OCT Image Quality Assessment with Joint Classification and Segmentation

Wu Yushan[1,2], Christina Bornberg[1], Jacqueline Chua[1], Leopold Schmetterer[1], Damon Wong[1]

[1] Singapore Eye Research Institute, Singapore National Eye Centre, [2] The Chinese University of Hong Kong

Purpose

- This work aimed to evaluate the use of deep learning-based multi-task models for the automated evaluation of image quality in optical coherence tomography (OCT) enface images of the optic nerve head.

Dataset

- 1291 enface images were obtained from a prototype SS-OCT system (Zeiss PlexElite SS-OCT) with an optic nerve head (ONH) centered 6mm x 6mm imaging protocol.
- Image quality is assessed according to a grading system that factored in the presence of artifacts and the overall strength of the signal.
- The grading scale comprised four distinct levels ranging from Grade 0 (an unusable image) to Grade 3 (an image with no quality concerns), with class sizes of 358, 327, 257, and 349 images for Grades 0 to 3, respectively.
- Segmentation masks consisting of locations of the ONH and regions with signal strength reductions (SSR) not attributed to pathology (consisting of "shadow" and "break") were manually delineated by a human grader.
- Segmentation masks includes 5 categories, i.e. background, optic nerve head, signal loss artefacts (shallow and dark), and motion artefacts.

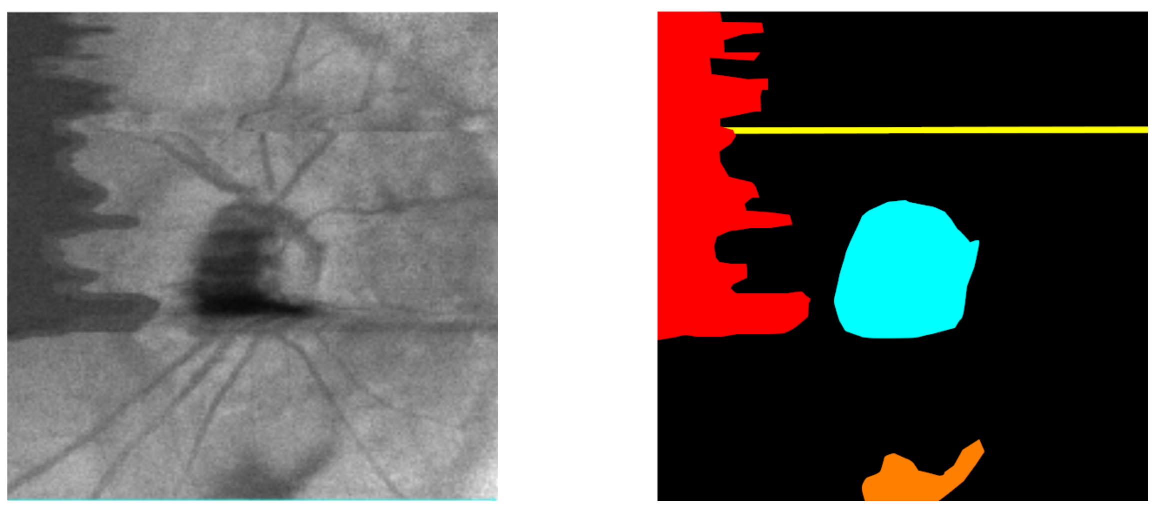

Figure 1. Examples of OCT enface image with the corresponding segmentation mask

Methods

- The dataset was split into 905 images for training, 193 images for validation and 193 images for testing.
- Image quality assessment was based on segmentation model frameworks, e.g. FPN and U-Net, with encoder backbone implemented using various well-eatablished convolution neural networks (CNNs), e.g. ResNet and EfficientNet, to extract image features.[1, 2, 3, 4]
- An additional classification head of was added to the pipeline to record the image quality classification results, together with the original output of predicted segmentation masks.

