Robust Localization of Retinal Lesions via Weakly-supervised Learning

Ruohan Zhao, Qin Li, Jane You

Contact: csrzhao@comp.polyu.edu.hk
Diabetic Retinopathy (DR) - Disease

- In 2014 there have been 415m adults living with diabetes. About 145m (35%) had some form of diabetic retinopathy (DR). Among these 45m (11%) had vision-threatening DR. In 2040 about 642m adults will have diabetes.

- About 7m of people with diabetes are blind due to DR.

- Low- and middle-income countries account for about 75% of the global diabetes cases. But medical infrastructure is lacking to identify and treat this disease.

- There are no early symptoms, but early detection and treatment can reduce the risk of vision loss by 95%.
Diabetic Retinopathy (DR)

- Hemorrhages
- Optic Disc
- Soft exudates
- Red small dots
- Hard exudates
Traditional diabetic eye screening lasts about 30 minutes. Within six weeks, they will receive results. Requires a well-trained clinician to manually evaluate color fundus photographs of the retina. Equipment are highly demanded in rural areas.

Motivation

- Traditional diabetic eye screening lasts about 30 minutes.
- Within six weeks, they will receive results.
- Requires a well-trained clinician to manually evaluate color fundus photographs of the retina.
- Equipment are highly demanded in rural areas.

Motivation

- Traditional diabetic eye screening lasts about 30 minutes.
- Within six weeks, they will receive results.
- Requires a well-trained clinician to manually evaluate color fundus photographs of the retina.
- Equipment are highly demanded in rural areas.

Retinal image

Classification

Healthy

Unhealthy

How?

Lesions localization

Only image-level label

**Method**

1. **Classification:** The network is trained in *only labels* on image level (No DR, Referable DR).

   - Non-referable DR (No DR)
     - NO DR
     - Mild
   - DR
     - Moderate
     - Severe
     - Proliferative DR
Method

2. Proposed Framework

Algorithm 1: Training and localizing procedure

Training procedure:

Input: Training Data \( I = \{(I_i, c_i)\}_{i=1}^{N} \)

Output: Network parameters \( \theta \)

While training is not convergent:

- Use mix-training \( I_2, c_2 \leftarrow \text{Mix-Training}(I_1, I_2, c_1, c_2) \);
- Get classification score \( Y \leftarrow \text{Classifier}(I_2) \);
- Update \( \theta \leftarrow \text{BinaryCrossEntropy}(Y, c) \);

Localization procedure:

Input: The inference image \( I_t \)

Output: localization map \( L \)

\( A_n^k, Y \leftarrow \text{Feed-forward}(I_t) \);

For different layers \( n \):

\( L_n \leftarrow \text{Grad-CAM}^{++}(A_n^k, Y) \);

\( L \leftarrow \text{Aggregation}(L_n) \)
Method

3. Mix-training strategy. Promote the capacity of capturing diversified lesions.

We choose two random training samples, denoted as \( \{I_1, I_2\} \in \mathbb{R}^{W \times H \times 3} \) with their corresponding image labels \( \{c_1, c_2\} \). We then crop a random patch from \( I_1 \) and overlaid the cropped patch to the corresponding region of \( I_2 \), to synthesize a new training sample \( \hat{I}_2 \).

\[
\hat{I}_2 = M \odot I_1 + (1 - M) \odot I_2
\]

\[
r_x \sim U(0, W), \quad r_y \sim U(0, H),
\]

\[
r_w = \lambda \cdot W, \quad r_h = \lambda \cdot H,
\]

\[
\hat{c} = \begin{cases} 
\lambda c_1 + (1 - \lambda) c_2 & c_1 \in DR, c_2 \in NDR \\
\hat{c} & \text{others}
\end{cases}
\]
Method

4. Inference method

The feature map extracted by the classifier reflects the parts of the fundus that are investigated by the classification network for assigning a label. To leverage feature maps from multi-layers and classification score, Grad-CAM++ can then be used to derive the localization map. We firstly feed the image $I_t$ into the network and obtain classification score $Y$ of referable fundus. Denote the feature map at the $n^{th}$ convolution layer of unit $k$ as $A_n^k$.

$$L_n = \text{ReLU} \left( \sum_k w_k \cdot A_n^k \right)$$

$$w_k = \sum_i \sum_j \alpha_{ij}^k \cdot \text{ReLU} \left( \frac{\partial Y}{\partial A_i^k} \right)$$

Result

Test data set: DiaretDB1 dataset [3]

- High resolution images used for testing
- Lesions marked by four experts
- Regions with more than 75% confidence among the experts are considered as acceptable.

TABLE I: Performance evaluation at lesion-level with other methods on DIARETDB1.

<table>
<thead>
<tr>
<th>Method</th>
<th>Red lesion</th>
<th>Bright lesion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Microaneurysm</td>
<td>Hemorrhages</td>
</tr>
<tr>
<td></td>
<td>Sen% FPI</td>
<td>Sen% FPI</td>
</tr>
<tr>
<td>Chudzik et al.</td>
<td>64.1 8</td>
<td>-</td>
</tr>
<tr>
<td>Seoud et al.</td>
<td>63.9 8</td>
<td>-</td>
</tr>
<tr>
<td>Quellec et al.</td>
<td>61 10</td>
<td>71 10</td>
</tr>
<tr>
<td>Gondal et al.</td>
<td>52 1.5</td>
<td>91 1.5</td>
</tr>
<tr>
<td>Ours</td>
<td>63.4 4.2</td>
<td>95.9 2.6</td>
</tr>
</tbody>
</table>

Result

**TABLE II**: Sensitivity % at image-level on the DIARETDB1 dataset. The best is shown in bold.

<table>
<thead>
<tr>
<th>Method</th>
<th>MAs</th>
<th>HEs</th>
<th>Soft Exudates</th>
<th>Hard Exudates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liu <em>et al.</em></td>
<td>-</td>
<td>-</td>
<td>83.0</td>
<td>83.0</td>
</tr>
<tr>
<td>Zhou <em>et al.</em></td>
<td>-</td>
<td>94.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zhao <em>et al.</em></td>
<td>-</td>
<td><strong>98.1</strong></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Quellec <em>et al.</em></td>
<td>-</td>
<td>94.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gondal <em>et al.</em></td>
<td>50</td>
<td>97.2</td>
<td>90.9</td>
<td><strong>100</strong></td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td><strong>68.9</strong></td>
<td><strong>97.5</strong></td>
<td><strong>92.2</strong></td>
<td><strong>98.5</strong></td>
</tr>
</tbody>
</table>

Result

(a) Fundus images (b) Ground-truth (c) Segmented results