HHS Public Access

Author manuscript

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2020 May 20.

Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2020 February ; 11316: . doi:10.1117/12.2549362.

Evaluation of Convolutional Neural Networks for Search in 1/f $^{2.8}$ Filtered Noise and Digital Breast Tomosynthesis Phantoms

Aditya Jonnalagadda[1,4], **Miguel A. Lago**[2,4], **Bruno Barufaldi**[3], **Predrag R. Bakic**[3], **Craig K. Abbey**[2], **Andrew D. Maidment**[3], **Miguel P. Eckstein**[1,2]

[1]Department of Electrical & Computer Engineering, UC Santa Barbara, Santa Barbara, CA, USA

[2]Department of Psychological & Brain Sciences, UC Santa Barbara, Santa Barbara, CA, USA

[3]Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA

[4]These authors contributed equally to this work

Abstract

With the advent of powerful convolutional neural networks (CNNs), recent studies have extended early applications of neural networks to imaging tasks thus making CNNs a potential new tool for assessing medical image quality. Here, we compare a CNN to model observers in a search task for two possible signals (a simulated mass and a smaller simulated micro-calcification) embedded in filtered noise and single slices of Digital Breast Tomosynthesis (DBT) virtual phantoms. For the case of the filtered noise, we show how a CNN can approximate the ideal observer for a search task, achieving a statistical efficiency of 0.77 for the microcalcification and 0.78 for the mass. For search in single slices of DBT phantoms, we show that a Channelized Hotelling Observer (CHO) performance is affected detrimentally by false positives related to anatomic variations and results in detection accuracy below human observer performance. In contrast, the CNN learns to identify and discount the backgrounds, and achieves performance comparable to that of human observer and superior to model observers (Proportion Correct for the microcalcification: CNN = 0.96; Humans = 0.98; CHO = 0.84; Proportion Correct for the mass: CNN = 0.98; Humans = 0.83; CHO = 0.51). Together, our results provide an important evaluation of CNN methods by benchmarking their performance against human and model observers in complex search tasks.

1. INTRODUCTION

The use of model observers for image quality assessment has been extensively used in the field of medical imaging for computer-generated and anatomical background and signals located at one or a few specified locations.[1–4] Among these observers, the most used models include the Ideal Observer (IO) and the Channelized Hotelling Observer (CHO). The IO is restricted to very specific situations in which background and signal statistics are known. The Ideal Observer is used to calculate the upper bound performance for a perceptual task and used to benchmark human performance.[5–9] With more realistic phantom simulations or real anatomical backgrounds, the IO is not computationally tractable and researchers rely on

Publication History: This work has not been published in any other conference/journal yet but we have plans to submit it to a journal.

approximations like the CHO model which incorporates a feature extraction stage through a set of linear channels.[10] When these channels mimic the early visual processing of the human visual system the CHO model can more accurately predict human performance than an ideal observer.[11]

Convolutional Neural Networks (CNN) have recently been applied to imaging tasks[12,13] and have proved to be a good approximation to the Ideal Observer[14,15] in simple Background-Known-Exactly and Location-Known-Exactly tasks. Here, we extend such work and compare CNN's performance to human and model observers for the search of two noise and possible signals (a simulated mass and a smaller simulated micro-calcification) embedded in $1/f^{2.8}$ filtered single slices of Digital Breast Tomosynthesis (DBT) virtual phantoms.

In this research, we extend previous evaluations of CNNs in medical imaging relevant tasks to complex search tasks, a variety of background and comparisons to standard model observers and humans. The results are novel and informative in understanding the potential contributions that CNNs might have to the field of medical image quality.

2. METHODS

2.1 Experiment Design

We considered two search tasks with different backgrounds: filtered noise and backgrounds generated from a virtual breast digital phantom. The task was a yes/no detection task with one of two different possible targets. Targets were present in 50% of the trials with a 50% probability of being one or the other. The first version of the experiment used an image generated with correlated Gaussian noise (size 1024×820 pixels) filtered to match the noise power spectrum of mammograms $1/f^{2.8}$.[16] In this case, the two targets were a microcalcification-like signal (MCALC, a bright sphere of 6 voxels diameter) and a mass-like signal (MASS, Gaussian blob of about 25 voxels diameter).

