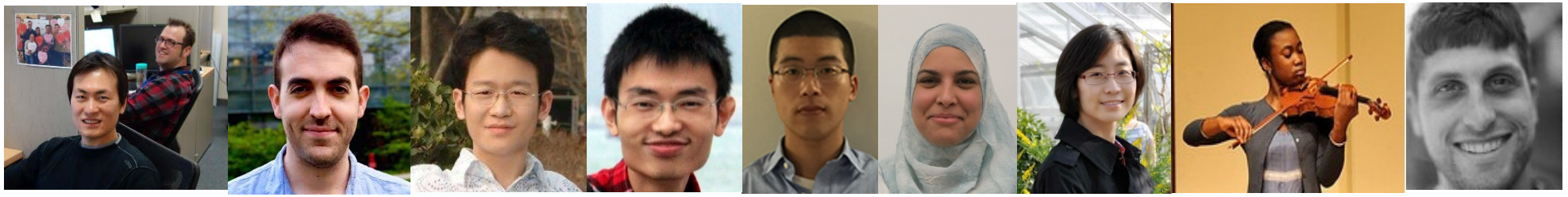
Deep Learning for Medical Imaging and Clinical Informatics: Preventive and Precision Medicine Perspectives

Le Lu (PhD)

Ping An Technology US Research Labs (Bethesda, Maryland)
Adjunct Faculty, Johns Hopkins University & Medical Institutions
lelu@cs.jhu.edu; llu7@jhmi.edu; lvle746@pingan.com

Talk slides used for GTC Taiwan 2018; CGMH 2018; SNMMI 2018



Q1: Do deep learning and deep neural networks help in medical imaging or medical image analysis problems? (Yes)

- ✓ **Deep CAD:** Lymph node application package (52.9% → 85%, 83%) and many CAD Applications
- ✓ **Deep Segmentation → Precision Medicine in Radiology & Oncology: Pancreas** segmentation application package (~53% → 81.14% in Dice Coefficient) and beyond (prostate segmentation, …)
- ✓ **Deep Lung** (Interstitial Lung Disease) Application **Package** + **DL Reading Chest X-ray**; **Pathological Lung Segmentation**, …
- ✓ **Unsupervised category discovery using looped deep pseudo-task optimization (mapping large-scale radiology database with category meta-labels) → Learning from PACS!**
- ✓ **A large-scale Chest X-ray database (with NLP based annotation): Dataset and Benchmark (CVPR 2017; CVPR 2018: TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays)**
- ✓ **Deep Lesion Graph in the wild: Relationship Learning and Organization of Significant Radiology Image Findings in a Diverse Large-scale Lesion Database (CVPR 2018)**

- **Updates & Publications can be downloaded: www.cs.jhu.edu/~lelu; https://clinicalcenter.nih.gov/drd/staff/le_lu.html;**

Perspectives

- Why the previous or current computer-aided diagnosis (CADe; CADx) systems are not particularly successful yet? **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, collaborative** 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 (segmentation) biomarkers** in precision medicine to **better assist human doctors to make more precise decisions;** b) **patient-level similarity retrieval system** for personalized diagnosis/therapy treatment: show by examples!

Three Key Problems (I)

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 (TMI 2016; MICCAI 2014); **MICCAI Young Researcher Publication Impact Award finalist!**
- Empirical analysis on **Lymph node detection** and **interstitial lung disease** (ILD) classification using CNN (TMI 2016b); prostate CAD (ISBI 2017; MICCA 2018), Vascular Calcification plaque (ISBI 2017), COLITIS (ISBI 2016), …
- 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); ISBI 2017 …

➢ **Clinical Impacts**: producing various *high performance* "second or first reader" **CAD use cases** and applications → effective imaging based prescreening (triage) tools on a cloud based platform for large population preventive profiling

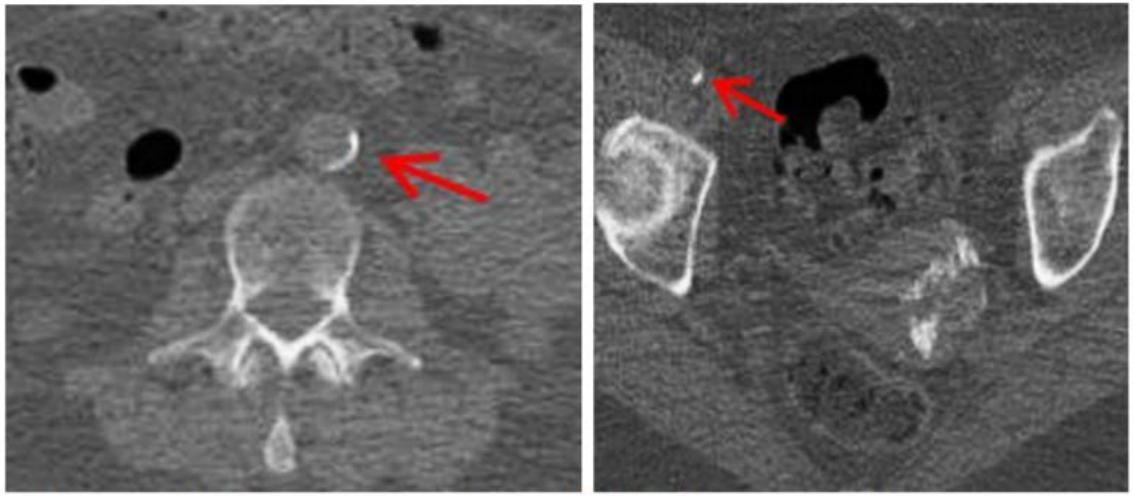
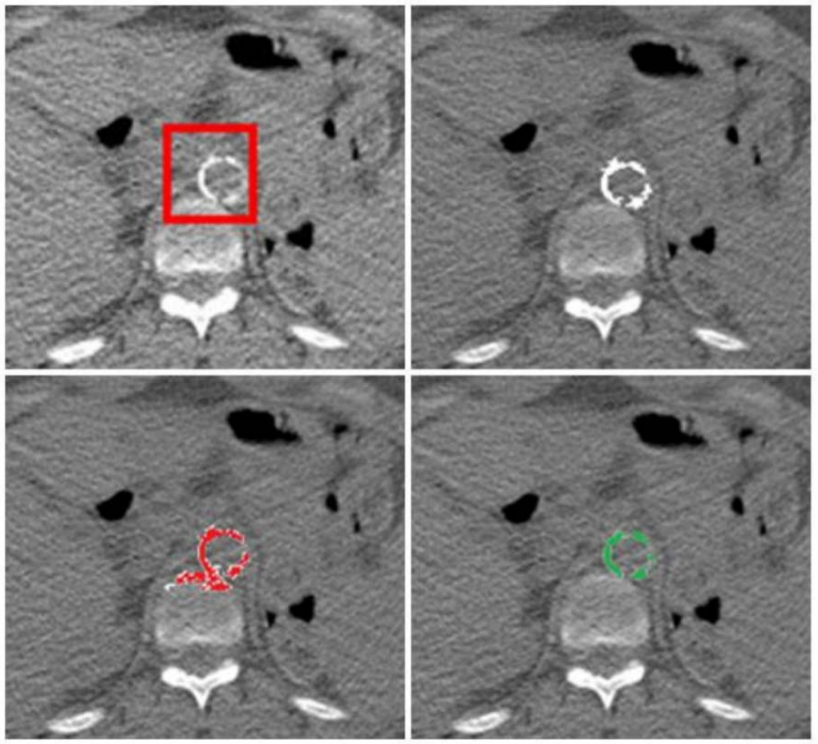
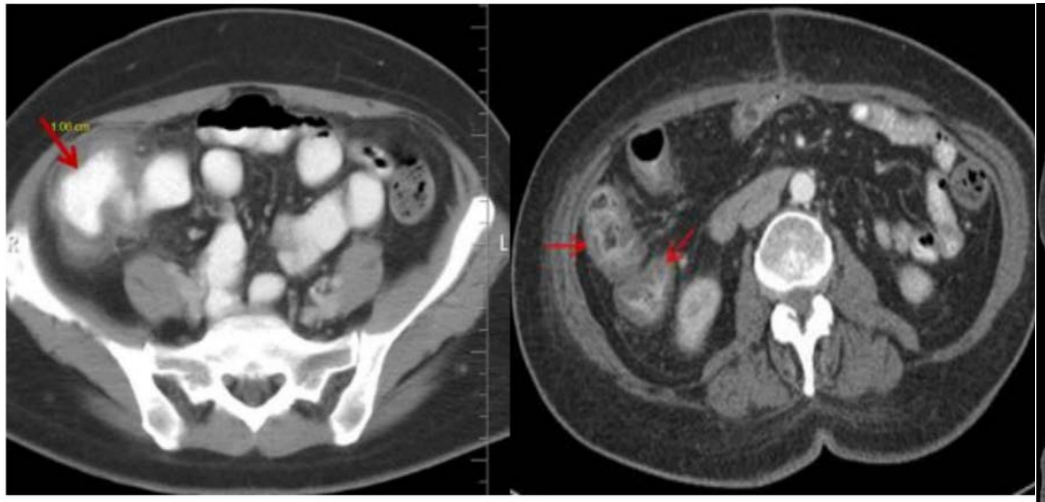
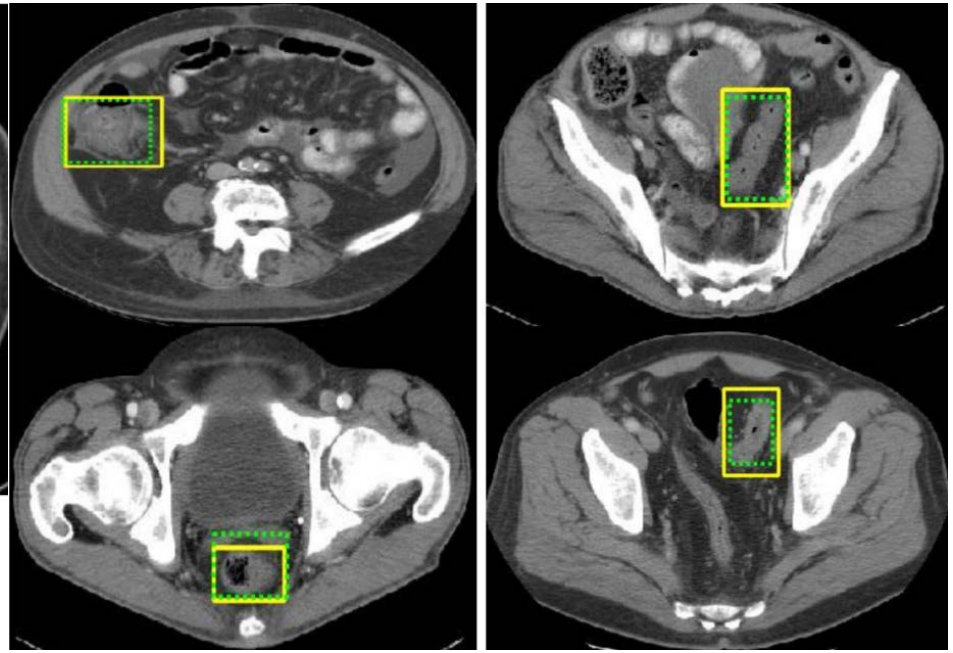


Figure 1. Examples of calcified plaques (red arrows) on abdominal (left) and pelvic (right) CT scans.



Atherosclerotic Vascular Calcification Detection and Segmentation on Low Dose Computed Tomography Scans …, Liu et al., IEEE ISBI 2017





***Detecting the undetectables?**

***Fitting in practical/real clinical settings in the wild??**

COLITIS DETECTION ON COMPUTED TOMOGRAPHY USING REGIONAL CONVOLUTIONAL NEURAL NETWORKS, Liu et al., IEEE ISBI 2016

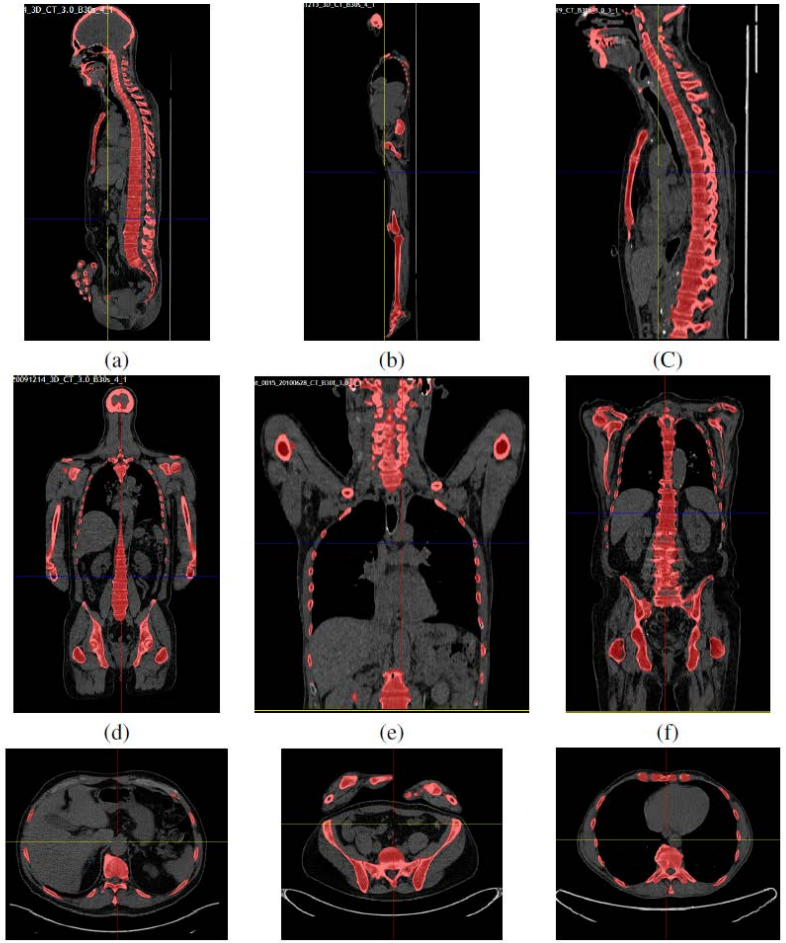
Three Key Problems (II)

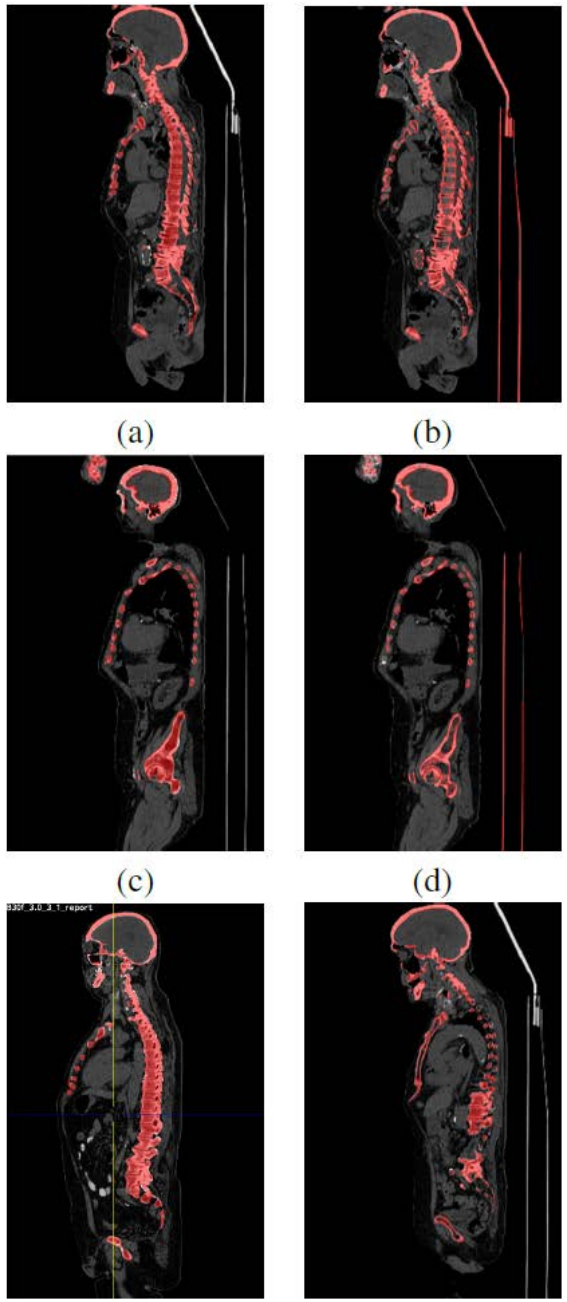
Semantic Segmentation in Medical Image Analysis

- "DeepOrgan" for **pancreas segmentation** (MICCAI **2015**) via scanning superpixels using multi-scale deep features ("Zoom-out") and probability map embedding.
- Deep segmentation on **pancreas** and **lymph node clusters** with Holistically-nested neural networks as building blocks to learn unary (segmentation **mask**) and pairwise (labeling segmentation **boundary**) CRF terms + spatial aggregation or + structured optimization, ISBI **2016**, MICCAI **2016**.
- The focus of recent MICCAI **2016-2018** papers since this is a much needed task → **Small** datasets; **(de-)compositional** representation is still the key. Scale up to thousands, thousands of patients if not more than that; weakly supervised segmentation → Effective and Efficient Precision Biomarkers, even predicting the future tumor growth (DL tumor growth model prediction MICCAI **2017**, TMI **2018**, MICCAI **2018**)!

➢ **Clinical Impacts**: semantic segmentation can help compute clinically more accurate and desirable precision imaging bio-markers or measurements → precision imaging personalized treatment and therapy → less guess more doing (RECIST 1.1)…

Results on PET-CT Patient Datasets (pathological …)





***robust, precision/accuracy, speed!**

Towards whole Body precision measurements or computable precision imaging biomarkers

→

"Robust Whole Body 3D Bone Masking via Bottom-up Appearance Modeling and Context Reasoning in Low-Dose CT Imaging", Lu et al., IEEE WACV 2016

→

Bone Mineral Density (BMD) scores, Muscle/Fat volumetric measurements in whole body or arbitrary FOV imaging … lung nodules, bone lesions, head-and-neck radiation sensitive organs, segmenting flexible soft anatomical structures for precision medicine, all clinically needed!

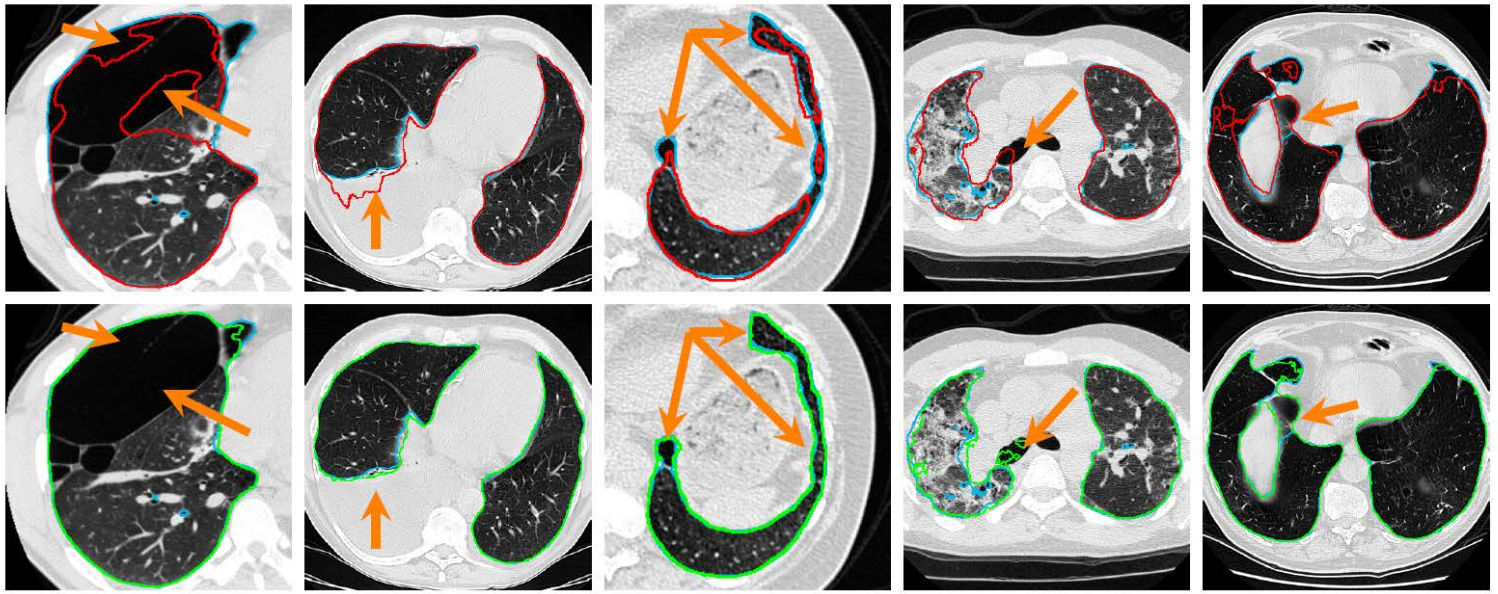


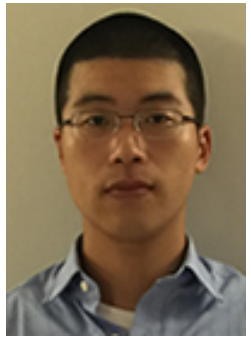
Fig. 2: Example masks of HNN and P-HNN, depicted in red and green, respectively. Ground truth masks are rendered in cyan. (a) HNN struggles to segment the pulmonary bullae, whereas P-HNN captures it. (b) Part of the pleural effusion is erroneously included by HNN, while left out of the P-HNN lung mask. (c) P-HNN is better able to capture finer details in the lung mask. (d) In this failure case, both HNN and P-HNN erroneously include the right main bronchus; however, P-HNN better captures infiltrate regions. (e) This erroneous ground-truth example, which was filtered out, fails to include a portion of the right lung. Both HNN and P-HNN capture the region, but P-HNN does a much better job of segmenting the rest of the lung.

