WEAKLY-SUPERVISED LOCALIZATION OF DIABETIC RETINOPATHY LESIONS IN RETINAL FUNDUS IMAGES

JAN M. KÖHLER, WALEED M. GONDAL
BOSCH CENTER FOR ARTIFICIAL INTELLIGENCE (BCAI), STUTTGART, GERMANY

RENÉ GRZESZICK, GERNOT A. FINK
DPTM. OF COMPUTER SCIENCE, TU DORTMUND UNIVERSITY, GERMANY

MICHAEL HIRSCH
MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS, TÜBINGEN, GERMANY

CONTACT: JAN.KOEHLER@DE.BOSCH.COM
Motivation: Working on retinal images to classify for diabetic retinopathy.

- Given a retinal image we want to classify the entire image for diabetic retinopathy.
- Additionally the individual lesions in the image responsible for the classification should be highlighted to increase trust of medical experts.
- The training of the algorithm has only image-level labels, i.e. no information about pixel-wise lesions (semi-supervised object localization).

Image source: [1]
Diabetic retinopathy (DR): Some facts

- 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.

- 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%.

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

Sources: [2, 3, 4]

Jan Köhler | Bosch Center for Artificial Intelligence (BCAI) | 20/Sept/2017

© Robert Bosch GmbH 2017. All rights reserved, also regarding any disposal, exploitation, reproduction, editing, distribution, as well as in the event of applications for industrial property rights.
IEEE International Conference on Image Processing 2017

Why Bosch is in this topic? Bosch Eye Care Solution*

- Bosch India provides a non mydriatic, handheld camera for fundus imaging.
- It is easy to use, portable and suitable for mass screening.
- The main target market is mass screening in developing countries.
- Mass screening will be supported by machine learning algorithms for classification and object localization.
- Bosch Center for Artificial Intelligence (BCAI) supports with algorithmic development.

Method: Class Activation Maps with Global average pooling

1. **GAP**: From each of the K activation maps $A^k \in R^{u \times v}$ (width $u$ and height $v$) the average value is calculated:

   \[ \text{GAP}: R^{u \times v} \rightarrow R, \quad A^k \rightarrow \frac{1}{UV} \sum_x \sum_y A^k_{xy} \]

2. **Weights $w^c_k$**: The GAP layer is fully connected to output neurons via $w^c_k$, $c \in \{\text{healthy, unhealthy}\}$

   \[ y^c = \sum_k w^c_k \text{GAP}(A^k) \]

   Weights $w^c_k$ encode the importance of each feature map $A^k$ with respect to class $c$.

3. **Training**: The network is trained in a weakly-supervised fashion, i.e. only labels on image level (healthy, unhealthy) are available. No label information about any lesion type is given.

Based on Zhou et al. „Learning Deep Features for Discriminative Localization“ [5]
Method: Class Activation Maps with Global average pooling

4. **Localization map L**: Given an image, the weighted sum of the activation maps of the last convolutional layer forms the localization map.

\[ L^c = \sum_k w_k^c A^k \]

5. **Upscaling**: The localization map is bilinear upsampled to image resolution

6. **Overlaying**: The original image is overlaid with the localization map

Based on Zhou et al., „Learning Deep Features for Discriminative Localization“ [5]
From localization map to True positive / False positive / false negative

- Localization Map
- Prediction Map
- Ground Truth Map
- Resultant Map

All regions > 0.65 of maximum value are marked as a lesion.

Overlapped region (TP)
Falsely predicted region (FP)
Region not covered (FN)
The network structure

- Network and training configuration:
  - VGG-16 based architecture
  - Batch normalization
  - L2 Regularization with weight decay factor of 0.0005
  - Gradient descent with momentum 0.8
  - Initial learning rate of 0.01 decayed by 1% after each epoch
  - 150 epochs for training
Experiments: Data sets used

Training data set:
Kaggle data set on diabetic retinopathy [6]

- 88,702 images of which 80% are used for training and 20% for validation
- No information about lesions given
- Collected in a clinical setting with high-end, stationary equipment.
- Five stages of diabetic retinopathy: For classification the first two classes were grouped into non-referable DR and the remaining three classes into referable DR.

