Radiology Report Generation using Multi-Modal Prototype Network

Data Science Institute
COLUMBIA UNIVERSITY

Amrutha Varshini Sundar, Andrew Schaefer, Ayush Sinha
Prabha Kiranmai Vasireddy, Navjot Singh, Vijay Kalmath
Hemant Palivela, Sining Chen

Data Science Capstone Project
with Accenture

Introduction and Overview

Radiography is the most well established imaging method to diagnose diseases. The ability to automatically generate accurate medical reports from diagnostic images saves time and efforts of medical experts and aids the increase of healthcare accessibility.

A multi-modal deep learning model can embed extracted visual (from x-ray images) and textual (from text reports) features in common space and learn the cross-modal patterns to generate a textual report that describes the radiology image. This project analyzes the SOTA system for this task XProNet[1].

Dataset

The IU-Xray dataset contains 7,470 chest x-ray images with 3,955 corresponding medical reports published by Indiana University.

Biases:

- Most reports are normal, having no findings.
- Even in abnormal reports, affected regions only exist in small parts of the text & image.

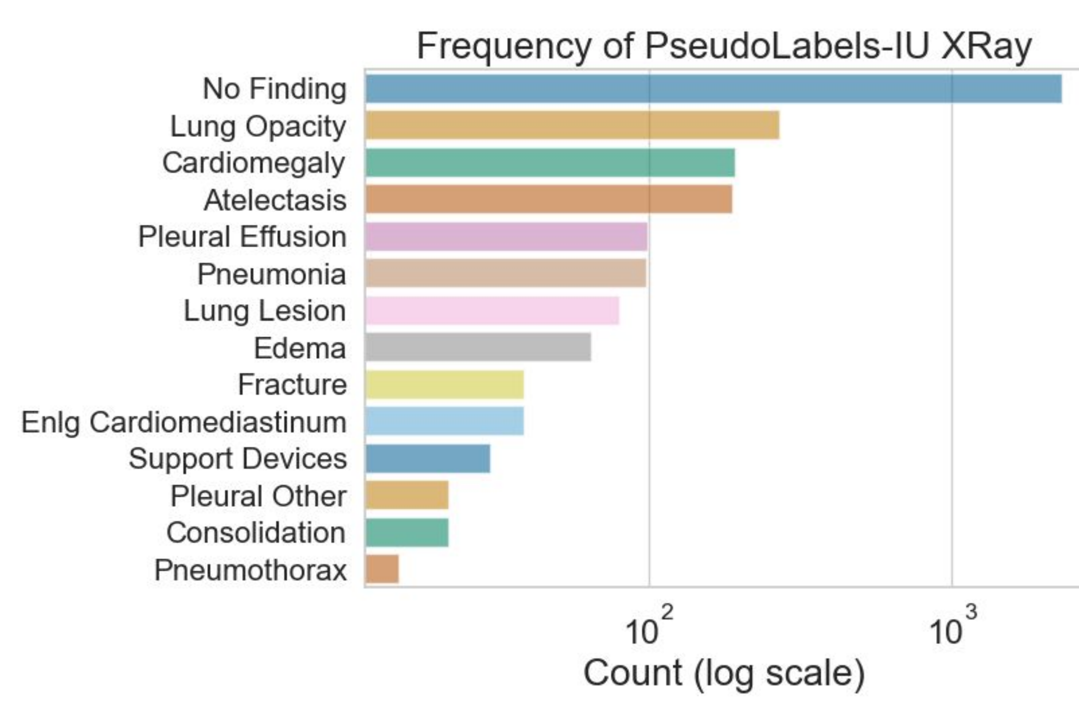
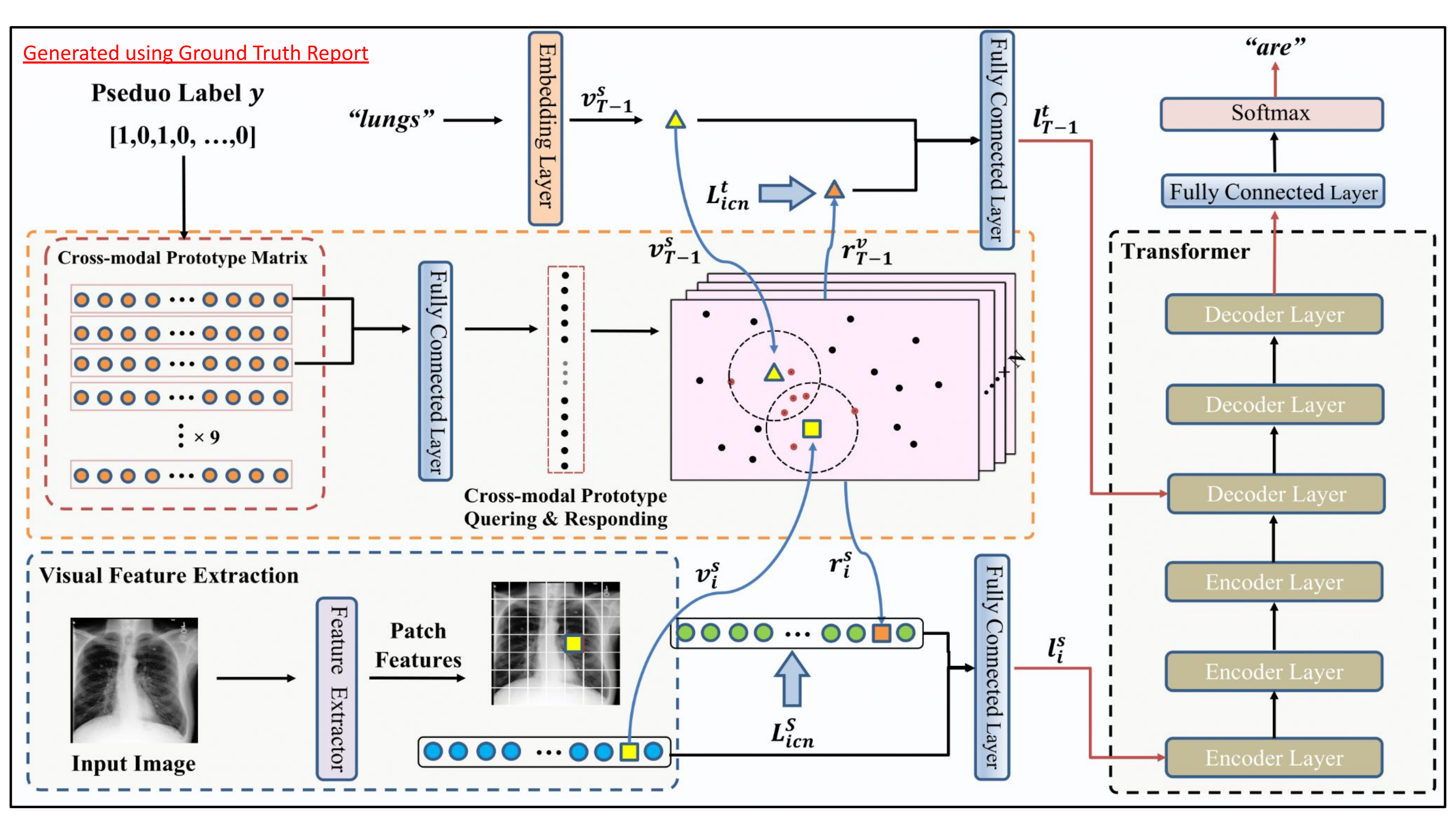
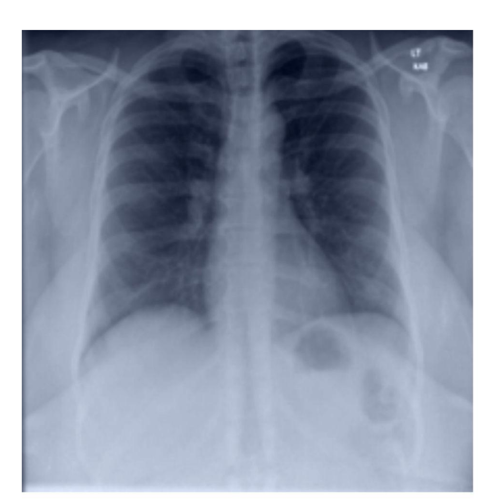


