J Med Syst (2019) 43: 197
https://doi.org/10.1007/s10916-019-1258-9

SYSTEMS-LEVEL QUALITY IMPROVEMENT

DeepLNAnno: a Web-Based Lung Nodules Annotating System for CT Images

Sihang Chen[1] · **Jixiang Guo**[1] · **Chengdi Wang**[2] · **Xiuyuan Xu**[1] · **Zhang Yi**[1] · **Weimin Li**[2]

Received: 21 December 2018 / Accepted: 27 March 2019 / Published online: 22 May 2019

Abstract

Lung cancer is one of the most common and fatal types of cancer, and pulmonary nodule detection plays a crucial role in the screening and diagnosis of this disease. A well-trained deep neural network model can help doctors to find nodules on computed tomography(CT) images while requiring lots of labeled data. However, currently available annotating systems are not suitable for annotating pulmonary nodules in CT images. We propose a web-based lung nodules annotating system named as DeepLNAnno. DeepLNAnno has a unique three-tier working process and loads of features like semi-automatic annotation, which not only make it much easier for doctors to annotate compared to some other annotating systems but also increase the accuracy of the labels. We invited a medical group from West China Hospital to annotate the CT images using our DeepLNAnno system, and collected a large number of labeled data. The results of our experiments demonstrated that a usable nodule-detection system is developed, and good benchmark scores on our evaluation data are obtained.

Keywords Annotating system · Neural network · Medical application · Medical data collection

Introduction

Lung Cancer is now the most common incident cancer and the leading cause of cancer death in China [2]. Lung cancer manifests itself as a pulmonary nodule in its early stage [3], thus detecting nodules is crucial for reducing mortality. Low-dose chest computed tomography (CT) images are often used to check for nodules, but there are several limitations. First, large hospitals typically perform thousands of CT studies every day, and a patient may wait almost a week for his or her medical report. Second, identifying the nodules is subjective and requires well-honed expertise which may not be available in underdeveloped areas. By developing a computer-aided nodule detection system, we aim to improve the ease and efficiency of nodule detection on CT images, particularly benefitting underdeveloped areas.

The most essential part of developing a computer-aided nodule detection system is to train a model which detects suspicious nodules correctly and efficiently, which can be concluded as a computer vision task. Computer vision tasks today require large datasets as small datasets can be easy to overfit for training and therefore only work for some selected cases. Luckily, as the development of the Internet and storage technologies, computer scientists can have access to large datasets much easier than before. However, there are still some challenges, as most of the artificial intelligence or machine learning require annotated data with detailed labels where some of the datasets may lack (e.g. human faces [7, 12, 13]).

One of the most common approaches of developing a nodule-detection system is to divide the task into two steps, the candidate generation step and the false positive reduction step [9], while both steps require detailed labels like the world coordinates of every nodule on a single CT study and their corresponding qualitative values. Although there are indeed some open datasets for CT images like LIDC-IDRI [1] or the dataset provided by Kaggle for the 2017 Data

---

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

✉ Weimin Li
weimi003@scu.edu.cn

[1] Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, 610065, People's Republic of China

[2] Department of Respiratory and Critical Care Medicine, West China School/West China Hospital, Sichuan University, Chengdu, Sichuan Province 610041, People's Republic of China

Science Super Bowl, two major obstacles make it still difficult for us to directly use these datasets to develop a well-performed computer-aided system. First, many factors could affect the performance of our neural network model like the differences in imaging equipment, the variations between different human races, the annotation standards etc. Besides, currently available datasets often lack detailed labels which we need to develop our nodule-detection system. Under these circumstances, a more well-fitted private dataset is in need, and a re-annotation task must be done to train a better model for detecting nodules.

Considering existing annotating tools, there are famous ones like LabelMe [11], or more specific medical segmentation tools like ITK-SNAP [15]. LabelMe is a web-based online image annotation tool developed by MIT and is widely used by researchers from all over the world to annotate their own data. Although it has some extra features like polygon-drawing and real-world 3D coordinate extraction, it is not suitable for annotating CT images. First CT images often come in DICOM (Digital Imaging and Communications in Medicine) format, which is not just an image or file format but an all-encompassing data transfer, storage, and display protocol built and designed to cover all functional aspects of contemporary medicine [6]. In other words, CT images come with an incompatible format with LabelMe. Besides, LabelMe is developed for the purpose of annotating natural images, so annotators can only assign every RoI with a single label using LabelMe while we need more than one label for each nodule, so it is unfeasible to use LabelMe or any other existing web-based annotating tools to annotate CT images. ITK-SNAP is a famous open source application intended to segment structures in 3D medical images, and it gets the job done to read DICOM files and make rectangle annotations on that. However, ITK-SNAP can not save the coordinates of the annotation automatically, meaning that the annotators have to write down the coordinates of the nodule areas every time they draw a rectangle, which is very time-consuming.

We build DeepLNAnno, a web-based system for annotating raw CT images to aid in building a more accurate and robust nodule detection system. It adopts a novel three-tier annotating procedure which makes the whole annotating process reasonable and efficient. It also provides loads of features like semi-automatic annotation, which make it much easier for annotators to perform annotation on CT images to mark out the nodules. The rest of the paper is organized as follows. A detailed description of our DeepLNAnno system including our unique working process and loads of features is presented in Section "DeepLNAnno". The real world usages of our system including experiment results are presented in Section "Real-world usages". Some thoughts about future work and system improvements are presented in Section "Future work". Finally, in Section "Conclusions", a conclusion of our DeepLNAnno system is presented.

DeepLNAnno

Several features distinguish DeepLNAnno from other annotating platforms, rendering it more suitable for annotating CT images.

