

Crowdsourcing Lung Nodules Detection and Annotation

Saeed Boorboor¹, Saad Nadeem¹, Ji Hwan Park¹, Kevin Baker², and Arie Kaufman¹

¹Department of Computer Science, Stony Brook University, Stony Brook, NY, USA

²Department of Radiology, Stony Brook Medicine, Stony Brook NY, USA

ABSTRACT

We present crowdsourcing as an additional modality to aid radiologists in the diagnosis of lung cancer from clinical chest computed tomography (CT) scans. More specifically, a complete workflow is introduced which can help maximize the sensitivity of lung nodule detection by utilizing the collective intelligence of the crowd. We combine the concept of overlapping thin-slab maximum intensity projections (TS-MIPs) and cine viewing to render short videos that can be outsourced as an annotation task to the crowd. These videos are generated by linearly interpolating overlapping TS-MIPs of CT slices through the depth of each quadrant of a patient's lung. The resultant videos are outsourced to an online community of non-expert users who, after a brief tutorial, annotate suspected nodules in these video segments. Using our crowdsourcing workflow, we achieved a lung nodule detection sensitivity of over 90% for 20 patient CT datasets (containing 178 lung nodules with sizes between 1-30mm), and only 47 false positives from a total of 1021 annotations on nodules of all sizes (96% sensitivity for nodules >4mm). These results show that crowdsourcing can be a robust and scalable modality to aid radiologists in screening for lung cancer, directly or in combination with CAD algorithms. For CAD algorithms, the presented workflow can provide highly accurate training data to overcome the high false-positive rate (per scan) problem. We also provide, for the first time, analysis on nodule size and position which can help improve CAD algorithms.

1. PURPOSE

Lung cancer is the leading cause of cancer-related deaths in both women and men around the world. Despite its serious outlook, early screening and detection of pulmonary nodules, the precursors of lung cancer, using chest CT scans can dramatically increase the survival rate of lung cancer patients. Nodule detection and chest CT analysis is dependant on the interpreting radiologist and factors such as fatigue, human perception error, image quality and noise, and turn-around time expectations can increase the likelihood of nodule misreads. Moreover, the requirement to search for nodules on every clinical chest CT scan, including those acquired in emergent settings, has resulted in an increase in the volume of scans and thus the number of CT images that require radiologist review. Double reading, computer-aided detection (CAD) and visualization techniques have been proposed to facilitate the screening and to expedite the decision-making process for the radiologists. Even though CAD systems are useful tools for eventual diagnosis, current CAD algorithms have high false positive rates per scan making them less useful for lung nodule detection. Therefore, beyond image-based visualization techniques, CAD and double reading modalities are not extensively used in clinical practice due to limited resources, cost-effectiveness, and lack of general applicability.

Crowdsourcing is a trending modality where simple cognitive tasks, such as image and video annotation, are outsourced to a pool of untrained individuals from an online community. Recent studies on crowdsourcing have shown promising results, especially in the medical domain, with applications in polyp and false positive detection in virtual colonoscopy, annotation and reference correspondence generation in endoscopic images, disease classification, anatomy measurement in CT scans, etc. We leverage the collective intelligence of the crowd in our context of lung nodule detection and annotation to overcome the problems with other modalities.

2. METHOD

Dataset: The crowdsourcing data consists of 20 anonymized chest CT patient scans, from the publicly-available LIDC database, containing 178 nodules between 1-30mm allowing for diverse characteristics, as shown in Table 1. Nodules >4mm were annotated by 5 expert radiologists and nodules <4mm, which are difficult to find, were annotated by at least 2 of these experts, providing ample expert annotations for validation.

