# AUTOMATED IMAGE QUALITY ASSESSMENT FOR ANTERIOR SEGMENT OPTICAL COHERENCE TOMOGRAPH

*Boyu Chen[1], Ameenat L. Solebo[2], Paul Taylor[1]*

[1]Institute of Health Informatics, University College London, UK
[2]Great Ormond Street Institute of Child Health, University College London, UK

## ABSTRACT

Optical Coherence Tomography (OCT) is a technique for diagnosing eye disorders. Image quality assessment (IQA) of OCT images is essential, but manual IQA is time consuming and subjective. Recently, automated IQA methods based on deep learning (DL) have achieved good performance. However, few of these methods focus on OCT images of the anterior segment of the eye (AS-OCT). Moreover, few of these methods identify the factors that affect the quality of the images (called "quality factors" in this paper). This could adversely affect the acceptance of their results. In this study, we define, for the first time to the best of our knowledge, the quality level and four quality factors of AS-OCT for the clinical context of anterior chamber inflammation. We also develop an automated framework based on multi-task learning to assess the quality and to identify the existing of quality factors in the AS-OCT images. The effectiveness of the framework is demonstrated in experiments.

***Index Terms***— Optical coherence tomograph, Image quality assessment, Multi-task learning, Deep learning

## 1. INTRODUCTION

Optical coherence tomography (OCT) is widely used in ophthalmic research and healthcare for eye disorders. More recently, it has been shown to provide a novel disease metric for anterior uveitis, a significant cause of morbidity, enabling noninvasive objective quantification of the presence of inflammatory cells in the front of the eye [1, 2].

The utility of OCT images is dependent on their quality, which could be negatively impacted by machine or subject-related factors [3]. Manual image quality assessment (IQA) is time-consuming and subjective, and whilst automated IQA addresses these limitations. As traditional IQA methods for OCT based on objective image parameters [4-6] may fail to address the issues of fidelity and suitability to clinical context [7, 8], recent studies [7-10] apply deep learning (DL) in this task.

However, these DL methods are designed for retinal OCT, which captures images from the posterior part of the eye to detect eye disorders such as diabetic retinopathy and age-related macular degeneration. Few DL methods have focused on IQA of AS-OCT. AS-OCT captures images from the anterior segment of the eye, where diseases such as anterior uveitis occur [1, 2]. It can also be used for the angle-closure assessment for glaucoma evaluation [11, 12].

Moreover, few of these DL methods identify the factors that affect image quality (henceforth termed "quality factors"). Unlike an ophthalmologist, "black-box" methods trained on OCT images offer no explanation of their results, which could adversely affect their acceptance [13]. Thus, it could be helpful to show the quality factors when IQA methods provide an assessment result.

To that end, we develop a novel definition for image quality level and quality factors for AS-OCT for the clinical context of the detection of anterior chamber inflammation. We also develop an automated framework to assess image quality and identify these quality factors. This framework uses multi-task learning to obtain knowledge of image quality and quality factors from AS-OCT images at the same time. Experiments show that learning knowledge of quality factors can improve the performance of IQA.

## 2. METHOD

### 2.1. Image Quality Level and Quality Factors

Retinal image quality characteristics are anchored in anatomical context which is missing in images of other anatomical areas such as the anterior chamber. For example, artefacts caused by eyelash may appear in the image of AS-OCT but will not appear in retinal OCT. To that end, we define three quality levels and four quality factors for AS-OCT for the clinical context of the detection ('ruling in' and 'ruling out') and quantification of anterior chamber inflammation.

The three quality levels are: '**Good**', '**Limited**' and '**Poor**'. 'Good' (Fig.1.a) is an image with a faithful view of the whole anterior chamber of the eye, without image artefact. 'Limited' (Fig.1.b) is an image with quality of reduced fidelity, but which still contains view sufficient to detect critical features of clinical interest. These images can be used to 'rule in' disease, but they cannot be used to 'rule

out' disease as inflammatory features may be obscured by artefact or missing due to cropping. 'Poor' (Fig.1.c) is an image with quality that cannot be used, due to the absence of areas free from artefact.

