Computer Vision and Image Understanding 214 (2022) 103303

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

MTCD: Cataract detection via near infrared eye images

Pavani Tripathi [a], Yasmeena Akhter [b], Mahapara Khurshid [b], Aditya Lakra [a], Rohit Keshari [a], Mayank Vatsa [b,*], Richa Singh [b]

[a] IIIT-Delhi, New Delhi, India
[b] IIT Jodhpur, Rajasthan, India

ARTICLE INFO

Keywords:
Iris
Cataract
Biometrics
Classification
Deep learning
Multitask Learning

ABSTRACT

Globally, cataract is a common eye disease and one of the leading causes of blindness and vision impairment. The traditional process of detecting cataracts involves eye examination using a slit-lamp microscope or ophthalmoscope by an ophthalmologist, who checks for clouding of the normally clear lens of the eye. The lack of resources and unavailability of a sufficient number of experts pose a burden to the healthcare system throughout the world, and researchers are exploring the use of AI solutions for assisting the experts. Inspired by the progress in iris recognition, in this research, we present a novel algorithm for cataract detection using near-infrared eye images. The NIR cameras, which are popularly used in iris recognition, are of relatively low cost and easy to operate compared to ophthalmoscope setup for data capture. However, such NIR images have not been explored for cataract detection. We present deep learning-based eye segmentation and multitask network classification networks for cataract detection using NIR images as input. The proposed segmentation algorithm efficiently and effectively detects non-ideal eye boundaries and is cost-effective, and the classification network yields very high classification performance on the cataract dataset.

1. Introduction

Cataract is an age-related ocular disorder in which the eye lens develops a cloudy layer due to the breaking down of proteins in the eye, which makes it opaque, leading to blurry vision. Both eyes of a person can have a different level of cataract and can develop at the different or same time. It is one of the most common eye diseases and is one of the primary causes of blindness (Pascolini and Mariotti, 2012). According to the *National Blindness and Visual Impairment Survey of India 2015–19*, people above the age of 50 years may develop blindness due to cataract. The condition contributes to the 66.2% blindness cases, 80.7% of severe visual impairment cases, and 70.2% moderate visual impairment cases in this age group. According to Murthy et al. (2008a), in India, 50%–80% of bilateral blindness cases can be attributed to cataract. These numbers demonstrate the need of detecting and correcting cataract in time.

The current process for cataract detection involves using a slit-lamp or an ophthalmoscope for capturing the eye images, and an ophthalmologist examines and tests the eyes of the patient to diagnose the presence of a cataract. While this is the *gold standard*, the rate of blindness, particularly in remote rural areas, is more than the trained ophthalmologists and resources (Murthy et al., 2008b). On the other hand, for biometrics authentication, the low cost near infra-red (NIR) cameras are used in iris recognition. These cameras are available in different form factors, i.e. single eye and dual eye, and they are easy to use. We postulate that eye images obtained from these cameras can help design low cost, accessible, and easy-to-use solutions for cataract detection.

As shown in Fig. 1, NIR eye images provide iris and pupil region which can be utilized to explore whether these images are useful for cataract detection. However, these samples also highlight the challenges involved in designing an automated algorithm. As shown in Fig. 2, the captured images may not be ideal because of: (i) drooping eyelids due to old age, (ii) inadequate camera-to-eye distance and angle, and (iii) excessive contraction or dilation of pupil due to other medical conditions (or ongoing medications such as blood thinner). Therefore, as the first step, designing an efficient segmentation algorithm that segments the iris and pupil regions from the acquired non-ideal images is important. Once the iris and pupil are segmented, the proposed approach involves designing the feature extraction and classification algorithm to differentiate between healthy eye images and images with cataract. In the feature extraction and classification stage as well, the primary challenges are irregular shape and size of iris and pupil. Depending on the kind of occlusion present in the eye image, the angle of capture, and the shape of iris/pupil, the segmented iris and pupil regions can be of different shapes. The classification

* Corresponding author.
E-mail address: mvatsa@iitj.ac.in (M. Vatsa).

https://doi.org/10.1016/j.cviu.2021.103303
Received 6 April 2021; Received in revised form 29 September 2021; Accepted 4 October 2021
Available online 19 October 2021

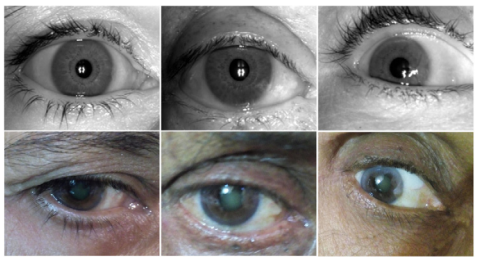

Fig. 1. Showcasing the affected samples of pre and post cataract from NIR spectrum(top row) and visible spectrum (bottom row).

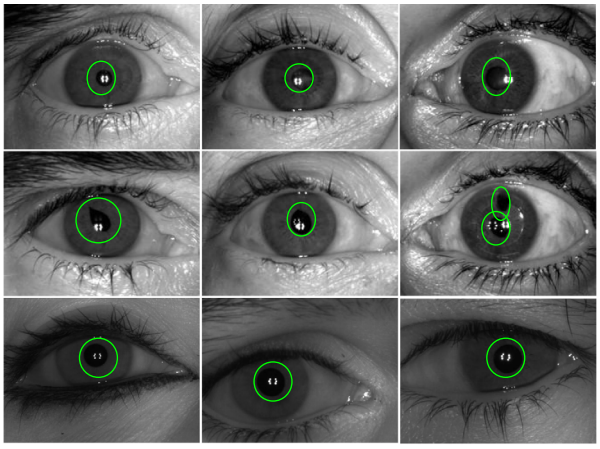

Fig. 2. Showcasing the visual differences in pupil and iris in the pre and post cataract samples. Top row shows the images with cataract and before surgery. Bottom row shows images with cataract removed after surgery.

algorithm should account for these variabilities and perform accurate classification.

To address the above-mentioned challenges, this research presents an automated algorithm, termed as MTCD, for cataract detection from NIR images. As shown in Fig. 3, the input NIR eye image is first processed using the proposed hierarchical pyramid network termed as $PyramidNet$ to segment iris and pupil patterns from images of eyes acquired in unconstrained environments, in the presence of five different covariates, viz. at-a-distance, clouding for pupil due to cataract, punctured iris due to cataract surgery, and excessive contraction or dilation. After post-processing, the segmented eye image (with iris and pupil boundaries) are then used by a multitask deep learning approach that performs two tasks: the first task classifies the image as *healthy* or *unhealthy*, and the second task classifies the images to one of three classes: *pre-cataract*, *post-cataract*, and *others*. The class 'others' consists of samples that are neither suffering from cataract nor have undergone surgery. The results of the proposed cataract detection algorithm are demonstrated on the publicly available IIITD Cataract Surgery dataset (Nigam et al., 2019). We further evaluate the performance of the proposed $PyramidNet$ on four challenging eye (iris) datasets that comprise the covariates mentioned above. Since cataract is assessed in the presence of eye drops used for dilating the pupil, we have also prepared a Pupil Dilation dataset comprising images before and after the use of eye drops.[1] The results on different datasets show that the proposed algorithm yields the best performance in terms of both computation efficiency and memory requirements.