- Ease of access. DeepLNAnno is a web-based system. Any user with a legitimate account and access to an Internet connection through a usable computer may annotate CT images.
- Ease of use. Labeling a study of CT images simply requires drawing rectangles using a mouse or trackpad to assign corresponding location values and qualitative values to nodules. The corresponding values are saved in the database every time a rectangle is drawn or a value placeholder is filled.
- Three-tier annotating process. DeepLNAnno adopts a novel annotating process which can save the time of medical professionals with superior expertise as well as ensuring the accuracy.
- Semi-automatic annotation. Annotators can start their annotations based on the results given by our model to avoid missing nodules to some extent.
- Fast data export. Our built-in scripts help the system to export the annotated data quickly. The annotated data is saved in the format required for training models.

Three-tier working process

DeepLNAnno adopts a three-tier working process which not only shortens the time for annotating but also ensures the accuracy of the labels. First, Ph.D. candidates in medicine are assigned with a certain number of CT studies which they annotate from the start. Second, doctors with at least five years of clinical experience conduct a first-round review and potentially modify the annotated results to increase accuracy. Third, advanced medical experts and professors perform a final review to determine if the annotations are entirely correct. They then submit the final labels which are used to train the detection models. With our three-step structure, data scientists can acquire labels of the CT images while doctors and experts save time. Figure 1 shows the overall flow of the DeepLNAnno process.

Data preprocessing and database design

DICOM is the most common data format for CT studies but it is not easily used by a web application. CT images in DICOM format contain extra information along with

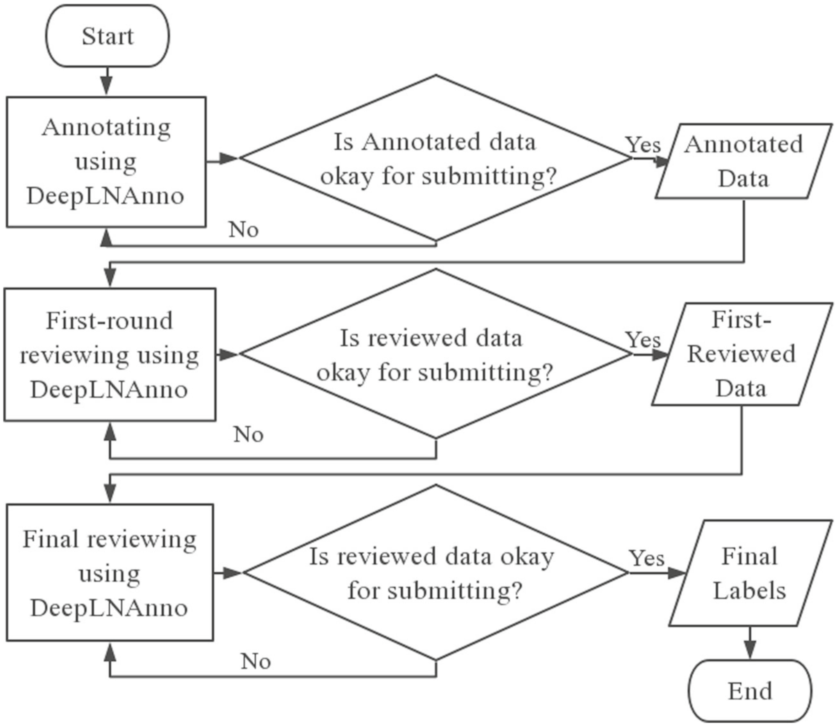


Fig. 1 The overall working process of DeepLNAnno

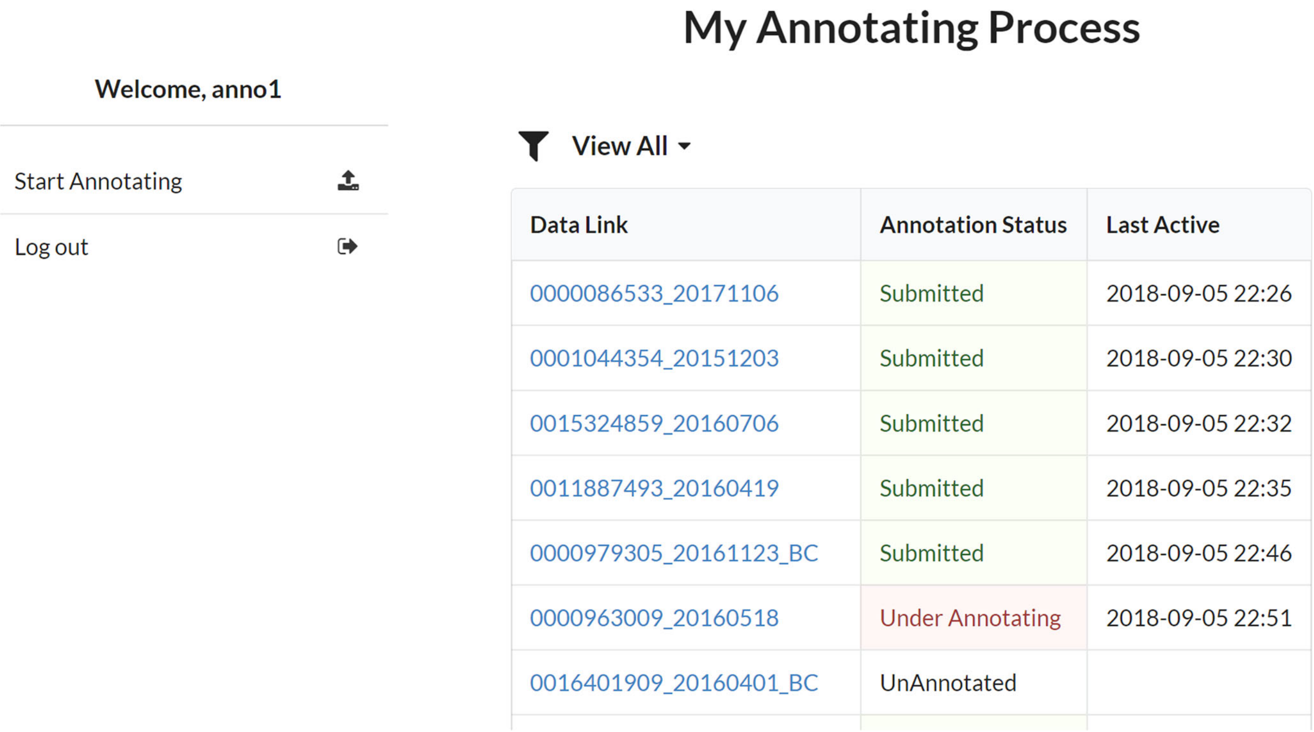
the image content, including the patient name, sex, and address etc. This private and sensitive information requires additional safeguards. Therefore, we first develop a tool for doctors to keep only non-sensitive information such as the image data values and the series IDs of raw CT images. Then, we build another tool to convert these anonymous CT images to the proper format which can be put into our DeepLNAnno system. We adopt a non-relational database for storing annotation records to avoid unnecessary redundancy.

