Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Assistive lesion-emphasis system: an assistive system for fundus image readers

Samrudhdhi B. Rangrej
Jayanthi Sivaswamy

Assistive lesion-emphasis system: an assistive system for fundus image readers

Samrudhdi B. Rangrej* and Jayanthi Sivaswamy
International Institute of Information Technology, Hyderabad, India

Abstract. Computer-assisted diagnostic (CAD) tools are of interest as they enable efficient decision-making in clinics and the screening of diseases. The traditional approach to CAD algorithm design focuses on the automated detection of abnormalities independent of the end-user, who can be an image reader or an expert. We propose a reader-centric system design wherein a reader's attention is drawn to abnormal regions in a least-obtrusive yet effective manner, using saliency-based emphasis of abnormalities and without altering the appearance of the background tissues. We present an assistive lesion-emphasis system (ALES) based on the above idea, for fundus image-based diabetic retinopathy diagnosis. Lesion-saliency is learnt using a convolutional neural network (CNN), inspired by the saliency model of Itti and Koch. The CNN is used to fine-tune standard low-level filters and learn high-level filters for deriving a lesion-saliency map, which is then used to perform lesion-emphasis via a spatially variant version of gamma correction. The proposed system has been evaluated on public datasets and benchmarked against other saliency models. It was found to outperform other saliency models by 6% to 30% and boost the contrast-to-noise ratio of lesions by more than 30%. Results of a perceptual study also underscore the effectiveness and, hence, the potential of ALES as an assistive tool for readers. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.4.2.024503](https://doi.org/10.1117/1.JMI.4.2.024503)]

Keywords: computer-assisted diagnostic; saliency; selective enhancement; color fundus image; convolutional neural network; gamma correction.

Paper 16277RR received Jan. 5, 2017; accepted for publication May 8, 2017; published online May 24, 2017.

1 Introduction

Recent advances in medical imaging have enabled the involvement of a variety of practitioners in the diagnosis process. In the past, medical experts who examined patients also analyzed images; the shortage of highly trained experts has led to the creation of new types of services. One such service allows images to be sent to a central location where they are read and observations/diagnosis are recorded and returned. Centers offering such services are called reading centers¹ and they are staffed with readers and some experts. Readers are trained only to examine images and write reports whereas experts are medically trained and, hence, can also diagnose based on evidence found in images and any available history of a patient maintained by the facility. Reading center-like settings play an important role in “screening” and “triage.”

Clinical screening is aimed at identifying individuals who may be at risk for a disease since early detection is preferred in effective disease management. Breast cancer screening for women in the age group of 45 to 54 years is one such example.² In resource-constrained settings, screening is done in camps by a field team, and the images are brought/transmitted to reading centers for experts to analyze and recommend further in-depth examination at a base hospital.³

Triage on the other hand is a practice followed by clinics to prioritize patients for experts' attention. A trained practitioner orders a preliminary test, a reader (semiexpert) analyzes the images, and the report is used to decide the priority of a patient. This practice helps to make the work-flow efficient and speedup

the diagnostic process. Acute stroke triage is an example of this.⁴

Image reading is a tedious task yet requires precision as it is critical for diagnosis. Fatigue or inattention causes readers to miss inconspicuous/subtle lesions leading to underreporting and incorrect diagnosis. Computer-assisted diagnostic (CAD) tools aim at addressing this problem. CAD tools draw readers' attention to abnormal regions typically by displaying augmented circles/markers on the abnormal regions.⁵ More recently, retrieving and displaying similar past cases along with the diagnosis have been shown to be effective.⁶ Augmentation-based assistance can potentially clutter an image especially when abnormalities are present in abundance and when different types of abnormalities are also proximal in the image. Retrieval-based assistance requires a large amount of storage to store previous cases and heavy computations to calculate similarity scores with large datasets. We argue that an alternate solution is to draw a reader's attention by emphasizing the abnormalities locally and making them more prominent while leaving the background tissue unaltered. This is motivated by the fact that the visual system draws attention to salient locations characterized by distinctive features, such as color and orientation.^{7,8} Boosting the contrast of such salient locations has been shown (in the case of natural scenes) to attract one's attention.^{9–11} Lesion-emphasis is both computationally efficient and clutterless. In this paper, we take this alternate reader-centric approach and propose an assistive lesion-emphasis system (ALES), which employs saliency of a region to determine the amount of its emphasis. The ALES concept is demonstrated for retinal abnormalities called diabetic

*Address all correspondence to: Samrudhdi B. Rangrej, E-mail: rangrej.bharat@research.iit.ac.in

retinopathy (DR), an eye disease observed in patients suffering from diabetes. Abnormalities specific to DR include two major types of lesions: lipid leakage called hard exudate (HE) and blood leakage called hemorrhage (HM). In color retinal images, the former appears as a bright blob while the latter appears as a dark splotch. The proposed ALES for DR has two stages: (i) saliency computation and (ii) lesion-emphasis.

Technical contributions toward the development of ALES are as follows: (a) approach to image reading. An assistive solution is proposed based on saliency modeling. (b) A convolutional neural network (CNN)-based saliency model with a loss function. Proposed architecture is inspired by the Itti-Koch model,⁷ which is extended and adapted to fundus images for DR analysis. (c) Saliency-based lesion-emphasis, this aids a least-obtrusive, yet effective assistance, as only lesions are emphasized. The paper is organized as follows. The design of the two stages is presented followed by their assessment.

