Computational assessment of visual search strategies in volumetric medical images

Gezheng Wen
Avigael Aizenman
Trafton Drew
Jeremy M. Wolfe
Tamara Miner Haygood
Mia K. Markey

Computational assessment of visual search strategies in volumetric medical images

Gezheng Wen,[a,b] Avigael Aizenman,[c] Trafton Drew,[d] Jeremy M. Wolfe,[c,e] Tamara Miner Haygood,[b] and Mia K. Markey[f,g,*]

[a]University of Texas at Austin, Department of Electrical and Computer Engineering, 107 West Dean Keeton, Austin, Texas 78712, United States
[b]University of Texas MD Anderson Cancer Center, Department of Diagnostic Radiology, 1515 Holcombe Boulevard, Houston, Texas 77030, United States
[c]Brigham and Women's Hospital, Department of Surgery, 75 Francis Street, Boston, Massachusetts 02115, United States
[d]University of Utah, Department of Psychology, 380 S 150 E Beh S, Salt Lake City, Utah 84112, United States
[e]Harvard Medical School, Department of Ophthalmology and Radiology, 64 Sidney Street, Cambridge, Massachusetts 02139, United States
[f]University of Texas at Austin, Department of Biomedical Engineering, 107 West Dean Keeton, Austin, Texas 78712, United States
[g]University of Texas MD Anderson Cancer Center, Department of Imaging Physics, 1515 Holcombe Boulevard, Houston, Texas 77030, United States

Abstract. When searching through volumetric images [e.g., computed tomography (CT)], radiologists appear to use two different search strategies: "drilling" (restrict eye movements to a small region of the image while quickly scrolling through slices), or "scanning" (search over large areas at a given depth before moving on to the next slice). To computationally identify the type of image information that is used in these two strategies, 23 naïve observers were instructed with either "drilling" or "scanning" when searching for target T's in 20 volumes of faux lung CTs. We computed saliency maps using both classical two-dimensional (2-D) saliency, and a three-dimensional (3-D) dynamic saliency that captures the characteristics of scrolling through slices. Comparing observers' gaze distributions with the saliency maps showed that search strategy alters the type of saliency that attracts fixations. Drillers' fixations aligned better with dynamic saliency and scanners with 2-D saliency. The computed saliency was greater for detected targets than for missed targets. Similar results were observed in data from 19 radiologists who searched five stacks of clinical chest CTs for lung nodules. Dynamic saliency may be superior to the 2-D saliency for detecting targets embedded in volumetric images, and thus "drilling" may be more efficient than "scanning." © 2016 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.3.1.015501]

Keywords: visual search; saliency map; eye tracking; chest computed tomography; diagnostic error.

Paper 15200R received Oct. 15, 2015; accepted for publication Nov. 23, 2015; published online Jan. 6, 2016.

1 Introduction

With the increasing use of consecutive cross-sectional medical imaging modalities in clinical practice [e.g., computed tomography (CT) and digital breast tomosynthesis], radiologists often need to scroll through a stack of images to get a three-dimensional (3-D) impression of anatomical structures and abnormalities. It is important to understand how radiologists search through such volumetric image data, and how they deploy their visual attention. Identifying efficient search strategies can improve diagnostic accuracy and training of early career radiologists.

Eye tracking can be used to record an observer's gaze positions over an image to elucidate what the observer looks at and when he/she looks at it. This provides an objective method for examining overt visual search patterns. Consequently, eye tracking has been used in a wide range of experiments examining radiologists' search strategies (e.g., lung cancer on conventional radiographs,[1,2] breast masses on mammograms,[3,4] and lesions on brain CT[5]). Visual search is guided by visual attention, which is the cognitive process of selectively attending to a region or object while ignoring the surrounding stimuli.[6] One factor guiding attention is "bottom-up," stimulus-driven

saliency. An item or location that differs dramatically from its neighbors is "salient" and more-salient items will tend to attract attention. Visual saliency maps are topographic maps that represent the conspicuity of objects and locations.[7] If the regions of an image that are identified as salient in a saliency map coincide with the regions on which an observer's eyes fixate, this suggests that the saliency map accurately predicts the observer's deployment of attention.

There are numerous psychophysical and computational models in the literature for generating saliency maps (e.g., Refs. 7–9), and they have been used to address problems such as human scene analysis,[10] video compression,[11] and object tracking.[12] Saliency maps from different labs differ in how they incorporate image and observer characteristics (e.g., color and intensity contrast,[13] visual motion,[14] and biological models[15]). Comparing the success with which various salience models predict eye movements is one source of information describing the image information that is used in observers' search strategies.[16] Thus, e.g., if a saliency map that includes contrast information is a better predictor than a map that does not, we may conclude that contrast is a factor guiding search. As a major goal of medical imaging is to help human observers (e.g., radiologists) locate abnormal targets (e.g., nodules) amidst normal regions (e.g., normal lung tissues), it is reasonable to think that a modality

*Address all correspondence to: Mia K. Markey, E-mail: mia.markey@utexas.edu

will be more effective if it makes targets more salient. In assessing new technology, it would be helpful to have a saliency model that will accurately predict the intrinsic saliency of targets as displayed by the modality. Thus, it would be valuable to examine saliency maps of medical images and to assess the degree to which they capture radiologists visual search patterns.

A limited number of previous studies have assessed the role of saliency maps with medical images (e.g., Refs. 5, 17, and 18), and they have shown positive associations between saliency and observers fixations. Most prior work has examined the correlation between saliency and eye movements in two-dimensional (2-D) medical images. Much less is known about the guidance of eye movements in 3-D volumes of medical image data such as chest CT images viewed in "cine" or "stack" mode. In a recent experiment,[19] 24 radiologists read five stacks of chest CT slices for lung nodules that could indicate lung cancer. While they searched, their eye positions and movements were recorded at 500 Hz. Based on the patterns of eye movements in this dataset, we (TD, JMW) suggested that radiologists could be classified into two groups according to their dominant search strategy: "drillers," who restrict eye movements to a small region of the lung while quickly scrolling through slices; and "scanners," who search over the entirety of the lungs while slowly scrolling through slices. Drillers outperformed scanners on a variety of metrics (e.g., lung nodule detection rate, proportion of the lungs fixated), although the study was not large enough to determine if these differences were reliable and if, indeed, drilling might be superior to scanning as a strategy. In this follow-up study, we further evaluate the two search strategies with the main objective of understanding what image information is actually used when drilling or scanning search strategies are employed. Under the hypothesis that visual saliency maps can predict radiologists' search strategies (e.g., Ref. 20), we examined the relationship of eye movements to computed saliency when radiologists view 3-D volumes of images. We compared saliency maps to observers' gaze distributions collected in two eye-tracking studies using different observer populations and different stimuli. In experiment 1, trained naïve observers searched for target letter T's in faux lung CTs. In experiment 2, experienced radiologists searched for lung nodules in chest CTs. Both studies were approved by Partners Human Research Committee (Boston, Massachusetts), on behalf of the Institutional Review Board of Brigham and Women's Hospital.