User interface and functionality

The UI structure of DeepLNAnno can be divided into two parts: the working interface and the administrating interface. The working interface allows the annotators and reviewers to perform their tasks, forming the most important part of the DeepLNAnno system. Administrators use the administrating interface to handle tasks such as creating or editing users, viewing annotation progress and assigning new tasks for users etc.

Figure 2 shows the task-selection page, where a logged-in user can enter the annotating/reviewing page by simply clicking on the CT study he or she wants to annotate or review on the page. A detailed view of the annotating/reviewing page is illustrated in Fig. 3. On the left side, there is an image display window, a progress bar and two navigation buttons. On the right side, there's a table showing all the extra information of the nodules that should be filled by the annotators. One can drag the slider on the progress bar, click on the navigation buttons or use a keyboard to switch between slices. Annotators use the mouse to draw a rectangle on the image as soon as they found a nodule on the displayed slice, and adjust the size or the position of that rectangle. Every time a rectangle is drawn, a new row of data will appear on the right-side table, which comes with a given area ID and the slice number of the CT study. Annotators will have to provide some extra information like the nodule ID given by themselves, the relative position of the nodule and the qualitative value of the nodule if there's a record in their pathological database. Every operation that annotators or reviewers make will be saved in the database automatically. When annotators or reviewers think all the nodules of one CT study are correctly annotated, they can submit the case, or if they feel like moving on to the next study without submitting the current one, they can click the Temp-Save button to temporarily save the annotation and visit back later.

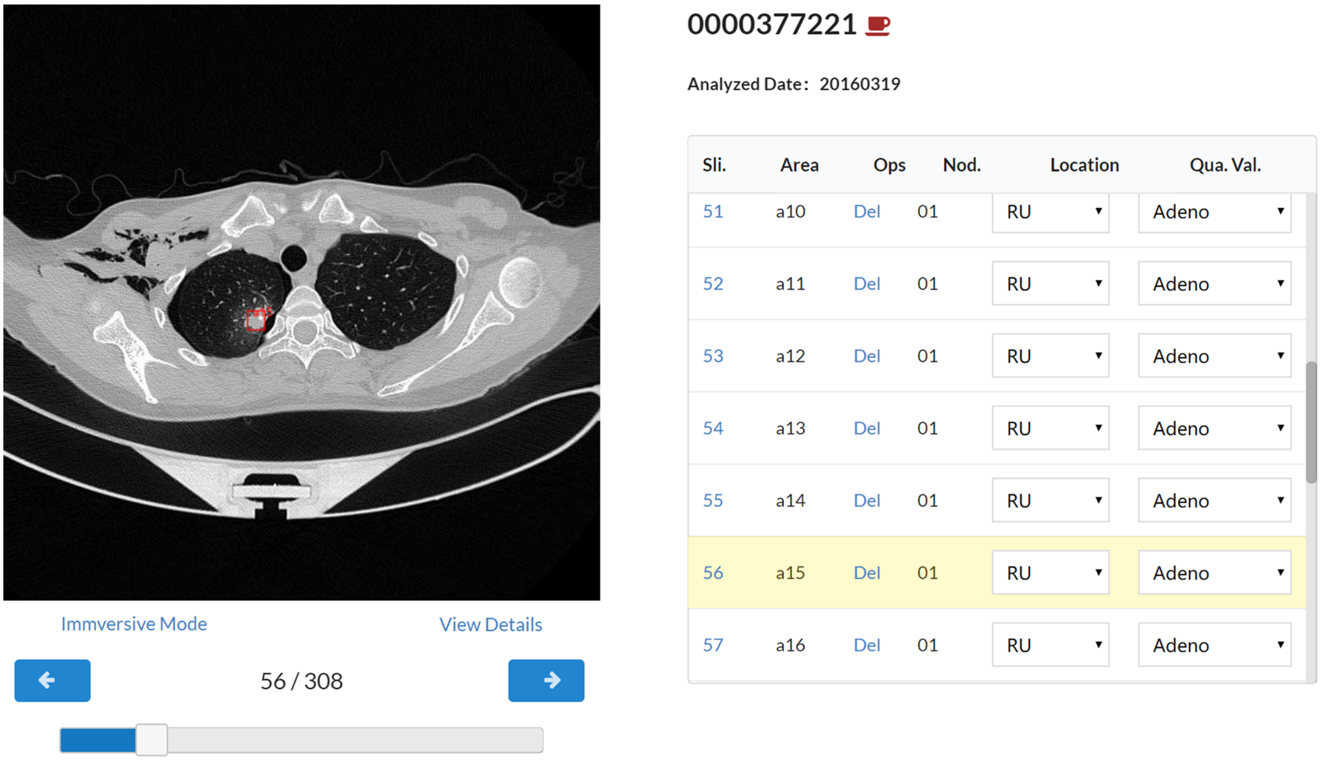
Figures 4 and 5 show the administrating interface of DeepLNAnno. As illustrated in Fig. 4, on the view-total page, an administrator can see all the studies in the database and their current status, whether it is untouched, is been working on or is already submitted. For every row in the table, there're labels to directly show which annotator or reviewer has handled this study. When administrators enter one CT study by clicking on the corresponding row in the table, they can see all the nodules annotated or reviewed



