IBM Research

Automatic Diagnosis of Pulmonary Embolism Using an Attention-guided Framework: A Large-scale Study

Luyao Shi[1], Deepta Rajan[2], Shafiq Abedin[2], Manikanta Srikar Yellapragada[3], David Beymer[2], and Ehsan Dehghan[2]

[1]Yale University
[2]IBM Research
[3]New York University

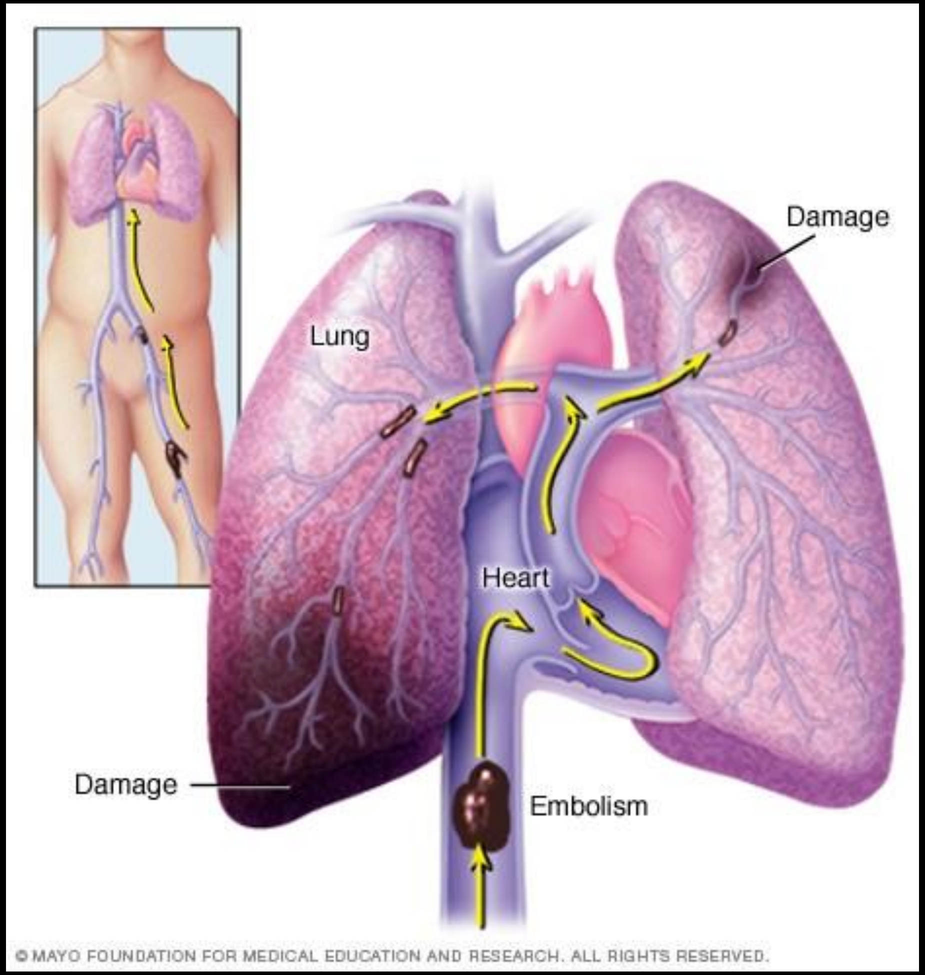
MIDL
Montréal 2020

About pulmonary embolism (PE)

- **Causes**: a clump of material, most often a blood clot, gets wedged into an artery in patients' lungs. These blood clots most commonly come from the deep veins of patients' legs.

- **Mortality**:
  - About 100,000 deaths/year in US.
  - 1 of 4 people who have a PE die without warning.
  - 10 to 30% of people will die within one month of diagnosis.

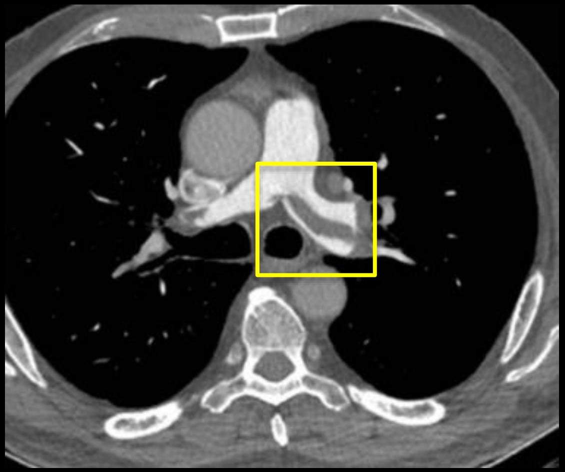
- Prompt recognition of the diagnosis and immediate initiation of therapeutic action is important.

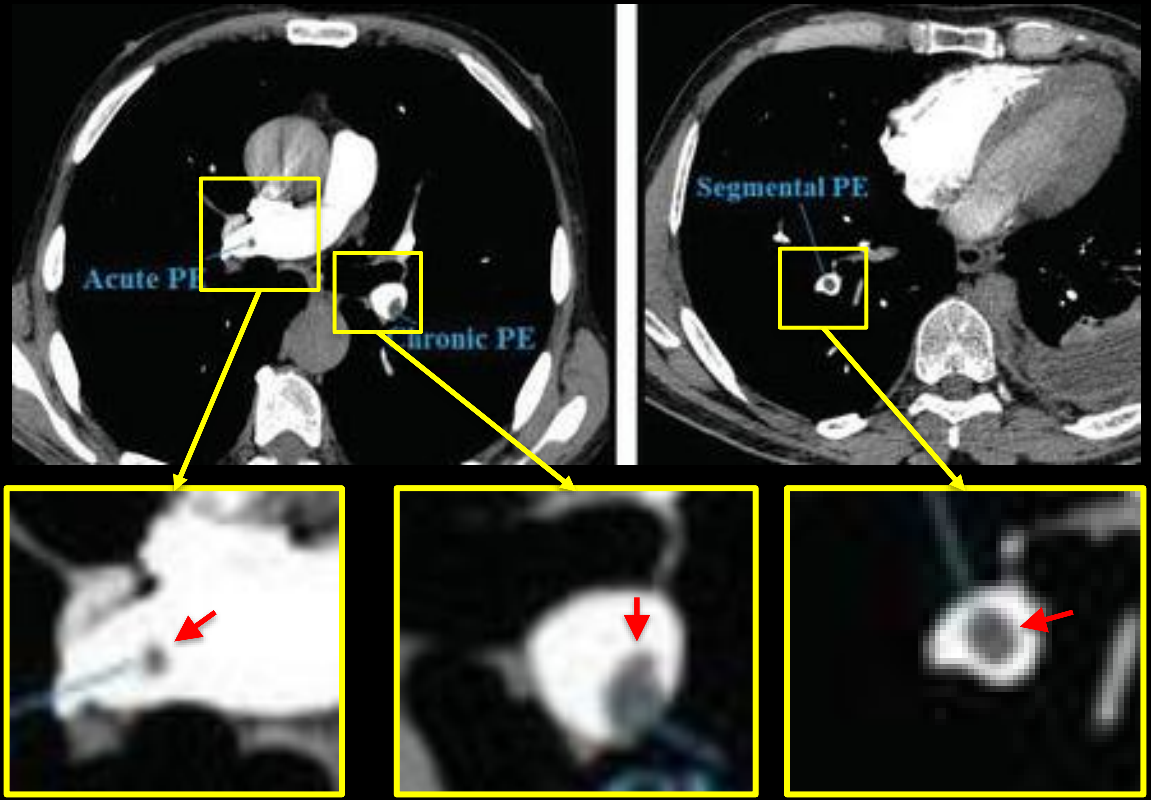


About pulmonary embolism (PE)

- Contrast Enhanced Chest CT is the preferred method of diagnostic imaging in patients with a clinical risk score indicative of PE.