[1] To the best of our knowledge, this is the only dataset in the research community and it will be released to the research community.

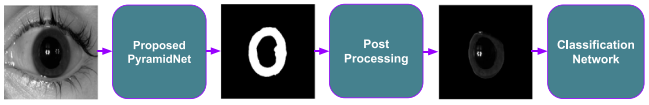

Fig. 3. Proposed Pipeline of the MTCD Approach (Best viewed in color). Architecture of Segmentation Network and Classification Network are shown in Fig. 4 and Fig. 7, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Related work

Literature review is divided into two subsections: (i) cataract related and (ii) iris and pupil segmentation related.

2.1. Literature on cataract prediction

There are limited efforts in automating cataract detection. Srivastava et al. (2014) have proposed a method to grade the nuclear cataract slit-lamp images using gray level image gradients. Yang et al. (2016) used an ensemble approach on models by exploiting three independent feature sets; wavelet, sketch, and texture-based features for grading the cataract fundus images. Ran et al. (2018) have extracted features from fundus images using a three-layer deep convolution neural network (CNN) and random forests (RF) to grade the cataract. They have demonstrated that RF improves grading accuracy. Pratap and Kokil (2019) have used pre-trained AlexNet for feature extraction for fundus image and support vector machines (SVM) for classifying images in different categories of cataract.

Zhang et al. (2019) have implemented a framework to grade the cataract into six levels using feature fusion approach obtained via ResNet18 model and handcrafted GLCM features. Xu et al. (2019b) aimed at grading cataracts from slit-lamp photos using Faster-RCNN to locate the nuclear region and finally used ResNet101 to grade the samples. Xu et al. (2019a) have used the deep model to learn useful features directly from input fundus images for grading the cataract and employed the deconvolution network method to investigate how CNN characterizes cataract layer-by-layer. Grammatikopoulou et al. (2019) have proposed an approach for semantic segmentation in cataract surgery videos. Zhang et al. (2020) proposed GraNet, a CNN-based model, by introducing a point-wise convolution method to learn high-level features for the classification of nuclear cataract from anterior segment optical coherence tomography (AS-OCT) images. To the best of our knowledge, no work has been reported which utilizes images acquired in the NIR spectrum. The proposed work aims to use NIR eye images as the input to cataract classification.

2.2. Literature on iris and pupil segmentation

The literature on iris and pupil segmentation in iris biometrics is very rich. Starting with pre-deep learning approaches such as Daugman (1993), Vatsa et al. (2008) and Zhang et al. (2010) to learning-based approaches (Zhao and Kumar, 2015; Radman et al., 2017), most of the algorithms focus on near-ideal eye imaging. In the recent literature, Convolutional Neural Network (CNN) based approaches are more prevalent. These approaches provide an end-to-end mechanism to search for optimal iris and pupil boundaries. Since segmentation requires the model to correctly segment very fine regions such as iris pixels occluded by eyelashes and specular reflections present in the pupil or iris, the targeted models are designed for segmentation in non-cooperative scenarios (Liu et al., 2016; Arsalan et al., 2017; Lakra et al., 2018; Hofbauer et al., 2019; Hu et al., 2019; Wang et al., 2020). To the best of our knowledge, there is no segmentation algorithm designed for eye images affected due to cataract or post-cataract surgery. Existing algorithms which work well on normal eyes but may not work properly due to the artifacts due to cloudy pupil (or any other medical

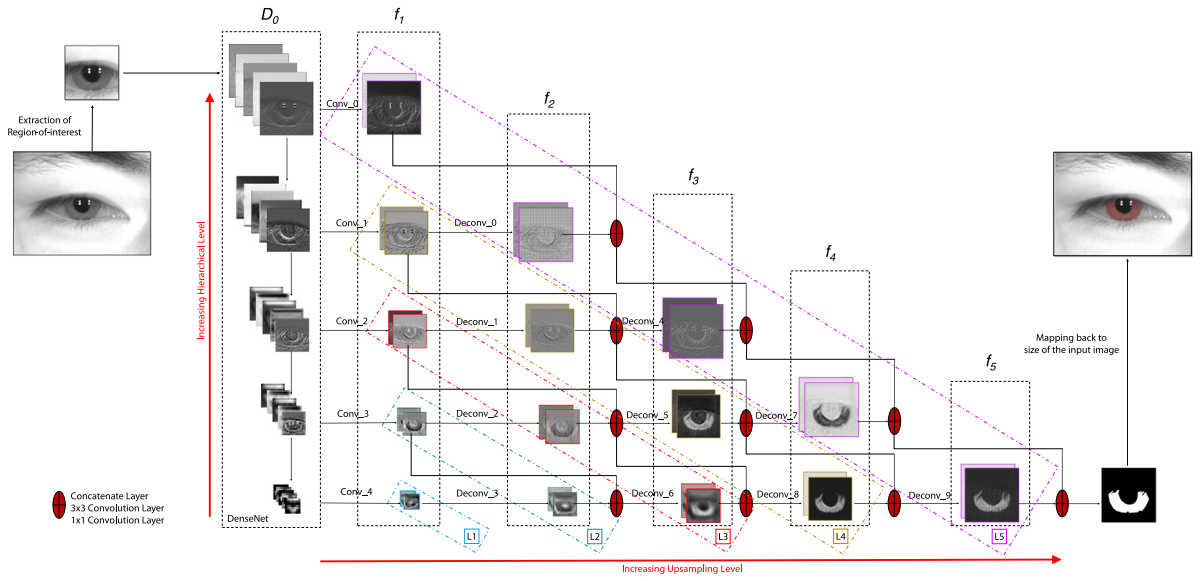

Fig. 4. Presents the proposed architecture, *PyramidNet* for iris and pupil segmentation in an unconstrained environment. The dotted boxes represent the pyramid structure. The upsampling level increases in the x-direction, and the hierarchical level increases in the y-direction. The intermediate feature maps in L1–L5 levels present the different information stored in each map which results in preserving the fine and global structure of the iris and pupil in the final output. (Best viewed in color). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

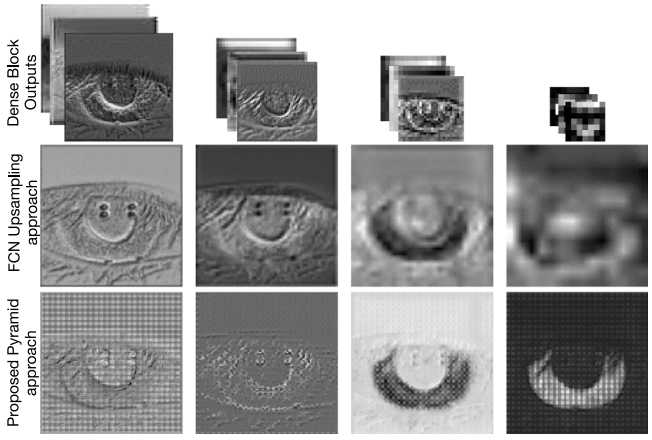

Fig. 5. Illustrating the difference in the local information present in the intermediate outputs when the feature maps are directly upsampled compared to the when the upsampling is done in the proposed pyramid like fashion.

conditions), small punctures or irregularities in iris that may have resulted due to cataract surgery, or pupil may be medically dilated. In this research, one of the contributions is proposing a segmentation algorithm that helps to segment iris and pupil boundaries in medically affected eyes.

