A crowdsourcing framework for retinal image semantic annotation and report documentation with deep learning enhancement

Jiahui Shao, Jin Li, Weizheng Kong, Shifan Liu, Junyi Wu, Huiqun Wu*

Department of Medical Informatics, Medical School of Nantong University, Nantong, China, 226001, People's Republic of China.

*Corresponding author’s
Telephone number: +86-513-85051891
Fax number: +86-513-85051820
E-mail address: wuhuiqun@ntu.edu.cn;

Abstract. To propose and implement a crowdsourcing framework for retinal image annotations to improve the annotation efficiency. In this study, open-source Bluelight was taken as backbone of the front end for online manual retinal image annotation for image semantic annotation and report documents, and based on that intelligent annotation and classification with deep learning (DL) was supplemented. For DL modules, we trained Mask-RCNN model to explicitly label the area of optic disc and macula. Furthermore, we trained Inception V3 model to classify diabetic retinopathy (DR) and normal retina. Then, we used Flask as the backend serving DL models. Finally, the implementation of interoperable annotation reports documentation and retrieval were conducted based on Lucene. The crowdsourcing framework was specially designed for professional doctors and computer researchers who have the ability to annotate. It efficiently and quickly completed the annotation of the retinal image and the macular area, and at the same time classified DR. Under this Browser/Server architecture, the tool achieved good cross-platform performance. In particular, the framework could provide annotation report documents to facilitate the optimization of subsequent DL models. Such crowdsourcing framework and reports documentation for retina semantic annotation could improve the effect of annotation and classification and worth further improvement and clinical validation.

1. Introduction

In recent years, medical image processing and computer aided detection attracts more and more investigators. Artificial intelligence learns a large number of medical image annotation data sets, which will help doctors perform clinical diagnosis and propose treatment plans. The deep learning (DL) technique is currently widely applied in medical image diagnosis. With the help of image analysis and DL, many diseases can be detected and treated early. In the field of medicine and healthcare, DL is mainly used in medical image analysis. The DL system has shown strong diagnostic performance in detecting various diseases, including tuberculosis in chest radiographs[1]. In training DL models of retinal image, neural networks have become the most popular artificial intelligence technology in modern medicine[2]. As an end-to-end learning framework, large sample size training dataset with annotation is crucial for the model training.

Data annotation is the process of processing images, voices, texts and other data through classification, identification, etc., and annotation the characteristics of the object as the basic material
for machine learning. Data annotation is the key link for most artificial intelligence algorithms to operate effectively. Data annotation is to label the data that needs to be recognized and distinguished by the machine, and then let the computer continuously learn the characteristics of these data, and finally realize that the computer can recognize it autonomously. Medical image annotation is the area annotation and classification annotation of medical image, which are mostly used to assist clinical diagnosis. However, this work is often a huge workload, boring and time-consuming process especially for the annotation of medical images.

To address above challenges, crowdsourcing[3] has been invested in the annotation of retinal image[4], as well as text annotations in radiology reports[5], or to realize the annotation of individual objects [6]. Crowdsourcing benefits from the combination of multiple contributors and depends on the level of information to be collected[7]. Crowdsourcing can be applied to many fields, such as bioomics image tagging, disaster management [8], land use data utilization[9]. In this study, we proposed a crowdsourcing framework to make annotation for retina image with recent advances in DL[10,11] such as Mask-RCNN model to label the area of optic disc(OD) and macula in retinal image and Inception V3 to classify diabetic retinopathy(DR) and normal retina.

2. Method

2.1. The framework of the proposed DL-empowered annotation for retinal image

The design of the framework is to reduce the burden of primary medical care through DL-empowered annotation and classification, and reduce the time doctors spend on manual annotation (Figure 1). To better conformance to the clinical imaging workflow, in which Digital Imaging and Communications in Medicine (DICOM)[12] was utilized to standardize the format the healthcare imaging and communications across the network. The images were saved in the cloud database in the form of DICOM. In the process of manual annotation of retinal image, Bluelight (https://github.com/cylab-tw/bluelight) was taken as backbone of the front end to retrieve those DICOM retinal image for online manual annotation.

![Figure 1. The diagram of the proposed framework for DL-empowered annotation of retinal image.](image)

The manual annotation tools could make the outline border fit the anatomic area on the fundus image to complete the semantic annotation. After the annotation of retinal images, the annotation results could be saved and exported in structured report.

2.2. Automatic annotation with DL models

The Mask R-CNN framework can realize object detection and classification[13]. We trained our model based on 400 images labeled by Labelme and fine-tuned a Mask R-CNN with a inception_v2
framework[14]. In this section, we will first introduce our retinal annotation model. Then, we will introduce our classification model. In the experiment of disease classification on retinal image, the Inception-V3 model was selected to classify retinal image[15]. Compared with other neural network models, the biggest characteristic of Inception network is that the convolution operation between the middle layer and the layer of the neural network was extended, such as VGG, AlexNet[16]. Therefore, we used the model to classify DR and normal retina.