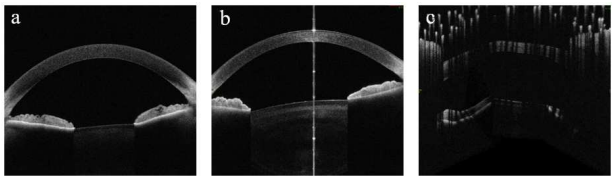


**Fig. 1.** Examples of three AS-OCT images with different quality level.

The four quality factors are: **1)** '**Eyelash'** (Fig.2.a) that is artefactual hyper and hyporeflective signal within the anterior chamber created by eyelash. **2)** '**Glare**' (Fig.2.b) that is a hyperreflective line of signal which occurs due to the fixation light from the AS-OCT camera. **3)** '**Left-cropped**' (Fig.2.c) and **4)** '**Right-cropped**' (Fig.2.d) that describe the anterior chamber cross sectional image is cropped at the left or right edge of the image. Images with good quality do not contain any quality factors, while limited or poor images contain at least one quality factor.

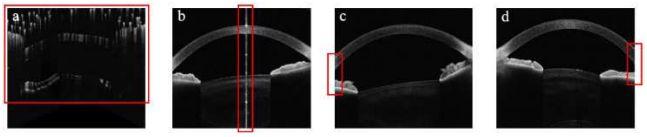


**Fig. 2.** The examples of four quality factors.

### 2.2. Automated Framework for AS-OCT IQA

We propose an automated IQA framework based on multi-task learning (MTL) [14] to assess the quality and identify the quality factors for AS-OCT images. To inject the knowledge of image quality from quality factors into our framework, we apply hard parameter sharing MTL [14]. It uses shared hidden layers to obtain shared knowledge from all the tasks (i.e., IQA and identifying different quality factors) and uses different output layers for separate tasks. Our framework (Fig.3) consists of feature extraction module based on shared hidden layers and assessment module based on output layers for separate tasks.

Feature extraction module is flexible and uses different convolutional neural networks (CNN), such as VGG [15], ResNet [16] and DenseNet [17], as feature extractor to learn shared knowledge from images. It outputs the feature maps extracted by the feature extractor. Assessment module classifies the quality level and identifies quality factors of the AS-OCT images. It flattens the extracted feature maps and send them into different fully connected (FC) layer blocks, each of which has 128 neural units. Among these blocks, one is used to predict the image quality, and its loss ($Loss_{quality}$) is multi-class cross entropy loss. The other $N$ blocks are used to identify the existing of quality factors, and the loss of the i-th block ($Loss_i$) is binary cross entropy loss. The final loss of the framework is:

$$Loss = w_{quality} * Loss_{quality} + (1 - w_{quality})/N * (\sum Loss_i) \quad (1)$$

where $w_{quality}$ is set as 0.6, empirically.

