SAM session

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.

Toward a standard for the evaluation of PET-Auto-Segmentation methods following the recommendations of AAPM task group No. 211: Requirements and implementation

Beatrice Berthon
Institut Langevin, ESPCI Paris, PSL Research University, CNRS UMR 7587, INSERM U979, Paris 75012, France

Emiliano Spezi[a)]
School of Engineering, Cardiff University, Cardiff, CF24 3AA, United Kingdom

Paulina Galavis
Department of Radiation Oncology, Langone Medical Center, New York University, New York, NY 10016, USA

Tony Shepherd
Turku PET Centre, Turku University Hospital, Turku 20521, Finland

Aditya Apte
Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

Mathieu Hatt, and Hadi Fayad
INSERM, UMR 1101, LaTIM, IBSAM, UBO, UBL, Brest 29609, France

Elisabetta De Bernardi
Medicine and Surgery Department, University of Milano-Bicocca, Monza 20900, Italy

Chiara D. Soffientini
Department of Electronics Information and Bioengineering, Politecnico di Milano, Milano 20133, Italy

C. Ross Schmidtlein
Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

Issam El Naqa
Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48103, USA

Robert Jeraj
School of Medicine and Public Health, University of Wisconsin, Madison, WI 53705, USA

Wei Lu
Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

Shiva Das
Department of Radiation Oncology, University of North Carolina, Chapel Hill, NC 27599, USA

Habib Zaidi
Division of Nuclear Medicine & Molecular Imaging, Geneva University Hospital, Geneva CH-1211, Switzerland

Osama R. Mawlawi
Department of Imaging Physics, MD Anderson Cancer Center, Houston, TX 77030, USA

Dimitris Visvikis
INSERM, UMR 1101, LaTIM, IBSAM, UBO, UBL, Brest 29609, France

John A. Lee
IREC/MIRO, Université catholique de Louvain (IREC/MIRO) & FNRS, Brussels 1200, Belgium

Assen S. Kirov
Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

(Received 2 August 2016; revised 7 April 2017; accepted for publication 15 April 2017; published xx xxxx xxxx)

Purpose: The aim of this paper is to define the requirements and describe the design and implementation of a standard benchmark tool for evaluation and validation of PET-auto-segmentation (PET-AS) algorithms. This work follows the recommendations of Task Group 211 (TG211) appointed by the American Association of Physicists in Medicine (AAPM).
Methods: The recommendations published in the AAPM TG211 report were used to derive a set of required features and to guide the design and structure of a benchmarking software tool. These items included the selection of appropriate representative data and reference contours obtained from

1 Med. Phys. 0 (0), xxxx 0094-2405/xxxx/0(0)/1/xx

 1

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

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.

Kirov AS, Fanchon LM. Pathology-validated PET image data sets and their role in PET segmentation. Clin Transl Imaging. 2014;2: 253–267.

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

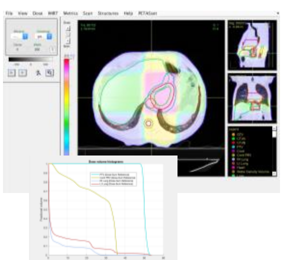
Within an image: according to differences in image reconstruction and noise levels across different realisations of that image

Hatt M, Lee J, Schmidtlein CR, et al. Classification and evaluation strategies of auto-segmentation approaches for PET: report of AAPM Task Group No. 211. Med Phys. 2017;44:e1–e42.

Implementation of requirements

PETASset
Platform

CERR: Computational Environment for Radiotherapy Research



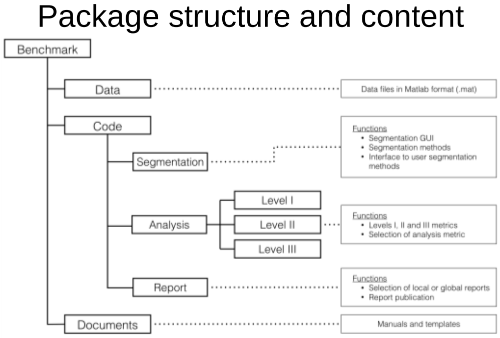
http://www.cerr.info
https://github.com/adityaapte/CERR

Washington University in St.Louis

Memorial Sloan Kettering Cancer Center

Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. Med Phys. 2003;30:979–985.

