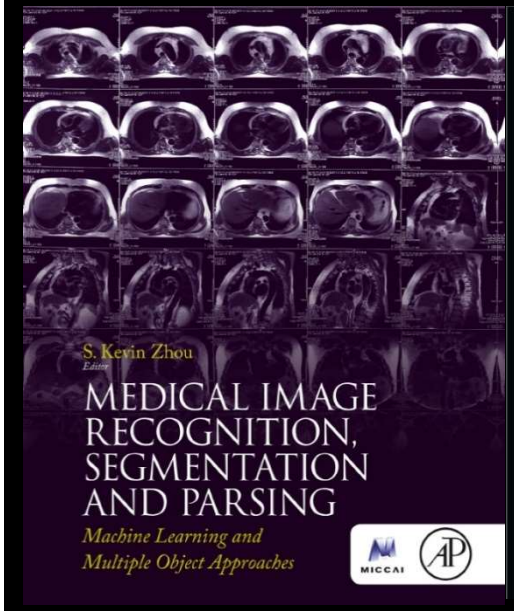
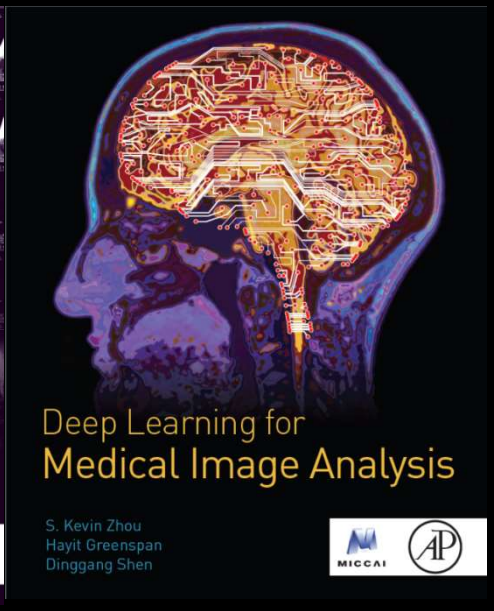
MISS 2018
Medical Imaging Summer School





Machine Learning + Knowledge:

Medical image recognition, segmentation and parsing

Dr. S. Kevin Zhou, 2018/08/03

Disclaimer

SIEMENS Healthineers

中国科学院 CHINESE ACADEMY OF SCIENCES

Institute of Computing Technology

Princeton

Beijing

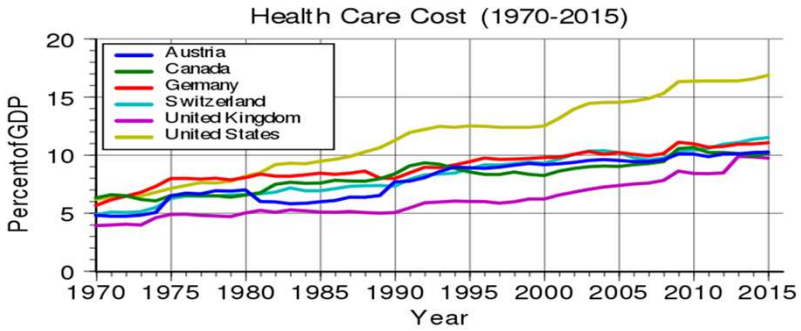
~14 years

2 months

Talk outline

- Overview of medical image parsing
- Medical learning + knowledge
- Deep learning + knowledge

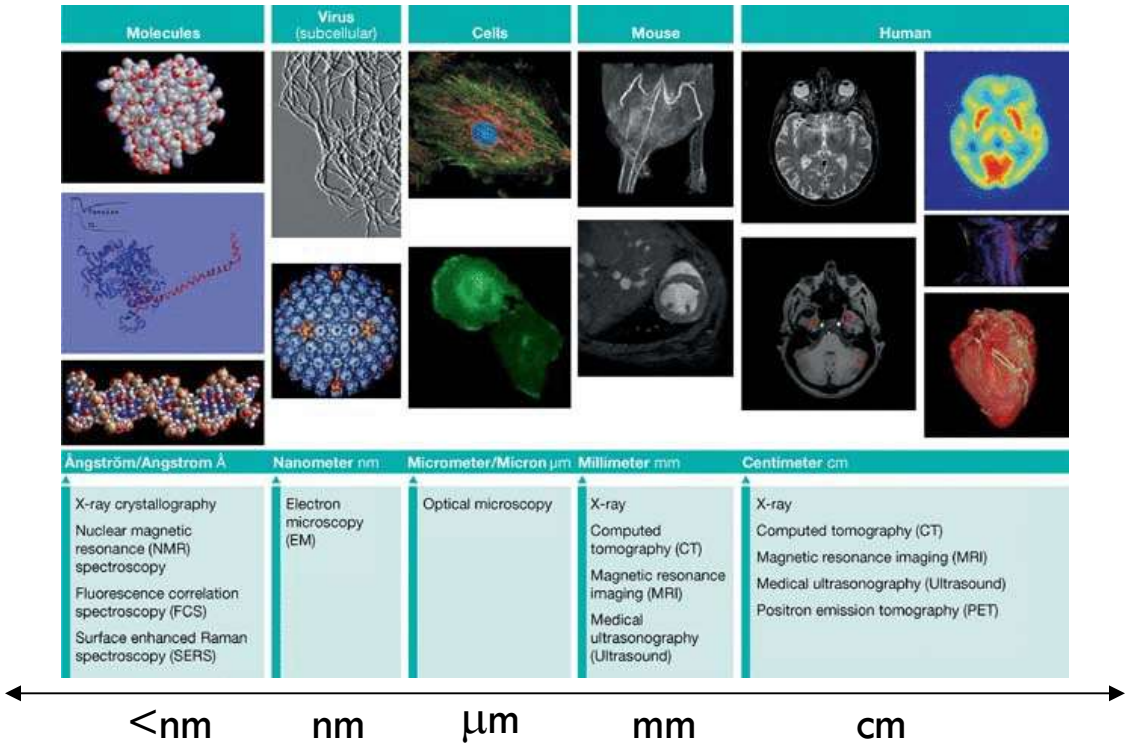
Medical imaging



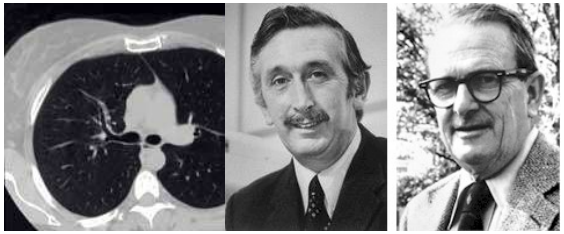
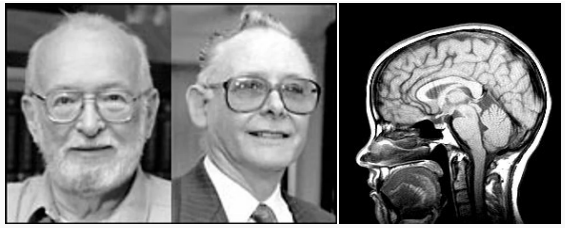
90% of all medical data are images

Medical imaging trend #1:
Multi-modality and multiscale

Multi-modality and multiscale

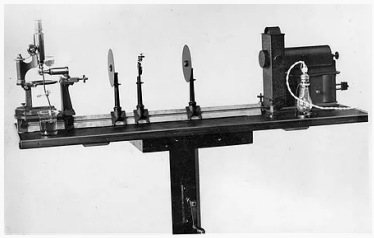


Nobel prizes & clinical imaging

X-Ray, 2D Wilhelm C Roentgen The Nobel Prize in Physics 1901	
Computed Tomography (CT), 3D Sir Godfrey Hounsfield and Dr. Alan Cormack The Nobel Prize in Physiology or Medicine 1979	
Magnetic Resonance Imaging (MRI), Soft tissue Paul Lauterbur and Sir Peter Mansfield The Nobel Prize in Physiology or Medicine 2003	

Nobel prizes & microscopy imaging

Ultra-microscope (under the wavelength of light 500nm)
Richard Zsigmondy
The Nobel Prize in Chemistry 1925



Phase-contrast microscope (colorless and transparent biological materials)
Frits Zernike
The Nobel Prize in Physics 1953

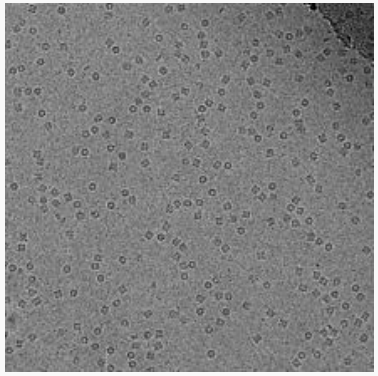
Electron microscope (greatly improves the resolution and expands the borders of exploration, 0.05nm resolution)
Ernst Ruska
The Nobel Prize in Physics 1986

Nobel prizes & microscopy imaging

Scanning tunneling microscope (3D images of surface objects at the atomic level, 0.1nm lateral and 0.01nm depth resolutions)
Gerd Binnig and Heinrich Rohrer
The Nobel Prize in Physics 1986

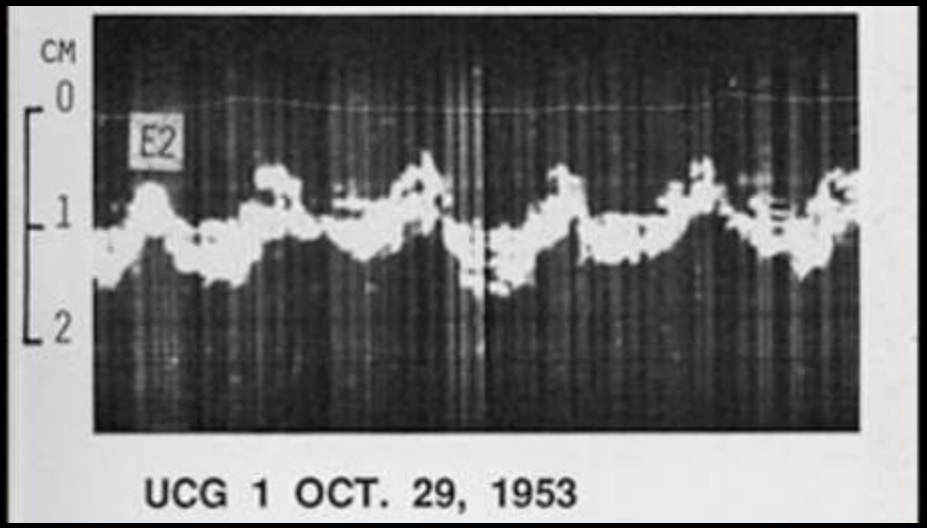
Super-resolved fluorescence microscopy (optical microscopy into the nanodimension)
Robert Eric Betzig, Stefan Hell and William E. Moerner
The Nobel Prize in Chemistry 2014

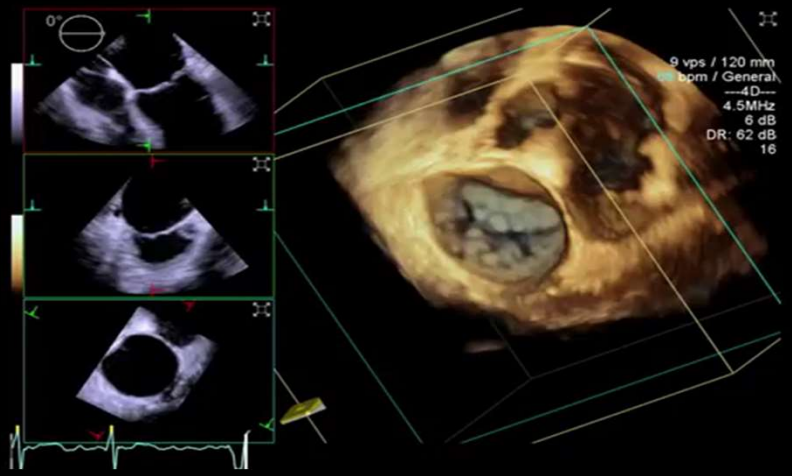
Cryo-electron microscopy (HR structure determination of biomolecules)
Jacques Dubochet, Joachim Frank and Richard Henderson
The Nobel Prize in Chemistry 2017



Medical imaging trend #2:

Increasing information density





https://www.youtube.com/watch?v=oJEhf6uF7hw

MRI: increasing information density

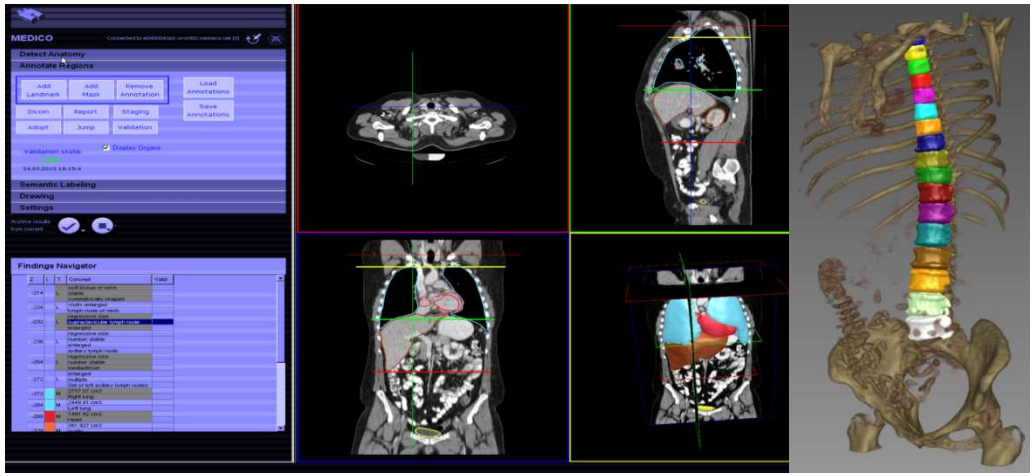
https://www.siemens.com/press/de/feature/2017/healthineers/2017-06-cinematic-rendering.php

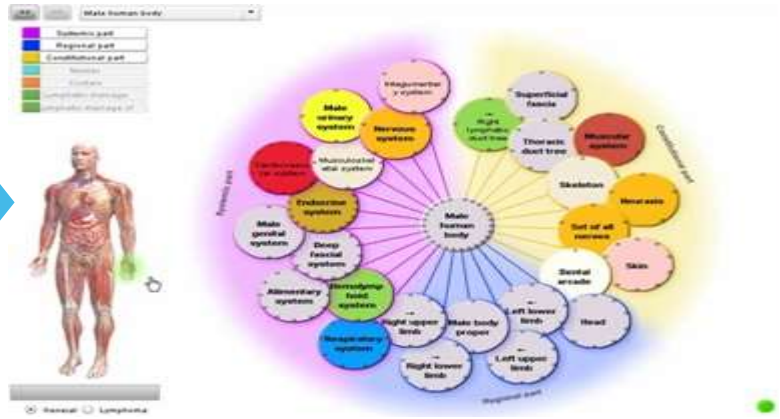
CT: increasing information density

https://www.dotmed.com/news/story/27660?p_begin=1

Medical imaging trend #3: AI

Holy grail: Medical image parsing





Medical image parsing

- Assigning semantic labels to pixels or voxels
- Unifying detection, segmentation, and parsing

Foundational model of anatomy (FMA)

- ~75,000 classes and over 120,000 terms; over 2.1M relationship instances

Kahn's Jr.'s radiology Gamuts

- 2,613 findings & 23,373 conditions

Technology benefits

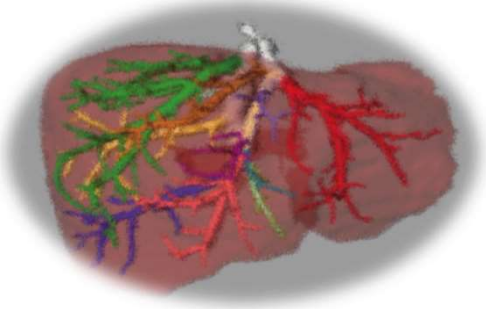
Imaging Scanner

- Personalized
- Consistent
- Fast
- Less radiation

Screening & Diagnosis

- Structured reading
- Clinical quantification
- Streamlined workflow
- Semantic reporting

