ALES: an assistive system for fundus image readers

by

rangrej.bharat , Jayanthi Sivaswamy

in

Journal of Medical Imaging

Report No: IIIT/TR/2017/-1

Centre for Visual Information Technology
International Institute of Information Technology
Hyderabad - 500 032, INDIA
April 2017

ALES: an assistive system for fundus image readers

Samrudhdhi B. Rangrej[a,*], **Jayanthi Sivaswamy**[a]
[a]International Institute of Information Technology, Hyderabad, India, 500032.

Abstract. Computer assisted diagnosis (CAD) tools are of interest as they enable efficient decision making in clinics and screening of diseases. Traditional approach to CAD algorithm design focuses on automated detection of abnormalities independent of the end-user who can be an image reader or an expert. We propose a novel, reader-centric system design wherein a readers 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.[1] The CNN is used to fine-tune standard low-level filters and learn new 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.

Keywords: Computer Assisted Diagnostic (CAD), saliency, selective enhancement, color fundus image, Convolutional Neural Network, gamma correction.

*Samrudhdhi B. Rangrej, rangrej.bharat@research.iiit.ac.in

1 Introduction

Recent advances in medical imaging has enabled the involvement of a variety of practitioners in the diagnosis process. While in the past medical experts who examined patients also analyzed images, the shortage of highly trained experts has led to 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. Centres offering such service are called Reading centers[2] 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 some disease since early detection is preferred in effective disease management. Breast cancer screening for women in the age group of 45-54 years is one such example.[3] 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.[4]

Triage on the other hand is a practice followed by clinics to prioritize patients for experts' attention. A trained practitioner orders preliminary tests, a reader (semi-expert) 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 speed up the diagnostic process. Acute stroke triage is an example of this.[5]

Image reading is a tedious task yet requires precision as it is critical to diagnosis. Fatigue or inattention causes readers to miss inconspicuous/subtle lesions leading to under-reporting 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.[6] More recently, retrieving and displaying similar past cases along with the diagnosis has been shown to be effective.[7] 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 like color and orientation.[1,8] 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 a novel Assistive Lesion Emphasizing System or ALES, which employs saliency of a region to determine the amount of its emphasis. The ALES concept is demonstrated for retinal abnormalities called Diabetic 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 and blood leakage called hemorrhage. 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 (ii) lesion-emphasis.

Technical contributions towards the development of ALES are as follows: (a) Novel approach to image reading. An assistive solution is proposed based on saliency modeling. (b) A Convolutional Neural Network (CNN) based saliency model with a novel loss function. Proposed architecture is inspired by the Itti-Koch model[1] 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 are presented followed by their assessment.

2 Saliency Computation

2.1 Background

Computational modeling of visual saliency has been a subject of research for long. Existing computational models range from biologically plausible ones[1,12] 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] image/video compression.[25] Such models however are

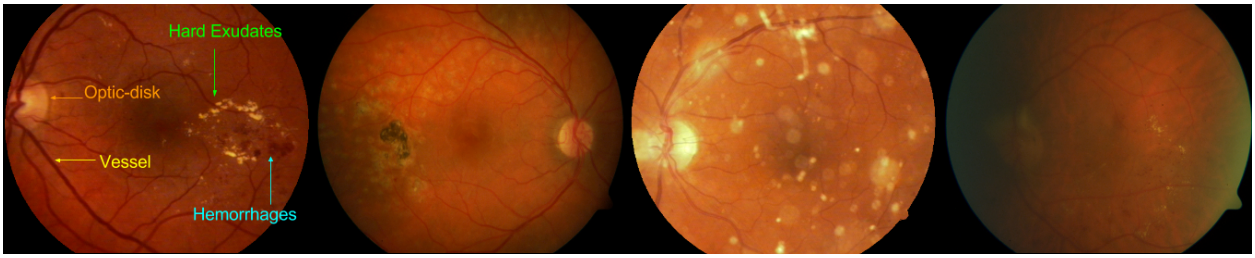

Fig 1 From left to right: Retinal image with DR lesions, dark artifacts, bright artifacts and varying illumination.

developed to compute saliency for natural images where: (a) there are only few objects of interest (b) the target objects are mostly in the center and (c) the background is free of clutter/texture. 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 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 3D 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 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 non-uniform illumination (see Fig 1). A good saliency model has to learn discriminate artifacts from true lesions and reject the former. Computational saliency models have been reported for only hard exudate.[31,32] Our aim is to develop saliency models for both hard exudate and hemorrhage. This is done using a Convolutional Neural Network(CNN) inspired by the Itti-Koch saliency model.[1] In this model, center-surround 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 center-surround filters (ii) learn new filters and (iii) learn the weights for combining the feature maps.

CNN is a type of feed forward 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 which performs filtering of activations with weights; maxpooling which downsamples the image by retaining the maxima in a local neighborhood and an activation function which applies a non-linear transformation on the intensity values of an image. We use the Rectified Linear Unit (ReLU) as an activation function.[36]

2.2.1 CNN Architecture

The architecture of the proposed model has five stages as shown in Fig 2. It is carefully designed to share similarities with 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 center surround difference pyramid and Stage 5 is identical to final normalization and combination of all the maps. Parallel fine-tunning of Itti-Koch filters and learning of new filters is carried out at

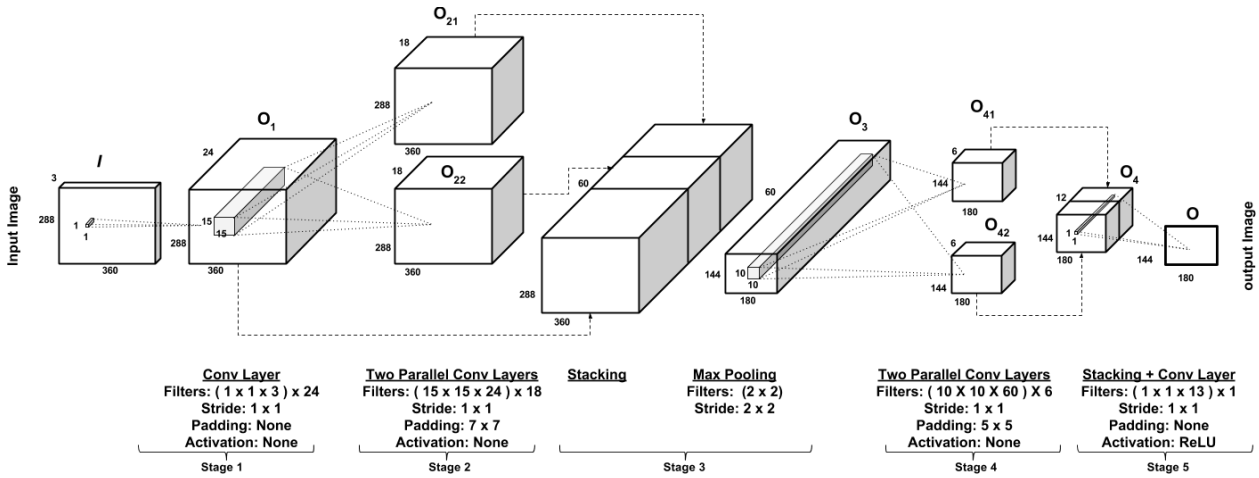

Fig 2 Proposed architecture with stage-wise description of types of layers, filter size, padding size, stride and activation function.

each stage. Model architecture is described below.

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.

$$O_1 = W_{color} * I + b_{color} \tag{1}$$

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 2-D orientation filters of resolution 20° were generated. Each filter was repeated and stacked to generate W_{orient}.

