POINT/COUNTERPOINT

Resolution modeling enhances PET imaging

Adam M. Alessio, Ph.D.
Department of Radiology, University of Washington, Seattle, Washington 98195
(Tel: 206-543-2419; E-mail: aalessio@u.washington.edu)

Arman Rahmim, Ph.D.
Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287
(Tel: 443-857-2080; E-mail: arahmim1@jhmi.edu)

Colin G. Orton, Ph.D., Moderator

OVERVIEW

One of the methods frequently employed to enhance PET images is resolution modeling (RM). Resolution modeling is known to visually enhance images. Some argue, however, that such improvements are deceptive and that RM leads to degradations elsewhere, whereas others claim that the enhancements are real and overall beneficial. This is the premise debated in this month's Point/Counterpoint.

Arguing for the Proposition is Adam M. Alessio, Ph.D. Dr. Alessio is Research Associate Professor in the Department of Radiology, and Adjunct Associate Professor of Bioengineering and Mechanical Engineering, at the University of Washington, Seattle. He received his Ph.D. in Electrical Engineering from the University of Notre Dame and has been at UW since 2003. Dr. Alessio's research focuses on tomographic image reconstruction for PET and CT systems. He is involved in numerous translational research projects for topics including cardiac perfusion imaging, radiation dose optimization for PET and CT, accurate system modeling, and statistical estimation of parametric images.

Arguing against the Proposition is Arman Rahmim, Ph.D. Dr. Rahmim is Assistant Professor in the Department of Radiology and Radiological Science, and Chief Physicist in the Section of High Resolution Brain PET Imaging, Johns Hopkins University School of Medicine. He received his Ph.D. in Medical Physics from the University of British Columbia, British Columbia, Canada, in 2005 and has been at Johns Hopkins since. His research interests include resolution modeling, whole-body parametric imaging, and 4D image reconstruction in dynamic as well as cardiac- and/or respiratory-gated PET imaging. He is certified by the ABSNM in Nuclear Medicine Physics and Instrumentation.

FOR THE PROPOSITION: Adam M. Alessio, Ph.D.

Opening Statement

Image generation in a wide-variety of fields, from microscopy[1] to astronomy,[2] employs resolution modeling techniques to reduce the degradations inherent in their imperfect imaging systems. In all of these applications, resolution degradation is a consequence of many factors including the physics of the signal, the sensors, and the electronics. In PET, the physical degradations directly related to spatial resolution loss, in order of origination from signal to final measurement,

include random positron range, photon pair noncollinearity, attenuation, intercrystal penetration, intercrystal scatter, detector inefficiencies, and electronics mispositioning. In total, a PET system has a spatially variant resolution loss. Resolution modeling and compensation techniques have been proposed for over two decades to account for the limitations in localization of radio-tracer distributions.[3]

To support the claim that "resolution modeling enhances PET imaging," it is important to clarify the term "enhances." In general, image restoration attempts to improve the image signal and/or reduce noise. In practice, restoration (or enhancement) needs to improve the task-based image quality (not simply signal or noise). It is well appreciated that two common tasks in clinical PET imaging include detection and quantification. In specific circumstances, resolution modeling in PET can genuinely improve hot-feature detection[4] and quantification.[5] Methods that enhance an image rarely provide improvement in all metrics and tasks—for example, detection may improve at the expense of quantitative accuracy. At the risk of oversimplifying the issue in the interest of a terse argument, my view is that most resolution modeling techniques provide some contrast enhancement with some apparent noise reduction (although true noise is minimally changed).[6] Both trends lead to demonstrably better detection performance,[4] leaving little room to question the assertion that resolution modeling enhances detection.