NSERC Fellow

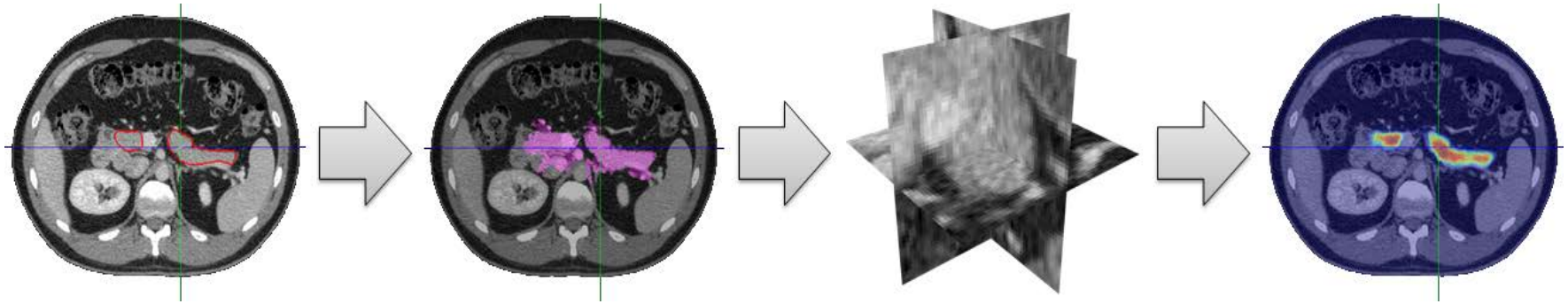
MICCAI 2017 Young Scientist Award runner-up

Asst. Professor
Nagoya Uni.,
Japan

A Roadmap of **Bottom-up Deep** Pancreas Segmentation: from Image Patch, Region, to Holistically-nested CNNs (HNN), P-HNN, Convolutional LSTM (context), …



ISTP Fellow,
2012-2014



P-ConvNet

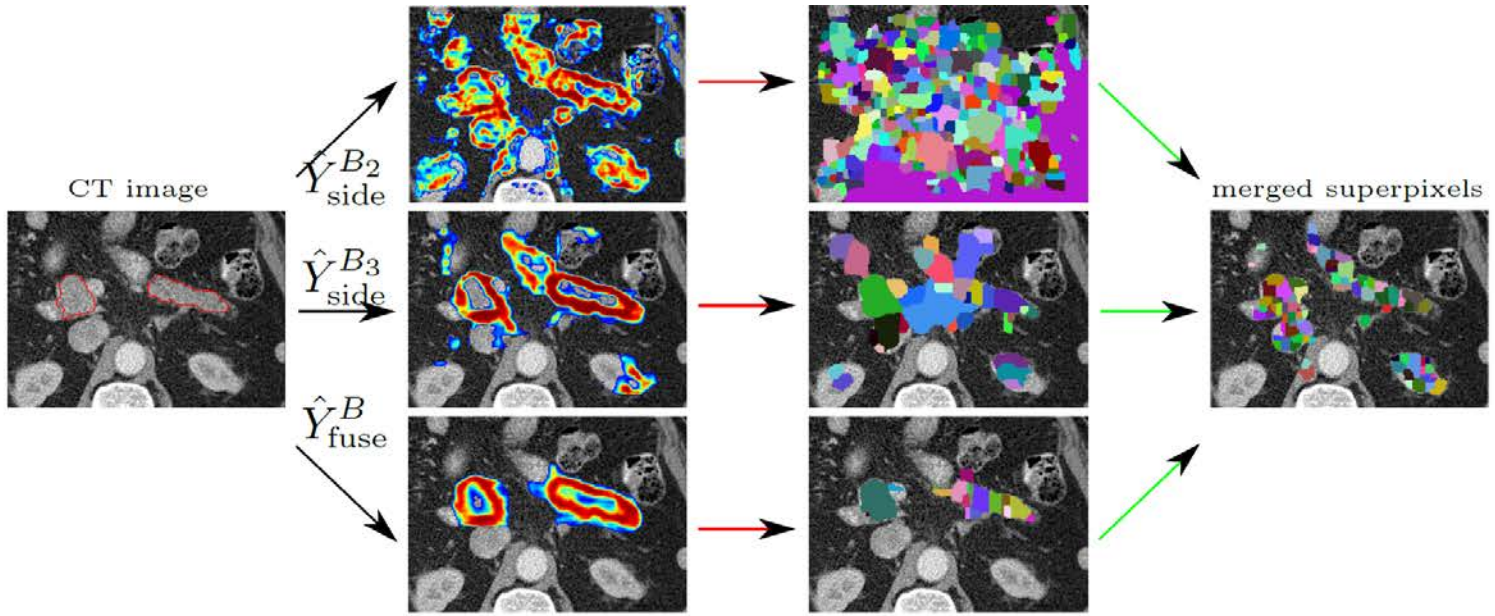
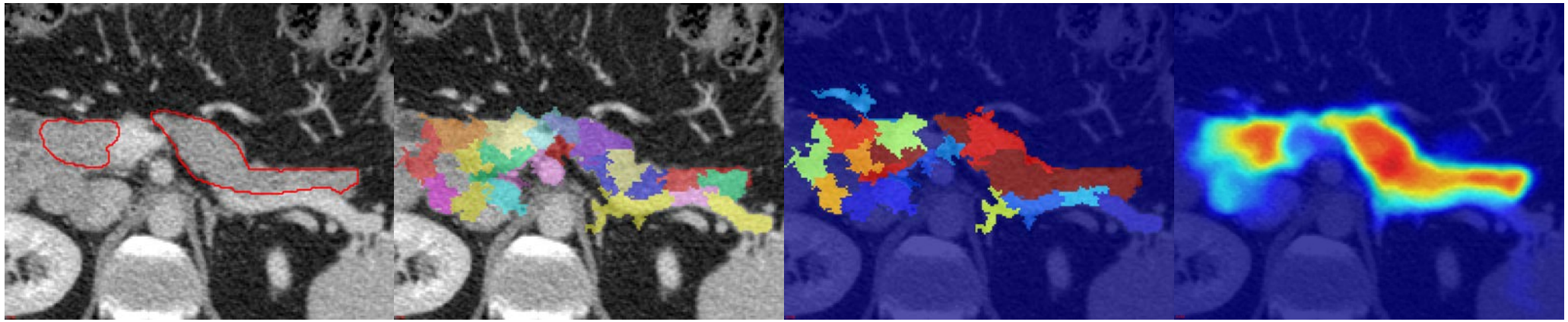
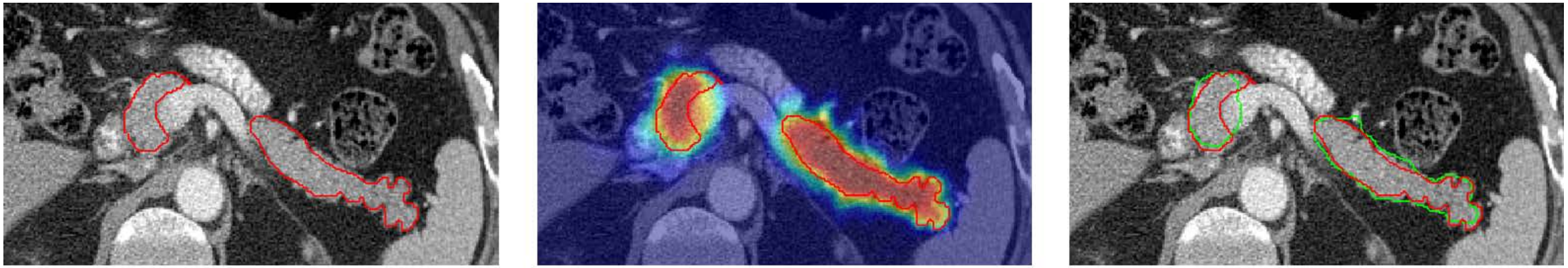


Fig. 2: "Multiscale Combinatorial Grouping" (MCG) [16] on three different scales of learned boundary predication maps from **HNN-B**: $\hat{Y}^{B_2}_{\text{side}}$, $\hat{Y}^{B_3}_{\text{side}}$, and $\hat{Y}^{B}_{\text{fuse}}$ using the original CT image as input (shown with ground truth delineation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

Improved pancreas segmentation accuracy over previous state-of-the-art work in Dice coefficients: from 50~60% to ~84%; ASD: from 5~6mm to 0.7mm; computational time from 3 hours to <2 minutes!





a) b) c)

a) The manual ground truth annotation (in red outline)
b) The probability map
c) The final segmentation (in green outline)

DSC=82.7%.

H. Roth, et al., **DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation**. MICCAI (1) 2015: 556-564

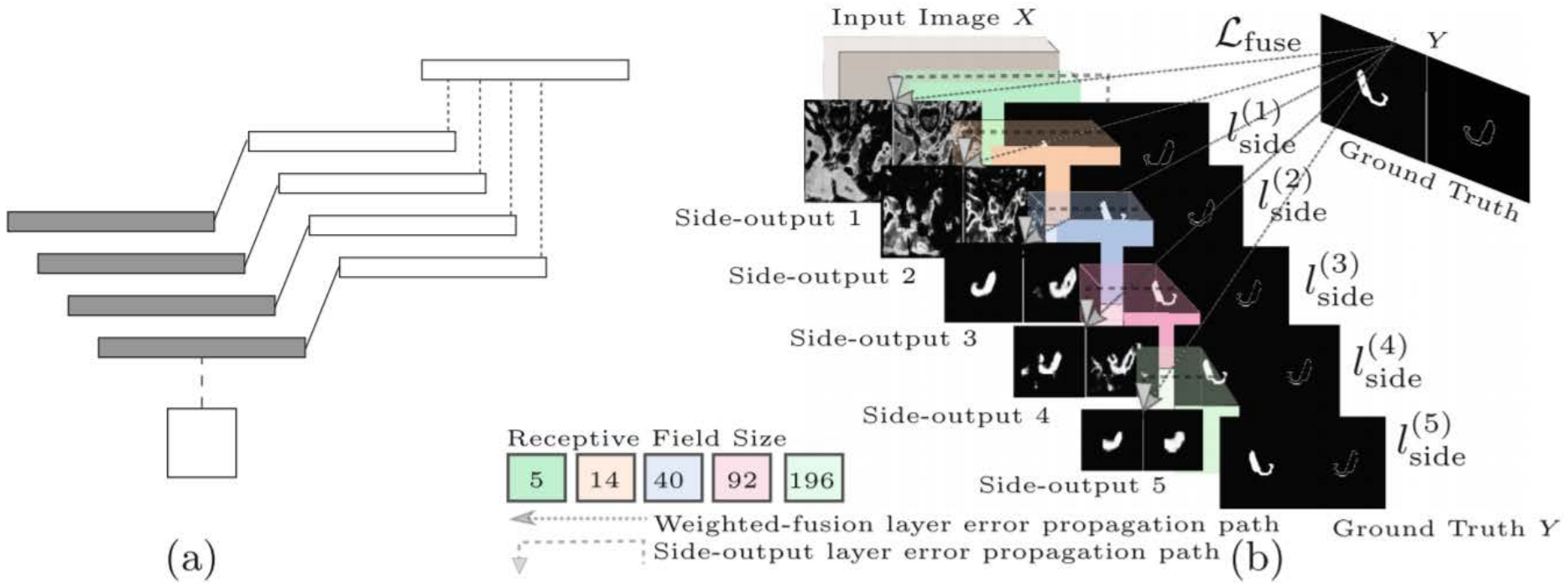
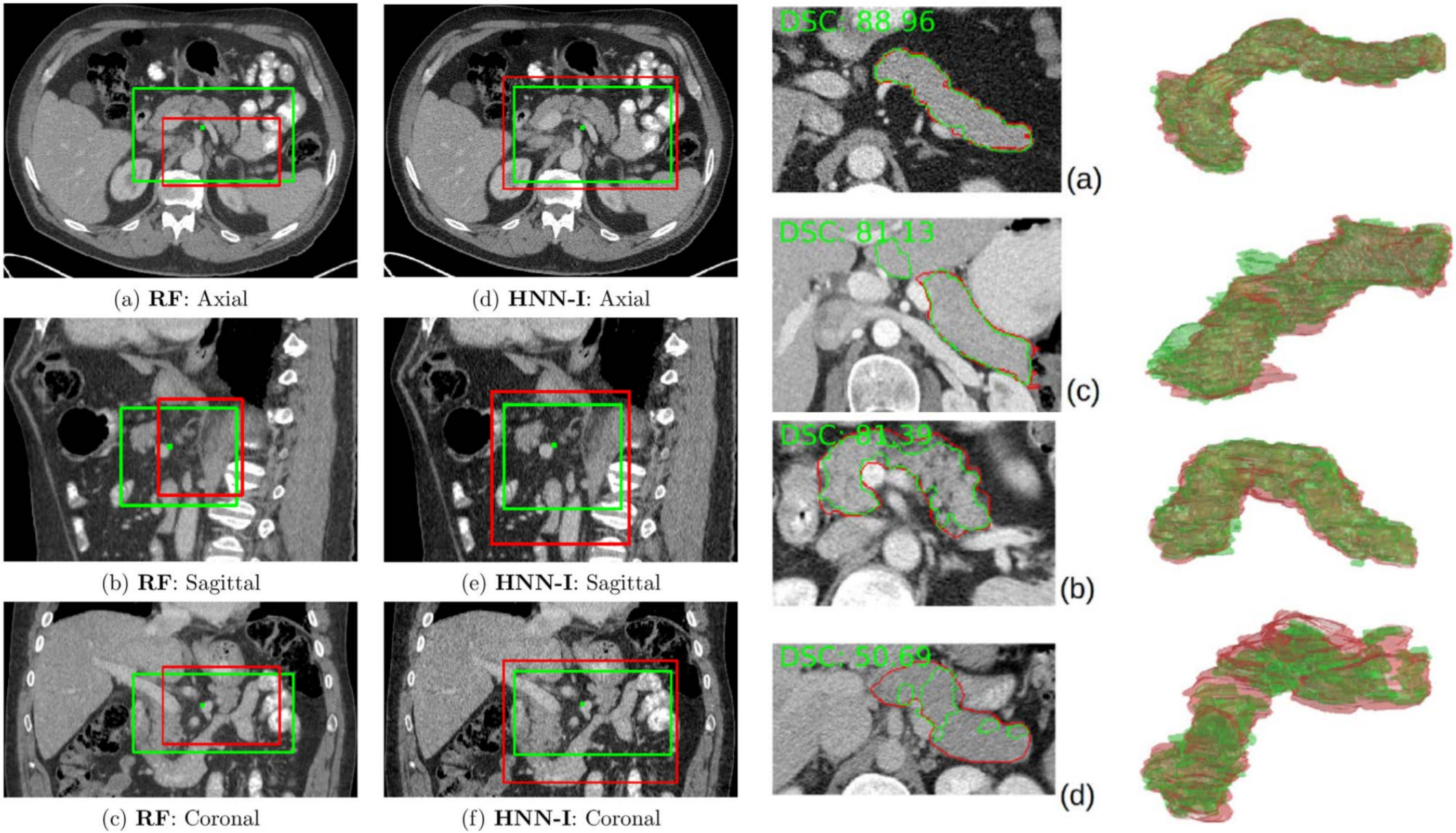
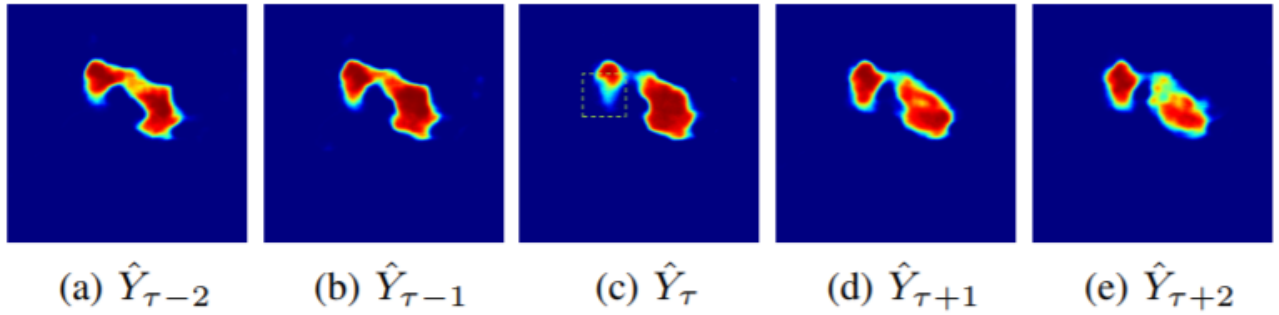


Fig. 1. Schematics of (a) the holistically-nested nets, in which multiple side outputs are added, and (b) the **HNN-I/B** network architecture for both interior (left images) and boundary (right images) detection pathways. We highlight the error back-propagation paths to illustrate the deep supervision performed at each side-output layer after the corresponding convolutional layer. As the side-outputs become smaller, the receptive field sizes get larger. This allows HNN to combine multi-scale and multi-level outputs in a learned weighted fusion layer (Figures adapted from [11] with permission).

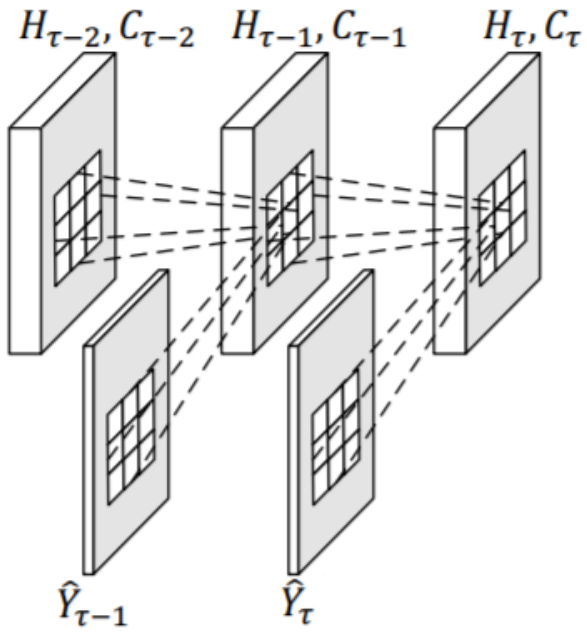
H. Roth, et al., "**Spatial Aggregation of Holistically-Nested Networks for Automated Pancreas Segmentation**", MICCAI, 2016



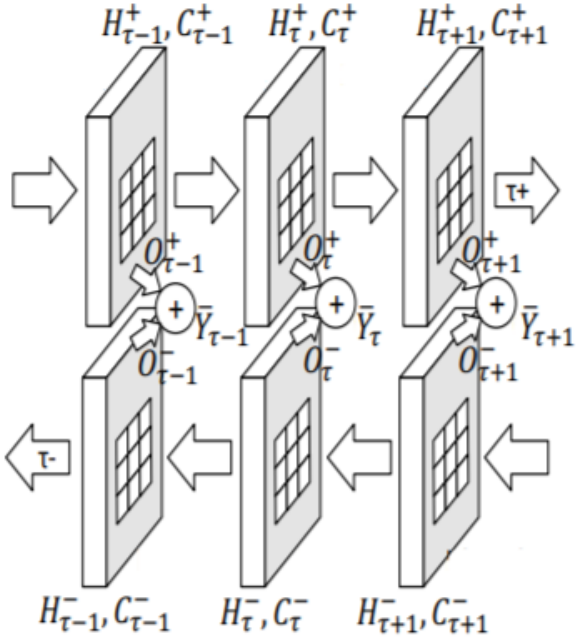
H. Roth, et al., **Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation**, Medical Image Analysis, 45 (2018) 94–107



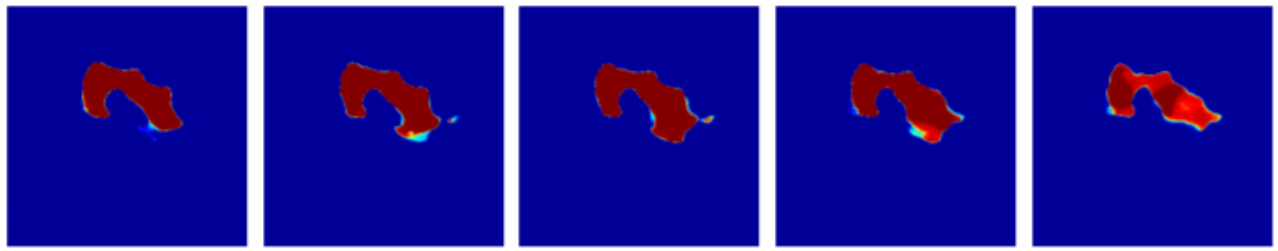
(a) $\hat{Y}_{\tau-2}$ (b) $\hat{Y}_{\tau-1}$ (c) $\hat{Y}_{\tau}$ (d) $\hat{Y}_{\tau+1}$ (e) $\hat{Y}_{\tau+2}$



(f) CLSTM



(g) Bi-direction CLSTM



(h) $\bar{Y}_{\tau-2}$ (i) $\bar{Y}_{\tau-1}$ (j) $\bar{Y}_{\tau}$ (k) $\bar{Y}_{\tau+1}$ (l) $\bar{Y}_{\tau+2}$

Fig. 4: The main construction units of the proposed RNN sub-network and its input/output segmentation sequence. The sequence of CNN sub-network outputs is shown in the first row (a-e), is taken as the input of the bi-direction CLSTM (g), which is an RNN architecture composed of 2 layers of CLSTM (f) working in opposite directions. The third row (h-l) presents the corresponding output sequence, which is sharp and clean. Note that the missing pancreatic part in $\hat{Y}_{\tau}$ (c), in the green dashed box, is recovered by shape continuity modeling in $\bar{Y}_{\tau}$ (j). For visual clearity, we ommit the input $\hat{Y}_{(\cdot)}$ in the bi-direction CLSTM (g), which is same as in (f).

J. Cai, et al., **Pancreas Segmentation in CT and MRI Images via Domain Specific Network Designing and Recurrent Neural Contextual Learning**, MICCAI 2017 https://arxiv.org/abs/1803.11303

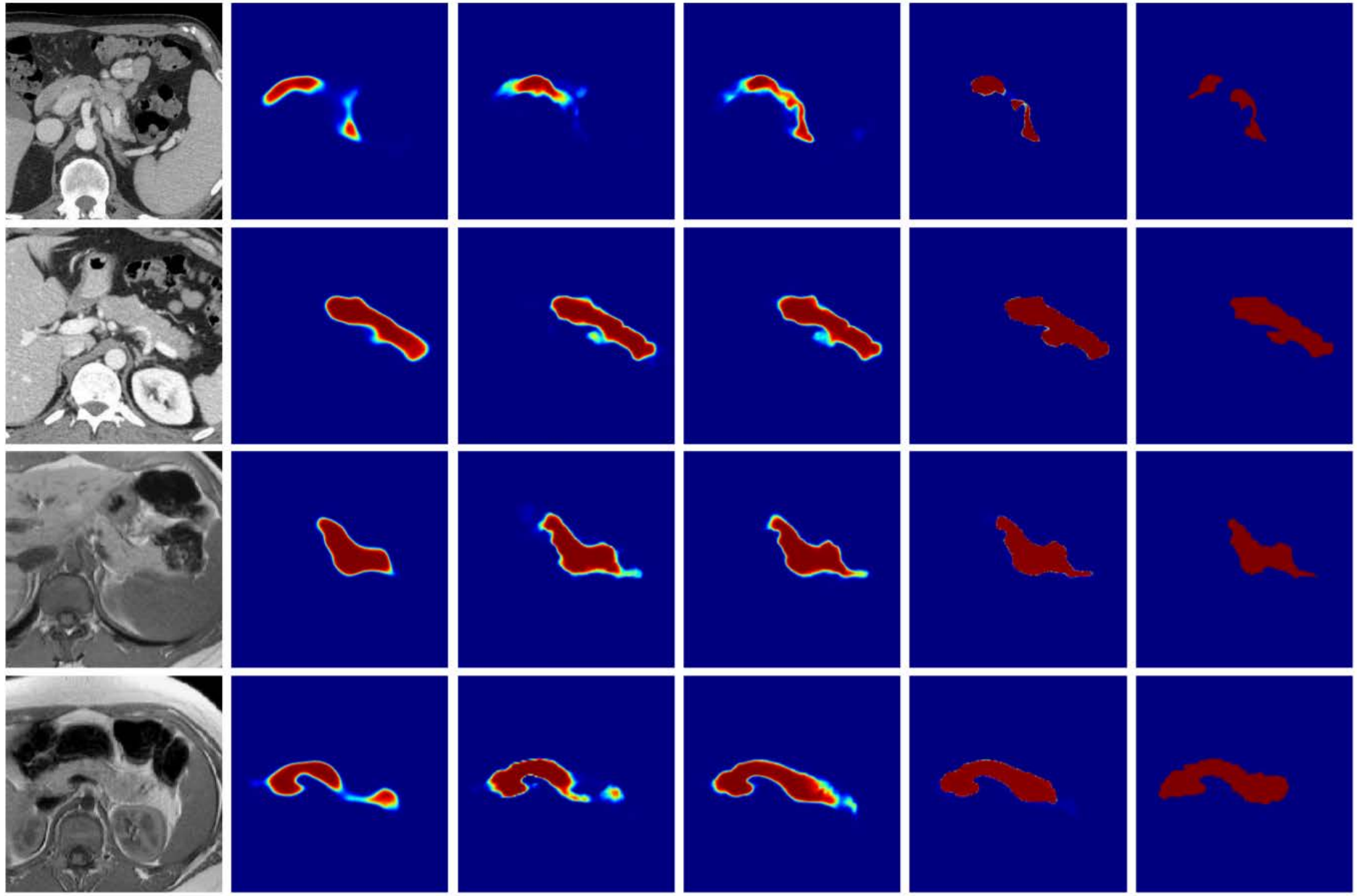


Fig. 8: Examples of output probability map: columns from left to right are the input CT/MRI image, results from HNN [12], UNET [16], the proposed PNet-MSA sub-network, and the full CNN-RNN ("PNet-MSA+BiRNN"), and the ground truth. Our model delivers the most clear probability maps which preserve detailed pancreatic boundaries.

Three Key Problems (III)

Interleaved or Joint Text/Image Deep Mining at large unconstrained environment or data sources → "large data, weak labels" (~216K 2D key image slices extracted from >60K unique patient studies)

- Interleaved Text/Image Deep Mining on a Large-Scale Radiology Image Database (IEEE 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**)
- Learning to Read Chest X-Rays: Recurrent Neural Cascade Model for Automated Image Annotation, (IEEE CVPR **2016**)
- Unsupervised Category Discovery via Looped Deep Pseudo-Task Optimization Using a Large Scale Radiology Image Database, IEEE WACV **2017**
- ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases, IEEE CVPR **2017**; CVPR **2018**, MICCAI **2018**

➢ **Clinical Impacts**: eventually to build an automated 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.

Learning to Read Chest X-ray using Deep Neural Networks (a little more like humans' interpretation?)

