Clinically Accurate Chest X-Ray Report Generation

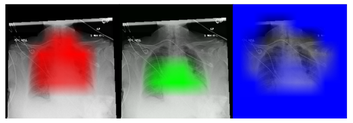
Guanxiong Liu*[1,3] Tzu-Ming Harry Hsu*[2] Matthew McDermott[2] Willie Boag[2] Wei-Hung Weng[2]
Peter Szolovits[2] Marzyeh Ghassemi[1,3]
[1]University of Toronto, [2]Massachusetts Institute of Technology, [3]Vector Institute

Full Paper

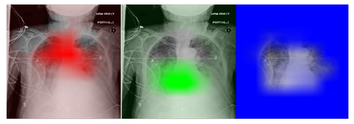
Motivation & Contributions

- A radiology report requires ***language fluency*** and ***clinical accuracy*** to be clinically helpful
- Previous works only focused on optimizing language metrics such as log-likelihood and CIDEr
- We propose a novel reward CCR to be used in reinforcement learning scenario to achieve both goals

Generated Reports and Attention Maps



ap upright and lateral views of the chest. there is moderate cardiomegaly. there is no pleural effusion or pneumothorax. there is no acute osseous abnormalities.



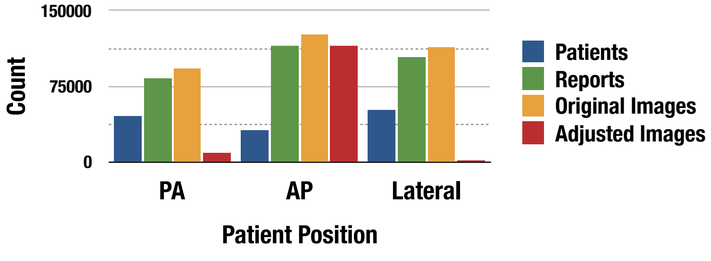
as compared to the previous radiograph, there is no relevant change. tracheostomy tube is in place. there is a layering pleural effusions. NAME bilateral pleural effusion and compressive atelectasis at the right base. there is no pneumothorax.

- CCR proves to increase clinical efficacy without sacrificing language metrics
- Attention maps are generated as a byproduct

Data: MIMIC-CXR

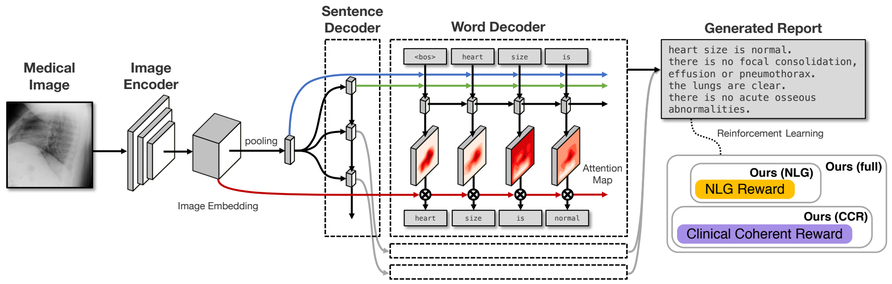
```
EXAMINATION: CHEST (PORTABLE AP)
INDICATION: History: 70M with intubated
FINDINGS:
   Endotracheal tube is in standard position. [**Doctor First
   Name **] enteric tube courses below the left hemidiaphragm
   with tip off the inferior borders of the film. …
IMPRESSION:
   1. New extensive subcutaneous [**Doctor First Name 21**]
   within the neck and chest. New right fourth and fifth rib
   fractures anteriorly. …
```

† *This chart demonstrates an internal alpha version of the dataset*



Methods: Clinically Coherent Reward

Overview



- **Image Encoder CNN** (DenseNet-121) first encodes the image as **image embedding map**
- **Sentence Decoder RNN** then generates the sentence **topics** and stop signals
- **Word Decoder RNN** generates words dependent on the **topic**, **image embedding map**, and **averaged image embedding** by the *visual sentinel model* which uses attention

Reinforcement Learning

- **Clinically Coherent Reward** evaluates the concordance of disease labels between the generated and true reports

beam search generated report / *true report* / *inner product of probabilities*

$$r_{\text{CCR}}(\mathbf{Z}, \mathbf{Z}^*) = \sum_t r_{\text{CCR},t}(\mathbf{Z}, \mathbf{Z}^*) \equiv \sum_t \sum_{s \in \{+,-\}} p(s|l_t(\mathbf{Z})) \cdot p(s|l_t(\mathbf{Z}^*))$$

for all labels / *for all disease states*

- **REINFORCE** approximates the gradient based on how much more reward the current model has over the baseline

losses to minimize / *CIDEr score* / *greedily generated report*

$$\mathcal{L}_{\text{NLG}}(\theta) = -\mathbb{E}_{(u,\mathbf{Z}) \sim p_\theta(u,\mathbf{Z})}[r_{\text{NLG}}(\mathbf{Z}, \mathbf{Z}^*) - r_{\text{NLG}}(\mathbf{Z}^g, \mathbf{Z}^*)]$$

