What do radiologists look for? Advances and limitations of perceptual learning in radiologic search

Robert G. Alexander*

Department of Ophthalmology, SUNY Downstate Health
Sciences University, Brooklyn, NY, USA ✉

Stephen Waite*

Department of Radiology, SUNY Downstate Health
Sciences University, Brooklyn, NY, USA ✉

Stephen L. Macknik

Department of Ophthalmology, SUNY Downstate Health
Sciences University, Brooklyn, NY, USA ✉

Susana Martinez-Conde

Department of Ophthalmology, SUNY Downstate Health
Sciences University, Brooklyn, NY, USA ✉

Supported by guidance from training during residency programs, radiologists learn clinically relevant visual features by viewing thousands of medical images. Yet the precise visual features that expert radiologists use in their clinical practice remain unknown. Identifying such features would allow the development of perceptual learning training methods targeted to the optimization of radiology training and the reduction of medical error. Here we review attempts to bridge current gaps in understanding with a focus on computational saliency models that characterize and predict gaze behavior in radiologists. There have been great strides toward the accurate prediction of relevant medical information within images, thereby facilitating the development of novel computer-aided detection and diagnostic tools. In some cases, computational models have achieved equivalent sensitivity to that of radiologists, suggesting that we may be close to identifying the underlying visual representations that radiologists use. However, because the relevant bottom-up features vary across task context and imaging modalities, it will also be necessary to identify relevant top-down factors before perceptual expertise in radiology can be fully understood. Progress along these dimensions will improve the tools available for educating new generations of radiologists, and aid in the detection of medically relevant information, ultimately improving patient health.

Introduction

Current models of medical image perception are incomplete and demonstrate significant gaps in the current understanding of radiologic expertise (see Waite et al., 2019 for a review). That is to say, we do not precisely know what an expert radiologist does: currently, radiologists achieve peak expertise only after years of trial-and-error training, during which they acquire their skillset through veiled principles that are yet to be articulated. Mentors provide feedback about mistakes, guidance for what is benign versus malignant, and other conceptual, factual, and procedural information. However, pattern recognition is difficult to teach (Kellman & Garrigan, 2009), and expertise in viewing radiologic images is therefore gained largely as a function of the number of images read, rather than through explicit instruction and understanding (Krupinski, Graham, & Weinstein, 2013; Nodine & Mello-Thoms, 2010). The result is a knowledge base that has not translated into concrete methods of training derived from critical perceptual features. Determining the precise features that radiologists use to discriminate abnormalities in medical images—and designing innovative heuristics for trainees that enable efficient learning of informative features—would help optimize performance in the field.

Error rates in radiology

The long-term goal of most studies of radiologist performance is to reduce error. Although it is difficult to determine the precise error rates in current clinical practice, it is clear that reductions in error rate would improve patient care. The diagnostic error rate in a typical clinical practice (comprising both normal

and abnormal image studies) is between 3% and 4% (Borgstede, Lewis, Bhargavan, & Sunshine, 2004; Siegle et al., 1998), which translates into approximately 40 million interpretive errors per year worldwide (Bruno, Walker, & Abujudeh, 2015). This error rate is substantially higher—approximately 30%—when all images contain abnormalities (Berlin, 2007; Rauschecker et al., 2020; see Waite et al., 2017 for a review). Detection/omission errors account for 60% to 80% of interpretive errors (Funaki, Szymski, & Rosenblum, 1997; Rosenkrantz & Bansal, 2016). Thus faulty perception is the most important source of interpretive error in diagnostic imaging (Berlin, 2014; Donald & Barnard, 2012; Krupinski, 2010).

Radiologic image viewing as a visual search task

Radiologic image viewing is essentially a specialized visual search task: the first step in medical imaging is the detection of medically relevant information in an image (e.g., nodules in a chest x-ray [CXR]), by searching for abnormalities amid normal anatomy and physiology (e.g., normal lung tissue). Visual features play a critical role in such tasks: search is generally faster and targets easier to detect and recognize when target features are dissimilar to those of the background and nontarget distractors (Alexander, Nahvi, & Zelinsky, 2019; Alexander & Zelinsky, 2011; Alexander & Zelinsky, 2012; Duncan & Humphreys, 1989; Ralph, Seli, Cheng, Solman, & Smilek, 2014; Treisman, 1991).

Outside the field of radiology, researchers often decompose visual search tasks into two stages: the initial detection of various simple target features across the visual field, and the subsequent deployment of attention and gaze to specific objects, in which more complex visual information is available on foveation (Alexander & Zelinsky, 2012; Alexander & Zelinsky, 2018; Castelhano, Pollatsek, & Cave, 2008; Wolfe, 1994a). Radiologic search tasks are similarly thought to involve an initial feature-processing step across the visual field, followed by foveation of specific objects (Krupinski, 2011; Kundel, Nodine, Conant, & Weinstein, 2007; Nodine & Kundel, 1987; Sheridan & Reingold, 2017).

Although there is a consensus on a few features (color, motion, orientation, and size) used by the human visual system to guide attention and eye movements during search tasks, the exploration of other potential features has been limited due to dissenting opinions, insufficient data, and alternative explanations for observed data patterns (Wolfe & Horowitz, 2004; Wolfe & Horowitz, 2017). Further, even if the visual system *can* use a feature, it does not mean that the feature *will* be used (Alexander et al., 2019). As a result, the features that guide search in any given context—including medical image viewing—are not yet fully known.

Types of errors in radiologic search

Some authors have proposed that errors in medical image perception can be best understood as stemming from different aspects of the task—leading to three general types of errors: search, recognition, and decision errors (Kundel, Nodine, & Carmody, 1978). "Scanning errors" (also called "search errors") result from failures in the first stage of search. Specifically, peripheral information fails to guide the observer's gaze to a relevant location, and high-resolution foveal vision does not assist interpretation because the observer never looks at the location directly (Doshi et al., 2019; Holland, Sun, Gackle, Goldring, & Osmar, 2019; Ukweh, Ugbem, Okeke, & Ekpo, 2019). In "recognition errors," abnormalities are foveated too briefly for the observer to correctly recognize them (Holland et al., 2019). Depending on the imaging modality, the foveation time that suffices to prevent recognition errors varies from 500 to 1000 ms (Hamnett & Jack, 2019; Holland et al., 2019). "Decision-making errors" occur when the observer either fails to recognize relevant features or actively dismisses them, despite foveating an abnormality for a relatively long period of time (Baskaran et al., 2019; Holland et al., 2019). Roughly one-third of omission errors falls under each of the earlier mentioned three categories (Kundel et al., 1978).