2 Materials and Methods

2.1 Experimental Data

2.1.1 Experiment 1: faux lung computed tomography search study

In experiment 1, we monitored the eye position of radiologically naïve observers as they searched through volumetric stacks of simulated approximations to lung CTs. The simulated lung CT images were comprised of 3-D blocks of $1/f^2$ noise and were split into 60 discrete depth levels. The size of each image slice was 500×500 pixels. Target letter T's and distractor L's were inserted into the stack of image slices, and were allowed to spread ± 2 depth levels from the image slice they were placed in, giving them some thickness. T's and L's were at their maximum opacity in the central slice with slices becoming more transparent 1 or 2 slices away from that center. Each volume contained a total of 12 items and a flat hazard function was used to determine how many (0 to 4) T's appeared per image. Targets were absent from 50% of all image volumes. The T's and L's were 20×20 pixels in extent. Targets were separated by at least 10×10 pixels and ± 8 depth levels. An amorphous shape was also inserted into the center of the noise. This shape acted similarly to an anatomical structure (e.g., heart) in a lung CT. This simulated "heart" grew smaller as subjects scrolled through the stack of images. This gave observers an idea of where they were in depth: an anatomical cue that could be used in a similar manner to radiologists' use in clinical lung CT cases. An example stimulus slice is shown in Fig. 1. Observers' task was to find and mark all instances of the target T and to ignore all instances of the distractor L. A "hit" on the target T was defined as a mouse click that occurred within a circle of radius 20 pixels around the center location of the T and within two slices in depth (i.e., a total of five slices would be accepted). Observers controlled the scrolling speed using a scroll bar to move up and down in depth through the stack of noise. Each observer viewed 1 practice trial and 20 experimental trials. Trial order was randomized between observers. Observers were allowed unlimited time to search and could choose when to move on to a new trial. The experiment was implemented in MATLAB® (MathWorks, Natick, Massachusetts) using Psychtoolbox version 3.0.9.[21,22]

Forty naïve observers (26 females, 14 males; mean age = 25.69 years, range = 18 to 51) participated in the study. All observers had normal or corrected-to-normal vision. All
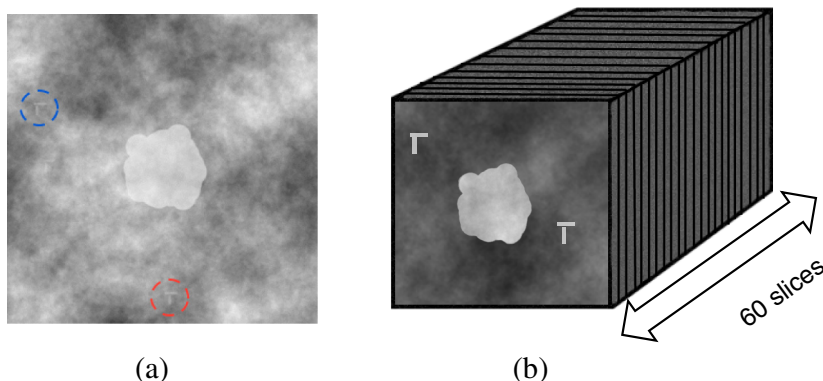

Fig. 1 (a) An example stimuli slice that contains one target T (red circle), one distractor L (blue circle), and a simulated heart at the center; (b) schematic of a simulated 3-D volume with 60 slices in depth.

participants gave informed consent and were compensated for their time. Subjects were randomly assigned to one of three experimental conditions: driller ($n = 14$), scanner ($n = 12$), or no instruction (NI) ($n = 14$). Subjects in the driller and scanner conditions were instructed with a specific search strategy, while subjects in the NI condition were told to choose any search strategy that could be used to find the T's quickly and accurately. All subjects viewed an instructional video at the beginning of the experiment. The video's introduction (which was identical across conditions) guided subjects through a trial, explaining the detection task and how to use the scroll bar to move between depth levels. The critical manipulation in instruction was in the second half of the video, which instructed observers how to search the images: (1) subjects in the driller condition were taught to split the image into four quadrants and to isolate their search to one section of the image at a time; (2) subjects in the scanner condition were taught to thoroughly search each image level that they stopped at. For subjects in the NI condition, no search strategy, including drilling or scanning was described, and they were free to use any search strategy they preferred.

An EyeLink1000 desktop eye-tracker (SR Research, Ottawa, Canada) was used to sample the X and Y position of the eye at 1000 Hz. To minimize eye-tracker error, observers' heads were immobilized in a chin rest at 65 cm from the monitor. The recorded eye position was then coregistered offline with slice/depth plane, which provided the possibility to visualize 3-D scan paths. The monitor for displaying the images was a 19" Mitsubishi Diamond Pro 991 TXM CRT monitor (Mitsubishi Electric, Tokyo, Japan) with a resolution of 1024×768 pixels. To avoid additional factors that could affect performance in the study, the same fixed-display parameters were used for all subjects, and observers were not allowed to change the settings. All experiments were conducted in a darkened room with consistent ambient lighting roughly 15 lux (1 lux = 1 lumen per square meter).

2.1.2 *Experiment 2: chest computed tomography search study*

We also reanalyzed the experimental data from our prior study (TD, JMW) with radiologists.[19] In that study, 24 experienced radiologists were asked to read five stacks of clinical chest CT slices drawn from the Lung Image Database Consortium,[23] and the search task was to identify lung nodules that could indicate lung cancer. Radiologists' eye positions and movements were sampled with the same eye-tracking device at 500 Hz. Similar data collection methods as described in Sec. 2.1.1 were used in this prior study, and described in detail in Ref. 19. Based on the qualitative analysis results presented in Ref. 19, there were a total of 19 drillers and 5 scanners, of which 3 drillers and 2 scanners were excluded from the present analysis due to incomplete eye tracking data and/or click data. As a result, there were a total of 16 drillers and 3 scanners.

2.2 *Saliency Maps*

We now describe two saliency models used to assess observers' search strategy through 3-D volumes. The two saliency models were designed to capture different forms of image characteristics that may be used by observers in their search. The first of these is a dynamic motion saliency model. The second is a more standard model that only considers 2-D image information. For each experimental trial, saliency maps were computed using the two saliency models. These were then compared with the ground truth of the eye-tracking data.

2.2.1 *Three-dimensional dynamic motion saliency*

Anecdotally, experienced radiologists often say that when quickly scrolling back and forth through 3-D image volumes, they are attracted to objects flitting in and out of the visual field, i.e., popping out. For example, they perceive that lung nodules tend to pop in and out of visibility while anatomical structures like blood vessels persist, gradually changing and shifting position across many slices.[19,24] A large number of psychology studies have shown that the human visual system can easily identify motion in a scene (e.g., Refs. 25–27), and that the human brain takes advantage of its sensitivity to optical flows (i.e., the pattern of apparent motion of objects in a visual scene caused by the relative motion between an observer and the scene). For example, optical flows help to control human walking.[28–31] Moreover, abrupt appearance of an object effectively captures observers' attention (e.g., Refs. 32 and 33). Therefore, we developed a dynamic saliency model that was designed to incorporate the unique dynamics produced when observers search through stacks of medical image slices. Unlike viewing a static image or a video that plays at a fixed frame rate, more complexity is added by the fact that observers have control of such as scrolling direction and speed.

Inspired by research on salient motion detection in crowd scenes,[34] our dynamic motion saliency map starts with an optical flow estimation using the classical Horn–Schunck algorithm.[35] The objective was to ascribe higher saliency to motion flows that deviate from the normal dominant flows, which is expected to reflect the observation that nodules pop out from anatomical backgrounds. As observers can scroll two directions in depth (i.e., move forward from the first slice to the 60th slice, or move backward from the 60th slice to the first slice), two optical flow fields were estimated to capture different motion dynamics.

Let us denote $G_{x,y,j}$ as the intensity at pixel location (x, y) at slice j, and denote $M_{x,y,j}$ and $P_{x,y,j}$ as the magnitude (i.e., length) and phase (i.e., direction) of the associated flow vectors (V_x, V_y)

$$M_{x,y,j} = \sqrt{V_x^2 + V_y^2}, \tag{1}$$

$$P_{x,y,j} = \arctan\left(\frac{V_x}{V_y}\right). \tag{2}$$

At each slice j, two pairs $(M_{x,y,j}^F, P_{x,y,j}^F)$ and $(M_{x,y,j}^B, P_{x,y,j}^B)$, where F is short for forward and B is short for backward, were computed to indicate the direction and motion velocity in that direction for forward and backward scrolling, respectively. Intuitively, if a region of the slice has irregular motion direction and/or velocity magnitude that significantly deviates from the rest of the slice, this region is considered to be salient. To obtain the motion saliency maps S_j^F and S_j^B for the slice, we employed a similar method to that proposed in Ref. 34 on magnitude images M_j^F and M_j^B. For the purpose of illustration, let us focus on M_j^F, keeping in mind that the same steps were applied to M_j^B. We first computed the Fourier transform of M_j^F and obtained the real and imaginary components of the spectrum

$$R(f) = \Re[\mathcal{F}(M_j^F)], \tag{3}$$

$$I(f) = \Im[\mathcal{F}(M_j^F)]. \tag{4}$$

Then, we calculated the log spectrum $L(f)$ and applied a local 3×3 averaging filter g on $L(f)$ to obtain a smoothed log spectrum $L'(f)$ that estimated background motion components

$$L(f) = \log[R(f)], \tag{5}$$

$$L'(f) = g(f) * L(f). \tag{6}$$

Finally, we performed the inverse Fourier transform on the spectrum residual[9] [i.e., difference between the log spectrum $L(f)$ and the smoothed version $L'(f)$] to obtain the final saliency map S_j^F in the spatial domain

$$S_j^F = \mathcal{F}^{-1}[e^{L(f)-L'(f)+I(f)}]. \tag{7}$$

To combine the two sets of dynamic information captured by S_j^F and S_j^B, we took the maximum of saliency values in S_j^F and S_j^B at each pixel location (x, y) to form the final 3-D dynamic motion saliency map $S_j^{3\text{-D}}$.