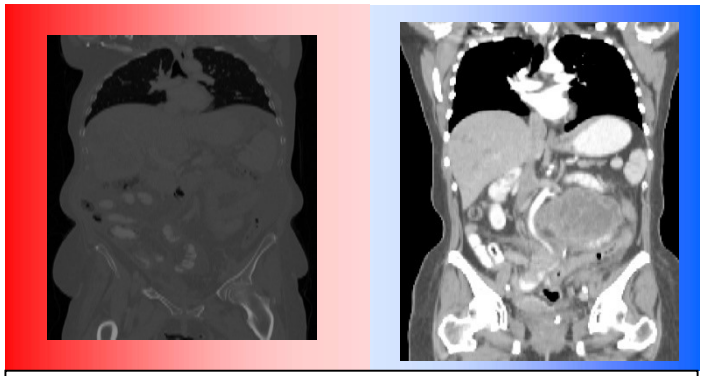
Therapy & Surgery



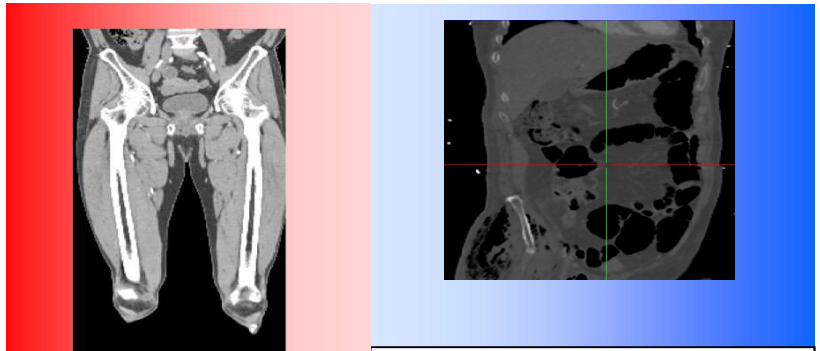
- Surgery planning
- Therapy prediction
- Therapy monitoring

Challenges

Image and shape variations



Contrast



Pathologies

- Deformed, missing, misleading context

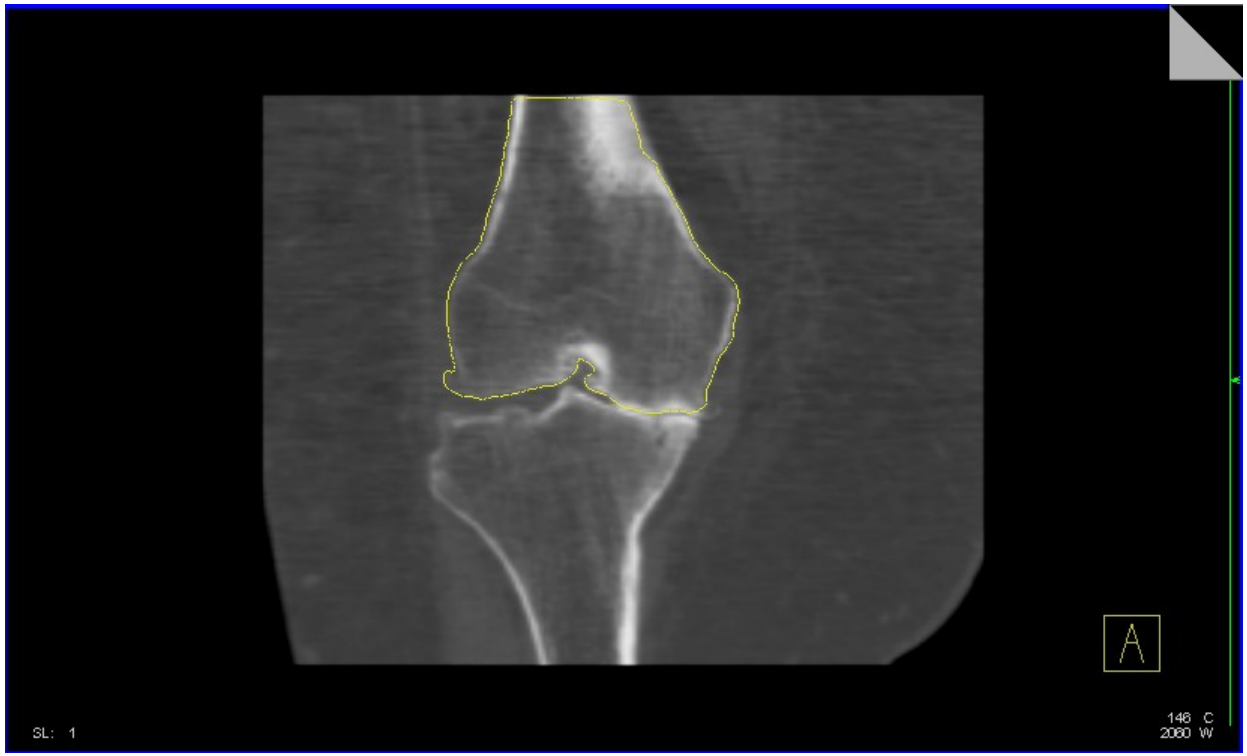
Context

Body Portions

- Narrow FOV , Severe Occlusion

Appearance variation

Touching Bones, Femur vs. Tibia



Appearance variation

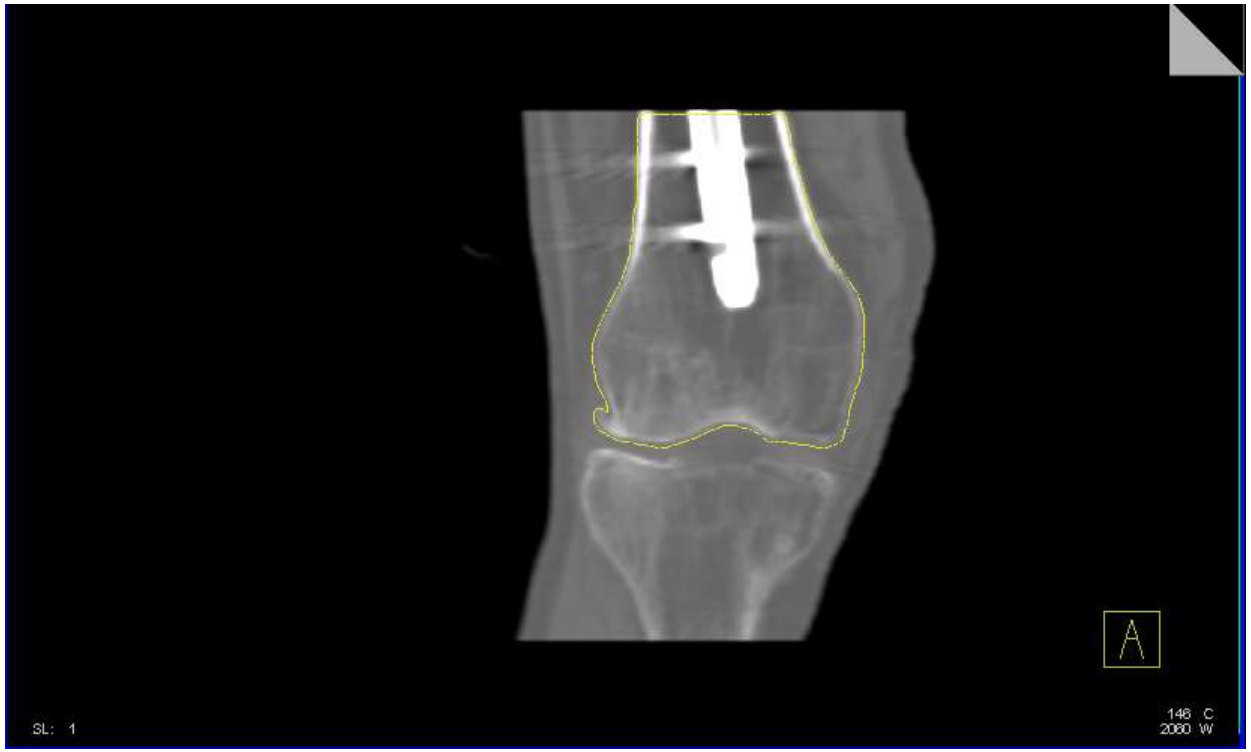
Pathology, Osteoporosis

Appearance variation

Broken Femur

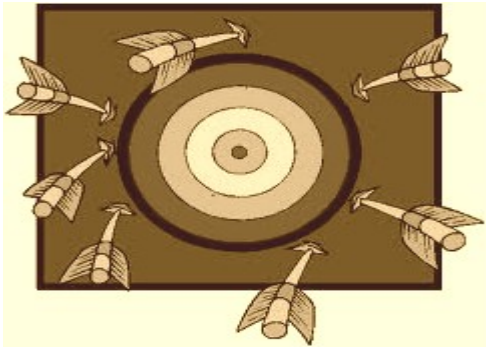
Appearance variation

Metallic Implants



Additional challenges

Robustness:
No outlier



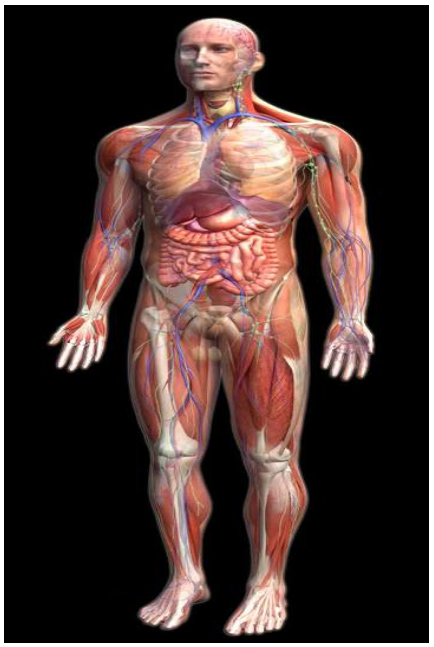
Accuracy:
Within inter-user variability

Speed:
Less than a few seconds

Opportunities

Large Amount of Datasets

+



Knowledge

=

Machine Learning + Knowledge

Machine learning + knowledge

Machine learning

- Supervised learning
  - Deep neural network
  - SVM, boosting, etc.
- Unsupervised learning
- Reinforcement learning

Knowledge

- Human anatomy
- Sensor physics
- Geometry
- Prior constraints
- Etc.

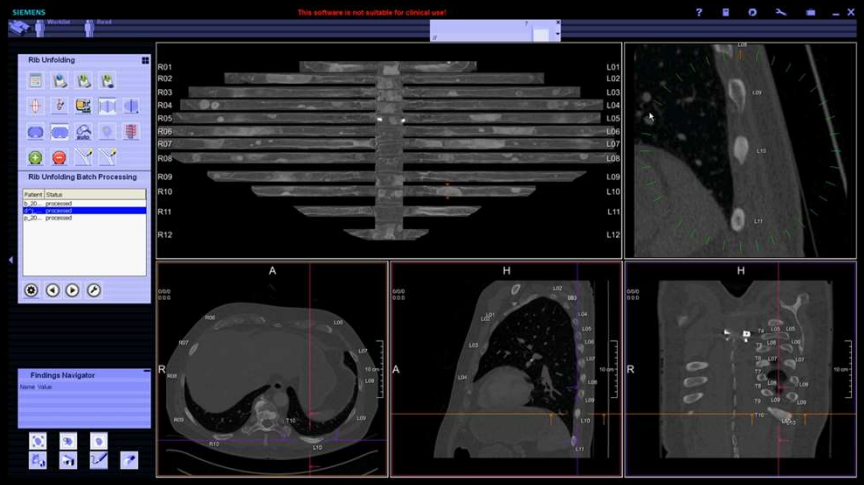
Talk outline

- Overview of medical image parsing
- Medical learning + knowledge
- Deep learning + knowledge

Rib unfolding

SIEMENS Healthineers

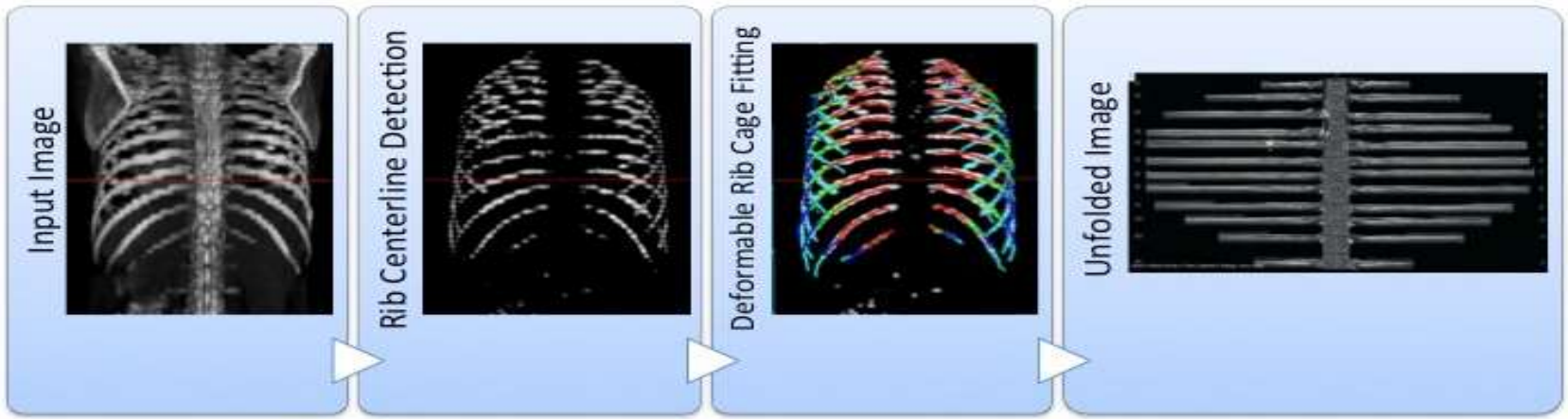
Oscar of invention



2 times faster, 10% more fracture detection

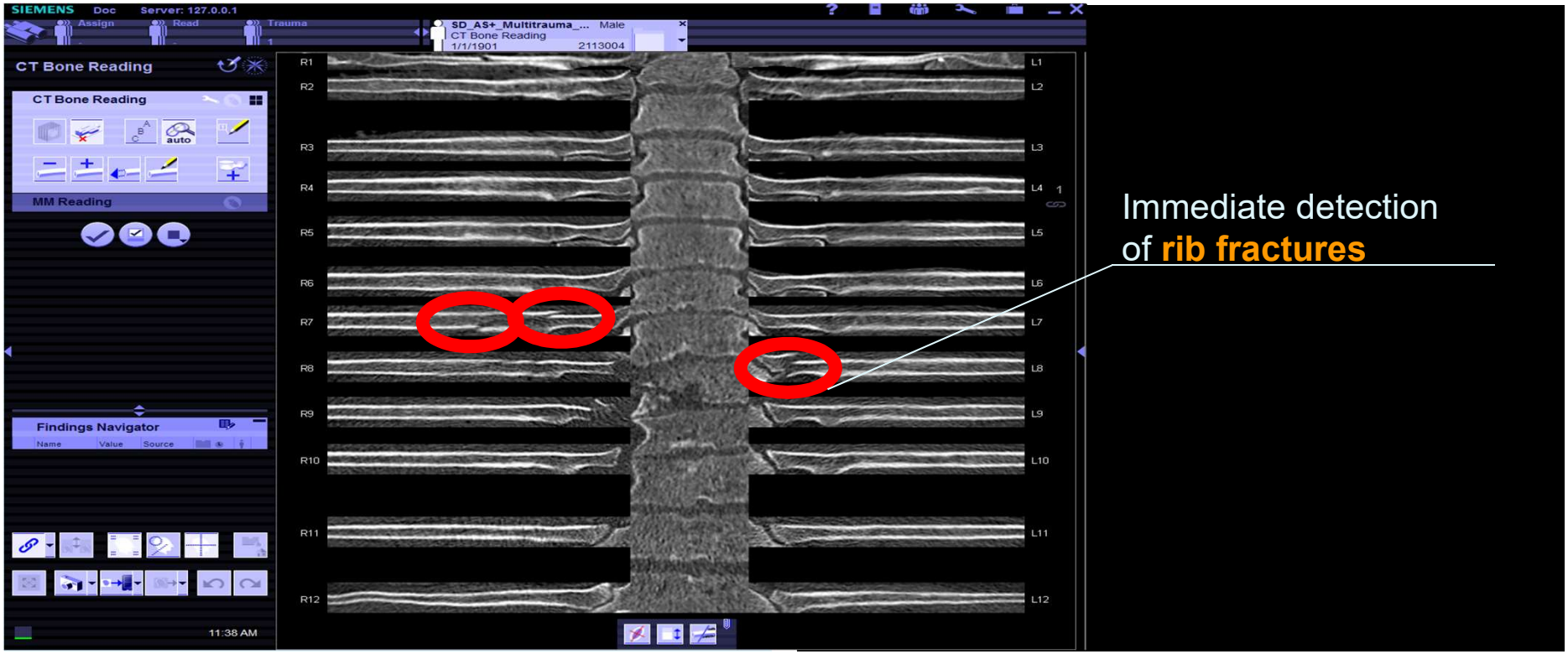
* Ringl H et al. The ribs unfolded - a CT visualization algorithm for fast detection of rib fractures: effect on sensitivity and specificity in trauma patients. Eur Radiol 2015; 25:1865-74