3. Proposed MTCD approach

The broad pipeline of the proposed MTCD algorithm is shown in Fig. 3. The eye image acquired from the NIR camera is given as input to the segmentation network. The segmented image is then used by the multitask network for classification. In this pipeline, the segmentation algorithm has to be robust to address real-world challenges such as specular reflections, eyelashes, de-pigmentation, irregularities due to cataract and cataract-removal surgery. In this section, we present the proposed PyramidNet for iris and pupil boundaries segmentation followed by the classification network.

3.1. Proposed PyramidNet for iris and pupil segmentation

Fig. 4 presents a diagrammatic representation of the proposed algorithm. The input image is processed by the proposed algorithm which produces its binary mask. This mask is multiplied with the original input image to extract the region of interest with iris and pupil boundaries.

The proposed algorithm uses DenseNet (Huang et al., 2017) as the backbone network. Let the input image be I_0 and $P_t(\cdot)$ be the non-linear transformation of the tth layer. Input to the tth layer is a concatenation of all the feature maps from the preceding layers, $I_0, I_1, \ldots, I_{t-1}$, i.e.,

$$I_t = P_t(c(I_0, I_1, \ldots, I_{t-1})) \tag{1}$$

where, $c(I_0, I_1, \ldots, I_{t-1})$ denotes the concatenation of the feature-maps produced in layers $0, \ldots, (t_{t-1})$. To allow down-sampling of the feature maps, the DenseNet architecture has been divided into multiple densely connected blocks known as *dense blocks*. We represent the set of these *dense blocks* as D_0^i where the range of i is from 1 to the total number of *dense blocks*. For the task of image classification, DenseNet is trained using categorical cross-entropy loss function. In the proposed method, DenseNet has been used in the *Pyramid Structure* for iris and pupil segmentation.

Upsampling using Pyramid Structure: Deep learning architectures (Arsalan et al., 2017, 2018; Lakra et al., 2018; Liu et al., 2016; Long et al., 2015), directly upsample the intermediate outputs to the size of the final predicted mask resulting in a coarse mask. Upsampling one resolution up, fusing with the previous intermediate output, and continuing the upsampling process in this manner preserves the finest details. For instance, if the feature map of size $n \times n$ is directly upsampled to $4n \times 4n$, then the local structure is not fully preserved. However, if the $n \times n$ feature map is first upsampled to $2n \times 2n$ followed by an upsampling to the size, $4n \times 4n$, then the maximum local structure is preserved. We refer to this kind of upsampling procedure as upsampling in a pyramidic manner. Fig. 5 presents the difference in the feature maps fused to create the final output.

As shown in Fig. 4 we first reduce the number of channels of each dense block, D_0^i to two and consider the segmentation as a two-class semantic segmentation problem, viz. iris class and background class. This creates the first *deep pyramid* structure. Each deep pyramid is

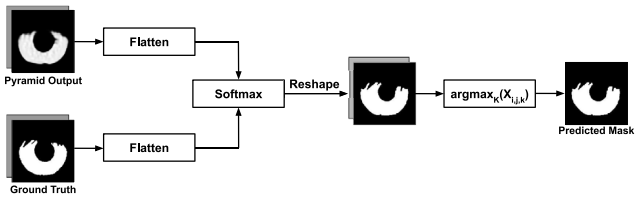

Fig. 6. Illustrating the process of calculating the loss that is back-propagated through the segmentation network.

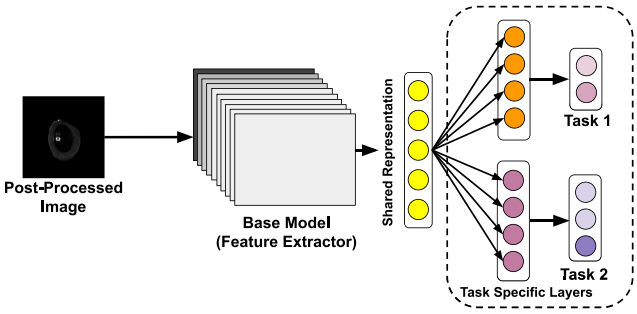

Fig. 7. Multitask classification network for cataract classification.

represented as f_j^i where i denotes the hierarchy of the feature maps in the y-direction, and j represents the upsampling level in the x-direction. Stacking multiple such *deep pyramids* creates a *structural pyramid* where each level is represented as *Lr*, where r is equal to the number of *deep pyramids*.

Deep Pyramid: As represented in Eq. (2), the output of each of the dense blocks is convolved with a 1×1 kernel to reduce the number of channels to two. This convolution operation sets the beginning point of our upsampling path and creates the first *deep pyramid*, symbolized as f_1^i where the range of i is from 1 to the number of outputs in a hierarchical level, in this case, the maximum value of i is equal to the number of blocks present in the base architecture. Mathematically, f_1^i represents the set of feature maps present in this level, $[f_1^1, \ldots, f_1^{nBlocks}]$, where *nBlocks* is equivalent to the number of dense blocks in the base architecture.

$$f_1{}^i = Conv_{1\times1}(D_0{}^i),\ where, i \in [1, \ldots, nBlocks] \tag{2}$$

The next *deep pyramid*, whose set of feature maps is represented as $[f_2^1, \ldots, f_2^{nBlocks-(j-1)}]$ utilizes $[f_1^1, \ldots, f_1^{nBlocks}]$. It is mathematically defined as:

$$f_2{}^i = f_1{}^i \odot Deconv(f_1{}^{i+1}),\\ where, i \in [1, \ldots, nBlocks - 1] \tag{3}$$

where the $\odot$ symbol denotes a set of fusion operations to combine the feature maps. After deconvolution, the upsampled features maps are concatenated with the features maps of the previous hierarchical level. After this, a 3×3 convolution filter is applied to this two-channel output. This convolution operation is done for two reasons. Firstly, it reduces the aliasing effect that may have occurred due to upsampling of lower resolution feature maps. Secondly, it helps in removing the noise present in the higher resolution feature maps. Due to the concatenation operation, the number of channels in the fused output increases from two to four. To reduce the number of channels back to two, we apply 1×1 convolution on the fused output. We continue fusing the outputs of each of the *deep pyramid* until the hierarchy level becomes the same as the number of blocks. Mathematically, every *deep pyramid* can be defined as:

$$f_j{}^i = f_{j-1}{}^i \odot Deconv(f_{j-1}{}^{i+1}) where,\\ j \in [2, \ldots, nBlocks], i \in [1, \ldots, nBlocks - (j-1)] \tag{4}$$

where i denotes the hierarchy of the feature maps in the y-direction, and j represents the upsampling level in the horizontal upsampling path. Due to the fusion of feature maps, the number of hierarchy levels keeps decreasing as we move forward in the horizontal upsampling path.

