Assessing Perceptual Noise Level Defined by Human Calibration and Image Rulers

Lisa Lei
kllei@stanford.edu

Introduction

Motivation

- Capture diagnosis related perceptual quality
- SNR of MRIs increases with the number of repetitively acquired data: find the optimal point to stop a scan
- Reconstruction evaluation and training loss

Solution

- Train a CNN to learn how radiologists perceive noise in medical images
- Statistical quantity to perceptual score by human calibration
- Adjustable pass/fail thresholds defined by image rulers

Dataset

- Raw data from the Stanford Lucile Packard Children's Hospital
- Inject white Gaussian noise to generate 4 noisier versions per slice → sum-of-squares reconstruction
- 1250 (slices) x5 (versions) images for training, 91x2 for test
- Train set label: select a minimal acceptable version (Fig. 1)
- Test set label: match to a image ruler (Fig. 2)

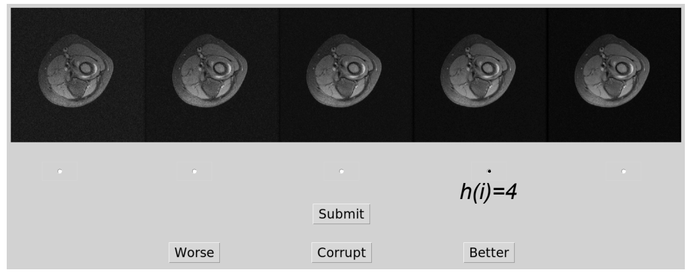

Figure 1. Graphic interface for training set labeling.

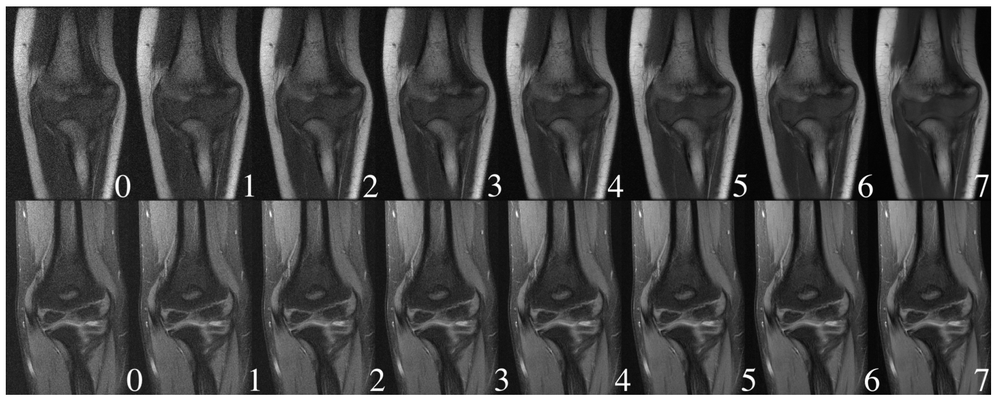

Figure 2.Two image rulers: top one for non-FS scans and bottom one for FS scans.

Models

Baseline

- SVM with polynomial kernel
- Binary classification on the default threshold

CNN model

- 4 conv with average pooling + FC
- Calibration $\mu = \frac{1}{n}\sum_{i=1}^{n} y_i^{h(i)}$

$$y_i^v = IEDD(x_i^v),\ \hat{y}_i^v = y_i^v + \mu - y_i^{h(i)}$$

- Objective: $\min_{\theta_D} \sqrt{\sum_{i}^{n} = 1 \frac{(D(x_i;\theta_D) - y_i)^2}{n}}$
- Ruler defined threshold: pass if $D(x;\theta_D) < \frac{D(x_{ri};\theta_D) + D(x_{rj};\theta_D)}{2}$

Discussion and Future Work

We propose a framework for noise-related perceptual quality assessment and examine one of its immediate application. The test method degenerates the problem to binary classification so that our results, while showing effectiveness, do not impress. But we keep our approach flexible and generalizable. *Future work*:

- Expand dataset
- Try adding more types of ruler
- Use for tuning the regularization strength for compressed-sensing reconstruction

Results

Classification accuracies in percentage.

Table I. Comparison of eight methods on the default threshold; effectiveness of using ruler defined thresholds

Threshold type	Chen [3]	Liu [2]	IEDD [1]	SVM
Ruler defined	71.98%	74.73%	79.67%	NA
Single best	61.53%	68.13%	69.78%	88.5%
	NN - SNR	NN - cSNR	NN - IEDD	NN - cIEDD
Ruler defined	80.21%	82.97%	86.26%	**89.01%**
Single best	81.87%	82.97%	80.77%	87.91%

Table II: Robustness on varying thresholds

Threshold	IEDD	NN - IEDD	NN - cIEDD
3 | 4	78.02%	81.32%	88.46%
4 | 5	79.67%	86.26%	89.01%
5 | 6	81.32%	82.42%	85.16%

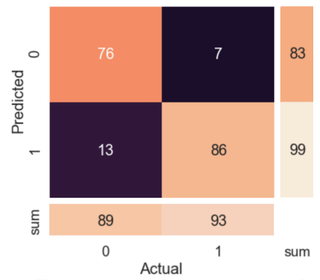

Figure 3. Confusion matrix of proposed method (89.01%).

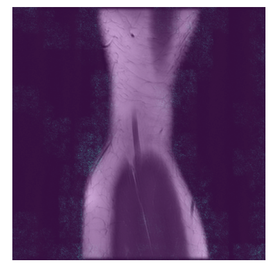

Figure 4. Saliency map on a test sample.

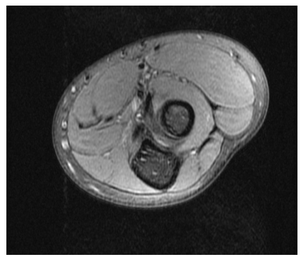

Figure 5. A false positive sample.

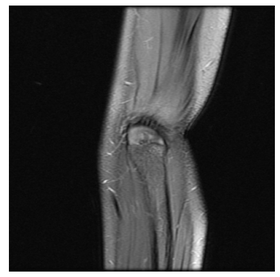

Figure 6. A false negative sample.

References [1] Ponomarenkoa M, Gaponb N, Voroninb V, Egiazariana K. Blind estimation of white Gaussian noise variance in highly textured images. CoRR, vol. abs/1711.10792, 2017.
[2] Liu X, Tanaka M, Okutomi M. Noise level estimation using weak textured patches of a single noisy image. 19th IEEE International Conference on Image Processing, Orlando, FL, 2012, pp. 665-668.
[3] Chen G, Zhu F, Heng P. A.. An Efficient Statistical Method for Image Noise Level Estimation. IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 477-485.