O_{21} and O_{22} are computed as follows.

$$O_{21} = W_{orient} * O_1 + b_{orient} \tag{2}$$

$$O_{22} = W_{rnd1} * O_1 + b_{rnd1} \tag{3}$$

The maps O_1, O_{22} and O_{21} are stacked in stage 3 and maxpooling 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,

$$O_3 = maxpooling([O_1, O_{21}, O_{22}]) \tag{4}$$

Stage 4 also has two parallel convolutional layers which operate on O_3 to produce O_{41} and O_{42} as two independent outputs. Both layers have 6 filters of size $(10 \times 10 \times 60)$. The first convolutional layer is initialized with center-surround (CS) filters and second with random filters. Weight initialization for W_{CS} is done using six 2-D center-surround filters which are generated as follows:

$$CS1 = \pm(G_1 - G_4) \tag{5}$$

$$CS2 = \pm(G_2 - G_5) \tag{6}$$

$$CS3 = \pm(G_3 - G_6) \tag{7}$$

$$CS4 = \pm(G_1 - G_5) \tag{8}$$

$$CS5 = \pm(G_2 - G_6) \tag{9}$$

$$CS6 = \pm(G_3 - G_7) \tag{10}$$

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

$$O_{41} = W_{CS} * O_3 + b_{CS} \tag{11}$$

$$O_{42} = W_{rnd2} * O_3 + b_{rnd2} \tag{12}$$

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, a ReLU activation is applied to get the desired output, which is a final gray scale image O of size (144×180).

$$O_4 = [O_{41}, O_{42}] \tag{13}$$

$$O_5 = W_{combination} * O_4 + b_{combination} \tag{14}$$

$$O = \max(0, O_5) \tag{15}$$

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 which aims to minimize a loss function which compares the system output with ground truth. 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 ground truth (GT) saliency values are in the range $[0, 1]$. Hence, we define a new loss function as follows.

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

Here, the tuple (x, y) denotes the (GT saliency, output) pixel pair; N is the total number of pixels; β is a weight used to handle class imbalance. α controls the threshold y_0 such that, a low loss is achieved for 2 conditions: (i) low GT saliency value ($x \in [0, 0.5]$) and a sub-threshold output (ii) a high GT saliency ($x \in [0.5, 1]$) and a supra-threshold 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).

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

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 hard exudate and hemorrhage saliency separately. We denote hard exudate (HE) and hemorrhage (HM) specific saliency models as S_{HE} and S_{HM} respectively. Data and computational resources used for training of CNN is discussed in the Material section. Computed saliency maps are used to emphasize lesions locally as described in the following section.

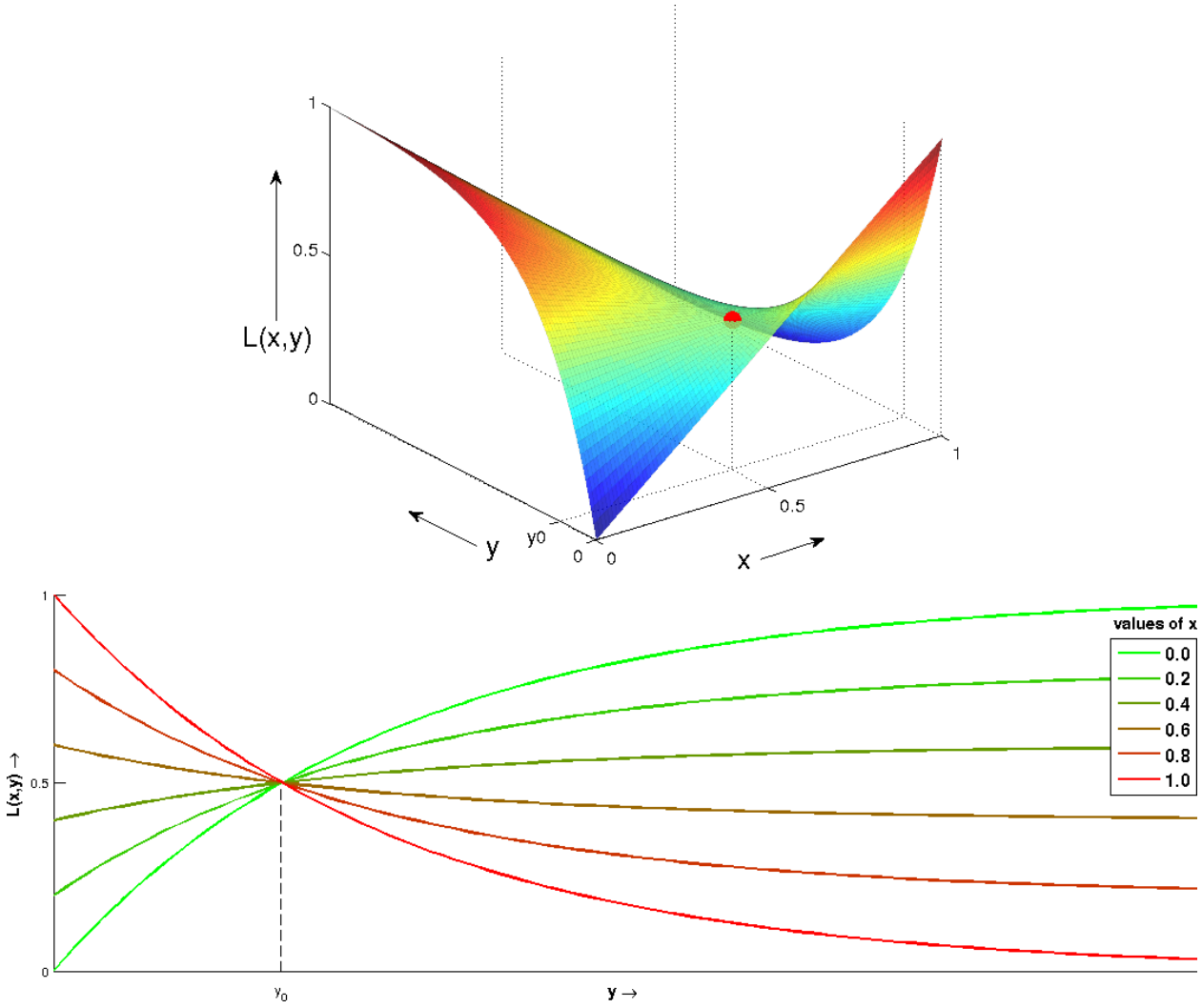

Fig 3 Loss function in the absence of class-imbalance ($\beta = 1$). (top) 3D view of a loss function. Saddle point is shown in red color. (bottom) 2D view of a loss function. Above surface is sampled at different values of x. y-projection of saddle point $(0.5, y_0)$ is also shown.

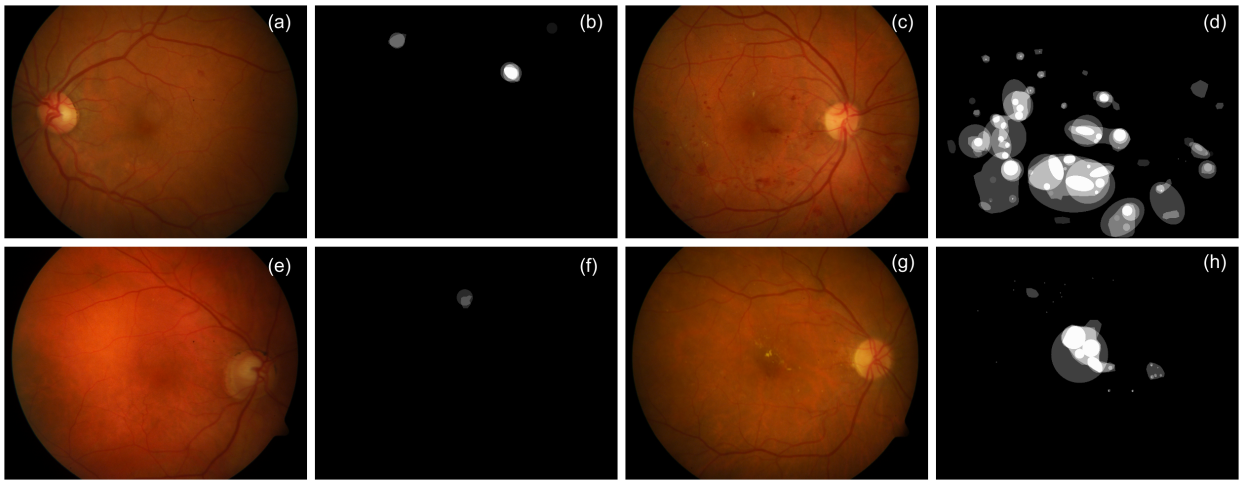

Fig 4 Original image and GT: (a,b) DR stage 2 (c,d) DR stage 3 (e,f) DME stage 2 (g,h) DME stage 3.

3 Lesion-Emphasis

3.1 Background

In DR reading centres, readers scrutinize images and assign a DR stage to the image using the ETDRS standard.[37] Staging is based on the following guidelines: (1) DR grade (on a scale 1-4) is proportional to the number of dark lesions (b) diabetic macular edema (DME) grade (scale 1-3) is proportional to the distance between macula and nearest hard exudate(see Fig 4). The grade determines the type of advice given to a subject being screened with some requiring immediate referral. Thus, a failure 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. Contrast-enhanced lesion will successfully draw a reader's attention and hopefully reduce the rate of misdiagnoses.