[Shin et al., IEEE CVPR 2016]

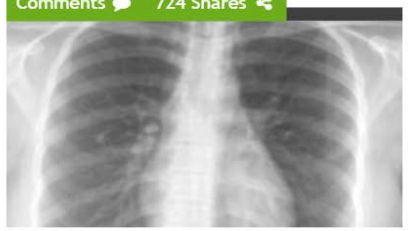
Lung diseases killing 4 million people every year, in comparison to Nearly **1.3 million people** die in **road crashes** each year!

Statistics from internet …

NVIDIA ACCELERATED COMPUTING | Getting Started | Downloads | Training | Ecosystem | F

NEWS CENTER | News | Research | Events

Comments 724 Shares

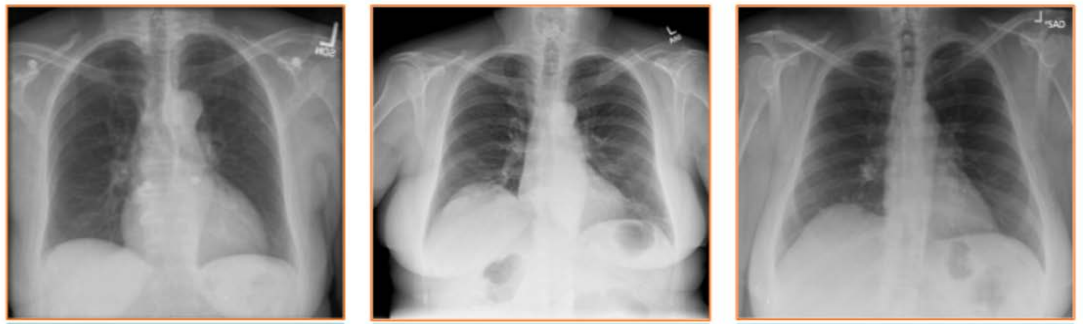


Detecting and Labeling Diseases in Chest X-Rays with Deep Learning

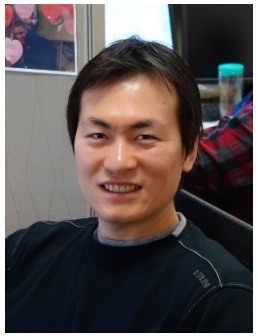
April 14, 2016

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.

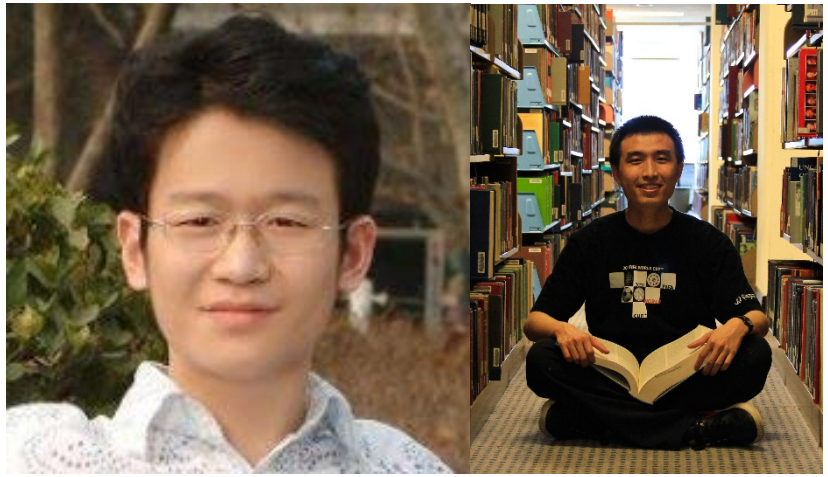
Accelerated by Tesla GPUs, the team trained their convolutional neural networks on a publicly available radiology dataset of chest x-rays and reports to describe the characteristics of a disease, such as location, severity and the affected organs.

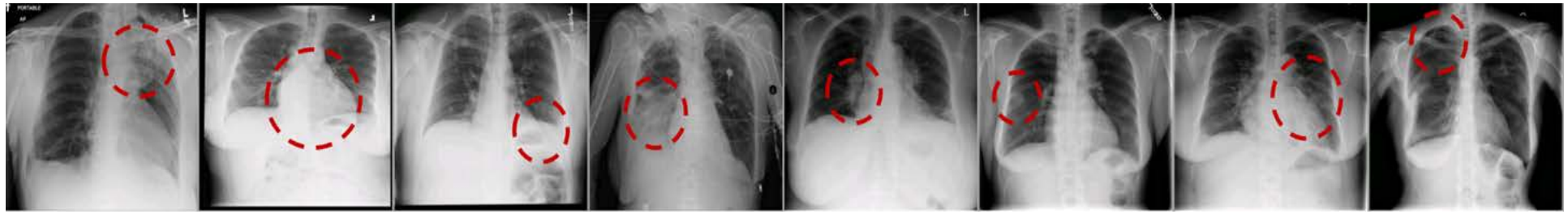


Examples of annotation generations (light green box) compared to true annotations (yellow box) for input images in the test set.



ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases

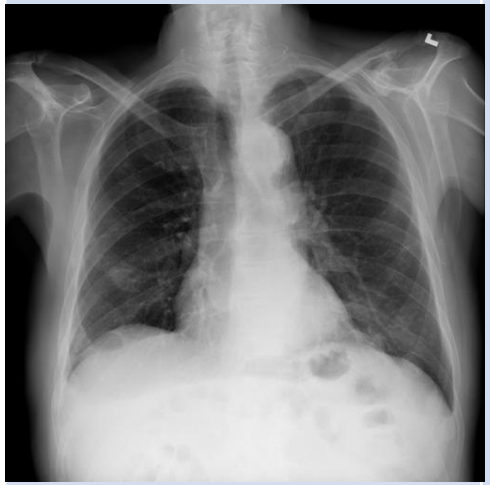




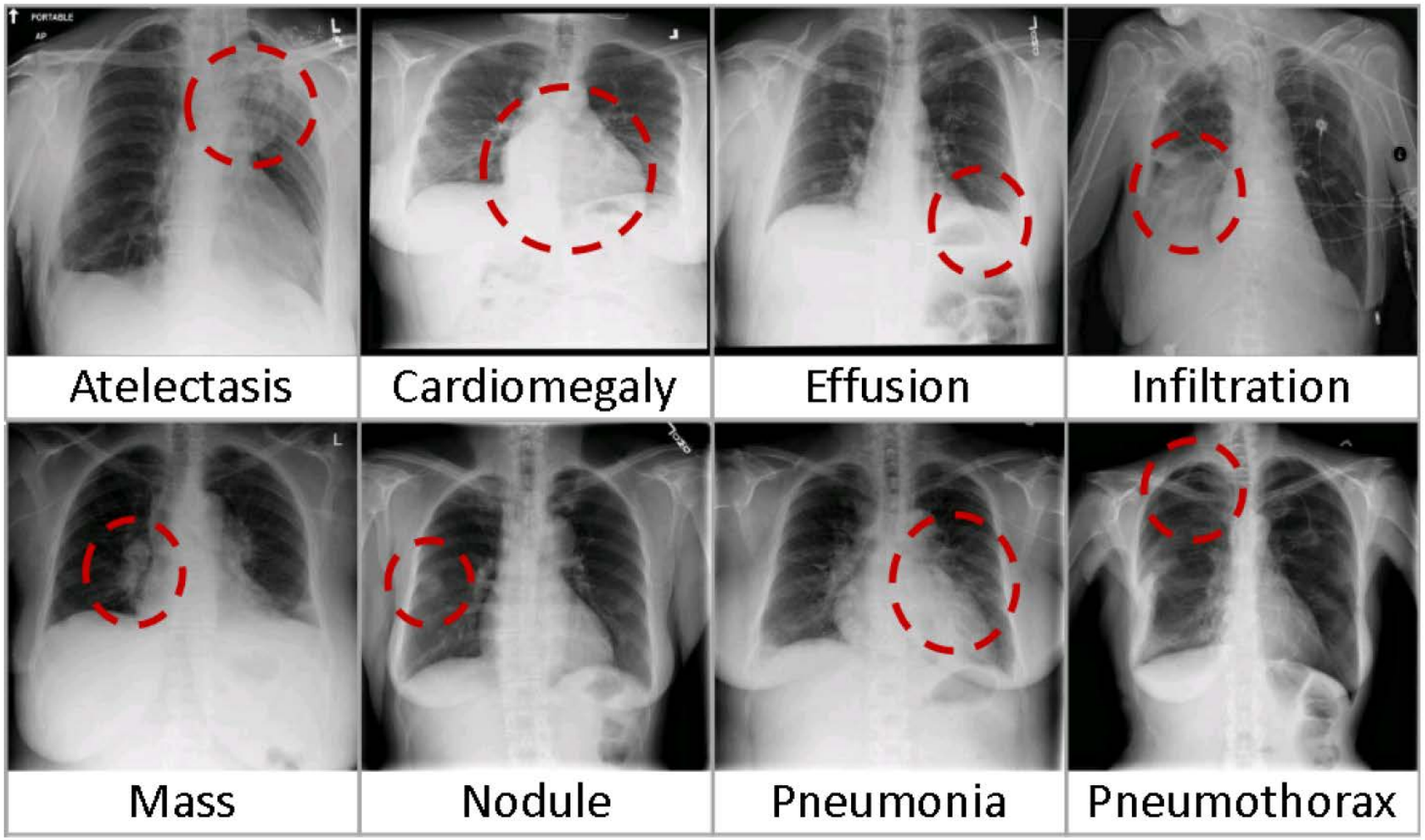
Motivation

1. Build a large-scale Chest X-rays dataset to facilitate the data-hungry deep learning paradigms
2. Critical/common disease patterns (predefined by radiologist)
3. Multiple labels (disease patterns) for images
4. Radiological report for each image
5. Small number of bounding boxes (outline the location of disease symptoms) for each disease category
6. Potential applications:
    A. Disease detection/classification
    B. Disease localization
    C. Automatic radiological generation
    …

A Sample Entry

Image	Report	Label	Bounding Box (optinal)
	findings: pa and lateral views of the chest demonstrate significantly improved bilateral lower lung field interstitial markings compatible with linear atelectasis. unchanged right 9th rib fracture peripherally. unchanged ossification left coracoacromial ligament. the cardiac and mediastinal contours are stable. impression: improved bilateral lower lung field linear atelectasis.	atelectasis	−<questions> <img_id>00154220_01.jpg</img_id> <class>1</class> <boxes/> <object_name>atelectasis</object_name> <image_url>images/00154220_01.jpg</image_url> <task>db</task> </questions> −<output> −<answer> <x>225.08474576271186</x> <y>547.0192167637712</y> <w>86.77966101694915</w> <h>79.1864406779661</h> </answer> <time>44616</time> <eval>neutral</eval> <img_id>00154220_01.jpg</img_id> </output>

8 Common Thorax Diseases



Disease Label Extraction: Two-stage NLP of Radiology Reports

Stage 1: Pathology Detection

- **DNorm** is used to map every mention of keywords in a report to a unique concept ID in the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT), a standardized vocabulary of clinical terminology for the electronic exchange of clinical health information.
- Another ontology-based approach, **MetaMap**, is adopted for the detection of Unified Medical Language System (UMLS) Metathesaurus.
- The results of DNorm and MetaMap are merged

Pneumonia	
C0032285	pneumonia
C0577702	basal pneumonia
C0578576	left upper zone pneumonia
C0578577	right middle zone pneumonia
C0585104	left lower zone pneumonia
C0585105	right lower zone pneumonia
C0585106	right upper zone pneumonia
C0747651	recurrent aspiration pneumonia
C1960024	lingular pneumonia
Pneumothorax	
C0032326	pneumothorax
C0264557	chronic pneumothorax
C0546333	right pneumothorax
C0546334	left pneumothorax

Sample SNOMED-CT concepts

Two-stage NLP of Radiology Reports

Stage 2: Removal of negation and uncertainty

- Rule out those negated pathological statements and uncertain mentions of findings and diseases
- Defined the rules on the dependency graph, by utilizing the dependency label and direction information between words, e.g.

Rule	Example
Negation	
no$\leftarrow * \leftarrow$ DISEASE	No acute pulmonary disease
$* \rightarrow$ *prep_without* $\rightarrow$ DISEASE	changes without focal airspace disease
clear/free/disappearance $\rightarrow$ *prep_of* $\rightarrow$ DISEASE	clear of focal airspace disease, pneumothorax, or pleural effusion
$* \rightarrow$ *prep_without* $\rightarrow$ evidence $\rightarrow$ *prep_of* $\rightarrow$ DISEASE	Changes without evidence of acute infiltrate
no $\leftarrow$ *neg* $\leftarrow$ evidence $\rightarrow$ *prep_of* $\rightarrow$ DISEASE	No evidence of active disease
Uncertainty	
cannot $\leftarrow$ *md* $\leftarrow$ exclude	The aorta is tortuous, and cannot exclude ascending aortic aneurysm
concern $\rightarrow$ *prep_for* $\rightarrow *$	There is raises concern for pneumonia
could be/may be/...	which could be due to nodule/lymph node
difficult $\rightarrow$ *prep_to* $\rightarrow$ exclude	interstitial infiltrates difficult to exclude
may $\leftarrow$ *md* $\leftarrow$ represent	which may represent pleural reaction or small pulmonary nodules
suggesting/suspect/... $\rightarrow$ *dobj* $\rightarrow$ DISEASE	Bilateral pulmonary nodules suggesting pulmonary metastases

Two-stage NLP of Radiology Reports

Stage 2: Removal of negation and uncertainty

- Rule out those negated pathological statements and uncertain mentions of findings and diseases
- Defined the rules on **the Dependency Graph via Parsing**, by utilizing the dependency label and direction information between words, e.g.

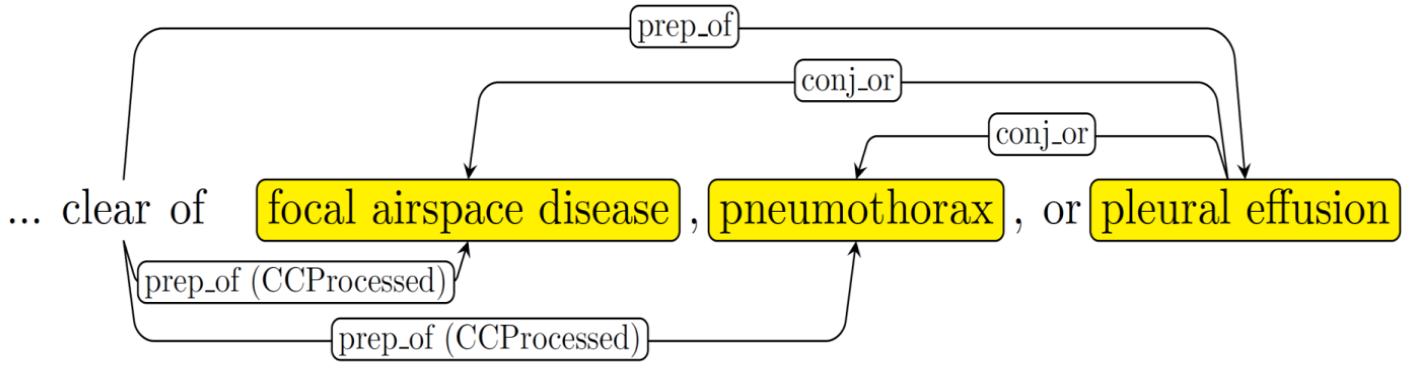
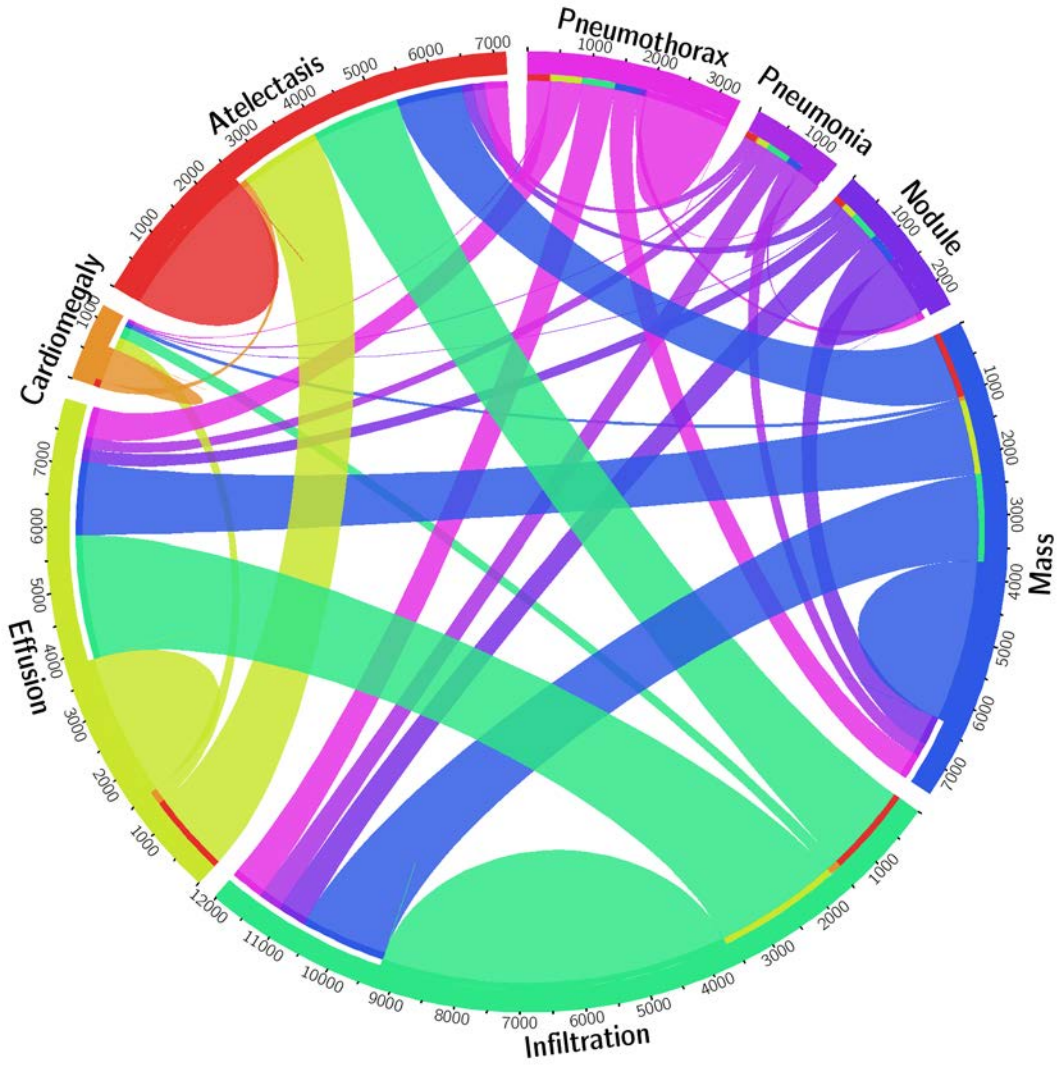


Image preparation

1. Select frontal view images only
2. Convert to 8-bit RBG images by using default window level and width
3. Resize image to 1024 * 1024
4. Totally 108,948 frontal view X-ray images of 32,717 unique patients

Disease Category Statistics

Item #	OpenI	Ov.	ChestX-ray8	Ov.
Report	2,435	-	108,948	-
Annotations	2,435	-	-	-
Atelectasis	315	122	5,789	3,286
Cardiomegaly	345	100	1,010	475
Effusion	153	94	6,331	4,017
Infiltration	60	45	10,317	4,698
Mass	15	4	6,046	3,432
Nodule	106	18	1,971	1,041
Pneumonia	40	15	1,062	703
Pneumothorax	22	11	2,793	1,403
Normal	1,379	0	84,312	0



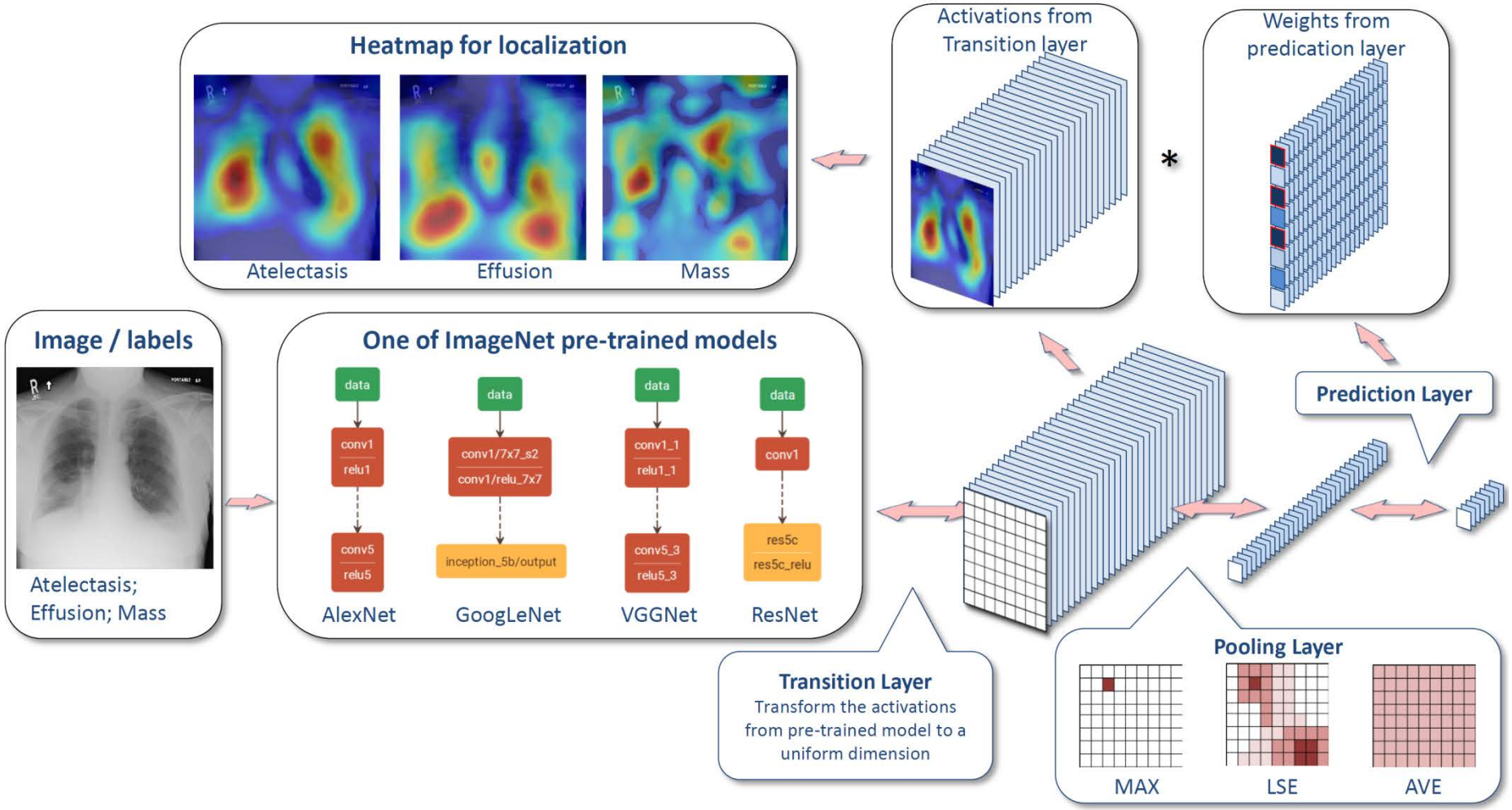
Evaluation of Disease Labeling

Item #	OpenI	Ov.	ChestX-ray8	Ov.
Report	2,435	-	108,948	-
Annotations	2,435	-	-	-
Atelectasis	315	122	5,789	3,286
Cardiomegaly	345	100	1,010	475
Effusion	153	94	6,331	4,017
Infiltration	60	45	10,317	4,698
Mass	15	4	6,046	3,432
Nodule	106	18	1,971	1,041
Pneumonia	40	15	1,062	703
Pneumothorax	22	11	2,793	1,403
Normal	1,379	0	84,312	0

Disease	MetaMap			Our Method		
	P /	R /	F	P /	R /	F
Atelectasis	0.95 /	0.95 /	0.95	0.99 /	0.85 /	0.91
Cardiomegaly	0.99 /	0.83 /	0.90	1.00 /	0.79 /	0.88
Effusion	0.74 /	0.90 /	0.81	0.93 /	0.82 /	0.87
Infiltration	0.25 /	0.98 /	0.39	0.74 /	0.87 /	0.80
Mass	0.59 /	0.67 /	0.62	0.75 /	0.40 /	0.52
Nodule	0.95 /	0.65 /	0.77	0.96 /	0.62 /	0.75
Normal	0.93 /	0.90 /	0.91	0.87 /	0.99 /	0.93
Pneumonia	0.58 /	0.93 /	0.71	0.66 /	0.93 /	0.77
Pneumothorax	0.32 /	0.82 /	0.46	0.90 /	0.82 /	0.86
Total	0.84 /	0.88 /	0.86	0.90 /	0.91 /	0.90

