# Visual Perception Based Assistance for Fundus Image Readers

Thesis submitted in partial fulfillment
of the requirements for the degree of

*MS by Research*
*in*
*Electronics & Communication Engineering*

by

Samrudhdhi B. Rangrej
201432631
rangrej.bharat@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500 032, INDIA
July 2017

To my amazing parents

International Institute of Information Technology

Hyderabad, India

## CERTIFICATE OF AUTHORSHIP

I, Samrudhdhi B. Rangrej, declare that the thesis, titled "Visual Perception Based Assistance for Fundus Image Readers", and the work presented herein are my own. I confirm that this work was done wholly or mainly while in candidature for a research degree at IIIT-Hyderabad.

10/07/2017

Date

Signature of the Candidate

International Institute of Information Technology

Hyderabad, India

# CERTIFICATE

It is certified that the work contained in this thesis, titled "Visual Perception Based Assistance for Fundus Image Readers" by Samrudhdhi B. Rangrej, has been carried out under my supervision and is not submitted elsewhere for a degree.

10 July 2017

Date

Adviser: Prof. Jayanthi Sivaswamy

# Acknowledgments

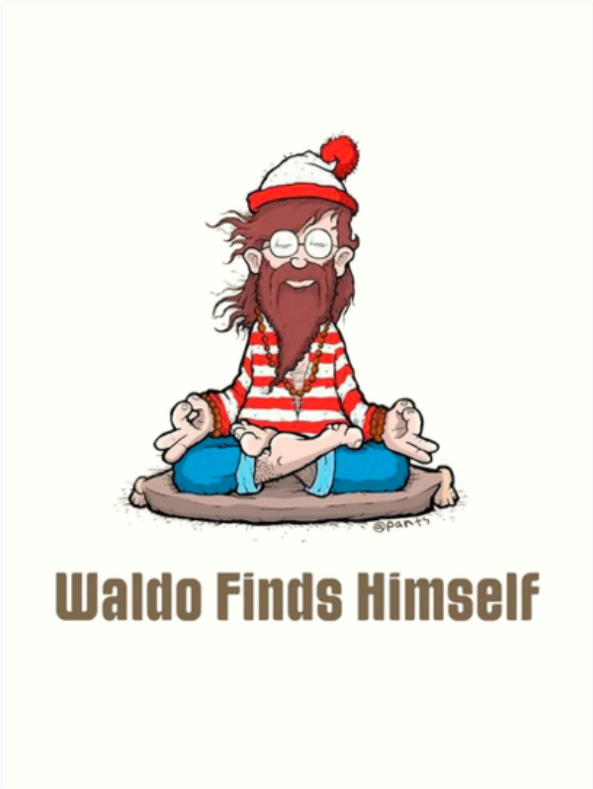
For transforming me from agitated struggler to a calm fighter, I will always be grateful to my advisor Prof. Jayanthi Sivaswamy. I will never forget her face when she advised me,

> "you may not get optimal things in life, but you will always find suboptimal choice; you just have to make some adjustments. Believe in yourself and don't be disheartened, because life only rewards to fighters".

I haven't only gained academic knowledge from her but also a new perspective to interpret things differently. I am also thankful to Dr. Priyanka Srivastava for her valuable inputs and feedbacks for eye-tracking study.

I am very fortunate to earn precious friendship of my seniors, Tabish, Palash, Ujjwal and Nishit. I am indebted to them for helping me out with research and most importantly, keeping my spirit up in the time of failure. I am thankful to my batch-mates, Karthik and Raghav for many fruitful discussions and to Arunava and Chetan (both are pursuing PhD, hence share some commonality) for philosophical advice about life. I also thank Pujitha, Jahnvi, Aabhas and all CVITians for making my experience a memorable one.

I also thank *Waldo*, who himself was lost but helped me find my way through research! My presentations are incomplete without mentioning *"Where is Waldo?"*, so how can my thesis be complete without mentioning it! I think all graduate students (obviously, including me) are like Waldo at the beginning; lost! But what happens at the end?




credit: Josh Mecouch(@ pants)

# Abstract

Diabetic Retinopathy (DR) is a condition where individuals with diabetes develop a disease in the inner wall of eye known as retina. DR is a major cause of visual impairments and early detection can prevent vision loss. Use of automatic systems for DR diagnosis is limited due to their lower accuracy. As an alternative, *reading-centers* are becoming popular in real-world scenarios. Reading center is a facility where retinal images coming from various sources are stored and trained personals(who might not be experts) analyze them. In this thesis we look at techniques to increase efficiency of DR image-readers working in reading centers.

The first half of this thesis aims at identifying efficient image-reading technique which is both fast and accurate. Towards this end we have conducted an eye-tracking study with medical experts while they were reading images for DR diagnosis. The analysis shows that experts employ mainly two types of reading strategies: dwelling and tracing. Dwelling strategy appears to be accurate and faster than tracing strategy. Eye movements of all the experts are combined in a novel way to extract an optimal image scanning strategy, which can be recommended to image-readers for efficient diagnosis. In order to increase the efficiency further, we propose a technique where saliency of lesions can be boosted for better visibility of lesions. This is named as an *Assistive Lesion Emphasis System(ALES)* and demonstrated in the second half of the thesis. ALES is developed as a two stage system: saliency detection and lesion emphasis. Two biologically inspired saliency models, which mimic human visual system, are designed using unsupervised and supervised techniques. Unsupervised saliency model is inspired from human visual system and achieved 10% higher recall than other existing saliency models when compared with average gazemap of 15 retinal experts. Supervised saliency model developed as deep learning based implementation of biologically inspired saliency model proposed by Itti-Koch(Itti, L., Koch, C. and Niebur, E., 1998) and achieves 10% to 20% higher AUC compared

to existing saliency model when compared with manual markings. Saliency maps generated by these models are used to boost the prominence of lesions locally. This is done using two types of selective enhancement techniques. One technique uses multiscale fusion of saliency map with original image and other uses spatially varying gamma correction; both increases CNR of lesions by 30%. One saliency model and one selective-enhancement technique are clubbed together to illustrate two complete ALESs. DR diagnosis done by analyzing ALES output using optimal strategy should presumably be faster and accurate.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

"The whole is greater than the sum of it's parts."

– *Aristotle*

An alliance between two individuals, where each benefit from the participation of other, is called *mutualism*. An example of mutualism is a symbiotic relationship where two species, who are susceptible otherwise, form a resilient association. This thesis is based on concept of mutualism between human reader and computerized diagnosis in a setting called *reading-center*.

In the past medical experts who examined patients also analyzed images. But the shortage of highly trained experts necessitates lessening the burden of image analysis from experts. This has given rise to new type of services called *reading-centers*[1]. Reading centers are facilities with computers and large data storage. Medical images coming from various sources are collected and stored here. These are staffed with readers and some experts. Readers, who acts as pseudo-experts, 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(e.g. Doheny Image Reading Center (DIRC) [2]). Manual image reading is a lengthy process. Hence, computerized diagnosis has been investigated in last three decades. Modern computing power has made computerized diagnosis faster than manual diagnosis. Yet, it has not achieved human accuracy. In this thesis we advocate a mutual design where computerized diagnosis, rather than working independently, assists image-reader. This is called Computer Assisted Diagnosis(CAD) and achieves best of computerized and manual

diagnosis; it is fast as well as accurate. Hence, as the Aristotle said, "The whole is greater than the sum of it's parts."

Reading center like settings play an important role in *screening* and *triage*. Since early detection is preferred in effective disease management, clinical screening is aimed at identifying individuals who may be at risk of developing a disease. 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].

This thesis looks at a disease called diabetic retinopathy(DR) and aims to develop *diagnostic strategy* and *assistive tools* for DR image readers. The proposed solutions can be employed in reading centers to increase throughput of DR screening.

## 1.1 Diagnosis of Diabetic Retinopathy

Diabetic Retinopathy (DR) is a condition where individuals with diabetes develop a disease in retina. Individuals in the age group of 20 to 74 years and with Type-I and Type-II diabetes have higher risk of developing DR [6]. By 2014, approximately 422 million people in whole world live with diabetes [7], out of which 1/3 people have DR. India alone has 99 million people with diabetes, out of which 1/4 people have DR; which amounts to nearly 25 million DR patients [8]. Diabetic retinopathy, if left untreated, causes visual impairments and in some cases permanent blindness. Diabetic retinopathy along with age-related macular degeneration(a type of DR) is cause of 2% cases of visual impairment and permanent blindness (see Figure 1.1).

Early detection of DR can prevent vision loss. Hence, many countries have started DR screening. Early Treatment of Diabetic Retinopathy Study(ETDRS)[9] started by U.S. government is one such example. DR screening is implemented by arranging camps at remote sites or at local

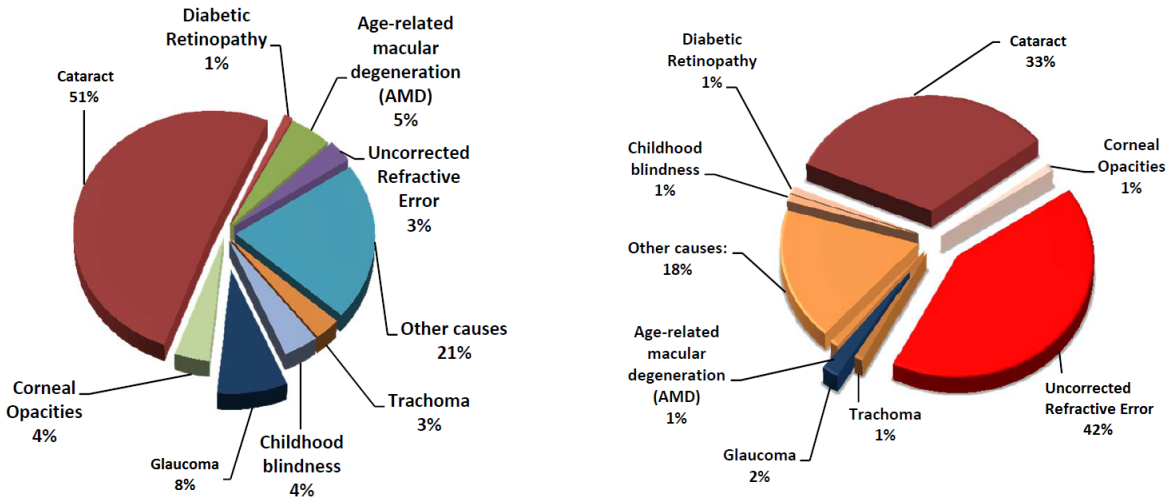



Figure 1.1: Relative contribution of diseases in causing (right) blindness and (left) visual impairment [source:WHO].

hospitals(e.g. Diabetic RetinaScreen [10]). Screening camps invite adults who are at a risk of developing DR. Trained practitioner at the camp performs dilated eye test on the patients. This involves dilation of pupil and capturing a photograph of retina using non-invasive fundus camera, either table mounted or hand-held. A fundus photograph of normal retina is shown in Figure 1.2. Abnormal Retina has lesions called hemorrhage and hard exudate; hemorrhage is blood leakage and appears as dark lesion, hard exudate is lipid leakage and appears as bright lesion (see Figure 1.3, right most image). Images captured at camps are brought to reading-centers where readers analyze these images and recommends appropriate consultation. For timely detection and care, it is very important that reading centers have high throughput. This is done by developing efficient *diagnostic strategy* and *assistive tools* for DR image readers.

## 1.2 Diagnostic Strategy

Increasing throughput of reading center is possible if readers have efficient (i.e. fast) and accurate diagnosis process. This is done by training readers to read images like experts. But, the relationship between accuracy of diagnosis and expertise has been observed to be very complex [11]. This indicates that accuracy of diagnosis depends on multiple aspects opposed to just one, namely expertise. Understanding of these aspects is long overdue. In the first half of this

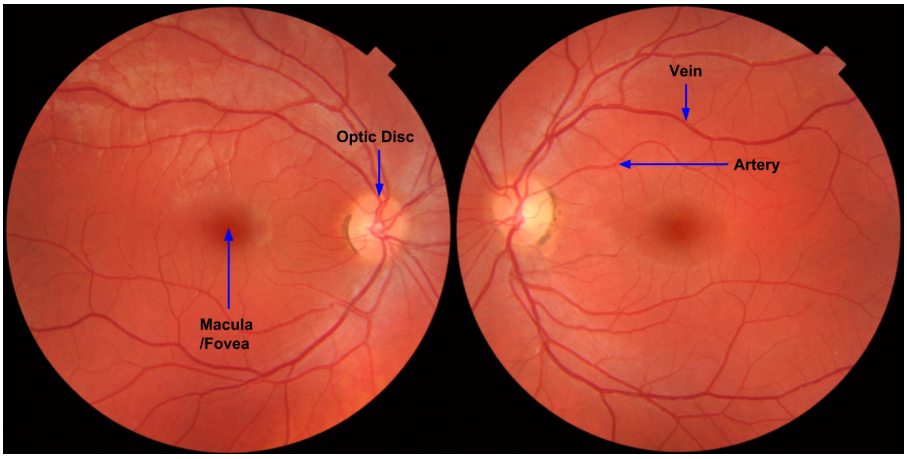



Figure 1.2: Color fundus photograph. (right) retina of right eye (left) retina of left eye.

thesis, we aim at examining the best practices and strategies used for DR diagnosis from fundus images. We examine the role of expertise in the diagnosis process and seek to understand its manifestation if any, at both conscious and subconscious levels. We also aim to develop an optimal strategy which can be used by fundus readers for efficient diagnosis.

## Contributions

An eye tracking study has been conducted while various practitioners were diagnosing DR from fundus images. Eye movements are later analyzed to understand the strategies used for diagnosis. Major contributions of this work are as follows.

- *Eye-tracking study.* A large scale eye tracking study has been conducted with 56 participants; out of which 44 subjects were medical experts coming from 6 different hospitals and remaining 12 subjects were engineering students from our institute who served as novices. Collected eye-tracking data can be used for analyzing diagnosis process.

- *Understanding of diagnostic strategy.* Eye movements are analyzed to understand the diagnostic strategies. We have designed new metric called coefficient of scanning($CS$). Based on $CS$, strategies used by various practitioners are categorized into: *dwelling* and *tracing*. It is found that *dwelling* leads to faster and more accurate diagnosis than *tracing*.

- *Optimal scanning strategy.* Gaze patterns of retinal experts are quantized into nine retinal zones as suggested by ETDRS [12] and combined using transition matrix. An optimal scanning strategy is extracted from above analysis. Optimal strategy recommends readers weightage and sequence in which nine retinal zones should be scanned for fast and accurate diagnosis.

## 1.3 Assistive Tools

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. As a solution, computerized diagnosis methods are developed. But due to various challenges in automatic image analysis, like artifacts and nonuniform illumination (see Figure 1.3), accuracy of these methods is low; which is not acceptable. 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 [13]. 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. We propose an alternate solution to draw a reader's attention by emphasizing 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 [14, 15]. Boosting the contrast of such salient locations has been shown (in the case of natural scenes) to attract one's attention [16, 17, 18]. Lesion-emphasis is both computationally efficient and clutterless. In the second half of this thesis, we take this alternate reader-centric approach and propose a novel CAD, which employs saliency of a region to determine the amount of its emphasis.

### Contributions

We developed an 'Assistive Lesion Emphasis System(ALES)' for better visibility of abnormalities. ALES has two stages, saliency computation and lesion-emphasis. First stage is more import

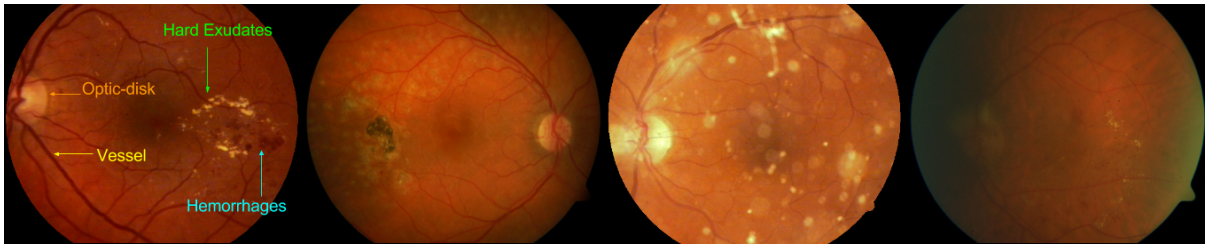



Figure 1.3: From left to right: Retinal image with DR lesions, dark artifacts, bright artifacts and varying illumination.

as accuracy of ALES depends on saliency. Two designs of ALES have been demonstrated here, one with unsupervised and other with supervised saliency models. Both ALESs have different and interchangeable lesion-emphasis stages.

- ALES-design:1
  - *Unsupervised saliency model (ALES1).* This is inspired from human visual system. Spatially-varying erosion and dilation(SED) operations are used to mimic Gaussian-shaped human fixation. Retinal Ganglion cells are mimicked using center-surround filters. Spatio-topic fusion is done by nonlinear combination of mean and variance maps.
  - *Lesion emphasis (ALES1).* Saliency maps generated by above model is fused with original image at multiple scales. This enhances lesions locally. Fusion is governed by a mixing parameter. A reader can change value of mixing parameter interactively to control the degree of enhancement.
- ALES-design:2
  - *Supervised saliency model (ALES2).* This is a deep neural network based implementation of an existing, biologically inspired saliency model by Itti and Koch [15]. Neural network fine-tunes standard filters for DR lesions and also learns new filters. Training of this network is done using novel loss function, custom-designed to handle range mismatch problem between output saliency and ground-truth.
  - *Lesion emphasis (ALES2).* Gamma correction, with constant value of gamma, enhances image globally by stretching dynamic range of intensity values. Spatially-varying gamma correction is achieved by defining parameter gamma as a function

of saliency. This modifies dynamic range of local intensities, resulting into selective enhancement of salient regions.

## 1.4 Organization of Thesis

Remaining thesis is organized as follows. Chapter 2 describes eye-tracking study and related analysis. Chapter 3 and 4 discuss ALES with unsupervised and supervised techniques respectively. Conclusions are presented in Chapter 5.

# Chapter 2

## Optimal Search Strategy for Readers

### *An Eye Tracking Study*

> "Now, the value of an idea has nothing whatsoever to do with the sincerity of the man who expresses it. Indeed, the probabilities are that the more insincere the man is, the more purely intellectual will the idea be, as in that case it will not be coloured by either his wants, his desires, or his prejudices."
>
> – *Oscar Wilde*

A popular health care initiative to address DR has been screening via a physical eye examination. Images captured at screening sites are sent to reading centres where trained readers scrutinize them and give preliminary diagnosis. This forms the basis for referral to retina experts for more comprehensive physical eye exam. In order to increase reach of DR screening and scaling up the number of people who can be screened, readers at reading-centers should be acquainted with efficient image reading technique. In this chapter we aim to examine the role of expertise in DR diagnosis process from the fundus image. Further, we also aim to evaluate the best practices and strategies used for DR diagnosis from fundus image. At the end we extract an optimal search strategy which can be recommended to readers for efficient diagnosis.

## Background

*Eye-tracking* is a popular technique which is used extensively to investigate visual perception in a range of tasks. Given that both 2D and 3D images are used in diagnosis, eye tracking studies

aimed at understanding medical image perception have been reported on CT images [19, 20, 21], X-ray [22, 23], microscopy [24, 25, 26] and mammography [27, 28]. Visual perception of fundus images has not been investigated and to the best of our knowledge, this is the first attempt to investigate visual search for DR lesions in fundus images.

In medical image perception specifically screening process, eye-tracking enables us to evaluate the conscious and subconscious aspects of perception. Conscious aspect can be understood as what and where to look. This involves domain knowledge, which directs the search for specific abnormalities in the specific location(s) while ignoring other normal anatomical structures. Subconscious aspect is understood as how to look. This involves the gaze pattern, dwelling and time delay which are difficult to teach or learn. Given the importance of evaluation of eye-movement in the diagnostic process, the current study examined this conscious and subconscious aspects during screening of diabetic retinopathy.

Retinal anatomy consists of three major structures: Optic Disk, vessel network, and macula/fovea. First two are irrelevant for DR detection and is ignored by practitioners with a short training. The macula is responsible for color vision with high acuity and hence is of specific clinical interest. The location of lesions relative to macula represents severity of the disease and this knowledge is gained only from medical training. Given the vital role of knowledge in identifying the possible location and related diagnostic value, it becomes necessary to evaluate the search strategies as a function of knowledge in DR.

## 2.1 Eye Tracking Experiment

### Stimuli Images

A dataset of 145 abnormal and 50 normal fundus images was obtained from LV Prasad Eye Insitute, Hyderabad [29]. A balanced (normal cases:abnormal cases $\sim$ 1) subset of 40 images was selected by senior retina consultant from this dataset.

## Participants

In order to maintain diversity in the participants, retina experts were invited from 5 different eye hospitals [1]. A total of 44 retina experts accepted the invitation and participated in the study. Retina experts were classified into 3 categories, namely, consultants, fellows and residents or optometrist (see Table 2.1). 12 engineering students from our home institute were also included to fulfill the role of novices. They were given a short training with 10 example fundus images.

| Level of expertise | Number of participants | Range of experience (years) |
|---|---|---|
| Consultant | 13 | 7-18 |
| Fellow | 17 | 3-7 |
| Resident/Optometrist | 14 | 1-12 |
| Novices | 12 | 0 |

Table 2.1: Description of Participants.