Existing work on enhancement of retinal images are based on illumination/contrast correction,[38–42] contourlet[43] and histogram equalization and matching.[44,45] These methods are primarily

aimed as a pre-processing stage for CAD development and none have been aimed at readers or experts. Hence, they often introduce textures and colour 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, non-linear, contrast stretching method. Gamma correction is defined as,

$$I_C(x, y) = I_O(x, y)^{\gamma} \tag{18}$$

Here $I_C(x, y)$ and $I_O(x, y)$ (Normalized between 0 and 1) are corresponding pixels from corrected and original images respectively. 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

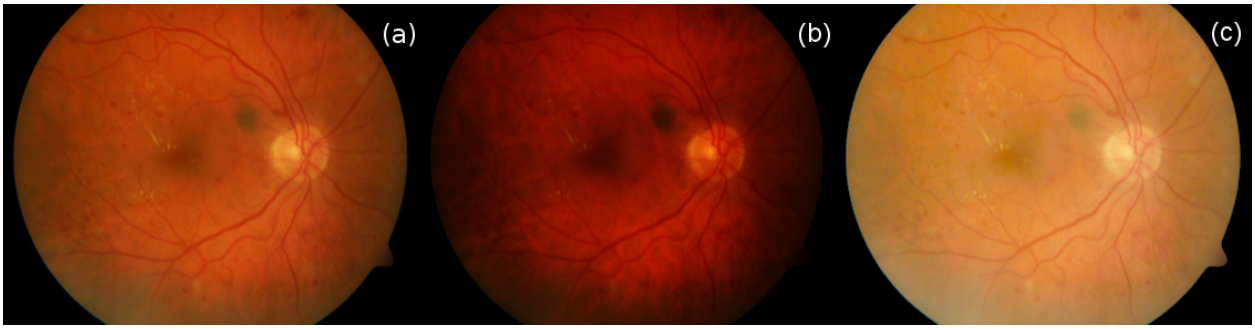

Fig 5 Gamma correction. (a) original image (b) corrected image with $\gamma = 2$ (d) corrected image with $\gamma = 0.5$.

values to lower range which results in darkening of the entire image, while $\gamma < 1$ pushes intensity values to higher range and thus brightening of 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.

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

where, a and b are 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 hemorrhage 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 hard exudate 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 DMED[51] datasets were used for training and testing the proposed saliency computation stage. Pre-processing 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. Fig 8(b) shows the result of preprocessing on a sample image Fig 8(a). All images were downsampled to size 288 x 360 and normalized to have zero mean value and unit variance. The size was chosen to minimize both distortion of the image (aspect ratio) and the computational cost.

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 4 experts as a heatmap (leftmost image in Fig.6) and the latter providing marking from one expert as a binary map. In order 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). In order to retain hard exudates (hemorrhages) 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 hemorrhages 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. 122 (of 134) images with hard exudates and 72 (of 85) images with hemorrhages 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

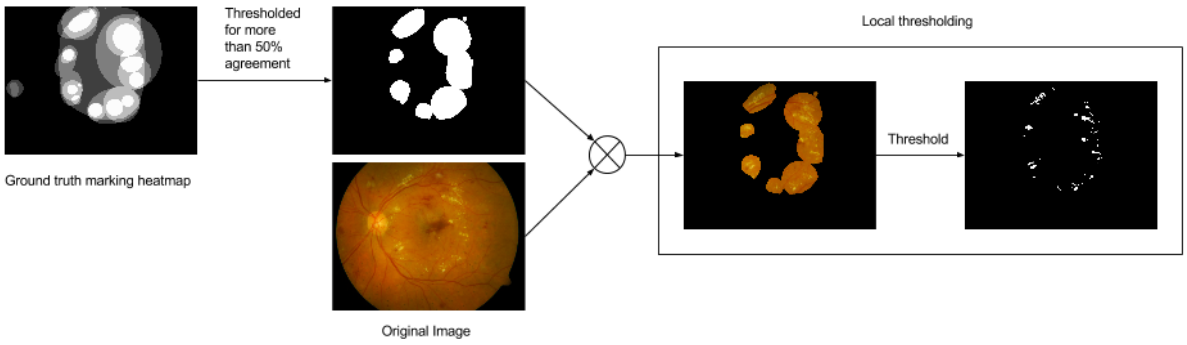

Fig 6 Procedure to obtain lesion-level ground truth from regional marking.

a variety of transformations. Random rotation between 0 to $30°$, random vertical shift between 0 to 57 pixels (20% of height of an image), random horizontal shift between 0 to 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 4GB of RAM for 10000 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 approximately 5 days. Cross-validation was not performed due to excessive training time.

Table 1 Dataset description.

Datasets	DIARETDB1	DRiDB	DMED
Total Number of Images	89	50	169
Images containing Hard Exudate	48	32	54
Images containing Hemorrhages	54	31	-

5 Evalution of ALES

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

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

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 while center-surround filters changed considerably during training (see Fig 7(a)(b)). Fig 7(b) shows progression in tuning of 3 channels of the center-surround filters of S_{HE}. It can be seen that the pattern of tuned filters are similar to difference of multiple ($>$ 2) Gaussian filters. Fig 7(c-d) shows how 3 filters (single channel only) with random initialization changes during training in stage 2 and 4. The sample filters from stage 2 can be expected to give 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,[1] SR,[18] AIM,[14] GBVS,[16] Torralba,[55] Judd[19] and *Rare*.[56] Among these Itti-Koch