An interesting debate is whether resolution modeling enhances quantification. In PET quantification, we need both accurate and reproducible estimates of activity concentrations. The NEMA IEC phantom is commonly used to measure contrast recovery curves (CRC) for a system showing increasing partial volume errors for features below 2-3 cm. One holy grail in PET imaging is to develop a system and image generation method with a flat CRC curve (no partial volume effect errors with no size-dependent bias) with small error bars (reproducible). It has been shown that resolution modeling can cause unpredictable edge artifacts and these artifacts are generally exaggerated when the resolution model overestimates degradations.[7] Modest resolution models, which do not try to recover more frequency content that the sampling will support, have manageable edge artifacts and lead to genuine, albeit modest, contrast-to-noise improvements.[8] Current resolution modeling methodology is far from achieving the flat CRC curve. Future methods with better system modeling and convergent algorithms hold promise to improve the CRC further. These comments lead to a tempered statement that appropriate application of current resolution modeling techniques enhances, but does not solve all the challenges with, detection and quantification in PET imaging.

AGAINST THE PROPOSITION:
Arman Rahmim, Ph.D.

Opening Statement

Resolution modeling in PET has attracted considerable interest especially in the past decade.[3] Unlike post-reconstruction partial volume correction (PVC) methods, RM models resolution-degrading phenomena within the reconstruction. It is a natural approach as one aims to design the system matrix to faithfully reproduce the true probabilities of detection, and is an attractive alternative to a range of PVC methods that make simplifying assumptions. In fact, RM produces images that are clearly enhanced visually, but it is my contention that RM is remarkable in its ability to deceive!

RM improves resolution (and contrast), and it is unfortunately not uncommon to see studies only characterizing this aspect. RM also reduces noise when defined as intensity variations within a region-of-interest (ROI), i.e., image roughness (σ_{spatial}). An alternative noise metric that assesses reproducibility is the ensemble standard deviation of ROI mean uptake (σ_{ensemble}). RM has been shown to reduce voxel variances but increase intervoxel correlations.[6,9] The first effect decreases both σ_{spatial} and σ_{ensemble}, while the latter further decreases σ_{spatial}, but shifts σ_{ensemble} in the opposite direction.[6] Subsequently, σ_{spatial} is reduced in RM, but σ_{ensemble} can increase especially for small ROIs.[10] This explains why RM can generate images assessed visually to be of higher quality, as it enhances contrast and reduces σ_{spatial}. However, it can degrade reproducibility and thus adversely impact quantitative imaging tasks as in pharmacokinetic imaging[10] or treatment response monitoring. RM may actually improve reproducibility for the Maximum Standardized Uptake Value (SUVmax) (Ref. [11]) (which we attribute to reduced voxel variances) though increasing its range of values across the population (similar to PVC), and can degrade reproducibility for SUVmean, especially for small volumes (similar to PVC).[12]

In detection tasks, another note of caution is in order. Dual-metric resolution (contrast) vs. noise trade-off analyses commonly depict improved curves for RM whether noise is defined as σ_{spatial} or σ_{ensemble} (though to a lesser extent in the latter case, as explained above). Nonetheless, as demonstrated recently,[9] such simplified analyses do not properly capture the impact of the modified noise texture in PET images. In fact, detection task performance can be expressed as a function of the noise power spectrum (NPS), which is amplified at mid-frequencies with RM and competes against the RM-enhanced modulation transfer function (MTF). One then must not make any conclusions of RM superiority based on dual-metric analysis and appropriate task-based performance assessment is required. A few detection studies have been performed for RM in PET,[13,14] and the results indicated *statistically significant* improvements for the designed studies, especially in the presence of time-of-flight. Whether or not RM will be *clinically significant* is another question. I believe that RM has the definitive potential to improve PET imaging in the context of diagnostic imaging, especially in oncology, but is likely to degrade performance in other contexts. By no means do I intend to discourage the application of RM, but wish to draw attention to its strengths and pitfalls, and to encourage research into its usage in a balanced and thoughtful manner.

Rebuttal: Adam M. Alessio, Ph.D.

My colleague raises valid concerns that resolution modeling (RM) can be deceptive. RM must be assessed with

rigorous methods that analyze more than the simple metrics of: Resolution or quantification defined with hot features in a cold background; Noise defined as voxel-to-voxel variance; and Detection defined as contrast to noise. Some papers employ these simple metrics and report overly optimistic performance with RM.