## Material

Experiments were carried out in dark room with constant environment condition. Tobii X2-30 eye tracker (sampling rate 30Hz) with a 15.6" display screen were used. Stimuli (images) were down-scaled to fit the display size. An image classification tool was developed in MATLAB 8.2 (see Figure 2.1) and a small tutorial was given to all participants to ensure easy usage of the software.

## Experiment Design

The instruction given to all participants is shown in Figure 2.2. Eye tracking was conducted during each trial with prior consent from all the participants. A 5-point calibration was performed at an interval of 5 images. The response time, response category and eye position $(x, y)$ with timestamp were stored for each trial. The session lasted for approximately 20 minutes. Eye positions were classified into fixation, saccade, and glissade using adaptive algorithm proposed

[1] LV Prasad Eye Insititute, Hyderabad [29]; Narayana Nethralaya, Bangalore [30]; Neoretina Eyecare Institute, Hyderabad [31]; Anand Eye Institute, Hyderabad [32]; Medivision Eye Care Centre, Hyderabad [33], Centre for Sight [34], Hyderabad


Figure 2.1: Example display from the software tool developed for the experiment.

| |
|---|
| You will be shown a retinal image on the screen.<br><br>Your task is to decide whether the image is normal or abnormal as fast as possible.<br><br>The images are deemed to be abnormal if they have hard exudates and/or hemorrhages and/or cotton wool spots.<br><br>You are requested to report your decision by pressing the corresponding buttons: Normal or Abnormal for a particular image. |

Figure 2.2: Instructions given to participants.

by [35]. Eye positions classified as glissade were removed from the data and the remaining data were registered to actual image size (in pixels).

## 2.2 Definitions

Lets begin with definitions of some quantities used for analysis of eye tracking data.

**Total track length.** The Euclidean distance between two consecutive fixations is defined as track length in *pixel* unit. The sum of all the track lengths for an image $j$ for an $i^{th}$ participant is the total track length $l_{ij}$. $l_{ij}$ depends on the subjective behavior of the participant as well as

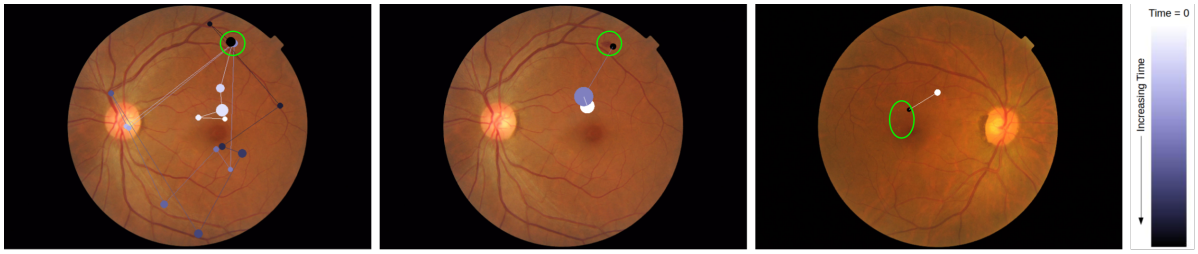



Figure 2.3: Total track length for sample images. Lesion locations are (a)-(b) Gaze-tracks for two different subjects for an image with a lesion (marked with green circles) in the periphery. (c) Gaze-track for an image with more centrally located lesion.

the location of the lesion in the image. Figure 2.3 illustrates these points. Figure 2.3 (a)-(b) show tracks for 2 subjects on one image. It is evident that one subject takes a very long route whereas other takes direct route to the lesion location indicated by a green circle. Assuming that the visual search starts from the center, image with lesion(s) in the peripheral region will have larger $l_{i,j}$ than an image with lesion(s) at the center. This is evident from Figure 2.3(b)-(c).

**Standardized track length.** The effect of location of lesion on $l_{ij}$ has to be nullified in order to capture inter-subject variability. The normalized track length is defined as standardized track length and computed as follows.

$$Standardized\ track\ length\ (L) = \frac{l_{ij}}{\sum_i l_{ij}} \tag{2.1}$$

$L$ is ratio of total track length and sum of total track length, hence is unitless quantity. It takes values between 0 and 1.

**Total dwell duration.** The time spent on each fixation is defined as dwell duration which is in *micro-seconds*. The sum of dwell duration for each fixation (including revisits) is considered as total dwell duration. Total dwell duration of $i^{th}$ participant for $j^{th}$ image is denoted as $d_{ij}$. $d_{ij}$ also depends on the subjective behavior of the participant and the conspicuity of lesions in the image. Figure2.4 illustrates this point. Figure2.4(a)-(b) shows that dwell pattern of two different subjects where one subject has fewer fixations than the other and hence lower total dwell duration. Figure2.4(b)-(c) on the other hand illustrates the dependency of $d$ on conspicuity. Here, an image with subtle lesions is seen to lead to longer dwell duration per fixation in a subject than an image with prominent lesions.

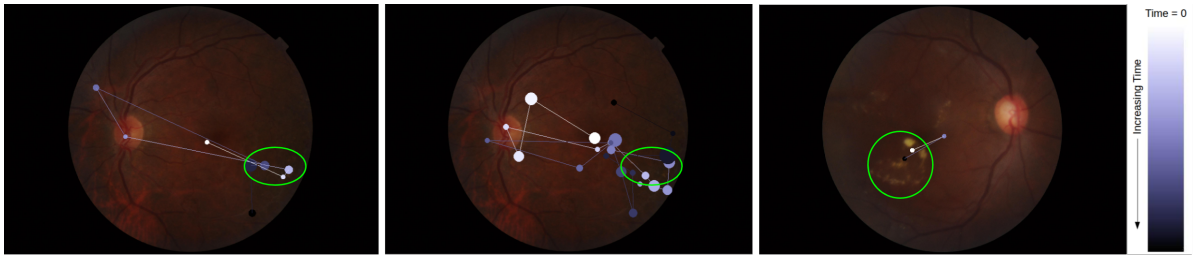



Figure 2.4: Comparison of dwell duration. Dwell duration is proportional to the radius of the disk representing fixation. Lesion locations are marked with green ovals. (a)-(b) Gaze-track for two different subjects for an image. Very different dwell pattern shows total dwell duration depends on subjective behavior. (b)-(c) Gaze-track for two different stimuli images. Gaze-track for image with subtle(prominent) lesion has more(less) total dwell duration.

**Standardized dwell duration.** The effect of conspicuity of lesion on $d_{ij}$ has to be nullified in order to capture inter-subject variability. The normalized total dwell duration termed as standardized dwell duration is defined as follows.

$$Standardized\ dwell\ duration\ (D) = \frac{d_{ij}}{\sum_i d_{ij}} \tag{2.2}$$

standardized dwell duration is also unitless quantity as it is the ratio of total dwell duration and summation of total dwell duration. It takes on a value between 0 to 1.

**Coefficient of scanning**. This is defined as the ratio of standard dwell duration and standard track length.

$$Coefficient\ of\ scanning\ (CS) = \frac{D}{L} \tag{2.3}$$

Depending on the value of $D$ and $L$, $CS$ can assume any positive value. Theoretically, $CS$ can assume value=$\infty$ when $L$ is zero. However, this is possible only if there are less than or equal to one fixations for an image and this is practically nearly impossible. Hence, practically $CS$ can assume any large($< \infty$) positive value. Coefficient of scanning represents the balance between track length and dwell duration and thus can provide insights into search strategies of subject underlying the Normal/Abnormal decision task.

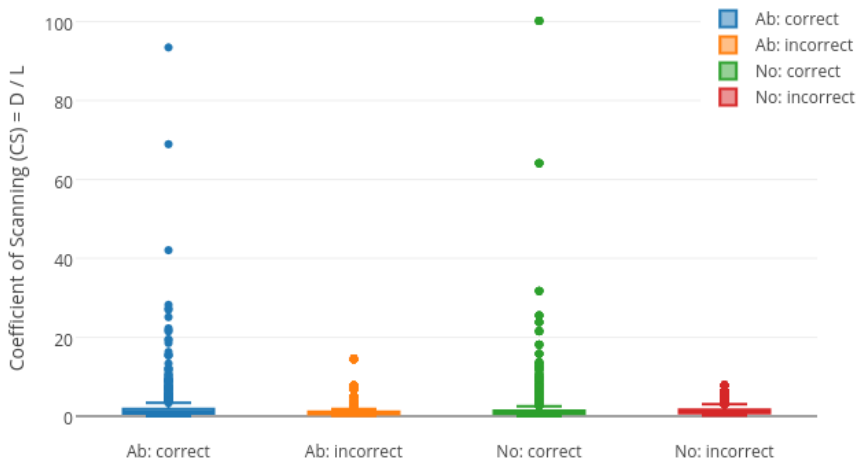



Figure 2.5: $CS$ values for 2240 responses.

## 2.3 Dwelling vs Tracing

The responses of the 56 participants for 40 images results in a total of 2240 responses. These were divided into 4 categories: (i) correct response for abnormal case (True Positive) (ii) incorrect response for abnormal case (False Negative) (iii) correct response for normal case (True Negative) (iv) incorrect response for normal case (False Positive). $CS$ was computed for each of these categories and the results are shown in Table 2.2 and Figure 2.5. $CS$ is seen to be significantly higher for True positive/negative categories than False positive/negative categories for both Normal and Abnormal cases. The average $CS$ value for the four groups is 1.5.

| Image category: Response | $CS$ | p-value |
|---|---|---|
| Abnormal: Correct | $2.1 \pm 5.0$ | $8.97 \times 10^{-7}$ |
| Abnormal: Incorrect | $1.2 \pm 1.6$ | |
| Normal: Correct | $1.6 \pm 4.5$ | $4.07 \times 10^{-4}$ |
| Normal: Incorrect | $1.4 \pm 1.5$ | |

Table 2.2: Average Coefficient of Scanning for various image category and response pair.

This average $CS = 1.5$ value was used to analyze the responses. Responses with $CS <= 1.5$ implies $D <= 1.5L$, i.e. Dwelling is less but the gaze-track is long. Hence, we call this search

strategy as *Tracing*. On the other hand, responses with $CS > 1.5$ has $D > 1.5L$, i.e. gaze-track is short but dwelling is more. We call this scanning strategy as *dwelling*.

### Performance Analysis of *Dwelling* and *Tracing* strategies

*Dwelling* always leads to correct response (see Table 2.2). Accuracy achieved by various subject groups with *dwelling* strategy is shown in Figure 2.6. Most of the participants have accuracy $> 80\%$. Achieved accuracy is thus high regardless of the expertise level of the participant ($p = 0.5219$).

*Tracing* strategy, on the other hand, leads to lower accuracy and analysis also shows that expertise affects the performance with this strategy. Specifically, accuracy decreases with levels of expertise (p-value $= 4.8 \times 10^{-4}$) (see Figure 2.7). In order to understand this trend, further analysis of the gaze pattern was done.

*Tracing* is characterized by higher standardized track length. Ideally, longer track length should be a result of greater spatial coverage. However, longer track length can also be due to revisits to regions scrutinized earlier. Therefore, we computed the correlation coefficient between track length and number of revisits and found it to be 0.8. This implies, the number of revisits increases with track length (see Figure 2.8). This can also be seen from the sample case shown in Figure 2.3(a) where a subject had longer gaze-track only due to revisits. Based on the above analysis, we can conclude that revisits appears to be beneficial for subjects to make a correct decision only if they have higher expertise.

Average response time with *dwelling* and *tracing* is 4.6s and 5.4s, respectively. Thus, *dwelling* appears to be an efficient strategy for accurate and fast decision, regardless of a subject's expertise level.

## 2.4 Discussion

The results indicate that readers use mainly two kinds of strategies while reviewing retinal images, namely tracing and dwelling. Tracing involves low dwell time for a particular region

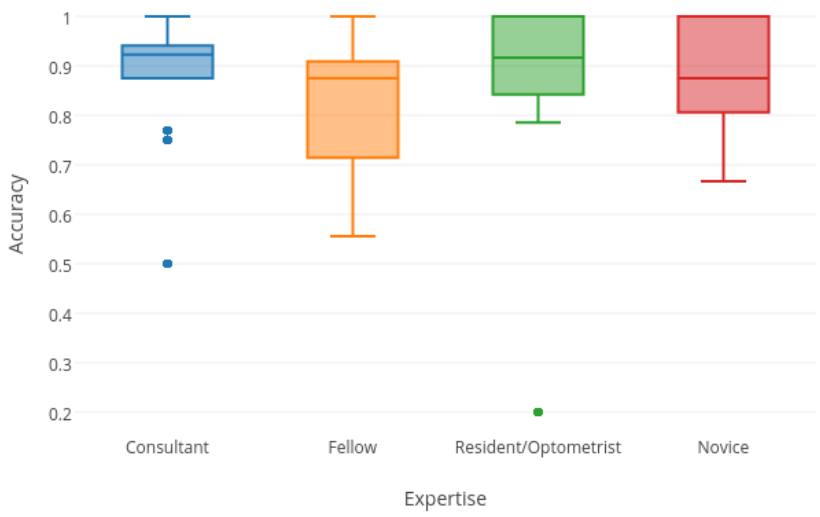



Figure 2.6: Accuracy achieved with *dwelling* strategy.

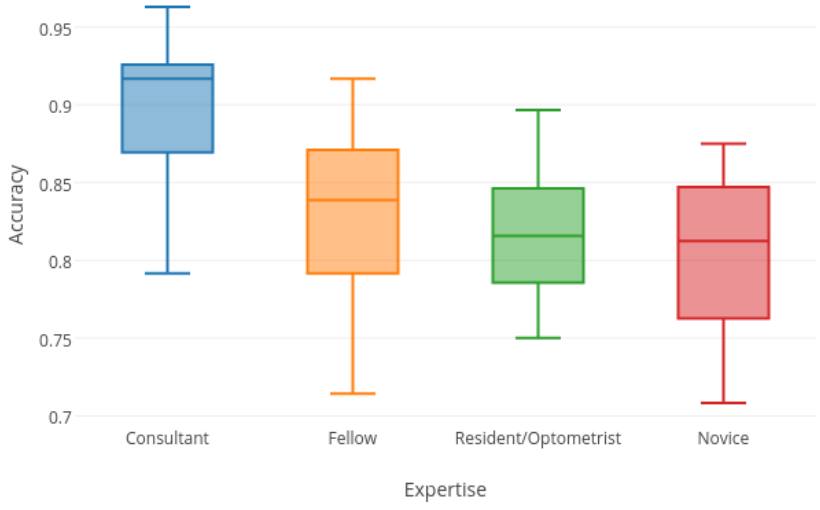



Figure 2.7: Accuracy achieved with *tracing* strategy.

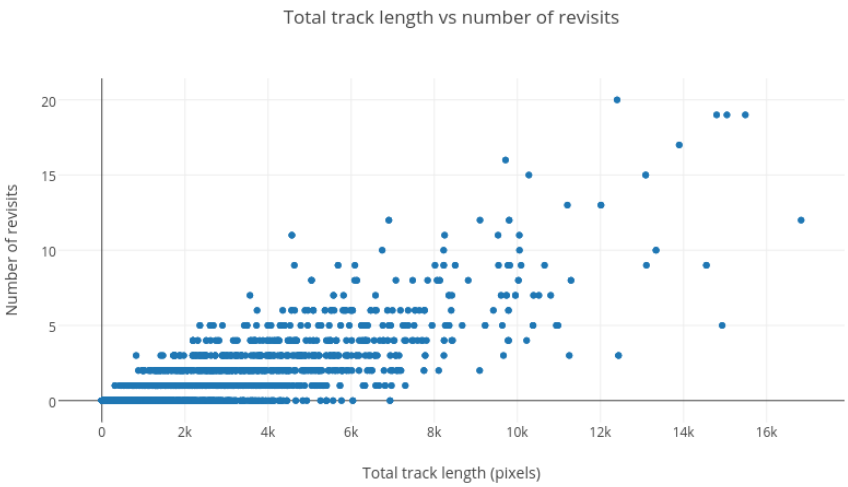



Figure 2.8: Correlation between track length and revisits.

and leads to a reader reviewing the region over multiple revisits. The net effect in Tracing is marked by long gaze-track and slow response. However, revisiting a region aids experts more than semi-experts and novices as indicated by the comparatively lower accuracy of novices or semi-experts than that of experts when employing tracing as strategies. It can be argued that revisiting enables experts to gather more evidence to support decision making but might not help semi experts or novices who lack the ability to gather and assess information in the region.

On the other hand, dwelling strategy involves longer dwell time for a region which is better suited to gather adequate information for decision making. This is affirmed by the fact that participants employing dwelling strategy have higher accuracy and response time. The long scrutiny enables full attention to a region which should be beneficial to a participant in overcoming any lack of experience and expertise, which might be why the high accuracy of decision is achieved by all expertise-groups. As dwelling requires scrutinizing regions for a longer time, the decision is possible with less number of reviews and consequently, the track length and response time is low. This makes dwelling an accurate and fast strategy.

The result suggests that dwelling has an advantage over tracing strategy. This trend is particularly evident in the case of novices and semi-experts, whose detection accuracy drops during tracing strategy. Current results extend previous finding[36] on 3D scanning pattern, which

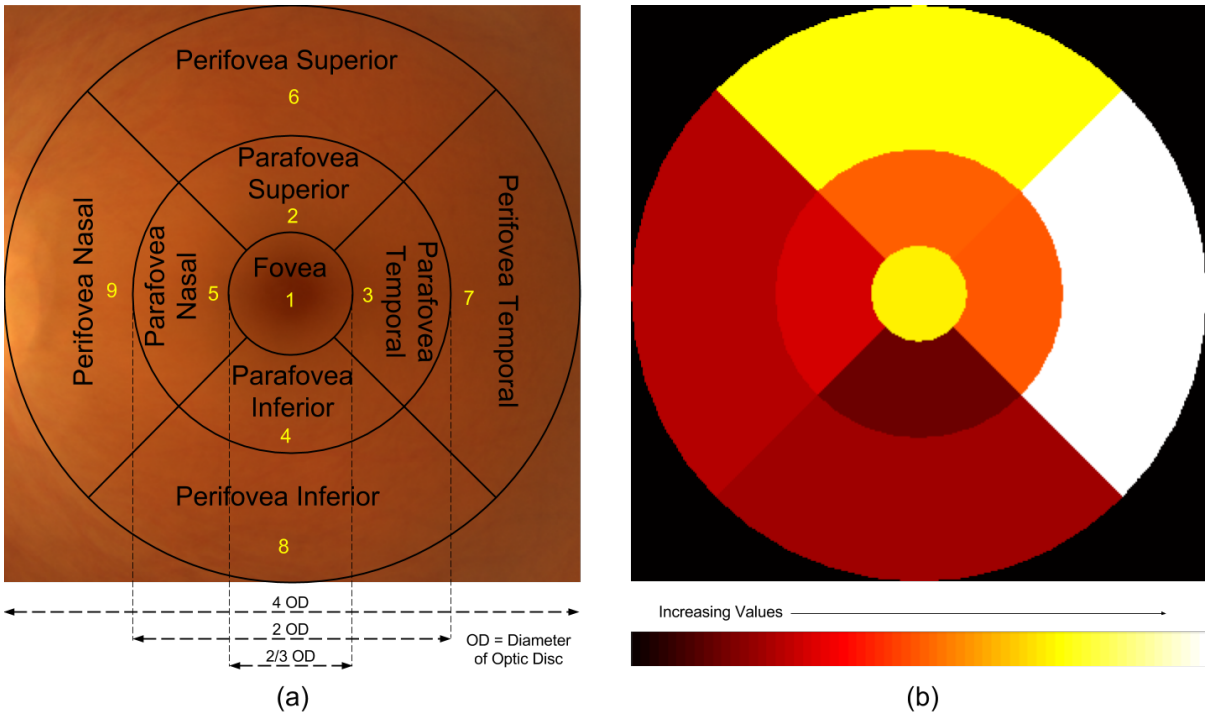



Figure 2.9: (a) Retinal zones recommended by ETDRS (b) Dwell map

shows a difference between drilling and scanning strategy. Their results show higher detection accuracy and smaller saccadic amplitude with drilling vs. scanning strategies.

Next we analyze search patterns and extract an optimal strategy.

## 2.5 Search Pattern Analysis

The ETDRS standards [12] specifies 9 zones of a fundus image (see Figure 2.9(a)) for review during diagnosis of diabetic retinopathy. In this section, eye movement patterns for *correct response* are analyzed in terms of the gaze transitions between these zones and zone-wise dwelling. The purpose of this analysis is to gain insights into best practices (search pattern) that underpin good performance.

