GTCx

MELBOURNE | Oct.4-5, 2016

DEEP NEURAL NETWORKS IN RADIOLOGY: PREVENTATIVE AND PRECISION MEDICINE PERSPECTIVES

Le Lu, PhD, NIH-CC, Oct. 26th, 2016 (GTC DC Talk DCS16103)

NIH National Institutes of Health
Turning Discovery Into Health

PRESENTED BY
NVIDIA

GTCx

OUTLINES

What does they mean precisely for preventative and precision medicine perspectives in radiology or medical imaging?

Do deep learning and deep neural networks help in medical imaging or medical image analysis problems? (Yes)

- Lymph node application package (52.9% → 85%, 83%)
- Pancreas application package (~53% → 81.14% in Dice Similarity Coefficient)
- Lung (Interstitial Lung Disease) application package + DL reading chest x-ray
- Unsupervised category discovery using looped deep pseudo-task optimization (mapping large-scale radiology database with category meta-labels)

COMPLEXITY & COMPOSABILITY

The Partnership of the Future

Microsoft's CEO explores how humans and A.I. can work together to solve society's greatest challenges.

By Satya Nadella

From left, visually impaired Microsoft developer Saqib Shaikh stands next to CEO Satya

http://www.slate.com/articles/technology/future_tense/2016/06/microsoft_ceo_satya_nadella_humans_and_a_i_can_work_together_to_solve_society.html

GTCx

PERSPECTIVES

Why the previous or current computer-aided diagnosis (CADx) systems are not particularly successful? **Integrating machine decisions is not easy for human doctors**: Good doctors hate to use; bad doctors are confused and do not know how to use? --> **Human-machine collaborative decision making process**

> Make machine decision more **interpretable** is very critical for the collaborative system --> learning mid-level attributes or embedding?

Preventive medicine: what human doctors cannot do (in very large scales: millions of general population, at least not economical): → **first-reader population risk profiling** …?

Precision Medicine: a), **new imaging biomarker** in precision medicine to **better assist human doctors to make more precise decisions;** b), **patient level similarity retrieval system** for diagnosis/therapy systems: show by examples!

APPLICATION FOCUS: CANCER IMAGING

Cancer Type	Lung (Bronchus)	Colorectal	Pancreatic	Breast (Female-Male)	Prostate
Estimated New Cases	224,390	134,490	53,070	246,660 – 2,600	180,890
Estimated Deaths	158,080	49,190	41,780	40,450 – 440	26,120

Cancer Facts and Figures 2016. Atlanta, Ga: American Cancer Society, 2016. Last accessed February 1, 2016.
http://www.cancer.gov/types/common-cancers

OVERVIEW: THREE CATEGORIES OF KEY PROBLEMS (I)

1. **Computer-aided Detection (CADe) and Diagnosis (CADx)**

Lung, Colon pre-cancer detection; Bone and Vessel imaging (6 years of industrial R&D at Siemens Corporation and Healthcare, 10+ product transfer; 13 conference papers in CVPR/ECCV/ICCV/MICCAI/WACV/CIKM, 12 US/EU patents, 27 Inventions)

Lymph node, colon polyp, bone lesion detection using Deep CNN + Random View Aggregation (http://arxiv.org/abs/1505.03046, TMI 2016a; MICCAI 2014a)

Empirical analysis on **Lymph node detection** and **interstitial lung disease** (ILD) classification using CNN (http://arxiv.org/abs/1602.03409, TMI 2016b)

Non-deep models for CADe using **compositional** representation (MICCAI 2014b) and **+mid-level cues** (MICCAI 2015b); **deep regression** based **multi-label** ILD prediction (*in submission*); missing label issue in ILD (ISBI 2016)

- **Clinical Impact**: producing various *high performance* “second or first reader” **CAD use cases** and applications → effective imaging based prescreening tools on a cloud based platform for large population

GTCx

OVERVIEW: THREE CATEGORIES OF KEY PROBLEMS (II)

2. Semantic Segmentation in Medical Image Analysis

"DeepOrgan" for **pancreas segmentation** (MICCAI 2015a) via scanning superpixels using multi-scale deep features ("Zoom-out") and probability map embedding http://arxiv.org/abs/1506.06448

Deep segmentation on **pancreas** and **lymph node clusters** with HED (Holistically-nested neural networks, Xie & Tu, 2015) as building blocks to learn unary (segmentation **mask**) and pairwise (labeling segmentation **boundary**) CRF terms + spatial aggregation or + structured optimization (The focus of three MICCAI 2016 papers since this is a much needed task → **Small** datasets; **(de-)compositional** representation is still the key.)

CRF: conditional random fields

- **Clinical Impact**: semantic segmentation can help compute clinically more accurate and desirable imaging bio-markers or precision measurement!

GTCx

OVERVIEW: THREE CATEGORIES OF KEY PROBLEMS (III)

3. **Interleaved or Joint Text/Image Deep Mining** on a Large-Scale Radiology Image Database → **"large"** datasets; **no labels (~216K 2D key images/slices extracted from >60K unique patients)**

Interleaved Text/Image Deep Mining on a Large-Scale Radiology Image Database (CVPR 2015, a proof of concept study)

Interleaved Text/Image Deep Mining on a Large-Scale Radiology Image Database for Automated Image Interpretation (its extension, JMLR, 17(107):1−31, 2016) http://arxiv.org/abs/1505.00670

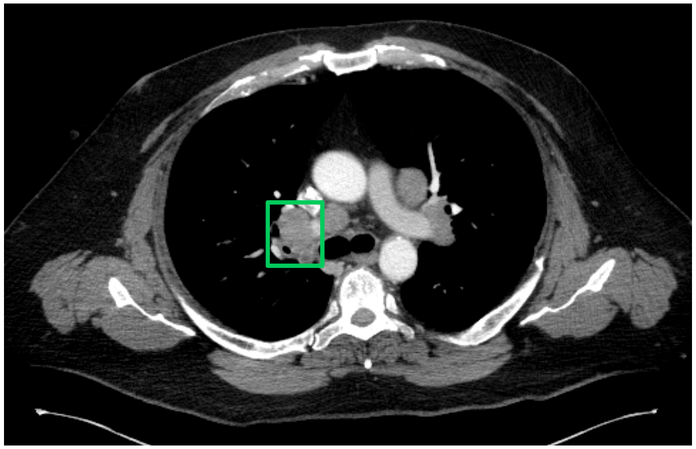
Learning to Read Chest X-Rays: Recurrent Neural Cascade Model for Automated Image Annotation, (CVPR 2016) http://arxiv.org/abs/1603.08486

Unsupervised Category Discovery via Looped Deep Pseudo-Task Optimization Using a Large Scale Radiology Image Database, http://arxiv.org/abs/1603.07965

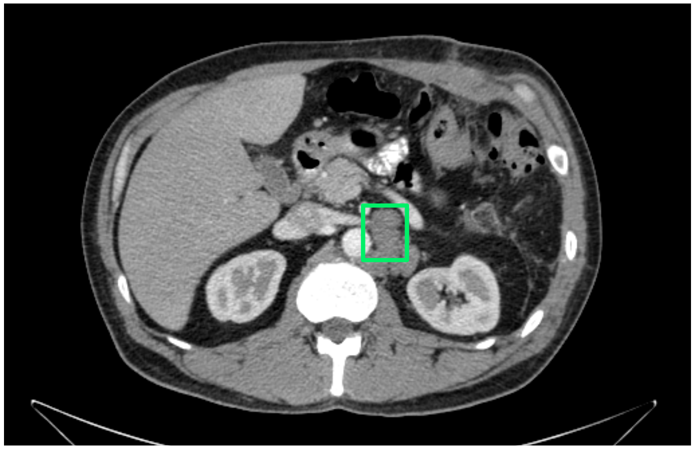
- **Clinical Impact**: eventually to build an automated programmable mechanism to parse and learn from hospital scale PACS-RIS databases to derive semantics and knowledge … has to be *deep learning* based since effective image features are very hard to be hand-crafted cross different diseases, imaging protocols and modalities.

(A.0) Automated Lymph Node Detection

- **Difficult due to large variations in appearance, location and pose.**
- **Plus low contrast against surrounding tissues.**



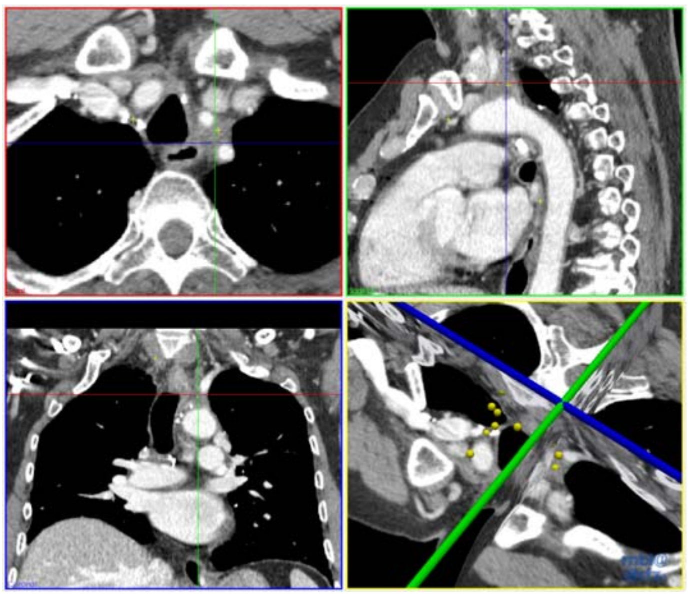
Mediastinal lymph node in CT

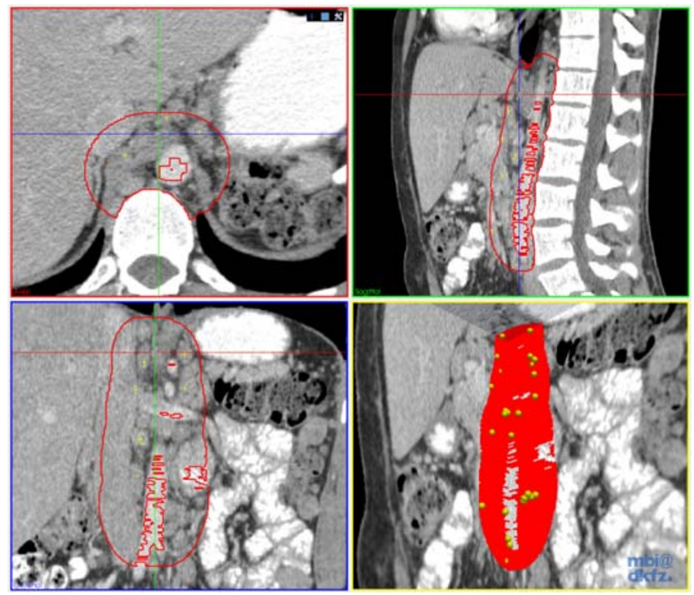


Abdominal lymph node in CT

Lymph Node Candidate Generation

- Mediastinum [J. Liu et al. 2014]
  - 388 lymph nodes in 90 patients
  - 3208 false-positives
    - 36 FPs per patient
- Abdomen [K. Cherry et al. 2014]
  - 595 lymph nodes in 86 patients
  - 3484 false-positives
    - 41 FPs per patient





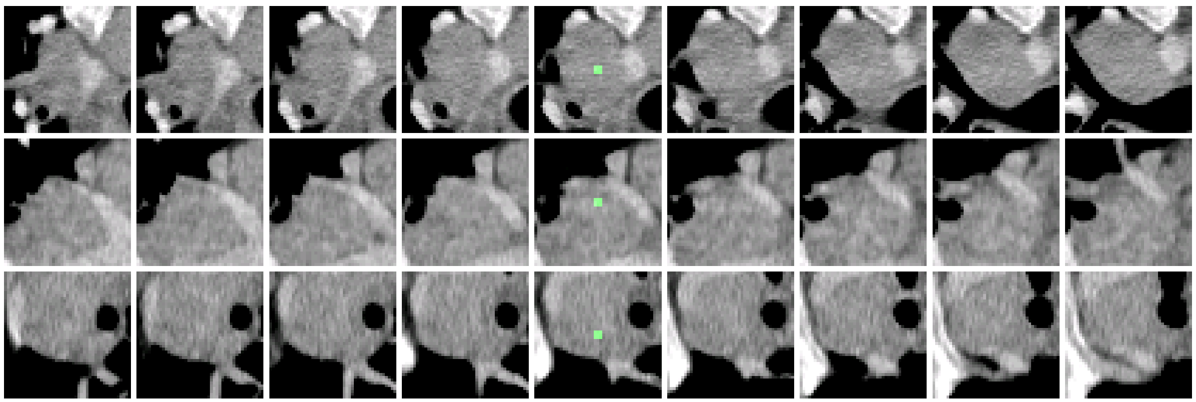
Shallow Models: 2D View Aggregation Using a Two-Level Hierarchy of Linear Classifiers **[Seff et al. MICCAI 2014]**

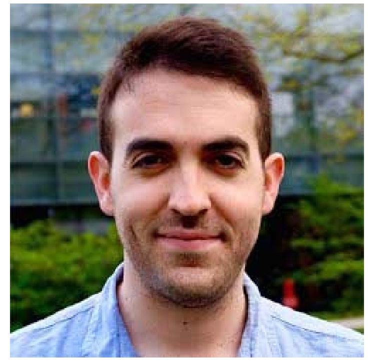
- VOI candidates generated via a random forest classifier using voxel-level features (not the primary focus of this work), for high sensitivity but also high false positive rates.
- **2.5D:** 3 sequences of orthogonal 2D slices then extracted from each candidate VOI (9 x 3 = 27 views).

Axial

Coronal

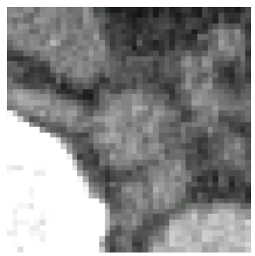
Sagittal





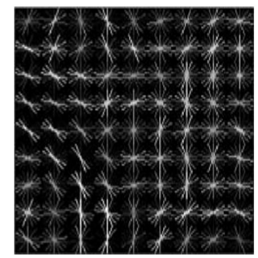
NDSEG Fellowship

HOG: Histogram of Oriented Gradients + LibLinear on recognizing 2D Views (state-of-the-art before DCNN)

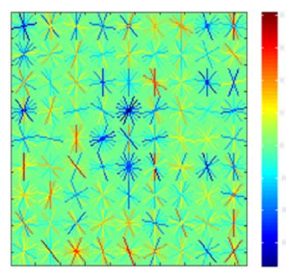


Abdominal LN axial slice.

HOG feature extraction



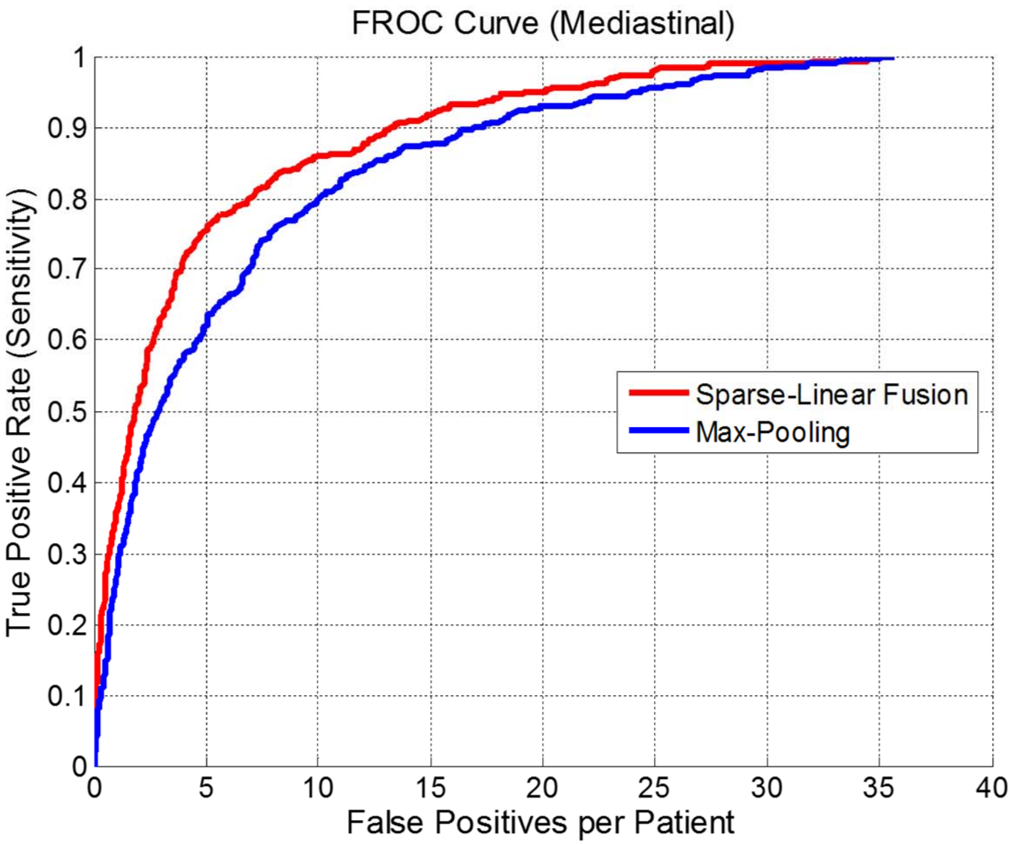
SVM training



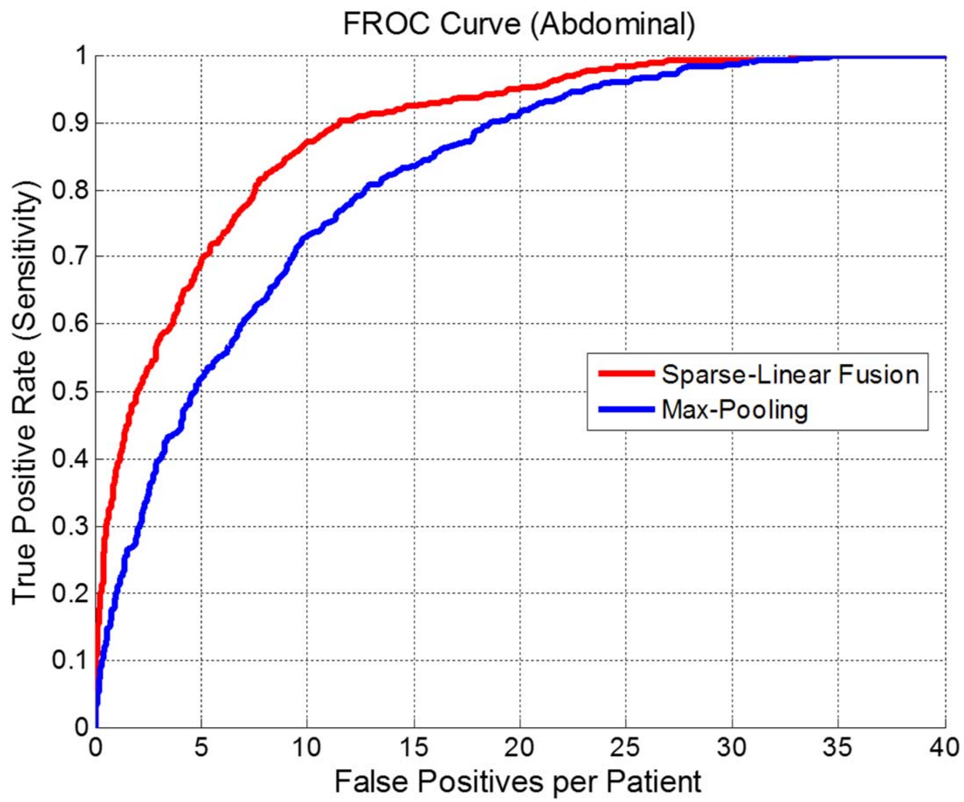
Resulting feature weights after training.