- PE can be visualized as perfusion defects.

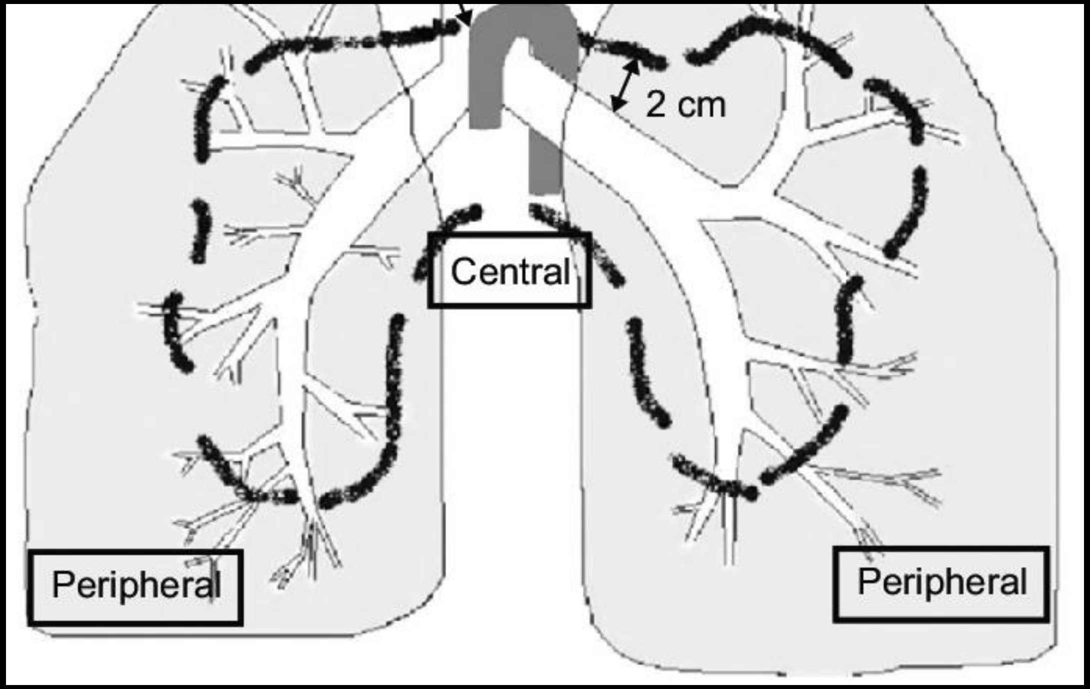




Abdel-Razzak M. Al-hinnawi. Computer-Aided Detection, Pulmonary Embolism, Computerized Tomography Pulmonary Angiography: Current Status (2018).

Motivation

- **Challenges**:
  - Increased probability of false-positive findings when the lesions involve peripheral pulmonary vascular regions.
  - Confounding factors:
    - Poorly filled vein with contrast media
    - Impacted bronchi or parenchymal disease
    - Lymphoid tissues around the vessels
    - Respiratory/cardiac motion artifacts
    - Image noise
  - PE detection/exclusion is quite time-consuming and dependent on the experience of the radiologist.
- **GOAL**: A deep learning-based computer-aided diagnosis (CAD) platform to detect PE with high accuracy.



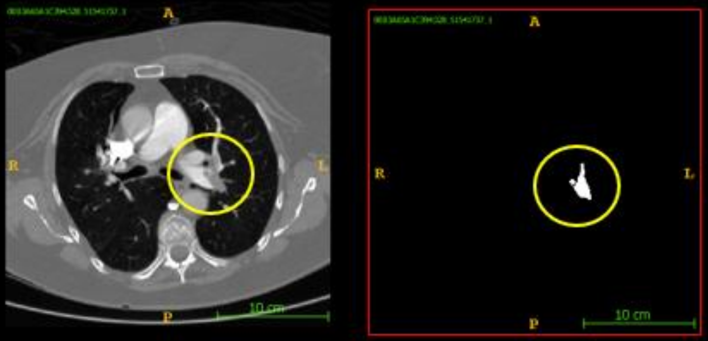
In-Hye Jung et al. Clinical outcome of fiducial-less Cyber Knife radiosurgery for stage I non-small cell lung cancer.(2015).

Training strategy

	Pros	Cons
Training with pixel-level annotated data	• Better interpretability • Higher accuracy	• Time consuming for radiologists • Limited availability • Less scalability
+		
End-to-end training with patient-level labels	• Largely available training data • Better scalability	• Less interpretability • Potentially worse performance

=

Hybrid Training

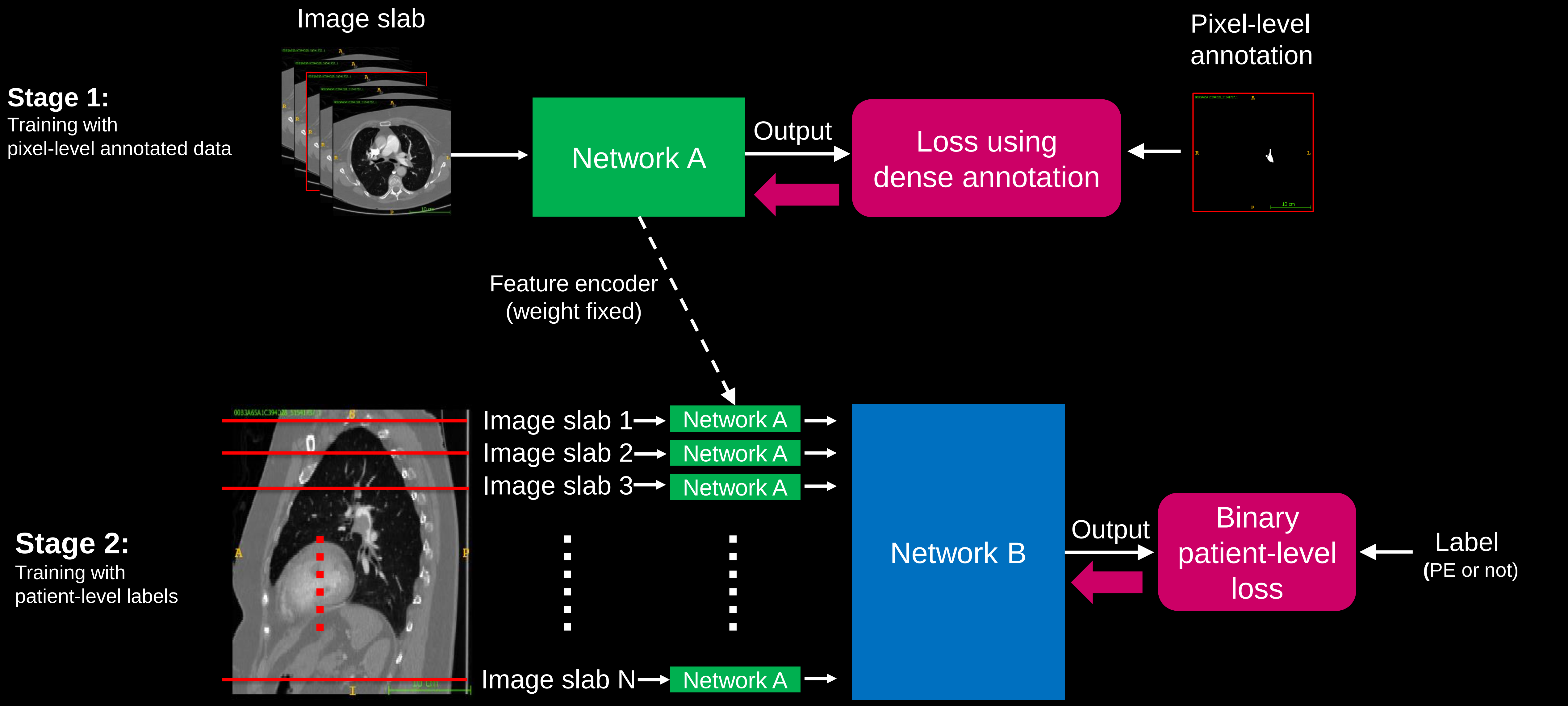


pixel-level annotated data

→ PE or not?

