GPU TECHNOLOGY CONFERENCE
April 4-7, 2016 | Silicon Valley

GPU POWERED SOLUTIONS IN THE SECOND KAGGLE DATA SCIENCE BOWL

PRESENTED BY
NVIDIA.

SECOND ANNUAL DATA SCIENCE BOWL

Massive online data science contest

Mon 14 Dec 2015 – Mon 14 Mar 2016

192 teams, 293 data scientists finished

$200,000 prize fund (top 3 teams)

DATA SCIENCE BOWL

Passion. Curiosity. Purpose.

Booz | Allen | Hamilton

kaggle

AGENDA

Competition overview

The winning solution

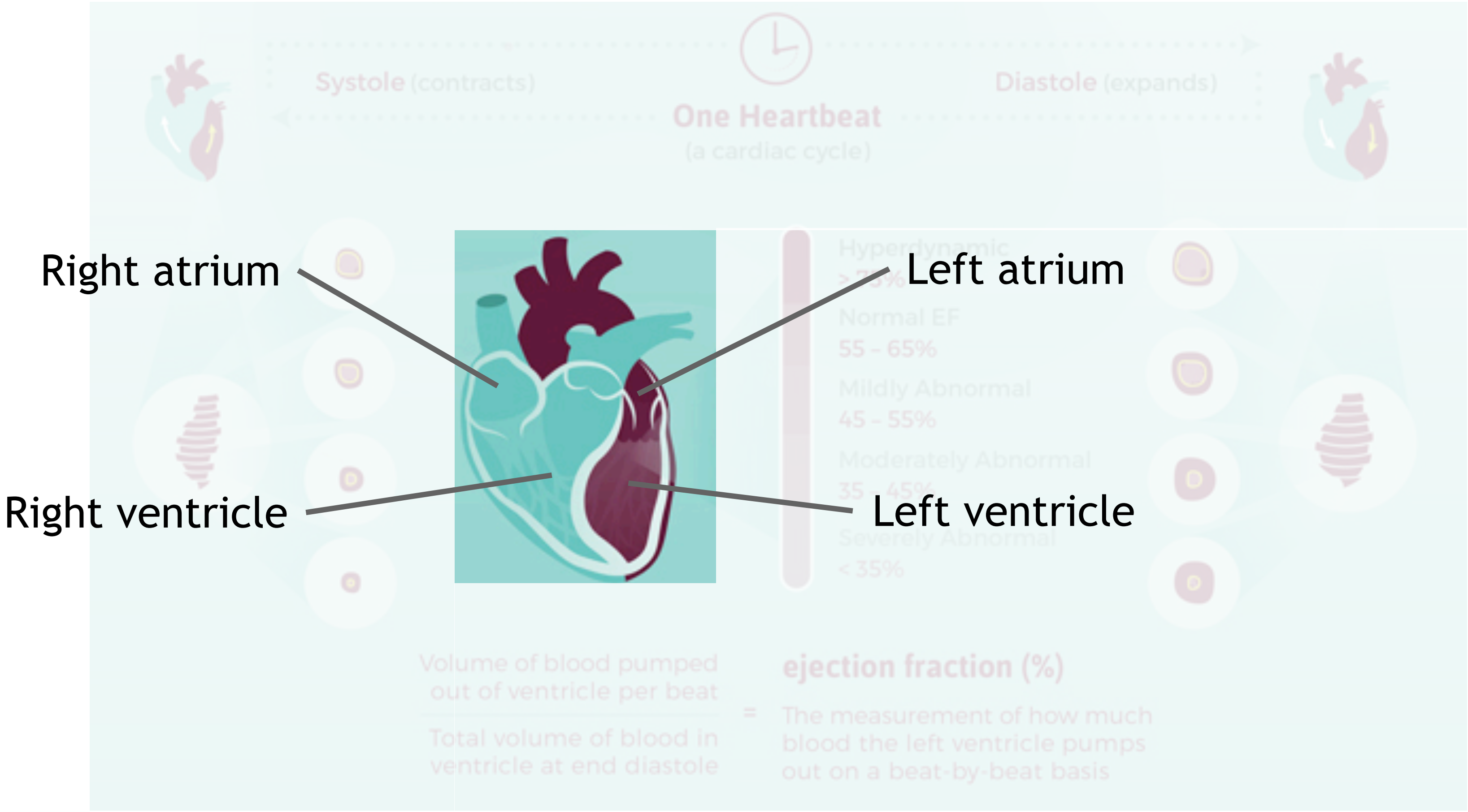
Presentation from competition organizers

Other successful approaches

COMPETITION OVERVIEW
"TRANSFORMING HOW WE DIAGNOSE HEART DISEASE"

COMPETITION ANATOMY

“The only unit of time that matters is heartbeats.” – Paul Ford



"…left ventricular ejection fraction (LVEF) is probably one of the single most important numerical values determined on an adult patient with heart disease"

"…low LVEF predicts in the patients that survive a heart attack are much more likely to die in the course of the next year than patients with normal LVEF"

"There are also diseases that cause a heart to enlarge before the LVEF changes… Thus, measurement of both LV volumes and the LVEF provide complimentary information that helps in the diagnosis of many patients with heart disease."

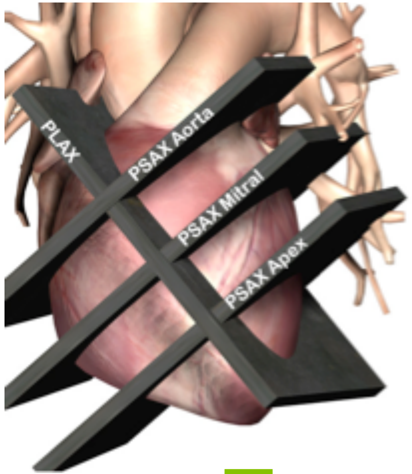
— Source: https://www.kaggle.com/c/second-annual-data-science-bowl/forums/t/19839/a-medical-perspective-on-the-quality-of-the-left-ventricular-volume-and

Andrew Arai, MD

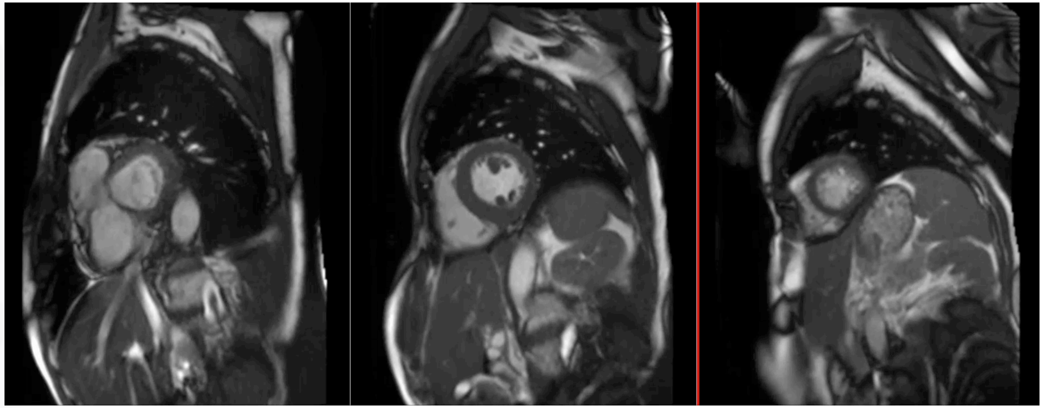
Cardiologist,
National Institutes of Health (NIH)

MEASURING EJECTION FRACTION

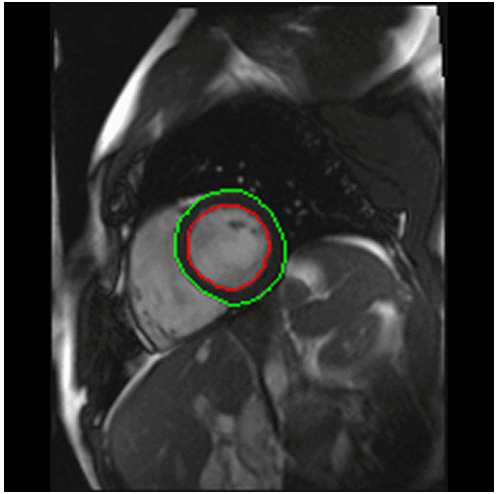
Magnetic Resonance Imaging (MRI) and expert annotation



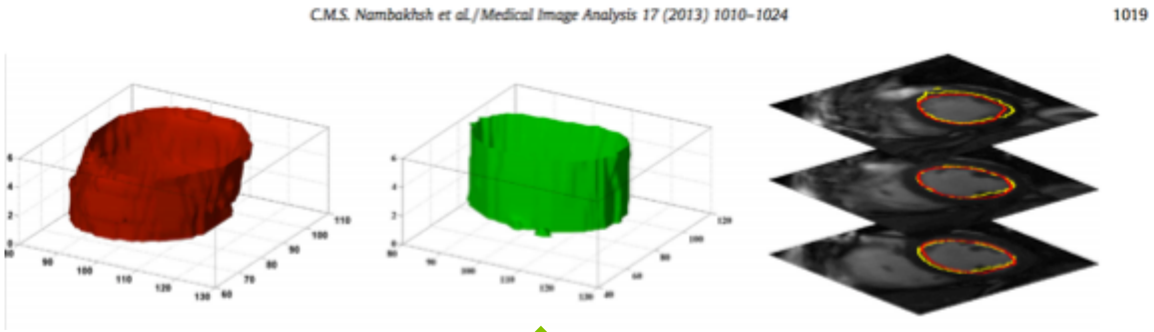
MRI imaging



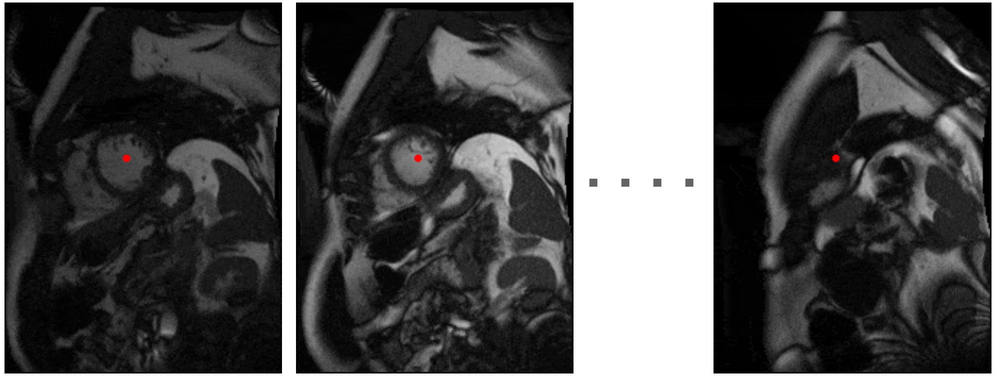
Manual annotation



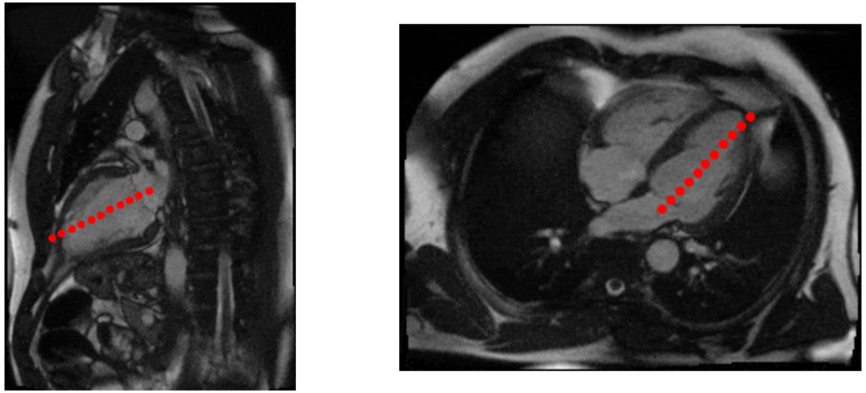
Software volume estimate



COMPETITION DATA



Short Axis (SAX) images:
varying # and locations of slices per patient, 30 timesteps



