Design, implementation and first results of the future standard for evaluation of PET-AS methods

Emiliano Spezi, Ph.D.
School of Engineering
Cardiff University
United Kingdom

This presentation is based on the paper “Toward a standard for the evaluation of PET-Auto-Segmentation methods following the recommendations of AAPM task group No. 211: Requirements and implementation” which was recently published in Medical physics. The paper is open access.

Background

PET imaging important tool in radiation oncology.

- Patient staging, prognosis, radiation therapy planning, therapy monitoring, and detection/prediction of recurrences or metastatic disease

Accurate delineation and reliable PET segmentation methods.
- The need for reliable PET-auto-segmentation (PET-AS) methods has been widely expressed

Reliable technique for routine clinical PET-AS?
- There is currently no established agreement on the most reliable PET-AS technique

How to assess PET-AS algorithms?
- The development of a standard benchmark has been recognised by many including AAPM TG211
Objectives

1. Review (a) requirements, (b) design and (c) implementation of a benchmark tool for the evaluation and the validation of PET-AS algorithms (PETASset)

2. Show the analysis and report tools available in PETAsset

3. Discuss future developments of the benchmark

Standard requirements

Usability and accessibility

- Easy to use and learn: intuitive GUI
- Comprehensive documentation
- Accessible to the public
- Extendable
Standard requirements
Types of Reference Contours

Absolute truth: only available for simulated images.

Single best estimate: surrogate of truth provided for physical phantom images and in the special case of patient images for which histopathology data are available.

Multiple equally best estimates: consensus manual expert delineations when no single delineation can be considered to be the best.


Standard requirements
Categories of accuracy metrics

Level I: metrics that assess the agreement in terms of volumetric properties such as the number of voxels in the VOI and the statistics of PET signal integrated over that volume.

Level II: metrics that quantify the geometric agreement including spatial matching between a particular PET-AS contour and the RC.

Level III: metrics that evaluate the clinical relevance of the disagreement between PET-AS contours and RCs.


Standard requirements
Robustness

Across datasets: governed by differences in anatomy and physiology

Within a dataset: resulting from differences in tumour volume shape & size between different patients

Within an image: according to differences in image reconstruction and noise levels across different realisations of that image

Implementation of requirements

PETASset Platform

CERR: Computational Environment for Radiotherapy Research


https://github.com/adityaapte/CERR
http://www.cerr.info

PETASset Package structure and content
PETASset Data

Images and reference contours

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Reference</th>
<th>Modality</th>
<th>Region</th>
<th>Number of CTs</th>
<th>Number of PETs</th>
<th>Number of studies</th>
<th>Number of CT/PET pairs</th>
<th>Outcome</th>
<th>Additional notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCLTHN</td>
<td>Daisne JF, Sibomana M, Bol A, Doumont T, Lonneux M, Grégory V.</td>
<td>Head &amp; Neck</td>
<td>UCL H&amp;N</td>
<td>7</td>
<td>2</td>
<td>14</td>
<td>7</td>
<td>2/7</td>
<td>Two voxel sizes/PET scans.</td>
</tr>
</tbody>
</table>

Files: UCLPTHN_01.mat.bz2 - UCLPTHN_02.mat.bz2 - UCLPTHN_03.mat.bz2 - UCLPTHN_04.mat.bz2 - UCLPTHN_05.mat.bz2 - UCLPTHN_06.mat.bz2 - UCLPTHN_07.mat.bz2

PETASset Data

Head & Neck

UCL H&N, 7 different PET scans each with 2 voxel sizes/PET scans.

Files: UCLPTHN_01.mat.bz2 - UCLPTHN_02.mat.bz2 - UCLPTHN_03.mat.bz2 - UCLPTHN_04.mat.bz2 - UCLPTHN_05.mat.bz2 - UCLPTHN_06.mat.bz2 - UCLPTHN_07.mat.bz2

PETASset Data

Lung

UCL Lung, 10 different CT and PET scans each with 2 voxel sizes/PET scans.

Files: UCLPTLU_01.mat.bz2 - UCLPTLU_02.mat.bz2 - UCLPTLU_03.mat.bz2 - UCLPTLU_04.mat.bz2 - UCLPTLU_05.mat.bz2 - UCLPTLU_06.mat.bz2 - UCLPTLU_07.mat.bz2 - UCLPTLU_08.mat.bz2 - UCLPTLU_09.mat.bz2 - UCLPTLU_10.mat.bz2

PETASset Data

**Tumour shaped phantoms**

Milan physical body phantom with Zeolites, 11 different studies each with 6 different scans (acquisitions).

Files: MILPPAB_01.mat.bz2 - MILPPAB_02.mat.bz2 - MILPPAB_03.mat.bz2 - MILPPAB_04.mat.bz2 - MILPPAB_05.mat.bz2 - MILPPAB_06.mat.bz2 - MILPPAB_07.mat.bz2 - MILPPAB_08.mat.bz2 - MILPPAB_09.mat.bz2 - MILPPAB_10.mat.bz2 - MILPPAB_11.mat.bz2


PETASset Data

**Simulated**

MSKCC PET simulator [https://github.com/CRossSchmidtlein/PETSTEP](https://github.com/CRossSchmidtlein/PETSTEP)

5 RC geometries/ 2 reconstructions/ 5 acquisition instances.

Evaluate the robustness of a PET-AS method as we introduce differences in the same image by varying the image reconstruction and noise levels.

