Content based retrieval of retinal OCT scans using twin CNN

MAHUA NANDY PAL[1,*] , SHUVANKAR ROY[1] and MINAKSHI BANERJEE[2]

[1]CSE Department, MCKV Institute of Engineering, Howrah, West Bengal 711204, India
[2]CSE Department, RCC Institute of Information Technology, Kolkata, West Bengal 700015, India
e-mail: mahua.nandy@gmail.com; shuvankarroy2@gmail.com; mbanerjee23@gmail.com

Abstract. Retinal imaging helps to detect retinal and cardiovascular abnormalities. Among these abnormalities, Diabetic Macular Edema (DME) and Age Related Macular Degeneration (AMD), both are frequent retinal degenerative diseases leading to blindness. Content based retinal OCT scan retrieval process makes use of characteristic features to retrieve similar Optical Coherent Tomography (OCT) scans, index-wise, from a database with minimal human intervention. A number of existing methods take care of segmentation and identification of retinal landmarks and pathologies from OCT volumes. As per the literature survey, till date, no papers are there which deal with the retrieval of retinal OCT scans. In this work, we propose a retrieval system for retinal OCT scans which extracts feature maps of both query and database samples from the layer of deep convolutional neural network and compares for their similarity. The Twin network comparison approach exploits deep features without the resource, space and computation exhaustive network training phase. Most of the techniques involving deep network implementation suffer from the drawbacks of data augmentation and resizing. These requirements have been eliminated automatically as part of the Twin network implementation procedure. The system successfully retrieves retinas with similar symptoms from the database of differently affected and unaffected OCT scans. We evaluated different variations of retrieval performances like AMD-Normal, DME-Normal, AMD-DME, AMD-DME-Normal, etc. Execution time optimization has also been achieved as the network used is comparatively shallow and network training is not required. The system retrieves similar scans from a dataset of abnormal and normal OCT scans with a mean average precision of 0.7571 and mean reciprocal rank of 0.9050. Considering all possible variations of retrieval, we achieved overall mean average precision and mean reciprocal rank of 0.631167 and 0.829607, respectively which are also quite notable with rank thresholds of 3, 5 and 7. Experiments show that the method is noticeably successful in retrieving similar OCT volumes. Per image mean average retrieval time is 8.3 sec. Automatic retrieval of retinal OCT volumes for the presence of a particular ailment can help ophthalmologists in the mass screening process.

Keywords. Retinal OCT scan; diabetic macular edema (DME); age related macular degeneration (AMD); convolutional neural network (CNN); mean average precision (MAP); mean reciprocal rank (MRR).

1. Introduction

Optical Coherent Tomography (OCT) is a non-invasive technique capturing retinal tissue images. The cross-sectional views of retinal tissues are obtained from this imaging modality. Some of the ailments, prevalent in affected retinal OCT scans are DME (Diabetic Macular Edema) and AMD (Age related Macular Degeneration). A fast, accurate, and reliable method for retrieval of affected samples will help greatly in improving the healthcare screening process at an early disease stage and with less manpower involvement. The proposed system retrieves relevant retinal OCT scans from the database, by comparing the features of the query and the OCT scan from the database within the network layer itself using similarity measurement. This technique can be utilized to implement a computer-aided diagnostic system leading to a decision regarding the manifestation of disease symptoms in OCT samples. But the main challenges coupled with OCT scans are intrinsic speckle noise, low optical contrast from place to place, variable individual B-scan resolutions, variable number of B-scans per volume, etc.

Among different deep convolution based 2D image retrieval works, Tzelepi *et al* [1] can be mentioned for general image retrieval. They obtained the feature representations from convolutional layers using max-pooling, and subsequently, they adapted and retrained the network,

*For correspondence

Published online: 25 August 2021

to produce more efficient and dense image descriptors. Similar type of image retrieval applications are also discussed in [2] and [3] by Saritha *et al* and Wan *et al* respectively. Saritha *et al* in [2] used a deep belief network (DBN) to extract the features. Wan *et al* [3] presented a comprehensive study on the application of deep learning in content based image retrieval.

Simple deep learning based approaches suffer from the challenge of network learning with few samples. Overcoming this challenge is highlighted by Koch *et al* [4] and by Vinyals *et al* [5], though both the works dealt with the problem domain of character recognition. Koch *et al* [4] exploited discriminative features to optimize the predictive power of the network on the Omniglot dataset. Vinyals *et al* [5] implemented one-shot learning on ImageNet and Omniglot. They reported an accuracy of 93.2% on ImageNet and 93.8% on Omniglot.

In [6] by Chung *et al*, to learn image representations with less supervision, the authors used a deep Twin Siamese CNN (SCNN) architecture that can be trained with only binary image pair information. They evaluated the learned image representations for content-based retrieval of two- dimensional medical images using a publicly available fundus image dataset. This method was associated with image resizing and a huge number of samples. They resized the images to 224×224 before feeding 2D images to the deep learning pipeline. They evaluated their work with mean average precision (MAP) and mean reciprocal rank (MRR). Reported results are the maximum of 67% MAP and 77% MRR in case of 2D retinal fundus images with only two classes for retrieval – normal and severe Diabetic Retinopathy. Some different domain applications of the Twin network are as follows. In Li *et al* [7], A Siamese neural network-based severity score measurement system has been proposed. The system automatically detects COVID-19 pulmonary disease severity in chest X-Ray images. [8–10] are different applications of Siamese network by Ramachandra *et al*, Zhang *et al* and Yin *et al,* respectively. [8] and [9] implemented real time visual tracking with a deeper Twin Siamese network and [10] tried to find anomalies in videos with a variation of the Siamese network.

We also considered several very recent deep learning supported works involving OCT images for our survey, though the works are not retrieval based, they are classification based works. These works have been reviewed analytically before implementation of the OCT retrieval system as no prior research work has addressed OCT retrieval. Wang *et al* in [11] utilized linear configuration pattern (LCP) based features along with Correlation-based Feature Subset (CFS) for OCT classification purpose and reported 99.3% overall accuracy. Public 3-class OCT dataset Duke was used for work evaluation. Lee *et al* in [12] represented a 2-class OCT classification method of normal and AMD affected images. According to them, initially, 2.5 million OCT images were extracted and 50,000 from each class have been selected. 80839 images

were used for deep neural net training purpose. Network weight initialization was done using the Xavier algorithm. The network architecture comprises of 13 convolution layers, 4 maxpool layers and three fully connected layers. They distinguished AMD from normal OCT samples. They represented the system efficiency through accuracy and area under the ROC curve as the evaluation metrics. They declared an area under the ROC curve of 92.78% with an accuracy of 87.63% considering image level. At the macula level, the reported area under the ROC curve is 93.83% with an accuracy of 88.98%. and at patient level, the area under the ROC curve of 97.45% with an accuracy of 93.45%. Maximum sensitivity and specificity with optimal cutoffs were 92.64% and 93.69%, respectively. Another classification work has been reported by Lu *et al* [13] for the classification of normal and abnormal retinal OCT images. They tried to exploit the benefits of transfer learning in this work. ResNet is previously trained on ImageNet. Four ResNet classifiers are further trained independently for retinal abnormalities and their outputs are combined to make the final decision [13]. They imported eye images from Wuhan University Eye Center, deidentified them and labeled them by eye experts. The dataset contained four abnormality classes (serous macular detachment, cystoid macular edema, macular hole, and epiretinal membrane). The test set consists of 300 normal and 537 abnormal images. Kaymak *et al* [14] in 2018, proposed four class classification of retinal OCT images – wet AMD, dry AMD, DME and healthy. In this case also, the concept of transfer learning was utilized. Pretrained AlexNet architecture is used for model generation. Model training and testing were performed with 83484 and 1000 images, respectively. Wang *et al* [15] in 2019, experimented over four CNN architecture – VGG-16, Inception-V3, ResNet-18, ResNet-50. These nets are initially trained on ImageNet. They claimed that ResNet-50 is the best model among them. They resized the images in preprocessing phase. They reported the maximum accuracy, sensitivity and specificity of 96.25%, 97% and 98.98%, respectively. They tested their model on a public 4-class Mendeley OCT dataset for binary classification of abnormal and normal classes. They achieved 99.5% test AUC and 99.85% test sensitivity though the results are obtained on generalized class definitions.