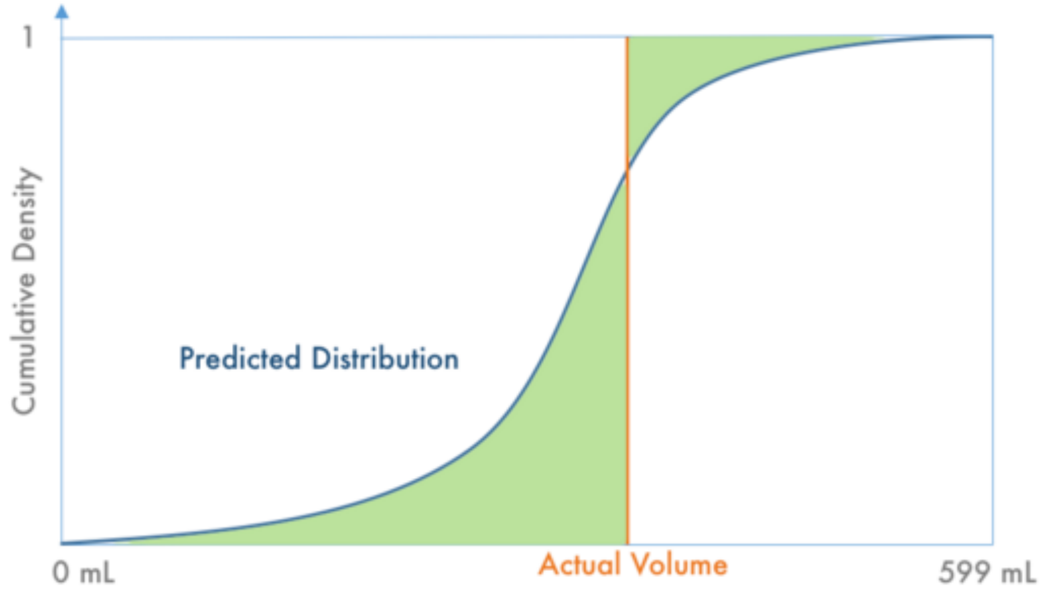
Long Axis (LAX) images: not all patients

- **DICOM file format:**
  - 16-bit images
  - Metadata
    - Patient Age
    - Patient Sex
    - Pixel Spacing
    - Slice Location (not all patients)
    - Various imaging geometry parameters relative to patient
    - Various imaging parameters
- **Two labels for whole patient study:**
  - Systole volume
  - Diastole volume

OBJECTIVE FUNCTION

Continuous Ranked Probability Score

$$C = \frac{1}{600N} \sum_{m=1}^{N} \sum_{n=0}^{599} (P(y \leq n) - H(n - V_m))^2$$



COMPETITION TIMELINE

Model upload deadline

Phase 1:

500 Training patients
200 test patients (no ground truth)
Hand labeling allowed for training data

Phase 2:

500 + 200 training patients
440 new test patients
No model modifications

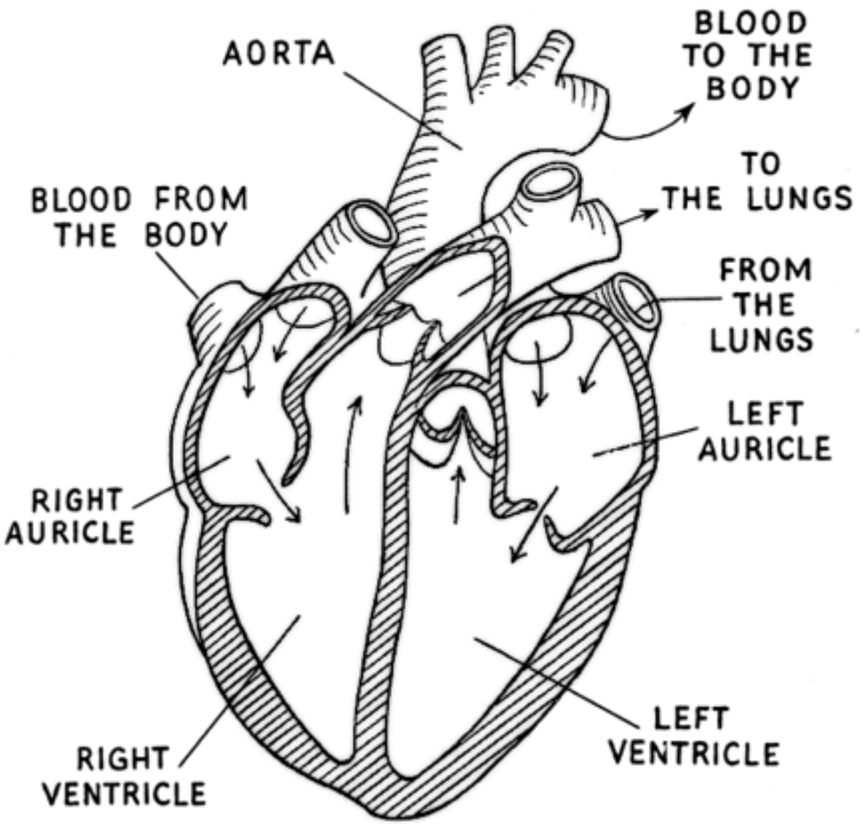
14 Dec — 7 Mar — 14 Mar

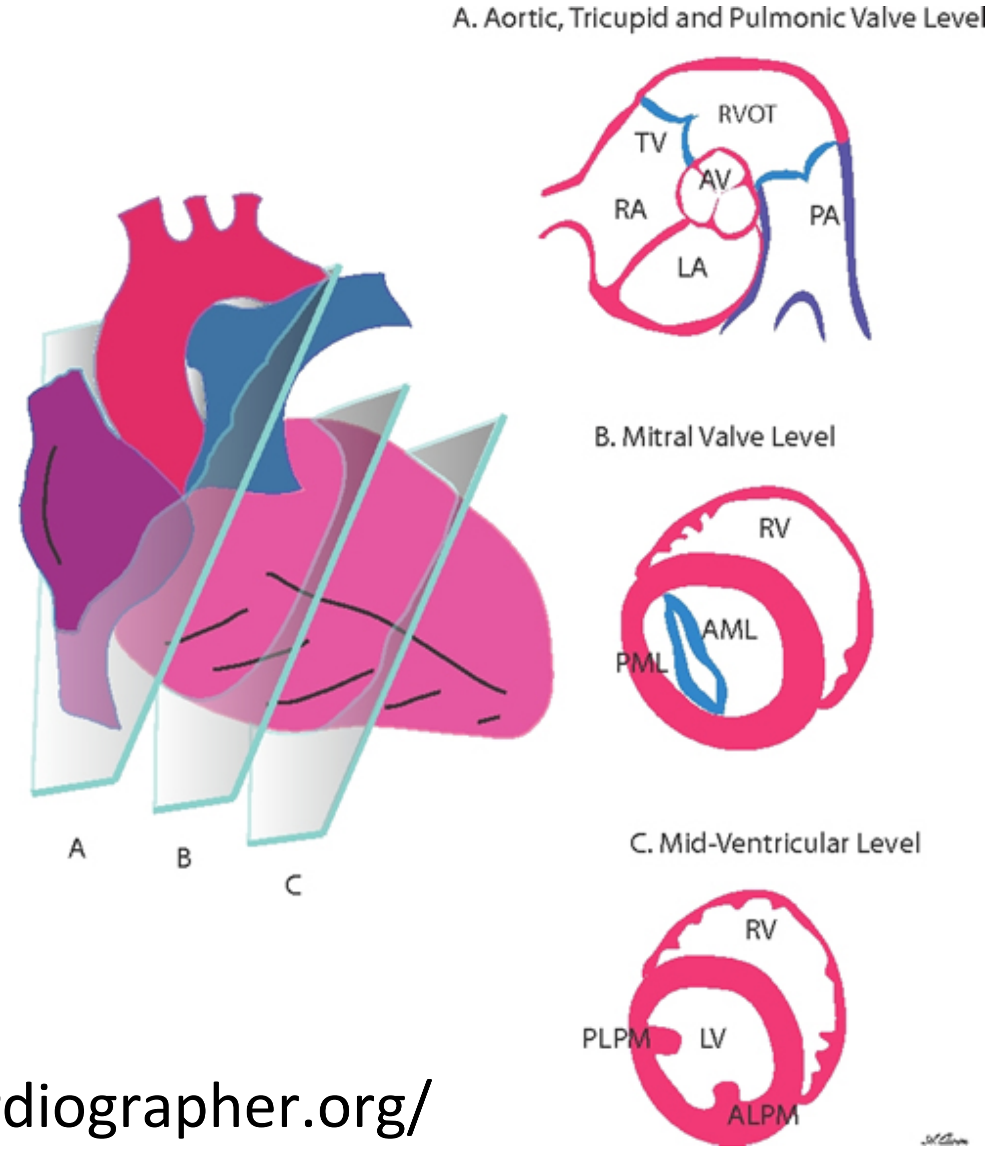
THE WINNING SOLUTION

TEAM: TENCIA & WOSHIALEX

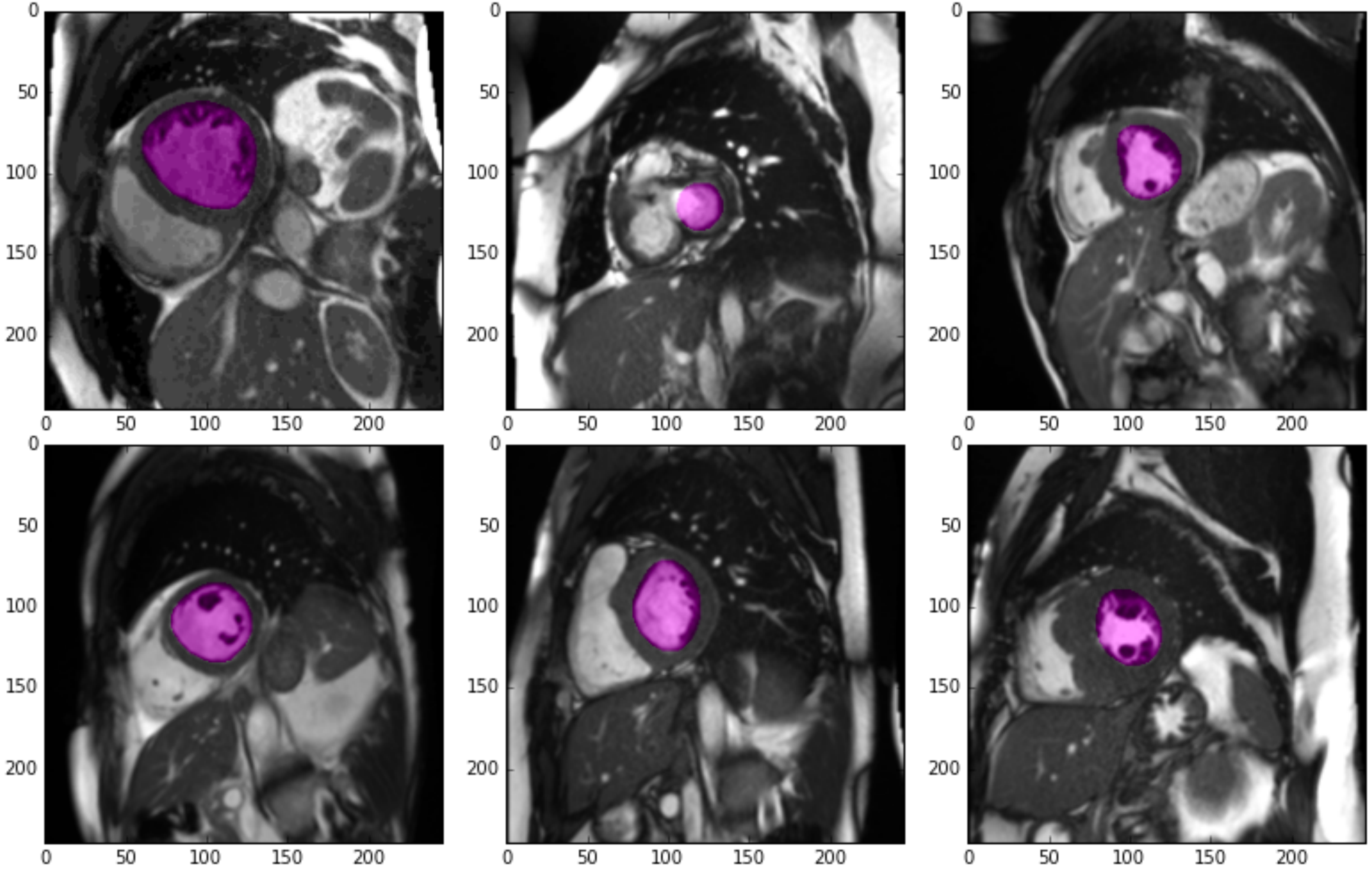
Heart Left Ventricle Volumes from MRI images