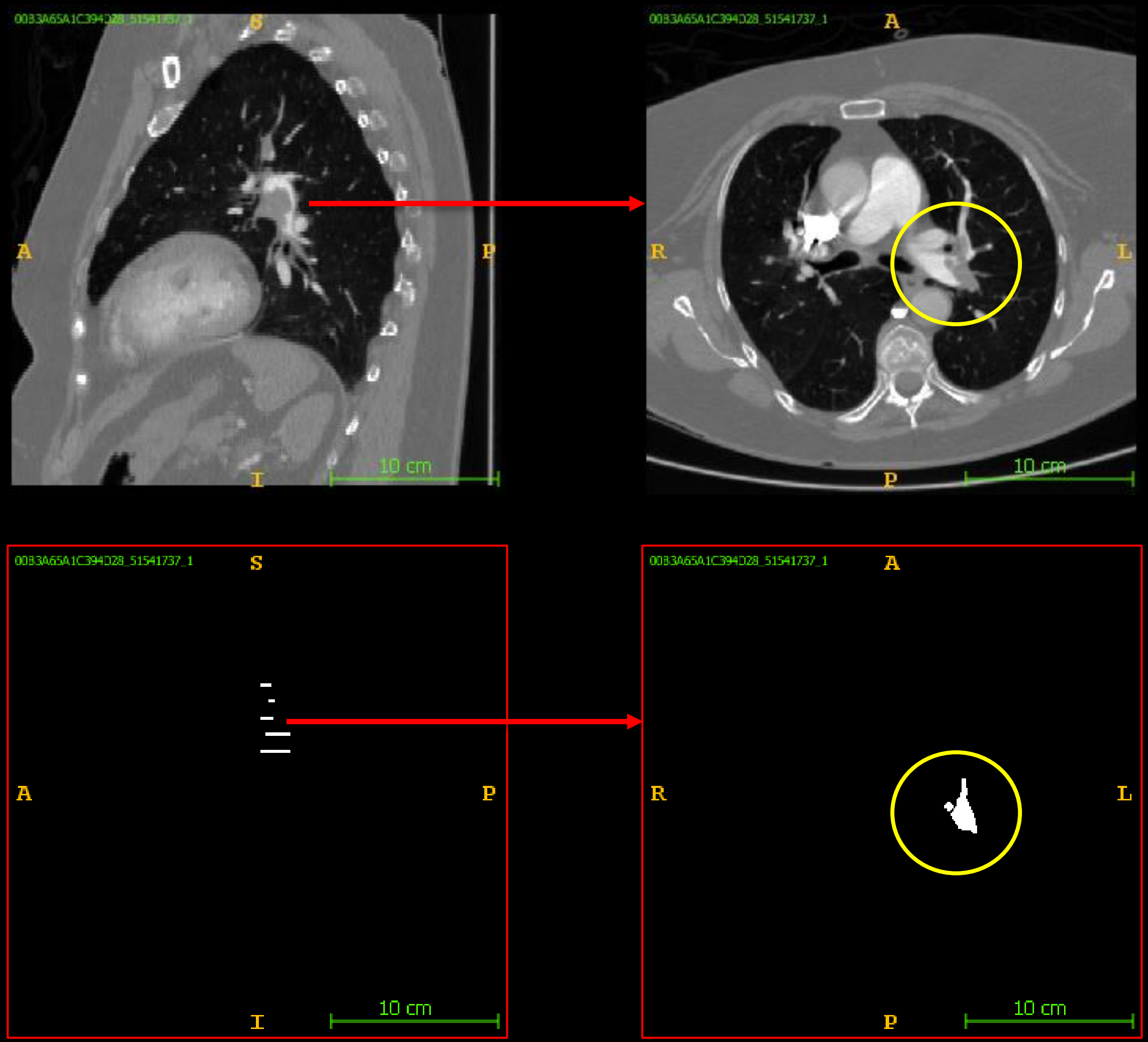
patient-level label

Hybrid Training Overview



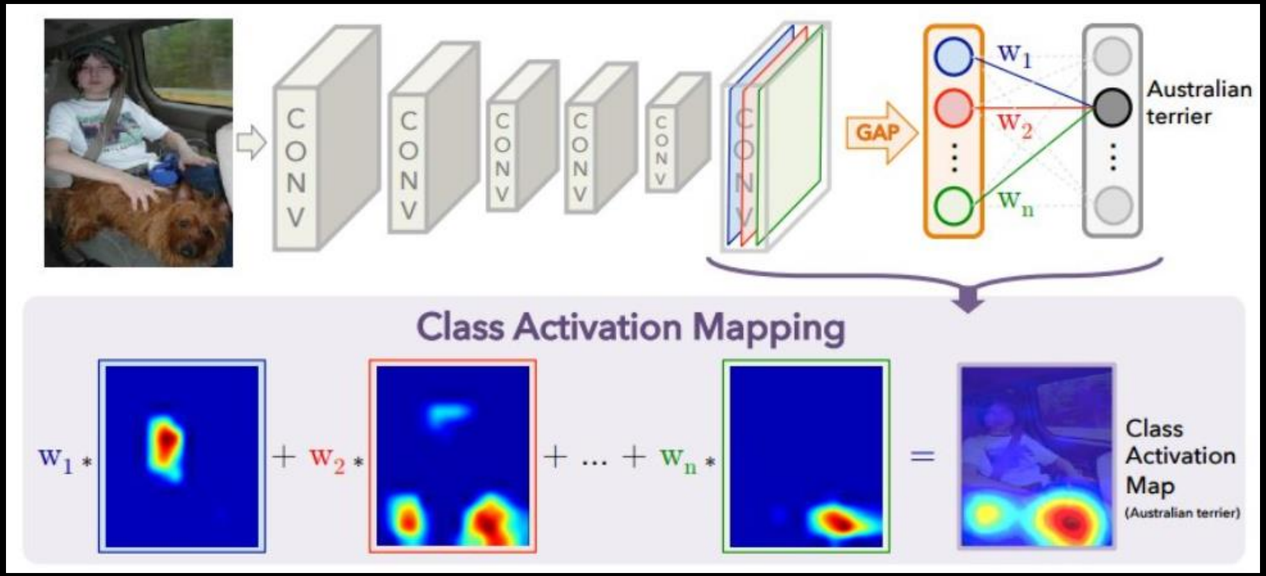
Stage1: training with pixel-level annotated data

- Pixel-level annotations every 10mm
- **Goal**: train an image encoding network that focus its attention on PE



