Proceedings Paper:
Xu, Y, Schebesch, F, Ravikumar, N et al. (1 more author) (2019) Detection of Unseen Low-Contrast Signals Using Classic and Novel Model Observers. In: Handels, H, Deserno, TM, Maier, A, Maier-Hein, KH, Palm, C and Tolxdorff, T, (eds.) Informatik aktuell. BVM 2019: Bildverarbeitung für die Medizin, 17-19 Mar 2019, Lübeck, Germany. Springer Vieweg , pp. 212-217. ISBN 978-3-658-25325-7

https://doi.org/10.1007/978-3-658-25326-4_47

Detection of Unseen Low-Contrast Signals Using Classic and Novel Model Observers

Yiling Xu[1], Frank Schebesch[1], Nishant Ravikumar[1], Andreas Maier[1]

[1]Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
yiling.xu@fau.de

Abstract. Automatic task-based image quality assessment has been of importance in various clinical and research applications. In this paper, we propose a neural network model observer, a novel concept which has recently been investigated. It is trained and tested on simulated images with different contrast levels, with the aim of trying to distinguish images based on their quality/contrast. Our model shows promising properties that its output is sensitive to image contrast, and generalizes well to unseen low-contrast signals. We also compare the results of the proposed approach with those of a channelized Hotelling observer (CHO), on the same simulated dataset.

1 Introduction

Medical image analysis techniques have been largely applied to clinical diagnosis in a variety of imaging modalities, including mammography. An important auxiliary task for any such application is to evaluate the image quality. Human observers are the ideal reference, but too costly for frequent studies and in many cases unavailable.

As a surrogate, mathematical model observers are popular among task-based image quality assessment since 90s [1]. In a detection task, model observers are trained to distinguish between signal-present and signal-absent images, and its performance is used to assess image quality. However, a mathematical model observer requires prior knowledge about the signal, which is challenging when dealing with low-contrast images [2].

While classic model observers follow the concepts given by Barrett et al. [1], in recent years there have been attempts to use other algorithmic concepts [3]. With the advent of wide-spread use of deep neural networks, Alnowami et al. [4] proposed a deep learning-based model observer and highlighted its promising performance on both clinical and simulated mammography images. However, the model in this paper contains 5 convolutional layers and more than 570 kernels, requiring significant computational resources, more training data and a long training time. This motivates us to employ emerging deep learning techniques to design a more compact model, with fewer trainable parameters. Our goal is to train a network with affordable cost that generalizes well, and is able to identify the presence or absence of signals in unseen lower contrast images.

2 Materials and Methods

• Data Set. To train, validate and test the different designs of model observers we used a synthetic database generated using an image creation pipeline from the Conrad framework [5]. The simulation is a re-implementation of parts of a toolbox that has been suggested previously for studies of model observers [6]. Here, a Gaussian shaped signal is modeled by

$$s(x, y; A, s) = A\, e^{-\frac{x^2+y^2}{2s^2}},$$

and its contrast and size are parameterized by A and s, respectively. For a noisy background two separate random image components are drawn from a normal distribution. Noise structure is simulated for both components by convolution of each with a cone-filter, represented as;

$$c(x, y) = \begin{cases} 0 & \text{if } |x| < 3\Delta x \wedge |y| < 3\Delta y \\ \frac{1}{\sqrt{x^2+y^2}} & \text{otherwise} \end{cases}$$

where, $\Delta x \times \Delta y$ denotes the pixel dimensions. One of the components is considered the background structure as it would appear as tissue in clinical images. The other serves as noise that would originate from an image acquisition process. Both components are weighted with constant factors and merged to background images b_i, $i = 1, ..., N$ such that signal-present and signal-absent images $I_i^{(p)}$ and $I_i^{(a)}$, respectively, can be created as;

$$I_i^{(p)} = s_i + b_i, \qquad I_i^{(a)} = b_i.$$

All images are of size 200×200 pixels. We simulate 5 groups of images, each group comprising 240 signal-present images and 240 signal-absent images. Signal intensity of the signal-present images is controlled to span 5 different levels such that images in group 1 are of the highest contrast, while images in group 5 are of the lowest contrast. As shown in (Fig. 1), signals in group 5 are hardly visible to the human eye. All signal-absent images contain only background information and are statistically similar. In order to prevent the neural network from learning only mean or variance features, we normalize each image in a pre-processing step, so that they all have the same mean and variance. We evenly assign one-sixth of the data set to a testing set, and the rest are further employed as the training and validation pool.