As shown in Fig. 4 it can be observed that each *deep pyramid* contains varied information. The feature map set of the highest hierarchical level has the maximum resolution. It contains maximum noise along with very fine details of the iris. The last hierarchical level feature map set of least resolution contains minimum noise and preserves the maximum global iris and pupil structure. Hence, when these feature maps are fused to create the next *deep pyramid*, the maximum amount of noise is removed while keeping the local and global iris and pupil structures intact. Further, the total computation cost while adding the feature maps is minimal. Consequently, accurate masks can be obtained without introducing too many overhead parameters to the base network.

Structural Pyramid: Fusing all the *deep pyramids* in the proposed manner creates a *structural pyramid*. Each level of *structural pyramid* contains feature maps of the same resolution and is represented as *Lr*, where r is equivalent to the number of *deep pyramids*. It can be visually seen from Fig. 4 that each set of feature maps in a particular *structural pyramid* level presents different information towards the final prediction. For instance, in level *L5* (represented in Magenta), some feature maps preserve the edge information, whereas others preserve the global structure of the iris and pupil, resulting in an accurate mask.

Iris and Pupil Mask Prediction: Once we have only one set of feature maps in the *deep pyramid*, it is flattened and softmax is applied to obtain per-pixel classification, i.e.

$$P(y = j|\Theta^{(i)}) = \frac{exp\left(\Theta^{(i)} x_{(p,q)}\right)}{\sum_{j=0}^{k} exp\left(\Theta_k^{(i)} x_{(p,q)}\right)} \tag{5}$$

where, k represents the number of classes, viz. two in our case and $\Theta^{(i)}$ symbolizes the softmax parameters and the probability map, $P(y = j|\Theta^{(i)})$ is achieved. To get the eye mask, each pixel is allocated the channel with the highest probability. Fig. 6 illustrating the process of predicting the mask. Finally, a binary morphological post-processing is performed where the mask is first dilated, followed by erosion operation. Finally, the eroded output is multiplied with the original image to generate a region of interest, i.e. eye region only.

3.2. Cataract classification using multitask learning

For the given problem of cataract detection, a segmented eye can be healthy or unhealthy and if unhealthy, it can be a cataract or any other disorder. The cataract affected eye may further be categorized into pre-cataract surgery or post-cataract surgery. In this research, we present this problem as a multitask learning problem with the following two tasks:

- Task 1 (T1): the first task is to classify the input image into one of the classes: $y_{T_1} \in \{healthy, unhealthy\}$
- Task 2 (T2): second task is to classify the input image into three classes, i.e. $y_{T_2} \in \{pre\text{-}cataract, post\text{-}cataract, others^2\}$.

Multitask learning can be accomplished in various ways, such as joint learning of multiple related tasks Liu et al. (2019) and learning auxiliary tasks to support main task (Liebel and Körner, 2018). Due to availability of limited number of images in the dataset, we have performed transfer learning. Pre-trained ResNet50 (He et al., 2016) is used

[2] The 'others' class consists of samples which are neither affected by cataract nor by surgery.

as the base model and it is trained for learning feature representations for cataract detection.

Fig. 7 illustrates the block diagram of the proposed multitask network. To train this network, joint optimization of the losses pertaining to these two tasks are performed. The final loss function is computed as the weighted sum of two classification losses. We have used the binary cross-entropy (BCE) loss and categorical cross-entropy (CCE) loss for Task 1 and Task 2, respectively. The two individual losses and the final loss are defined as follows:

$$BCE = -y_{T_1} log(p) - (1 - y_{T_1}) log(1 - p) \tag{6}$$

$$CCE = -\sum_{i=1}^{3} y^i_{T_2} log(y^i_{T_2}) \tag{7}$$

$$Finalloss = \lambda * BCE + CCE \tag{8}$$

where, i is the class index and p is the class probability.

3.3. Implementation details

This section provides the implementation details of the proposed approach.

Segmentation Network: The proposed segmentation architecture, $PyramidNet$ utilizes the DenseNet model with 43 convolution layers and is trained from scratch using the CASIAv4-distance dataset,[3] UBIRISv2 (Proença and Alexandre, 2005) and IIITD Cataract Surgery dataset (Nigam et al., 2019). The model is trained for 60 epochs using adaptive moment estimation (Kingma and Ba, 2014), *Adam* optimizer with initial learning rate of 0.001. During training, contrast normalization and flip operations are used to augment the dataset size by 10 times. For contrast normalization, 5 different contrast factors have been used. Size of the input images for all the datasets in the NIR spectrum is 640×480. The ROI is extracted using SegDenseNet (Lakra et al., 2018). After extraction of ROI the size of the image reduces to 224×224 which is then fed into the proposed $PyramidNet$.

Classification Network: For training the classification network, we have used IIITD Cataract Surgery, IIITD alcohol and (the proposed) pupil dilation datasets. The cataract samples (pre-cataract and post-cataract surgery) are considered as unhealthy for Task 1 and then two separate classes in Task 2. The other two datasets are used as the healthy class (more details about the dataset are in the next section). For Task 1 and Task 2, we have used sigmoid and softmax activation functions, respectively. For feature extraction, transfer learning concept is utilized where pre-trained (on ImageNet dataset) ResNet50 is used as the base model and fine-tuning is performed on the train sets of the above mentioned datasets. As shown in Fig. 4, a global average pooling (GAP) layer and two fully connected (FC) layers are added on the pre-trained ResNet50 model. These two fully connected layers are added for the two classification tasks, Task 1 and Task 2. The best results are obtained with a model trained on 100 epochs with a learning rate of 0.00001, $\lambda = 0.5$, Adam as an optimizer, and a batch size of 4 on NVIDIA V100 32GB GPU. To achieve better generalization, we have also performed data augmentation with contrast normalization by various factors and flip operations, which increased the dataset size by five times.

4. Datasets

The proposed deep learning based segmentation and classification method is evaluated on three datasets, viz., IIITD Cataract Surgery (Nigam et al., 2019), IIITD Alcohol (Arora et al., 2012), and on the proposed Pupil Dilation dataset. These datasets are chosen since they comprise various covariates of eye image, making them suitable choices for evaluating the efficiency of the proposed models.

[3] http://biometrics.idealtest.org/dbDetailForUser.do?id=4

Table 1
Characteristics of the proposed Pupil Dilation dataset.

Characteristics	Pupil Dilation
Sensor	Vista Sensor
Environment	Indoor
Sessions	Two
No. of individuals	88
No. of images	276 (pre) and 276 (post)
Resolution	640 × 480
Challenges	Excessive dilation due to the administered eyedrops.

