

Training PET/CT Encoders with Masked Language Modeling on Reports

Sabri Eyuboglu

Introduction

Acquiring well-structured labels is a significant bottleneck in developing computer vision models that encode learning for radiology. Usually, physicians manually label each exam, either by reading the accompanying report or by re-interpreting the scan. The process is painstaking and time-intensive, which limits the size and number of datasets that we can train models on. Each exam is accompanied by an imaging report written by a radiologist at the time of the study. These reports often the only record of an exams findings. This begs the question: could we bypass the labeling process and train scan encoder directly on the report itself?

We propose a novel approach for training scan encoders with reports alone. We devise two multi-modal tasks that demand strong encodings of the input scan: radiology masked language modeling and scan/report mismatch detection. Our transformer-based model then attends to both the masked report text and the image encoding to make predictions.

We perform a case-study on PET/CT exams from Stanford Hospital.

Tasks

Radiology Masked Language Modeling

IDEA: Mask words indicative of underlying pathology.

FDG uptake in the **MASK** lymph nodes

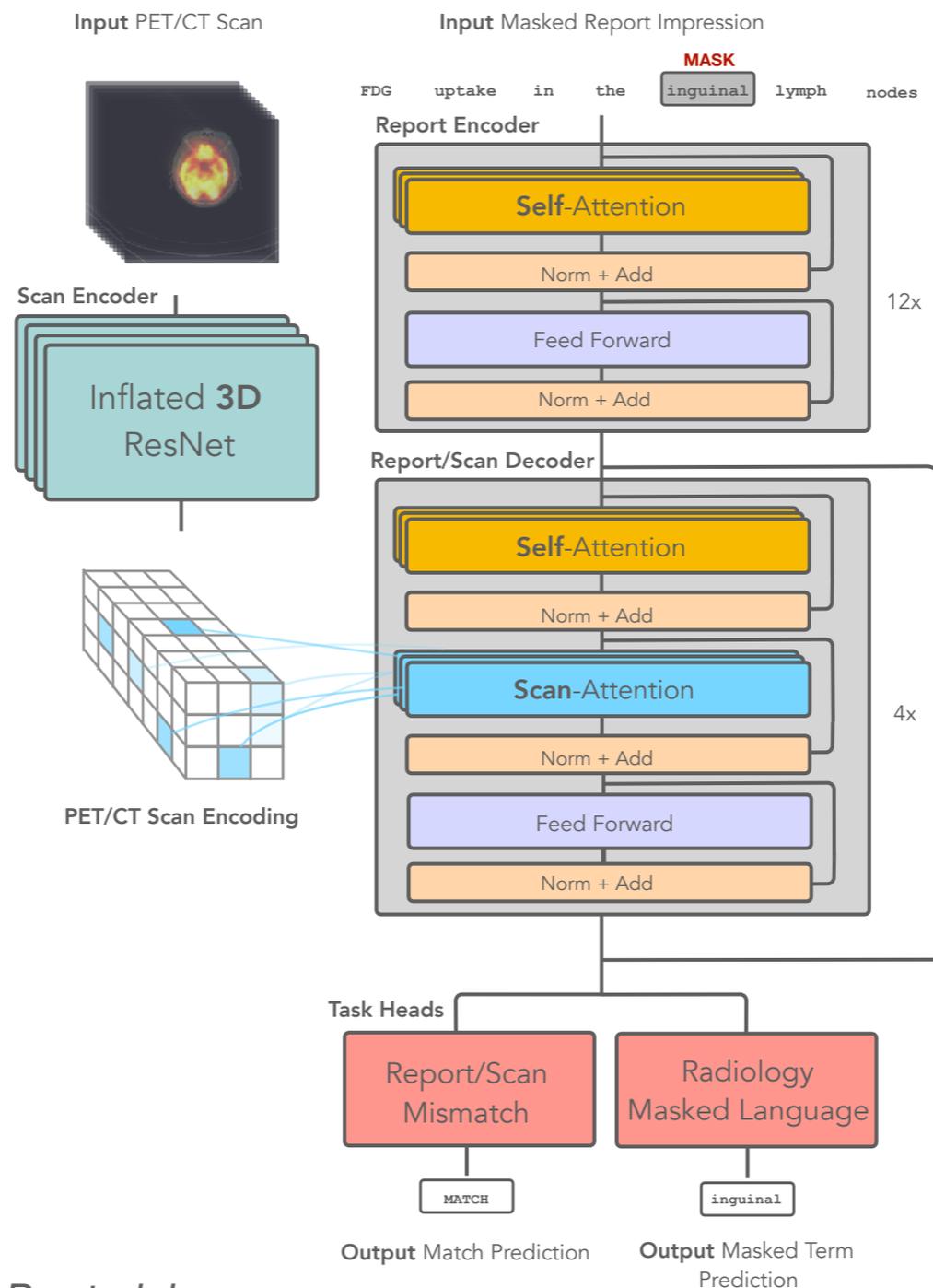
Report/Scan Mismatch Prediction

IDEA: 50% of the time, match a report with a randomly chosen exam. Task is to detect when reports are mismatched.