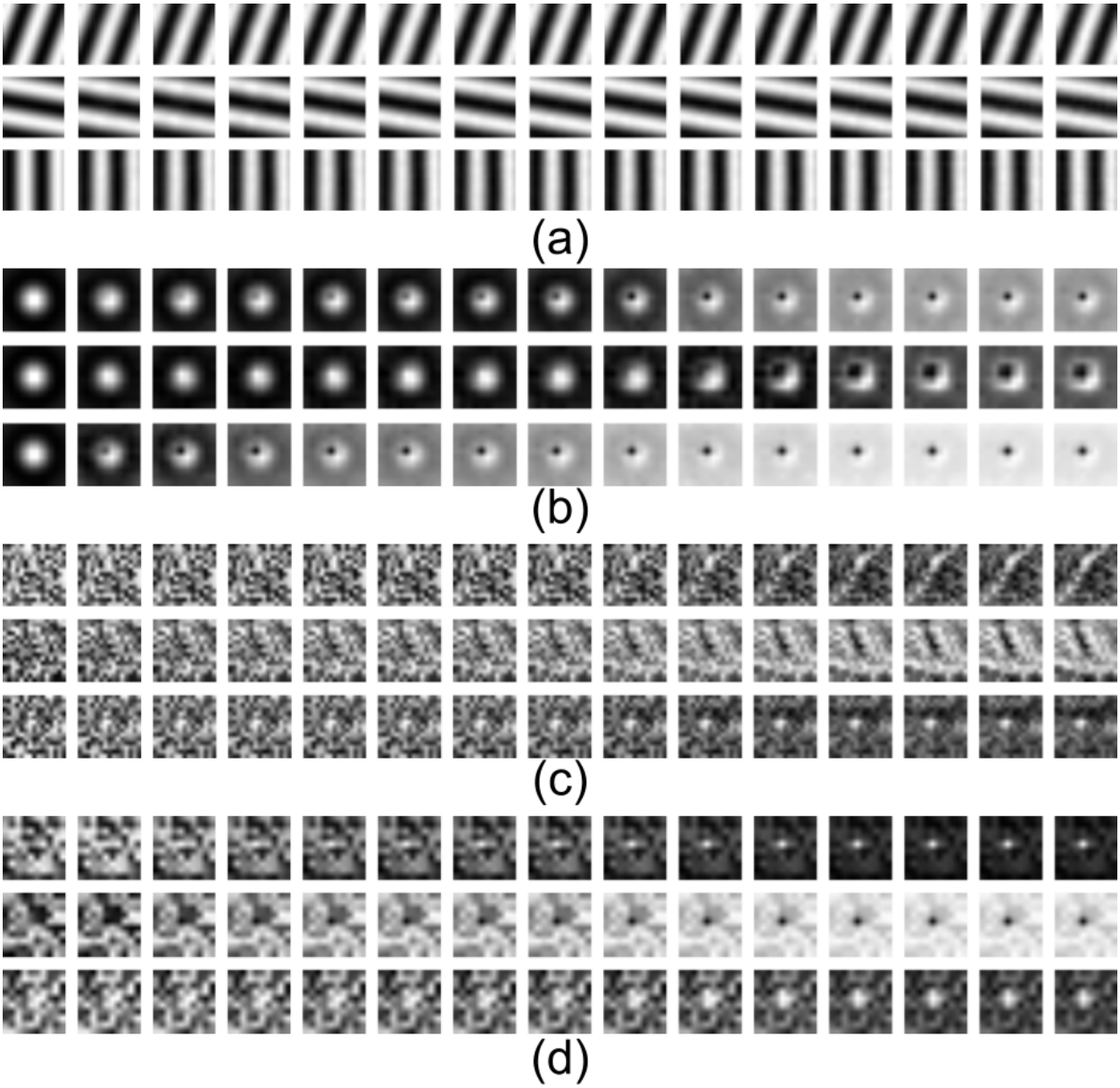

Fig 7 Evolution of the filters during training of hard exudates saliency. (a) 3 channels of W_{orient} from stage 2 (b) 3 channels of $CS3$ filters from stage 4 (c) 3 channels of random filters from stage 2 (d) 3 channels of random filters from stage 4.

is biologically plausible, SR is spectral analysis based, AIM and Torralba are information-theoretic, GBVS is graph based, Judd is pattern classification and *Rare* is based on top-down bias. Saliency maps for 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 hard exudates and Fig 9 for hemorrhages. 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 hemorrhages 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 retina is also important for normal vs 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 hyper-reflective region.

Quantitative evaluation was performed on both abnormal and normal images (see Table 3). Normal images were taken from DRiDB[50] and DIARETDB0.[57] The metrics used for the evaluation are: receiver operating characteristic curves (ROC), area under the ROC curve (AUC), false positive rate(FPR) vs saliency and precision vs 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 nega-

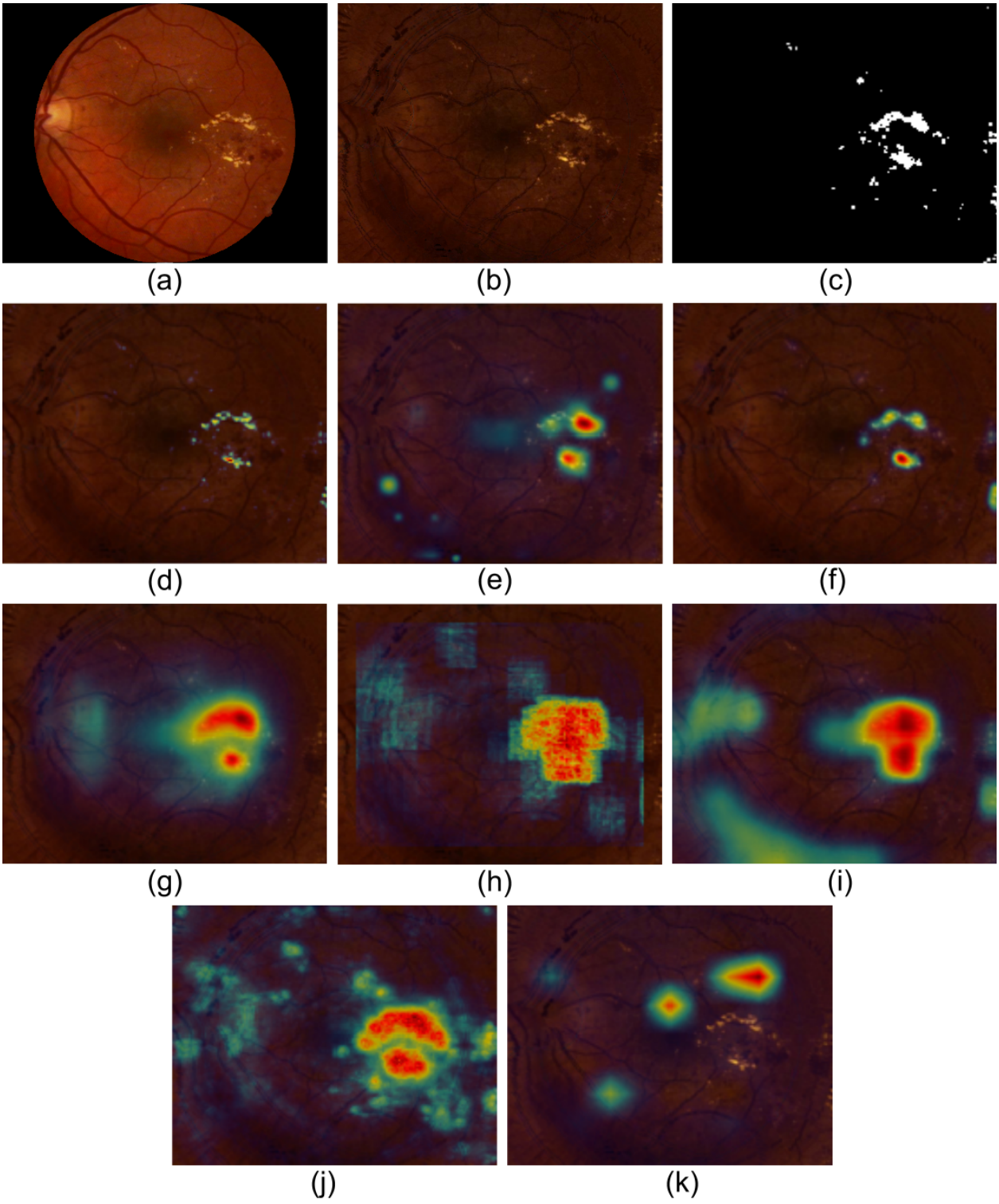

Fig 8 Hard exudate saliency. (a) original color fundus image (b) pre-processed image (c) ground truth. Computed saliency maps of (d) Proposed (e) Itti-Koch (f) SR (g) GBVS (h) AIM (i) Rare (j) Torralba (k) Judd.