This work proposes a content based retrieval system for retinal OCT volumes. Feature maps are extracted and compared from the deep layer of CNN. During network training, tensors grow inside the network which augments high memory requirement. To cope with increasing memory requirements, input data is usually resized. Twin network does not require resizing as the implementation characteristic itself eliminates the requirement of exhaustive network training. Data resizing in biomedical applications leads to significant information loss. Twin Siamese CNN architecture is used for comparing feature encodings. OCT database is prepared to obtain a representative frame

from each of the volumes. These representative frames are fed into the network with maximum permissible resolution as per the data availability. Absolute distances between intermediate feature representations are computed and samples are retrieved in distance sorted order. Gradations of retrieved scans are verified for quantitative evaluation of system efficiency. To the best of our knowledge, OCT retrieval work is hardly explored previously. So, the contributions of the work can summarly be represented as:

- No literature as of date explored OCT retrieval and application of Twin network for this purpose.
- The work proposes the elimination of computationally expensive and resource exhaustive network training by the application of Twin network for retrieval purpose.
- Twin CNN retrieval excludes the requirement of data resizing which prevents loss of medically significant information.
- The application has been evaluated by different categories of retrievals like DME-Normal, AMD-Normal and DME-AMD-Normal and AMD-DME, etc.

One excel sheet containing category-wise retrieval result summary and one execution time sheet have been provided as supplementary materials.

2. Preliminaries

2.1 *Diabetic macular edema (DME) and age related macular degeneration (AMD)*

Diabetes mellitus is considered to be an epidemic by 2025, more than 300 million people are estimated to be affected by it worldwide. Thus the physiological problem arises due to long-standing diabetes mellitus certainly will increase. Diabetic macular edema (DME) is one of the complications of diabetes mellitus that leads to vision loss accompanied by loss of quality of life. Similarly, age related macular degeneration (AMD) is another prevalent complication that arises due to the occurrence of diabetes mellitus. Diabetes mellitus is responsible for the incidence and progression of both the complications through altering hemodynamics, increasing oxidative stress, accumulating advanced glycation end products, etc. [16]. DME and AMD, mainly affect the retinal macular region. Both are accompanied by macular edema and associated with aggressive inflammatory restoration procedure that aggravates disease progression. This process directly leads to the breakdown of the blood-retinal barrier (BRB). Though the underlying causes may be different, the key pathophysiological process in DME and AMD is characterized by the breakdown of the blood-retinal barrier, inflammatory processes and an increase in vascular permeability. Though DME and AMD indications may be compared with each other they are very different from personal perception. In [16], Das *et al* discussed the inflammatory conditions of AMD, DME, etc.

2.2 *Twin CNN*

A convolutional neural network (CNN) is an artificial neural network having one or more convolution layer/s in between the input and output layer. The neurons of the fully connected layers are interconnected with weighted connections. The impulse or signal passes through these layers finding the probability distribution at the output layer. The advantage of using deep CNN is that the network itself is capable of framing feature characterizations of the input patterns. No hand-crafted features are required to efficiently represent the input patterns. The disadvantage of using CNN is that they require a huge number of labeled samples for efficient and successful training to generate output probability. Particularly in medical applications, collecting huge labeled samples is quite difficult and in many cases impossible. One-shot learning helps to remove this disadvantage of a large number of sample requirement through the implementation of a Twin network. Thus, the Twin network is responsible for evaluating the reference sample and test samples using the same weight and bias initialization. The intermediate vectors become comparable to each other. Thus the absolute difference between two feature encodings is considered as the measurement of similarity score. The less the value of absolute difference the more similar the representative scans are.

2.3 *Content based retrieval*

Content based retrieval is retrieving similar objects based on the visual content of the data. It is an active and current research field. In this era of digital imaging, ever-growing visual data are handled through content based retrieval. Retrieval system helps in developing automated medical diagnostic systems, as well. Figure 1 outlines a general retrieval system.

3. Evaluation metrics

The most important and widely used retrieval system evaluation metrics are Mean Average Precision (MAP) and Mean Reciprocal Rank (MRR).

For average precision (AP), we set rank threshold k = 3, 5 and 7 for our application and computed percent relevant in top k to find precision. MAP is the mean of average precisions at different ranks.

$$MAP = 1/N \sum_{i=1}^{N} AP \qquad (1)$$

In MRR, the rank position of the first relevant retrieval sample is considered and the mean of reciprocal of rank is computed across multiple queries.
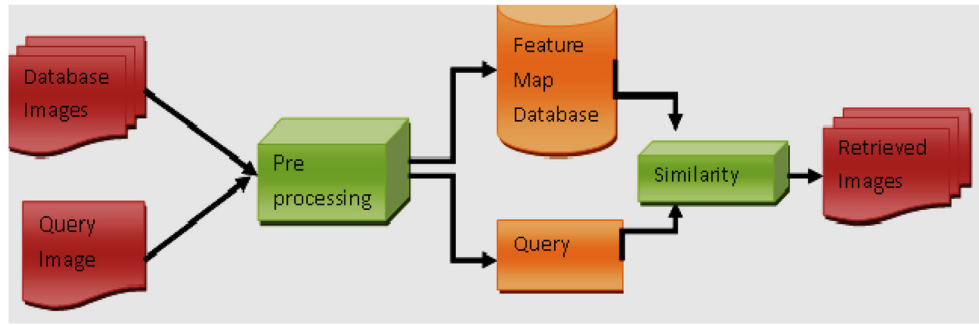
Figure 1. Outline of content based retrieval process.

$$MRR = 1/N \sum_{i=1}^{N} \frac{1}{rank_i} \qquad (2)$$

where N is the number of query samples and $rank_i$ is the rank of the first similar retrieval in the ith query sample.

4. Implementation requirements

The system has been implemented using Intel Core i5-6500CPU @ 3.20 GHz, 6M Cache, up to 3.6 GHz, 12 GB DDR4 RAM and NVIDIA GeForce GTX 1660 SUPER GPU having 1408 CUDA cores and 6GB of GDDR6 memory. Software specification for the work is as follows, Python 3.8.6 with NVIDIA CUDA 10.1. Different required side-loaded modules are Tensorflow 2.3.1, Keras 2.4.3, Pillow 8.0.1, PyQt 5.15.1, etc.

5. Dataset description

Duke dataset [17] for retinal OCT volume has been used for work evaluation. This dataset is frequently used among researchers. This dataset consists of volumetric scans acquired from 45 patients: 15 normal patients, 15 patients with dry AMD, and 15 patients with DME using Spectralis SD-OCT (Heidelberg Engineering Inc.). Variety of B-scan resolutions (Height $\times$ Width) that are available in the

dataset are (496 $\times$ 512), (496 $\times$ 1024), (496 $\times$ 768). Figures 2, 3 and 4 provide some AMD affected, some DME affected and some normal OCT B-scans. Individual scans are in .tiff format. Retinal B-scans are 2D and having a horizontal span of retinal cross-sectional tissues. We verified the result in another dataset [18] as well. In Mendeley dataset [18], images are labeled into 4 categories- CNV, DME, DRUSEN, and NORMAL.

6. Proposed methodology

The proposed methodology follows three distinct phases: namely, database preparation, model implementation and similarity measurement-retrieval. The process flow diagram is shown in figure 5. All the phases of the proposed method are discussed in detail in sub-sections 6.1, 6.2 and 6.3 .