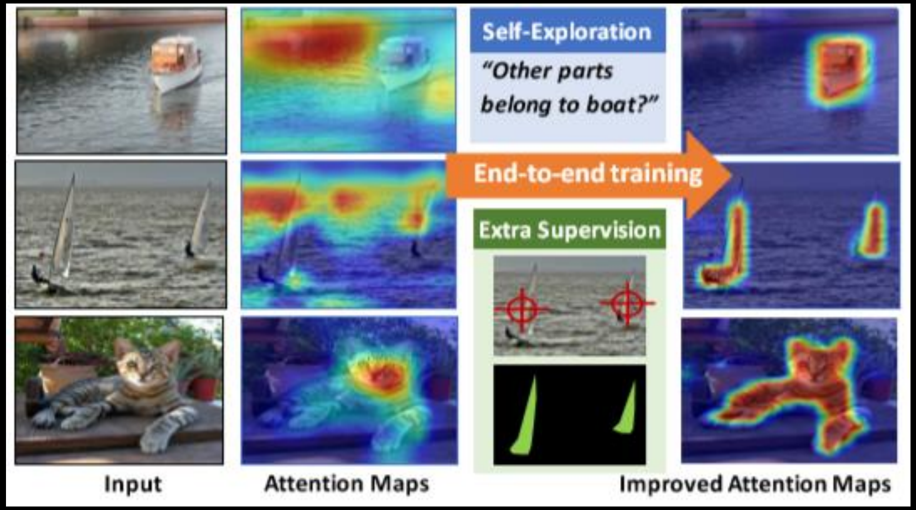
Attention map

- Class activation map (CAM): indicates the discriminative image regions used by the CNN to identify a particular class.



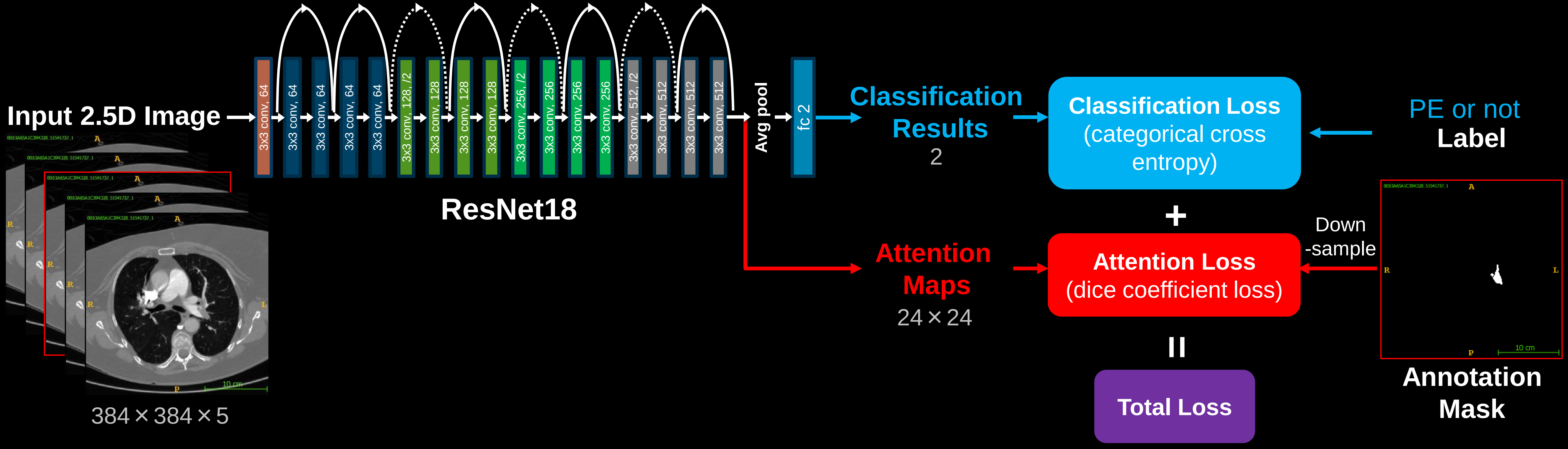
Zhou et al. Learning Deep Features for Discriminative Localization. (2016)

- Guided attention inference networks (GAIN): supervise the attention maps while training the network.



Li et al., Tell Me Where to Look: Guided Attention Inference Network. (2018)

Stage1: attention-guided training



- Resample volumetric images (bilinear interpolation): slice thickness [0.5mm, 5mm] → 2.5mm
- 10,388 slabs (5 slices) of annotated pairs from 1,670 positive volumetric images
- Same amount of negative slabs randomly sampled from 593 negative volumetric images
- Image cropped to center 384 × 384, [-1024HU,500HU] → [0, 255]
- 80% training, 20% validation
- Training epochs: 100 (save the model with the highest val. acc.)