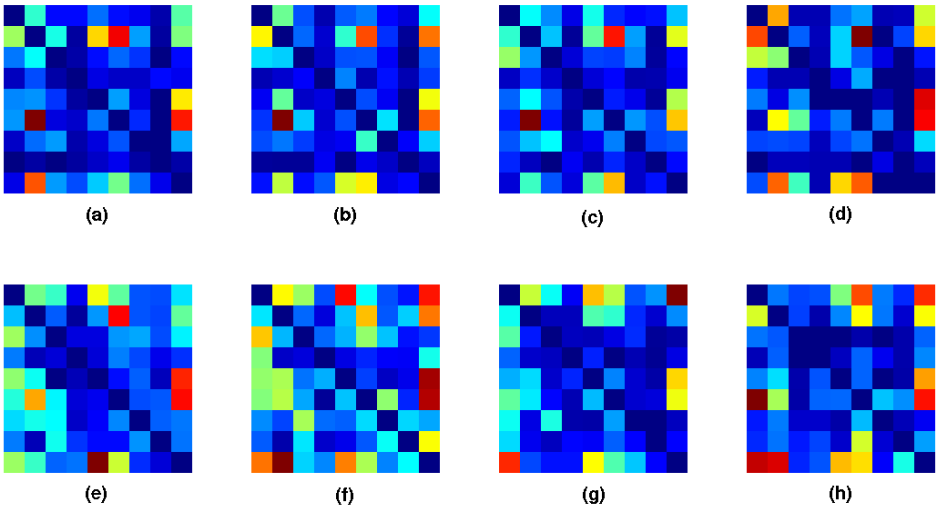



Figure 2.10: Average transition matrices for different expertise-groups. Top/bottom row: abnormal/normal cases. Left to right: consultants, residents, fellows, novices.

### 2.5.1 Transition Pattern

Transition is defined as the movement of eye-gaze from one zone to another. A matrix with the number of transitions from zone $j$ to zone $i$ as entry $(i, j)$ is defined as the transition matrix($TM$) [37]. Each retinal zone was assigned a number as shown in Figure 2.9(a). The transition matrix for abnormal and normal cases were analyzed separately. Computed average transition matrix for medically trained participants is shown in Figure 2.11. This matrix was not observed to be significantly different across expertise-groups with medical training (abnormal: p=0.0804, normal: p=0.2712), whereas, average $TM$ for novices is significantly different from the other expertise-groups (abnormal: p=0.0022, normal: p=0.0041).

### 2.5.2 Dwell map

A dwell map showing average time spent on dwelling in a zone was also constructed. This was done by first summing the dwell duration for fixations falling in particular zone. This sum is defined to be the zone-wise total dwell duration. Total dwell duration of $k^{th}$ zone for $i^{th}$

participant and $j^{th}$ image is denoted by $d_{ijk}$. Next, an average zone-wise total dwell duration for expertise-group $E$ was computed as follows.

$$d_k(E) = \frac{1}{ij} \sum_{j,i \in E_j} d_{i,j,k} \tag{2.4}$$

Here, $E_j$ is the set of participants who have given correct response for $j^{th}$ image and belong to expertise-group $E$. Difference between $d_k(E)$ for the four subject groups was not significant (abnormal: p=0.5726, normal: p=0.9375). Average dwell map for all participants in shown in Figure 2.9(b).

Zone-level analysis suggests that for correct response, expertise level has no influence on the dwell time whereas it does have an influence of the transition pattern. This motivates the next section where we try to extract an optimal transition pattern for scanning the retinal image.

## 2.6 Recommendation

### 2.6.1 Optimal Transition Pattern

The analysis was done after excluding the responses from novices. Trials for novices are ignored to avoid false practices. $TM$ is calculated for each of these trials and averaged separately into two classes to derive $TM(normal)$ and $TM(abnormal)$ as shown in the Figure2.11(a) and (b).

In order to extract an optimal transition which can be used for both normal and abnormal images, the transition matrices for these two classes are combined as follows.

$$TM_{avg} = \frac{TM(normal) + TM(abnormal)}{2} \tag{2.5}$$

Average transition matrix $TM_{avg}$ is shown in Figure 2.11(c).

Entries in $(i, j)$ and $(j, i)$ positions in $TM_{avg}$ shows number of transitions from $j$ to $i$ and from $i$ to $j$ respectively. Net number of transitions from $j$ to $i$ is $(i, j) - (j, i)$. Net transition matrix ($TM_{net}$) is defined as follows.

$$TM_{net} = TM_{avg} - TM_{avg}^T \tag{2.6}$$

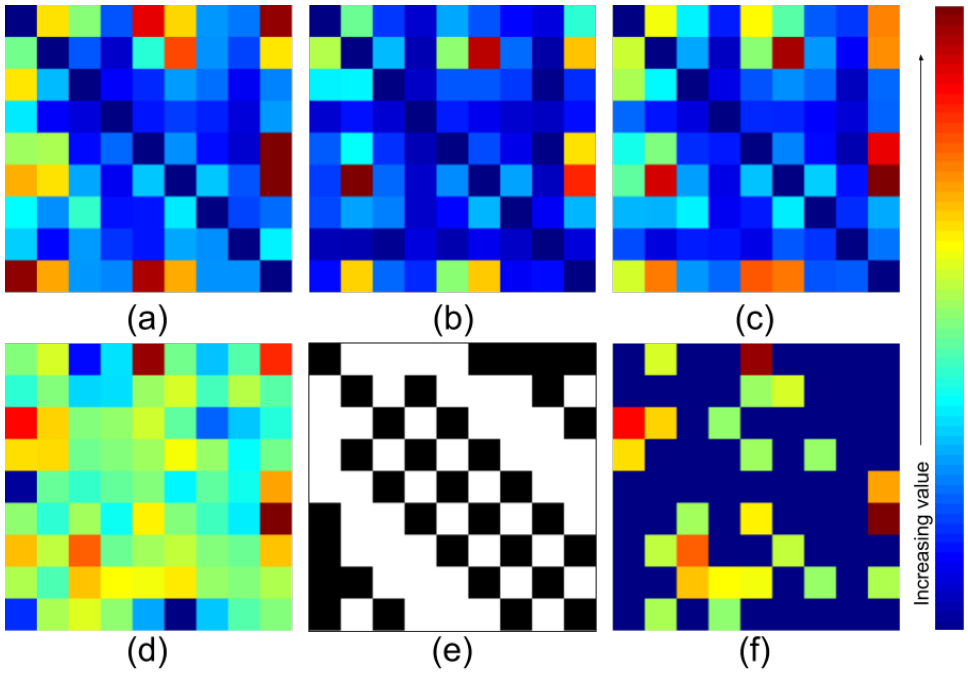



Figure 2.11: Average transition matrix. Transition matrix (a) for normal cases (b) for abnormal cases (c) average transition matrix $TM_{avg}$ (d) net transition matrix $TM_{net}$ (e) binary neighborhood matrix $NM$ (e) conditioned net transition matrix $TM_c$.

Where, $TM_{avg}^T$ is transpose of $TM_{avg}$. $TM_{net}$ is shown in Figure 2.11(d).

The aim is to extract an optimal transition pattern which can be used for assessing both normal and abnormal cases. Visual search for abnormal case is terminative as search ends as soon as a region with an abnormality is located whereas for normal case, search is exhaustive as it continues until all the regions are reviewed (see Figure 2.12). Weak transitions in $TM_{net}$ might have only occurred during exhaustive search and not during terminative search. Optimal transition pattern should progress from strongest to weakest transition. This ensures that important transitions happen at the start and as soon as an abnormality is located, search can be terminated. Also optimal transition pattern should have least number of revisits. In order to maintain easy and smooth eye movement, transitions between neighboring zones are given a higher priority.

In order to satisfy the last characteristic, a binary neighborhood matrix $NM$ is designed. As shown in Figure2.11(e), $(i, j)$ element of $NM$ has value equal to 1 if $i^{th}$ and $j^{th}$ zones share a

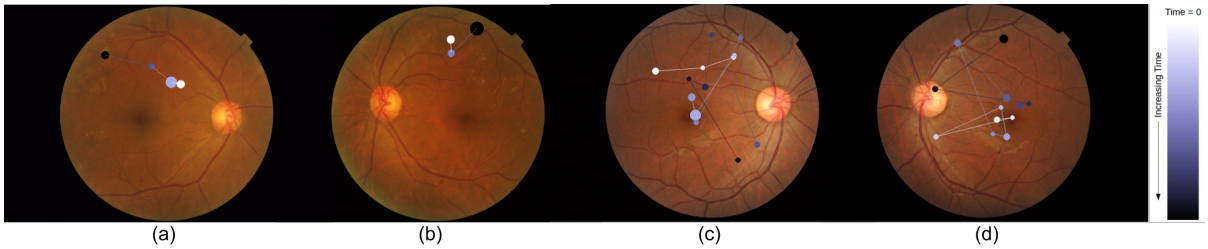



Figure 2.12: Gaze-pattern of one subject for 4 images. (a),(b) visual search for abnormal cases is terminative (c),(d) visual search for normal cases is exhaustive.

common boundary, else the value is 0. $TM_{net}$ is multiplied with $NM$ to allow transitions only between neighboring zones. Also $TM_{net}$ is anti-symmetric matrix, i.e. $(i, j) = -(j, i)$. To avoid duplicate transitions all elements with negative values are set to zero. The final constrained matrix $TM_c$ (see Figure2.11(f)) is defined as follows.

$$TM_c = TM_{net} \circ H(TM_{net}) \circ NM \tag{2.7}$$

Where $\circ$ signifies Hadamard product and $H(.)$ is Heaviside step function.

Optimal transition pattern is extracted from $TM_c$ using algorithm 1. First a function $Neig(x)$ is defined which returns neighbors of zone $x$ as an output. Next a set of unvisited zones $Z$ is defined. Initially all zones are unvisited. Nine zones can be traversed using eight transitions without any revisit. So arrays $S$ and $D$ of length=8 are defined which will store source and destination of the transition respectively. The pair of zones with strongest net transition is initialized as the first source and destination. A search for remaining transition pattern is done such that above stated characteristics are achieved. The destination of previous transition becomes source of next transition. The inner *If* condition searches for next destination in three hierarchical levels. The unvisited zone which has highest net transitions from current source is selected as the destination. Here only neighboring zone will be selected as $TM_c$ is constrained with $NM$. If such zone is not found than the neighboring unvisited zone which has strong potential destination is selected. If unvisited neighboring zone is not found than any unvisited zone with strong potential destination is selected. Once a zone is selected as destination, it is removed from set $Z$. At the end of $7^{th}$ transition only one zone is left unvisited, which is selected as the last destination. The final $S$ and $D$ traces optimal transition pattern.

**Algorithm 1:** Algorithm to extract optimal gaze-track from constrained net transition matrix $TM_c$

Define $Neig(x)$ = neighbors of zone x;

Define $Z = 1:9$;

Define $[S]_{8\times1}$ and $[D]_{8\times1}$;

Initialize $(D(1), S(1)) = \text{argmax}_{(i,j)}\, TM_c$;
Remove D(1) and S(1) from Z;

**for** $k \in 2:6$ **do**
    $S(k) = D(k-1)$;

    **if** $\max_{i\in Z} TM_c(i, S(k)) > 0$ **then**
        $D(k) = \text{argmax}_{i\in Z}\, TM_c(i, S(k))$;
    **else**
        $Z_{Neig} = Neig(S(k)) \cap Z$;
        **if** $\sim isempty\{Z_{Neig}\}$ **then**
            $D(k) = \text{argmax}_{j\in Z_{Neig}}\{\max_{i\in Z}(TM_c(i,j))\}$;
        **else**
            $D(k) = \text{argmax}_{j\in Z}\{\max_{i\in Z}(TM_c(i,j))\}$;
        **end**
    **end**
    Remove $D(k)$ from $Z$;
**end**
$S(8) = D(7)$;
$D(8) = Z$;

Optimal transition pattern achieved with Algorithm 1 is shown in Figure 2.13(a). Transition pattern successfully captured human preference for the horizontal over vertical and top over bottom direction. Transitions are graceful (not abrupt) as conceived while neighborhood constraint is introduced. Due to similarities with human bias and gracefulness, optimal transition pattern is easy to follow by any practitioner. Transitions are spiraling out, visiting the zones near to macula first and then progressing towards peripheral zones. This allows the search to be efficient, fast and terminative in nature.

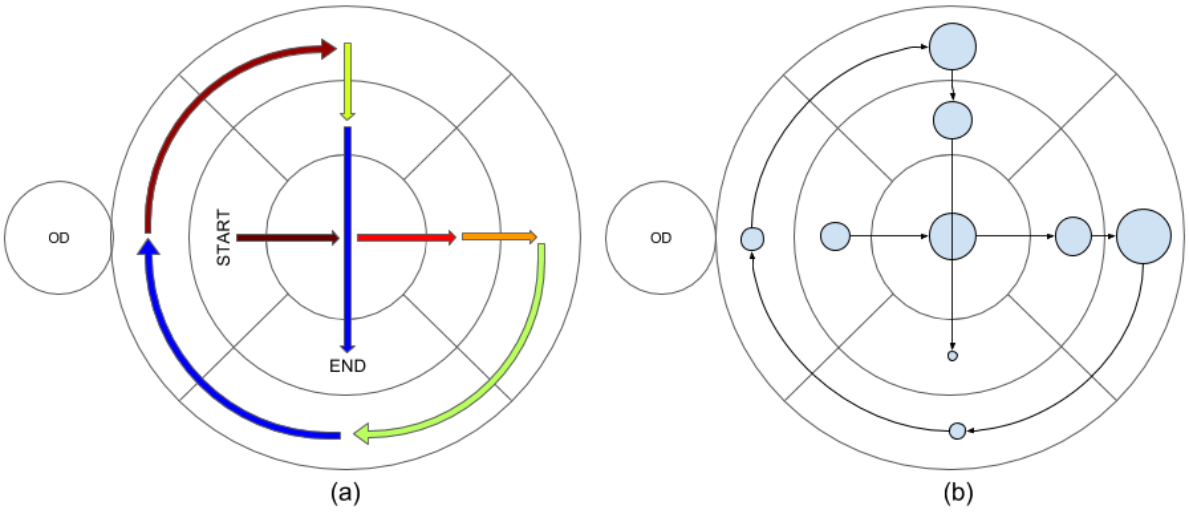



Figure 2.13: (a) Optimal transition pattern (b) Optimal gaze-track. Scanning strategy is guided by optimal transition pattern and dwelling by dwell map. The radius of circle represents the amount of required dwelling.

### 2.6.2 Optimal dwell map

As the task is to decide the category (normal or abnormal) as soon as possible, the visual search for abnormal images is terminative in nature. i.e. once a lesion is located in certain zone, response was made without scrutinizing other zones (see Figure 2.12(a),(b)). In the absence of lesions, visual search for correctly classified normal cases is more exhaustive in nature (see Figure 2.12(c),(d)). This is because of the cautiousness to avoid false negatives. In order to achieve the dwell map which does not depend on the location of lesions, trials for normal cases where medically trained participants registered correct responses are considered. Novices are ignored to avoid false practices. Final dwell map(see Figure2.9(b)) is computed as follows.

$$d_k = \frac{1}{ij} \sum_{j \in N, i \in S_j} d_{i,j,k} \tag{2.8}$$

Where, $S_j$ is a set of medical experts who have given correct response for $j^{th}$ image. Figure2.9(b) shows that dwelling is more in (a) superior zones than inferior zones and (ii) temporal zones than nasal zones. Fovea/macula is the central dark zone which is responsible for color vision. Presence of hard exudate in this zone can cause impairments in color vision. So fovea is scrutinized for longer duration. Parafovea zones are dwelt on less than perifovea zones. This can be due

to two reasons. First, parafovea zones are near to macula and have dark background which provides higher contrast for hard exudates and as a result abnormalities pop out prominently. This phenomenon requires less dwell duration to identify lesions. Second, parafovea zones have smaller area than perifovea zones. So number of fixations in parafovea zones is less and hence total dwell duration is less. Optimal transition pattern along with optimal dwell map generates an optimal gaze-track as shown in Figure 2.13(b). This search pattern is recommended to image reader for fast and accurate diagnosis of DR.

## 2.7 Conclusion

An eye tracking study has been conducted with the aim to understand image reading strategies which lead to good performance in diagnosing retinal images. The analysis of eye-tracking data indicates that readers use mainly two kinds of strategies while reviewing retinal images, namely *tracing* and *dwelling*. *Tracing* involves low dwell time for a particular region and leads to a reader reviewing the region over multiple revisits. As a net effect, *Tracing* is marked by long gaze-track and slow response. On the other hand, *dwelling* strategy involves longer dwell time for a region which is better suited to gather adequate information for decision making. This is affirmed by the fact that subjects employing dwelling strategy have higher accuracy and faster response time. Analysis of dwell-map and transition matrix suggests that scanning strategy used by medically trained practitioner agrees well, but novices(with no medical training) have different search pattern. An optimal search pattern is extracted from eye tracking data of medically trained participants. Optimal search pattern takes various human factors into account, which makes it efficient and easy to follow. We recommend image readers in reading-centers to use this search pattern for fast and accurate diagnosis.

# Chapter 3

## Assistive Lesion Emphasis System

### *An Unsupervised Approach*

> "Everyone believes in the law of errors [Gaussian]: the mathematicians, because they think it is an experimental fact; and the experimenters, because they suppose it is a theorem of mathematics."
>
> – *Gabriel Lippmann*

Image reading is a tedious task, yet requires high accuracy. Tedium of reader can lead to incorrect diagnosis. To address this problem, computerized diagnosis has been explored. But, low accuracy of computerized diagnosis limits its employability in reading centers. Hence, as mentioned in Chapter 1, we advocate a reader-centric approach where computerized diagnosis works as an *assistant* to a reader. We propose a design where Computer Assisted Diagnosis(CAD) detects salient regions and allows a reader to selectively highlight them. A reader can make decision based on highlighted regions. In this chapter, above mentioned scheme is implemented to design a CAD for diagnosis of DR. Developed CAD is a two stage system: computation of saliency and highlighting of salient regions.

Existing computational saliency models have been developed for general (natural) images and hence may not be suitable for medical images. This is due to the variety of imaging modalities and the requirement of the models to capture not only normal but also deviations from normal anatomy. We present a biologically inspired model for colour fundus images. The proposed model uses *spatially-varying* morphological operations to amplify lesions locally and combines an ensemble of resultant maps to generate the saliency map. Detected salient regions can be

highlighted in many ways. We propose an elegant approach for highlighting, named interactive selective enhancement(ISE). Computed saliency maps are mixed/fused with original image at multiple scales to selectively enhance only salient regions without altering the background. A reader can vary the amount of mixing interactively and can control degree of emphasis. Saliency model and ISE collectively make 'Assistive Lesion Emphasis System(ALES)'.

ALES is validated stage-wise. The saliency model is validated against an average Human Gaze map of 15 experts and found to have 10% higher recall (at 100% precision) than four leading saliency models proposed for natural images. The F-score for match with manual lesion markings by 5 experts was 0.4 (as opposed to 0.532 for gaze map) for our model and very poor for existing models. ISE was found to boost contrast to noise ratio of a lesion by $\sim 30\%$.

## 3.1 Saliency Computation

### 3.1.1 Background

Human attention is attracted by most prominent or visually salient objects in a scene. Saliency of an object is modulated by the task at hand [38] and we selectively attend to most informative regions of visual field while ignoring some unimportant regions. Much effort has been made to understand task-specific visual attention and perception, resulting in cognitive modeling of visual saliency [39]. Computational modeling of the same has been of interest to computer vision community for years [40] leading to their use in segmentation [41], object recognition [42], retrieval [43], image/video compression [44] and context aware image editing [45]. Many such applications, including ALES, are of interest in medical domain as well.

Medical experts look for specific type of abnormalities often at specific locations, ignoring irrelevant areas and artifacts, while reviewing images for diagnosis. Deriving saliency models for medical images is a difficult task due to the variability in modalities, anatomy, type of diseases and artifacts. Very little work has been done to devise computational models of saliency for medical images [46, 47, 48]. Saliency has been used in x-ray image classification[49], segmentation and registration of MRI [50, 51] and anatomical plane classification from fetal ultrasound [52].

Computational models of visual attention are broadly classified into 2 groups: bottom-up and top-down [40]. Bottom-up models are stimuli driven, whereas top-down models are intention, task or goal driven and based on prior knowledge. Depending on the medical modality, general bottom-up models may or may not be successful in explaining how experts review images. Existing top-down models for general images are inappropriate as the tasks are different. Saliency models need to be designed for specific modalities and type of lesions. We propose a saliency model for diabetic retinopthay analysis from colour fundus images of eyes. Our model handles normal and abnormal images with bright lesions (hard exudates, cotton wool spots) and dark lesions (hemorrhages, microaneurysms).

Taxonomy of computational saliency models includes ones which are biologically plausible [15] or based on spectral-analysis [53], information- and decision-theory [54], pattern classification [55], graphs [56], etc. The proposed model is biologically motivated and is based on morphological processing. The model has been evaluated against *gaze maps* of retina specialists as well as manual lesion markings and compared with four existing bottom up saliency models developed for natural images.

### 3.1.2 Method

#### Motivation

Human fixations follow a Gaussian distribution [57]. Visual information near a gaze-point is attended to more than those that are away. In terms of visual processing, this implies that information in a region proximal to a gaze-point is given higher importance than that are not. The inspection and understanding of the entire visual scene is done gradually as we move our eyes. For example, given a 'Where is Waldo?' puzzle, we first fixate on one specific cartoon. This creates an image on retina such that projection of fixated cartoon lies on macula, a central region of retina which has densely packed photo-sensors(cones). As a result, the fixated cartoon is perceived with higher resolution than other regions of the puzzle. Then we gradually scan other parts of puzzle by fixating on different cartoons one after another. Every time we fixate on a cartoon, we analyze it for being a possible 'Waldo'(see Figure 3.1).