Traditionally, decision-making errors have been considered "cognitive" errors (in which the abnormality is visually detected but the meaning or importance of the finding is not correctly understood or appreciated), and both scanning and recognition errors have been considered "perceptual" errors (in which an abnormality is not observed) (Kundel, 1989). However, all three kinds of errors may result from perceptual failures, and all three kinds of errors might reflect expectation or cognitive biases, making it sometimes impossible to determine whether a given search, recognition, or decision-making error is perceptual or cognitive in nature. Indeed, both foveal and peripheral search performance might rely on the same perceptual features, rather than on different features for peripheral versus foveal search (Maxfield & Zelinsky, 2012; Nakayama & Martini, 2011; Zelinsky, Peng, Berg, & Samaras, 2013; but see Alexander & Zelinsky, 2018). Thus an observer looking specifically for dark, oriented bands might fail to foveate a bright, round lesion for the same reasons they might fail to recognize it if foveated.

Visual search in 3D volumetric imaging

Searching through CXRs and other two-dimensional (2D) images is superficially similar to search tasks conducted in traditional laboratory settings. However, search through three-dimensional (3D) volumetric

images involves a qualitatively different process than that through 2D images. When reading a computed tomography (CT) or a magnetic resonance imaging scan, a radiologist must scroll through a stack of images—thin slices of the 3D volume of an organ (Nakashima, Komori, Maeda, Yoshikawa, & Yokosawa, 2016). When searching for lung nodules in an image stack from a chest CT, a common strategy is to restrict one's eye movements to a small region of the image, while quickly scrolling (i.e., "drilling") through the stack (Drew et al., 2013). An alternative strategy is to change depth more slowly, and make eye movements across a larger area of the image (Drew et al., 2013) (Figure 1). Similar patterns have been found in searches through digital breast pathology images, in which observers "zoom" in and out of a single image, rather than scroll through image stacks (Mercan, Shapiro, Brunyé, Weaver, & Elmore, 2018). These strategies may depend on the body part imaged, with "scanning" being more likely in studies of larger body parts: when intending to search a small anatomic region, it makes little sense to make large eye movements. Thus radiologists typically adopt a "driller" strategy when viewing CTs of the abdomen and pelvis (Kelahan et al., 2019). Although some searches through these anatomic regions involve slower changes in depth—and those radiologists might be characterized as "scanners" (as opposed to "drillers")—their search patterns still qualitatively resemble the "driller" pattern shown in Figure 1. Regardless, and despite any differences in strategy, the vast amount of 3D data that radiologists must scrutinize effectively prevents the careful foveation of each image region within a CT stack (Eckstein, Lago, & Abbey, 2017; Miller et al., 2015). Therefore because some image regions are only seen peripherally on some slices, peripheral vision is especially important in searches of CT images.

As peripheral vision cannot provide the kind of fine spatial discrimination that characterizes foveal vision, detectability of certain lesions can differ between searches through 3D or 2D images. For instance, Eckstein et al. (2017) found higher detectability for calcifications in 2D single slice images, and relatively improved detection of masses in 3D volumetric imaging, lending support to the notion that observer performance in 2D search tasks might not generalize to that in volumetric searches. To address this possibility, Wen et al. (2016) used a novel dynamic 3D saliency approach to model naive observers' gaze distributions in images designed to mimic lung CTs. They found that the dynamic 3D saliency model predicted gaze distributions better than the traditional saliency approach, especially for observers who scrolled quickly in depth. Wen et al. (2016) proposed that the success of the dynamic saliency model might result from the human visual system's sensitivity to optic flow (the apparent motion of objects caused by the relative
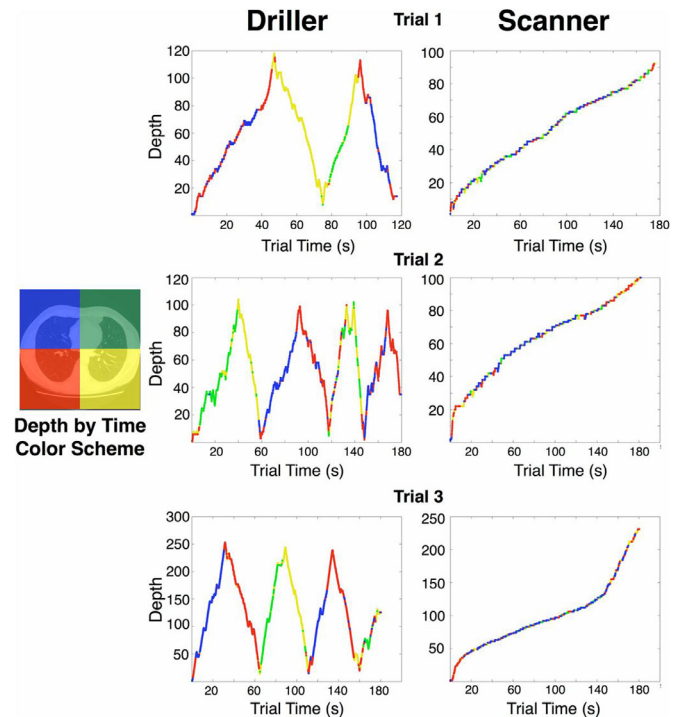

Figure 1. Description of 3D scan paths from Drew et al. (2013), who recorded eye position in each quadrant (left panel) as observers scrolled through CT scans in depth. Color indicates the quadrant of the image the radiologist was looking at during a given time in the trial. "Depth" on the y-axis refers to the 2D orthogonal slice of the scan currently viewed. In this study, radiologists looking for nodules on chest CTs could be characterized into two groups based on their search strategies. "Drillers," such as the radiologist whose data appear in the middle column, tend to look within a single region of an image while quickly scrolling back and forth in depth through stacks of images. "Scanners," such as the radiologist whose data appear in the right column, scroll more slowly in depth, and typically do not return to depths that they have already viewed. Scanners make more frequent eye movements to different spatial locations on the image, exploring the current 2D slice in greater detail. Note that although scanners spend more time than drillers making saccades per slice, neither scanners nor drillers visit all four quadrants of the image on every slice. Thus some regions of some slices may never be viewed foveally by either group. (Reprinted from Drew et al., 2013).

motion between observer and scene). However, evidence from other studies indicates that optic flow may not preattentively guide attention or eye movements (Wolfe & Horowitz, 2004; Wolfe & Horowitz, 2017).

Gaining expertise in medical image analysis

Expert radiologists efficiently direct their gaze to clinically relevant information using learned features

in their peripheral vision (Kundel, 2015). That is, expert radiologists are better at "search," finding abnormalities faster than novices because they need fewer eye movements to foveate an abnormality that they first detect peripherally (Drew et al., 2013; Manning, Ethell, Donovan, & Crawford, 2006). The initial stage of peripheral processing typically produces the greatest differences between experts and novices in radiologic search (Drew et al., 2013; Manning et al., 2006). Researchers have therefore begun to ask whether performance differences between expert and novice radiologists may be the result of underlying differences in search strategies (Brams et al., 2020; Wood, 1999; Wood et al., 2013).