The concern about the *clinical significance* of RM improvements is valid. A detection performance evaluation by Kadrmas *et al.*[4] demonstrated, through observer studies on over 400 measured phantom images, that the area under the ROC increases by ∼30% when using better RM. While this was a statistically significant improvement, the authors acknowledged that they could not conclude about the clinical significance of such a gain. In medical imaging, we rarely perform studies that prove real clinical improvements because these often require numerous patient exams, application in multiple sites, and knowledge of patient outcomes (all challenging!). We usually test our methods with much more limited evaluations. I would advocate that many of these limited evaluations, while not necessitating full clinical trials, should be improved.

In PET image generation, the customary approach for proving a method is to show improvement in a couple of reasonable metrics. This leads to the common expectation that performance on all other fronts stays consistent, i.e., the method only helps. For clinical acceptance, we need to quantify the good *and* bad performance of our methods to ensure they are applied in the appropriate context. In the future, I hope that clinical PET practice will use images tailored for the task at hand, as opposed to trying to garner all the necessary information from a single image. In the clinic, RM methods incorporated into convergent algorithms could provide more consistent, accurate quantification, but may not be appropriate for detection due to noise correlations. Conversely, RM methods designed to accentuate hot features of clinically relevant sizes could be used solely for tumor detection tasks. In this context, RM offers the potential to enhance clinical PET.

Rebuttal: Arman Rahmim, Ph.D.

Flattening the CRC curve is valuable but may incur other costs. What my esteemed colleague refers to as true noise [i.e., $\sigma_{ensemble}$, or the coefficient-of-variation (COV) when expressed as a percentage] may change little with RM,[2] or instead be amplified twofold[4] or even more[6] in small ROIs for some RM implementations. As such, the issue of reproducibility in quantitative imaging tasks merits special attention. This is a reason some sites with the HRRT scanner (including ours) pursuing quantitative pharmacokinetic imaging have discontinued usage of RM.

RM also results in increased mean uptake variability across the population,[15] attributed to true intersubject differences that are less suppressed in RM but may also be partly due to the degraded reproducibility.

Another issue is the impact of RM on the predictive and prognostic value of PET: PVC had no significant effect on the prediction of response following treatment[16] and in fact de-

graded performance in two studies.[17,18] This was attributed to the fact that PVC (similar to what RM does) removes the volume information implicit in SUVmean values, increasing them by greater amounts for complete-responders (which are associated with smaller tumors) than for partial-/nonresponders, thus actually diminishing intergroup differences. One could easily imagine that RM produces a similar detrimental effect on the discrimination power, though it is very meaningful to investigate explicit usage of volume information in addition to corrected SUV values.

Finally, I note that I would have again taken the counterpoint position if the proposition was instead that "RM does *not* enhance PET imaging" (!) for these two statements are not logical complements, and there is a third real one, namely that RM may enhance PET imaging in certain tasks and degrade others. The community needs to achieve careful and comprehensive assessment of these issues and propose solutions appropriately sensitive to the various imaging tasks.

[1] J. R. Swedlow, J. W. Sedat, and D. A. Agard, "Deconvolution in optical microscopy," in *Deconvolution of Images and Spectra*, 2nd ed., edited by P. Jansson (Academic Press, San Diego, CA, 1997), pp. 284–309, http://www.msg.ucsf.edu/agard/Publications/BookChapters/Bk-14-Swedlow.pdf.

[2] J. L. Starck, E. Pantin, and F. Murtagh, "Deconvolution in astronomy: A review," Publ. Astron. Soc. Pac. **114**, 1051–1069 (2002).

[3] A. Rahmim, J. Qi, and V. Sossi, "Resolution modeling in PET imaging: Theory, practice, benefits, and pitfalls," Med. Phys. **40**, 064301 (15pp.) (2013).

[4] D. J. Kadrmas, M. E. Casey, N. F. Black, J. J. Hamill, V. Y. Panin, and M. Conti, "Experimental comparison of lesion detectability for four fully-3D PET reconstruction schemes," IEEE Trans. Med. Imaging **28**, 523–534 (2009).

