AN ASSISTIVE ANNOTATION SYSTEM FOR RETINAL IMAGES

Ujjwal, Arunava Chakravarty, Jayanthi Sivaswamy

CVIT, IIIT-Hyderabad, India

ABSTRACT

Annotated data is critical for the development of many computer assisted diagnostic (CAD) algorithms. The process of manual annotation is very strenuous, time-consuming and an expensive component in CAD development. In this paper, we propose the idea of an interactive Assistive Annotation System (AAS) aimed at helping annotators by automatically marking possible regions of interest for further refinement by an annotator. We propose an unsupervised, biologically inspired method for *bright* lesion annotation. The performance of the proposed system has been evaluated against region-level ground truth in DiaretDB1 dataset and was found to have a sensitivity of 60% at 7 false positives per image. Preliminary testing was also done on public datasets which do not provide any lesion level annotations. A visual assessment of the obtained results affirm a good agreement with lesions visible in images. The system with a simple modification is shown to have the potential to handle dark lesion annotation, which is a significantly more challenging problem. Thus, the proposed system is a good starting point for exploring the AAS idea for retinal images. Such systems can help extend the use of many existing datasets by enriching the image-level annotations with localised information.

***Index Terms*—** Groundtruth annotation, color fundus images, assistive annotation, GMP.

1. INTRODUCTION

Ground truth (GT) annotations of medical images play a key role in CAD systems development. Most CAD systems are supervised and hence require considerable amount of annotated data for training and validation. Unavailability of enough annotated data for training may lead to overfitting of the system leading to poor cross-training performance.

While image-level annotations are comparatively easy, acquiring structural annotations for medical landmarks and abnormalities is often a tedious enterprise. It requires precious time of multiple medical experts to provide a consensus based gold standard. While the nature and quality of these annotations fundamentally affect the design and validation of CAD systems, the task, due to its fully manual nature, is prone to human errors resulting in inconsistencies and large inter-observer variance.

To overcome the above difficulties, we envision an Assistive Annotation System (AAS): a fast, simple and interactive semi-automatic system that generates annotations with minimal False Positives (FP) in real time and allows manual annotators to further refine the annotations based on specialized domain and case-specific knowledge. AAS is not aimed at substituting human annotators but rather assisting them. An accurate AAS can significantly reduce the burden on human annotators by making their task more of verification and less of annotation. It can expedite the annotation process thereby allowing annotation over larger datasets and an efficient use of experts time.

While the computer vision community has invested efforts for developing automatic and semi-automatic GT generation tools, similar efforts in medical image analysis domain have been limited [1, 2, 3]. One obvious reason lies in the need for specialized medical domain knowledge required for the annotations.

To the best of our knowledge there have not been any attempts at developing effective AAS systems. From an image analysis perspective, this task seems similar to candidate extraction for lesions. However candidate extraction stages in retinal CAD systems do not exhaustively focus on performance levels suited for annotation assistance and usually result in very high FPs [4].

In this paper, we propose a novel framework for annotation of bright and dark lesions in fundus images. Bright lesions (BL) such as hard exudates and drusen are associated with diabetic retinopathy (DR) and age-related macular degeneration, respectively[5]. Dark lesions such as haemorrhages and microaneurysms are also of great significance as they are early indicators of DR. Grading of DR severity is evaluated based on the number, type and location of these lesions necessitating accurate lesion-level GT for proper diagnosis. We propose a fast, unsupervised, biologically inspired method for lesion annotation that extends the Generalized Motion Pattern (GMP) [6] image representation. While GMP has been shown to be effective in detection of lesion at an image level, we extend it for the task of lesion localization and detection by aggregating evidence from multiple localized GMPs. The proposed method has been validated against manual markings on the public dataset of DiaretDB1 [7].

This work is partly funded by the Department of Science and Technology, Govt. of India, under Grant DST/INT/NL/Biomed/P(3)/2011(G)

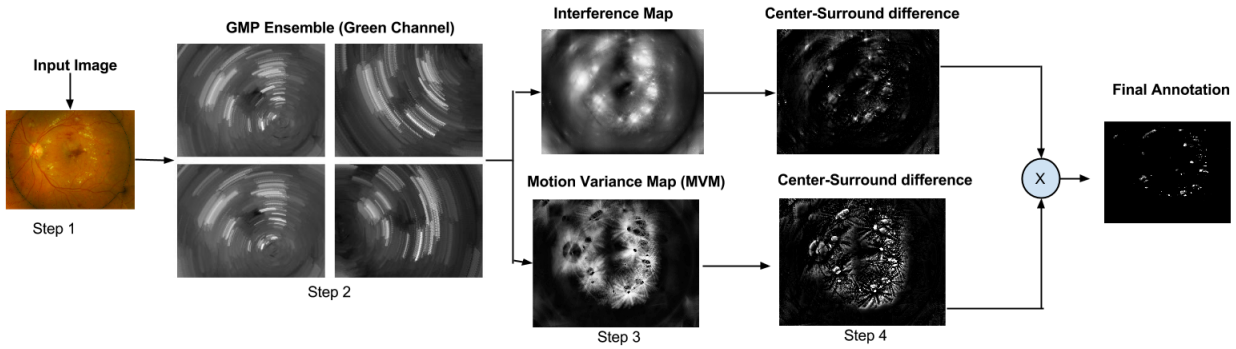

Fig. 1. Proposed AAS pipeline (with intermediate results) for generating annotations for bright lesions.