PETASset
Package structure and content

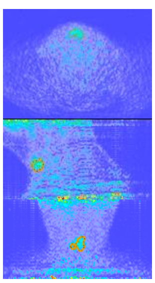


PETASset Data
Images and reference contours

Dataset	Reference	Center	Data type	Anatomical region	Number of studies	Number of series/study	Number of structures/ series	Reference contour	CT data	Additional features
UCLPTLU	Wanet et al.[11]	Université catholique de Louvain	Patient	Lung	10	2	1	Specimen	Yes	2 voxel sizes/PET scan
UCLPTHN	Daisne et al.[12]		Patient	H&N	7	1	1	Specimen	No	–
MILPPAB	Zito et al.[13]	Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico	Phantom	Lung & Pelvis	11	6	1	CT	No	Different acquisition instances
BRENPHN	Hatt et al.[14]	LaTIM, INSERM	Phantom	H&N	6	1	1	Simulation	No	Heterogeneous (2 RC contours)
BRENPLU			Phantom	Lung	2	1	1	Simulation	No	Heterogeneous (2 RC contours)
SIMPTLU	Berthon et al.[16]	MSKCC/ Cardiff University	Patient	Lung	10	5	1	Simulation	No	5 RC geometries/ 2 reconstructions/ 5 acquisition instances
SIMPTHN			Patient	H&N	10	5	1	Simulation	No	5 RC geometries/ 2 reconstructions/ 5 acquisition instances
SIMPTAB			Patient	Pelvis	10	5	1	Simulation	No	5 RC geometries/ 2 reconstructions/ 5 acquisition instances

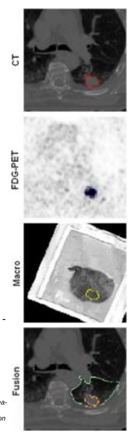
66 studies

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

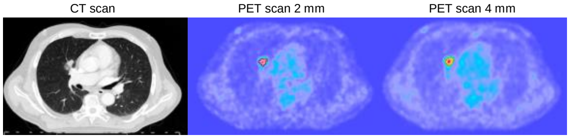


Daisne J-F, Sibomana M, Bol A, Doumont T, Lonneux M, Gr egoire V. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. Radiother Oncol. 2003;69:247–250.
Wanet M, Lee JA, Weynand B, et al. Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: a comparison with threshold-based approaches, CT and surgical specimens. Radiother Oncol. 2011;98:117–125.

PETASset Data
Lung

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

CT scan | PET scan 2 mm | PET scan 4 mm



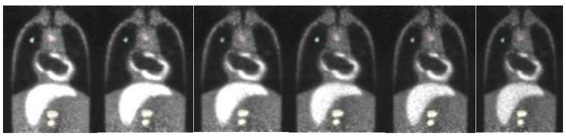
Files: UCLPTLU_01.mat.bz2 - UCLPTLU_02.mat.bz2 - UCLPTLU_03.mat.bz2 - UCLPTLU_04.mat.bz2 - UCLPTLU_05.mat.bz2 - UCLPTLU_06.mat.bz2 - UCLPTLU_07.mat.bz2 - UCLPTLU_08.mat.bz2 -UCLPTLU_09.mat.bz2 - UCLPTLU_10.mat.bz2

and B, et al. Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: a comparison with threshold-based approaches, CT and surgical specimens. Radiother Or

PETASset Data
Tumour shaped phantoms

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

PET scan 1 PET scan 2 PET scan 3 PET scan 4 PET scan 5 PET scan 6

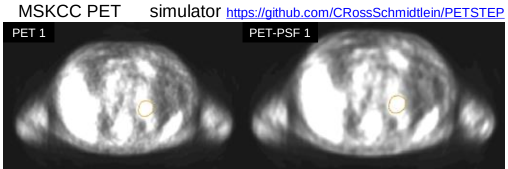


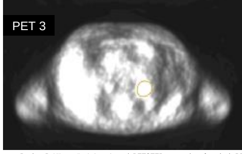
Files: MILPPAB_01.mat.bz2 - MILPPAB_02.mat.bz2 - MILPPAB_03.mat.bz2 - MILPPAB_04.mat.bz2 - MILPPAB_05.mat.bz2 - MILPPAB_06.mat.bz2 - MILPPAB_07.mat.bz2 - MILPPAB_08.mat.bz2 - MILPPAB_09.mat.bz2 - MILPPAB_10.mat.bz2 - MILPPAB_11.mat.bz2

Zito F, De Bernardi E, Soffientini C, et al. The use of zeolites to generate PET phantoms for the validation of quantification strategies in oncology. Med Phys. 2012;39:5353–5361.