• Classic Model Observer. A frequently studied task for model observers is the detection task which in its basic form involves the detection of a known signal at a known location. Based on the performance of a model observer on a selected data set a certain degree of image quality can be assessed [7]. The channelized Hotelling observer (CHO) is a group of observers that produce its decision metric from a feature description of the image data. The well-known Laguerre-Gauss channels [8] produce a set of rotationally-invariant features. Thus well applicable for Gaussian signal shapes [9], we compute scores for all tested contrast levels

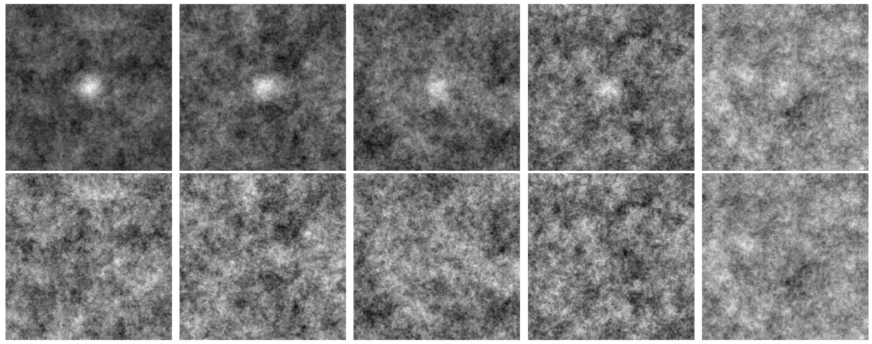

Fig. 1. From left to right: Examples images from highest to lowest contrast (groups 1-5) with signal present (first row), and signal absent (second row). The images shown above are aimed for better visualization, while the actual data are normalized as described in the text.

using such a classic model observer, namely, a CHO with 10 Laguerre-Gauss channels. The width parameter that controls the area which is effectively taken into account for the feature computation is selected as $4s$, where s is the size parameter of the signal.

The detectability index SNR_λ is associated with a model observer via its test statistic λ [7]. It is a figure of merit of how separable the two groups are w.r.t. λ and is defined as:

$$\mathrm{SNR}_\lambda = \frac{\mathrm{E}(\lambda|\text{signal}) - \mathrm{E}(\lambda|\text{no signal})}{\sqrt{0.5\,(\mathrm{Var}(\lambda|\text{signal}) + \mathrm{Var}(\lambda|\text{no signal}))}}.$$

We choose this model observer and its figure of merit for comparison with the proposed approach using a neural network.

• Neural Network Architecture and Training Strategy. In this paper, we propose a convolutional neural network with 2 convolutional layers. The first convolutional layer has 6 kernels of size 11×11, followed by a rectified linear unit (ReLu) activation, batch normalization and an 8×8 max-pooling with 2×2 stride. The second convolutional layer has 3 kernels of size 5×5, followed by ReLu activation, batch normalization and a 2×2 max-pooling with 2×2 stride. The output is then passed to two fully connected layers to reduce the dimension to 2 corresponding to the number of classes. A softmax layer is employed as the final classification layer, with a cross entropy loss function (see Fig. 2). This neural network serves as a model observer and predicts a score between 0 to 1.0 as to how likely the image contains a signal. In validation a classification is achieved based on a probability threshold of 0.5.