ML + Rib cage model fitting

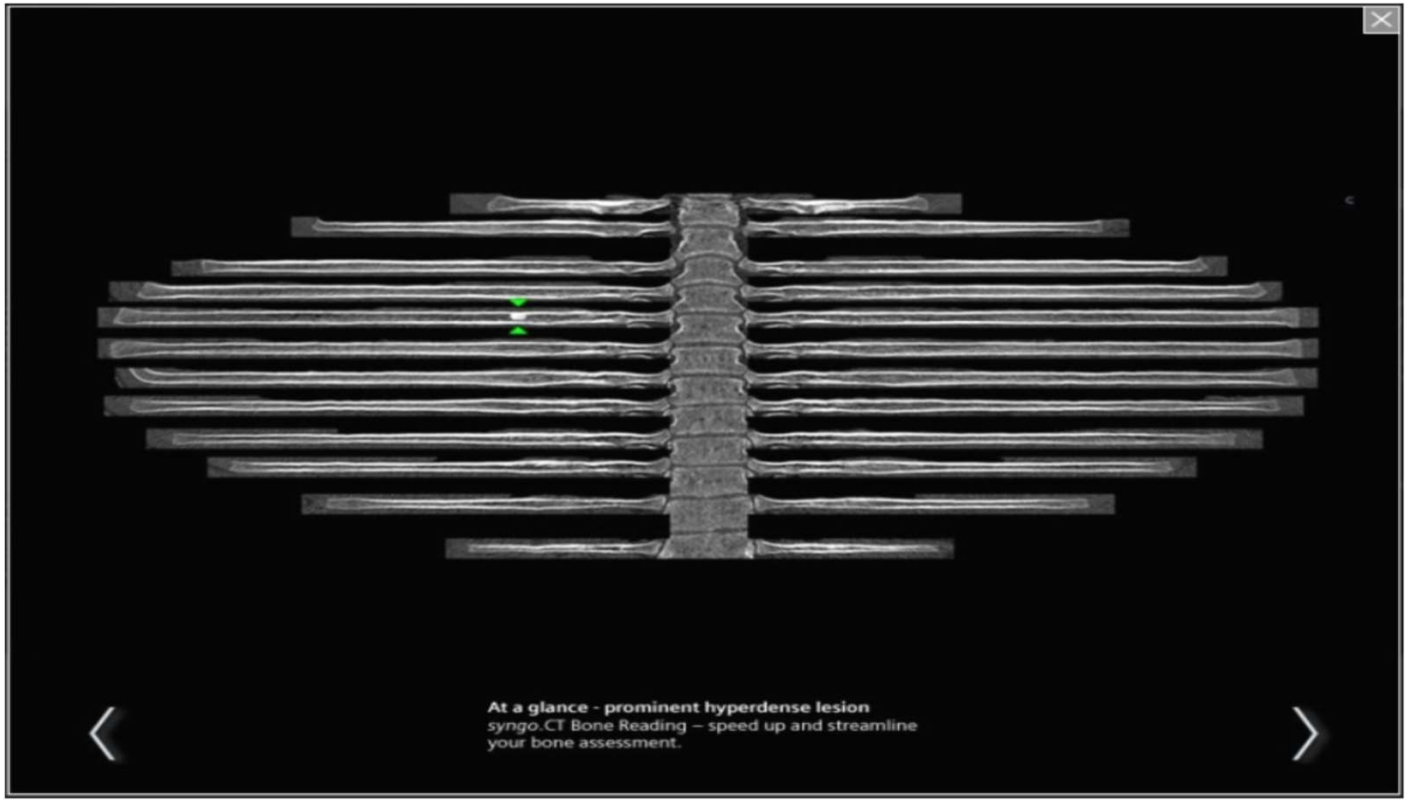


* Wu et al, “A learning based deformable template matching method for automatic rib centerline extraction and labeling in CT images,” CVPR 2012.

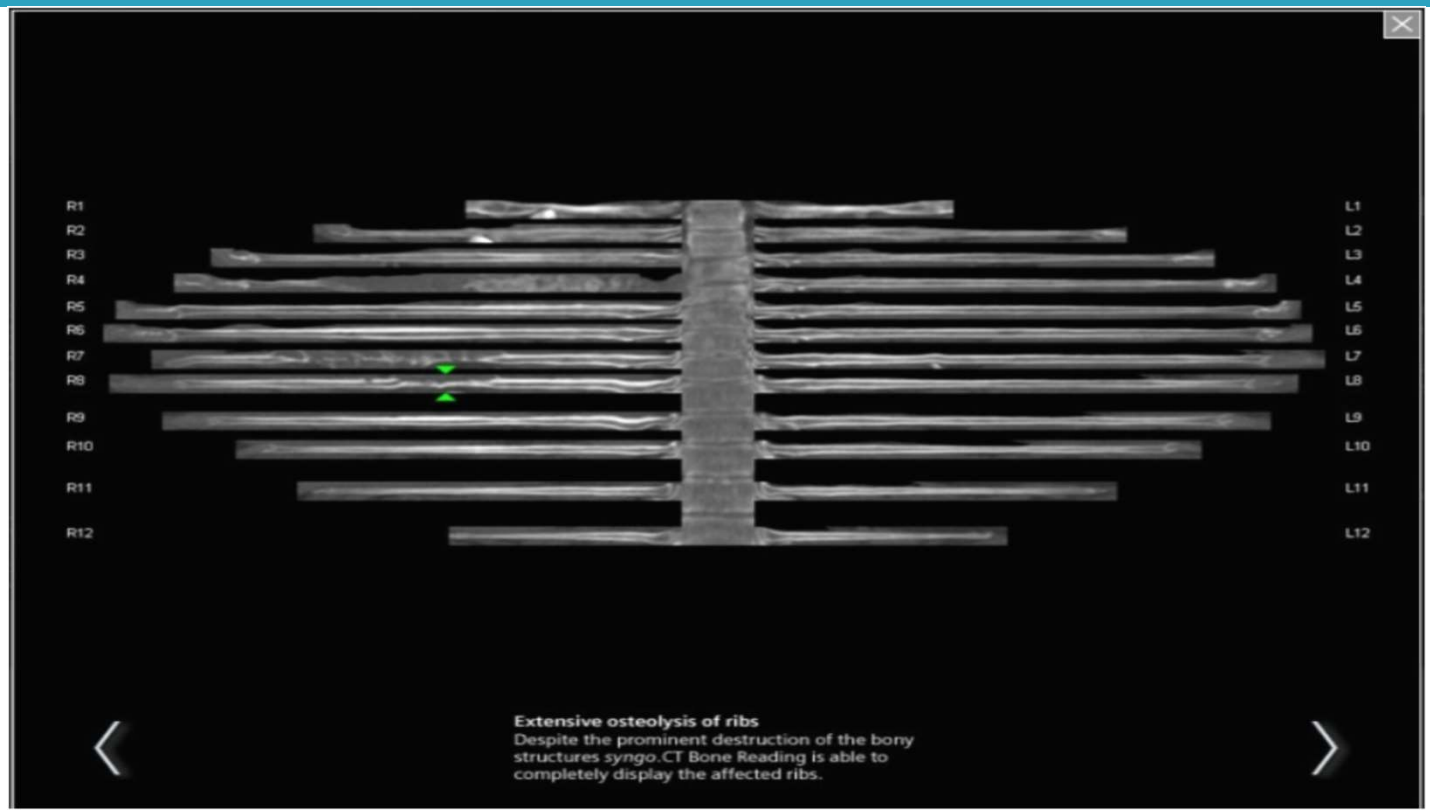
Fractures


SIEMENS
Doc
Server: 127.0.0.1
Assign
Read
Trauma
SD_AS+_Multitrauma_...
Male
CT Bone Reading
1/1/1901
2113004
CT Bone Reading
CT Bone Reading
auto
MM Reading
Findings Navigator
Name
Value
Source
11:38 AM
R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
L1
L2
L3
L4
L5
L6
L7
L8
L9
L10
L11
L12
Immediate detection
of rib fractures

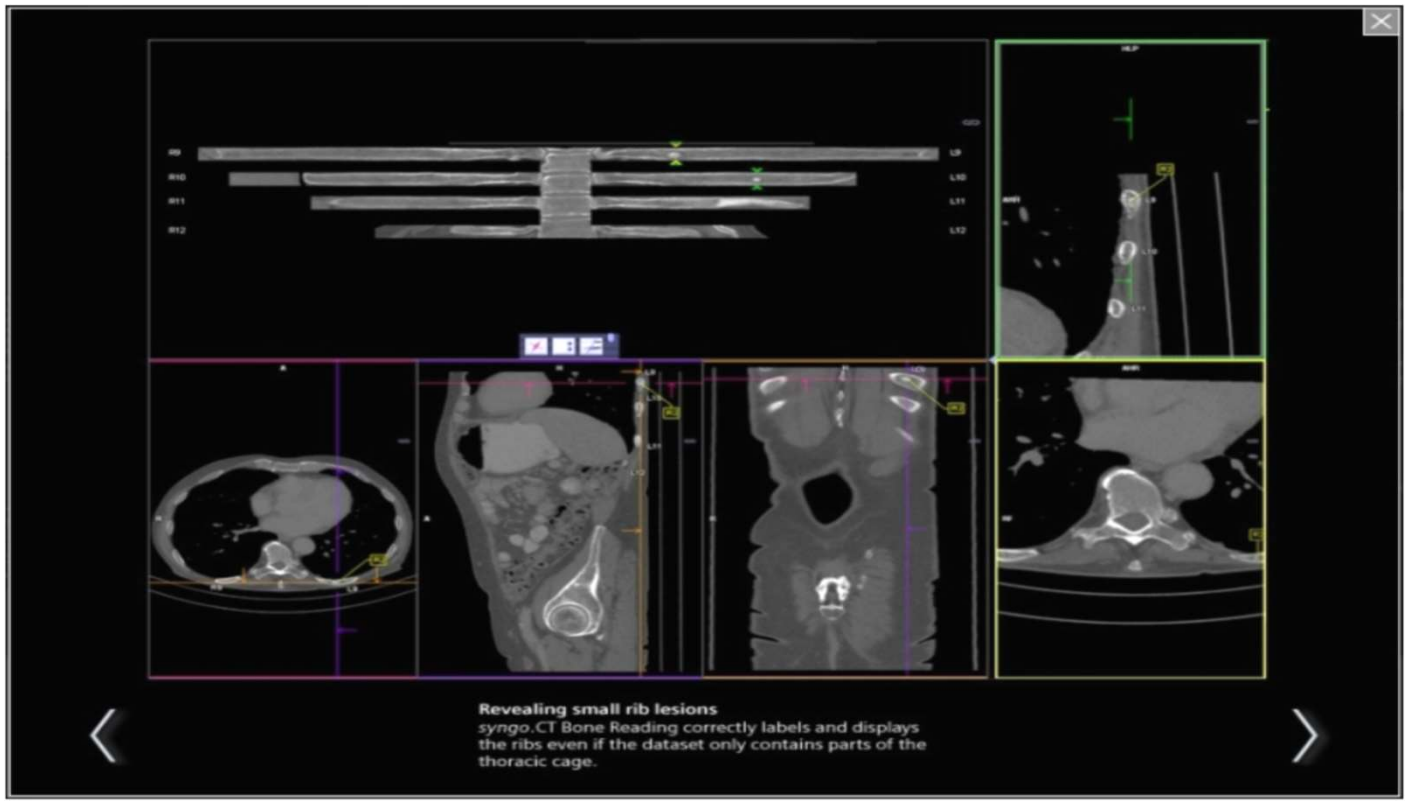
Hyperdense lesion



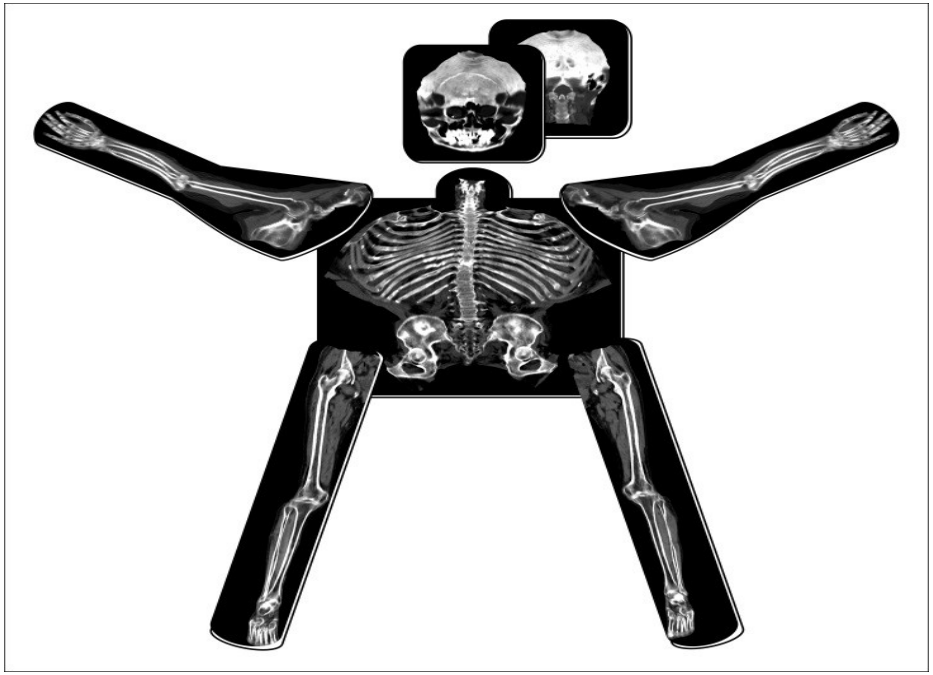
Osteolysis


R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
L1
L2
L3
L4
L5
L6
L7
L8
L9
L10
L11
L12
Extensive osteolysis of ribs
Despite the prominent destruction of the bony structures syngo.CT Bone Reading is able to completely display the affected ribs.

Incomplete rib cage

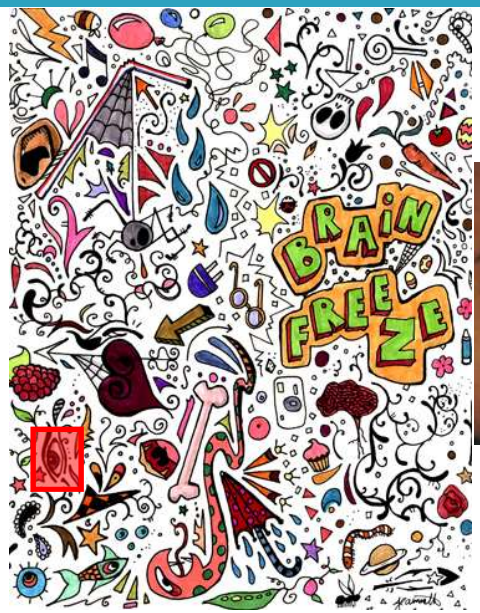


Skeleton unfolding

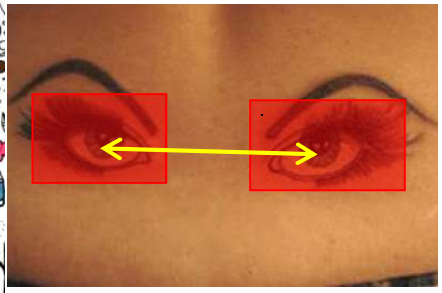


- Patent US9558568 B2: Visualization method for a human skeletonfrom a medical scan
- The system is currently under development and is not for sale. Its future availability cannot be guaranteed.

