

CXR Repair Notes

▼ Main Takeaways

- ▷ Uses a contrastive learning approach to match images with text reports by creating encodings of each and then matching closest image/text pairs
- ▷ Reframes report creation task as a report retrieval task as opposed to a image captioning/text generation task by making use of the fact that there is a limited domain diagnoses
- ▷ Claims to achieve SOTA with F1 score of 0.352

▼ Previous work

- ▷ Used image encoder, text decoder setup: the main faults are that usual text generation evaluation metrics don't work well in this scenario because they don't check for accurate/proper diagnosis (instead they're looking for matching words) and reports are longer than standard image captions
- ▷ Other approaches have tried breaking down report generation into two steps: generate sentence ideas and then translate these ideas into actual sentences

▼ Setup

- ▷ Retrieves both closest sample report from corpus and closest set of sentences from corpus
- ▷ Baseline retrieval model just finds the closest sample image to input and then returns its report
- ▷ Clinical efficacy of reports is computed using an accurate diagnosis labeller model that makes diagnosis based on report; this result is compared to true diagnosis
- ▷ Also compare similarity of generated report and true report using hidden layer of labeller as an encoding
- ▷ Still use natural language metrics like BLEU as one metric among many despite the weakness of these metrics in the context of radiology report generation

▼ Data

- ▷ Use MIMIC CXR dataset to train contrastive image learning, get reports, and create baseline retrieval
- ▷ Evaluate on MIMIC CXR and CheXpert datasets
- ▷ Evaluated based on impressions section of radiology report since this is where the doctor's diagnosis is
- ▷ Qualitatively, the model produces right ideas with slightly different word choice whereas other methods produce incorrect ideas; this is because natural language metrics aren't aligned with factual accuracy
- ▷ Baseline retrieval method is already pretty good - which makes sense

▼ Variations

- ▷ Better to have contrastive learning that's pretrained on general natural language + image pairs and then finetuned on radiology reports rather than only trained on radiology reports
- ▷ Sentence approach is better than whole report approach because there is a larger space of model answers: the whole report approach requires an exact match previous patient in corpus while sentence approach allows you to pick out lines based on certain characteristics