In order to train the network, we employ the Adam optimizer, with the learning rate set to $1e^{-5}$. The neural network is trained to a maximum of 300 epochs, and after 100 epochs, an early stopping criterion is called if the validation loss has not improved in the preceding 20 epochs. To evaluate the robustness of the neural network and to detect signals in unseen lower contrast images, we always leave out group 5 from the training and validation data sets. A 5-fold cross validation is carried out and consequently, five different sets of weights are trained for our network. We employ these as five individual models in the test phase, using samples from group 5 in the test set.

Besides the probability output from the network itself as a measure of detection confidence, these outputs were also considered as decision scores. By this analogy, SNR_λ can be evaluated for the neural network as well, which allows to qualitatively compare both methods for different contrast levels.

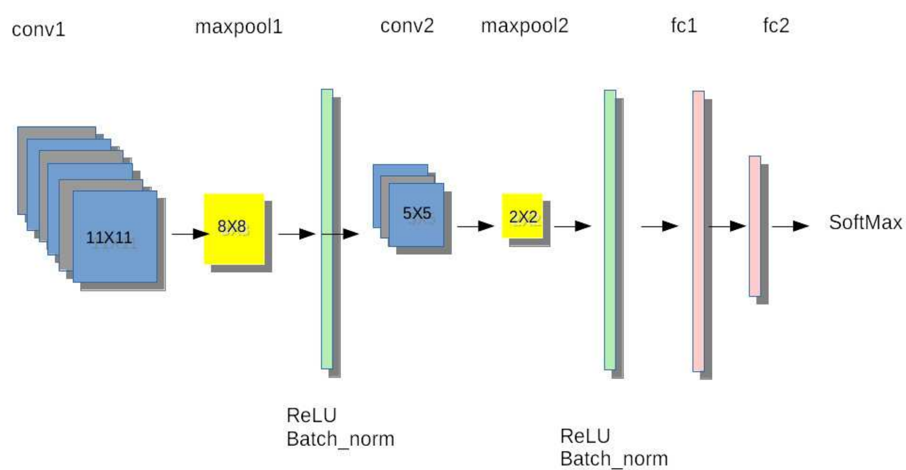

Fig. 2. Illustration of the neural network structure

3 Results

We test both classic and neural network based model observers on the test data samples which are separated randomly before the training takes place. As we do 5-fold cross validation in training, we further combine the results from 5 different weight sets by averaging the softmax scores sample-wise. In (Tab. 1) we show its classification accuracy on the test data with adjusted threshold. The accuracy and sensitivity show very good performance in the first 3 groups and decrease as contrast levels go lower. (Fig. 3) shows typical training curves in our experiments, where in this case early stopping criterion was called at epoch 140.

(Tab. 2) shows SNR_λ values for both classic model observer and our proposed neural network. It can be seen that both models have high SNR_λ on higher contrast levels, and lower SNR_λ on contrast level 4 and 5. Our proposed neural network's ability to distinguish signal-present images from signal-absent images shows a comparable decreasing trend to the classic model observer. Furthermore,