Tencia Lee & Qi Liu
April 2016





wpclipart.com http://echocardiographer.org/

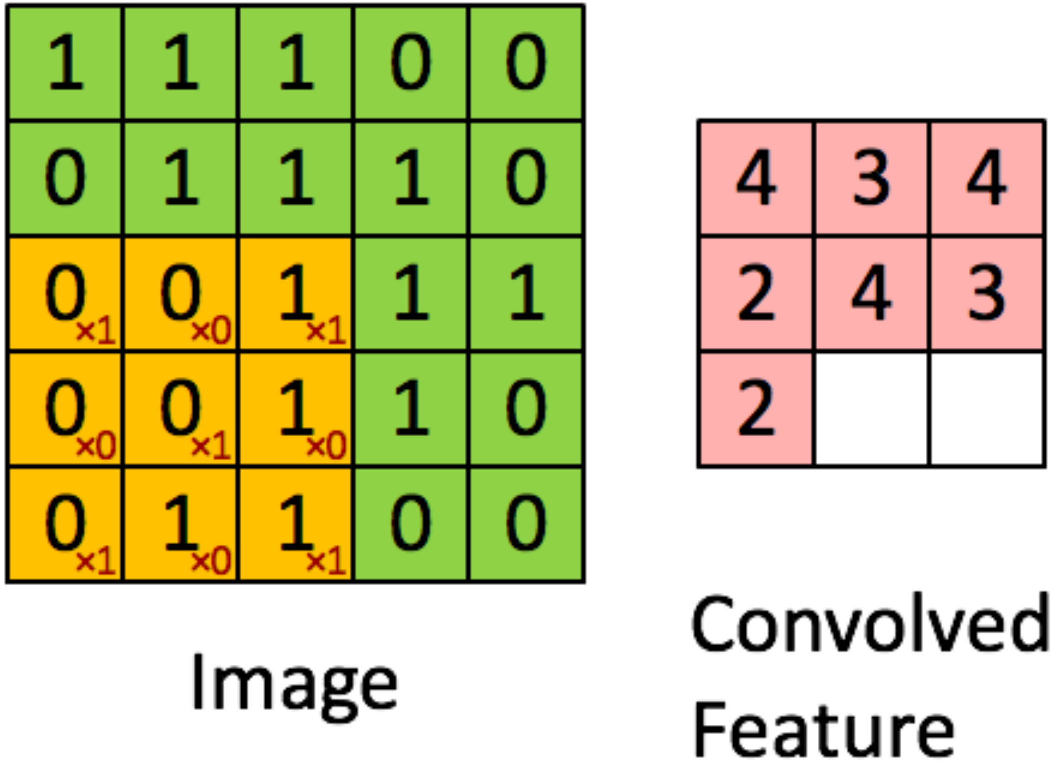


kaggle.com

The challenges

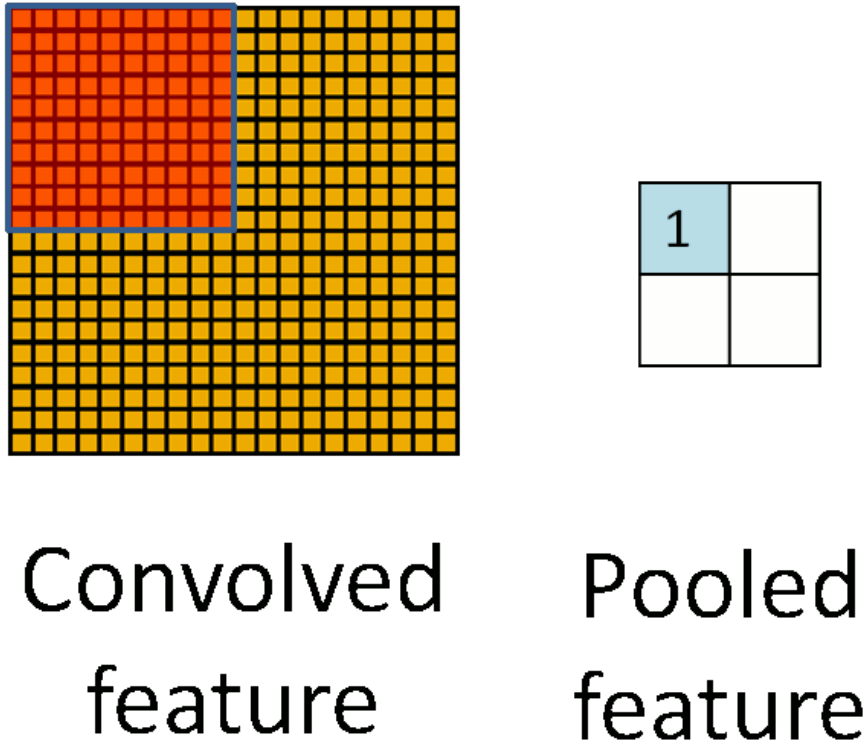
- Dirty data: mislabeled images, badly organized directories
- Only 700 images in segmentation training set
- 150,000 images to be segmented (500 training patients, ~300 each), coming from a completely different set of MRIs
- Some were dark, partly obscured, had odd artifacts along the edges, or significantly different from the segmentation set
- Ground truths are human-segmented and can be wrong

Convolution operation on image

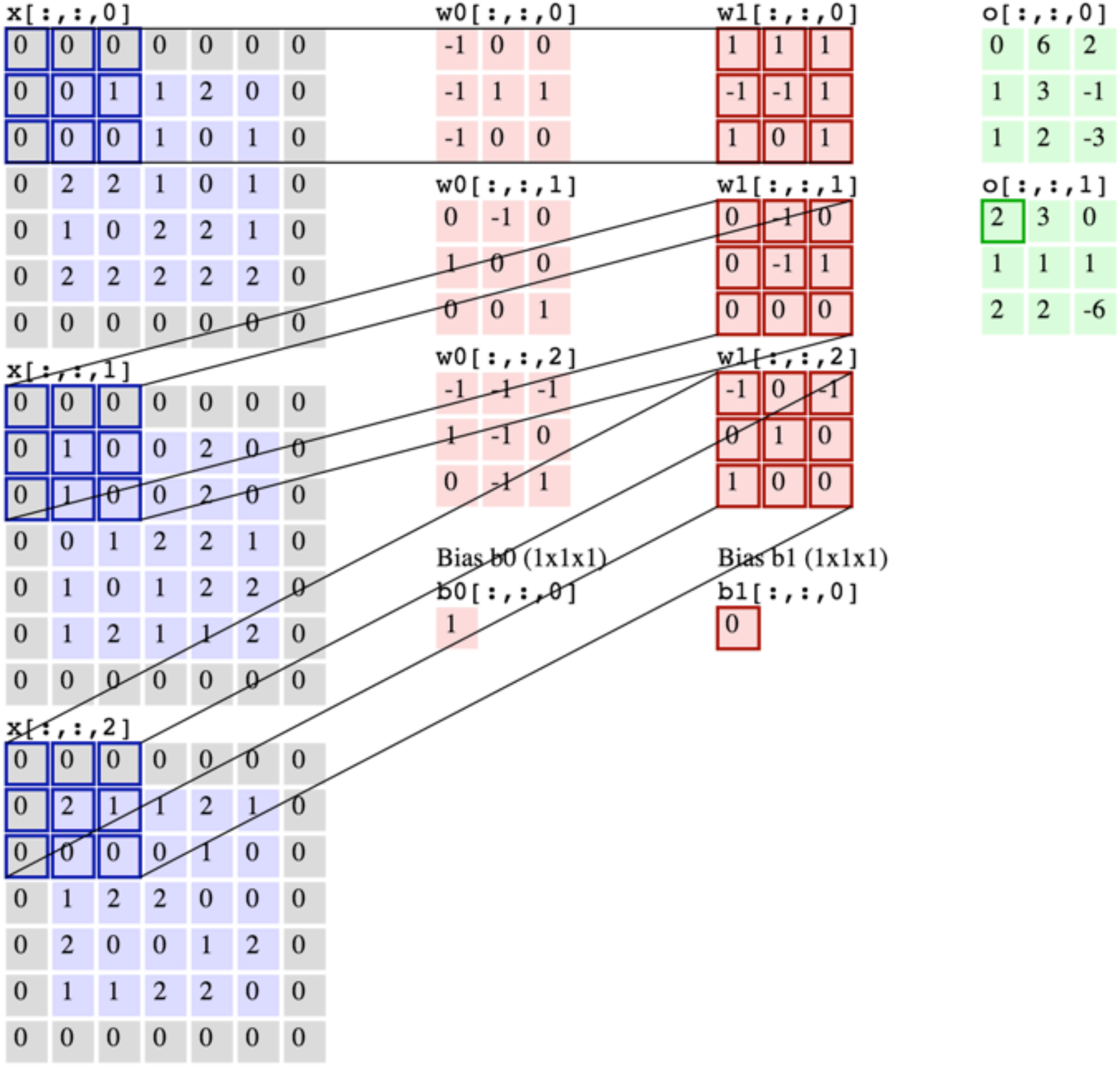


http://ufldl.stanford.edu/tutorial/

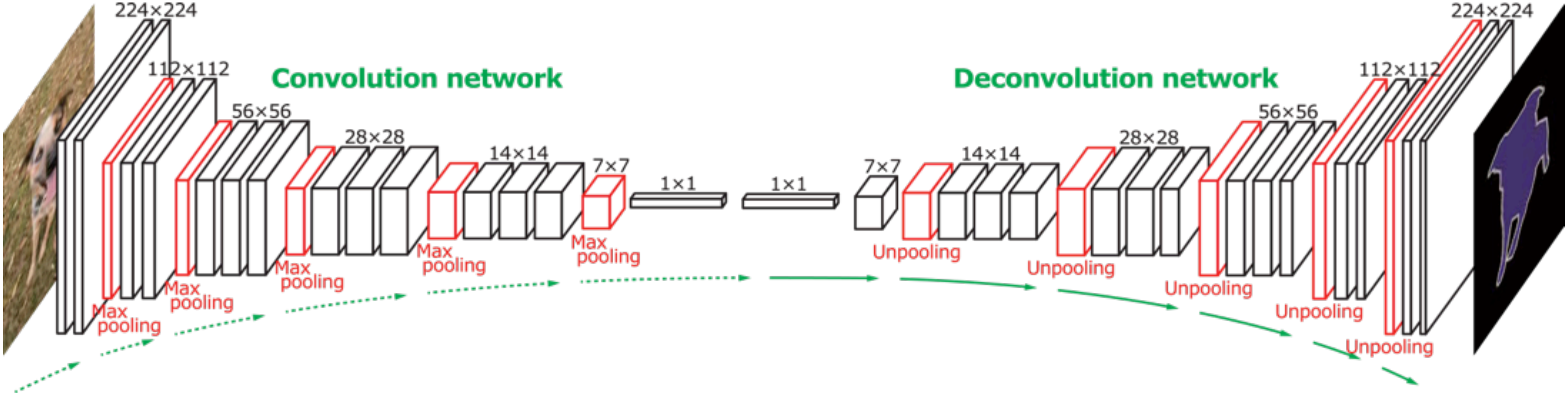
Pooling operation on convolved features



http://ufldl.stanford.edu/tutorial/



cs231n.github.io



http://cvlab.postech.ac.kr/

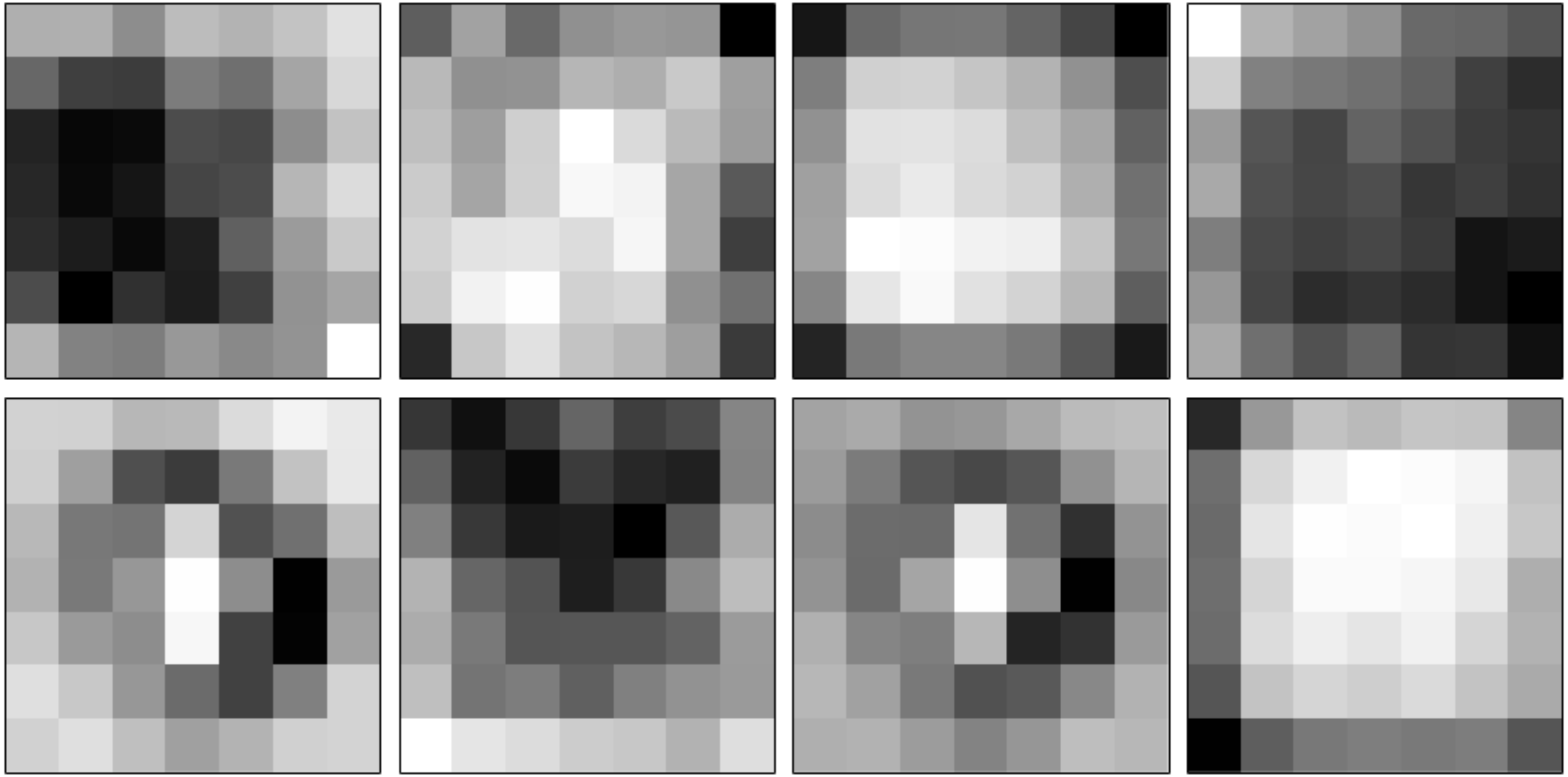
Layer Op / Type	# Filters / Pool / Upscale Factor	Filter Size	Padding	Output Shape
Input				(b, 1, 246, 246)
Conv + BN + ReLU	8	7	valid	(b, 8, 240, 240)
Conv + BN + ReLU	16	3	valid	(b, 16, 238, 238)
MaxPool	2			(b, 16, 119, 119)
Conv + BN + ReLU	32	3	valid	(b, 32, 117, 117)
MaxPool	2			(b, 32, 58, 58)
Conv + BN + ReLU	64	3	valid	(b, 64, 56, 56)
MaxPool	2			(b, 64, 28, 28)
Conv + BN + ReLU	64	3	valid	(b, 64, 26, 26)
Conv + BN + ReLU	64	3	full	(b, 64, 28, 28)
Upscale	2			(b, 64, 56, 56)
Conv + BN + ReLU	64	3	full	(b, 64, 58, 58)
Upscale	2			(b, 64, 116, 116)
Conv + BN + ReLU	32	7	full	(b, 32, 122, 122)
Upscale	2			(b, 32, 244, 244)
Conv + BN + ReLU	16	3	full	(b, 16, 246, 246)
Conv + BN + ReLU	8	7	valid	(b, 8, 240, 240)
Conv + sigmoid	1	7	full	(b, 1, 246, 246)