2 Saliency Computation

2.1 Background

Computational modeling of visual saliency has been a subject of research for awhile. Existing computational models range from biologically plausible ones^{7,12} to information- and decision-theoretic,^{13,14} graphical,^{15,16} spectral-analysis,^{17,18} pattern classification based,^{19–21} etc. These models have been employed in many computer vision tasks, such as object recognition,²² image tracking and retrieval,²³ segmentation,²⁴ and image/video compression.²⁵ Such models, however, are developed to compute saliency for natural images where (a) there are only a few objects of interest, (b) the target objects are mostly in the center, and (c) the background is free of clutter/textures. Medical images do not fit this category. Also, medical image analysis, being highly domain-specific area, requires separate attention to each modality/disease. For example, the model developed to generate saliency of a tumor in brain MRI will not work for the DR lesions in color fundus images. Hence, various task-specific saliency models have been developed for different applications, including medical image classification and retrieval,²⁶ plane identification from three-dimensional (3-D) ultrasound,²⁷ registration of dynamic renal MR images,²⁸ prostate MRI segmentation,²⁹ and saliency modeling for glioblastoma multiforme tumor.³⁰

Our interest lies in developing a saliency model for DR lesions, which can be used for lesion-emphasis in ALES. Saliency computation for DR images is a challenging task due to artifacts and nonuniform illumination (see Fig. 1). A good saliency model has to learn to discriminate artifacts

from true lesions and reject the former. Computational saliency models have been reported for only HE.^{31,32} Our aim is to develop saliency models for both HE and HM. This is done using a CNN inspired by the Itti-Koch saliency model.⁷ In this model, center-surround (CS) difference maps are computed in the color, intensity, and orientation dimensions at different scales using pyramids, and a linear combination of these maps is defined as the saliency. The proposed CNN architecture is derived from this model.

2.2 Method

The task-specific saliency has been traditionally modeled as a weighted combination of low-level feature maps, where weights are learned from prior knowledge.^{33,34} A neural network-based extension of Itti-Koch model has already been shown to be effective in handling normalization and feature competition with biologically plausible dynamics.³⁵ Our approach is to use a CNN to (i) fine-tune standard orientation and CS filters, (ii) learn new filters, and (iii) learn the weights for combining the feature maps.

CNN is a type of feedforward neural network, which is biologically inspired. The important feature of CNN is local connectivity and weight sharing, i.e., each neuron in the current layer is locally connected to a small set of neurons from the previous layer. Synaptic weights used for the local connectivity are same for all the neurons, which enables convolution property. The architecture we propose has three building blocks: a convolutional layer that performs filtering of activations with weights, maxpooling that downsamples the image by retaining the maxima in a local neighborhood, and an activation function that applies a nonlinear transformation on the intensity values of an image. We use the rectified linear unit (ReLU) as an activation function.³⁶

2.2.1 Convolutional neural network architecture

The architecture of the proposed model has five stages as shown in Fig. 2. It is carefully designed to share similarities with a standard Itti-Koch saliency model. Similarities are as follows: stage 1 is equivalent to color/intensity pyramid, stage 2 serves the purpose of orientation pyramid, stage 3 facilitates further computation by stacking the feature maps of previous layers, stage 4 models the CS difference pyramid, and stage 5 is identical to final normalization and combination of all the maps. Parallel fine-tuning of Itti-Koch filters and learning of new filters are carried out at each stage. Model architecture is described below.

Fig. 1 From left to right: Retinal image with DR lesions, dark artifacts, bright artifacts, and varying illumination.

Fig. 2 Proposed architecture with stage-wise description of types of layers, filter size, padding size, stride, and activation function.

An image I of size $(288 \times 360 \times 3)$ forms the input to stage 1. Stage 1 has one convolutional layer with 24 filters, each of size $(1 \times 1 \times 3)$ and produces an activation map O_1 as output. These filters learn 24 different color transformations

$$O_1 = W_{\text{color}} * I + b_{\text{color}}. \quad (1)$$

Stage 2 has two parallel convolutional layers, each operating on O_1 to produce two independent output activation maps O_{21} and O_{22} . Each convolutional layer has 18 filters of size $(15 \times 15 \times 24)$. The first convolutional layer is initialized with orientation filters and the second one is initialized with random filters. Weight initialization for W_{orient} is done as follows. Eighteen two-dimensional (2-D) orientation filters of resolution 20 deg were generated. Each filter was repeated and stacked to generate W_{orient} . O_{21} and O_{22} are computed as follows:

$$O_{21} = W_{\text{orient}} * O_1 + b_{\text{orient}}, \quad (2)$$

$$O_{22} = W_{\text{rnd1}} * O_1 + b_{\text{rnd1}}. \quad (3)$$

The maps O_1 , O_{22} , and O_{21} are stacked in stage 3 and max-pooling in (2×2) neighborhood is applied to generate a feature map O_3 of size $(144 \times 180 \times 60)$. Using $[\cdot, \cdot, \cdot]$ notation for stacking

$$O_3 = \max \text{pooling}([O_1, O_{21}, O_{22}]). \quad (4)$$

Stage 4 also has two parallel convolutional layers that operate on O_3 to produce O_{41} and O_{42} as two independent outputs. Both layers have six filters of size $(10 \times 10 \times 60)$. The first convolutional layer is initialized with CS filters and second with random filters. Weight initialization for W_{CS} is done using six 2-D CS filters, which are generated as follows:

$$\text{CS1} = \pm(G_1 - G_4), \quad (5)$$

$$\text{CS2} = \pm(G_2 - G_5), \quad (6)$$

$$\text{CS3} = \pm(G_3 - G_6), \quad (7)$$

$$\text{CS4} = \pm(G_1 - G_5), \quad (8)$$

$$\text{CS5} = \pm(G_2 - G_6), \quad (9)$$

$$\text{CS6} = \pm(G_3 - G_7), \quad (10)$$

where G_n is a Gaussian filter with mean 0 and variance n . In these equations, the positive sign is used for HE while the negative sign is used for HM saliency. Each CS filter was repeated and stacked to make W_{CS} . O_{41} and O_{42} are computed as follows:

$$O_{41} = W_{\text{CS}} * O_3 + b_{\text{CS}}, \quad (11)$$

$$O_{42} = W_{\text{rnd2}} * O_3 + b_{\text{rnd2}}. \quad (12)$$

Stacking of O_{41} and O_{42} generates O_4 in stage 5. A convolutional layer with a filter of size $(1 \times 1 \times 12)$ operates on the stack O_4 to produce a single image O_5 . This stage learns the final weighted combination of all feature maps. Finally, an ReLU activation is applied to get the desired output, which is a final grayscale image O of size (144×180)

$$O_4 = [O_{41}, O_{42}], \quad (13)$$

$$O_5 = W_{\text{combination}} * O_4 + b_{\text{combination}}, \quad (14)$$

$$O = \max(0, O_5). \quad (15)$$

The ReLU activation function, unlike “sigmoid” or “tanh,” is linear in the positive range thus ensuring linear mapping (no saturation) for positive saliency while clipping the negative saliency to zero as desirable.