2.2.2 Two-dimensional saliency

To deploy a more standard salience model, we computed a saliency map for each 2-D slice using graph-based visual saliency (GBVS).[8] GBVS is a 2-D bottom-up saliency model that attempts to predict saliency solely based on the image features of the slice. It has been demonstrated to perform well in predicting observers fixations when viewing 2-D medical images (e.g., chest x-ray images and retinal images[18]). To reduce the GBVS-induced bias toward the central region of the image,[8] we scaled the computed saliency map by performing element-wise multiplication on an inverse-bias map. Let us denote $S_j^{2-\text{D}} = \mathbf{B} \odot \text{GBVS}(G_j)$ as the computed 2-D saliency map for slice j, where $\mathbf{B}$ is the inverse-bias map and $\odot$ denotes element-wise multiplication.

2.2.3 Top-down factor

In addition to the image-driven features, human visual attention is also influenced by goal-driven top-down features.[16] In the context of medical images, top-down features correspond to the knowledge and expertise of radiologists, and prior expectation on the task. Hence, for both models, in order to incorporate some of the role of top-down knowledge, we also performed automatic segmentations before applying the saliency models to exclude some image regions from being salient. Specifically, since observers were informed that no target or distractor would appear within the heart region of the faux lung CTs, we segmented the heart on each slice so that the regions inside the simulated heart were not allowed to be salient. The heart segmentation was achieved by masking out the largest connected region identified by Canny edge detection.[36] Similarly, since lung nodules can only occur in the lung regions of chest CTs, we performed lung segmentation on each slice to restrict the salient regions to inside the two lungs.

2.3 Evaluation

To quantify the correlation between the computed saliency maps and observers' visual attention, we conducted a receiver operating characteristic (ROC) analysis. We considered the saliency map, S, as a binary classifier, which classified a given percent of image pixels as fixated and the rest as not fixated.[16] The fixations of the observers were treated as the ground truth. The fixated area for each fixation was a circle of radius 20 pixels (i.e., 1 degree of visual angle[37]) around the fixation. A pixel was counted as a hit if the saliency map classified it as fixated, and it was actually fixated by the observer. A pixel was counted as an FA if the saliency map classified it as fixated but it was never fixated by the observer. Changing thresholds of the saliency value resulted in different classifiers, which yielded different pairs of sensitivity and specificity. In this way, an ROC curve was constructed, and area under the ROC curve (AUC) was calculated as the figure of merit to indicate how well observers' fixations overlapped with the saliency map. For each observer, a single mean AUC value was obtained by averaging the AUCs of each experimental trial. AUCs of drillers and scanners achieved by the two saliency models were compared in order to identify which type of image information was used by the two groups of observers.

3 Results and Discussions

3.1 Experiment 1: Faux Lung Computed Tomography Search Study

3.1.1 Observers' search strategies

As described in Sec. 2.1.1, the subjects in the faux lung CT search study were grouped into three different conditions and were given different instructions on which visual search strategies to use. As the primary interest of the study was to compare drilling and scanning strategies, NIs, who did not receive either drilling or scanning instruction, were not relevant and thus excluded from further analysis. (More details are available in Appendix) We first checked observers' 3-D scan paths to determine whether their search strategies matched the received instructions. We divided each slice into four quadrants of equal sizes, and drew depth-by-time plots with eye position indicated by color-coded quadrants. A subject is qualitatively classified as a driller if they hold a constant eye position on one subsection of the slice while scrolling through depth, while a scanner tends to search the entire slice before scrolling in depth.[19] The depth-time plots in Fig. 2 show that the subjects followed our instructions and employed the corresponding visual search strategies: drillers drilled, while scanners scanned. Hence, we were confident in our ability to classify observers' search strategies based on the training instructions. One driller and two scanners who had either poor eye-tracking calibration or unreasonable behavioral data [e.g., extremely low hit rates and high false-alarm (FA) rates] were excluded from the analysis. As a result, there were a total of 13 drillers and 10 scanners.

3.1.2 Saliency maps

Given the distinct designs of the two saliency models described in Sec. 2.2, the resulting saliency maps are expected to highlight different useful image information that observers might utilize in their search strategies. Figure 3 shows example saliency maps overlapped upon a stimulus slice. It can be observed that the 2-D
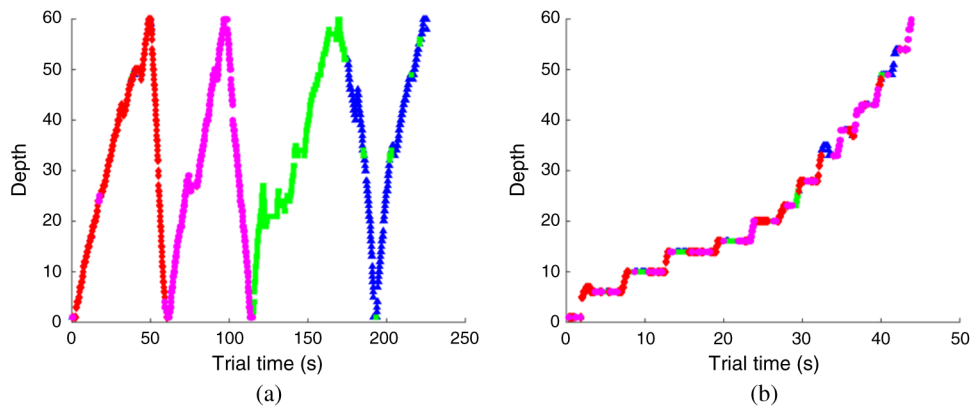
(a) (b)

Fig. 2 Example depth-time plots of observers' eye positions. The four colors indicate four quadrants of the slice. (a) Drillers focused on one quadrant at a time, and they had multiple runs of scrolling in both forward and backward viewing directions to cover all the quadrants. (b) Scanners went over multiple quadrants before they moved to the next slice.
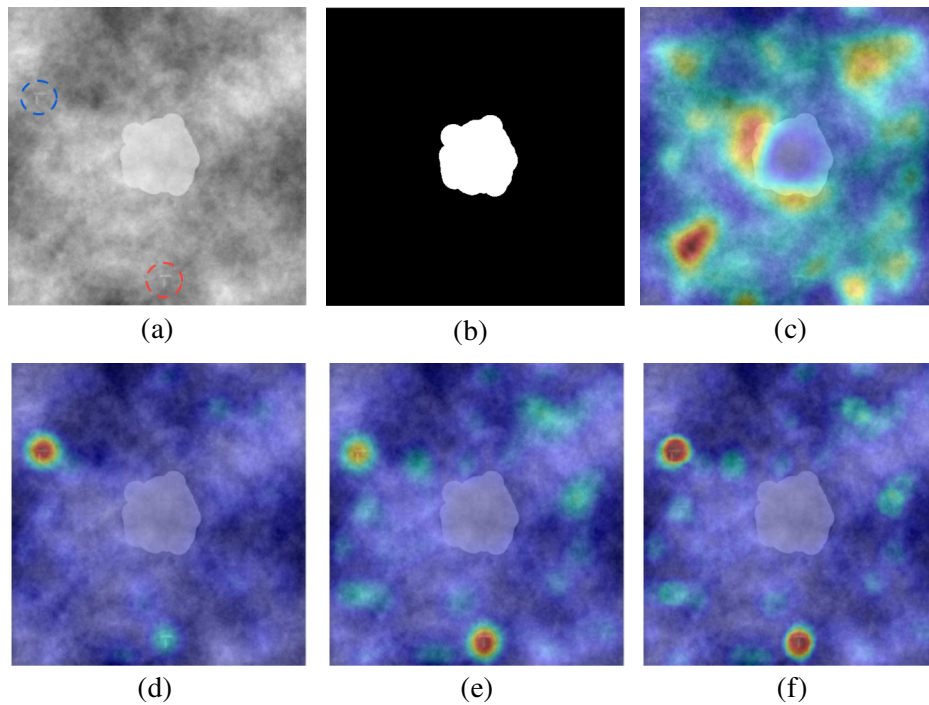

