

# Conclusion to the Special Series on 2D and 3D Imaging: Perspectives in Human and Model Observer Performance

Claudia Mello-Thoms,<sup>a</sup> Craig K. Abbey,<sup>b</sup> and Elizabeth A. Krupinski<sup>c</sup>

<sup>a</sup>University of Iowa, Department of Radiology, Iowa City, Iowa, United States

<sup>b</sup>University of California, Santa Barbara, Department of Psychological and Brain Sciences, Santa Barbara, California, United States

<sup>c</sup>Emory University, Department of Radiology and Imaging Sciences, Atlanta, Georgia, United States

Medical imaging is quickly expanding from the analysis of a single 2D image to the reading of hundreds (even thousands) of stacked 2D images, in 3D imaging. While visual search and image perception are relatively well understood in 2D imaging, such understanding is in its infancy in 3D images. Models of image perception that explain lesion detection in 2D images (such as the holistic model of Nodine and Kundel<sup>1</sup>) cannot be directly applied to 3D images, as there is no initial “gist” to be obtained from the first slice in a stack, because the lesion of interest may not be visible at that point. Model observers, developed to explain and mimic human behavior, also cannot be immediately translated from one domain into the other.

The objectives of this two-part special series were to characterize the differences between the use of 2D and 3D imaging as it relates to understanding human and computer decision-making processes. Part 1 of the series ([Volume 7, Issue 5](#)) included two contributions: “[2D CNN versus 3D CNN for false-positive reduction in lung cancer screening](#)” by Juezhao Yu and colleagues, and “[Eye tracking reveals expertise-related differences in the time-course of medical image inspection and diagnosis](#)” by Tad T. Brunyé and colleagues. Yu et al. studied whether 3D convolution neural networks (CNN) were superior to 2D CNNs to reduce false positives in lung cancer screening using low dose computed tomography. They found that the 3D-CNN model has a better performance in false-positive reduction, but the improvement is limited, and training requires substantial computational resources. Brunyé et al. used eye-position tracking to study the reading of breast histopathology slides by 92 pathology residents and attending pathologists and explored the differences in visual search behavior between these two groups. They found that detection rates were similar, but faculty members had more and longer durations on critical regions, and year of residency predicted odds of fixating critical regions.

Part 2 of the series ([Volume 8, Issue 4](#)) included six contributions. Ye Li and colleagues, in “[DeepAMO: a multi-slice, multi-view anthropomorphic model observer for visual detection tasks performed on volume images](#),” presented a multislice, multiview anthropomorphic model observer (DeepAMO) for visual detection tasks performed in volumetric images showing that it has good potential to reproduce absolute observer performance not just relative ranks. Kehn E. Yapp, Patrick Brennan, and Ernest Ekpo, in “[Endodontic disease detection: digital periapical radiography versus cone-beam computed tomography](#),” presented a systematic review on the relatively novel topic of the use of digital periapical radiography (a 2D image) versus cone-beam computed tomography (a 3D image) for the detection of endodontic disease, concluding that the literature is quite mixed and therefore a good area for more study. Craig K. Abbey, Miguel A. Lago, and Miguel P. Eckstein, in “[Comparative observer effects in 2D and 3D localization tasks](#),” compared how human observers localize targets masked by noise and clutter as they scroll through a 3D image versus when they read a single 2D slice – not surprisingly, there were efficiency differences that varied as a function of target and background characteristics. Karim El Khoury and colleagues., in “[Improved 3D U-Net robustness against JPEG 2000 compression for male pelvic organ segmentation in radiotherapy](#),” analyzed and compared 3D and 3D U-Nets versus JPEG 2000 compression for the segmentation of male pelvic organs in radiotherapy, with the 3D U-Net being 50% more robust. Lauren H. Williams and colleagues, in “[Characteristics](#)

of expert search behavior in volumetric medical image interpretation,” examined whether radiologists use the holistic model of Kundel and Nodine (derived for 2D images) when reading volumetric images (appears as if not as much), and whether observers with different levels of expertise relied on drilling or scanning search strategies when reading chest computed tomography looking for lung nodules (drilling is associated with better accuracy). Finally, Lago, Abbey, and Eckstein, in “Medical image quality metrics for foveated model observers,” showed how the effects of peripheral vision in 3D search can be explained (to first order) by distance to the nearest fixation and target detection performance measured as a function of retinal eccentricity.

Overall, the contributions to the special series revealed a great deal about 2D versus 3D perception and decision-making with medical images using a wide variety of image types, tasks, and approaches. We hope that the studies reported may lead to future ones that further elucidate the perceptual and cognitive mechanisms underlying 3D image interpretation. It was interesting to see how different 3D image viewing and interpretation really is from 2D – sometimes more efficient sometimes less, sometimes more accurate sometimes not. These studies illustrate very nicely how much we know and how more we still need to know in order to better assist clinicians to use the wide variety of complex images to improve patient care.

## References

1. C.F. Nodine and H.L. Kundel, “The cognitive side of visual search in radiology,” in *Eye Movements: from Physiology to Cognition*, J. K O’Regan and A. Levy-Schoen, Eds., pp. 573–582, Elsevier Science Publishers BV, North-Holland, New York (1987).