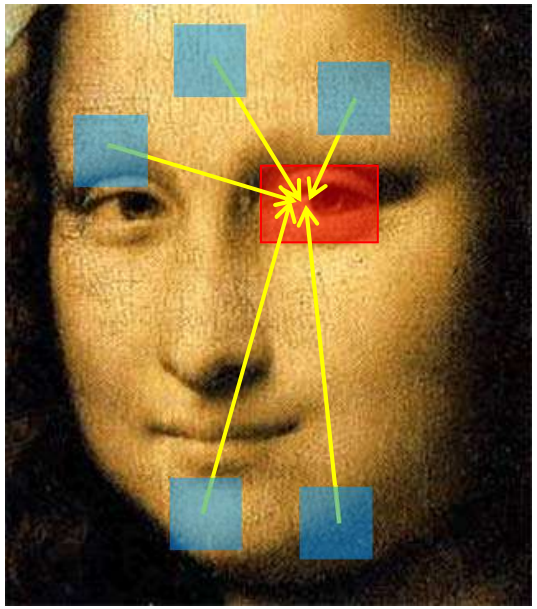
Context



Unitary / Local Context



Pairwise Context



Holistic / Global Context

Classification-based object detection

[Viola & Jones'01]

- Model
  - An object O at translation $t_{x:y}$ and with scale s: $O = [\, t_{x:y}, s \,]$
  - Detector D: $P(O \mid I) = P(+1 \mid I[O])$
- Optimization (exhaustive scanning)

$$\text{argmax } P(+1 \mid I[t_{x:y}, s])$$

- Computation

$$|t_x| * |t_y| * |s| * \tau(D)$$

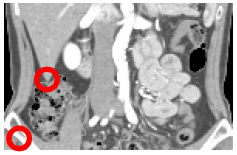
Unitary / Local Context

Submodular landmark detection

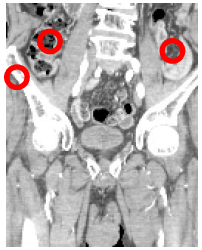
[Liu et al. CVPR'10]

Detection of 238 landmarks in ~1s

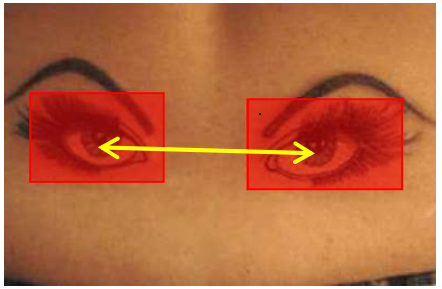
Skull Base
Lung Top



Liver, Hip



Hip, Kidney



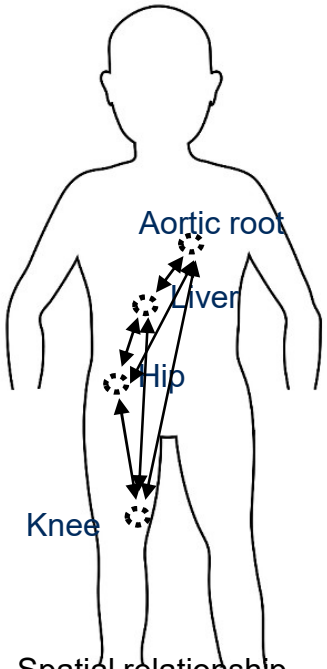
Pairwise Context

Trachea

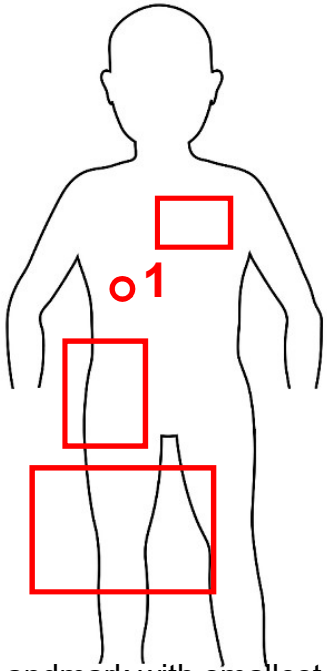
Liver, Sternum

Knee

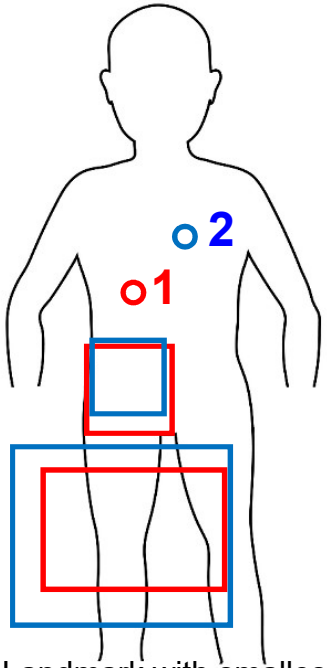
“Greedy search” for fast detection



Spatial relationship between landmarks provided by training data



Landmark with smallest search range: Aortic root



Landmark with smallest search range: Liver

Submodular maximization

$$\text{Index set}: T_{(1):(n)} = \{t_{(1)} \prec t_{(2)} \prec \ldots \prec t_{(n)}\}$$

$$\text{Search range}: \Omega[t_i \mid T_{(1):(n)}] = \bigcap_{t_{(j)} \in T_{(1):(n)}} \Omega[t_i \mid t_{(j)}]$$

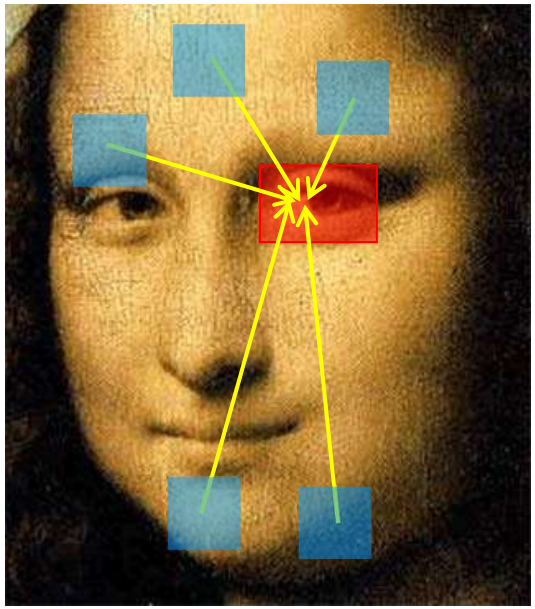
$$\arg\max_{T_{(1):(N)}} F(T) = -V(\Omega[t_{(1)}]) - \sum_{k=2}^{N} V(\Omega[t_{(k)} \mid T_{(1):(k-1)}])$$ **submodular**

- Theorem: If F is a submodular, nondecreasing function and $F(\phi) = 0$, then the greedy algorithm finds a set T' such that $F(T') \geq (1 - 1/e) \max F(T)$

- Approximation reaches at least 63% of optimal solution (off-line bound)

Context integration: combining classification and regression [Lay et al. IPMI 13]

- Idea
  - Integrating local and global contexts
  - Combining classification and regression
  - Sparse scanning + verification
- Performance
  - Fast speed
  - High accuracy
  - Good scalability to multiple objects



Step 1: multiple landmark

$$P(O_{1:N} \mid V) \approx P_L(O_{1:N} \mid V) P_G(O_{1:N} \mid V)$$

Local Context

$$P_L(O_{1:N} \mid V) = \prod_n P_L(O_n \mid V)$$

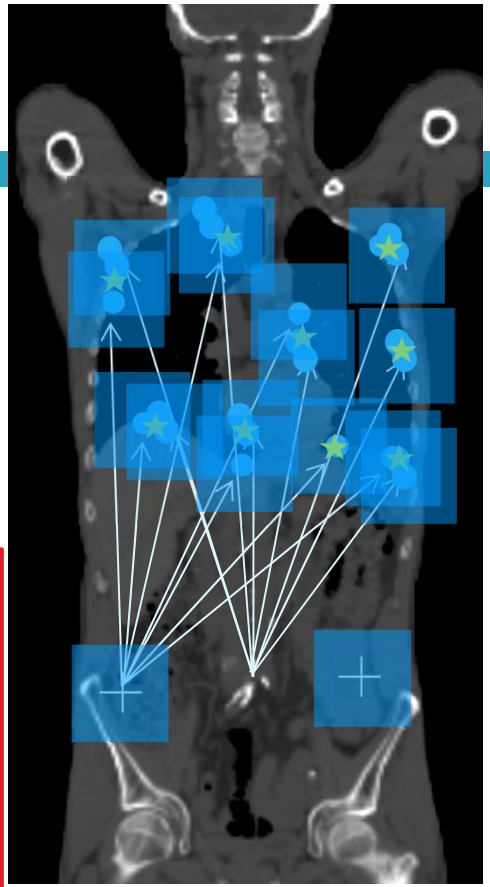
$$P_L(O_n \mid I) \approx P_L(+1 \mid I[O_n])$$

Global Context

$$P_G(O_{1:N} \mid V) = \sum_{u \in \Omega} P_G(O_{1:N} \mid V[u]) P(V[u])$$

$$= \mid \Omega \mid^{-1} \sum_{u \in \Omega} P_G(O_{1:N} \mid V[u])$$

$$P_G(O_{1:N} \mid V[u]) \approx K^{-1} \sum_{k=1}^{K} \delta(O_{1:N} - u - dO^k_{1:N}(V[u]))$$



Step 2: boundary

- Boundary initialization
  - Robust shape alignment

$$\sum_{j=1}^{D} \psi(|x_j - T(\sum_{m=1}^{M} \lambda_m S_{m,j}; \beta)|^2)$$

β - Transformation parameters

$\lambda = [\lambda_1, \lambda_2, ..., \lambda_M]$ - Shape blending weights

  - Typically better than bounding box-based initialization (more DOF)

- Boundary refinement
  - Active Shape Model (ASM)
  - Discriminative boundary models

Rapid multi-organ segmentation

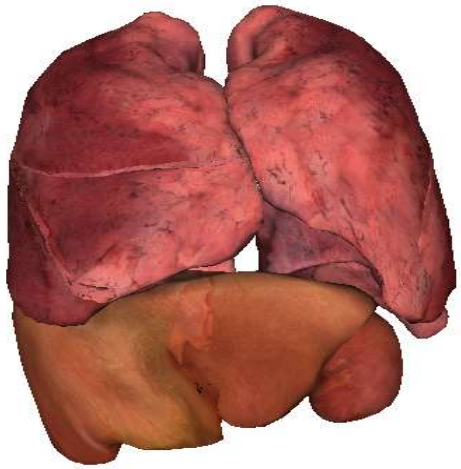
Accuracy

- ~ inter-user variability

Running time:

- 1-2 seconds

syngo.via Radiation Therapy Suite

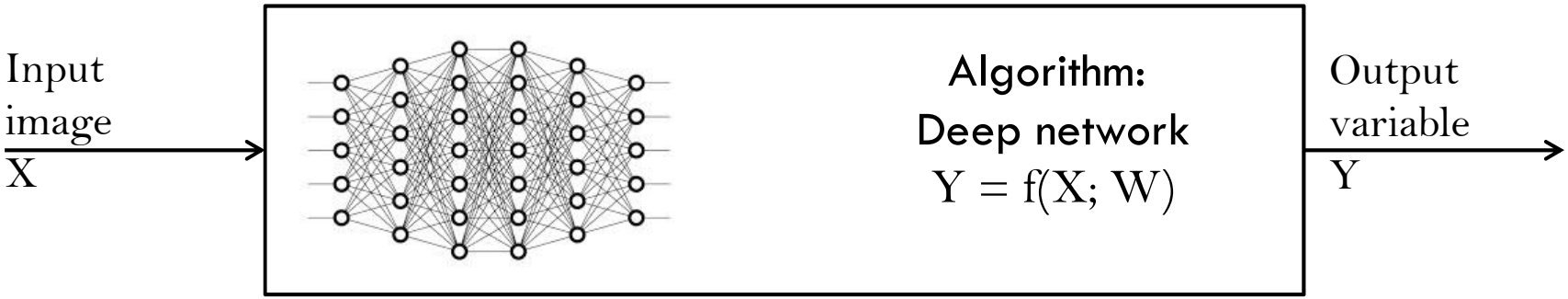


* US Patent 7949173, “Method and system for regression-based object detection in medical Images”
* Lay et al, “Rapid multi-organ segmentation using context integration and discriminative models,” IPMI 2013.

Talk outline

- Overview of medical image parsing
- Medical learning + knowledge
- Deep learning + knowledge

Deep learning

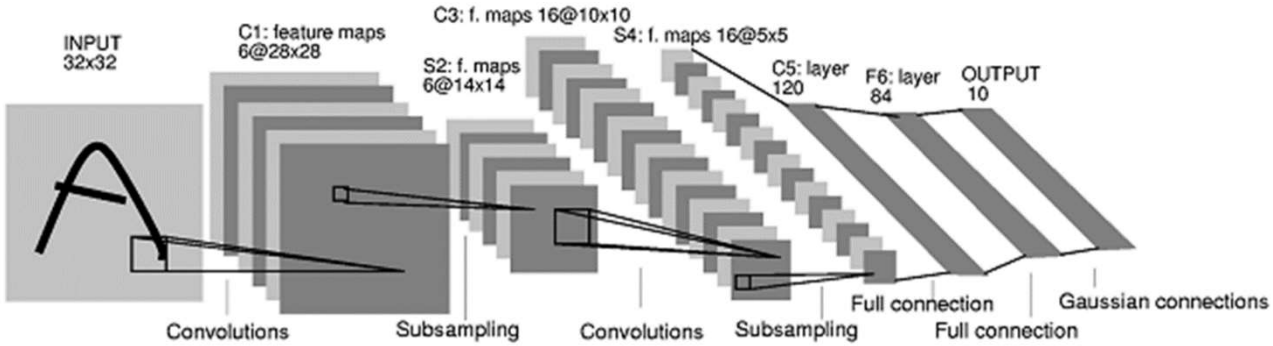


Learning:

$$\arg\min_W \Sigma_i \text{Loss}(Y_i, f(X_i; W)) + \text{Reg}(W)$$

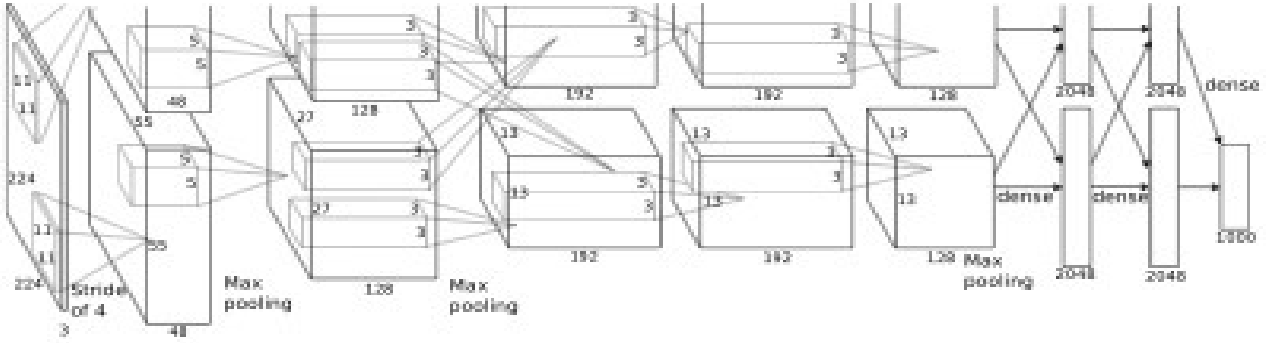
$$\arg\min_W \Sigma_i \text{Loss}(Y_i, f(X_i; W))$$

LeNet



AlexNet

(2012 ImageNet winner)

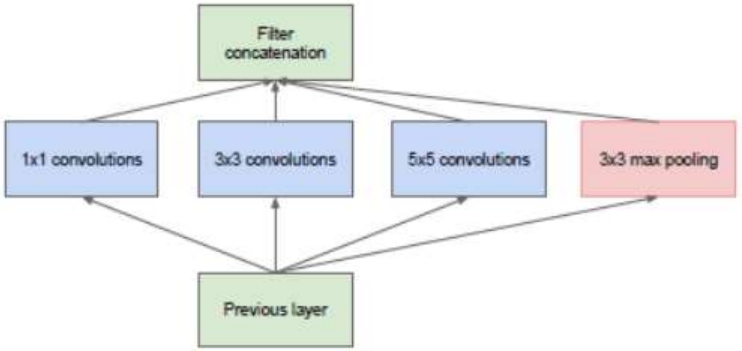


* Filter group

$$\arg\min_W \Sigma_i \, \mathrm{Loss}(Y_i, \, f(X_i; W)) + \mathrm{Reg}(W)$$

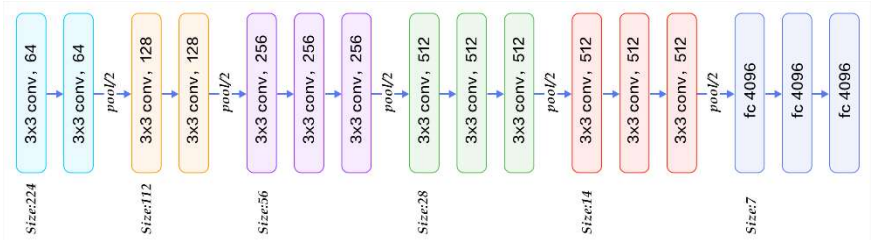
- Advanced network architecture:
  - Inception (2014 ImageNet winner), VGG
  - ResNet (2015 ImageNet winner), ResNeXt, DenseNet
  - Dual path network (DPN)
  - Attention, Squeeze and excitation (SENet) (2017 ImageNet winner)
  - Deep supervision, Deep context-aware network
  - Image-to-image: FCN, Unet, SegNet, etc.