Note that a unified, compact HOG model is trained, regardless of axial, coronal, or sagittal views, or unifying view orientations.

Lymph Node Detection FROC Performance



Lymph Node Detection FROC Performance



- Enriching HOG descriptor with other image feature channels, e.g., mid-level semantic contours/gradients, can further lift the sensitivity for 8~10%!
- About 1/3 FPs are found to be smaller lymph nodes (short axis < 10 mm).

MAKE SHALLOW TO GO DEEPER VIA MID-LEVEL CUES?

[SEFF ET AL. MICCAI 2015]

We explore a learned transformation scheme for producing enhanced semantic input for HOG, based on LN-selective visual responses.

Mid-level semantic boundary cues learned from segmentation.

All LNs in both target regions are manually s

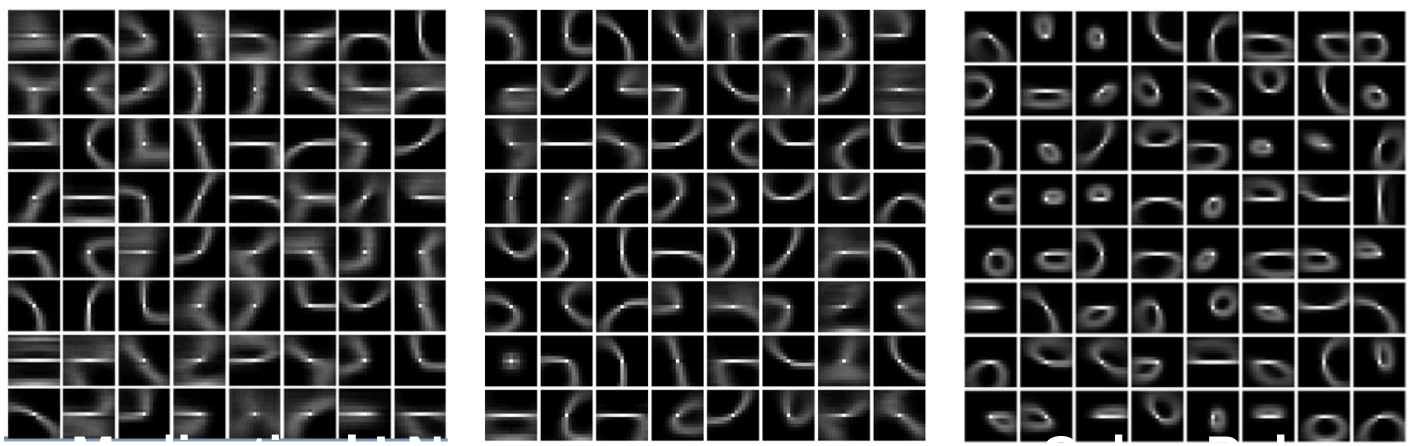
Target region	# Patients	# LNs
Mediastinal	90	389
Abdominal	86	595

LEARNING SUPERVISED BOUNDARY CUES VIA SKETCH TOKENS [CVPR'13]

Extract all patches (radius = 7 voxels) centered on a boundary pixel

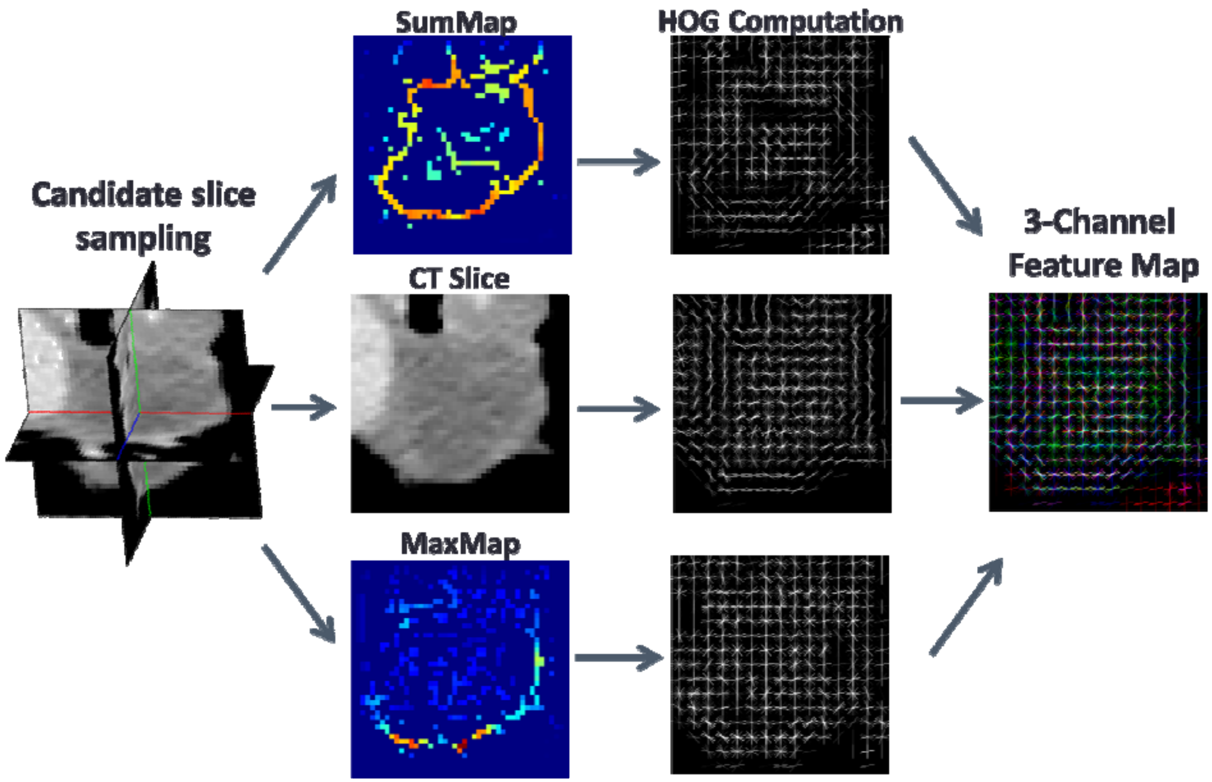
Cluster into "sketch token" classes using k-means with k = 150

A random forest is trained for sketch token classification for input CT patches



MULTI-CHANNEL HOG FEATURE MAP CONSTRUCTION

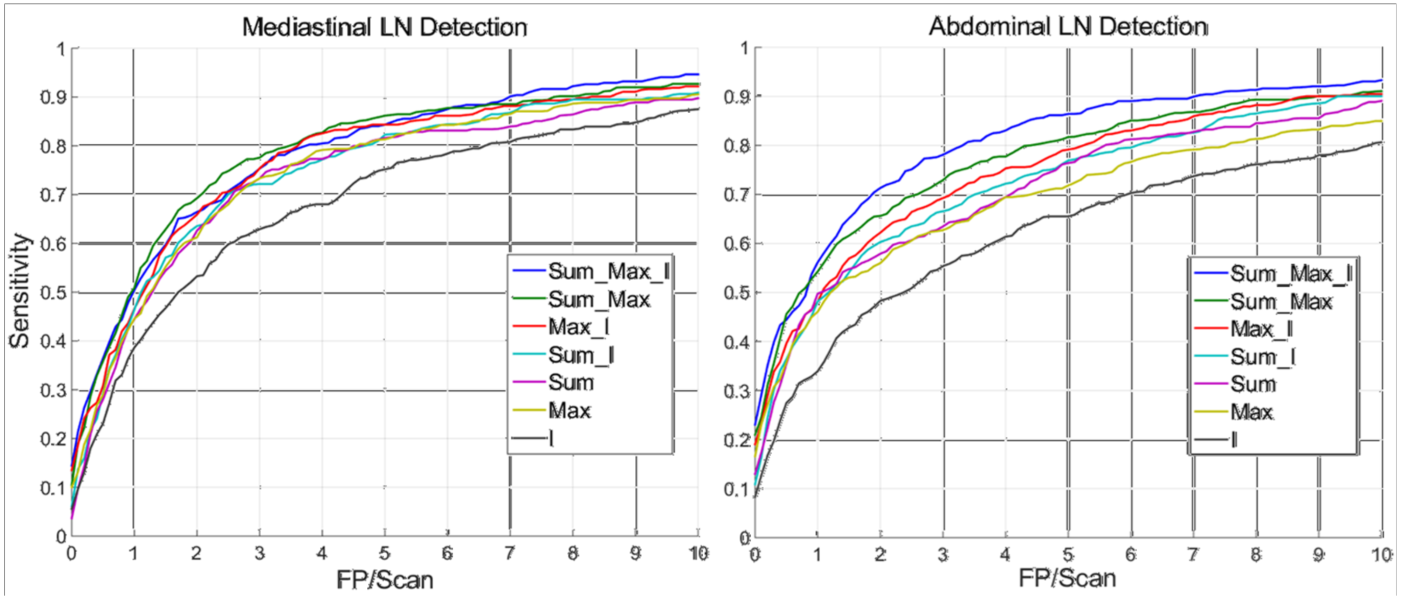
An enhanced, 3-channel feature map:



SINGLE TEMPLATE RESULTS

Top performing feature sets (*Sum_Max_I* and *Sum_Max*) exhibit 15%-23% greater recall than the baseline HOG at low FP rates (e.g. 3/FP scan).

- Our system outperforms the state-of-the-art deep CNN system (Roth et al., 2014) in the mediastinum, e.g. 78% vs. 70% at 3 FP/scan.



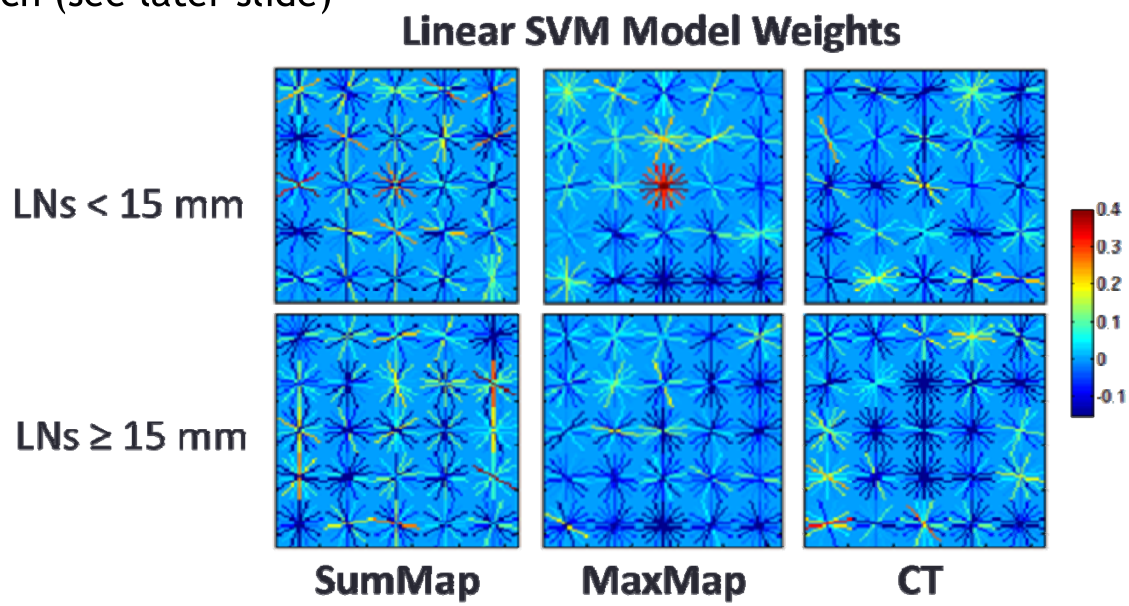
Six-fold Cross-Validation FROC curves are shown for the two body regions

GTCx

CLASSIFICATION

A linear SVM is trained using the new feature set; A HOG cell size of 9x9 pixels gives optimal performance.

Separate models are trained for specific LN size ranges to form a mixture-of-templates-approach (see later slide)



Visualization of linear SVM weights for the abdominal LN detection models

Mixture Model Detection Results

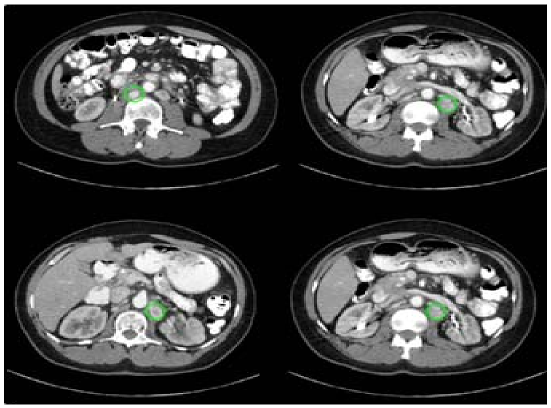
Wide distribution of LN sizes invites the application of size-specific models trained separately.

LNs > 20 mm are especially clinically relevant

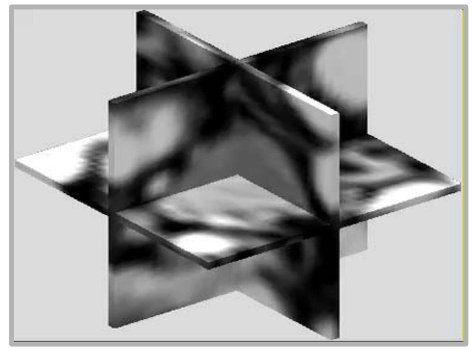
Single template and mixture model performance for abdominal models

(A.1) Deep models: Random Sets of Convolutional Neural Network Predictions via Compositional Representation

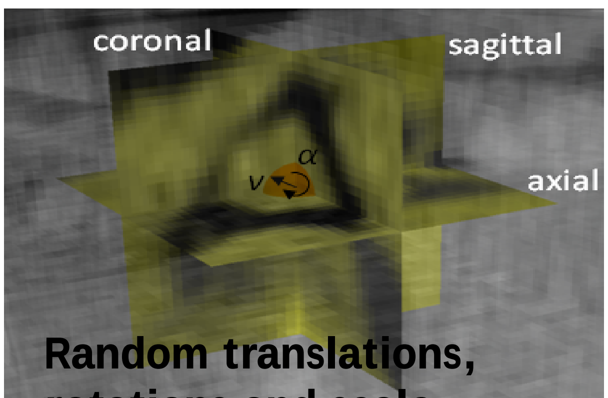
[Roth et al. MICCAI 2014, Shin et al. TMI 2016; Roth et al. TMI 2016]



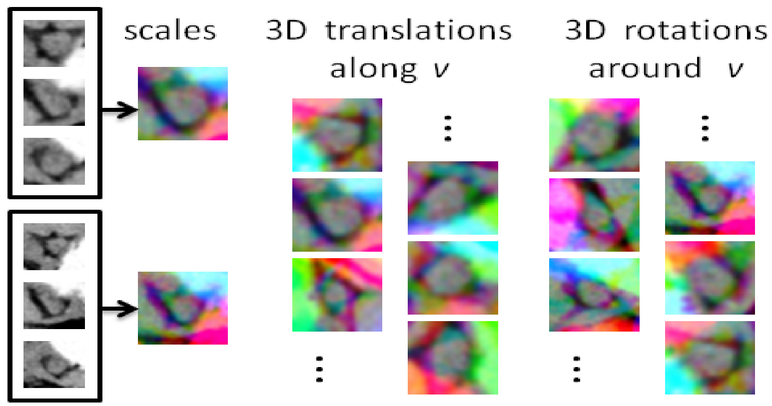
Application to appearance modeling and detecting lymph node



