A study on the impact of statistical weights on lesion detection performance in iterative CT reconstruction

V. Haase, A. Griffith, Z. Guo, M. B. Oktay, K. Hahn, H. Schoendube, K. Stierstorfer, F. Noo

***Abstract*—Iterative CT reconstruction with the penalized least-square model may offer significant gains in terms of image quality at equal dose, and may thereby allow either dose reduction or improved diagnostic. In this work, we are interested in evaluating image quality improvements that result from using statistical weights in this model. Image quality is assessed in terms of lesion detection with unknown location, using the principles of LROC analysis with human observers. Reconstruction without and with statistical weights are compared for two penalties: a quadratic penalty, and an edge-preserving penalty. Interestingly, our study failed showing any major improvements due to the use of weights. Furthermore, it was even observed that performance with weights could be even worse, possibly due to the utilization of weights leading to disturbing discretization errors. Because there are a lot of degrees of freedom in our experimental set-up, it should not be concluded that statistical weights are not useful. However, we can state that improvements are not straightforward and may depend on many aspects including the task and also anatomical location and variability. This observation is valuable from a computational viewpoint since using statistical weights generally leads to long reconstruction times; if weights can be ignored or simplified in some settings, reconstruction times can be largely improved for these settings.**

I. Introduction

Iterative CT reconstruction using advanced statistical models is currently a hot topic of research. For diagnostic CT imaging, the main advantage that is sought from such reconstruction is improved image quality, which may either be used to increase the role of CT in medicine and clinical research or to reduce dose to the patient for conventional CT scans.

Among a number of options, the penalized least-square model with statistical weights [1] is popular. In this model, the statistical weights represent the variance of the measurements. These weights are applied in the data fidelity term to enable accounting for different noise levels across measurements. To regularize the reconstruction, the data fidelity term is further balanced with a penalty term that typically constrains differences between neighboring voxel values.

Designing an efficient iterative algorithm to solve the penalized weighted least-square reconstruction problem turns out to be highly difficult, particularly for convergence within 1 HU from the desired solution. The wide dynamic range of statistical weights is largely responsible for this situation. In this work, we are interested in evaluating the image quality improvement brought by the statistical weights, in comparison with using the same model but without weights. The evaluation is carried out in fanbeam geometry, and task-based assessment of image quality is employed with human observers.

V. Haase, K. Hahn, H. Schoendube and K. Stierstorfer are with Siemens Healthcare GmbH, Forchheim, Germany. A. Griffith, Z. Guo, M. B. Oktay and F. Noo are with the Department of Radiology and Imaging Sciences, University of Utah, Salt Lake City, Utah, USA. The concepts presented in this paper are based on research and are not commercially available. This work was partially supported by Siemens Medical Solutions, USA.

II. Evaluation methodology

A. Problem formulation

Let $\underline{x}$ be the vector of image pixel values to be reconstructed and let $\underline{b}$ be the vector of CT measurements. The reconstruction is defined as the minimizer of

$$\Phi(\underline{x}, \underline{b}) = \Phi_1(\underline{x}, \underline{b}) + \beta\, \Phi_2(\underline{x}) \tag{1}$$

with positivity constraint on the entries of $\underline{x}$. Here, $\Phi_1(\underline{x}, \underline{b})$ is the data fidelity term, and $\Phi_2(\underline{x})$ is the penalty (regularization) term. The hyper-parameter $\beta > 0$ controls the balance between the two terms. The data fidelity term is

$$\Phi_1(\underline{x}, \underline{b}) = \left\| \mathbf{W}^{-1/2}(\mathbf{A}\underline{x} - \underline{b}) \right\|^2 \tag{2}$$

where $\mathbf{A}$ is the forward projector, and $\mathbf{W}$ is a diagonal matrix with entries equal to the exact variance of the measurements. When we use the statistical weights, we normalize them so that their mean value for measurements associated with the central ray is equal to one. This normalization enables a fair comparison between imaging without and with statistical weights. The regularization term penalizes differences between neighboring pixel values through a potential function. We consider both a quadratic potential and the edge-preserving FAIR potential.

The desired reconstruction was robustly computed using the iterative coordinate descent (ICD) method [2]. The last iteration was chosen as that for which all pixel values changed by less than $0.0001/\mathrm{cm}$. All iterations were initiated with a zero image to prevent any bias from an initial reconstruction.

B. Data simulation and image reconstruction

Reconstructions were performed from computer-simulated fanbeam data of the FORBILD head phantom. A realistic fanbeam CT geometry was used; see Table 1. Several noisy data sets were created to simulate repeated scans. The noise was based on Poisson statistic using a realistic body-size bowtie filter. The beam was monochromatic. A low energy of

focal spot size	0.12 cm × 0.09 cm
anode angle	7°
source trajectory radius	57 cm
source to detector distance	104 cm
number of detector elements	672
angular detector width	0.001231 radians
detector row height	0.128 cm
number of projections per turn	1160
incoming number of photons per ray:	
for quadratic penalty	180,000
for edge-preserving penalty	90,000

TABLE I
DATA SIMULATION PARAMETERS.