PETASset Data

**Simulated**

INSERM Brest. Numerical phantom, 6 H&N and 2 lung studies to simulate heterogeneous tracer uptake.

Lung files: BRENPLU_01.mat.bz2 - BRENPLU_02.mat.bz2
H&N files: BRENPHN_01.mat.bz2 - BRENPHN_02.mat.bz2 - BRENPHN_03.mat.bz2 - BRENPHN_04.mat.bz2 - BRENPHN_05.mat.bz2 - BRENPHN_06.mat.bz2

PETASset Code

Segmentation extended

Add name of algorithm here

Number of the scan to process

Bounding box (e.g. from initialisation)

PETASset Code

Analysis

Level I metrics are used to provide basic and essential information on the delineated VOI (a) Volume, (b) Mean uptake value, (c) Maximum uptake value, (d) Centre of mass

Level II metrics are used to quantify the similarity between the PET-AS contour and the RC.

\[
\begin{align*}
DSC &= \frac{2 \times (A \cap R)}{|A| + |R|}, \quad \text{range } [0, 1] \\
S &= \frac{|A \cap R|}{|A|}, \quad \text{range } [0, 1] \\
RD &= \max \left( \frac{1}{N} \sum_{i=1}^{N} \min(a_i - h), \frac{1}{N} \sum_{i=1}^{N} \min(h - a_i) \right), \quad \text{range } [0, +\infty) \\
PFPV &= \frac{|A \setminus R|}{|A|}, \quad \text{range } [0, 1] \\
DTPV &= \left( |A \cap R| - (A \cap R) \right), \quad \text{range } [0, +\infty)
\end{align*}
\]

PETASset Code

Level I Analysis

Level II Analysis

PETASset Code

PETASset Code

Analysis

Level I metrics are used to provide basic and essential information on the delineated VOI (a) Volume, (b) Mean uptake value, (c) Maximum uptake value, (d) Centre of mass

Level II metrics are used to quantify the similarity between the PET-AS contour and the RC.

\[
\begin{align*}
DSC &= \frac{2 \times (A \cap R)}{|A| + |R|}, \quad \text{range } [0, 1] \\
S &= \frac{|A \cap R|}{|A|}, \quad \text{range } [0, 1] \\
RD &= \max \left( \frac{1}{N} \sum_{i=1}^{N} \min(a_i - h), \frac{1}{N} \sum_{i=1}^{N} \min(h - a_i) \right), \quad \text{range } [0, +\infty) \\
PFPV &= \frac{|A \setminus R|}{|A|}, \quad \text{range } [0, 1] \\
DTPV &= \left( |A \cap R| - (A \cap R) \right), \quad \text{range } [0, +\infty)
\end{align*}
\]
PETASset Code

Level II Analysis

PETASset Code

Local report

Local report: designed to summarise the performance of PET-AS methods for a single study and a selection of metrics. The structured report contains

(i) PETASset analysis details

(ii) Level I analysis

(iii) Level II analysis

PETASset Code

Local report example
PETASset Code

Local report example

Global report: designed to include the performance of PET-AS methods across several cases.

<table>
<thead>
<tr>
<th>Method</th>
<th>Volume</th>
<th>Mean SEV</th>
<th>Mean SEV</th>
<th>DSC</th>
<th>S</th>
<th>PPF</th>
<th>SD (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLAIR</td>
<td>21 ± 13</td>
<td>5.6 ± 0.1</td>
<td>8.2 ± 0.0</td>
<td>0.26 ± 0.07</td>
<td>0.78 ± 0.08</td>
<td>0.09 ± 0.02</td>
<td></td>
</tr>
<tr>
<td>GMH</td>
<td>21 ± 25</td>
<td>5.8 ± 0.3</td>
<td>8.2 ± 0.0</td>
<td>0.26 ± 0.07</td>
<td>0.78 ± 0.08</td>
<td>0.09 ± 0.02</td>
<td></td>
</tr>
<tr>
<td>PTV</td>
<td>21 ± 30</td>
<td>5.9 ± 0.3</td>
<td>8.2 ± 0.0</td>
<td>0.26 ± 0.07</td>
<td>0.78 ± 0.08</td>
<td>0.09 ± 0.02</td>
<td></td>
</tr>
</tbody>
</table>

Conclusions

PETASset was designed and built following AAPM TG211 report which identified the need for developing a standard evaluation framework designed for the assessment of PET-AS algorithms.

PETASset includes a shared database of reference images and contours used in published articles. We expect this database to grow over time.

PETASset allows users to evaluate segmentation methods by either importing segmentation contours produced by external applications, or by coding a new segmentation method in the benchmark platform.

Future work includes the design and implementation of metrics to evaluate the clinical implications of contour accuracy in radiotherapy treatment planning (Level III Analysis).
Acknowledgments

AAPM TG211 members
http://www.aapm.org/org/structure/default.asp?committee_code-TG211

This work was supported in part through NIH/NCI grant R01CA172638 (WL) and NIH/NCI Cancer Center Support grant P30 CA008748 (CRS, WL, ASK), BB and ES acknowledge support from Cancer Research Wales grant No. 7061 and 2476. HZ acknowledges support from the Swiss National Science Foundation under grant SNF 31003A-149957 and the Swiss Cancer Research Foundation under grant KFS-3855-02-2016. JAL is a Research Associate with the Belgian F.R.S.-FNRS.

The work on this report was funded in part by the American Association of Physicists in Medicine.