Image Quality in Optoacoustic Imaging

A Data Story for Scientific Visualization

Melle van der Brugge

May 18, 2025

Target Audience: Researchers in biomedical imaging and machine learning
Suggested Place of Publication: Nature Biomedical Engineering – Visual Abstract / Elsevier Graphical Abstract or as demo dashboard on GitHub Pages

The Data Story

Background

Photoacoustic imaging (PAI) is a hybrid modality that combines optical contrast with ultrasound resolution by using the photoacoustic effect: tissue is illuminated with pulsed light, and the resulting acoustic waves are measured. These waves are used to reconstruct biomedical images with rich structural and functional information.

This project evaluates the quality of reconstructed PAI images using various configurations (PA2–PA7) and compares them to the ground truth (PA1). The goal is to assess how different image quality metrics behave across configurations and datasets, and to highlight how visualizations can aid interpretation.

Datasets

- **KneeSlice**: In vivo imaging of a human knee, showing anatomical structures.
- **Phantoms**:
  - *DoranssoPhantom*: Simulated phantom with known optical targets.
  - *BreastPhantom*: Mimics breast tissue for clinical benchmark testing.
- **Transducers**:
  - *7MHz*: Low-frequency sensor for deeper tissue penetration.
  - *9MHz*: High-frequency sensor for better spatial resolution.
- **SmallAnimal**:
  - *Tissue*: Images from normal tissue.
  - *Tumor*: Mouse tumor data for oncological imaging study.

Message

Not all image quality metrics behave consistently across datasets or conditions. By visualizing differences in performance, correlation, and prediction quality, this dashboard encourages a more nuanced selection of metrics for PAI. In particular, it shows that higher configurations (PA6, PA7) yield superior results, but some metrics capture that better than others.

Metrics Used

- **Structural Similarity Metrics**: SSIM, MSSSIM, IWSSIM
- **Error-Based Metrics**: PSNR, VIF, GMSD, MSGMSD
- **No-Reference Metrics**: BRISQUE, CLIP-IQA
- **Other Metrics**: FSIM, HAARPSI, UQI, S3IM, TV

Key Visualizations

- **Line plots**: Metric comparisons across configurations (PA2–PA7)
- **Image Grids**: Visual reconstruction differences (PA1–PA7) across datasets
- **Heatmap**: Correlation matrix between metrics
- **Radar Chart**: Summary of all metric performances per configuration

