THE IMPORTANCE OF DATA QUALITY AND TRUE NOISE

Paul Kinahan, PhD

Vice Chair of Radiology Research, University of Washington

Chair, AAPM Research Committee

Chair, AAPM Metrology and Standards Subcommittee

- AI-based algorithms in medical imaging can offer advantages, but in general we do not know when they are wrong
- There can also be unintended consequences that are only apparent when the entire imaging chain is reviewed
- We need clinically relevant, clinically feasible, and metrology-based assessments

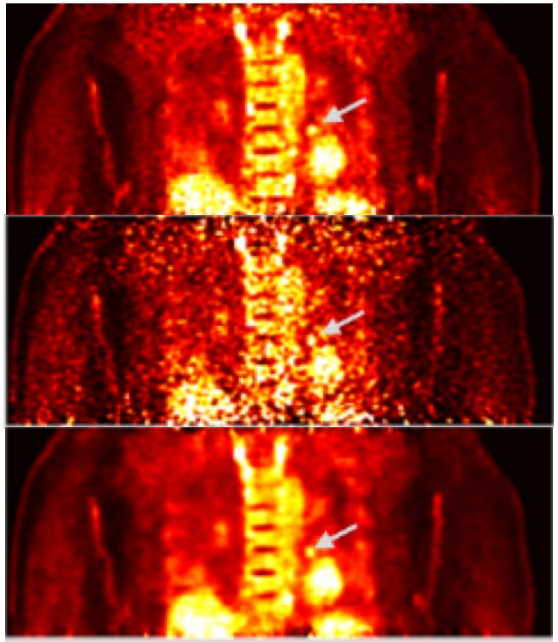
THE NEED FOR METRICS FOR AI-BASED METHODS

- Linear image reconstruction and processing methods allow the use of tools like MTF and NPS for an general assessment of image quality
- For non-linear AI-based methods ROC studies provide quantitative assessment of image quality, but these can be overly task-specific and are often not feasible
- The importance of proper noise texture and noise effects across multiple realizations is typically not appreciated
- Quantitation when AI methods are used is unknown in general
- Metrics should be developed through evidence-based multi-center multi-vendor comparisons and achieve consensus. E.g. though AAPM reports "A methodology for image quality evaluation of advanced CT systems" Wilson et al. *Medical Physics* 2013

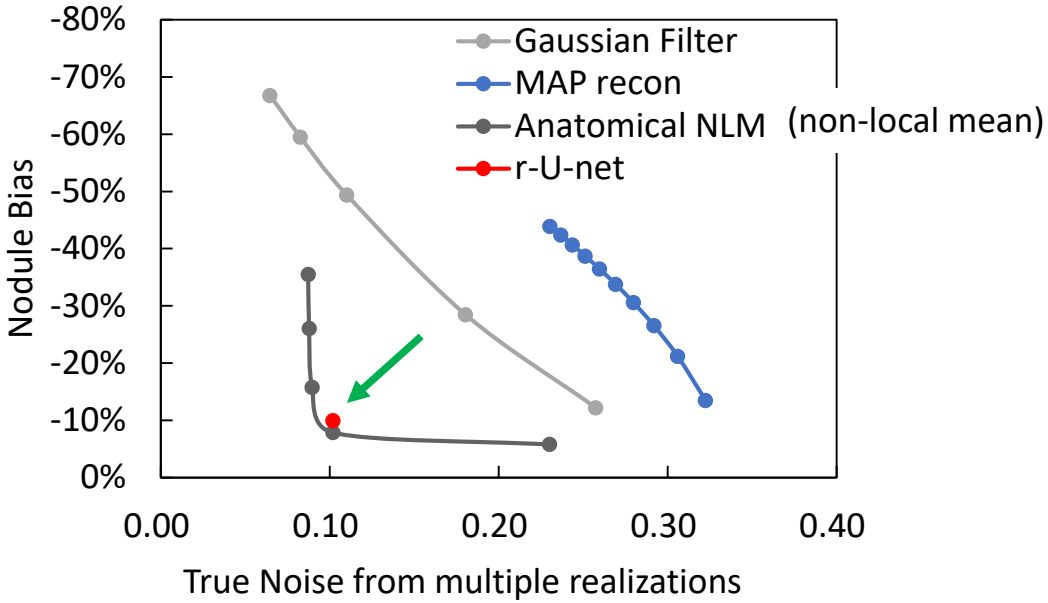
Specific example: AI-based image Denoising

