

Semantic Scholar

MedICaT: A Dataset of Medical Images, Captions, and Textual References

Sanjay Subramanian, Lucy Lu Wang, Sachin Mehta, Ben Bogin, Madeleine van Zuylen, Sravanthi Parasa, Sameer Singh, Matt Gardner, Hannaneh Hajishirzi



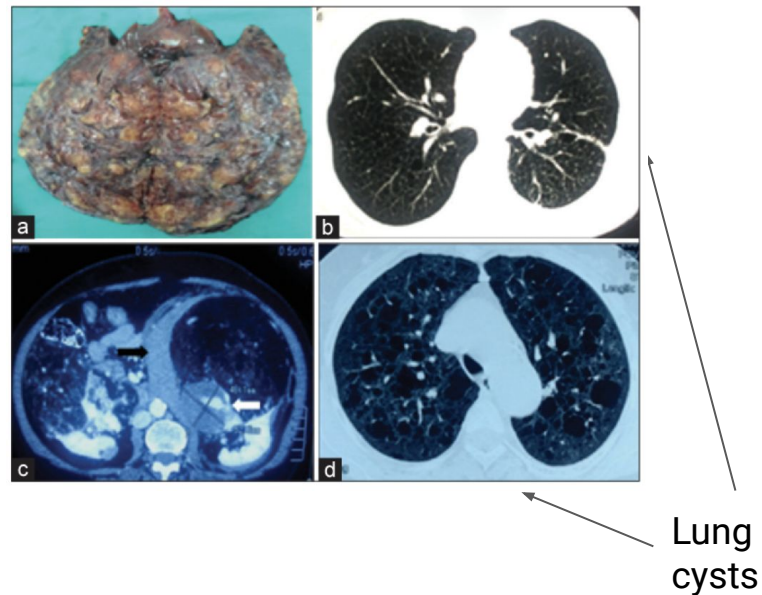
Papers are a useful source of medical images

- How can images in medical papers be useful?



Papers are a useful source of medical images

- How can images in medical papers be useful?
 - Retrieving images representative of a specific medical concept
 - Additional training data for medical image classification



Papers are a useful source of medical images

- How can images in medical papers be useful?
 - Retrieving images representative of a specific medical concept
 - Additional training data for medical image classification
 - Research paper reading experience - show figures with associated text

Corpus ID: 222310689

MedICaT: A Dataset of Medical Images, Captions, and Textual References

Sanjay Subramanian, Lucy Lu Wang, +6 authors Hamaneh Hajjshirzi · Published 2020 · Computer Science · ArXiv

Understanding the relationship between figures and text is key to scientific document understanding. Medical figures in particular are quite complex, often consisting of several subfigures (75% of figures in our dataset), with detailed text describing their content. Previous work studying figures in scientific papers focused on classifying figure content rather than understanding how images relate to the text. To address challenges in figure retrieval and figure-to-text alignment, we introduce... CONTINUE READING

[View PDF on arXiv](#) [Save to Library](#) [Create Alert](#) [Cite](#) [Launch Research Feed](#)

ABSTRACT FIGURES AND TABLES 24 REFERENCES RELATED PAPERS

Figures and Tables from this paper.

Figures and Tables



Figure 1

Dataset Statistics	MedICaT
Number of papers	131,048
Number of figures	231,980
Avg. length (words)	1.2
Avg. length (img. pixels)	2.6
Avg. width (img. pixels)	74.2
Avg. height (img. pixels)	391
Avg. reference length (tokens)	40.3
Avg. longest matching (tokens) w/ ref.	238
Avg. longest matching (tokens) w/ ref. & text	398
Fig. without captions	916
Fig. with subfigures*	376

Table 1



Figure 2

Model	Validation Top 1	Top 5
ResNet + alignment baseline	49.0	69.9
ResNet + alignment + ResNet + alignment	79.1	93.6
ResNet	-	-
Caption reference results		
ResNet + CTR + nearest neighbor	47.6	63.9
ResNet + CTR + nearest neighbor + CTR	71.6	88.6
ResNet + embedding + CTR	74.7	91.9
Using modified subfigures		
ResNet + CTR + nearest neighbor	46.1	60.0
ResNet + CTR + nearest neighbor + CTR	66.4	81.1
ResNet + embedding + CTR	69.9	83.2

Table 2

[View All 7 Figures & Tables](#)