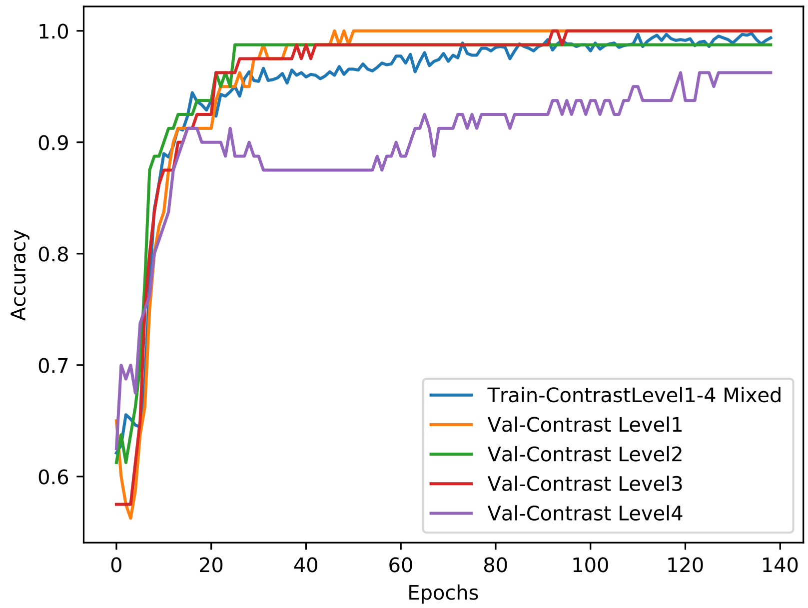

Fig. 3. Train and validation accuracy along training epochs

Table 1. Performance of the combined neural network model on test data

Contrast Level	Accuracy	Accuracy(CMO)	Sensitivity	Specificity
1 (highest)	1.0	1.0	1.0	1.0
2	1.0	1.0	1.0	1.0
3	1.0	1.0	1.0	1.0
4	0.975	1.0	0.975	0.975
5 (lowest)	0.9	0.95	0.9	0.9

The result of our novel model observer suggests the presence of a significant difference between level 1-3, and the two lowest contrast levels, than within level 1-3, which can be seen as alternative reference to human performance (see Fig. 1).

4 Discussion

In this paper, we propose a neural network as a model observer. The network is trained with simulated images of four different contrast levels and tested on images with similar contrast, as well as an unseen lower contrast. The results highlight the model's potential for predicting human performance qualitatively as its classification performance declines with the image contrast, while still being able to detect signals in some of the very low-contrast images. Furthermore, the

Table 2. SNR_λ on test results from both model observers

Contrast Level	Classic Model Observer	Proposed Neural Network
1 (highest)	10.7	10.5
2	8.0	10.3
3	6.5	9.4
4	5.0	6.1
5 (lowest)	3.4	2.0

performance is comparable to a classic model observer. Compared with other techniques, our model requires little knowledge about the signal and demands only reasonable training time. We believe our proposed model observer can also be applicable to clinical data, to assess image quality, among other tasks. For instance, training a network on simulated high- and low-contrast images and then evaluating it on real mammography images, to distinguish between high and low contrast samples, may provide a means for automatically detecting lesions and microcalcifications, which will be a topic of future work.

References

1. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. vol. 90; 1993. p. 9758–9765.
2. Damien R, Alexandre B, Julien O, FranÃ§ois B, Francis V. Objective assessment of low contrast detectability in computed tomography with Channelized Hotelling Observer. Physica Medica. 2016 January;32:76–83.
3. Kalayeh MM, Marin T, Brankov JG. Generalization Evaluation of Machine Learning Numerical Observers for Image Quality Assessment. IEEE Transactions on Nuclear Science. 2013 June;60(3):1609–1618.
4. Alnowami MR, Mills G, Awis M, Awis M, Elangovanr P, Patel M, et al. A deep learning model observer for use in alterative forced choice virtual clinical trials. 2018 November;.
5. Maier A, Hofmann H, Berger M, Fischer P, Schwemmer C, Wu H, et al. CONRAD – A Software Framework for Cone-beam Imaging in Radiology. Medical Physics. 2013 November;40(11):111914–1–8.
6. Wunderlich A. IQmodelo; 2016. [Online; accessed 28-October-2018]. https://github.com/DIDSR/IQmodelo.
7. He X, Park S. Model Observers in Medical Imaging Research. Theranostics. 2013;3(10):774–786.
8. Gallas BD, Barrett HH. Validating the Use of Channels to Estimate the Ideal Linear Observer. Journal of the Optical Society of America A: Optics, Image Science, and Vision. 2003 September;20(9):1725–1738.
9. Schebesch F, Maier A. Towards Optimal Channels for a Detection Channelized Hotelling Observer. In: Hoeschen MSC, editor. 3rd Conference on Image-Guided Interventions & Fokus Neuroradiologie; 2017. p. 12–13.