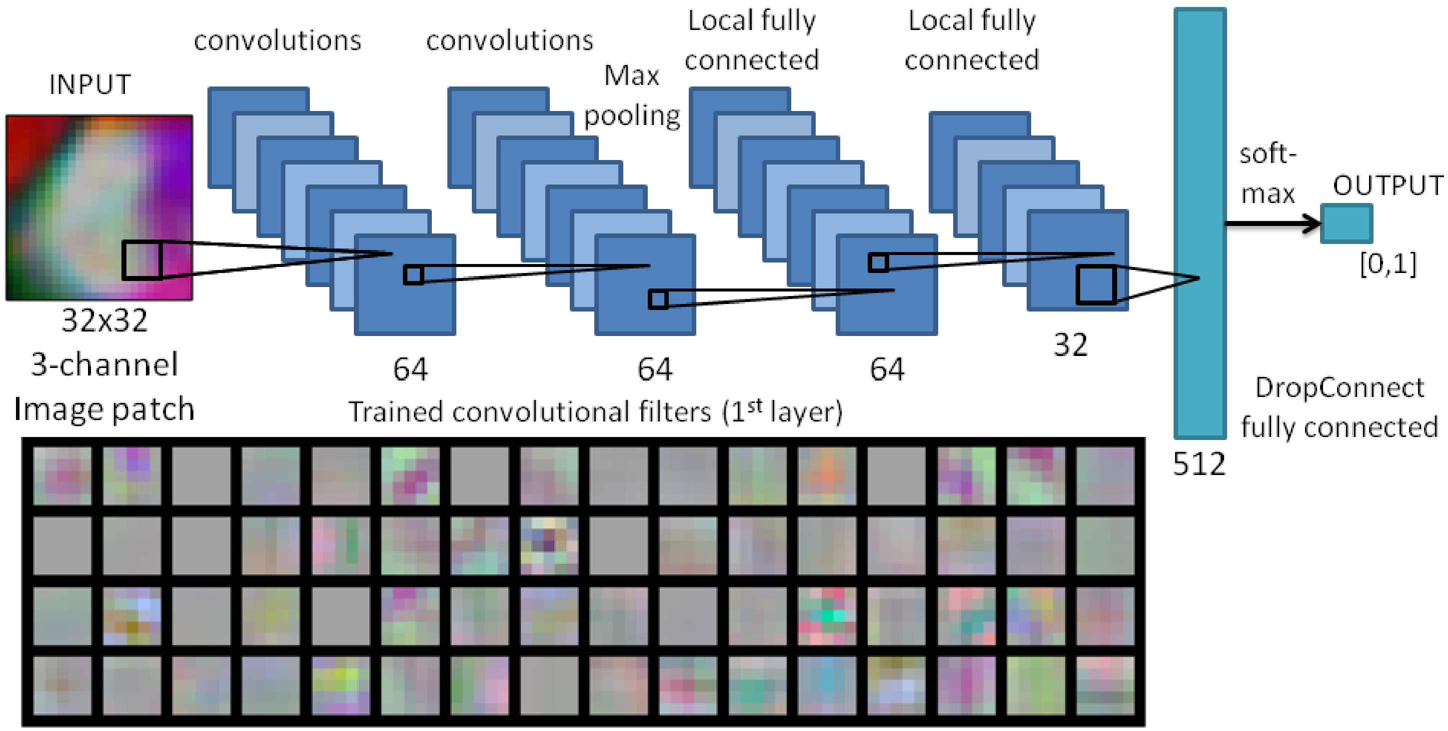
Asst. Prof. Nagoya University (now)



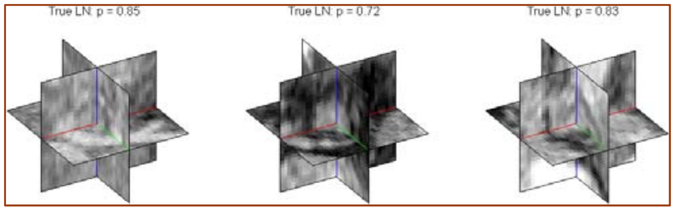
Random translations, rotations and scale

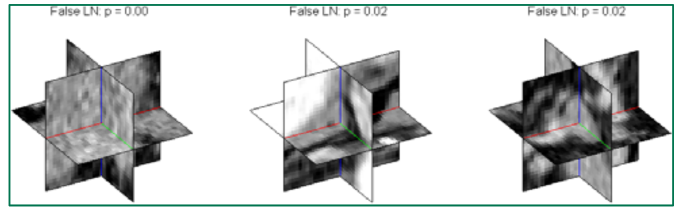


Convolutional Neural Network Architecture



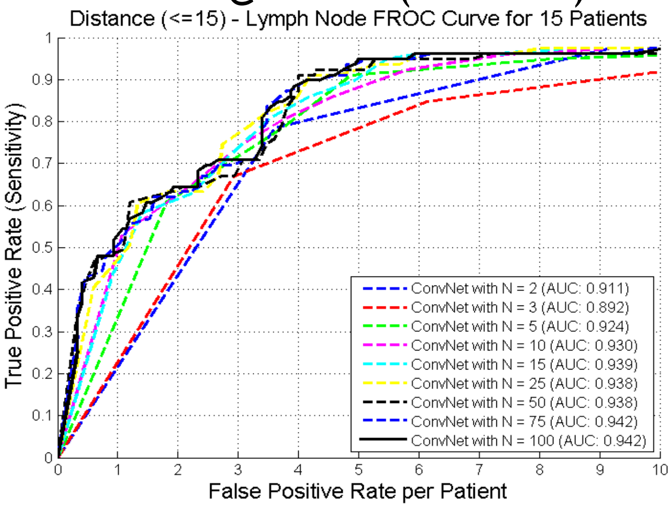
Experimental Results (~100% sensitivity but ~40 FPs/patient at candidate generation step; then 3-fold CV with data augmentation)





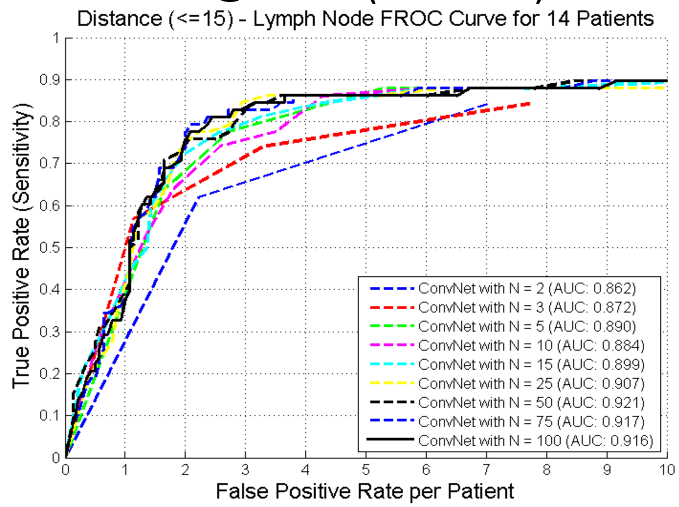
- Mediastinum

71% @ 3 FPs (was 55%)



- Abdomen

83% @ 3 FPs (was 30%)

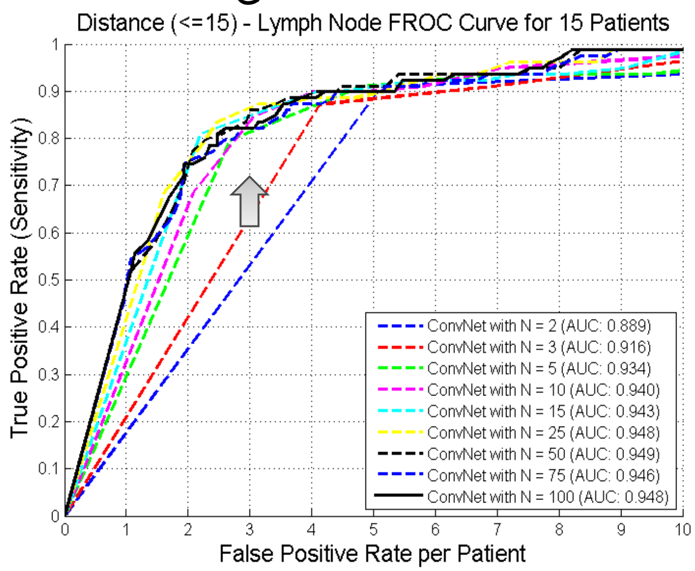


Experimental Results (cont.)

Training mediastinum and abdomen Jointly!

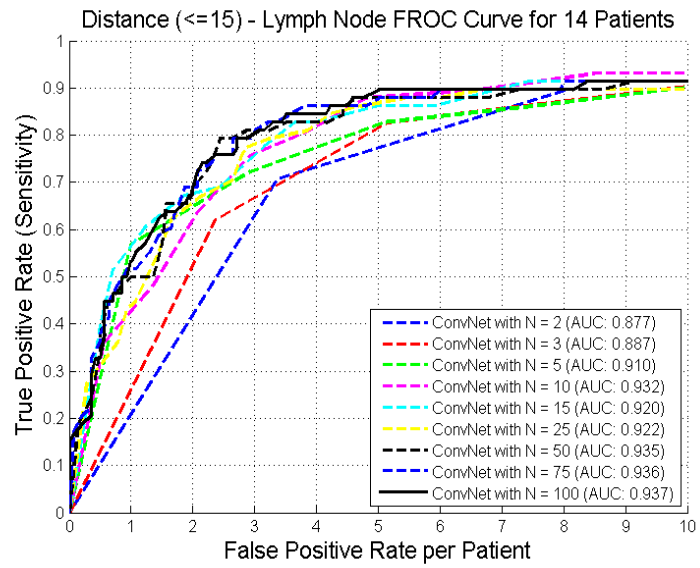
- Mediastinum

82% @ 3 FPs



- Abdomen

80% @ 3 FPs



GTCx

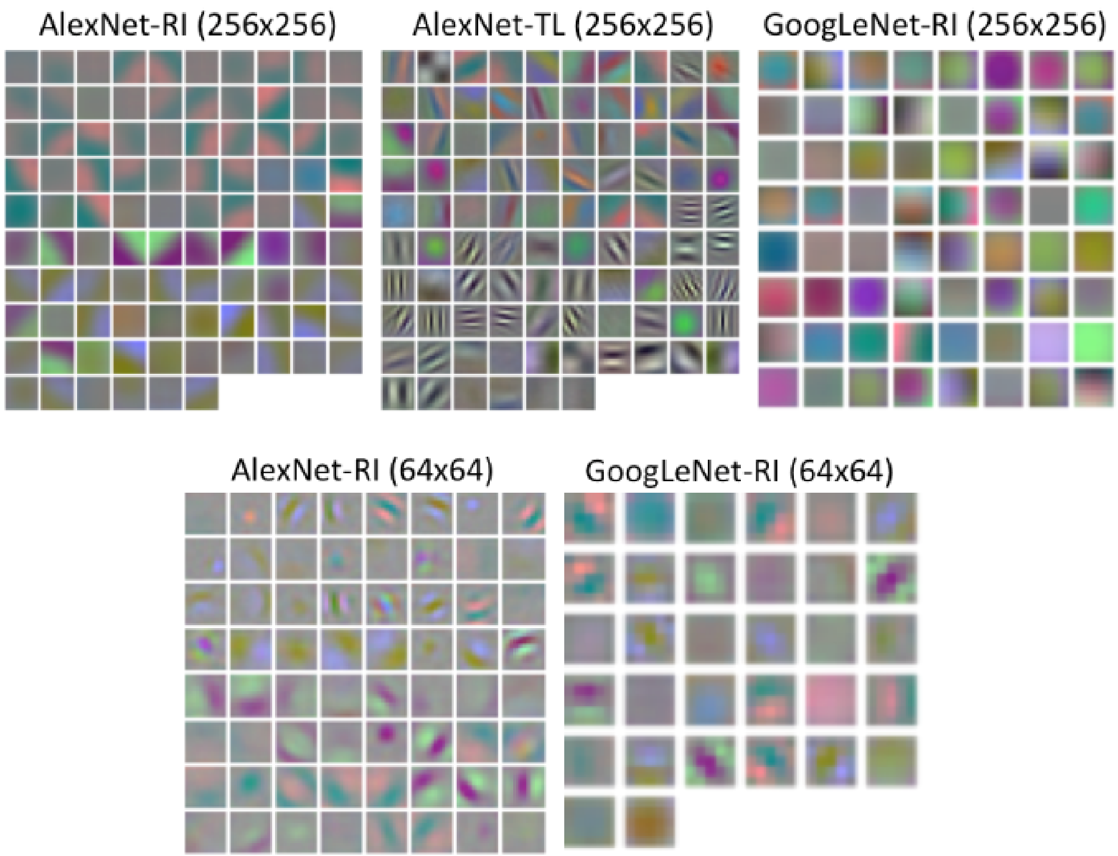
WHERE ARE WE? COMPARING TO PREVIOUS WORK (CAD 1.0 OR 2.0)

1) The previous state-of-the-art work is (Feulner et al., MedIA, 2013) which shows **52.9% sensitivity at 3.1 FP/Vol on 54 Chest CT scans or 60.9% recall at 6.1 FP/Vol**. 2) "In order to compare **the automatic detection results with the performance of a human**, we did an experiment on the **intra-human observer variability**. Ten of the CT volumes were annotated a second time by the same person a few months later. The first segmentations served as ground truth, and the second ones were considered as detections. TPR and FP were measured in the same way as for the automatic detection. 3) The TPR was 54.8% with 0.8 false positives per volume on average. **While 0.8 FP is very low, a TPR of 54.8% shows that finding lymph nodes in CT is quite challenging also for humans**."

Method	Body Region	Number CT Vol.	Size (mm)	TP Criterion	TPR (%)	FP/Vol.
Kitasaka et al. (2007)	Abdomen	5	>5.0	Overlap	57.0%	58
Feuerstein et al. (2009)	Mediastinum	5	>1.5	Overlap	82.1%	113
Dornheim (2008)	Neck	1	>8.0	Unknown	100%	9
Barbu et al. (2010)	Axillary	101	>10.0	In box	82.3%	1.0
Feulner et al. (2013)	Mediastinum	54	>10.0	In box	52.9%	3.1
Intra-obs. Var. (Human)	Mediastinum	**10**	>10.0	In box	**54.8%**	**0.8**

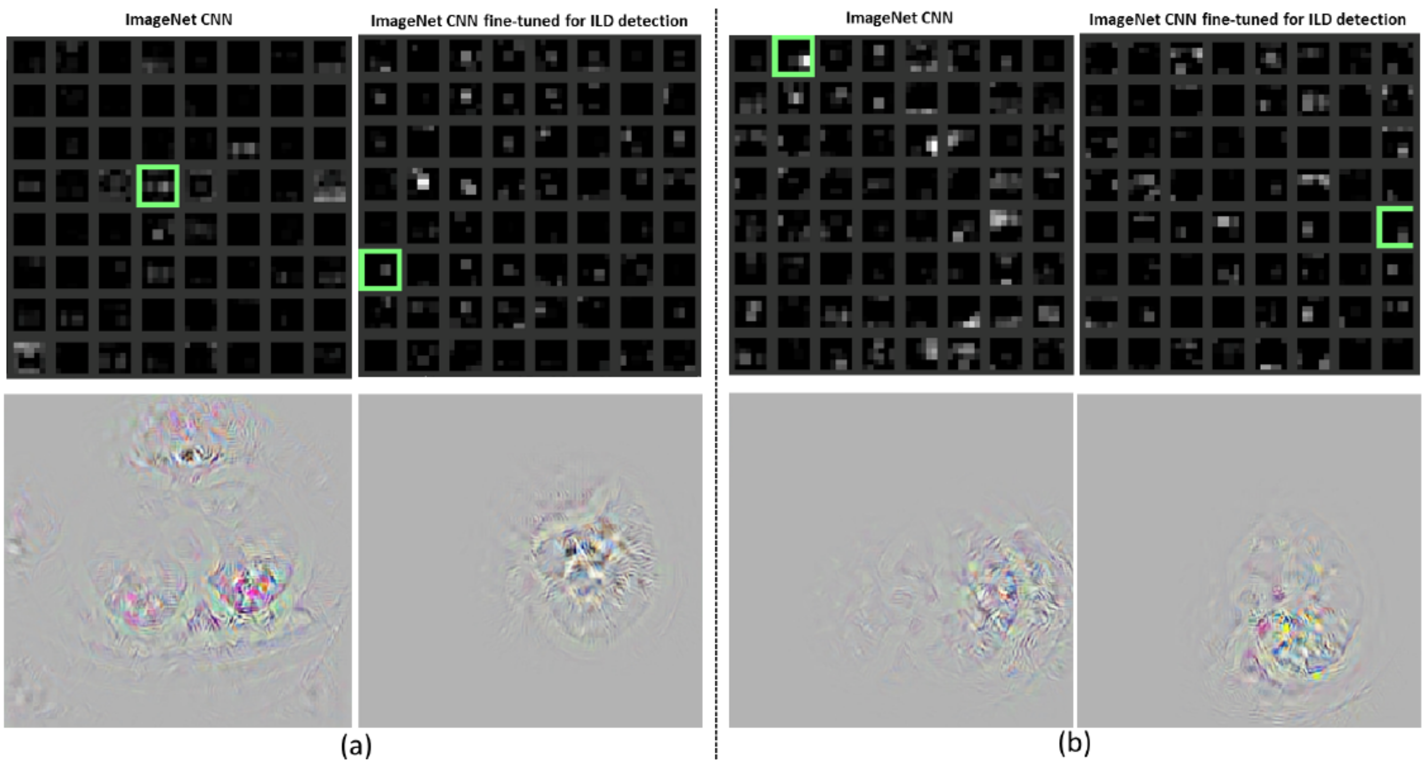
- Table reproduced from Table 3, **Feulner et al., "Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior"**, MedIA, 17(2): 254-270 (2013).
- Note that Barbu et al. (2010) is not directly comparable to other papers since Axillary lymph nodes are easier to detect.
- **Our "comparable" result is 82% @ 3FP …**

VISUALIZATION ON TRANSFER LEARNING (LEARNED FROM THORACO-ABDOMINAL LNS) [SHIN ET AL., TMI 2016]



KRIBB Fellowship

BETTER LOCALIZATION AFTER FINE-TUNING?