Inception & VGG



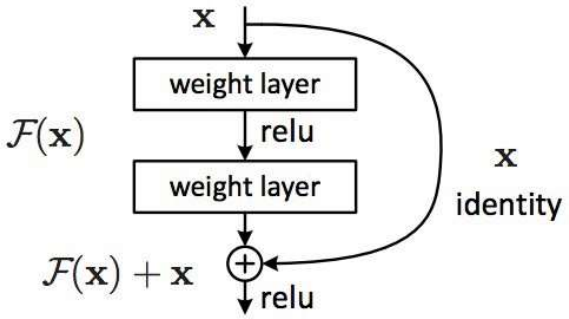
(a) Inception module, naïve version

Inception
[arXiv: 1409.4842]

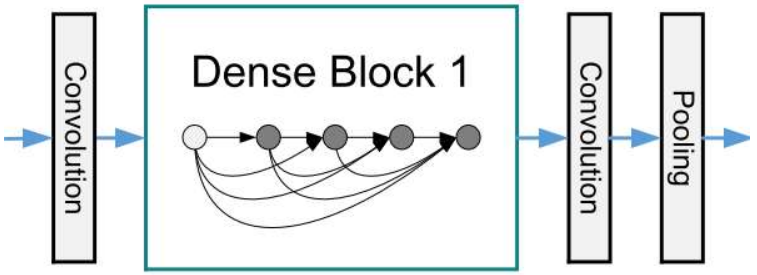


VGG Net
[arXiv: 1409.1556]

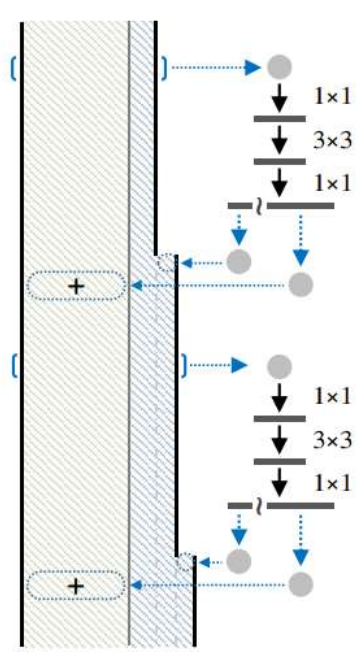
Skip connection



ResNet [arXiv:1512.03385]
ResNeXt [arXiv:1611.05431]

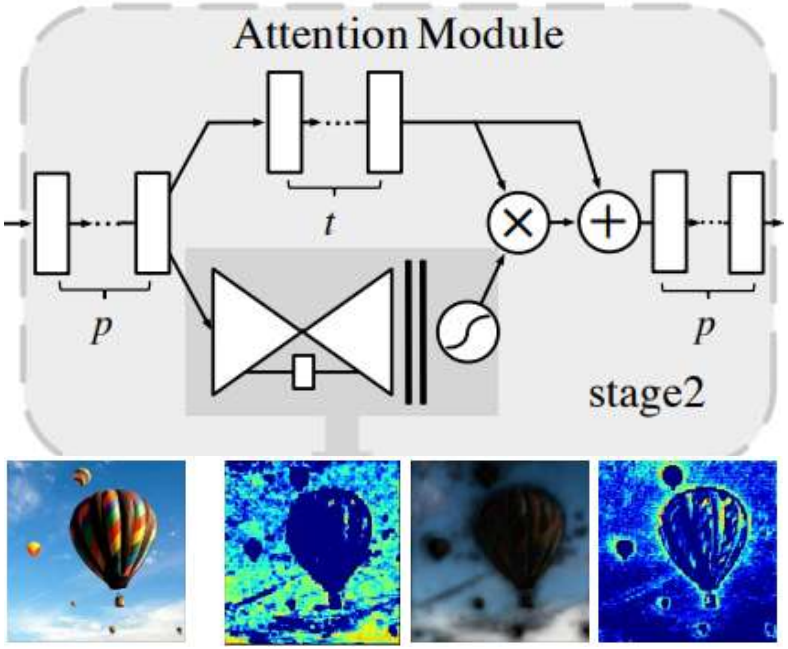


Dense Net
[arXiv:1608.06993]



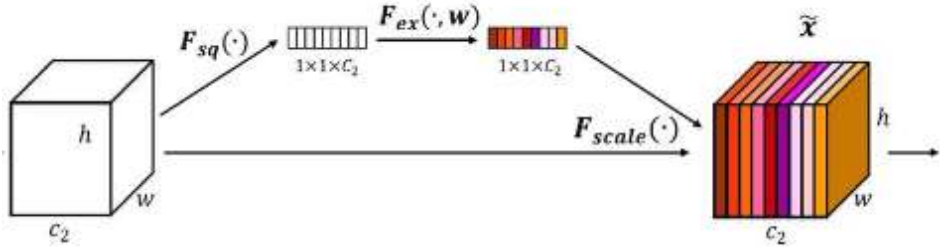
Dual Path Network
[arXiv:1707.01629]

Attention



Spatial attention

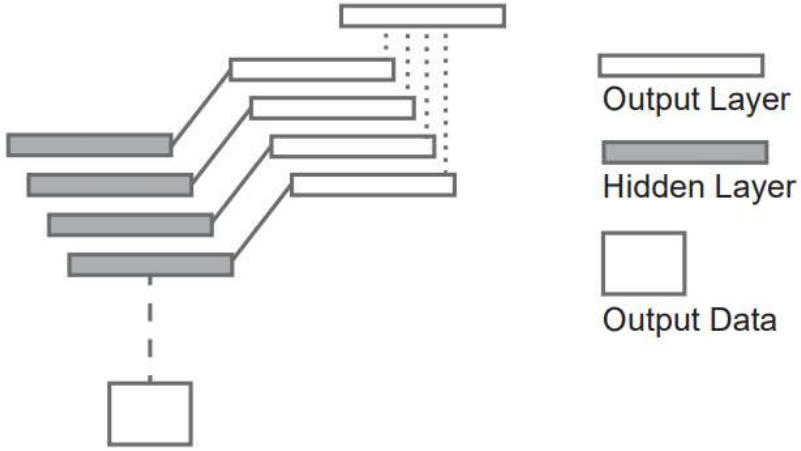
Residual attention network [arXiv:1704.06904]



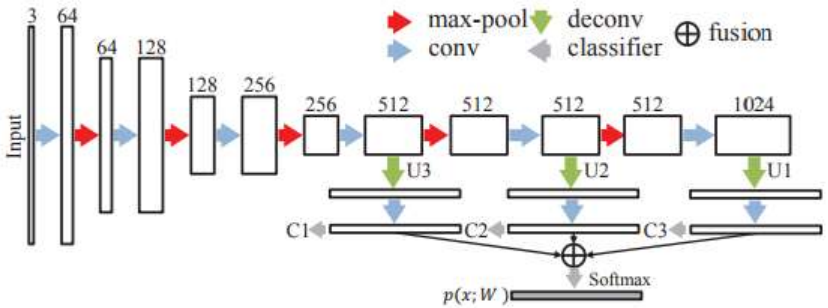
Channel-wise attention

Squeeze and excitation (SENet) [arXiv:1709.01507]

Multiscale fusion



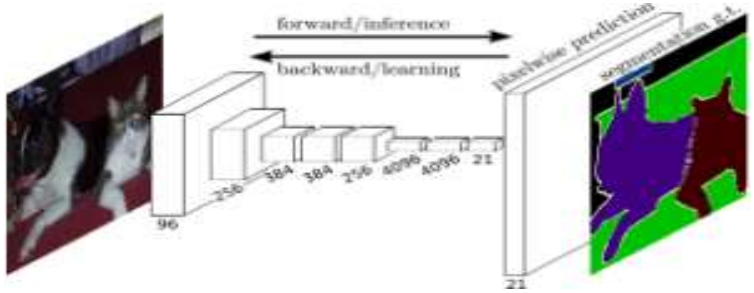
Deep supervision
[arXiv:1504.06375}



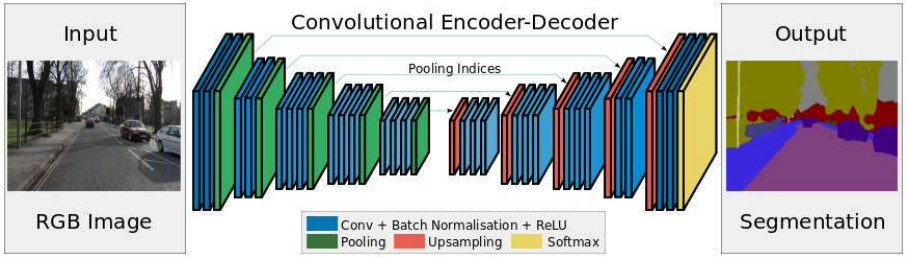
Deep context aware network
[arXiv:1604.02677]

Deep image-to-image network (DI2IN)

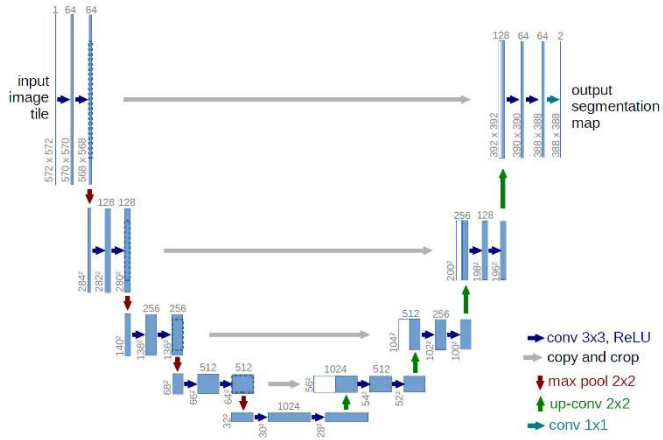
input: image, output: image, pixel-based prediction



FCN [arXiv: 1411.4038]



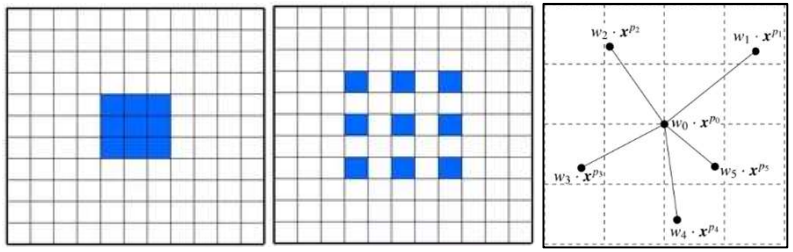
SegNet [arXiv: 1511.00561]



U-Net
[arXiv: 1505.04597]

Convolution & activation function

- Convolution
  - Regular
  - Dilated [arXiv:1511.07122]
  - Active [arXiv:1703.09076]



- Activation function
  - Sigmoid
  - RELU, leaky RELU
  - Scaled exponential linear units (SELU) [arXiv: 1706.02515]

$$\mathrm{selu}(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leqslant 0 \end{cases}$$

$\arg\min_W \Sigma_i \underline{\text{Loss}}(Y_i, f(X_i; W))$

- Loss
  - Classification: cross entropy
  - regression: L2, L1 etc.
  - Ranking: scoring function

- Structured loss

$$\text{Loss}(Y_{1:N}, f(X_{1:N}; W))$$

- Image-based loss (incl. adversarial)

$$\text{Loss}(Y_{1:M*N}, f(X_{1:M*N}; W))$$

- Multitask loss

$$\text{Loss}_1 + \text{Loss}_2 + \ldots + \text{Loss}_k$$

$\arg\min_W \Sigma_i \text{Loss}(Y_i, f(X_i; W))$

Representation: problem-specific, easy-for-learning

- Recognition
  - Multiple labels

  Classification (image-level)

- Landmark/lesion center detection
  - Point coordinate

  Regression (image-level)
  - Heat map:

  Regression (pixel-level)

- Object detection
  - Box

  regression (pixel-level)

- Segmentation/parsing
  - Mask

  classification(pixel-level) + image-based loss (dice, etc.)

$\arg\min_W -\Sigma_i \text{Loss}(Y_i, f(X_i; W))$

- Optimization
  - Back propagation
  - "Vanishing gradient"
  - SGD, ADAM, etc.
  - Batch normalization (BN)
- Memory
  - # of layers
  - # of feature maps
  - # of samples in a mini batch

Deep Learning

What society thinks I do

What my friends think I do

What other computer scientists think I do

What mathematicians think I do

What I think I do

```
In [1]:
import keras
Using TensorFlow backend.
```

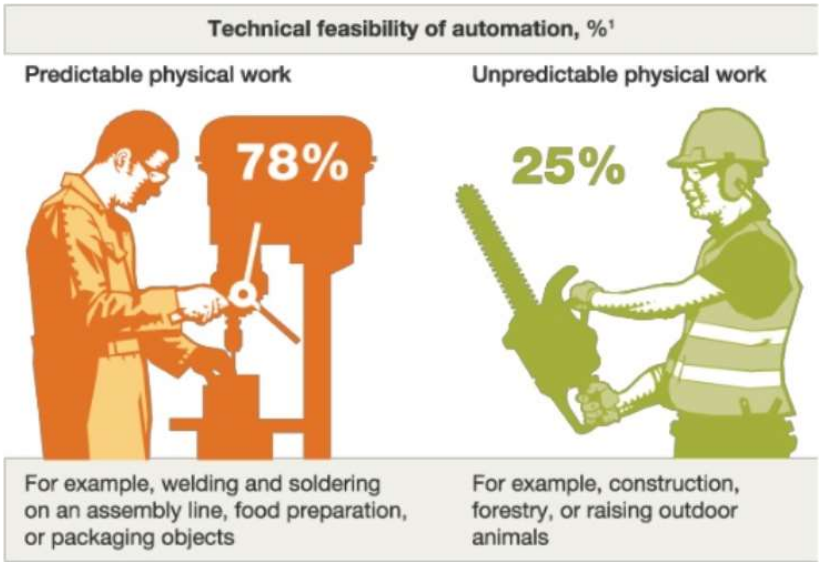
What I actually do

Deep neural nets = “pasta machine”

Ingredients

Pasta

Which jobs are at risk?



Which jobs are at risk?