Moreover, we deployed the model to achieve automatic annotation. The images trained in this paper were of size 512*512 pixels. The running environment of the experimental algorithm was configured as: Python3.7, Tensorflow1.14, windows10, Pycharm. Before training the model, we utilized Labelme to preprocess the dataset of retinal image manually. After that, we need to edit the pbtxt file to set the parameters of the macula and OD. Two training batches were provided, one was used as the training image of the data, and the other was used as the validation image of the model. We periodically checked the image data and deleted some ambiguous annotations from the training set.

2.3. The annotation reports documentation and retrieval
In this study, the annotation reports were documented in an interoperable XML format which could further mapped into DICOM structured reports for exchanges. Besides, we used Lucene indexing and search technology[17] to retrieve the XML file results containing retinal image annotations. The initial search term set entered by the user was expanded by the concept similarity and relevance integrated in the domain ontology to generate expanded search terms and corresponding weights. The annotation files were collected and submitted to Lucene search engine to for search.

3. Results

3.1. The semantic retinal image annotation and classification
As the Figure 2 illustrated, the retinal image from cloud eye-PACS could be retrieved and displayed on the front-end annotation web page where image annotation tools were shown in toolbars for annotation. The annotater could retrieve the retinal image from the worklist and make some manual annotations, and save the results in the local computer in xml file. The user could save the image annotation marks as an xml file for further use.

For DL enhancement, we deployed the trained DL models on a Flask server with a link on the frontend bluelight browser to launch the DL applications. Figure 3 showed the automatic annotation of DR and semantic annotation of OD and macula annotation process by the DL models.
3.2. The interoperable annotation reports documentation and retrieval

Relevant clinical information available for the retinal image in the study included clinical information (such as Date, PatientsName, PatientID, and the StudyUID). This information can realize the interpretation and reuse of retinal image and data sets, since it contains the basic information regards to annotation of image, and it provides image analysis data that can be used for machine learning. Comprehensive use of open-source Apache Lucene[18] database, it is a searchable database for management and retrieval of text information. Lucene has been used in medical and biological information retrieval projects[19]. Lucene's text search function could be used to search for key words or phrases in an article, similar to keyword search in various literature databases. Lucene could also search for information in data tables, and retrieved DICOM structured reports. It could greatly improve the speed of information retrieval and analysis, and increased the number of effective information acquisitions. The deployment of Lucene for annotation report was shown in Figure 4.

4. Discussion

Medical image annotation often requires talents with professional background. In addition, experts sometimes need to preprocess the image in advance, and adjust the color of the image to make the detection result more accurate. Currently, medical image annotation lacked a unified standard annotated image data set, mainly because the conventional method of collecting image tags from ordinary users through Google search could not be applied to the field of medical image[20]. A medical professional required a certain amount of time to access and process the images that were produced in real-time by the scanners in the hospital networks[21]. The image classification task in computer vision aimed to distinguish many categories from prominent objects in the image [22]. DL models had been used in eye
imaging technologies, mainly retinal photography, to identify major eye diseases including DR, glaucoma, and age-related macular degeneration. However, the application of DL model to intelligently recognize retinal image often requires the installation of client programs. The user cannot directly use the DL model to achieve the annotation of retinal image on the computer without the client installed. It remains challenging to develop an intelligent annotation platform for retinal image on the web. In the semantic annotation of the macula and OD area of the retinal image, we chose the Mask-RCNN approach as it could annotate both structures efficiently.

Although crowdsourcing tasks should not be too complicated, our research was not inconsistent with this view, that was, when solving images affected by more controlled settings, it could realize the research of complex tasks [23]. Existing annotation tools often support common image formats (such as png), and Bluelight support annotations for images in DCM formats. Zou et al. designed an image annotation tool for glaucoma, specifically for the OD [24]. Compared with them, our automatic annotation has a significant advantage in time. Ye et al proposed ‘A Crowdsourcing Framework for Medical Data Sets’ and established a keyword search system [25].

We not only proposed the crowdsourcing framework based on DL-enhanced retinal image semantic annotation and report documents, but also realized the function of semantic retrieval. In the following work, we plan to integrate various models into the same platform, so as to realize the comprehensive annotation platform of retinal image. In this case, however, the classification results should be based on the details of a particular location in the image. In addition, Flask is a lightweight web application framework written in Python. Not only is the framework simpler and less logical, but it also has good scalability. Compared with the Django framework, it is more flexible and can be deployed quickly by deploying the model and uploading image to display the results of annotation and classification on the web page. Nevertheless, there might be inaccurate problems in training classifiers from images, and we also encountered inaccurate results or misjudgment problems in the experiment. In addition, performance steadily improved with the use of more training samples and the framework will show its advantage if the iteration of annotation process increased. In the future, we will continue to improve the efficiency of automatic annotation by training more data for DL models and extracting lesion types from structured reports, thereby increasing the accuracy of the model and providing evidence for clinical trials through DICOM structured reports.

