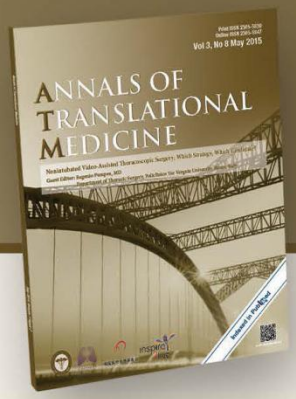
ANNALS OF TRANSLATIONAL MEDICINE



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

Peer Review File

Article information: http://dx.doi.org/10.21037/atm-20-7436

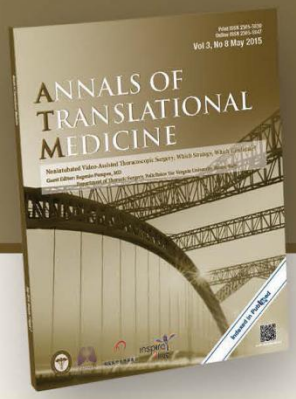
Reviewer Comments

Zheng et al. developed and compared the performance of two DL algorithms, one supervised (using traditional methods) and one semi-supervised (using generative adversarial networks, GANs), to detect angle closure (defined as iridotrabecular contact, ITC) in anterior segment OCT (AS-OCT) images. Reference labels of angle closure were provided by three glaucoma specialists on a large AS-OCT image dataset, although only a small subset of labelled data was used in model training. Both models were tested on two small independent datasets, with the GANs semi-supervised model producing marginally better performance. Overall, the premise of the study is intriguing, and its results could be relevant for developing future deep learning models. However, I have questions about study methodology and suggestions to strengthen the study's results.

Major points:

1. The premise of this study is intriguing, that a model developed using small labeled dataset and large unlabeled dataset can produce better performance than using the small labeled dataset alone. However, it would have made sense to also train a supervised model using all of the data, especially since it was labeled by 3 graders. This would allow the reader to better contextualize the performance of the semi-supervised model and the benefits of this approach.

ANNALS OF TRANSLATIONAL MEDICINE



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

Response: We agree with the reviewer that it would have made sense to also train a supervised model using all of the data. This would allow the reader to better contextualize the performance of the semi-supervised model. We have added a deep learning model which trained on full datasets (33,864 ACA images in total). We have incorporated results in Table 2 and elaborated on these in the abstract, methods, results and discussion section. The figure 1 was revised as well.

Changes in the text:

Under “Abstract” section, page 3, line 59, the following sentence was modified.

- “We further developed two supervised deep learning (DL) models training on the same supervised dataset and the whole dataset separately.”

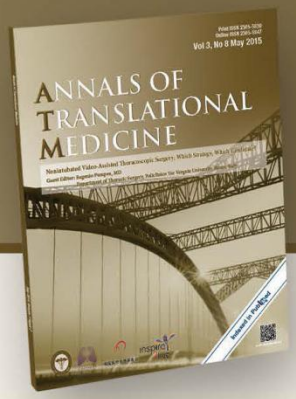
Under “Abstract” section, page 3, line 76, the following sentence was modified.

- “For closed-angle detection using clinician grading of AS-OCT imaging as the reference standard, the semi-supervised GANs model showed comparable performance, with AUCs of 0.97 (95% CI, 0.96-0.99) and 0.98 (95% CI, 0.94-1.00), compared with the supervised DL model (using the whole dataset) (AUCs of 0.97 [95% CI, 0.96-0.99], and 0.97 [95% CI, 0.94-1.00]). When training on the same samll supervised dataset, the semi-supervised GANs achieved better performance than the supervised DL model (AUCs of 0.90 [95% CI: 0.84-0.96], and 0.92 [95% CI, 0.86-0.97]).”

Under “Methods” section, page 8, line 283, the following sentence was revised.

- “After image grading and preprocessing, we further built 3 training datasets for semi-supervised GAN and DL models development: (1) a fully supervised dataset (33,864

ANNALS OF TRANSLATIONAL MEDICINE



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

ACA images with labels); (2) a small supervised dataset (20, 50, 100, 200, 400 and 1000 ACA images randomly selected from the whole dataset); and (3) an unsupervised dataset (32,864 ACA images without labels) (Figure 1)."