Confidential: SPIE Medical Imaging 2018

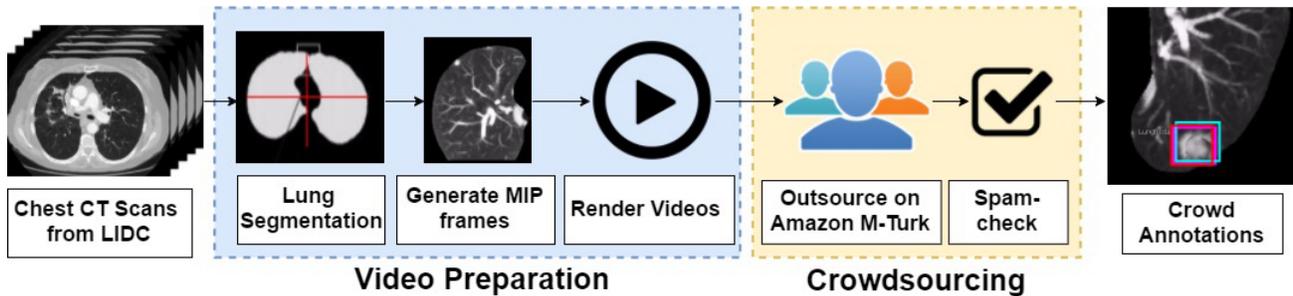


Figure 1. A pipeline of our proposed experimental study: patient chest CT scans are acquired from the LIDC; overlapping TS-MIP videos of segmented lungs are rendered and outsourced to the crowd for lung detection and annotation

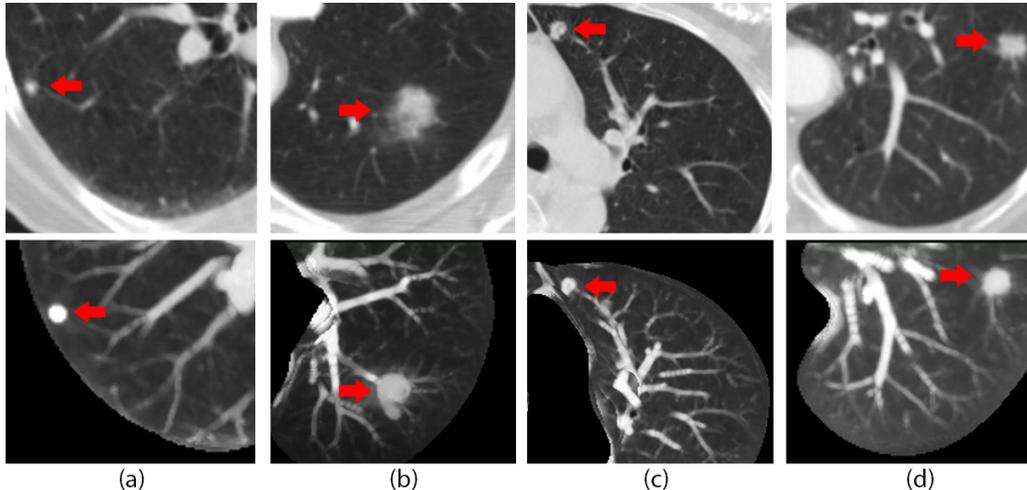


Figure 2. A comparison between a standard CT slice view (top), a TS-MIP of a lung's quadrant as used in this paper (bottom) for: (a) Pleural located nodule; (b) Vessel-attached nodule; (c) Hilar based nodule and (d) Central nodule.

MIP Video Preparation: Prior studies have shown an improvement in lung nodule detection by less-experienced readers when using sliding TS-MIPs in contrast to simple cine viewing of raw CT axial images. Additionally, based on feedback from our in-house user study with 20 non-expert subjects (17 males and 3 females, ages 28 ± 4), on the effective frame size for cine viewing, we learnt that, screening for nodules throughout the entirety of both lungs simultaneously was a frustrating task, prone to distraction by the anatomy outside the lung parenchyma. With input from expert radiologists, we were able to overcome these challenges by (1) automatically segmenting the lung region from the CT scans, (2) dividing the lungs into four quadrants, (3) generating running TS-MIPs of 5 consecutive CT slices for each quadrant (Figure 2), (4) rendering a video sequence of 3 frames per second (fps) for each quadrant, and (5) presenting individual quadrant videos to the crowd for lung nodule detection and annotation.

Crowd Annotations: To reach out to a large and diverse crowd base, we used Amazon Mechanical Turk (MTurk), an internet-based crowdsourcing platform, where cognitive tasks can be distributed to reliable crowd workers at a modest cost. The annotation user interface was integrated into MTurk through an annotation tool called Vatic. For this study, we allotted 10 workers per video on MTurk and awarded \$0.30 to every worker who passed our quality review process, described below. When a crowd worker voluntarily accepts our advertised task, he/she is first introduced to a tutorial about the task. The tutorial consists of three parts: (a) short TS-MIP videos of three different nodule types and sizes, (b) a step-by-step instruction set on how to use the interactive annotation tool, and (c) the review process by which the workers will be paid.