Data Link	Annotation Status	Last Active
0000086533_20171106	Submitted	2018-09-05 22:26
0001044354_20151203	Submitted	2018-09-05 22:30
0015324859_20160706	Submitted	2018-09-05 22:32
0011887493_20160419	Submitted	2018-09-05 22:35
0000979305_20161123_BC	Submitted	2018-09-05 22:46
0000963009_20160518	Under Annotating	2018-09-05 22:51
0016401909_20160401_BC	UnAnnotated	

Fig. 2 The task-selection page of the working interface

by the annotators or reviewers just like an annotator or a reviewer. The only difference is that administrators don't get the authority to make any modifications, which guarantees the stability and correctness of the annotated labels. In addition, administrators can see how many studies each annotator or reviewer has annotated or reviewed and assign

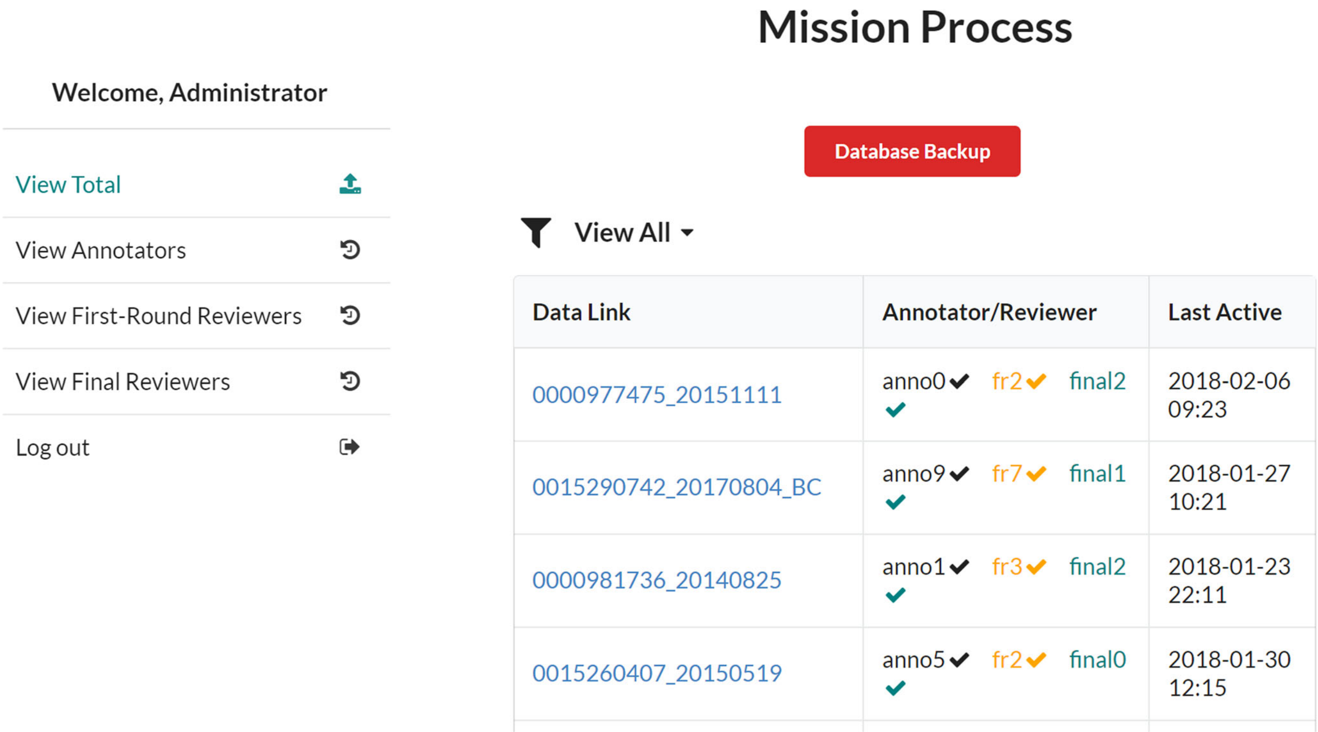


0000377221

Analyzed Date: 20160319

Sli.	Area	Ops	Nod.	Location	Qua. Val.
51	a10	Del	01	RU	Adeno
52	a11	Del	01	RU	Adeno
53	a12	Del	01	RU	Adeno
54	a13	Del	01	RU	Adeno
55	a14	Del	01	RU	Adeno
56	a15	Del	01	RU	Adeno
57	a16	Del	01	RU	Adeno