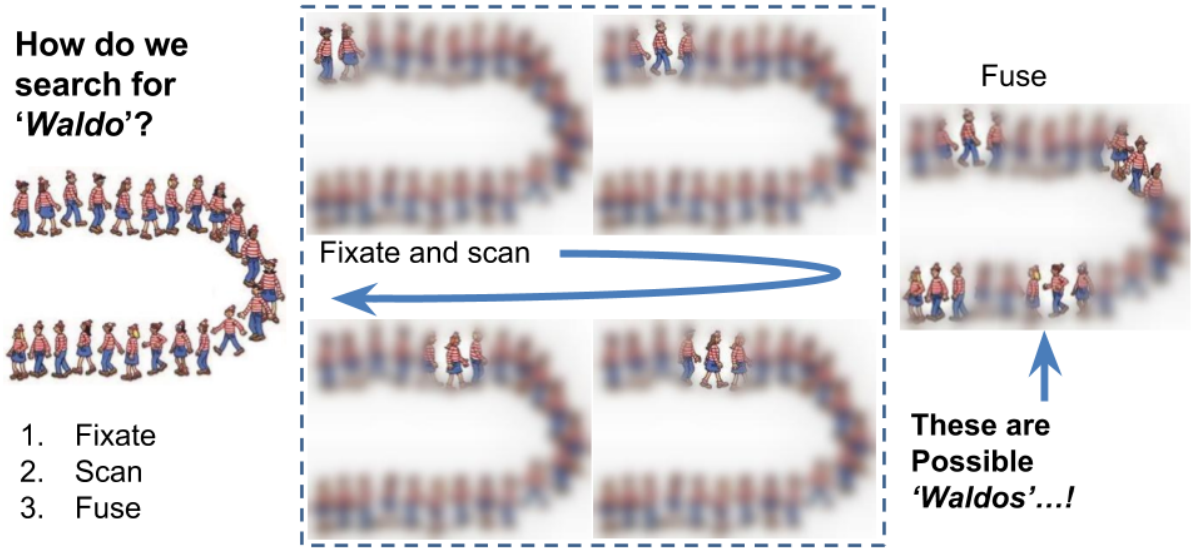



Figure 3.1: Visual scene analysis. [right image: 'The Waldo Watchers', source: http://waldo.wikia.com]

The whole process of visual scene perception can be divided into three processing blocks. First is a Gaussian shaped fixation which projects visual scene on retina with variable resolution. Second is sensing of projection image by retinal ganglion cells. Third is spatio-topic combination of projections sensed after multiple fixations. Use this concept, we propose an approach to saliency computation for a specific task, namely to detect abnormalities (bright and dark lesions) in a given image. Here, fixations are mimicked by *spatially-varying* Gaussian-fashioned morphological operations, Ganglion cells by center-surround filters and spatio-topic fusion by non-linear combination of mean and variance maps.

The processing pipeline has 3 stages: (1) Preprocessing (2) Generation of an ensemble of pre-saliency maps (3) ensemble integration to produce the final saliency map. The details are presented next.

## Preprocessing

Given a colour fundus image, all processing was restricted to the green channel. Illumination correction is performed using method shown in [58]. The vessel network and the optic-disk were detected and inpainted [59]. The fundus was extended to cover the mask region [60].

### Pre-saliency map generation

Let us consider a gaze-point $p$ in the image. Our strategy is to boost the prominence of lesions or abnormalities in the image based on their proximity to $p$. Specifically, boosting is higher for proximal as opposed to distant lesions. Boosting is achieved using spatially varying morphological processing.

Given an image $I$ and a gaze-point $p = (a, b)$, the origin is shifted to $p$. Resulting image $I_s(x, y) = I(x - a, y - b)$ which is denoted as $I_s(r, \theta)$ in polar coordinates. For any given $\theta \in (-\pi, \pi]$, we denote the 1-D image $I_\theta(r) = I_s(r, \theta)$. Consider a structuring element,

$$b(\rho) = \begin{cases} 0, & \rho \in D_b \\ -\infty, & \text{otherwise} \end{cases} \tag{3.1}$$

where, $D_b$ is domain of structuring element $b$. Grayscale dilation and erosion of $I_\theta$ with $b$ is expressed respectively as,

$$(I_\theta \oplus b)(r) = max\{I_\theta(r - \rho) \mid \rho \in D_b\} \tag{3.2}$$

$$(I_\theta \ominus b)(r) = min\{I_\theta(r + \rho) \mid \rho \in D_b\} \tag{3.3}$$

In order to introduce spatially-varying processing, the domain $D_b$ is made to vary with $r$ and hence the dilation/erosion is defined as follows,

$$D_b(r_n) \equiv [-f(r_n), f(r_n)] \tag{3.4}$$

$$(I_\theta \oplus b)(r_n) = max\{I_\theta(r_n - \rho) \mid \rho \in D_b(r_n)\} \tag{3.5}$$

$$(I_\theta \ominus b)(r_n) = min\{I_\theta(r_n + \rho) \mid \rho \in D_b(r_n)\} \tag{3.6}$$

We have chosen $f(r) = \lambda G_\sigma(r)$, where $G_\sigma(r)$ is a Gaussian distribution with zero mean and variance $\sigma^2$. $\lambda$ is a free parameter which controls the domain length. Eq. 3.5 and 3.6 can

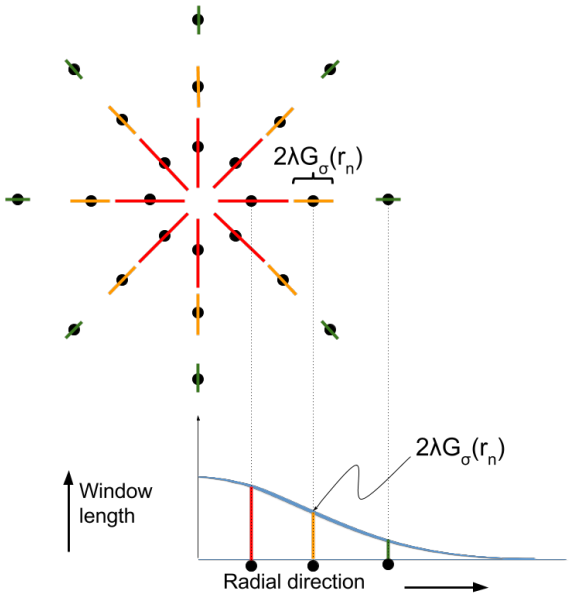



Figure 3.2: Window length at different points in the image.

also be interpreted as rank filtering with a filter/window whose length is a function of $r_n$. At any point $r_n$, a window centred at $r_n$ and of length $2\lambda G_\sigma(r_n)$ is used to do ranking operation. The above $1D$ operation (dilation/erosion) is performed on $I_\theta(r)$, $\forall\theta \in (-\pi, \pi]$. i.e. in each direction. Figure 3.2 shows windows of varying length along different directions about a gaze point $P : (r, \theta)$. Windows for $I_{\theta=0}(r)$ are shown to be derived from Gaussian distribution.

Processed images are shifted back to get the final pre-saliency (PS) maps: $\mathrm{PS}_q^i(x, y) = I_q(x + a_i, y + b_i)$ where $q$ represents dilation or erosion operation. Dilation (erosion) will boost the saliency of bright (dark) lesions.

We illustrate this idea with a phantom in Figure 3.3. Here, an image patch is modeled as gray-scale texture with idealized lesions appearing as dots of appropriate colour: white dot (hard exudate or HE), white blurred spot (cotton wool spot or CWS) and dark dot (hemorrhage or HM) (Figure 3.3a). Four gaze-points $p_i$ are selected on the three lesions and background (Figure 3.3b). The PS maps obtained with different $p_i$ are dramatically different as seen in Figure 3.3c-j. Overall, it can be seen that a lesion is spatially extended while the background remains unchanged, in the PS map. It is noteworthy that when $p$ is on the background (shown

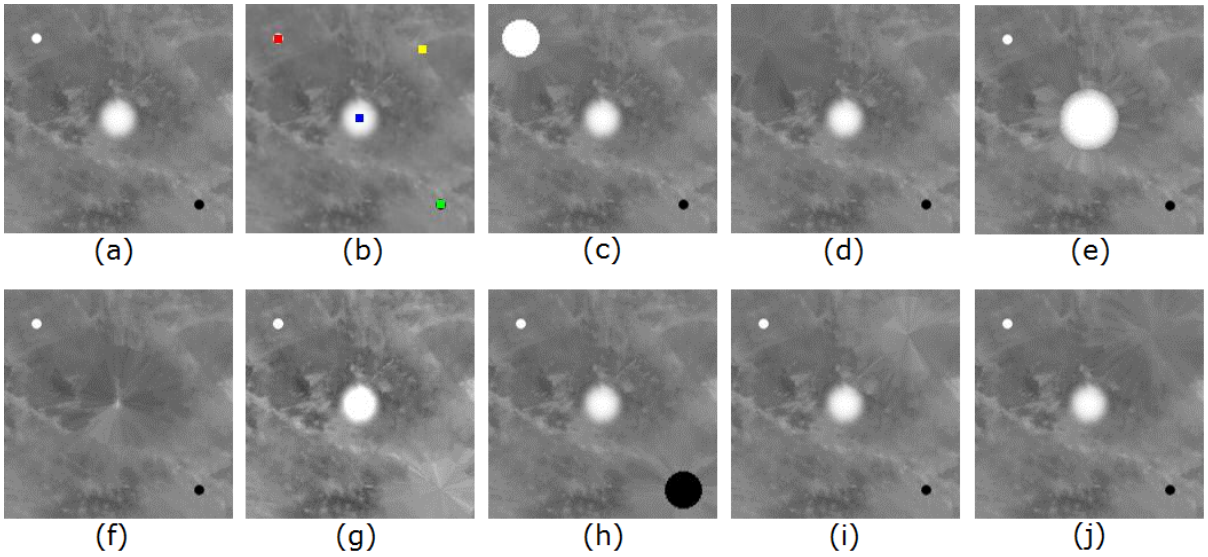



Figure 3.3: Pre-saliency maps for a phantom image (a) at four gaze-points considered (b). The maps with dilation - erosion for the red, blue, green and yellow gaze-points are shown in (c)-(d),(e)-(f),(g)-(h) and (i)-(j) respectively.

in yellow), both dilation and erosion has no effect since the lesions are not proximal enough to $p$ in any direction. Hence, the original image and PS are almost identical in appearance. Similar behaviour is shown in real images next.

A sample fundus image patch with a HE and HM is shown in Figure 3.4 along with the PS maps for dilation/erosion derived with three $p_j$ (shown in green, blue and black). In Figure 3.3 and 3.4, the $p_j$ were selected on true lesions deliberately to show the effect of the proposed spatially-varying morphological processing. In reality, the lesion locations are unknown. Hence, a set of $\{p_j \mid j = 1, 2, 3, ..., J\}$ at randomly selected locations are used to generate an ensemble of PS maps. Since both bright and dark lesions are of interest, both erosion and dilation are applied separately at each $p_j$, to obtain $2J$ PS maps. A judicious combination of these maps can help derive the desired saliency map, which is explained next.

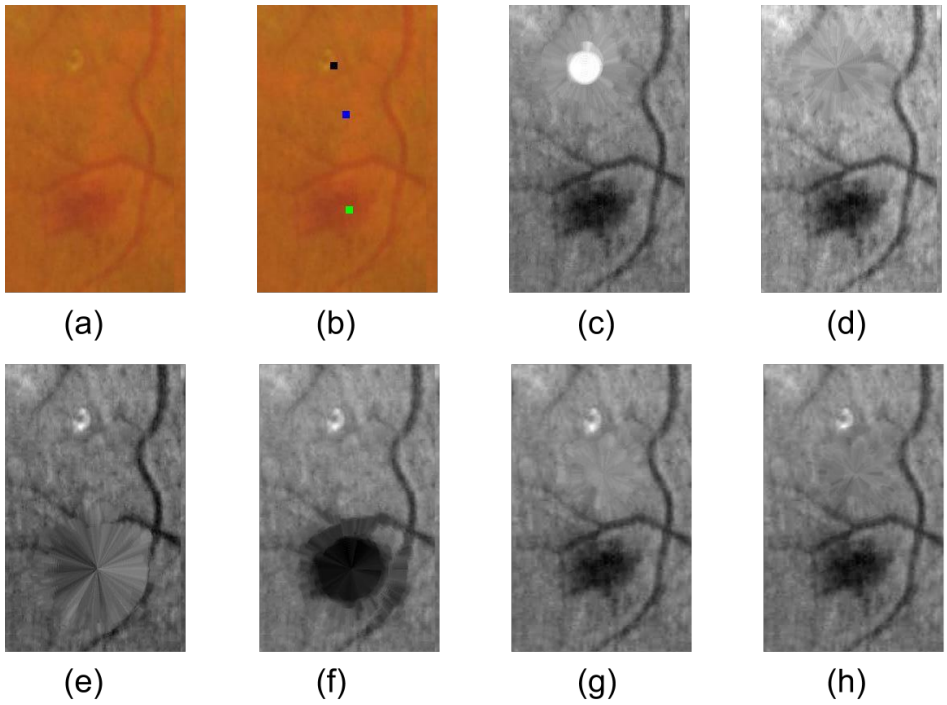



Figure 3.4: Pre-saliency maps for a fundus image patch (a) at three gaze-points (b). The maps with dilation - erosion for the black, green and blue gaze-points are shown in (c)-(d);(e)-(f) and (g)-(h) respectively.

**Integration**

In order to combine the $J$ PS maps, we follow the strategy used in [61]. The $J$ maps are summed to create a combined map $\mathrm{C}_q$. The variance at every pixel location is also computed to derive a variance map $\mathrm{V}_q$.

$$\mathrm{C}_q = \sum_{i=1}^{J} \mathrm{PS}_q^i \tag{3.7}$$

$$\mathrm{V}_q = \mathrm{Variance}(\mathrm{PS}_q^i); i \in [1, 2, 3, ..., J] \tag{3.8}$$

These two maps provide evidence for a location to be salient. An explicit Evidence map ($\mathrm{E}_q$) is computed by exponential weighting of $\mathrm{C}_q$ by $\mathrm{V}_q$ as,

$$\mathrm{E}_q = \mathrm{C}_q \times e^{(\tau \times \mathrm{V}_q)}, \tau \in \mathbb{R} \tag{3.9}$$

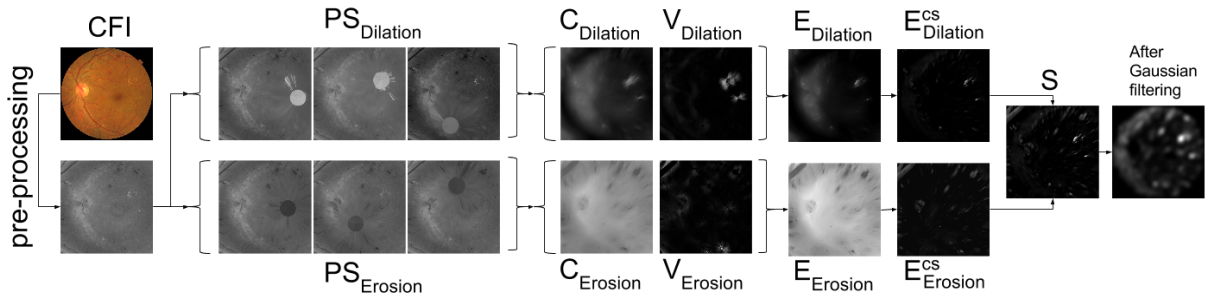



Figure 3.5: Proposed model for computing saliency with intermediate results.

Here, $\tau$ helps control the contribution of the variance at a pixel to the final evidence. A negative (positive) value of $\tau$ is chosen for bright (dark) lesion. Higher absolute value of $\tau$ boosts the saliency of even less prominent lesions. Separate $\mathrm{E}_q$s, one for dilation and other for erosion, are extracted.

#### Saliency map computation

A Center surround (CS) filter is applied to handle the variable conspicuity of lesions. This is applied to both $\mathrm{E}_q$ separately and they are combined to get the Saliency map ($S$) as,

$$S = max\{\mathrm{E}^{cs}_{dilation}, \mathrm{E}^{cs}_{erosion}\} \tag{3.10}$$

where, $\mathrm{E}^{cs}_q$ is CS filtered $\mathrm{E}_q$ and $max\{.\}$ is a pixel-wise max operation. $S$ is finally smoothed with a Gaussian filter. The proposed pipeline is shown in Figure 3.5 with all intermediate results.

## 3.2 Interactive Selective Enhancement

### 3.2.1 Background

In DR reading centres, readers scrutinize images and assign a DR stage to the image using the ETDRS standard(see Figure 3.6) [9]. 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

| Measure | Score | Observable Findings |
|---|---|---|
| **ICDR severity level** | | |
| No apparent retinopathy | 0 | No abnormalities (Level 10 ETDRS) |
| Mild non-proliferative diabetic retinopathy | 1 | Microaneurysm(s) only (Level 20 ETDRS) |
| Moderate non-proliferative diabetic retinopathy | 2 | More than just microaneurysm(s) but less than severe non-proliferative diabetic retinopathy (Level 35, 43, 47 ETDRS) |
| Severe non-proliferative diabetic retinopathy | 3 | Any of the following: > 20 intra-retinal haemorrhages in each of 4 quadrants, definite venous beading in ≥2 quadrants, prominent intra-retinal microvascular abnormalities in ≥1 quadrant, or no signs of proliferative retinopathy. (Level 53 ETDRS: 4-2-1 rule) |
| Proliferative diabetic retinopathy | 4 | One or more of the following: neovascularization and/or vitreous or preretinal haemorrhages. (Levels 61, 65, 71, 75, 81, 85 ETDRS) |
| **Macular oedema severity level** | | |
| No macular oedema | 0 | No exudates and no apparent thickening within 1 disc diameter from fovea |
| Macular oedema | 1 | Exudates or apparent thickening within 1 disc diameter from fovea |

Abbreviations: ETDRS, Early Treatment Diabetic Retinopathy study; ICDR, International Clinical Diabetic Retinopathy

doi:10.1371/journal.pone.0139148.t001

Figure 3.6: ETDRS guideline for grading diabetic retinopathy and diabetic macular edema.

Figure 3.7). 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.

Assistive Lesion Emphasis System(ALES) uses above computed saliency maps to enhance salient regions. In order to do this, we next consider the problem of selective enhancement of color fundus images. Existing enhancement techniques include single channel(mostly green channel) enhancement or color enhancement which uses global information. Such methods alter perceptual quality of an image [62](see Figure 3.8(b)). Such enhanced images may be useful for

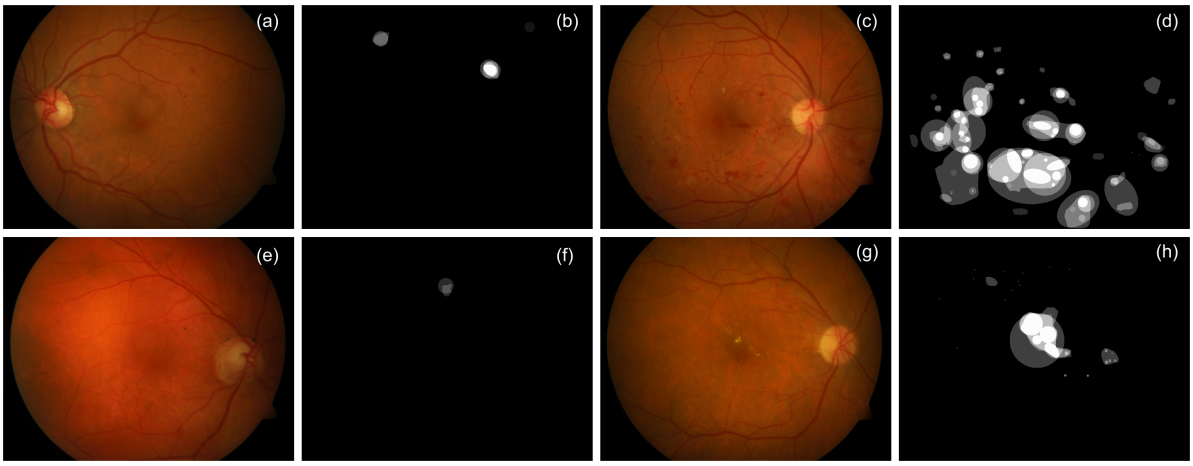



Figure 3.7: Original image and GT: (a,b) DR stage 2 (c,d) DR stage 3 (e,f) DME stage 2 (g,h) DME stage 3.

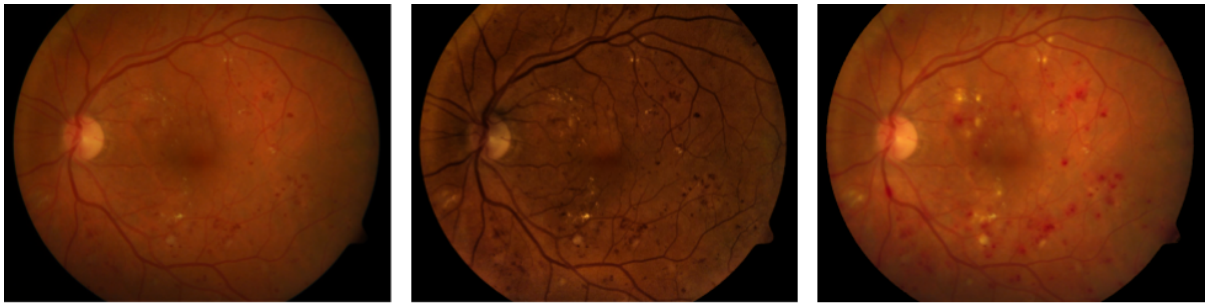



Figure 3.8: Fundus image enhancement (a) original image (b) global enhancement, which is not appropriate for ALES (c) selective enhancement, which is desirable for ALES.

computational purposes but unsuitable for visual presentation in real word scenario where a human image reader is diagnosing images. Fully-automatic enhancement also does not allow a reader to apply minor adjustment as is often desirable. For example, radiologists routinely vary the window-center and window-length to adjust brightness and contrast of CT images for better visibility [63].