FAILURE CASES

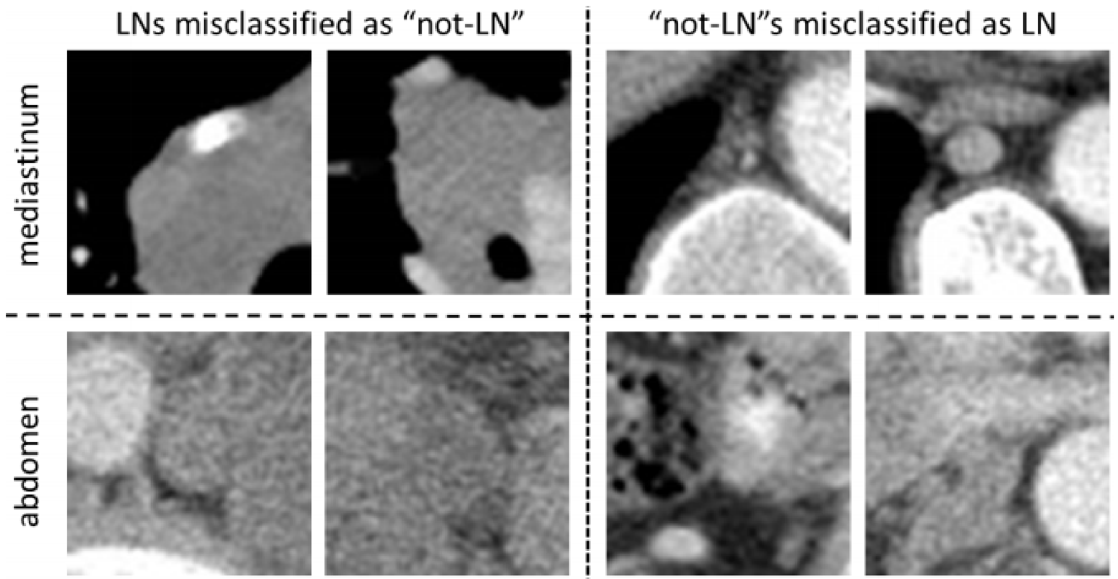
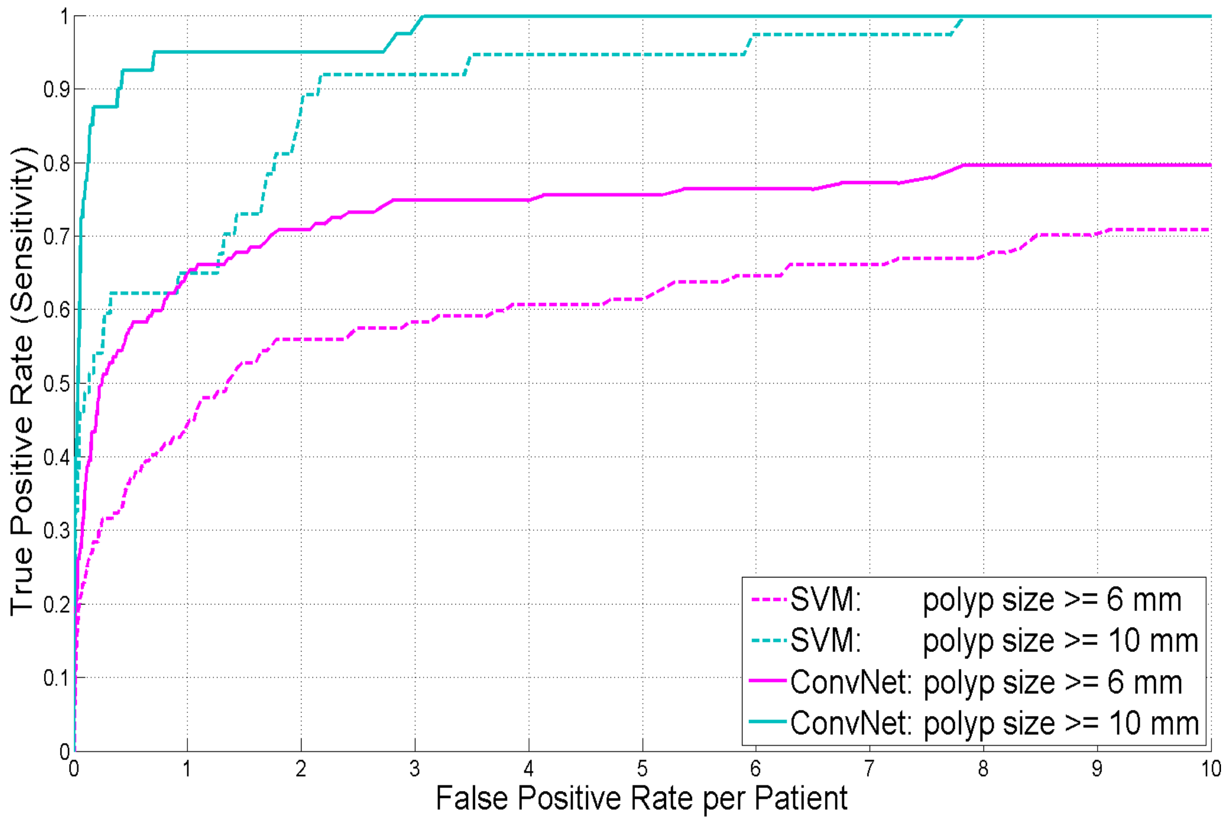


Fig. 9. Examples of misclassified lymph nodes (in axial view) of both false negatives (**Left**) and false positives (**Right**). Mediastinal LN examples are shown in the **upper** row, and abdominal LN examples in the **bottom** row.

Generalizable? Colon CADe Results using a deeper CNN on 1186 patients (or 2372 CTC volumes) via fine-tuning AlexNet



[**SVM** baseline] Summers, et a., Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population, Gastroenterology, vol. 129, no. 6, pp.1832–1844, 2005. 1,186 patients with prone and supine CTC images (394/792 patients; 79/173 polyps training/testing split)

GTCx

References:

1. "A New 2.5D Representation for Lymph Node Detection using Random Sets of Deep Convolutional Neural Network Observations", MICCAI 2014

2. "2D View Aggregation for Lymph Node Detection using a Shallow Hierarchy of Linear Classifiers", MICCAI 2014

3. “Detection of Sclerotic Spine Metastases via Random Aggregation of Deep Convolutional Neural Network Classifications", (Oral), MICCAI Spine Imaging Workshop 2014

4. "Leveraging Mid-Level Semantic Boundary Cues for Computer-Aided Lymph Node Detection", MICCAI 2015

5. "An Analysis of Robust Cost Functions for Deep CNN in Computer-aided Diagnosis", MICCAI DLMIA workshop 2015

6. "Anatomy-specific Classification of Medical Images using Deep Convolutional Nets", IEEE ISBI 2015

7. "Improving Computer-aided Detection using Convolutional Neural Networks and Random View Aggregation", IEEE Trans. on Medical Imaging, 2016

8. "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning", IEEE Trans. on Medical Imaging, 2016

SEMANTIC SEGMENTATION: TOP-DOWN OR BOTTOM-UP PARADIGMS?

(A.2) MORE ACCURATE IMAGING BIO-MARKER USING LYMPH NODE VOLUME, INSTEAD OF DIAMETER BASED *RECIST CRITERION*? [NOGUES ET AL., MICCAI 2016, US PATENT APPLICATION: 62/345,606]

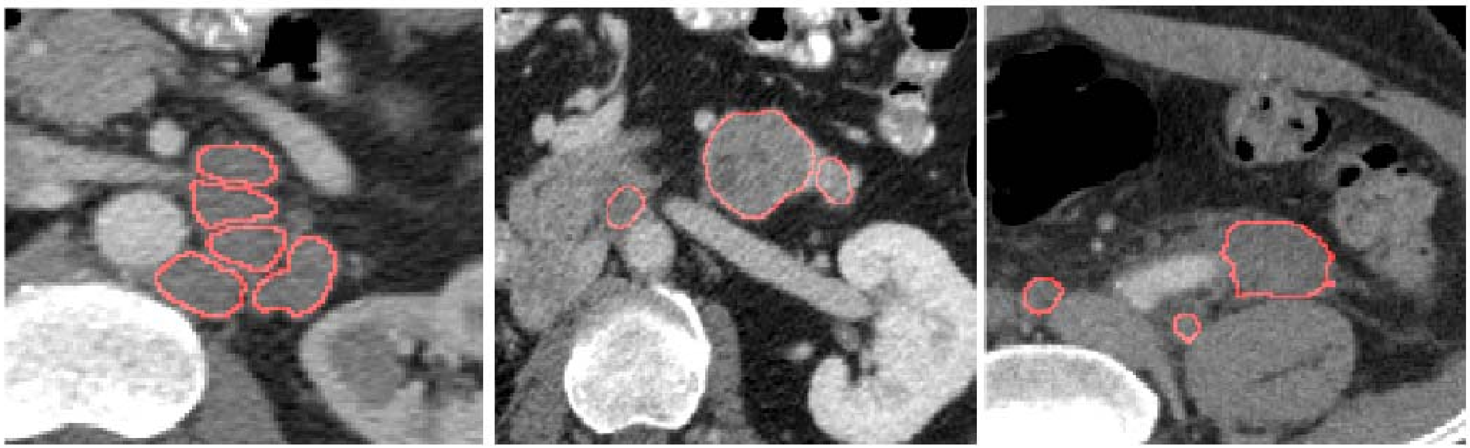
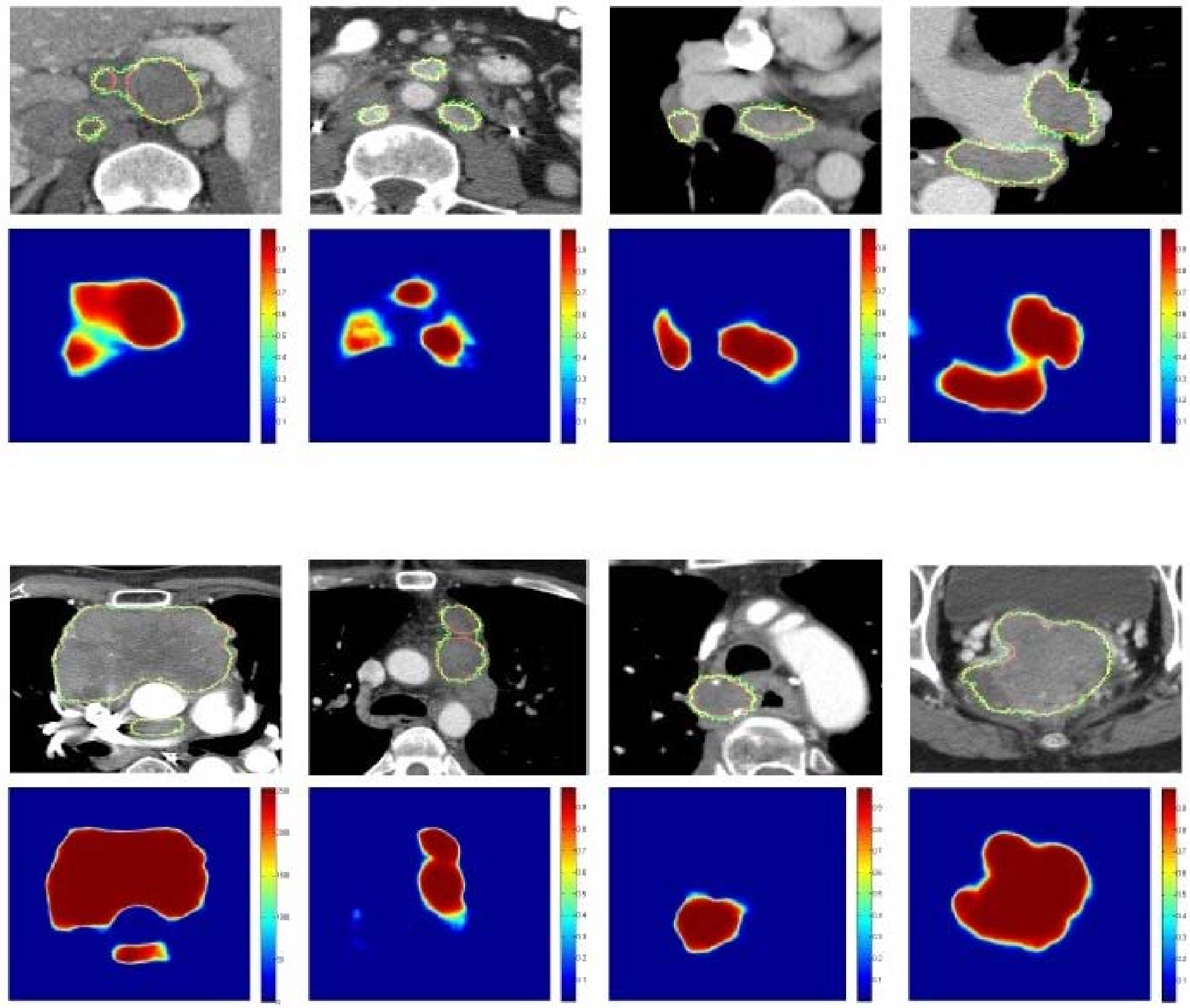


Fig. 1. Examples of thoracoabdominal lymph node clusters in CT images with ground truth (red) boundaries.

MICCAI 2016 Student Travel Award

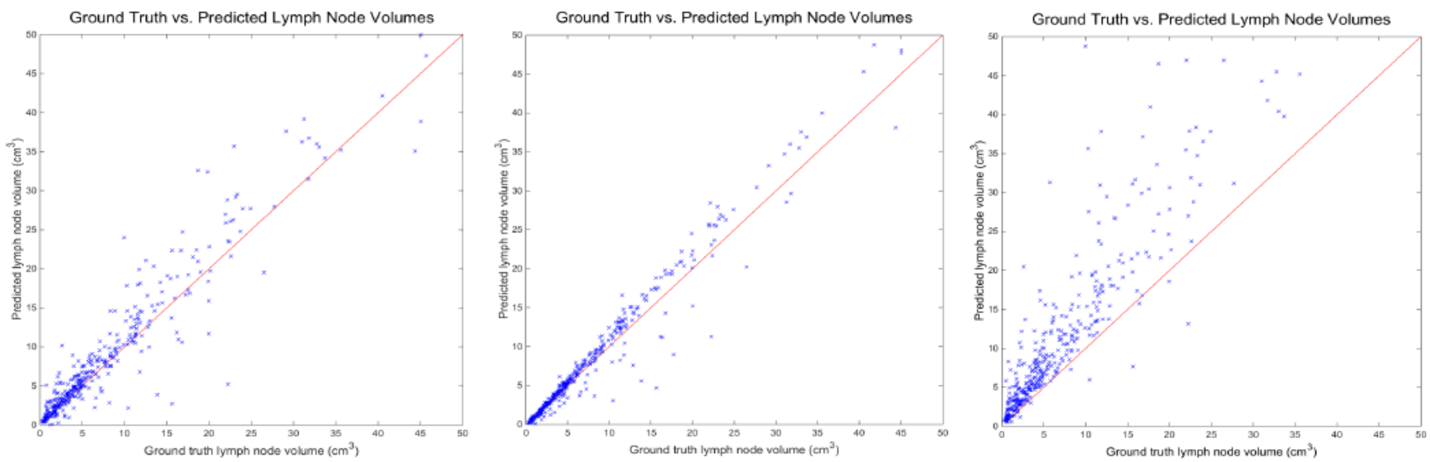


Computing much more precise imaging based biomarkers, meta-information for precision medicine decision making, beyond the current Clinical Protocol … (Oncology Example) → Mining reliable, high accuracy, clinically-relevant meta-measurements from unstructured, noisy low-level imaging data

AUTOMATIC LYMPH NODE CLUSTER SEGMENTATION USING HOLISTICALLY-NESTED NEURAL NETWORKS AND STRUCTURED OPTIMIZATION IN CT IMAGES

Method	Evaluation Metric		
	Mean DSC (%)	Mean IoU (%)	Mean RVD (%)
HNN-A	73.0 ± 17.6	60.1 ± 18.8	32.2 ± 46.3
MGO	82.1 ± 9.6	70.6 ± 11.9	13.7 ± 13.1

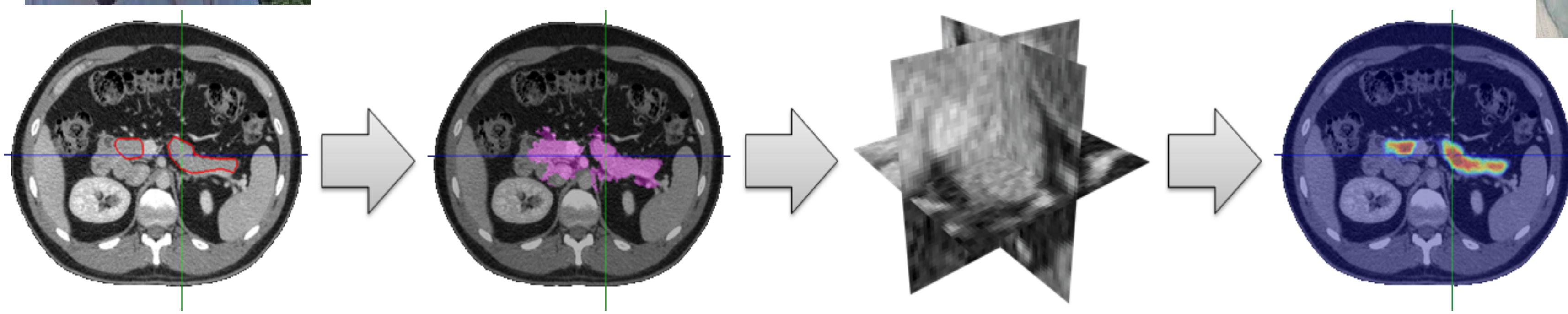
Table 1. Evaluation of segmentation accuracy : HNN-A and MGO



(a) HNN Predictions (b) Matrix Global Optimization (c) Graph Cuts

https://wiki.cancerimagingarchive.net/display/Public/CT+Lymph+Nodes annotated datasets are publicly available)

(A.2) A ROADMAP OF BOTTOM-UP DEEP PANCREAS SEGMENTATION: PATCH, REGION, HOLISTIC CNNS [FARAG ET AL., IEEE TIP 2017]

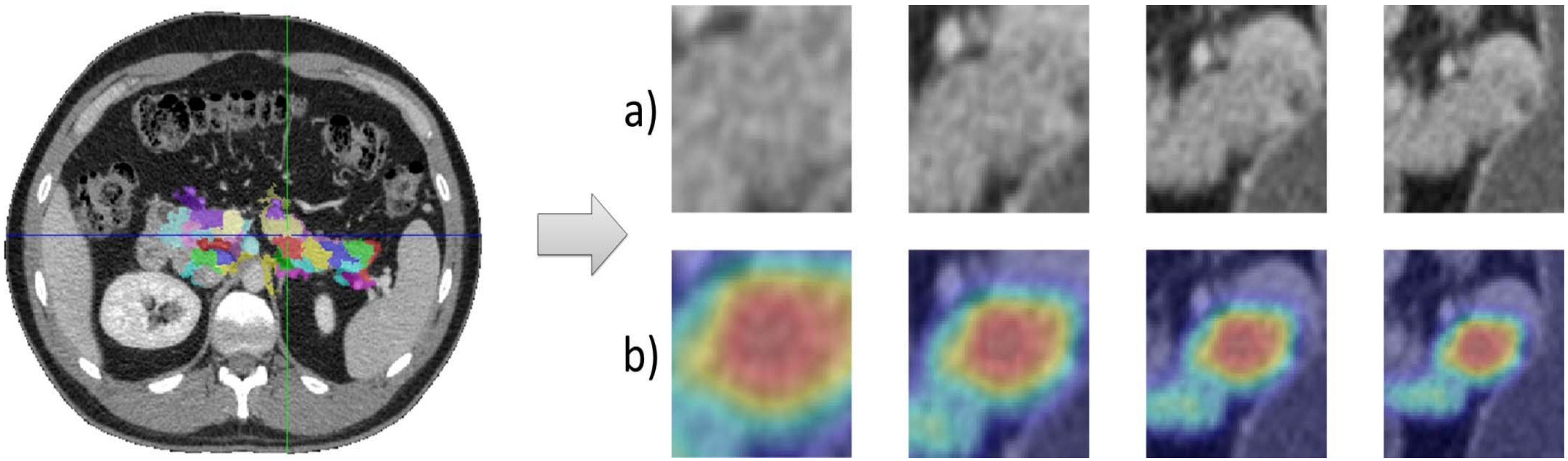


Ground truth　　Random Forest　　2.5D Patch　　ConvNet prob.

