Eye-gaze Guided Alignment between Image and Text
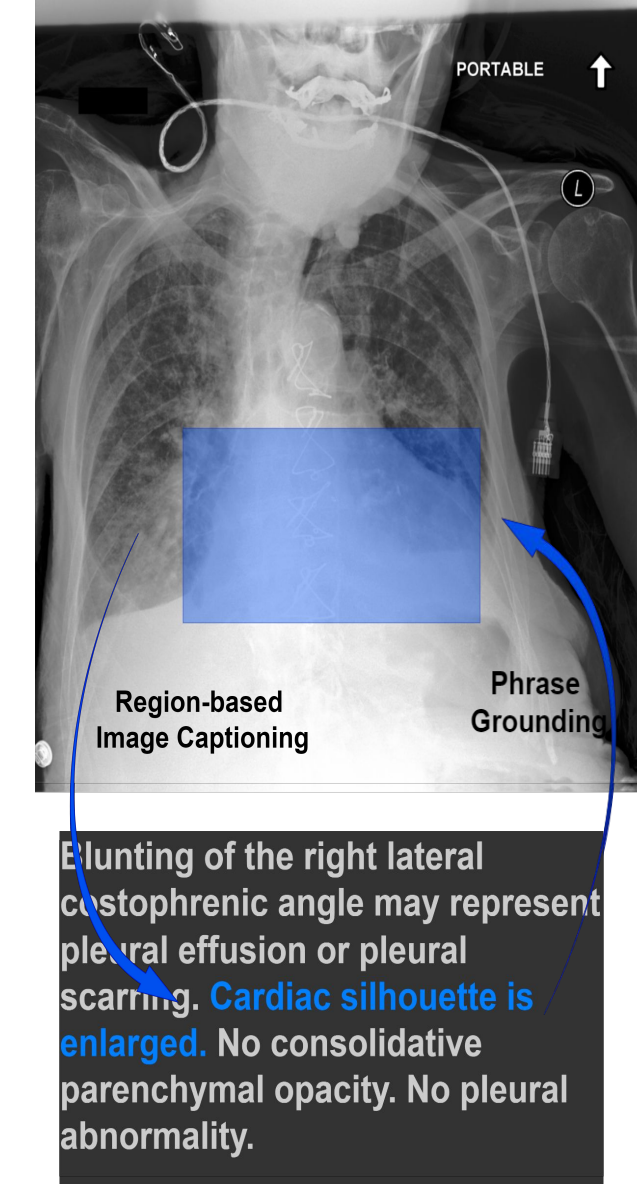
Elham Ghelichkhan, Tolga Tasdizen
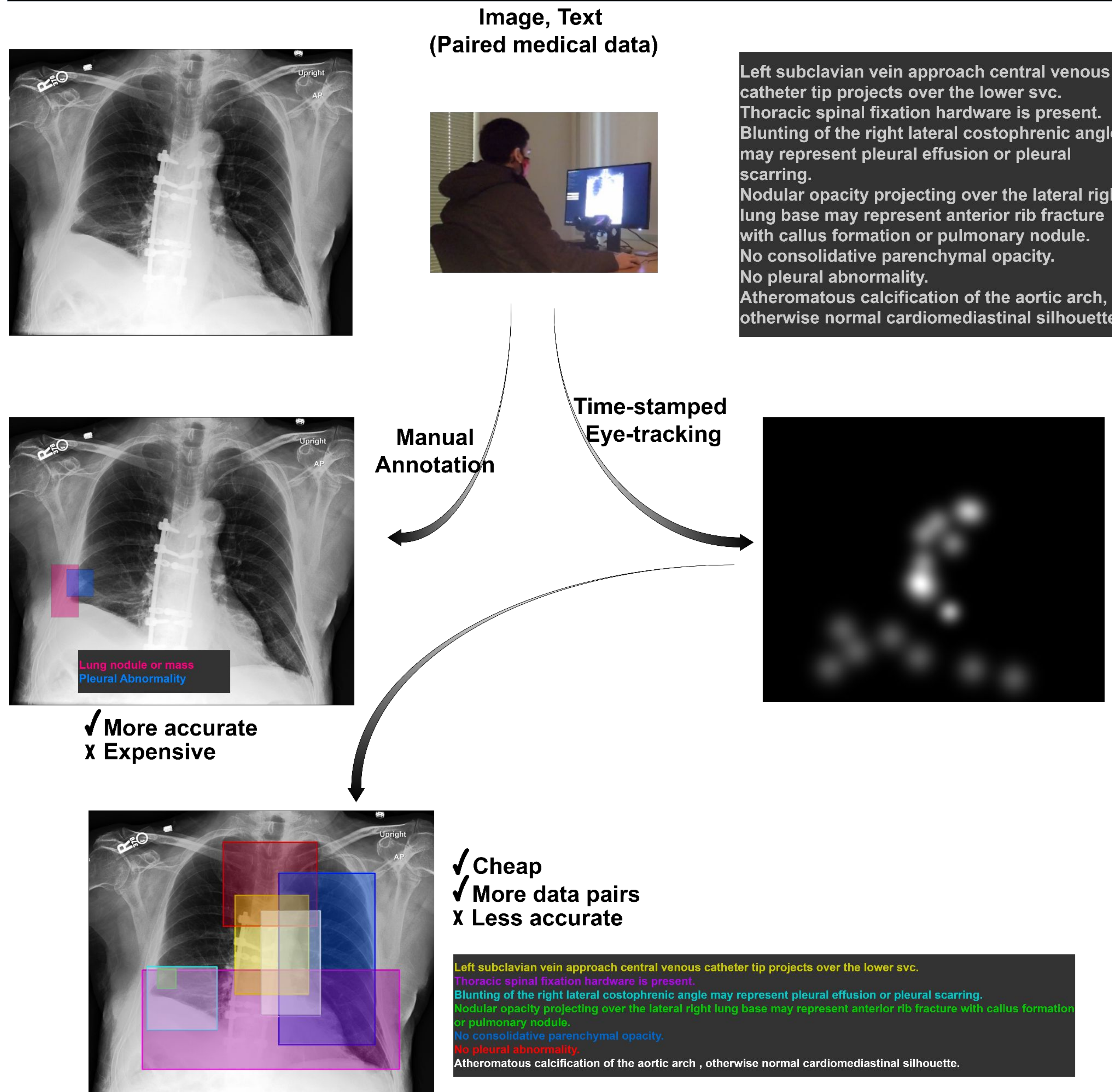
Introduction

