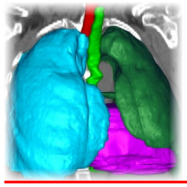
Introduction to the Auto-Segmentation Grand Challenge

Jinzhong Yang, PhD

MD Anderson Cancer Center



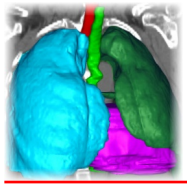
Organizers and Major Contributors

- Jinzhong Yang (MD Anderson Cancer Center)
- Greg Sharp (Massachusetts General Hospital)
- Mark Gooding (Mirada Medical)
- Harini Veeraraghavan (Memorial Sloan Kettering Cancer Center)
- Wouter van Elmpt (Maastro Clinic)
- Samuel Armato III (University of Chicago)
- Keyvan Farahani (NIH/NCI)
- Andre Dekker (Maastro Clinic)
- Tim Lustberg (Maastro Clinic)
- Justin Kirby (NIH/NCI)
- Kirk Smith (TCIA)
- Tracy Nolan (TCIA)
- Jayashree Kalpathy-Cramer (Harvard University)
- Artem Mamonov (Harvard University)
- Andrew Beers (Harvard University)

Sponsors

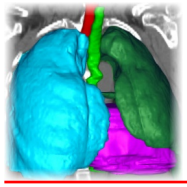
AMERICAN ASSOCIATION of PHYSICISTS IN MEDICINE

THE CANCER IMAGING ARCHIVE



Disclosures

- I received funding/salary support (not related to this work) from the CPRIT RP110562, NIH/NCI UH2 CA202665, NIH/NCI U19 CA021239, NIH/NCI R01 CA218148, MDACC Institutional Research Grant, and Master Research Agreement with Varian Medical Systems.

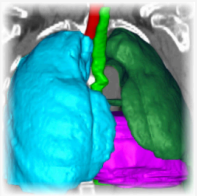


Introduction

- Autoseg 2013 workshop
  - Time: March 8, 2013; Location: Boston, MA
  - https://na-mic.org/wiki/Autoseg_2013
- Autoseg 2015 workshop
  - In conjunction with 2015 World Congress on Medical Physics & Biomedical Engineering
  - Time: June 7, 2015; Location: Toronto, Canada
  - https://na-mic.org/wiki/Autoseg_2015
- 2017 AAPM Thoracic Auto-segmentation Challenge
  - http://autocontouringchallenge.org

AAPM CHALLENGES

My Competitions



AAPM Thoracic Auto-segmentation Challenge

Organized by MarkGooding - Current server time: Aug. 2, 2017, 7:58 a.m. UTC

Previous	▶ Current	Next
Training Phase	Pre-AAPM Challenge	AAPM Live Challenge
May 18, 2017, midnight UTC	June 19, 2017, midnight UTC	Aug. 2, 2017, midnight UTC

Learn the Details | Phases | Participate | Results | Forums | Team

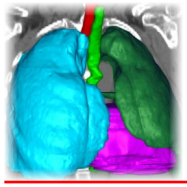
- Overview
- Contouring Guidelines
- Evaluation
- Submission Guidelines
- Live Challenge Infomation
- Terms And Conditions
- Organizers

Overview

Numerous auto-segmentation methods exist for Organs at Risk in radiotherapy. The overall objective of this auto-segmentation grand challenge is to provide a platform for comparison of various auto-segmentation algorithms when they are used to delineate organs at risk (OARs) from CT images for thoracic patients in radiation treatment planning. The results will provide an indication of the performances achieved by various auto-segmentation algorithms and can be used to guide the selection of these algorithms for clinic use if desirable. The challenge is made up of multiple phases:

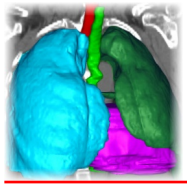
Phase 1 will conducted via this website in advance of the AAPM meeting. 12 test images will be provided and results will be submitted online. The 3 top place contestants in this phase will be invited to present at the challenge symposium at AAPM.