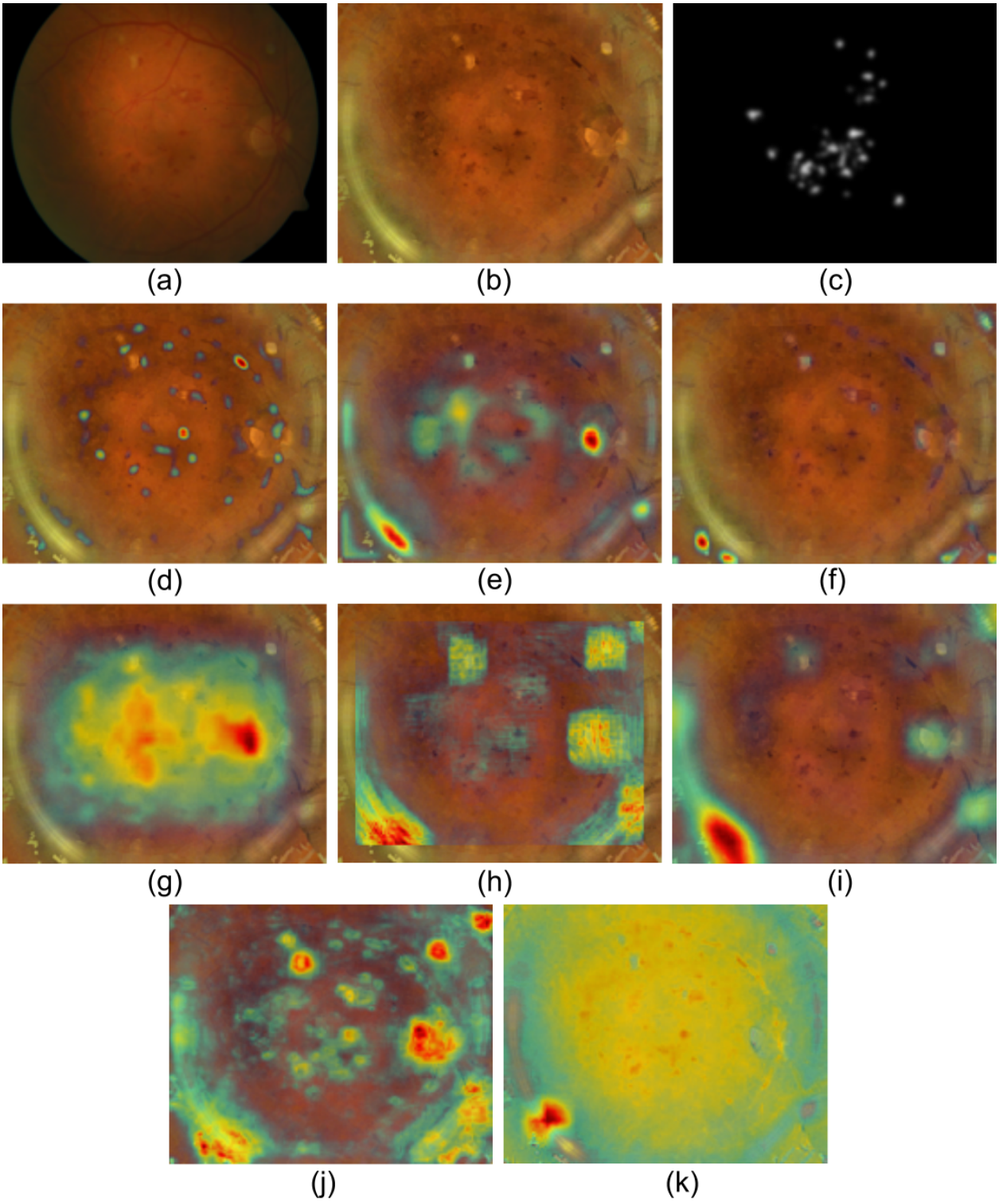

Fig 9 Hemorrhage saliency. (a) original color fundus image (b) pre-processed image (c) Gaussian convolved ground truth. Computed saliency maps of (d) Proposed (e) Itti-Koch (f) SR (g) GBVS (h) AIM (i) Rare (j) Torralba (k) Judd.

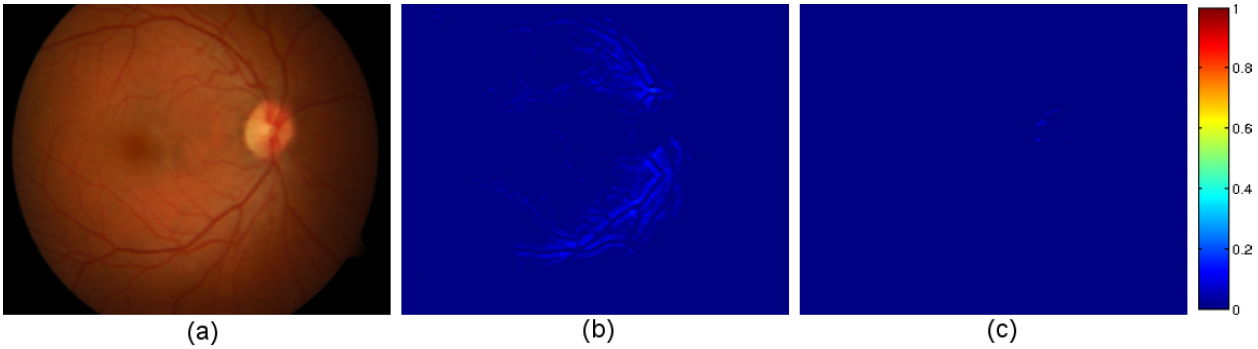

Fig 10 Predicted saliency for normal cases. (a) Normal color fundus image and saliency maps for (b) hard exudate (c) hemorrhage.

tive(FN). Background pixels, if correctly/incorrectly detected with reference to GT, are considered true negative(TN)/false positive(FP).

Table 3 Number of images in the test set.

Type of abnormality	Abnormal images	Normal images
Hard exudate	12	33
Hemorrhage	13	33

ROC is a graphical representation of achieved True Positive Rate ($TPR = \frac{TP}{TP+FN}$) vs False Positive Rate ($FPR = \frac{FP}{FP+TN}$) for varying threshold values. The ROC and AUC are presented in Fig 11 and Table 4. AUC values are reported for whole test set as well as a balanced test set (with equal number of normal and abnormal images). The results show that proposed model outperforms all other models. It can be seen that the Judd saliency model has nearly same AUC value as S_{HM}. This can be explained as follows. In normal cases (absence of any hemorrhage), the model successfully rejects background pixels as non-salient regions. Since the test set is skewed towards normal images, the overall performance is good and almost at par with S_{HM}. However, this model's use of multi-scale 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

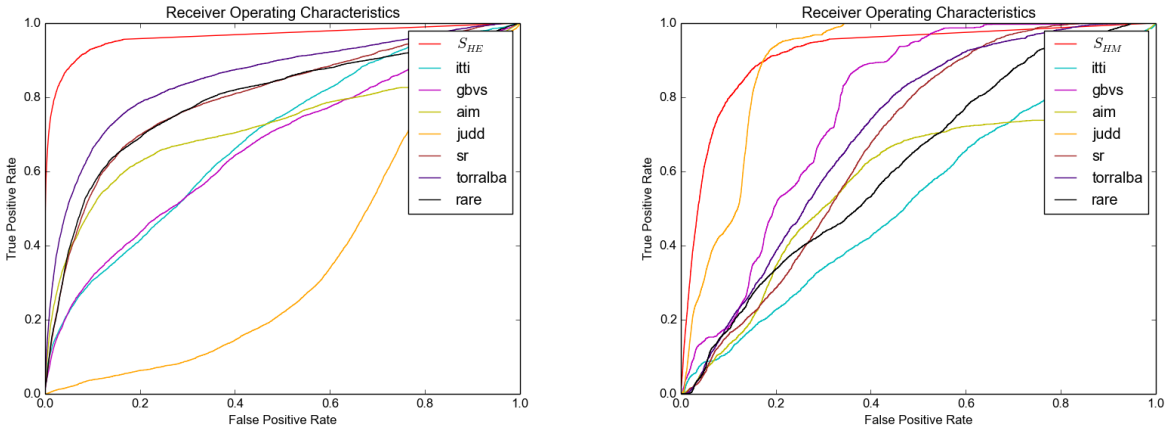

Fig 11 Receiver Operating Characteristics(ROC). (a) hard exudate saliency (b) hemorrhage saliency.

is compromised for abnormal images. This can be verified from analysis of False Positive Rate discussed next. Hence, the AUC value of Judd model on balanced test set is less than that of whole test set.

Table 4 Comparison of AUC scores.

	Hard Exudate		Hemorrhage	
	whole test set	balanced test set	whole test set	balanced test set
S_{HE}/S_{HM}	**0.962**	**0.959**	**0.918**	**0.923**
SR	0.799	0.829	0.678	0.622
Torralba	0.853	0.867	0.708	0.683
rare	0.794	0.791	0.621	0.591
Itti-Koch	0.688	0.675	0.526	0.523
AIM	0.723	0.728	0.589	0.575
GBVS	0.667	0.664	0.774	0.677
Judd	0.367	0.383	0.896	0.827