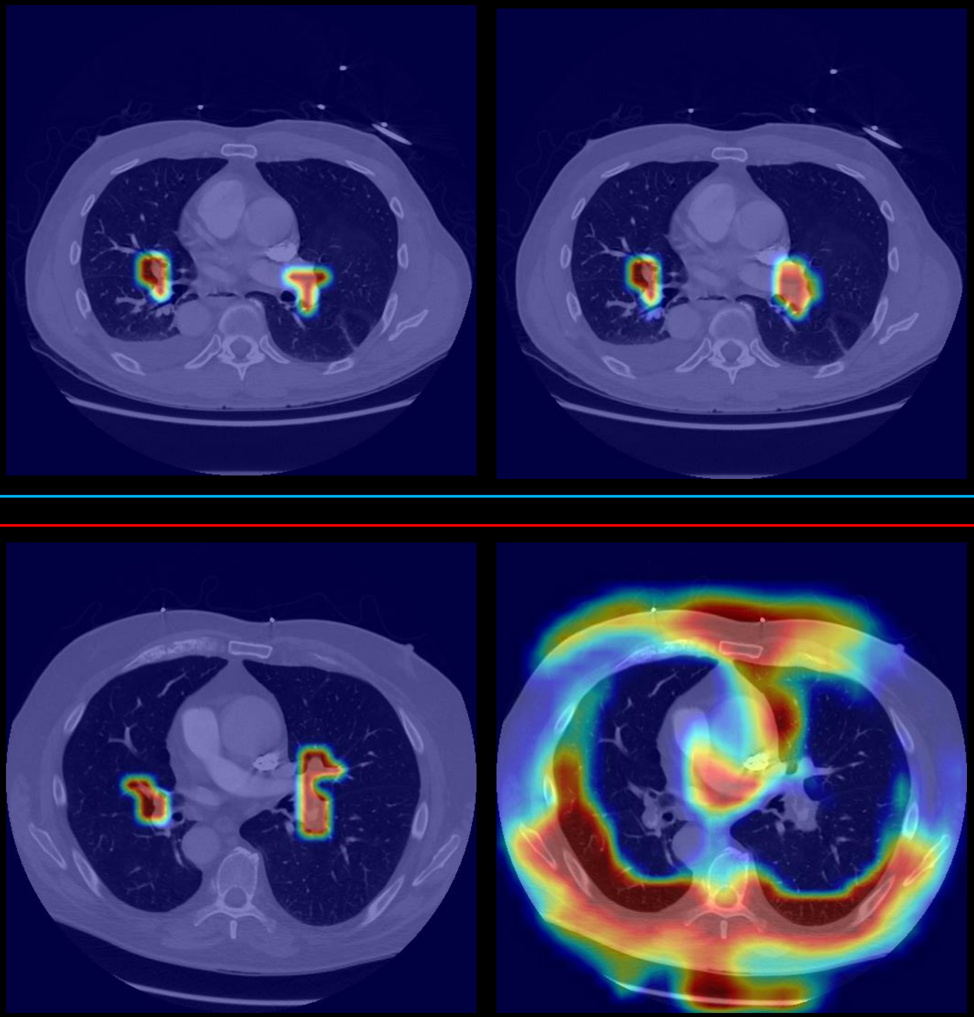
Stage1: results on the validation set

Example Attention Maps

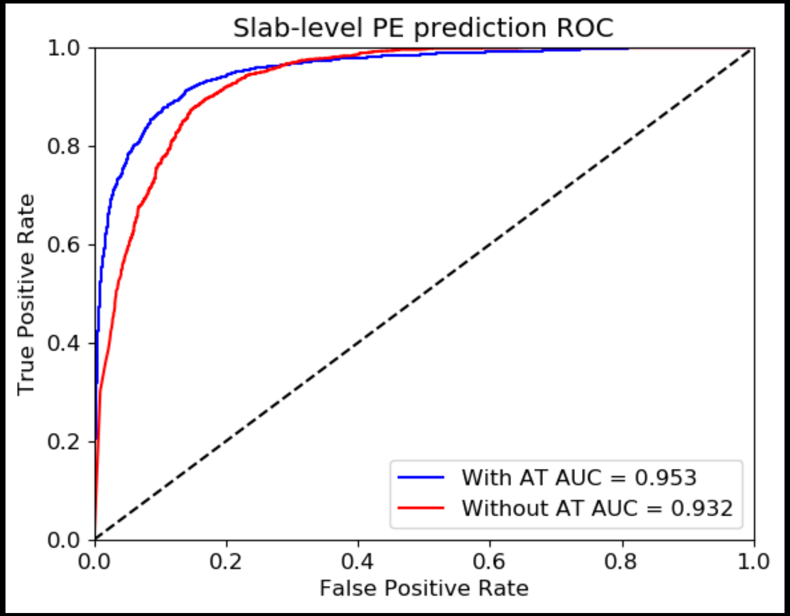
Annotation Mask (down-sampled) | Attention Map