The second version of the experiment involved the use of a single-slice Digital Breast Tomosynthesis (DBT) virtual phantoms (size 2048×1792 pixels) using the OpenVCT virtual breast imaging tool from the University of Pennsylvania.[17] OpenVCT generates full phantom DBT images including different tissues (skin, Cooper's ligaments, adipose and glandular) in a realistic manner. Two targets were simulated as a single microcalcification and a mass in the phantom. Additionally, for DBT images, we designed a psychophysics experiment in order to know human performance. On each trial, 6 observers searched only for a given type of signal. Each participant participated in 300 trials per signal type. Figure 1 describes the procedure of the experiment for a single trial for each signal-type condition. Observers responded whether the signal was present or absent. The signals were present in 50% of the trials.

2.2 Model observers

Three different model observers were implemented for the correlated noise images: 1) the Ideal Observer (IO), which optimally uses the visual information to calculate posterior probabilities and achieves upper bound performance;[18] 2) the Non-Prewhitening model observer with Eye filter (NPWE), which uses the signal as template with an eye filter that

attenuates spatial frequencies based on the human Contrast Sensitivity Function;[19,20] and 3) the Channelized Hotelling Observer (CHO) which processes the stimulus with channels and combines them linearly optimally.[1,21] On the other hand, for DBT phantoms, the IO is mathematically non-tractable and thus we applied only the CHO model observer.

2.2.1 Ideal Observer Model—The ideal observer uses statistical information from the noise field and the target to build an optimal template.[18] The ideal observer template $\mathbf{w}$ was convolved (*) with the image $\mathbf{g}$ thus calculating a decision variable $\lambda = \mathbf{w_{IO}} * \mathbf{g}$. To build the template we used the inverse of the covariance matrix of the noise field $\mathbf{K_g}$ and the signal $\mathbf{s}$: $\mathbf{w_{IO}} = \mathbf{K_g}^{-1}\mathbf{s}$.

2.2.2 Non-prewhitening with Eye Filter Model—The non-prewhitening observer with eye filter (NPWE) makes use of a mathematical representation of the target (matched filter) and an eye filter that considers the contrast sensitivity function of the human visual system.[19] The spatial filter used the following expression.

$$E(\rho) = \rho^{\alpha} exp(-\beta\rho^{\gamma})$$

Where $\alpha = 1.4$, $\beta = 0.013$ and $\gamma = 2.6$ and ρ is the radial spatial frequency in cycles per degree[2].

The model observer's template $\mathbf{w_{NPWE}}$ and decision variable λ are constructed as follows.

$$\hat{\mathbf{s}} = FFT(\mathbf{s})$$
$$\mathbf{w_{NPWE}} = FFT^{-1}\left(\widehat{\mathbf{E}}^2\mathbf{s}\right)$$
$$\lambda = \mathbf{w_{NPWE}} * \mathbf{g}$$

Where *FFT* is the fast Fourier transform and the carat symbols ^ refer to the frequency domain.

2.2.3 Channelized Hotelling Observer—The channelized hotelling observer (CHO) was built using Gabor channels (8 orientations and 6 spatial frequencies: 0.5, 1, 2, 4, 8 and 16 cycles per degree)[21]. The set of channels $\mathbf{T}$ is used to extract different features from the signal and build the template $\mathbf{w_{CHO}}$ and decision variable λ as follows.

$$\mathbf{v} = \mathbf{T}^t\mathbf{s}$$
$$\mathbf{K_v} = \mathbf{T}^t\mathbf{K_g}\mathbf{T}$$
$$\mathbf{w_{CHO}} = \mathbf{T}\mathbf{K_v}^{-1}\mathbf{v}$$
$$\lambda = \mathbf{w_{CHO}} * \mathbf{g}$$