We propose a semi-automatic solution with Interactive Selective Enhancement (ISE). ISE enhances salient regions locally, with minimum alteration to perceptual quality(see Figure 3.8(c)). ISE provides interactive control over degree of enhancement i.e. reader can vary parameters to improve results generated by default parameter settings and can observe changes in real-time.

ISE however, does not aim to correct non-uniform illumination or blur. Details are provided next.

### 3.2.2 Method

Given an image $I$ and evidence maps $E^{cs}_{dilation}$ and $E^{cs}_{erosion}$, enhanced image $X$ is computed over multiple scales and fused as follows,

$$X_k = \sum_i \Big\{ [(1-\alpha)I_k - \alpha((a_k\, E^{cs}_{erosion} \times I_k) * G_i)] + [(1-\beta)I_k + \beta((b_k\, E^{cs}_{dilation} \times I_k) * G_i)] \Big\} \quad (3.11)$$

Here, the first and second terms help enhance dark and bright lesions respectively. $k$ is index of color channel and $i$ is scale. $a$ and $b$ control color shade in the enhanced image. $\alpha$ and $\beta$ are mixing parameters which control the degree of enhancement. $\alpha$, $\beta$, $a$, $b$ are control parameters which can be varied by the reader. $E^{cs}_{erosion}$(or $E^{cs}_{dilation}$) masks out all the background and picks only dark (or bright) lesions from $I$, which is then mixed with $I$ over multiple scales. Negative sign in the first term effects a darkening of the salient regions. Similarly, positive sign in the second term has the effect of brightening the salient regions.

## 3.3 Results

Assistive Lesion Emphasis System(ALES) is validated stage-wise.

### 3.3.1 Evaluation of saliency model

The proposed Spatially-varying Erosion and Dilation (or SED) model for saliency was tested on images collected from a local eye hospital. Along with the images, manual marking (GT) of lesion regions was collected from 5 retina experts. The model was evaluated in two ways: i) against gaze maps(GM) derived from eye tracking and ii) GT. An eye tracking experiment was performed on 15 retina experts while they were reviewing the images exhaustively. To avoid human fatigue and hence low accuracy of eye tracking data, dataset size was limited to 10

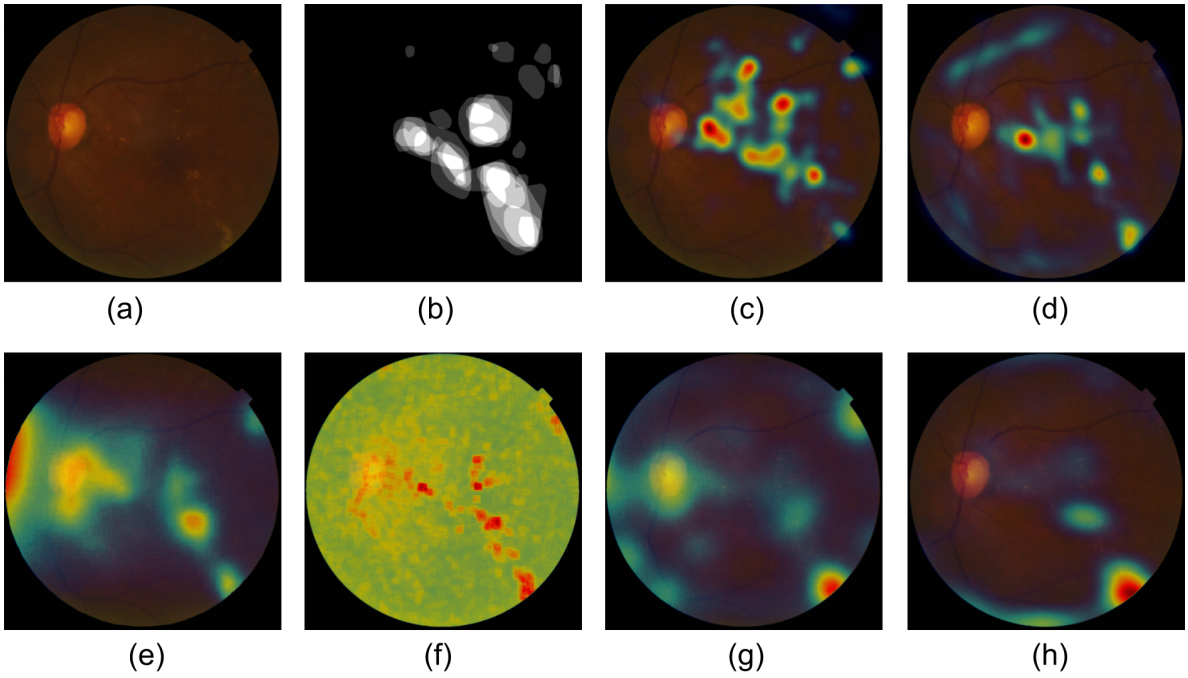



Figure 3.9: Comparison of saliency models for abnormal fundus image (a) Sample fundus image (b) Markings from 5 experts (c) GM map(d) SED (Our model) (e) GBVS (f) AIM (g) IK (h) SR

images, all containing DR lesions. The GM was collected separately for normal images. GMs of all experts were denoised and averaged to derive an average GM for each image. The proposed model was also benchmarked against 4 different bottom up models of saliency. The preprocessed images (section 3.1.2) were taken as input for all the saliency models for a fair comparison. The models taken for comparison were based on different approaches: SR [53], Itti-Koch (IK) [15], GBVS [56] and AIM [54]. Of these models, only the AIM model is based on learning (from patches of a large number of *natural* images).

A sample abnormal image and its softmap GT are shown in Figure 3.9a-b. The GM and the computed saliency maps are presented in Figure 3.9c-h. SED and AIM appear to be similar to both GT as well as GM, though there are some false positive saliency regions as well. The saliency maps from GBVS, IK and SR are sparse and do not have as much overlap with the GT and GM. The saliency maps were also computed for images without any abnormalities i.e.

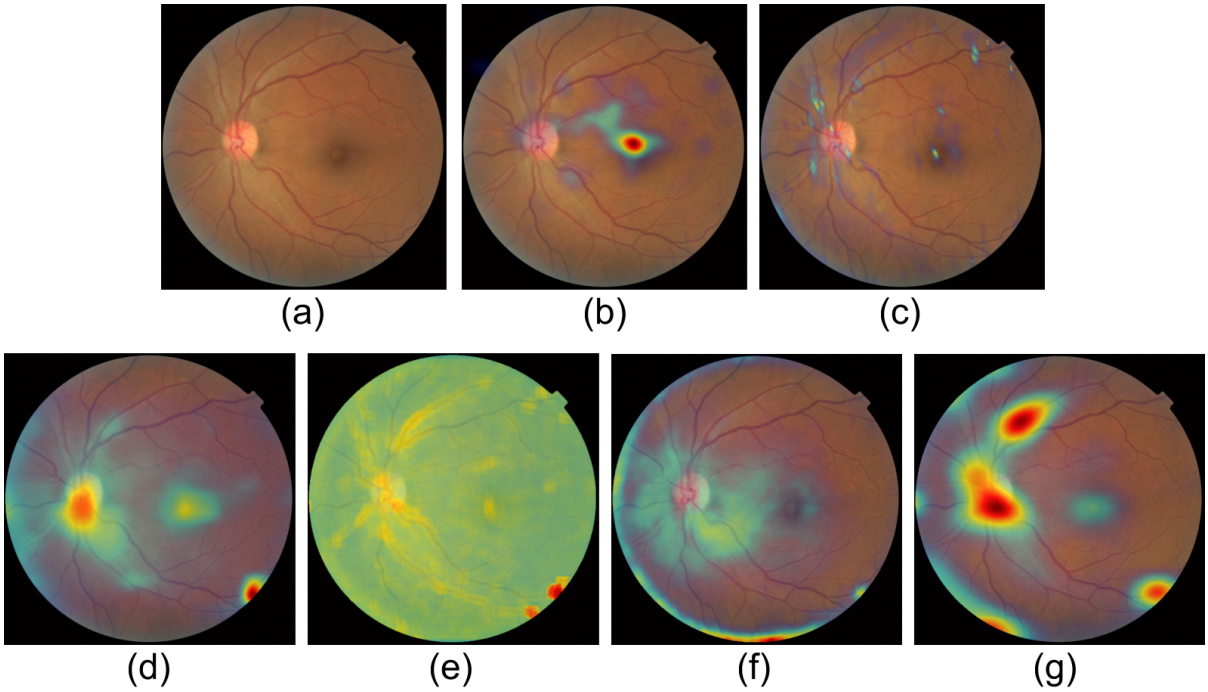



Figure 3.10: Comparison of saliency models for normal fundus image. (a) Original fundus image (b) Average gaze map (c) SED (Our model) (d) GBVS (e) AIM (f) IK (g) SR

*normal* images. These are shown for a sample image for all models in Figure 3.10. The GM for Figure 3.10a is shown in Figure 3.10b.

**Evaluation against Ground truth**

In the medical domain, unlike general vision, high level (or top down) knowledge is used during image scrutiny as the end goal is diagnosis. Hence, we evaluated all the models against GT. It is quite possible that both overt and covert attentions are used in determining if a region has abnormalities or not. This may result into occasional fixations on nom-lesion regions and no fixation on actual lesion regions. Hence, it is of interest to determine the degree of overlap between the GM and GT maps also. The GT softmap was thresholded at 50% agreement to generate a binary map. All saliency maps and GMs were thresholded from 0-90% of saliency in steps of 10% to generate the PR plot and F-score vs %saliency plot presented in Figure

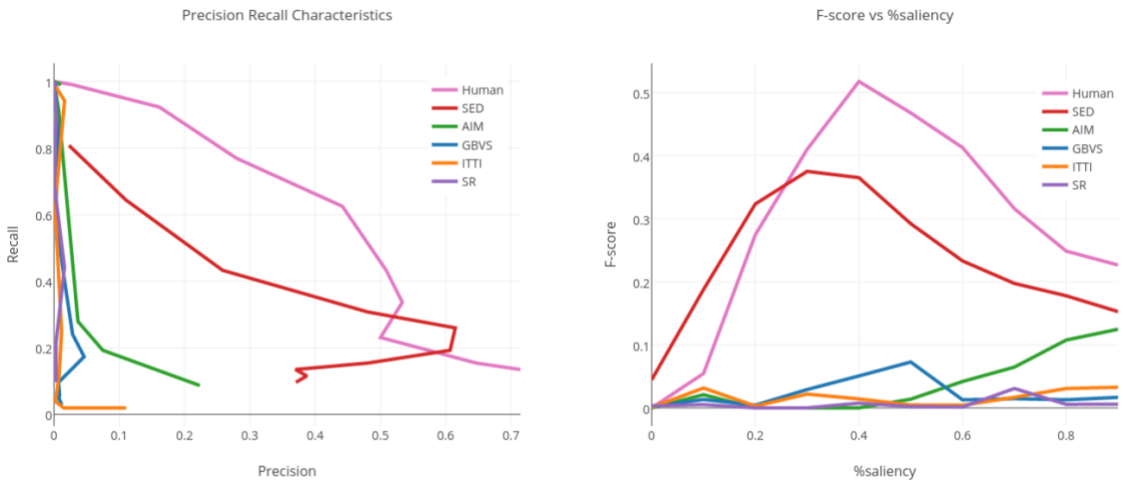



Figure 3.11: Comparative performance of various saliency models against ground truth (a) Precision-Recall characteristics (b) F-score for saliency models.

3.11. The plots reveal that SED and GM outperform all saliency models. This underscores the fact that a saliency model designed for natural images is not appropriate for medical images especially in helping to detect abnormalities.

### Evaluation against Gaze-map

We also compared the saliency maps with GMs. Precision ($P$) and recall ($R$) were computed by thresholding $S$ in the 0-90% range in steps of 10%. The resultant plots are presented in Figure 3.12(a). It can be observed that for $P < 0.2$ the $R$ values of existing models show an increasing trend whereas for $P > 0.2$ the trend is a decreasing one. In contrast, SED shows an increasing trend for all $R \in [0, 1]$. At $P = 1$, SED outperforms the existing methods by 10%. This relative improvement is also seen in F-score which is presented in Figure 3.12(b).

Figure 3.10(b) reveals that experts scrutinize even normal cases with a set of fixations before making a diagnosis. The GM also indicates the macula to be the only region with significant foveation. This is to be expected as macula is responsible for sharp colour vision and is hence a danger zone; any presence of abnormality here calls for swift intervention. Although GM for

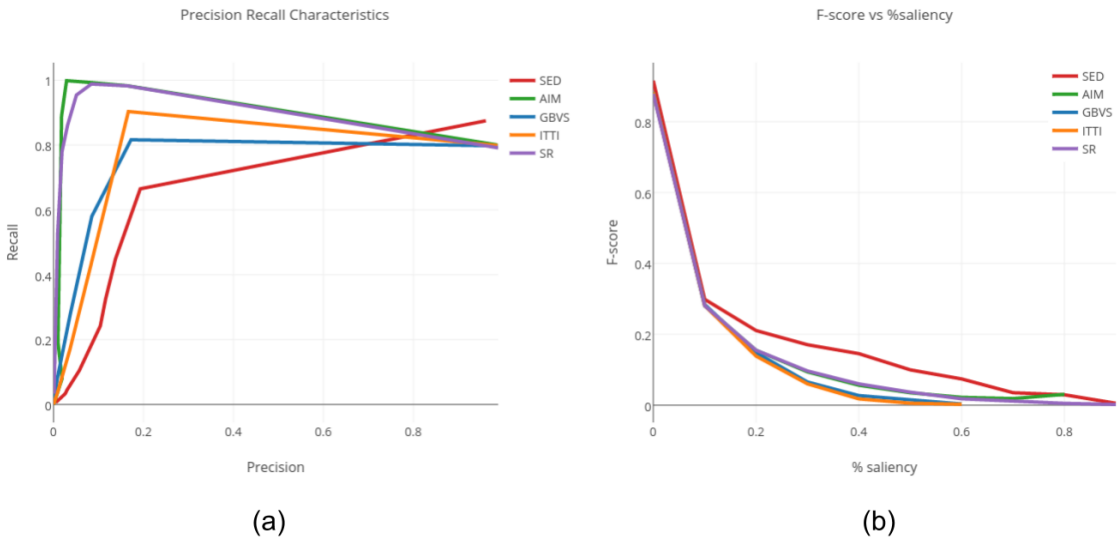



Figure 3.12: Comparative performance of various saliency models against gaze maps (a) Precision-Recall characteristics (b) F-score for saliency models.

normal images has greater than zero saliency, it is expected that computed saliency map has zero value. It can be seen that only SED is able to reject normal regions and generate a very sparse saliency map as desirable. The scrutiny of normal images is guided by more complex knowledge about the normal anatomy, danger zones and an expert's scrutiny style.

### 3.3.2 Evaluation of Interactive Selective Enhancement

The publicly available DiaretDB1[64] dataset is chosen for validation as it contains both bright and dark lesions and also provides lesion-level GT. We use 47 abnormal images out of given 89 images for the evaluation. A set of experiments are reported here for different settings of mixing parameters. $a$ and $b$ are fixed for all the experiments. First, mixing parameter $\alpha$ and $\beta$ are varied one at a time while the fixed parameter is set to zero. Varying values of $\alpha$ with $\beta = 0$ results in enhancement of only dark lesions. Similarly, varying values of $\beta$ with $\alpha = 0$ results in enhancement of only bright lesions. Figures 3.13 and 3.14 show qualitative results of IES for dark and bright lesions respectively. An excessive enhancement of dark (bright) lesions is seen to relatively brighten (darken) the background. If, both $\alpha$ and $\beta$ are varied simultaneously, this

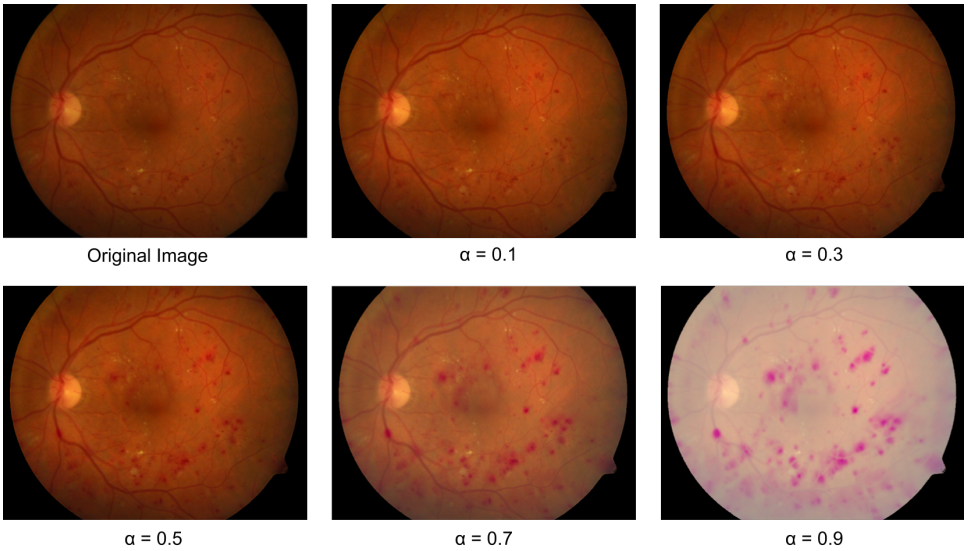



Figure 3.13: Selective enhancement of dark lesions for different $\alpha$ ($\beta = 0$).

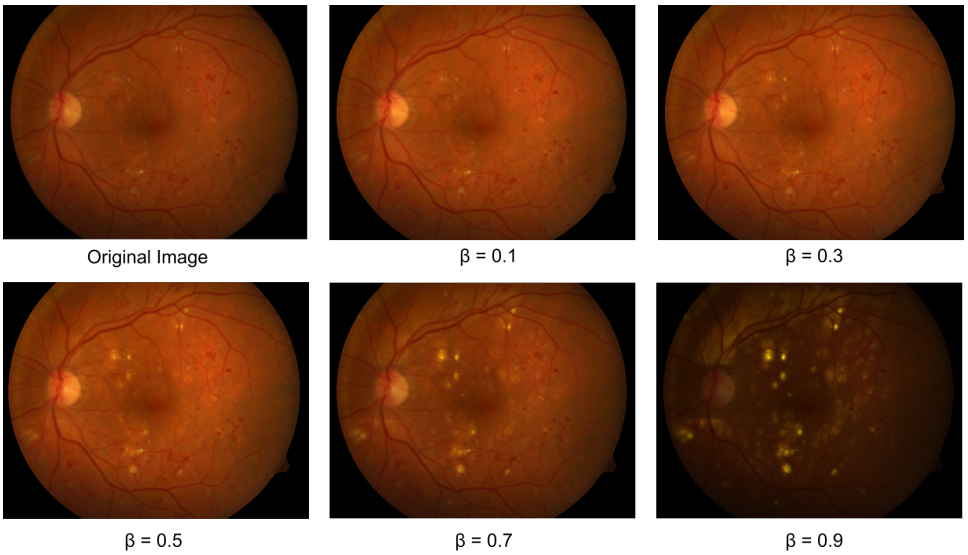



Figure 3.14: Selective enhancement of bright lesions for different $\beta$ ($\alpha = 0$).

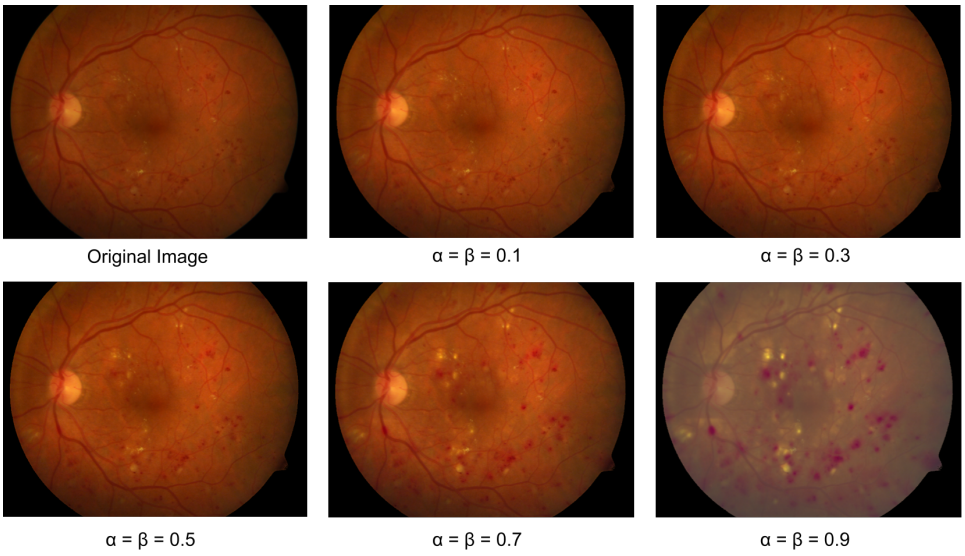



Figure 3.15: Balanced enhancement of both bright and dark lesions for varying values of $\alpha$ and $\beta$.

results in a balanced enhancement of both dark and bright lesions. Figure 3.15 shows this on a sample image.