Sørensen-Dice Coefficient

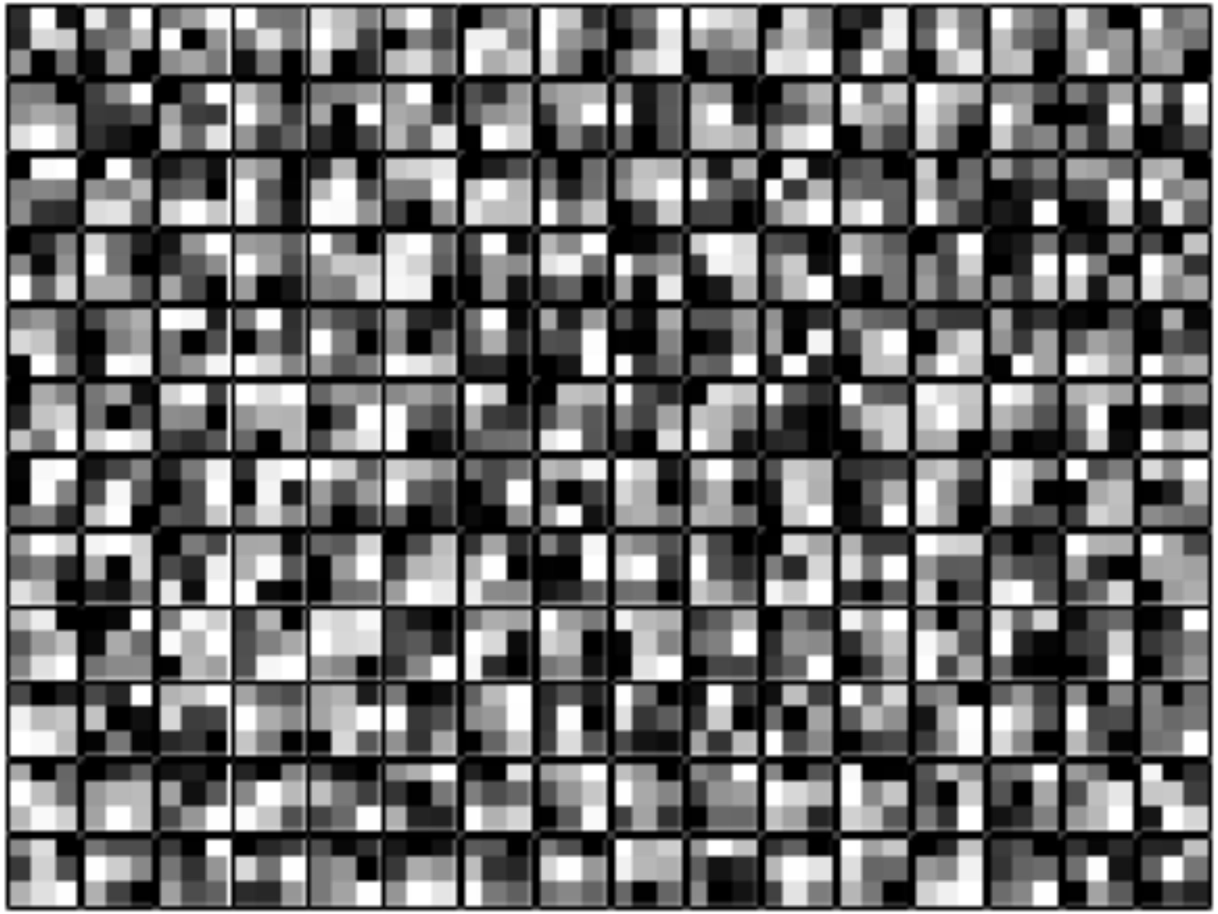
$$Loss = -\frac{s + 2\sum_{i,j} pred_{ij} \cdot target_{ij}}{s + \sum_{i,j}(pred_{ij} + target_{ij})}$$

- classes are very unbalanced
- 97% of pixels in input are not part of ventricle
- more robust than binary cross-entropy