We also show a qualitative results on images from other unmarked datasets to demonstrate its effectiveness.

The proposed AAS pipeline (see figure 1) for detecting bright lesions is as follows: The input image, which is assumed to be labelled as abnormal, is preprocessed to handle non-uniform illumination [8] and extend the circular field of view [9] to avoid artifacts that occur due to the circular mask. Subsequent processing is limited to the green channel of the input image as it has the best contrast for both dark and bright lesions. A rotational motion is induced in the preprocessed image about a set of randomly chosen pivot points to derive an ensemble of GMPs (section 2.1). These are integrated to obtain two maps: an interference pattern map and a map of variance of pixel values across the GMPs (section 2.2). Local maxima are found (at multiple scales) in these two maps and combined to obtain the final soft map in which high pixel values indicate lesion locations (section 2.4). Detection of dark lesions proceeds in a similar manner. These steps are described in detail next.

2. PROPOSED METHOD

2.1. Random GMP Ensemble

Given an image I, its GMP representation I_{GMP} is defined as

$$I_{GMP}(\vec{r}) = f(\{I(T_j(\vec{r})|1 \leq j \leq N\})) \quad (1)$$

Here $\vec{r}$ denotes a pixel location, T denotes a set of N rigid transformations applied to I to produce a set of N resultant images $\{I(T_j(\vec{r})|1 \leq j \leq N\}$. Finally, these images are combined into the GMP signature using a coalescing function $f(.)$ that maps the set of pixel intensities at each location $\vec{r}$ across the set of transformed images to a scalar value.

Ensemble GMP generation: For a given image I, we denote the GMP generated by inducing a rotation motion about a pivot P_k as $I^{(k)}_{GMP}$. An ensemble of GMPs is generated when K pivots are chosen: $C = \{I^{(k)}_{GMP}|1 \leq k \leq K\}$. In our work, each $I^{(k)}_{GMP}$ was generated using eq. 1, with T as rotation by an angle in the range $[-\theta, \theta]$ degrees in steps of $1\,\mathrm{deg}$. The coalescing function f was chosen as max. Since the GMP generation process serves to spatially extend a lesion along the direction of motion which is now along an arc, the value of θ determines the length of the arc over which it is extended. We now define two quantities fundamental to our design : i) an *Interference map* I_{int}, which is derived by combining K GMPs generated with different pivots. $I_{int} \triangleq \sum_k I^{(k)}_{GMP}$ and ii) a *Variance map* $I_{var} \triangleq Var(C)$, where Var denotes the variance of pixel values across the GMPs in the ensemble C. These two maps capture complementary information essential to the detection of lesions as explained next.

1. At true lesion locations, the pixel value will be high in every $I^{(k)}_{GMP} \in C$. Hence I_{int} will be a maximum and I_{var} will be a minimum at such locations.
2. At locations *not in* the neighbourhood of any lesion, the pixel value will be low in every $I^{(k)}_{GMP} \in C$. Hence, both I_{int} and I_{var} will be low at such locations.
3. At locations *within* the neighbourhood of any lesion, the pixel value can be high in every $I^{(k)}_{GMP} \in C$ depending on the lesion configuration in the given image. Hence, I_{var} will be high at such locations. It should be noted that the neighbourhood size over which this holds will depend on the angle of rotation θ.

Figure 1 illustrates these observations.

2.2. Evidence Gathering and Aggregation

As discussed above, bright lesions are characterized by local maxima in I_{int} and local minima in I_{var}. In our work, these extrema were found using a multi-scale (8 dyadic scales), center-surround difference operation [10]. The derived extrema maps O_{int} and O_{var} capture evidence for the presence of lesion in a given location. They are aggregated into an Evidence map I_E for lesion as:

$$I_E \triangleq O_{int} \times O_{var} \quad (2)$$

I_E will have a strong response at lesion locations. While the multiplication operation effectively suppresses the response at

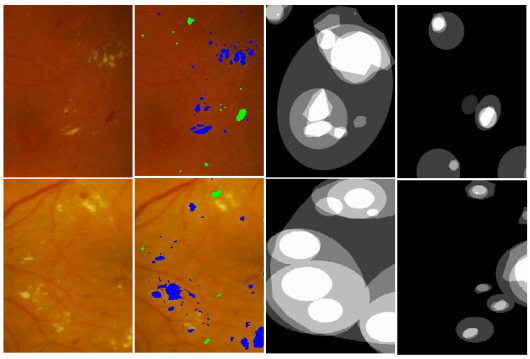

Fig. 2. Sample results on image patches from Diaretdb1. Left to right: original; original with AAS results overlaid (blue and green colours denote bright and dark lesions, respectively); GT markings for bright and GT markings for dark lesions.