2.3 Convolutional Neural Network

The Convolutional Neural Network (CNN) developed was based on Mask R-CNN[22,23] (Region-based CNN). It was pre-trained to do instance segmentation for objects from the

MSCOCO image dataset. This network is an extension to Faster R-CNN[24] originally trained to do object classification and localization. Both Mask R-CNN and Faster R-CNN have two stages. The first stage is the same for both networks and utilizes a Region Proposal Network (RPN), which proposes candidate object bounding boxes. The second stage of Faster R-CNN operates on these candidates using RoIPool[25], pooling convolutional features at candidate locations and performs object classification and bounding-box regression. On the other hand, the second stage of Mask R-CNN has an additional pathway to output a binary mask for each Region of Interest (RoI). The method used by Mask R-CNN to separate segmentation from object classification contributes to better performance.

The MSCOCO dataset had 81 different classes including the background. Mask-RCNN implementation[26] was adapted for detecting cell nuclei in divergent images as part of the 2018 Data Science Bowl challenge. We reduced the number of classes to two. ResNet-50[27] was used as the backbone of this network. Anchor sizes in the RPN network are based on the expected tumor size. By default, the anchor is a square and has five possible side lengths, (8, 16, 32, 64, 128) pixels. For some tasks, these prior sizes are modified based on dataset statistics. For training, the network requires the signal image along with a binary mask corresponding to each of the signal locations. For the filtered noise images, we have access to the exact signal information and the Gaussian noise that was added to the signal. We thresholded this signal with a Gaussian noise mask in order to generate the binary mask segmentation ground truth. For the phantoms, we do not know the exact noise mask added to the signal. Therefore, we used a bounding box around the signal location in order to generate the binary segmentation mask where the bounding box size is determined by the original signal size used in all phantoms.

Data were separated into training, validation and test subsets. For filtered noise backgrounds, the split was 1000 images (training), 200 (validation) and 200 (test) respectively. For phantoms, the split was 468 (training), 52 (validation) and 52 (test) respectively. There was no overlap between different sets. We initialized the network weights with ResNet-50 classification network weights. We started with training the final layers for a few epochs (40 for filtered noise backgrounds, 20 for phantoms) and after that, we trained the whole network for the remaining epochs (80 for filtered noise backgrounds, 20 for phantoms) with a learning rate of 0.001. For the filtered noise backgrounds, we had additional training for the whole network for 40 epochs with a learning rate of 0.0001. After training for a predefined number of epochs, we chose the model corresponding to the epoch which gave us the least validation loss. We tested the model on signal present and signal absent images.

2.4 Figures of merit for model and human performance

For the model observers (IO, CHO, and NPWE), we constructed a ROC curve using, for each trial, the highest template response across locations. We obtained the Proportion Correct (PC) by choosing the corresponding optimal decision threshold (PC maximizing) from the ROC curves for each model.

The network outputs multiple detections within the same image with associated probability values for signal presence. We pick the maximum value within the image as the probability of the signal's presence in the image. Finally, we built a ROC curve with the maximum

probabilities across each of the images. We chose the optimal decision threshold value that maximized the PC. This is the PC that is reported as the CNN performance (see Results).

For the filtered noise images, in order to compare model efficiency, we calculated a PC for each target amplitude (contrast). Then, we took an amplitude that matched CNN and model observer performance and calculated the efficiency using the following formula:

$$efficiency = \frac{amplitude_{model1}{}^2}{amplitude_{model2}{}^2}$$

We note that such procedure to use efficiency for the DBT phantoms was not possible given the computational cost of generating the phantoms with varying signal amplitudes. For the DBT phantoms, we simply compared the performance of CNN to that of human observers.

3. RESULTS

Figure 2 shows the corresponding amplitudes for the signals in order to achieve the same performance (PC of 0.6 for MCALC and 0.77 for MASS) for the CNN and the corresponding model observer (IO, CHO or NPWE). We tuned the signal amplitude to achieve similar performance for the four models. The efficiencies relative to the IO for the CHO, NPWE, and CNN were: 0.16, 0.09, 0.88 for the MCALC and 0.9, 1, 0.9 for the MASS, respectively).