With Attention Training

Without Attention Training

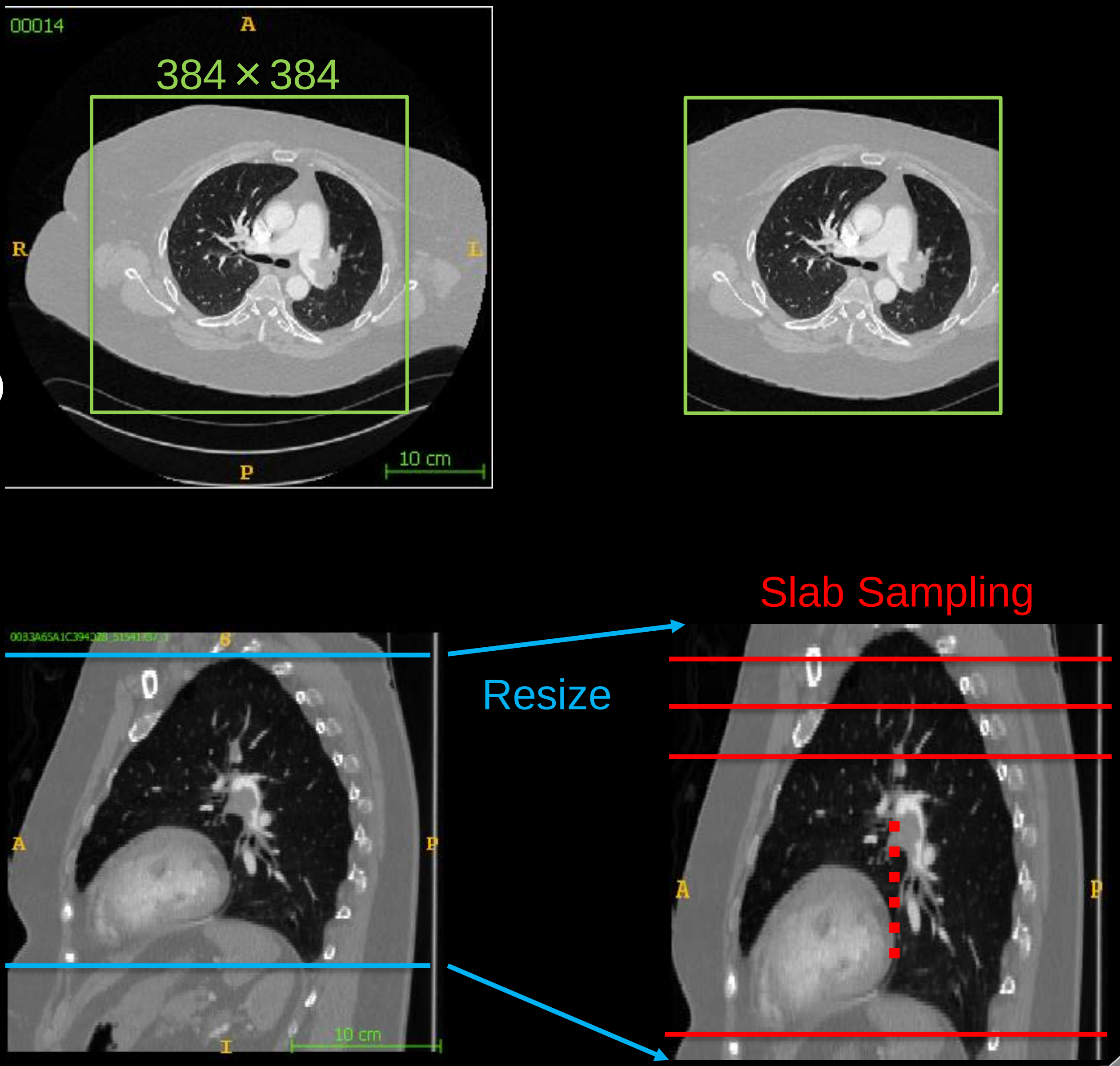


ResNet’s slab-level PE prediction result on the validation data

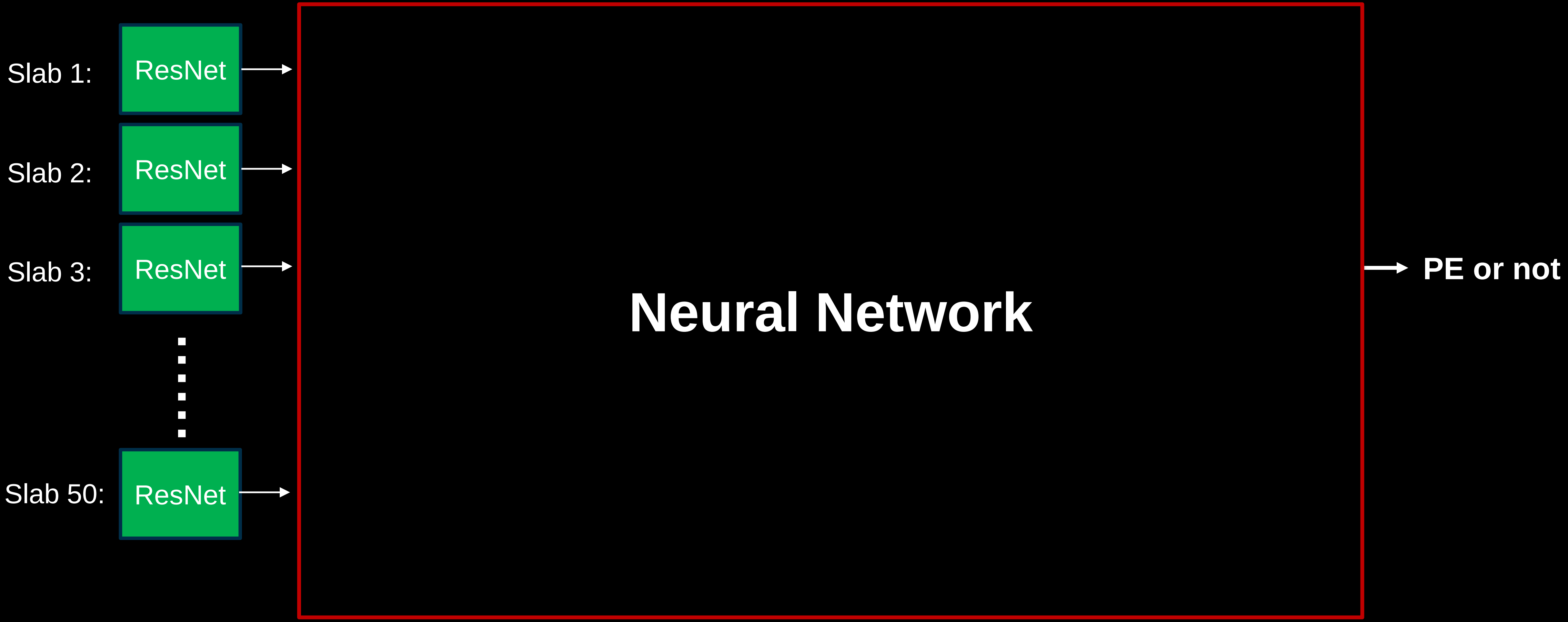


Stage2: training data and image pre-processing

- **Data-preprocessing:**
  - Image cropped to center 384 × 384, [-1024HU,500HU] → [0-255]
  - Identify lung regions using lung mask (produced by in-house lung segmentation method) resize to 200 slices, then sample 50 slabs
  - ? × 512 × 512 → 50 × 384 × 384 × 5

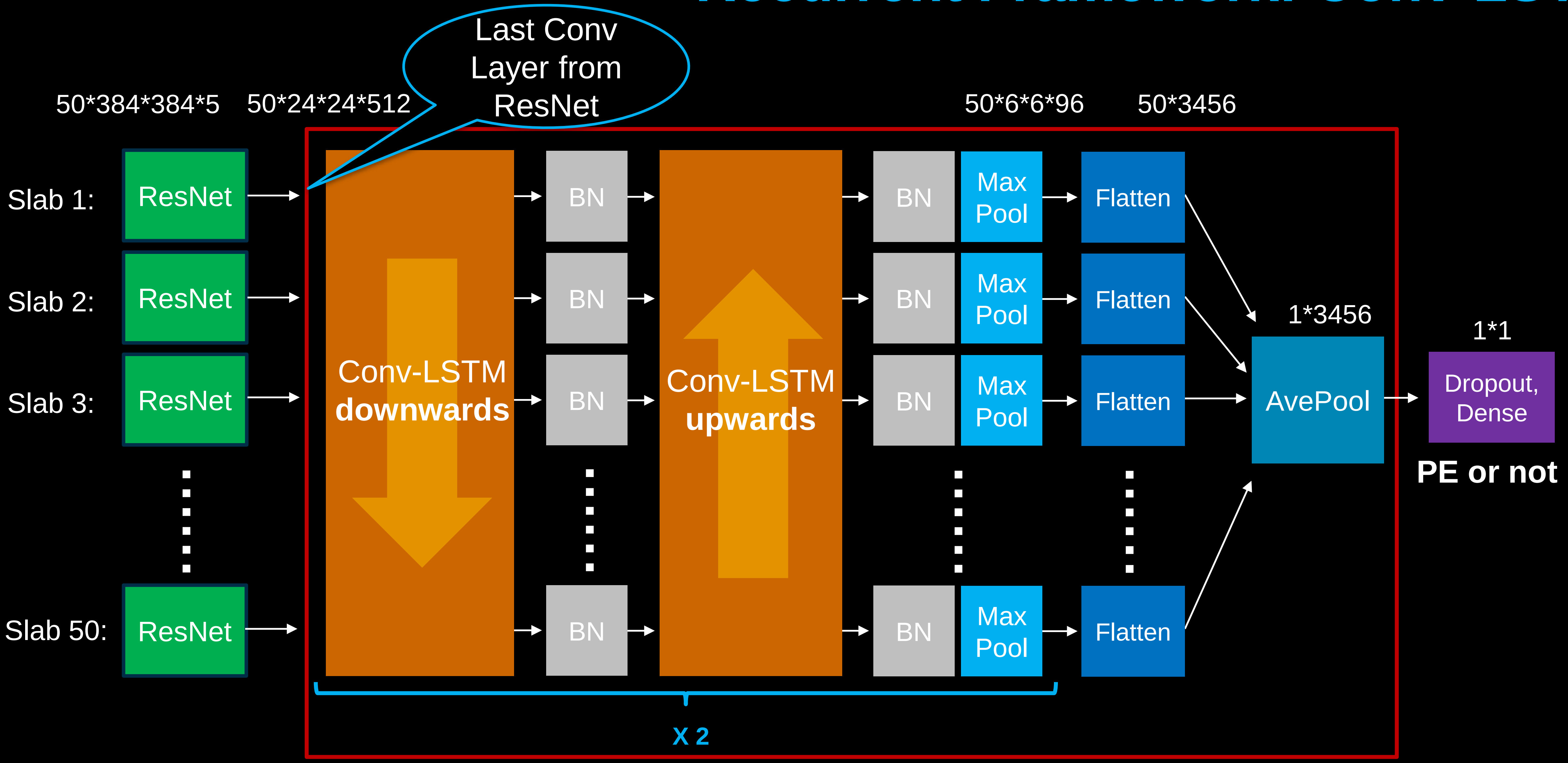


Stage2: training with patient-level labeled data

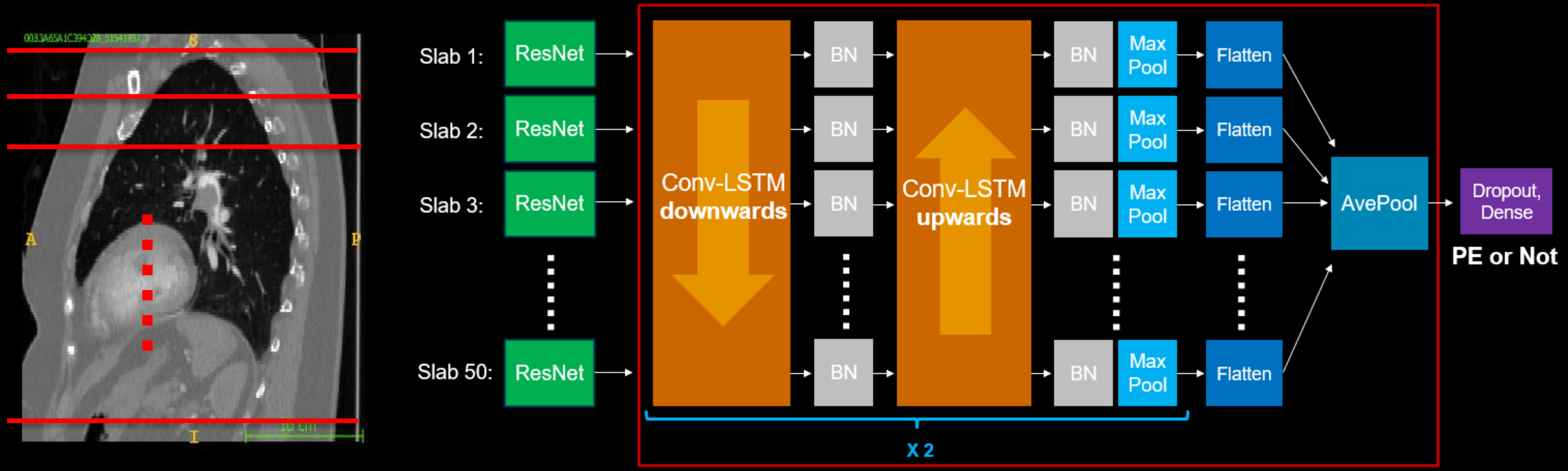


Stage2: training with patient-level labeled data

Recurrent Framework: Conv-LSTM



Stage2: training parameters



- Classification loss: Binary cross entropy (BCE)
- Optimizer: Adam optimizer
- Learning rate: 10^{-4}
- Training epochs: 50 (save the model with the highest val. acc.)