6.1 *Database preparation*

In the Duke dataset, the number of B-scans or slices in individual OCT volumes are not equal for different patients. Individual slice resolution also varies. Thus, to characterize a representative slice all the B-scans of a volume are cropped to minimum resolution from both sides equally and then preprocessed by applying morphological tophat transformation, contrast limited adaptive histogram
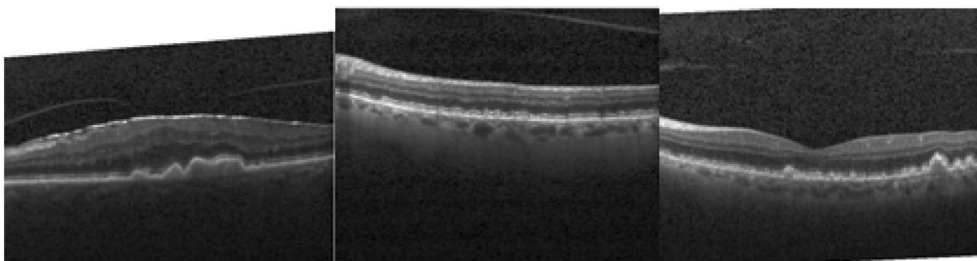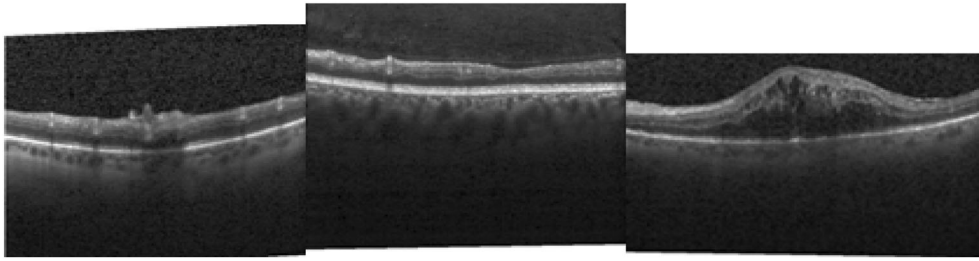

Figure 2. Original AMD affected scans.

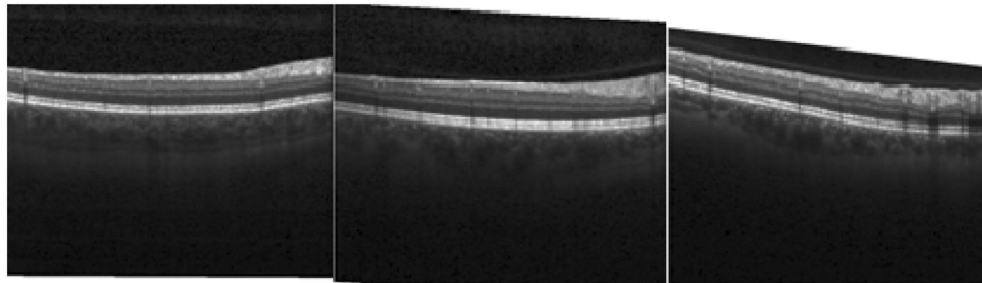**Figure 3.** Original DME affected scans.

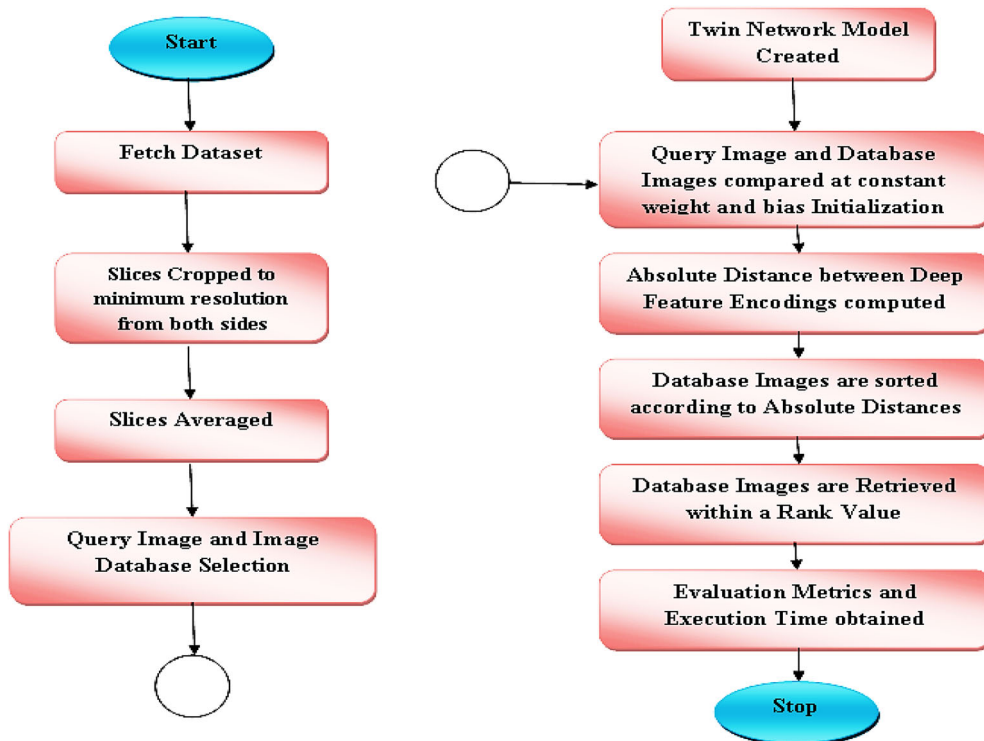**Figure 4.** Original normal OCT scans.

Figure 5. Process flow diagram.

equalization (CLAHE) and adaptive gamma correction (AGC) subsequently. These enhanced B-scans are averaged to maintain the

standard and are converted to non-overlapping patches of size 48×48 which are inputs to the network. Resizing operation of B-scans was avoided to prevent loss of information inside the region of interest.

6.2 *Model implementation*

We utilized successfully the benefit of one-shot learning in the application of retrieval of affected and normal retinal OCT scans using Convolutional Neural Network (SCNN). Based on the similarity measurement, OCT images are retrieved from the database where both normal and differently affected OCT images are available at the same time. Initially, we adopted a deep CNN model which was generated following U-type connections [19]. This architecture is quite efficient in extracting biomedical image features. We assumed and compared a comparatively simple I-shaped CNN architecture for Twin network based retrieval work, with this architecture also.

Based on the initial evaluation results, further experiments are followed with a simpler CNN architecture. The

pictorial representation of detailed Twin CNN comparison has been represented with all the network information provided as indiagram labels in figure 6.

6.3 *Similarity measurement and retrieval*

In a deep CNN model, feature maps are refined in convolution layers. Patch feature representations, available after the 4th convolution layer are observed for actual experimentation as a deeper convolutional layer is expected to represent a more effective representation map capable of capturing image characteristics precisely. Thus, feature maps are available at fully connected layer. The absolute difference between individual image feature encodings is computed by summing up patch siamese distances. To compare two representative scans, the absolute difference between feature encodings serves to produce the similarity score between query and database scans. Detailed Keras architecture of the