Figure 3 shows proportion correct detection for the two signals for human observers, the CHO model, and the CNN in the single slice of the DBT phantoms. Due to a lack of access to the exact statistical properties of the signal and backgrounds, the ideal observer implementation is not possible. We did not calculate the NPWE model which also requires knowledge of the mathematical function describing the signal luminance modulation through space. The interaction between CNN/CHO model performance and signal type for single slice DBT images was opposite to the filtered noise background. For the DBT images, the CHO model performed better relative to the CNN for the microcalcification rather than the masses.

4. DISCUSSION

For the filtered noise images, the efficiency of the CHO and the NPWE was low for the microcalcification but higher for the mass signals. This reflects the bottleneck in the NPWE and CHO models to access high spatial frequency information for the microcalcification signal. The CHO model has a set of spatial frequency channels with the highest channel at 32 cycles/degrees. The contrast sensitivity function of the NPWE also has a drop-off in sensitivity for high spatial frequencies. CNN does not have such a bottleneck and thus outperforms these two model observers for such small signals. For the masses which are larger and do not contain as much power in high spatial frequencies, the CHO and NPWE performed similar to the IO and also better than the CNN.

In terms of the comparison across signal types. All models performed better with the microcalcification signal relative to the mass signal. Models required less signal amplitude

to detect the microcalcification than the mass signal. This result is explained by the differences in shape for each target and interaction with the noise background. While the microcalcification has very sharp and clear edges, the mass blends with the background, thus making it more confusing with a background that is already low-frequency noise. This follows the trend seen in previous publications.[28,29]

One limitation of the filtered noise backgrounds is that they are a statistically stationary process and do not contain anatomical structures.[30] To evaluate the CNN with a more realistic anatomical background we investigated model performance for search in single slice DBT phantoms. For the phantom images, IO calculation is not possible. We benchmarked CNN performance against the CHO model and human observer performance. Our results show a steep deterioration of the CHO model performance model compared to human observers, mainly because of the inhomogeneity of these backgrounds, which can look very different in different regions of the phantom image. In particular, the CHO had difficulty detecting searching for the mass signal. For the microcalcification, the CHO model achieved better performance but still falls below human observer performance. In contrast, the CNN performance is comparable to human performance and seems to cope better with the non-stationary anatomical structured in the background of the DBT phantoms.

5. CONCLUSIONS

Convolutional neural networks can approximate ideal observers for more complex search tasks. In addition, CNNs have the attribute of learning to discount potential distractors related to anatomical backgrounds which are a limitation of traditional model observers[31]. Together, the results extend previous results with CNNs to visual search and a variety of backgrounds and show the potential of CNNs for image quality evaluation.

REFERENCES

1. Barrett HH, Yao J, Rolland JP & Myers KJ Model observers for assessment of image quality. Proceedings of the National Academy of Sciences 90, 9758–9765 (1993).
2. Zhang Y, Pham B & Eckstein MP Evaluation of JPEG 2000 encoder options: human and model observer detection of variable signals in x-ray coronary angiograms. IEEE transactions on medical imaging 23, 613–632 (2004). [PubMed: 15147014]
3. Favazza CP, Fetterly KA, Hangiandreou NJ, Leng S & Schueler BA Implementation of a channelized Hotelling observer model to assess image quality of x-ray angiography systems. JMI 2, 015503(2015). [PubMed: 26158086]
4. Yu L et al. Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: impact of radiation dose and reconstruction algorithms. Med Phys 40, 041908(2013). [PubMed: 23556902]
5. Abbey CK & Eckstein MP Classification images for detection, contrast discrimination, and identification tasks with a common ideal observer. J Vis 6, 335–355 (2006). [PubMed: 16889473]
6. Barlow HB & Reeves BC The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. Vision Research 19, 783–793 (1979). [PubMed: 483597]
7. Geisler WS Contributions of ideal observer theory to vision research. Vision Research 51, 771–781 (2011). [PubMed: 20920517]
8. Braje WL, Tjan BS & Legge GE Human efficiency for recognizing and detecting low-pass filtered objects. Vision Research 35, 2955–2966 (1995). [PubMed: 8533334]