(a) (b) (c)

(d) (e) (f)

Fig. 3 Example stimulus slice, and calculated 2-D saliency maps and dynamic saliency maps, overlapped upon the slice. Saliency maps are represented as heat maps, and the color indicates the saliency value at that location: red is more salient than blue. (a) The stimuli with one target T (red dashed circle) and one distractor L (blue dashed circle). (b) Heart segmentation: any position within the mask was excluded from being salient. (c) 2-D saliency map: neither T or L is salient. (d) Forward dynamic saliency map: L is salient while T is not salient. (e) Backward dynamic saliency map: T is salient while L is not. (f) Fused dynamic saliency map: both T and L are salient. (d) and (e) An example that the two directions of scrolling may provide different dynamic information when observers go back and forth through the volume in depth. (c) The 2-D saliency map did not do well in identifying objects of interest for this particular example.

saliency map in Fig. 3(c) is visually different from the dynamic saliency maps in Figs. 3(d)–3(f). In this particular example, neither object of interest (i.e., target T or distractor L) was highlighted by the 2-D saliency map, while the dynamic saliency maps performed better. Comparing the saliency regions in Figs. 3(d) and 3(e), we can see that only the distractor L is relatively salient in (d) while only the target T is relatively salient in

(e). This agrees with our expectations that the two saliency maps S_j^F and S_j^B capture different dynamic information contained in forward and backward scrolling, which may be one of the benefits of the drilling strategy. That may also explain why driller observers go back and forth multiple times through each quadrant of the slices in depth. As shown in Fig. 3(e), after fusing the two dynamic saliency maps, the final dynamic saliency map

highlights both objects. Table 1 summarizes the mean saliency of the objects in the 2-D saliency map and the dynamic saliency map. Note that the saliency values have been normalized into the range of [0,1] for easy interpretation. Given the mean saliency of 0.71 for target T's and 0.67 for distract L's, it is clear that most of the objects of interest were highly salient in the dynamic saliency map. With the mean of saliency around 0.5, the 2-D saliency map was less effective than the dynamic saliency map in highlighting the objects of interest [$t(251) = 15.8$, $p < 0.01$]. All statistical tests used in the study were two-sided tests unless otherwise noted. Hence, it is reasonable for us to hypothesize that it would be easier for the observers who utilized more dynamic saliency to identify T's and L's in their search process than the observers who relied more on the 2-D saliency.

a. Did saliency maps predict observers' fixations?

We computed AUCs to evaluate the value of saliency maps in terms of predicting observers' fixations. The goal of these evaluations was to identify which type of information drillers and scanners primarily relied on during their visual search. Recall that the 3-D dynamic saliency map captured the dynamics of scrolling through the stack of slices while the 2-D saliency map was built purely upon 2-D image characteristics and no interactions across adjacent slices were incorporated. We hypothesize that the dynamic saliency map is a better predictor of the drilling search strategy, while the 2-D saliency map is a better predictor of the scanning search strategy. That is equivalent to the assumption that drillers use more 3-D dynamic information while scanners use more 2-D information.

Table 2 summarizes the AUCs achieved by the dynamic saliency map and the 2-D saliency map for both drillers and scanners. We performed mixed repeated measures ANOVA to compare the effects of the between-subject factor "search strategy," and the within-subject factor "saliency model" on the AUCs. The results suggested that there was a statistically significant interaction between the search strategy and the saliency model [$F(1,21) = 81.6$, $p < 0.01$]. Hence, it is valuable to conduct follow-up *post hoc* tests to determine whether there were any simple main effects, and if there were, what these effects were.[38]

- AUC with the dynamic saliency: drillers versus scanners

Table 2 Average AUCs achieved by the dynamic saliency map (top row of the table) and the 2-D saliency map (bottom row of the table) for drillers and scanners, respectively. There was a significant interaction between the factor search strategy and the factor saliency model [$F(1,21) = 81.6$, $p < 0.01$]. For the dynamic saliency map, the AUC for the drillers is statistically higher than the AUC for scanners [$t(21) = 2.46$, $p = 0.03$]. For the 2-D saliency map, the AUC for the scanners is statistically higher than the AUC for drillers [$t(21) = 8.61$, $p < 0.01$].

	Drillers	Scanners
AUC (dynamic saliency)	0.69	0.65
AUC (2-D saliency)	0.63	0.74

The top row of Table 2 summarizes the AUCs achieved by the dynamic saliency map for both drillers and scanners. The mean AUC was obtained as an average over all observers in the group and all the 20 trials of the images they read. The subjects were grouped into drillers and scanners based on the training instructions they received. The mean AUC for the 13 drillers was 0.69 while the mean AUC for the 10 scanners was 0.65. A two-sample *t*-test showed that the AUC for the drillers were statistically higher than the AUC for the scanners [$t(21) = 2.46$, $p = 0.03$], which suggests that the drillers used more dynamic motion information than the scanners. Figure 4(a) plots the AUC distributions for the drillers (i.e., red diamonds) and the scanners (i.e., blue squares). It is clear that most of the AUCs for the drillers are located at higher positions of the plot than the AUCs for the scanners.

- AUC with the 2-D saliency: drillers versus scanners

The bottom row of Table 2 summarizes the AUCs achieved by the 2-D saliency map for both drillers and scanners. The mean AUC for the 10 scanners was 0.74 while the mean AUC for the 13 drillers was only 0.63. Similarly, the result of a two-sample *t*-test showed that the AUC for the scanners was statistically higher than the AUC for the drillers [$t(21) = 8.61$, $p < 0.01$], which suggests that the scanners used more 2-D saliency information than the drillers. Figure 4(b) plots the AUC distributions for the drillers (i.e., red diamonds) and the scanners (i.e., blue squares). Similarly, it is clear that all of the AUCs for the scanners are located at higher positions than the AUCs for the drillers.

- AUC with the dynamic saliency versus AUC with the 2-D saliency

As another way to assess the effectiveness of the saliency models for explaining search behaviors, we compared the AUCs within each search strategy group (Table 2). In other words, for drillers, we compared the AUCs obtained by predicting their fixations using the dynamic saliency maps to the AUCs obtained by predicting their fixations using the 2-D saliency maps. Paired *t*-tests were used to identify which of the two saliency models achieved

Table 1 Average saliency of objects (i.e., targets T's and distractors L's) in the dynamic saliency maps and the 2-D saliency maps. With an average saliency around 0.7, the objects were highly salient in the dynamic saliency maps. However, the 2-D saliency maps were not as effective in highlighting those objects of interest.