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

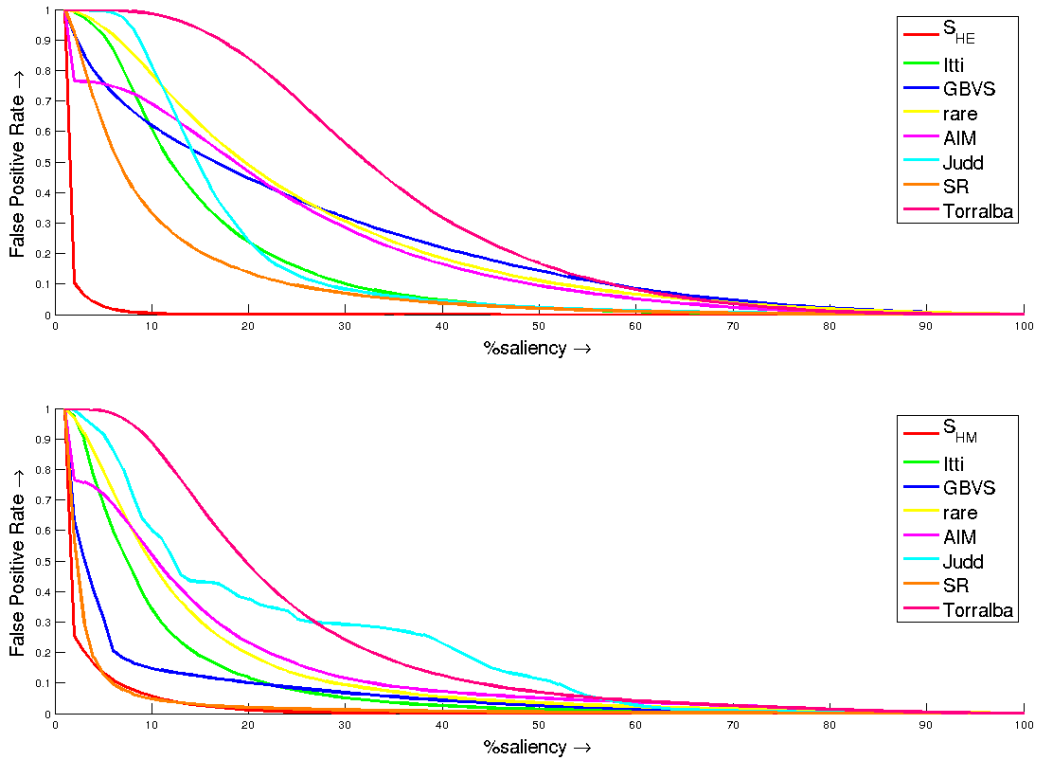

Fig 12 Flase positive rate vs saliency. (top) hard exudates (bottom) hemorrhages.

The FP Rate ($FPR = \frac{FP}{FP+TN}$) for low saliency values may be a better metric for this purpose. FPR was computed(on whole test set) 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(non-lesion) 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 $<<$ 0.01). In the case of hemorrhages, Torralba and Judd model have higher FPR(in the low saliency range) than other models which indicates its inability to handle non-lesion regions.

FPR demonstrates how well a saliency model can reject background pixels. Precision ($\frac{TP}{TP+FP}$) on the other hand 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

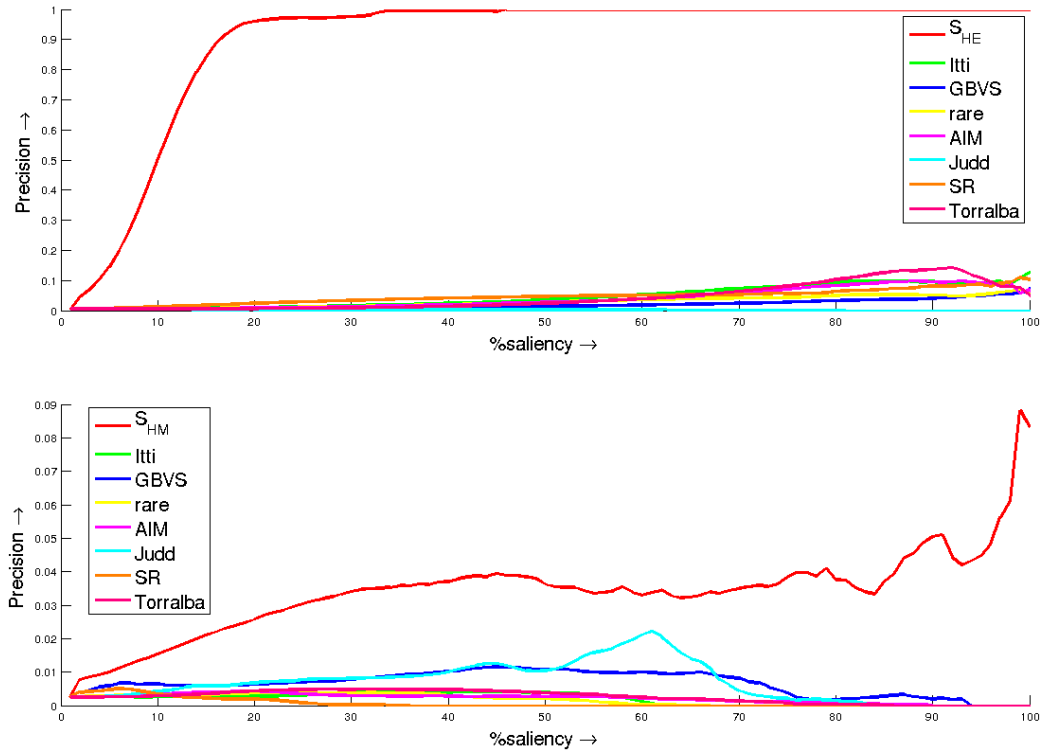

Fig 13 Positive Predictive Rate/Precision vs saliency. (top) hard exudates (bottom) hemorrhages.

correctly detected lesion-pixels among all the detected pixels (and hence precision) can be expected to increase. The precision vs saliency plots(computed on whole test set) for the two types of lesions are shown in Fig. 13. The wide difference in precision levels for S_{HE} and S_{HM} attest to the fact that detection of hemorrhages is more challenging; the low precision caused by confusion between vessel fragments (persisting after inpainting) and hemorrhages. Further, it can also be seen that both S_{HE} and S_{HM} outperform other saliency models (Wilcoxon signed-rank test: p-value $<<$ 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 (hard exudate) and blue (hemorrhages). It can be observed that by and large, the background and non-lesion regions, including blood vessels, are non-salient. Some high saliency

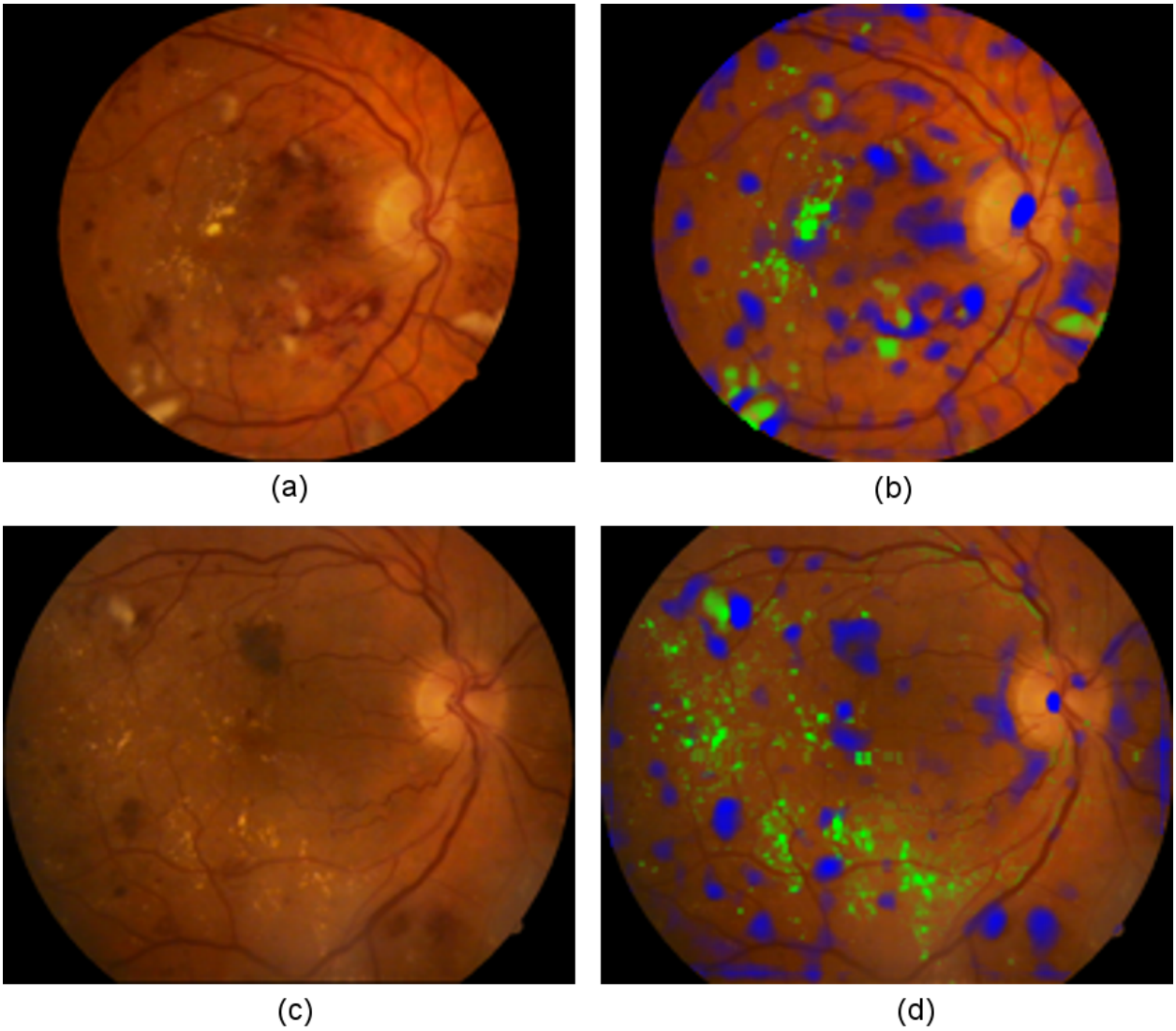

Fig 14 Combined saliency for hard exudate and hemorrhage. (a,c) Original images with lesions (b,d) combined saliency maps for hard exudate (green) and hemorrhage (purple) shown overlaid on the original image.