Pupil Dilation Dataset: The proposed dataset contains images showcasing variations due to Pupil Dilation. Tropicacyl Plus, a prescription drug used to treat paralysis of the ciliary muscle and dilate pupils before and after ophthalmic surgery, is used to create the dataset. The dataset consists of 528 images acquired from human subjects before and after the medicine is administered by the ophthalmologist. The pupil dilation dataset contains 528 images, 264 pre-eyedrop treatment and 264 post-eyedrop treatment images of 44 subjects. Fig. 8 shows sample images acquired pre and post eyedrop treatment. Table 1 presents various characteristics of the images. To the best of our knowledge, this is the first dataset of its kind and is released to the research community via http://iab-rubric.org/resources.html. For experiments, 50 samples are used for testing while the remaining form the training set. After augmentation, the number of training samples is 1815.

IIITD Cataract Surgery Dataset contains 880 samples from 132 individuals, 440 each representing catarct and post cataract surgery samples (represented as pre and post cataract surgery). 100 samples from both the classes are kept in the test set and the remaining comprise the train set. After augmentation, the number of training samples is 4080.

IIITD Alcohol Dataset Arora et al. (2012) studied the effect of alcohol on pupil dilation/constricts. The pupil dilates/constricts due to intake of alcohol which results in affecting the iris recognition performance. Also, it is clearly shown in Fig. 8 (row c and d) how the alcohol can affect the size of the pupil which in turn can affect the iris recognition. More details about this dataset can be found in Arora et al. (2012). This dataset contains 440 images pertaining to 110 subjects. Of these, 50 randomly selected samples are used for testing, while the remaining comprise the training set. After applying the augmentation, the number of training samples is 1170.

Data Preparation: For learning the segmentation model, we have pre-trained the model on the CASIAv4-distance and UBIRISv2 (Proença and Alexandre, 2005) datasets and then IIITD Alcohol and IIITD Cataract Surgery datasets are used for fine-tuning. For learning the cataract classification model, the ImageNet pre-trained base model is used and then IIITD Alcohol, IIITD Cataract Surgery, and the proposed Pupil Dilation datasets are used. Data augmentation is applied so as to minimize the data imbalance problem. IIITD Alcohol and Pupil Dilation datasets belong to one class, and the IIITD Cataract Surgery belongs to the other class, thus making the overall data balanced. Ground truth segmentation masks for iris and pupil have been manually annotated using Adobe Photoshop. We will release the proposed database, annotations, and train–test partition details via http://iab-rubric.org/.

5. Experimental results

The performance of the proposed MCTD approach is presented in two parts, (i) segmentation and (ii) classification. The effectiveness of the algorithm is compared by varying the base model and comparing the results with existing algorithms. We have also performed an ablation study to demonstrate the effectiveness of various components of the algorithm.

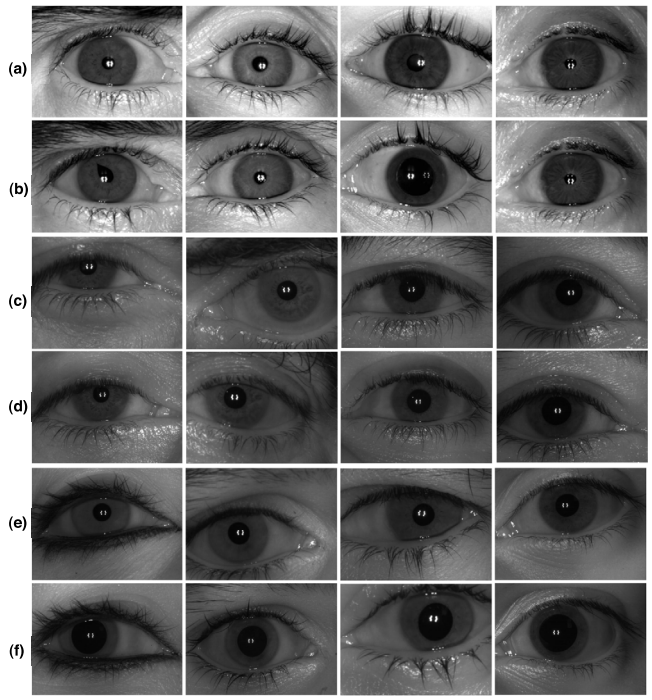

Fig. 8. Sample images: (a) and (b) are pre and post cataract surgery; (c) and (d) are pre and post alcohol; (e) and (f) are pre and post pupil dilation from the Pupil Dilation dataset.

5.1. Segmentation performance

The performance of the segmentation algorithm is measured using the average classification error rate proposed in the NICE-I competition (NICE.I).

$$Error = \frac{1}{N \times m \times n} \sum_{i,j=1}^{m,n} M_G^{te'}(i,j) \oplus M_P^{te'}(i,j) \tag{9}$$

where, $M_G^{te'}$, $M_P^{te'}$, N, m and n denote the ground truth mask, the predicted mask, total number of test samples, height, and width of the mask, respectively. The logical exclusive-OR operator calculates the correspondent disagreeing pixels' proportion between the ground truth and the predicted segmentation mask. We compare results with a non-deep learning method (Zhao and Kumar, 2015) and two deep learning methods: IrisParseNet (Wang et al., 2020), and SegDenseNet (Lakra et al., 2018).

Fig. 9 shows the sample results on the IIITD Cataract Surgery dataset where the masks are overlaid on the iris and pupil regions. These examples show that the proposed algorithm is able to detect the fine boundaries of iris and pupil region. Table 2 presents segmentation errors obtained from the proposed algorithm and the existing algorithms. The percentage error has reduced by 21.4%, 11.9%, and 31.8% (from the next best performing model on these datasets) on the IIITD Cataract Surgery, IIITD Alcohol, and Pupil Dilation datasets, respectively compared to existing techniques. It is observed that the proposed method yields state-of-the-art accuracies on all these datasets. Further, Fig. 10 compares the performance across the methods based on the classification accuracy. It can be observed that the proposed method is able to classify each iris pixel more accurately compared to existing deep learning methods for iris segmentation.

Fig. 11 shows sample masks generated by the proposed method and comparison with existing algorithms on the three datasets. As can be visually observed, the proposed method can predict very accurate masks, implying that it preserves both global and fine structures of the iris and pupil. The first row shows how the model can segment the iris region even when it is severely occluded by reflection. Further, all the masks predicted by the proposed method have fine-details, such as removing areas secluded by fine eyelashes. Also, unlike the SegDenseNet (Lakra et al., 2018), the proposed method can predict the mask for sample images of the IIITD Cataract dataset, which contains bubbles. It is our assertion that the proposed method can overcome this because upsampling in pyramid fashion preserves both the local and global structures. Further, as shown in Fig. 11, when the contrast difference between the iris and the sclera region is extremely low, the proposed algorithm is still able to detect the boundaries. It can be directly observed that both SegDenseNet (Lakra et al., 2018) and Zhao and Kumar (2015) fail to segment the boundaries correctly. However, $PyramidNet$ can handle these cases with great precision because it restores the information in a pyramid manner. The fine edge information and global structure present in the *structural pyramid* $L5$ when fused can accurately predict the boundaries even when the contrast difference is extremely low.