40kV was used to accentuate variations in noise level across the CT measurements, and thus the effect of statistical weights.

The image was defined on a grid of square pixels of 0.0375 cm sidelength. Matrix A was based on the distance-driven method [3]. When present, the lesion was always within the low-contrast brain-tissue region of the phantom, with no overlap with the bones. The lesion was a 7 mm diameter disk with random contrast varying between 20 and 30 HU.

Note that the incoming number of photons is different from the evaluation with quadratic penalty to the evaluation with edge-preserving penalty. In this work, we did not attempt to evaluate benefits offered by the edge-preserving penalty over the quadratic penalty. The reader interested in this question will find related results in [4]. Our focus here is assessment of the role played by the statistical weigths under two settings: (i) reconstruction with quadratic penalty and moderate data noise, (ii) reconstruction with edge-preserving penalty and high noise.

C. Task-based assessment of image quality

Image quality was assessed in terms of lesion detectability with unknown location, using LROC analysis. The area under the LROC curve, called AUC, was used as figure-of-merit. AUC was directly evaluated using an alternative forced choice experiment that involved two human observers (readers) for reconstructions with quadratic penalty, and three human observers for reconstructions with edge-preserving penalty.

Each observer participated in two sessions, within each of which the observer evaluated images reconstructed without and with statistical weights. Each session included 400 cases (40 training plus 160 testing cases for each method). Random numbers were used to avoid any bias from case ordering presentation. All observers read the cases in the same dimmed room (10 lux), on a medical-grate monitor calibrated according to the ACR Technical Standard for Electronic Practice.

III. RESULTS

Figure 1 displays the main results obtained from the LROC studies for reconstruction with quadratic penalty and moderate data noise. Figure 2 displays the main results obtained from the LROC studies for reconstruction with edge-preserving penalty and high data noise. In these two figures, the left column shows the session-averaged AUC result obtained for each reader, and the right column shows the main and the difference in reader-and-session-averaged performance. The error bars for the plot on the left column correspond to individual 95% confidence intervals, whereas the error bars for the plot on the right column correspond to joint 95% confidence intervals (based on Bonferoni inequality and fixed reader effects). For reconstruction with quadratic penalty and moderate data noise, we observe that the reader-and-session-averaged performance is slightly worse for reconstruction with statistical weights than for reconstruction without weights. For reconstruction with edge-preserving penalty and high data noise, we observe that the reader-and-session-averaged performance is, with 95% confidence, essentially the same for both reconstruction without and with statistical weights. Note that the size of the error bars in the bottom plots indicate that the statistical accuracy for these observations is fairly strong. More details on the methodology and image appearance can be found in a related conference record [5].

IV. DISCUSSION AND CONCLUSION

In this work, we reported results of two LROC studies with human observers. These studies aimed at evaluating improvements resulting from the use of statistical weigths in penalized least-square CT reconstruction. Interestingly, the study failed showing any major improvements due to the use of such weights. Furthermore, it was observed that performance with weights could be even worse, possibly due to the utilization of weights leading to disturbing discretization errors. Because there are a lot of degrees of freedom in our experimental set-up, it should not be concluded that statistical weights are not useful. However, we can state that improvements are not straightforward and may depend on many aspects including the task and also anatomical location and variability. Our results are valuable from a computational viewpoint since using statistical weights generally leads to long reconstruction times; if weights can be ignored or simplified in some settings, reconstruction times can be largely improved for these settings.

REFERENCES

[1] Thibault, J.-B., Sauer, K. D., Bouman, C. A., and Hsieh, J., "A three-dimensional statistical approach to improved image quality for multislice helical ct," *Medical physics* **34**(11), 4526–4544 (2007).

[2] Bouman, C. A. and Sauer, K., "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Transactions on image processing* **5**(3), 480–492 (1996).

[3] De Man, B. and Basu, S., "Distance-driven projection and backprojection in three dimensions," *Physics in Medicine and Biology* **49**, 2463 – 2474 (2004).

[4] Noo, F., Hahn, K., and Guo, Z., "Importance of the grayscale in early assessment of image quality gains with iterative CT reconstruction,"

[5] Haase, V., Griffith, A., Guo, Z., Oktay, M. B., Hahn, K., Schoendube, H., Stierstorfer, K., and Noo, F., "Penalized least-square CT reconstruction without and with statistical weights: effect on lesion detection performance with human observers," *Proceedings of the 14th International Meeting on Fully Three-dimensional Image Reconstruction in Radiology and Nuclear Medicine* , 831–835 (2017).

