Predicting Analyst Performance at Acquisition Quality Assessment from Visual Scanning of Brain Image Montages

Steven R. Gomez*

Department of Computer Science
Brown University

Abstract

We present an analysis of MRI quality inspection behaviors from real neuroscience research workflows. We find two major strategies among analysts for inspecting MRI sets and that performance of the inspection (i.e., the accuracy of classifying an acquisition as "good" or "bad") can be measured and modeled with parameters including *total slices inspected in the series* and *average time per slice*. A quantitative evaluation showed that model predictions are within $X\%$ error. Finally, we discuss how an analysis and model of inspection performance helps guide the design and development of new imaging QA tools for neuroscientists.

1 Introduction

Assessing the quality of an MRI acquisition is often the first step neuroscience workflows that utilize brain imaging. The reason for this is that scans with errors may complicate or invalidate insights gained from that data or derived data, like white matter tractography. Errors in scans occur for a variety of reasons, e.g., if the scanned subject moves during the acquisition, so this quality check is justified on a regular basis.

In this paper, we present an analysis of an imaging quality assessment task (henceforth called *image inspection*) that revealed two major strategies among experts. We also describe and quantitatively evaluate a predictive model of human performance (accuracy) at image inspection based on user behaviors observed during the tasks, including the number of slices inspected per set and the duration of those inspections.

Modeling image inspection is important in the medical domain for 1) **understanding how humans do visual search with medical images**, and 2) **characterizing risk/uncertainty** in an imaging analysis.

Contributions

1. An analysis of image inspection strategies observed in real neuroscience workflows. We identify two major strategies – *fast searching* and *slow searching*.

2. A predictive model of assessment accuracy given an analyst's inspection strategy, and a quantitative evaluation of the model using imaging *n* experts. We also collected anecdotal feedback from these participants that describes what they look in individual slices and over a series.

2 Related Work

We build off previous work in visual search behaviors in medical imaging, task and workflow analysis, and predictive modeling.

Human perception skills directly affect the utility of medical imaging in diagnostics and image quality assessment [5, 2]. Our work tries to quantify the success perception-interaction strategies for assessing a set of medical images under time constraints.

Workflow evaluation, low-level activity coding [1], and task analysis [4] are tools for identifying bottlenecks or performance breakdowns in human-computer interfaces. In a classic example, Project Ernestine evaluated two workflows for a telephone operator system using GOMS-family cognitive models [3]. It accurately predicted that an interface disrupting the critical path of a workflow can significantly hinder human performance (task completion speed) with that workflow. While this paper does not provide a GOMS analysis of image inspection, it does characterize core strategies for analyzing image mosaics, which could lead to new presentation tools for MRI scans that facilitate quality assessment. Furthermore, we can extract some principles for effectively displaying or navigating through multiple medical images, incorporating general guidelines for multi-view visualization [6].

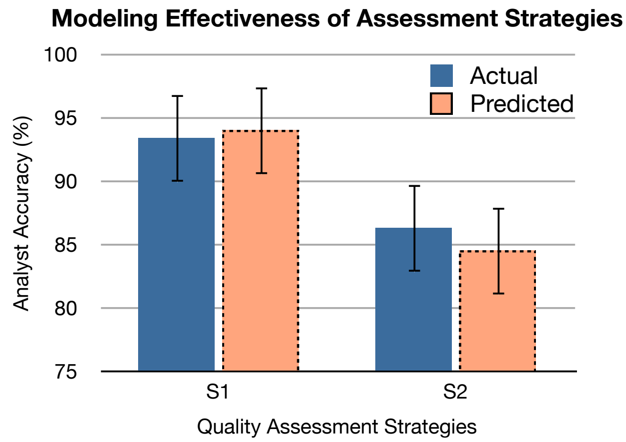

Figure 1: Analyst performance predictions at MRI quality assessment task. We identified two major strategies for completing the assessment task and are able to predict accuracy performance based on strategies used.

3 Methods

4 Results

We found two major strategies for image inspection – *fast searching* and *slow searching*. In fast searching, analysts were trying to look at many slices and use quick visual scanning to find any problems with slices that would quickly disqualify the data set. In slow searching, analysts took time to chose specific slices carefully and spent significant time with these selected slices. This was confirmed in anecdotal feedback obtained by questionnaire and by plotting total slices inspected against the mean time per slice, as shown in Figure 2.

The accuracy model was fit to inspection parameters [including total slices inspected, time per slice, and possibly some eye-tracking

*e-mail: steveg@cs.brown.edu

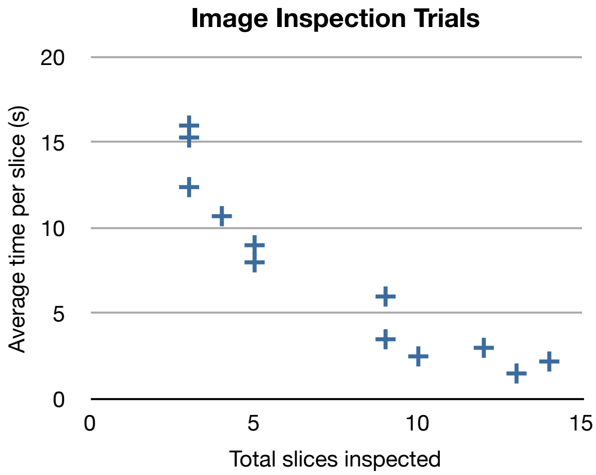

Figure 2: Clusters correspond to *fast* and *slow* inspection strategies.

measures, e.g., total saccades, average fixation time] and predictions on test data were within [SOME ACCURACY].

5 Conclusion

References

[1] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, 2005.

[2] M. S. Chesters. Human visual perception and roc methodology in medical imaging. *Physics in Medicine and Biology*, 37(7):1433, 1992.

[3] W. D. Gray, B. E. John, and M. E. Atwood. Project Ernestine: validating a GOMS analysis for predicting and explaining real-world task performance. *Hum.-Comput. Interact.*, 8(3):237–309, 1993.

[4] B. Kirwan and L. K. Ainsworth, editors. *A Guide to Task Analysis: The Task Analysis Working Group*. Taylor & Francis, 1992.

[5] E. A. Krupinski. The importance of perception research in medical imaging. *Radiation Med.*, 18(6):329–334, 2000.

[6] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the working conference on Advanced visual interfaces*, AVI ’00, pages 110–119, New York, NY, USA, 2000. ACM.