Unsupervised Multimodal Representation Learning across Medical Images and Reports

Tzu-Ming Harry Hsu, Wei-Hung Weng, Willie Boag, Matthew McDermott, and Peter Szolovits

{stmharry,ckbjimmy,wboag,mmd,psz}@mit.edu

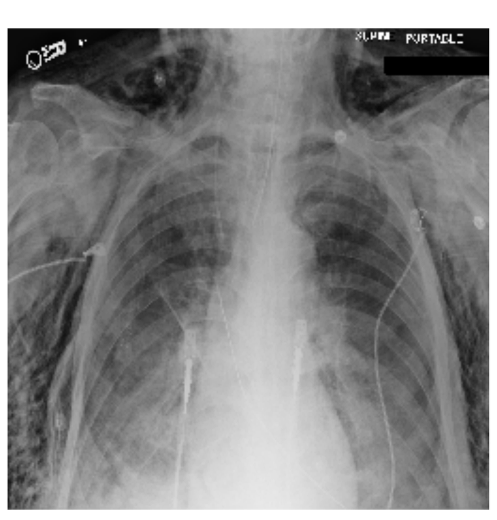
Motivation & Contributions

- Medical **reports** and **images** have been explored in the form of report generation, image annotation, image generation, and joint representation learning
- Parallel image/report pairs are not always feasible
- We investigate the effect of using semi-supervised algorithms in learning joint embedding spaces on the **MIMIC-Chest X-ray*** dataset
- We show that, on large scale, unsupervised methods achieve comparable results on the metrics for retrieval

MIMIC Chest X-ray Dataset

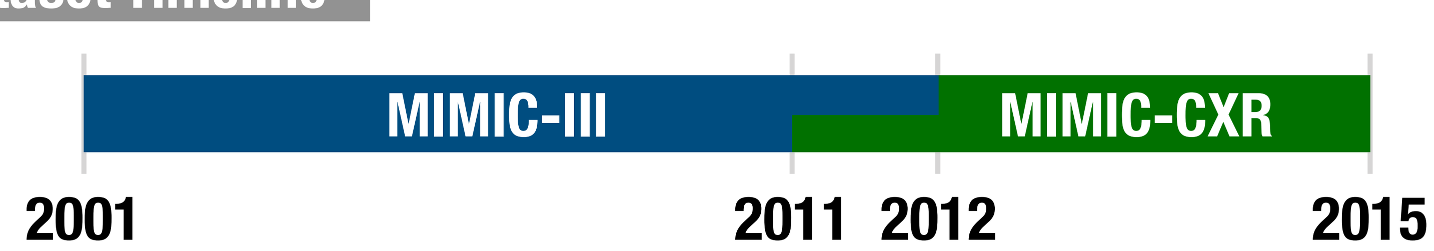
- The MIMIC Chest X-ray (MIMIC-CXR) consists of 473,057 chest X-ray images and 206,563 corresponding radiology reports from 63,478 patients admitted to critical care units at Beth Israel Deaconess Medical Center.