The Annotation Tool: Using the interactive annotation tool, the workers can toggle video controls to watch the short overlapping TS-MIP video segments. While watching the videos, a suspected nodule can be annotated by pausing the video and drawing a bounding box around its region (Figure 3). We also added a small image of a gorilla in our videos which the workers had to annotate; this helped us distinguish the random clickers from the

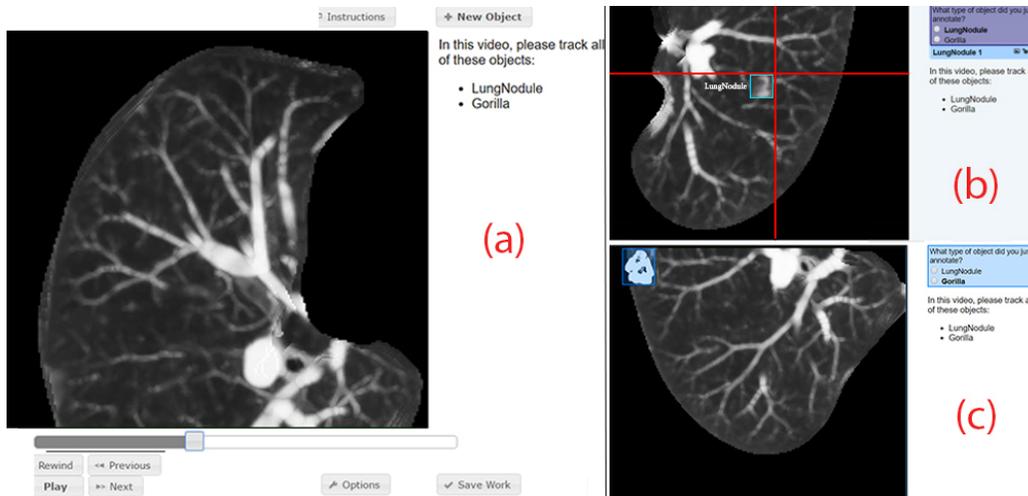


Figure 3. Screen shots of the Vatic annotation tool.(a) The main interface. (b) An example of drawing a bounding box and annotating a nodule. (c) An example of a gorilla that needs to be annotated for quality review.

Table 1. Number of nodules detected by the crowd

Nodule Size (mm in diameter)	No. of Nodules (Ground Truth)	No. of Nodules Detected	Sensitivity
≤ 4	91	78	85.7%
> 4 and ≤ 6	30	28	93.3%
> 6 and ≤ 8	18	17	94.1%
> 8 and ≤ 10	15	15	100%
> 10	24	23	95.8%
Total	178	161	90.4%

proper workers; this helped us distinguish random clickers from proper workers.

3. RESULTS

For the 80 outsourced videos, we allotted 10 workers per video on MTurk. As the workers had the liberty to work on more than one video, we collected the annotations from 143 unique workers. Five crowd attempts were discarded as spam based on our quality review.

The 800 completed tasks contained a total of 1021 annotations. We analyzed the results by comparing the worker annotation with the ground truth annotations. If a worker’s bounding box overlapped with an expert radiologist’s annotation by more than 60%, we considered the nodule detected. Out of 178 nodules in the 20 patient CT chest scans, the crowd detected 161 of these nodules, with total of 47 false-positive annotations, resulting in an overall nodule detection sensitivity of 90.4%. The sensitivities based on the nodule sizes, as shown in Table 1, are as follows: 85.7% for small-sized nodules (≤ 4 mm in diameter), 95.2% for medium-sized nodules (>4 mm and ≤ 10 mm in diameter), and 95.8% in case of large-sized nodules (>10 mm in diameter). Moreover, out of the 10 workers, small-, medium- and large-sized nodules were successfully detected by a mean of 4.2, 8.7, and 9.8 workers respectively, with a median of 4 (σ 2.4), 9 (σ 1.8), and 10 (σ 0.54) respectively.