A quantitative evaluation of IES was performed by computing the contrast to noise ratio (CNR) for varying values of mixing parameters. CNR is defined as

$$CNR = \frac{|m_f - m_b|}{\sigma_b} \tag{3.12}$$

where, $m_f$ and $m_b$ are mean intensity of foreground (lesions in our case) and background respectively. $\sigma_b$ is the standard deviation of background intensity. Figure 3.16 shows the relation between CNR and mixing parameter which is strictly increasing. Varying mixing parameter from 0 to 1 increases CNR by $\sim$ 30% for both dark and bright lesions (from 4.04 to 5.8 for bright lesions and from 3.36 to 4.75 for dark lesions).

Figures 3.13, 3.14 and 3.15 suggest that certain range of $\alpha$ and $\beta$ give the best perceptual quality. It is possible to dynamically compute the optimum parameters of ISE for a given image and use it as default settings subject to minor adjustments by reader.

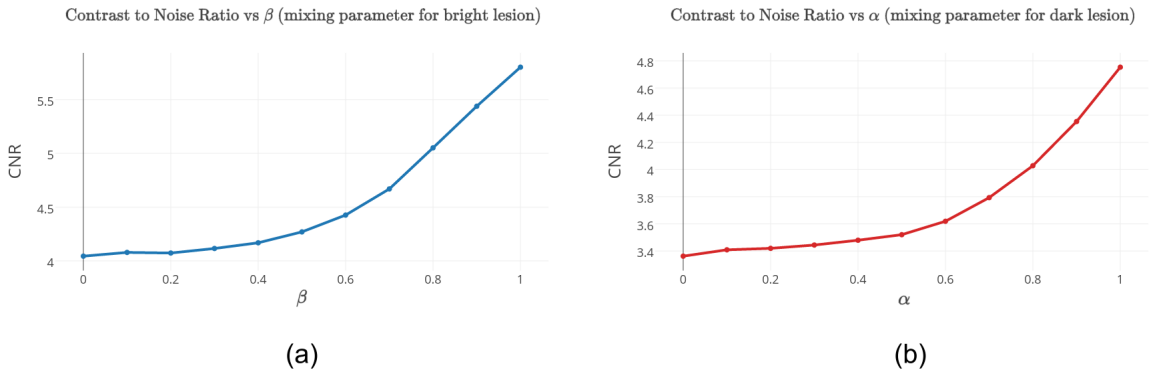



Figure 3.16: Contrast to noise ratio as a function of mixing parameter: (a) dark lesions (b) bright lesions.

## 3.4 Conclusion

We proposed a novel approach where computerized diagnosis assists human reader to reduce fatigue while DR image analysis. We named it Assistive Lesion Emphasis System(ALES) and is developed as two-stage system: saliency computation and selective emphasis. Saliency computation is done with novel spatially-varying erosion & dilation(SED) and lesion emphasis is done using interactive selective enhancement(ISE). ALES offers an efficient and highly interactive environment for image readers in reading centres.

Saliency computation has not received much attention in medical imaging. We presented a saliency model which is biologically inspired and is based on combining results of spatially-varying morphological processing centred at randomly chosen gaze points. The poor performance of saliency models such AIM, GBVS etc., in contrast to the good performance of our model in predicting gaze points and identifying abnormal regions, underscore the point that generic saliency models are unsuitable for the medical domain. As opposed to the strategy we used where the gaze points were chosen randomly, a *guided* spatial arrangement of gaze-points would be an interesting variant worth exploring. We also developed interactive and selective enhancement of regions of abnormalities. The results shows that contrast of lesions can be boosted significantly with increasing values of mixing parameters. For maximum perceptual quality, optimization of mixing parameters can be explored.

# Chapter 4

## Assistive Lesion Emphasis System

### *A Supervised Approach*

> "In other words then, if a machine is expected to be infallible, it cannot also be intelligent. There are several mathematical theorems which say almost exactly that. But these theorems say nothing about how much intelligence may be displayed if a machine makes no pretense at infallibility."
>
> – *Alan Turing*

Computer Assisted Diagnosis (CAD) tools are of interest as they enable efficient decision making in reading-centers for screening of diabetic retinopathy. We proposed a novel, reader-centric design named 'assistive lesion emphasis system (ALES)' in Chapter 1 and discussed an unsupervised approach in chapter 3. ALES draws reader's attention to abnormal regions in a least-obtrusive yet effective manner, using saliency-based emphasis of abnormalities *and* without altering appearance of background tissues. In this chapter we demonstrate yet another approach for development of ALES. Here, lesion-saliency is *learnt* in supervised manner using a convolutional neural network (CNN), inspired by the saliency model of Itti and Koch [15]. Training of CNN is done using novel loss function which is designed to handle range mismatch between intensity values of output saliency and ground truth. Network is designed and trained to fine-tune standard low-level filters and learn new high-level filters for deriving a lesion-saliency map. Computed saliency map is then used to perform lesion-emphasis via a spatially-variant version of gamma correction. This is done by defining parameter gamma as function of saliency. ALES is validated next in stage-wise fashion, similar to previous chapter. Proposed saliency

model has been evaluated on public datasets and benchmarked against other saliency models. Our saliency model was found to outperform other saliency models by 6 to 30% and lesion-emphasis technique was found to boost contrast to noise ratio of lesions by more than 30%. At the end of chapter we describe a perception study conducted in order to evaluate effectiveness of ALES in reading-center-like setting. Results of a perception study proves that ALES is an effective assistive tool for readers.

## 4.1 Saliency Computation

### 4.1.1 Background

Computational modeling of visual saliency has been a subject of research for long. Existing computational models range from biologically plausible ones [15, 65] to information- and decision-theoretic [66, 54], graphical [67, 56], spectral-analysis [41, 53], pattern classification based [55, 68, 69], etc. Out of these models, most of information/decision-theoretic and pattern classification based models are supervised in nature. However they are 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 [49], plane identification from 3D ultrasound [52], registration of dynamic renal MR images [51], prostate MRI segmentation [50] and saliency modeling for Glioblastoma multiforme tumor [47].

Our interest lies in developing supervised saliency model for DR lesions which can be used for lesion-emphasis in ALES. Learning saliency for DR images is a challenging task due to artifacts and non-uniform illumination (see Figure 1.3 in Chapter 1). A good saliency model has to *learn* discriminate artifacts from true lesions and reject the former. Supervised saliency model has been reported for only hard exudate [48]. 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 [15]. In Itti-Koch's saliency 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.

### 4.1.2 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 [70, 71]. 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 [72]. 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 [73].

#### CNN Architecture

The architecture of the proposed model has five stages as shown in Figure 4.1. 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 normal-

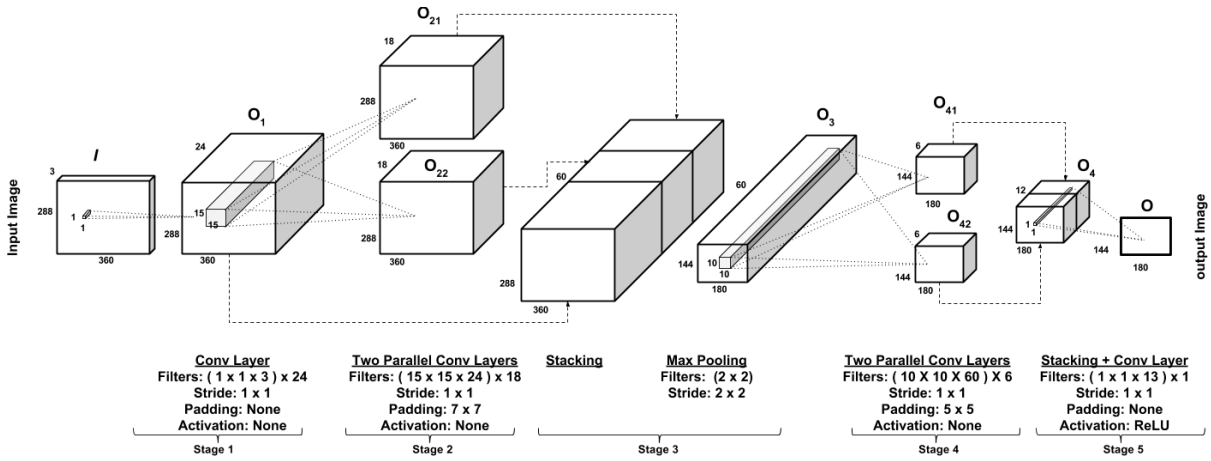



Figure 4.1: Proposed architecture with stage-wise description of types of layers, filter size, padding size, stride and activation function.

ization and combination of all the maps. Parallel fine-tunning of Itti-Koch filters and learning of new filters is carried out at 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.

$$O_1 = W_{color} * I + b_{color} \tag{4.1}$$

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.

$$O_{21} = W_{orient} * O_1 + b_{orient} \tag{4.2}$$

$$O_{22} = W_{rnd1} * O_1 + b_{rnd1} \tag{4.3}$$

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

$$O_3 = maxpooling([O_1, O_{21}, O_{22}]) \tag{4.4}$$

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:

$$CS1 = \pm(G_1 - G_4) \tag{4.5}$$

$$CS2 = \pm(G_2 - G_5) \tag{4.6}$$

$$CS3 = \pm(G_3 - G_6) \tag{4.7}$$

$$CS4 = \pm(G_1 - G_5) \tag{4.8}$$

$$CS5 = \pm(G_2 - G_6) \tag{4.9}$$

$$CS6 = \pm(G_3 - G_7) \tag{4.10}$$

where $G_n$ is a Gaussian filter with mean 0 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.

$$O_{41} = W_{CS} * O_3 + b_{CS} \tag{4.11}$$

$$O_{42} = W_{rnd2} * O_3 + b_{rnd2} \tag{4.12}$$

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 $(144 \times 180)$.

$$O_4 = [O_{41}, O_{42}] \tag{4.13}$$

$$O_5 = W_{combination} * O_4 + b_{combination} \tag{4.14}$$

$$O = \max(0, O_5) \tag{4.15}$$

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.

**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.

$$L(X, Y) = \frac{1}{N} \sum_{x \in X, y \in Y} \beta x e^{-\alpha y} + (1 - x)(1 - e^{-\alpha y}) \tag{4.16}$$

Here, the tuple ($x$, $y$) denotes the (GT saliency, output) pixel pair; $N$ is the total number of pixels; $\beta$ is a weight used to handle class imbalance. $\alpha$ 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 Figure 4.2).

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).

$$y_0 = \frac{1}{\alpha} \log(\beta + 1) \tag{4.17}$$

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 $\alpha$.

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.

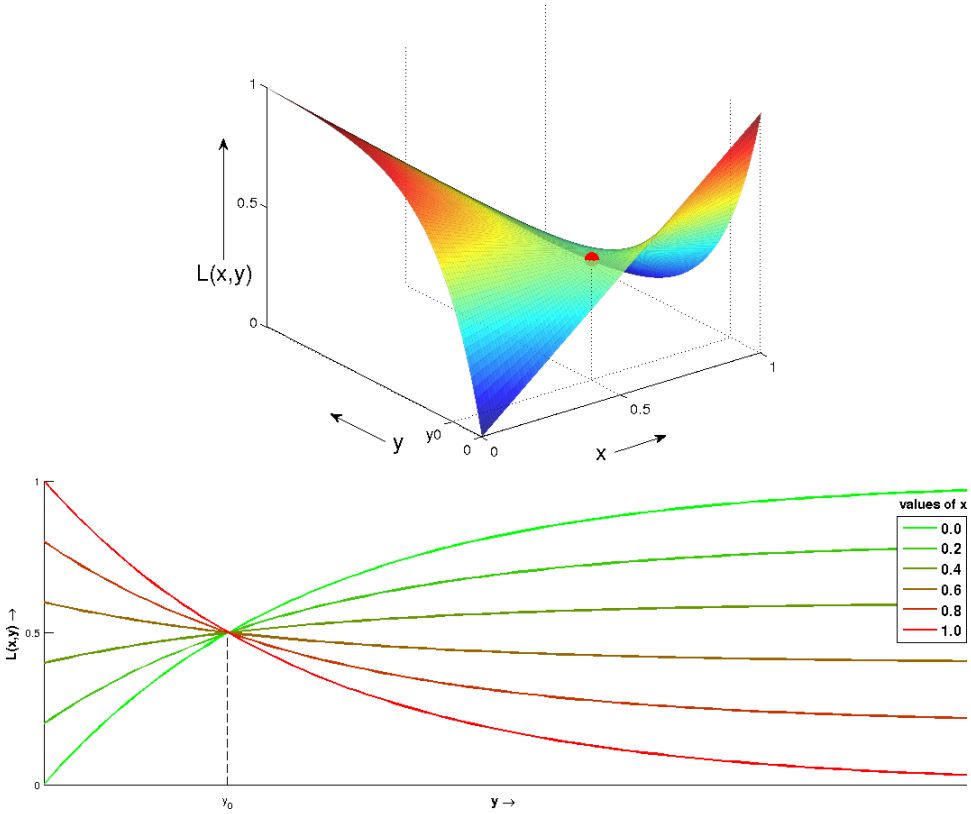



Figure 4.2: Loss function in the absence of class-imbalance ($\beta = 1$). (top) 3D view of a loss function. Saddle point is shown in red color. (bottom) 2D view of a loss function. Above surface is sampled at different values of $x$. y-projection of saddle point $(0.5, y_0)$ is also shown.

## 4.2 Lesion-Emphasis

### 4.2.1 Background

Readers in DR reader-centers analyze fundus images by following the guidelines provided by ETDRS [9]. The staging is done based on location of hard exudate and number of hemorrhages as described in section 4.2. Hence, ALES is designed to apply local enhancement by increasing contrast of lesions, which can successfully draw readers' attention and reduce diagnostic error.

Existing work on enhancement of retinal images are based on illumination/contrast correction [74, 75, 76, 77, 78], contourlet [79] and histogram equalization and matching [80, 81]. 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 [18], luminance/chrominance adjustment based on saliency [17], iterative addition of point variation values [82], de-emphasis of background texture [83] and saliency weighted luminance correction [84]. 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.

### 4.2.2 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,

$$I_C(x,y) = I_O(x,y)^{\gamma} \tag{4.18}$$

Here $I_C(x,y)$ and $I_O(x,y)$ (Normalized between 0 and 1) are corresponding pixels from corrected and original images respectively. $\gamma$ 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 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 Figure 4.3. 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.

$$I_C(x,y) = I_O(x,y)^{\left(1-\frac{S_{HE}(x,y)}{a}+\frac{S_{HM}(x,y)}{b}\right)} \tag{4.19}$$

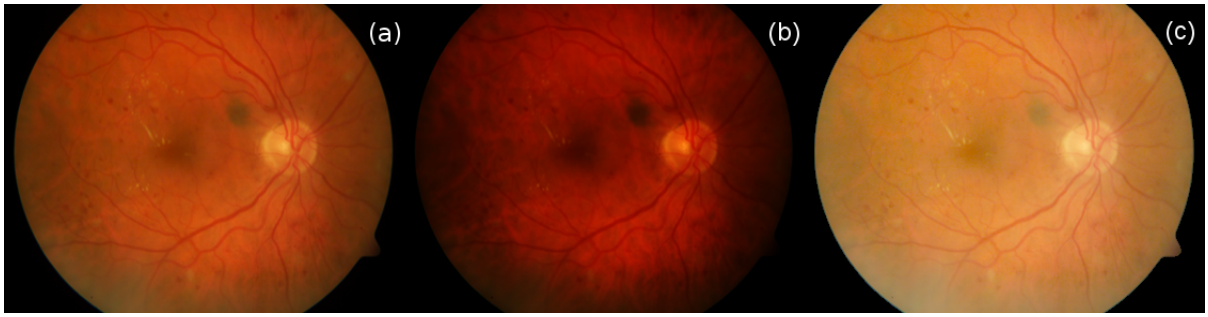



Figure 4.3: Gamma correction. (a) original image (b) corrected image with $\gamma = 2$ (d) corrected image with $\gamma = 0.5$.

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.

## 4.3 Material

The publicly available DIARETDB1 [64], DRiDB [85] and DMED [86] datasets were used for training and testing the proposed saliency computation stage. Pre-processing consisted of illumination correction [58]; fundus extension to remove the black mask region [60]; detection of vessels [59] and Optic-disk using circular Hough transform. The last two were subsequently inpainted to reduce false detections. Figure 4.6(b) shows the result of preprocessing on a sample image Figure 4.6(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 Figure4.4) and the latter providing

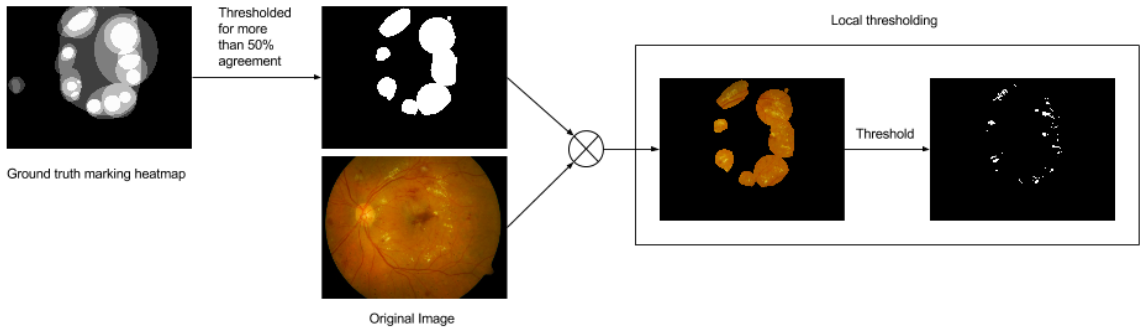



Figure 4.4: Procedure to obtain lesion-level ground truth from regional marking.

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 Figure4.4). 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 Figure4.4); this was finally thresholded to derive the final binary, lesion-level GT (rightmost image in Figure4.4). In order to retain hard exudates (hemorrhages) in the final GT, we retain pixels above (below) a threshold. This GT was downsampled to $144 \times 180$ 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 4.1 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 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. 4.16 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 4.2. Training time was approximately 5 days. Cross-validation was not performed due to excessive training time.

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

Table 4.1: Dataset description.

| Parameters | $S_{HE}$ | $S_{HM}$ |
|---|---|---|
| L2 regularization | 0.01 | 0.01 |
| Learning Rate | 0.0005 | 0.0005 |
| Nesterov momentum | 0.7 | 0.6 |
| Decay | $5 \times 10^{-8}$ | $1 \times 10^{-4}$ |
| $\beta$ | 225 | 111 |
| $\alpha$ | 500 | 500 |
| Batch size | 8 | 8 |

Table 4.2: Parameter values used for training.

## 4.4 Results

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

### 4.4.1 Saliency Computation

#### 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 Figure 4.5(a)(b)). Figure 4.5(b) shows progression in tuning of 3 channels of the center-surround filters

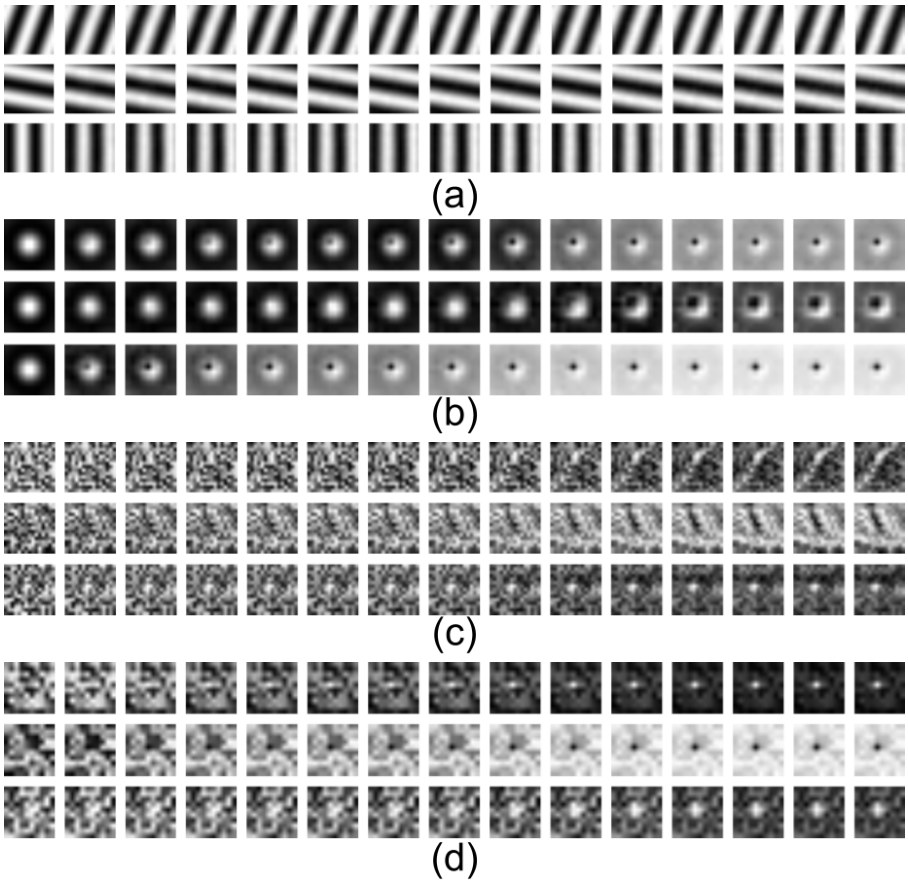



Figure 4.5: Evolution of the filters during training of hard exudates saliency. (a) 3 channels of $W_{orient}$ from stage 2 (b) 3 channels of $CS3$ filters from stage 4 (c) 3 channels of random filters from stage 2 (d) 3 channels of random filters from stage 4.

of $S_{HE}$. It can be seen that the pattern of tuned filters are similar to difference of multiple ( $> 2$) Gaussian filters. Figure 4.5(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.

