SAM session : TH-AB-702-3

Components of a standard and a procedure for evaluation of PET-AS methods

Assen S Kirov, Ph.D.
Department of Medical Physics
Memorial Sloan-Kettering Cancer Center
New York

Disclosure: No conflict of interest

Objectives

I. Introduction of an evaluation standard
- components
- performance criteria

II. Discussion of the limitations and dependencies of the PET segmentation process

III. Acceptance and implementation of PET auto-segmentation algorithms

Evaluation criteria for segmentation algorithms

- Accuracy
- Precision
- Efficiency

Udupa *et al,* A framework for evaluation of image segmentation algorithms, Computerized Medical Imaging and Graphics, 30, (2006) 75-87

- Robustness

Hatt *et al,* 2011, 'PET functional volume delineation: a robustness and repeatability study', *Eur J Nucl Med Mol Imaging*, vol. 38, no. 4, pp. 663-72.

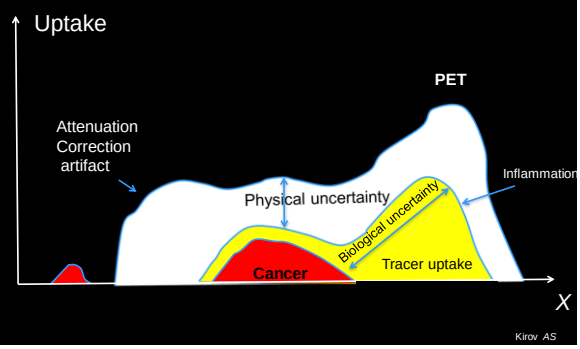
What should be "true" (reference) volume ?

(a) The avid volume in the PET image

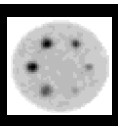
(b) The activity distribution

(c) The biological quantity of interest

What should be "true" (reference) volume ?



Benchmark Image sets:
A. Phantoms
1. Simple



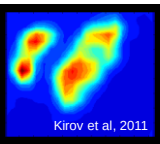
Advantages:

- Exact representation of the scanner resolution, image noise and other image artifacts
- Ground truth accurately known
- Easy to generate and use

Disadvantages:

- The objects have simplistic and unrealistic shape and activity distribution
- Most with few exceptions have cold walls

Benchmark Image sets:
A. Phantoms
2. Realistic


Kirov et al, 2011

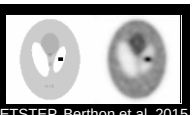
Advantages:

- Exact representation of the scanner resolution, image noise and other image artifacts
- Capable to produce lesion shapes corresponding to actual tumors
- Known ground truth

Disadvantages:

- Difficult to generate inhomogeneous activity
- Labor intensive
- Experimental uncertainties

Benchmark Image sets:
B. Simulated Phantoms
1. Forward Projected images


PETSTEP, Berthon et al, 2015

Advantages:

- Flexibility in phantom design
- Precise knowledge of the reference object
- Computationally cheap

Disadvantages:

- Scatter count distributions and noise are usually less accurately modeled
- Detailed physics and system information ignored

Benchmark Image sets:
B. Simulated Phantoms
1. Monte Carlo


Le Miatre et al, 2009

Advantages:

- Realistic count distributions
- Precise knowledge of the reference object
- Scanner-specific information

Disadvantages:

- Computationally expensive
- Model requires extensive up front experience

Benchmark Image sets: C. Clinical images

Ga-68 DOTA-JR11

Advantages:

- Exact representation of the scanner resolution, image noise and other image artifacts
- Real activity distributions

Disadvantages:

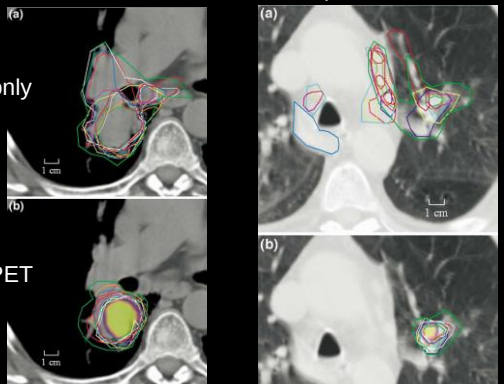
- Uncertainties in the knowledge of the reference object, even with histopathology reference