Test data set:
DiaretDB1 data set [1]

- 89 high resolution images used for testing
- Lesions marked by four experts.
- Regions with more than 75% consensus among the experts are considered as positive.
Results for object localization on diaretDB1 – example images

Ground truth:
- Yellow -- Hard Exudates
- Blue -- Soft Exudates
- Red -- Hemorrhages

Our result:
- Green -- Our predicted lesion regions

Results for object localization on diaretDB1 – example images


Ground truth: Yellow -- Hard Exudates  Blue -- Soft Exudates  Red -- Hemorrhages
Our result: Green -- Our predicted lesion regions
IEEE International Conference on Image Processing 2017
Results for object localization on diaretDB1 – example images


Ground truth: Yellow -- Hard Exudates  Blue -- Soft Exudates  Red -- Hemorrhages
Our result: Green -- Our predicted lesion regions
Results for object localization on diaretDB1 – example images


Ground truth: Yellow -- Hard Exudates  Blue -- Soft Exudates  Red -- Hemorrhages
Our result: Green -- Our predicted lesion regions
Lesion detection - performance at image level

Image level sensitivity [%]

<table>
<thead>
<tr>
<th>Method</th>
<th>Type</th>
<th>H*</th>
<th>HE*</th>
<th>SE*</th>
<th>RSD*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhou et al.[7]</td>
<td>S.</td>
<td>94.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Liu et al.[8]</td>
<td>S.</td>
<td>-</td>
<td>83.0</td>
<td>83.0</td>
<td>-</td>
</tr>
<tr>
<td>Haloi et al.[9]</td>
<td>S.</td>
<td>96.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mane et al.[10]</td>
<td>S.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>96.4</td>
</tr>
</tbody>
</table>

Ours (50% Overlap) | W. S. | 97.2 | 93.3 | 81.8 | 50
Ours (OnePixel Overlap) | W. S. | 97.2 | 100  | 90.9 | 50

** S. = supervised, W. S. = weakly-supervised

Image level sensitivity

- Detecting at least one lesion of type T on an image is counted as a True Positive.
- A lesion is detected if either one pixel or 50% of its pixels are covered by the prediction map.
- Sensitivity = \[\frac{\text{Number of images lesion type } T \text{ is detected}}{\text{Number of total images showing lesion type } T}\]

Binary image classification (healthy vs. unhealthy) yields 93.6% sensitivity and 97.6% specificity on DiaretDB1 dataset with AUC of 0.954.
Performance at lesion level

Lesion level sensitivity

- Detecting at least one pixel (OnePixel Overlap) or 50% of the pixels (50% Overlap) of a lesion of type T is counted as a True Positive.
- A False Positive (FP) is a predicted region not containing any lesion type or a predicted region with mIOU<0.5.
- Sensitivity = \( \frac{\text{Number of detected regions of lesion type } T}{\text{Number of total regions of lesion type } T} \)

<table>
<thead>
<tr>
<th>Type</th>
<th>Method</th>
<th>Hemorrhages</th>
<th>Hard Exudates</th>
<th>Soft Exudates</th>
<th>RSD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SE %</td>
<td>SE %</td>
<td>SE %</td>
<td>SE %</td>
</tr>
<tr>
<td>W.S.</td>
<td>Quellec et al. [11]</td>
<td>71</td>
<td>80</td>
<td>90</td>
<td>61</td>
</tr>
<tr>
<td>S.</td>
<td>Dai et al. [12]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>29</td>
</tr>
<tr>
<td>W.S.</td>
<td>Ours (50% Overlap)</td>
<td>72</td>
<td>47</td>
<td>71</td>
<td>21</td>
</tr>
<tr>
<td>W.S.</td>
<td>Ours (OnePixel Overlap)</td>
<td>91</td>
<td>87</td>
<td>89</td>
<td>52</td>
</tr>
</tbody>
</table>

Lesion level sensitivity [\%]

FROC curves for all four lesion types
Summary

- Given only image level labels we could identify lesion regions which are important for the CNN for classification of retinal images.
- The network is trained for binary classification and classification accuracy is high, though introducing a GAP layer, with sensitivity of 93.6% and specificity of 97.6% on the DiaretDB1 test data.
- The sensitivity for detecting the lesion regions is beating or competitive to supervised methods.
- Red small dots are hard to localize. A reason might be the small resolution of the feature maps which are the basis of the localization map.
IEEE International Conference on Image Processing 2017

References


Thank you

Remarks?

Questions?

Contact: Jan.Koehler@de.bosch.com