is seen in the middle of the optic disc 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 more visible in the processed results which makes counting of dark lesion easier. Hard exudates which are close to macula as per GT are not clearly visible in the original images but are prominent after

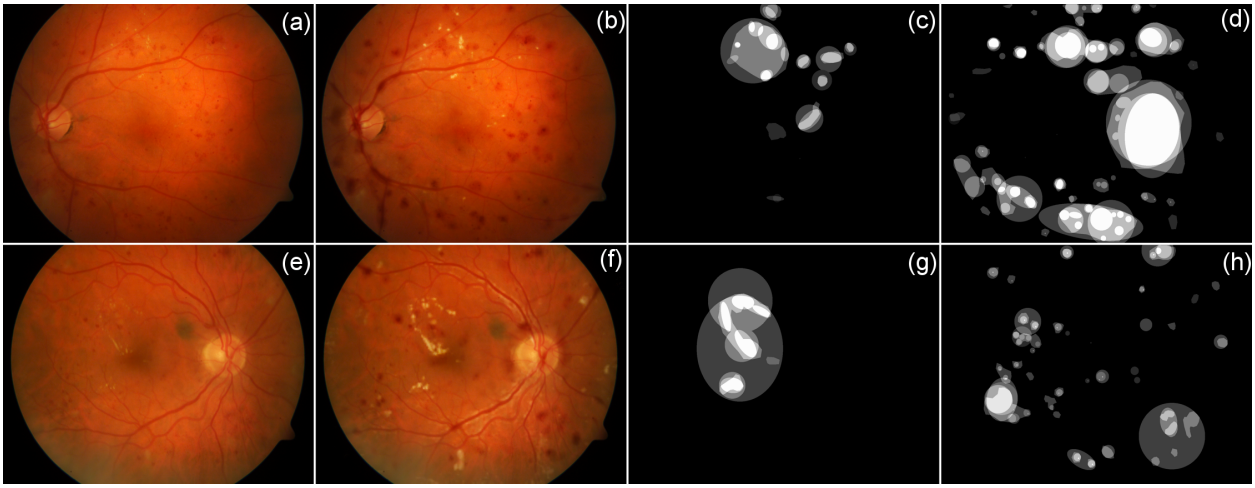

Fig 15 ALES output for abnormal images. (a)(e) original images (b)(f) corrected images (c)(g) hard exudate GT (d)(h) hemorrhage GT.

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.

$$CNR = \frac{|m_f - m_b|}{\sigma_b} \tag{20}$$

where, m_f and m_b are mean intensity of foreground (lesions) and background respectively. σ_b is the standard deviation of background intensity. CNR was computed for images from two datasets containing both hemorrhage and hard exudate. 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 the Fig 16. ALES can be seen to introduce minimal artifacts, as desirable.

Table 5 Average contrast-to-noise ratio.

	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

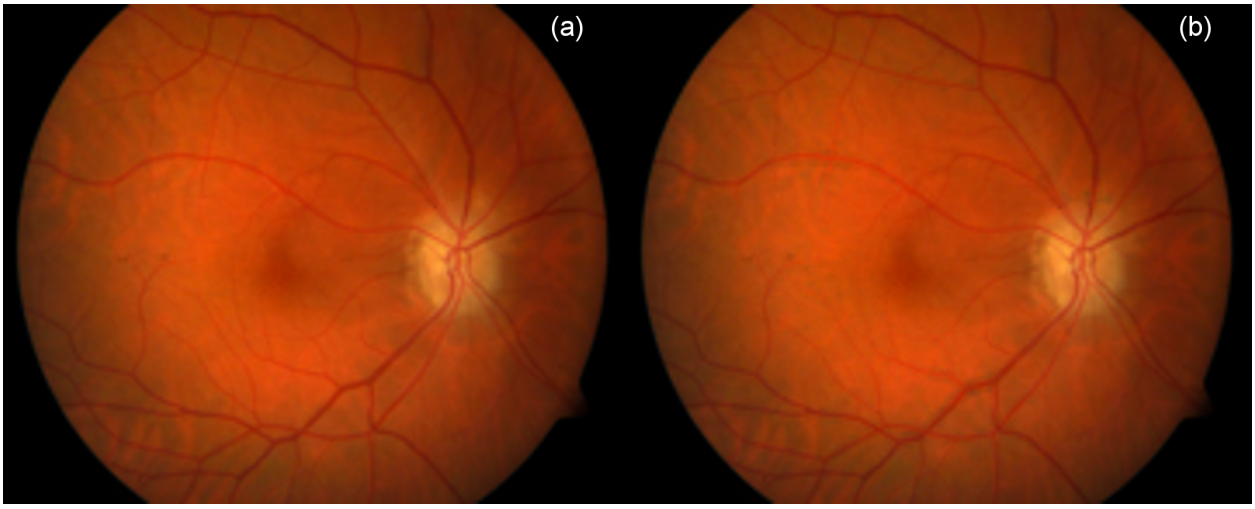

Fig 16 ALES output for normal image. (a) original image (b) corrected image.

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 global-level decision about an image which is classification of an image as a normal vs abnormal case. The second study was aimed at measuring the effectiveness of ALES for local-level decision about an image such as locating all hemorrhages and a hard exudate 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 the experiments were selected by a local expert. 30 pairs of (original and ALES output) images, with equal number of normal and abnormal cases, were used in the first study. 10 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

12 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 Lenovo monitor of size 1366×768 pixels and the interval between the response for an image and the display of next image was 2s.

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 hemorrhages; (ii) click on the hard exudate which is closest to macula. Hemorrhages are of interest in detecting DR. Macula is the site of high acuity colour vision and therefore presence of hard exudates 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 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 hard exudates 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 hard exudate 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 hard exudate 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 hemorrhages. Hence, we measure the Sensitivity = $\frac{TP}{TP+FN}$; TP is the number of correctly detected and FN is the number of undetected hemorrhages. Table 7 also shows that ALES is effective for DR detection as the sensitivity has increased significantly with ALES.

Table 6 Average accuracy and response time for abnormal vs normal classification task in Study 1.

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

7 Conclusions and Future Work

We conceptualized the role of CAD in reading centres/triage to be different from the prevailing paradigm. In our view, the role could be to emphasize the abnormalities and leave the background

Table 7 Performance for local level decision task in Study 2.

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

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 2-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 seconds and produce a result where the background is unaltered and lesions are made prominent (up to 30% improvement in the CNR). Thus, we conclude that ALES can be an effective and computationally efficient tool employable in reading centers. The results of our perception studies attested to the effectiveness of ALES. However, a more rigorous evaluation needs to be done in a clinical setting.

Acknowledgments

This work was supported by the Dept. of Electronics and Information Technology, Govt. of India under Grant: $DeitY/R\&D/TDC/13(8)/2013$.

References

1 L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence* (11), 1254–1259 (1998).

2 G. Lotz, T. Peters, E. Zrenner, *et al.*, "A domain model of a clinical reading center-design and implementation," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 4530–4533, IEEE (2010).

3 K. C. Oeffinger, E. T. Fontham, R. Etzioni, *et al.*, "Breast cancer screening for women at average risk: 2015 guideline update from the american cancer society," *Jama* **314**(15), 1599–1614 (2015).

4 J. M. G. Wilson, G. Jungner, *et al.*, "Principles and practice of screening for disease.," *World Health Organization. Public Health Paper* (34) (1968).

5 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).

6 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," *Journal of clinical imaging science* **2** (2012).

7 J. Wang, Y. Yang, M. N. Wernick, *et al.*, "An image-retrieval aided diagnosis system for clustered microcalcifications," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 1076–1079, IEEE (2016).