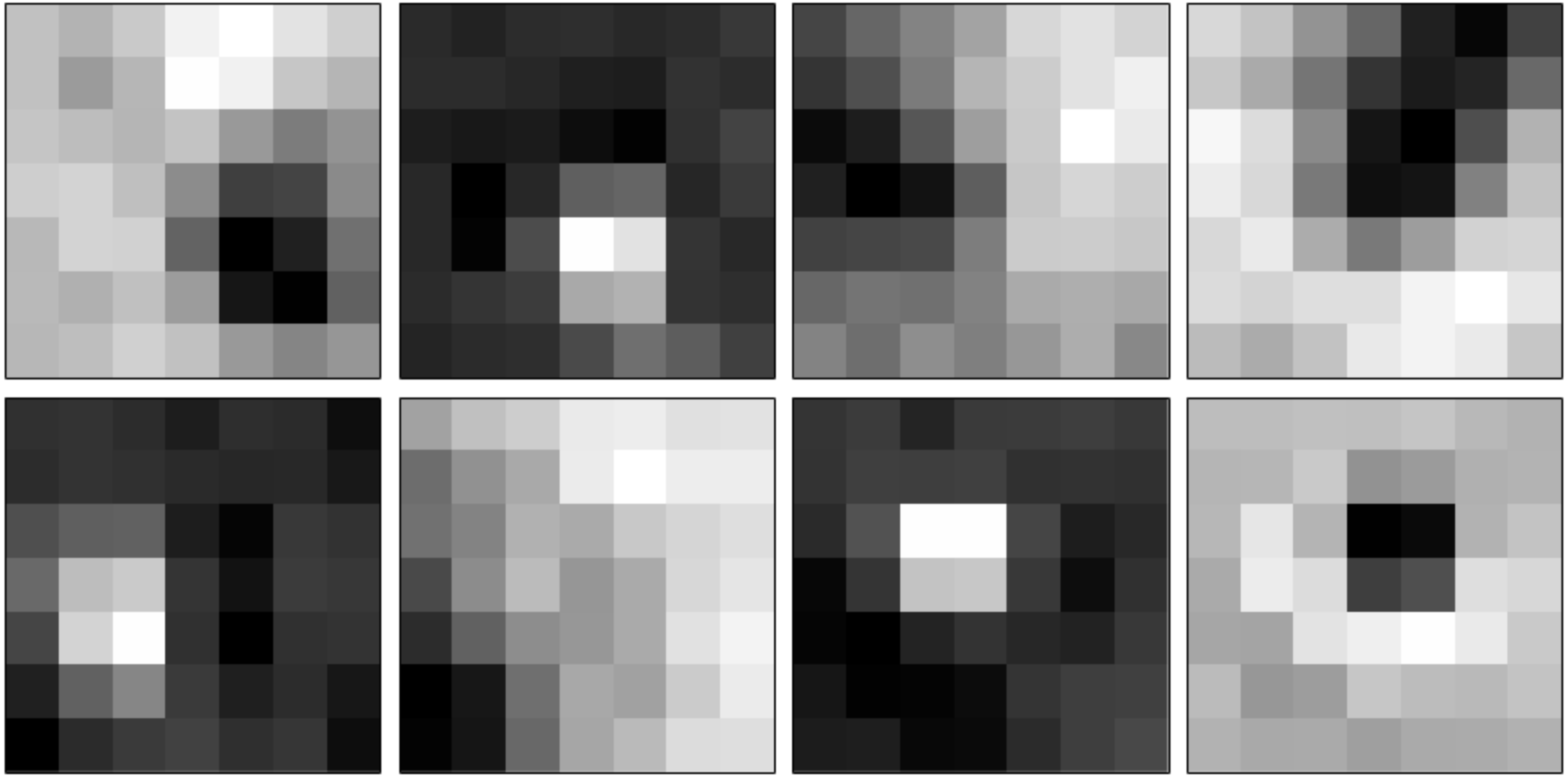
Top Layer Filter Weights

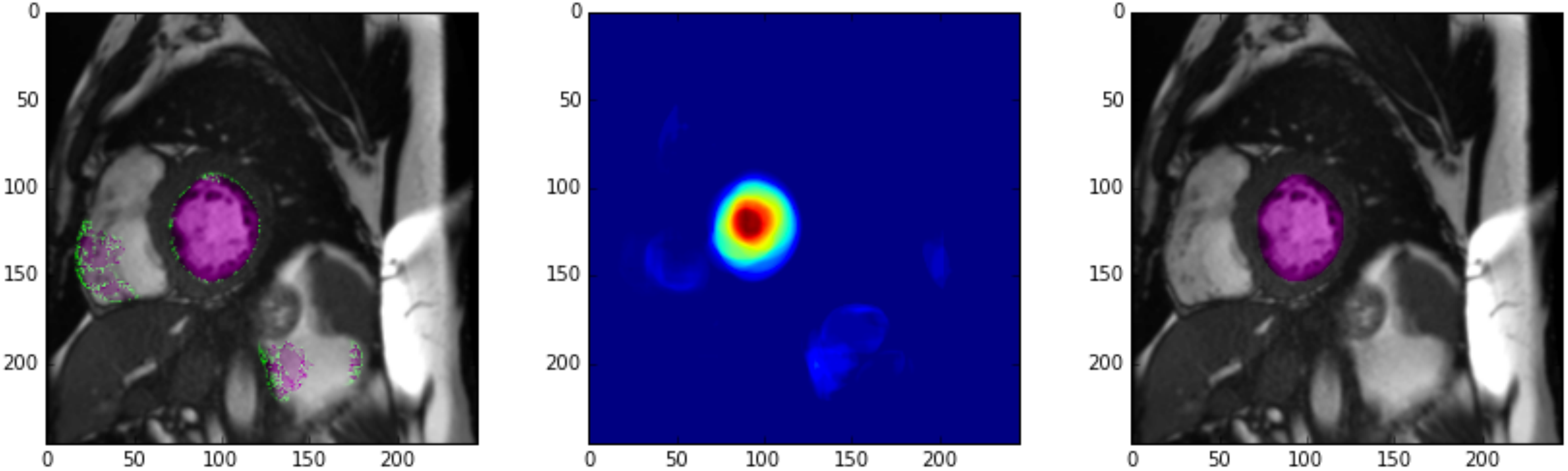


Middle Layer Filter Weights

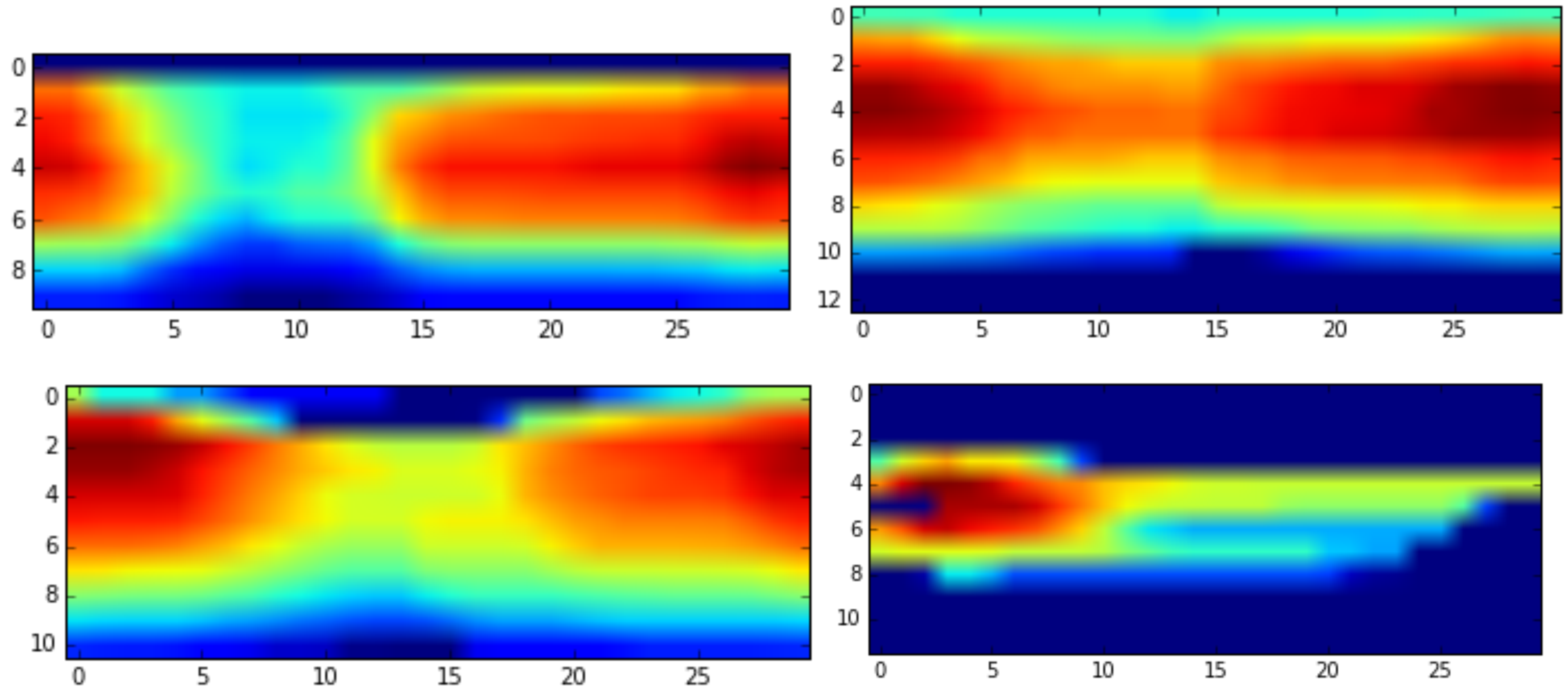


Bottom Layer Filter Weights




0
50
100
150
200
0
50
100
150
200
0
50
100
150
200
0
50
100
150
200
0
50
100
150
200
0
50
100
150
200

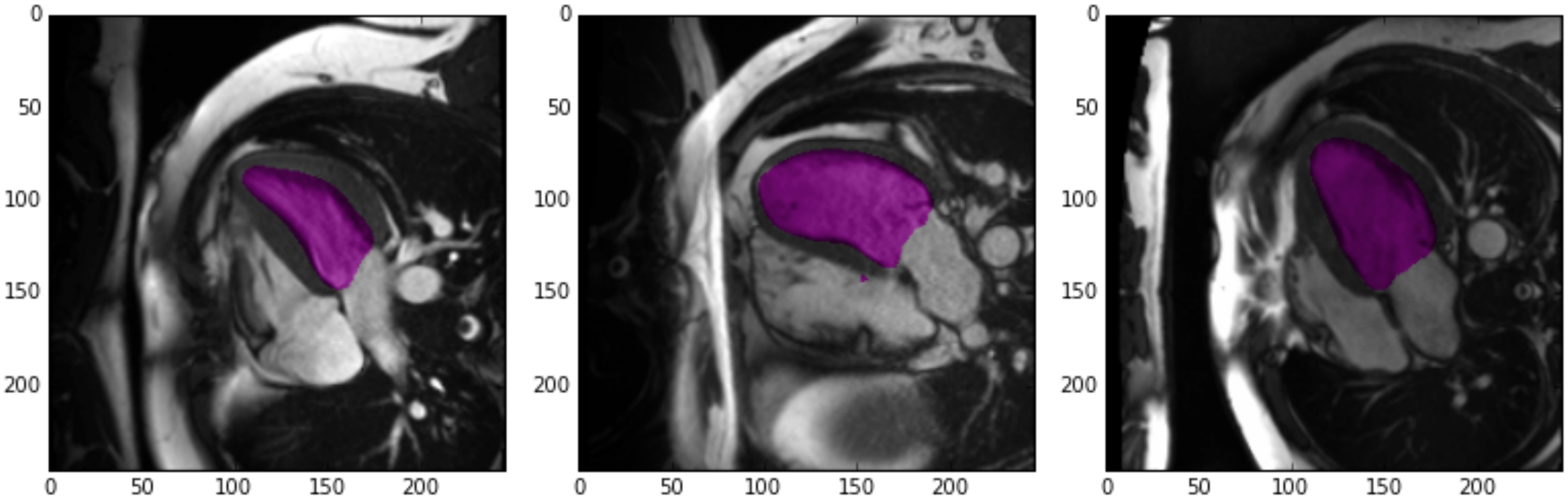
Heat maps - area, height x time

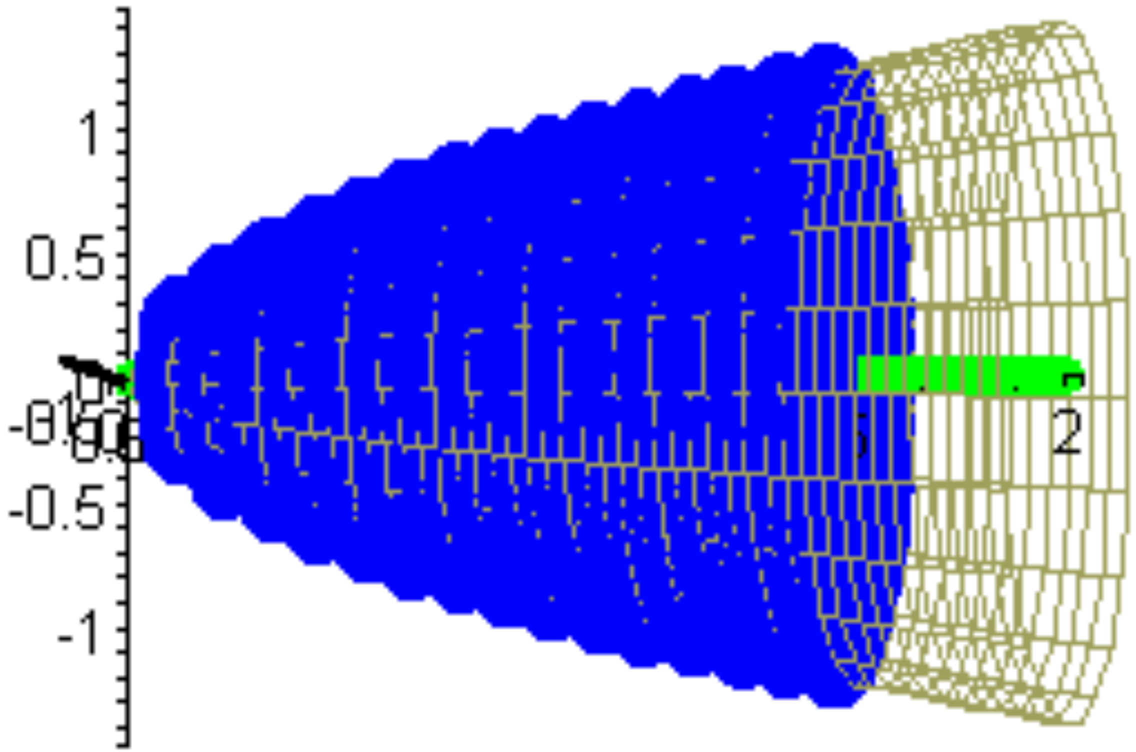


Other Models

- One-slice: segmentation net -> single slice area -> volume
- Age-sex prior: age and gender -> volume
- Four-chamber view:
    - hand-labeled 736 four-chamber view DICOMs
    - trained segmentation net to find cross-sectional area
    - calculated volume by rotating area around main axis

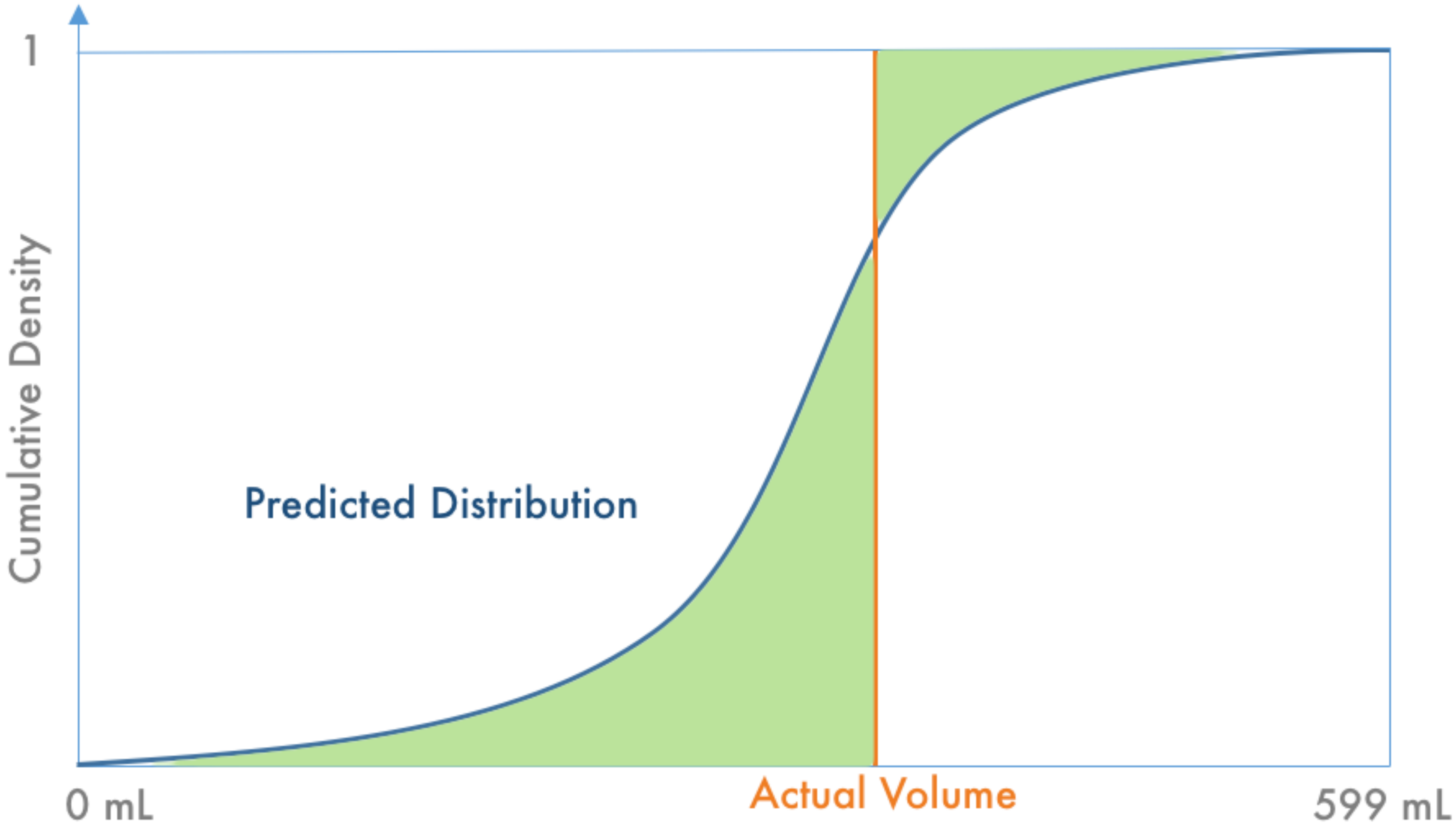
Four-chamber view model




1
0.5
-0.5
-1
2

Linear ensembling

- Very simple method for combining many CNN models as well as other models
- Optimized linear weights on each model to minimize CRPS score
- Filtered CNN models by whether all times have a certain # of nonzero areas
- When CNN fails, use 4-chamber + one-slice.
- When 4-chamber + one-slice fails, use age-sex model.