Twin network is provided in figures 7(a) and (b). Figure 7(a) represents architecture details of an individual network of the twin part of the model separately. Figure 7(b) represents the overall architecture of the whole Twin network model. Sequential block of figure 7(b) has been expanded to show the detailed sequential architecture
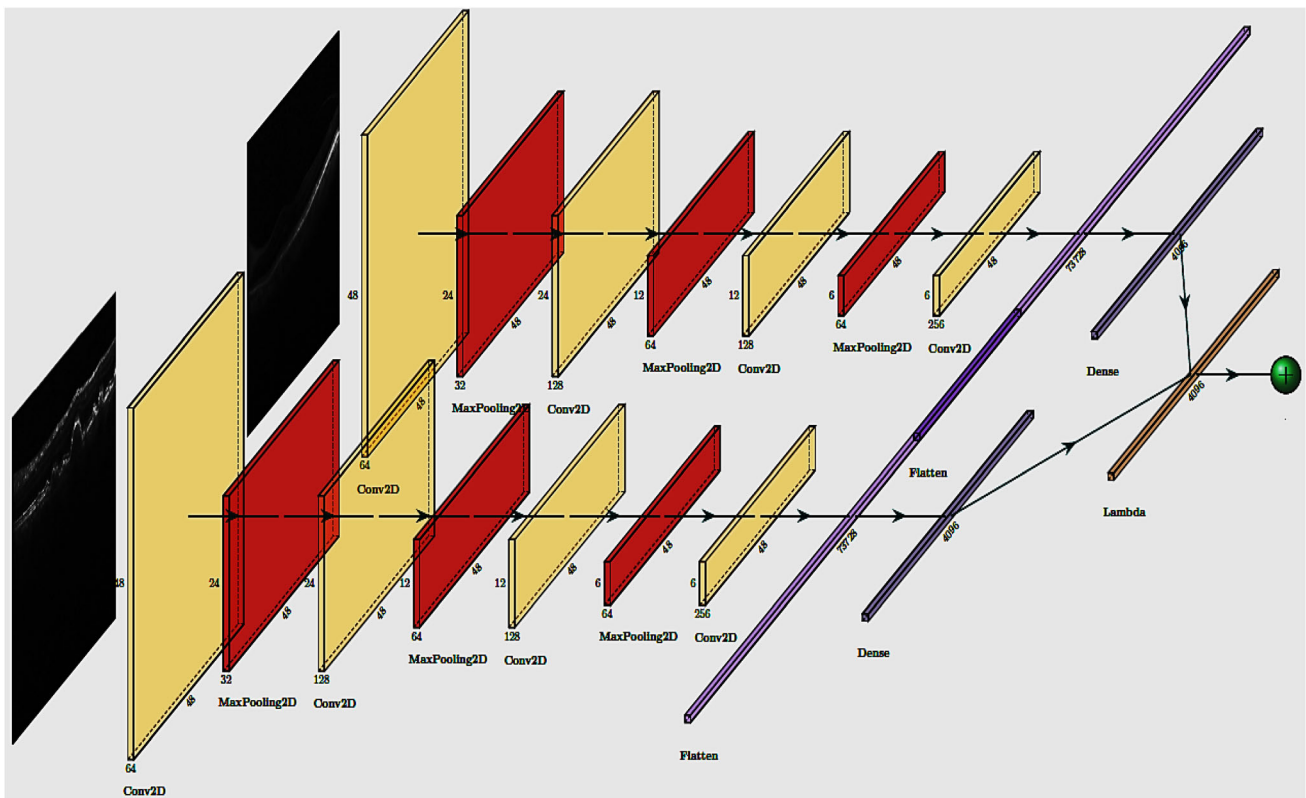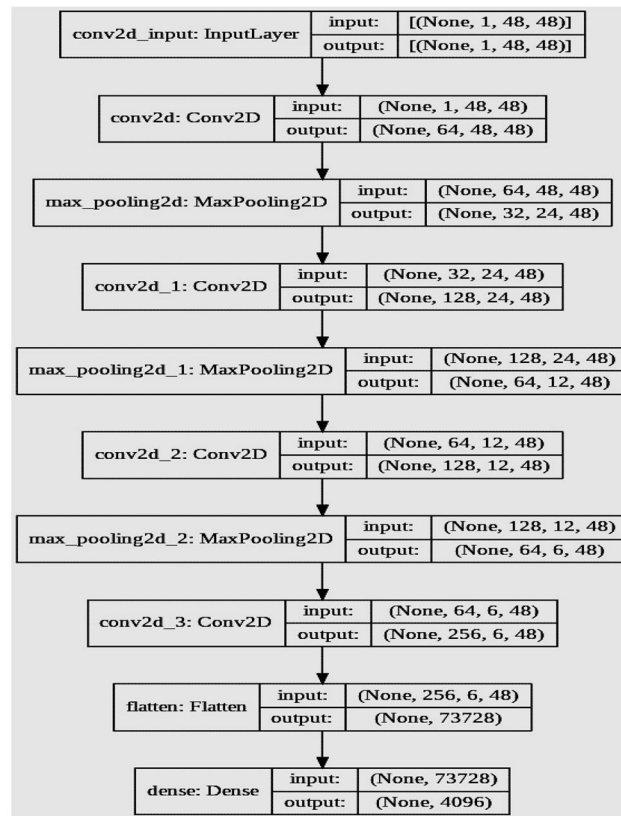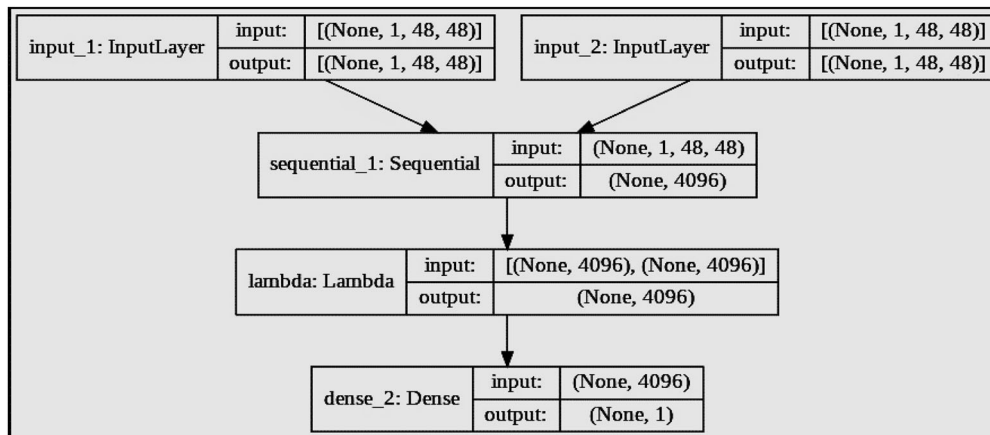

Figure 6. Detailed Twin CNN comparison model.

(a)

(b)

Figure 7. (**a**) Architecture details of an individual network of the twin part of the model. (**b**) Overall architecture of the whole Twin network model.

in figure 7(a) through which input_1 and input_2 proceeds parallelly.

In the individual Twin part, four consecutive 2D convolution layers with 64, 128, 128 and 256 filters are used, respectively. Consecutive convolution layers are interleaved by a maxpool layer. This pipeline is followed by successive flatten and dense layers. Input patches from both the scans are parallelly passed through the above mentioned Twin network layers and the absolute difference between the outputs of the dense layers is computed. The estimated similarity score is obtained as output by passing the absolute difference through subsequent dense and sigmoid layers. In this process, images are retrieved according to sorted similarity scores.

7. Experimental results

7.1 *Effectiveness comparison between U-shaped and I-shaped twin CNN*

As U-shaped CNN [19] is efficient in pixel based segmentation of biomedical images, initially DME-Normal Twin CNN retrieval results are computed for U-shaped and simple I-shaped networks. As per the experimental findings, it was observed that the I-shaped network is more effective in case of pair-wise intermediate feature comparison. Table 1 shows the evaluation comparison results considering an arbitrary seed value at a given point of time.

7.2 *Experiments with I-shaped CNN*

A total of 3231 B-scan images were extracted initially from the Duke dataset. B-scans are of variable resolutions. In Duke, individual patient data contains variable number of B-scans as well.

7.2a *Database preparation.* Duke slices were cropped to the resolution of 496×512 to avoid information loss in the area of interest and then averaged to obtain individual representative slices. Averaged slices of three different classes are shown in figures 8 and 9.

7.2b *Model implementation.* Retrieval performance depends heavily on the algorithm used to represent the minute features of the images. But it is really difficult to extract finer details of the contour, shape and structure of the images manually. Challenges present in this relevant field are low contrast, uneven illumination, presence of speckle noise, presence of lesions, etc. To overcome these challenges, we tried an application of a convolutional neural network (CNN) to extract features of both affected and normal slices. CNN is an extremely powerful machine learning algorithm that is capable of extracting feature characteristics by deep convolution. Further, to reduce exhaustive training complexity and a large volume of training data requirements, a Twin network has been proposed for successful comparison between OCT scans. We propose a comparatively shallow network comprising of only four convolution layers for this purpose.
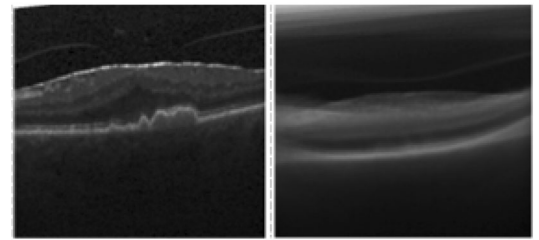
The comparison exhibits less time, space and computational complexity. A schematic diagram for the Twin model is in figure 10.