Collaborators: Geoff Angus, Dr. Matt Lungren, Dr. Guido Davidzon, Dr. Bhavik Patel, Yuhao Zhang, Jared Dunnmon,

Architecture

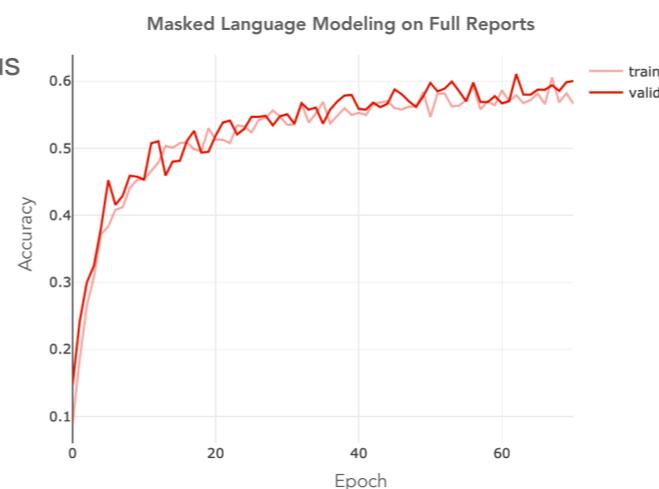


Pre-training

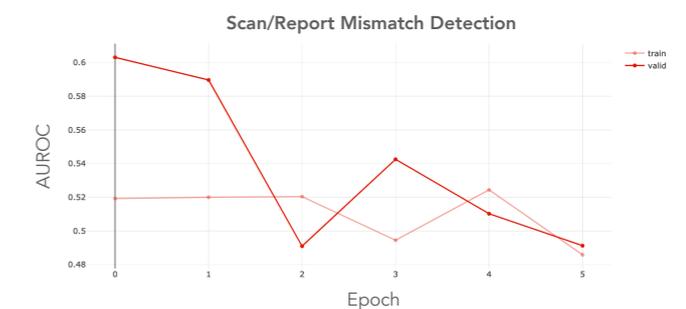
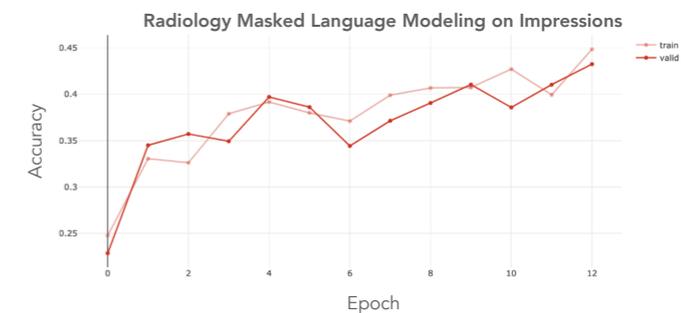
We begin with **BERT**_{base} pre-trained on BookCorpus and English Wikipedia.

ISSUE: **BERT** word-piece tokenizer splits up common words in PET/CT reports.

We train a new word-piece tokenizer, and further pre-train the model with traditional MLM on full PET/CT reports.



Quantitative Analysis



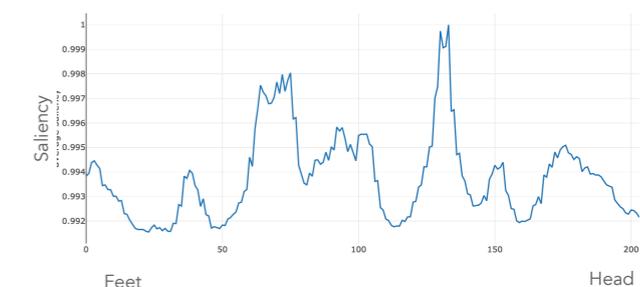
Qualitative Analysis

Error Analysis

Focus of increased radio tracer uptake in the **upper** left lobe. Focus of increased radio tracer uptake in the **upper** left lobe. Prediction: **lower**

Anterior mediastinal mass with ... Prediction: **Anterior mediastina**

Saliency Maps



Attention Probabilities from a Scan-Attention Head