Paired Image and Report

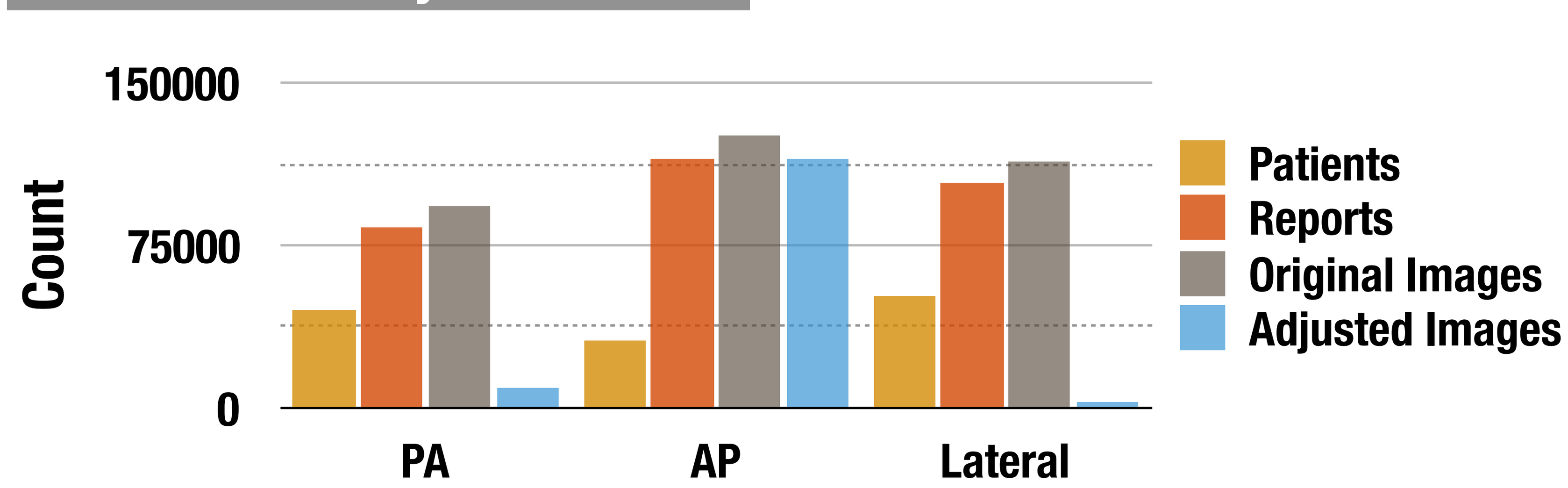


```
EXAMINATION: CHEST (PORTABLE AP)
INDICATION: History: 70M with intubated
FINDINGS:
  Endotracheal tube is in standard position.
  [**Doctor First Name **] enteric tube courses
  below the left hemidiaphragm with tip off the
  inferior borders of the film. …
IMPRESSION:
  1. New extensive subcutaneous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

Dataset Timeline

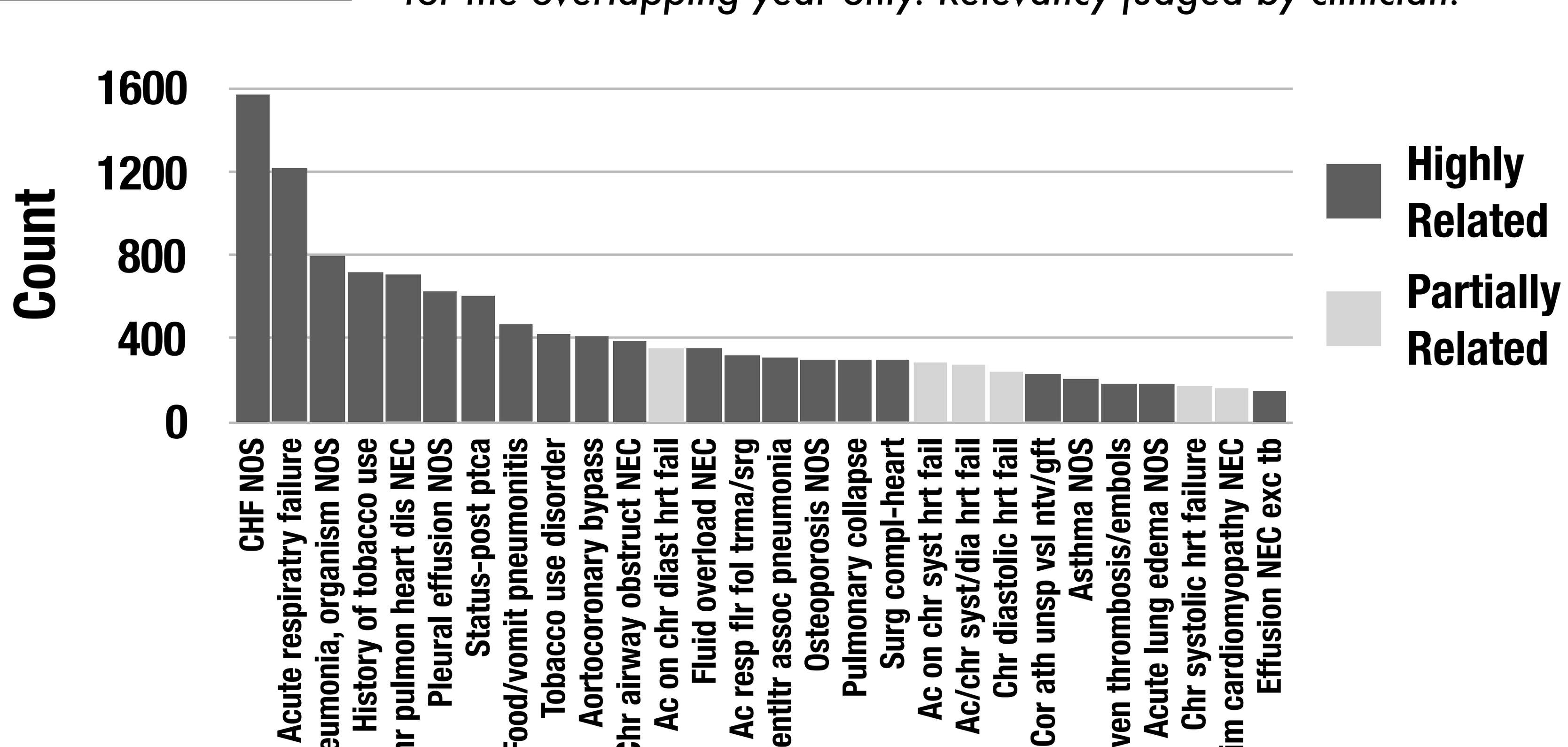


MIMIC-CXR X-ray Breakdown

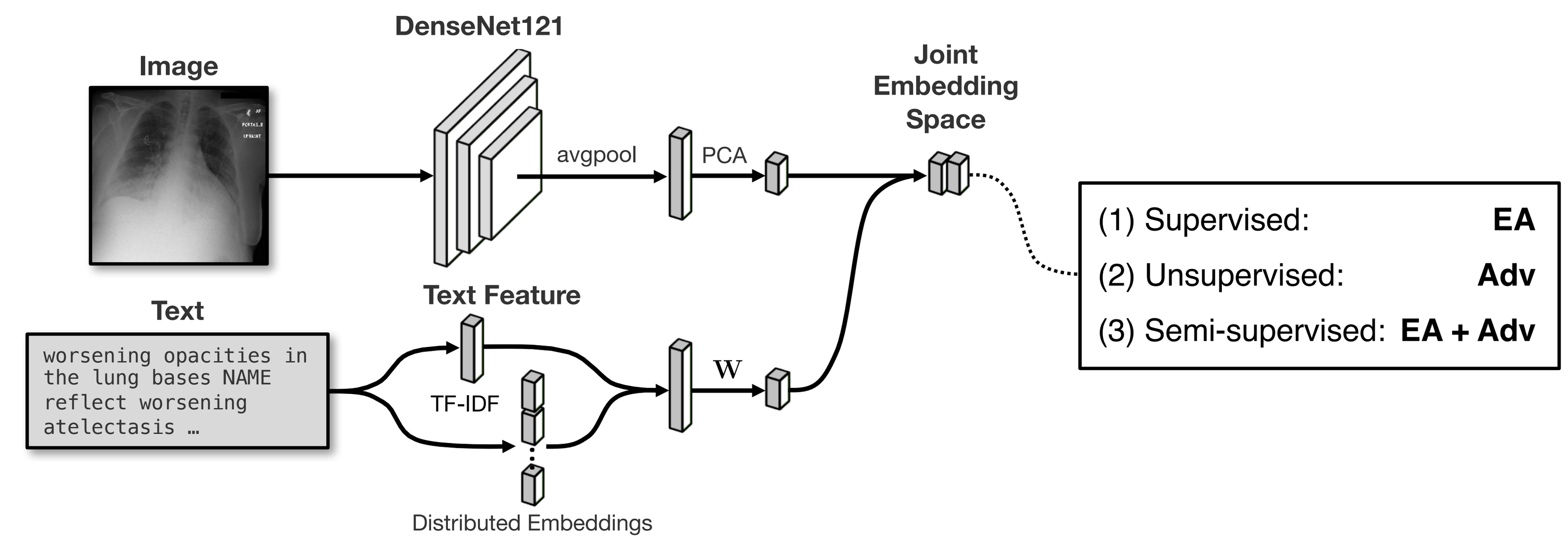


ICD-9 Codes

*sourced from MIMIC-III for patients with MIMIC-CXR records for the overlapping year only. Relevancy judged by clinician.