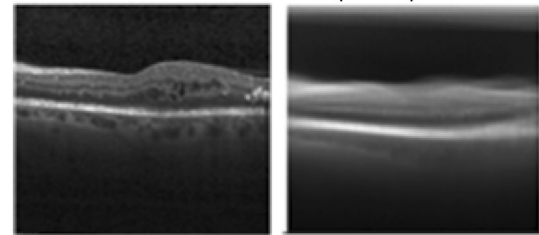**Table 1.** Evaluation comparison between U-shaped CNN and I-shaped CNN.

DME-normal retrieval	MAP	MRR
U-shaped CNN	0.59999	0.69166
I-shaped CNN	0.61666	0.80000
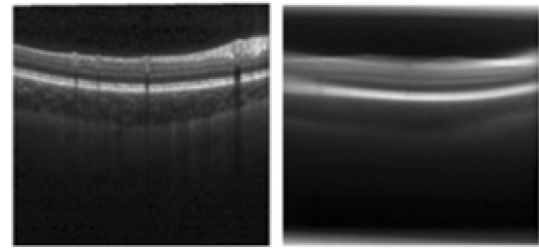
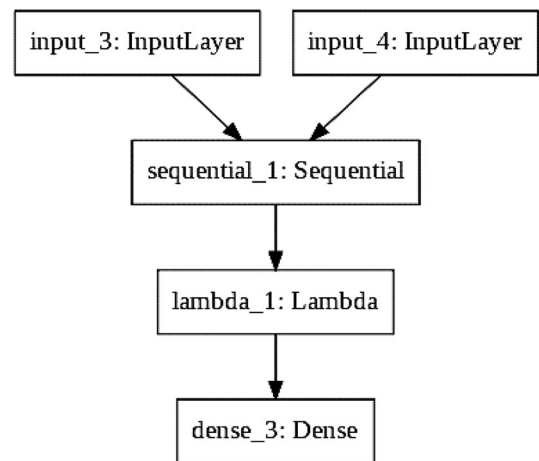**Figure 8.** An original and averaged version of Normal OCT (Normal1).

AMD affected OCT (AMD1)

DME affected OCT (DME5)

Figure 9. An original and averaged version of affected OCT.

Figure 10. Twin model overview.

7.2c *Seed value optimization of Twin CNN retrieval.* To tackle random initialization of weight and bias of the network, and for successful comparison, the paired network

computations are executed in the controlled environment of weight and bias initialization. To organize non-determinism, the seed at the starting point of the sequence is made predictable across the model generation pipeline. Thus, randomness has made predictable as much as possible. We tested the experimentation with the initial fire of randomization in NumPy and TensorFlow from 0 to 2. Experimental results are uploaded as supplementary materials in an excel file.

7.2d *Similarity measurement and retrieval.* Feature maps of database image patches and query image patches are compared from the deep convolution layer of CNN architecture. We assume that the last convolution layer is capable of representing the most significant and refined feature maps, hence we considered the intermediate representation after the fourth convolution layer for computing the absolute element wise difference between the query patch and the database image patch. Similar weight initialization guarantees that similar images are possibly be retrieved. Flatten feature encodings are passed through a dense layer before computing the absolute difference. A dense layer with a sigmoid function is used to produce the probability of being similar. Thus, maximum similar scans are retrieved following summed-up patch siamese distances. The rank thresholds have been set to 3, 5 and 7. Thus the system retrieves three, five or seven most similar retinal B-scans and ailment gradations are observed for evaluation of the proposed retrieval system. System evaluation has been done by MAP and MRR. Best MAP has been obtained considering seed value = (2,2) and best MRR value has been obtained using seed value = (1, 2) for randomization initialization in NumPy and TensorFlow respectively. The summaries of four categories of retrieval results are shown in the form of average precision and reciprocal rank in Tables 2, 3, 4, 5.

Graphical representation of MAP and MRR result summary for all four retrieval variations are presented in figures 11 and 12, respectively.

Overall average MAP and MRR metrics of the proposed method for OCT retrieval are 0.631167 and 0.829607, respectively. Overall evaluation metric values are quite noticeable. If we consider only abnormal and normal

Table 2. Evaluation metrics for DME-normal retrieval.

DME-normal retrieval		
	AP Seed = (2,2)	RR Seed = (1,2)
K=3	0.75862	0.862068
K=5	0.72413	0.877586
K=7	0.72413	0.883333
Mean value	0.735626667	0.874329

**Table 3.** Evaluation Metrics for AMD-Normal Retrieval.

AMD-Normal Retrieval		
	AP Seed = (2,2)	RR Seed = (1,2)
K=3	0.804597	0.93103
K=5	0.77241	0.93793
K=7	0.75862	0.93793
Mean Value	**0.7785423**	**0.93563**

Table 4. Evaluation metrics for DME-AMD retrieval.

DME-AMD retrieval		
	AP Seed = (2,2)	RR Seed = (1,2)
K=3	0.533333	0.75555
K=5	0.513333	0.77222
K=7	0.49999	0.77698
Mean Value	0.515552	0.76825

Table 5. Evaluation Metrics for AMD-DME-Normal Retrieval.

AMD-DME-normal retrieval		
	AP Seed = (2,2)	RR Seed = (1,2)
K=3	0.5303	0.72348
K=5	0.47727	0.74507
K=7	0.47727	0.75211
Mean value	0.49494667	0.74022
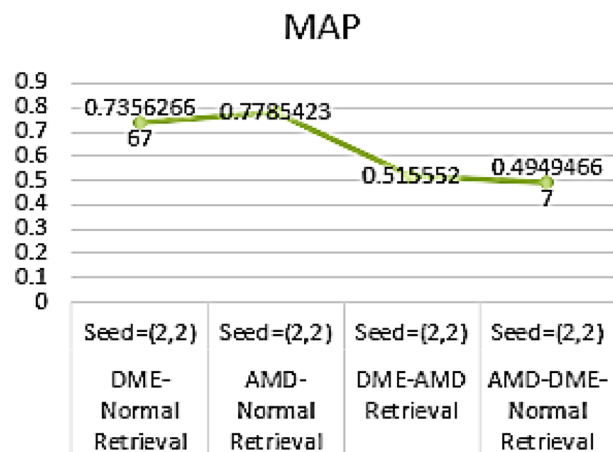
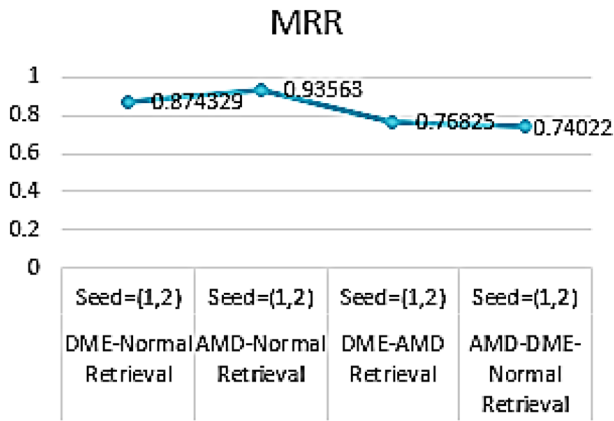**Figure 11.** Graphical representation of MAP values.

MRR

Figure 12. Graphical representation of MRR values.

retrieval for the Duke dataset, our system demonstrates MAP of 0.7571 and MRR of 0.9050.