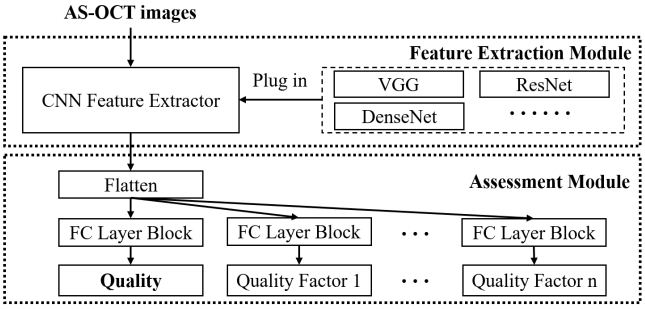


**Fig. 3.** The architecture of the IQA framework.

### 2.3. Implementation

Our framework is implemented in PyTorch. Its implementations of VGG16, ResNet18, ResNet34, ResNet50, and DenseNet121 are used as feature extractors in the feature extraction module. Image data are divided into training, validation, and testing sets (70%, 15%, and 15%, respectively). We ensure the ratios of quality labels are the same over the three sets. All images are resized to 256*256 pixels. Rotation and brightness are used as data augmentation. Adam optimizer is used to optimize the framework. Learning rate, batch size and epoch are 0.001, 64 and 200, respectively. To balance the number of quality labels in each training epoch, $N_{smaple}$ (set as 512 empirically) images of good, limited and poor labels are randomly extracted (with replacement) from training set, respectively, and combined as the training data in each epoch. To encourage the framework to learn general knowledge of quality, the framework is only trained based on pure IQA task in the first $P$ (set as 10 empirically) epochs. This means that only the feature extractor and the FC layer block for predicting quality can be optimized, and the loss is $Loss_{quality}$ only. After $P$ epochs, the framework uses the $Loss$ in equation (1) as the loss to optimise all the blocks. Early stopping is applied to avoid overfitting. The training will be halted if the loss value in validation set does not decrease after 30 epochs.

**Table 1.** The number/proportions of quality factors in the limited, poor, and total AS-OCT images.

| Quality Factor | Limited (1827) | Poor (405) | Total (2825) |
|---|---|---|---|
| Eyelash | 730/40.0% | 291/71.9% | 1021/36.1% |
| Glare | 177/9.7% | 6/1.5% | 183/6.5% |
| Left-cropped | 350/19.2% | 6/1.5% | 356/12.6% |
| Right-cropped | 1009/55.2% | 34/8.4% | 1043/36.9% |

## 3. EXPERIMENTS AND RESULTS

### 3.1. Dataset

The experimental dataset includes 2,825 AS-OCT images (593 '**Good**', 1,827 '**Limited**' and 405 '**Poor**'), from 45 eyes of 24 patients, collected from a specialist hospital in London. Each image was annotated by at least two senior ophthalmic clinicians. They then came to a final agreement for the label. Table 1 shows the number and proportions of each quality factor in limited, poor and total images.

### 3.2. Results

To show that learning knowledge of quality factors can improve IQA, we train the framework with MTL (trained for the IQA and quality factor identification tasks) and without MTL (purely trained for the IQA task). To obtain a robust result, we repeat the experiment 10 times. The training, validation and testing sets are randomly split from original dataset in each round. As IQA is multi-class classification, we use macro-averaged precision, recall and F1 as evaluation metrics in each round and calculate the average result in the testing datasets over 10 rounds. Table 2 shows the IQA performance of the framework (with and without MTL) based on different feature extractors.

**Table 2.** The IQA performance with and without MTL.

| Model | MTL | Precision | Recall | F1 |
|---|---|---|---|---|
| VGG16 | No | 0.853 | 0.856 | 0.852 |
| | Yes | **0.872** | **0.895** | **0.881** |
| ResNet18 | No | 0.861 | **0.876** | 0.865 |
| | Yes | **0.864** | 0.873 | **0.866** |
| ResNet34 | No | 0.856 | **0.865** | 0.854 |
| | Yes | **0.863** | 0.860 | **0.855** |
| ResNet50 | No | 0.844 | 0.846 | 0.836 |
| | Yes | **0.861** | **0.850** | **0.847** |
| DenseNet121 | No | 0.839 | 0.855 | 0.831 |
| | Yes | **0.863** | **0.876** | **0.865** |

Table 2 shows that VGG16 achieves the best performance. It reaches macro-averaged precision, recall, and F1 of 0.872, 0.895 and 0.881, respectively. We use it as an example model to further show the performance of our framework. Table 3 shows the averaged result over 10 rounds in the testing datasets for predicting each quality level and each quality factor (binary classification). We use precision, recall and F1 for binary classification as evaluation metrics. Table 4 shows the confusion matrix of the quality level classification (averaged result over 10 rounds in the testing datasets).

### 3.3. Discussion

Table 2 indicates that learning knowledge of quality factors can improve IQA performance. In most cases, models based on MTL outperform equivalent models without MTL. This indicates that future study can consider finding relevant "quality factors" to improve the IQA performance.