Fig. 3 The annotating/reviewing page of the working interface

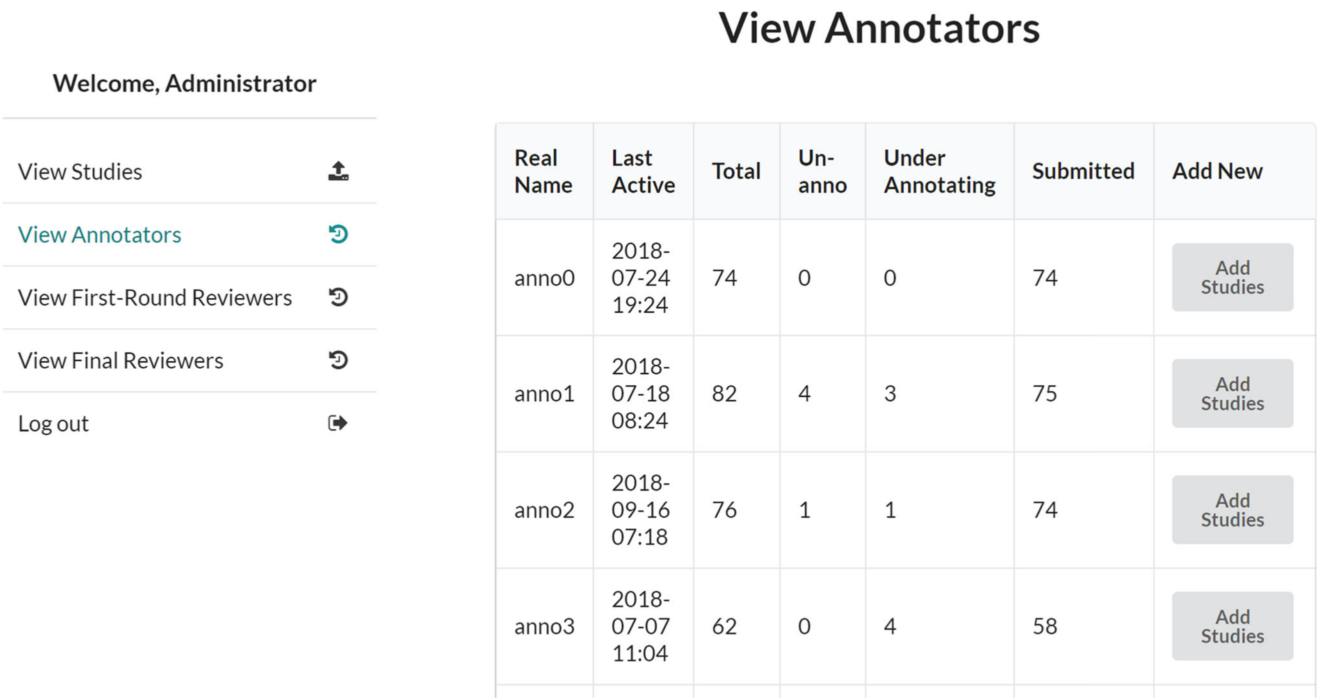


Data Link	Annotator/Reviewer	Last Active
0000977475_20151111	anno0✓ fr2✓ final2 ✓	2018-02-06 09:23
0015290742_20170804_BC	anno9✓ fr7✓ final1 ✓	2018-01-27 10:21
0000981736_20140825	anno1✓ fr3✓ final2 ✓	2018-01-23 22:11
0015260407_20150519	anno5✓ fr2✓ final0 ✓	2018-01-30 12:15

Fig. 4 The view-total page of the administrating interface

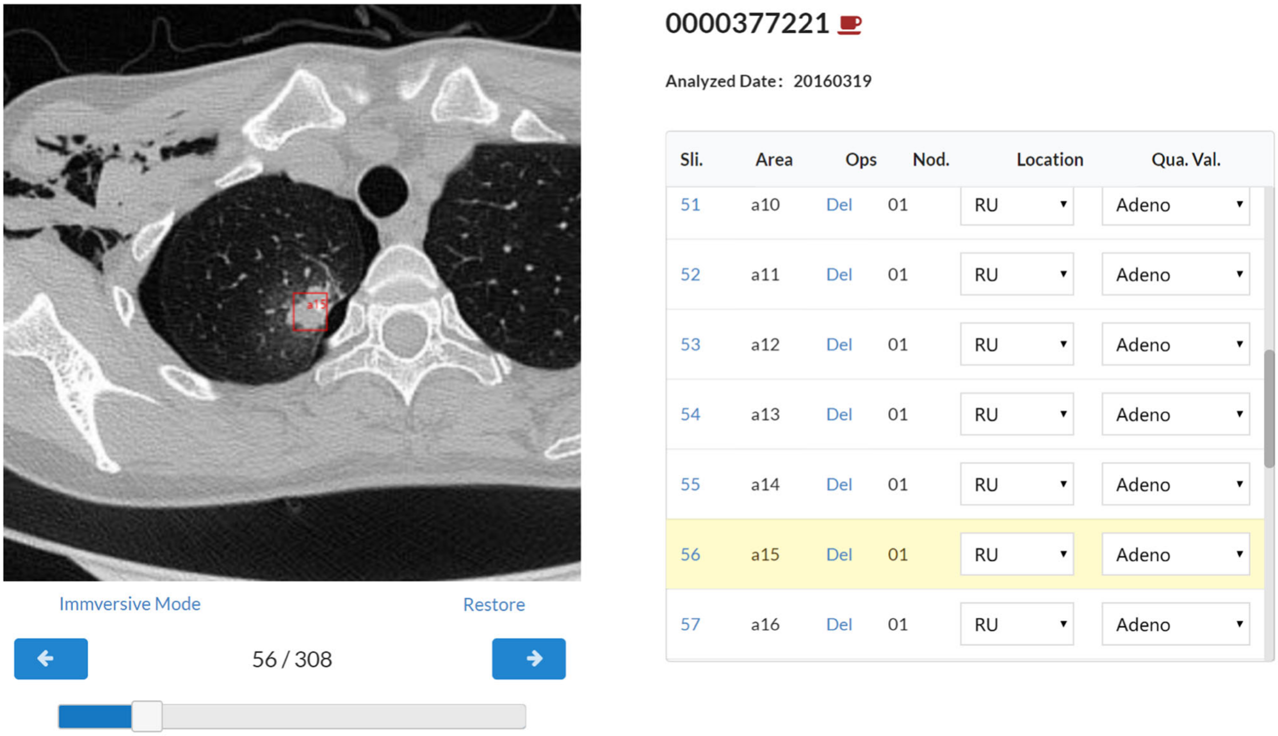
new CT data to them if needed on the view-annotator/view-reviewer page shown in Fig. 5.

