Overview of MiReportor: Generating Reports for Multimodal Medical Images

Xuwen Wang, Hetong Ma, Zhen Guo and Jiao Li
Institute of Medical Information and Library,
Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
li.jiao@imicams.ac.cn

Abstract

This demo paper presents a brief introduction of MiReportor, a computer-aided medical imaging report generator, which leverages a unified framework of medical image understanding and generation to predict readable descriptions for medical images, and assists radiologists in imaging reports writing.

1 Introduction

In the intelligent-assisted diagnosis scenario, computers are required to present reliable interpretation of medical imaging findings. Medical Imaging Report Generation (MIRG) integrates advanced technologies such as computer vision and natural language processing for identifying critical information from medical images and giving reasonable explanations (Messina, 2022). This demo paper presents a brief overview of MiReportor (**M**edical **i**maging **Report** generat**or**), a prototype system designed for computer-aided imaging report writing, open accessed by
http://mireportor.com

MiReportor generates fluent imaging reports in both Chinese and English and provides human-computer interaction service for radiomics researchers on image reading, report reviewing and editing.

2 System Overview

The initial design of MiReportor was derived from the unified framework of medical image understanding and generation that we proposed earlier (Wang, 2019). It takes multimodal medical images, such as CT, X-ray, Ultrasound, etc. as input, and predicts related semantic labels as well as brief readable descriptions of radiology findings. In the recent work, we refer to the latest progress on vision-language representation (Feng, 2022) and update our workflow (see Figure 1).

2.1 Medical Image-Text Representation

Well pre-trained medical image-text representation is the basis for generating good descriptions. We selected the visual-language model BLIP (Li, 2022) as our backbone network to build a joint representation model of medical images and texts. We collected open source datasets containing parallel medical image and texts such as ROCO (Pelka, 2018) [1] and MedICaT (Subramanian, 2020)[2], nearly 299K medical image -text pairs for pre-training. We also extracted millions of medical image-text pairs from biomedical literatures and utilized the data bootstrapping strategy suggested by BLIP to filter noisy image-text pairs for optimizing the representation model.

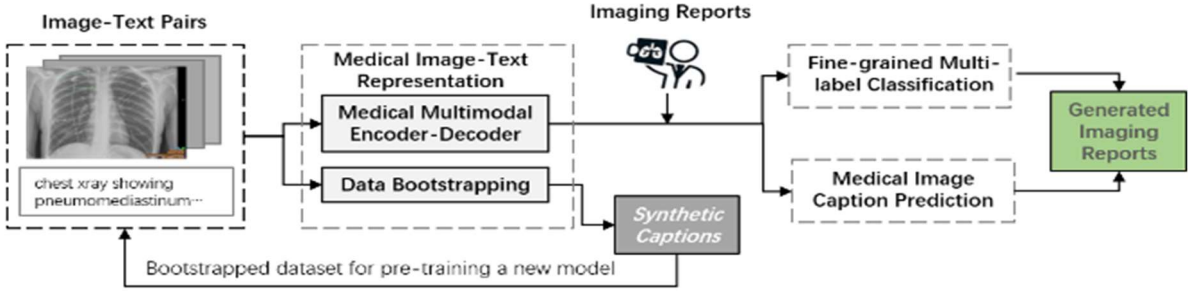

Figure1: Workflow of MiReportor.

[1] https://github.com/razorx89/roco-dataset

[2] https://github.com/allenai/medicat

Proceedings of the 16th International Natural Language Generation Conference: System Demonstrations, pages 1–3
September 11–15, 2023.

2.2 Fine-grained Multi-label Classification

We measured the potential semantic association between medical images and reports by computing their cross-modal semantic similarities. By referring to medical knowledge systems such as UMLS and MeSH, we performed secondary data annotation on medical image datasets according to medical terms and their semantic types. Then we constructed a transfer learning-based fine-grained multi-label classification model to identify key semantic concepts related to medical images (Wang, 2021).