We also compare the performance of the proposed algorithm with the FCN architecture (Long et al., 2015). In this approach, a deconvolution operation has been used to upscale the image and combine with the previous layer output feature maps. However, in the $PyramidNet$ architecture, the second and third blocks of the DenseNet (as shown in Fig. 4) are utilized in two ways. This results in multiple feature maps of the same resolution. Combining these incorporates both the coarse and fine structures of iris and pupil in the segmentation process. The difference between the proposed $PyramidNet$ and FCN-8s outputs has been shown in Fig. 9. The results for FCN-8s are computed using our implementation of FCN-8s. On the cataract dataset, FCN yields 1.35% segmentation error and $PyramidNet$ achieves 0.77% segmentation error. This comparison shows that for iris and pupil boundary segmentation, it is imperative to combine feature maps at each upsampling level.

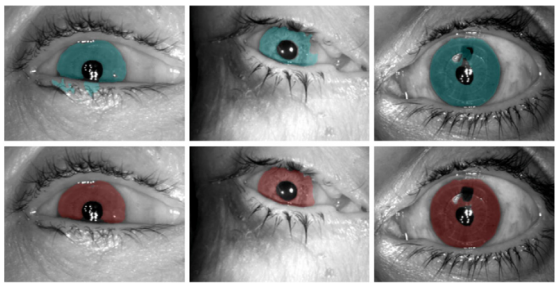

Fig. 9. Illustrating the segmentation output by FCN-8s (first row) and the proposed $PyramidNet$ (second row) algorithms on the IIITD Cataract Surgery dataset. The masks are overlaid on the images to visually demonstrate segmentation with respect to the iris and pupil boundaries. The results demonstrate that the proposed algorithm yields finer boundaries compared to FCN-8s approach.

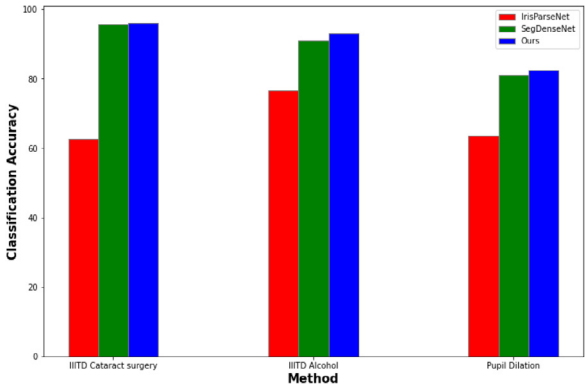

Fig. 10. Showcasing the classification accuracy of existing and proposed segmentation methods on the datasets used in the paper.

Table 2
Comparisons of the proposed and existing iris segmentation techniques using average segmentation error (%). For fair comparison no post-processing is performed for Wang et al. (2020).

Method	IIITD Cataract Surgery	IIITD Alcohol	Pupil Dilation
IrisParseNet (Wang et al., 2020)	9.87	3.06	8.16
(Zhao and Kumar, 2015)	6.28	8.51	7.67
SegDenseNet (Lakra et al., 2018)	0.98	1.42	3.46
Proposed Method	**0.77**	**1.25**	**2.36**

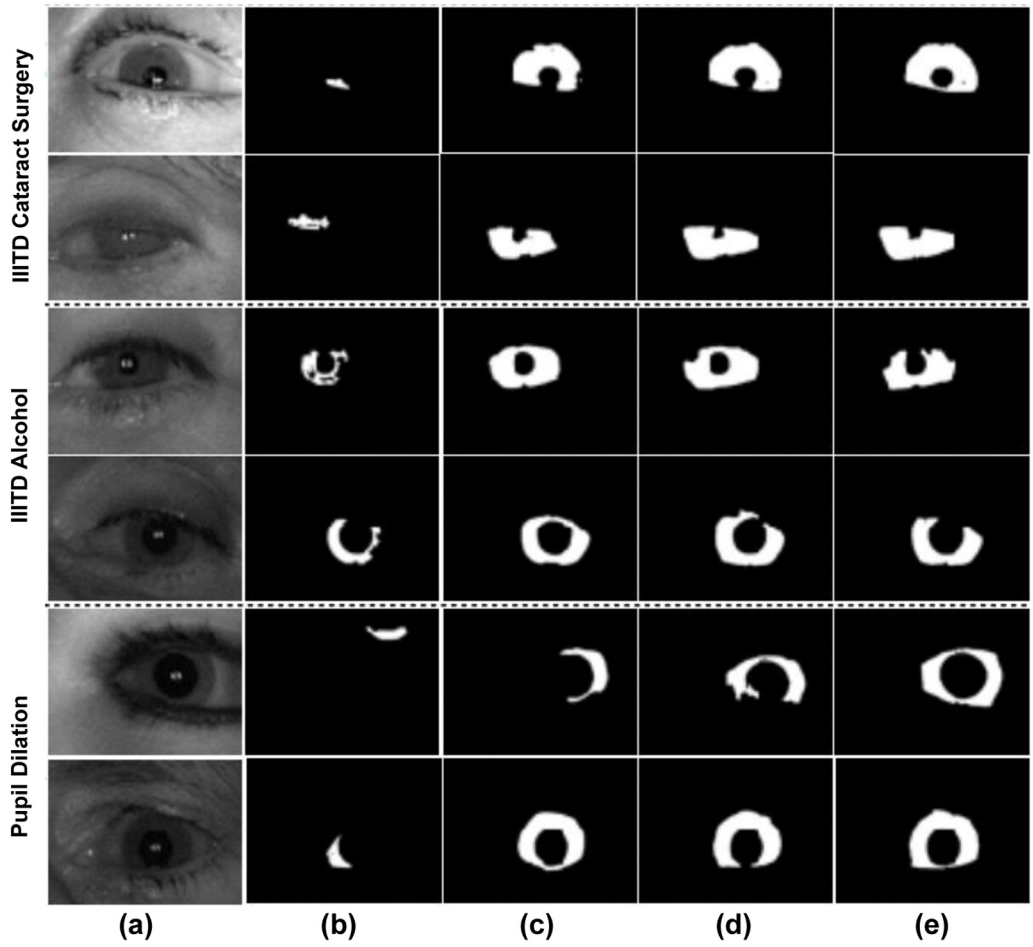

Fig. 11. Showcasing the results of iris segmentation on multiple datasets. (a) The input image; masks obtained by (b) Zhao and Kumar (2015) method, (c) SegDenseNet (Lakra et al., 2018) (the next best performing deep learning approach), (d) proposed PyramidNet, and (e) ground truth.

To further show the efficacy of each component of $PyramidNet$ architecture, ablation study is performed. In the proposed method the number of *structural* levels is equivalent to the number of dense blocks present in the base architecture. To understand the effect of each *structural* pyramid level on the final output, we have computed the segmentation error. In our ablative study, the least segmentation error is achieved when all the dense blocks are used for building the *structural* pyramid. This is so because for iris and pupil segmentation both global and fine structures must be preserved. It is our assertion that maximum amount of fine structure is preserved by the output of the first level of the *structural* pyramid and the maximum global information is stored in the last i.e. fifth level of the *structural* pyramid. We have observed that there is a small decrease in segmentation error if the feature maps obtained at *structural* pyramid level $L3$ are directly upsampled to the size of the output image compared to upsampling of feature maps of $L2$ level. However, on upsampling the feature map obtained at $L5$ level, there is a significant decrease in the segmentation error because the maximum amount of local information/the finest details of the iris and pupil are preserved in it.