5. Conclusion
In this article, we introduced the intelligent annotation and intelligent classification of retinal image, which required minimal manual work and could realize automatic classification by uploading images. Such crowdsourcing framework and reports documentation for retina semantic annotation could improve the effect of annotation and classification and worth further improvement and clinical validation.

Acknowledgments
This work was supported by the grant from National Key R&D Program of China (2018YFC1314900, 2018YFC1314902), Science and Technology Project of Nantong City(MS12020037), Nantong “226 Project”, Excellent Key Teachers in the “Qing Lan Project” of Jiangsu Colleges and Universities, Jiangsu Students' Platform for innovation and entrepreneurship training program(201910304108Y); and Jiangsu postgraduate research and innovation program (KYCX20_2836).

References
[1] Lakhani P, Sundaram B. (2017) Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. J. Radiology, 284(2):574-582.
[2] Kahn CE Jr. (1994) Artificial intelligence in radiology: decision support systems. J. Radiographics, 14(4):849-61.
[3] Patten SB. (2015) The Wisdom of crowds (Vox Populi) and antidepressant use. J. Clinical Practice & Epidemiology in Mental Health, 11(1):1-3.
[4] Mitry D, Zutis K, Dhillon B, et al. (2016) The accuracy and reliability of crowdsource annotations of digital retinal images. J. Transl Vis Sci Technol, 5(5):6.

[5] Cocos A, Qian T, Callison-Burch C, et al. (2017) Crowd control: Effectively utilizing unscreened crowd workers for biomedical data annotation. J. Journal of Biomedical Informatics, 69:86-92.

[6] Gurari D, Sameki M, Bette M. (2016) Investigating the influence of data familiarity to improve the design of a crowdsourcing image annotation system. In: HCOMP. Austin. pp. 59-68

[7] Grote A, Schaadt NS, Forestier G, et al. (2019) Crowdsourcing of histological image labeling and object delineation by medical students. J. IEEE Trans Med Imaging, 38(5):1284-1294.

[8] Boccardo P, Pasquali P. (2012) Web mapping services in a crowdsourcing environment for disaster management: state-of-the-art and further development. J. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 39:543-548.

[9] See L, Comber A, Salk C, Fritz S, et al. (2013) Comparing the quality of crowdsourced data contributed by expert and non-experts. J. PLoS One, 8(7):e69958.

[10] Medley D O, Santiago C, Nascimento J C. (2019) Deep Active Shape Model for Robust Object Fitting. J. IEEE Transactions on Image Processing, (99):1-1

[11] Ren S, He K, Girshick R, Sun J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. J. IEEE Trans Pattern Anal Mach Intell, 39(6):1137-1149.

[12] Pianykh O S. (2012) Digital imaging and communications in medicine (DICOM). Springer, Berlin.

[13] He K, Gkioxari G, Dollar P, et al. (2020) Mask R-CNN. J. IEEE Trans Pattern Anal Mach Intell, 42(2):386-397.

[14] Zhou LQ, Wu XL, Huang SY, et al. (2020) Lymph node metastasis prediction from primary breast cancer US images using deep learning. J. Radiology, 294(1):19-28.

[15] Szegedy C, Vanhoucke V, Ioffe S, et al. (2015) Rethinking the inception architecture for computer vision. C. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 2818-2826.

[16] Krizhevsky A, Sutskever I, Hinton G E. (2017) Imagenet classification with deep convolutional neural networks. J. Communications of the ACM, 60(6): 84-90.

[17] Müller HM, Van Auken KM, Li Y, et al. (2018) Textpresso Central: a customizable platform for searching, text mining, viewing, and curating biomedical literature . J. BMC Bioinformatics, 19(1): 94.

[18] Milosavljevic B, Boberic D, Surla D. (2010) Retrieval of bibliographic records using Apache Lucene. J. Electronic Library, 28(4):525-539.

[19] Spat S, Cadonna B, Rakovac I, et al. (2008) Enhanced information retrieval from narrative German-language clinical text documents using automated document classification. J. Stud Health Technol Inform, 136:473–478.

[20] Yan, K, Wang X, Lu L, et al. (2018) DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. J. J Med Imaging (Bellingham), 5(3): p. 036501.

[21] Kathiravelu P, Sharma A, Purkayastha S, et al. (2020) Developing and Deploying Machine Learning Pipelines against Real-Time Image Streams from the PACS. J. Proceedings of Machine Learning Research, 1-14.

[22] Russakovsky O, Deng J, Su H, et al. (2014) ImageNet large scale visual recognition challenge. J. International Journal of Computer Vision, 1-42.

[23] Staffelbach M, Sempolinski P, Kijewski-Correa T, et al. (2015) Lessons learned from crowdsourcing complex engineering tasks. J. PLoS One, 10(9):e0134978.

[24] Zou B J, Guo Y D, Chen Z L, et al. (2018) BGIDB: A fundus ground truth building tool with automatic DDLS classification for glaucoma research. J. Journal of Central South University, 25(9):2058-2068.

[25] Ye C, Coco J, Epishova A, et al. (2018) A crowdsourcing framework for medical data sets. J. AMIA Jt Summits Transl Sci Proc, 2017:273-280.