7.2e User interface. A user interface for content based OCT retrieval from AMD-DME-Normal, AMD-DME, AMD-Normal, DME-Normal databases has been developed which takes the query input and the path of the OCT scan database as user inputs from medical experts and shows the content-

based retrieval results along with the retrieved categories sorted by minimum Euclidean distance. Apart from retrieval results, this interface shows Average Precision (AP) and Reciprocal Rank (RR) for rank, K = 3, 5, 7 and system evaluation metrics Mean Average Precision (MAP) and Mean Reciprocal Rank (MRR) also. Images of system execution for four different retrieval variations with the proper user interface have been shown in figures 13, 14, 15, 16.

7.2f Implementation level execution time requirement. Best MAP values are obtained considering seed values = (2, 2) and best MRR values are obtained considering seed values = (1, 2). Execution time requirements are also encouraging. Database preparation, similarity measurement and retrieval times are 8.3987 sec and 8.3452 sec per image considering seed values = (2, 2) and (1, 2) respectively. So overall database preparation, similarity measurement and retrieval time per image is 8.372 sec. Table 6 shows a summary of different retrieval timings.

7.2g System verification using Another dataset. The system performance is further verified with the Mendeley dataset [18]. 15 arbitrarily selected volumes of each of the four categories of OCT are considered for experimentation. The
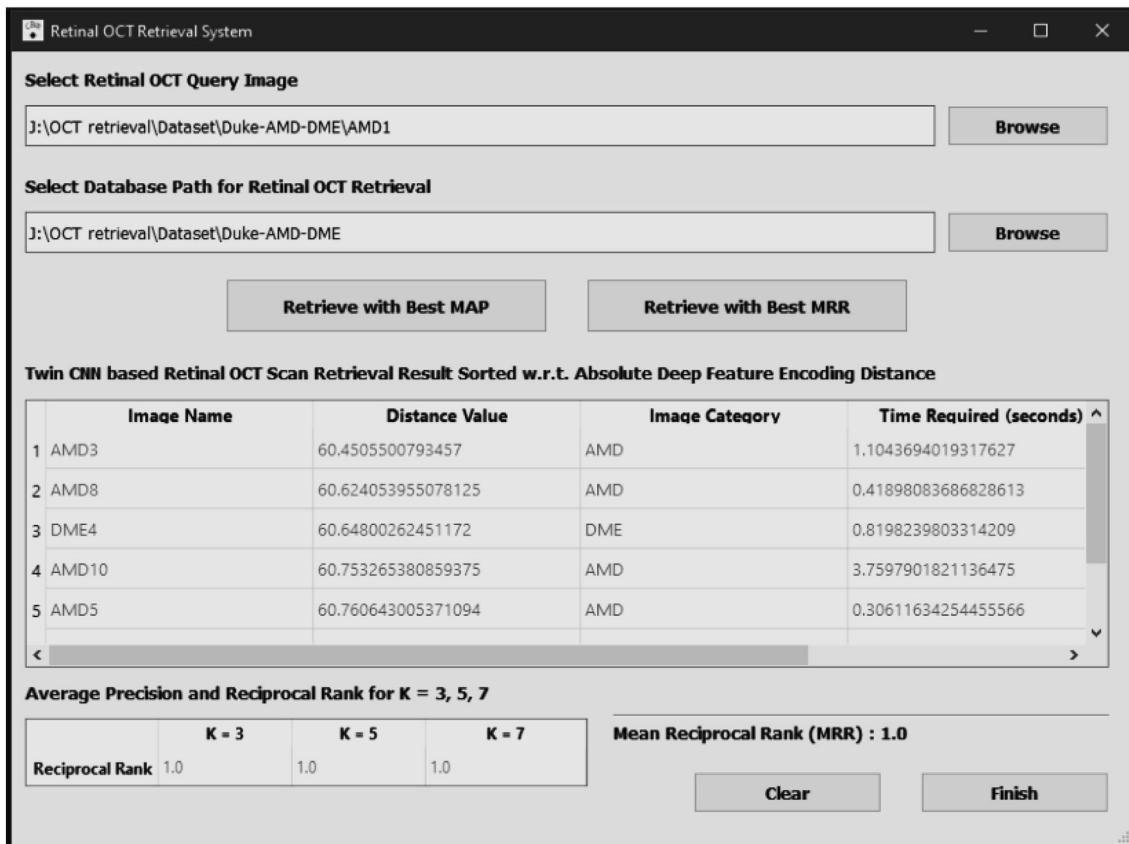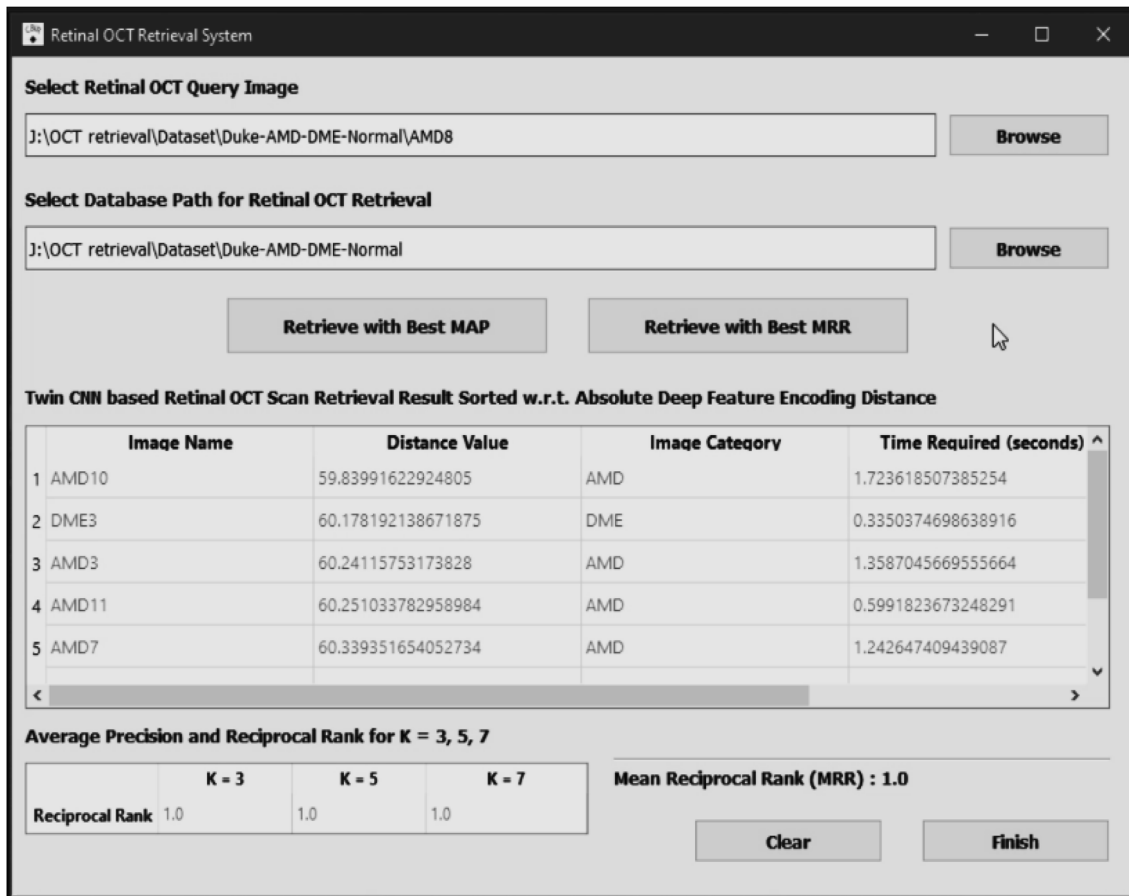