Figure 1. Frequency of disease pseudo-labels, generated using CheXPert

XProNet Architecture and Performance





Ground Truth Report:

lungs are clear bilaterally . cardiac and mediastinal silhouettes are normal . pulmonary vasculature is normal . no pneumothorax or pleural effusion . no acute bony abnormality

Epoch 1	the lungs are normal .
Epoch 10	the heart is normal in size . the mediastinum is unremarkable . the lungs are clear .
Epoch 30	heart size normal . lungs are clear . xxxx are normal . no pneumonia effusions edema pneumothorax adenopathy nodules or masses

Figure 2. XProNet Architecture (left) and Example Report (right)

Methodology	BLEU-1	BLEU-2	BLEU-3	METEOR	ROUGE-L
ResNet-101	0.3867	0.2439	0.1663	0.1677	0.3347
EfficientNet	0.4381	0.2679	0.1852	0.1680	0.3360
Balanced IU-XRay	0.2665	0.1669	0.1186	0.1255	0.3169

Table 1. Performance metrics of XProNet

Novel Inference Pipeline with Pseudo-label Generation

Currently, pseudo-labels are generated using the ground truth report, rendering XProNet non-productionizable. We built a model that generates pseudo-labels using visual features from the diagnostic image(s), achieving an average test F1-score of 0.65.

