Diagnosing diabetic retinopathy from images of the eye fundus

Lucy Wang, Amelie Schäfer {lwang8, amesch}@stanford.edu

https://youtu.be/PzuAppZm9R8

Introduction
- Diabetic retinopathy (DR) is leading cause of blindness in working aged adults
- Need for regular screening without specialist in rural areas
- Use convolutional neural network to classify eye fundus images into 5 disease stages

Dataset
- 3662 labeled 3-channel images from APTOS 2019 Blindness detection challenge on Kaggle [1]
- 5 classes for different disease stages
- Imbalanced class prevalence
- Split into train/dev/test: 70/15/15 preserving class distribution

Model
- MobileNetV2
- Avg Pooling
- FC 5 neurons
- Softmax
- 128 x 128 raw/preprocessed
- 224 x 224 raw/preprocessed

Methods
- Categorical cross entropy loss
- Adam optimizer
- Categorical accuracy
- Quadr. Weighted kappa (QWKP)
- Learning rate
- Image size, image preprocessing [4]
- Unfreeze weights
- Weighted loss
- Data augmentation
- 3-channel image
- 128 x 128 raw/preprocessed
- 224 x 224 raw/preprocessed

Model optimization
- Model weights frozen
- LR: 0.0001
- Categorical accuracy
- Training set
- 0.8246
- Development set
- 0.7267

- Model weights unfrozen
- LR: 0.0001
- Categorical accuracy
- Training set
- 0.9767
- Development set
- 0.9875

- Weighted loss
- LR: 0.0001
- Accuracy QWKP
- Training set
- 0.6324
- Development set
- 0.5827

- Data augmentation
- Rotation, flipping
- Preprocessed 224 x 224
- LR: 0.0001
- Weighted loss
- Accuracy QWKP
- Training set
- 0.5827
- Development set
- 0.4873

Results

<table>
<thead>
<tr>
<th></th>
<th>Categorical accuracy</th>
<th>Categorical accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training set</td>
<td>Development set</td>
</tr>
<tr>
<td>Initial</td>
<td>LR: 0.001</td>
<td>LR: 0.001</td>
</tr>
<tr>
<td>Raw images 224 x 224</td>
<td>0.8246</td>
<td>0.7267</td>
</tr>
<tr>
<td>Model weights frozen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw images 224 x 224</td>
<td>0.9767</td>
<td>0.9875</td>
</tr>
<tr>
<td>Unfreeze model weights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw images 224 x 224</td>
<td>0.9767</td>
<td>0.9875</td>
</tr>
<tr>
<td>Weighted loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw images 224 x 224</td>
<td>0.6324</td>
<td>0.5827</td>
</tr>
<tr>
<td>Data augmentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR: 0.0001</td>
<td>0.8066</td>
<td>0.8066</td>
</tr>
<tr>
<td>Accuracy QWKP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training set</td>
<td>0.5827</td>
<td>0.5827</td>
</tr>
<tr>
<td>Development set</td>
<td>0.4873</td>
<td>0.4873</td>
</tr>
</tbody>
</table>

Conclusion
- Challenges: small amount of data, class imbalance
- Strategies: transfer learning, data augmentation, weighted loss
- Performance comparable to lower field of kaggle competitors

Future work
- Cropping background
- L2 regularization
- More data augmentation
- Deeper analysis of saliency maps and class activation maps
- Collect more data

References

Model interpretation
- Saliency maps

Desirable behavior
- No DR
- Mild DR
- Moderate DR
- Severe DR
- Proliferate DR

Undesirable behavior
- No DR
- Mild DR
- Moderate DR
- Severe DR
- Proliferate DR

Desirable behavior

Undesirable behavior

No DR
Mild DR
Moderate DR
Severe DR
Proliferate DR