PETASset Data
Simulated

MSKCC PET simulator https://github.com/CRossSchmidtlein/PETSTEP





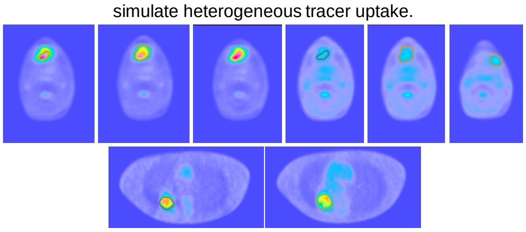
5 RC geometries/ 2 reconstructions/ 5 acquisition instances.

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

Berthon B, Häggström I, Apte A, et al. PETSTEP: generation of synthetic PET lesions for fast evaluation of segmentation methods. Phys Med. 2016;31:969–980.

PETASset Data
Simulated

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

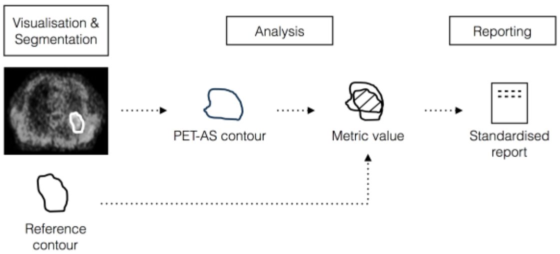


Lung files: BRENPLU_01.mat.bz2 - BRENPLU_02.mat.bz2
H&N files: BRENPHN_01.mat.bz2 - BRENPHN_02.mat.bz2 - BRENPHN_03.mat.bz2 - BRENPHN_04.mat.bz2 - BRENPHN_05.mat.bz2 - BRENPHN_06.mat.bz2

Hatt M, Cheze le Rest C, Descourt P, et al. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. Int J Radiat Oncol Biol Phys. 2010;77:301–308.