Figure 13. Sample retrieval for Mean Reciprocal Rank from AMD-DME database.
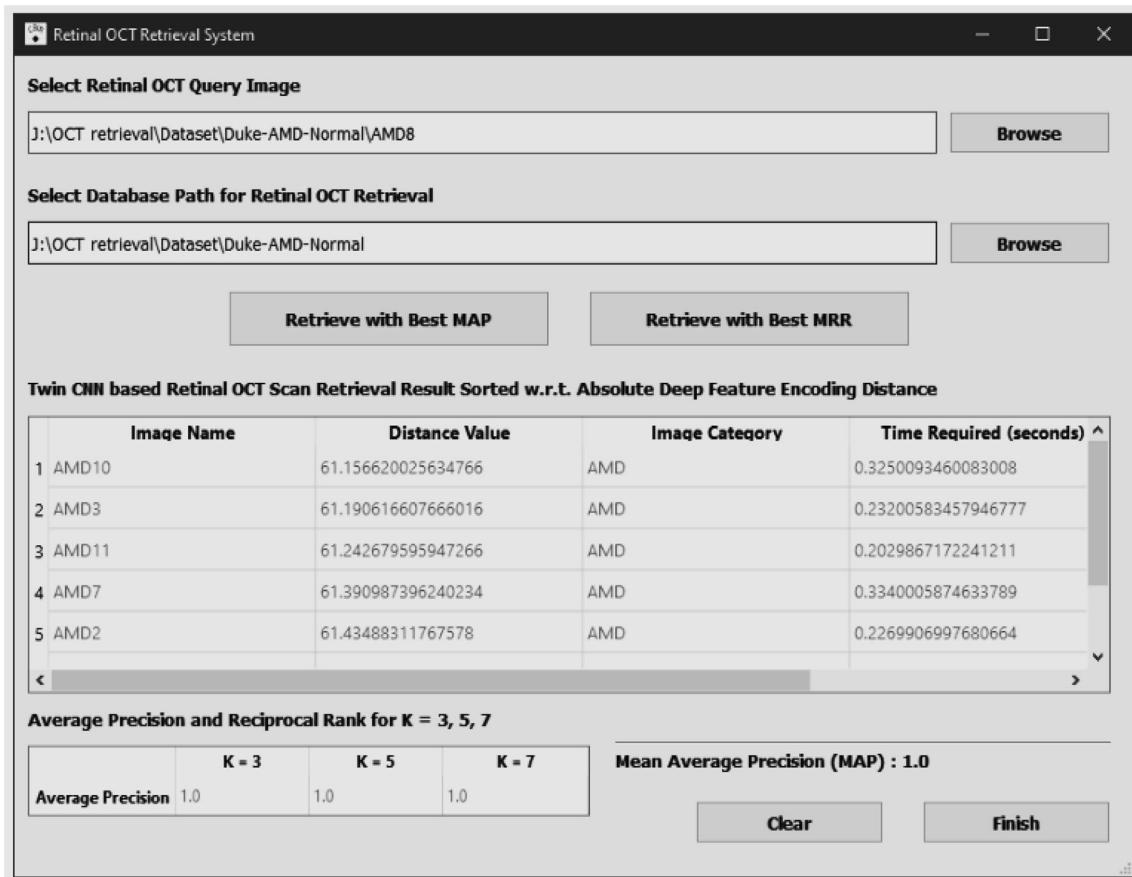
Figure 14. Sample retrieval for Mean Reciprocal Rank from AMD-DME-Normal database.

four different categories available in Mendeley dataset [18] are CNV, DME, DRUSEN and Normal. We have experimented with the most crucial four-class retrieval of OCT samples. System evaluation has been done by MAP and MRR with seed value = (2, 2) for MAP and seed value = (1, 2) for MRR as decided from seed value optimization previously. Considered Rank thresholds (K) are 3, 5 and 7. The results are depicted in table 7. As expected, critical four class retrieval performance is quite good. If we consider broader class retrieval of abnormal and normal classes only, the system performance improves. In that case, the system receives MAP of 0.72311 and MRR of 0.8756.

8. Discussion

Twin CNN based retrieval considering AMD and DME affected OCT and normal OCT dataset have been presented in this paper. We observed an impressive results while retrieving retinal OCT scans (AMD affected, DME affected and Healthy) through deep CNN feature map comparison. Usually, deep learning techniques are extremely computationally exhaustive and the training phase is very time

consuming making use of a large amount of training data with expensive GPUs. Most of the deep learning based processes require data augmentation for successful network training. In our approach, one-shot learning has been utilized to overcome these drawbacks. The system captures and compares pair-wise image characteristics from the convolution layer representation of the deep network. Similar weight and bias fastening between pair networks ensure that e representations of similar OCT might not be very different. Image resizing has been avoided as it may incur the cost of losing important information in regions of interest in OCT scans. Data augmentation is also not required as Twin CNN refrains from model training as part of its implementation criteria. Data encodings are compared at the deep CNN level of Twin CNN directly. Thus, the storage requirement for database featurerepresentation maps is also eliminated. Implementation of a quite shallow Twin CNN and pair-wise comparison for retrieval work helps in reducing both the time and computational complexities of the work. In this process, the system is capable of retrieving similar OCTs from the database. Retinal OCT retrieval system proficiency is evaluated using the most widely used retrieval metrics, MAP and MRR. Overall

Figure 15. Sample retrieval for Mean Average Precision from AMD-Normal database.

average MAP and MRR of the system are 0.631167 and 0.829607, respectively. If we consider only abnormal and normal retrieval, our system demonstrates MAP of 0.7571 and MRR of 0.9050 which are pretty good results. The mean average time for dataset preparation, similarity measurements and retrieval is 8.3 sec per query image with partly GPU/CPU execution. These are the specific highlights of the system implementation.

As per our knowledge, to date, no papers are available on retinal OCT retrieval. Thus, we are unable to provide any comparison table for previous retrieval evaluation metrics in the same domain. Instead, we verified the performance of OCT retrieval by arbitrarily taking 15 representative scans of each of the four different categories available in the Mendeley dataset [18] also. The system reflects MAP of 0.72311 and MRR of 0.8756 while experimented with the Mendeley dataset [18] if we consider broader category retrieval performance of abnormal and normal classes only. As Twin CNN dynamically compares two scans within the network, similar images should be less distant from each other. It is quite expected that the proposed dynamic Twin distance calculation does not vary much while tested on the

different datasets as it avoids creating a trained model as part of its implementation principle.

Execution time can further be improved by shifting the operation entirely to GPU execution and by applying CuPy implementation [20]. To utilize NVIDIA's CUDA architecture fully, we may shift the implementation to Python's CuPy interface with NVIDIA CUDA. Thus, full utilization of GPU acceleration has been achieved while computation within the neural network grows. CuPy is equivalent to NumPy interface efficient in handling array structure required for computation in deep learning but with the strength of increased speed gained from parallel computing on GPU [20]. As expected from one of our previously executed experiments regarding CuPy execution, we observed almost four-fold execution speed up in intermediate layer map extraction and almost two-fold speed up in similarity measurement and retrieval while executing the same experiments with the CuPy interface. These may be considered as the future scope of the work.