Search tasks that are initially performed slowly and inefficiently can become faster and efficient with practice, as the searchers learn task-relevant features (Frank, Reavis, Tse, & Greenlee, 2014; Sireteanu & Rettenbach, 2000; Steinman, 1987). Thus performance in one study's initially slow search task—searching for red-green bisected disks among green-red bisected disks—improved dramatically across eight training sessions spread over eight different days, and this improvement was still present when the same subjects were retested 9 months later (Frank et al., 2014). Radiologists similarly become faster at searching for abnormalities as they gain expertise (Krupinski, 1996; Kundel et al., 2007; Nodine, Kundel, Lauver, & Toto, 1996.

Perceptual learning in radiology

Like any other perceptual skill, the ability to detect radiologic abnormalities can improve through perceptual learning, that is, experience-induced improvements in the way perceptual information is extracted from stimuli. Perceptual learning techniques have been developed to accelerate the acquisition of perceptual expertise in domains as varied as flight training and mathematics, as well as in histopathology and surgery (Guerlain et al., 2004; Kellman, 2013; Kellman & Kaiser, 2016; Kellman, Massey, & Son, 2010; Krasne, Hillman, Kellman, & Drake, 2013). One example from radiology is that novice mammography film readers' sensitivity toward low-contrast information in x-rays improves with increasing practice viewing x-ray images (Sowden, Davies, & Roling, 2000; Sowden, Rose, & Davies, 2002).

Several studies have demonstrated transfer of learning from trained images to new images. For instance, performance improvements resulting from training through exposure to CXR images can transfer from positive contrast images to negative contrast images, and vice versa (Sowden et al., 2000). Recent research has shown that training to ascertain if abdominal CTs are consistent with appendicitis can transfer both to previously unseen abdominal CT images and to different image orientations (Johnston et al., 2020). Another recent study (Li, Toh, Remington, & Jiang, 2020) presented novice nonradiologist participants with pairs of CXRs, one of which always contained a tumor. Across four sessions, observers practiced discriminating which CXR contained the tumor and locating the tumor within the image. Perceptual learning resulted in discrimination performance improvements both for old images and for novel images.

Transfer of learning to novel images has also been demonstrated when training to identify bone fractures on pelvic radiographs (Chen, HolcDorf, McCusker, Gaillard, & Howe, 2017). Studies on the effects of image variability have produced mixed results: whereas one study found that greater variability of training images led to greater transfer effects (Chen et al., 2017), another study found comparable performance after training with either a larger number of images or more repetitions of a smaller number of images (Li et al., 2020).

The earlier mentioned evidence of limited transfer notwithstanding improvements in detection or discrimination of visual stimuli are usually constrained to the particular tasks (Ball & Sekular, 1987; Saffell & Matthews, 2003) or to the features (Ball & Sekular, 1987; Fahle, 2005) used during training (see Karni & Sagi, 1991; Sagi, 2011; Watanabe & Sasaki, 2015 for reviews). Similarly, the perceptual skills that radiologists develop over the course of their training are restricted to specific radiologic image perception tasks. Indeed, radiologists are no better at performing nonradiologic search tasks than nonradiologists are (Nodine & Krupinski, 1998). Thus radiologic expertise does not arise from general perceptual improvements, but instead results from the learning of features and task demands specific to radiologic search tasks. This is consistent with findings from laboratory studies suggesting that perceptual learning primarily improves radiologic search as a result of the learning of task-relevant visual features (as opposed to the learning of other task demands [Frank et al., 2014]).

Perceptual learning regimens have been shown to improve performance even in medical students who had already seen thousands of images prior to perceptual training (Krasne et al., 2013; Sowden et al., 2000). Further, even small amounts of practice in a relatively short interval can produce significant improvements in radiologic performance. For example, Krasne et al. (2013) used web-based perceptual and adaptive learning modules to enhance histopathology pattern recognition and image interpretation in a test group of medical students. The many short classification trials in the learning modules were combined with a continuous assessment of accuracy and reaction time, which was used both to track progress and to adapt trials to focus

perceptual learning on the categories of patterns that needed the most practice. The training led to improved accuracy and reaction times from pretest to posttest, with a delayed posttest (6–7 weeks later) showing that much of this learning was retained. In a different perceptual learning study with other radiologic tasks, improvement was appreciated after just a few hours of training (Johnston et al., 2020).

Detailed feedback may be particularly important for the perceptual learning of radiologic features: in people with no prior experience viewing mammographic images, sensitivity to lesions with complex visual structures only improved when feedback about both response correctness and correctness of the identified *location* of the lesion was provided (Frank et al., 2020). In such conditions, performance improvements were significantly retained 6 months after training. Using a different radiologic task (Johnston et al., 2020) similarly found stronger perceptual learning when participants received feedback about both accuracy and location, compared with accuracy alone.

Training to recognize the often-subtle diagnostic features in radiology may thus benefit more from specific feedback during instruction than from learning the simple features that are often used in perceptual learning tasks (e.g., orientation). In particular, subtle structural feature layouts may be both difficult to perceive and vary substantially across different radiologic images. Although observers can be successfully trained in perceptual learning of difficult- or impossible-to-perceive patterns (Seitz, Kim, & Watanabe, 2009; Watanabe, Náñez, & Sasaki, 2001), explicit feedback can help radiologists learn how to better resolve subtle critical features (Seitz, 2020). This does not necessarily mean that feedback must be provided after each image: blocked feedback (i.e., explicit messages indicating the percent of images that were correctly diagnosed) has been shown to boost perceptual learning (Seitz et al., 2010).

The main caveat for this general approach is that developing the perceptual skills needed for successful detection of abnormalities requires practice within the correct task context and with the correct set of features. Thus the key is: how do we determine the features that expert radiologists use when searching through medical images?

Saliency models and guiding features in radiologic images

In visual search, eye movements are often directed to the most "salient" or "informative" regions in an image (McCamy, Otero-Millan, Di Stasi, Macknik, & Martinez-Conde, 2014; Otero-Millan, Troncoso, Macknik, Serrano-Pedraza, & Martinez-Conde, 2008). Salient regions are thus a reasonable place for radiologists to explore first when searching medical images for clinically relevant abnormalities. Several studies have therefore attempted to use computational models of saliency to specify the features radiologists use when searching for abnormalities. Tests of saliency models may provide insights into the importance of different features in radiologic image viewing. For instance, if a saliency model fails to accurately predict human performance, it may be that the model neglects relevant visual features. Conversely, if a model identifies salient lesions that radiologists miss, the model may rely on overlooked image features that radiologists might incorporate in future searches.