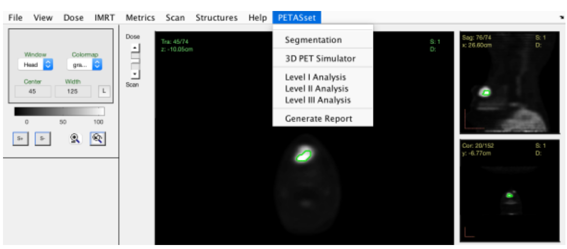
PETASset Code

Workflow



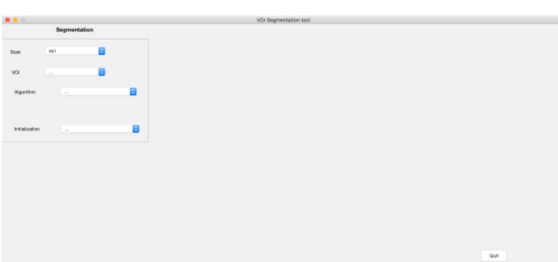
PETASset Code

Segmentation

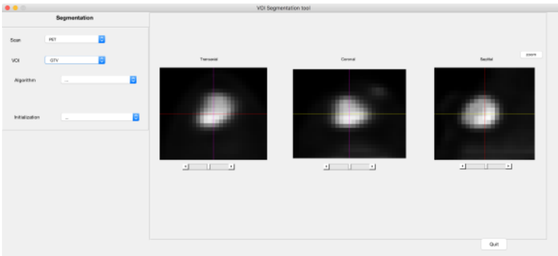


PETASset Code

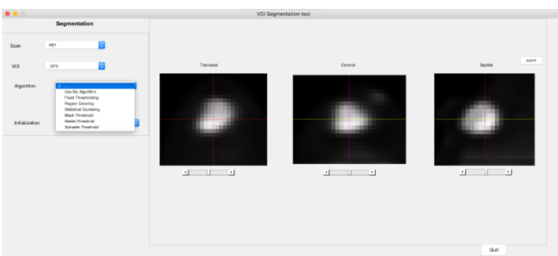
Segmentation



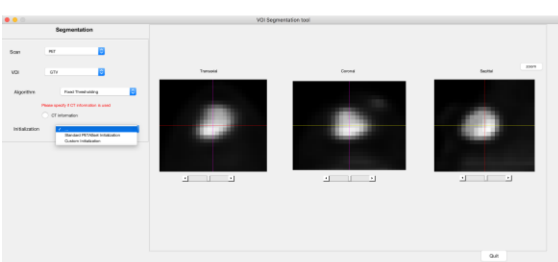
PETASset Code
Segmentation



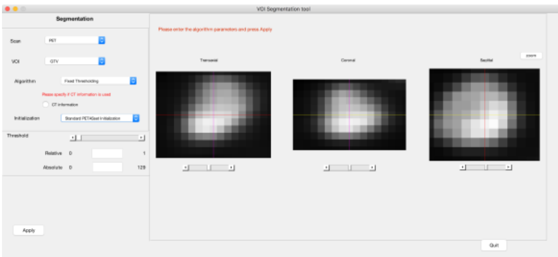
PETASset Code
Segmentation



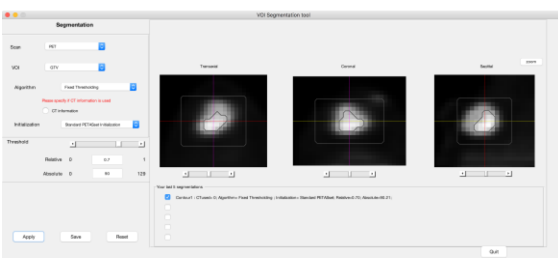
PETASset Code
Segmentation



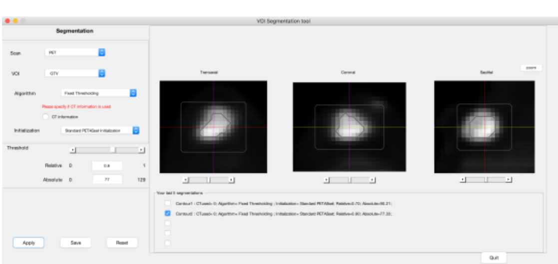
PETASset Code
Segmentation



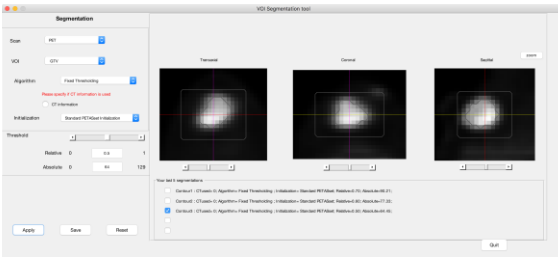
PETASset Code
Segmentation



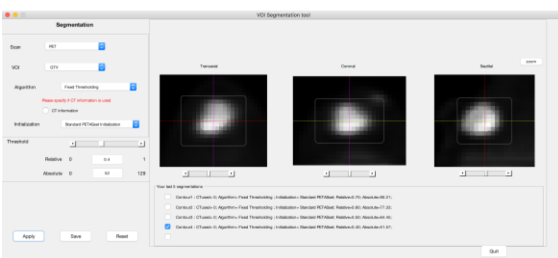
PETASset Code
Segmentation



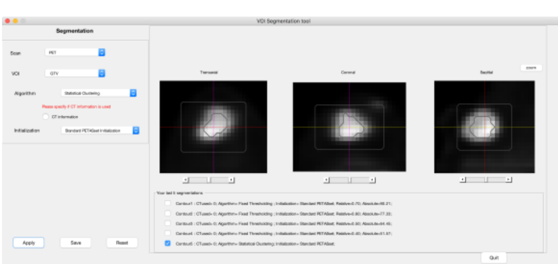
PETASset Code
Segmentation



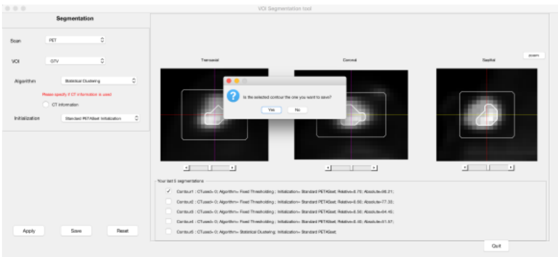
PETASset Code
Segmentation



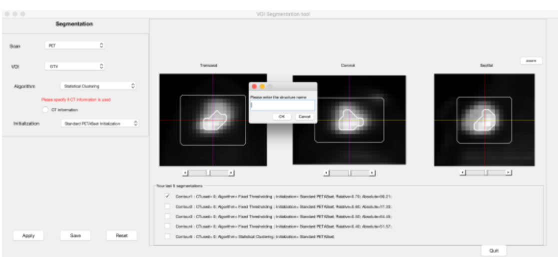
PETASset Code
Segmentation



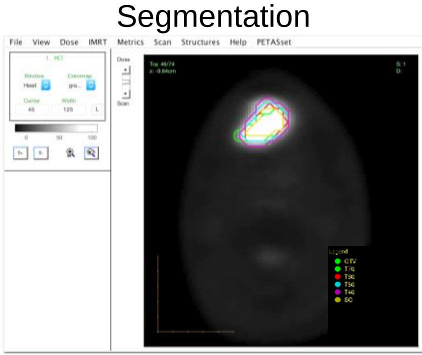
PETASset Code
Segmentation



PETASset Code
Segmentation



PETASset Code
Segmentation



PETASset Code

Segmentation extended

```
function [Mask,algorithmName,parameterNames,parameterValues]=YourSegmentationAlgorithm_PETASset(ScanNum,Box)
% This is a template for custom segmentation algorithm implementation into
% the BM
% N.B.: The arguments to this script are:
%     - Scan No
%     - bounding box corresponding to initilisation chosen
%
% Author: B. Berthon
% email: BerthonB@cardiff.ac.uk

%------------- BEGIN CODE --------------

if ~exist('planC')
   global planC
end
indexS = planC{end};

% Image corresponding to the whole scan
Image=planC{indexS.scan}(ScanNum).scanArray;

% Image to segment (cropping the scan around the structure to delineate)
cImage=Image(Box(1):Box(2),Box(3):Box(4),Box(5):Box(6));

algorithmName='Name of your method'; % please replace

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% your code here
% input: cImage (scan cropped around the structure to delineate)
% output:
%   - cMask: a binary mask (1 in the tumour, 0 outside), the same size as cImage