1
Cumulative Density
Predicted Distribution
0 mL
Actual Volume
599 mL

kaggle.com

Tools used

- Python
- Deep Learning: Theano, Lasagne
- Data handling: Fuel, HD5py
- Image processing: OpenCV, Scikit-image

2ND AND 3RD PLACE APPROACHES

2ND PLACE: TEAM KUNSTHART

Data Science Lab at Ghent University, Belgium

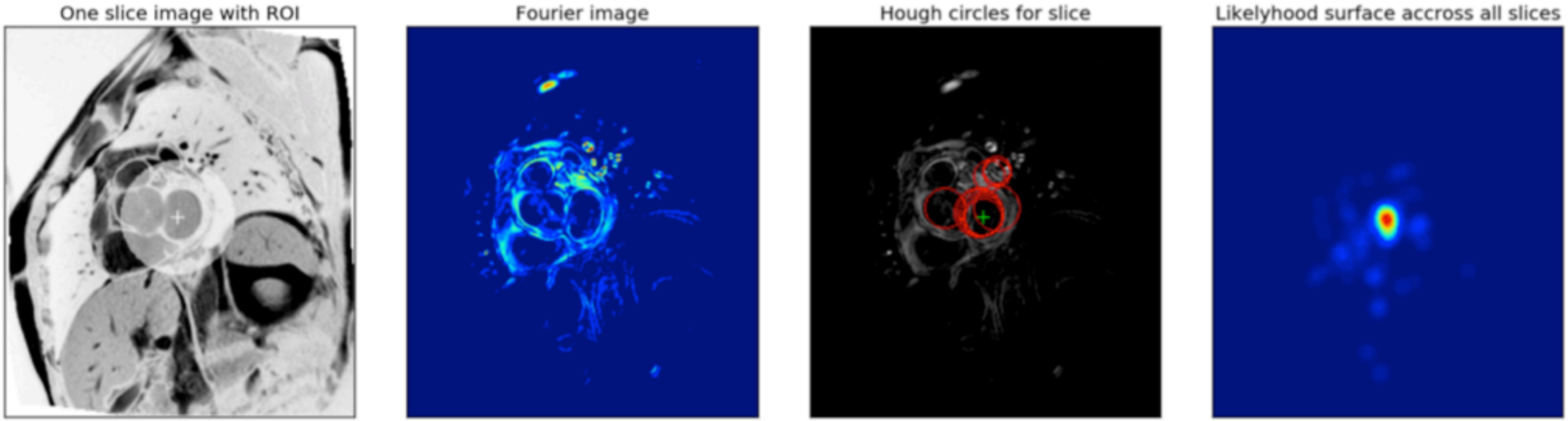
PhD students: Ira Korshunova, Jeroen Burms and Jonas Degrave

Professor Joni Dambre

3 members of Team “Deep Sea”, winners of the First Data Science Bowl

2ND PLACE: TEAM KUNSTHART

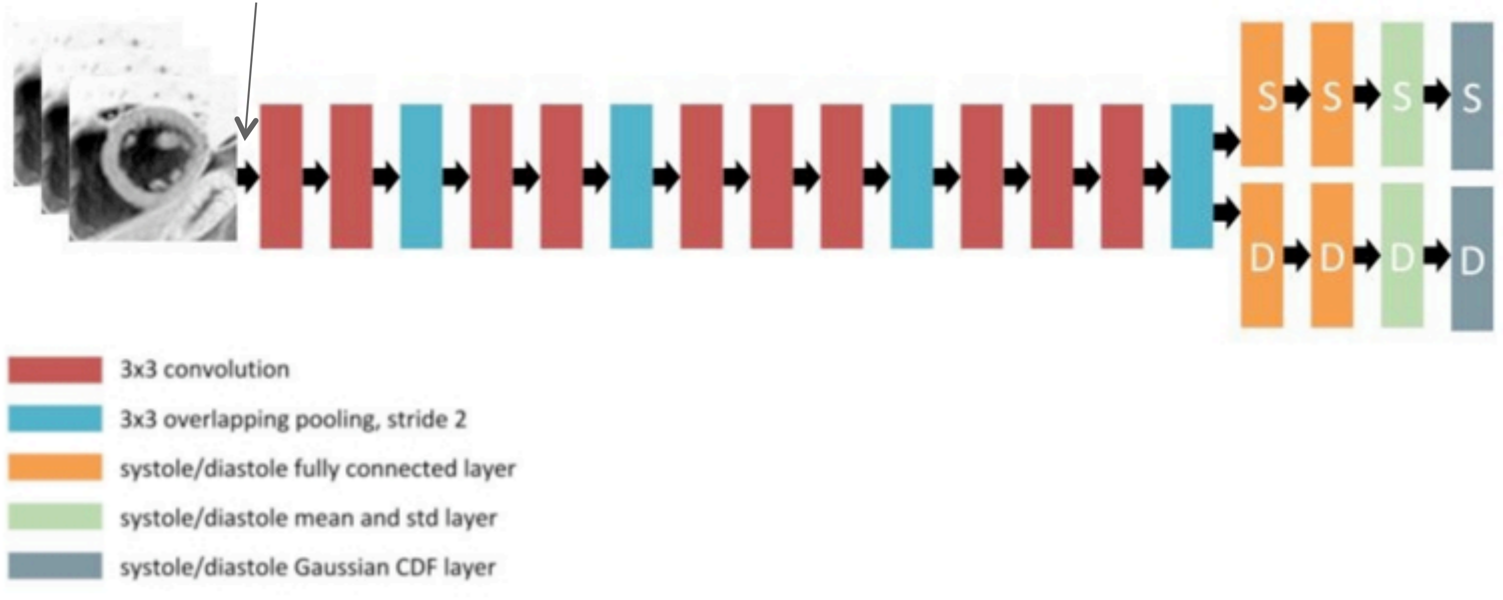
Stage 1: ROI extraction



2^{ND} PLACE: TEAM KUNSTHART

Stage 2: Single Slice Convolutional Neural Networks

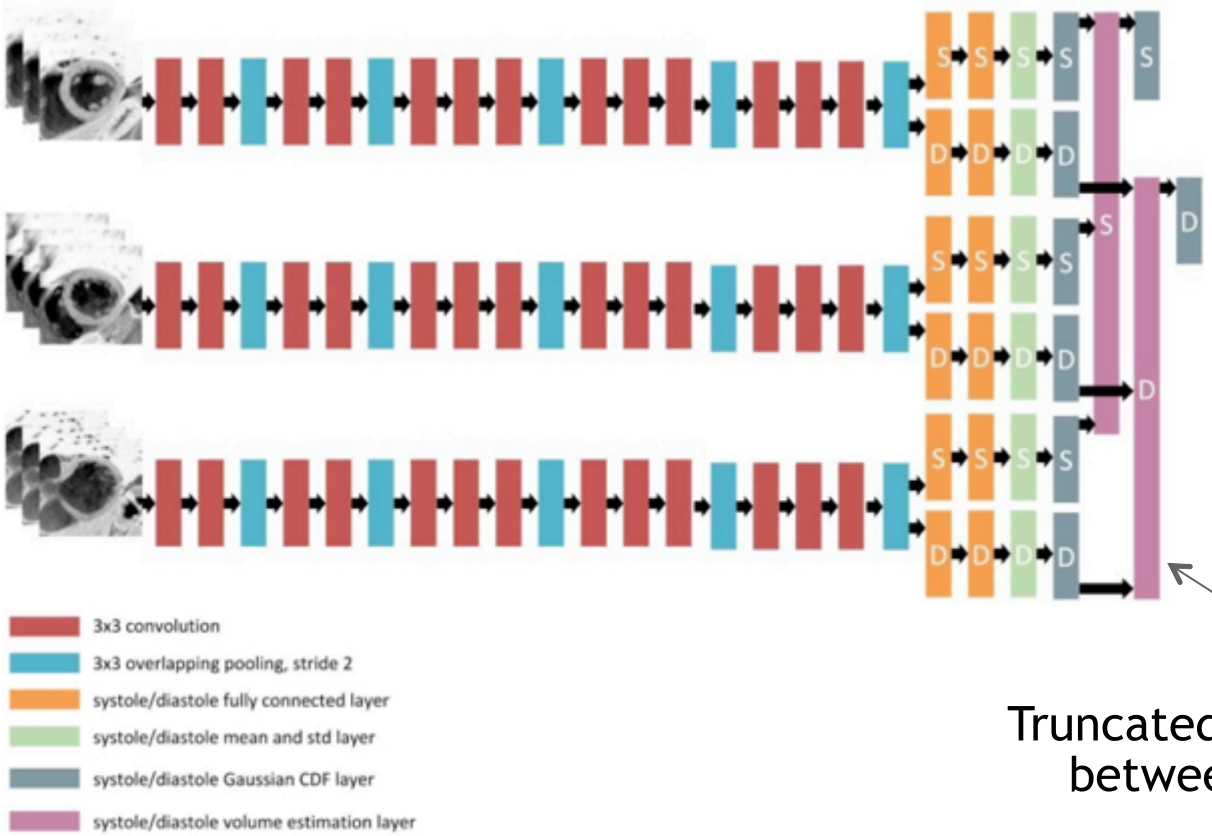
Train and test time augmentation

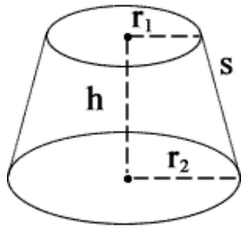


Multiple models trained for single SAX slices and 2-Ch and 4-Ch stacks

2ND PLACE: TEAM KUNSTHART

Stage 3: Patient Convolutional Neural Networks





Truncated cone volume estimate between consecutive slices

2^{ND} PLACE: TEAM KUNSTHART

Stage 4: Model ensembles

~250 total models trained

Error was dominated by small number of outliers

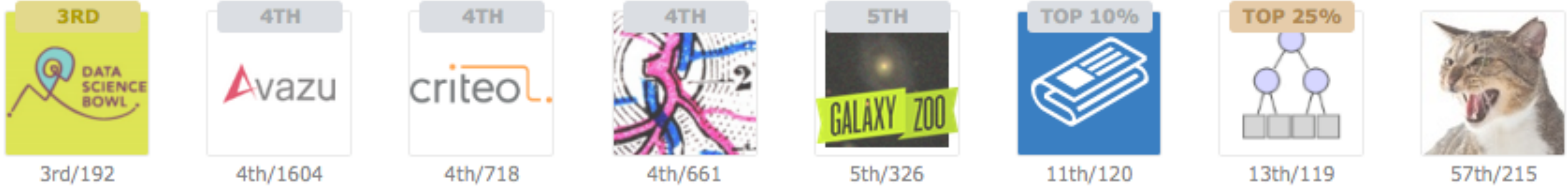
Setup framework so that each individual model could be selectively applied to each patient based on heuristics

Implemented two different ensembling strategies: ~75% patients received a ‘personalized’ ensemble