2.2.2 Loss function

Training a CNN is an unconstrained optimization problem that aims to minimize a loss function, which compares the system output with ground truth (GT). Conventional loss functions for regression assume same numeric range for both. However, in the present case, the ReLU activation allows $O \in [0, \infty)$, whereas

the GT saliency values are in the range of $[0,1]$. Hence, we define a new loss function as follows:

$$L(X, Y) = \frac{1}{N} \sum_{x \in X, y \in Y} \beta x e^{-\alpha y} + (1-x)(1 - e^{-\alpha y}), \quad (16)$$

where the tuple (x, y) denotes the (GT saliency, output) pixel pair, N is the total number of pixels, and β is a weight used to handle class-imbalance. α controls the threshold y_0 such that, a low loss is achieved for two conditions: (i) low GT saliency value ($x \in [0, 0.5]$) and a subthreshold output and (ii) a high GT saliency ($x \in [0.5, 1]$) and a suprathreshold output (see Fig. 3).

The loss function has a saddle point at $(x, y) = (0.5, y_0)$. The threshold value y_0 is found by substituting $x = 0.5$ in $L(X, Y)$ [or differentiating $L(X, Y)$ w.r.t. x and equating to 0].

$$y_0 = \frac{1}{\alpha} \log(\beta + 1). \quad (17)$$

Ideally, zero GT saliency should correspond to nearly zero output values, whereas it suffices to have high GT saliency

($x = 1$) correspond to a large range of output values. This is achievable with the threshold y_0 tending to zero or equivalently, very large α .

The proposed CNN was trained for HE and HM saliency separately. We denote HE- and HM-specific saliency models as S_{HE} and S_{HM} , respectively. Data and computational resources used for training of CNN are discussed in Sec. 5.1.2. Computed saliency maps are used to emphasize lesions locally as described in Sec. 3.

3 Lesion-Emphasis

3.1 Background

In DR reading centers, readers scrutinize images and assign a DR stage to the image using the Early Treatment of Diabetic Retinopathy Study (ETDRS) standard.³⁷ Staging is based on the following guidelines: (a) DR grade (on a scale 1 to 4) is proportional to the number of dark lesions and (b) diabetic macular edema (DME) grade (scale 1 to 3) is proportional to the distance between macula and nearest HEs (see Fig. 4). The grade determines the type of advice given to a subject being screened with some requiring immediate referral. Thus, a failure

Fig. 3 Loss function in the absence of class-imbalance ($\beta = 1$). (a) 3-D view of a loss function. Saddle point is shown in red color. (b) 2-D view of a loss function. Above surface is sampled at different values of x . y -projection of saddle point $(0.5, y_0)$ is also shown.

Fig. 4 Original image and GT: (a, b) DR stage 2, (c, d) DR stage 3, (e, f) DME stage 2, and (g, h) DME stage 3.

of a reader to attend to “all” dark lesions or the bright lesion “nearest” to the macula can have serious implications as it leads to an incorrect stage assignment to the image. The approach taken in ALES, therefore, is to increase the local contrast of the lesions and make them more prominent. A contrast-enhanced lesion will successfully draw a reader’s attention and hopefully reduce the rate of misdiagnoses.

Existing work on the enhancement of retinal images is based on illumination/contrast correction,^{38–42} contourlet,⁴³ and histogram equalization and matching.^{44,45} These methods are primarily aimed as a preprocessing stage for CAD development and none have been aimed at readers or experts. Hence, they often introduce textures and color shifts. These methods are developed to correct global variations and improve contrast at a global (rather than local) level. ALES aims to emphasize the lesion locally without altering the global statistics of the image and do this by using saliency information. Saliency-based local enhancement techniques have been reported for natural images. These include optimization to match object and target saliency,¹¹ luminance/chrominance adjustment based on saliency,¹⁰ iterative addition of point variation values,⁴⁶ de-emphasis of background texture,⁴⁷ and saliency-weighted luminance correction.⁴⁸ Most of these techniques are computationally complex. Since ALES is aimed at readers in screening or triage scenario, the lesion-emphasis needs to be done with a simple, fast, and computationally efficient method.

3.2 Method

We propose a spatially varying method that achieves lesion-specific emphasis by modifying a global contrast stretching method. This is done by choosing a parametric, spatially invariant method and allowing the parameter to be a function of the local saliency. We start with gamma correction, which is a well-known spatially invariant, nonlinear, contrast stretching method. Gamma correction is defined as

$$I_C(x, y) = I_O(x, y)^\gamma, \quad (18)$$

where $I_C(x, y)$ and $I_O(x, y)$ (normalized between 0 and 1) are the corresponding pixels from corrected and original images, respectively. And γ is the “global” parameter. This typically is used to match the dynamic contrast of an image to that of a display device. A choice of $\gamma > 1$ pushes intensity values to a lower range, which results in the darkening of the entire image, whereas $\gamma < 1$ pushes intensity values to a higher range thus, brightening the image. Gamma correction on the sample fundus image can be seen in Fig. 5. It can be observed that global correction fails to emphasize the lesions locally.

The above operation can be made to be spatially varying by defining gamma as a function of the saliency at a point as follows:

Fig. 5 Gamma correction: (a) original image, (b) corrected image with $\gamma = 2$, and (c) corrected image with $\gamma = 0.5$.

$$I_C(x, y) = I_O(x, y) \left[1 - \frac{S_{HE}(x, y)}{a} + \frac{S_{HM}(x, y)}{b} \right], \quad (19)$$

where a and b are the normalizing parameters. Ideally, the background pixels should have zero saliency in both S_{HE} and S_{HM} and, hence, $\gamma = 1$ for such pixels, which implies no correction. Pixels from regions containing HM should have $S_{HE}(x, y) = 0$ and $S_{HM} > 0$, so $\gamma > 1$ resulting in a darkening of the region. Pixels from regions containing HE will have $S_{HE}(x, y) > 0$ and $S_{HM} = 0$, so $\gamma < 1$ should lead to a brightening of the region.