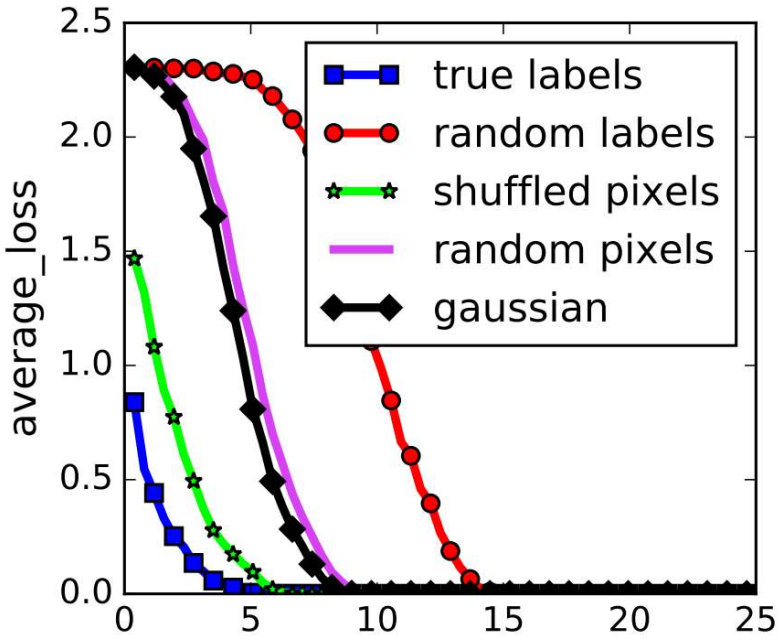
- AI scientist

Deep NN = "equalizer"

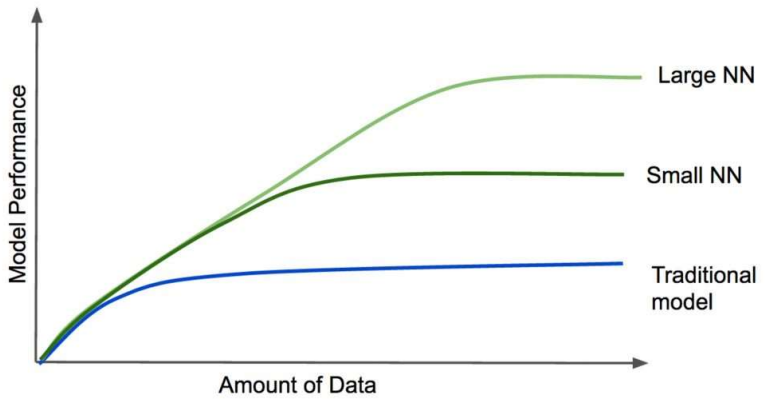
Even worse: Auto ML!

Deep neural nets = "super memorizer"

- "state-of-the-art convolutional networks for image classification trained with stochastic gradient methods *easily fit a random labeling* of the training data."
  [Zhang et al. ICLR2017]



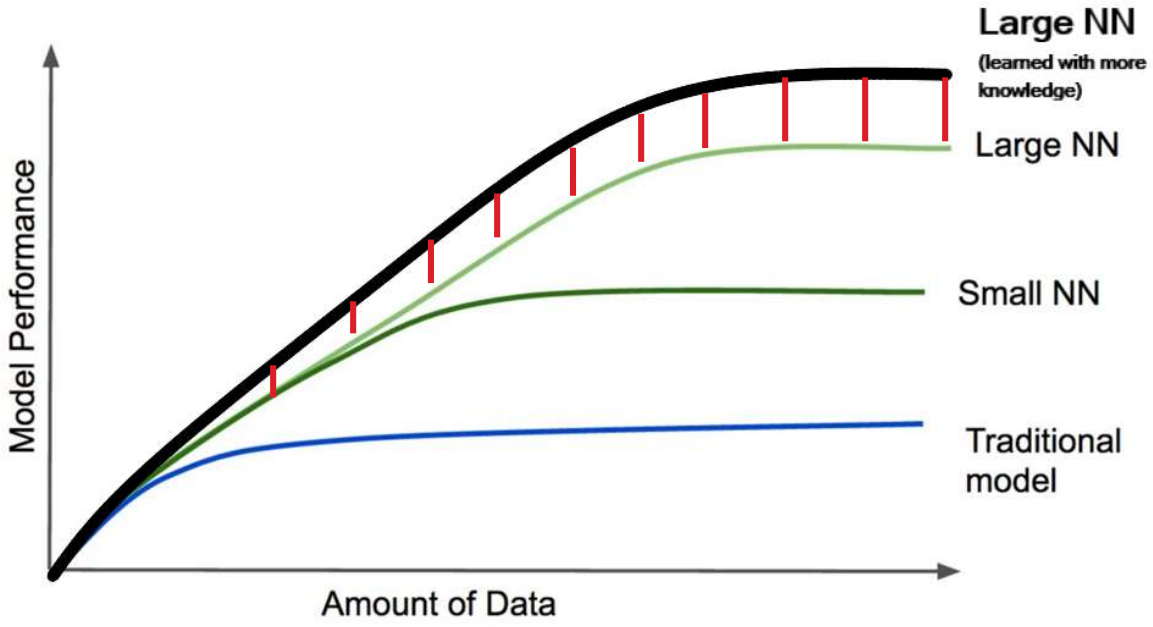
Performance vs amount of data



Recipe for performance improvement:

- Increasing data
- Increasing model capacity
- Repeat the above

Creating a 'knowledge gap'



Why it might work?

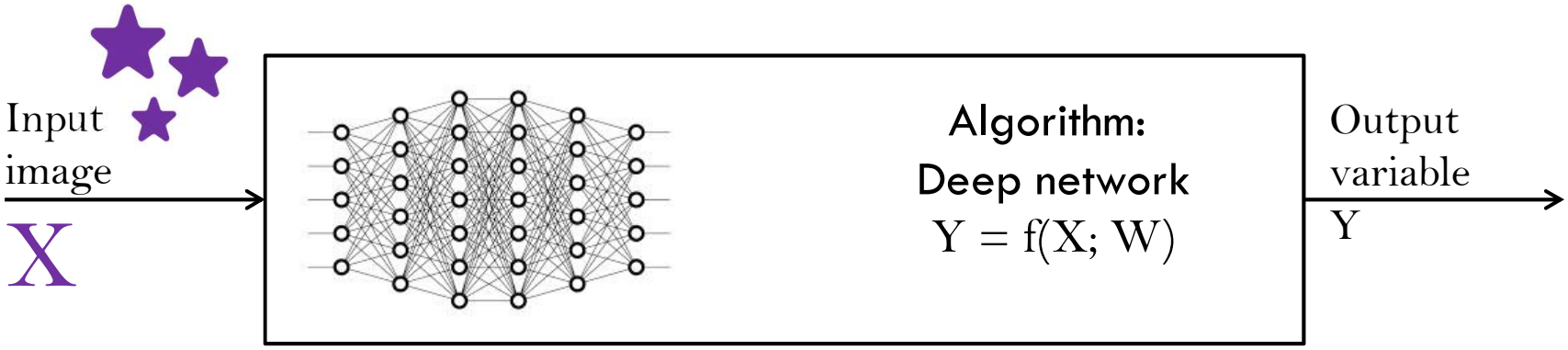
- Making the pattern more uniquely defined
- Seeing more examples
- Making problems more learnable
- Exploiting known information rather than brute force learning

Deep learning with knowledge fusion

Knowledge fusion

- Input
- Output
- Algorithm

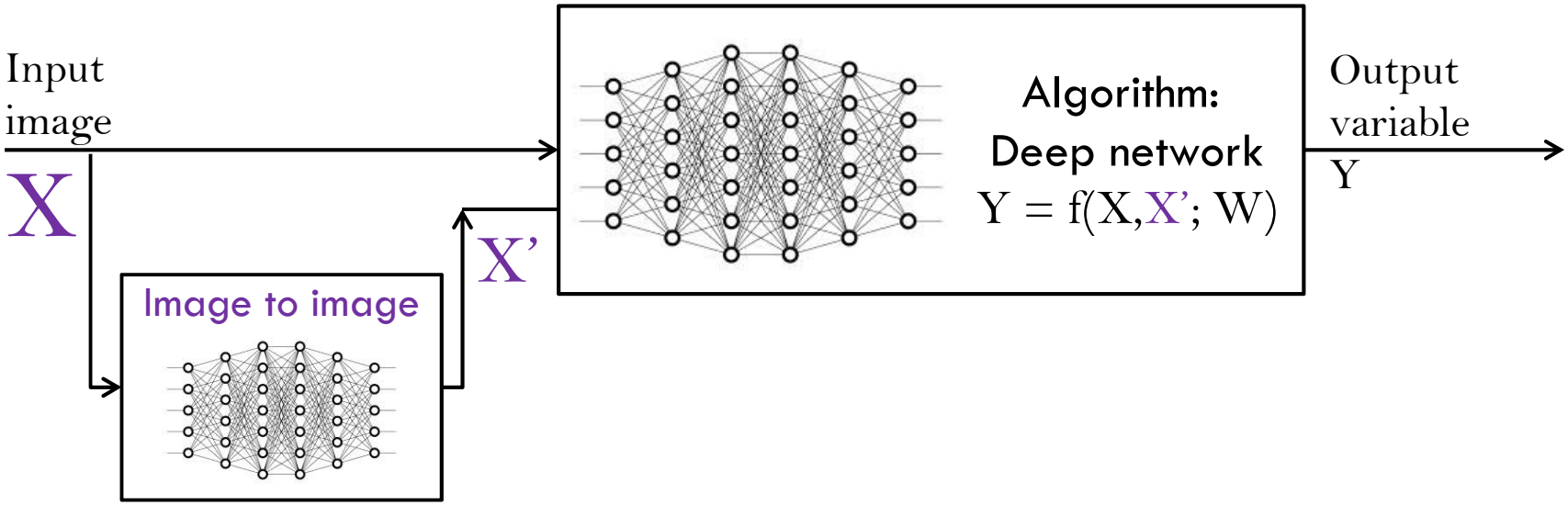
Knowledge in input



Knowledge in input

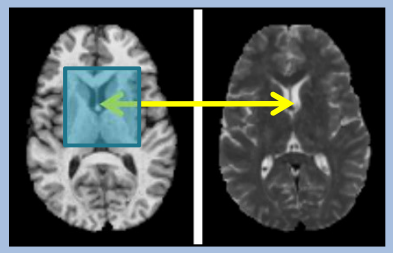
- Multi-modal inputs (RGBD, MR T1+T2, etc.)
- Synthesized inputs
- Other inputs

Synthesized inputs

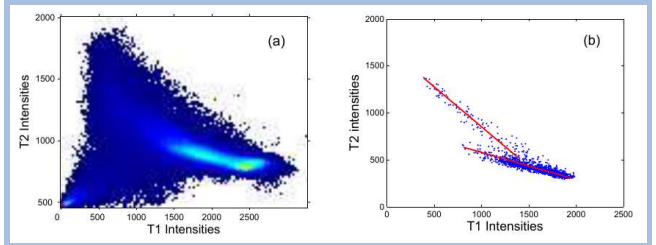


Supervised cross-domain image synthesis using location-sensitive deep network (LSDN) [MICCAI'2015]

Cross-domain image synthesis



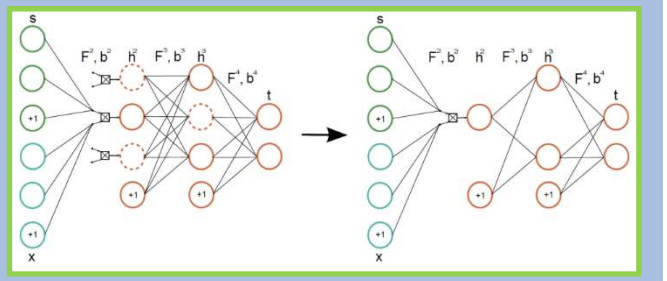
The importance of spatial info.



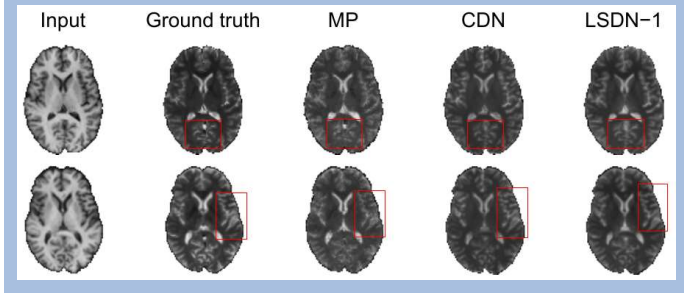
Whole image

Small region 10^3 voxels

Location-sensitive deep network (LSDN)



Accurate result



- Nguyen, et al. Cross-Domain Synthesis of Medical Images Using Efficient Location-Sensitive Deep Network, MICCAI 2015
- Vemulapalli, et al. Unsupervised Cross-modal Synthesis of Subject-specific Scans, ICCV 2015

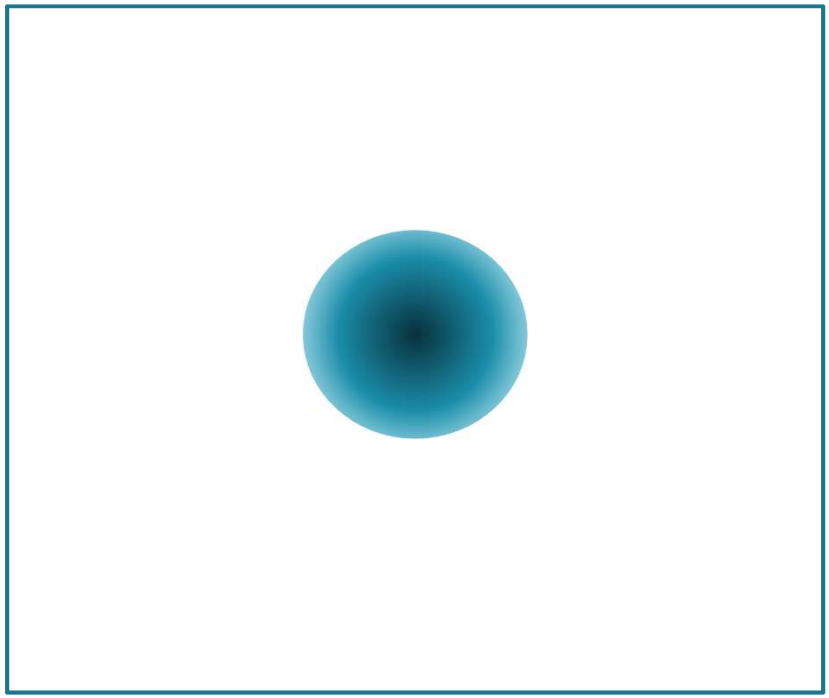
Knowledge in output

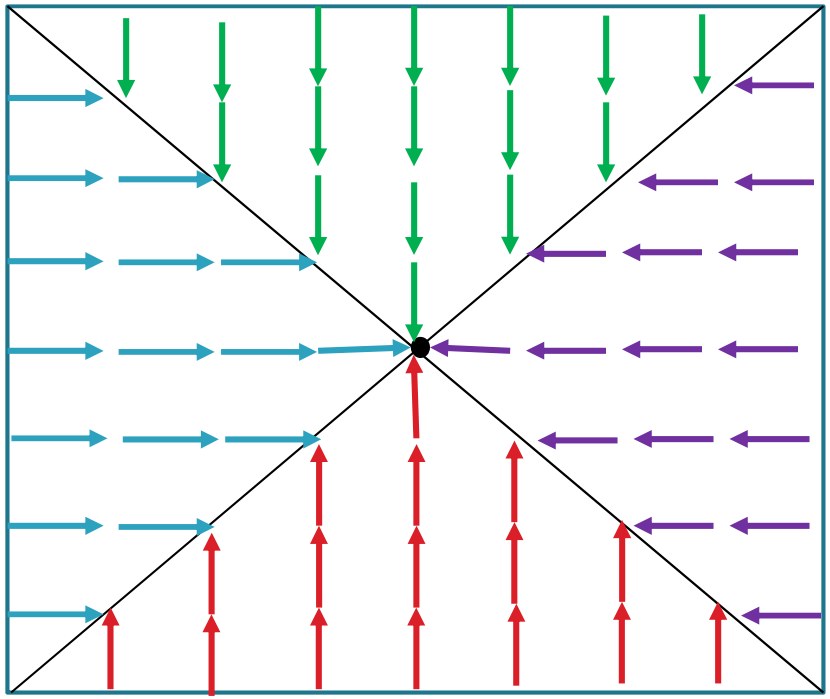
Knowledge in output

- New representation
- Multitask
- More priors

Landmark representation:

spatially local vs distributed



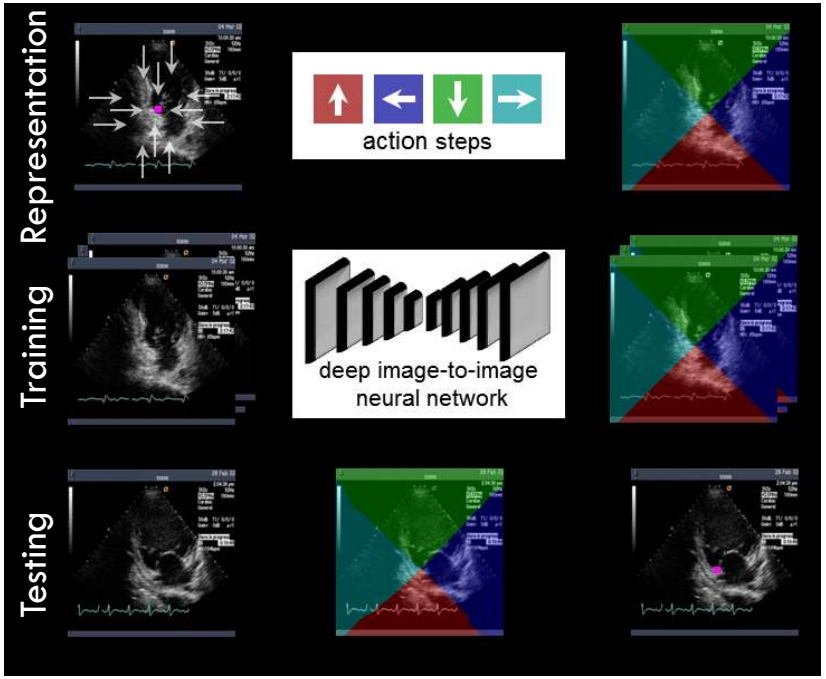


- Xu et al., Supervised Action Classifier: Approaching Landmark Detection as Image Partitioning, MICCAI 2017.