P-ConvNet

GTCx

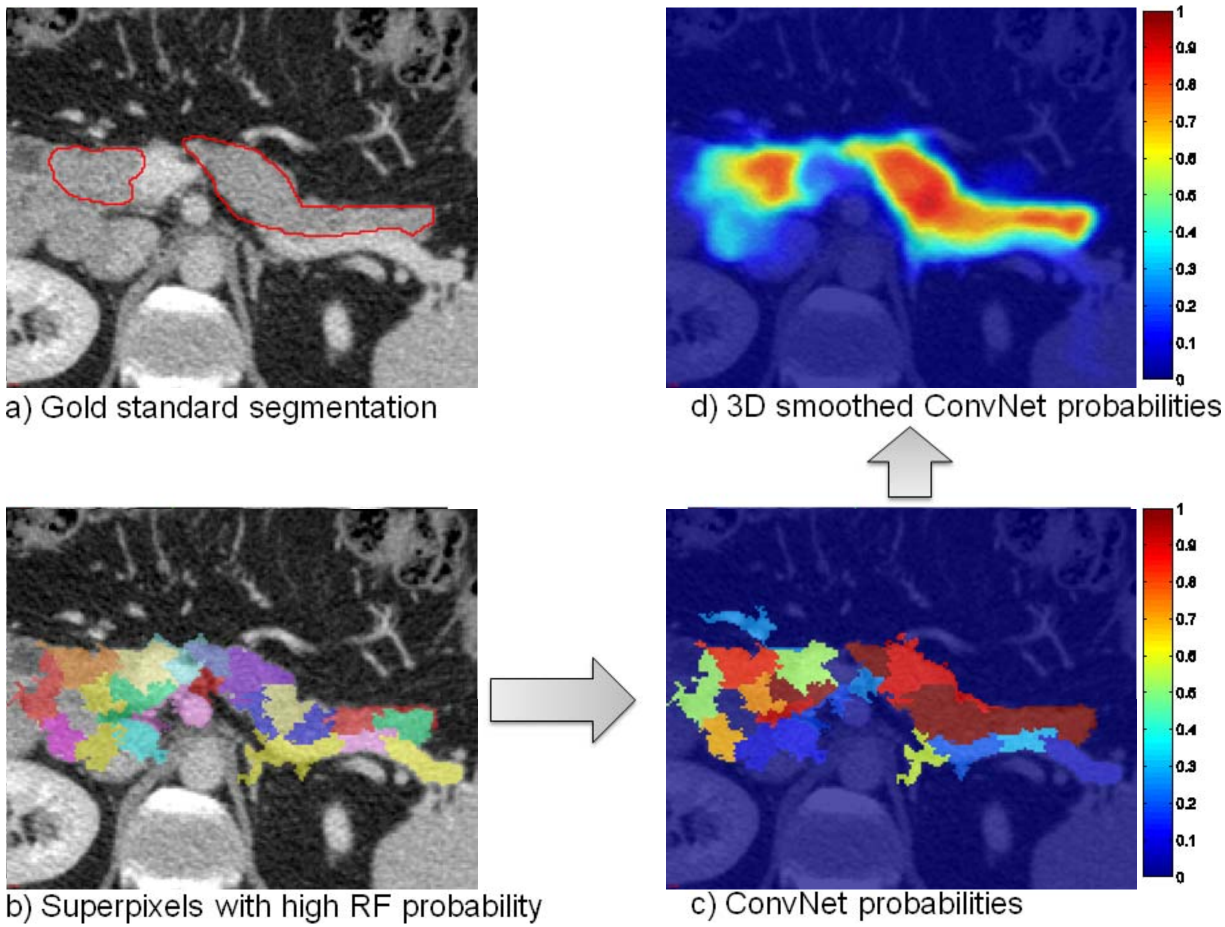
MULTI-SCALE "ZOOM-OUT" R-CONVNET



Zoom-out in dual spaces

DEEPORGAN: R²-CONVNET VIA TWO-CHANNEL ENCODING

[ROTH ET AL., MICCAI 2015]



~68% Dice score

~27% Dice score

~57% Dice score

GTCx

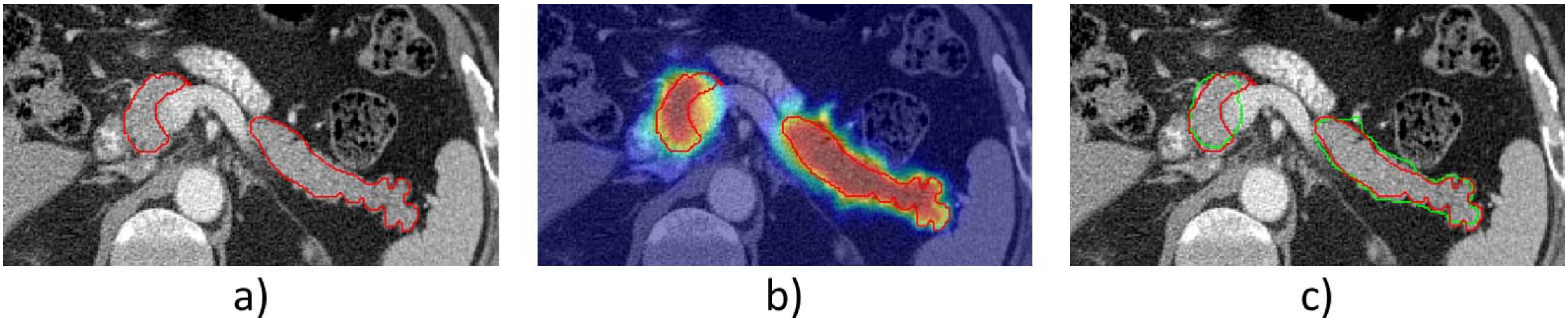
4-FOLD CV PERFORMANCE ON 82 CT SCANS

Table 1: **4-fold cross-validation**: optimally achievable DSCs, our initial candidate region labeling using S_{RF}, DSCs on $P(x)$ and using smoothed $G(P(x))$, and a CRF model for structured prediction (best performance in bold).

DSC (%)	Opt.	$S_{RF(x)}$	$P_0(x)$	$G(P_0(x))$	$P_1(x)$	$G(P_1(x))$	$P_2(x)$	$G(P_2(x))$	$CRF(P_2(x))$
Mean	80.5	26.1	60.9	69.5	56.8	62.9	64.9	**71.8**	68.2
Std	3.6	7.1	10.4	9.3	11.4	16.1	8.1	**10.7**	4.1
Min	70.9	14.2	22.9	35.3	1.3	0.0	33.1	**25.0**	59.6
Max	85.9	45.8	80.1	84.4	77.4	87.3	77.9	**86.9**	74.2

- Averaged surface-surface distances: **0.94+/-0.6mm** (p<0.01) with *R^2-ConvNet* from **1.46+/-1.5mm** if just *P-ConvNet* is applied.
- Previous state-of-the-art: [46.6% to 69.1%] DSC, all under **LOO (Leave-one-patient-out)**.

AN ABOVE-AVERAGE EXAMPLE

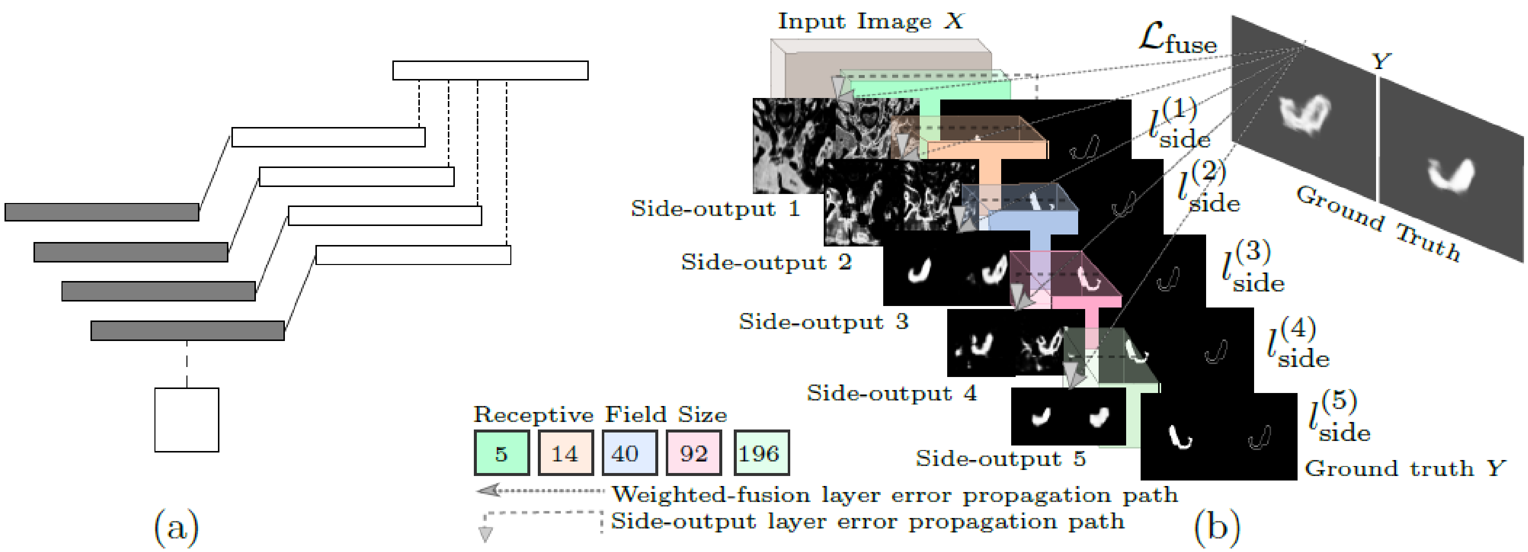


a) The manual ground truth annotation (in red outline)
b) The ***G(P2(x))*** probability map
c) The final segmentation (in green outline) at p_2=0.6

DSC=82.7%.

SPATIAL AGGREGATION OF HOLISTICALLY-NESTED NETWORKS FOR AUTOMATED PANCREAS SEGMENTATION

[ROTH, ET AL., MICCAI 2016, US PATENT APPLICATION: 62/345,606]



Saining Xie and Zhuowen Tu, "***Holistically-Nested Edge Detection***", ICCV 2015 (**Marr Prize Honorable Mention**).
Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu, "***Deeply-Supervised Nets***", AISTATS 2015.

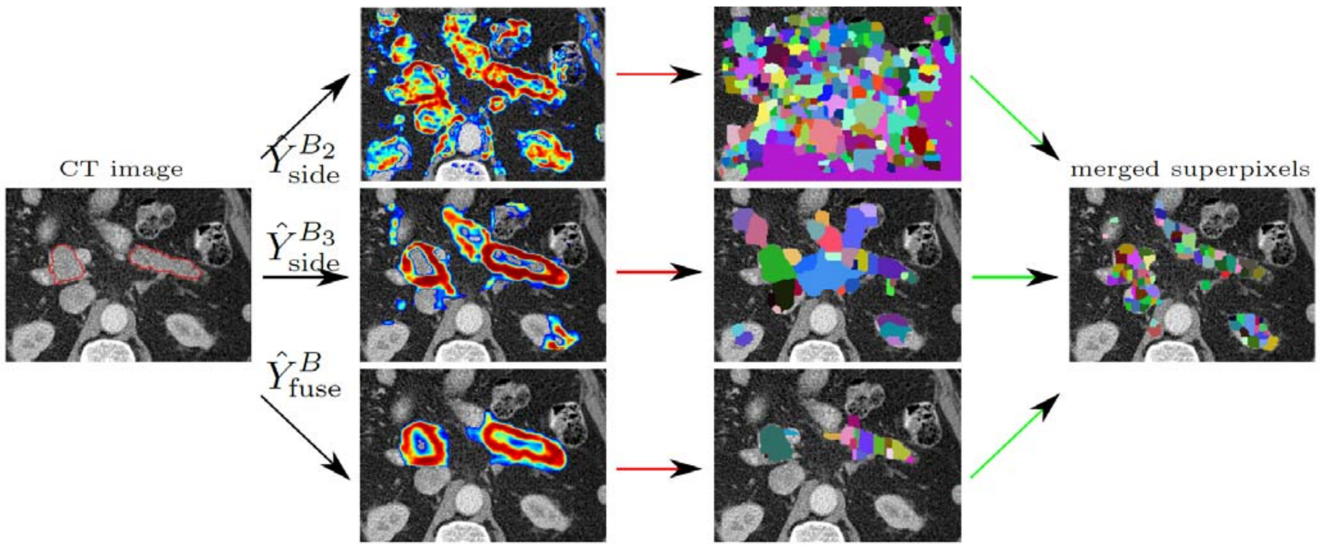


Fig. 2: "Multiscale Combinatorial Grouping" (MCG) [16] on three different scales of learned boundary predication maps from **HNN-B**: $\hat{Y}^{B_2}_{\text{side}}$, $\hat{Y}^{B_3}_{\text{side}}$, and $\hat{Y}^{B}_{\text{fuse}}$ using the original CT image as input (shown with ground truth delineation of pancreas). MCG computes superpixels at each scale and produces a set of merged superpixel-based object proposals. We only visualize the boundary probabilities $p > 10\%$.

GTCx

MORE ACCURATE AND FASTER?
[ROTH, ET AL. MICCAI 2016]

Table 1: **4-fold cross-validation**: The DSC and average minimum distance (Dist) performance of our implementation of [6], optimally achievable superpixels, **HNN-I**, and **HNN-RF** spatial aggregation, and DCRF (best performance in bold).