4 Material

The publicly available DIARETDB1,⁴⁹ DRiDB,⁵⁰ and DMED⁵¹ datasets were used for training and testing the proposed saliency computation stage. Preprocessing consisted of illumination correction,⁵² fundus extension to remove the black mask region,⁵³ detection of vessels,⁵⁴ and optic disk using circular Hough transform. The last two were subsequently in-painted to reduce false detections. Figure 8(b) shows the result of preprocessing on a sample image Fig. 8(a). All images were downsampled to size 288×360 . This size has been chosen to be half the size of DRiDB images, which has the lowest resolution. This should minimize both distortion of the image (aspect ratio) and the computational cost. Images were normalized to have zero mean value and unit variance.

The chosen datasets provide different types of lesion markings whereas our CNN training requires a lesion-level GT. Only DMED provides lesion markings (pixel level). Both DIARETDB1 and DRiDB datasets provide markings as regions around lesion(s) with the former providing markings of four experts as a heatmap (leftmost image in Fig. 6) and the latter providing marking from one expert as a binary map. To derive consistency in GT across the datasets, the markings were processed as follows (see Fig. 6). The heatmap was thresholded at 50% agreement to derive a binary mask. This mask was multiplied with the original image to extract the lesion regions (second image from right in Fig. 6); this was finally thresholded to derive the final binary, lesion-level GT (rightmost image in Fig. 6). To retain HEs (HMs) in the final GT, we retain pixels above (below) a threshold. This GT was downsampled to 144×180 for training. Training for S_{HE} was done directly with the derived GT whereas for S_{HM} a Gaussian was convolved with the GT as HMs are more diffused in appearance.

Table 1 presents the details of the datasets used. Since the aim is to derive a saliency model for abnormalities, only abnormal images were used in training. One hundred and twenty two

Table 1 Dataset description.

Datasets	DIARETDB1	DRiDB	DMED
Total number of images	89	50	169
Images containing HE	48	32	54
Images containing HMs	54	31	—

Table 2 Parameter values used for training.

Parameters	S_{HE}	S_{HM}
L2 regularization	0.01	0.01
Learning rate	0.0005	0.0005
Nesterov momentum	0.7	0.6
Decay	5×10^{-8}	1×10^{-4}
β	225	111
α	500	500
Batch size	8	8

(of 134) images with HEs and 72 (of 85) images with HMs were used for training. Training and testing were done on whole images. Given that the dataset size is not large, (i) the training set size was chosen to be larger to ensure a variety of data for learning and (ii) online data augmentation was done using a variety of transformations. Random rotation between 0 deg and 30 deg, random vertical shift between 0 and 57 pixels (20% of height of an image), random horizontal shift between 0 and 72 pixels (20% of width of an image), and occasional horizontal/vertical flips are used for augmentation. Training was done on NVIDIA GTX 970 GPU, with 4 GB of RAM for 10,000 epochs by minimizing the loss function in Eq. (16) using a stochastic gradient descent optimizer. We experimented with a number of learning schemes and finally determined the suitable values of parameters as given in Table 2. Training time was ~ 5 days. Cross-validation was not performed due to excessive training time.

Fig. 6 Procedure to obtain lesion-level GT from regional marking.

5 Evaluation of Assistive Lesion-Emphasis System

Assessment of ALES is done stage-wise for both abnormal and normal cases.

5.1 Saliency Computation

5.1.1 Evaluation of trained filters

CNN generates output by convolving a set of filters with the input. These filters are key components for computation of activation maps for saliency. Evolution of filters was assessed qualitatively over the period of training. It was observed that the orientation filters underwent very small changes, whereas CS filters changed considerably during training [see Figs. 7(a) and 7(b)]. Figure 7(b) shows progression in tuning of three channels of the CS filters of S_{HE} . It can be seen that the pattern of tuned filters is similar to the difference of multiple (>2) Gaussian filters. Figures 7(c)–7(d) show how three filters (single channel only) with random initialization change during training in stage 2 and 4. The sample filters from stage 2 can be expected to give a higher response for linear structures, bifurcations, and

bright spots (with dark top), respectively. The sample filters from stage 4 are similar to Gaussian filters or blob detectors.

5.1.2 Evaluation of saliency

The performance of the saliency models was evaluated against seven existing computational saliency models: Itti–Koch,⁷ SR,¹⁸ AIM,¹⁴ GBVS,¹⁶ Torralba,⁵⁵ Judd,¹⁹ and Rare.⁵⁶ These state-of-the-art saliency models are selected such that each model represents a class of saliency models, i.e., biologically plausible ones, information- and decision-theoretic, graphical, spectral-analysis based, and pattern classification based, as mentioned in Sec. 2.1. Itti–Koch model is biologically inspired and combines orientation, intensity, and color information in a purely stimulus-driven manner. SR model extracts spectral residual in the log-spectrum of an image as variation from natural image statistics; computed spectral residual is converted to a spatial domain saliency map. AIM is based on information theory and uses joint likelihood between basis coefficients (derived using filters) and the density of these coefficients as a measure of self-information. GBVS is a graph-based model that forms activation maps on a set of features and normalize

Fig. 7 Evolution of the filters during training of HEs saliency: (a) three channels of W_{orient} from stage 2, (b) three channels of CS3 filters from stage 4, (c) three channels of random filters from stage 2, and (d) three channels of random filters from stage 4.

them to highlight conspicuity. Torralba uses perceptual dimensions, such as naturalness, openness, roughness, expansion, and ruggedness, to represent dominant spatial structure of a scene. Judd's model is based on pattern classification and uses feature-based classification of various regions of a scene. The rare model uses bottom-up features capturing structure rarity and a top-down feature derived from mouse-clicks on the image by participants in an experiment. Saliency maps for all these existing models were computed using publicly available codes using default parameter settings.

A sample image, its GT, and the computed saliency maps are shown in Fig. 8 for HEs and Fig. 9 for HMs. Ideally, a computed saliency map should appear sparse and similar to the GT as this would be ideal for lesion-emphasis in the next stage of ALES. It can be seen from the computed maps in Fig. 8, only S_{HE} , SR, and Torralba have either of these desired characteristics. Of these, the map with Torralba is the least sparse. In general, it is more difficult to differentiate HMs from the background and blood vessel fragments. This is seen from the fact that,

of all the computed maps in Fig. 9, only those with ours and SR models have the desired features.