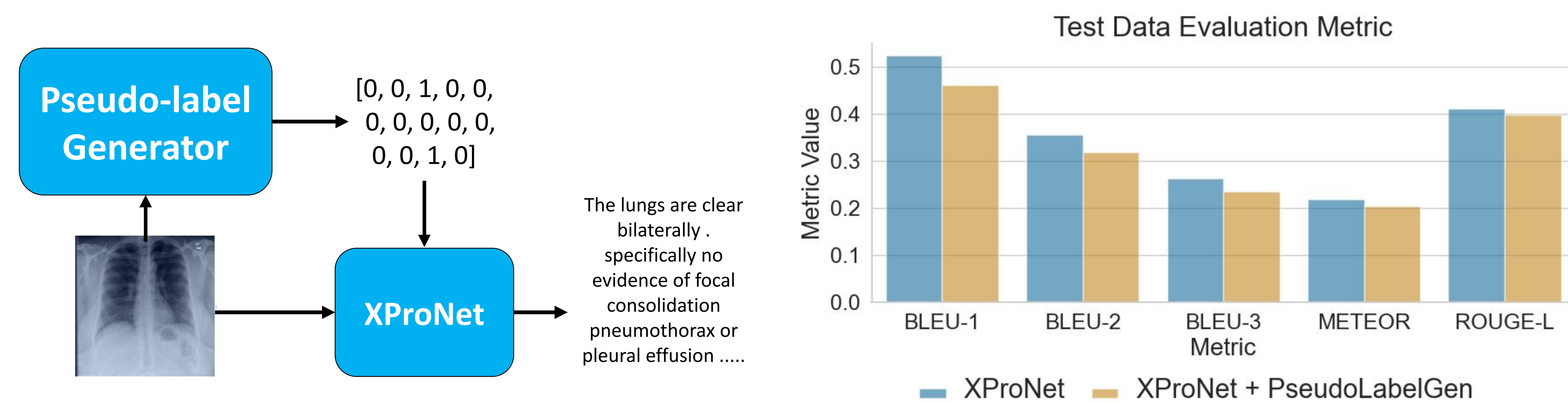


Figure 3. Inference PIpeline (left) and Performance on Test Data (right)

XProNet Drawbacks

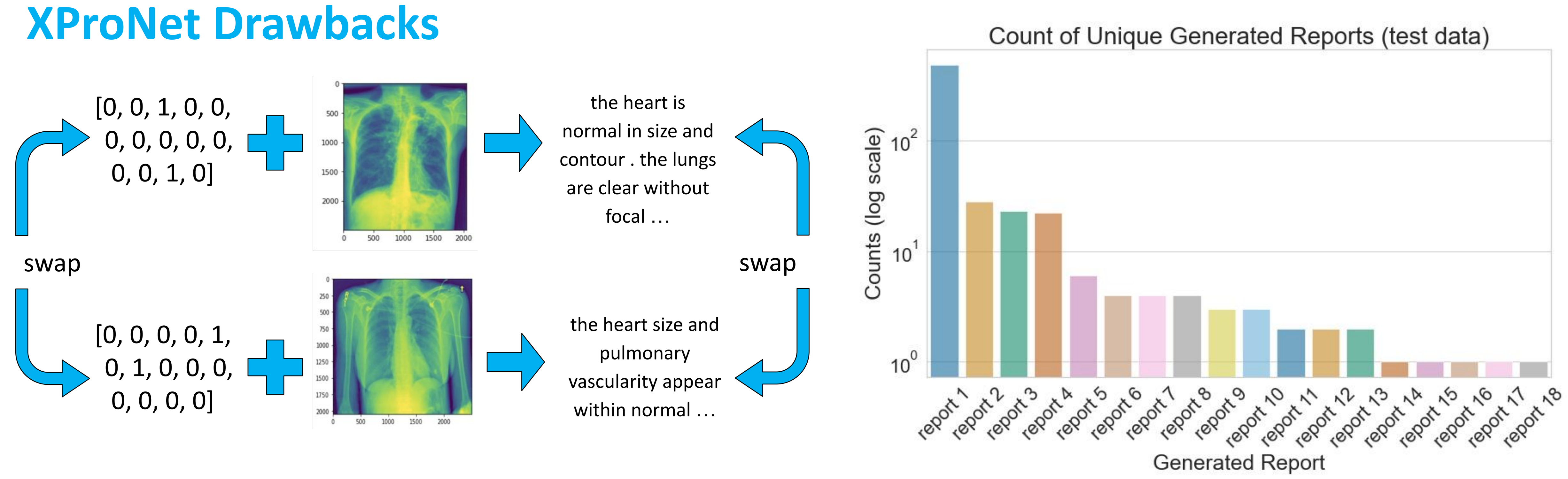


Figure 4. Pseudo-label swap illustration (left) and Frequency of unique generated reports (right)

- Each pseudo-label group tends to correspond to one unique generated report
- # Unique Generated Reports: 18 vs. # Unique Ground Truth: 396
- # Unique Words in Generated Reports: 37 vs. # Unique Words in Ground Truth: 634

Conclusion and Future Work

XProNet performs significantly better than previous SOTA and with our novel pseudo-label generator, it can now be productionized. Analysis and investigation show that XProNet suffers from over-emphasizing on the pseudo-labels in report generation. Future work is required to address this and allow for more diversification of reports.

References

[1] Wang J., Bhalerao A. and He Y.: Cross-modal Prototype Driven Network for Radiology Report Generation. In: European Conference on Computer Vision 2022. arXiv preprint arXiv:2207.04818