Under "Methods" section, page 12, line 456, the following sentences were revised.

- "To evaluate the performance of semi-supervised GANs for closed-angle detection, we trained two supervised DL models: the 1st supervised DL model (DL_Model_F) using the fully supervised training dataset and the 2nd supervised DL model (DL_Model_S) using the small supervised dataset."

Under "Results" section, page 19, the first paragraph, line 686, the following sentence was added.

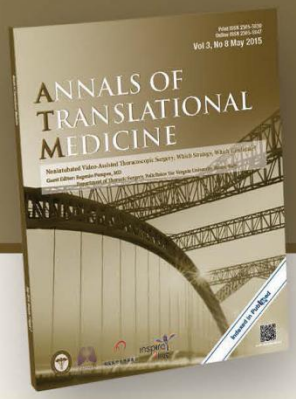
- "We then compared the diagnostic performance between the semi-supervised GANs, and two supervised DL models (DL_Model_S and DL_Model_F). As shown in Table 2 and Figure 3, the semi-supervised GANs model has performance close to the DL_model_F (training on whole dataset). The semi-supervised GANs achieved better performance than the supervised DL_model_S when training on the same small supervised training dataset."

Under "Discussion" section, page 23, the first paragraph, line 747, the following sentence was revised.

- "In two independent testing datasets, the semi-supervised GANs model's performance was acceptable when training with relatively small dataset (400 images)."

Table 2 and Figure 3 was modified as advised.

ANNALS OF TRANSLATIONAL MEDICINE



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

2. The rationale for including Visante image is unclear. These images differ quite a bit from CASIA images, and they are in the large minority in the complete dataset. How many were included in the training dataset? Also, the test datasets were comprised entirely of CASIA images. What purpose does including them serve other than to add noise to the training dataset?
Response: We strongly agree with the reviewer that Visante images may add noise to the whole training processes. We have removed Visante images from datasets and re-trained the models using datasets with only Casia images.
Changes in the text:
Under “Methods” section, page 7, the second paragraph, line 257, the following sentence was deleted.

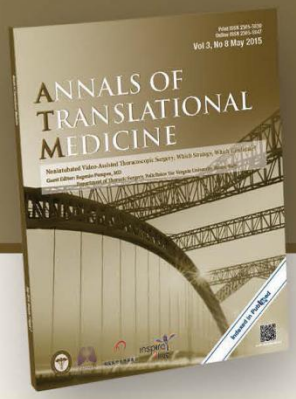
- “either one of the two AS-OCT modalities: Visante ASOCT (Model 1000 with a scan speed of 2000 A-scans per second, software version 2.1; Carl Zeiss Meditec) or”

Under “Methods” section, page 8, the first paragraph, line 278, the following sentence was deleted.

- “2,798 Visante AS-OCT images and”

Figure 1 was modified as advised.

2. How was the number of training images determined? 400 images could be a large number for a relatively simple task, which appears to be the case since the performance of the supervised model is not substantially worse than that of the semi-supervised model. Taking even fewer



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

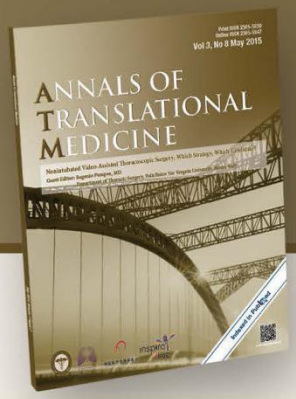
images may accentuate the benefit of the GANs approach, which currently is not statistically significant.

Response: We thank the reviewer for raising a valid point and understand the concern expressed. In current study, we conducted experiments by varying different shots of image dataset (20, 50, 100, 200, 400 and 1000 samples). When trained with a large supervised dataset (n = 1000, accuracy = 0.93 [95% CI, 0.91-0.95]), the semi-supervised GAN methods had only slight improvements over the same method trained with a smaller dataset (n = 400, accuracy = 0.92 [95% CI, 0.90-0.94]). We therefore chose 400 images for finial analysis. We have elaborated on the computer experiments in the methods section and added table 2 to describe the performance of semi-supervised GAN with different image samples as supervised datasets.
Changes in the text:
Under “Methods” section, page 8, the first paragraph, line 285, the following sentence was added.