The MedICaT Dataset

- Subfigure-subcaption annotations - 2000+ figures
 - Previous work -- e.g. compound figure segmentation (De Herrera et al., 2016) and subfigure labeling (You et al., 2011) -- is more limited

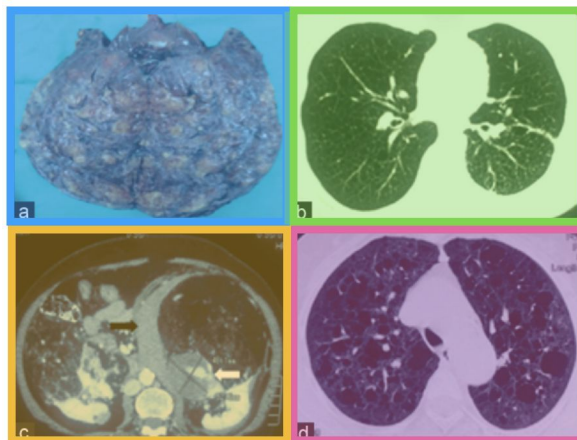
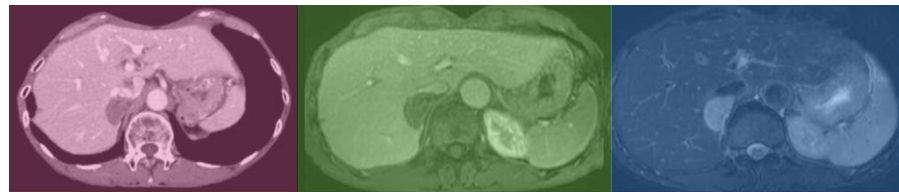


Figure 1: (a) Right renal angiomyolipoma (gross specimen postexcision). (b) High-resolution computed tomography chest images of Case 1 showing multiple variable sized cysts uniformly scattered in both lungs. (c) Computed tomography abdomen showing bilateral renal angiomyolipomas with fat densities, tortuous vessels, and pseudoaneurysm (white arrow). There is also the presence of perinephric hematoma (black arrow). (d) High-resolution computed tomography image of Case 2 showing bilateral lung cysts.



The MedICaT Dataset

- Subfigure-subcaption annotations - 2000+ figures
 - Previous work -- e.g. compound figure segmentation (De Herrera et al., 2016) and subfigure labeling (You et al., 2011) -- is more limited



The tumor (approximately 40mm in diameter) was hypovascular on enhanced computed tomography scan (right), indicated low intensity on T1-weighted MRI (center), and high intensity on T2-weighted or diffusion MRI (left). Dynamic study revealed peripheral enhancement on a late phase. The tumor located close to the inferior vena cava. MRI = magnetic resonance imaging.



The MedICaT Dataset

- Subfigure-subcaption annotations - 2000+ figures
 - Previous work -- e.g. compound figure segmentation (De Herrera et al., 2016) and subfigure labeling (You et al., 2011) -- is more limited
- Inline references - 150K+ figures
 - Provide additional context and descriptions for images

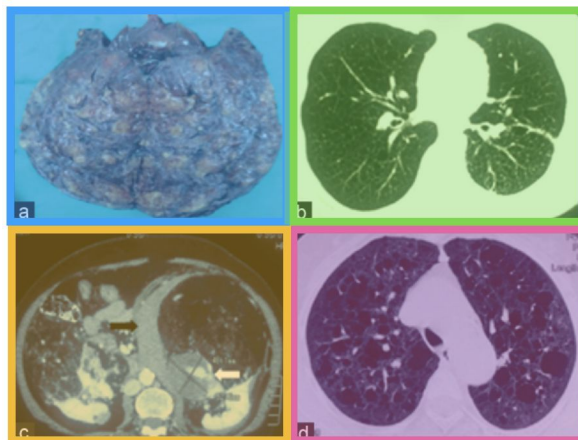


Figure 1: (a) Right renal angiomyolipoma (gross specimen postexcision). (b) High-resolution computed tomography chest images of Case 1 showing multiple variable sized cysts uniformly scattered in both lungs. (c) Computed tomography abdomen showing bilateral renal angiomyolipomas with fat densities, tortuous vessels, and pseudoaneurysm (white arrow). There is also the presence of perinephric hematoma (black arrow). (d) High-resolution computed tomography image of Case 2 showing bilateral lung cysts.



Corresponding inline reference:

"CT thorax revealed multiple variable sized thin-walled cysts scattered throughout the lung parenchyma bilaterally [**Figure 1d**]."