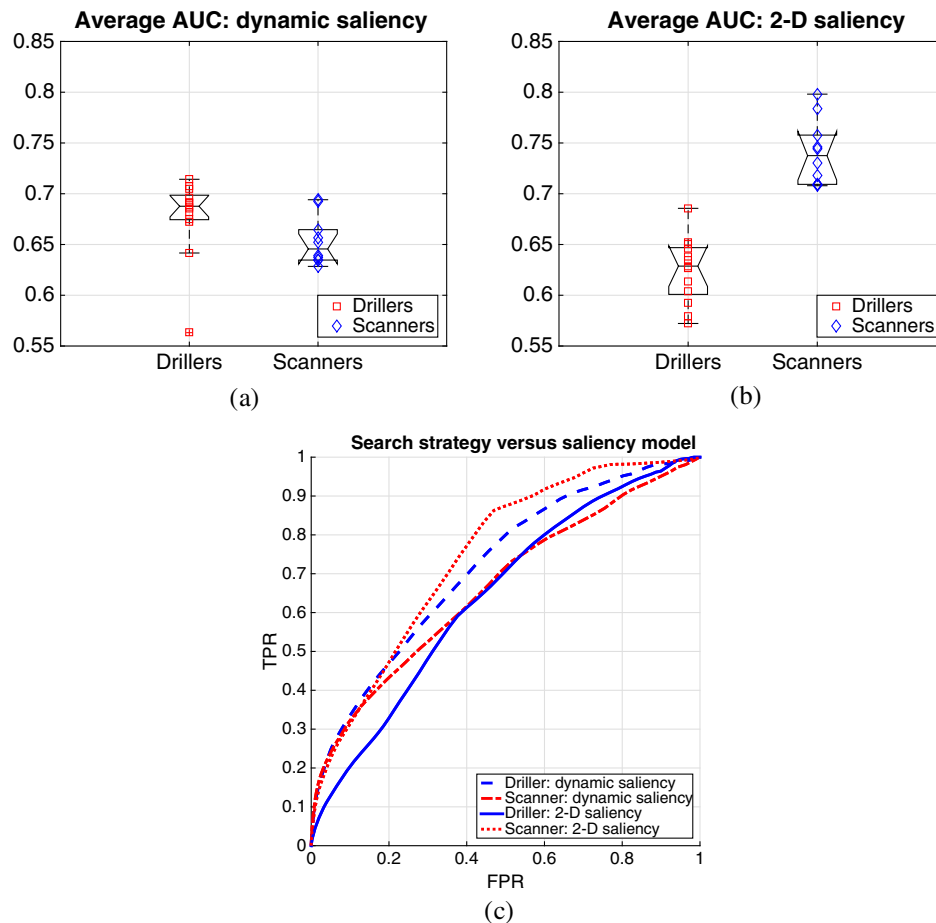|	Mean saliency of target T's	Mean saliency of distractor L's
Dynamic saliency map	0.71	0.67
2-D saliency map	0.52	0.50

Fig. 4 AUC scores on predicting fixations for 13 drillers (red squares) and 10 scanners (blue diamonds) using (a) the dynamic saliency map and (b) the 2-D saliency map, and (c) summary ROC curves. In (a) with the dynamic saliency map, most of the AUCs for the drillers are located at higher positions of the plot than the AUCs for the scanners, which suggests that drillers may have used more dynamic information than scanners did. In (b) with the 2-D saliency map, all the AUCs for the scanners are located at higher positions of the plot than the AUCs for the drillers, which suggests that scanners may have used more 2-D information than drillers. Summary ROC curves in (c) show the interactions between the two search strategies and the two saliency models. For the drillers, the ROC curve with the dynamic saliency (i.e., blue-dashed line) is above the ROC curve with the 2-D saliency (blue solid line). For the scanners, the ROC curve with the 2-D saliency (i.e., red-dotted line) is above the ROC curve (i.e., red dash-dotted line) with the dynamic saliency.

a higher AUC for each search strategy, respectively. For drillers, the AUC with the dynamic saliency was 0.69, which was significantly higher than the AUC with the 2-D saliency map 0.63 [$t(12) = 5.69$, $p < 0.01$]. For scanners, the AUC with the 2-D saliency map was 0.74, which was significantly higher than the AUC 0.65 with the dynamic saliency map [$t(9) = 5.96$, $p < 0.01$]. These results were consistent with our assumptions that drillers used more 3-D dynamic information while scanners used more 2-D information.

Figure 4(c) plots the summary ROCs that show the interactions between the two search strategies and the two saliency models. Data from the 13 drillers and the 10 scanners reading the 20 trials were used to generate these curves. Similar to the non-parametric method used in multireader and multicase studies for averaging multiple ROC curves,[39] each of the summary curves was obtained by computing the

arithmetic average of the sensitivity at each specificity value.[40] For the drillers, the ROC curve with the dynamic saliency is generally above the curve with the 2-D saliency, and hence has a higher AUC. On the other hand, for the scanners, the ROC curve with the 2-D saliency is generally above the curve with the dynamic saliency, and hence has a higher AUC.

When we compared the AUCs obtained with the two saliency models at the individual level, the results for 12 (out of 13) drillers, and all 10 scanners were consistent with the general conclusions (i.e., the drillers were better predicted by the dynamic saliency maps, while the scanners were better predicted by the 2-D saliency maps). We suspect that the one driller who was better predicted by the 2-D saliency map was using a different strategy than the other drillers, but we would need more data to support this assertion.

b. Does saliency influence observers' detection performance?

Based on the results in Sec. 3.1.2(a), we argue that the actions of drillers and scanners were best modeled by the dynamic saliency model and the 2-D saliency model, respectively, and thus drillers and scanners were utilizing different image information in their search. Moreover, as shown in Table 1 and Fig. 3, objects of interest may have different saliency values in the 2-D saliency map and the dynamic saliency map. Hence, we are interested in whether the saliency of objects influenced observers' detection performance. (For example, is a less-salient target T more likely to be missed by the observer?) Depending on whether the observer found and correctly marked the target, the target T's were grouped into miss T's and hit T's, where a mark was defined as a hit T if the subject made a click that occurred less than 20 pixels (in both X and Y coordinates) away from the true location of T. Marks on the distractor L's were considered FA L's. Any other mark that was not either on a target T or distractor L was considered an accidental click, and excluded from analysis. We calculated the average saliency within a circle of radius 20 pixels around the location of the objects. The saliency values were normalized into the range of [0,1] for easy interpretation.

Table 3 summarizes the average saliency of the found targets (hit T's), and the missed targets (miss T's) on the two saliency maps. As the drillers tend to use primarily the dynamic information while the scanners used primarily the 2-D information, the average saliency of objects was calculated with the dynamic saliency map for the drillers, and with the 2-D saliency map for the scanners, respectively. For the drillers, hit T's had an average dynamic saliency of 0.72, while miss T's had an average dynamic saliency of 0.66. For the scanners, hit T's had an average 2-D saliency of 0.55 while miss T's had an average 2-D saliency of 0.47. It can be observed that for both scanners and drillers, the saliency of hit T's was statistically higher than that of miss T's [$t(219) = 3.9$, $p < 0.01$ for the drillers, and $t(168) = 2.89$, $p < 0.01$ for the scanners]. This observation indicates a positive answer to our question: the saliency of targets influenced observers' detection performance, and less-salient targets were more likely to be missed by the observers, regardless of their search strategies. On the other hand, if we compare the average saliency of hit T's and FA on L's, we can observe that for drillers, the average saliency of hits and FAs were reliably different [$t(390) = 2.11$, $p = 0.02$], while for scanners, they were not statistically different [$t(241) = 0.68$, $p = 0.25$]. This suggests that for drilling and scanning search strategies, relying on different image information may result in different types of diagnostic errors.[19] For example, given the larger difference between the saliency of hit T's and miss T's, scanners may be more prone than drillers to search errors,[41] as they are more likely to miss less-salient targets.

3.2 Chest Computed Tomography Search Study

As saliency seemed to reflect the differences between drillers and scanners in the faux lung CT study, we conducted the same analysis on a previously reported chest CT search study.[19] The goal was to examine whether the strategy differences among experienced radiologists could be explained through saliency in a similar manner as in experiment 1.