Phase 2 will be conducted live at the AAPM. A further 12 test images will be provided for



Why Challenge

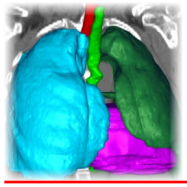
- Provide a platform for comparison of various auto-segmentation algorithms
- Provide a guideline for the selection of auto-segmentation algorithms for clinical use
- Provide benchmark data for evaluating/commissioning auto-segmentation tools



Challenge Format

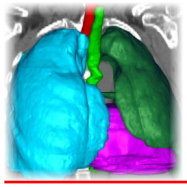
Training Phase | Pre-AAPM Challenge | AAPM Live Challenge

- Training phase (May 19 – Jun 20)
  - Download 36 training datasets with ground truth to train and optimize segmentation algorithms
- Pre-AAPM challenge (Jun 21 – Jul 17)
  - Perform segmentation on 12 off-site test datasets
- AAPM Live challenge (Aug 2)
  - Perform segmentation on 12 live test datasets and submit results within 2 hours at AAPM meeting



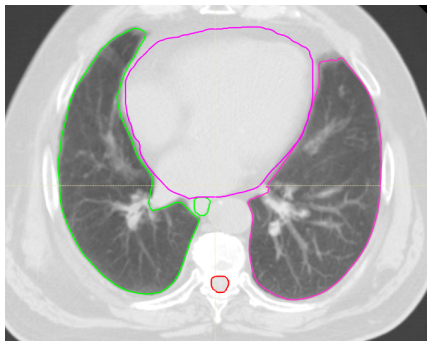
Benchmark Data

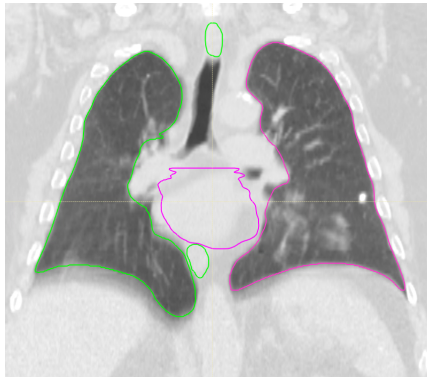
- 60 patients with their average 4DCT or free breathing CT scans
- Acquired from 3 institutions (20 each)
  - Maastro clinic
  - MD Anderson Cancer Center
  - Memorial Sloan Kettering Cancer Center
- Data divided into 3 groups, stratified per institution
  - 36 training datasets
  - 12 off-site test datasets
  - 12 live test datasets
- Data provided in DICOM-RT (both CT and RT Structure set), as commonly used in most commercial TPS

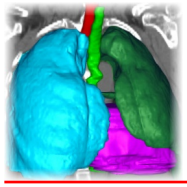


Contouring Guidelines

- Use RTOG 1106 contouring atlas as guideline
- Use standardized naming conventions (Santanam et al, IJROBP 2012)
- Included OARs
  - Serial-like organs (allow partially contouring)
    - Spinal cord (SpinalCord)
    - Esophagus (Esophagus)
  - Parallel-like organs (contour the whole organ)
    - Heart (Heart)
    - Lungs (Lung_L, Lung_R)
- Additional details: http://autocontouringchallenge.org



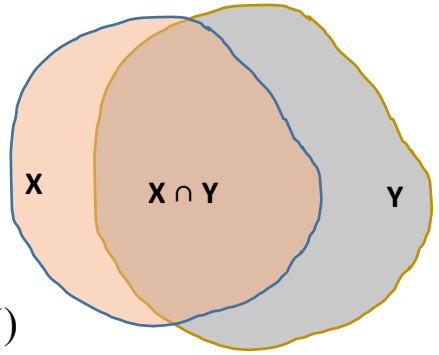




Evaluation Criteria

- Dice coefficient
  - A measure of relative overlap

$$D = \frac{2|X \cap Y|}{|X| + |Y|}$$



- Mean surface distance
  - Measure average distance of two contours in millimeters

$$\vec{d}_{avg}(X,Y) = \frac{1}{|X|}\sum_{x \in X} \min_{y \in Y} d(x,y) \quad d_{avg}(X,Y) = \frac{\vec{d}_{avg}(X,Y) + \vec{d}_{avg}(Y,X)}{2}$$