100% Dose PET image

10% Dose



10% Dose with CNN-based noise reduction



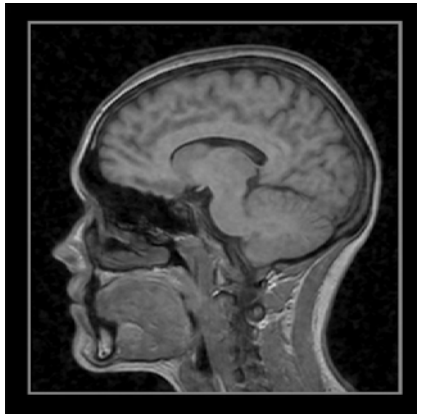
Message: deep learning based denoising can get to the best tradeoff faster

Courtesy Chi Liu, Yale

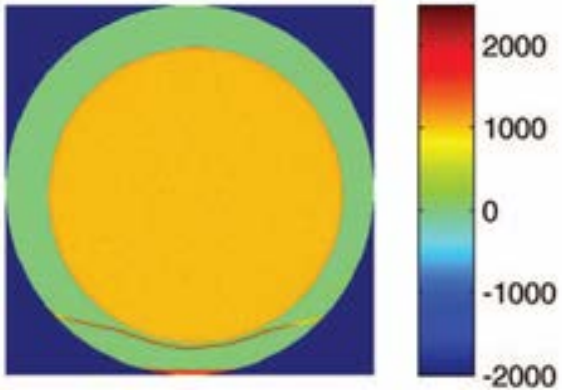
W. Lu, et al. PMB 2019

Example: Perceptions versus reality for AI-based image Denoising

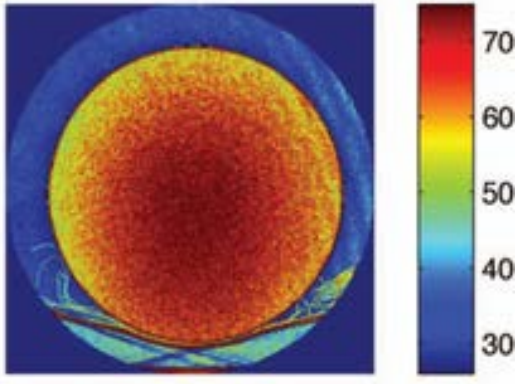
"Nvidia's new tech uses AI to denoise your images"
Digital Arts, July 10, 2018



Vaishnav, *Medical Physics* 2014



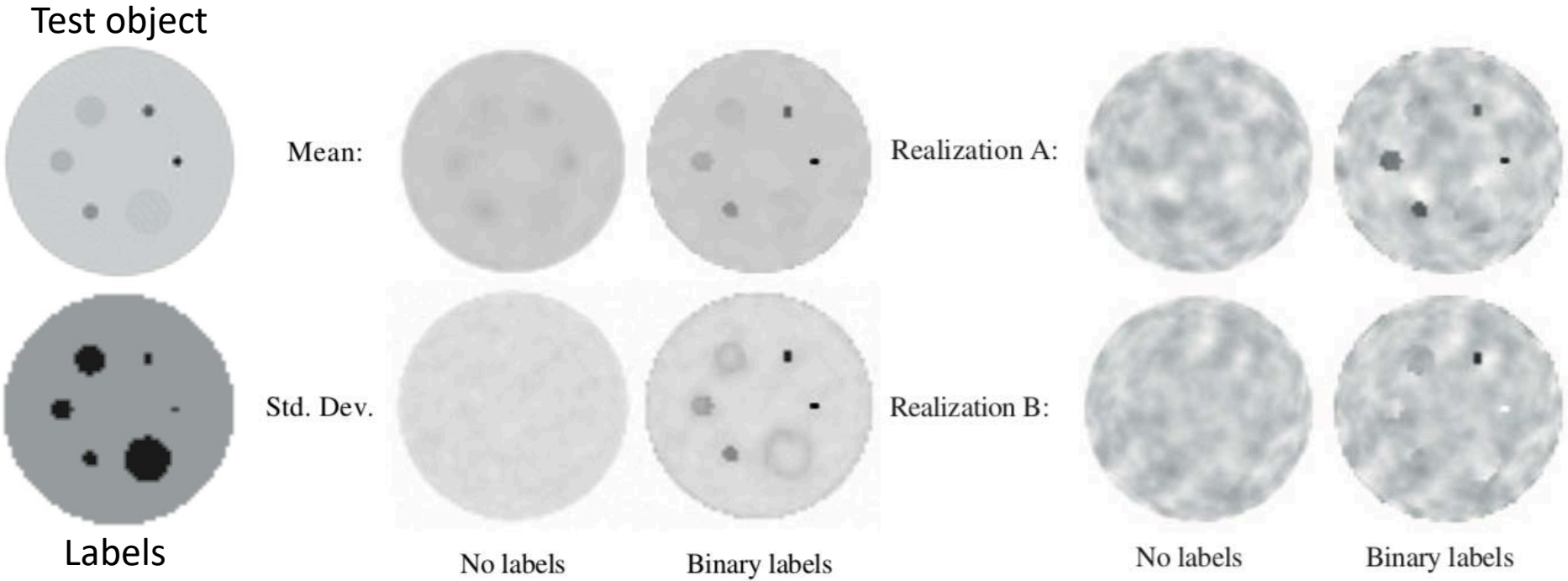
Apparent noise in a single CT image of the ACR phantom