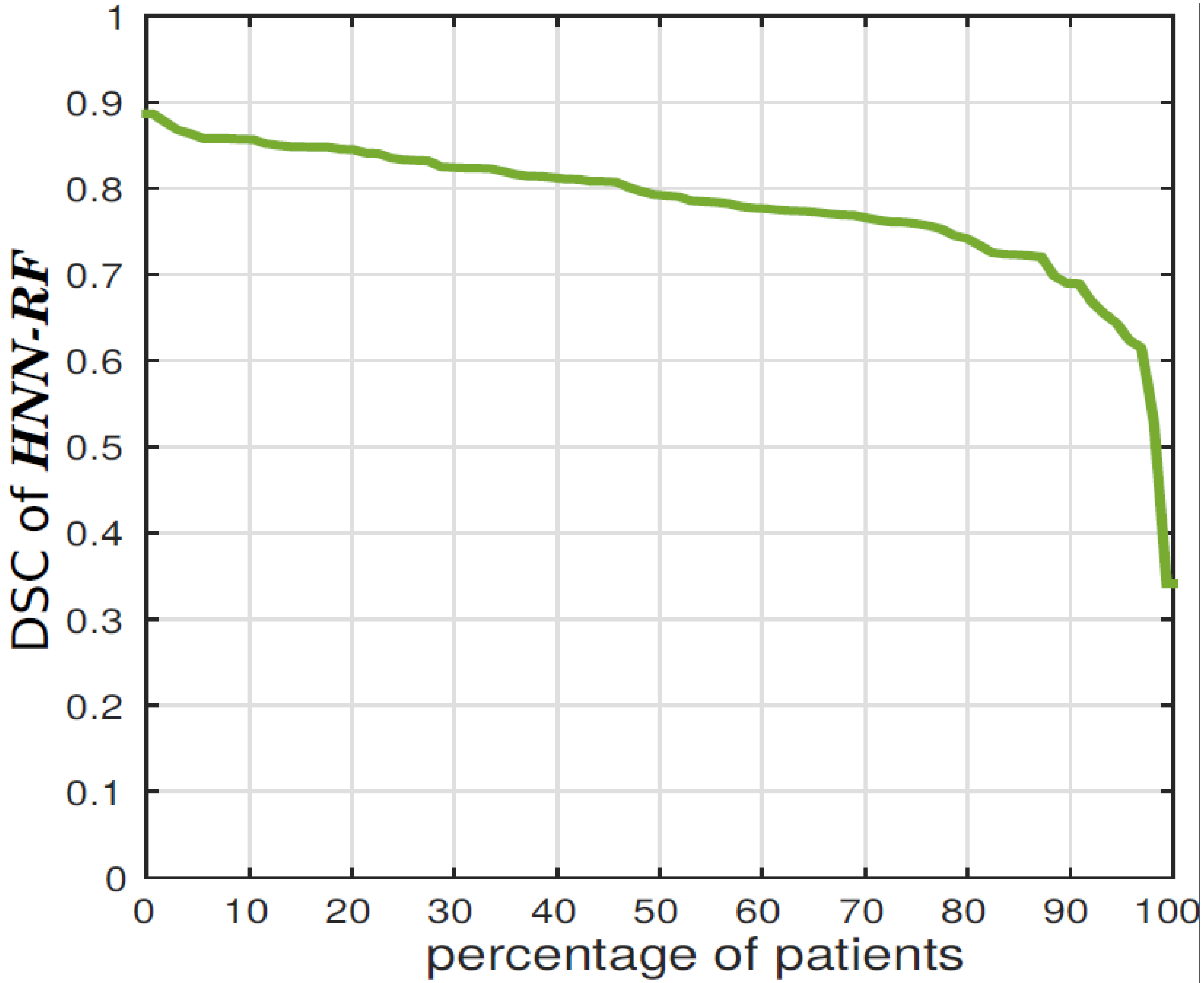
DSC[%]	[6]	Opt.	HNN-I	HNN-RF	DCRF
Mean	71.42	88.08	76.99	**78.01**	77.14
Std	10.11	2.10	9.45	**8.20**	10.58
Min	23.99	81.24	24.11	**34.11**	16.10
Max	86.29	92.00	87.78	**88.65**	88.30

Dist[mm]	[6]	Opt.	HNN-I	HNN-RF	DCRF
Mean	1.53	0.15	0.70	**0.60**	0.69
Std	1.60	0.08	0.73	**0.55**	0.76
Min	0.20	0.08	0.17	**0.15**	0.15
Max	10.32	0.81	5.91	**4.37**	5.71

https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT annotated datasets are publicly available) → https://www.synapse.org/#!Synapse:syn3193805/wiki/217789

Our newest results are (**81.4% +/- 7.3%**) in Dice and ~0.43 mm mean surface-to-surface distance with a stacked implementation of HNNs for both pancreas localization & segmentation.

Localize & Zoom better to see better or segment more accurately) → proper Zooming is related to scale & attention models


DSC of HNN-RF
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 10 20 30 40 50 60 70 80 90 100
percentage of patients

GTCx

References:

1. "DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation", MICCAI 2015
2. “A Bottom-up Approach for Automatic Pancreas Segmentation Abdominal CT Scans", Oral, MICCAI Abdominal Imaging Workshop 2014
3. "Spatial Aggregation of Holistically-Nested Networks for Automated Pancreas Segmentation", MICCAI 2016
4. Automatic Lymph Node Cluster Segmentation using Holistically-Nested Networks and Structured Optimization", MICCAI 2016
5. Pancreas Segmentation in MRI using Graph-based Decision Fusion on Convolutional Neural Networks", MICCAI 2016
6. "A Bottom-up Approach for Pancreas Segmentation Using Cascaded Superpixels and (Deep) Image Patch Labeling", to appear, IEEE Trans. Image Processing, 2016, arXiv:1505.06236, 2015
7. “Spatial Aggregation of Holistically-Nested Convolutional Neural Networks for Automated Pancreas Localization and Segmentation”, in preparation, 2016

MICCAI

GTCx

(A3) HOLISTIC ILD (*INTERSTITIAL LUNG DISEASE*) PREDICTION VIA MULTI-LABEL DEEP LEARNING (LOOKING FOR A CLINICALLY MORE DESIRABLE PROTOCOL TO ASSIST DECISION MAKING)

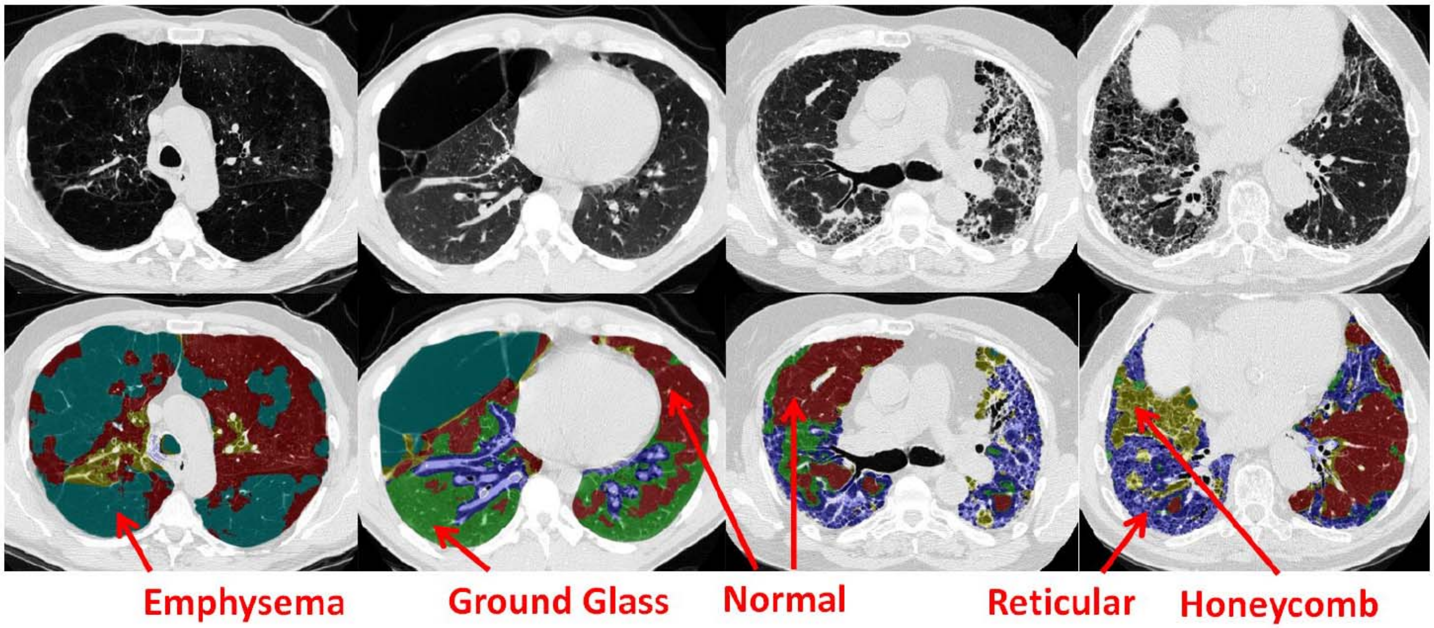
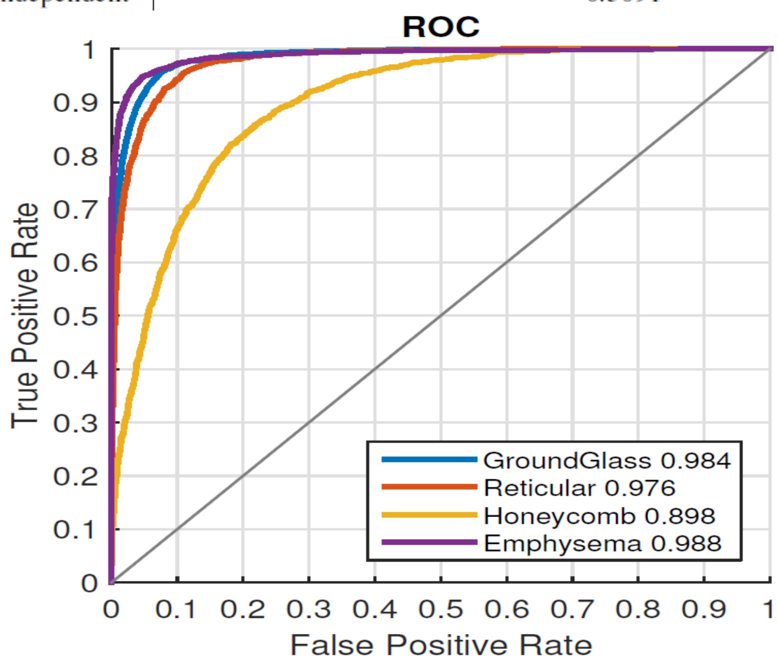


Fig. 1. Examples of ILD patterns. Every voxel in the lung region is labeled as healthy or one of the four ILD diseases: Ground Glass, Reticular, Honeycomb or Emphysema. The first row is the lung CT images. The second row is their corresponding labeling.

ISTP Fellow

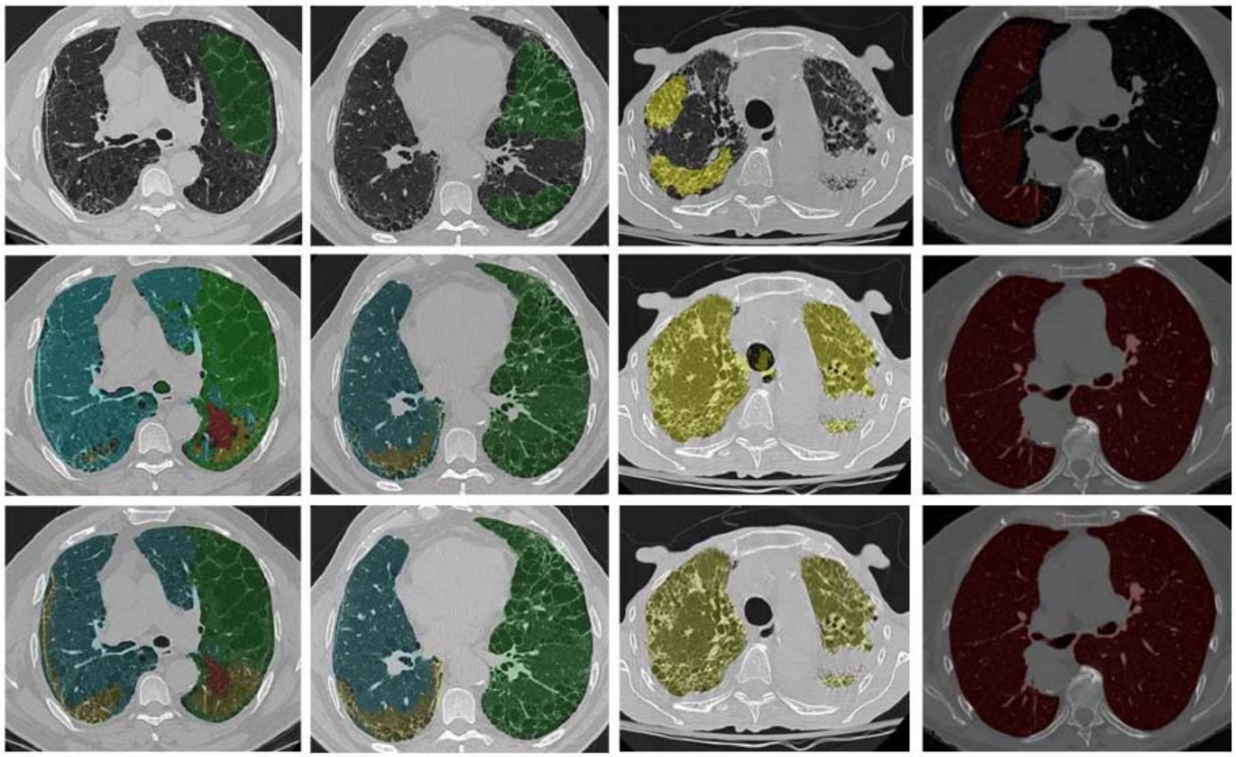
F SCORES

Disease	F-score					
	Ground Glass	Reticular	Honeycomb	Emphysema	Overall	Healthy
multilabel unbalanced testing (658 patients)	0.8750	0.6791	0.1651	0.8851	0.8185	
multilabel balanced testing (658 patients)	0.8960	0.7157	0.1615	0.9094	0.8393	
multilabel balanced training (658 patients)	0.9237	0.8810	0.9818	0.9639	0.9395	
multitask unbalanced testing (533 patients)	0.81	0.74	0.30	0.89		
multitask unbalanced training (533 patients)	0.86	0.78	0.49	0.88		
multitask balancing testing (533 patients)	does not improve					
multitask testing (658 patients) independent			0.2705	0.8616		0.9897
multitask training (658 patients) independent			1	0.9991		1
multitask testing (658 patients) independent			0.3091			



GTCx

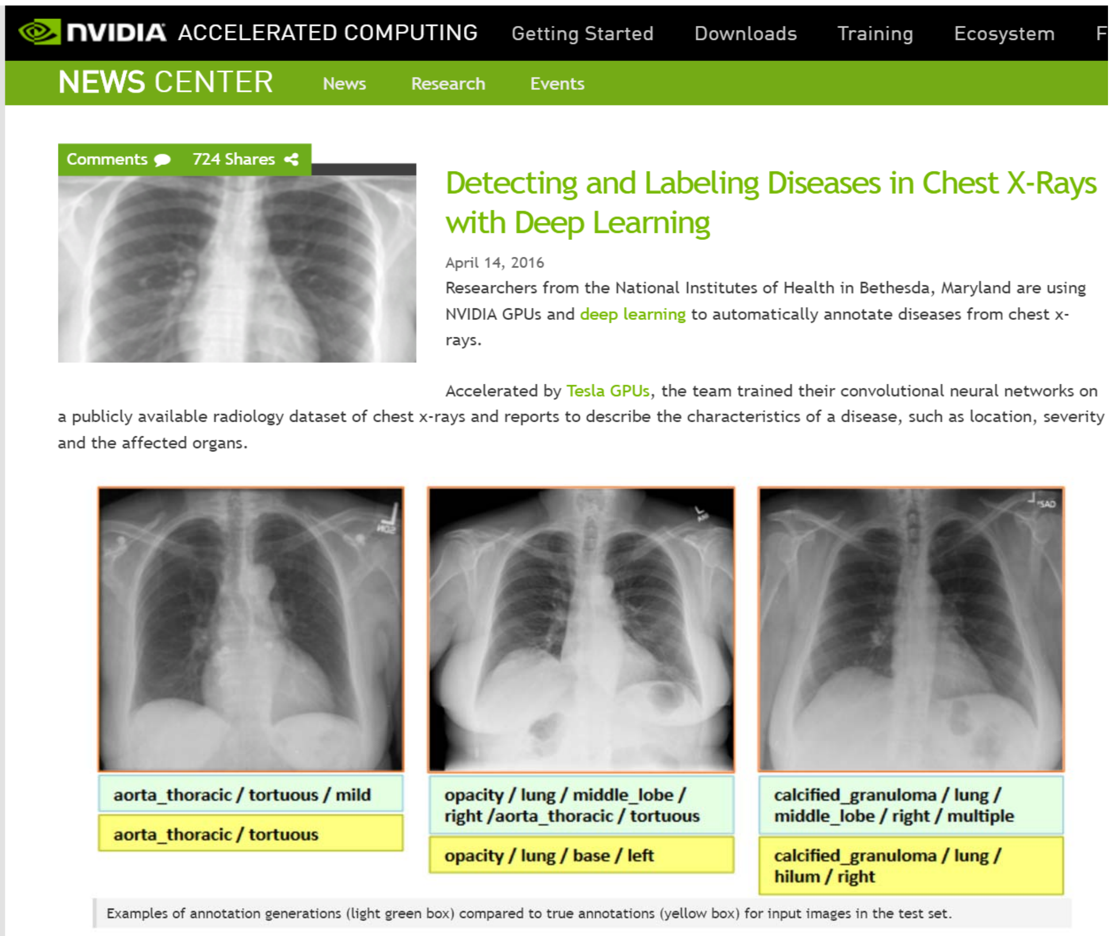
LABEL ANYTHING (THAT MATTERS) FROM EVERYWHERE?



Gao et al., “SEGMENTATION LABEL PROPAGATION USING DEEP CONVOLUTIONAL NEURAL NETWORKS AND DENSE CONDITIONAL RANDOM FIELD”, IEEE ISBI, 2016

LEARNING TO READ CHEST X-RAY USING DEEP NEURAL NETWORKS, (A LITTLE MORE LIKE HUMANS?)