There are also some extra features which make our system much easier to use. DeepLNAnno provides two extra image-viewing modes for annotators, making it easier to find and annotate nodules, named as the Zoom Mode and the Immersive Mode (Fig. 6). Users can switch to Zoom Mode to annotate small nodules much more



Real Name	Last Active	Total	Un-anno	Under Annotating	Submitted	Add New
anno0	2018-07-24 19:24	74	0	0	74	Add Studies
anno1	2018-07-18 08:24	82	4	3	75	Add Studies
anno2	2018-09-16 07:18	76	1	1	74	Add Studies
anno3	2018-07-07 11:04	62	0	4	58	Add Studies

Fig. 5 The view-annotator page of the administrating interface

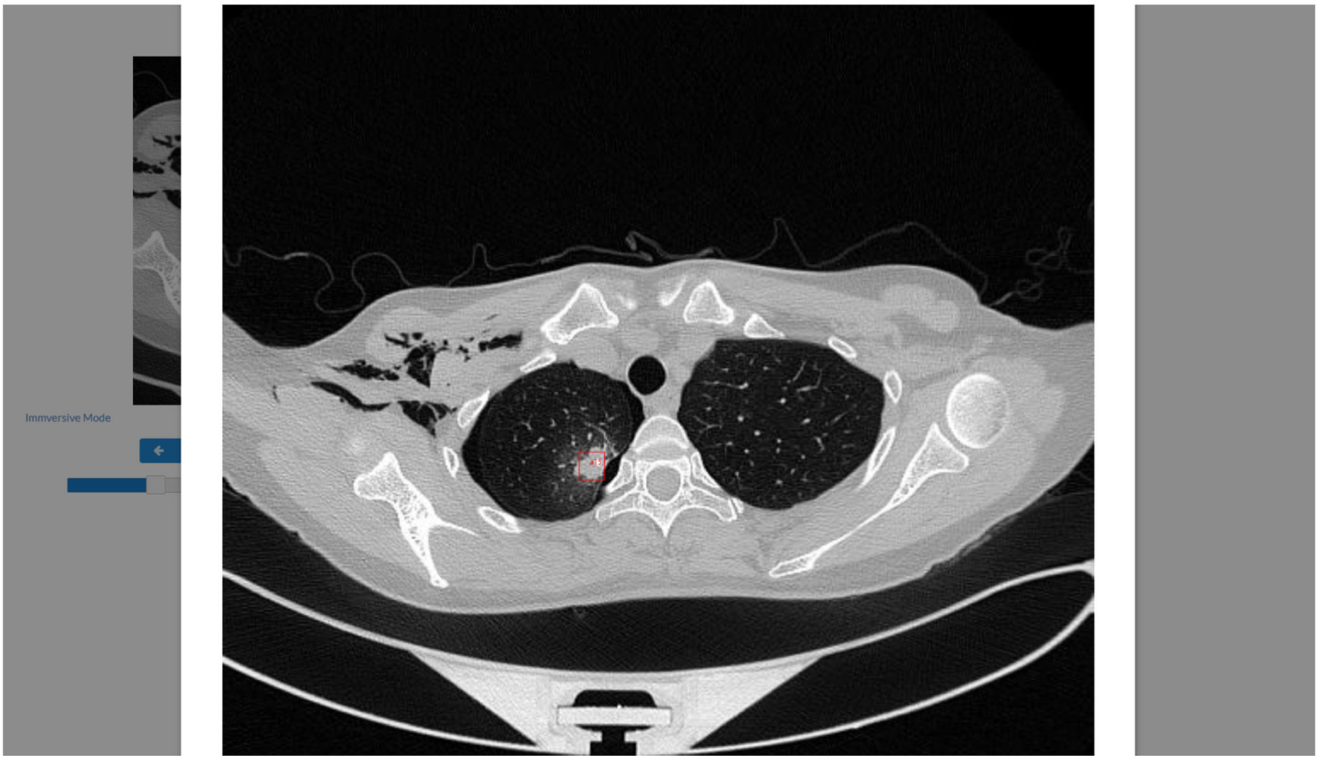


0000377221

Analyzed Date: 20160319

Sli.	Area	Ops	Nod.	Location	Qua. Val.
51	a10	Del	01	RU	Adeno
52	a11	Del	01	RU	Adeno
53	a12	Del	01	RU	Adeno
54	a13	Del	01	RU	Adeno
55	a14	Del	01	RU	Adeno
56	a15	Del	01	RU	Adeno
57	a16	Del	01	RU	Adeno

(a) Zoom Mode



(b) Immersive Mode

Fig. 6 Illustrations of the two extra image-viewing modes presenting the same slice as in Fig. 3

easily as it magnifies the original image four times. They can also switch to Immersive Mode where the displayed image will fill the entire web page, which makes it large enough for annotators or reviewers to find the nodules. A preloading strategy is employed to improve the unfriendly user experience caused by network latency. We have also developed scripts to quickly export annotation data from the database.