running average of reward

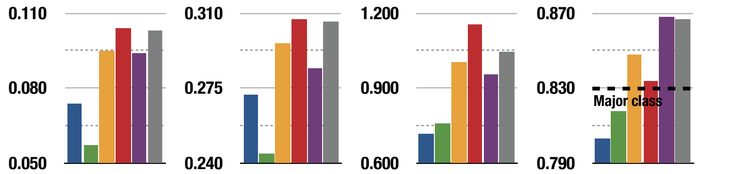
$$\mathcal{L}_{\text{CCR}}(\theta) = -\mathbb{E}_{(u,\mathbf{Z}) \sim p_\theta(u,\mathbf{Z})}\left[\sum_t r_{\text{CCR},t}(\mathbf{Z}, \mathbf{Z}^*) - \overline{r}_{\text{CCR},t}\right]$$

- The model is **(A)** first trained with conventional decoder log-likelihood (teacher forcing), **(B)** then fine-tuned with REINFORCE

Results

Automatic Evaluation on MIMIC-CXR

† *This chart corresponds to Table 2 in our paper*



statistical ***precision*** up to 4-gram

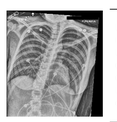
statistical ***recall*** on the longest common subsequence

TF-IDF based, ***SOTA*** NLG metric

compared to CheXpert labeler in 14 disease/devices

- **Noise-RNN:** Simple RNN language model with random initial state
- **1-NN:** The report of the most similar CXR in the training set
- **TieNet:** Trained with text decoder and classification loss
- **Ours (NLG):** Ours, but applies only the NLG reward for language fluency
- **Ours (CCR):** Ours, but applies only the CCR reward for clinical accuracy
- **Ours (full):** Ours with all the bells and whistles

Sample CXR Images and Reports

	Ground Truth	TieNet	Ours (full)
	cardiomegaly is moderate. bibasilar atelectasis is mild. there is no pneumothorax. a lower cervical spinal fusion is partially visualized. healed right rib fractures are incidentally noted.	ap portable upright view of the chest. there is no focal consolidation, effusion, or pneumothorax. the cardiomediastinal silhouette is normal. imaged osseous structures are intact.	pa and lateral views of the chest. there is mild enlargement of the cardiac silhouette. there is no pleural effusion or pneumothorax. there is no acute osseous abnormalities.
	as compared to the previous radiograph, the monitoring and support devices are unchanged. unchanged bilateral pleural effusions, with a tendency to increase, and resultant areas of atelectasis. the air collection in the bilateral soft tissues is slightly decreased. unchanged right picc line. no definite evidence of pneumothorax.	as compared to the previous radiograph, the patient has received a nasogastric tube. the course of the tube is unremarkable, the tip of the tube projects over the middle parts of the stomach. there is no evidence of complication, notably no pneumothorax. the other monitoring and support devices are constant. constant appearance of the cardiac silhouette and of the lung parenchyma.	as compared to the previous radiograph, there is no relevant change. tracheostomy tube is in place. there is a layering pleural effusions. NAME bilateral pleural effusion and compressive atelectasis at the right base. there is no pneumothorax.
	the course of the dobbhoff feeding catheter is unremarkable, and its tip is seen projecting over the stomach. there is no evidence of complications, specifically no pneumothorax. as compared to the prior radiograph dated DATE, there has been no other significant interval change.	ap portable upright view of the chest. overlying ekg leads are present. there is no focal consolidation, effusion, or pneumothorax. the cardiomediastinal silhouette is normal. imaged osseous structures are intact.	as compared to the previous radiograph, there is no relevant change. the endotracheal tube terminates approximately 3 cm above the NAME. the endotracheal tube extends into the stomach. there is no evidence of complications, notably no pneumothorax. there is no pleural effusion or pneumothorax.
	interval placement of a left basilar pigtail chest tube with improving aeration in the left mid to lower lung and near complete resolution of the pleural effusion. there are residual patchy opacities within the left mid and lower lung as well as at the right base favoring resolving atelectasis. no pneumothorax is appreciated on this semi upright study. heart remains stably enlarged. mediastinal contours are stably widened, although this NAME be related to portable technique and positioning. this can be better evaluated on followup imaging. no pulmonary edema.	as compared to the previous radiograph, the patient has been extubated. the nasogastric tube is in unchanged position. the lung volumes remain low. moderate cardiomegaly with minimal fluid overload but no overt pulmonary edema. no larger pleural effusions. no pneumonia.	ap upright and lateral views of the chest. there is moderate cardiomegaly. there is no pleural effusion or pneumothorax. there is no acute osseous abnormalities.

References

[Visual Sentinel] Lu et al. Knowing when to look: Adaptive attention via a visual sentinel for image captioning. In CVPR 2017.

[TieNet] Wang et al. TieNet: Text-image embedding network for common thorax disease classification and reporting in chest X-rays. In CVPR, 2018.