9. Eckstein MP, Whiting JS & Thomas JP Detection and discrimination of moving signals in Gaussian uncorrelated noise. 2712, 9–25 (1996).
10. Gallas BD & Barrett HH Validating the use of channels to estimate the ideal linear observer. J Opt Soc Am A Opt Image Sci Vis 20, 1725–1738 (2003). [PubMed: 12968645]
11. Burgess AE, Li X & Abbey CK Visual signal detectability with two noise components: anomalous masking effects. J Opt Soc Am A Opt Image Sci Vis 14, 2420–2442 (1997). [PubMed: 9291611]
12. Kupinski MA, Edwards DC, Giger ML & Metz CE Ideal observer approximation using Bayesian classification neural networks. IEEE Transactions on Medical Imaging 20, 886–899 (2001). [PubMed: 11585206]
13. Myers KJ, Anderson MP, Brown DG, Wagner RF & Hanson KM Neural network performance for binary discrimination tasks. Part II: effect of task, training, and feature preselection. in Medical Imaging 1995: Image Processing vol. 2434 828–837 (International Society for Optics and Photonics, 1995).
14. Zhou W & Anastasio MA Learning the ideal observer for SKE detection tasks by use of convolutional neural networks (Cum Laude Poster Award). in Medical Imaging 2018: Image Perception, Observer Performance, and Technology Assessment vol. 10577 1057719(International Society for Optics and Photonics, 2018).
15. Zhou W & Anastasio MA Learning the ideal observer for joint detection and localization tasks by use of convolutional neural networks. in Medical Imaging 2019: Image Perception, Observer Performance, and Technology Assessment vol. 10952 1095209(International Society for Optics and Photonics, 2019).
16. Burgess AE, Jacobson FL & Judy PF Human observer detection experiments with mammograms and power-law noise. Medical Physics 28, 419–437 (2001). [PubMed: 11339738]
17. Pokrajac DD, Maidment ADA & Bakic PR Optimized generation of high resolution breast anthropomorphic software phantoms. Medical Physics 39, 2290–2302 (2012). [PubMed: 22482649]
18. Burgess A, Wagner R, Jennings R & Barlow HB Efficiency of human visual signal discrimination. Science 214, 93–94 (1981). [PubMed: 7280685]
19. Burgess A Statistically defined backgrounds: performance of a modified nonprewhitening observer model. JOSA A 11, 1237–1242 (1994). [PubMed: 8189286]
20. Zhang Y JPEG 2000 encoder options on model observer performance in signal known exactly but variable tasks (SKEV). in Proceedings of SPIE 371–382 (2003). doi:10.1117/12.480078.
21. Eckstein MP & Whiting JS Lesion detection in structured noise. Acad Radiol 2, 249–253 (1995). [PubMed: 9419557]
22. He K, Gkioxari G, Dollar P & Girshick R Mask R-CNN. in 2961–2969 (2017).
23. Girshick R, Donahue J, Darrell T & Malik J Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. in 2014 IEEE Conference on Computer Vision and Pattern Recognition 580–587 (2014). doi:10.1109/CVPR.2014.81.
24. Ren S, He K, Girshick R & Sun J Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks in Advances in Neural Information Processing Systems 28 (eds. Cortes C, Lawrence ND, Lee DD, Sugiyama M & Garnett R) 91–99 (Curran Associates, Inc., 2015).
25. Girshick R Fast R-CNN. in 1440–1448 (2015).
26. GitHub - matterport/Mask_RCNN: Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. https://github.com/matterport/Mask_RCNN.
27. He K, Zhang X, Ren S & Sun J Deep Residual Learning for Image Recognition. in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 770–778 (2016). doi:10.1109/CVPR.2016.90.
28. Lago MA, Abbey CK & Eckstein MP A foveated channelized Hotelling search model predicts dissociations in human performance in 2D and 3D images. in Medical Imaging 2019: Image Perception, Observer Performance, and Technology Assessment vol. 10952 109520D(International Society for Optics and Photonics, 2019).
29. Eckstein MP, Lago MA & Abbey CK The role of extra-foveal processing in 3D imaging. in Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment vol. 10136 101360E(International Society for Optics and Photonics, 2017).