Methods



- Text Features $\mathbf{X}$: TF-IDF, GloVe embedding, DAN [Cer] sentence/paragraph embedding
- Embedding Alignment (EA)

$$\mathcal{L}_{\text{EA}}\left(\mathbf{X},\mathbf{Y}\right)=\left\|\mathbf{W}^{\top}\mathbf{X}-\mathbf{Y}\right\|_F^2$$

- Adversarial Domain Adaption (Adv)

$$\mathcal{L}_{\text{Adv}}^{D}\left(\mathbf{X},\mathbf{Y}\right)=\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim p(\mathbf{X},\mathbf{Y})}\left[-\log \underbrace{D}_{\text{discriminator}}\left(\mathbf{W}^{\top}\mathbf{x}\right)-\log\left(1-D\left(\mathbf{y}\right)\right)\right]$$

$$\mathcal{L}_{\text{Adv}}^{\mathbf{W}}\left(\mathbf{X},\mathbf{Y}\right)=\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim p(\mathbf{X},\mathbf{Y})}\left[-\log\left(1-D\left(\underbrace{\mathbf{W}^{\top}}_{\text{generator}}\mathbf{x}\right)\right)\right]$$

- Procrustes Refinement (Adv + Proc) [Grave]

$$\mathcal{L}_{\text{Proc}}\left(\mathbf{X},\mathbf{Y}\right)=\left\|\mathbf{W}^{\top}\mathbf{X}-\underbrace{\mathbf{P}}_{\text{correspondence}}\mathbf{Y}\right\|_F^2$$