Table 2. *Evaluation of image labeling results on OpenI dataset. Performance is reported using P, R, F1-score.*

Framework Overview



Multi-label Setting

- We define a 8-dimensional label vector

$$\mathbf{y} = [y_1, ..., y_c, ..., y_C], y_c \in \{0, 1\}, C = 8$$

- y_c indicates the presence with respect to according pathology in the image
- A all-zero vector represents the status of “Normal”.
- This definition transits the multi-label classification problem into a regression-like loss setting

Transition Layer

- A variety of CNN networks are adopted and integrated into the proposed framework, e.g. GoogLeNet and ResNet
- Transform the activations from previous layers into a uniform dimension of output $S \times S \times D, S \in \{8, 16, 32\}$
- Pass down the weights from pre-trained DCNN models in a standard form, which is critical for using this layers' activations to further generate the heatmap in pathology localization step

Multi-label Classification Loss Layer

- We first experiment 3 standard loss functions for the regression task instead of using the softmax loss for traditional multi-class classification model, i.e., Hinge Loss (HL), Euclidean Loss (EL) and Cross Entropy Loss (CEL)
- The model has difficulty learning positive because of one-hot-like image labeling strategy (sparse image labels) and the unbalanced numbers of pathology and “Normal” classes.
- Positive/negative balancing, e.g.

$$L_{W\text{-}CEL}(f(\vec{x}), \vec{y}) =$$
$$\beta_P \sum_{y_c=1} -\ln(f(x_c)) + \beta_N \sum_{y_c=0} -\ln(1 - f(x_c)),$$

Global Pooling Layer

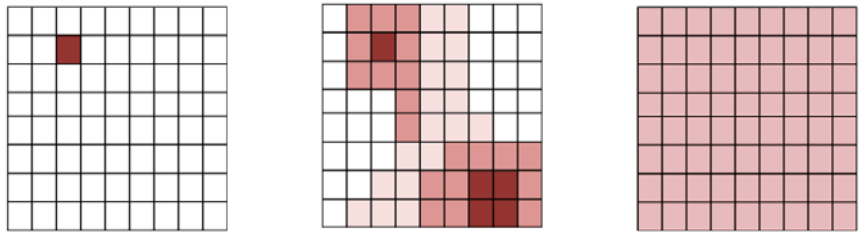
- The pooling layer plays an important role that chooses what information to be passed down.
- Besides the conventional max pooling and average pooling, we also utilize the Log-Sum-Exp (LSE) pooling, which is defined as

$$x_p = \frac{1}{r} \cdot \log \left[\frac{1}{S} \cdot \sum_{(i,j) \in \mathbf{S}} exp(r \cdot x_{ij}) \right]$$

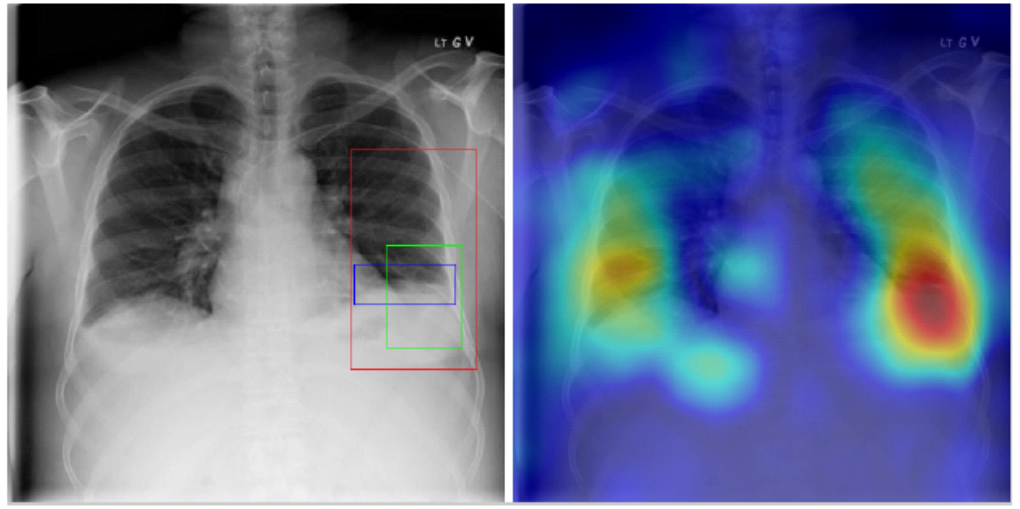
$$x_p = x^* + \frac{1}{r} \cdot \log \left[\frac{1}{S} \cdot \sum_{(i,j) \in \mathbf{S}} exp(r \cdot (x_{ij} - x^*) \right]$$

where $x^* = max\{|x_{ij}|, (i,j) \in \mathbf{S}\}$.

- By controlling the hyper-parameter r, the pooled value ranges from the maximum in S (when r -> 1) to average (r -> 0).

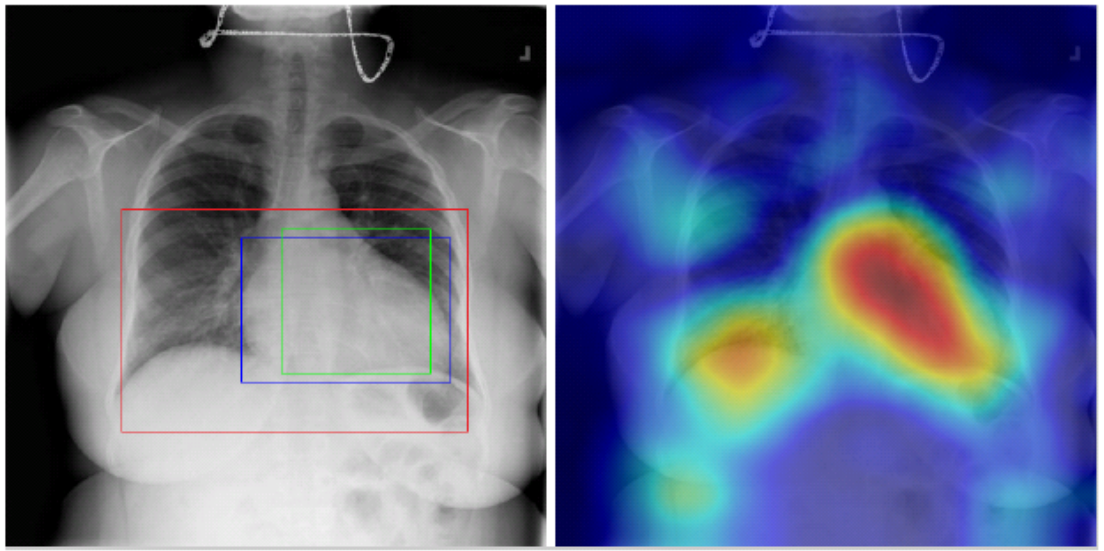


Disease Localization Results - Atelectasis (肺不張)

Radiology report	Keyword	Localization Result
findings include: 1. left basilar atelectasis/consolidation. 2. prominent hilum (mediastinal adenopathy). 3. left pic catheter (tip in atriocaval junction). 4. stable, normal appearing cardiomediastinal silhouette. impression: small right pleural effusion otherwise stable abnormal study including left basilar infiltrate/atelectasis, prominent hilum, and position of left pic catheter (tip atriocaval junction).	Effusion; Infiltration; Atelectasis	

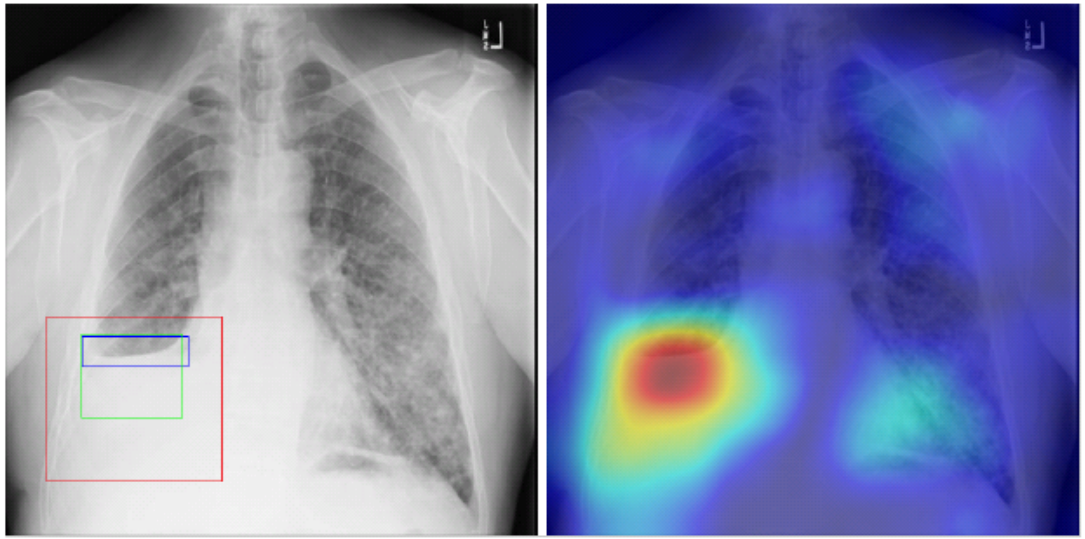
*Correct bounding box (in green), false positives (in red) and the ground truth (in blue)

Disease Localization Results - Cardiomegaly (心臟肥大)

Radiology report	Keyword	Localization Result
findings include: 1. cardiomegaly (ct ratio of 17/30). 2. otherwise normal lungs and mediastinal contours. 3. no evidence of focal bone lesion. dictating	Cardiomegaly	

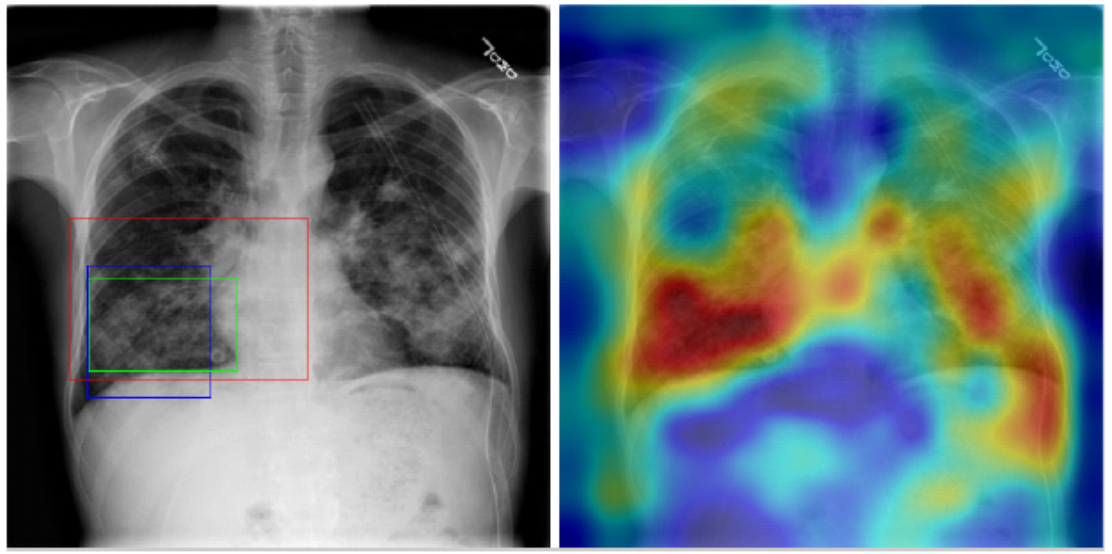
*Correct bounding box (in green), false positives (in red) and the ground truth (in blue)

Disease Localization Results - Effusion (肋膜積水)

Radiology report	Keyword	Localization Result
findings: no appreciable change since XX/XX/XX. small right pleural effusion. elevation right hemidiaphragm. diffuse small nodules throughout the lungs, most numerous in the left mid and lower lung. impression: no change with bilateral small lung metastases.	Effusion; Nodule	

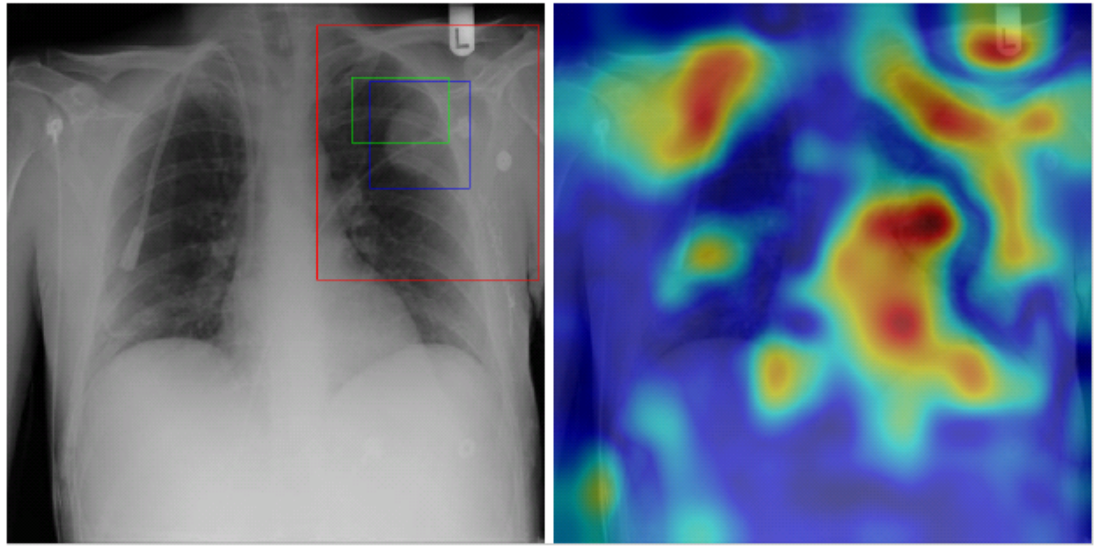
*Correct bounding box (in green), false positives (in red) and the ground truth (in blue)

Disease Localization Results - Infiltration (肺浸潤)

Radiology report	Keyword	Localization Result
findings: port-a-cath reservoir remains in place on the right. chest tube remains in place, tip in the left apex. no pneumothorax. diffuse patchy infiltrates bilaterally are decreasing. impression: infiltrates and effusions decreasing.	Infiltration	

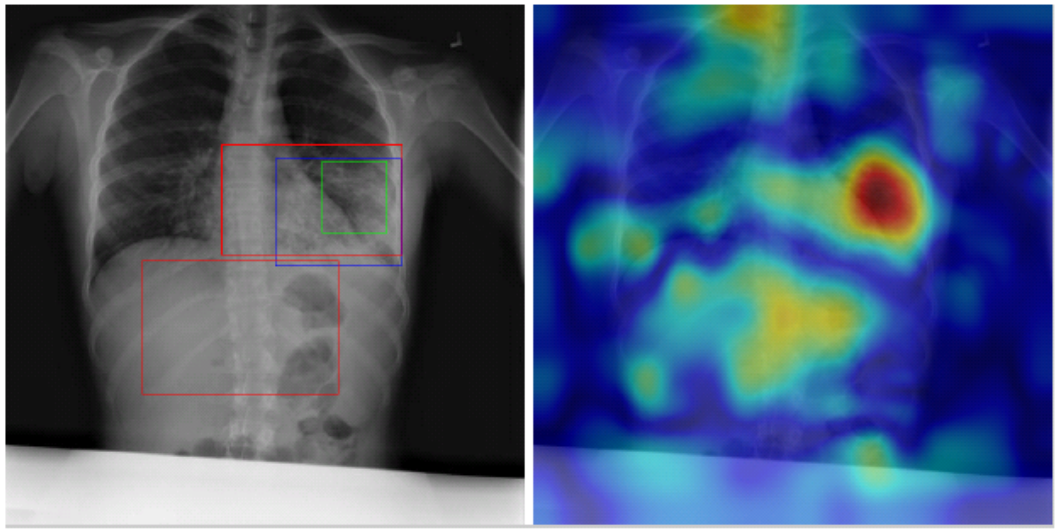
*Correct bounding box (in green), false positives (in red) and the ground truth (in blue)

Disease Localization Results - Mass (肿瘤)

Radiology report	Keyword	Localization Result
findings: right internal jugular catheter remains in place. large metastatic lung mass in the lateral left upper lobe is again noted. no infiltrate or effusion. extensive surgical clips again noted left axilla. impression: no significant change.	Mass	

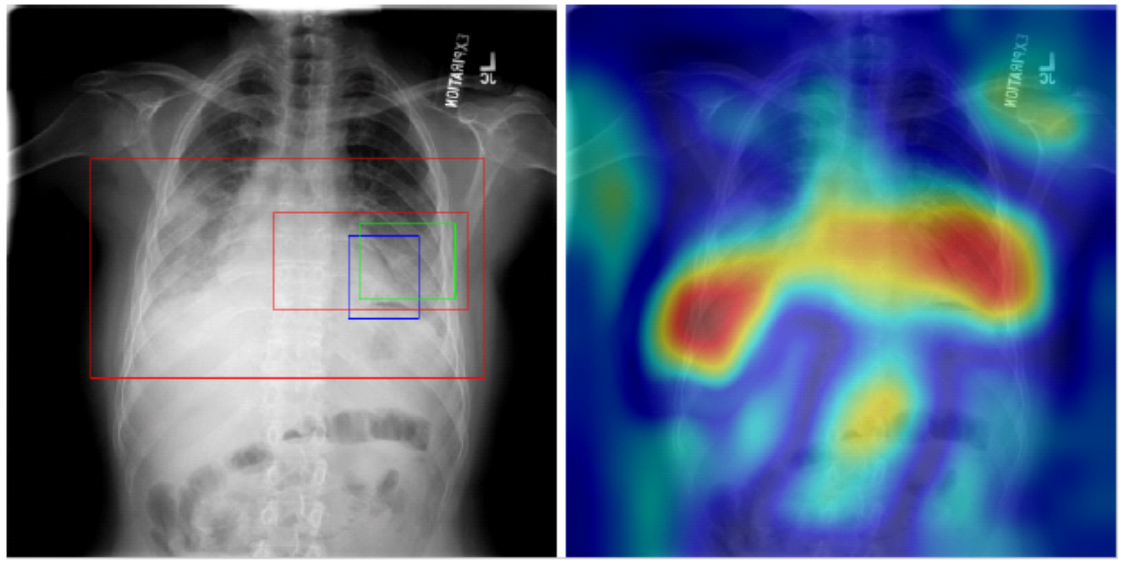
*Correct bounding box (in green), false positives (in red) and the ground truth (in blue)

Disease Localization Results - Pneumonia (肺炎)

Radiology report	Keyword	Localization Result
findings: unchanged left lower lung field infiltrate/air bronchograms. unchanged right perihilar infiltrate with obscuration of the right heart border. no evidence of new infiltrate. no evidence of pneumothorax the cardiac and mediastinal contours are stable. impression: 1. no evidence pneumothorax. 2. unchanged left lower lobe and left lingular consolidation/bronchiectasis. 3. unchanged right middle lobe infiltrate	Pneumonia; Infiltration	

*Correct bounding box (in green), false positives (in red) and the ground truth (in blue)

Disease Localization Results - Pneumothorax (肺氣腫)

Radiology report	Keyword	Localization Result
findings: frontal lateral chest x-ray performed in expiration. left apical pneumothorax visible. small pneumothorax visible along the left heart border and left hemidiaphragm. pleural thickening, mass right chest. the mediastinum cannot be evaluated in the expiration. bony structures intact. impression: left post biopsy pneumothorax.	Mass; Pneumothorax	

*Correct bounding box (in green), false positives (in red) and the ground truth (in blue)

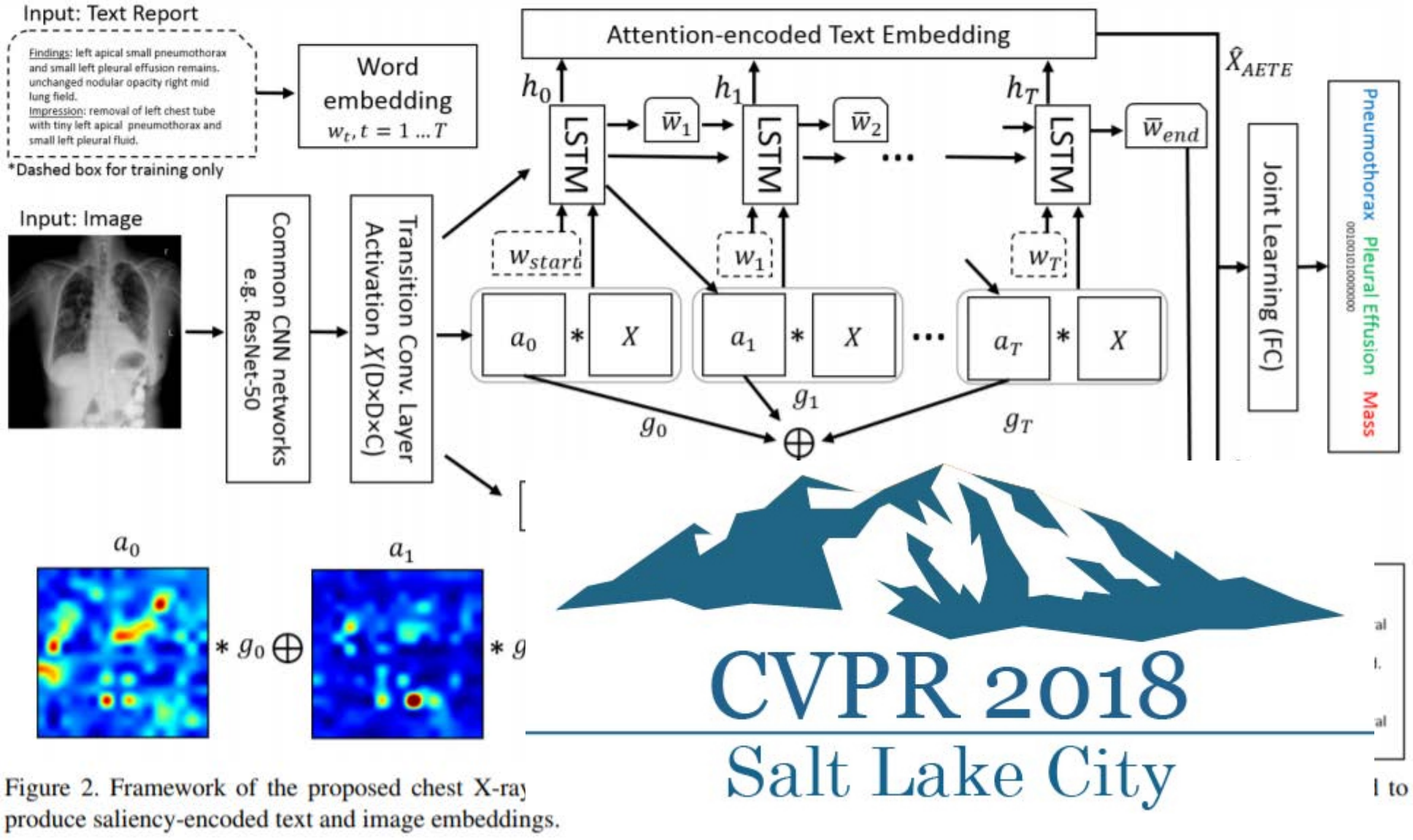
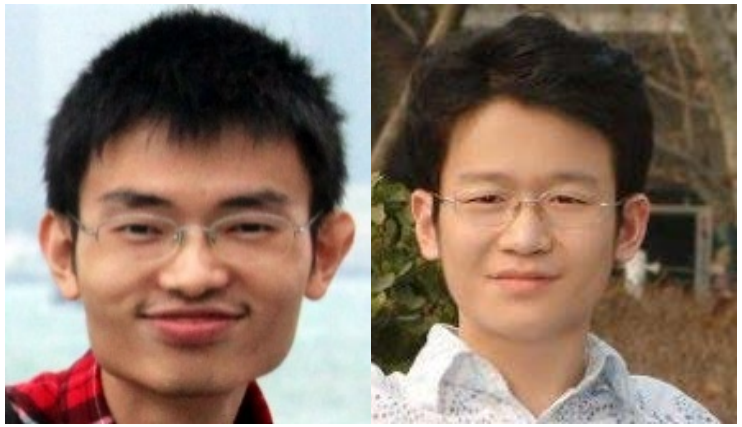


Figure 2. Framework of the proposed chest X-ray ... produce saliency-encoded text and image embeddings.