**Table 3.** The performance of VGG16 for predicting each quality level and each quality factor.

| Label | | Precision | Recall | F1 |
|---|---|---|---|---|
| **Quality Level** | Good | 0.801 | 0.912 | 0.851 |
| | Limited | 0.943 | 0.895 | 0.918 |
| | Poor | 0.872 | 0.879 | 0.874 |
| **Quality Factor** | Eyelash | 0.947 | 0.865 | 0.904 |
| | Glare | 0.870 | 0.788 | 0.819 |
| | Left-cropped | 0.865 | 0.831 | 0.841 |
| | Right-cropped | 0.949 | 0.898 | 0.921 |

Table 3 shows that our framework based on VGG16 has higher recall of 'Good' (0.912) compared with 'Limited' and 'Poor'. Table 4 provides more information about this result. Interestingly, no image labelled as 'Good' is predicted as 'Poor', and no image labelled as 'Poor' is predicted as 'Good'. The results could indicate that our framework has the potential to support clinical experts to retrieve images with good quality from AS-OCT dataset. Some studies (e.g., [18, 19]) screen suitable medical images based on traditional image-based parameters such as brightness, contrast, and blurriness. However, these parameters cannot be directly applied to screen AS-OCT images used for detecting anterior chamber inflammation. Table 3 also shows that our framework is able to identify the quality factors. The F1 of predicting 'Eyelash' and 'Right-cropped' achieve 0.904 and 0.921, respectively. The F1 of predicting 'Glare' and 'Left-cropped' are relatively low as they only represent 6.5% and 12.6% of the original dataset.

**Table 4.** The confusion matrix for predicting the quality level based on VGG16.

| Label | Predicted by the Framework | | |
|---|---|---|---|
| | Good | Limited | Poor |
| Good | 81.2 | 7.8 | **0.0** |
| Limited | 20.7 | 246.2 | 8.1 |
| Poor | **0.0** | 7.4 | 53.6 |

Our study has some limitations. We concluded and defined four quality factors from our dataset, but we believe the categories can be further extended. Although it may not be appropriate to directly extrapolate these factors to other AS-OCT data acquired for different clinical purposes (e.g., to examine corneal health or iris disorders), we believe these factors are likely to appear in other AS-OCT datasets. We also believe these factors could provide a novel insight for future study related to the IQA of AS-OCT.

Unlike IQA in natural images (e.g., [20, 21]) which have various data sources for evaluation, IQA datasets for medical image are usually limited. We have set up https://www.zooniverse.org/projects/lolasolebo/eyes-on-eyes, a Citizen Science project [22] that encourages the public to access and annotate our AS-OCT data.

Our framework looks at B scans rather than the volumes of an eye. This reflects the current practice of scan (rather than volume) based image analysis in AS-OCT [1, 2, 5, 11]. However, future work may need to consider developing cumulative quality score for volume scans.

VGG, even though achieve best performance in the framework, is slower than more modern alternative. Our framework is flexible and can make use of a variety of feature extractors. Better extractors can also be plugged in to the framework. This will be addressed in our future work.

## 4. CONCLUSION

Few DL-based IQA methods for OCT focus on AS-OCT. Moreover, few of these methods aim to identify the factors that affect the quality in the images, which could adversely affect the acceptance of results. This study aims to tackle these two problems. The contributions of this research are: (1) We define, for the first time to our best knowledge, the quality level and the quality factors for AS-OCT images for the clinical context of anterior chamber inflammation; (2) We propose an automated IQA framework that can assess image quality and identify the existing of quality factors for AS-OCT images. Experiments show that learning knowledge of quality factors can improve IQA performance, and that this framework could support clinical experts to screen suitable AS-OCT images for diagnosis or study.

## 5. ACKNOWLEDGEMENTS

We thank the children who provided data for this study and their families, and Mr. Harry Petrushkin, Consultant Ophthalmologist at Moorfields Eye Hospital, for his support in recruiting children to the image acquisition study.