- Orthogonal Regularization

$$\mathcal{R}_{\text{ortho}}=\beta\left\|\mathbf{W}^{\top}\mathbf{W}\odot\underbrace{\left(\mathbf{e}\mathbf{e}^{\top}-\mathbf{I}\right)}_{\text{off-diagonal terms}}\right\|_F^2$$

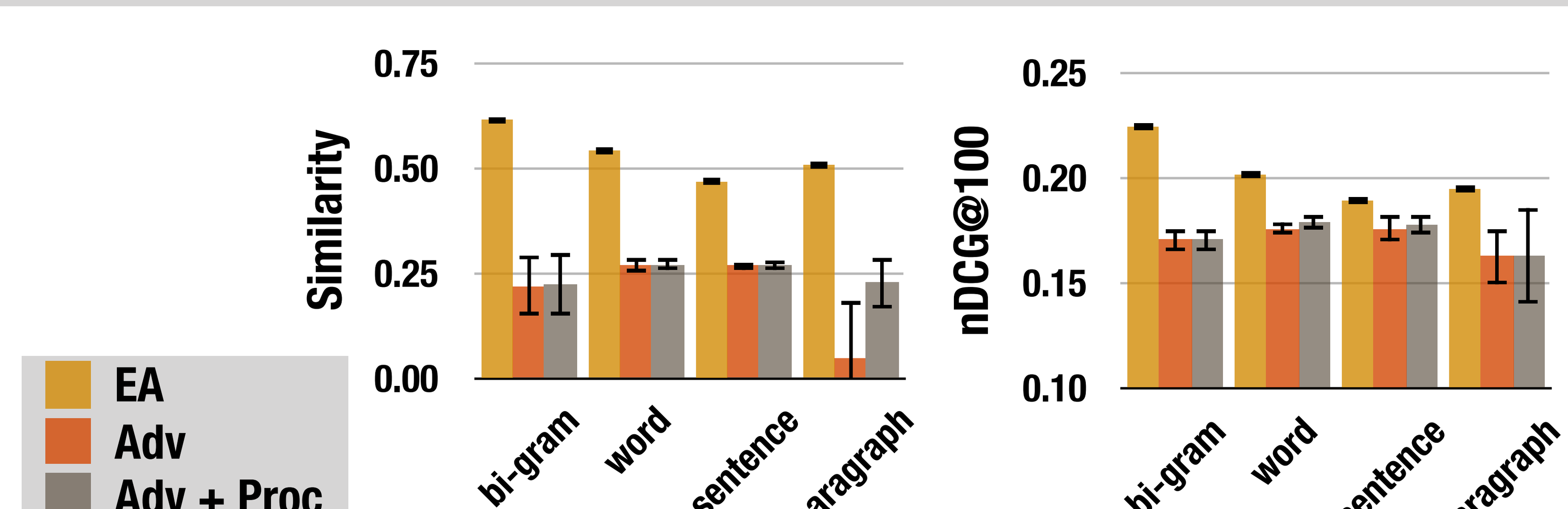
- Metrics for Retrieval Tasks

$$\text{MRR}=\frac{1}{|Q|}\sum_{q\in Q}\frac{1}{\underbrace{\text{rank}_q}_{\text{rank of the first related item}}}$$

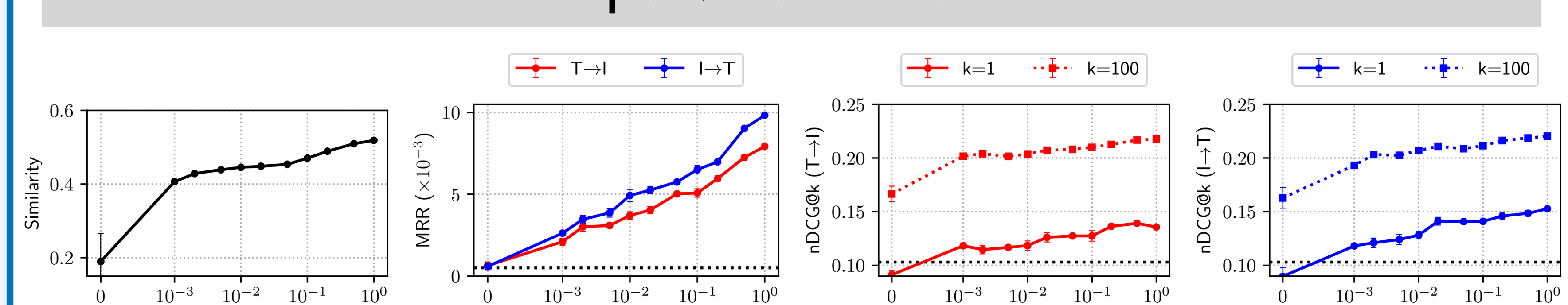
$$\text{nDCG@k}=\frac{1}{|Q|}\sum_{q\in Q}\frac{1}{\text{IDCG}_q}\sum_{p=1}^{k}\frac{2^{\overbrace{\text{rel}_{pq}}^{\text{relation coefficient in [0, 1]}}}-1}{\log_2(p+1)}$$

Results

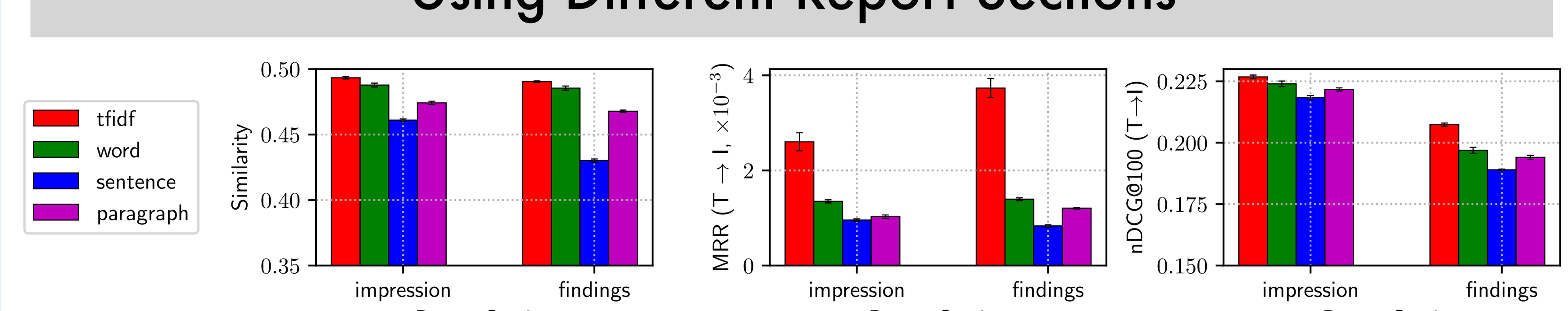
Supervision and Text Feature



Supervision Fraction



Using Different Report Sections



References

Grave et al. Unsupervised alignment of embeddings with Wasserstein procrustes. In arXiv, 2018.

Cer et al. Universal sentence encoder. In arXiv, 2018.