Figure 5 shows example saliency maps overlapped upon one slice from a clinical chest CT. Lung segmentations were performed first to restrict possible salient regions to inside the lungs. Similar to Fig. 3, in this particular example, the saliency of the nodules was relatively low in the 2-D saliency map [Fig. 5(b)] while the nodules were relatively salient in the dynamic saliency map [Fig. 5(c)]. Table 4 shows that the mean saliency of nodules was significantly higher in the dynamic saliency map than in the 2-D saliency map [$t(51) = 2.67$, $p < 0.01$]. Hence, search strategies utilizing dynamic saliency may be superior to strategies relying on 2-D saliency for finding nodules in clinical CT images. This sample of radiologists was not instructed as to how to search through the chest CT volumes, but interestingly, less than 1/6 of these radiologists adopted a "scanning" strategy. This observation may also support our hypothesis that incorporating the dynamics of scrolling through the stack of CT slices may be preferred in this particular type of search task.

a. Does saliency predict radiologists' fixations?

Table 5 summarizes the AUCs achieved by the dynamic saliency map and the 2-D saliency map for both drillers and scanners. We performed mixed repeated measures ANOVA to compare the effects of the between-subject factor "search strategy," and the within-subject factor "saliency model" on the AUCs. The results suggested that there was a statistically significant interaction between "search strategy" and "saliency model" [$F(1,17) = 8.47$, $p = 0.01$]. We then conducted follow-up analysis for any simple main effects.

- AUC with the dynamic saliency: drillers versus scanners

The top row of Table 5 summarizes the AUCs achieved by the dynamic saliency map for predicting radiologist fixations. The mean AUC was obtained as an average over all radiologists in the group and the five cases they read. The subjects were grouped as drillers and scanners based on the qualitative analysis presented in Ref. 19 The mean AUC for the drillers was 0.73 while the mean AUC for the
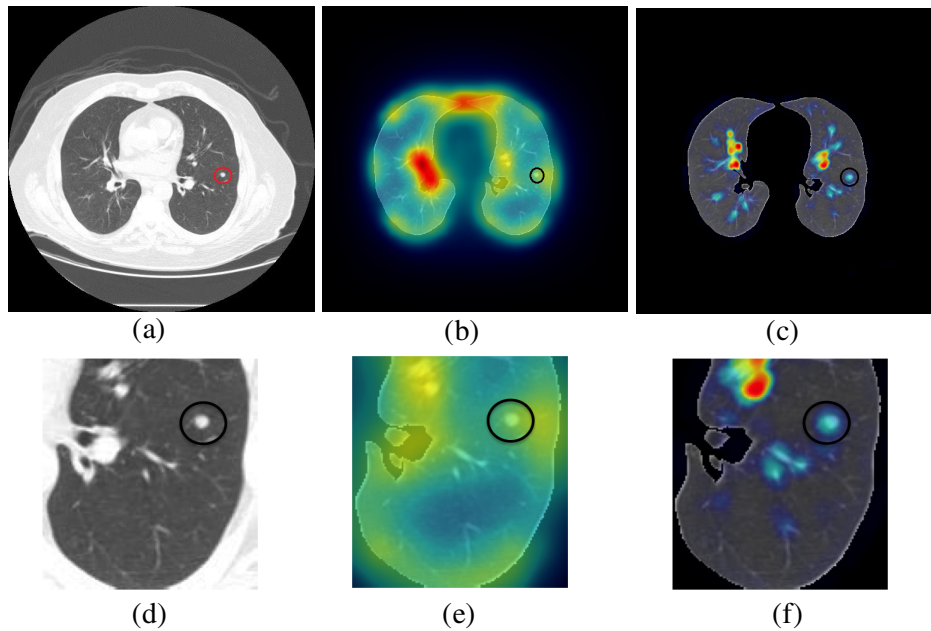
Table 3 Average saliency of the found targets (hit Ts) and the missed targets (miss Ts) on the dynamic saliency map (for drillers) and the 2-D dynamic saliency map (for scanners). The saliency values were normalized into the range of [0,1]. For both scanners and drillers, the saliency of hit T's is higher than that of miss T's. For drillers, the average saliency of hits and FAs was statistically different. For scanners, it was not statistically different.

	Hits	Misses	False alarms
Drillers (dynamic saliency)	0.72	0.66	0.65
Scanners (2-D saliency)	0.55	0.47	0.52

(a)   (b)   (c)

(d)   (e)   (f)

Fig. 5 Example slice of chest CTs, and 2-D saliency map and dynamic saliency map overlaid upon the slice. (a) The original slice with a red circle highlighting the nodule in the right lung; (b) 2-D saliency map with a black circle highlighted nodule; (c) 3-D dynamic saliency map with the black circle highlighted nodule; (d), (e), and (f) nodule-region magnified version of (a), (b), and (c), respectively. All the saliency maps are represented as heat maps. The saliency of the nodule is relatively low in (b) while the nodule is salient in (c). This shows that the dynamic saliency and the 2-D saliency highlighted different image information that may be used by radiologists in their search.

Table 4 Average saliency of nodules in the dynamic saliency map and the 2-D saliency map. The mean saliency of nodules in the dynamic saliency map was higher than that in the 2-D saliency map, and the two values were statistically different.

	Mean saliency of nodules
Dynamic saliency map	0.38
2-D saliency map	0.33

Table 5 Average AUCs achieved by the dynamic saliency map (top row) and the 2-D saliency map (bottom row).

	Drillers	Scanners
AUC (dynamic saliency)	0.73	0.70
AUC (2-D saliency)	0.71	0.77

scanners was 0.70. Figure 6(a) plots the AUC distributions for the drillers (i.e., red diamonds) and the scanners (i.e., blue squares). Although the AUC for the drillers was not statistically higher than the AUC for the scanners [$t(17) = 0.78$, $p = 0.44$] perhaps due to the underpowered sample size of radiologists (16 drillers and 3 scanners), the difference between the two AUCs suggests that the driller

radiologists might use more dynamic information than the scanner radiologists, consistent with our findings in the faux CT search study. Any more definitive answer awaits more data, in particular from more scanners.

- AUC with the 2-D saliency: drillers versus scanners

The bottom row of Table 5 summarizes the AUCs achieved by the 2-D saliency map. As before, the subjects were grouped as drillers and scanners. The mean AUC for the scanners was 0.77 while the mean AUC for the drillers was only 0.71. The AUC for the scanners was statistically higher than the AUC for the drillers [$t(17) = 3.83$, $p < 0.01$], which suggests that the scanner radiologists might use more 2-D information than the driller radiologists. Figure 6(b) plots the AUC distributions for the drillers (i.e., red diamonds) and the scanners (i.e., blue squares). It is clear that all of the AUCs for the scanners are located at higher positions than the AUCs for the drillers. This agrees with our findings in the faux CT search study.

- AUC with the dynamic saliency versus AUC with the 2-D saliency

We compared the AUCs within each search strategy group (Table 5) to assess the effectiveness of the saliency models for explaining search behaviors. In other words, for driller radiologists, we compared the AUCs obtained by predicting their fixations using the 2-D saliency maps to the AUCs obtained by predicting their fixations using the dynamic
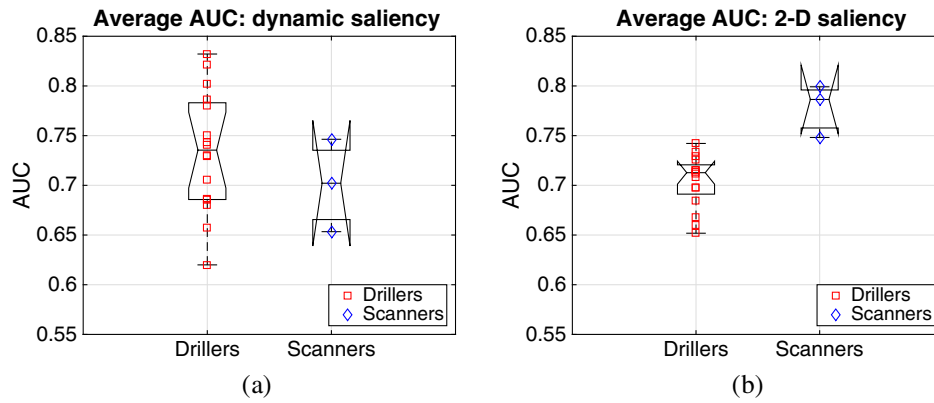
Fig. 6 AUC scores on predicting fixations for 16 drillers (red squares) and 3 scanners (blue diamonds) using the dynamic saliency map (a) and the 2-D saliency map (b). In (a) with the dynamic saliency map, most of the AUCs for the drillers are located at higher positions of the plot than the AUCs for the scanners, which suggests that drillers may have used more dynamic information than scanners did. In (b) with the 2-D saliency map, all the AUCs for the scanners are located at higher positions of the plot than the AUCs for the drillers, which suggests that scanners may have used more 2-D information than drillers.