Semi-automatic annotation

One big feature of the DeepLNAnno system is the semi-automatic annotation. Semi-automatic annotation has been applied in the field of machine learning since the early years of 21st Century [5, 14]. Russell et al. use semi-automatic labeling in their LabelMe system by first train an algorithm of a particular object class which has enough annotations, then apply the algorithm on unannotated images to perform detection or segmentation. The users then only have to provide feedback about whether the system has made a right detection or not, and then the detector will be trained again with the enlarged dataset. Russell et al. [8] Our system is inspired by the idea, and we put semi-automatic annotation feature in our system to improve the annotating experience as well as obtaining more accurate labels.

As depicted above, the working process of the system can be summarized as three different stages: the annotating, the first-round reviewing and the final reviewing. However in reality, during the annotating process, annotators are likely to miss a certain number of nodules, which is sort of acceptable clinically while missing nodules can influence a lot in final benchmarks and scores for a nodule detection system. More specifically, because all of our annotators are Ph.D. candidates, which can result in even more nodules missing. Our semi-automatic annotation feature can solve this problem to some extent. We first gather some of the CT images and have these images carefully annotated and reviewed by experts. Then we use these images to train a neural network model which can be pretty sensitive in detecting nodules on CT images. After training the model, we use the model to detect nodules on the unannotated images and save the results in the database before the manual annotation work even starts. Then when annotators try to annotate, they can always have a nice reference produced by a deep neural network, and avoid missing some nodules to some extent, and with the enlargement of the dataset, a better model for detecting nodules can be obtained. The experimental results about how semi-automatic annotation manages to outperform the traditional annotation are described in Section “Future work”.

Real-world usages

Annotated data collection

We worked with a medical group from West China Hospital, Sichuan University, who used DeepLNAnno to perform annotations on a pulmonary CT dataset they provided. They now have provided us with two different sets of data, named DeepLN-I and DeepLN-II respectively. The former one contains 1298 CT studies scanned from 208 patients, and the latter one has 204 studies and all of which are thin-layered images with a slice spacing of one millimeter.

They finished annotating and reviewing all the CT studies of DeepLN-I within one month using our three-tier working process and provided us with 2854 annotated nodules. We then used these annotated data to train our neural network model and perform nodule detection on DeepLN-II to make the semi-annotation process possible. They finished annotating and reviewing DeepLN-II with nodule references produced by our model trained on DeepLN-I.

Experients and results

Liao et al. propose a 3D deep neural network architecture which can effectively detect lung nodules from CT scans and the proposed method won the first place in the Data Science Bowl 2017 Competition held by Kaggle [4]. We trained our nodule detection model based on the open source code published by the authors to see if the annotated data is really useful and if the semi-annotation process can really contribute to the benchmarks. Two metrics are adopted to test how our models performed. The sensitivity criterion is the result of detected true positives divided by the number of real nodules. It is used to see if the model can find as many real nodules as possible, which is important as missing nodules can have a serious impact on the patient’s treatment. Besides, we used the Free-Response Receiver Operating characteristics(FROC) analysis as described in [10]. The overall FROC score of our model is defined as the average of the sensitivity at seven predefined false positive(FP) rates: 1/8, 1/4, 1/2, 1, 2, 4 and 8 FPs per scan.

Table 1 The differences in benchmarks with and without the semi-annotation

Metrics	DeepLN-I	DeepLN-II
Sensitivity	0.94	0.97
FP/Scan	19.39	15.33
FROC	0.65	0.71

We used the FROC criterion to evaluate the accuracy of our model for nodule detection.

As shown in Table 1, the model trained on DeepLN-I obtains 94% in sensitivity and 65% in FROC benchmark. With the help of semi-annotation, we managed to increase the benchmarks to 97% and 71% accordingly with the same method for training the model. So semi-annotation can not only make it easier to annotators or reviewers to find the nodules but also make the annotated data more accurate. We are working on our own method to improve our detection model, especially in the metric of FROC. Hopefully, in the near future, a much more suitable nodule detector for our data can be developed.

Moreover, our system makes it possible to give the annotated nodules pathological properties according to pathology reports. We have collected 1169 nodules with pathological labels from the two datasets. The pathological distribution of data is shown in Fig. 7. We used these pathological labels to perform some classification experiments and received astonishing results. We trained a simple 3D convolutional neural network with the two most numerous classes, the adenocarcinoma nodules and the benign nodules and received an evaluation accuracy of 87.5% and F1 score of 89%. We are pretty sure that more works can be done in the future to not only detect where the nodules are but also predict their pathological properties as the number of pathological labels increases.

Future work

A few new features can be added to our DeepLNAnno system to make it even better for annotating nodules on CT images, and other medical annotating systems may also benefit from these ideas.

Adjustable brightness and contrast

In medical image analysis, setting the right parameters can be very crucial for doctors can see these images more clearly and thus perform a better judgment for curing the diseases. For our particular case, when doctors try to find the nodules on CT images, in clinical practice, they often adjust the window width and window level to make the nodules easier to find. But that's based on the commercial softwares they've purchased and can be tricky to adopt to our web-based annotating system. Our system currently uses the default window-level-width values of the pulmonary window for our users and sometimes it may not be capable of handling particular studies. So maybe in the future, this feature can be added to DeepLNAnno for a better annotating experience.

Polygon annotation

DeepLNAnno is originally designed for the task of pulmonary nodule detection, so right now the only on-image annotation function is to draw a rectangle which can exactly wrap up a nodule. Maybe in the future, more accurate nodule segmentation task is demanding to make it possible for our computer-aided system to describe the morphological characteristics of the nodules or to provide a more accurate diameter of nodules which is very important for clinical practice. So polygon annotation feature should be added to our system for annotating the segmentation masks of nodules.