CXRs, the radiologic test most ordered in hospitals (Pirnejad, Niazkhani, & Bal, 2013), produce 2D images with overlapping structures that have different luminance or optical densities. The larger the difference in thickness or density in the anatomy between two structures (e.g., air and dense tissue), the larger the resulting difference in radiographic density or contrast. Salient regions in these images are therefore typically those that differ in density from the regions around them: for example, nodules differ in density from their surround, and can be salient on CXRs (Jampani, Ujjwal, Sivaswamy, & Vaidya, 2012). Alzubaidi, Balasubramanian, Patel, Panchanathan, and Black (2010) found that image regions that typically capture radiologists' gazes in CXRs are characterized by oriented edges and textures. Jampani, Ujjwal, et al. (2012) moreover found that, in the case of CXR's demonstrating pneumoconiosis (a lung disease caused by breathing in certain kinds of dust), saliency maps generated using the graph-based visual saliency (GBVS) computational model performed relatively well compared with radiologists' fixations. In a more recent and comprehensive study, Wen et al. (2017) evaluated 16 representative saliency models and ranked them by how well the saliency maps agreed with radiologists' eye positions during interpretation of CXR, CT scans, and positron emission tomography (PET) scans. Relative to CXR, CT scans produce higher resolution views of nodules and other abnormalities, as well as better visualization of soft tissue. Scrolling through 3D CT scans can produce feature dimensions not present in 2D images, such as apparent motion and optic flow. In PET, the customary use of a dye containing radioactive tracers yields stronger signals in the regions to which the tracer flows, and therefore higher saliency. Wen et al. (2017) found that the rank orders of the models over medical images were different from the benchmark rank-order over natural images, and that the models' performance differed across medical imaging modalities (Figure 2). This pattern is consistent with reports that radiologic search uses different features than other search tasks, and that the skills involved may differ across imaging
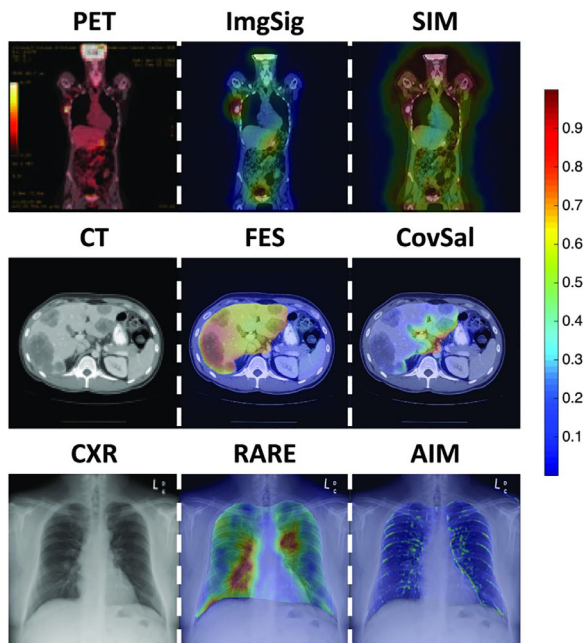
Figure 2. Examples of saliency models applied to PET, CT, and CXR. Saliency maps are represented as heat maps, with color indicating the saliency at that location: red is more salient than blue. The left column displays representative images. The middle column shows examples of saliency maps that *accurately* highlighted the regions of interest in the images. The right column shows examples of saliency maps that highlighted task-*irrelevant* regions in the images. Models with accurate predictions may provide insight into the features that radiologists use to view images, and models with inaccurate predictions may help narrow the list of potential features that need to be assessed. Image signature (ImgSig; Hou, Harel, & Koch, 2012), fast and efficient saliency (FES; Tavakoli, Rahtu, & Heikkilä, 2011), and RARE (Riche et al., 2013) were top-ranked models for PET, CT, and CXR. (Reprinted from Wen et al., 2017). SIM, Saliency by induction mechanisms; CovSal, Covariance saliency; AIM, Attention based on information maximization.

modalities (Gunderman, Williamson, Fraley, & Steele, 2001; Nodine & Krupinski, 1998). Further, certain saliency models matched the performance of individual radiologists better than that of others (Wen et al., 2017). For example, the "Saliency in Context" or "Salicon" model—which operates very differently from most saliency models—uses high-level semantics in deep neural networks to recognize objects, rather than relying solely on low-level feature differences (Huang, Shen, Boix, & Zhao, 2015). As one might expect, the Salicon model did not correlate strongly with any of the other saliency models tested, but it outperformed most other models in predicting the eye movements of two participants (a radiologist faculty member and a fellow) when viewing CT scans. However, Salicon ranked 15th or 16th among 16 models for predicting the eye movements of each of the four other radiologists,

suggesting that individual differences in gaze behavior and performance may result at least partly from different radiologists using different image features, or employing different task strategies, during visual search (Wen et al., 2017; see also Wen et al., 2016).

Visual features that have predicted the gaze patterns of expert radiologists in 2D medical image models include orientation, pixel intensity, and size (Alzubaidi et al., 2010; Jampani, Sivaswamy, & Vaidya, 2012). Intensity and orientation also predict the gaze patterns of expert radiologists in 3D volumetric images, as is optic flow, although the features relevant to 3D images may partly depend on the search strategies that radiologists adopt (e.g., "scanning" slices before scrolling in depth versus "drilling" more quickly in depth) (Wen et al., 2016; Wen et al., 2017).

We note that the earlier mentioned saliency models do not consider different task strategies. Instead, these models are "bottom-up," or solely driven by the low-level visual features of an image. The bottom-up modeling approach assumes that certain parts of an image are salient enough to be attended and looked at, regardless of the ongoing goals of the viewer (e.g., (Theeuwes, 2004). Bottom-up features are particularly important for the detection of unexpected or incidental findings. Yet, attention and eye movements are also affected by "top-down" information: recurrent feedback processing can bias the direction of eye movements and attention as a result of the viewer's goals and expectations (Alexander et al., 2019; Chen & Zelinsky, 2006; Folk, Remington, & Johnston, 1992). Top-down mechanisms are known to play powerful roles in search, by restricting exploration to image regions that are likely to contain targets, by preventing salient but task-irrelevant features from capturing as much attention as task-relevant features (Alexander & Zelinsky, 2012; Chen & Zelinsky, 2006; Folk et al., 1992; Wolfe, Butcher, Lee, & Hyle, 2003), and by "filling-in" missing information about search targets (Alexander & Zelinsky, 2018). In the case of radiologists, previous knowledge and expertise (including their a priori expectations about a task) can change the features they may use to accomplish their task goals. Further, the interpretation of medical imaging is highly task-dependent: radiologic expertise in one domain does not necessarily rely on the same skills as expertise in another domain (Beam, Conant, & Sickles, 2006; Elmore et al., 2016; Gunderman et al., 2001). Task demands can similarly change within a specific domain and therefore affect search behavior: for example, in patients with known renal cell cancer, radiologists may prolong their search of the lungs, looking carefully for nodules that might represent metastatic lesions. Thus "top-down" features likely play a large role in radiologic search, in which relevant information may not be salient from a low-level perspective. For instance, a fat-containing tumor may

resemble normal fat—and therefore fail to be detected by bottom-up models—but could be displaced spatially, appearing in a location where there should not be fatty tissue. Similarly, air in the lungs is normal, but air in the heart is abnormal. In addition, knowledge of the typical appearance of lesions and the normal appearance of the surrounding organ/tissue also aids radiologists in searching for lesions. Understanding the patient's clinical scenario can help ensure that the radiologist carefully examines the relevant regions of the image with the right features in mind. Consequently, including top-down information is a major challenge in developing accurate models of radiologic search.