True noise from multiple i.i.d. realizations

Message: Separate images do not tell the full story for noise

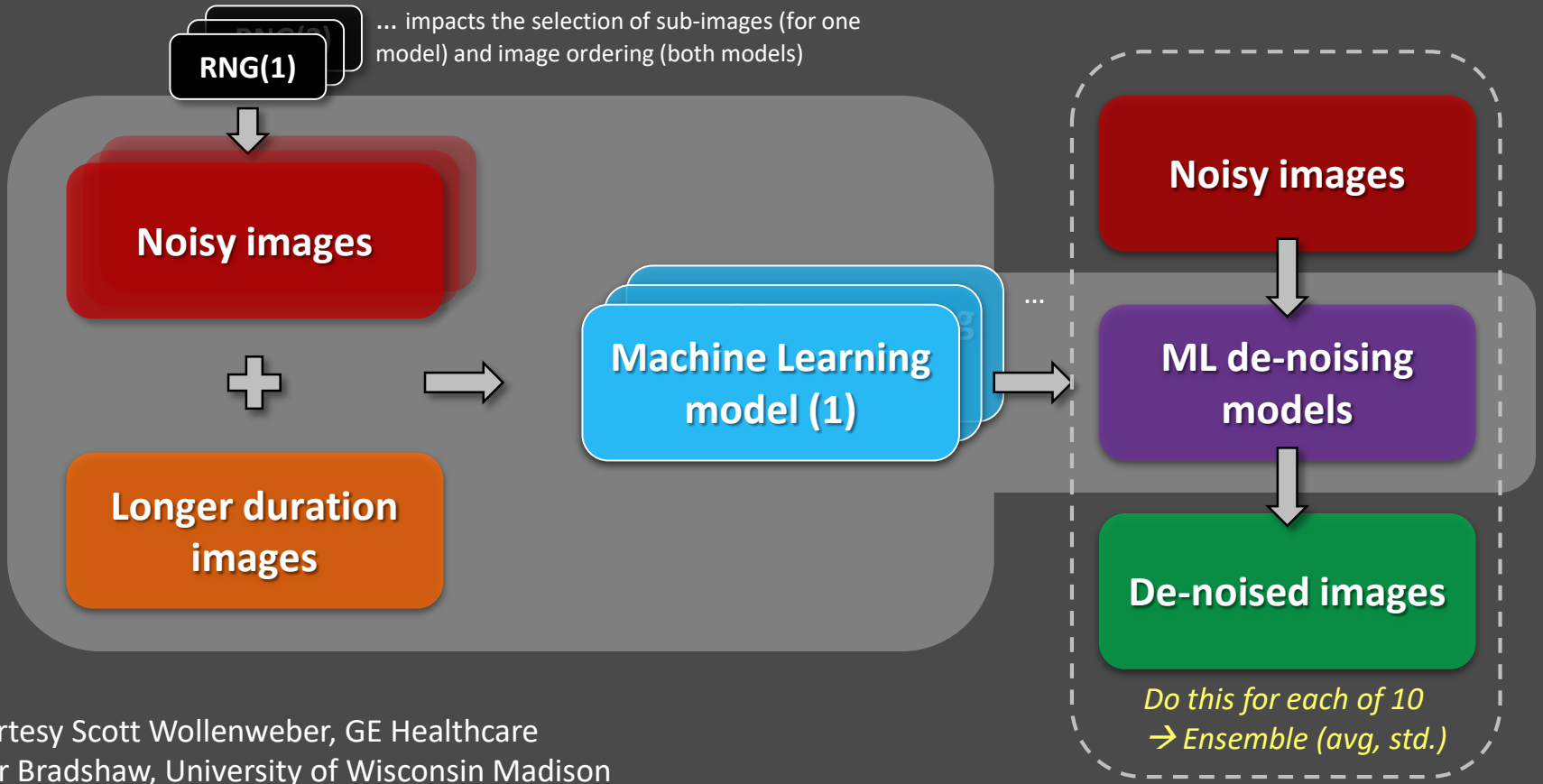
100 repeated simulations of a reconstructed test object with and without edge preserving labels