CVPR 2018
Salt Lake City

TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays, (Spotlight), CVPR 2018

Deep Lesion Graphs in the "Wild": Relationship Learning and Organization of Significant Radiology Image Findings in a Diverse Large-scale Lesion Database



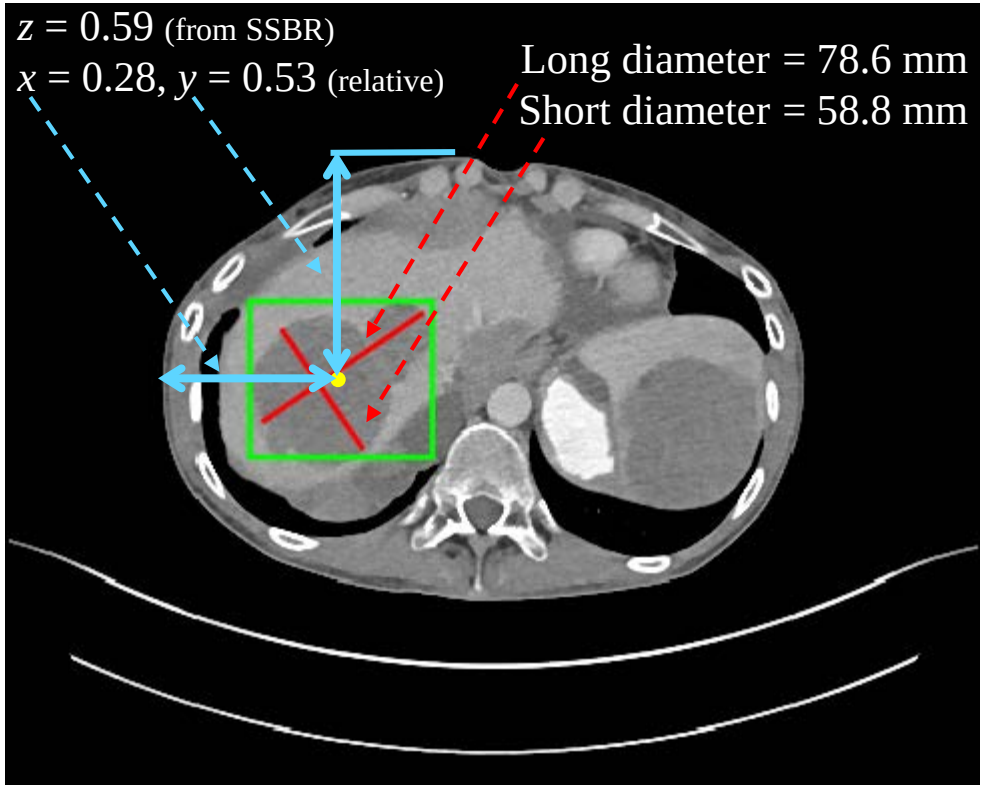
CVPR 2018
Salt Lake City

Background

- Large-scale datasets with diverse images and dense annotations are important for both computer vision and medical image
  - Crowd-sourcing can be used in to annotate computer vision datasets, but Medical Imaging requires considerate specialized knowledge & training
  - Mining Internet Images via Deep Learning can be used in computer vision to acquire self-annotations;
- Fortunately, like web data in computer vision, a vast amount of **loosely-labeled and largely untapped data source** does exist in the form of Picture Archiving and Communication Systems (PACS/RIS).
  - Similarly, can we mine the "unstructured but extremely informative" PACS?

Background

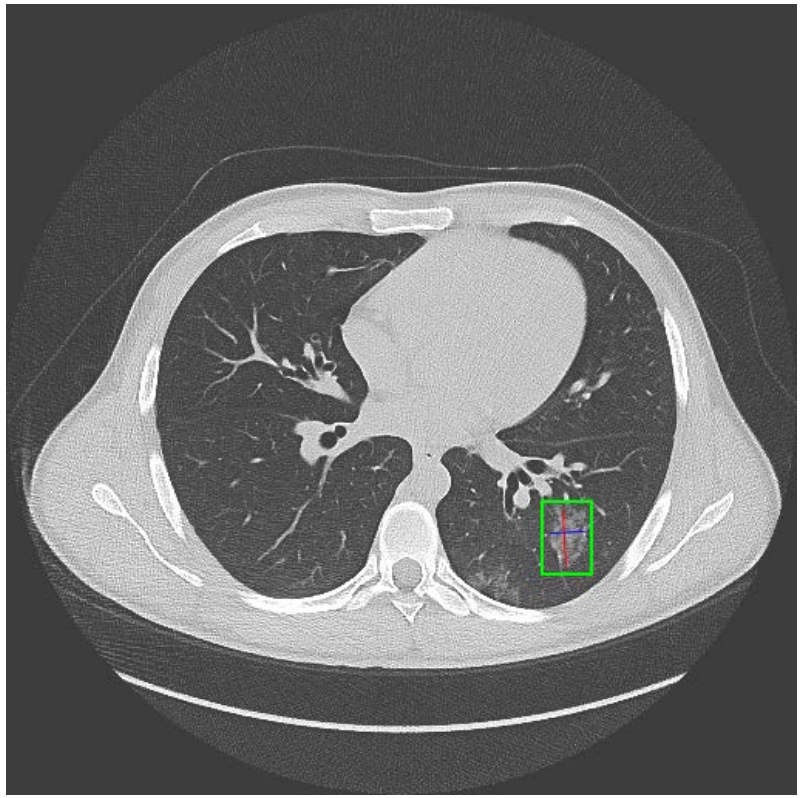
- Radiologists in their daily work may routinely mark and measure some significant abnormalities or “lesions” on radiology images
  - Collected over years and stored in hospitals’ PACS/RIS
  - Sometime known as “bookmarks”
  - Used to assess patients’ conditions or therapy responses



To the best of our knowledge, no prior work has been done on **learning “deep lesion” similarity graph & embedding** via imaging on such a “**large-scale**” comprehensive dataset with “**weak**” cues!

"DeepLesion" Dataset

- Mined from bookmarks (RECIST diameters) in NIH's PACS
- 32,120 axial CT slices from 10,594 studies of 4,427 unique patients
- 1–3 lesions in each image with size measurements (long-axis and short axis)
- 32,735 lesions altogether
- Convert any pair of diameters to a bounding-boxes
  - $(x_{min} - 5, y_{min} - 5, x_{max} + 5, y_{max} + 5)$
  - $x_{min} = \min(x_{11}, x_{12}, x_{21}, x_{22})$,
  - $x_{max} = \max(x_{11}, x_{12}, x_{21}, x_{22})$



Problem Definition

- Lesions in DeepLesion are basically unsorted and lack semantic labels, *e.g.*, lung nodule, mediastinal lymph node
- Our goal: understand and organize a large quantity of lesions, or oncology findings, by "**Automated Instance-level Similarity Modeling & Topology-discovery Mining**"
  1. discover their types and locations
  2. find similar lesions from a population of different patients, *i.e.*, content-based retrieval
  3. track the same lesions within the same patient's several longitudinal studies, *i.e.*, lesion instance matching or tracking among multiple studies

  https://www.healthdatamanagement.com/news/oncology-ai-runs-on-the-arterys-mica-medical-imaging-cloud-ai-platform
- Our approach: learning **"functional"** deep feature representations for each instance that capture & keep the similarity relationship in type, location, and size, etc.
- One of the early **scientifically-viable, practical** approaches on doing such AI/DL medical imaging task at scale! → precision imaging measurements for oncology …

Related Work

- **Deep Metric Learning**
  - **Siamese network** [1]
  - **Triplet network** – "weak deep learning, or DL with weak data self-regularization!"

$$\|f(A) - f(P)\|_2^2 + m < \|f(A) - f(N)\|_2^2$$

  - **Triplet Loss** with multiple labels with hierarchical structures [2]
    - Our method shares the similar spirit with [2], but we **lack** well-defined supervision cues (to define which pair is more similar than the other, and logically why?) given the significant radiology findings in the collected dataset
    - We proposed strategies to propose and leverage **weak cues**, e.g., self-supervised body part regressor and iterative refinement, etc.

[1] J. Bromley et al. Signature verification using a "siamese" time delay neural network, NIPS 1994
[2] X. Zhang et al. Embedding label structures for fine-grained feature representation, CVPR 2016

Related Work

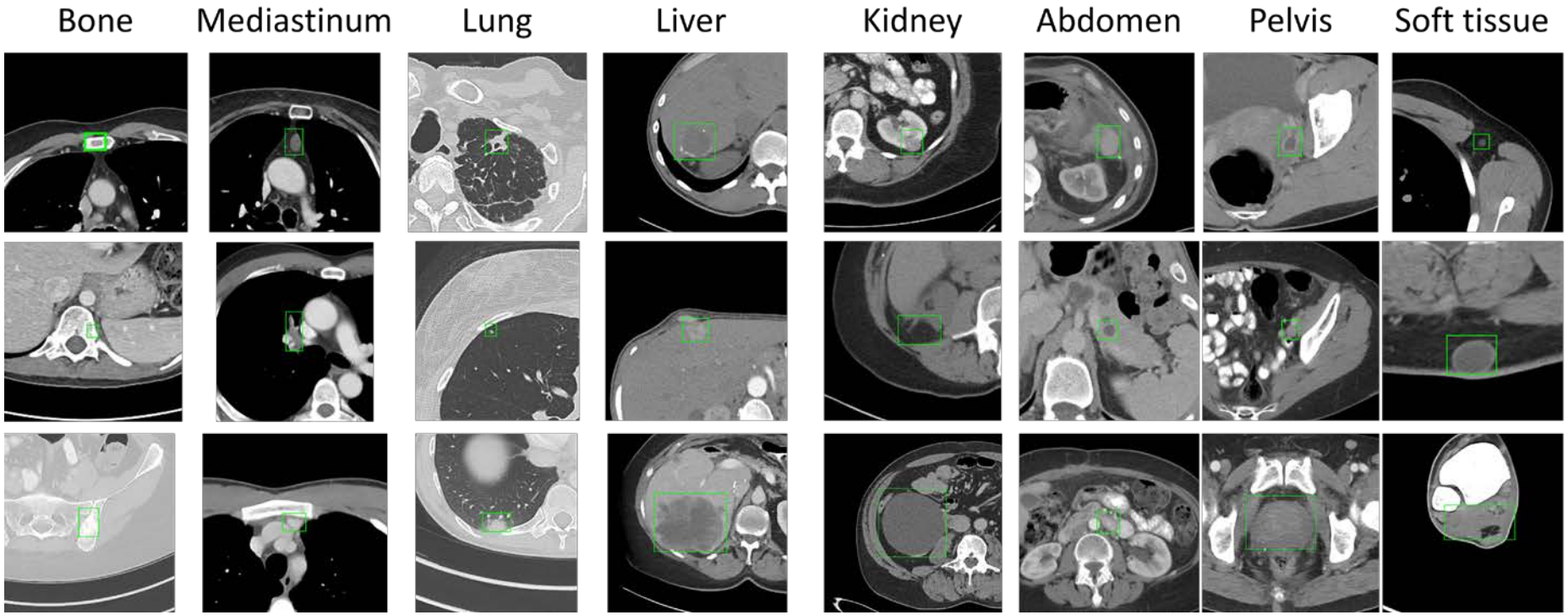
- Lesion Retrieval (**inter-patients**)
  - Existing methods typically focus on **one type of lesion** (*e.g.*, lung lesion or mammographic mass) and learn the similarity relationship based on **fully manually-annotated labels** (or radiology reports)
  - We learn deep lesion embedding on a large diverse dataset with **weak cues**, in a **self-regularization** manner when "**big data**" are available!
- Lesion Matching (across **intra-patient** multiple time points)
  - Existing work generally require organ segmentation or time-consuming non-rigid volumetric registration and focus on certain lesion types
  - Our lesion embedding is fast and can match **all "categories" of lesions**

Liu, Lu, Ye, Yu, Huang: Coarse-to-fine classification via parametric and **nonparametric models for computer-aided diagnosis**. ACM CIKM 2011

Bi, Wu, Lu, Liu, Tao, Wolf: AdaBoost on **low-rank PSD matrices for metric learning**. IEEE CVPR 2011

Supervision Cue (I): Lesion Type

- We randomly select 30% lesions and manually label them into 8 types: lung, abdomen, mediastinum, liver, pelvis, soft tissue, kidney, and bone
  - Coarse-scale attributes of the lesions



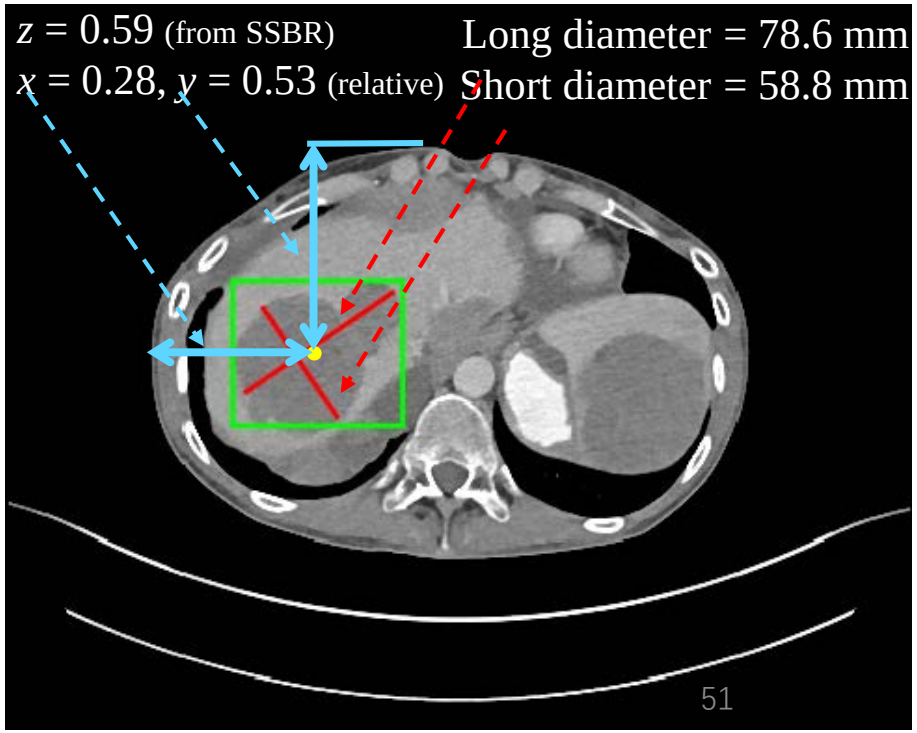
Supervision Cue (I): Lesion Type

- **Mediastinum**: mainly consists of lymph nodes (LNs) in the chest
- **Abdomen**: miscellaneous ones that are not in liver or kidney
- **Soft tissue**: lesions and LNs in the muscle, skin, fat, etc.

- Among the labeled samples, we randomly select 25% as training seeds to predict pseudo-labels, 25% as the validation set, and the other 50% as the test set
- We use labeled seed samples to train a classifier (RBF SVM) on ImageNet pre-trained deep image features and apply it to all unlabeled samples to get their pseudo-labels (may be noisy!)

Supervision Cue (II): Relative Body Location

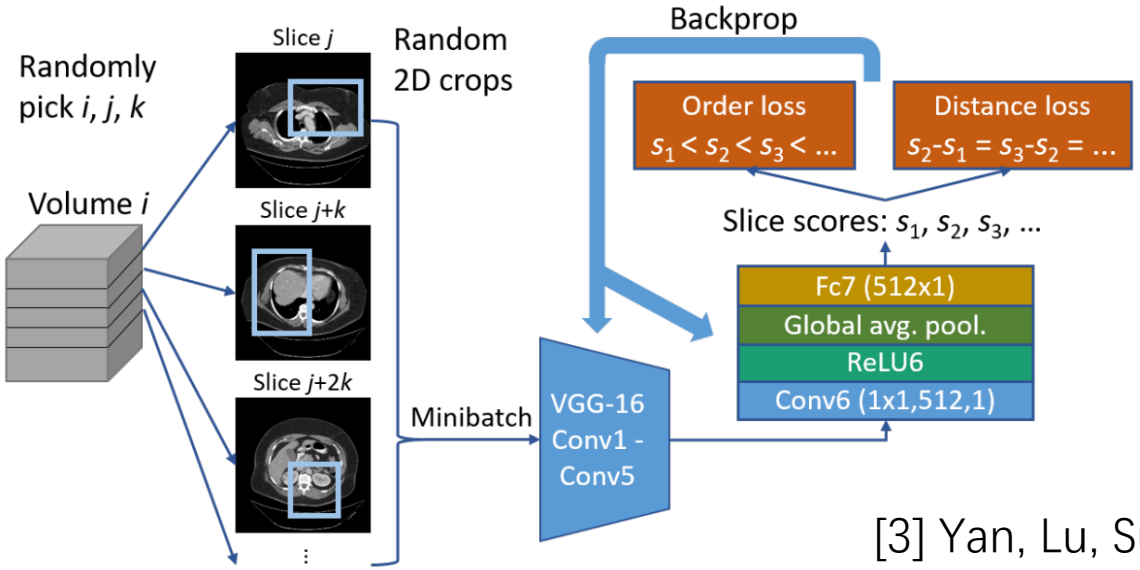
- *X* and *Y*: easy ☺
- *Z*: self-supervised body part regression or regressor (SSBR) [3]
- SSBR [3]
  - Intuition: volumetric medical images are intrinsically Upper/Down structured or ordered!
  - The superior-inferior slice order information (self-supervision) can be leveraged to learn an deep appearance-based *z predictor*



[3] Yan, Lu, Summers. Unsupervised Body Part Regression via Spatially Self-ordering Convolutional Neural Networks, ISBI 2018

Supervision Cue (II): Relative Body Location

- *h* is the sigmoid function, *g* is the smooth L1 loss
- The order loss and distance loss terms collaborate to push each slice score towards the correct direction relative to other slices



$$L_{\text{SSBR}} = L_{\text{order}} + L_{\text{dist}};$$

$$L_{\text{order}} = -\sum_{i=0}^{m-2} \log h\left(s_{j+k(i+1)} - s_{j+ki}\right);$$

$$L_{\text{dist}} = \sum_{i=0}^{m-3} g(\Delta_{i+1} - \Delta_i),$$

$$\Delta_i = s_{j+k(i+1)} - s_{j+ki},$$

[3] Yan, Lu, Summers. Unsupervised Body Part Regression via Spatially Self-ordering Convolutional Neural Networks, ISBI 2018

Figure 2. *Framework of the self-supervised body part regressor (SSBR).*

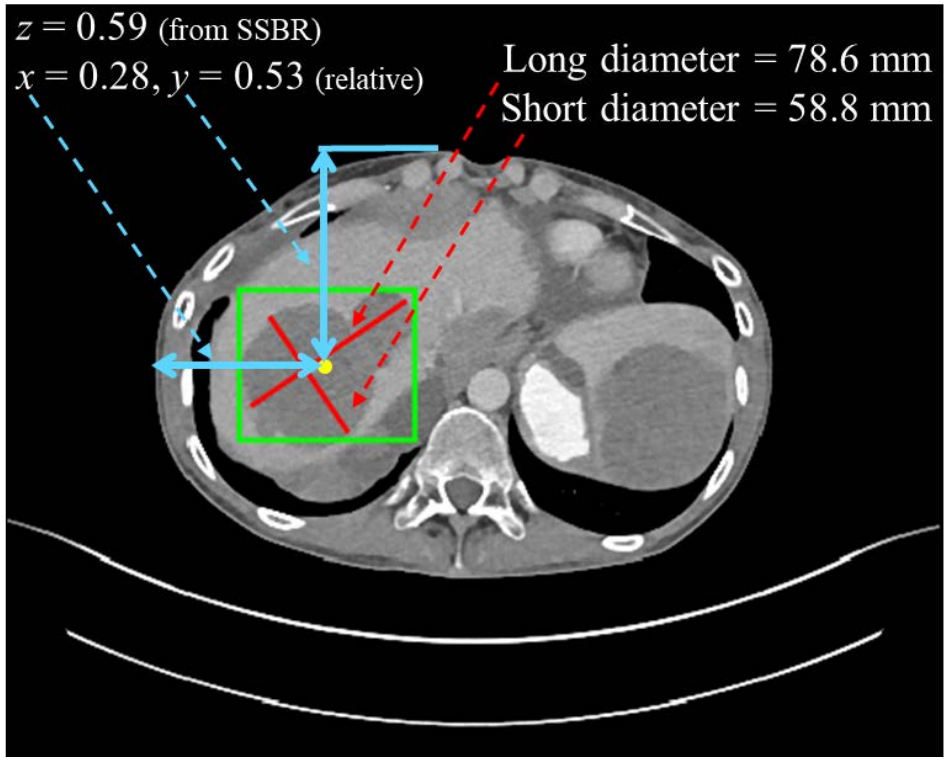
Supervision Cue (II): Relative Body Location

- In DeepLesion, some CT volumes are zoomed in on a portion of the body, *e.g.*, only the left half is shown
  - Technique: train SSBR on random crops of the axial slices

- **Data Augmentation**: SSBR does not perform well on rare body parts that are much less frequent in the training set, *e.g.*, head and legs → inconsistence!
  - Technique: train SSBR → examine the correlation coefficients (r) of slice scores and slice indices to find rare parts → train SSBR again on a resampled training set with hard volumes oversampled

Supervision Cue (III): Lesion Size

- Lengths of long and short axes of lesion diameters
- Has already been annotated and measured by radiologists
- Ranges from 0.2 to 343 mm with a median of 15.6 mm