- Hausdorff distance (95% Hausdorff distance)
  - Measure the 95% distance of all points in one contour to the other in millimeters

$$\vec{d}_{H95}(X,Y) = K_{95}\left(\min_{y \in Y} d(x,y)\right) \quad d_{H95}(X,Y) = \frac{\vec{d}_{H95}(X,Y) + \vec{d}_{H95}(Y,X)}{2}$$

$K_{95}(\cdot)$ is the 95th percentile

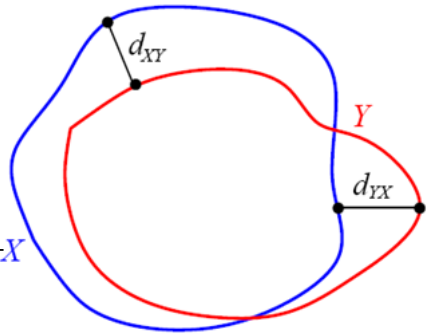
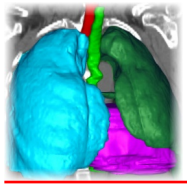
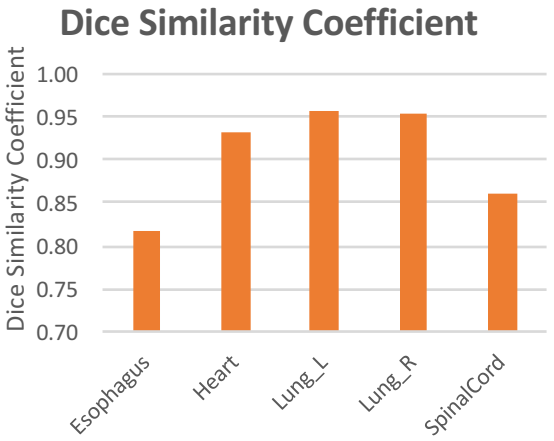


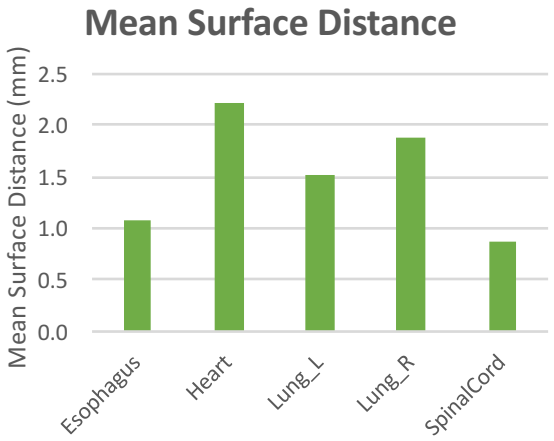
Image: wikipedia

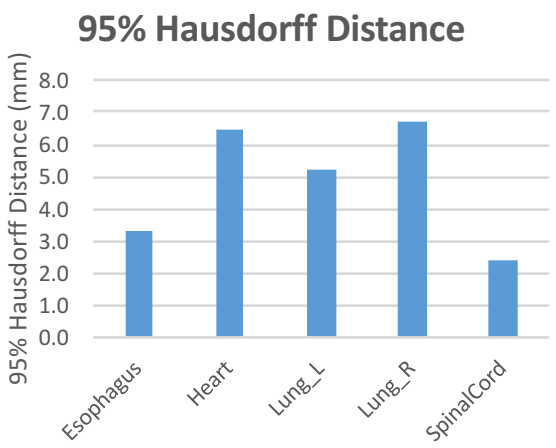


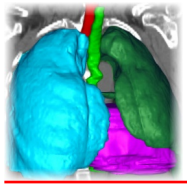
Evaluation Criteria (cont.)