Computational saliency for healthy retinas is also important for normal versus abnormal decisions. Hence, S_{HE} and S_{HM} were tested on normal cases. Sample images and the computed maps are shown in Fig. 10. The almost blue maps indicate that both models give virtually zero saliency values for the pixels representing healthy tissue. Weak responses in S_{HE} are seen occasionally near the peripapillary region, which is due to the presence of a hyperreflective region.

Quantitative evaluation was performed on both abnormal and normal images (see Table 3). A set of 33 normal images, used for testing both S_{HE} and s_{HM} , was taken from DRiDB⁵⁰ and DIARETDB0.⁵⁷ A set of 12(13) abnormal images, used for testing $S_{HE}(S_{HM})$, was taken from DRiDB,⁵⁰ DMED,⁵¹ and DIARETDB1⁴⁹ as mentioned in Sec. 4. The two sets of abnormal images, one for S_{HE} and another for S_{HM} , have a few common images as they contain both HEs and HMs. The metrics used for the evaluation are: receiver operating characteristic

Fig. 8 HE saliency: (a) original color fundus image, (b) preprocessed image, and (c) GT. Computed saliency maps of (d) proposed, (e) Itti-Koch, (f) SR, (g) GBVS, (h) AIM, (i) Rare, (j) Torralba, and (k) Judd.

Fig. 9 HM saliency: (a) original color fundus image, (b) preprocessed image, and (c) Gaussian convolved GT. Computed saliency maps of (d) proposed, (e) Itti-Koch, (f) SR, (g) GBVS, (h) AIM, (i) Rare, (j) Torralba, and (k) Judd.

Fig. 10 Predicted saliency for normal cases. (a) Normal color fundus image and saliency maps generated by (b) S_{HE} and (c) S_{HM} .

Table 3 Number of images in the test set of S_{HE} and S_{HM} .

Model	Abnormal images	Normal images	Total number of images
S_{HE}	12	33	45
S_{HM}	13	33	46

(ROC) curves, area under the ROC curve (AUC), false positive rate (FPR) versus saliency, and precision versus saliency. The binary lesion-level GT is used as the reference standard. In case of S_{HE} , GT is already binary hence used directly. In case of S_{HM} , GT is continuous hence thresholded at 0.5 value to obtain binary GT. Lesion-pixels, if correctly/incorrectly detected with reference to GT, are considered true positive (TP)/false negative (FN). Background pixels, if correctly/incorrectly detected with reference to GT, are considered true negative (TN)/false positive (FP).

ROC is a graphical representation of achieved true positive rate (TPR = $\frac{TP}{TP+FN}$) versus FPR (FPR = $\frac{FP}{FP+TN}$) for varying threshold values. The ROC and AUC are presented in Fig. 11 and Table 4. The results show that proposed model outperforms all other models. It can be seen that the Judd saliency model has nearly the same AUC value as S_{HM} . This can be explained as follows. In normal cases (absence of any HM), the model successfully rejects background pixels as nonsalient regions. Since the test set is skewed toward normal images, the overall performance is good and almost at par with S_{HM} . However, this model's use of multiscale analysis causes the saliency response to be spatially extended rather than highly localized as can be seen in Fig. 9(k). Consequently, the performance of Judd model is compromised for abnormal images. This can be verified from analysis of FPR discussed next.

As noted in Ref. 58, AUC aids the assessment of a model's ability to assign high saliency values to lesions, but it fails to give any insight into its handling of nonlesion regions, such as background and artifacts, where FP can be created. Low saliency should ideally correspond to nonlesion regions. The

Table 4 Comparison of AUC scores.

	HE	HM
S_{HE}	0.962	—
S_{HM}	—	0.918
SR	0.799	0.678
Torralba	0.853	0.708
Rare	0.794	0.621
Itti-Koch	0.688	0.526
AIM	0.723	0.589
GBVS	0.667	0.774
Judd	0.367	0.896

FP rate (FPR = $\frac{FP}{FP+TN}$) for low saliency values may be a better metric for this purpose. FPR was computed by thresholding the computed saliency with a step-size of 1% of the maximum saliency value and comparing with GT (see Fig. 12). Both S_{HE} and S_{HM} have a low FPR almost over the entire range of saliency values, which indicates they handle the background (nonlesion) regions well. On the other hand, barring SR, all existing models have a relatively high FPR for lower saliency range (Wilcoxon signed-rank test: p value $\ll 0.01$). In the case of HMs, Torralba and Judd model have higher FPR (in the low saliency range) than other models, which indicates its inability to handle nonlesion regions.

FPR demonstrates how well a saliency model can reject background pixels. On the other hand, precision ($\frac{TP}{TP+FP}$) helps assess how often a model correctly detects lesions. Precision is also known as positive predictive value in some literature. As saliency increases, the proportion of correctly detected lesion-pixels among all the detected pixels (and hence precision) can be expected to increase. The precision versus saliency plots for the two types of lesions are shown in Fig. 13. The wide

Fig. 11 ROC: (a) HE saliency and (b) HM saliency. In each case, the plots computed with seven existing saliency models are also shown for comparison.

Fig. 12 False positive rate versus saliency: (a) HEs and (b) HMs.

Fig. 13 Positive predictive rate/precision versus saliency: (a) HEs and (b) HMs.

Fig. 14 Combined saliency for HE and HM. (a, c) Original images with lesions and (b, d) combined saliency maps for HE (green) and HM (purple) shown overlaid on the original image.

difference in precision levels for S_{HE} and S_{HM} attests to the fact that detection of HMs is more challenging, the low precision caused by confusion between vessel fragments (persisting after inpainting) and HMs. Further, it can also be seen that both S_{HE} and S_{HM} outperform other saliency models (Wilcoxon signed-rank test: p value $\ll 0.01$).