Enhancing image-text alignment using eye-tracking data
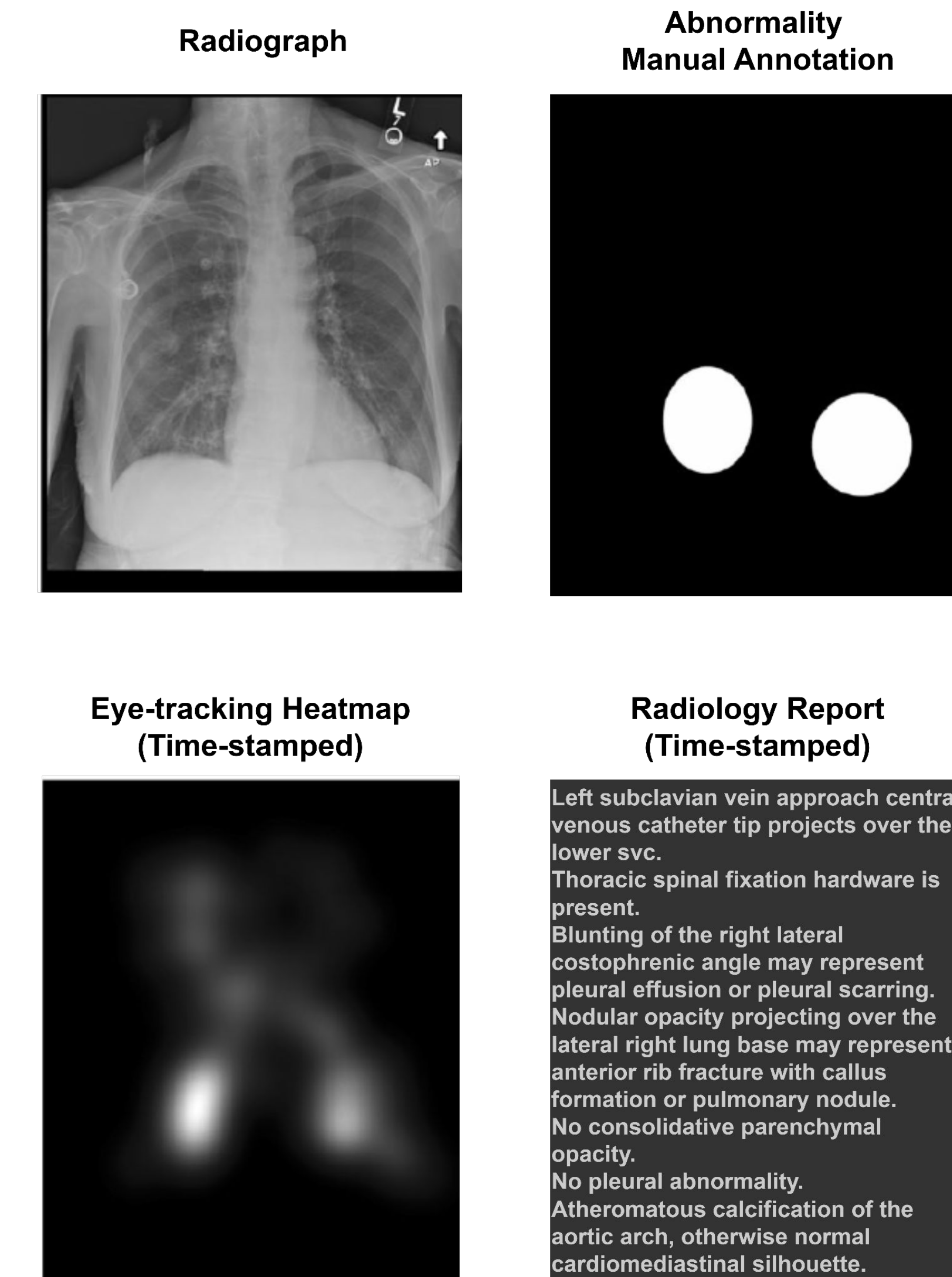
1. **Region-based image captioning-** provides initial medical