**Evaluation of Saliency**

The performance of the saliency models was evaluated against seven existing computational saliency models: Itti-Koch [15], SR [53], AIM [54], GBVS [56], Torralba [87], Judd [55] and *Rare* [88]. Among these Itti-Koch 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 Figure 4.6 for hard exudates and Figure 4.7 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 Figure 4.6, 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 Figure 4.7, 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 Figure 4.8. 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 4.3). Normal images were taken from DRiDB [85] and DIARETDB0 [89]. 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 negative(FN). Background pixels, if correctly/incorrectly detected with reference to GT, are considered true negative(TN)/false positive(FP).

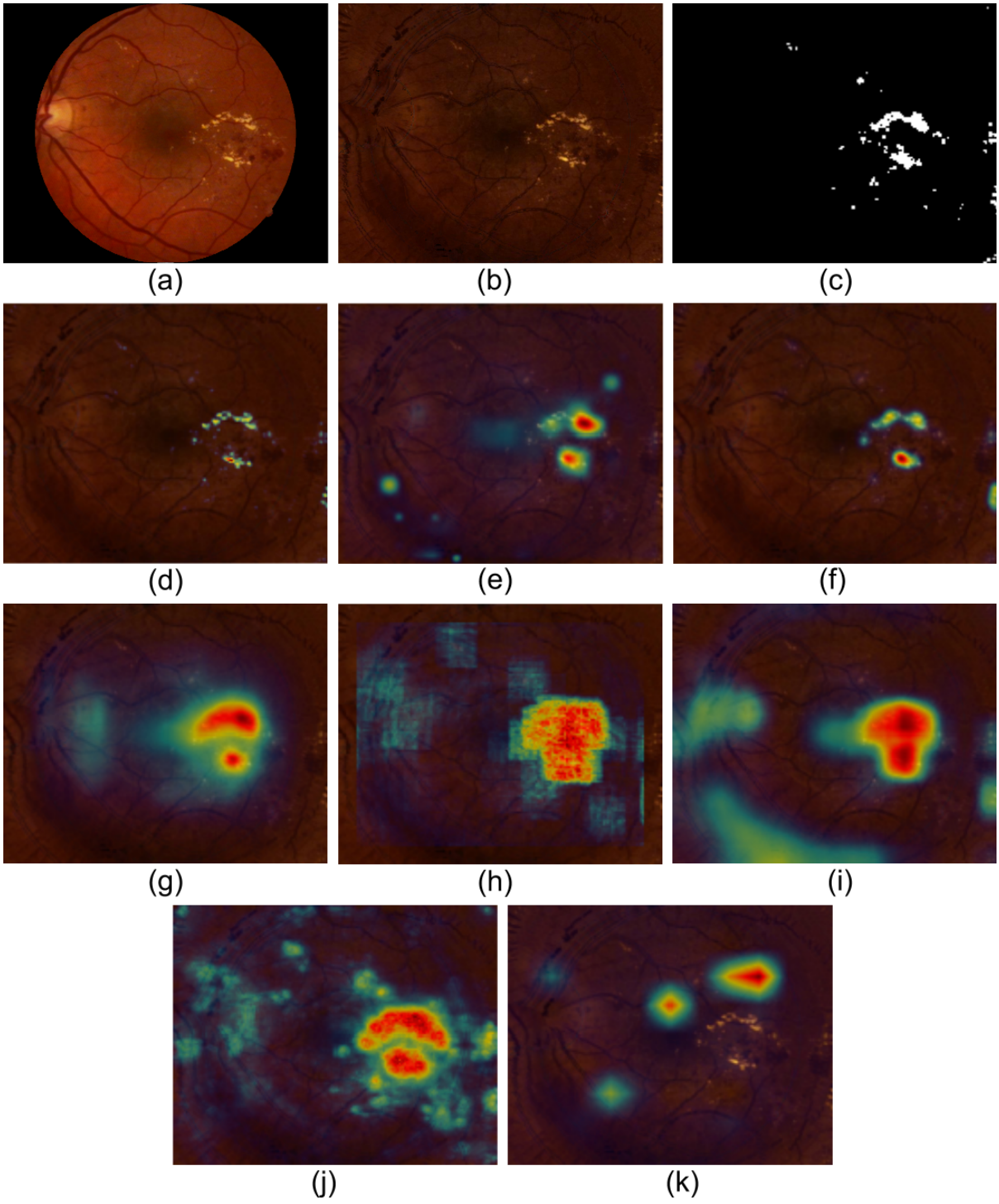



Figure 4.6: Hard exudate saliency. (a) original color fundus image (b) pre-processed image (c) ground truth. Computed saliency maps of (d) Proposed (e) Itti-Koch (f) SR (g) GBVS (h) AIM (i) Rare (j) Torralba (k) Judd.

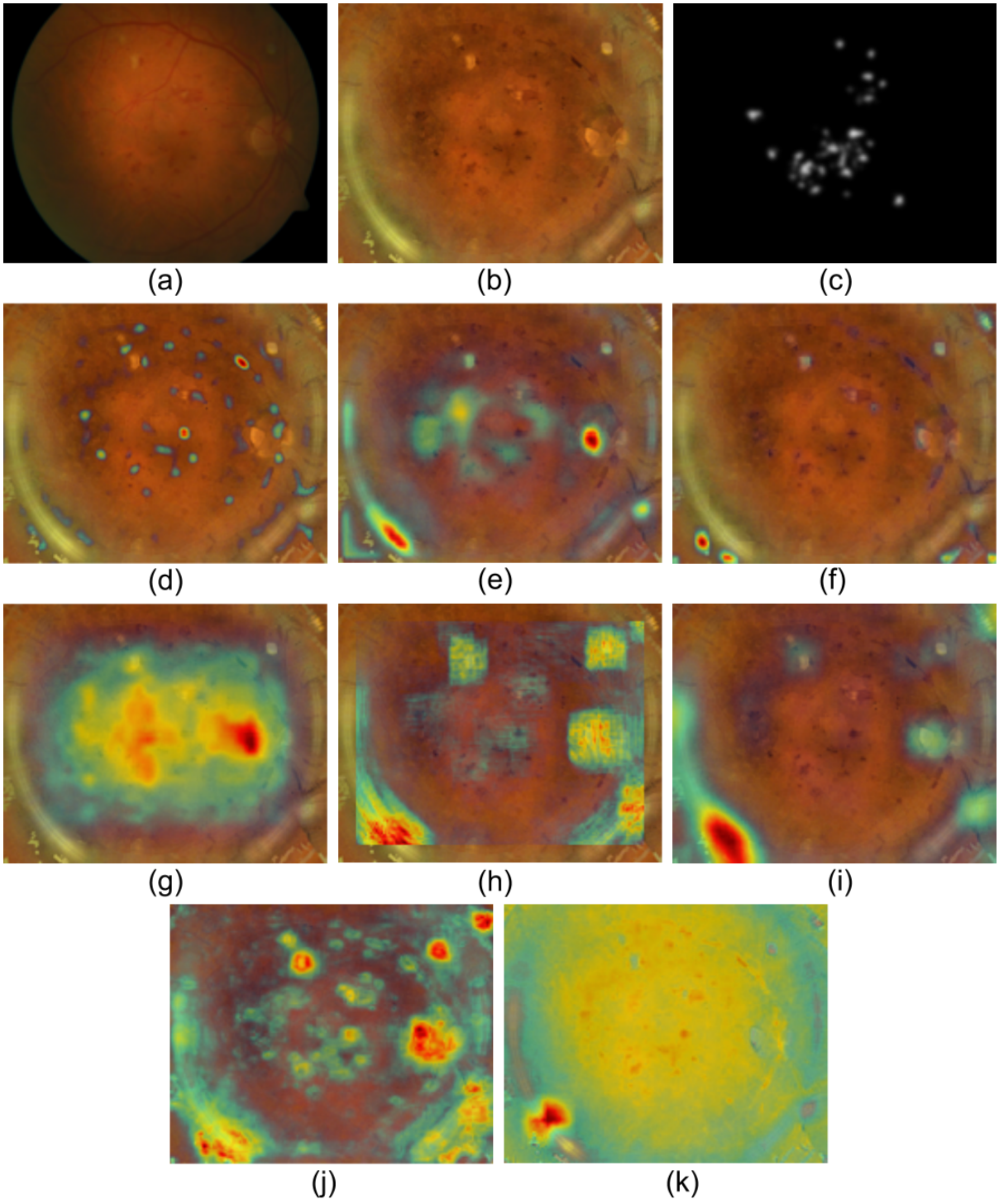



Figure 4.7: Hemorrhage saliency. (a) original color fundus image (b) pre-processed image (c) Gaussian convolved ground truth. Computed saliency maps of (d) Proposed (e) Itti-Koch (f) SR (g) GBVS (h) AIM (i) Rare (j) Torralba (k) Judd.

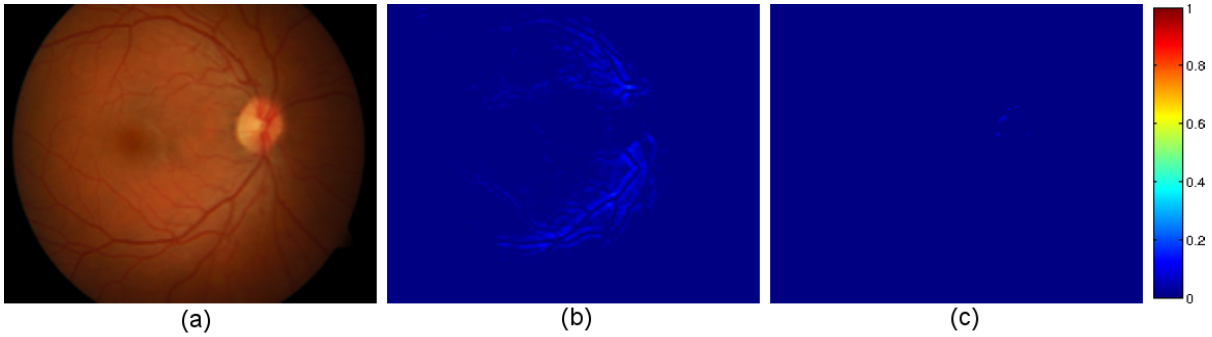



Figure 4.8: Predicted saliency for normal cases. (a) Normal color fundus image and saliency maps for (b) hard exudate (c) hemorrhage.

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

Table 4.3: Number of images in the test set.

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 Figure 4.9 and Table 4.4. 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 Figure 4.7(k). Consequently, the performance of Judd model 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.

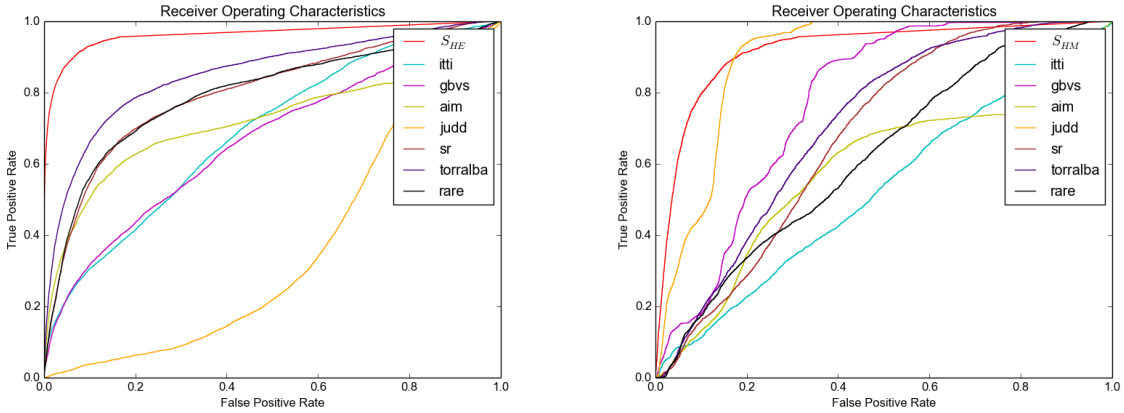



Figure 4.9: Receiver Operating Characteristics(ROC). (a) hard exudate saliency (b) hemorrhage saliency.

As noted in [90] 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. 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 Figure 4.10). 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 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 Figure 4.11. 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

| | 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 |

Table 4.4: Comparison of AUC scores.

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 Figure 4.12. 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 is seen in the middle of the optic disc where blood vessels converge, which is erroneous.

### 4.4.2 Lesion-emphasis

Qualitative results of ALES are shown in Figure 4.13 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 emphasis. It is notable that the image in Figure 4.13(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.

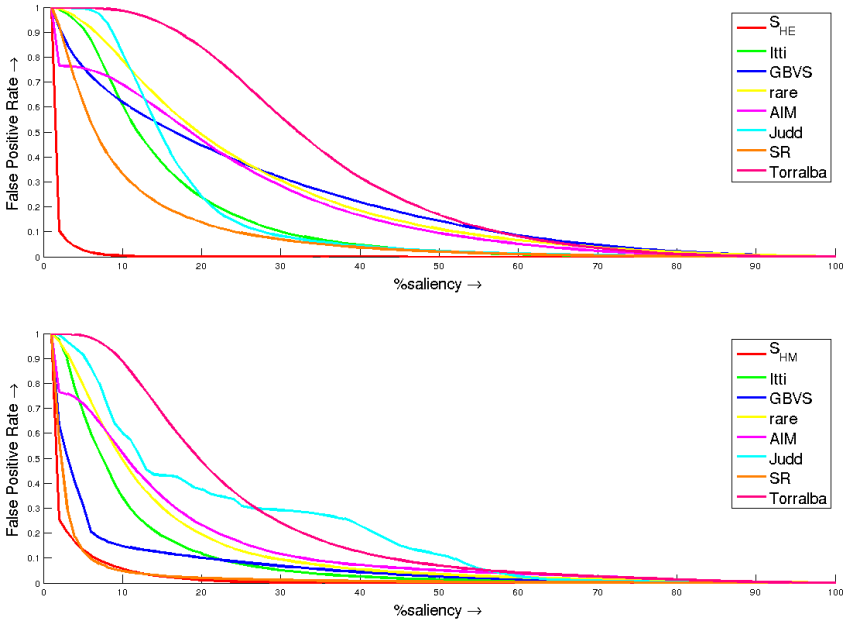



Figure 4.10: Flase positive rate vs saliency. (top) hard exudates (bottom) hemorrhages.

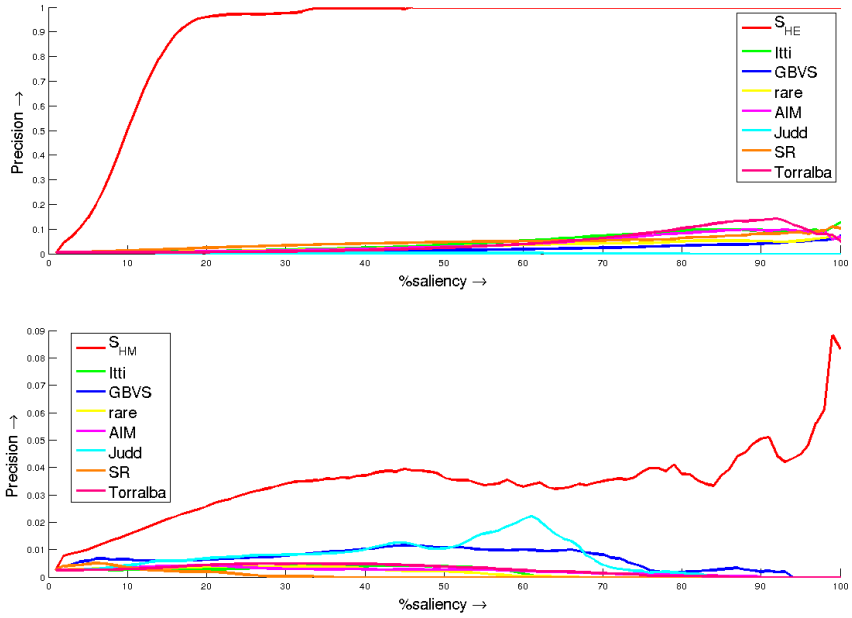



Figure 4.11: Positive Predictive Rate/Precision vs saliency. (top) hard exudates (bottom) hemorrhages.

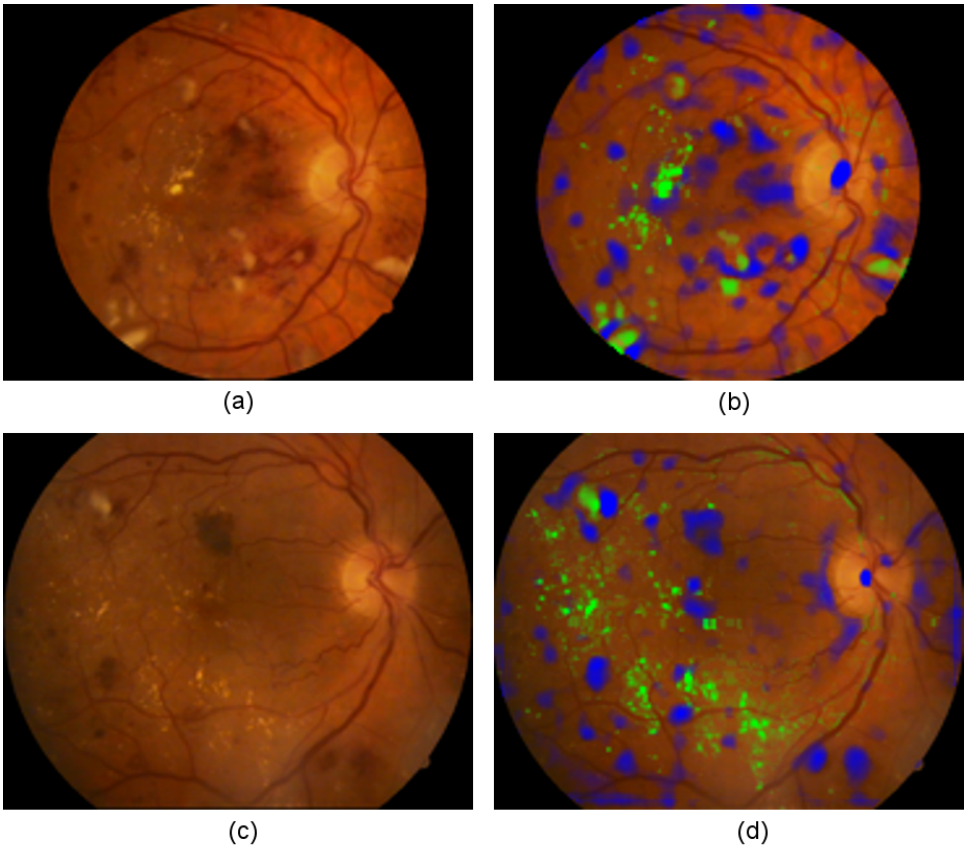



Figure 4.12: Combined saliency for hard exudate and hemorrhage. (a,c) Original images with lesions (b,d) combined saliency maps for hard exudate (green) and hemorrhage (purple) shown overlaid on the original image.

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

$$CNR = \frac{|m_f - m_b|}{\sigma_b} \tag{4.20}$$

where, $m_f$ and $m_b$ are mean intensity of foreground (lesions) and background respectively. $\sigma_b$ 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 4.5. 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 Figure 4.14. ALES can be seen to introduce minimal artifacts, as desirable.