- “(2) a small supervised dataset (20, 50, 100, 200, 400 and 1000 ACA images randomly selected from the whole dataset)”

Under “Results” section, page 17, the first paragraph, line 660, the following sentences were added.

- “For closed-angle detection using the clinician’s grading of AS-OCT imaging as the reference standard, we first evaluated the performance of semi-supervised GANs with decreasing supervised training samples of 1000, 400, 200, 100, 50 and 20. The results reveal



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

that the diagnostic capability of semi-supervised GANs underwent a performance degradation as the number of supervised datasets decreased (Table 1 in the Supplement). Specifically, when trained with a large supervised dataset (n = 1000, accuracy = 0.93 [95% CI, 0.91-0.95]), the semi-supervised GANs methods had only slight improvements over the same method trained with smaller dataset (n = 400, accuracy = 0.92 [95% CI, 0.90-0.94]). Considering the similar results, we only used the semi-supervised GANs trained with 400 supervised samples for further experiments."

Tabel 2 presented the diagnostic matrices of the semi-supervised GANs model for closed-angle detection with different training samples

3. Why are the test datasets so small? A larger number could help to shrink the confidence intervals so that the difference in performance between the two models becomes significant.

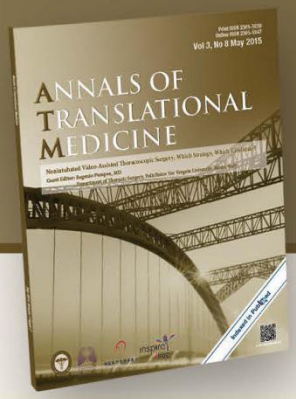
Response: We acknowledge the point made by the reviewer and enlarged the JSIEC testing dataset. The results have been incorporated in Figure 1, 3 and Table 1, 3 and elucidated. We have also elaborated on these in the results section.

Changes in the text:

Under "Methods" section, page 9, the first paragraph, line 313, the following sentence was modified.

- "The first testing dataset (hereinafter referred to as the "JSIEC testing dataset") was collected from an ongoing clinical trial (Chinese clinical trial registration number:

ANNALS OF TRANSLATIONAL MEDICINE



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

ChiCTR2000037892) in JSIEC between March 2019 and June 2020. The JSIEC testing dataset included 264 open-angle and 251 angle-closure ACA images."

Under "Results" section, page 14, the first paragraph, line 632, the following sentence was modified.

- "The JSIEC testing dataset consisted of 515 ACA images from 103 participants (mean age 63.8 years, 60.2% female)."

5. There are many grammatical mistakes. Please proof-read carefully before resubmitting.

Response: The manuscript has been professionally edited by a native English speaker.

Specific points:

Line 36: This is overly general and not always true (the supervised model in this study had adequate performance despite a small training dataset). Please reword.

Response: We accept the reviewer's suggestion and reword the sentence as advised.

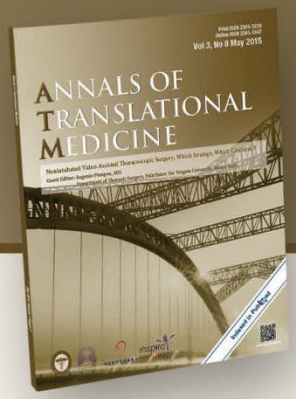
Changes in the text:

Under "abstract" section, page 3, the second paragraph, line 61, the following sentence was revised.

- "Semi-supervised learning algorithms can leverage an unlabeled dataset when labeling is limited or expensive to obtain."

Line 44: Provide definition of angle closure here.

ANNALS OF TRANSLATIONAL MEDICINE



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

Response: We thank the reviewer for the suggestion and added the definition of angle closure in our 'abstract.

Changes in the text:

Under "abstract" section, page 3, the second paragraph, line 68, the following sentence was added.

- "The closed angle was defined as iris-trabecular contact beyond the scleral spur in AS-OCT images."

Line 54: Based on the CI, it is questionable whether the GANs model produce significantly better performance.