http://irakorshunova.github.io/2016/03/15/heart.html

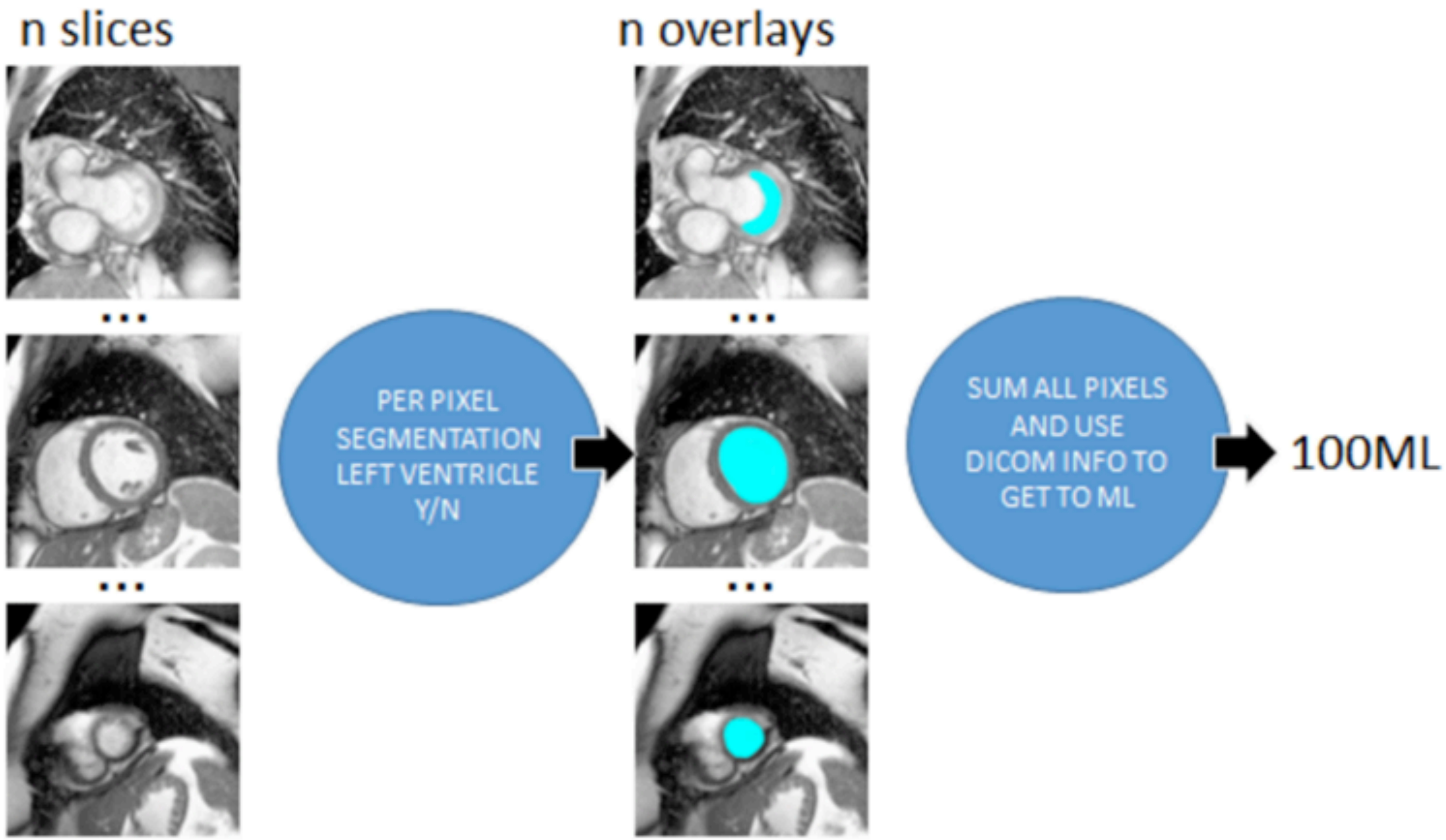
3RD PLACE: JULIAN DE WIT

Owner DWS Systemen, The Hague Area, Netherlands



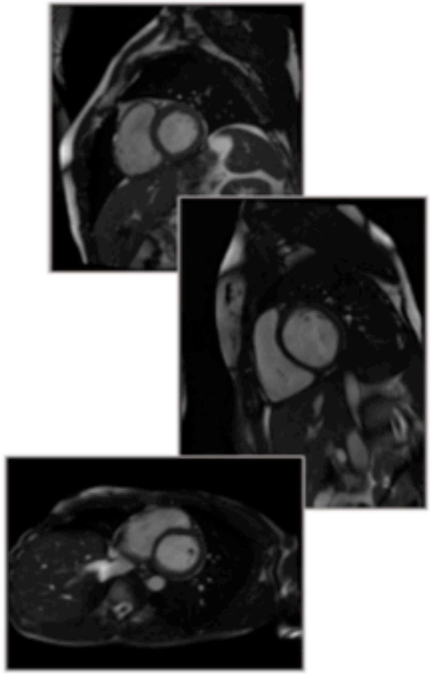
3RD PLACE: JULIAN DE WIT

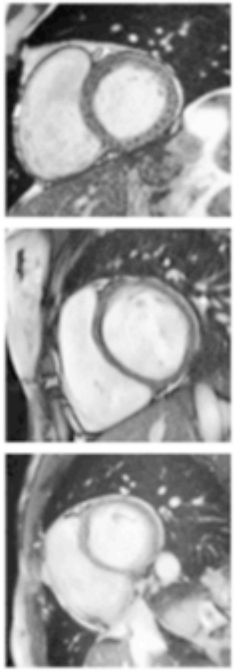
Idealized solution



3RD PLACE: JULIAN DE WIT

Stage 1: Pre-processing

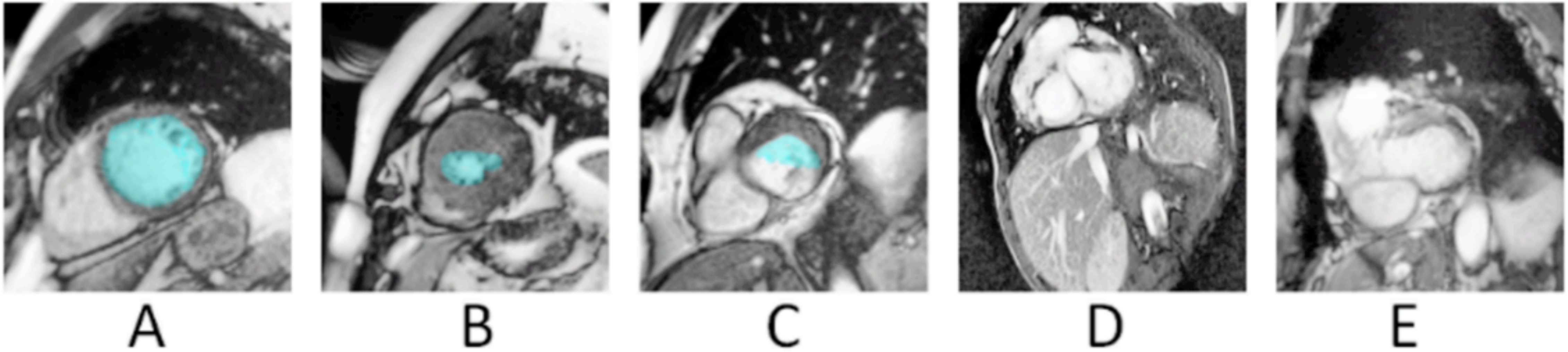




Pixel scaling
Contrast stretching
Crop to 180x180

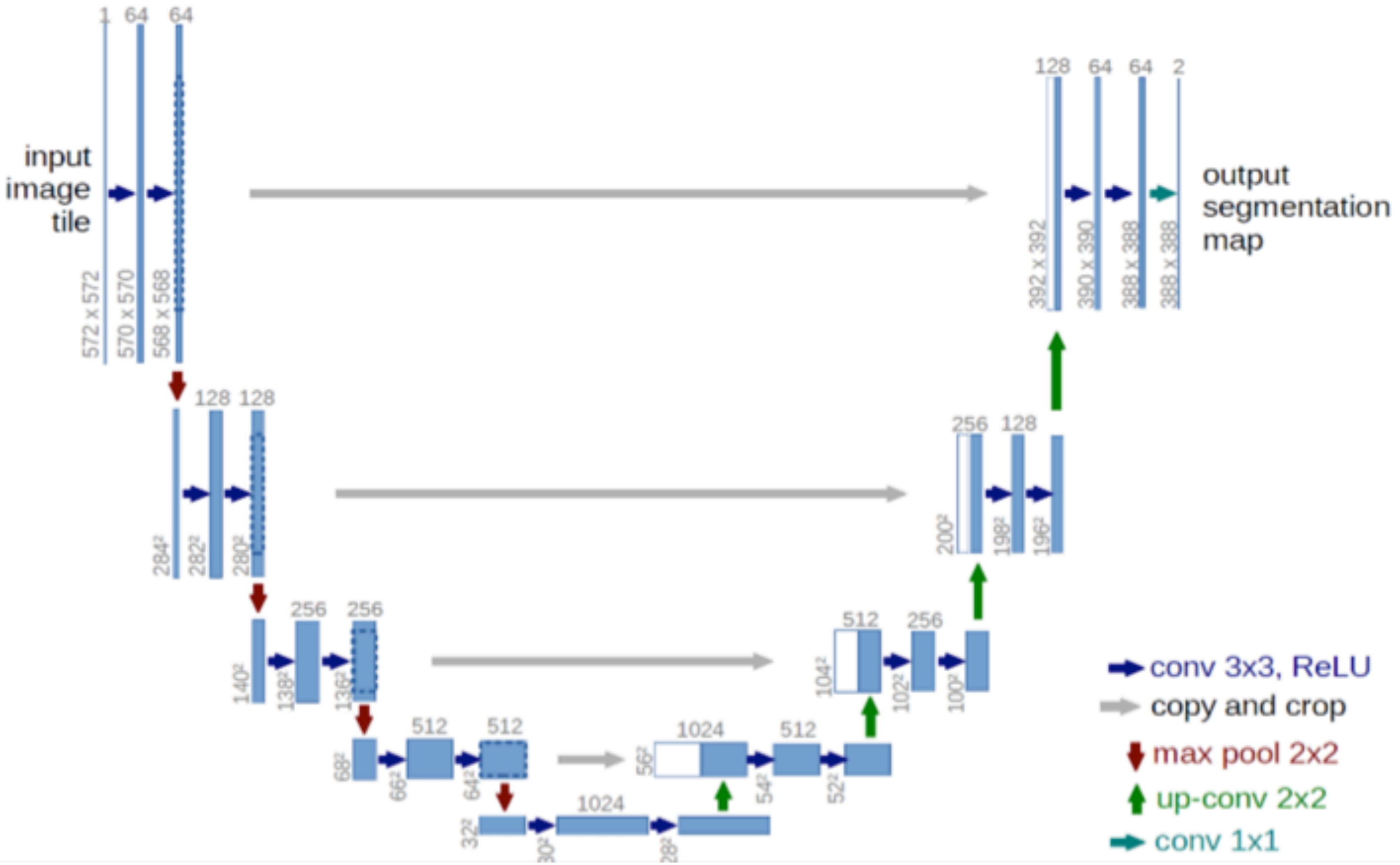
3RD PLACE: JULIAN DE WIT

Stage 2: Manual labeling



3RD PLACE: JULIAN DE WIT

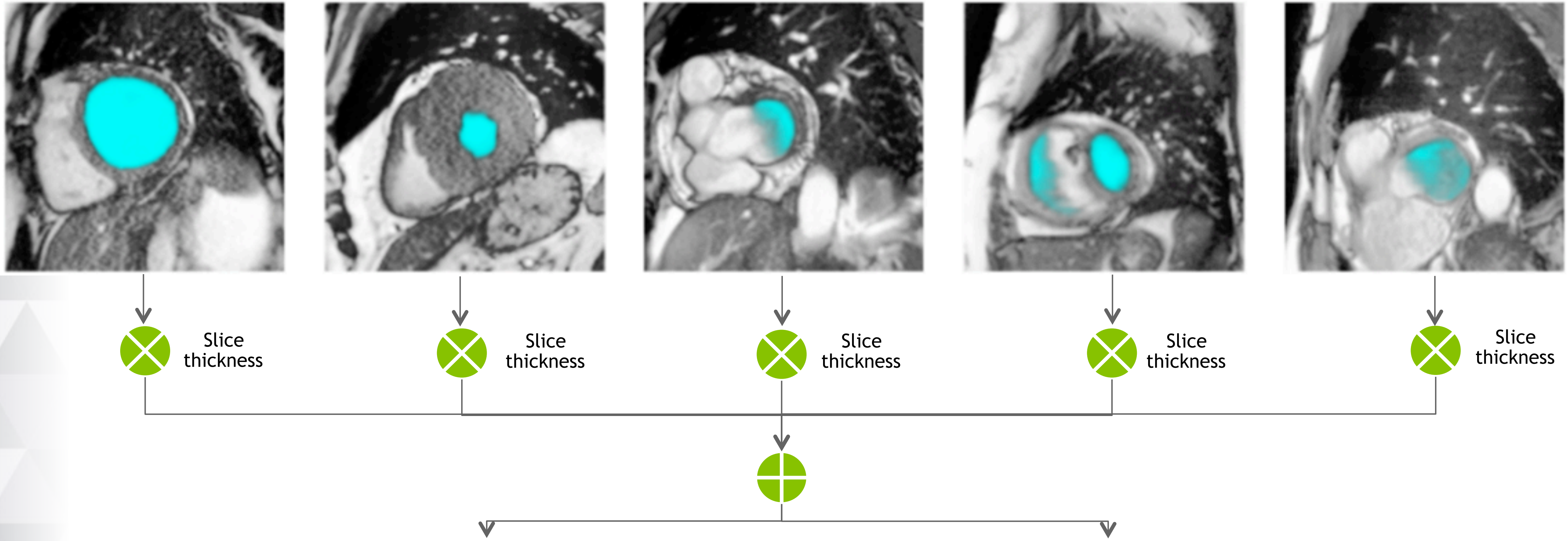
Stage 3: U-net segmentation architecture



Ronneberger, Fischer, Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv:1505.04597

3RD PLACE: JULIAN DE WIT

Stage 4: Integrating segmentations to volume estimates



$\min_{time}$ = systolic volume

$\max_{time}$ = diastolic volume

3RD PLACE: JULIAN DE WIT

Stage 5: Model calibration

Used gradient boosting regressor: raw volume estimates, segmented image features and metadata features