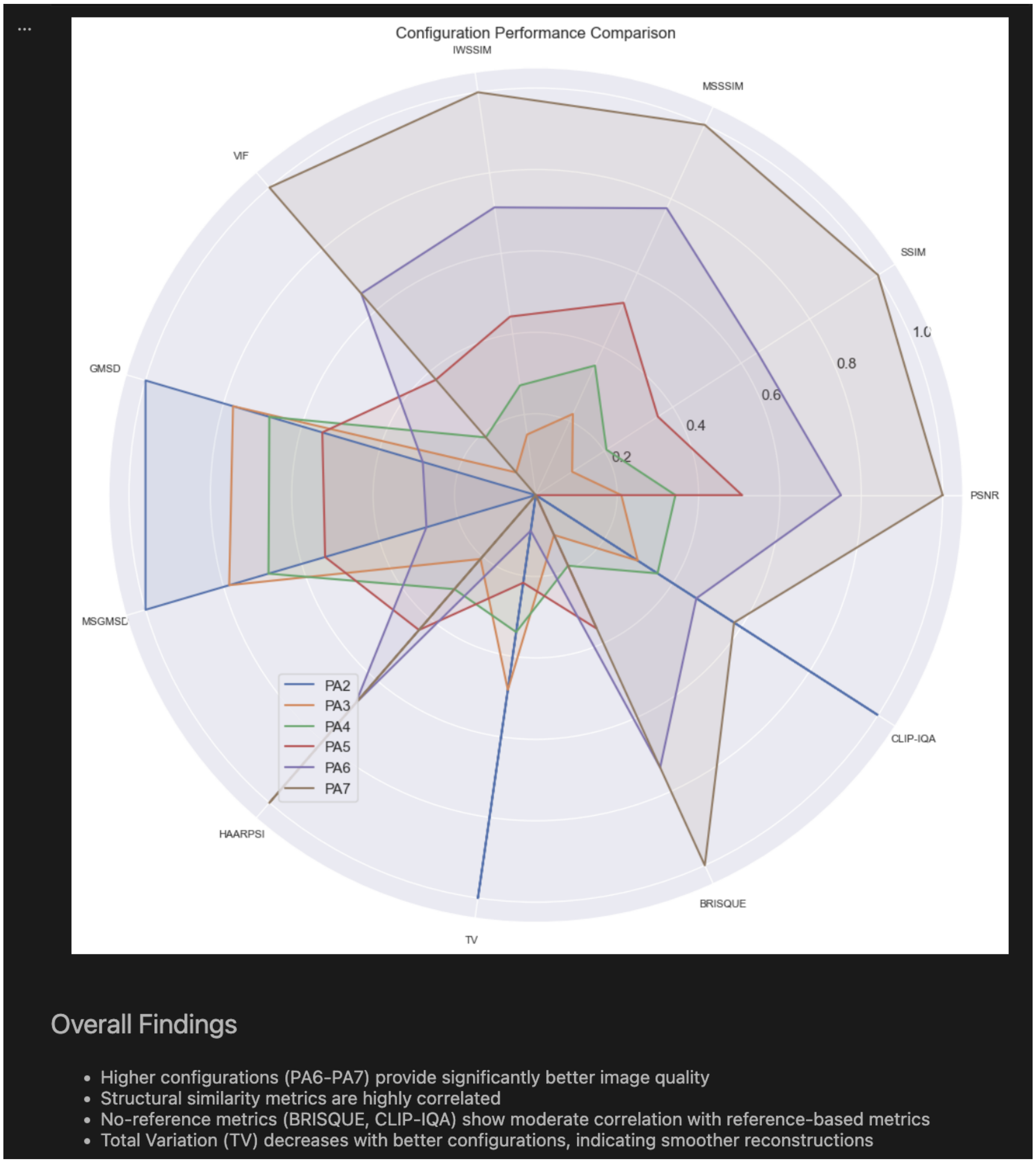


Figure 1: Illustrative layout of the interactive dashboard showing image examples, metric trends, and a performance radar chart.

References

- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612.
- Sheikh, H. R., & Bovik, A. C. (2006). Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2), 430–444.
- Zhang, L., Zhang, L., Mou, X., & Zhang, D. (2011). FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8), 2378–2386.
- Breger, A., Karner, C., Selby, I., et al. (2024). A Study on the Adequacy of Common IQA Measures for Medical Images. *MICAD 2024, Manchester, UK*. Retrieved from `https://arxiv.org/abs/2405.19224`
- Davoudi, N., Deán-Ben, X. L., & Razansky, D. (2019). Deep learning optoacoustic tomography with sparse data. *Nature Machine Intelligence*, 1, 453–460.

Code and Data

The full codebase and data can be accessed via the GitHub repository:
`https://github.com/MellevdB/optoacoustic-dashboard.git`

Earlier Version Data Story (Concept)

Name: Melle van der Brugge

Group: B

Date: May 18, 2025

Description of the data:

The data comes from my thesis research on optoacoustic imaging. It includes several biomedical imaging datasets collected using different sensor geometries and sampling sparsities. Each image in the dataset is reconstructed from raw optoacoustic signals. I apply both full-reference and no-reference image quality metrics to evaluate how image quality changes with sensor configurations and levels of sparsity.

Message that you want to tell with your data:

Not all image quality metrics are equally reliable in assessing quality changes across different imaging conditions. My data story shows how metrics like SSIM, VIF, and BRISQUE vary in sensitivity and usefulness depending on dataset and reconstruction sparsity. Understanding which metrics are robust helps guide the development of better models for medical imaging quality assessment.

Additionally, I trained a model to predict image quality scores based on these metrics—its predictions are visualized with a scatter plot.