- Final score takes into account observer contouring variability
  - Normal tissue contouring uncertainty could be as large as 2 cm (*IAEA Human Health Series No. 31, 2016*)
- 3 cases were manually contoured by multiple observers
- Mean score of these observers used as reference score (R)





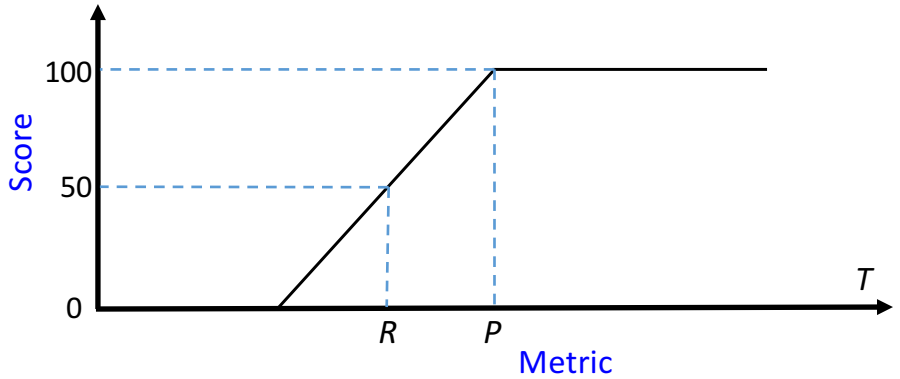




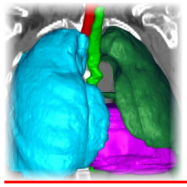
Evaluation Criteria (cont.)

- Perfect measure (P): Dice = 1, MSD/HD95=0
- Normalization of the score (T) for each metric, each structure

$$Score = max\left(50 + \frac{T - R}{P - R} * 50\ , 0\right)$$

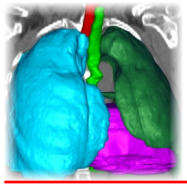


  - Score = 100; perfect segmentation
  - Score = 50; equivalent to average inter-observer reference
  - Score =0; below the reference by more than the perfect score above the reference
- Final score is the average of normalized scores of all metrics/structures



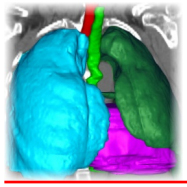
Timeline

- May 19, 2017: Release of training data
- Jun 21, 2017: Release of off-site test data
- Jul 17, 2017: Off-site test closed
- Jul 30 - Aug 03, 2017: AAPM Annual Meeting
  - Aug 02, 2017: Live challenge at AAPM
  - Aug 03, 2017: Challenge symposium at AAPM
- Aug 04, 2017: Online leader board available
- The challenge is never ended…



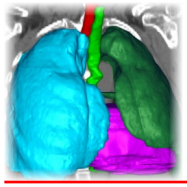
Participation Statistics

- Registrants: 110
- Offsite submissions: 31 from 11 participants (teams)



Off-site Competition Results

Results					
#	User	Entries	Date of Last Entry	Team Name	Difference ▲
1	elehanx	1	07/18/17		55.201336 (1)
2	paulaljabar	2	07/18/17		54.194330 (2)
3	xuefeng	2	07/18/17	uva-dl	52.617371 (3)
4	Bruno	4	07/18/17		46.412347 (4)
5	jwu	5	07/18/17		43.632788 (5)
6	aapm_qing	1	07/18/17	UVa_Radiology	43.218407 (6)
7	zamdborl	3	07/18/17		42.106869 (7)
8	brentvdh	4	07/18/17		36.133697 (8)
9	radonc	4	07/28/17		35.477495 (9)
10	ZhiweiZhai	1	07/18/17	ZhaiLUMC	31.273569 (10)



Presentations from Off-site Winners

- Xiao Han (Elekta Inc.)
  Automatic Thoracic CT Image Segmentation using Deep Convolutional Neural Networks

- Xue Feng (University of Virginia)
  A 3D UNet based thoracic segmentation framework using cropped images

- Bruno Oliveira (University of Minho)
  Automatic Multi-organ Segmentation in 3D Computed Tomography

- Paul Aljabar (Mirada Medical Ltd.)
  Circumscriptio ex machina: a step-change for auto-contouring