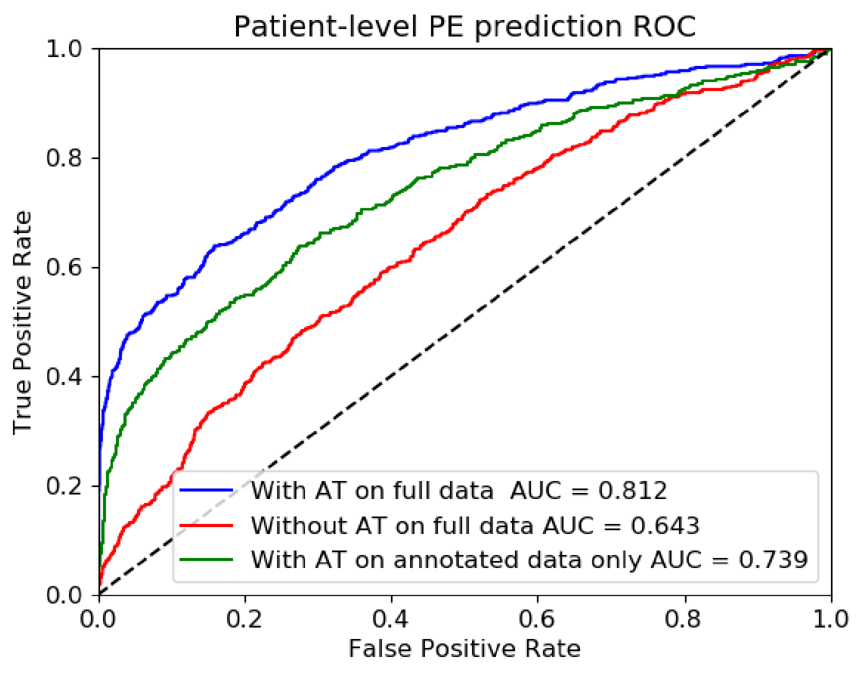
Stage2: patient-level inference results on testing data

IBM

- **Training data:**
  - Annotated Studies: 1670+, 593-
  - Labeled volumetric images: 4186+, 4603-
  - 80% training, 20% validation

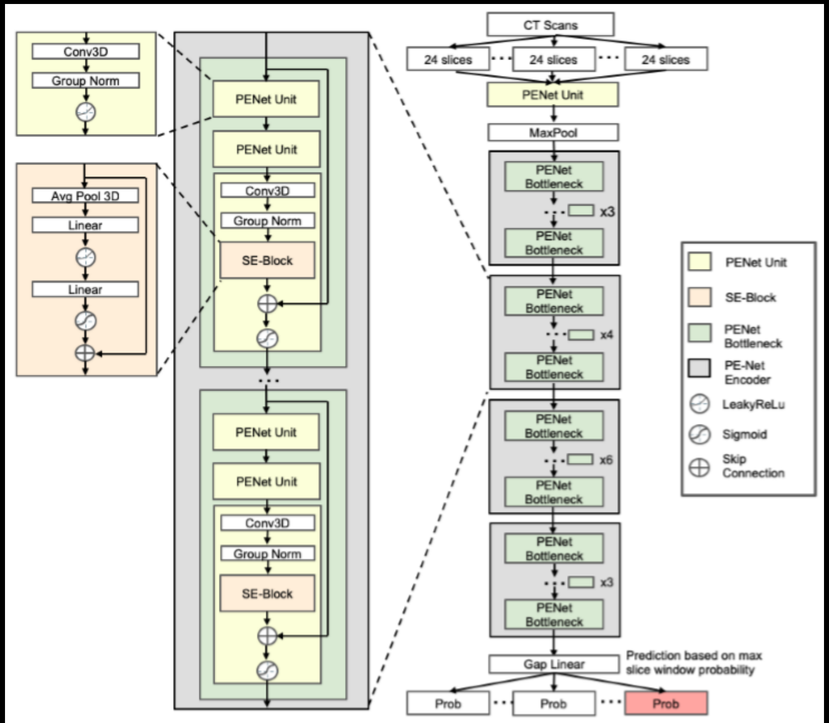
- **Testing data** (2160 total): 517+, 1643-

	Stage 1 Data	Stage 1 Loss	Stage 2 Data	AUC
Scenario 1	1670+, 593-	Atten. Loss + Cls. Loss	1670+, 593-	0.739
Scenario 2	1670+, 593-	Cls. Loss	1670+, 593- & 4186+,4603-	0.643
Scenario 3	1670+, 593-	Atten. Loss + Cls. Loss	1670+, 593- & 4186+,4603-	**0.812**



Comparison with state-of-the-art

PENet



- Training data was labeled on a slice level for the presence/absence of a PE

SC Huang, et al. PENet - a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging. (2019)

3D CNN

- Starts with an I3D model (3D CNN pretrained on video action recognition dataset)
- Demonstrated success in acute aortic syndrome detection
- Trained only on our patient-level labeled PE data

MS Yellapragada, et al. Deep Learning Based Detection of Acute Aortic Syndrome in Contrast CT Images.(2020)