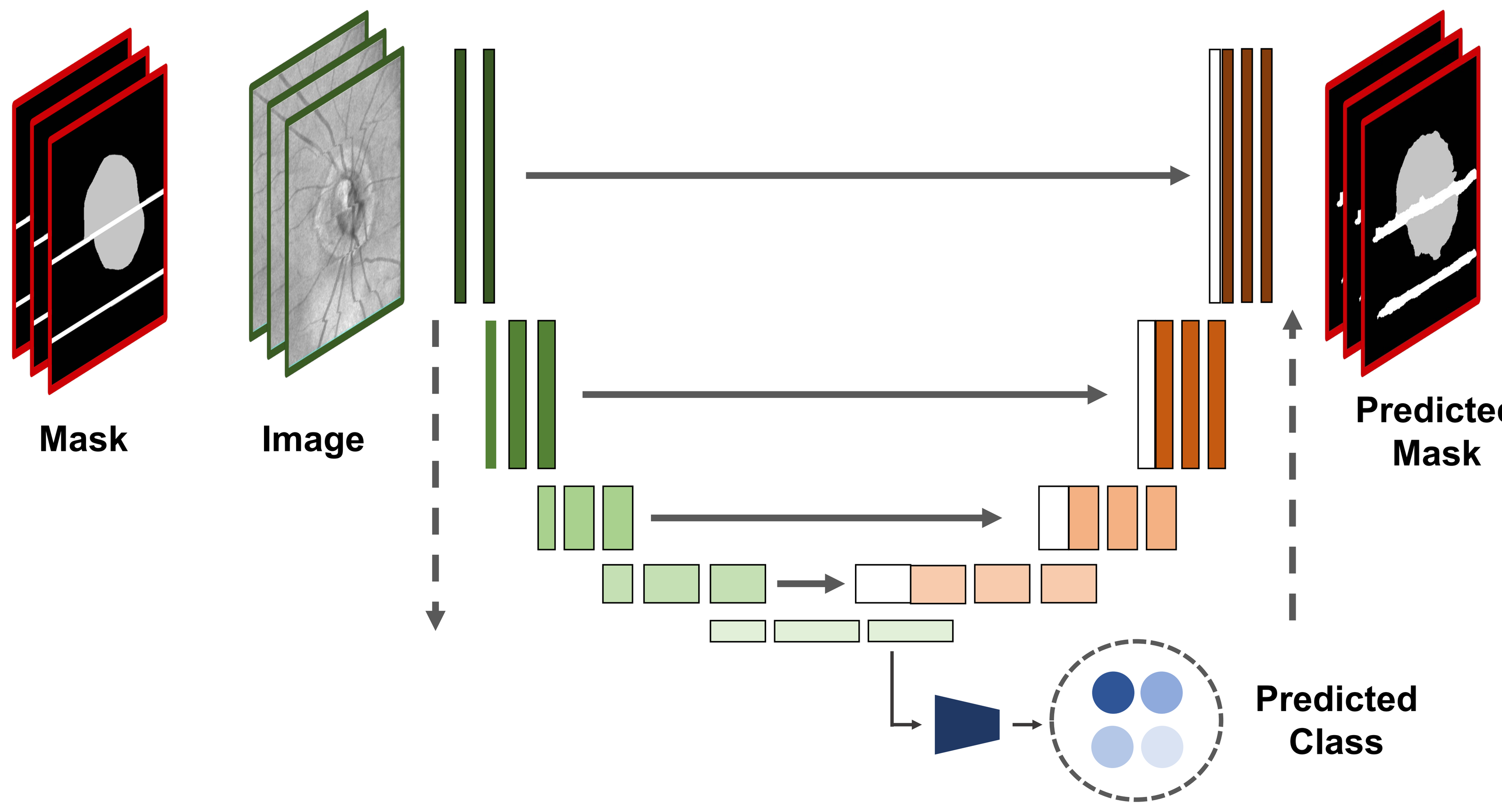

Figure 2. Illustration of multi-task pipeline incorporating segmentation and classification (U-Net as segmentation structure)

- Transfer learning from pre-trained models was utilized by minor revision on model structure and fine-tuning of the model weights using the training data.
- Various neighboring class combinations were explored to effectively retain correlation information between classes.
- Gradient-weighted class activation mapping (Grad-CAM) was used to provide visual explanations of important areas focused by the model.[5]