Clinical Images: The ground truth problem

Steenbakkers, IJROBP, 64, no.2, p. 435 ,2006

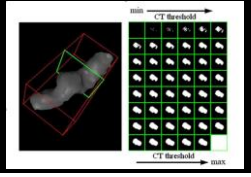
CT only

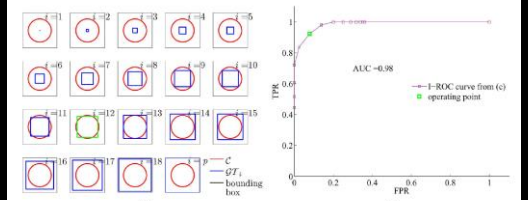
CT + PET



The ground truth problem: a series of monotonically ordered reference contours

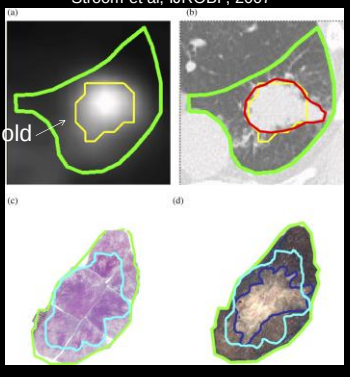
T. Shepherd, M. Teras, et al, TMI, 2012





Clinical Images: The ground truth problem

Stroom et al, IJROBP, 2007



PET:
42% threshold

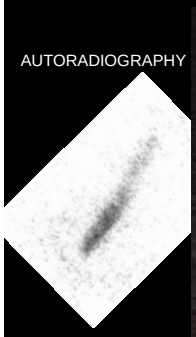
CT

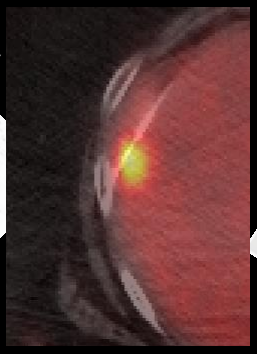
Micro-
scopic
slice
(4 um)

Macro-
+ micro-
scopic
contours

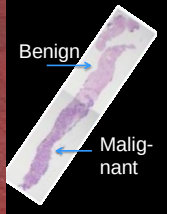
Clinical Images: The ground truth problem and PET/CT guided biopsies

AUTORADIOGRAPHY





PATHOLOGY
MICROGRAPH

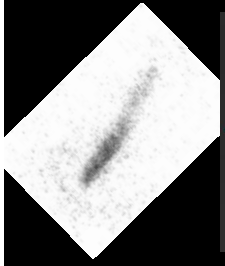


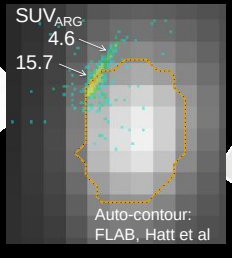
Kirov et al, SNM 2014

Fanchon et al, JNM 2016

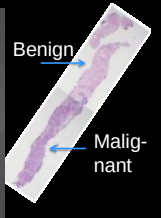
Clinical Images: The ground truth problem and PET/CT guided biopsies

AUTORADIOGRAPHY





PATHOLOGY
MICROGRAPH

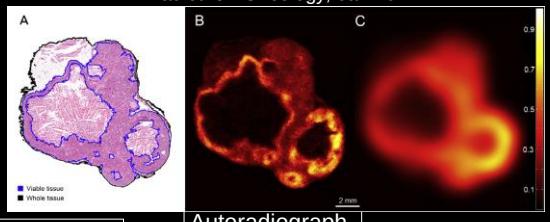


Kirov et al, EANM 2015

Fanchon et al, JNM 2016

An alternative approach to histopathological validation The "Synthetic PET" approach