Computational models may theoretically make use of such top-down information, and some have begun to do so. Jampani et al. (2012) found that although the lung regions cover only approximately 40% to 50% of the area in a typical 2D CXR, they contain approximately 84% of all fixations, which is indicative of their top-down importance. A modified saliency model that combined bottom-up and top-down saliency (and thus avoided unimportant image regions) performed better than the standard bottom-up GBVS model in predicting eye fixations (Jampani, Ujjwal, et al., 2012). To incorporate top-down factors in their saliency models, Wen et al., (2017) segmented relevant anatomic regions (i.e., the liver and aorta in CT images), so that other regions (e.g., the kidney on CT images) were not allowed to be salient.

Potential future approaches

As promising candidate features continue to be identified by models that accurately predict radiologist performance, further studies will need to confirm that these are the features radiologists do use, rather than features that are *correlated with* actual critical features. One approach to confirming that the correct features have been identified is to create faux radiologic images that are matched to real images in terms of such features. Recently, Semizer, Michel, Evans, and Wolfe (2018) took this approach, testing whether a texture model (the Portilla-Simoncelli texture algorithm; [Portilla & Simoncelli, 2000]) accurately captured features used by radiologists (for similar approaches in other domains; see Alexander, Schmidt, & Zelinsky, 2014; Rosenholtz, Huang, & Ehinger, 2012; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012). This study used the texture algorithm to generate faux images that perfectly matched real medical images in terms of the modeled texture features, but which were otherwise different. When no visible lesion was present in the image, radiologists' performance was comparable for real images and faux images, suggesting that both types of images were equivalent in terms of the critical features that radiologists use to render judgment.

However, when visible lesions were present, radiologists performed better with real than with faux images, suggesting that they used additional information other than texture to conduct the task (Semizer et al., 2018). One possibility is that the spatial relationships between image elements are also important: the Portilla-Simoncelli texture algorithm discards spatial relationships between local features, thereby removing the global or configural shape, which previous research has found to be important in the targeting of visual attention during search (Alexander et al., 2014).

We believe that the texture approach can prove fruitful in conjunction with the correct features or model. Although a simple grayscale object (e.g., a white bar on a black background) can be described by first-order cardinal image statistics—including contrast, spatial frequency, position, entropy, and orientation—none of these dimensions individually indicates that a radiologic image region is normal or abnormal. These dimensions may guide the targeting of eye movements (Wolfe, 1994a), but are not necessarily task relevant, and therefore do not provide a complete picture of the relationship between informativeness and ocular targeting (McCamy et al., 2014). The identification of some combinations of features has proven efficient in other contexts (Rappaport, Humphreys, & Riddoch, 2013; Wildegger, Riddoch, & Humphreys, 2015; Wolfe, 1994b; Wolfe et al., 1990), and may apply to efficient radiologic search. Krupinski, Berger, Dallas, and Roehrig (2003) found that individual features (signal-to-noise ratio, size, conspicuity, location, and calcification status) did not predict the gaze patterns of radiologists, but their combination did. Although Semizer et al. (2018) suggested that more than texture is involved in radiologic search, their model might have relied on the wrong features or combination of features. Because texture space is so large, with many potential dimensions and feature values along those dimensions, it remains unknown what values are important. However, just as normal images may be represented as statistical combinations of features (i.e., visual texture), abnormal images are theoretically detectable as deviations from such statistical combinations or textures.

Having said that, understanding of eye movements will be important for determining these features. One of the main results from the study of radiologic search is that experts can find tumors faster than novices—meaning that experts can find relevant features in their peripheral vision and direct their central vision to those features (Drew et al., 2013; Manning et al., 2006). A key difference between expert and novice radiologic search is therefore the ability to detect critical peripheral features. Where a general approach in radiology has been to determine the features that distinguish normal from abnormal image regions—a process that has aided the growth and development of computer aided

detection and diagnostic tools—we propose that future work should be directed to identifying the key features that experts locate in their visual periphery. Such features, critical for radiologists' peripheral advantage, have not yet been determined. Longitudinal studies of how gaze dynamics change as a function of expertise may be particularly valuable toward this goal. Once these features are known, follow-up studies may develop heuristics based on them, to help optimize radiologists' eye movements and foveation during radiologic screening. Many of the approaches outlined earlier use purely visual information, without considering the additional knowledge that radiologists rely on, such as the patient's history or the context in which a test was ordered. Although much can be achieved through visual-only approaches, current models are likely to constitute the first few steps along the way to future models that may combine bottom-up visual information with oculomotor behavior, top-down expectations, and other contextual data (e.g., patient history) to more fully characterize expert radiologists' perceptual experience. As current understanding of the features that radiologists use becomes more refined, it will become possible to design heuristics for radiologists in training, which may specifically target the perceptual learning of relevant image features. Such tools have the potential to enhance student instruction, decrease perceptual errors, and help improve patient health and well-being.

Conclusions

Perceptual errors in radiology are a significant contributor to patient harm (Waite et al., 2019; Waite et al., 2017). Educational and practical interventions to improve human perceptual and decision-making skills are therefore needed to improve diagnostic accuracy and to reduce medical error (Ekpo, Alakhras, & Brennan, 2018; Waite et al., 2020; Waite et al., 2019). However, the features used by expert radiologists during visual inspection of medical images are not yet well specified. Feature-based modeling approaches, including the saliency models discussed in this review, may help isolate such features—a practical matter critically important to the perceptual training of future radiologists, and ultimately to patient safety.

Keywords: attentional targeting, medical image perception, perceptual learning, radiology, visual search

Acknowledgments

Commercial relationships: none.
Corresponding author: Robert G. Alexander.
Email: robert.alexander@downstate.edu.
Address: Departments of Ophthalmology, Neurology, and Physiology/Pharmacology, SUNY Downstate Health Sciences University, Brooklyn, NY, USA.

*RGA and SW contributed equally to this work as co–first authors.

References

Alexander, R. G., Nahvi, R. J., & Zelinsky, G. J. (2019). Specifying the precision of guiding features for visual search. *Journal of Experimental Psychology: Human Perception & Performance, 45*(9), 1248–1264, doi:10.1037/xhp0000668.

Alexander, R. G., Schmidt, J., & Zelinsky, G. J. (2014). Are summary statistics enough? Evidence for the importance of shape in guiding visual search. *Visual Cognition, 22*(3–4), 595–609, doi:10.1080/13506285.2014.890989.

Alexander, R. G., & Zelinsky, G. J. (2011). Visual similarity effects in categorical search. *Journal of Vision, 11*(8):9, 1–15, doi:10.1167/11.8.9.

Alzubaidi, M., Balasubramanian, V., Patel, A., Panchanathan, S., & Black Jr, J. A. (2010, March). What catches a radiologist's eye? A comprehensive comparison of feature types for saliency prediction. In *Medical Imaging 2010: Computer-Aided Diagnosis. International Society for Optics and Photonics, 7624*, 76240W, doi:10.1016/j.visres.2011.12.004.