Category-wise retrieval results summary and execution timesheet are uploaded as supplementary material in excel files. The source code and video file of the work have been
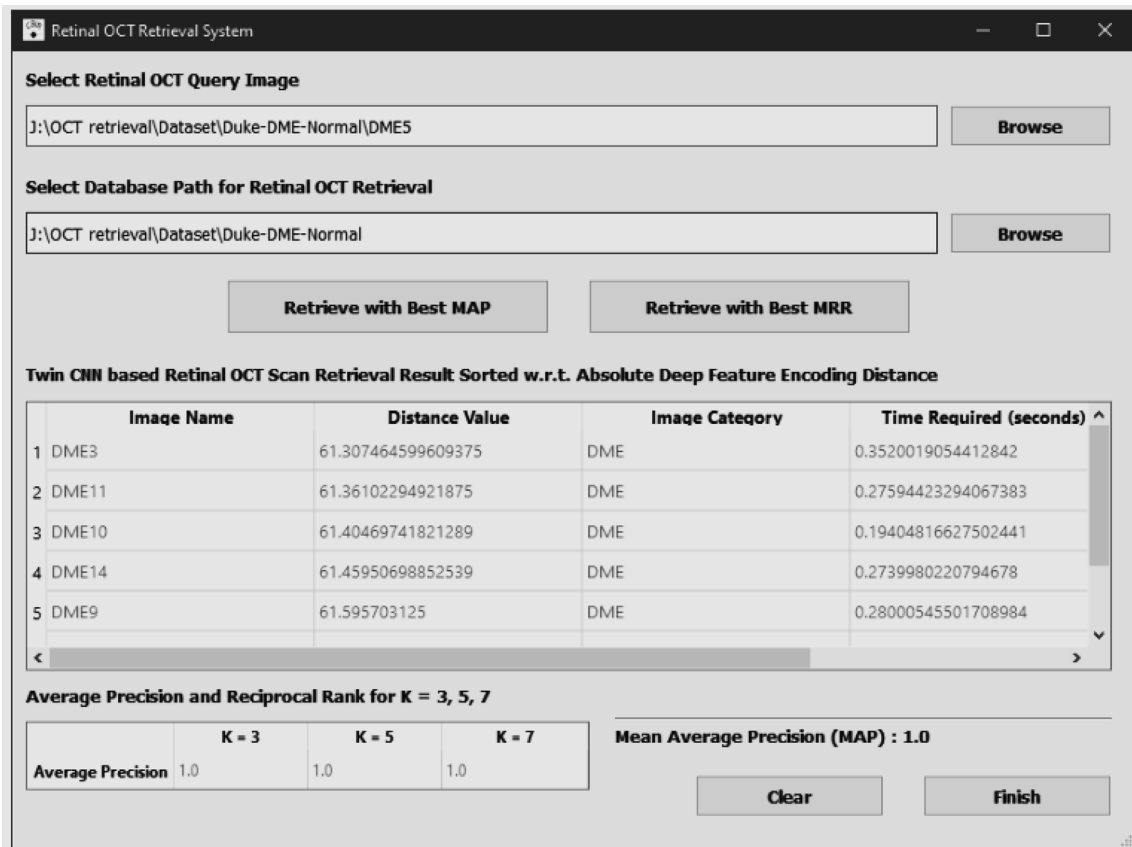
Figure 16. Sample retrieval for Mean Average Precision from DME-normal database.

Table 6. Duke execution time optimization at different convolution layers.

Retrieval categories	Total retrieval time(seed=2,2)	Average retrieval time(seed=2,2)	Total retrieval time(seed=1,2)	Average retrieval time(seed=1,2)
AMD-DME-normal	497.6706707	11.05934824	493.5794148	10.96843144
AMD-DME	218.5273552	7.284245173	225.8835175	7.529450583
AMD-normal	218.2457707	7.274859023	209.431468	6.981048934
DME-normal	239.2926657	7.976422191	237.0571992	7.901906641
Mean average Retrieval time		8.398718656		8.3452094

Table 7. Evaluation metrics for CNV-DME-DRUSEN-normal retrieval.

CNV-DME-DRUSEN-normal retrieval

	AP Seed = (2,2)	RR Seed = (1,2)
K=3	0.5131	0.7228
K=5	0.4927	0.7301
K=7	0.4555	0.7432
Mean values	0.4871	0.732033

uploaded to the following GitHub repository link: (https://github.com/shuvankarroy/OCT-retrieval).

9. Conclusion

Different retinal diseases severely affect ocular health in the developed world. Different imaging modalities of the retina help in detecting, diagnosing and providing treatment to these threats easily. We have presented a deep CNN based content based retinal image retrieval method to

provide a solution for retrieval of healthy, AMD and DME-affected retinal images. We evaluated our work on the Duke dataset and achieved an inspiring retrieval result as far as the method simplicity is concerned.

This work has a beneficial effect in the societal context as retrieval of affected retinal OCT scans is very important for early diagnosis and treatment of some fatal eye ailments. OCT retrieval work has the potential to eventually permit diagnostic image analysis in an automated and deterministic manner. Though OCT symptoms are very difficult to identify without human intervention, we generate impressive results with a robust retrieval system. This work contributes to the preventive measure of some permanent and critical changes associated with the progression of such diseases in the retina.

Acknowledgement

References

[1] Tzelepi M and Tefas A 2018 Deep convolutional learning for content based image retrieval. *Neurocomputing.* 275: 2467–2478.

[2] Saritha R R, Paul V and Kumar P G 2019 Content based image retrieval using deep learning process. *Cluster Comput.* 22: 4187–4200.

[3] Wan J, Wang D, Hoi SCH, Wu, P, Zhu J, Zhang Y and Li J 2014 Deep earning for content-based image retrieval: A comprehensive study. In: *Proceedings of the 22nd ACM International Conference on Multimedia*, pp. 157–166

[4] Koch G., Zemel R and Salakhutdinov R 2015 Siames neural networks for one-shot image recognition. In: *ICML Deep Learning Workshop* (vol. 2)

[5] Vinyals O, Blundell C, Lillicrap T and Wierstra D 2016 Matching networks for one shot learning. *Adv. Neural Inf. Process. Syst.* 29: 3630–3638.

[6] Chung YA and Weng WH 2017 Learning deep representations of medical images using siamese CNNs with application to content-based image retrieval. arXiv preprint arxiv:1711.08490

[7] Li MD, Arun NT, Gidwani M, Chang K, Deng, F, Little BP, Mendoza DP, Lang M, LeeSI, O'Shea A and Parakh A 2020 Automated assessment of COVID-19 pulmonary disease severity on chest radiographs using convolutional Siamese neural networks. medRxiv

[8] Ramachandra B, Jones M and Vatsavai R 2020 Learning a distance function with a Siamese network to localize anomalies in videos. In*: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2598–2607

[9] Zhang Z and Peng H 2019 Deeper and wider siamese networks for real-time visual tracking. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4591-4600

[10] Yin Z, Wen C, HuangZ, Yang F and Yang, Z 2019 SiamVGG-LLC: visual tracking using LLC and deeper siamese networks. In: *Proceedings of the IEEE 19th International Conference on Communication Technology*, pp. 1683–1687

[11] Wang Y, Zhang Y, Yao Z, Zhao R and Zhou F 2016 Machine learning based detection of age-related macular degeneration (AMD) and diabetic macular edema (DME) from optical coherence tomography (OCT) images. *Biomed Opt Express.* 7: 4928–4940.

[12] Lee C S, Baughman D M and Lee A Y 2017 Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. *Ophthalmol Retina.* 1: 322–327

[13] Lu W, Tong Y, Yu Y, Xing Y, Chen C and Shen Y 2018 Deep learning-based automated classification of multi-categorical abnormalities from optical coherence tomography images. *Transl Vis. Sci. Technol.* 7: 41–41.

[14] Kaymak S and Serener A 2018 Automated age-related macular degeneration and diabetic macular edema detection on oct images using deep learning. In: *Proceedings of the IEEE 14th International Conference on Intelligent Computer Communication and Processing*, pp. 265–269

[15] Wang J, Deng G, Li W, Chen Y, Gao F, Liu H, He Y and Shi G 2019 Deep learning for quality assessment of retinal OCT images. *Biomedi. Opt. Express.* 10: 6057–6072.

[16] Das U N 2016 Diabetic macular edema, retinopathy and age-related macular degeneration as inflammatory conditions. *Arch. Med. Sci.: AMS* 12: 1142.

[17] Dataset link:- http://www.duke.edu/∼sf59/Datasets/2014_BOE_Srinivasan.zip. Accessed on- 12 Oct 2020 (07:58 AM)

[18] Dataset link:- https://data.mendeley.com/datasets/rscbjbr9sj/2. Accessed on 28 March 2021 (07:30 AM)

[19] Ronneberger O, Fischer P and Brox T 2015 U-net: Convolutional networks for biomedical image segmentation. In: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241

[20] Nishino ROYUD and Loomis SHC 2017 CuPy: A NumPy-compatible library for NVIDIA GPU calculations. In: *Proceedings of the 31st Confernce on Neural Information Processing Systems*, p. 151