[5] V. Y. Panin, F. Kehren, C. Michel, and M. Casey, "Fully 3-D PET reconstruction with system matrix derived from point source measurements," IEEE Trans. Med. Imaging **25**, 907–921 (2006).

[6] S. Tong, A. M. Alessio, and P. E. Kinahan, "Noise and signal properties in PSF-based fully 3D PET image reconstruction: An experimental evaluation," Phys. Med. Biol. **55**, 1453–1473 (2010).

[7] C. C. Watson, "Estimating effective model kernel widths for PSF reconstruction in PET," *IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Valencia* (IEEE, Piscataway, NJ, 2011), pp. 2368–2374.

[8] S. Tong, A. M. Alessio, K. Thielemans, C. Stearns, S. Ross, and P. E. Kinahan, "Properties and mitigation of edge artifacts in PSF-based PET reconstruction," IEEE Trans. Nucl. Sci. **58**, 2264–2275 (2011).

[9] A. Rahmim and J. Tang, "Noise propagation in resolution modeled PET imaging and its impact on detectability," Phys. Med. Biol. **58**, 6945–6968 (2013).

[10] S. A. Blinder, K. Dinelle, and V. Sossi, "Scanning rats on the high resolution research tomograph (HRRT): A comparison study with a dedicated micro-PET," Med. Phys. **39**, 5073–5083 (2012).

[11] I. S. Armstrong, H. A. Williams, and J. C. Matthews, "Accuracy and variability of quantitative measurements using PET with time-of-flight information and resolution modelling," *2011 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Valencia* (IEEE, Piscataway, NJ, 2011), pp. 4167–4170.

[12] N. J. Hoetjes, F. H. P. van Velden, O. S. Hoekstra, C. J. Hoekstra, N. C. Krak, A. A. Lammertsma, and R. Boellaard, "Partial volume correction strategies for quantitative FDG PET in oncology," Eur. J. Nucl. Med. Mol. Imaging **37**(9), 1679–1687 (2010).

[13] D. J. Kadrmas, M. E. Casey, M. Conti, B. W. Jakoby, C. Lois, and D. W. Townsend, "Impact of time-of-flight on PET tumor detection," J. Nucl. Med. **50**, 1315–1323 (2009).

[14] J. Schaefferkoetter, M. Casey, D. Townsend, and G. El Fakhri, "Clinical impact of time-of-flight and point response modeling in PET reconstructions: A lesion detection study," Phys. Med. Biol. **58**, 1465–1478 (2013).

[15] I. Armstrong, M. Kelly, H. Williams, and J. Matthews, "Impact of PSF modelling and time-of-flight reconstruction on SUVmax, SUVpeak and total lesion glycolysis in lung lesions," J. Nucl. Med. **54**(Suppl. 2), Abstract 375 (2013), http://jnumedmtg.snmjournals.org/cgi/content/meeting_abstract/54/2_MeetingAbstracts/375.

[16] M. van Heijl, J. M. Omloo, M. I. V. Henegouwen, J. J. van Lanschot, G. W. Sloof, and R. Boellaard, "Influence of ROI definition, partial volume correction and SUV normalization on SUV-survival correlation in oesophageal cancer," Nucl. Med. Commun. **31**(7), 652–658 (2010).

[17] M. Hatt, A. Le Pogam, D. Visvikis, O. Pradier, and C. C. Le Rest, "Impact of partial-volume effect correction on the predictive and prognostic value of baseline F-18-FDG PET images in esophageal cancer," J. Nucl. Med. **53**(1), 12–20 (2012).

[18] J. A. Maisonobe, C. A. Garcia, H. Necib, B. Vanderlinden, A. Hendlisz, P. Flamen, and I. Buvat, "Comparison of PET metabolic indices for the early assessment of tumour response in metastatic colorectal cancer patients treated by polychemotherapy," Eur. J. Nucl. Med. Mol. Imaging **40**(2), 166–174 (2013) .