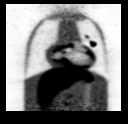
Marian Axente, … Jeffrey F. Williamson, Andrei Pugachev
Radiother. Oncology, Jan 2014



Excise and section lesion → Autoradiograph each section: tracer uptake in 3D → Simulated PET

Types of Images in the benchmark

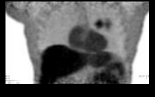
Physical Phantoms



F. Zito, E. De Bernardi, *et al,* Med Phys, 2012

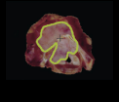
Simulated Phantoms

Berthon B et al,, Med. Phys 2015



Le Maitre, *et al,* IEEE proc. 2009

Clinical Images



Daisne, *et al,* Radiology. 2004

Evaluation Metrics –illustration and recommendations

1. Volume Difference B(segmented) - A(reference)
2. Barycenter Distance 
3. Jaccard Similarity Coefficient $\frac{|A \cap B|}{|A \cup B|}$
4. Dice Similarity Coefficient $\frac{2|A \cap B|}{|A| + |B|}$ 
5. Sensitivity $\frac{|A \cap B|}{|A|} = \frac{|TP|}{|TP| + |FN|}$
6. Positive Predictive Value $\frac{|A \cap B|}{|B|} = \frac{|TP|}{|TP| + |FP|}$
7. Hausdorff Distance $\mathrm{HD}(A,B) = \max\left\{\max_i \min_j \delta(a_i, b_j), \max_j \min_i \delta(a_i, b_j)\right\}$

Comparison of Contour Evaluation Metrics

Evaluation criteria	Location	Size	Shape	FN,FP	Complexity
Volume difference	no	yes	no	no	+
Barycenter distance	yes	no	no	no	++
Jaccard similarity coefficient	yes	yes	yes	no	++
Dice similarity coefficient (DSC)	yes	yes	yes	no	++
Hausdorff distance	yes	no	yes	no	+++
Sensitivity + Positive Predictive Value (PPV)	yes	yes	yes	yes	++

Uncertainties in PET Segmentation

A. Uncertainties in the PET image

Boellaard, JNM 2009:
- technical (specification, administration, time)
- physical (detection physics, reconstruction)
- biological (glucose level, inflammation, comfort)
  - tumor heterogeneity

B. Inaccuracies of auto-segmentation

C. Dependence on:
- scanner and protocol
- tracer type and isotope
- lesion type and body site
- segmentation task

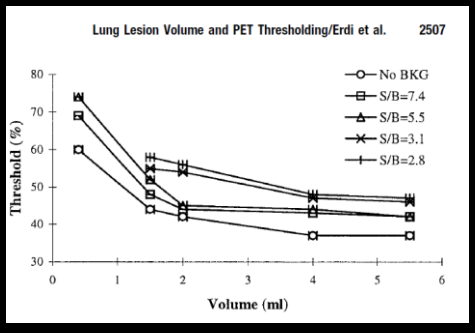
Segmentation requirements: Dependence on the task

Radiation therapy:

1. Target definition
   a) Tumor delineation
   b) “Dose painting”
      - aggressive region
      -”by numbers”

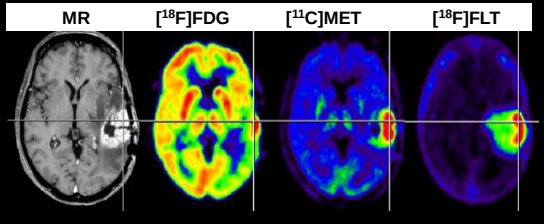
2. Treatment assessment
   a) by changes in volume
   b) by uptake changes in segmented vol.