## 6. REFERENCES

[1] X. Liu *et al.*, "Instrument-based tests for measuring anterior chamber cells in uveitis: a systematic review," *Ocular immunology and inflammation,* vol. 28, no. 6, pp. 898-907, 2020.

[2] S. Akbarali *et al.*, "Imaging‐Based uveitis surveillance in juvenile idiopathic arthritis: feasibility, acceptability, and diagnostic performance," *Arthritis & Rheumatology,* vol. 73, no. 2, pp. 330-335, 2021.

[3] L. Tanga *et al.*, "Evaluating the effect of pupil dilation on spectral-domain optical coherence tomography measurements and their quality score," *BMC ophthalmology,* vol. 15, no. 1, pp. 1-5, 2015.

[4] P. Barnum, M. Chen, H. Ishikawa, G. Wollstein, and J. Schuman, "Local quality assessment for optical coherence tomography," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2008: IEEE, pp. 392-395.

[5] S. I. Niwas *et al.*, "Complex wavelet based quality assessment for AS-OCT images with application to angle closure glaucoma diagnosis," *Computer Methods and Programs in Biomedicine,* vol. 130, pp. 13-21, 2016.

[6] D. Stein *et al.*, "A new quality assessment parameter for optical coherence tomography," *British Journal of Ophthalmology,* vol. 90, no. 2, pp. 186-190, 2006.

[7] J. Wang *et al.*, "Deep learning for quality assessment of retinal OCT images," *Biomedical optics express,* vol. 10, no. 12, pp. 6057-6072, 2019.

[8] M. Zhang, J. Y. Wang, L. Zhang, J. Feng, and Y. Lv, "Deep residual-network-based quality assessment for SD-OCT retinal images: preliminary study," in *Medical Imaging 2019: Image Perception, Observer Performance, and Technology Assessment*, 2019, vol. 10952: SPIE, pp. 269-274.

[9] J. Y. Wang, L. Zhang, M. Zhang, J. Feng, and Y. Lv, "Deep convolutional network based on rank learning for OCT retinal images quality assessment," in *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 2019, vol. 10953: SPIE, pp. 23-28.

[10] R. Wang *et al.*, "OCT image quality evaluation based on deep and shallow features fusion network," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020: IEEE, pp. 1561-1564.

[11] H. Fu *et al.*, "Age challenge: angle closure glaucoma evaluation in anterior segment optical coherence tomography," *Medical Image Analysis,* vol. 66, p. 101798, 2020.

[12] J. Hao *et al.*, "Hybrid variation-aware network for angle-closure assessment in As-Oct," *IEEE Transactions on Medical Imaging,* vol. 41, no. 2, pp. 254-265, 2021.

[13] A. Singh, J. J. Balaji, M. A. Rasheed, V. Jayakumar, R. Raman, and V. Lakshminarayanan, "Evaluation of Explainable Deep Learning Methods for Ophthalmic Diagnosis," *Clinical Ophthalmology (Auckland, NZ),* vol. 15, p. 2573, 2021.

[14] S. Vandenhende, S. Georgoulis, W. Van Gansbeke, M. Proesmans, D. Dai, and L. Van Gool, "Multi-task learning for dense prediction tasks: A survey," *IEEE transactions on pattern analysis and machine intelligence,* 2021.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.

[18] R. Han *et al.*, "Validating automated eye disease screening AI algorithm in community and in-hospital scenarios," *Frontiers in Public Health,* vol. 10, 2022.

[19] F. Li *et al.*, "A multicenter clinical study of the automated fundus screening algorithm," *Translational Vision Science & Technology,* vol. 11, no. 7, pp. 22-22, 2022.

[20] J. Wu and X. Di, "Integrating neural networks into the blind deblurring framework to compete with the end-to-end learning-based methods," *IEEE Transactions on Image Processing,* vol. 29, pp. 6841-6851, 2020.

[21] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 30, no. 1, pp. 36-47, 2018.

[22] M. L. Jones and H. Spiers, "The crowd storms the ivory tower," *Nature methods,* vol. 15, no. 8, pp. 579-580, 2018.