Construction of MedICaT

1. Identify open-access publications in PubMed Central



Construction of MedICaT

1. Identify open-access publications in PubMed Central
2. Collect captions and references for figures using S2ORC (Lo et al., 2020)

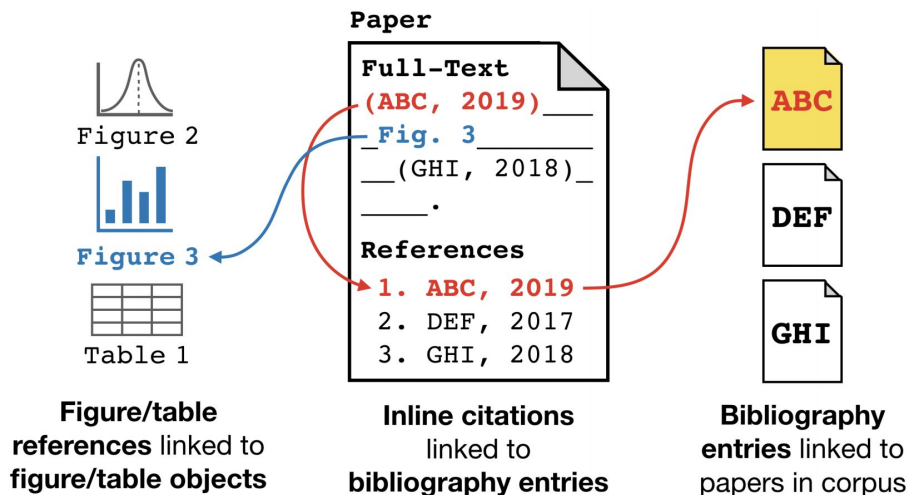
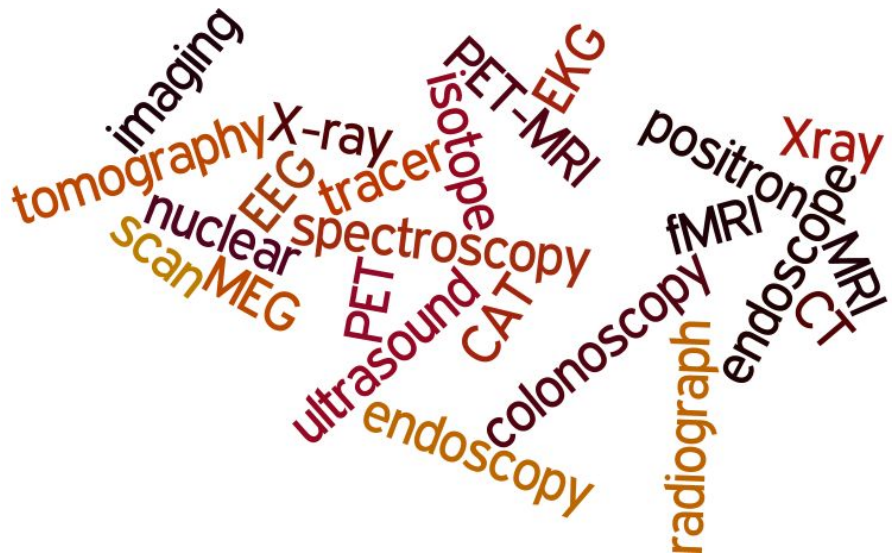


Image Source: Lo et al., 2020



Construction of MedICaT

1. Identify open-access publications in PubMed Central
2. Collect captions and references for figures using S2ORC (Lo et al., 2020)
3. Filter figures using keyword filters on captions



Construction of MedICaT

1. Identify open-access publications in PubMed Central
2. Collect captions and references for figures using S2ORC (Lo et al., 2020)
3. Filter figures using keyword filters on captions
4. Filter figures using a figure-type classifier trained on DocFigure (Jobin et al., 2019) so we are left with medical figures.