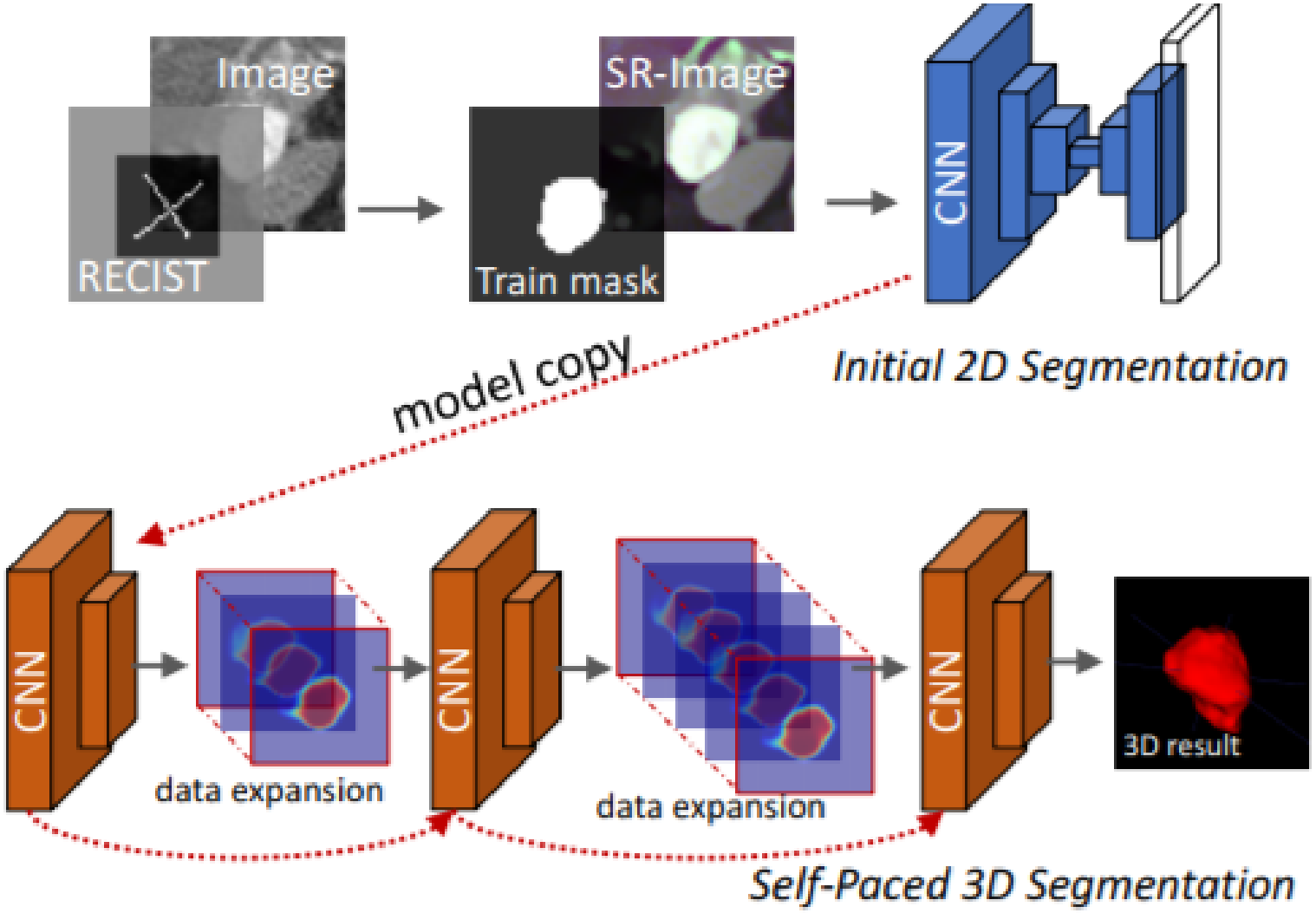
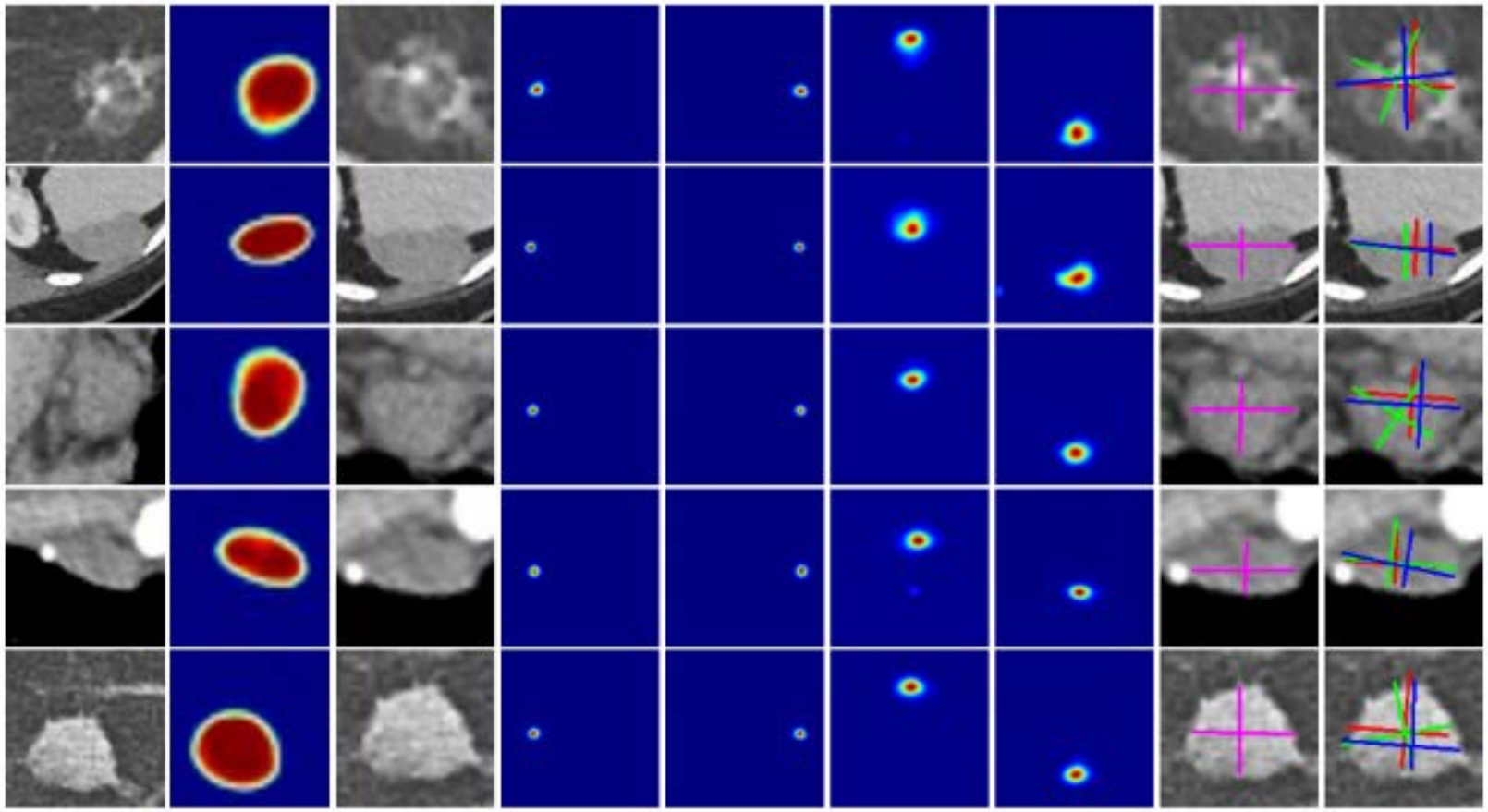


Figure 1. *Overview of the proposed weakly supervised self-paced segmentation with CNN (Sec.3) for 3D lesion segmentation.*

J. Cai, et al., "**Accurate Weakly-Supervised Deep Lesion Segmentation using Large-Scale Clinical Annotations: Slice- Propagated 3D Mask Generation from 2D RECIST**", MICCAI, 2018, arXiv:1801.08614,



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 3. Given the input test image (a), we can obtain the predicted lesion mask (b), the transformed image (c) from the STN, and the estimated keypoint heatmaps (d)-(g) from the SHN. From (d)-(g), we obtain the estimated RECIST (h), which is close to the annotations (i) labeled by radiologists. Red, green, and blue marks denote DL, R1,

Y. Tang, et al., "**Accurate Semi-Automatic RECIST Labeling on CT Scans with Cascaded Convolutional Neural Networks**", MICCAI, 2018,

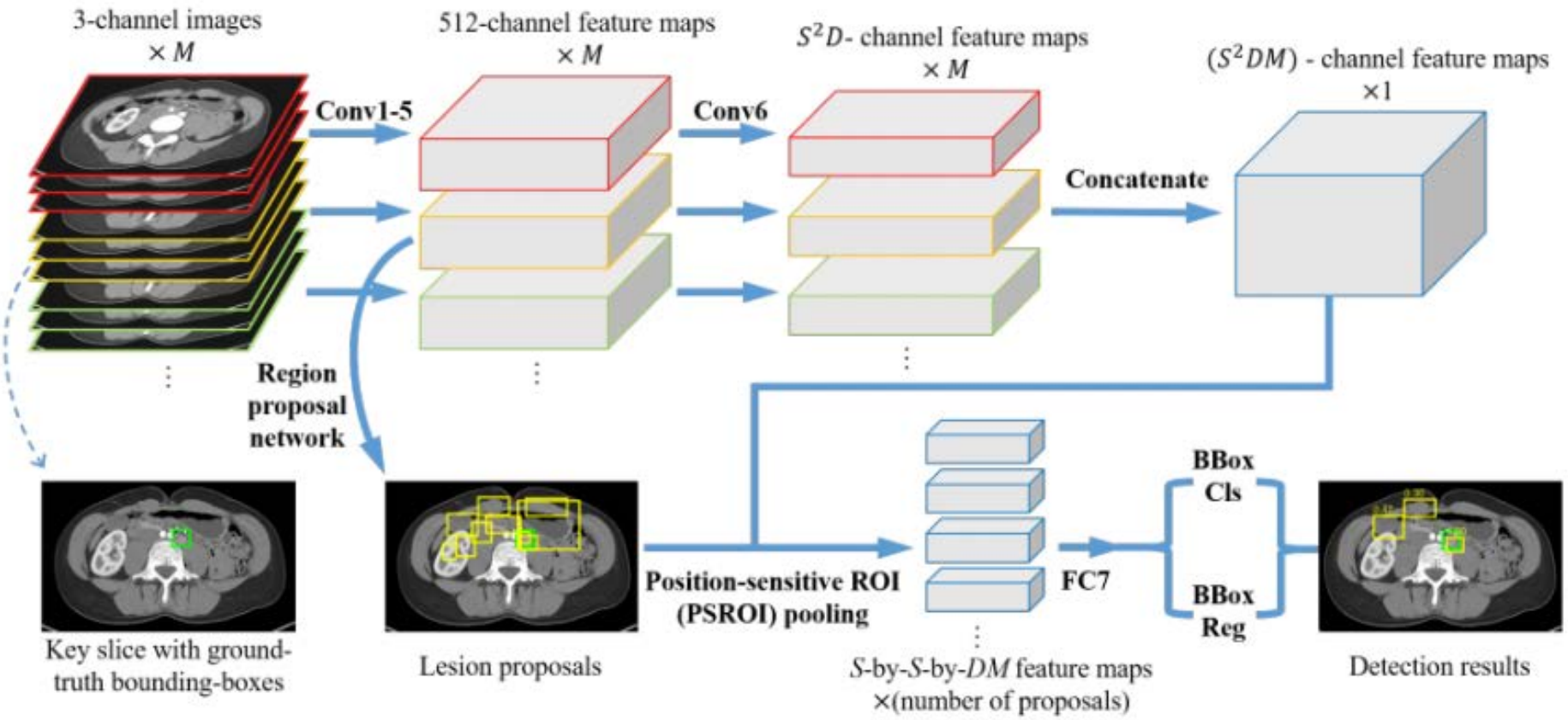
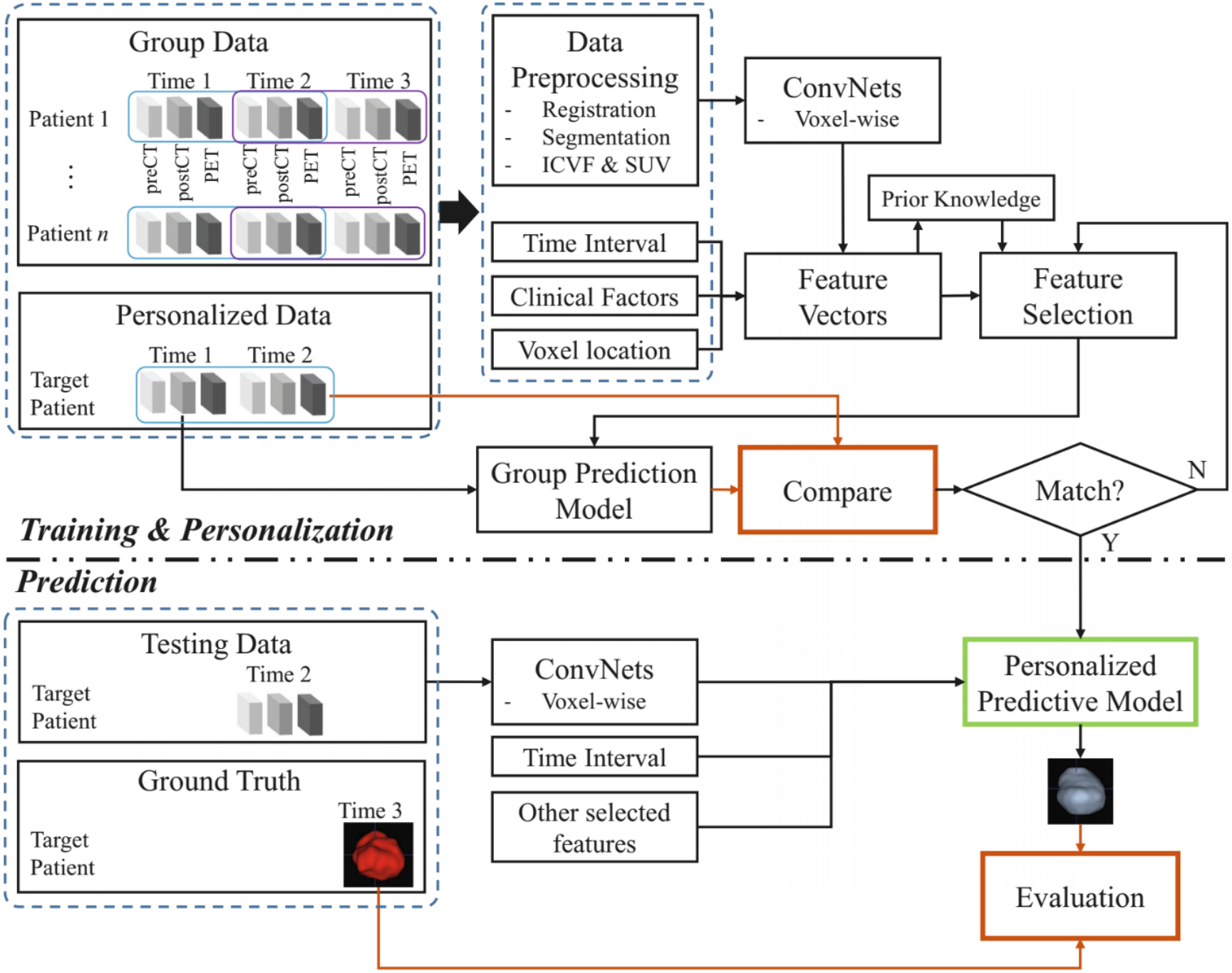
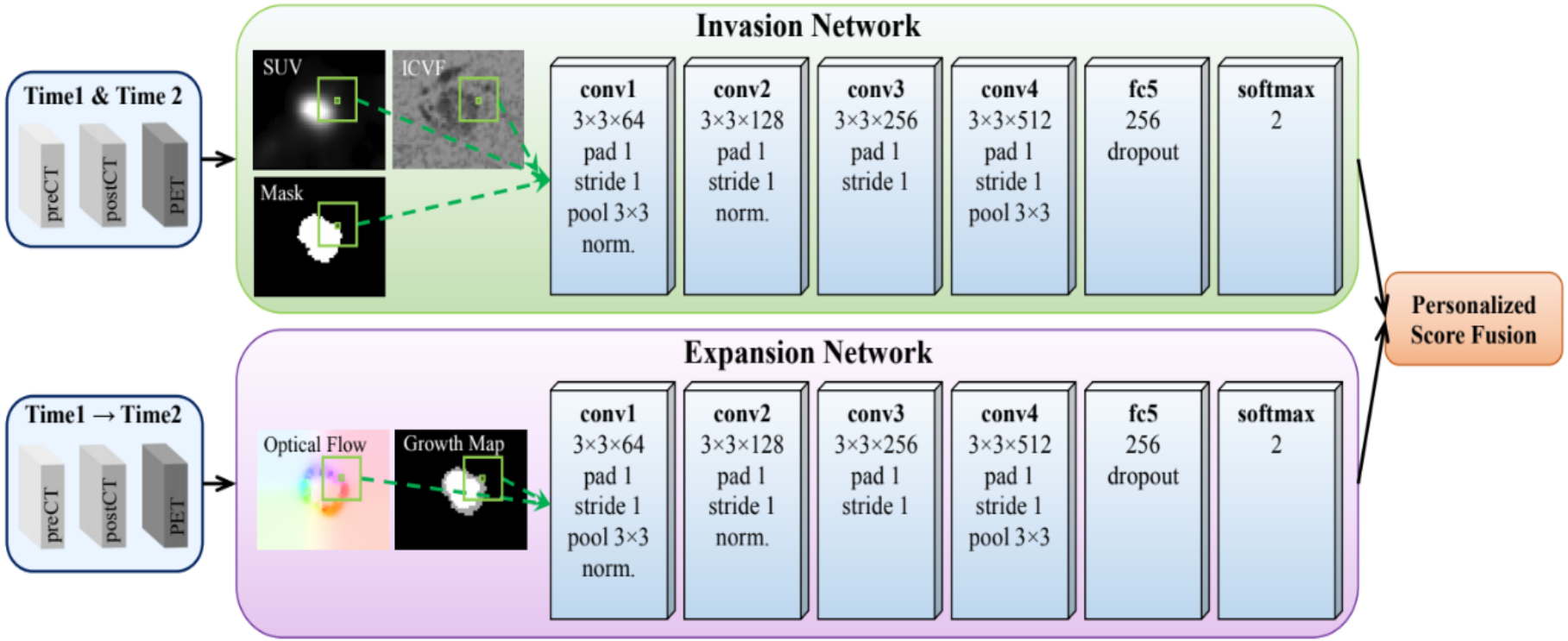


Fig. 1. The framework of 3D context enhanced region-based CNN (3DCE) for lesion detection.

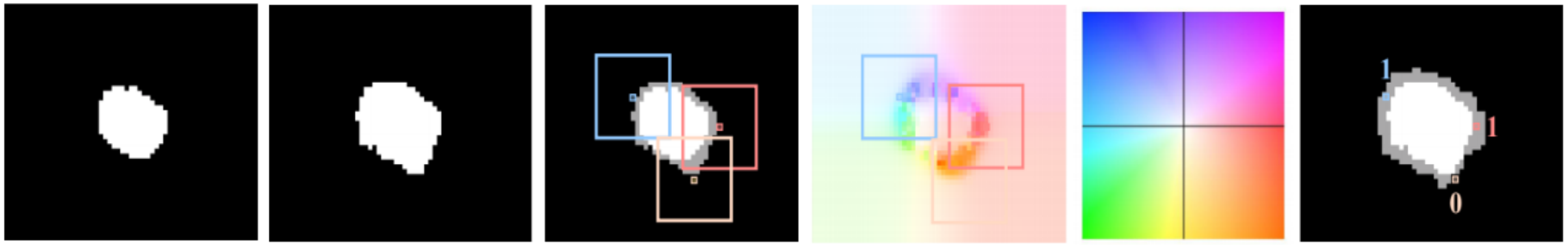
K. Yan, et al., "**3D Context Enhanced Region-based Convolutional Neural Network for End-to-End Lesion Detection**", MICCAI, 2018,



Personalized Pancreatic Tumor Growth Prediction via Group Learning, MICCAI 2017



architecture for late fusion of the invasion and expansion networks for predicting tumor growth.



(a) Tumor mask at time 1 (b) Tumor mask at time 2 (c) Tumor growth map (time1 → time2) (d) 3-channel optical flow (time1 → time2) (e) Flow field color coding (f) Tumor growth map (time2 → time3)

Convolutional **Invasion** and **Expansion** Networks for Tumor Growth Prediction, IEEE Trans. Medical Imaging, 2018

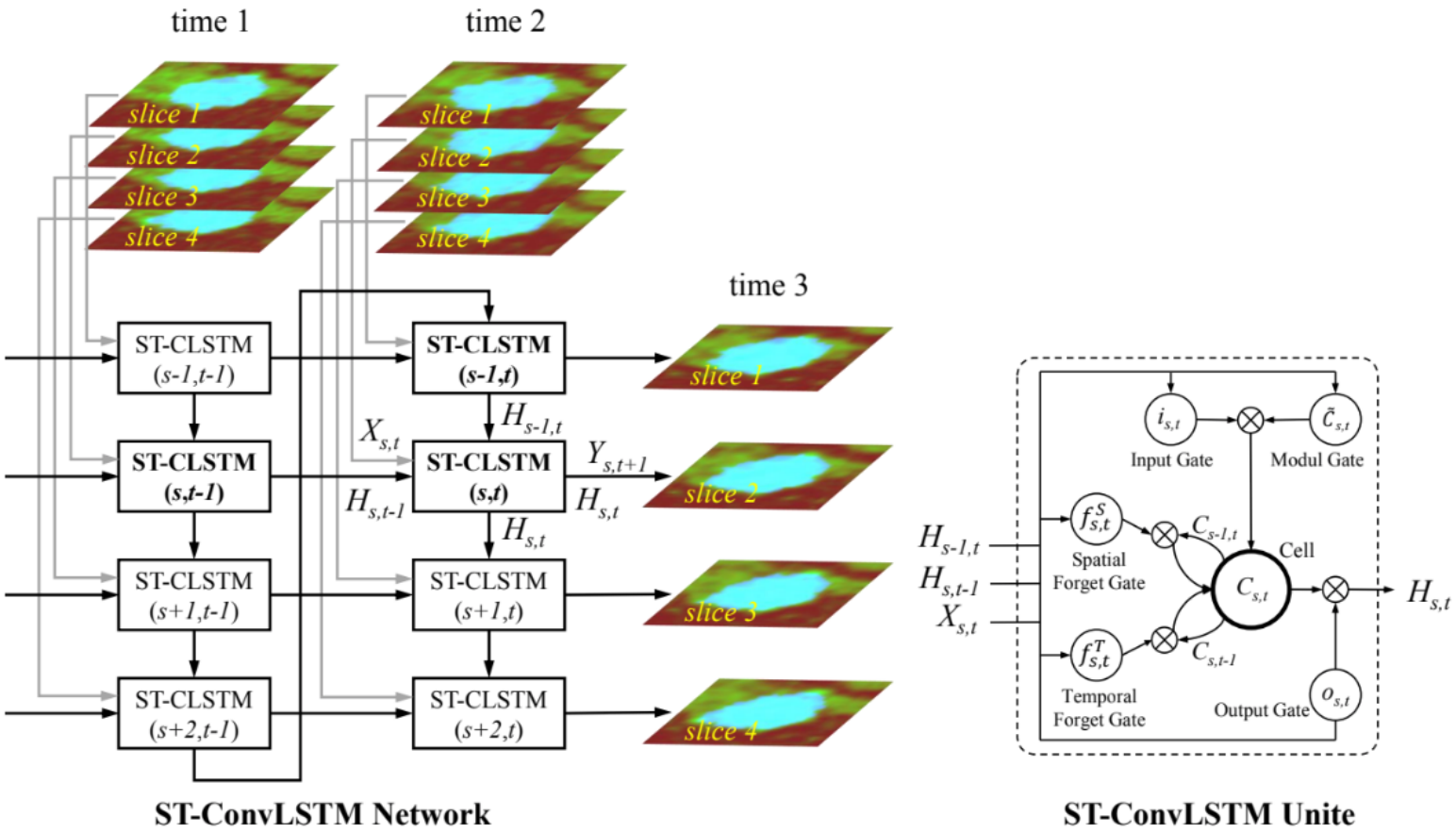


Fig. 1. Left: The proposed Spatial-Temporal Convolutional LSTM (ST-ConvLSTM, or ST-CLSTM) network for learning 4D longitudinal data to predict tumor growth. In this example, 3 time points each with 4 spatially adjacent image slices (each slice is a 3-channel color image) are shown. **Right:** The ST-ConvLSTM unit.

Spatial-Temporal Convolutional LSTMs: Learning 4D Longitudinal Data for Tumor Growth Prediction, In Submission, 2018

Build Triplet Network with Sequential Sampling

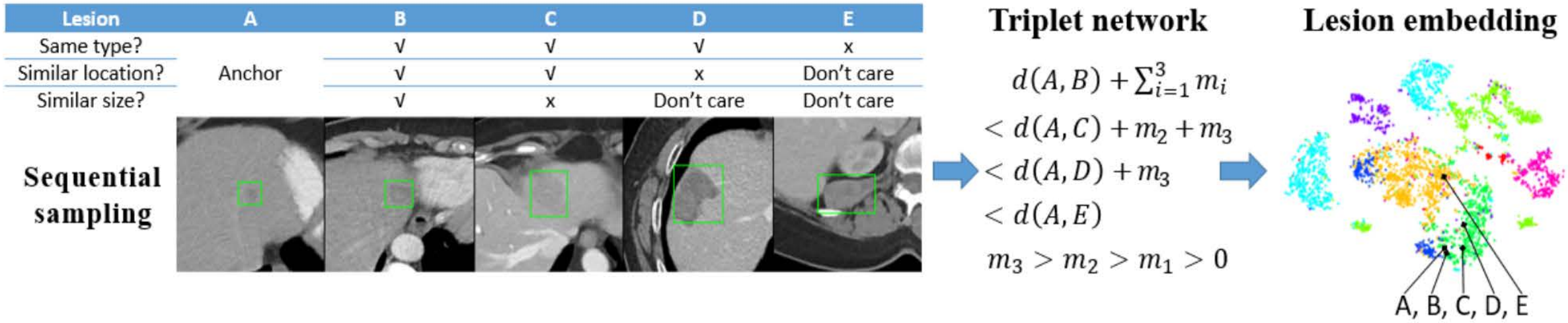


Figure 1. *The proposed framework. Using a triplet network, we learn a feature embedding for each lesion in our comprehensive DeepLesion dataset. Training samples A–E are selected with a sequential sampling strategy so as to make the embeddings respects similarity in type, location, and size.*

- With a flavor of "**Gibbs Sampling**"; sampling similarity orders according to each **marginal/conditional distribution** (e.g., body location similarity), iteratively → to comply similarity scores into the **joint distribution**
- **"Cue Priorities"**, **not all cues are created equally!** → DL or DNN is "paradigm shifting" or "super component"?
- **Ranking** a series of samples to make them **self-organize and move to** the right space in the feature space!

Triplet network with Sequential Sampling

- Do not use hard triplet mining because of the noisy cues
- Note that there is label noise in the 4th row of Fig. 3, where lesion *D* does not have the same type with *A* ~ *C* (soft tissue versus pelvis)

$$\|f(A) - f(P)\|_2^2 + m < \|f(A) - f(N)\|_2^2$$

where an anchor A, a positive sample P with the same label as A, and a negative sample N with a different label. f(.) is the embedding function to be learned and m is a predefined margin.