The two saliency models can be combined to derive a common saliency map for a given image. Examples of such maps are shown for sample images in Fig. 14. Here, the salient regions are color coded with green (HE) and blue (HMs). It can

be observed that by and large, the background and nonlesion regions, including blood vessels, are nonsalient. Some high saliency is seen in the middle of the optic disk where blood vessels converge, which is erroneous.

5.2 Lesion-Emphasis

Qualitative results of ALES are shown in Fig. 15 for sample images. It can be observed that dark lesions in poorly illuminated regions are not visible in the original images but are

Fig. 15 ALES output for abnormal images. (a) and (e) original images, (b) and (f) corrected images, (c) and (g) HE GT, and (d) and (h) HM GT.

Table 5 Average CNR.

	DIARETDB1		DRiDB	
	HE	HM	HE	HM
Original image	3.97	2.91	2.98	3.05
ALES image	5.67	4.02	4.02	4.63

more visible in the processed results, which makes counting of dark lesion easier. HEs, which are close to macula as per GT, are not clearly visible in the original images but are prominent after emphasis. It is notable that the image in Fig. 15(e) contains a dark artifact near the optic disk, which is not enhanced by ALES as desirable. Vessel regions are emphasized incorrectly, which may be undesirable for fully automatic CAD. A reader using ALES will know to ignore it.

Quantitative evaluation of ALES was done using contrast-to-noise ratio (CNR) of lesions

$$\text{CNR} = \frac{|m_f - m_b|}{\sigma_b}, \quad (20)$$

where m_f and m_b are the mean intensity of foreground (lesions) and background, respectively, and σ_b is the standard deviation of background intensity. CNR was computed for images from two datasets containing both HM and HE. CNR is computed based on the pixel values in foreground and background regions. Pixels marked as lesion region in GT were considered as foreground, and the remaining pixels in a bounding box enclosing the lesion region were considered as background. The CNR values are presented in Table 5. CNR is improved by more than 30% in each case. Lesions with improved local CNR should attract reader's attention more than the ones in the original image.

ALES output for a sample normal image is shown in Fig. 16. ALES can be seen to introduce minimal artifacts, as desirable.

6 Perception Studies

Two perception studies were conducted to measure the effectiveness of ALES in assisting readers. The first study was aimed at measuring the effectiveness for a global-level decision about an image, which is a classification of an image as a normal versus abnormal case. The second study was aimed at measuring the

effectiveness of ALES for a local-level decision about an image, such as locating all HMs and an HE in proximity of macula. This task is designed in accordance with the guidelines given by the ETDRS.³⁷

6.1 Stimuli

Stimuli (images) for both experiments were selected by a local expert. Thirty pairs of (original and ALES output) images, with equal number of normal and abnormal cases, were used in the first study. Ten pairs of abnormal images were used in the second experiment. Randomized studies were conducted in two sessions with a gap of 4 days to ensure least fatigue for participants.

One image from each pair (original and ALES output) was randomly selected and grouped as set **A**. The remaining images were grouped as set **B**. Hence, set **A** and **B** are mutually exclusive. Set **A** was used for the first session and set **B** for the second. This was done for both the studies.

6.2 Subjects

Twelve engineering student volunteers were recruited for the studies. They were given a brief introduction to DR using some images prior to the first session.

6.3 Experiment Design

For both studies, images were shuffled and displayed on a Lenovo monitor of size 1366×768 pixels, and the interval between the response for an image and the display of the next image was 2 s.

Study 1: Subjects were shown images and asked to press key "A" for abnormal and key "N" for normal to indicate their decisions about the images.

Study 2: Subjects were shown images of abnormal cases and were instructed to mark (i) all HMs and (ii) click on the HE, which is closest to macula. HMs are of interest in detecting DR. Macula is the site of high acuity color vision and therefore presence of HEs in its proximity is of interest to detect DME.

6.4 Results

The results for the first and second studies are shown in Tables 6 and 7, respectively. The results in Table 6 indicate that ALES is effective as the accuracy of global-level decision is significantly

Fig. 16 ALES output for normal image: (a) original image and (b) corrected image.

Table 6 Average accuracy and response time for abnormal versus normal classification task in study 1.

	Original image	ALES output	<i>p</i> value (Wilcoxon signed-rank test)
Accuracy	71%	78%	<0.05
Response time	5.00 s	4.47 s	<0.05

Table 7 Performance for local-level decision task in study 2.

	Original image	ALES output	<i>p</i> value (Wilcoxon signed-rank test)
Accuracy for DME	70.83%	79.17%	<0.05
Sensitivity for DR	41.56%	49.78%	<0.05

higher with the ALES output than with original images. ALES is also seen to cause a significant decrease in the response time.

DME requires immediate referral to a clinic, and as per ETDRS standards, this is assessed based on the proximity of HEs to macula. We assess DME by computing accuracy = $\frac{TP+TN}{\text{Total number of images}}$. TP is defined to be the number of images where a subject's HE marking is in the same retinal zone as given by GT and TN is the number of images where the subject has not marked any HE and the corresponding GT also indicates no sign of DME. Table 7 shows that the accuracy for detection of DME improves significantly with ALES.

As per ETDRS standards, determining the DR stage based on an image requires localizing and counting all HMs. Hence, we measure the sensitivity = $\frac{TP}{TP+FN}$; TP is the number of correctly detected and FN is the number of undetected HMs. Table 7 also shows that ALES is effective for DR detection as the sensitivity has increased significantly with ALES.

6.4.1 Limitations of the study

Results of the perception study suggest that ALES aids the reader in making an "accurate" and "fast" diagnosis. However, results are preliminary as the study was conducted on a relatively small set of cases and with participants who were not professional readers. Also, a commercial monitor without any color calibration was used for the display. Performance could vary depending on color calibration of a medical grade monitor. A limitation of the study design is that it did not ask participants to report FP/FN, if any were found. Reporting was not required as participants had no domain knowledge. Assessing the clinical performance of professional readers was beyond the scope of this work.