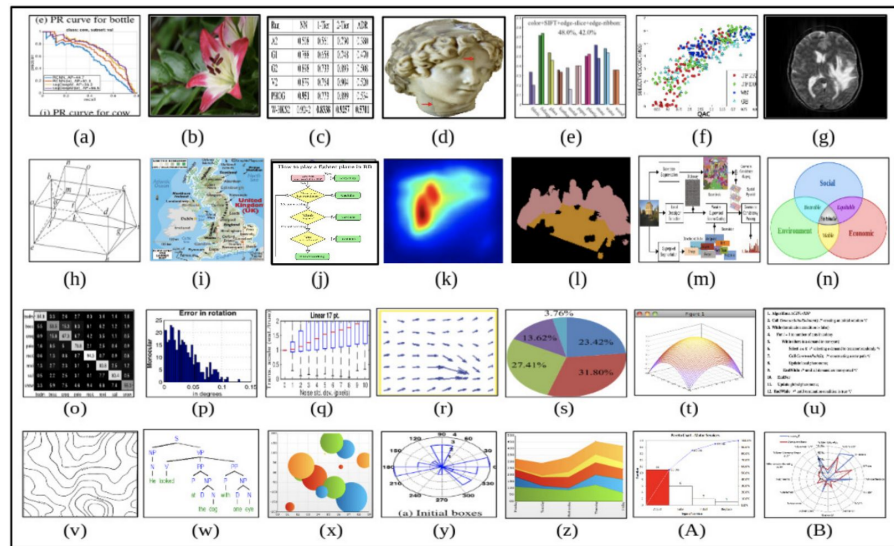
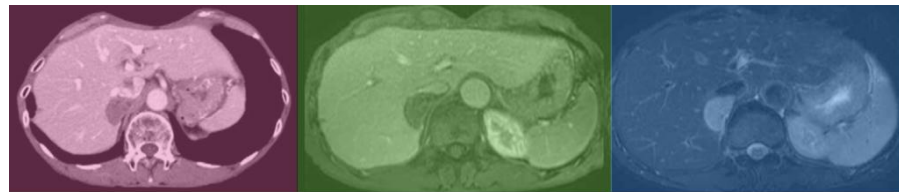


Image Source: Jobin et al., 2019



Subfigure-subcaption Alignment

- **Input:** A compound figure and the corresponding full caption
- **Output:** A box around each subfigure and the subcaption corresponding to each subfigure
- Challenges with how subfigures are referenced -- by letter, by spatial position, shared subcaptions, etc.
- Annotations: 7507 subcaptions for 2069 compound figures
 - Varied medical knowledge among annotators, but no physicians



The tumor (approximately 40mm in diameter) was hypovascular on enhanced computed tomography scan (right), indicated low intensity on T1-weighted MRI (center), and high intensity on T2-weighted or diffusion MRI (left). Dynamic study revealed peripheral enhancement on a late phase. The tumor located close to the inferior vena cava. MRI = magnetic resonance imaging.

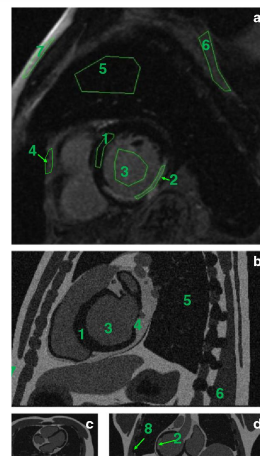


Fig. 3 Assignment of tissue values. ROIs were delineated in a clinical MR (a). Similar contrast values were imposed to the tissues in the attenuation image generated with the XCAT phantom. Panes b, c and d represent the sagittal, transaxial and axial views of the XCAT phantom, respectively, after conversion to MR-like contrast values

Image Sources: Ohkura et al., 2015;
Turco et al., 2016



Subfigure-subcaption Alignment Models

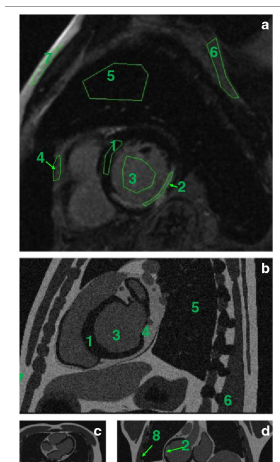
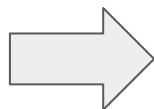


Fig. 3 Assignment of tissue values. ROIs were delineated in a clinical MR (a). Similar contrast values were imposed to the tissues in the attenuation image generated with the XCAT phantom. Panes b, c and d represent the sagittal, transaxial and axial views of the XCAT phantom, respectively, after conversion to MR-like contrast values