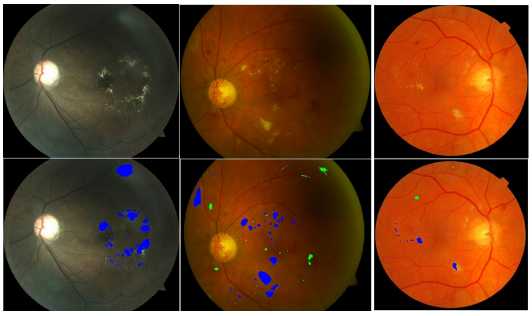

Fig. 3. Annotations generated by AAS on 3 sample images from DMED (left), Diaretdb0 (middle) and Messidor (right) datasets.

false locations, further post-processing is done to i) suppress the Optic Disc (OD) as it shares similar features as bright lesions and ii) remove pixels of value less than 0.1 from the normalized I_E and finally iii) smoothen with a Gaussian filter to obtain a softmap.

3. EXPERIMENTS AND RESULTS

The proposed AAS was evaluated on several public datasets, of which only one [7] provided regional markings for lesions from 4 experts. Even though the proposed AAS was designed for bright lesions (BL), we also investigated its potential for handling dark lesions (DL) by making simple modifications to the proposed method: the input image was complemented and vessels were masked out. Fig. 2 shows obtained annotations on sample regions with lesions from Diaretdb1. More sample results are provided in Fig. 5. A comparison with GT in the last column indicates a high level of agreement. It should be noted that AAS was not explicitly designed to segment the lesions, which explains the variable level of overlap between the annotations and the GT. Fig. 3 shows annotations obtained for sample images from public datasets (Diaretdb0 [11], Messidor [12], DMed [13]) for which *no* regional markings are available. Results show that despite intra-image variations and the unsupervised setting, the proposed AAS produces very good estimates of the regions which contain BL or DLs. Some false annotations for DL appear on vessel segments (Fig 3, column 2) highlighting the need for a more precise vessel map for masking. Nevertheless, the results are encouraging as the AAS designed for BL is also able to locate DLs after a simple modification.

Next, we analysed the influence of the parameter settings on the derived annotations, namely, N, the number of pivot points and θ, the angle of rotation. The Positive Predictive Value (PPV)) is the fraction of overlap, at a pixel level, between the GT and annotations. This is plotted against t, the threshold used to binarise I_E and obtain annotations in Fig. 4 (a) and (b) for Diaretdb1. The former shows the effect of varying N with $\theta = 10\,\mathrm{deg}$. A high value of N will statistically boost the evidence gathered in I_{var} and I_{int} albeit at an increased computational cost. The plots indicate that for $N > 50$, the improvement in PPV is only marginal. In Fig. 4 (b), the effect of varying θ is studied for $N = 150$. A low θ will result in a low spatial extension of a pixel and cannot adequately smooth out the background tissues while a high value will lead to excessive spatial extension and overlap across multiple lesions affecting the lesion localization in I_E. The plots in Fig. 4 (b) are consistent with this expectation as the PPV is lowest for $\theta = 20\,\mathrm{deg}$ for all t. Thus, we conclude that the choice of θ is more critical than N, when N is sufficiently large. A maximum PPV value of 0.65 is obtained. This translates into a huge savings in the effort to generate annotations. The question that remains is if this savings is at the cost of FPs which will have to be removed by manual intervention. This was examined by plotting the sensitivity (SN) of our AAS against FPs per image (*fppi*) or the FROC curve.

Since Diaretdb1 has GT from 4 experts, a majority concurred (3 out of the 4 experts) on lesion presence with $> 50\%$ confidence. The evaluation was performed for different extents of spatial overlap x between the result and GT. A GT region was declared to be detected (TP), if at least one connected component in our result had more than $x\%$ overlap with it, else it was declared as a False Negative (FN). All connected components in our result that did not have at least a $x\%$ overlap with the GT markings were taken as FP. θ and N were assigned their optimal values (10 and 150 respectively). FROC curves (Fig. 4(c)) for different x were obtained for BL by plotting $SN = \frac{TP}{TP+FN}$ against the average *fppi* by varying t. In addition to the trade off between the SN and *fppi*, the plots indicate that when maximal overlap is desired, the number of correctly marked annotations will be low as SN falls sharply for $x\,90\%$. For a moderate overlap setting (black curve), 60% of the BLs are correctly identified by the system for an *fppi* of 7. These results establish the applicability of the proposed system for assistive annotation.