7 Conclusions and Future Work

We conceptualized the role of CAD in reading centers/triage to be different from the prevailing paradigm. In our view, the role could be to emphasize the abnormalities and leave the background tissue unaltered. This concept can be used to design assistive solutions for readers in many ways. We took a reader-centric approach and presented a two-stage system (ALES) design focused on DR. The proposed ALES performs saliency

computation followed by lesion-emphasis, which was modeled using a spatially varying gamma correction. Starting with the bottom-up Itti-Koch model, we demonstrated that a CNN-based saliency model can be built by fine-tuning low-level filters and simultaneously learning new high-level filters. The proposed saliency model outperformed other state-of-the-art models for both bright and dark lesions for both abnormal and normal cases.

Assessment results of ALES indicate that it can successfully discriminate artifacts from true lesions and reject them. ALES is fast, as a given image can be processed in 5 s and produce a result where the background is unaltered and lesions are made prominent (up to 30% improvement in the CNR). Thus, we can conclude that ALES can be an effective and computationally efficient tool employable in reading centers. The results of our perception study support the potential of ALES. However, a rigorous evaluation needs to be done in a clinical setting.

A feature that can be incorporated in ALES is user-control over *a* and *b* in Eq. (19) to permit a reader to change the degree of lesion-emphasis for bright and dark lesions, respectively. This will help a reader to assess ambiguous regions at various levels of emphasis and, in turn, reduce FPs/FNs, if any. Such interactive ALES can work in real time; initial rendering, which requires computing saliency and lesion-emphasis (*a* and *b* set as normalizing factors), takes 5 s and rendering after varying values of *a* and *b*, takes 0.001 s as it requires only lesion-emphasis (for an image size = 288 × 360).

Disclosures

Samrudhdi B. Rangrej and Jayanthi Sivaswamy have no conflicts of interest, financial or otherwise.

Acknowledgments

This work was supported by the Department of Electronics and Information Technology, Government of India under Grant No. DeitY/R&D/TDC/13(8)/2013.

References

1. G. Lotz et al., "A domain model of a clinical reading center-design and implementation," in *2010 Annual Int. Conf. of the IEEE Engineering in Medicine and Biology*, pp. 4530–4533, IEEE (2010).
2. K. C. Oeffinger et al., "Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society," *J. Am. Med. Assoc.* **314**(15), 1599–1614 (2015).
3. J. M. G. Wilson and G. Jungner, *Principles and Practice of Screening for Disease*, World Health Organization, Geneva (1968).
4. M. Wintermark, H. A. Rowley, and M. H. Lev, "Acute stroke triage to intravenous thrombolysis and other therapies with advanced CT or MR imaging: Pro CT 1," *Radiology* **251**(3), 619–626 (2009).
5. S. V. Destounis, A. L. Arieno, and R. C. Morgan, "CAD may not be necessary for microcalcifications in the digital era. CAD may benefit radiologists for masses," *J. Clin. Imaging Sci.* **2**, 45 (2012).
6. J. Wang et al., "An image-retrieval aided diagnosis system for clustered microcalcifications," in *2016 IEEE 13th Int. Symp. on Biomedical Imaging (ISBI)*, pp. 1076–1079, IEEE (2016).
7. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 1254–1259 (1998).
8. A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognit. Psychol.* **12**(1), 97–136 (1980).
9. D. G. Albrecht and D. B. Hamilton, "Striate cortex of monkey and cat: contrast response function," *J. Neurophysiol.* **48**(1), 217–237 (1982).