30. Bochud FO, Abbey CK & Eckstein MP further investigation of the effect of phase spectrum on visual detection in structured backgrounds. 1999 273–283.

31. Eckstein MP The efficiency of reading around learned backgrounds. in Proceedings of SPIE 61460N-61460N–9 (2006). doi:10.1117/12.655750.

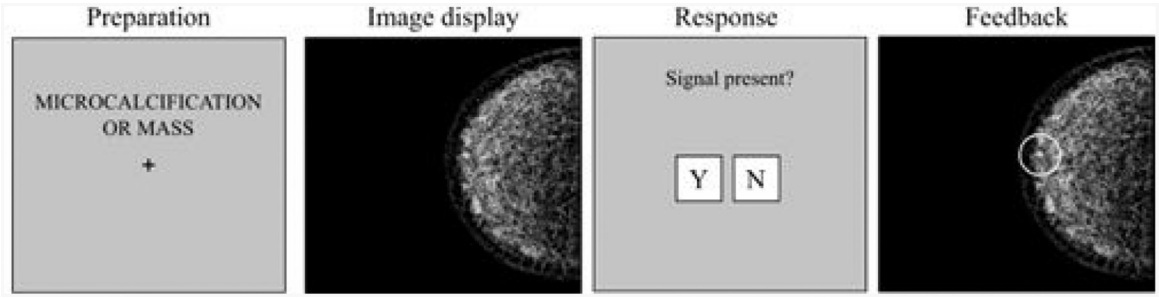

Figure 1:
Outline of the psychophysical search experiment with DBT phantoms

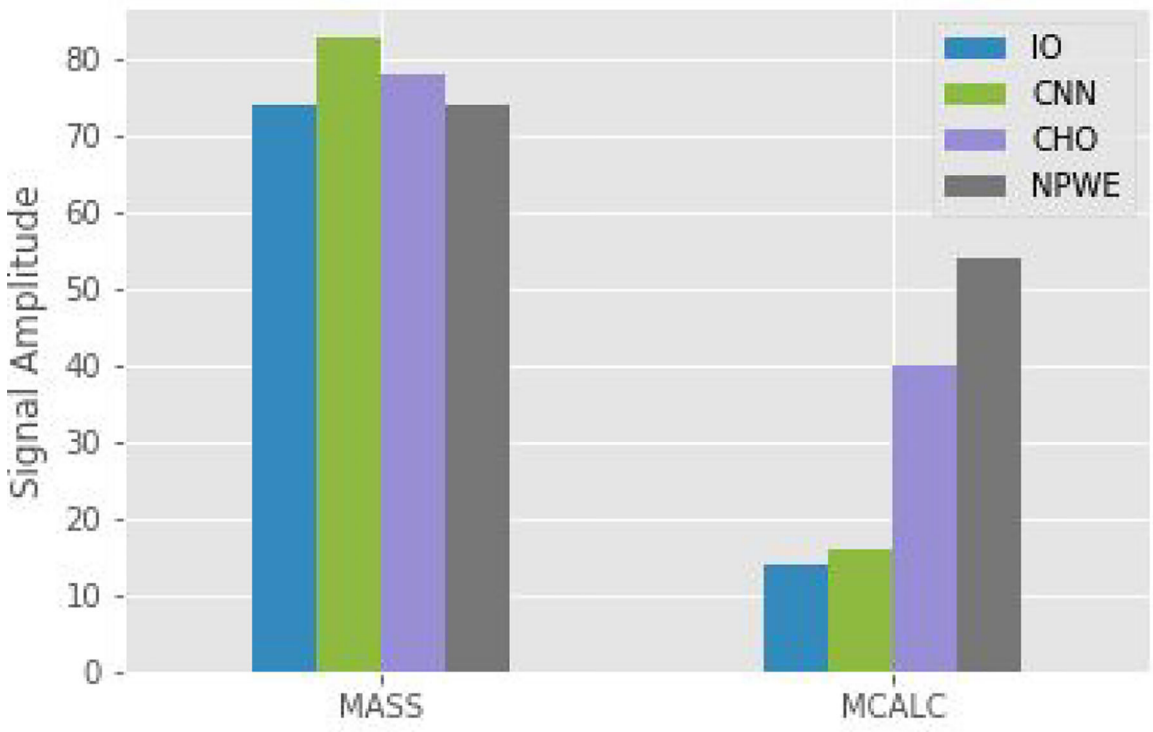