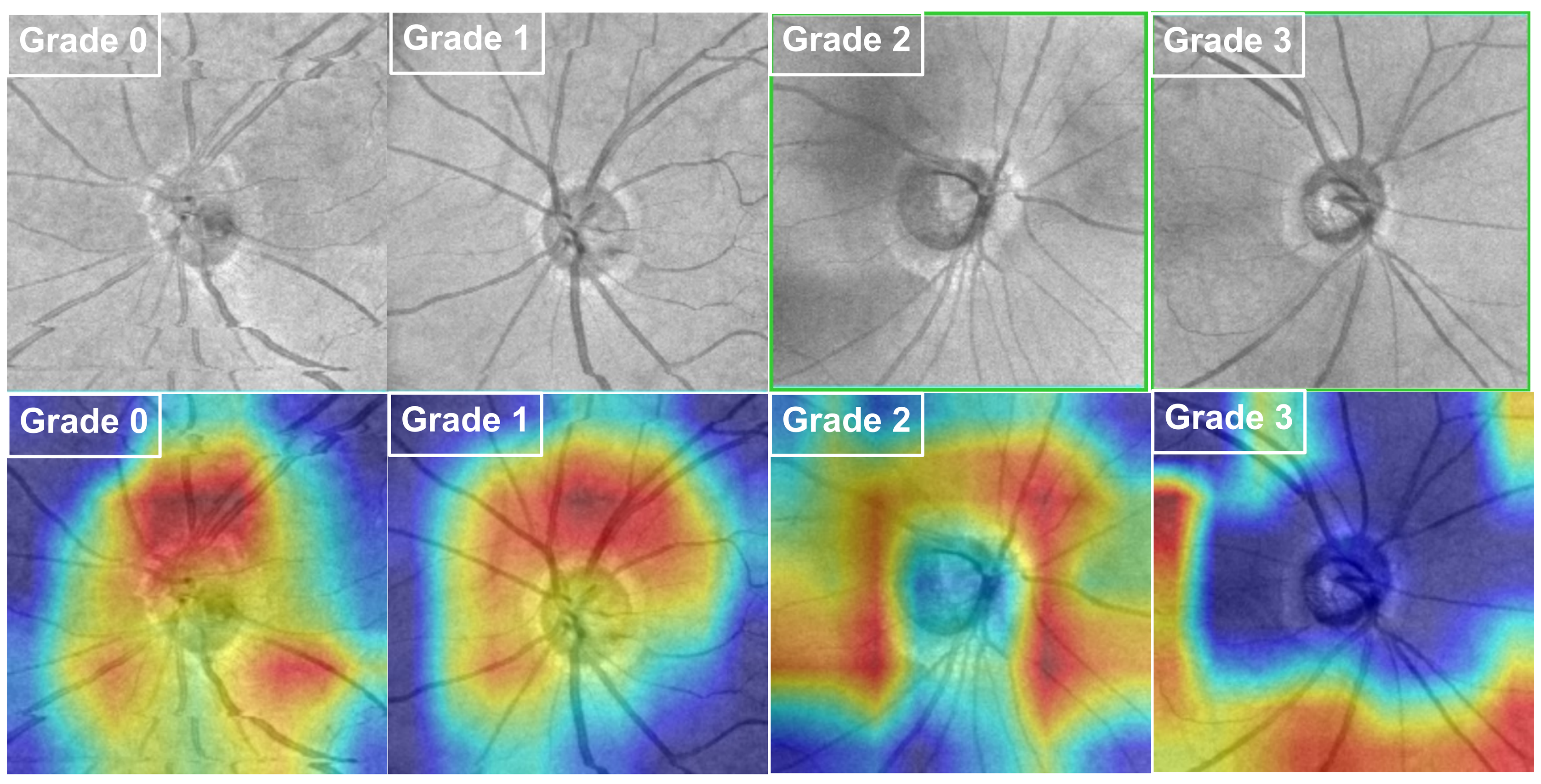

Figure 3. Top Row: Sample OCT images of different quality grades, with Grade 0 indicating poor quality and Grade 3 indicating optimal quality. Bottom Row: Corresponding Grad-Cam images

Results

- Amalgamating proximate classes to reduce the 4-class problem into two binary groupings with grades 0 to 2 as one group, and grade 3 as another, yielded improvement in the performance.
- Performance of multi-task models with joint classification and segmentation generally exceeds the performance of classification models without incorporating the segmentation task.
- Output segmentation masks from the pipeline showed successful detection of the optic nerve head and signal loss areas ("shadow"), but difficulty still existed in accurately locating motion artefacts due to their inconspicuousness.

Table 1: Summary of performance with different class grouping and model structure on test set

Image Class Grouping*	Model^	F1-score**
4 classes	EfficientNet-B0	91.13%
4 classes	FPN + ResNext101	93.80%
2 classes	EfficientNet-B0	96.10%
2 classes	FPN + ResNext101	98.10%

** Original grading of image quality was provided as 4 classes. "2 classes" refers to grouping grades 0 to 2 as one group representing any image quality deficiency and grade 3 as another representing no image quality deficiency*
^ EfficientNet-B0: single-task classification model; FPN + ResNext101: multi-task segmentation and classification model
*** F1-score is the harmonic mean of the precision and recall, evaluating the class-wise performance in multi-class scenarios.*

Conclusion

- Deep learning classification techniques can be used for OCT image quality assessment and regularization with the segmentation task of image artifacts can enhance classification performance.

References

1. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer International Publishing, 2015.
2. Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
3. He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
4. Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." *International conference on machine learning*. PMLR, 2019.
5. Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.