10. Y. Kim and A. Varshney, "Saliency-guided enhancement for volume visualization," *IEEE Trans. Visual Comput. Graphics* **12**(5), 925–932 (2006).
11. L.-K. Wong and K.-L. Low, "Saliency retargeting: an approach to enhance image aesthetics," in *IEEE Workshop on Applications of Computer Vision (WACV 2011)*, pp. 73–80, IEEE (2011).
12. S.-J. Park, J.-K. Shin, and M. Lee, "Biologically inspired saliency map model for bottom-up visual attention," in *Int. Workshop on Biologically Motivated Computer Vision*, pp. 418–426, Springer, New York (2002).
13. X. Hou and L. Zhang, "Dynamic visual attention: searching for coding length increments," in *Advances in Neural Information Processing Systems*, pp. 681–688 (2009).
14. N. D. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: an information theoretic approach," *J. Vision* **9**(3), 5 (2009).
15. C. Yang et al., "Saliency detection via graph-based manifold ranking," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3166–3173 (2013).
16. J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems*, pp. 545–552 (2006).
17. R. Achanta et al., "Frequency-tuned salient region detection," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2009)*, pp. 1597–1604, IEEE (2009).
18. X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2007)*, pp. 1–8, IEEE (2007).
19. T. Judd et al., "Learning to predict where humans look," in *IEEE 12th Int. Conf. on Computer Vision*, pp. 2106–2113, IEEE (2009).
20. X. Huang et al., "Salicon: reducing the semantic gap in saliency prediction by adapting deep neural networks," in *Proc. of the IEEE Int. Conf. on Computer Vision*, pp. 262–270 (2015).
21. S. S. Kruthiventi, K. Ayush, and R. V. Babu, "DeepFix: a fully convolutional neural network for predicting human eye fixations," arXiv preprint arXiv:1510.02927 (2015).
22. D. Walther et al., "Attentional selection for object recognition—a gentle way," in *Int. Workshop Biologically Motivated Computer Vision*, pp. 472–479, Springer, New York (2002).
23. T. Kadir and M. Brady, "Saliency, scale and image description," *Int. J. Comput. Vision* **45**(2), 83–105 (2001).
24. H. Jiang et al., "Automatic salient object segmentation based on context and shape prior," in *Proc. of the British Machine Vision Conf.*, pp. 110.1–110.12 (2011).
25. L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Trans. Image Process.* **13**(10), 1304–1318 (2004).
26. Z. Camlica, H. Tizhoosh, and F. Khalvati, "Medical image classification via SVM using LBP features from saliency-based folded data," arXiv preprint arXiv:1509.04619 (2015).
27. A. Kumar et al., "Plane identification in fetal ultrasound images using saliency maps and convolutional neural networks," in *IEEE 13th Int. Symp. on Biomedical Imaging (ISBI 2016)*, pp. 791–794, IEEE (2016).
28. D. Mahapatra and Y. Sun, "Registration of dynamic renal MR images using neurobiological model of saliency," in *5th IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro*, pp. 1119–1122, IEEE (2008).
29. D. Mahapatra and J. M. Buhmann, "Visual saliency-based active learning for prostate magnetic resonance imaging segmentation," *J. Med. Imaging* **3**(1), 014003 (2016).
30. S. Banerjee et al., "A novel GBM saliency detection model using multi-channel MRI," *PLoS One* **11**(1), e0146388 (2016).
31. V. Jampani et al., "Assessment of computational visual attention models on medical images," in *Proc. of the Eighth Indian Conf. on Computer Vision, Graphics and Image Processing*, p. 80, ACM (2012).
32. X. Zou et al., "Learning-based visual saliency model for detecting diabetic macular edema in retinal image," *Comput. Intell. Neurosci.* **2016**, 1–10 (2016).
33. V. Navalpakkam and L. Itti, "Modeling the influence of task on attention," *Vision Res.* **45**(2), 205–231 (2005).
34. S. Frintrop, G. Backer, and E. Rome, "Goal-directed search with a top-down modulated computational attention system," in *Joint Pattern Recognition Symp.*, pp. 117–124, Springer, New York (2005).
35. M. de Brecht and J. Saiki, "A neural network implementation of a saliency map model," *Neural Networks* **19**(10), 1467–1474 (2006).
36. V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. of the 27th Int. Conf. on Machine Learning (ICML 2010)*, pp. 807–814 (2010).
37. P. Mitchell et al., *Guidelines for the Management of Diabetic Retinopathy*, National Health and Medical Research Council, Canberra, Australia (2008).
38. C. Sinthanayothin et al., "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *Br. J. Ophthalmol.* **83**(8), 902–910 (1999).
39. M. Foracchia, E. Grisan, and A. Ruggeri, "Luminosity and contrast normalization in retinal images," *Med. Image Anal.* **9**(3), 179–190 (2005).
40. E. Grisan et al., "Model-based illumination correction in retinal images," in *3rd IEEE Int. Symp. on Biomedical Imaging: Nano to Macro*, pp. 984–987, IEEE (2006).
41. G. D. Joshi and J. Sivaswamy, "Colour retinal image enhancement based on domain knowledge," in *Sixth Indian Conf. on Computer Vision, Graphics and Image Processing*, pp. 591–598, IEEE (2008).
42. Y. Wang, W. Tan, and S. C. Lee, "Illumination normalization of retinal images using sampling and interpolation," *Proc. SPIE* **4322**, 500 (2001).
43. P. Feng et al., "Enhancing retinal image by the contourlet transform," *Pattern Recognit. Lett.* **28**(4), 516–522 (2007).
44. N. M. Salem and A. K. Nandi, "Novel and adaptive contribution of the red channel in pre-processing of colour fundus images," *J. Franklin Inst.* **344**(3), 243–256 (2007).
45. M. J. Cree, E. Gamble, and D. Cornforth, "Colour normalisation to reduce inter-patient and intra-patient variability in microaneurysm detection in colour retinal images," in *ARPS Workshop on Digital Image Computing (WDIC 2005)*, Brisbane, Australia, pp. 163–168 (2005).
46. A. Hagiwara, A. Sugimoto, and K. Kawamoto, "Saliency-based image editing for guiding visual attention," in *Proc. of the 1st Int. Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*, pp. 43–48, ACM (2011).
47. S. L. Su, F. Durand, and M. Agrawala, "De-emphasis of distracting image regions using texture power maps," in *Proc. of the 2nd Symp. on Applied Perception in Graphics and Visualization* (2005).
48. W.-M. Ke, C.-R. Chen, and C.-T. Chiu, "BiTA/SWCE: image enhancement with bilateral tone adjustment and saliency weighted contrast enhancement," *IEEE Trans. Circuits Syst. Video Technol.* **21**(3), 360–364 (2011).
49. T. Kauppi et al., "The DIARETDB1 diabetic retinopathy database and evaluation protocol," in *Proc. of the British Machine Vision Conf. (BMVC 2007)*, pp. 1–10 (2007).
50. P. Prentas et al., "Diabetic retinopathy image database (DRiDB): a new database for diabetic retinopathy screening programs research," in *8th Int. Symp. on Image and Signal Processing and Analysis (ISPA 2013)*, pp. 711–716, IEEE (2013).
51. L. Giancardo et al., "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Med. Image Anal.* **16**(1), 216–226 (2012).
52. G. D. Joshi and J. Sivaswamy, "Colour retinal image enhancement based on domain knowledge," in *Sixth Indian Conf. on Computer Vision, Graphics and Image Processing*, pp. 591–598, IEEE (2008).
53. L. Tang et al., "Splat feature classification with application to retinal hemorrhage detection in fundus images," *IEEE Trans. Med. Imaging* **32**(2), 364–375 (2013).
54. G. Azzopardi et al., "Trainable COSFIRE filters for vessel delineation with application to retinal images," *Med. Image Anal.* **19**(1), 46–57 (2015).
55. A. Torralba et al., "Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search," *Psychol. Rev.* **113**(4), 766–786 (2006).
56. M. Mancas, "Relative influence of bottom-up and top-down attention," in *5th Int. Workshop on Attention in Cognitive Systems (WAPCV 2008)*, pp. 212–226, Springer, New York (2008).
57. T. Kauppi et al., "DIARETDB0: evaluation database and methodology for diabetic retinopathy algorithms," in *Machine Vision and Pattern Recognition Research Group*, Lappeenranta University of Technology, Finland (2006).
58. Z. Bylinskii et al., "What do different evaluation metrics tell us about saliency models?" arXiv preprint arXiv:1604.03605 (2016).

Samrudhdhi B. Rangrej received her B.Tech in electronics and communication engineering from Nirma University, Ahmedabad, India. Currently, she is pursuing master of science by research at International Institute of Information Technology, Hyderabad, India. Her research interests include CAD algorithm development for retinal image analysis.

Jayanthi Sivaswamy received her PhD in electrical engineering from Syracuse University, New York, USA. Since 2001, she has been with IIIT, Hyderabad, India. Prior to that, she was at the University of Auckland, New Zealand. Her research interests include CAD algorithm development, neuroimage analysis, and image reconstruction.