Table 3
Characteristics of the models proposed for iris segmentation. Details for Wang et al. (2020) have been directly taken from the paper.

Algorithms	Model size (MB)	No. of parameters (M)	Test time (sec)
IrisParseNet (Wang et al., 2020)	119.0	31.28	0.15
SegDenseNet (Lakra et al., 2018)	57.30	8.00	0.024
Proposed PyramidNet	**11.9**	**0.92**	**0.017**

The proposed PyramidNet has significantly lower number of parameters compared to IrisParseNet (Wang et al., 2020) and SegDenseNet (Lakra et al., 2018)[4]. As shown in Table 3, the number of parameters has reduced by *30 times* and the size of the model has reduced by *10 times*. Further, the testing time of PyramidNet is also the least. It yields state-of-the-art results on three datasets and is the optimal model both in

[4] Parameters are calculated using our own implementation of Liu et al. (2016) methods

Table 4
Summarizes the performance of the proposed approach of multitask eye image classification by changing the segmentation algorithm. We used SegDenseNet (Lakra et al., 2018) as a baseline and replaced it with the proposed PyramidNet, which outperformed the baseline results. It is also evident from the table that the post-processing step on the segmentation masks of PyramidNet improves the overall performance of both the tasks.

Segmentation Algorithms		Accuracy (%)	Precision	Recall	F1 score
Baseline SegDenseNet	T1	97.34	0.96	0.97	0.97
	T2	92.34	0.92	0.92	0.92
PyramidNet	T1	100	1.0	1.0	1.0
	T2	95.67	0.96	0.96	0.96
PyramidNet + Post-Processing	T1	**100**	**1.0**	**1.0**	**1.0**
	T2	**96.67**	**0.97**	**0.97**	**0.97**

terms of computation cost and memory consumption. To be uniform, all the algorithms are implemented and run on the same machine, keeping all the configurations same.

5.2. Cataract classification

The cataract classification performance is reported in terms of the classification accuracy, precision, recall, and F1 score. The output of segmentation algorithm, i.e. segmented iris and pupil region, is used as input to the classification algorithm. For comparison, we have used SegDenseNet (Lakra et al., 2018) approach (2nd best segmentation approach — from Table 2). Further, in order to showcase the effect of binary morphological operations (post-processing) after the proposed *PyramidNet*, we have shown the results with and without post processing. Table 4 summarizes the results of the proposed multitask classification algorithm with three segmentation approaches. It can be clearly observed that PyramidNet yields improved performance compared to the baseline results of SegDenseNet. PyramidNet differentiates between the healthy and unhealthy classes with 100% accuracy. Further, PyramidNet with post-processing does not deteriorate the performance in Task T1 but improves the classification performance in Task T2. For differentiating between the diseased classes, i.e., task T2, PyramidNet with post-processing yields an error of only 3.3%. Analyzing the precision and recall, we have observed that both precision and recall of *healthy class* is 1. This result is due to the fact that there is no overlap between the samples of healthy and unhealthy classes. It is further supported by Table 5 (confusion matrix) that very few pre-cataract and post-cataract samples are misclassified into each other. Interestingly, among the remaining two classes, the precision of *post-cataract class* is lower than the *others* class, while the recall of the *post-cataract class* is higher than the *others* class. After post-processing, the overall performance and precision of post-cataract performance improves, however, the recall reduces marginally by 0.03.

Fig. 12 shows the tSNE plots of the healthy and unhealthy classes (Task 1), the first one is for the image space and the second one is for the feature space. It is observed that the affected class (pre and post cataract) is well distinguishable from the healthy class. Fig. 13 shows the sample results of the proposed method. In the experiments, for Task 2, we have observed that some of the pre-cataract and post-cataract samples are misclassified with each other (as shown in Table 5).

We next analyze the effect of base model, learning rate, and number of epochs:

Effect of changing the base model: For cataract classification, the performance of different deep learning models, viz. InceptionV3 (IV3) (Szegedy et al., 2016), VGG16 (Simonyan and Zisserman, 2014), ResNet50 (RN50), and DenseNet121 (DN121) are compared. As reported in Table 6, ResNet50 outperforms all other architectures for both the tasks and is an effective choice as a base model.

Changing the learning rate: In this experiment, the learning rate is varied from 0.01 to 0.000001. We observe that the learning rate of 0.00001 outperforms the others yielding 100% test accuracy for Task 1 and 96.67% accuracy for Task 2.

Changing the number of epochs: We have also evaluated the performance by varying the number of epochs, and reported the results. It is shown that 100 epochs with learning rate = 0.00001 yields the best results for this classification problem. If we increase the number of epochs by 20, the results remain the same, beyond which the model starts overfitting.

Table 5
Illustrates the confusion matrix for the two tasks.

	Predicted/Actual	Healthy	Unhealthy
Task 1	Healthy	1.0	0.0
	Unhealthy	0.0	1.0

	Predicted/Actual	Pre-Cataract	Post-Cataract	Others
Task 2	Pre-Cataract	0.96	0.4	0.0
	Post-Cataract	0.6	0.94	0.0
	Others	0.0	0.0	1.0

Table 6
Shows the F1 scores obtained by varying the pre-trained models on the two tasks T1 and T2.

Model	VGG16	IV3	DN121	RN50
T1	1.0	0.96	0.99	1.0
T2	0.94	0.84	0.91	**0.97**

Fig. 12. Illustrating the tSNE plot for Task 1: left plot shows the samples in the original image space and the right plot shows the samples in the feature space.

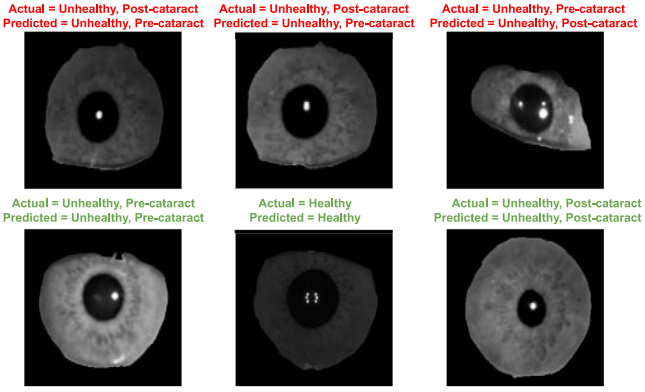

Fig. 13. Shows some correctly classified and misclassified samples from the dataset (best viewed in color). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Conclusion

Cataract is one of the primary causes of visual impairment worldwide and cataract surgery is the most common surgical intervention. Typically, the prognosis, regular monitoring, and the decision of

whether a patient should be taken up for surgery mostly depends on the discretion of the ophthalmologist. In resource constrained settings with limited experts, it is very important to have a clinical decision-support technique to improve sensitivity and specificity of cataract detection and monitoring. This paper presents a deep learning algorithm for cataract detection. To the best of our knowledge, this is the first work which proposes to use near infrared eye images, popularly used in iris biometrics, for cataract detection. A deep learning-based architecture, $Pyramid Net$, is proposed for segmenting iris and pupil boundaries where the model fuses the coarse and fine information extracted from convolution blocks at different levels in a pyramid-like fashion. The segmented iris and pupil regions are then used for cataract classification via a multi-task network. Experiments performed on the cataract dataset show that (i) effective cataract detection is possible in NIR domain, (ii) the proposed segmentation algorithm is effective in detecting iris and pupil boundaries even with challenging scenarios, and (iii) the overall cataract detection performance encourages such an approach to be used in automated decision support system. It is our assertion that the findings of this research and the availability of our datasets, will spur further research in this domain.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work of Mayank Vatsa was supported in part by the Swarnajayanti Fellowship by the Government of India.