8 A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive psychology* **12**(1), 97–136 (1980).

9 D. G. Albrecht and D. B. Hamilton, "Striate cortex of monkey and cat: contrast response function.," *Journal of neurophysiology* **48**(1), 217–237 (1982).

10 Y. Kim and A. Varshney, "Saliency-guided enhancement for volume visualization," *IEEE Transactions on Visualization and Computer Graphics* **12**(5), 925–932 (2006).

11 L.-K. Wong and K.-L. Low, "Saliency retargeting: An approach to enhance image aesthetics," in *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*, 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 *International Workshop on Biologically Motivated Computer Vision*, 418–426, Springer (2002).

13 X. Hou and L. Zhang, "Dynamic visual attention: Searching for coding length increments," in *Advances in neural information processing systems*, 681–688 (2009).

14 N. D. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of vision* **9**(3), 5–5 (2009).

15 C. Yang, L. Zhang, H. Lu, *et al.*, "Saliency detection via graph-based manifold ranking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3166–3173 (2013).

16 J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 545–552 (2006).

17 R. Achanta, S. Hemami, F. Estrada, *et al.*, "Frequency-tuned salient region detection," in

Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on, 1597–1604, IEEE (2009).

18 X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, 1–8, IEEE (2007).

19 T. Judd, K. Ehinger, F. Durand, *et al.*, "Learning to predict where humans look," in *Computer Vision, 2009 IEEE 12th international conference on*, 2106–2113, IEEE (2009).

20 X. Huang, C. Shen, X. Boix, *et al.*, "Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 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, L. Itti, M. Riesenhuber, *et al.*, "Attentional selection for object recognitiona gentle way," in *Biologically Motivated Computer Vision*, 472–479, Springer (2002).

23 T. Kadir and M. Brady, "Saliency, scale and image description," *International Journal of Computer Vision* **45**(2), 83–105 (2001).

24 H. Jiang, J. Wang, Z. Yuan, *et al.*, "Automatic salient object segmentation based on context and shape prior.," in *BMVC*, **6**(7), 9 (2011).

25 L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *Image Processing, IEEE Transactions on* **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, P. Sridar, A. Quinton, *et al.*, "Plane identification in fetal ultrasound images using saliency maps and convolutional neural networks," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 791–794, IEEE (2016).

28 D. Mahapatra and Y. Sun, "Registration of dynamic renal mr images using neurobiological model of saliency," in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, 1119–1122, IEEE (2008).

29 D. Mahapatra and J. M. Buhmann, "Visual saliency-based active learning for prostate magnetic resonance imaging segmentation," *Journal of Medical Imaging* **3**(1), 014003–014003 (2016).

30 S. Banerjee, S. Mitra, B. U. Shankar, *et al.*, "A novel gbm saliency detection model using multi-channel mri," *PloS one* **11**(1) (2016).

31 V. Jampani, J. Sivaswamy, V. Vaidya, *et al.*, "Assessment of computational visual attention models on medical images," in *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing*, 80, ACM (2012).

32 X. Zou, X. Zhao, Y. Yang, *et al.*, "Learning-based visual saliency model for detecting diabetic macular edema in retinal image," *Computational Intelligence and Neuroscience* **2016** (2016).

33 V. Navalpakkam and L. Itti, "Modeling the influence of task on attention," *Vision research* **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 *Pattern Recognition*, 117–124, Springer (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 *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 807–814 (2010).

37 P. Mitchell, S. Foran, E. Assistance, *et al.*, "Guidelines for the management of diabetic retinopathy," *National Health and Medical Research Council* (2008).

38 C. Sinthanayothin, J. F. Boyce, H. L. Cook, *et al.*, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *British Journal of Ophthalmology* **83**(8), 902–910 (1999).

39 M. Foracchia, E. Grisan, and A. Ruggeri, "Luminosity and contrast normalization in retinal images," *Medical Image Analysis* **9**(3), 179–190 (2005).

40 E. Grisan, A. Giani, E. Ceseracciu, *et al.*, "Model-based illumination correction in retinal images," in *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.*, 984–987, IEEE (2006).

41 G. D. Joshi and J. Sivaswamy, "Colour retinal image enhancement based on domain knowledge," in *Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on*, 591–598, IEEE (2008).

42 Y. Wang, W. Tan, and S. C. Lee, "Illumination normalization of retinal images using sampling and interpolation," in *Medical Imaging 2001*, 500–507, International Society for Optics and Photonics (2001).

43 P. Feng, Y. Pan, B. Wei, *et al.*, "Enhancing retinal image by the contourlet transform," *Pattern Recognition Letters* **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," *Journal of the Franklin Institute* **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 *WDIC2005 ARPS workshop on digital image computing, Brisbane, Australia*, 163–168 (2005).

46 A. Hagiwara, A. Sugimoto, and K. Kawamoto, "Saliency-based image editing for guiding visual attention," in *Proceedings of the 1st international workshop on pervasive eye tracking & mobile eye-based interaction*, 43–48, ACM (2011).

47 S. L. Su, F. Durand, and M. Agrawala, "De-emphasis of distracting image regions using texture power maps," in *APGV*, (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 Transactions on Circuits and Systems for Video Technology* **21**(3), 360–364 (2011).

49 T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, *et al.*, "The diaretdb1 diabetic retinopathy database and evaluation protocol.," in *BMVC*, 1–10 (2007).

50 P. Prentasic, S. Loncaric, Z. Vatavuk, *et al.*, "Diabetic retinopathy image database (dridb): a new database for diabetic retinopathy screening programs research," in *Image and Signal Processing and Analysis (ISPA), 2013 8th International Symposium on*, 711–716, IEEE (2013).

51 L. Giancardo, F. Meriaudeau, T. P. Karnowski, *et al.*, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Medical image analysis* **16**(1), 216–226 (2012).

52 G. D. Joshi and J. Sivaswamy, "Colour retinal image enhancement based on domain knowledge," in *Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on*, 591–598, IEEE (2008).

53 L. Tang, M. Niemeijer, J. M. Reinhardt, *et al.*, "Splat feature classification with application to retinal hemorrhage detection in fundus images," *IEEE Transactions on Medical Imaging* **32**(2), 364–375 (2013).

54 G. Azzopardi, N. Strisciuglio, M. Vento, *et al.*, "Trainable cosfire filters for vessel delineation with application to retinal images," *Medical image analysis* **19**(1), 46–57 (2015).

55 A. Torralba, A. Oliva, M. S. Castelhano, *et al.*, "Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search.," *Psychological review* **113**(4), 766 (2006).

56 M. Mancas, "Relative influence of bottom-up and top-down attention," in *Attention in cognitive systems*, 212–226, Springer (2008).

57 T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, *et al.*, "Diaretdb0: Evaluation database and methodology for diabetic retinopathy algorithms," *Machine Vision and Pattern Recognition Research Group, Lappeenranta University of Technology, Finland* (2006).

58 Z. Bylinskii, T. Judd, A. Oliva, *et al.*, "What do different evaluation metrics tell us about saliency models?," *arXiv preprint arXiv:1604.03605* (2016).

List of Figures

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

3 Loss function in the absence of class-imbalance ($\beta = 1$). (top) 3D view of a loss function. Saddle point is shown in red color. (bottom) 2D 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.

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

5 Gamma correction. (a) original image (b) corrected image with $\gamma = 2$ (d) corrected image with $\gamma = 0.5$.

6 Procedure to obtain lesion-level ground truth from regional marking.

7 Evolution of the filters during training of hard exudates saliency. (a) 3 channels of W_{orient} from stage 2 (b) 3 channels of $CS3$ filters from stage 4 (c) 3 channels of random filters from stage 2 (d) 3 channels of random filters from stage 4.

8 Hard exudate saliency. (a) original color fundus image (b) pre-processed image (c) ground truth. Computed saliency maps of (d) Proposed (e) Itti-Koch (f) SR (g) GBVS (h) AIM (i) Rare (j) Torralba (k) Judd.

9 Hemorrhage saliency. (a) original color fundus image (b) pre-processed image (c) Gaussian convolved ground truth. Computed saliency maps of (d) Proposed (e) Itti-Koch (f) SR (g) GBVS (h) AIM (i) Rare (j) Torralba (k) Judd.

10 Predicted saliency for normal cases. (a) Normal color fundus image and saliency maps for (b) hard exudate (c) hemorrhage.

List of Tables