% additional optional outputs:
%   - parameterNames: cell array of strings with the names of parameters used, e.g. {'paramater1', parameter2',...}
%   - parameterValues: array of parameter values, e.g. [2, 0.5]

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% returning Mask from cMask
Mask=zeros(size(Image));

%------------- END CODE --------------
```

Bounding box (e.g. from initialisation)

Number of the scan to process

Add name of algorithm here

Add code that outputs cMask and parameters

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 update value, (d) Centre of mass

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

$$DSC\,(A,B) = \frac{2 \times |A \cap B|}{|A| + |B|}, \quad \text{range } [0,1]$$

$$S = \frac{|A \cap B|}{|A|}, \quad \text{range } [0,1]$$

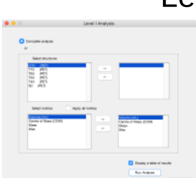
$$HD = \max\left(\frac{1}{N_A}\sum_i min_j\|a_i - b_j\|, \frac{1}{N_B}\sum_i min_j\|b_i - a_j\|\right), \quad \text{range } [0,+\infty)$$

$$PPV = \frac{|A \cap B|}{|B|}, \quad \text{range } [0,1]$$

$$DUV = (|A \cup B|) - (|A \cap B|), \quad \text{range } [0,+\infty)$$

PETASset Code

Level I Analysis



```
>> planC{indexS.structures}(1).analysisI
ans =
  struct with fields:

    Volume: 3.5944
       COM: [26.6236 -6.6497 -9.9011]
      Mean: 98.0081
       Max: 128.8421
```

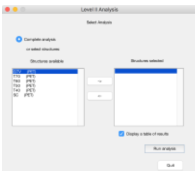
```
>> planC{indexS.structures}(2).analysisI
ans =
  struct with fields:

    Volume: 3.3669
       COM: [26.6351 -6.6527 -9.8223]
      Mean: 105.6920
       Max: 128.8421
```

Figure 3: Results of Analysis I

File Edit View Insert Tools Desktop Window Help

	Volume (mL)	Centre of Mass (COM) x	Centre of Mass (COM) y	Centre of Mass (COM) z	Mean	Max
GTV (PET)	3.5944	26.6236	-6.6497	-9.9011	98.0081	128.8421
T70 (PET)	3.3669	26.6351	-6.6527	-9.8223	105.6920	128.8421
T60 (PET)	5.7329	26.6673	-6.4997	-9.9280	96.2451	128.8421
T50 (PET)	8.3719	26.6896	-6.4221	-9.9863	88.1812	128.8421
T40 (PET)	11.0108	26.6673	-6.3994	-9.9606	81.1861	128.8421
SC (PET)	4.0494	26.7289	-6.3959	-9.9318	99.6171	128.8421