Comtat et al. PMB 2002

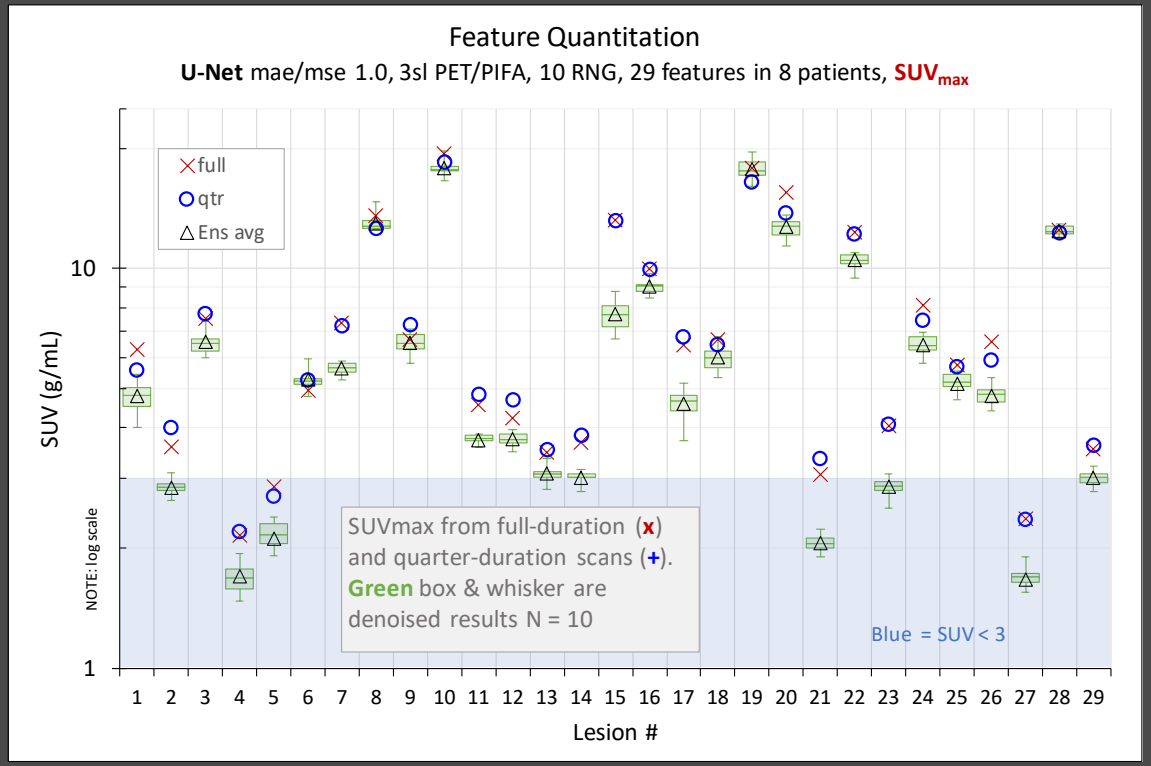
Message: Separate images do not tell the full story for detection

Other sources of variability, e.g. random number generation (RNG) of model parameters before training



Courtesy Scott Wollenweber, GE Healthcare
Tyler Bradshaw, University of Wisconsin Madison

Quantitative Image Feature Values in PET images: U-Net Denoising



Summary

For each rng (1-10), get average and std. of percent difference to long-duration result.

Calculate:

avg_e = average over (1-10) results,
std_e = average of stdev over (1-10)
$R(std_e)$ = range of stdev across ensemble

	1:30	Ens.
avg_e	0.0%	**-16.3%**
std_e	6.4%	**12.5%**
$R(std_e)$	-	5.3%

Courtesy Scott Wollenweber, GE Healthcare
Tyler Bradshaw, University of Wisconsin Madison

Message: Separate images do not tell the full story for quantitation

AMERICAN ASSOCIATION of PHYSICISTS IN MEDICINE
Improving Health Through Medical Physics

SUMMARY

- Machine intelligence based algorithms in medical imaging can offer many advantages, but in general we do not know when they are wrong. E.g. false-positives/negatives, quantitative errors
- The importance of noise, noise texture, and bias measures from multiple realizations is typically not appreciated or assessed
- There can be unintended consequences that are only apparent when the entire imaging chain is reviewed. E.g. AI-based image reconstruction -> AI-based denoising -> AI-based CAD, all from different vendors
- Metrology-based standards and terminology should be used, such as those used by QIBA, to avoid misinterpretation and improve rigor and reproducibility
- We need clinically relevant and feasible standardized assessments. Current approaches were developed for linear algorithms

Thanks to Scott Wollenweber, Chi Liu, Michael McNitt-Gray, Cynthia McCollough, William Hsu, Curt Langlotz, Berkman Sahiner, Ed Jackson, Nancy Obuchowski, Maryellen Giger, and Craig Abbey