SEA-Net: Squeeze-and-Excitation Attention Net for Diabetic Retinopathy Grading

Ziyuan Zhao¹, Kartik Chopra², Zeng Zeng¹*, and Xiaoli Li¹

¹ Institute for Infocomm Research (I2R), A*STAR, Singapore
² National University of Singapore, Singapore
Background

- Diabetic retinopathy (DR) is a common retinal disease that leads to blindness.

- In Singapore, around 1 out of 12 people aged from 19 to 69 years are affected by diabetes, and 43.5% among them suffer from different severity of DR *

- It Augments the blood pressure in small vessels and influence the circulatory

Challenges

Manual Inspection

- There are no early warning symptoms for DR.
- Difficulties in timely diagnosis and early treatment.
- DR grading also suffers from high intra- and inter-observer variability.
Challenges

Automatic Method

- Extracted features from photos are hand-crafted features.

- Feature localization and segmentation can not be well embedded into the whole DR detection framework.

- Most about Binary classification (DR / no DR)
SEA-Net: Squeeze-and-Excitation Attention Net

- **Attention Net** is extended from BiRA-Net* for spatial attention

- **SE blocks** are introduced to recalibrate channel-wise feature maps for fine-grained classification

* https://ieeexplore.ieee.org/document/8803074
Attention Net

- **ResNet-50** is implemented first for deep feature extraction ($I \rightarrow U$)

- Through a sequence of $1 \times 1$ convolution layers and pooling layers, the refined feature map is obtained ($U \rightarrow A$)

- The global average pooling (GAP) layer provides a receptive field of whole spatial extent

- An element-wise division is used followed by a softmax layer

- Output: $Output = GAP(A^l) \odot GAP(A^l \otimes U^l)$

<table>
<thead>
<tr>
<th>Feature Maps</th>
<th>Batch Norm 2048</th>
<th>Conv 1x1, 64</th>
<th>ReLU</th>
<th>Conv 1x1, 16</th>
<th>ReLU</th>
<th>Conv 1x1, 8</th>
<th>ReLU</th>
<th>Conv 1x1, 1</th>
<th>Sigmoid</th>
<th>Conv 1x1, 2048</th>
<th>Attention Maps A</th>
</tr>
</thead>
</table>

$A^l$ and $U^l$ are $l$-th attention map and $l$-th feature map. 
$\otimes$ and $\odot$ denote element-wise multiplication and element-wise division.
Squeeze-and-Excitation Block

- SE Block is borrowed from **Squeeze-and-Excitation Networks***

- To exploit channel dependencies and contextual information, we propose to incorporate the SE block into the proposed architecture

Squeeze-and-Excitation Block

- The positions of SE blocks in the network influence the performance of DR grading.
- To find the optimal position, we explore three different positions of SE blocks.

(a) SE Block before Attention Net

SE-AT-Net

(b) SE Block after Attention Net

AT-SE-Net

(c) SE Block with Attention Net

SEA-Net
Hybrid Loss Function

- Implement center loss to reduce the loss-accuracy discrepancy and get an improved convergence.

- The weighted cross entropy loss is used to alleviate data imbalance

\[
\mathcal{L}_{ct} = \frac{1}{2} \sum_{i=1}^{m} \| x_i - c_{y_i} \|_2^2 \quad \mathcal{L}_{ce} = \text{weight}_y \left( - \log \left( \frac{\exp(x[y])}{\sum_j \exp(x[j])} \right) \right)
\]

\[
\mathcal{L} = \mathcal{L}_{ce} + \lambda \mathcal{L}_{ct}
\]

\(\lambda\) is a scalar to control the strength of loss functions.
Dataset

- The retinal images are provided by EyePACS consisting of 35126 images. And each image is labeled as {0, 1, 2, 3, 4}, depending on the disease’s severity.

- Following the data distribution adopted by Maria A. Bravo et al, a balanced testing dataset of 1560 images was applied to our experiments for testing, and the rest were used for training.

Results

- The proposed framework outperforms other methods in all metrics.
- SE blocks are proved to be effective in the proposed methods.
- In SEA-Net, the SE blocks are placed alternatively with convolution layers, recalibrating the learned feature maps in an adaptive manner.

<table>
<thead>
<tr>
<th>Method</th>
<th>ACA</th>
<th>Marco-F1</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bravo et al.</td>
<td>50.51</td>
<td>50.81</td>
<td>-</td>
</tr>
<tr>
<td>BiRA-Net</td>
<td>54.31</td>
<td>57.25</td>
<td>-</td>
</tr>
<tr>
<td>AT-Net</td>
<td>54.42</td>
<td>49.51</td>
<td>86.99</td>
</tr>
<tr>
<td>SE-AT-Net</td>
<td>57.76</td>
<td>55.05</td>
<td>87.34</td>
</tr>
<tr>
<td>AT-SE-Net</td>
<td>5.83</td>
<td>58.92</td>
<td>87.21</td>
</tr>
<tr>
<td>SEA-Net</td>
<td>58.59</td>
<td>58.72</td>
<td>87.38</td>
</tr>
<tr>
<td>SEA-Net (λ = 0.1)</td>
<td>59.94</td>
<td>60.47</td>
<td>87.6</td>
</tr>
</tbody>
</table>
Results

- The proposed method is further improved with the proposed hybrid loss function.

- The proposed hybrid loss can learn better discriminative features, especially for confusing classes, i.e., class 0 and class 1.
Conclusion

- We proposed a novel deep learning architecture for DR grading.

- Spatial attention and channel attention are implemented to boost each other, recalibrating the attention maps adaptively.

- A hybrid loss function based on weighted cross entropy loss and center loss is implemented.

- Experimental results demonstrate the effectiveness of the proposed architecture.
THANKS

✉ Zhao_Ziyuan@i2r.a-star.edu.sg