Alexander, R. G., & Zelinsky, G. J. (2018). Occluded information is restored at preview but not during visual search. *Journal of Vision, 11*(4), 1–16, doi:10.1167/18.11.4.

Alzubaidi, M., Balasubramanian, V., Patel, A., Panchanathan, S., & Black Jr, J. A. (2010, March). What catches a radiologist's eye? A comprehensive comparison of feature types for saliency prediction. In Medical Imaging 2010: Computer-Aided Diagnosis (Vol. 7624, p. 76240W). *International Society for Optics and Photonics.*

Ball, K., & Sekular, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research, 27*, 935–965.

Baskaran, L., Maliakal, G., Al'Aref, S. J., Singh, G., Xu, Z., Michalak, K., Dolan, K., Gianni, U., van Rosendael, A., van den Hoogen, I., Han,

D., Stuijfzand, W., Pandey, M., Lee, B. C., Lin, F., Pontone, G., Knaapen, P., Marques, H., Bax, J., Berman, D., Chang, H., Shaw, L. J., . . . Min, J. K. (2020). Identification and Quantification of cardiovascular structures from CCTA: an end-to-end, rapid, pixel-wise, deep-learning method. *JACC: Cardiovascular Imaging, 13*(5), 1163–1171.

Beam, C. A., Conant, E. F., & Sickles, E. A. (2006). Correlation of radiologist rank as a measure of skill in screening and diagnostic interpretation of mammograms. *Radiology, 238*(2), 446–453, doi:10.1148/radiol.2382042066.

Berlin, L. (2007). Accuracy of diagnostic procedures: has it improved over the past five decades? *American Journal of Roentgenology, 188*(5), 1173–1178.

Berlin, L. (2014). Radiologic errors, past, present and future. *Diagnosis, 1*(1), 79–84.

Borgstede, J. P., Lewis, R. S., Bhargavan, M., & Sunshine, J. H. (2004). RADPEER quality assurance program: a multifacility study of interpretive disagreement rates. *Journal of the American College of Radiology, 1*(1), 59–65.

Brams, S., Ziv, G., Hooge, I. T., Levin, O., De Brouwere, T., & Verschakelen, J., . . .Helsen, W. F. (2020). Focal lung pathology detection in radiology: Is there an effect of experience on visual search behavior? *Attention, Perception, & Psychophysics, 82*, 2837–2850.

Bruno, M. A., Walker, E. A., & Abujudeh, H. H. (2015). Understanding and confronting our mistakes: The epidemiology of error in radiology and strategies for error reduction. *Radiographics, 35*(6), 1668–1676.

Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review, 15*(4), 795–801.

Chen, W., HolcDorf, D., McCusker, M. W., Gaillard, F., & Howe, P. D. (2017). Perceptual training to improve hip fracture identification in conventional radiographs. *PLoS One, 12*(12), e0189192.

Chen, X., & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research, 46*(24), 4118–4133.

Donald, J. J., & Barnard, S. A. (2012). Common patterns in 558 diagnostic radiology errors. *Journal of Medical Imaging and Radiation Oncology, 56*(2), 173–178, doi:10.1111/j.1754-9485.2012.02348.x.

Doshi, A. M., Huang, C., Melamud, K., Shanbhogue, K., Slywotsky, C., & Taffel, M., . . .Kim, D. (2019). Utility of an automated radiology-pathology

feedback tool. *Journal of the American College of Radiology, 16*(9), 1211–1217.

Drew, T., Vo, M. L., Olwal, A., Jacobson, F., Seltzer, S. E., & Wolfe, J. M. (2013). Scanners and drillers: Characterizing expert visual search through volumetric images. *Journal of Vision, 13*(10):3, 1–13, doi:10.1167/13.10.3.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433–458.

Eckstein, M. P., Lago, M. A., & Abbey, C. K. (2017). The role of extra-foveal processing in 3D imaging. *Proceedings of SPIE–the International Society for Optical Engineering, 10136*, 101360E, doi:10.1117/12.2255879.

Ekpo, E. U., Alakhras, M., & Brennan, P. (2018). Errors in mammography cannot be solved through technology alone. *Asian Pacific Journal of Cancer Prevention, 19*(2), 291–301, doi:10.22034/apjcp.2018.19.2.291.

Elmore, J. G., Cook, A. J., Bogart, A., Carney, P. A., Geller, B. M., & Taplin, S. H., . . .Miglioretti, D. L. (2016). Radiologists' interpretive skills in screening vs. diagnostic mammography: Are they related? *Clinical Imaging, 40*(6), 1096–1103, doi:10.1016/j.clinimag.2016.06.014.

Fahle, M. (2005). Perceptual learning: Specificity versus generalization. *Current Opinion in Neurobiology, 15*(2), 154–160, doi:10.1016/j.conb.2005.03.010.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology. Human Perception and Performance, 18*(4), 1030–1044.

Frank, S. M., Qi, A., Ravasio, D., Sasaki, Y., Rosen, E. L., & Watanabe, T. (2020). Supervised learning occurs in visual perceptual learning of complex natural images. *Current Biology, 30*(15), 2995–3000.e3.

Frank, S. M., Reavis, E. A., Tse, P. U., & Greenlee, M. W. (2014). Neural mechanisms of feature conjunction learning: Enduring changes in occipital cortex after a week of training. *Human Brain Mapping, 35*(4), 1201–1211.

Funaki, B., Szymski, G., & Rosenblum, J. (1997). Significant on-call misses by radiology residents interpreting computed tomographic studies: Perception versus cognition. *Emergency Radiology, 4*(5), 290–294, doi:10.1007/BF01461735.

Guerlain, S., Green, K. B., LaFollette, M., Mersch, T. C., Mitchell, B. A., & Poole, G. R., . . . Chekan, E. G. (2004). Improving surgical pattern recognition through repetitive viewing of video clips. *IEEE*

Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 34(6), 699–707.

Gunderman, R., Williamson, K., Fraley, R., & Steele, J. (2001). Expertise: Implications for radiological education. *Academic Radiology, 8*(12), 1252–1256, doi:10.1016/s1076-6332(03)80708-0.

Hamnett, H. J., & Jack, R. E. (2019). The use of contextual information in forensic toxicology: An international survey of toxicologists' experiences. *Science & Justice, 59*(4), 380–389, doi.org/10.1016/j.scijus.2019.02.004.

Holland, K., Sun, S., Gackle, M., Goldring, C., & Osmar, K. (2019). A qualitative analysis of human error during the DIBH procedure. *Journal of Medical Imaging and Radiation Sciences, 50*(3), 369–377.e1.

Hou, X., Harel, J., & Koch, C. (2012). Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(1), 194–201, doi:10.1109/TPAMI.2011.146.

Huang, X., Shen, C., Boix, X., & Zhao, Q. (2015). Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks. *Paper presented at the Proceedings of the IEEE International Conference on Computer Vision*, (pp. 262–270).