PETASset Code
Level II Analysis




```
>> planC{indexS.structures}(2).analysisII
ans =
  struct with fields:

                      DSC: [1×1 struct]
                        S: [1×1 struct]
                      PPV: [1×1 struct]
                       HD: [1×1 struct]
                       DU: [1×1 struct]
    ReferenceStructureType: 'Simulation template'
```

```
>> planC{indexS.structures}(2).analysisII.DSC
ans =
  struct with fields:

                     Name: 'Dice Similarity Coefficient'
                    Value: 0.8468
      ReferenceStructIndex: '1'
    ReferenceStructureName: 'GTV'
```

Figure 3: Results of Analysis II

File Edit View Insert Tools Desktop Window Help

	Dice	Sensitivity	Positive Predictive Value	Hausdorff Distance (cm)	Delineation Uncertainty (cm)
GTV (PET)	1	1	1	0	0
T70 (PET)	0.8468	0.7491	0.9738	0.1317	0.3570
T60 (PET)	0.7821	0.9029	0.6898	0.1698	0.3570
T50 (PET)	0.6626	0.9622	0.5053	0.2917	0.3570
T40 (PET)	0.5612	0.9798	0.3932	0.4079	0.3570
SC (PET)	0.7216	0.6961	0.7490	0.2062	0.3570

PETASset Code
Local report

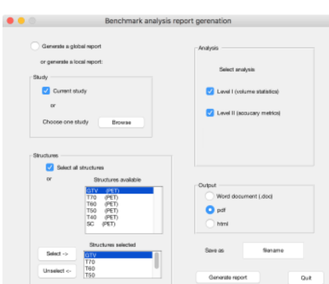
Local Report: designed to summarise the performance of PET-AS methods for a single study and a selection of metrics. The structured report contains

(i) PETASset analysis details

(ii) Level I analysis

(iii) Level II analysis

PETASset Code
Local report example



PETASset Code

Local report example

PETASset Code

Global report

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

TABLE III. Average Level I and Level II metric values calculated across the entire PETASset dataset and associated standard deviation.

	Level I Absolute metric error (% RC)			Level II			
Method	Volume	Max SUV	Mean SUV	DSC	S	PPV	HD (cm)
FLAB	27 ± 15	3.0 ± 12	6.3 ± 11	0.74 ± 0.07	0.69 ± 0.09	0.82 ± 0.09	0.25 ± 0.16
GMM	21 ± 25	5.0 ± 11	0.21 ± 10	0.76 ± 0.08	0.77 ± 0.08	0.78 ± 0.09	0.17 ± 0.12
FT50	60 ± 37	0.89 ± 11	3.7 ± 35	0.53 ± 0.08	0.43 ± 0.11	0.91 ± 0.10	0.30 ± 0.08
FT42	61 ± 70	0.36 ± 9.8	15 ± 20	0.64 ± 0.07	0.56 ± 0.09	0.88 ± 0.09	0.24 ± 0.08
RG	42 ± 21	0.18 ± 12	11 ± 18	0.68 ± 0.07	0.62 ± 0.10	0.85 ± 0.11	0.23 ± 0.10
KM	70 ± 163	2.7 ± 11	11 ± 58	0.73 ± 0.10	0.85 ± 0.05	0.69 ± 0.13	0.27 ± 0.20
GCM	39 ± 13	0.98 ± 9.6	9.0 ± 17	0.70 ± 0.06	0.65 ± 0.09	0.83 ± 0.09	0.19 ± 0.05
WT	42 ± 26	2.5 ± 11	3.3 ± 18	0.67 ± 0.07	0.63 ± 0.11	0.79 ± 0.10	0.22 ± 0.08
Range	21/70	0.18/5.00	0.21/15	0.53/0.76	0.43/0.85	0.69/0.91	0.17/0.30
Median (SD)	42 (± 17)	1.7 (± 1.6)	7.7 (± 4.9)	0.69 (± 0.07)	0.64 (± 0.13)	0.83 (± 0.07)	0.24 (± 0.04)
Agreement limits (example)	(0,59)	(0,3.3)	(0,12.6)	(0.62,1)	(0.51,1)	(0.76,1)	(0,0.28)

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 SNFN 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.