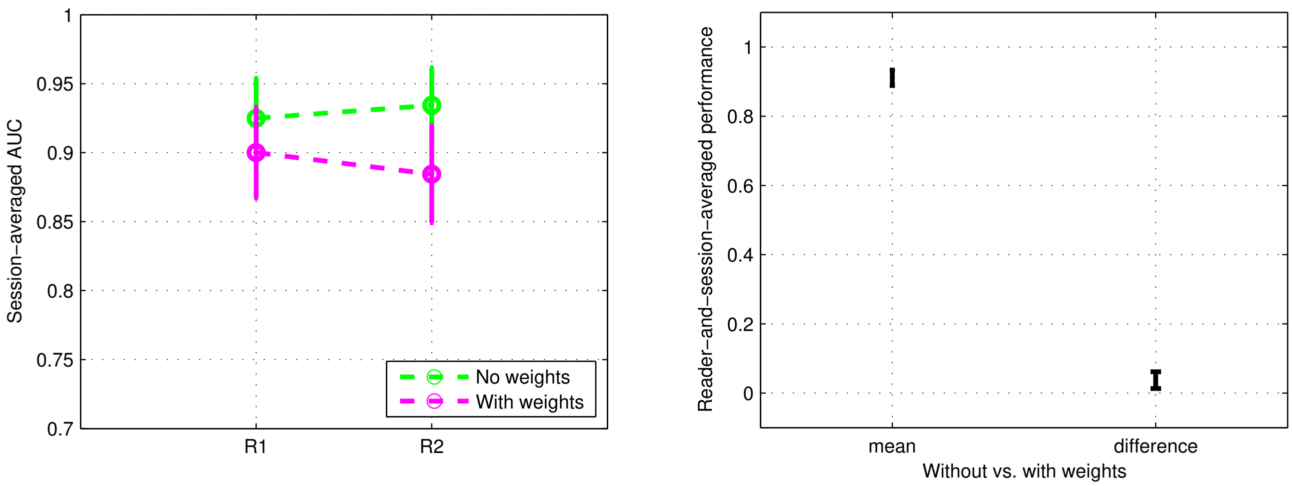

Fig. 1. Results from the LROC studies for reconstruction with quadratic penalty and moderate data noise. (left) AUC value for the two observers, as obtained after taking the mean over sessions; the error bars correspond to individual 95% confidence intervals. (right) Mean and difference in reader-and-session-averaged AUC between reconstruction without weights and reconstruction with weights; the error bars correspond to joint 95% (based on Bonferroni inequality and fixed-reader effects).

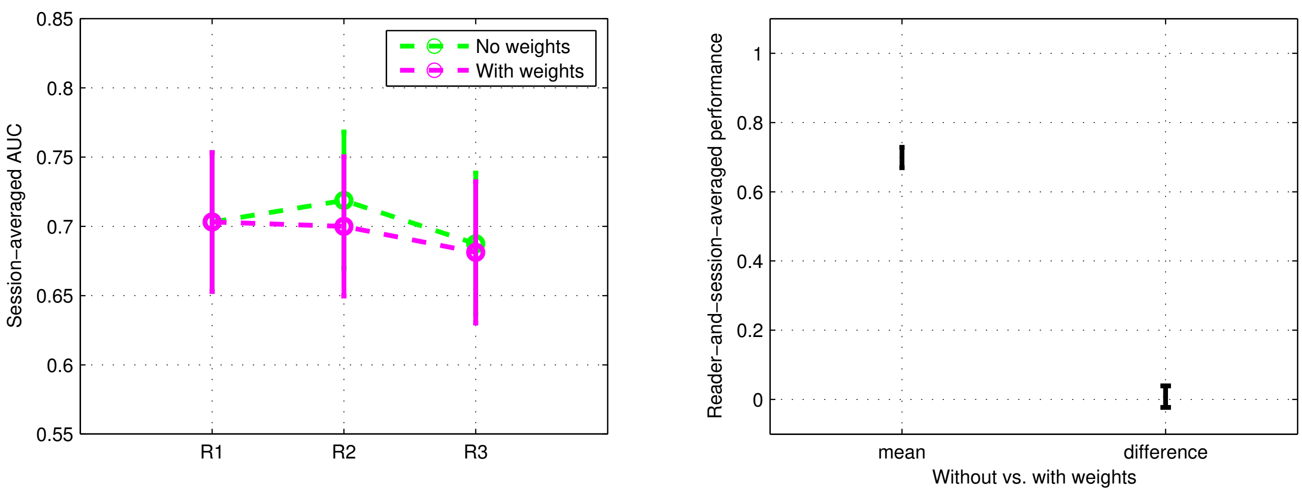

Fig. 2. Results from the LROC studies for reconstruction with edge-preserving penalty and high data noise. (left) AUC value for the three observers, as obtained after taking the mean over sessions; the error bars correspond to individual 95% confidence intervals. (right) Mean and difference in reader-and-session-averaged AUC between reconstruction without weights and reconstruction with weights; the error bars correspond to joint 95% (based on Bonferroni inequality and fixed-reader effects).