Regressed on the error in volume estimate rather than the volume

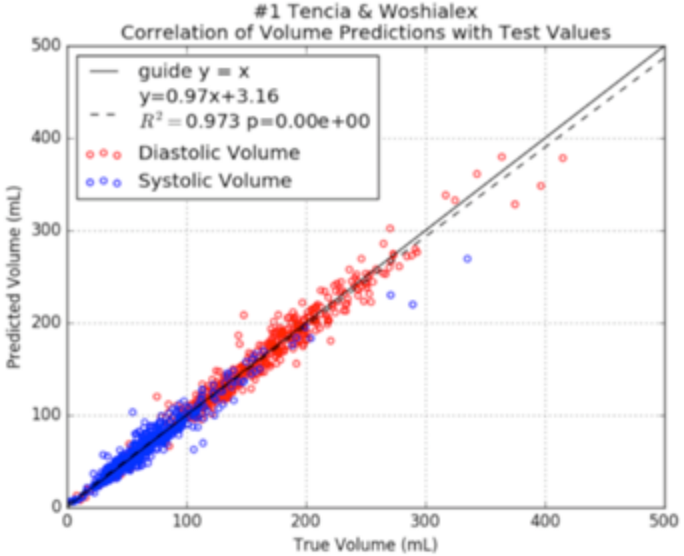
Used k-fold training to avoid overfitting

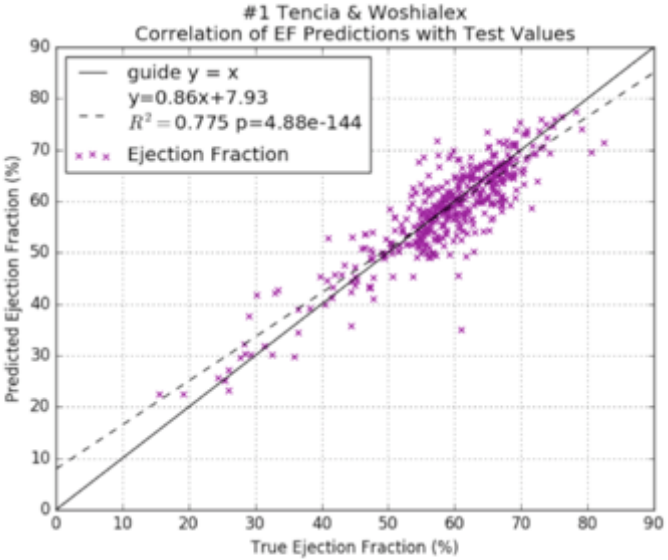
This calibration almost accounted for the difference between 3rd and 4th place

			Diastole (ml)		Systole (ml)	
Fold no	**Train n**	**Validate n**	**Raw mae**	**Cal mea**	**Raw mae**	**Cal mea**
fold0	558	140	9.76	8.81	8.42	7.11
fold1	558	140	10.72	10.33	8.02	7.07
fold2	558	140	10.40	9.07	9.25	7.40
fold3	558	140	10.87	9.47	8.96	8.00
fold4	560	138	8.94	8.44	7.63	6.25
average			**10.14**	**9.22**	**8.46**	**7.17**

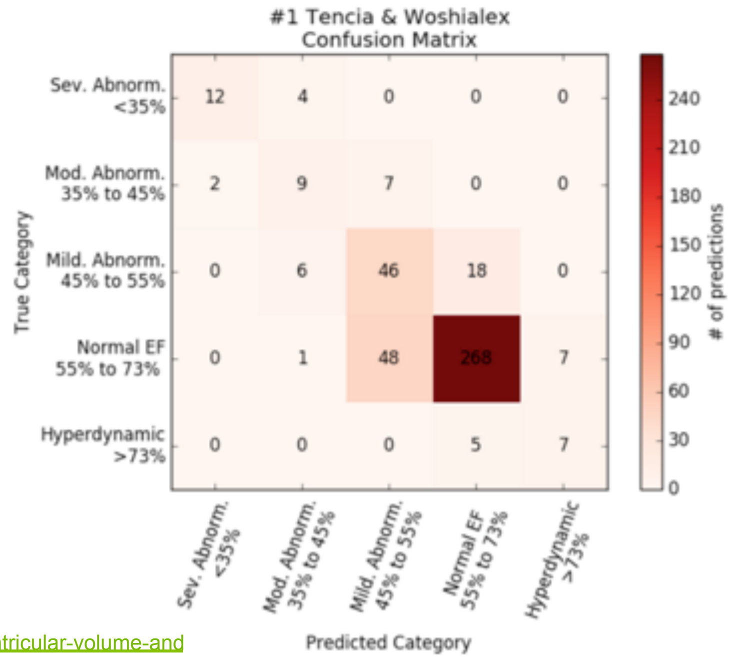
http://juliandewit.github.io/kaggle-ndsb/

MEDICAL SIGNIFICANCE





RMS Error		
Tencia Woshialex	Diastolic	12.02 mL
	Systolic	10.19 mL
	E Fraction(%)	4.88 %
kunsthart	Diastolic	13.65 mL
	Systolic	10.43 mL
	E Fraction(%)	6.99 %
JuliandeWit	Diastolic	13.63 mL
	Systolic	10.32 mL
	E Fraction(%)	5.04 %



https://www.kaggle.com/c/second-annual-data-science-bowl/forums/t/19840/winning-and-leading-teams-submission-analysis
https://www.kaggle.com/c/second-annual-data-science-bowl/forums/t/19839/a-medical-perspective-on-the-quality-of-the-left-ventricular-volume-and

BOOZE-ALLEN-HAMILTON & NVIDIA TEAM

BAH-NVIDIA TEAM

Jared Sylvester
Booz Allen Hamilton

Samantha Tracht
Booz Allen Hamilton

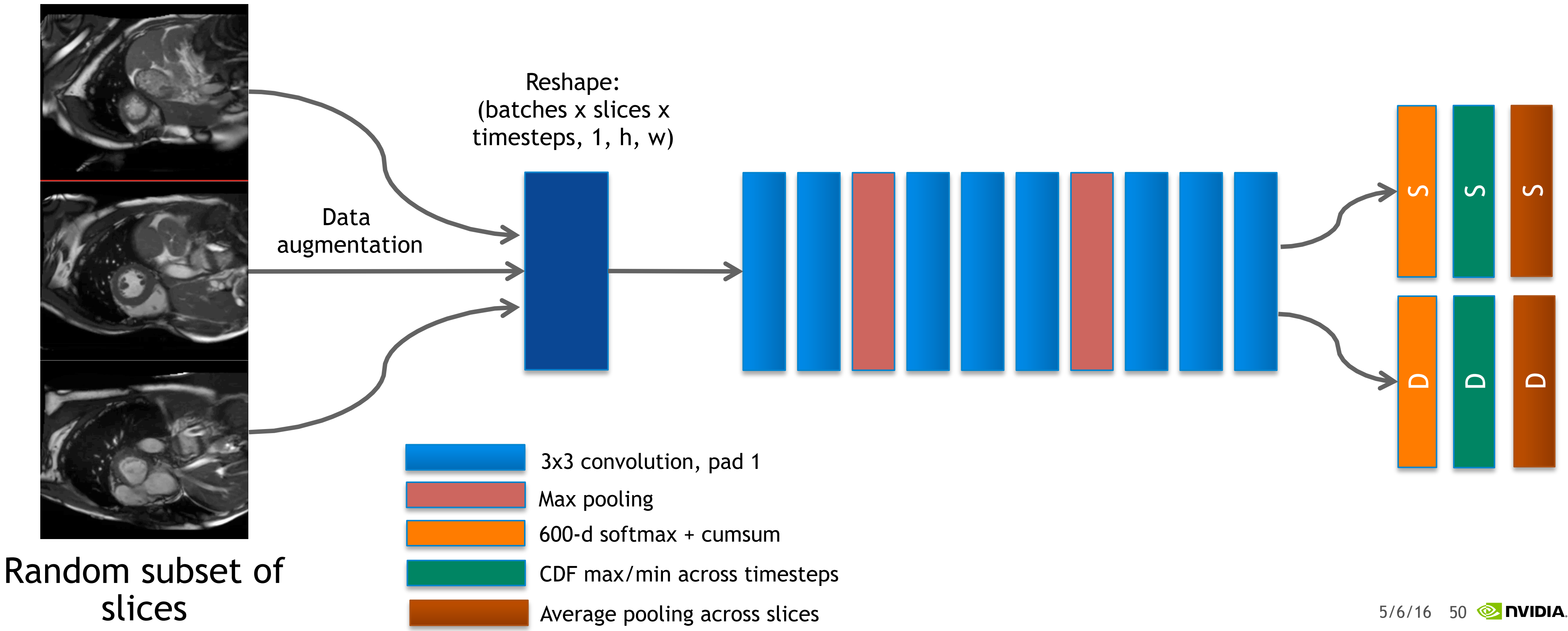
Peter VanMaasdam
Booz Allen Hamilton

Jon Barker
NVIDIA

Maxim Milakov
NVIDIA

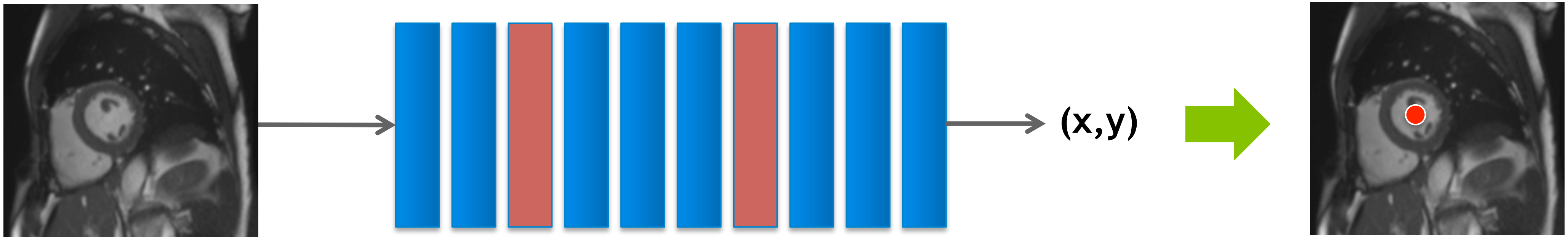
BAH-NVIDIA TEAM

Approach 1: Patient convolutional neural network

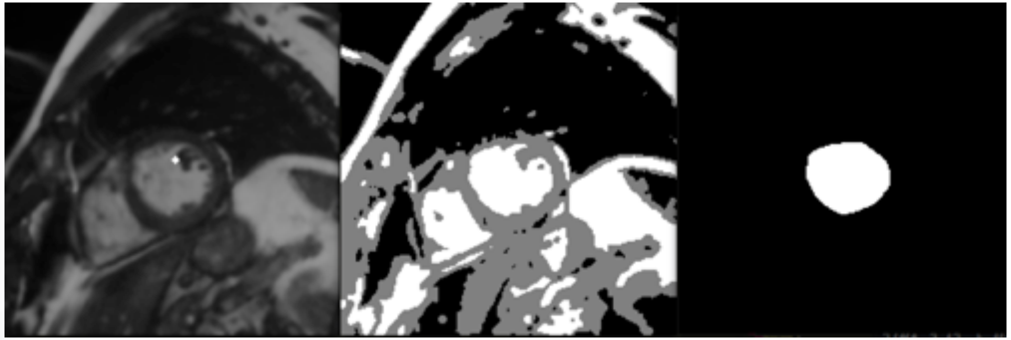


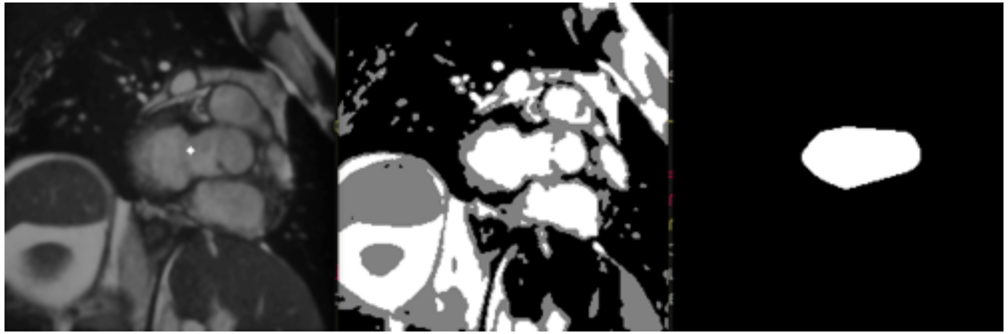
BAH-NVIDIA TEAM

Approach 2: Convnet localization and image segmentation



Single image





SUMMARY

Data science based - “no assumptions” - approach demonstrated medically significant approaches that could save valuable cardiologist time

Solutions converged on varied convolutional neural network based approaches

Outlier cases are incredibly important and average accuracy is not necessarily a sufficient metric in medical applications

Ensembles of diverse models are key to handling difficult edge cases

Distributed GPU training can enable rapid model iteration and ensemble training

GPU TECHNOLOGY CONFERENCE

April 4-7, 2016 | Silicon Valley

THANK YOU

JOIN THE CONVERSATION

#GTC16

JOIN THE NVIDIA DEVELOPER PROGRAM AT developer.nvidia.com/join

PRESENTED BY

NVIDIA