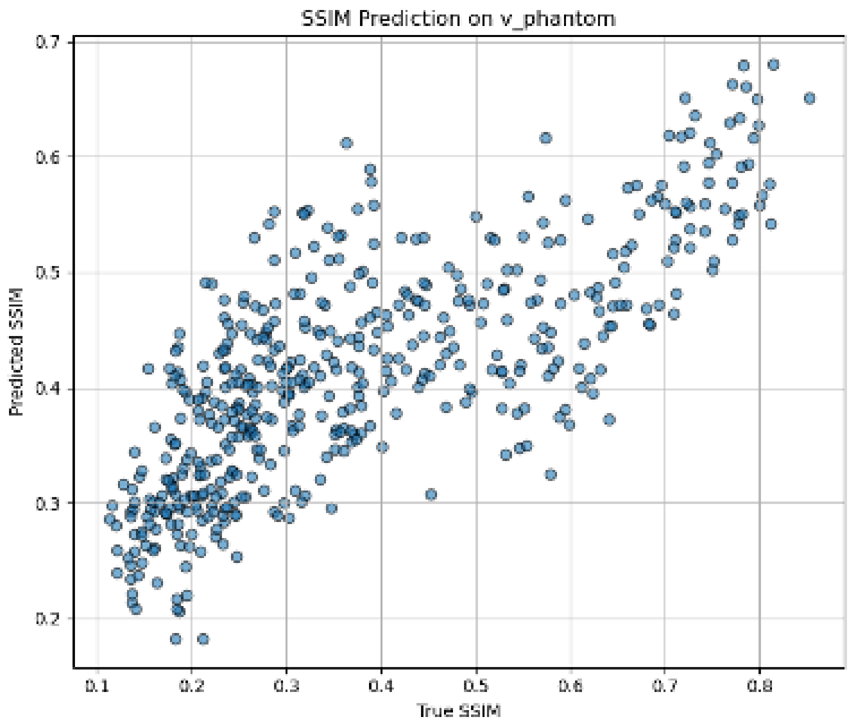
Relevant metrics and visual representations:

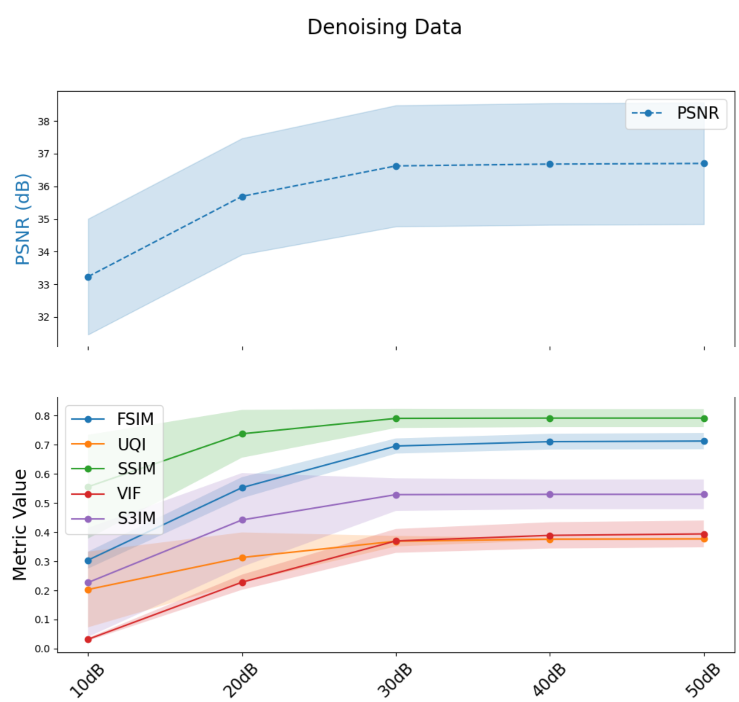
Metric or quantity shown	**Visual representation**
Average image quality vs. sparsity level	Line plot showing how quality drops with increased sparsity
Image quality variation per dataset	Bar plot comparing quality across datasets (e.g. SCD vs. SWFD)
Prediction error of trained model	Scatter plot of predicted vs. true quality scores

Type of data visualization(s): (e.g. donut plot, bubble plot, mosaic plot, bar graph, map, word cloud, interactive)

Line plots, bar graphs, box plots, scatter plots, and optional heatmaps.

Drawing: (see inserted sketches below)





Feedback and Response

Received Feedback:

- It is really helpful that you described how all the images are made. So readers would know how the image is connected to the original data, and it would be possible to reproduce them. Also, the plots are quite beautiful.
- Very clear what the data is about and the visualization is nice.
- It would be nice if there were some explanations to these words. People outside the field might not know the meanings of SSIM, VIF, BRISQUE, etc. Maybe consider at least putting the full form of the abbreviation.
- I would like to hear more about what the quality score is exactly and how it is determined.

Response to Feedback:

I've now added short explanations and full forms of key image quality metrics such as SSIM (Structural Similarity Index), VIF (Visual Information Fidelity), and BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) to make the story more accessible to non-experts. I also included a brief clarification on how the quality scores are computed—these are calculated using both full-reference and no-reference IQA metrics, which compare reconstructed images to a reference or estimate perceptual quality directly. This should make the visualization easier to interpret for a wider audience.