Jampani, V., Sivaswamy, J., & Vaidya, V. (2012). Assessment of computational visual attention models on medical images. *Paper presented at the Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing* (pp. 1–8).

Johnston, I. A., Ji, M., Cochrane, A., Demko, Z., Robbins, J. B., Stephenson, J. W., . . . Green, C. S. (2020). Perceptual learning of appendicitis diagnosis in radiological images. *Journal of Vision, 20*(8):16, 1–17, doi:10.1167/jov.20.8.16.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences of the United States of America, 88*, 4966–4970.

Kelahan, L. C., Fong, A., Blumenthal, J., Kandaswamy, S., Ratwani, R. M., & Filice, R. W. (2019). The radiologist's gaze: Mapping three-dimensional visual search in computed tomography of the abdomen and pelvis. *Journal of Digital Imaging, 32*(2), 234–240.

Kellman, P. J. (2013). Adaptive and perceptual learning technologies in medical education and training. *Military Medicine, 178*(10 Suppl), 98–106, doi:10.7205/MILMED-D-13-00218.

Kellman, P. J., & Garrigan, P. (2009). Perceptual learning and human expertise. *Physics of Life Reviews, 6*(2), 53–84, doi:10.1016/j.plrev.2008.12.001.

Kellman, P. J., & Kaiser, M. K. (2016). Perceptual learning modules in flight training. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 38*(18), 1183–1187, doi:10.1177/154193129403801808.

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science, 2*(2), 285–305, doi:10.1111/j.1756-8765.2009.01053.x.

Krasne, S., Hillman, J., Kellman, P. J., & Drake, T. (2013). Applying perceptual and adaptive learning techniques for teaching introductory histopathology. *Journal of Pathology Informatics, 4*(1), 34–34, doi:10.4103/2153-3539.123991.

Krupinski, E. A. (1996). Visual scanning patterns of radiologists searching mammograms. *Academic Radiology, 3*(2), 137–144.

Krupinski, E. A. (2010). Current perspectives in medical image perception. *Attention, Perception, & Psychophysics, 72*(5), 1205–1217, doi:10.3758/APP.72.5.1205.

Krupinski, E. A. (2011). The role of perception in imaging: Past and future. *Seminars in Nuclear Medicine, 41*(6), 392–400, doi:10.1053/j.semnuclmed.2011.05.002.

Krupinski, E. A., Berger, W. G., Dallas, W. J., & Roehrig, H. (2003). Searching for nodules: What features attract attention and influence detection? *Academic Radiology, 10*(8), 861–868, doi.org/10.1016/S1076-6332(03)00055-2.

Krupinski, E. A., Graham, A. R., & Weinstein, R. S. (2013). Characterizing the development of visual search expertise in pathology residents viewing whole slide images. *Human Pathology, 44*(3), 357–364, doi:10.1016/j.humpath.2012.05.024.

Kundel, H. L. (1989). Perception errors in chest radiography. In Seminars in Respiratory Medicine. *Thieme Medical Publishers, Inc. 10*(3), 203–210.

Kundel, H. L. (2015). Visual search and lung nodule detection on CT scans. *Radiology, 274*(1), 14–16, doi:10.1148/radiol.14142247.

Kundel, H. L., Nodine, C. F., & Carmody, D. (1978). Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative Radiology, 13*(3), 175–181.

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology, 242*(2), 396–402.

Li, Z. S., Toh, Y. N., Remington, R. W., & Jiang, Y. V. (2020). Perceptual learning in the identification of lung cancer in chest radiographs. *Cognitive Research: Principles and Implications, 5*(1), 4.

Manning, D., Ethell, S., Donovan, T., & Crawford, T. (2006). How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography, 12*(2), 134–142.

Maxfield, J. T., & Zelinsky, G. J. (2012). Searching through the hierarchy: How level of target categorization affects visual search. *Visual Cognition, 20*(10), 1153–1163, doi:10.1080/13506285.2012.735718.

McCamy, M. B., Otero-Millan, J., Di Stasi, L. L., Macknik, S. L., & Martinez-Conde, S. (2014). Highly informative natural scene regions increase microsaccade production during visual scanning. *The Journal of Neuroscience, 34*(8), 2956–2966, doi:10.1523/JNEUROSCI.4448-13.2014.

Mercan, E., Shapiro, L. G., Brunyé, T. T., Weaver, D. L., & Elmore, J. G. (2018). Characterizing diagnostic search patterns in digital breast pathology: Scanners and drillers. *Journal of Digital Imaging, 31*(1), 32–41.

Miller, W. T., Jr, Marinari, L. A., Barbosa, E., Jr, Litt, H. I., Schmitt, J. E., & Mahne, A., . . .Akers, S. R. (2015). Small pulmonary artery defects are not reliable indicators of pulmonary embolism. *Annals of the American Thoracic Society, 12*(7), 1022–1029.

Nakashima, R., Komori, Y., Maeda, E., Yoshikawa, T., & Yokosawa, K. (2016). Temporal characteristics of radiologists' and novices' lesion detection in viewing medical images presented rapidly and sequentially. *Frontiers in Psychology, 7*, 1553, doi:10.3389/fpsyg.2016.01553.

Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research, 51*(13), 1526–1537.

Nodine, C. F., & Krupinski, E. A. (1998). Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. *Academic Radiology, 5*(9), 603–612.

Nodine, C. F., & Kundel, H. L. (1987). The cognitive side of visual search in radiology. J. K. O'Regan, & A. Levy-Schoen (Eds.), *Eye Movements from Physiology to Cognition* (pp. 573–582). New York: Elsevier.

Nodine, C. F., Kundel, H. L., Lauver, S. C., & Toto, L. C. (1996). Nature of expertise in searching mammograms for breast masses. *Academic Radiology, 3*(12), 1000–1006.

Nodine, C. F., & Mello-Thoms, C. (2010). The role of expertise in radiologic image interpretation. In:E., Samei, & E., Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques* (pp. 139–156). Cambridge: Cambridge University Press.

Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., & Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: Foundations for a common saccadic generator. *Journal of Vision, 8*(14):21, 21–18, doi:10.1167/8.14.21/8/14/21/.

Pirnejad, H., Niazkhani, Z., & Bal, R. (2013). Clinical communication in diagnostic imaging studies. *Applied Clinical Informatics, 4*(04), 541–555.

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision, 40*(1), 23.

Ralph, B. C., Seli, P., Cheng, V. O., Solman, G. J., & Smilek, D. (2014). Running the figure to the ground: Figure-ground segmentation during visual search. *Vision Research, 97*, 65–73.

Rappaport, S. J., Humphreys, G. W., & Riddoch, M. J. (2013). The attraction of yellow corn: Reduced attentional constraints on coding learned conjunctive relations. *Journal of Experimental Psychology: Human Perception and Performance, 39*(4), 1016.