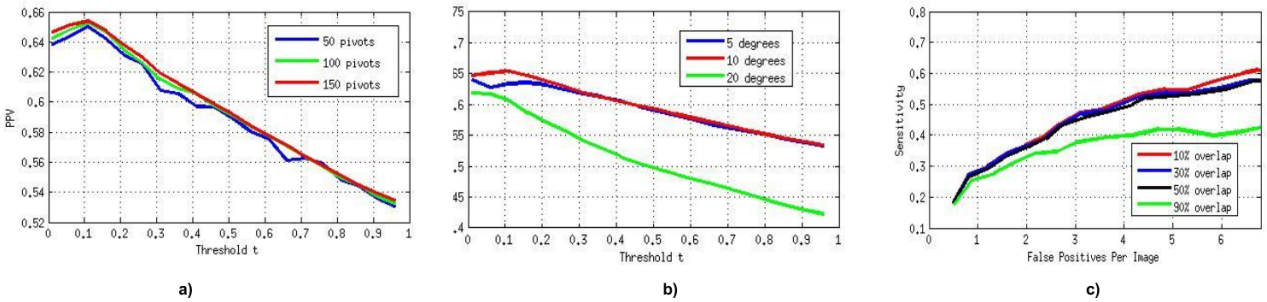

Fig. 4. Annotation evaluation: (a) PPV vs different N, (b) PPV vs θ and (c) FROC plot for bright lesions.

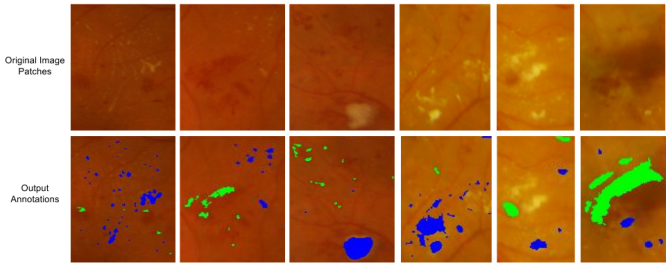

Fig. 5. Additional results on image patches from DiaretDB1[7]

4. CONCLUSION

A simple and biologically inspired method for building an AAS is proposed for assisting human experts in annotating retinal lesions. The time for annotating an image from Diaretdb1 is 50 seconds with an unoptimised, MATLAB implementation on a $i7$ processor with 8 cores. Qualitative results of evaluation of AAS show a good performance across labeled and unlabeled images. Quantitative results of also demonstrate the potential savings to be had by experts in the annotation process with the AAs. We argue that an $fppi < 10$ that was shown to be attainable is acceptable since they can be manually removed whereas the alternative is to derive *all* the annotations manually which is also not guaranteed to be 100% accurate as can be generally seen from the inter-observer variations in GT markings. The methodology presented here paves the way to work towards building annotation assistance systems which can be readily used by human annotators to reduce the burden of large-scale annotations.

5. REFERENCES

[1] "Imageclef - the clef cross language image retrieval track," `www.imageclef.org`.

[2] "Liver ct annotation task," Available at `www.imageclef.org/2014/liver`.

[3] Maier-Hein et al., "Can masses of non-experts train highly accurate image classifiers?," in *MICCAI 2014*. 2014.

[4] Balint Antal et al., "An ensemble-based system for microaneurysm detection and diabetic retinopathy grading," *IEEE TBME*, 2012.

[5] Michael D et.al Abràmoff, "Retinal imaging and image analysis," *Biomedical Engineering, IEEE Reviews in*, 2010.

[6] K Sai Deepak et al., "Automatic assessment of macular edema from color retinal images," *IEEE TMI*, 2012.

[7] V. Kalesnykiene et.al, "Diaretdb1 diabetic retinopathy database and evaluation protocol," .

[8] Marco Foracchia et al., "Luminosity and contrast normalization in retinal images," *Medical Image Analysis*, 2005.

[9] Li Tang et al., "Splat feature classification with application to retinal hemorrhage detection in fundus images," *IEEE TMI*, 2013.

[10] Laurent Itti and Christof Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision research*, 2000.

[11] "Diaretdb0: Evaluation database and methodology for diabetic retinopathy algorithms," *Lappeenranta University of Technology, Finland*, 2006.

[12] "Kindly provided by the messidor program partners (see http://messidor.crihan.fr)".," .

[13] Luca Giancardo et al., "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Medical Image Analysis*, vol. 16, no. 1, pp. 216–226, 2012.