[SHIN ET AL., CVPR 2016, US PATENT APPLICATION: 62/302,084]



NVIDIA ACCELERATED COMPUTING Getting Started Downloads Training Ecosystem

NEWS CENTER News Research Events

Comments 724 Shares

Detecting and Labeling Diseases in Chest X-Rays with Deep Learning

April 14, 2016

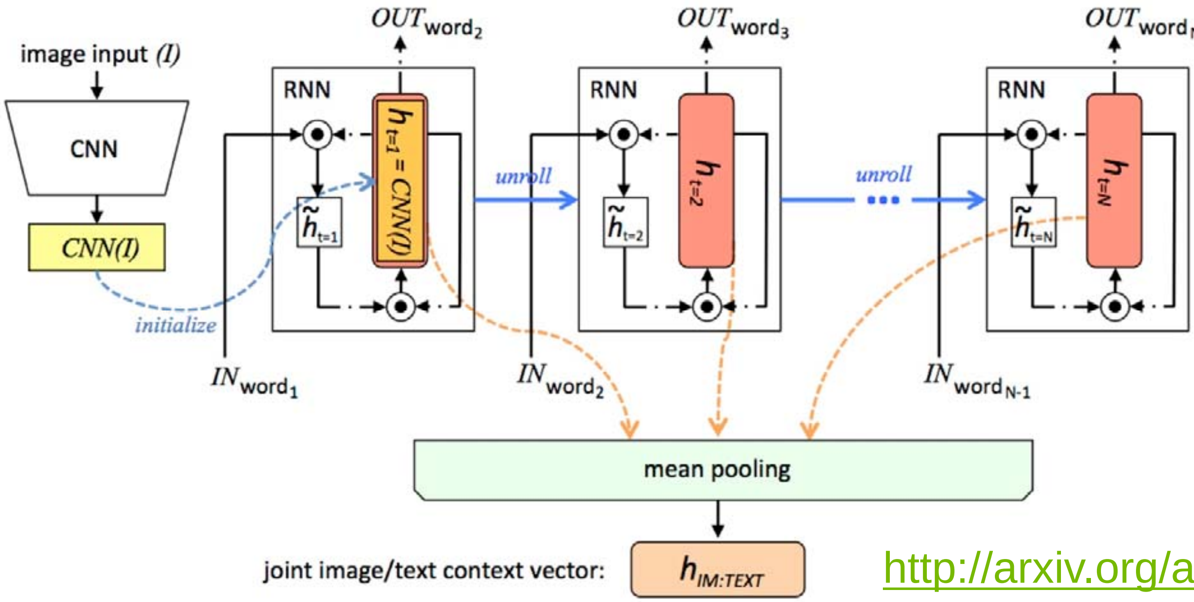
Researchers from the National Institutes of Health in Bethesda, Maryland are using NVIDIA GPUs and deep learning to automatically annotate diseases from chest x-rays.

Accelerated by Tesla GPUs, the team trained their convolutional neural networks on a publicly available radiology dataset of chest x-rays and reports to describe the characteristics of a disease, such as location, severity and the affected organs.

aorta_thoracic / tortuous / mild	opacity / lung / middle_lobe / right /aorta_thoracic / tortuous	calcified_granuloma / lung / middle_lobe / right / multiple
aorta_thoracic / tortuous	opacity / lung / base / left	calcified_granuloma / lung / hilum / right

Examples of annotation generations (light green box) compared to true annotations (yellow box) for input images in the test set.

Towards more accurate, "human-like" image annotation, using MeSH



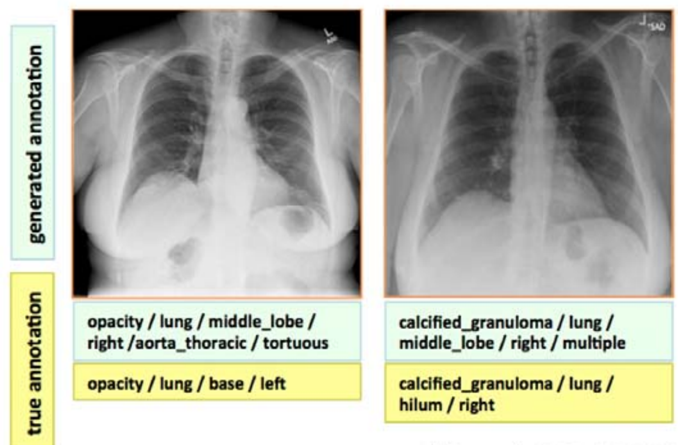
http://arxiv.org/abs/1603.08486

Comparison: None
Indication: Burmese male has complete TB treatment
Findings: Both lungs are clear and expanded with no infiltrates. Basilar focal atelectasis is present in the lingula. Heart size normal. Calcified right hilar XXXX are present
Impression: No active disease.

MeSH

Major
Pulmonary Atelectasis / lingula / focal
Calcinosis / lung / hilum / right

CHEST 2V FRONTAL/LATERAL XXXX, XXXX XXXX PM



Shin et al, IEEE CVPR 2016

GTCx

References:

1. "Holistic Classification of CT Attenuation Patterns for Interstitial Lung Diseases via Deep CNNs", MICCAI DLMIA workshop 2015

2. "SEGMENTATION LABEL PROPAGATION USING DEEP CONVOLUTIONAL NEURAL NETWORKS AND DENSE CONDITIONAL RANDOM FIELD", IEEE ISBI, 2016

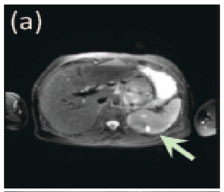
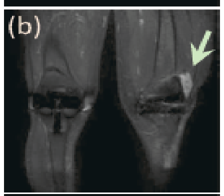
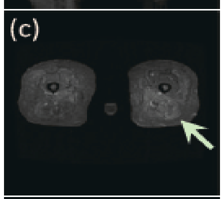
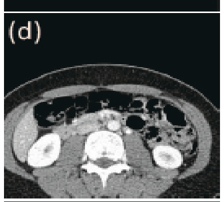
3. "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning", IEEE Trans. on Medical Imaging, 2016

4. Multi-label Deep Regression and Unordered Pooling for Holistic Interstitial Lung Disease Detection", MICCAI-MLMI 2016.

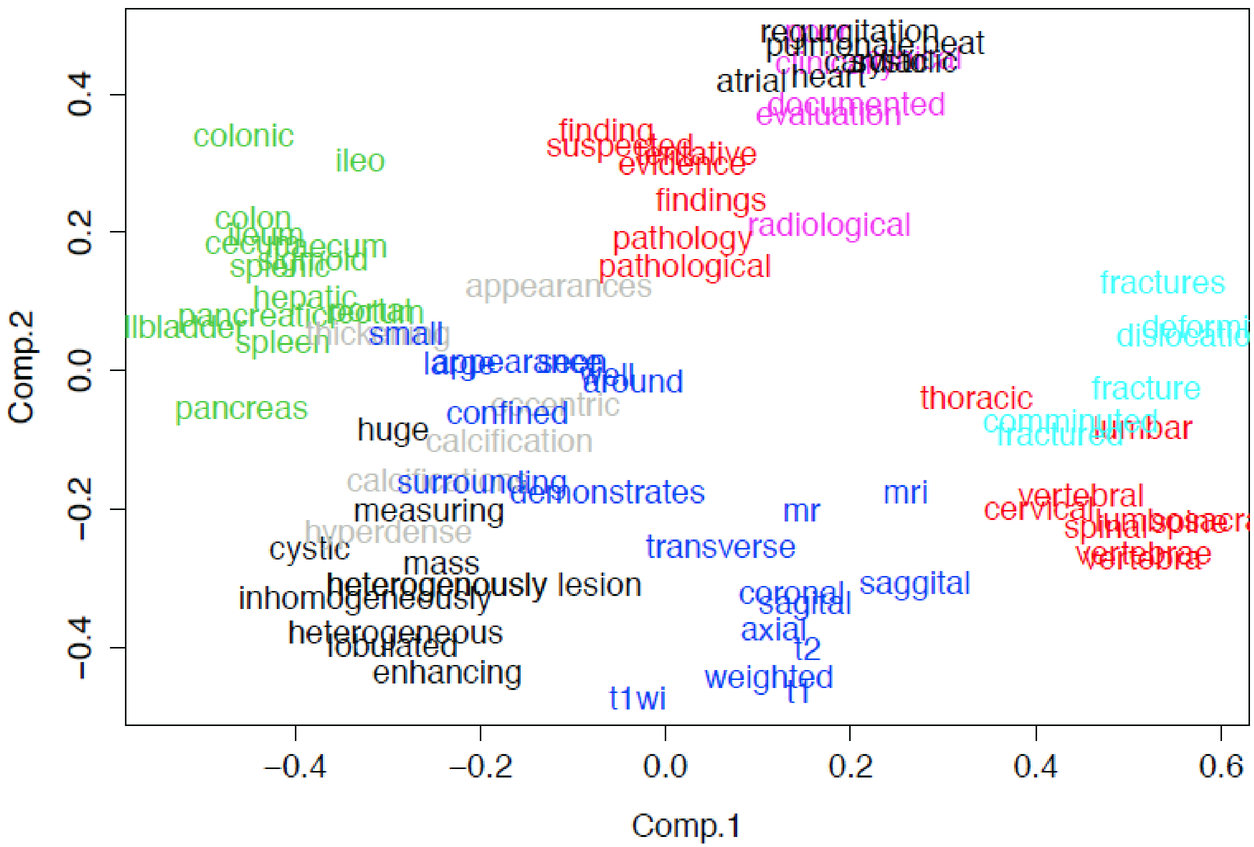
5. "Learning to Read Chest X-Rays: Recurrent Neural Feedback Model for Automated Image Annotation", IEEE CVPR, 2016

6. “Holistic Interstitial Lung Disease Detection using Deep Convolutional Neural Networks: Multi-label Learning and Unordered Pooling”, in preparation, 2016

(A.4) INTERLEAVED TEXT/IMAGE DEEP MINING ON A LARGE-SCALE RADIOLOGY DATABASE (780K/62K PATIENTS) FOR AUTOMATED IMAGE INTERPRETATION

Input image	Generated key-words	Disease detection	Original text
(a) 	originating effusion upper *avg distance 0.14*	label: cyst cyst: 0.999 no cyst: 2.24e-05 disease: 1.54e-05 gallstone: 5.32e-07 hydronephrosis: 3.48e-07	2 multiple clip artifacts indicative of previous surgery in the left abdominal wall and left retroperitoneum about the kidney 3 in the upper abdomen non enhancing well defined foci of high signal intensity on t2 weighted images consistent with cysts one about a centimeter at the left renal splenic interface series 501 image 19 the other less than 5 mm in the periphery of the right kidney series 501 image 12 4 multiple gallstones
(b) 	susceptibility findings tibialis *avg distance 0.20*	label: abscess abscess: 0.663 infection: 0.103 osteochondromatosis: 0.037 synovitis: 0.032 cyst: 0.026	... for example series 701 image 12 and series 401 image 27 with findings suggesting minimally enhancing rim laterally for example series 1101 image 21 may ... the findings suggest a fluid collection with ... the location suggests possibility of a synovial collection synovial thickening as the appearance is nonspecific correlation with clinical findings is recommended regarding the possibility of an infection abscess
(c) 	basal fasciitis findings *avg distance 0.31*	label: myositis myositis: 0.996 fasciitis: 0.002 tenosynovitis: 0.002 lymphedema: 1.30e-05 no myositis: 2.84e-06	images were obtained of both thighs including stir scans findings include 1 areas of slight increase in signal intensity in some muscles on the stir scan more apparent on the left than the right for example series 4 image 13 the left hamstrings and vastus medialis consistent with myositis 2 no evidence of gross fatty infiltration of the muscles
(d) 	anterior effusion renal *avg distance 0.34*	label: cyst cyst: 0.709 lymphocele: 0.120 no gallstone: 0.050 syndrome: 0.020 pyelonephritis: 0.016	adrenal glands 1.2 mm lower right kidney focus e.g series 3 image 63 possibly due to cyst no evidence of pleural effusion splenomegaly hydronephrosis calcification in gallbladder or kidneys or definite adrenal mass or calcification
(e) 	subclavian effusion hairy *avg distance 0.20*	label: osteophyte osteophyte: 0.472 disease: 0.207 gynecomastia: 0.098 no hydronephrosis: 0.034 pneumothorax: 0.028	history lymphoma restaging chest subcentimeter right apex lung cavity series 921780 image 11 unchanged since xx/xx/xxxx spine osteophytes no evidence of pleural or pericardial effusion bulky axilla mediastinum or hilum adenopathy or lung mass or infiltrate

INTERLEAVED TEXT/IMAGE DEEP MINING ON A LARGE-SCALE RADIOLOGY DATABASE



Example words embedded in the vector space using Open Source RNN based Google Word-to-Vector modeling (visualized on 2D), trained from 1B words in 780K radiology reports and 0.2B from OpenI: an open access biomedical image search engine; http://openi.nlm.nih.gov .

~1.2 billion words with OpenI

"cyst"

cysts	0.799191
hydatid	0.734686
cystic	0.701855
unilocular	0.654273
tailgut	0.639764
nonparasitic	0.621647
epidermoid	0.604492
lipoma	0.588372
cheesy	0.586947
multiloculated	0.584199
pearly	0.583126
multilocular	0.582670
lesion	0.579009
tgdc	0.578533
multiseptate	0.575851

~1.2 billion words with OpenI

"heart"

cardiac	0.672690
respiratory	0.644453
beat	0.642630
pressure	0.558879
murmur	0.551323
systolic	0.548490
pericardial	0.538957
dobutamine	0.537429
intracardiac	0.533799
great	0.532735
rate	0.531352
beats	0.524729
atrial	0.524052
tachycardia	0.521093
minute	0.520249

~1.2 billion words with OpenI

"brain"

hemisphere	0.684149
hemispheric	0.668626
cerebellum	0.663902
whole	0.661564
regions	0.647632
mri	0.646674
structural	0.638171
neuroanatomical	0.636563
crinion	0.626951
in	0.626707
parasaggital	0.618392
illustration	0.610440
striatal	0.609282
brains	0.607442
behavioral	0.606803

~1.2 billion words with OpenI

"liver"

hepatic	0.764163
spleen	0.683242
cirrhotic	0.664428
cirrhosis	0.664262
hcc	0.656473
portal	0.610437
hepatocellular	0.603930
parenchyma	0.597169
splenic	0.579957
hepatomegaly	0.573687
tumor	0.571135
abdomen	0.559092
hepatectomy	0.556156
bclc	0.546798
subcapsular	0.542745

~1 billion words reports only

"cyst"

cysts	0.768382
septated	0.586067
polyp	0.583761
simple	0.534717
septation	0.500951
parapelvic	0.500877
incidental	0.500760
small	0.487211
cystic	0.477632
pole	0.471933
multiseptated	0.469851
polyps	0.464380
exophytic	0.459088
hyperdense	0.457558
mucous	0.448427

~1 billion words reports only

"heart"

lungs	0.526600
mediastinum	0.517008
consolidating	0.486605
pa	0.449816
chest	0.433362
infiltrates	0.428404
hyperinflated	0.413326
cardiomegaly	0.410785
hyperlucent	0.400836
pectus	0.396142
great	0.395712
ectatic	0.394560
shifted	0.389205
ray	0.389091
infiltrate	0.387224

~1 billion words reports only

"brain"

t1	0.615066
mri	0.595027
sagittal	0.580841
flair	0.565445
t2	0.555053
axial	0.554040
spgr	0.520954
weighted	0.502047
technique	0.487768
astrocytoma	0.480527
gbm	0.476956
gradient	0.476593
oligodendroglioma	0.465892
postcontrast	0.463686
3d	0.458123

~1 billion words reports only

"liver"

spleen	0.759884
gallbladder	0.648075
hepatomegaly	0.642022
gallstones	0.611837
pancreas	0.608356
gallstone	0.606063
steatosis	0.601081
dome	0.594812
portal	0.570008
ascites	0.551869
hepatosplenomegaly	0.540501
hepatic	0.537453
cirrhosis	0.530389
fatty	0.522134
kidneys	0.515252

UNSUPERVISED LOOPED DEEP PSEUDO-TASK OPTIMIZATION

[WANG ET AL. ARXIV 2016, WACV 2017, US PATENT APPLICATION: 62/302,096]

Cluster #5

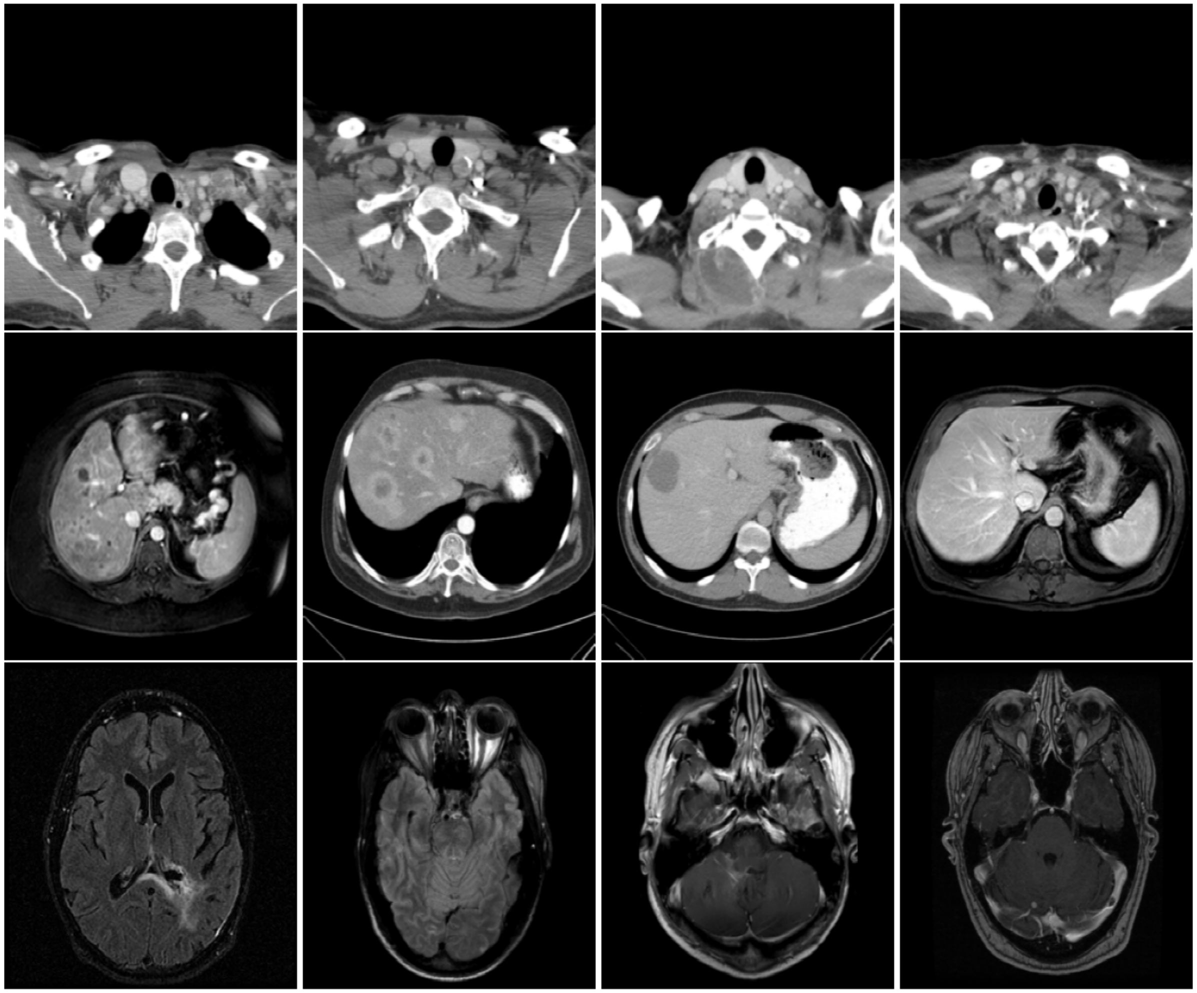
Word	Frequency
neck	656
adenopathy	343
thyroid	295
lymph	292
supraclavicular	236
nodes	218
mass	203
enhancing	96
bulky	77
paratracheal	76