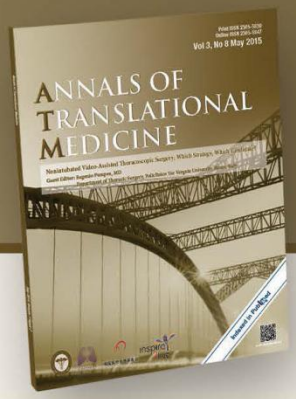
Response: We acknowledge the point made by the reviewer. As mentioned above, we enlarged the JSIEC testing dataset. We also agree with the review that it is questionable whether the GANs model produce significantly better performance. We have revised these phrases to convey the message.

Under "abstract" section, page 3, line 80, the following sentence was added.

- "When training on the same small supervised dataset, the semi-supervised GANs achieved performance at least as well as, if not better than, the supervised DL model (AUCs of 0.90 [95% CI: 0.84-0.96], and 0.92 [95% CI, 0.86-0.97])."

Line 72: There are now DL algorithms that automate the detection of the scleral spur. Also, the scleral spur is more easily detectable in modern OCT devices. Please mention here.

ANNALS OF TRANSLATIONAL MEDICINE



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

Response: We agree with the reviewer that the message has not been conveyed effectively. We have made appropriate changes in the introduction.

Under “introduction” section, page 5, line 179, the following sentence was revised.

- “For example, to evaluate ACA, clinicians have to manually mark the scleral spur, although some researchers had proposed the algorithm to detect the scleral spur and modern AS-OCT modality (swept source) can image the scleral spur more easily.”

The following reference was added as well.

- ‘Xu BY, Chiang M, Pardeshi AA, Moghimi S, Varma R. Deep neural network for scleral spur detection in anterior segment OCT images: The Chinese American eye study. Transl Vis Sci Technol. 2020;9(2):1-10. doi:10.1167/tvst.9.2.18’

Line 80: Mention that this algorithm was developed to detect ITC.

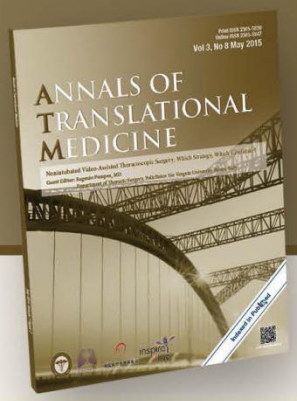
Response: We thank the reviewer for the suggestion and revised the text in our ‘introduction’.

Changes in the text:

Under “Introduction” section, page 5, the second paragraph, line 188, the following sentence was modified.

- “More recently, Fu et al. proposed a supervised DL system to detect iridotrabecular contact (ITC) with high accuracy (an area under the receiver operating characteristic curve [AUC] of 0.96) training on 8,270 AS-OCT ACA images.”

ANNALS OF TRANSLATIONAL MEDICINE



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

Line 111: Define the modified Shaffer grading system. Also, why is this important since gonioscopic angle status was not an exclusion criteria or outcome of the study?

Response: We agree with the reviewer and added definition of the modified Shaffer grading system in our manuscript.

Changes in the text:

Under “Methods” section, page 7, the second paragraph, line 254, the following sentences were modified.

- “The angle was graded using the modified Shaffer grading system: grade 0, no structures visible; grade 1, Schwalbe’s line visible; grade 2, anterior trabecular meshwork visible; grade 3, posterior trabecular meshwork or scleral spur visible; grade 4, ciliary body visible.”

The following reference was added as well.

- ‘Foster PJ, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. Br J Ophthalmol 2002;86(2):238–242.’

Line 117: Should read “one”, not “on”.

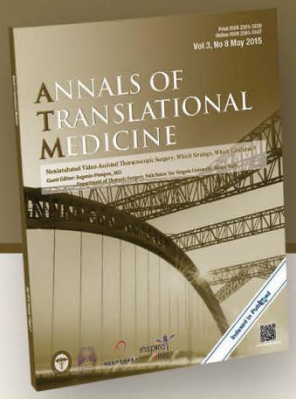
Response: We thank the reviewer for the suggestion and have made the appropriate changes.

Changes in the text:

Under “Methods” section, page 7, the second paragraph, line 260, the following sentence was revised.

- “AS-OCT imaging and gonioscopy were performed in one day.”

ANNALS OF TRANSLATIONAL MEDICINE



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

Line 120: Was any degree of ITC sufficient? Other studies have used 1/3 the width of the trabecular meshwork.