saliency maps. The same analysis was performed for the scanners. For drillers, the AUC with the dynamic saliency was not statistically higher than the AUC with the 2-D saliency $[t(15) = 1.26, \ p = 0.22]$. Similarly, for scanners, the AUC with the 2-D saliency was not statistically higher than the AUC with the dynamic saliency $[t(2) = 1.97, \ p = 0.18]$. While not all of these comparisons yielded statistically reliable differences, the trends are broadly consistent with the faux lung CT data: scanners tend to use primarily 2-D information in their search, whereas drillers appear to use more dynamic information. Again, any more definitive answer awaits more data, in particular from more scanners.

Another interesting observation from Fig. 6 is that for driller radiologists, the AUC with the dynamic saliency maps [i.e., red squares in Fig. 6 (a)] has higher variability than that with the 2-D saliency map [i.e., red squares in Fig. 6(b)]. This may be caused by the differences in areas of expertise and years in practice (e.g., radiologists with more experience with chest CT may be more adapted to the drilling strategy than less experienced radiologists). It is also possible that some of the driller radiologists were able to extract other types of dynamic information that were not incorporated in the current design of the dynamic saliency model.

Table 6 Average saliency of the found nodules (hits) and the missed nodules (misses). The values of saliency were normalized into the range of [0,1]. For both scanners and drillers, the saliency of the found nodules was higher than that of the missed nodules.

	Hits	Misses	False alarms
Drillers (dynamic saliency)	0.40	0.35	0.44
Scanners (2-D saliency)	0.44	0.33	0.43

Clearly, further research is necessary to better understand the causes of such high variability.

b. Does saliency influence observers' detection performance?

Nodules were categorized into misses, hits, and FAs based on their behavioral response. We then calculated the average saliency within a circle of radius 20 pixels around the center location of each nodule. From Table 6, for both drillers and scanners, the found nodules (hits) had a statistically higher average saliency than the missed nodules $[t(873) = 5.06, \ p < 0.01$ for drillers and $t(105) = 2.69, \ p < 0.01$ for scanners]. These results are consistent with our conclusions in the faux lung CT search study that the saliency of nodules did influence radiologists' detection performance, and that utilizing different search strategies may result in different diagnostic errors.

4 Conclusions and Future Work

With the increasing usage of volumetric medical imaging in clinical practice, radiologists have to develop suitable search strategies for reading volumetric images. We investigated two search strategies, drilling and scanning, by comparing observers' gaze positions with computational saliency maps. The results of the faux lung CT study with trained naïve observers and the chest CT study with experienced radiologists both led to the following conclusions. (1) Drillers rely on 3-D dynamic motion saliency more than scanners do. (2) Scanners use more 2-D saliency than drillers do. (3) Drillers rely on 3-D dynamic saliency more than 2-D saliency, while scanners rely on 2-D saliency more than 3-D dynamic saliency. (4) The saliency of items being searched for does have an impact on an observers' detection performance. One of the key limitations of the study was that there were not enough "scanner" radiologists in experiment 2, so our experiment 2 conclusions for radiologist searchers must be more limited than our findings from experiment 1 with naïve subjects. Nevertheless, the two experiments show the same basic pattern of results.

As has been previously reported, drillers found more nodules and covered more lung tissue during their search in the same
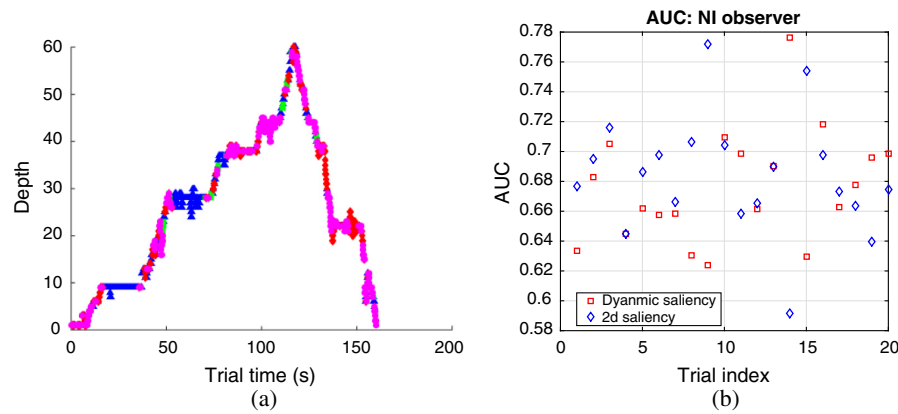
Fig. 7 (a) Example depth-time plots of NI observers' eye positions. The four colors indicate four quadrants of the slice. (b) Example AUCs achieved by the dynamic saliency maps (red squares) and the 2-D saliency maps (blue diamonds) in experimental trials, for predicting an NI's fixations. NI observers tend to use a mix of drilling and scanning search strategy, and their search behavior varied on a trial-by-trial basis.

amount of time as scanners in this dataset.[19] One of the reasons that the drilling strategy was used by the majority of radiologists is that radiologists may have realized scanning is not the best strategy for finding nodules in volumetric image data, and have adjusted their search strategies accordingly. This aligns with our hypothesis that the drilling strategy, which incorporates more 3-D dynamic information, may be superior to the scanning strategy for this type of search task. However, it is possible that other clinical tasks (e.g., searching for signs of pneumonia) may still rely more on 2-D image information, and may therefore elicit more scanning behavior.

In the future, other advanced saliency models might improve the predictions of radiologists' fixations, and thus provide stronger evidence for which type of image information is being used by radiologists in their search strategies. Beyond simple bottom-up salience, a more complete account of search strategies would include the top-down knowledge used by expert radiologists to guide their attention. Moreover, instead of simply assessing where drillers and scanners fixate, it would be interesting to investigate the sequence of radiologists' fixations. It seems unlikely that the sequence would simply find radiologist fixating locations in the descending order of salience. Future work could also explore the utility of neurobiological modeling of visual attention to describe observers' visual search strategies.[16]

It is important to note that our faux lung CT search study with instructed naïve observers probably fails to capture some important aspects of a clinical chest CT search task performed by experienced radiologists. Certain critical features in lung CTs/medical images may not have been appropriately captured by the volumetric noise. Hence, addition assessments with radiologists in real clinical settings with various clinical tasks would be valuable to achieve a more definitive assessment of our hypothesis and observations, especially for those conclusions that suffer from an underpowered sample size of radiologists and experimental cases.

Appendix: No Instruction Observers

In experiment 1, observers in the NI condition did not receive either drilling or scanning instruction, and they were told to choose any search strategy that could be used to find the T's

quickly and accurately. The depth-by-time plots of their 3-D scan paths [e.g., Fig. 7(a)] show that they tend to use a mix of drilling and scanning strategies, and their search strategies varied on a trial-by-trial basis [e.g., Fig. 7(b)]. For example, for one particular NI, the AUCs achieved by the dynamic saliency were higher in 14 (out of 20) trials than the counterpart AUCs achieved by the 2-D saliency; while for another NI, the AUCs achieved by the dynamic saliency were higher in only four trails. As no instruction on efficient search strategies was provided to the NIs, these naïve observers' search behavior was not consistent in terms of what image information they relied on for deploying their visual attention. As a result, both intraobserver and interobserver variability in NIs search strategies were extremely high and difficult to interpret.