Landmark detection using DI2IN + supervised action map [MICCAI'2017]

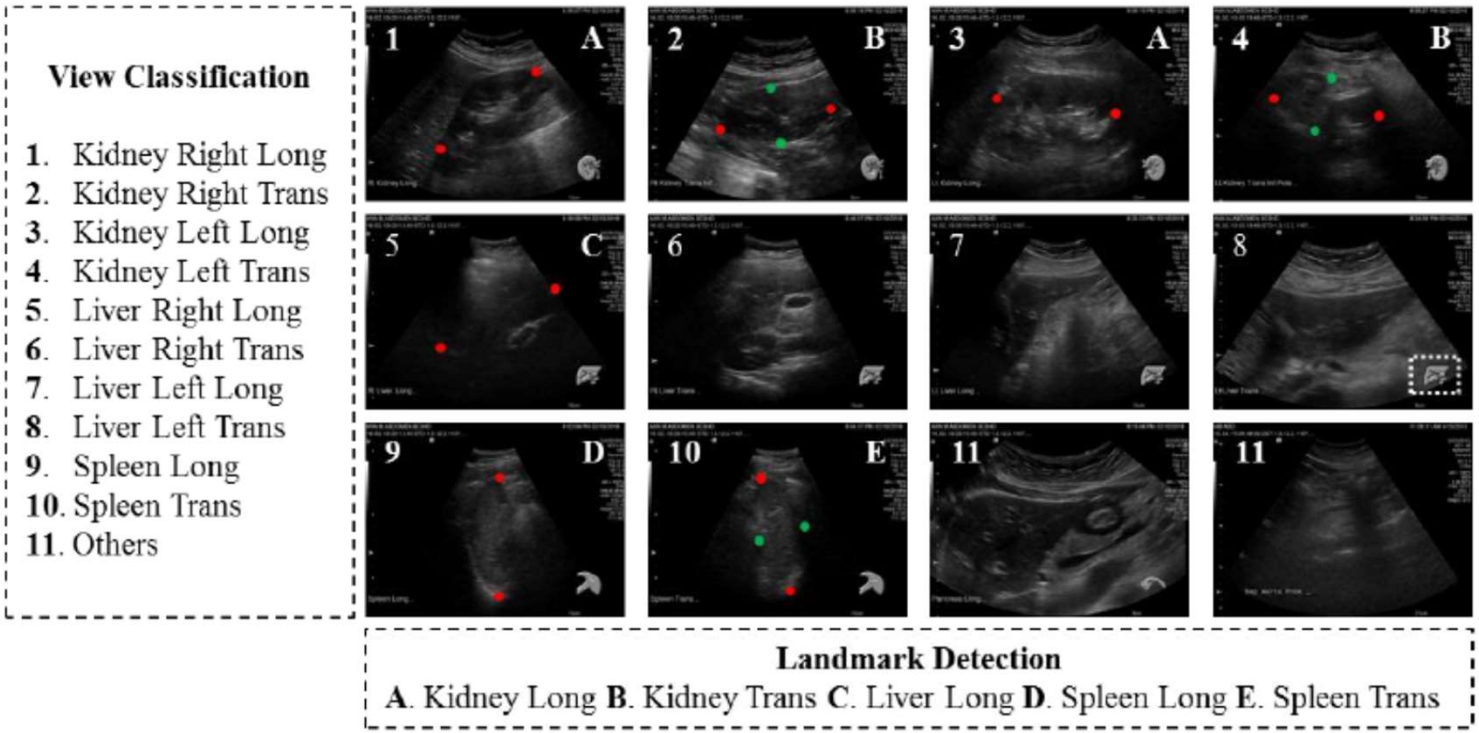
- Novel representation -- supervised action map
- Deep image2image network (DI2IN)

		PBT		DRL		I2I		SAC	
		lmk1	lmk2	lmk1	lmk2	lmk1	lmk2	lmk1	lmk2
CA	mean	10.45	13.85	7.69	10.02	6.73	9.02	**6.31**	**8.01**
	50%	5.74	8.11	5.43	7.63	5.00	6.40	**4.35**	**5.88**
	80%	11.11	16.18	9.33	13.73	8.54	11.40	**7.54**	**10.83**
OB	mean	59.23	130.66	29.99	32.45	30.07	21.97	**14.94**	**16.76**
	50%	35.31	139.49	11.69	13.17	5.39	6.08	**4.85**	**5.91**
	80%	109.84	193.64	43.98	45.76	13.34	15.54	**11.76**	**13.67**

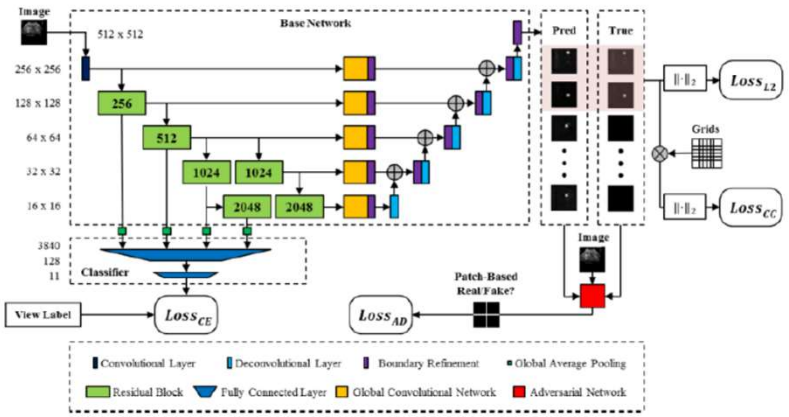


- Xu et al., Supervised Action Classifier: Approaching Landmark Detection as Image Partitioning, MICCAI 2017.

View classification and landmark detection for abdominal ultrasound images



Simultaneous view classification and landmark detection for abdominal ultrasound images



View classification

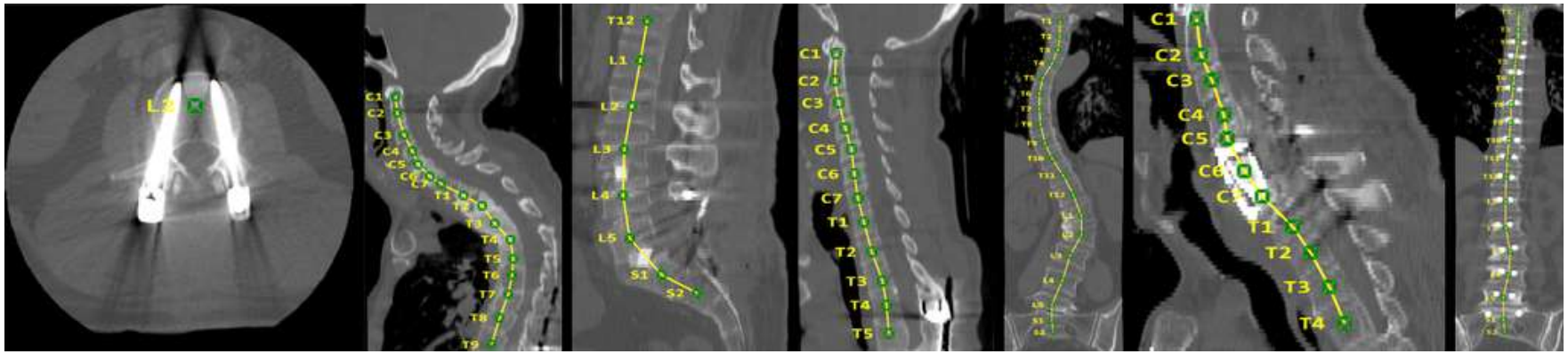
- MTL: 85.29%, STL: 81.22%, Human: 78.87%

Measurement

	KL_LA	KT_LA	KT_SA	LL_LA	SL_LA	ST_LA	ST_SA
Human	4.500	5.431	4.283	5.687	6.104	4.578	4.543
PBT [11]	11.036	9.147	8.393	11.083	7.289	9.359	12.308
SFCN	7.044	7.332	5.189	10.731	8.693	91.309	43.773
MGCN_R	**4.278**	**4.426**	**3.437**	**6.989**	**3.61**	**7.923**	**7.224**

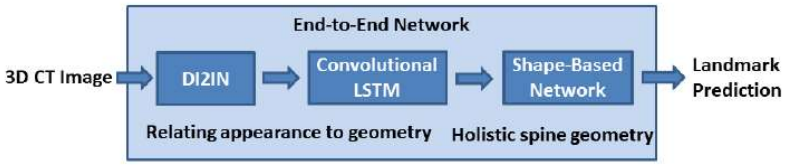
- Xu et al., Less is More: Simultaneous View Classification and Landmark Detection for Abdominal Ultrasound Images, MICCAI 2018 (accepted)

Vertebral landmark detection



Vertebral landmark detection using deep image2image recurrent network [MICCAI'2017]

- DL + shape constraints
- Reducing failure rate by 14% compared to best method on public benchmarking data set

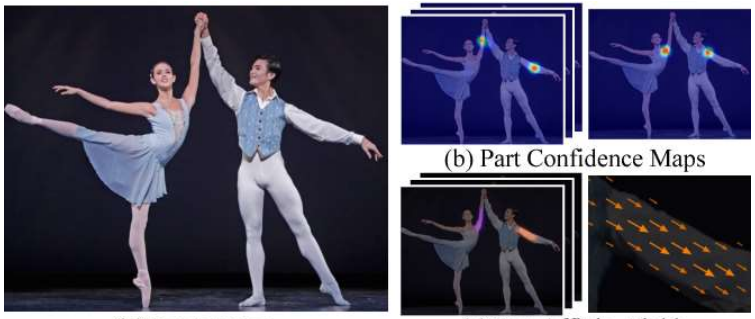


Region	Method	Set 1			Set 2		
		Mean	Std	Id.Rates	Mean	Std	Id.Rates
All	Glocker *et al.* [2]	12.4	11.2	70%	13.2	17.8	74%
	Suzani *et al* [4]	18.2	11.4	-	-	-	-
	Chen *et al.* [3]	-	-	-	8.8	13.0	84%
	Our method	10.6	**8.7**	78%	8.7	8.5	85%
	Our method +1000	**9.0**	8.8	**83%**	**6.9**	**7.6**	**89%**

- Yang et al., Deep Image-to-Image Recurrent Network with Shape Basis Learning for Automatic Vertebra Labeling in Large-Scale 3D CT Volumes, MICCAI 2017

Multi-person pose estimation [Cao et al CVPR2017]

Figure 1. **Top:** Multi-person pose estimation. Body parts belong-



(a) Input Image

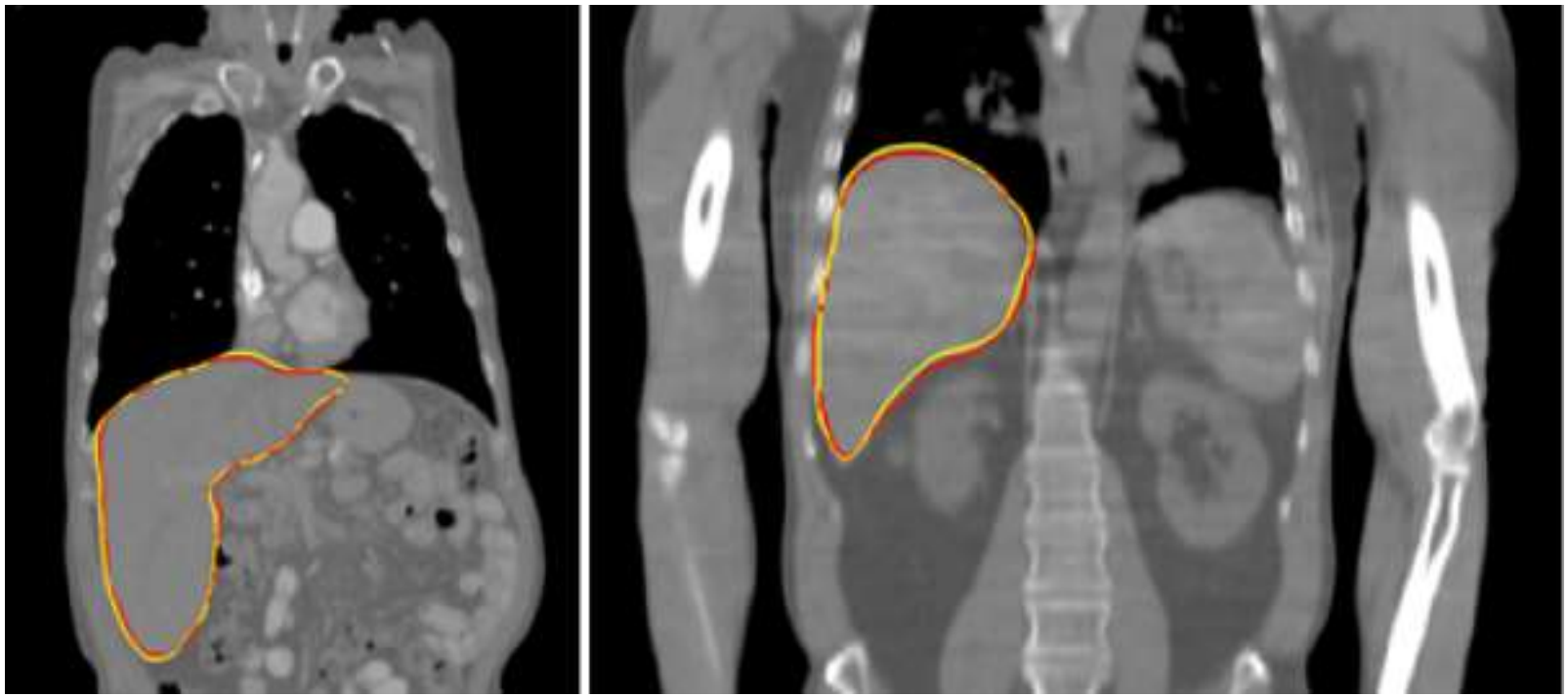
(b) Part Confidence Maps

(c) Part Affinity Fields

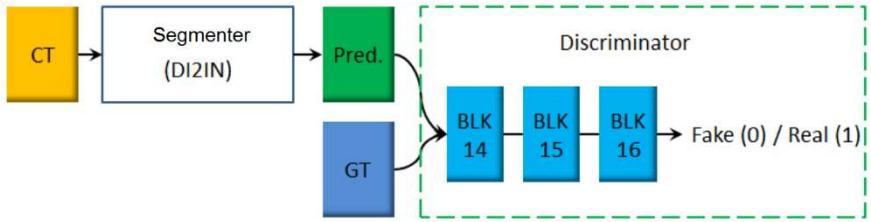
(d) Bipartite Matching

(e) Parsing Results

Organ contouring



Adversarial image2image network for organ contouring [MICCAI'2017]



- Using image2image network and adversarial shape prior
- Liver segmentation: 34% error reduction when using 1000 CT data sets