Response: We thank the reviewer for seeking clarity on the issue of ITC. In the current study, we were unable to measure ITC (in um) for all images in our datasets. Manually grading ITC in 1/3 or 1/2 may lead to interobserver variability. We have added these phrases and acknowledged the limitations in the discussion section.

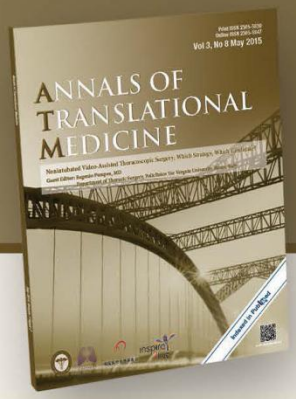
Changes in the text:

Under “Discussion” section, page 25, the second paragraph, line 1085, the following sentences were revised.

- “Third, Fu et al. reported a supervised DL system to detect angle closure with high accuracy using ITC of 1/3 the width of the trabecular meshwork14. As different reference standards were used, our study cannot be directly comparable to Benjamin and Fu’s studies. Additionally, gonioscopy was not the reference standard in this study. Future work is warranted to develop semi-supervised GANs for detecting eyes with gonioscopic angle closure.”

Line 215: How were measurements obtained? This is not described in the methods. Also, measurements between the Visante and CASIA may not be directly comparable and should not be grouped in this manner.

ANNALS OF TRANSLATIONAL MEDICINE



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

Response: We thank the reviewer for the query. We have added more details in the methodology section to describe the measurement. As mentioned previously, we removed the Visante images from our datasets as suggestion.

Changes in the text:

Under "Methods" section, page 10, line 334, the following sentences were added.

- "We used a customized software (Anterior Segment Analysis Program (ASAP)) to measure AS-OCT parameters between the JSIEC and NUHS testing datasets. The parameters included: AOD 750 (the length of the line segment between the cornea and iris at a 750 μm distance from the scleral spur), ACD (the anterior chamar depth), and TISA750 (the trabecular-iris space area at 750 um anterior to the scleral spur)."

The following reference was added as well.

- 'Zheng C, Cheung CY, Aung T, et al. In vivo analysis of vectors involved in pupil constriction in Chinese subjects with angle closure. Investig Ophthalmol Vis Sci. 2012;53(11):6756-6762. doi:10.1167/iovs.12-10415'

Line 247: "Those" not "that".

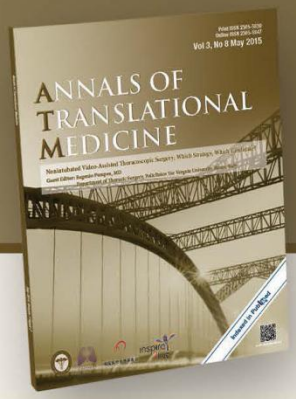
Response: We thank the reviewer for the suggestion and have made the appropriate changes.

Changes in the text:

Under "Discussion" section, page 23, line 704, the following sentence were revised.

- "In two independent testing datasets, the semi-supervised GANs model's performance was acceptable when training with relatively small dataset (400 images)."

ANNALS OF TRANSLATIONAL MEDICINE



Notice: Please give your response to the comments point-by-point as shown in the following format.
【At the END of each reply/response from you, please DO describe how you responded to the reviewer comment in the text, e.g., "we added some data ....(see Page xx, line xx)" or "we have modified our text as advised (see Page xx, line xx)"】.

Comment 1: **********
Reply 1: **********
Changes in the text: **********

Comment 2: **********
Reply 2: **********
Changes in the text: **********

Line 253: I would say that DL models often benefit from being trained on large datasets, rather than they require them. The cited studies used large image datasets, but did not show that smaller datasets would not have sufficed.

Response: We thank the reviewer for the suggestion and have made the appropriate changes.

Changes in the text:

Under “Discussion” section, page 23, the second paragraph, line 753, the following sentence were revised.

- “The DL models often benefit from being trained on large datasets.”

279: “Compared”, not “comparable”.

Response: We thank the reviewer for the suggestion and change ‘comparable’ to ‘compared’.

Changes in the text:

Under “Discussion” section, page 24, the second paragraph, line 919, the following sentence were revised.

- “Although our study cannot be directly compared to previous studies due to different reference standards of closed angles, our results still showed the potential application of this technique with excellent diagnostic performance.”