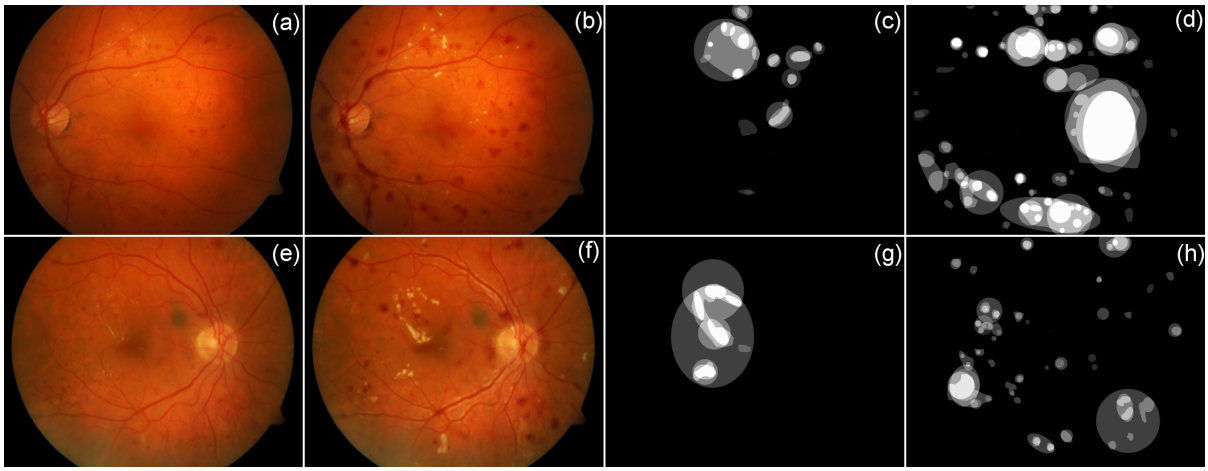



Figure 4.13: ALES output for abnormal images. (a)(e) original images (b)(f) corrected images (c)(g) hard exudate GT (d)(h) hemorrhage GT.

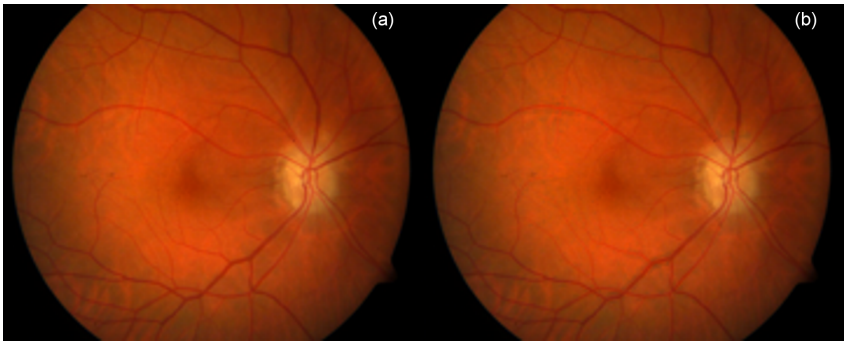



Figure 4.14: ALES output for normal image. (a) original image (b) corrected image.

| | 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 |

Table 4.5: Average contrast-to-noise ratio.

## 4.5 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 [9].

### 4.5.1 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.

### 4.5.2 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.

### 4.5.3 Experiment design

For both studies, images were shuffled and displayed on Lenovo monitor of size $1366 \times 768$ 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.

### 4.5.4 Results

The results for the first and second studies are shown in Tables 4.6 and 4.7 respectively. The results in Table 4.6 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 4.7 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 4.7 also shows that ALES is effective for DR detection as the sensitivity has increased significantly with ALES.

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

Table 4.6: 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 for DME | 70.83% | 79.17% | $< 0.05$ |
| Sensitivity for DR | 41.56% | 49.78% | $< 0.05$ |

Table 4.7: Performance for local level decision in Study 2.

## 4.6 Conclusions

We presented supervised approach for designing ALES for 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 can be done in a clinical setting.

# Chapter 5

# Conclusions

"That's one small step for a man, one giant leap for mankind."

– *Neil Armstrong*

We have looked at specific disease called diabetic retinopathy(DR). DR is a leading cause of visual impairments and permanent blindness. Mass screening is an important initiative for early detection and cure of DR. Screening is done by arranging camps and diagnosing adults at risk of developing DR. In order to increase the reach and throughput of mass screening, a service called reading-center has come to existence. Image reading-centers are crucial part of screening diabetic retinopathy. Images coming from various sources are analyzed by readers working in reading centers and the reports are used to recommend further consultation. Image reading is a lengthy process which leads to fatigue and subsequently misdiagnoses. we aimed at developing assistance for DR image readers. First we analyzed diagnostic strategies used by retinal experts with various level of experience and derived a recommendation for fast and accurate diagnosis. In order improve performance of DR readers further, we proposed a design where computerized diagnosis work as an assistant to a reader. The design of computer assisted diagnosis(CAD) was done using assistive lesion emphasis system or ALES. ALES was developed based on visual perception and it draws readers' attention to lesion in least obtrusive way.

In order to understand best practices and derive a recommendation, we conducted an eye tracking study with 57 participants while they were diagnosing DR images. We collected their eye movements and other behavioral data like response time, and analyzed them to understand strategy used for DR diagnosis. A new quantity named coefficient of scanning($CS$) was designed and

based on that diagnostic strategies were classified into two bins, *tracing* and *dwelling*. *Tracing* refers to the strategy where image-reader scrutinizes all regions by visiting them multiple times; but with superficial examination in each visit. *Tracing* strategy leads to slower decision making. Also, while using *tracing* strategy diagnostic accuracy of semi-experts and novices suffers more than experts. *Dwelling*, on the other hand, refers to the strategy where image-reader scrutinizes one region thoroughly before moving to next region. This eliminates necessity of revisiting to same region multiple time, making *dwelling* a fast strategy. *Dwelling* also leads to accurate diagnosis irrespective of the expertise-level of a reader. Hence *dwelling* can be marked as most efficient strategy. We also examined gaze-patterns of medically trained participants and found them very similar. This patterns are combined to extract an optimal search strategy. Optimal search strategy recommends image-readers to scan retinal zones in a particular weighted sequence. This pattern takes care of visual biases and hence presumed to be easily practicable. Also, the optimal strategy has properties of terminative search, which allows one to stop scanning the image as soon as an abnormality is located. This can hopefully make diagnosis faster.

In order to boost the performance of readers further we have developed saliency based lesion emphasis system called ALES. ALES was developed as two stage system with saliency and selective-enhancement as two stages. Two designs of ALES were presented. One uses unsupervised saliency model and other supervised saliency model. Both ALES employed different(but interchangeable) techniques for lesion emphasis. Unsupervised saliency model used spatially-varying erosion and dilation(SED) operations to mimic human fixation and center-surround filters to mimic ganglion cells in the retina. SED was validated against an average Gaze-map of 15 experts and found to have 10% higher recall than four leading saliency models proposed for natural images. Supervised saliency model was a CNN based implementation of existing biologically inspired saliency model and is trained with novel loss function. Computed saliency has 10%(for hemorrhages) and 20%(for hard exudate) higher AUC than existing models. Saliency maps were used for selective enhancement of lesion. The lesion emphasis was done in two ways, one with multiscale fusion of saliency map and original image and another with spatially varying gamma correction. Both ALESs improved CNR of lesions by 30%. A perception study also validated the effectiveness of ALES. We suggest that DR diagnosis done with ALES maps rather

than original images and scanning them with recommended optimal strategy will prove to be an efficient practice.

Solutions offered in this thesis are just a starting point of a long journey for developing reader-centric CAD. Using the findings of this thesis one can also develop an artificially intelligent reader(AI-reader). For example, a neural network of AI-reader sees an output of ALES and scan this output in a sequence as recommended by optimal search pattern. As optimal strategy is terminative in nature, a network can stop scanning image as soon as an abnormality is located. This can save lot of computation and make the diagnosis faster. The CAD developed this way will have ability to read images in *human-like* fashion and might replace human reader with sufficient training. This can create a big impact on automated diagnosis of future generation.

# Publications

## Related Publications

- Rangrej, Samrudhdhi B., Jayanthi Sivaswamy, and Priyanka Srivastava.
  *"Investigating Visual Search for Diagnosis of Diabetic Retinopathy."*
  Under preparation

- Rangrej, Samrudhdhi B., and Jayanthi Sivaswamy.
  *"ALES: an assistive system for fundus image readers"*
  Under review (Journal of Medical Imaging)

- Rangrej, Samrudhdhi B., and Jayanthi Sivaswamy.
  *"A biologically inspired saliency model for color fundus images."*
  Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing. ACM, 2016.

## Other Publication

- Karthik G.*, Rangrej, Samrudhdhi B.* and Jayanthi Sivaswamy.
  *"A deep learning framework for segmentation of retinal layers from OCT images."*
  Under review (20th International Conference on Medical Image Computing and Computer Assisted Intervention 2017)

* equal contribution

# Bibliography


[1] G. Lotz, T. Peters, E. Zrenner, and R. Wilke, "A domain model of a clinical reading center-design and implementation," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 4530–4533, IEEE, 2010.

[2] D. E. Institute, "Doheny image reading center (dirc)," 2016. [Online; accessed 26-January-2017].

[3] K. C. Oeffinger, E. T. Fontham, R. Etzioni, A. Herzig, J. S. Michaelson, Y.-C. T. Shih, L. C. Walter, T. R. Church, C. R. Flowers, S. J. LaMonte, *et al.*, "Breast cancer screening for women at average risk: 2015 guideline update from the american cancer society," *Jama*, vol. 314, no. 15, pp. 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*, no. 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*, vol. 251, no. 3, pp. 619–626, 2009.

[6] D. E. Singer, D. M. Nathan, H. A. Fogel, and A. P. Schachat, "Screening for diabetic retinopathy," *Annals of Internal Medicine*, vol. 116, no. 8, pp. 660–671, 1992.

[7] W. H. Organization *et al.*, "Global report on diabetes," 2016.

[8] S. S. Gadkari, Q. B. Maskati, and B. K. Nayak, "Prevalence of diabetic retinopathy in india: The all india ophthalmological society diabetic retinopathy eye screening study 2014," *Indian journal of ophthalmology*, vol. 64, no. 1, p. 38, 2016.

[9] P. Mitchell, S. Foran, E. Assistance, and J. Foran, "Guidelines for the management of diabetic retinopathy," *National Health and Medical Research Council*, 2008.

[10] E. A. G. N. R. S. Committee *et al.*, "Framework for the development of a diabetic retinopathy screening programme for ireland," *Naas: Health Service Executive*, 2008.

[11] D. Mitry, T. Peto, S. Hayat, J. E. Morgan, K.-T. Khaw, and P. J. Foster, "Crowdsourcing as a novel technique for retinal fundus photography classification: Analysis of images in the epic norfolk cohort on behalf of the ukbiobank eye and vision consortium," *PloS one*, vol. 8, no. 8, p. e71154, 2013.

[12] E. Y. Chew, M. L. Klein, F. L. Ferris, N. A. Remaley, R. P. Murphy, K. Chantry, B. J. Hoogwerf, and D. Miller, "Association of elevated serum lipid levels with retinal hard exudate in diabetic retinopathy: Early treatment diabetic retinopathy study (etdrs) report 22," *Archives of ophthalmology*, vol. 114, no. 9, pp. 1079–1084, 1996.

[13] 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*, vol. 2, 2012.

[14] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive psychology*, vol. 12, no. 1, pp. 97–136, 1980.

[15] 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*, no. 11, pp. 1254–1259, 1998.

[16] D. G. Albrecht and D. B. Hamilton, "Striate cortex of monkey and cat: contrast response function.," *Journal of neurophysiology*, vol. 48, no. 1, pp. 217–237, 1982.

[17] Y. Kim and A. Varshney, "Saliency-guided enhancement for volume visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 925–932, 2006.

[18] 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*, pp. 73–80, IEEE, 2011.

[19] D. V. Beard, R. E. Johnston, O. Toki, and C. Wilcox, "A study of radiologists viewing multiple ct scans using an eyetracking device," 1989.

[20] S. Ellis, X. Hu, L. Dempere-Marco, G. Yang, A. Wells, and D. Hansell, "Thin-section ct of the lungs: eye-tracking analysis of the visual approach to reading tiled and stacked display formats," *European journal of radiology*, vol. 59, no. 2, pp. 257–264, 2006.

[21] H. Matsumoto, Y. Terao, A. Yugeta, H. Fukuda, M. Emoto, T. Furubayashi, T. Okano, R. Hanajima, and Y. Ugawa, "Where do neurologists look when viewing brain ct images? an eye-tracking study involving stroke cases," *PloS one*, vol. 6, no. 12, p. e28928, 2011.

[22] H. L. Kundel, C. F. Nodine, and E. A. Krupinski, "Searching for lung nodules: Visual dwell indicates locations of false-positive and false-negative decisions.," *Investigative radiology*, vol. 24, no. 6, pp. 472–478, 1989.

[23] K. S. Berbaum, E. A. Brandser, E. Franken, D. D. Dorfman, R. T. Caldwell, and E. A. Krupinski, "Gaze dwell times on acute trauma injuries missed because of satisfaction of search," *Academic radiology*, vol. 8, no. 4, pp. 304–314, 2001.

[24] E. A. Krupinski, A. A. Tillack, L. Richter, J. T. Henderson, A. K. Bhattacharyya, K. M. Scott, A. R. Graham, M. R. Descour, J. R. Davis, and R. S. Weinstein, "Eye-movement study and human performance using telepathology virtual slides. implications for medical education and differences with experience," *Human pathology*, vol. 37, no. 12, pp. 1543–1556, 2006.

[25] E. A. Krupinski, A. R. Graham, and R. S. Weinstein, "Characterizing the development of visual search expertise in pathology residents viewing whole slide images," *Human pathology*, vol. 44, no. 3, pp. 357–364, 2013.

[26] T. Jaarsma, H. Jarodzka, M. Nap, J. J. van Merriënboer, and H. P. Boshuizen, "Expertise in clinical pathology: combining the visual and cognitive perspective," *Advances in Health Sciences Education*, vol. 20, no. 4, pp. 1089–1106, 2015.

[27] C. Mello-Thoms, "Perception of breast cancer: eye-position analysis of mammogram interpretation," *Academic radiology*, vol. 10, no. 1, pp. 4–12, 2003.

[28] H. L. Kundel, C. F. Nodine, E. F. Conant, and S. P. Weinstein, "Holistic component of image perception in mammogram interpretation: gaze-tracking study 1," *Radiology*, vol. 242, no. 2, pp. 396–402, 2007.

[29] L. V. P. E. Institute, "L v prasad eye institute," 2017. [Online; accessed 27-January-2017].

[30] N. Nethralaya, "Narayana nethralaya," 2017. [Online; accessed 27-January-2017].

[31] Neoretina, "Neoretina," 2016. [Online; accessed 27-January-2017].

[32] A. E. Institute, "Anand eye institute," 2015. [Online; accessed 27-January-2017].

[33] medivisioneyecare.com, "Medivision eye care centre," 2015. [Online; accessed 27-January-2017].

[34] centreforsight, "centre for sight," 2017. [Online; accessed 3-February-2017].

[35] M. Nyström and K. Holmqvist, "An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data," *Behavior research methods*, vol. 42, no. 1, pp. 188–204, 2010.

[36] T. Drew, M. L.-H. Võ, and J. M. Wolfe, "The invisible gorilla strikes again sustained inattentional blindness in expert observers," *Psychological science*, p. 0956797613479386, 2013.

[37] I. Hooge and G. Camps, "Scan path entropy and arrow plots: capturing scanning behavior of multiple observers," 2013.

[38] A. L. Yarbus, *Eye movements during perception of complex objects*. Springer, 1967.

[39] J. M. Wolfe, K. R. Cave, and S. L. Franzel, "Guided search: an alternative to the feature integration model for visual search.," *Journal of Experimental Psychology: Human perception and performance*, vol. 15, no. 3, p. 419, 1989.

[40] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013.

[41] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*, pp. 1597–1604, IEEE, 2009.

[42] D. Gao and N. Vasconcelos, "Discriminant saliency for visual recognition from cluttered scenes," in *Advances in neural information processing systems*, pp. 481–488, 2004.

[43] W. Wang, Y. Song, and A. Zhang, "Semantics-based image retrieval by region saliency," in *International Conference on Image and Video Retrieval*, pp. 29–37, Springer, 2002.

[44] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE TIP*, vol. 19, no. 1, pp. 185–198, 2010.

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

[46] V. Jampani, J. Sivaswamy, V. Vaidya, *et al.*, "Assessment of computational visual attention models on medical images," in *ICVGIP*, p. 80, ACM, 2012.

[47] S. Banerjee, S. Mitra, B. U. Shankar, and Y. Hayashi, "A novel gbm saliency detection model using multi-channel mri," *PloS one*, vol. 11, no. 1, 2016.

[48] X. Zou, X. Zhao, Y. Yang, and N. Li, "Learning-based visual saliency model for detecting diabetic macular edema in retinal image," *Computational Intelligence and Neuroscience*, vol. 2016, 2016.

[49] 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.

[50] D. Mahapatra and J. M. Buhmann, "Visual saliency-based active learning for prostate magnetic resonance imaging segmentation," *Journal of Medical Imaging*, vol. 3, no. 1, pp. 014003–014003, 2016.

[51] 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*, pp. 1119–1122, IEEE, 2008.

[52] A. Kumar, P. Sridar, A. Quinton, R. K. Kumar, D. Feng, R. Nanan, and J. Kim, "Plane identification in fetal ultrasound images using saliency maps and convolutional neural

networks," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 791–794, IEEE, 2016.

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

[54] N. D. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of vision*, vol. 9, no. 3, pp. 5–5, 2009.

[55] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Computer Vision, 2009 IEEE 12th international conference on*, pp. 2106–2113, IEEE, 2009.

[56] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, pp. 545–552, 2006.

[57] Q. Zhao and C. Koch, "Learning a saliency map using fixated locations in natural scenes," *Journal of vision*, vol. 11, no. 3, p. 9, 2011.

[58] 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*, pp. 591–598, IEEE, 2008.

[59] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable cosfire filters for vessel delineation with application to retinal images," *Medical image analysis*, vol. 19, no. 1, pp. 46–57, 2015.

[60] L. Tang, M. Niemeijer, J. M. Reinhardt, M. K. Garvin, and M. D. Abràmoff, "Splat feature classification with application to retinal hemorrhage detection in fundus images," *IEEE Transactions on Medical Imaging*, vol. 32, no. 2, pp. 364–375, 2013.

[61] A. Ujjwal, J. Sivaswamy, *et al.*, "An assistive annotation system for retinal images," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1506–1509, IEEE, 2015.

[62] 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," 2005.

[63] N. E. M. Association *et al.*, "Digital imaging and communications in medicine (dicom), part 14: Gray-scale standard display function," 2001.

[64] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, and J. Pietilä, "The diaretdb1 diabetic retinopathy database and evaluation protocol.," in *BMVC*, pp. 1–10, 2007.

[65] 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*, pp. 418–426, Springer, 2002.

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

[67] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3166–3173, 2013.

[68] X. Huang, C. Shen, X. Boix, and Q. Zhao, "Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 262–270, 2015.

[69] 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.

[70] V. Navalpakkam and L. Itti, "Modeling the influence of task on attention," *Vision research*, vol. 45, no. 2, pp. 205–231, 2005.

[71] S. Frintrop, G. Backer, and E. Rome, "Goal-directed search with a top-down modulated computational attention system," in *Pattern Recognition*, pp. 117–124, Springer, 2005.

[72] M. de Brecht and J. Saiki, "A neural network implementation of a saliency map model," *Neural Networks*, vol. 19, no. 10, pp. 1467–1474, 2006.

[73] 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)*, pp. 807–814, 2010.

[74] C. Sinthanayothin, J. F. Boyce, H. L. Cook, and T. H. Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *British Journal of Ophthalmology*, vol. 83, no. 8, pp. 902–910, 1999.

[75] M. Foracchia, E. Grisan, and A. Ruggeri, "Luminosity and contrast normalization in retinal images," *Medical Image Analysis*, vol. 9, no. 3, pp. 179–190, 2005.

[76] E. Grisan, A. Giani, E. Ceseracciu, and A. Ruggeri, "Model-based illumination correction in retinal images," in *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.*, pp. 984–987, IEEE, 2006.

[77] 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*, pp. 591–598, IEEE, 2008.

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

[79] P. Feng, Y. Pan, B. Wei, W. Jin, and D. Mi, "Enhancing retinal image by the contourlet transform," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 516–522, 2007.

[80] 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*, vol. 344, no. 3, pp. 243–256, 2007.

[81] 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*, pp. 163–168, 2005.

[82] 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*, pp. 43–48, ACM, 2011.

[83] S. L. Su, F. Durand, and M. Agrawala, "De-emphasis of distracting image regions using texture power maps," in *APGV*, 2005.

[84] 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*, vol. 21, no. 3, pp. 360–364, 2011.

[85] P. Prentasic, S. Loncaric, Z. Vatavuk, G. Bencic, M. Subasic, T. Petkovic, L. Dujmovic, M. Malenica-Ravlic, N. Budimlija, and R. Tadic, "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*, pp. 711–716, IEEE, 2013.

[86] L. Giancardo, F. Meriaudeau, T. P. Karnowski, Y. Li, S. Garg, K. W. Tobin, and E. Chaum, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Medical image analysis*, vol. 16, no. 1, pp. 216–226, 2012.

[87] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search.," *Psychological review*, vol. 113, no. 4, p. 766, 2006.

[88] M. Mancas, "Relative influence of bottom-up and top-down attention," in *Attention in cognitive systems*, pp. 212–226, Springer, 2008.

[89] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, H. Uusitalo, H. Kälviäinen, and J. Pietilä, "Diaretdb0: Evaluation database and methodology for diabetic retinopathy algorithms," *Machine Vision and Pattern Recognition Research Group, Lappeenranta University of Technology, Finland*, 2006.

[90] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand, "What do different evaluation metrics tell us about saliency models?," *arXiv preprint arXiv:1604.03605*, 2016.