Method	ASD (mm)			
	Mean	Std	Max	Median
Ling *et al.* (400) [5]	2.95	5.07	37.45	2.01
DI2IN (400)	2.38	1.31	10.35	2.0
DI2IN-AN (400)	2.09	0.94	7.94	1.88
DI2IN (1000)	2.15	0.81	6.51	1.95
DI2IN-AN (1000)	**1.95**	0.75	**6.48**	**1.81**

- Yang et al., Automatic Liver Segmentation Using an Adversarial Image-to-Image Network, MICCAI 2017

Segmentation representation

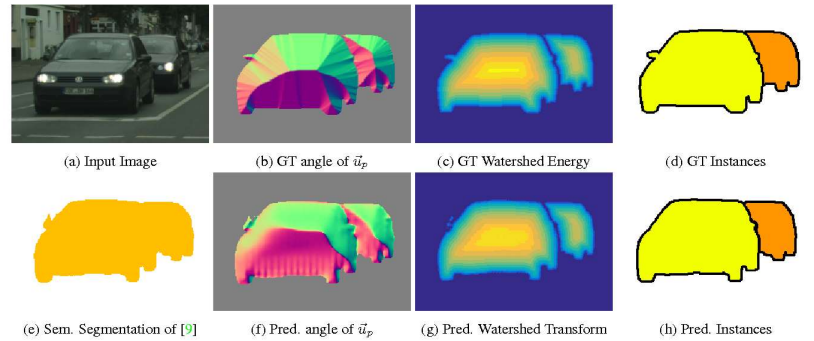
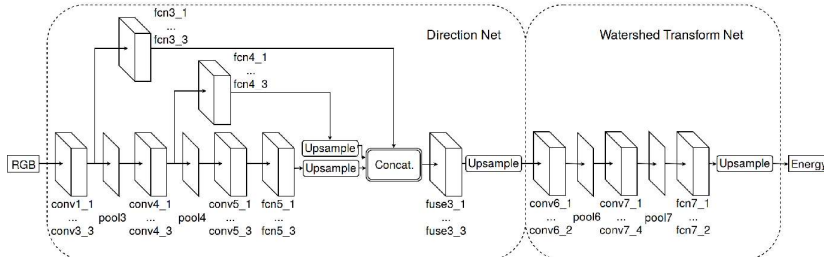
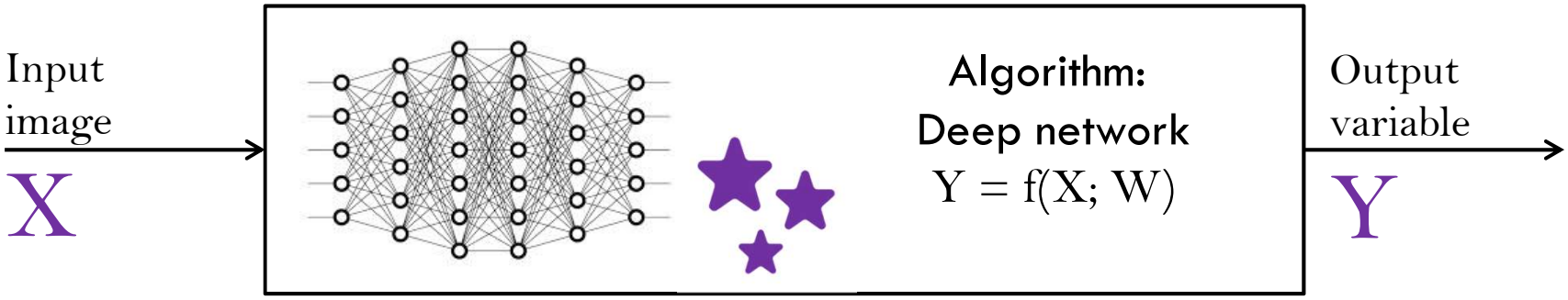


Figure 3: Steps in our pipeline. Note that f) shows the output of the DN module after end-to-end fine-tuning.



Deep Watershed Transform for Instance Segmentation, ArXiv 1611.08303

Knowledge in algorithm

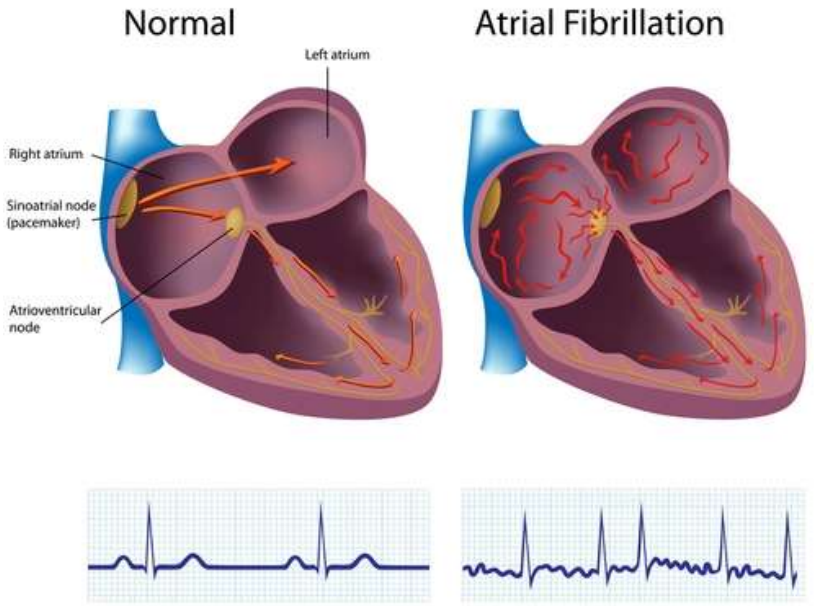


Knowledge in algorithm

- Network design (no more brute force)

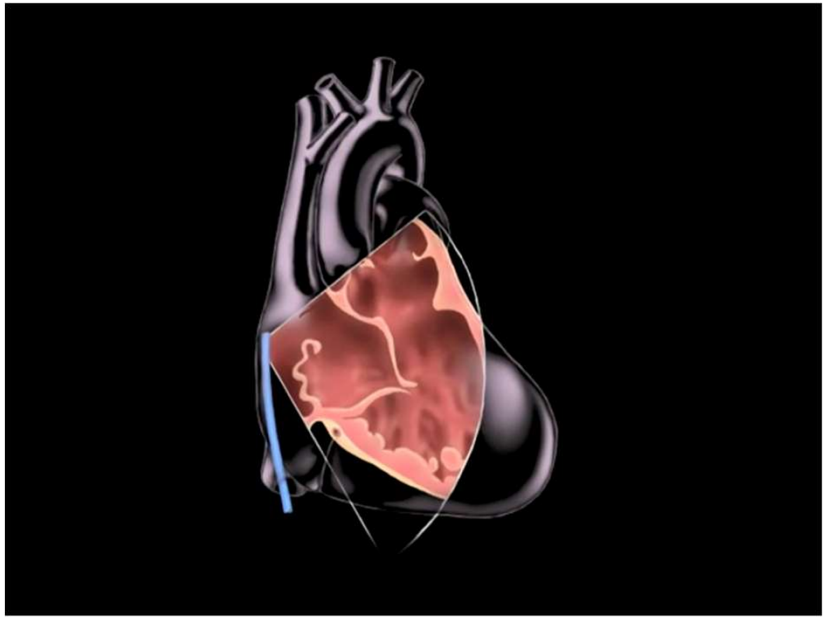
Atrial fibrillation

- Irregular, often rapid heart beat
- ~2.7–6.1M people in the US. With the aging of the U.S. population, this number is expected to increase.
- ~2% of young people (<65) and ~9% of senior people have AFib.
- AFib costs the United States about $6 billion each year.

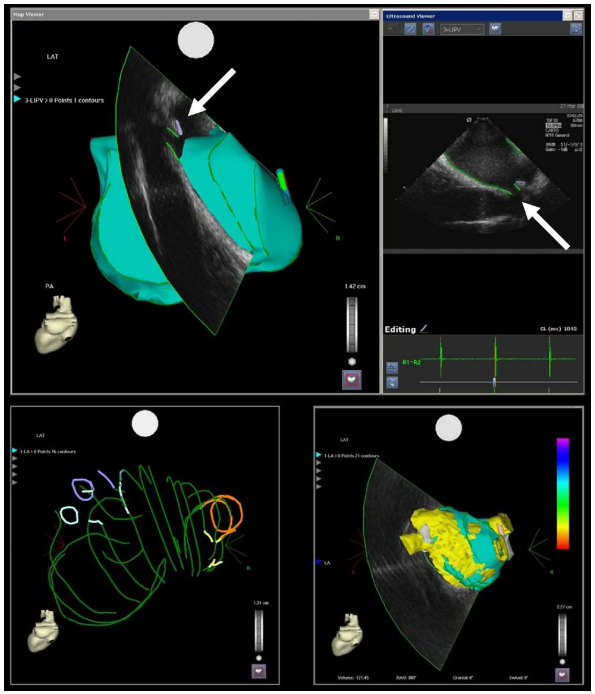


https://www.cdc.gov/dhdsp/data_statistics/fact_sheets/images/afib_heart.jpg

Cardiac ablation using intracardiac echocardiograph (ICE)



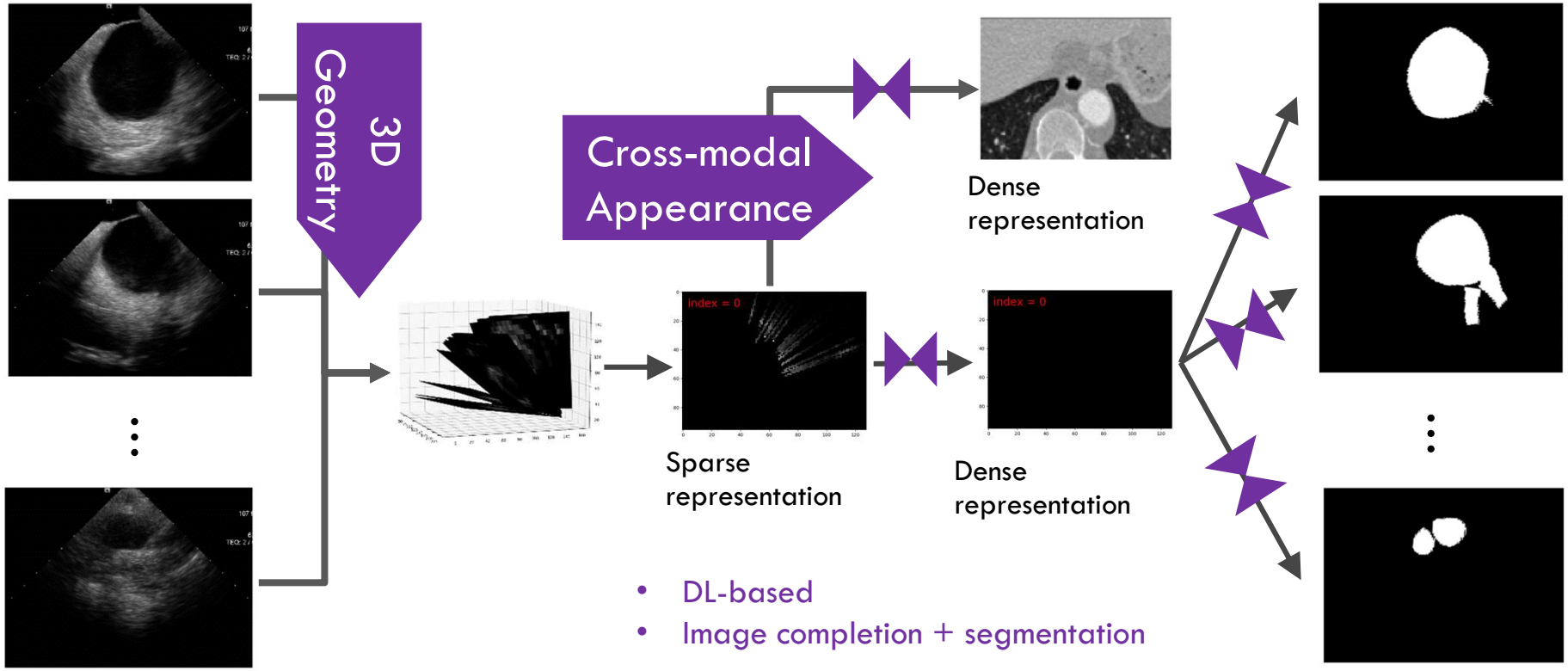
https://www.youtube.com/watch?v=Z03dJ8TG5_8&t=295s



http://imaging.onlinejacc.org/content/jimg/2/4/498/F2.large.jpg

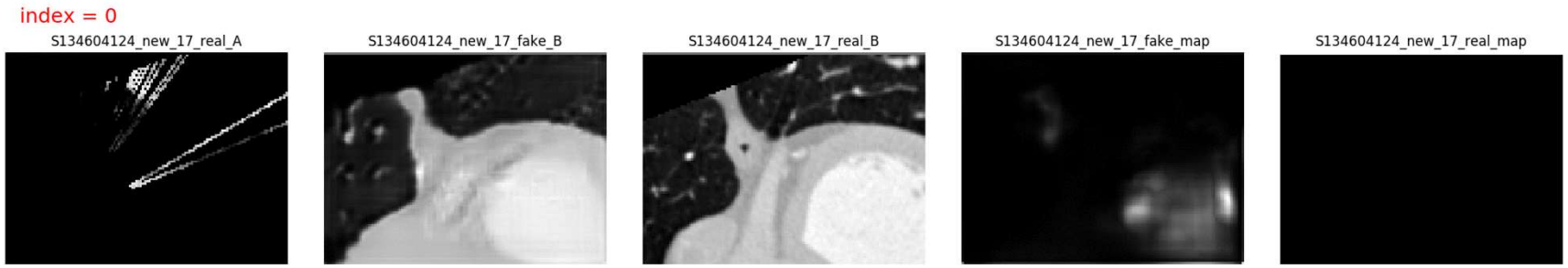
ICE auto contouring:

A knowledge-fused DL algorithm



- DL-based
- Image completion + segmentation

Results



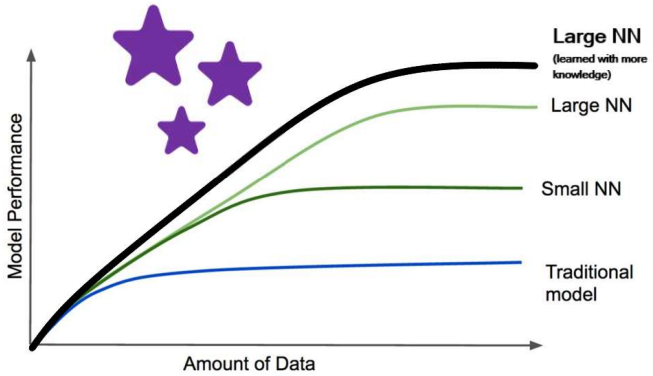
	LA	LAA	LIPV	LSPV	RIPV	RSPV	Total
2D only	0.926	0.443	0.553	0.483	0.549	0.242	0.872
3D only	0.907	0.363	0.546	0.418	0.603	0.403	0.853
2D + 3D	0.942	0.658	0.706	0.620	0.718	0.395	0.898

3D network without appearance knowledge doesn't converge!

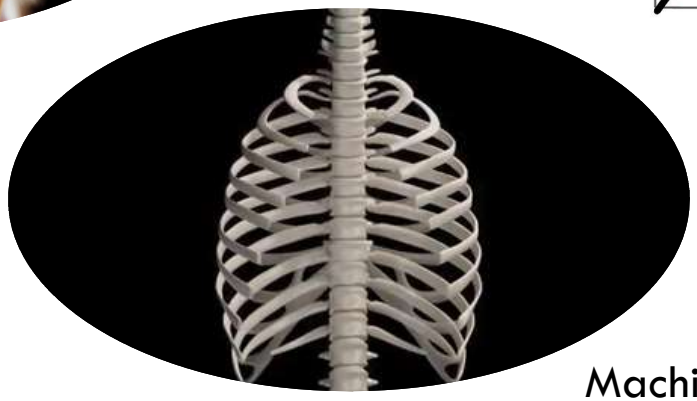
- Liao et al. More knowledge is better: Cross-domain volume completion and 3D+2D segmentation for intracardiac echocardiography contouring, MICCAI 2018 (accepted)

Recap

Medical imaging trends



'Knowledge' gap



Machine learning + knowledge