References

1. D. Manning, S. Ethell, and T. Donovan, "Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph," *Br. J. Radiol.* **77**(915), 231–235 (2014).
2. H. L. Kundel, C. F. Nodine, and L. Toto, "Searching for lung nodules: the guidance of visual scanning," *Invest. Radiol.* **26**(9), 777–781 (1991).
3. H. L. Kundel et al., "Holistic component of image perception in mammogram interpretation: gaze-tracking study 1," *Radiology* **242**(2), 396–402 (2007).
4. C. Mello-Thoms, "How much agreement is there in the visual search strategy of experts reading mammograms?" *Proc. SPIE* **6917**, 691704 (2008).
5. H. Matsumoto et al., "Where do neurologists look when viewing brain CT images? An eye-tracking study involving stroke cases," *PLoS One* **6**(12), e28928 (2011).
6. M. Carrasco, "Visual attention: the past 25 years," *Vision Res.* **51**(13), 1484–1525 (2011).
7. L. Itti and C. Koch, "Computational modelling of visual attention," *Nat. Rev. Neurosci.* **3**(3), 194–203 (2001).
8. J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems*, B. Schökopf, J.C. Platt, and T. Hofmann, Eds., pp. 545–552, MIT Press, Cambridge, Massachusetts (2006).
9. X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'07)*, pp. 1–8, IEEE (2007).

10. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11), 1254–1259 (1998).
11. L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Trans. Image Process.* **13**(10), 1304–1318 (2004).
12. G. Zhang et al., "Visual saliency based object tracking," in *Computer Vision–ACCV 2009*, pp. 193–203, Springer (2010).
13. M. Cheng et al., "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(3), 569–582 (2015).
14. A. Belardinelli, F. Pirri, and A. Carbone, "Motion saliency maps from spatiotemporal filtering," in *Attention in Cognitive Systems*, L. Paletta and J. K. Tsotsos, Eds., pp. 112–123, Springer, Berlin Heidelberg (2009).
15. J. K. Tsotsos et al., "Modeling visual attention via selective tuning," *Artif. Intell.* **78**(1), 507–545 (1995).
16. A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(1), 185–207 (2013).
17. M. Alzubaidi et al., "What catches a radiologist's eye? A comprehensive comparison of feature types for saliency prediction," *Proc. SPIE* **7624**, 76240W (2010).
18. V. Jampani et al., "Assessment of computational visual attention models on medical images," in *Proc. of the Eighth Indian Conf. on Computer Vision, Graphics and Image Processing*, p. 80, ACM (2012).
19. T. Drew et al., "Scanners and drillers: characterizing expert visual search through volumetric images," *J. Vision* **13**(10), 3 (2013).
20. L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Res.* **40**(10), 1489–1506 (2000).
21. D. H. Brainard, "The psychophysics toolbox," *Spatial Vision* **10**, 433–436 (1997).
22. D. G. Pelli, "The videotoolbox software for visual psychophysics: transforming numbers into movies," *Spatial Vision* **10**(4), 437–442 (1997).
23. S. G. Armato, III 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).
24. S. E. Seltzer et al., "Spiral CT of the chest: comparison of cine and film-based viewing," *Radiology* **197**(1), 73–78 (1995).
25. S. K. Rushton, M. F. Bradshaw, and P. A. Warren, "The pop out of scene-relative object movement against retinal motion due to self-movement," *Cognition* **105**(1), 237–245 (2007).
26. G. Johansson, "Visual motion perception," *Sci. Am.* **232**, 76–88 (1975).
27. G. Johansson, "Spatio-temporal differentiation and integration in visual motion perception," *Psychol. Res.* **38**(4), 379–393 (1976).
28. I. E. Holliday and T. S. Meese, "Optic flow in human vision: MEG reveals a foveo-fugal bias in V1, specialization for spiral space in hMSTs, and global motion sensitivity in the IPS," *J. Vision* **8**(10), 17 (2008).
29. F. P. Redlick, M. Jenkin, and L. R. Harris, "Humans can use optic flow to estimate distance of travel," *Vision Res.* **41**(2), 213–219 (2001).
30. W. H. Warren et al., "Optic flow is used to control human walking," *Nat. Neurosci.* **4**(2), 213–216 (2001).
31. P. A. Warren and S. K. Rushton, "Evidence for flow-parsing in radial flow displays," *Vision Res.* **48**(5), 655–663 (2008).
32. S. Yantis and J. Jonides, "Abrupt visual onsets and selective attention: voluntary versus automatic allocation," *J. Exp. Psychol.* **16**(1), 121 (1990).
33. S. E. Christ and R. A. Abrams, "Abrupt onsets cannot be ignored," *Psychonomic Bull. Rev.* **13**(5), 875–880 (2006).
34. C. C. Loy, T. Xiang, and S. Gong, "Salient motion detection in crowded scenes," in *2012 5th Int. Symp. on Communications Control and Signal Processing (ISCCSP)*, pp. 1–4, IEEE (2012).
35. B. K. Horn and B. G. Schunck, "Determining optical flow," *Proc. SPIE* **0281**, 319–331 (1981).
36. J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.* (6), 679–698 (1986).
37. O. Le Meur and T. Baccino, "Methods for comparing scanpaths and saliency maps: strengths and weaknesses," *Behav. Res. Methods* **45**(1), 251–266 (2013).
38. D. Howell, *Statistical Methods for Psychology*, Cengage Learning, Boston, Massachusetts (2012).
39. W. Chen and F. Samuelson, "The average receiver operating characteristic curve in multireader multicase imaging studies," *Br. J. Radiol.* **87**(1040) (2014).
40. T. Fawcett, "An introduction to roc analysis," *Pattern Recognit. Lett.* **27**(8), 861–874 (2006).
41. H. L. Kundel, C. F. Nodine, and D. Carmody, "Visual scanning, pattern recognition and decision-making in pulmonary nodule detection," *Invest. Radiol.* **13**(3), 175–181 (1978).

Gezheng Wen is a PhD candidate in electrical and computer engineering at the University of Texas at Austin. He received his BS degree from the University of Hong Kong in 2011 and received his MS degree from the University of Texas at Austin in 2014. He is a research trainee in diagnostic radiology at the University of Texas MD Anderson Cancer Center. His research interests include medical image quality assessment, model observer, and visual search of medical images.

Avigael Aizenman is a research assistant at the Brigham and Women's Hospital Visual Attention Lab. She received her BA degree from Brandeis University in 2013. Her research interests include visual attention and visual search, with a specific focus on search in medical images.

Trafton Drew is an assistant professor in the Psychology Department at the University of Utah. He received his BS degree from the University of North Carolina at Chapel Hill. He received his MS and PhD degrees in cognitive psychology from University of Oregon. His lab's (http://aval.psych.utah.edu/) current interests include medical image perception, eye-tracking, and the neural substrates of working memory. He is a member of the Medical Image Perception Society, Vision Sciences Society, and the Psychonomic Society.

Jeremy M. Wolfe is a professor of ophthalmology and radiology at Harvard Medical School and director of the Visual Attention Lab at the Brigham and Women's Hospital. He received his AB degree in psychology from Princeton University and a PhD in psychology from the Massachusetts Institute of Technology. His research focuses on visual search with particular interest in socially important search tasks such as cancer and baggage screening. He was Psychonomic Society Chair and fellow of the Society of Experimental Psychologists, APS, and APA Divisions 1, 3, 6, 21.

Tamara Miner Haygood is an associate professor in diagnostic radiology at the University of Texas MD Anderson Cancer Center. She received her PhD in history from Rice University in 1983 and the MD from the University of Texas Health Science Center at Houston in 1988. She is the author of more than 65 peer-reviewed journal papers and is a member of SPIE. Her research interests include visual recognition memory for imaging studies and efficiency of interpretation.

Mia K. Markey is a professor of biomedical engineering and engineering foundation endowed faculty fellow in engineering at the University of Texas at Austin as well as adjunct professor of imaging physics at the University of Texas MD Anderson Cancer Center. She is a fellow of the American Association for the Advancement of Science (AAAS) and a senior member of both the IEEE and SPIE.