References

Arora, S.S., Vatsa, M., Singh, R., Jain, A., 2012. Iris recognition under alcohol influence: A preliminary study. In: IAPR International Conference on Biometrics, pp. 336–341.

Arsalan, M., Hong, H.G., Naqvi, R.A., Lee, M.B., Kim, M.C., Kim, D.S., Kim, C.S., Park, K.R., 2017. Deep learning-based iris segmentation for iris recognition in visible light environment. Symmetry 9 (11), 263.

Arsalan, M., Naqvi, R.A., Kim, D.S., Nguyen, P.H., Owais, M., Park, K.R., 2018. Irisdensenet: Robust iris segmentation using densely connected fully convolutional networks in the images by visible light and near-infrared light camera sensors. Sensors 18 (5), 1501.

Daugman, J.G., 1993. High confidence visual recognition of persons by a test of statistical independence. IEEE Trans. Pattern Anal. Mach. Intell. 15 (11), 1148–1161.

Grammatikopoulou, M., Flouty, E., Kadkhodamohammadi, A., Quellec, G., Chow, A., Nehme, J., Luengo, I., Stoyanov, D., 2019. CaDIS: Cataract dataset for image segmentation. arXiv preprint arXiv:1906.11586.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Hofbauer, H., Jalilian, E., Uhl, A., 2019. Exploiting superior CNN-based iris segmentation for better recognition accuracy. Pattern Recognit. Lett. 120, 17–23.

Hu, J., Hui, Z., Xiao, L., Liu, J., Li, X., He, Z., Li, L., 2019. Seg-edge bilateral constraint network for iris segmentation. In: IAPR International Conference on Biometrics.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Lakra, A., Tripathi, P., Keshari, R., Vatsa, M., Singh, R., 2018. Segdensenet: Iris segmentation for pre and post cataract surgery. IEEE Int. Conf. Pattern Recognit..

Liebel, L., Körner, M., 2018. Auxiliary tasks in multi-task learning. arXiv preprint arXiv:1805.06334.

Liu, N., Li, H., Zhang, M., Liu, J., Sun, Z., Tan, T., 2016. Accurate iris segmentation in non-cooperative environments using fully convolutional networks. In: IAPR International Conference on Biometrics.

Liu, X., Suganuma, M., Okatani, T., 2019. Joint learning of multiple image restoration tasks. arXiv preprint arXiv:1907.04508.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.

Murthy, G., Gupta, S.K., John, N., Vashist, P., 2008a. Current status of cataract blindness and vision 2020: the right to sight initiative in India. Indian J. Ophthalmol. 56 (6), 489–494.

Murthy, G., Gupta, S.K., John, N., Vashist, P., 2008b. Current status of cataract blindness and vision 2020: the right to sight initiative in India. Indian J. Ophthalmol. 56 (6), 489.

NICE.I, - Noisy iris challenge evaluation - Part I, http://nice1.di.ubi.pt/index.html.

Nigam, I., Keshari, R., Vatsa, M., Singh, R., Bowyer, K., 2019. Phacoemulsification cataract surgery affects the discriminative capacity of iris pattern recognition. Sci. Rep..

Pascolini, D., Mariotti, S.P., 2012. Global estimates of visual impairment: 2010. Br. J. Ophthalmol. 96 (5), 614–618.

Pratap, T., Kokil, P., 2019. Computer-aided diagnosis of cataract using deep transfer learning. Biomed. Signal Process. Control 53, 101533.

Proença, H., Alexandre, L.A., 2005. UBIRIS: A noisy iris image database. In: International Conference on Image Analysis and Processing. Springer, pp. 970–977.

Radman, A., Zainal, N., Suandi, S.A., 2017. Automated segmentation of iris images acquired in an unconstrained environment using HOG-SVM and GrowCut. Digit. Signal Process..

Ran, J., Niu, K., He, Z., Zhang, H., Song, H., 2018. Cataract detection and grading based on combination of deep convolutional neural network and random forests. In: International Conference on Network Infrastructure and Digital Content. IC-NIDC, IEEE, pp. 155–159.

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

Srivastava, R., Gao, X., Yin, F., Wong, D.W., Liu, J., Cheung, C.Y., Wong, T.Y., 2014. Automatic nuclear cataract grading using image gradients. J. Med. Imaging 1 (1), 014502.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.

Vatsa, M., Singh, R., Noore, A., 2008. Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing. IEEE Trans. Syst. Man Cybern. 38 (4), 1021–1035.

Wang, C., Muhammad, J., Wang, Y., He, Z., Sun, Z., 2020. Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition. IEEE Trans. Inf. Forensics Secur. 15, 2944–2959.

Xu, X., Zhang, L., Li, J., Guan, Y., Zhang, L., 2019a. A hybrid global-local representation CNN model for automatic cataract grading. IEEE J. Biomed. Health Inf. 24 (2), 556–567.

Xu, C., Zhu, X., He, W., Lu, Y., He, X., Shang, Z., Wu, J., Zhang, K., Zhang, Y., Rong, X., et al., 2019b. Fully deep learning for slit-lamp photo based nuclear cataract grading. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 513–521.

Yang, J.-J., Li, J., Shen, R., Zeng, Y., He, J., Bi, J., Li, Y., Zhang, Q., Peng, L., Wang, Q., 2016. Exploiting ensemble learning for automatic cataract detection and grading. Comput. Methods Programs Biomed. (ISSN: 0169-2607) 124, 45–57.

Zhang, H., Niu, K., Xiong, Y., Yang, W., He, Z., Song, H., 2019. Automatic cataract grading methods based on deep learning. Comput. Methods Programs Biomed. (ISSN: 0169-2607) 182, 104978.

Zhang, X., Sun, Z., Tan, T., 2010. Texture removal for adaptive level set based iris segmentation. In: IEEE International Conference on Image Processing, pp. 1729–1732.

Zhang, X., Xiao, Z., Higashita, R., Chen, W., Yuan, J., Fang, J., Hu, Y., Liu, J., 2020. A novel deep learning method for nuclear cataract classification based on anterior segment optical coherence tomography images. In: IEEE International Conference on Systems, Man, and Cybernetics, pp. 662–668.

Zhao, Z., Kumar, A., 2015. An accurate iris segmentation framework under relaxed imaging constraints using total variation model. In: IEEE International Conference on Computer Vision.