We further analyzed the sensitivities based on the nodule location and its attachment within the lung region (Table 2). A nodule can be located at the lung periphery or inside the lung region, and it can be attached either to the pleural region (the outer periphery of the lungs), the hilar region (the inner periphery of the lungs that contains the major bronchi, the pulmonary arteries and veins), the vessels, or the central region.

We also analyzed the nodule misreads by the crowd workers (Tables 1 and 2). Our analysis shows that apart from a single miss, the crowd managed to detect all the remaining non-peripheral medium- to large-sized nodules and approximately 95% of the small-sized non-peripheral nodules. However, for the peripheral-located nodules, we observed that for medium- and large-sized nodules, the crowd missed the nodules attached to the hilar region

Confidential: SPIE Medical Imaging 2018

Table 2. Further evaluation of 178 nodules based on location and attachment

Variable		≤4mm	4–6mm	6–8mm	8–10mm	>10mm
Location	Peripheral	42/53 (79.3%)	14/15 (93.3%)	8/9 (88.9%)	8/8 (100%)	13/14 (92.9%)
	Non-peripheral	36/38 (94.7%)	14/15 (93.3%)	9/9 (100%)	7/7 (100%)	10/10 (100%)
Attachment	Pleural	27/32 (84.4%)	12/12 (100%)	5/5 (100%)	6/6 (100%)	8/8 (100%)
	Vessel	31/35 (88.6%)	10/11 (90.9%)	6/6 (100%)	0/0	3/3 (100%)
	Hilar	3/7 (42.9%)	1/2 (50%)	1/2 (50%)	3/3 (100%)	3/4 (75%)
	Central	17/17 (100%)	5/5 (100%)	5/5 (100%)	6/6 (100%)	9/9 (100%)

of the lungs. This was due to the similar color intensities of the bronchi, pulmonary arteries and veins that made these nodule seem like part of the structure within the hilar region; a non-expert worker is not trained enough to distinguish between these structures. The same observation holds true for small-sized nodules as only 43% of the small hilar-attached nodules were detected by the crowd.

We further noticed that for small-sized nodules, the crowd had a low sensitivity for nodules attached to the vessels or to the pleural region of the lungs. The faint grayish taint of these small-sized nodules, attached to the bright branching vessels, made it difficult for a non-expert user to spot in the MIP videos. Moreover, the artifacts formed at the boundaries of the overlapping MIPs due to the growing and shrinking of the lungs in axial traversal resulted in the occlusion of small nodules attached to the pleural region of the lungs. For our 38-nodule patient dataset, the crowd missed only 2 small-sized nodules and had 5 false positives. For the dataset with only 1 nodule (> 10mm in diameter), the crowd successfully annotated the single nodule; however, 9 regions were marked as false positives.

4. NEW WORK TO BE PRESENTED

We leverage the combination of crowdsourcing and MIPS, for the first time, to design and evaluate a novel study for accurate lung nodule annotations from non-experts. We achieve state-of-the-art >90% sensitivity and only 47 false positives from a total of 1021 annotations on nodules of all sizes (96% sensitivity for nodules >4mm); latest 2016 CAD achieved 94% sensitivity and an average of 8 false positives per scan for nodules >4mm. CAD is not used in clinical practice due to the high false-positive rate. The presented workflow can provide highly accurate training data to CAD algorithms to overcome this high false-positive rate problem. We also provided, for the first time, analysis on nodule size and position which can help improve CAD algorithms. Moreover, the use of minimal and automated preprocessing steps in our framework allows for rapid scaling of our crowdsourcing application to hundreds of thousands of lung datasets without any overhead.

5. CONCLUSION

In this work, we presented a crowdsourcing framework for lung nodule detection and annotation to aid radiologists in lung cancer diagnosis. We presented short videos of lungs to 143 unique untrained workers, who after a brief tutorial, annotate 20 chest CT patient datasets containing 178 pulmonary nodules of varying sizes with a detection sensitivity of over 90%. Several interesting insights were presented to help build on this work and open up avenues to incorporate crowdsourcing as an additional tool along with CAD, double reading, etc. in the clinical workflow to assist radiologists in the critical task of lung nodule screening.

6. SUBMISSION STATUS

This is the first and only submission of this work for publication and presentation. It has not been and is not being submitted elsewhere for publication or presentation.