interpretation given a medical image region.
2. **Phrase grounding-** provides image region where the given sentence is referring to.
   - ➤ Improving risk stratification
   - ➤ Decreasing diagnostic process time
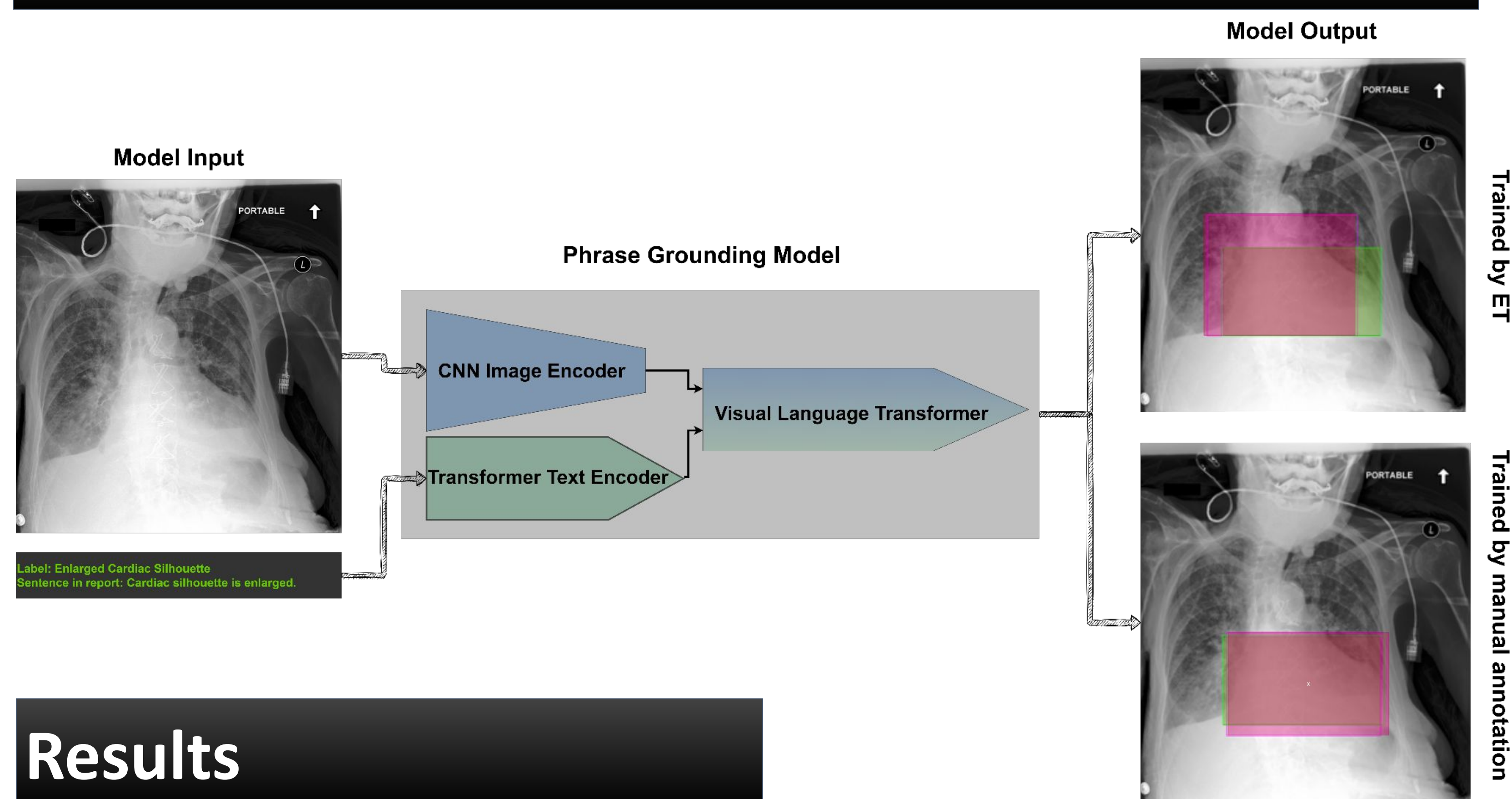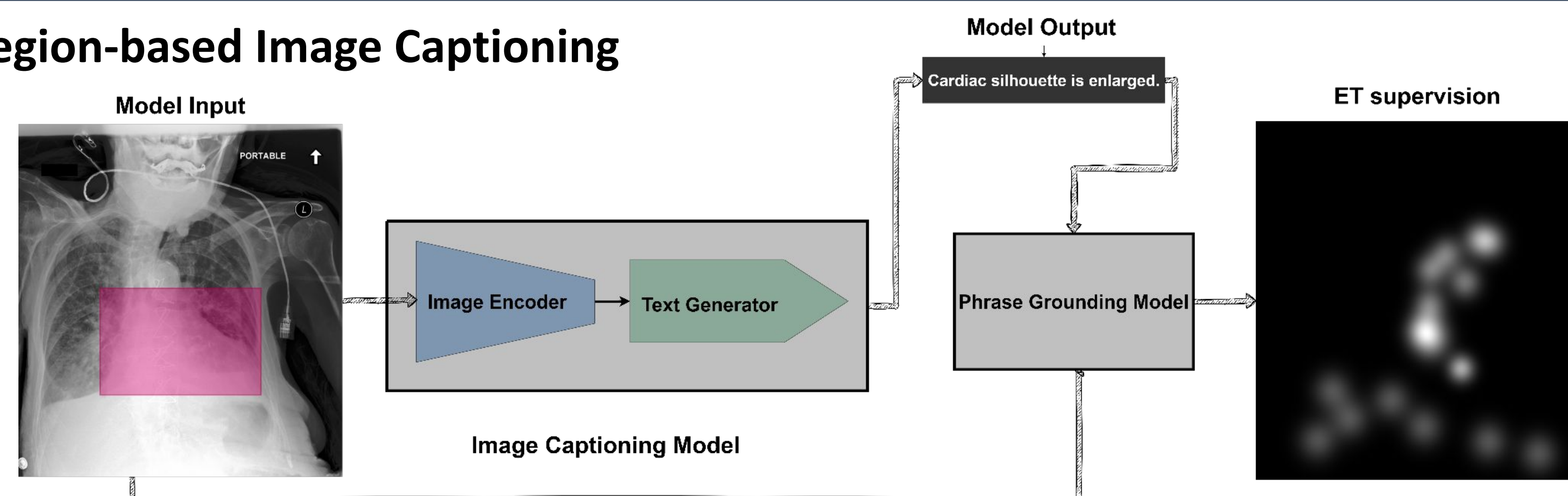   - ➤ Improving treatment planning