Text-only CRF Tagger



Faster R-CNN



SciBERT+CRF



Subfigure-subcaption Alignment Models

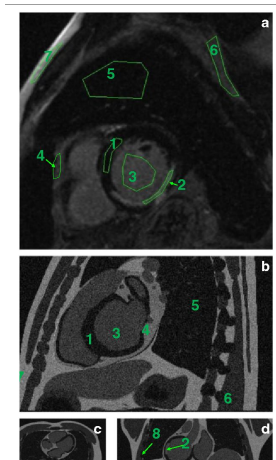
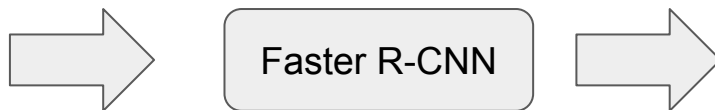
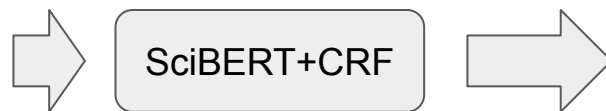
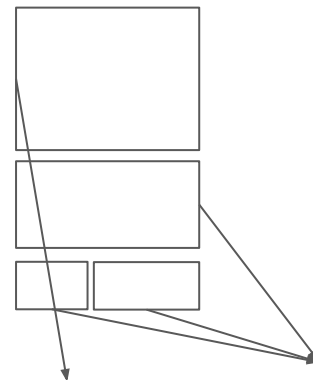


Fig. 3 Assignment of tissue values. ROIs were delineated in a clinical MR (a). Similar contrast values were imposed to the tissues in the attenuation image generated with the XCAT phantom. Panes b, c and d represent the sagittal, transaxial and axial views of the XCAT phantom, respectively, after conversion to MR-like contrast values

Text-only CRF Tagger



Alignment heuristic



ROIs were delineated in a clinical MR (a). Similar contrast values were imposed to the tissues in the attenuation image generated with the XCAT phantom.

Panes b, c and d represent the sagittal, transaxial and axial views of the XCAT phantom, respectively, after conversion to MR-like contrast values



Subfigure-subcaption Alignment Models

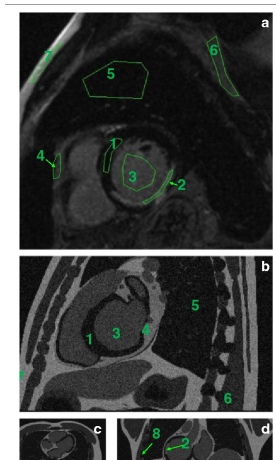
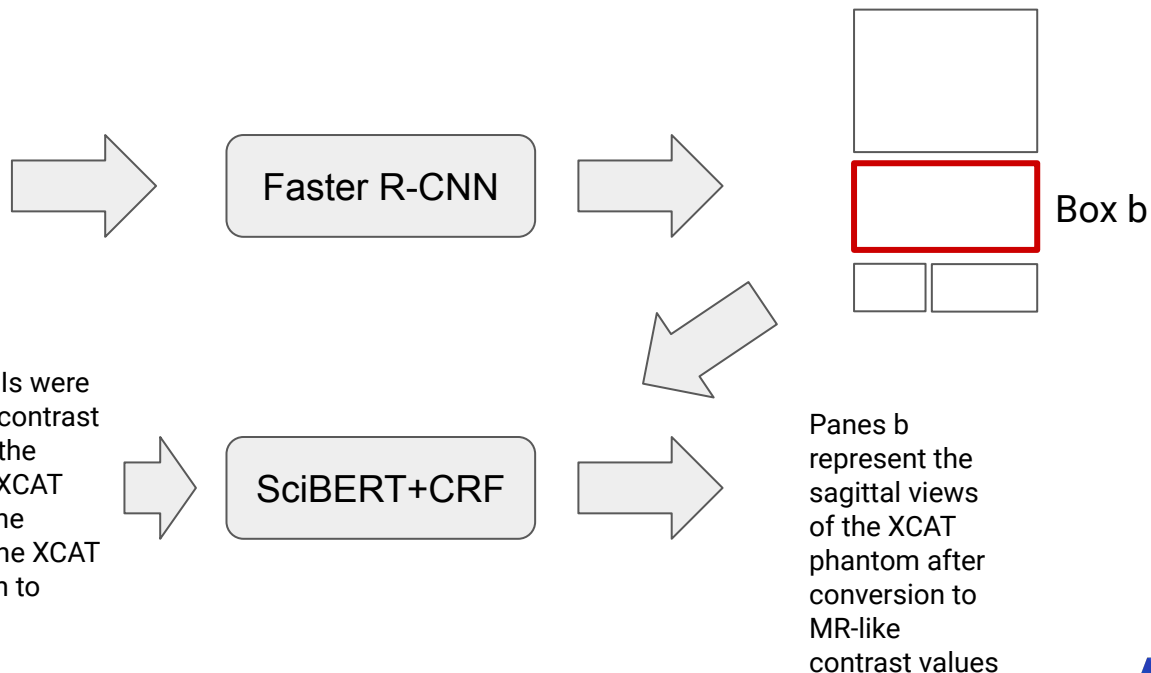