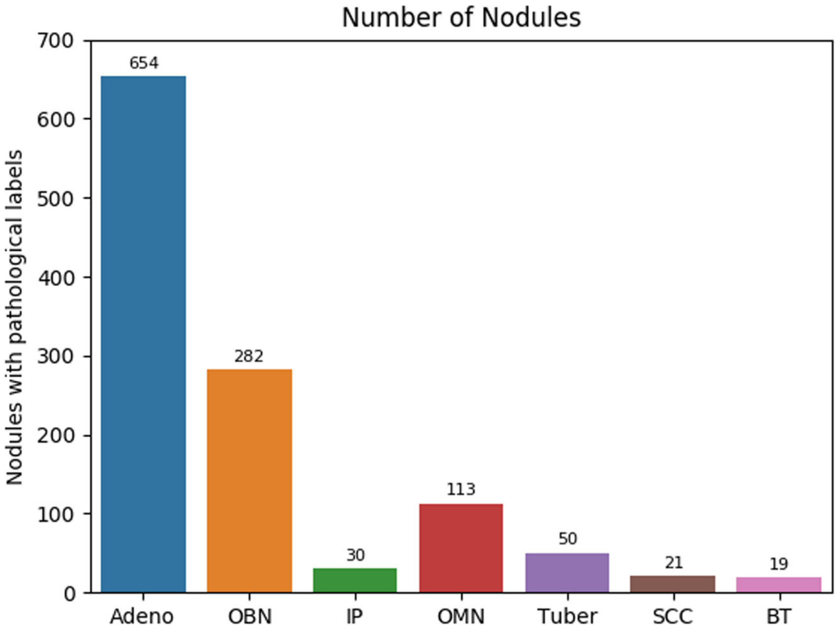


Fig. 7 The number of nodules of each pathological category. The above bar chart shows the number of the nodules of each pathological category, the abbreviations of the x-axis are: adenocarcinoma, other benign nodules, inflammatory pseudotumor, other malignant nodules, tuberculosis, squamous cell carcinoma and benign tumors

Conclusions

The development of an automatic nodule detection system on CT images can improve medical outcomes for lung cancer patients. A vast number of annotated CT studies are required to train more accurate and robust detection models. We develop DeepLNAnno, a web-based system aimed at increasing the efficiency and accuracy of annotating CT images. We also introduce a novel three-tier annotation process and a semi-automatic annotation feature to boost up efficiency while increasing accuracy. DeepLNAnno is scalable and can easily be transformed to annotate other medical data in the DICOM format. We are currently developing a tool on top of DeepLNAnno to annotate 3D ultrasound images. Once the associated patent applications are completed, we may open the source code of DeepLNAnno for data scientists and developers who are interested in constructing their own annotating systems.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant 61432012 and U1435213 and by the Science and Technology Project of Chengdu under Grant 2017-CY02-00030-GX.

Funding This study was funded by the National Natural Science Foundation of China under Grant 61432012 and U1435213 and by the Science and Technology Project of Chengdu under Grant 2017-CY02-00030-GX.

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Armato, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A. et al., The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Med. Phys.* 38(2):915–931, 2011.
2. Chen, W., Zheng, R., Baade, P. D., Zhang, S., Zeng, H., Bray, F., Jemal, A., Yu, X. Q., and He, J., Cancer statistics in china, 2015. *CA Cancer J. Clin.* 66(2):115–132, 2016.
3. Jacobs, C., van Rikxoort, E. M., Twellmann, T., Scholten, E. T., de Jong, P. A., Kuhnigk, J. M., Oudkerk, M., de Koning, H. J., Prokop, M., Schaefer-Prokop, C. et al., Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. *Med. Image Anal.* 18(2):374–384, 2014.
4. Liao, F., Liang, M., Li, Z., Hu, X., and Song, S., Evaluate the malignancy of pulmonary nodules using the 3d deep leaky noisy-or network. arXiv:1711.08324, 2017.
5. Marques, O., and Barman, N., Semi-automatic semantic annotation of images using machine learning techniques. In: *International Semantic Web Conference*, pp. 550–565: Springer, 2003.
6. Pianykh, O. S., *Digital imaging and communications in medicine (DICOM): a practical introduction and survival guide*. Berlin: Springer, 2009.
7. Rowley, H. A., Baluja, S., and Kanade, T., Human face detection in visual scenes. In: *Advances in Neural Information Processing Systems*, pp. 875–881, 1996.
8. Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T., Labelme: a database and web-based tool for image annotation. *Int. J. Comput. Vis.* 77(1-3):157–173, 2008.
9. Setio, A. A. A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., van Riel, S. J., Wille, M. M. W., Naqibullah, M., Sánchez, C. I., and van Ginneken, B., Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks. *IEEE Trans. Med. Imaging* 35(5):1160–1169. 10.1109/TMI.2016.2536809, 2016.
10. Setio, A. A. A., Traverso, A., De Bel, T., Berens, M. S., van den Bogaard, C., Cerello, P., Chen, H., Dou, Q., Fantacci, M. E., Geurts, B. et al., Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *Med. Image Anal.* 42:1–13, 2017.
11. Torralba, A., Russell, B. C., and Yuen, J., Labelme: online image annotation and applications. *Proc. IEEE* 98(8):1467–1484, 2010.
12. Torralba, A., and Sinha, P., Detecting faces in impoverished images. Tech. rep. MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB, 2001.
13. Turk, M., and Pentland, A., Eigenfaces for recognition. *J. Cogn. Neurosci.* 3(1):71–86, 1991.
14. Wenyin, L., Dumais, S. T., Sun, Y., Zhang, H., Czerwinski, M., and Field, B. A., Semi-automatic image annotation. In: *Interact*, Vol. 1, pp. 326–333, 2001.
15. Yushkevich, P. A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J. C., and Gerig, G., User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31(3):1116–1128, 2006.