Cluster #23

Word	Frequency
liver	524
abdomen	337
enhancement	217
mass	198
lesion	168
lobe	161
adenopathy	119
lesions	109
segment	58
bulky	45

Cluster #64

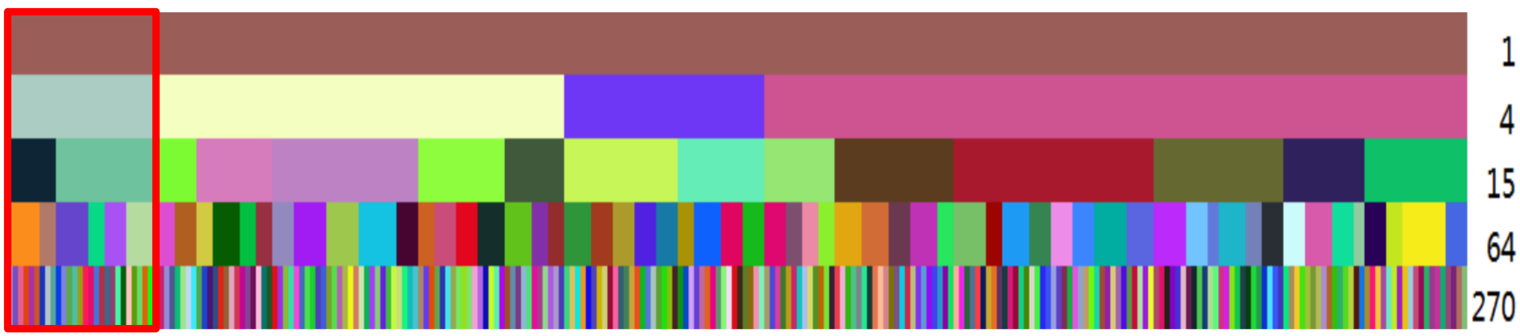
Word	Frequency
enhancement	277
cerebellar	193
lesion	192
lobe	186
flair	173
hemisphere	155
mass	134
abnormal	119
frontal	115
cerebellum	113



RSNA 2016 Best Paper Award!

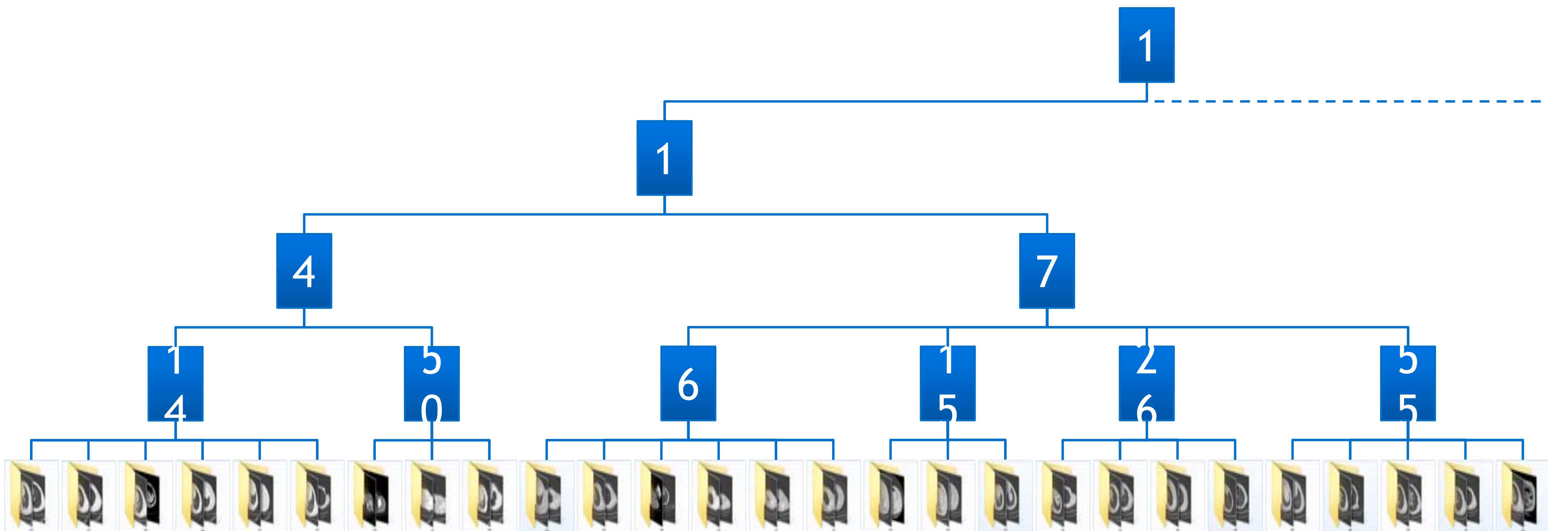
FIVE-LEVEL HIERARCHICAL CATEGORIZATION

- Form a hierarchical category tree (ontology semantics?) of (270, 64, 15, 4, 1) different class labels from bottom (leaf) to top (root). The random color coded category tree is shown.
- *Model Selection* embedded for a large scale radiology image database (215,786 key images from 61,845 unique patients)



A SAMPLE BRANCH OF CATEGORY HIERARCHY

The high majority of images in the clusters of this branch are verified as CT Chest scans by radiologists.



GTCx

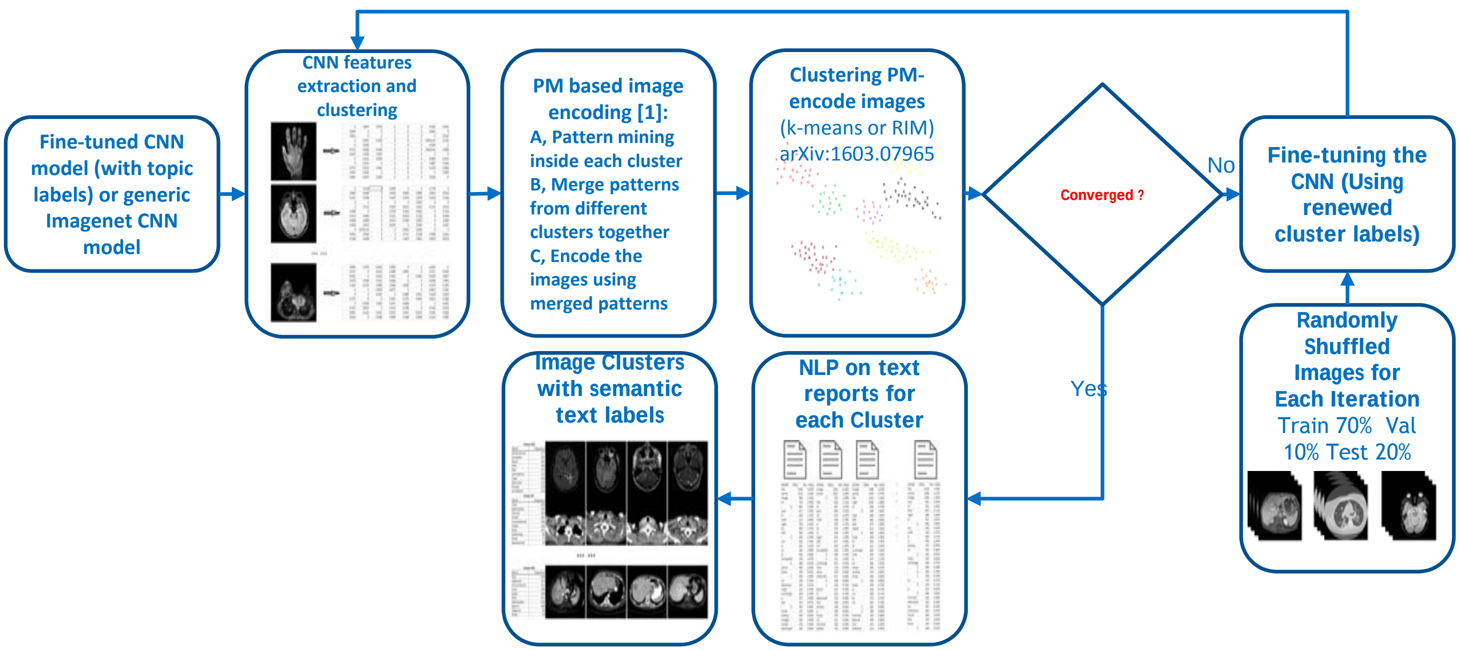
*With “Radiologist-in-the-loop” Protocol to build an annotated Large-scale Radiology Image Database → Flickr 30K, MS COCO …?

** Significantly better quantitative classification performance than [Shin et al., CVPR 2015; Shin et al. JMLR 2016], in recognizing learned categories!!

[Shin et al. CVPR 2015] may be the first work demonstrating the benefits of transfer learning from ImageNet to ~0.22M radiology key image database.

Cluster Number	Radiologist's Impression (Label)	Withih-Class Coherence or Consistency
1	Abdomen or pelvis CT with disease	High
2	Abdomen or pelvis CT with disease	Moderately High
3	Abdomen or pelvis CT with disease (predominantly bowel)	Moderately High
4	Chest CT with pulmonary disease	High
5	Neck or chest CT with tumor or lymphadenopathy	High
6	Abdomen or pelvis CT with disease	High
7	Chest CT containing mass	Moderately High
8	Chest CT with pulmonary disease (predominantly pleural)	Moderately High
9	Body MRI (chest, abdomen, or pelvis) with mass/lesion	High
10	Chest or abdomen CT (chest, abdomen, or pelvis) with lesion in hepatic dome	High
11	Chest CT with lesions (predominately pleural)	High
12	Abdomen CT with lesions (predominately hepatic or renal)	High
13	Abdominal or Pelvis MRI with tumor	High
14	Lower extremity CT with disease	Moderately Low (2 internal sub-classes)
15	Chest CT with nodule/mass	High
16	Abdominal MRI with tumor (predominately hepatic)	Moderately High
17	Abdomen CT with hepatic lesions	High
18	Pelvis CT with lesions/tumor	High
19	Chest CT with pulmonary disease (predominately pleural)	High
20	Chest CT with tumor	High

FRAMEWORK OF LDPO-PM:
[WANG ET AL. ARXIV:1603.07965]



[1] Mid-level Deep Pattern Mining. CVPR 2015; arxiv:**1506.06343**

Scene Recognition Dataset

MIT Indoor-67
(indoor scenes, 67 categories, 15620 images)

Building-25
(Architecture Style, 25 categories, 4794 images)

Scene-15
(Both indoor and outdoor, 15 categories, 4485 images)

GTCx

Results- Learned Features for Supervised classification

* Unsupervised feature representation learning on MIT Indoor-67(67 categories, 15620 images)

Method	Accuracy (%)	Comment
D-patch [1]	38.10	2012
D-parts [2]	51.40	2013
DMS [3]	64.03	2013
MDPM-Alex [4]	64.12	2015
MDPM-VGG [4]	76.95	2015
MetaObject [5]	78.90	2015
LDPO-PM-Alex*	**63.68**	Our unsupervised method
LDPO-PM-VGG*	**72.52**	Our unsupervised method
FC (CaffeRef) [4]	57.74	CNN FC feature
FC (VGG) [4]	68.87	CNN FC feature
CONV-FV (CaffeRef) [6]	**69.70**	Fisher Vector (supervised)
CONV-FV (VGG) [6]	**81.00**	Fisher Vector (supervised)

GTCx

References:

1. "Interleaved Text/Image Deep Mining on a Large-Scale Radiology Image Database", IEEE CVPR 2015
2. "Interleaved Text/Image Deep Mining on a Large-Scale Radiology Image Database for Automated Image Interpretation", Journal of Machine Learning Research, arXiv:1505.00670, 2016
3. "Unsupervised Category Discovery via Looped Deep Pseudo-Task Optimization Using a Large Scale Radiology Image Database", arXiv:1603.07965, 2016
4. "Unsupervised Joint Mining of Deep Features and Image Labels for Large-scale Radiology Image Categorization and Scene Recognition", …, 2016

GTCx

TAKE-HOME-MESSAGES

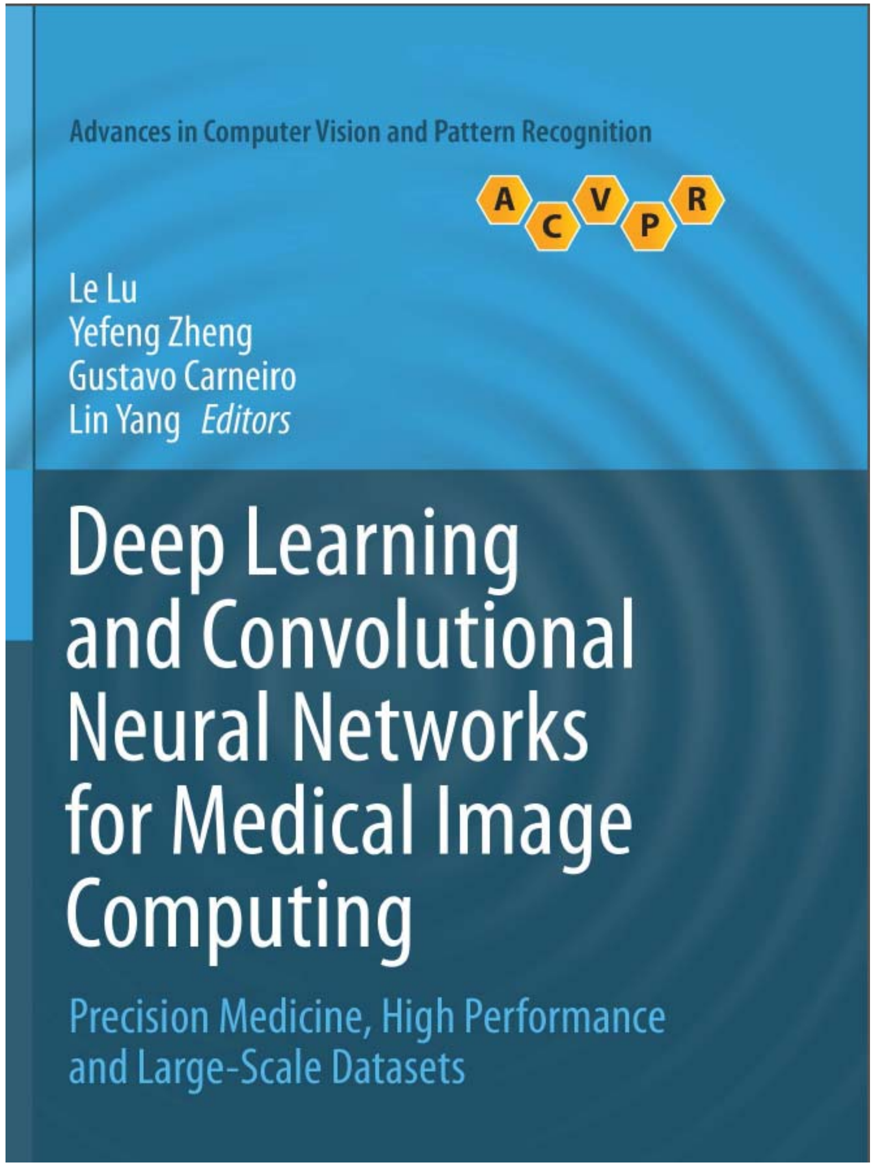
- There exists the exact mapping of semantic object detection, object segmentation & parsing, and image-text captioning problems towards corresponding medical imaging tasks

- Preventative and precision medicine in radiology are feasible means to advance healthcare through improved quantitative performance on hard and important clinical problems.

- It is time to empower deep learning or deep neural networks under novel visual representations to solve previously poorly performed yet critical issues (lymph node, pancreas, chest X-ray, unconstrained ILD prediction, etc.) from doctors' wish-list; and work with them to make new diagnosis protocols!

- **Key Technical Elements:** Compositional & Hierarchical Visual Representations, Structured Prediction & Optimization, Heterogeneous Visual Cues (Boundary, etc.) integration, CNN Architectures & Loss Functions, Sequential vs. End-to-end Training …

GTCx


Advances in Computer Vision and Pattern Recognition
A C V P R
Le Lu
Yefeng Zheng
Gustavo Carneiro
Lin Yang Editors
Deep Learning and Convolutional Neural Networks for Medical Image Computing
Precision Medicine, High Performance and Large-Scale Datasets

Thank you & our amazing trainees, collaborators!

Radiology and Imaging Sciences
National Institutes of Health Clinical Center

le.lu@nih.gov; rms@nih.gov

Thanks NIH Intramural Research Program (NIH-IRP) for support and NVidia for donating Tesla K40 and Titan X GPUs! NIH FARE awards (2014,2015, 2016), KRIBB Fellowship, NDSEG Fellowship, MICCAI student travel award 2016, RSNA trainee research prize 2016, …

NIH National Institutes of Health
Turning Discovery Into Health