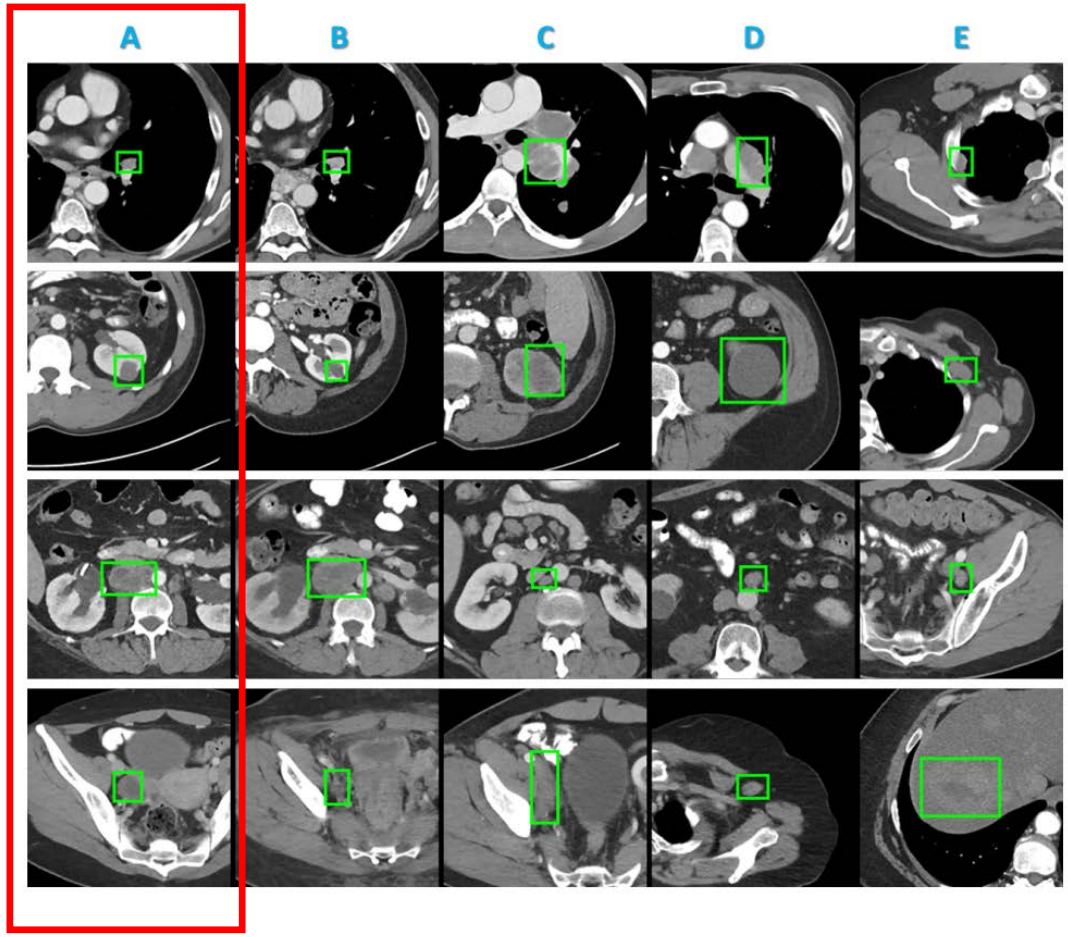


Figure 3. *Sample training sequences. Each row is a sequence. Columns 1–5 are examples of lesions A–E in Fig. 1, respectively.*

Triplet Network with Sequential Sampling

- Joint Loss function
  - A selected sequence of 5 instances can be decomposed into three triplets: {*ABC*, *ACD* and *ADE*} ; Joint Loss →

$$L = \frac{1}{2S}\sum_{i=1}^{S}\left[\max(0, d_{AB}^2 - d_{AC}^2 + m_1) + \max(0, d_{AC}^2 - d_{AD}^2 + m_2) + \max(0, d_{AD}^2 - d_{AE}^2 + m_3)\right]$$

$$m_3 > m_2 > m_1 > 0$$

- Iterative refinement learning
  - With the learned similarity embedding, we can retrain the lesion type classifier to get "cleaner" pseudo-labels (using deep embedding features), then fine-tune the triplet network with a lower learning rate

Network Architecture

- Backbone: VGG-16
- Multi-scale, multi-crop
- Output: a deep **1408D** feature embedding vector for each lesion instance (@various of image sizes/dimensions, @1 mm/pixel)

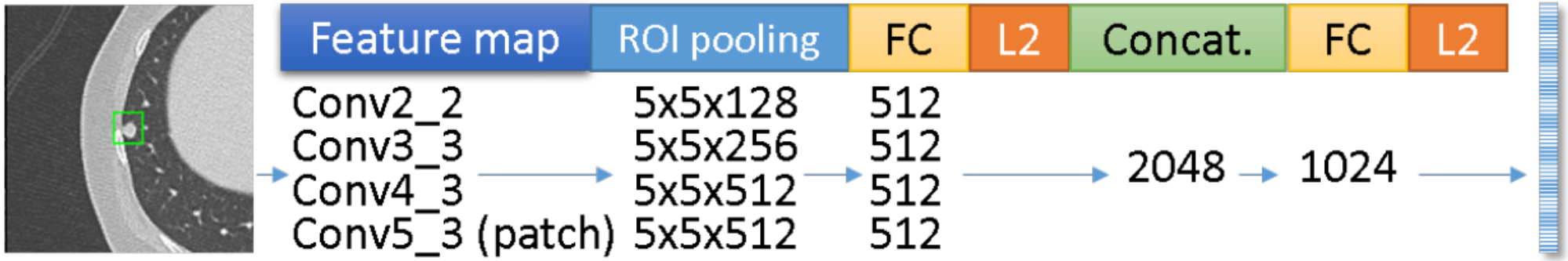


Figure 4. *Network architecture of the proposed triplet network.*

Lesion Organization: Retrieval & Matching

- **Inter-patient** content-based **lesion retrieval**: finding nearest neighbors of query lesions → data-learned similarity score permitting!
- **Intra-patient lesion matching**: **graph-based edge pruning**

Algorithm 1 *Intra-patient lesion matching*

Input: Lesions of the same patient represented by their embeddings; the study index s of each lesion; intra-study threshold T_1; inter-study threshold T_2.

Output: Matched lesion groups.

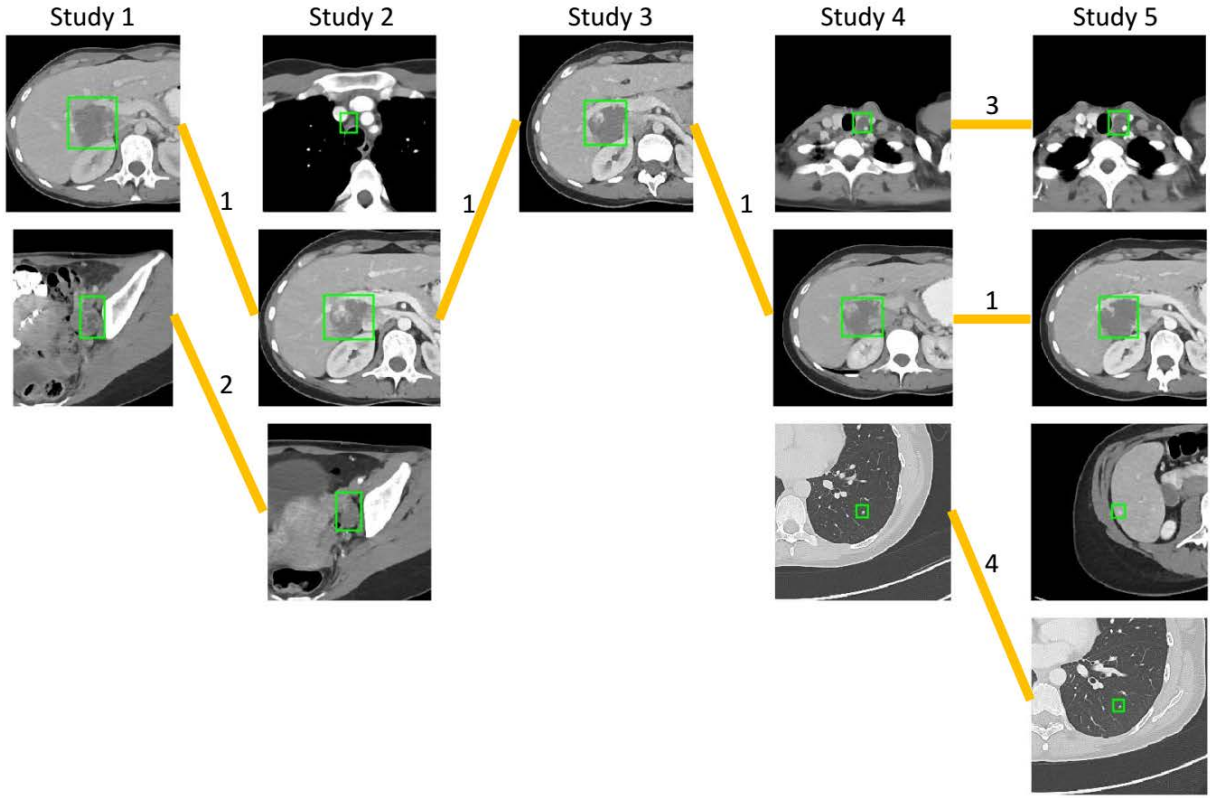
1: Compute an intra-patient lesion graph $G = \langle V, \mathcal{E} \rangle$, where V are nodes (lesions) and $\mathcal{E}$ are edges. Denote d_{ij} as the Euclidean distance between nodes i, j.

2: **Merge** nodes i and j if $s_i = s_j$ and $d_{ij} < T_1$.

3: **Threshold**: $\mathcal{E} \leftarrow \mathcal{E} - \mathcal{D}, \mathcal{D} = \{\langle i, j \rangle \in \mathcal{E} | d_{ij} > T_2\}$.

4: **Exclusion**: $\mathcal{E} \leftarrow \mathcal{E} - \mathcal{C}, \mathcal{C} = \{\langle i, j \rangle \,|\, \langle i, j \rangle \in \mathcal{E}, \langle i, k \rangle \in \mathcal{E}, s_j = s_k, \text{and } d_{ij} \geq d_{ik}\}$.

5: **Extraction**: Each node group with edge connections is considered as a matched lesion group.



Implementation Details: Image Preprocessing

- Rescale image intensity to floating-point numbers in [0, 255] using a single windowing (-1024 ~ 3071 HU)
- Resize spacing to 1 mm/pixel
- Crop a patch with 50 mm padding around each lesion's bounding-box
- Use 3 neighboring slices (interpolated at 2 mm inter-slice intervals) to encode 3D information
- No data augmentation was used

Implementation Details: Training Schemes

- The maximum value of each dimension of the locations and sizes is normalized to 1

$$m_1 = 0.1, m_2 = 0.2, m_3 = 0.4$$

- 24 five-instance sequences per mini-batch
- SGD with a learning rate of 0.002, reduced to 0.0002 at iteration 2K#, converges in 3K iterations
- To train SSBR, we used 800 random unlabeled CT volumes of 420 subjects from *DeepLesion* patient population

Experiments

- Visualization of DeepLesion: projecting lesion densely-connected hyper-graph into a 2D map (*t-SNE*) !
- *X*- and *Y*-axes of the scatter map correspond to the *X*- and *Z*-coordinates of the relative body location of each lesion → organized by lesion type, then location!
- Illustration of the distribution diversity of *DeepLesion* dataset (https://arxiv.org/abs/1710.01766)

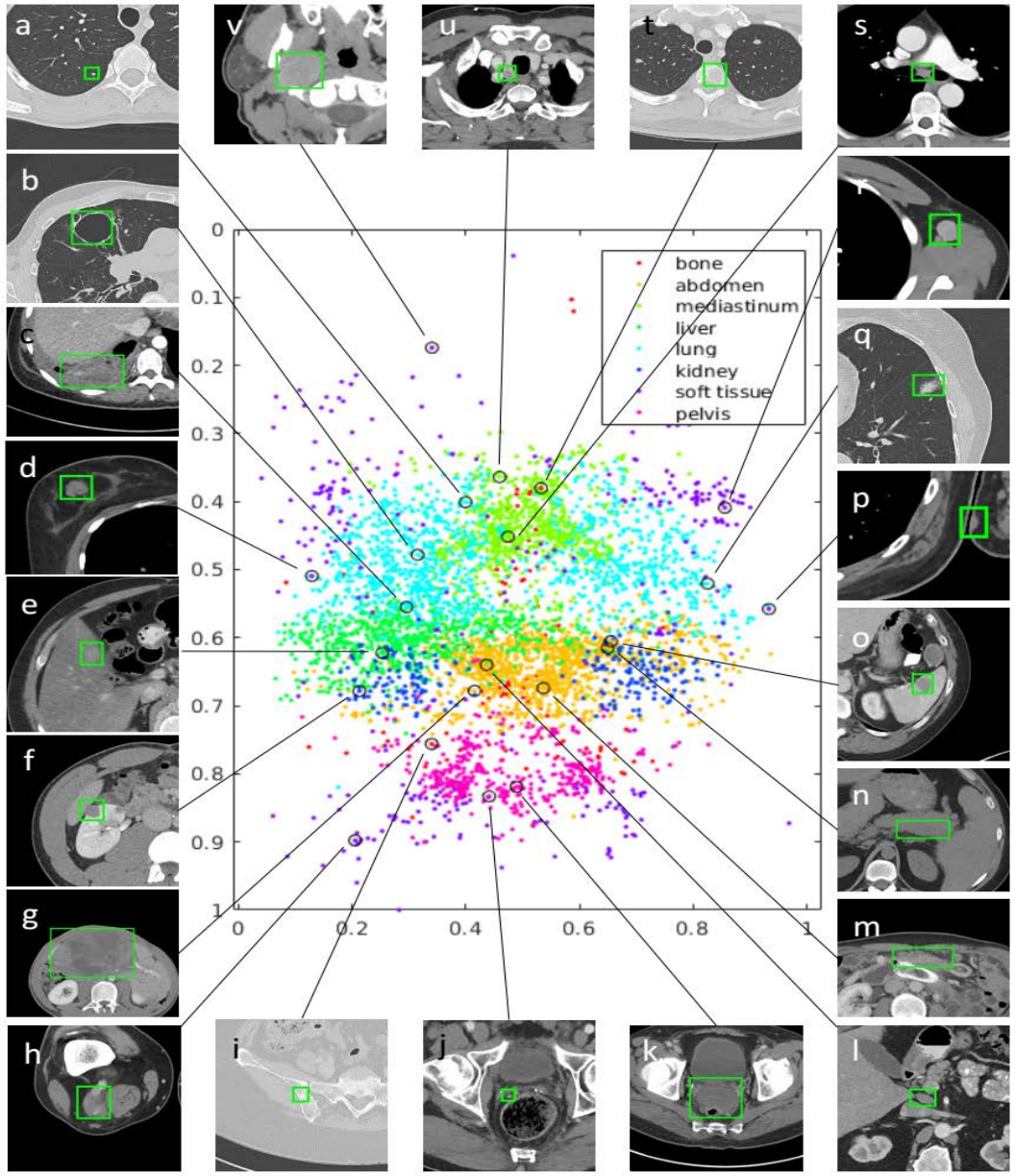


Figure 3. Visualization of the DeepLesion dataset (test set). The x- and y-axes of the scatter map correspond to the x- and z-coordinates of the relative body location of each lesion, respectively. Therefore, this map is similar to a frontal view of the human body. Colors indicate the manually labeled lesion types. Sample lesions are exhibited to show the great diversity of DeepLesion, including: a. lung nodule; b. lung cyst; c. costophrenic sulcus (lung) mass/fluid; d. breast mass; e. liver lesion; f. renal mass; g. large abdominal mass; h. posterior thigh mass; i. iliac sclerotic lesion; j. perirectal lymph node (LN); k. pelvic mass; l. periportal LN; m. omental mass; n. peripancreatic lesion; o. splenic lesion; p. subcutaneous/skin nodule; q. ground glass opacity; r. axillary LN; s. subcarinal LN; t. vertebral body metastasis; u. thyroid nodule; v. neck mass.

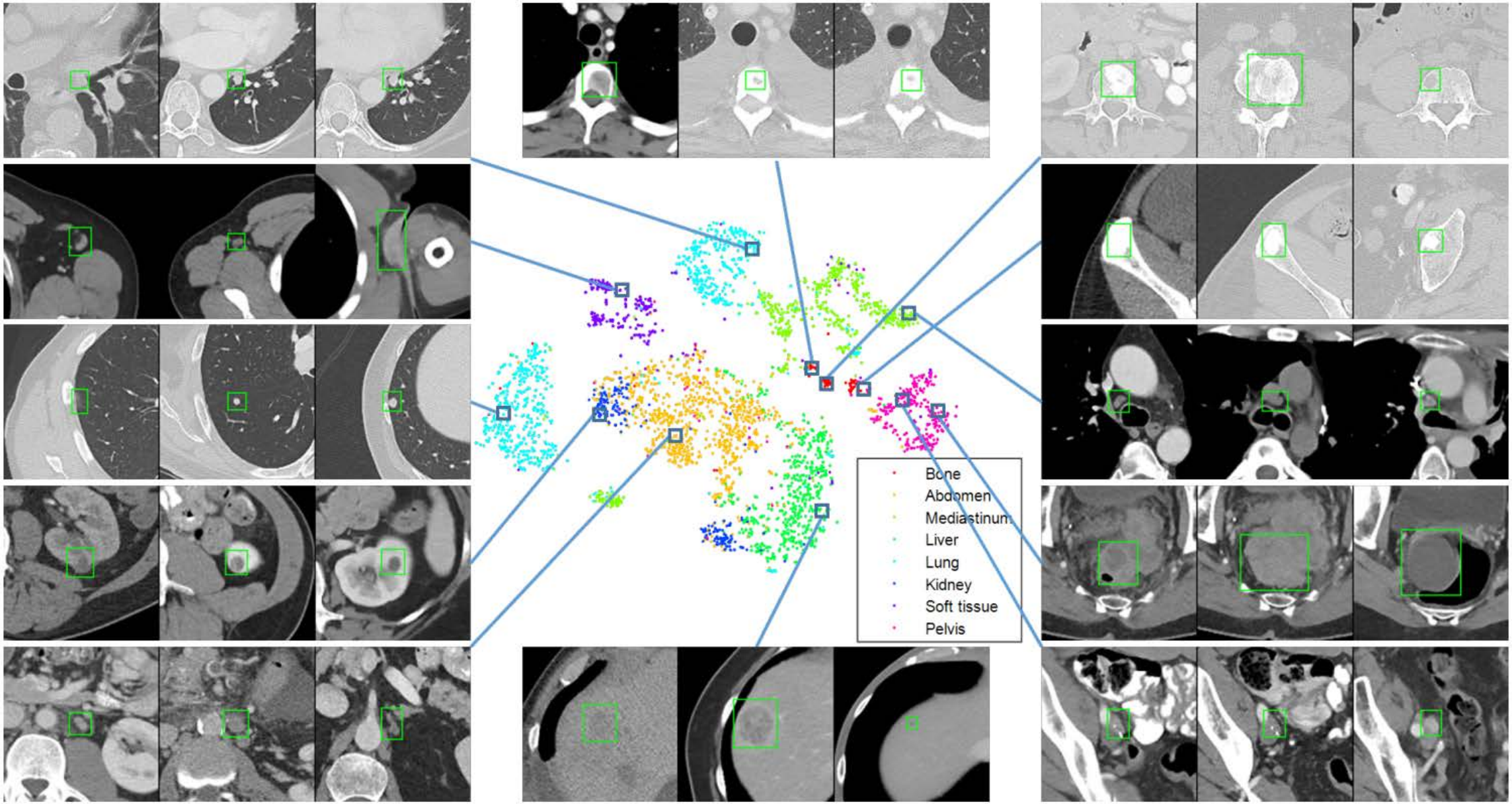


Figure 5. *t-SNE visualization of the lesion embeddings on the test set (4,927 samples) of DeepLesion. Colors indicate the manually labeled lesion types. We also split the samples to 128 clusters using K-means and show 3 random lesions in 12 representative clusters. We did not choose to show closest samples because they are very similar. Best viewed in color.*

Experiments: Lesion Retrieval

Multi-scale deep lesion appearance vector via Triplet Network to encode lesion type, location and size (thus sub-types)!

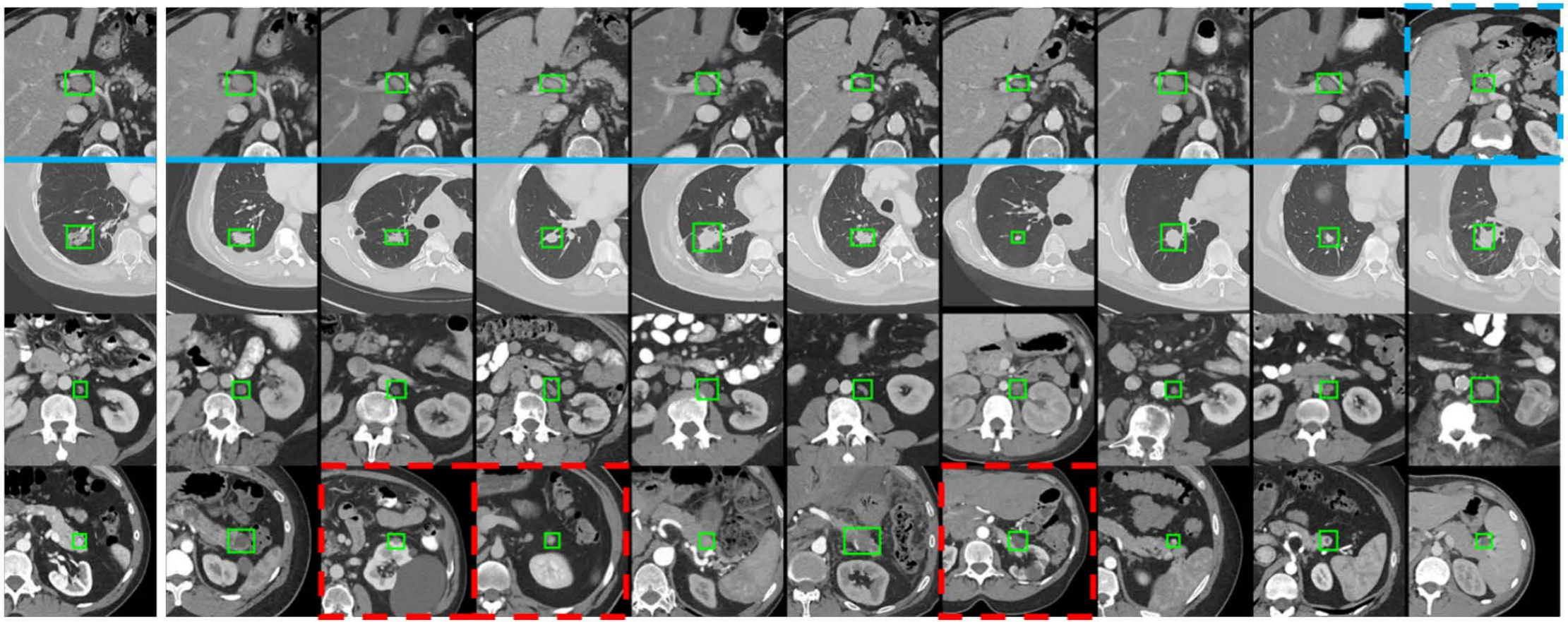


Figure 6. *Examples of query lesions (first column) and the top-9 retrieved lesions on the test set of DeepLesion. In the first row, the blue dashed box marks the lesion from a different patient than the query one, whereas the other 9 are all from the same patient. In rows 2–4, we constrain that the query and all retrieved lesions must come from different patients. Red dashed boxes indicate incorrect results, see text.*

Experiments

- Quantitative Results of lesion **retrieval** (average retrieval error of continuous labels, ARE), **clustering** and **classification**

$$\text{ARE} = \frac{1}{K}\sum_{i=1}^{K} \|\mathbf{t}^Q - \mathbf{t}_i^R\|_2$$

Feature representation	Average retrieval error			Clustering		Classification
	Type	Location	Size	Purity	NMI	Accuracy
Baseline: Multi-scale ImageNet feature	15.2	9.6	6.9	58.7	35.8	86.2
Baseline: Location feature	22.4	**2.5**	8.8	51.6	32.6	59.7
Triplet with type	8.8±0.2	10.8±0.2	5.7±0.1	84.7±2.8	71.5±1.7	89.5±0.3
Triplet with location	13.0±0.1	6.5±0.1	6.2±0.1	61.1±4.4	39.5±4.3	87.8±0.5
Triplet with type + location	8.7±0.2	7.2±0.1	6.0±0.1	81.3±4.7	68.0±2.4	89.9±0.3
Triplet with type + location + size	**8.5±0.1**	7.2±0.0	**5.1±0.0**	**86.0±3.9**	**72.4±4.6**	**90.5±0.2**
w/o Multi-scale feature: conv5	11.5±0.2	7.1±0.1	6.3±0.0	79.8±0.6	64.8±1.2	86.6±0.4
w/ Multi-scale feature: conv5 + conv4	9.7±0.2	7.0±0.0	5.4±0.1	82.4±3.3	67.9±2.2	89.0±0.6
w/o Iterative refinement	8.7±0.2	7.3±0.0	5.2±0.1	85.4±2.8	69.8±2.0	90.2±0.2

Table 1. *Evaluation results on the test set (4,927 samples) of DeepLesion. For retrieval, we compute the average retrieval error (%) in type, location, and size of the top-5 retrieved lesions compared to the query one. For clustering, lesions are clustered to 8 groups using K-means to calculate the purity and NMI (%). For classification, we train a 8-way softmax classifier on the seed labeled samples and apply it on the test set. The CNN in each method was trained 5 times using different random seeds. Mean results and standard deviations are reported.*

Experiments: Analysis & Insights

- When location and size are added as supervision cues, our network performs best on lesion-type retrieval; even better than when only lesion-type is used as the cue.
- **Location** and **size** provide important complementary information in learning lesion similarity embedding
- When only coarse-scale features (conv5, conv4) are used, **location** ARE is slightly better because location mainly relies on higher-level context information
- Fusing fine-level features (conv3, conv2) significantly improves **type** and **size** prediction, indicating details are important in these aspects
- **Iterative lesion feature/type refinement** [4] also helps!