Rauschecker, A. M., Rudie, J. D., Xie, L., Wang, J., Duong, M. T., & Botzolakis, E. J., . . . Bryan, R. N. (2020). Artificial intelligence system approaching neuroradiologist-level differential diagnosis acc uracy at brain MRI. *Radiology, 295*(3), 626–637.

Riche, N., Mancas, M., Duvinage, M., Mibulumukini, M., Gosselin, B., & Dutoit, T. (2013). Rare2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing: Image Communication, 28*(6), 642–658.

Rosenholtz, R., Huang, J., & Ehinger, K. A. (2012). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology, 3*, 13.

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision, 12*(4):14, 1–17, doi:10.1167/12.4.14.

Rosenkrantz, A. B., & Bansal, N. K. (2016). Diagnostic errors in abdominopelvic CT interpretation: Characterization based on report addenda. *Abdominal Radiology, 41*(9), 1793–1799.

Saffell, T., & Matthews, N. (2003). Task-specific perceptual learning on speed and direction discrimination. *Vision Research, 43*(12), 1365–1374.

Sagi, D. (2011). Perceptual learning in vision research. *Vision Research, 51*(13), 1552–1566.

Seitz, A. R. (2020). Perceptual expertise: How is it achieved? *Current Biology, 30*(15), R875–R878, doi.org/10.1016/j.cub.2020.06.013.

Seitz, A. R., Kim, D., & Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron, 61*(5), 700–707, doi:10.1016/j.neuron.2009.01.016.

Seitz, A. R., Protopapas, A., Tsushima, Y., Vlahou, E. L., Gori, S., & Grossberg, S., …Watanabe, T. (2010). Unattended exposure to components of speech sounds yields same benefits as explicit auditory training. *Cognition, 115*(3), 435–443.

Semizer, Y., Michel, M. M., Evans, K. K., & Wolfe, J. M. (2018). Texture as a diagnostic signal in mammograms. Proceedings of the 40th Annual Meeting of the Cognitive Science Society, Madison, WI: Cognitive Science Society, pp. 1043–1048.

Sheridan, H., & Reingold, E. M. (2017). The holistic processing account of visual expertise in medical image perception: A review. *Frontiers in Psychology, 8*, 1620, doi:10.3389/fpsyg.2017.01620.

Siegle, R. L., Baram, E. M., Reuter, S. R., Clarke, E. A., Lancaster, J. L., & McMahan, C. A. (1998). Rates of disagreement in imaging interpretation in a group of community hospitals. *Academic Radiology, 5*(3), 148–154.

Sireteanu, R., & Rettenbach, R. (2000). Perceptual learning in visual search generalizes over tasks, locations, and eyes. *Vision Research, 40*(21), 2925–2949.

Sowden, P. T., Davies, I. R., & Roling, P. (2000). Perceptual learning of the detection of features in x-ray images: a functional role for improvements in adults' visual sensitivity? *Journal of Experimental Psychology. Human Perception and Performance, 26*(1), 379–390.

Sowden, P. T., Rose, D., & Davies, I. R. (2002). Perceptual learning of luminance contrast detection: Specific for spatial frequency and retinal location but not orientation. *Vision Research, 42*(10), 1249–1258.

Steinman, S. B. (1987). Serial and parallel search in pattern vision? *Perception, 16*(3), 389–398.

Tavakoli, H. R., Rahtu, E., & Heikkilä, J. (2011). Fast and efficient saliency detection using sparse sampling and kernel density estimation. In Scandinavian conference on image analysis (pp. 666–675). Berlin, Heidelberg: Springer.

Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. *Psychonomic Bulletin & Review, 11*(1), 65–70.

Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 17*(3), 652–676.

Ukweh, O. N., Ugbem, T. I., Okeke, C. M., & Ekpo, E. U. (2019). Value and diagnostic efficacy of fetal morphology assessment using ultrasound in a poor-resource setting. *Diagnostics, 9*(3), 109.

Waite, S., Farooq, Z., Grigorian, A., Sistrom, C., Kolla, S., & Mancuso, A., . . .Macknik, S. L. (2020). A review of perceptual expertise in radiology-How it develops, how we can test it, and why humans still matter in the era of Artificial Intelligence. *Academic Radiology, 27*(1), 26–38, doi:10.1016/j.acra.2019.08.018.

Waite, S., Grigorian, A., Alexander, R. G., Macknik, S. L., Carrasco, M., Heeger, D. J., . . . Martinez-Conde, S. (2019). Analysis of perceptual expertise in radiology – Current knowledge and a new perspective. *Frontiers in Human Neuroscience, 13*, 213, doi:10.3389/fnhum.2019.00213.

Waite, S., Scott, J., Gale, B., Fuchs, T., Kolla, S., & Reede, D. (2017). Interpretive error in radiology. *American Journal of Roentgenology, 208*(4), 739–749.

Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413*(6858), 844–848, doi:10.1038/35101601.

Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: Toward a comprehensive theory. *Annual Review of Psychology, 66*, 197–221, doi:10.1146/annurev-psych-010814-015214.

Wen, G., Aizenman, A., Drew, T., Wolfe, J. M., Haygood, T. M., & Markey, M. K. (2016). Computational assessment of visual search strategies in volumetric medical images. *Journal of Medical Imaging, 3*(1), 015501–015501.

Wen, G., Rodriguez-Nino, B., Pecen, F. Y., Vining, D. J., Garg, N., & Markey, M. K. (2017). Comparative study of computational visual attention models on two-dimensional medical images. *Journal of Medical Imaging, 4*(2), 025503, doi:10.1117/1.Jmi.4.2.025503.

Wildegger, T., Riddoch, J., & Humphreys, G. W. (2015). Stored color–form knowledge modulates perceptual sensitivity in search. *Attention, Perception, & Psychophysics, 77*(4), 1223–1238.

Wolfe, J. M. (1994a). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1*(2), 202–238.

Wolfe, J. M. (1994b). Visual search in continuous, naturalistic stimuli. *Vision Research, 34*(9), 1187–1195.

Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search

for feature singletons. *Journal of Experimental Psychology. Human Perception and Performance, 29*(2), 483–502.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*(6), 495–501, doi:10.1038/nrn1411nrn1411.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour, 1*, 0058.

Wolfe, J. M., Yu, K. P., Stewart, M. I., Shorter, A. D., Friedman-Hill, S. R., & Cave, K. R. (1990). Limitations on the parallel guidance of visual search: Color× color and orientation× orientation conjunctions. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 879.

Wood, B. P. (1999). Visual expertise. *Radiology, 211*(1), 1–3, doi:10.1148/radiology.211.1.r99ap431.

Wood, G., Knapp, K. M., Rock, B., Cousens, C., Roobottom, C., & Wilson, M. R. (2013). Visual expertise in detecting and diagnosing skeletal fractures. *Skeletal Radioogyl, 42*(2), 165–172.

Zelinsky, G. J., Peng, Y., Berg, A. C., & Samaras, D. (2013). Modeling guidance and recognition in categorical search: Bridging human and computer object detection. *Journal of Vision, 13*(3):30, 1–20, doi:10.1167/13.3.30.