Figure 2:
Signal amplitudes for the same proportion correct for the ideal observer (IO), convolutional neural network (CNN), channelized hotelling observer (CHO) and non-prewhitening observer with eye filter (NPWE) in correlated noise.

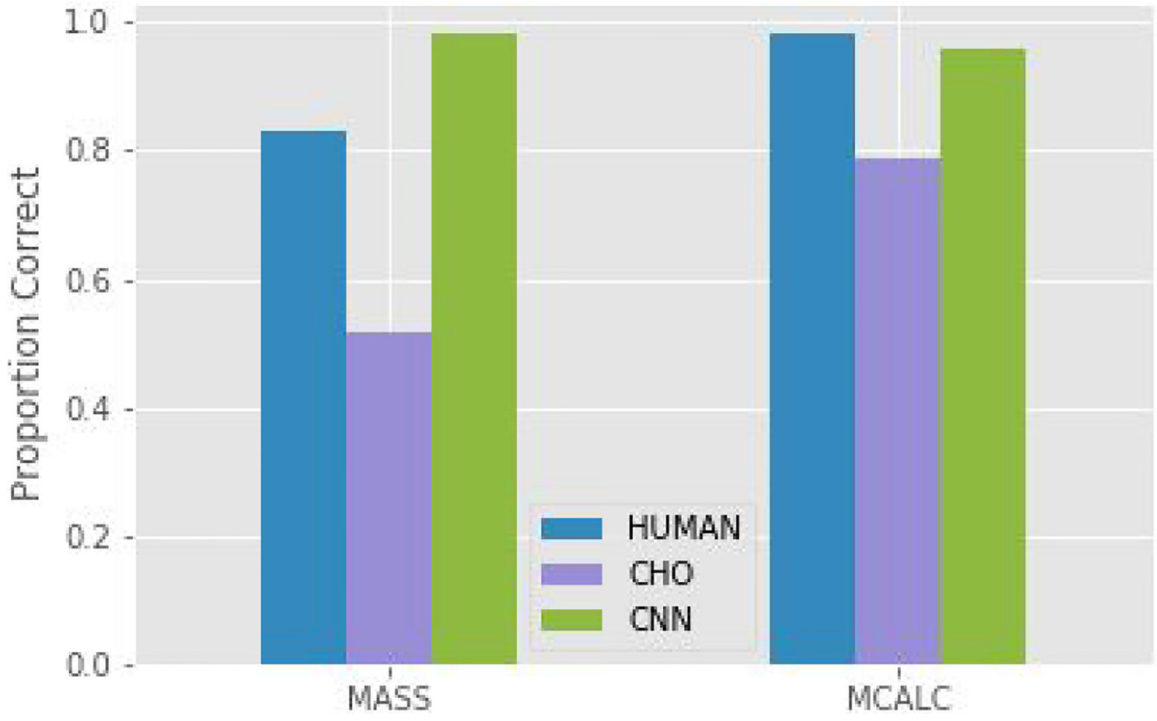

Figure 3:
Proportion correct (PC) for human observers, channelized hotelling observer (CHO) and convolutional neural network (CNN) in DBT phantoms.