296–3297.
- [77] G. Papandreou, I. Kokkinos, and P. Savalle, "Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 390–399.
- [78] Z. Wu, C. Shen, and A. van den Hengel, "Bridging category-level and instance-level semantic image segmentation," 2016, *arXiv:1605.06885*.
- [79] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell, "Understanding convolution for semantic segmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 1451–1460.
- [80] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.
- [81] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with Atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.
- [82] G. Wang, P. Luo, L. Lin, and X. Wang, "Learning object interactions and descriptions for semantic image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5235–5243.
- [83] S. Kong and C. Fowlkes, "Recurrent scene parsing with perspective understanding in the loop," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 956–965.
- [84] S. D. Jain, B. Xiong, and K. Grauman, "FusionSeg: Learning to combine motion and appearance for fully automatic segmentation of generic objects in videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2126.
- [85] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in *Wavelets*. Cham, Switzerland: Springer, 1990, pp. 286–297.
- [86] A. Giusti, D. C. Cireşan, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Fast image scanning with deep max-pooling convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 4034–4038.
- [87] X. Zhang, T. Wang, and Y. Yang, "Hyperspectral images classification based on multi-scale residual network," 2020, *arXiv:2004.12381*.
- [88] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
- [89] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 31, 2017, pp. 1–22.
- [90] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1529–1535, Aug. 2010.
- [91] Z. Zhao and A. Kumar, "An accurate iris segmentation framework under relaxed imaging constraints using total variation model," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3828–3836.
- [92] G. Huo, D. Lin, and M. Yuan, "Multi-source heterogeneous iris segmentation method based on lightweight convolutional neural network," *IET Image Process.*, vol. 17, no. 1, pp. 118–131, 2023, doi: [10.1049/ipr2.12621](https://doi.org/10.1049/ipr2.12621).
- [93] C. YT, S. TK, L. YH, and K. WGCW, "Effectiveness evaluation of iris segmentation by using geodesic active contour (GAC)," *J. Supercomput.*, vol. 76, pp. 1–14, Mar. 2018, doi: [10.1007/s11227-018-2450-2](https://doi.org/10.1007/s11227-018-2450-2).
- [94] T. Huijuan, Z. Jiahao, L. Jianxin, L. Jiawei, and D. Linlin, "A low-quality iris image segmentation algorithm based on SRN-UNet," *Acta Photonica Sinica*, vol. 51, Jan. 2022, Art. no. 210006.
- [95] L. Dong, Y. Liu, and X. Zhu, "PFSegIris: Precise and fast segmentation algorithm for multi-source heterogeneous iris," *Algorithms*, vol. 14, no. 9, p. 261, Aug. 2021.
- [96] G. Huo, D. Lin, M. Yuan, Z. Yang, and Y. Niu, "Heterogeneous iris segmentation method based on modified U-Net," *J. Electron. Imag.*, vol. 30, no. 6, pp. 1–12, Nov. 2021.

[97] Z. Ruiyi and S. Wenzhong, "PI-UNet: A neural network model for heterogeneous iris segmentation," *Comput. Eng. Appl.*, vol. 57, pp. 223–229, Jan. 2021.

[98] H. Hofbauer, E. Jalilian, and A. Uhl, "Exploiting superior CNN-based iris segmentation for better recognition accuracy," *Pattern Recognit. Lett.*, vol. 120, pp. 17–23, Apr. 2019.

[99] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, and T. Tan, "Accurate iris segmentation in non-cooperative environments using fully convolutional networks," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2016, pp. 1–8, doi: 10.1109/ICB.2016.7550055.

[100] S. Ahmad and B. Fuller, "Unconstrained iris segmentation using convolutional neural networks," 2018, *arXiv:1812.08245*.

[101] Y. Yang, P. Shen, and C. Chen, "A robust iris segmentation using fully convolutional network with dilated convolutions," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2018, pp. 9–16.

[102] M. Sardar, S. Banerjee, and S. Mitra, "Iris segmentation using interactive deep learning," *IEEE Access*, vol. 8, pp. 219322–219330, 2020.



JASEM RAHMAN MALGHEET received the bachelor's degree from the University of Babylon, Iraq, and the M.Sc. degree from Benha University, Egypt. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM).



NORIDAYU BT MANSHOR received the Ph.D. degree from Universiti Sains Malaysia (USM), in 2014. Currently, she is a Lecturer with the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM). Her main research interests include image processing, computer vision, and pattern recognition.



LILLY SURIANI AFFENDEY received the Bachelor of Computer Science degree from Universiti Pertanian Malaysia, in 1991, the M.Sc. degree in computing from the University of Bradford, U.K., in 1994, and the Ph.D. degree from Universiti Putra Malaysia, in 2007. She is currently an Associate Professor with the Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia. Her current research interests include multimedia databases, video content-based retrieval, data science, and big data analytics.



ALFIAN BIN ABDUL HALIN (Member, IEEE) received the B.Sc. degree in information technology from Universiti Teknologi MARA, Malaysia, in 2000, the Master of Multimedia Computing degree from Monash University, Australia, in 2004, and the Ph.D. degree in computer science from Universiti Sains Malaysia, in 2011. From 2000 to 2001, he was an Executive with the Information Technology Department, Matsushita Display Devices Corporation, Malaysia. In 2001, he joined the Multimedia Department (JMM), Faculty of Computer Science and Information Technology (FSKTM), Universiti Putra Malaysia (UPM), where he has been a Senior Lecturer, since 2011. His research interests include semantic analysis and understanding of multimedia content using image/audio processing, pattern recognition, and machine learning techniques. He holds memberships in several local entities, namely, PECAMP (Information Retrieval and Knowledge Management Society), Malaysia, the Digital Information and Computation Research Group (JMM), and the Road Safety Research Centre (UPM). He is also an Associate Member of two faculty research groups, namely, the Computer Graphics, Vision and Visualization Research Group, FSKTM, and Intelligent Computing, FSKTM.

• • •