[4] Wang et al., Unsupervised Joint Mining of Deep Features and Image Labels for Large-scale Radiology Image Categorization and Scene Recognition, IEEE WACV 2017

Experiments: Classification

- The most confusing types are mediastinum/lung lesions, and abdomen/liver/kidney lesions, since some of them are similar in both appearance and location.

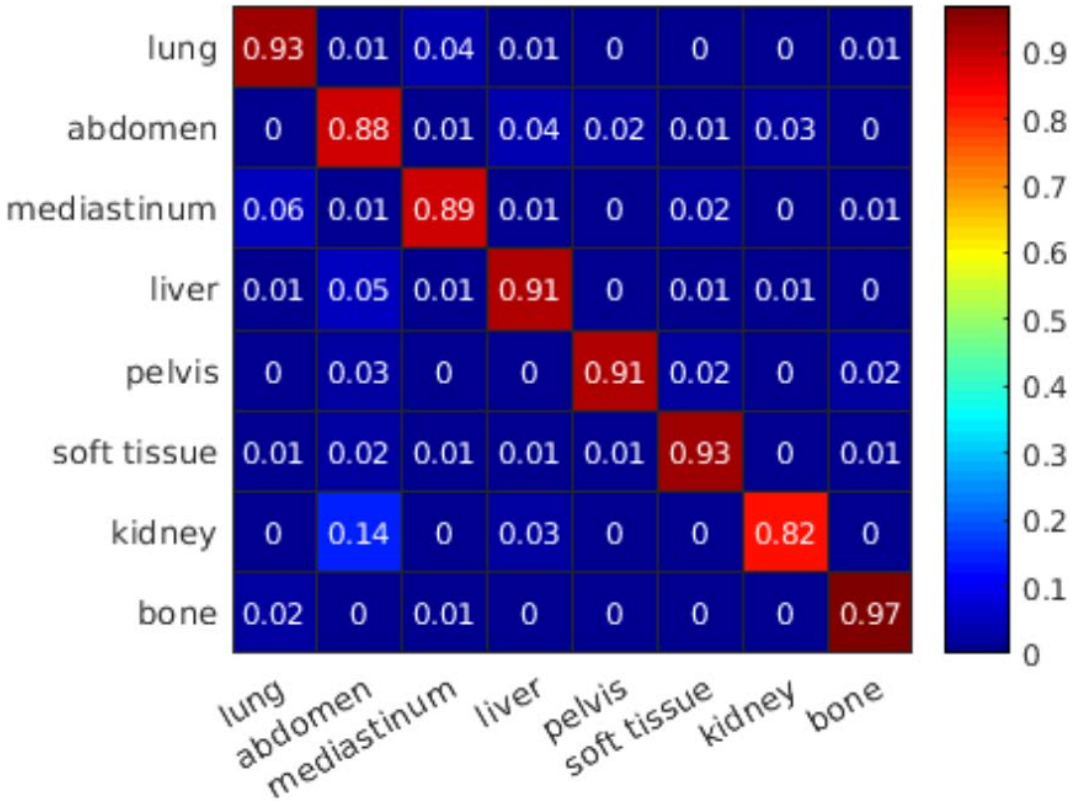


Figure 7. *The confusion matrix of lesion classification.*

Experiments: Lesion Matching

- Manually grouped 1313 lesions from 103 patients in DeepLesion to **593 groups** for evaluation
- 1–11 lesions per group
- A true positive decision assigns two lesions of the same instance to the same group, and a false positive decision assigns two lesions of different instances to the same group, etc. → **Good quantitative lesion matching accuracy!**

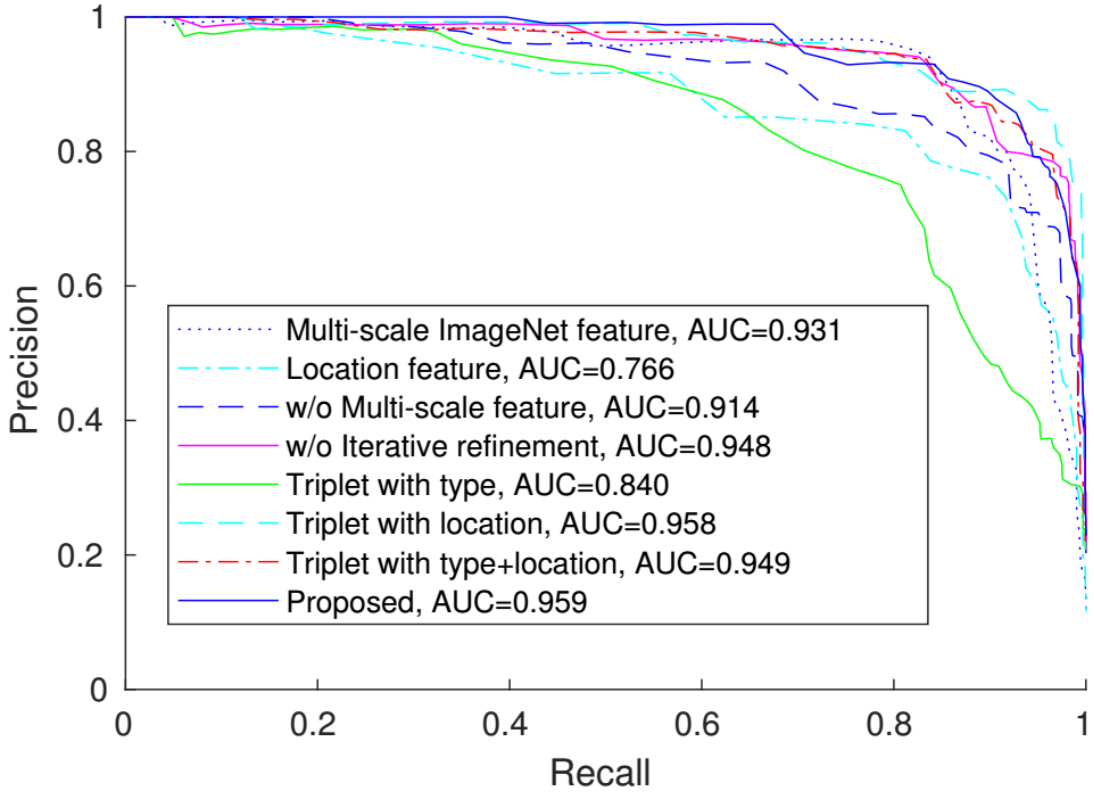


Figure 8. *Precision-recall curves of various methods on the intra-patient lesion matching task using DeepLesion. The area-under-curve (AUC) values are shown in the legends.*

Longitudinal Lesion Matching

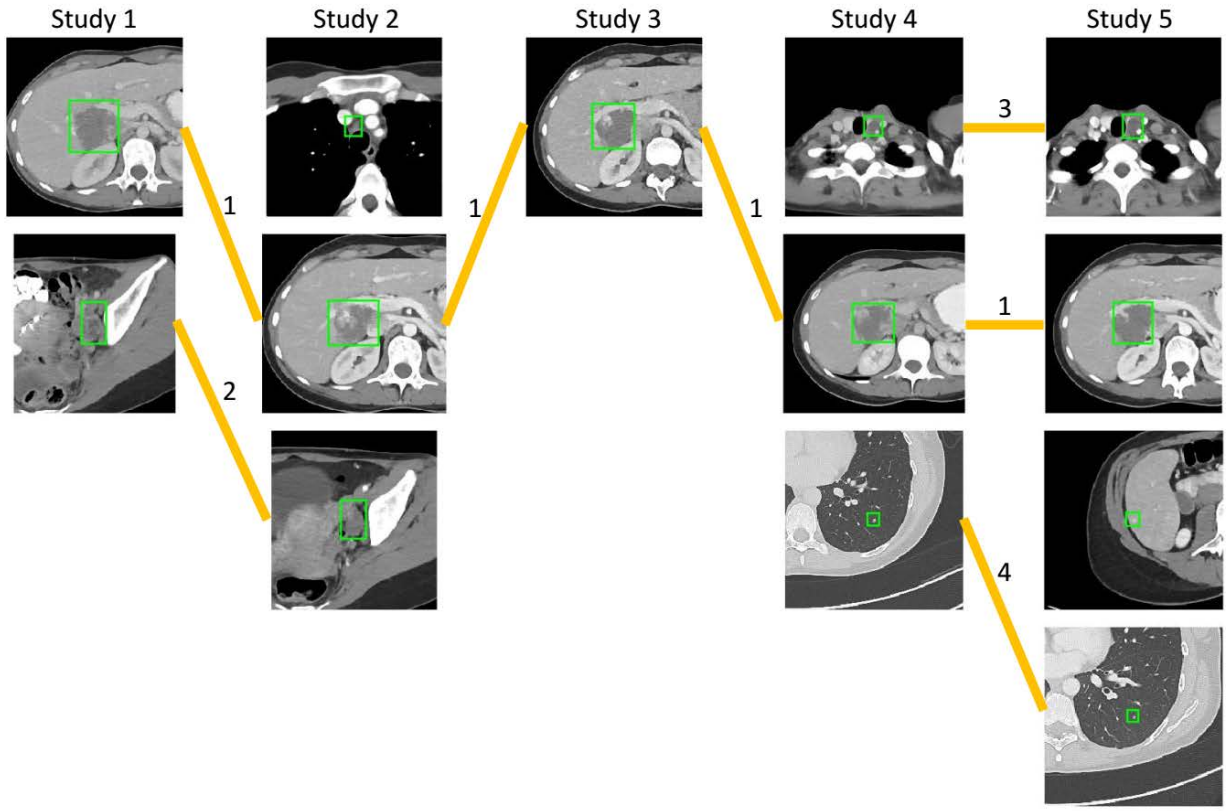


Figure 5. All lesions of a sample patient in DeepLesion. Lesions in each study (CT examination) are listed in a column. Not all lesions occur in each study, because the scan ranges of each study vary and radiologists only mark a few target lesions. We group the same lesion instances to sequences. Four sequences are found and marked in the figure, where the numbers on the connections represent the lesion IDs.

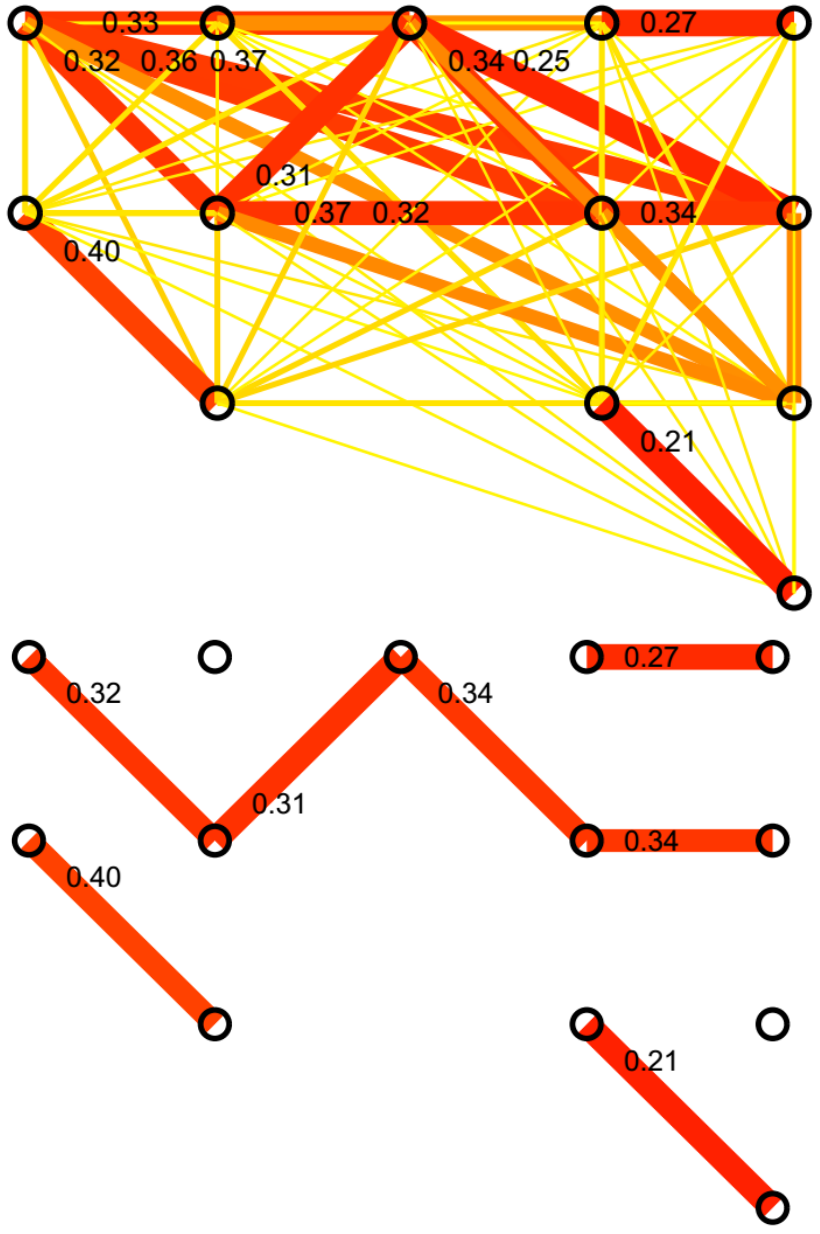
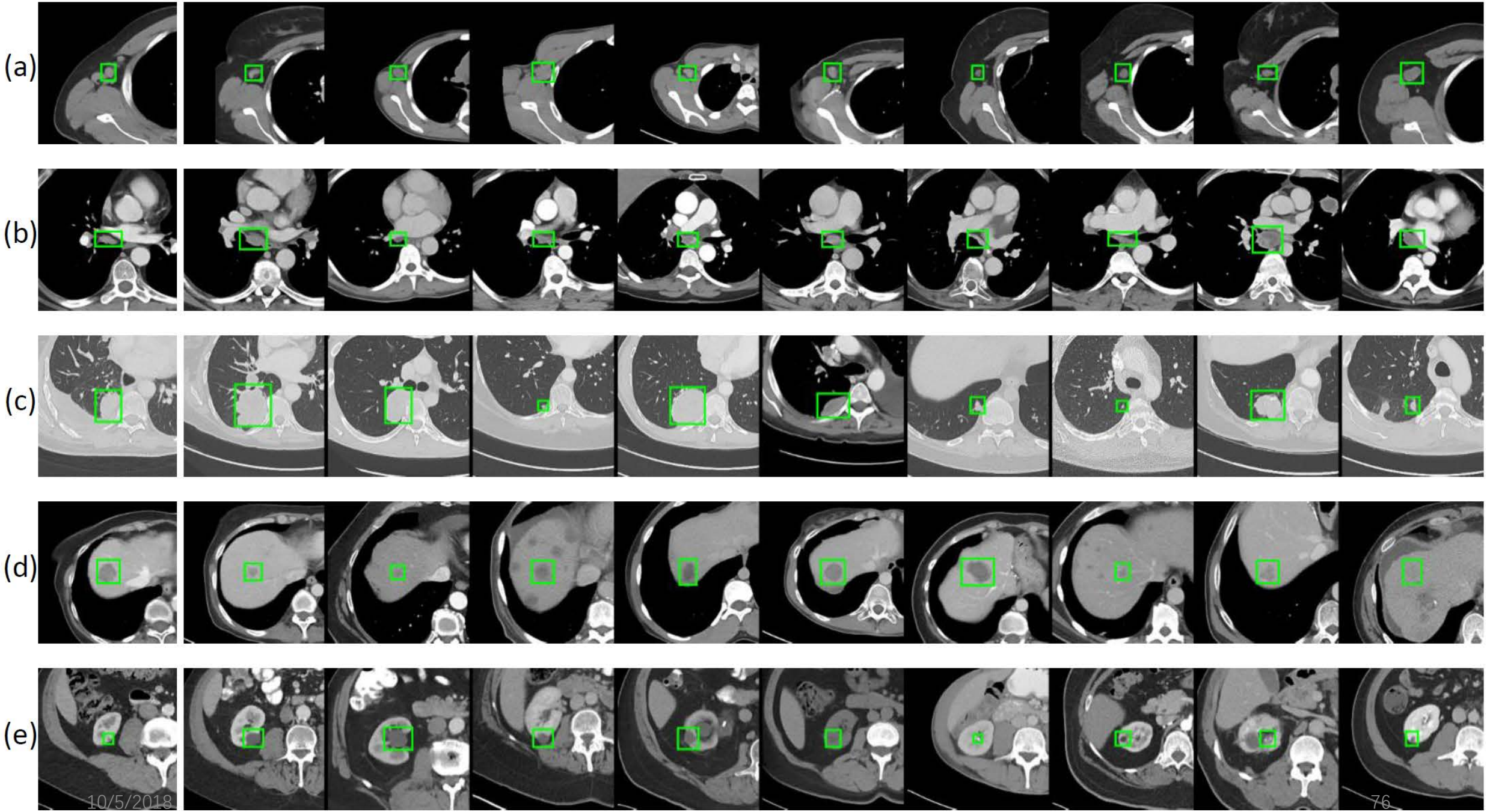


Figure 7. The final lesion sequences found by processing the lesion graph in Fig. 6 using Algo. 1 in the paper. They are the same with the ground-truth in Fig. 5.


(a)
(b)
(c)
(d)
(e)

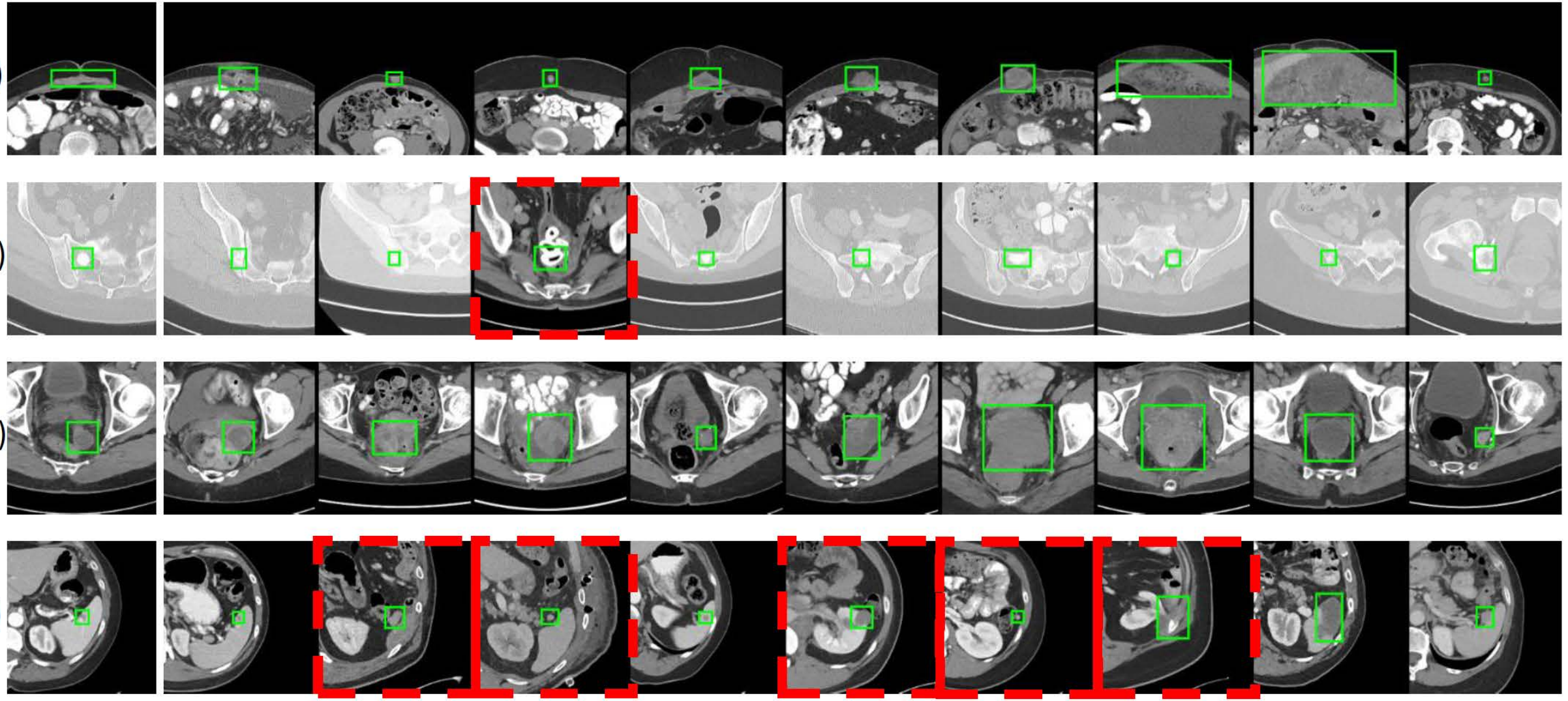


Figure 4. More examples of query lesions (first column) and the top-9 retrieved lesions on the test set of DeepLesion. We constrain that the query and all retrieved lesions must come from different patients. Red dashed boxes indicate incorrect results. The lesions in each row are: (a) Right axillary lymph nodes; (b) subcarinal lymph nodes; (c) lung masses or nodules near the pleura; (d) liver lesions near the liver dome; (e) right kidney lesions; (f) lesions near the anterior abdomen wall; (g) lesions on pelvic bones except the one in the red box, which is a peripherally calcified mass. (h) inferior pelvic lesions; (i) spleen lesions except the ones in red boxes.

Summary

- We present a large-scale and comprehensive dataset, **DeepLesion**, including significant radiology image findings mined from PACS
  - can be used for multi-category lesion detection, retrieval, classification, segmentation, …, **the first study of its kind!**
- **Lesion Graph Embedding** is learned with a triplet network to model their similarity relationship in type, location, and size
  - The only manual efforts needed are the class labels of some seed images
  - **Non-parametric deep** radiology instance/knowledge representation
- Promising results are obtained in (a) inter-patient content-based lesion retrieval and (b) intra-patient lesion matching, qualitatively and quantitatively.

[5] Liu, Lu, Ye, Yu, Huang: Coarse-to-fine classification via parametric and **nonparametric models for computer-aided diagnosis**. ACM CIKM 2011

Acknowledgement

- **Scientific Collaborators:** Holger Roth, Xiaosong Wang, Ke Yan, Youbao Tang, Mingchen Gao, Jianhua Yao, Yifan Peng, Hoochang Shun, Ling Zhang, Adam Harrison, Jinzheng Cai, Jiamin Liu, Ari Seff, Isabella Nogues, Amal Farag, Ziyue Xu, Yuxing Tang, Zhiyong Lu, Jing Xiao, Lin Yang, Nathan Lay
- **Clinical Collaborators:** Lauren Kim (MD), Mohammadhad Bagheri (MD), Ronald Summers (MD), Evrim Turkbey (MD)
- NIH Intramural Research Program, NVIDIA GPU grants