Methods

Model Input — Phrase Grounding Model (CNN Image Encoder, Transformer Text Encoder, Visual Language Transformer) — Model Output (Trained by ET, Trained by manual annotation)

Label: Enlarged Cardiac Silhouette
Sentence in report: Cardiac silhouette is enlarged.

Motivation

Image, Text (Paired medical data)

Left subclavian vein approach central venous catheter tip projects over the lower svc.
Thoracic spinal fixation hardware is present.
Blunting of the right lateral costophrenic angle may represent pleural effusion or pleural scarring.
Nodular opacity projecting over the lateral right lung base may represent anterior rib fracture with callus formation or pulmonary nodule.
No consolidative parenchymal opacity.
No pleural abnormality.
Atheromatous calcification of the aortic arch, otherwise normal cardiomediastinal silhouette.

Manual Annotation | **Time-stamped Eye-tracking**

✓ More accurate
✗ Expensive

Lung nodule or mass
Pleural Abnormality

✓ Cheap
✓ More data pairs
✗ Less accurate

Dataset: REFLACX

Radiograph | **Abnormality Manual Annotation**

Eye-tracking Heatmap (Time-stamped) | **Radiology Report (Time-stamped)**

Left subclavian vein approach central venous catheter tip projects over the lower svc.
Thoracic spinal fixation hardware is present.
Blunting of the right lateral costophrenic angle may represent pleural effusion or pleural scarring.
Nodular opacity projecting over the lateral right lung base may represent anterior rib fracture with callus formation or pulmonary nodule.
No consolidative parenchymal opacity.
No pleural abnormality.
Atheromatous calcification of the aortic arch, otherwise normal cardiomediastinal silhouette.