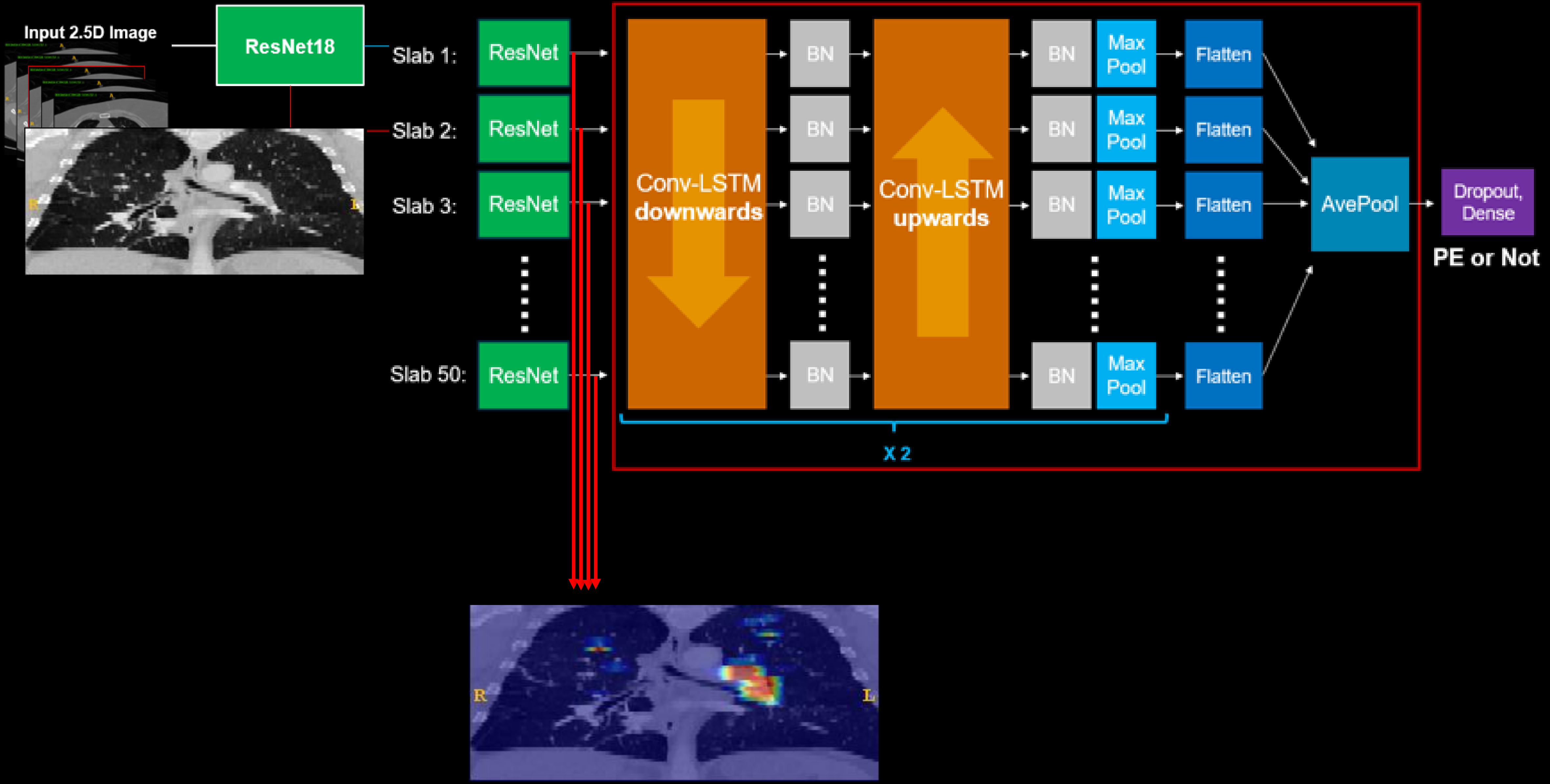
Comparison with state-of-the-art

Approach	Testset description: Size	Clinical sites	Img. protocols	AUC	Accuracy
PENet (int.)	198	Single	Single	0.79	0.74
PENet (ext.)	227	Single	Single	0.77	0.67
3D CNN	2160	Multiple	Mixed	0.787	0.727
Proposed	2160	Multiple	Mixed	**0.812**	**0.781**

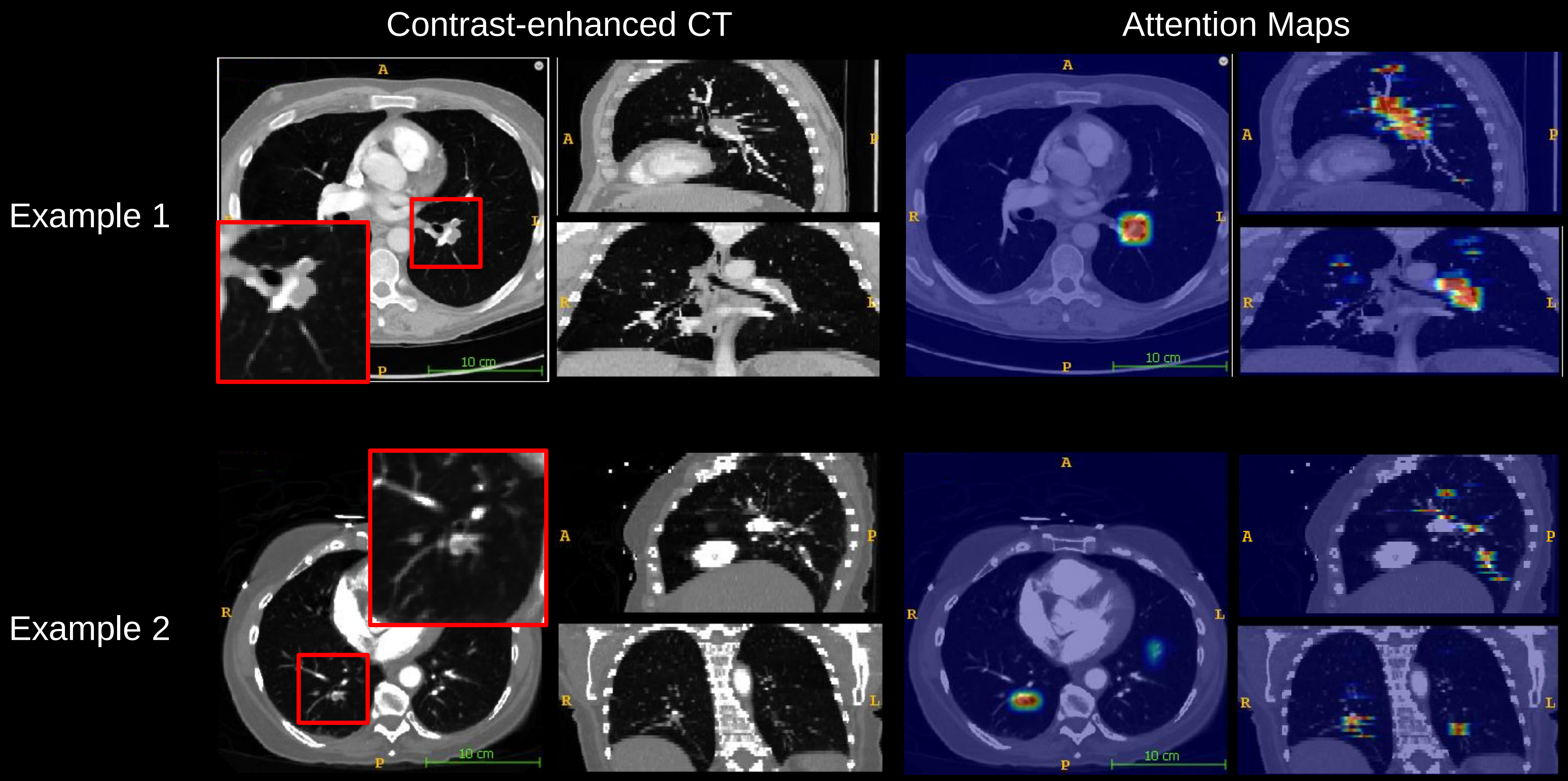
Mixed protocols:

- Contrast-enhanced Chest CT vs. CT pulmonary angiogram
- Different dose levels (noise level)
- Different image reconstruction kernels
- Slice thickness: 0.5mm-5mm

Auxiliary output – PE localization



Auxiliary output – PE localization



Future Work

- Using more efficient network structures (e.g. DenseNet) to replace ResNet18.
- In Stage1, the weights of classification loss and attention loss can be optimized (currently 1:1).
- Fully end-to-end training where the weights of ResNet can also be updated.

Summary

- We introduced a deep learning model to detect PE on volumetric contrast-enhanced CT scans using a 2-stage hybrid training strategy
  - Training with attention loss on pixel-level annotated data improves the network's localization ability
  - Second-stage convolution-LSTM networks reduce false positives on patient-level prediction
- Our evaluation involves the largest number of patient studies among all the research studies on automatic PE detection.
- Achieved state-of-the-art PE detection, while providing attention maps for radiologists as references.
- Applicable to other detection problems where the availability of volumetric imaging data exceeds radiologists' capacity to manually delineate ground truth.

Acknowledgement

IBM

Thank you!

Q&A

IBM Research

We would like to thank **Yiting Xie**, **Benedikt Graf** and **Arkadiusz Sitek** from IBM Watson Health Imaging for helping us generate the 3D CNN results.