Fig. 3 Assignment of tissue values. ROIs were delineated in a clinical MR (a). Similar contrast values were imposed to the tissues in the attenuation image generated with the XCAT phantom. Panes b, c and d represent the sagittal, transaxial and axial views of the XCAT phantom, respectively, after conversion to MR-like contrast values

Text+Box Embedding CRF Tagger



Subfigure-subcaption Alignment Results

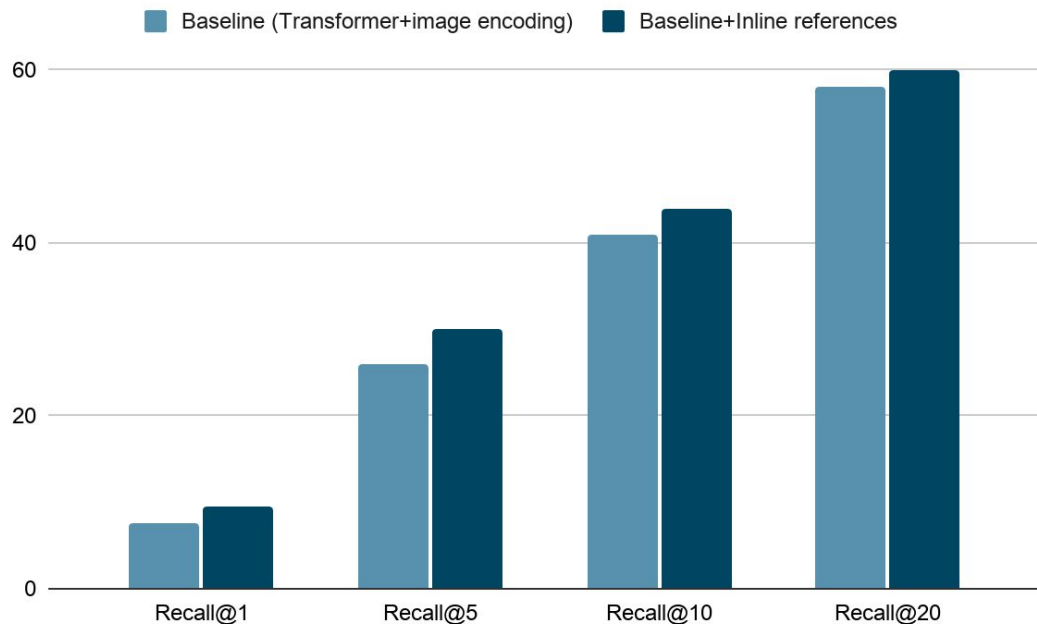
Metric: F1 between set of tokens in gold subcaption and set of tokens in predicted subcaption

Model	Validation F1	Test F1
Oracles + alignment heuristic	88.0	90.8
Oracles (single-span) + alignment heuristic	78.1	75.5
Gold subfigure oracle		
Text-only CRF <small>w/o pretrained weights</small>	47.8	43.3
Text-only CRF <small>w/ pretrained weights</small>	71.0	66.9
Text+Box Embedding CRF	74.1	71.9
Using predicted subfigures		
Text-only CRF <small>w/o pretrained weights</small>	44.7	40.3
Text-only CRF <small>w/ pretrained weights</small>	66.4	61.3
Text+Box Embedding CRF	69.9	67.5



Data Augmentation with Inline References

We also demonstrate how inline references can be used as additional training data for an image-text matching model.



Summary and Future Work

We introduce MedlCaT, a dataset of 217K medical images from 131K OA papers, with subfigure-subcaption annotations and inline figure references

Data and code at github.com/allenai/medicat

Future work:

- Can unsupervised pre-training on the full dataset help with subfigure-subcaption alignment?
- Can paired images and text in MedlCaT help with medical image classification? (subfigure-subcaption alignment should help)
- Can inline references be used in a document understanding task?