If tumor changes…
=> Threshold methods are not consistent



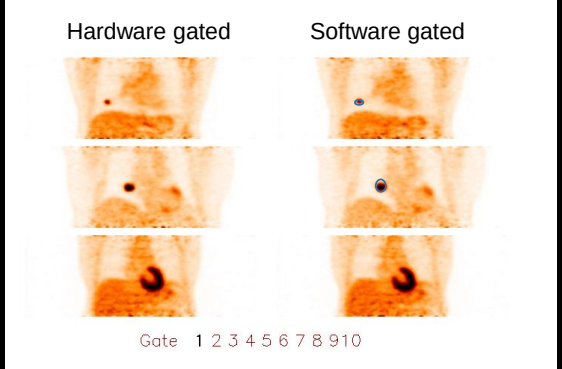
Erdi et al, Cancer, suppl. 1997

Dependence on tracer type



Courtesy Joseph Osborne, MD, MSKCC

Effect of motion on segmentation



Courtesy Adam Kesner

Kesner et al, Radiology, 281, 2016

How to evaluate a PET segmentation tool

Step	1. Vendor acceptance	2. Basic	3. Phase II	4. Phase III	5. Impact
Objective	Proper functioning of software	Accuracy, Repeatability Robustness	ARR -> realistic shapes and uptake	ARR -> clinical images	Evaluation of clinical impact
Datasets	Vendor	Simple uniform, objects repeated	Irregular shape, non-uniform phantoms no cold wall	Clinical images	Clinical images, treatment plans and follow-up
Ground truth	Vendor	CT	High res. CT or digital voxel level accuracy.	Digitized histopathology and/or manual delineations	Treatment outcome data
Metrics	Vendor	Volume errors, DSC	DSC, Sensitivity, PPV, HD	DSC, Sensitivity, PPV, HD, Statistical	Outcome

Evaluation of PET Auto Segmentation: Summary

I. The evaluation standard:

Criteria: Accuracy, Precision, Robustness, Efficiency

Image sets: Different Phantoms and Clinical images

Figures of merit: Sensitivity, PPV, Hausdorff distance

II. Acceptance and evaluation

Multistep: Vendor, Basic, Realistic, Impact

III. PET Segmentation Limitations

Biological phenomena – heterogeneity

Dependencies – scanner, protocol, tracer,motion

Physician review and editing is imperative

Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211

Med. Phys. 2017 online

Mathieu Hatt
INSERM, UMR 1101, LaTIM, University of Brest, IBSAM, Brest, Fr

Michael P. MacManus
Peter MacCallum Cancer Centre, Melbourne, Australia

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

Osama R. Mawlawi
MD Anderson Cancer Center, Houston, TX 77030, USA

Charles R. Schmidtlein
Memorial Sloan Kettering Cancer Center, New York, NY 10065, U.

Ursula Nestle
Universitätsklinikum Freiburg, Freiburg 79106, Germany

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

Andrei B. Pugachev
University of Texas Southwestern Medical Center, Dallas, TX

Curtis Caldwell
Sunnybrook Health Sciences Center, Toronto, ON M4N 3M5, Cana

Heiko Schöder
Memorial Sloan Kettering Cancer Center, New York, NY 100

Elisabetta De Bernardi
University of Milano-Bicocca, Monza, Italy

Tony Shepherd
Turku University Hospital, Turku 20521, Finland

Wei Lu
Memorial Sloan Kettering Cancer Center, New York, NY 10065, U.

Emiliano Spezi
School of Engineering, Cardiff University, Cardiff, Wales, Un

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

Dimitris Visvikis
INSERM, UMR 1101, LaTIM, University of Brest, IBSAM, Br

Xavier Geets and Vincent Gregoire
Université catholique de Louvain (IREC/MIRO) & FNRS, Brussels

Habib Zaidi
Geneva University Hospital, Geneva CH-1211, Switzerland

Robert Jeraj
University of Wisconsin, Madison, WI 53705, USA

Assen S. Kirov[a)]
Memorial Sloan Kettering Cancer Center, New York, NY 100