Results

Train/ Test	mIoU	Acc.
Manual bbox/ Manual bbox	45.4	66.67
ET bbox/ Manual bbox	23.65	33.2
ET bbox/ ET bbox	36.07	55.85

Future Works

➤ **Region-based Image Captioning**

Model Input — Image Captioning Model (Image Encoder, Text Generator) — Model Output: Cardiac silhouette is enlarged. — Phrase Grounding Model — ET supervision

➤ **Phrase Grounding**

Model Input — CNN Image Encoder, Transformer Text Encoder, Visual Language Transformer — ET supervision — Model Output

Label: Enlarged Cardiac Silhouette
Sentence in report: Cardiac silhouette is enlarged.
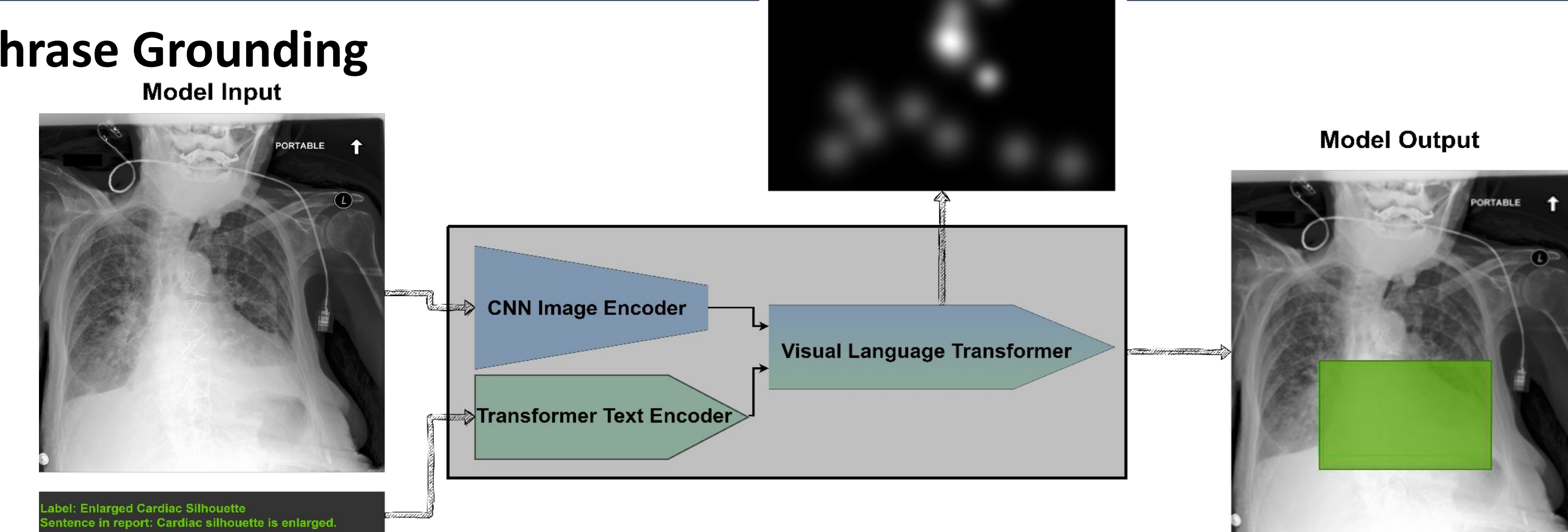
References

[1] Ricardo Bigolin Lanfredi, Mingyuan Zhang, William F Auffermann, Jessica Chan, Phuong-Anh T Duong, Vivek Srikumar, Trafton Drew, Joyce D. Schroeder, and Tolga Tasdizen. Reflacx, a dataset of reports and eye-tracking data for localization of abnormalities in chest x-rays. Scientific data, 9(1):350, 2022.
[2] Johnson, A. E. W. et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific Data 6, 317, https://doi.org/10.1038/s41597-019-0322-0 (2019).
[3] Zhihao Chen, Yang Zhou, Anh Tran, Junting Zhao, Liang Wan, Gideon Su Kai Ooi, Lionel Tim-Ee Cheng, Choon Hua Thng, Xinxing Xu, Yong Liu, et al. Medical phrase grounding with region-phrase context contrastive alignment. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 371–381. Springer, 2023.

THE UNIVERSITY OF UTAH

www.sci.utah.edu