2.3 Medical Multimodal Encoder-Decoder

Since general vision-language model is too large to be applied under low resources, we refer to Liu (2021)'s work on Multi-stage Pre-training and proposed an improved Medical Multimodal Encoder-Decoder (MMED) adapted to the medical scenarios. To capture the alignment of medical images and multi-grained texts, MMED was pre-trained in multiple stages with different training tasks and optimizing objectives. More details about MMED are under review for publication, and we will update this module in the future version.

2.4 Medical Image Caption Prediction

Interpretable descriptions of medical images are the basic composition of semi-structured imaging reports. We developed multiple caption prediction models that generate hierarchical texts for multi-modal medical images, including semantic labels, image sentence topics and coherent sentence descriptions. To obtain accurate reports for specific anatomical parts and imaging types, we fine-tuned caption models based on different open-sourced datasets of real medical imaging reports, such as MIMIC-CXR (Johnson, 2019), Chest X-ray (Demner-Fushman, 2016), etc. One of them is TMRGM (Wang, 2021)[3], a chest X-ray report generation model. Further, by connecting the efficient Aliyun translation interface service, multilingual reports can be output.

3 Evaluation

Considering the various linguistic and visual characteristics of different groups of people, we manually annotated a sentence template library of chest X-ray image reports. We used TMRGM to generate image reports for healthy and patient groups respectively. We validated the performance of chest X-ray caption prediction based on the IU Chest X-ray Dataset, see Table 1. An example of X-ray report generated by our system is illustrated in Figure 2. The current demo deployed both TMRGM and a BLIP-based Chest X-ray report generation model. Users can compare and choose the one suitable to their data. More report generation models will be integrated for other imaging types and body parts in the future.

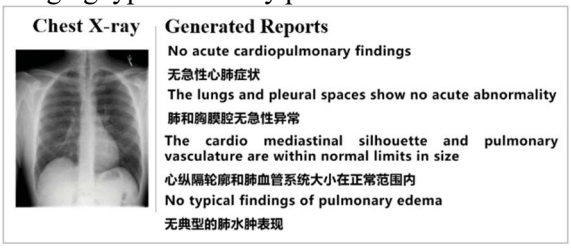

Figure2: An example of imaging report generated by MiReportor.

Method	B1	B2	B3	B4	MT	RG	CD
TieNet (Wang, 2018)	0.286	0.160	0.104	0.074	0.108	0.226	--
CoAtt (Jing, 2018)	0.303	0.181	0.121	0.084	0.132	0.249	0.175
Adapt-att	0.378	0.255	0.185	0.138	0.162	**0.316**	**0.387**
BLIP	0.394	0.232	0.154	0.109	0.167	0.315	0.257
TMRGM	**0.419**	**0.281**	**0.201**	**0.145**	**0.183**	0.280	0.359

Table 1: Preliminary results of Chest X-ray Report Generation, in which B1 to B4 refer to BLEU score, MT refers to METEOR, RG refers to ROUGE, and CD refers to CIDEr.

4 Conclusions

This paper briefly introduces MiReportor, a prototype system for interactive generation of medical imaging reports in both Chinese and English. Experiments based on public chest X-rays revealed the ability of computers on understanding and interpreting medical images. It facilitates the human-computer collaboration practice of imaging

[3] https://github.com/zhangyudoc/TMRGM

diagnosis, which may contribute to the efficient communication between radiologist, clinicians, and patients.

Acknowledgments

This work has been supported by the National Natural Science Foundation of China (Grant No. 61906214), the Beijing Natural Science Foundation (Grant No. Z200016), the CAMS Innovation Fund for Medical Sciences (CIFMS, Grant No.2021-I2M-1-056）

References

Pablo Messina, Pablo Pino, Denis Parra, Alvaro Soto, Cecilia Besa, Sergio Uribe, Marcelo andía, Cristian Tejos, Claudia Prieto, Daniel Capurro (2022). A survey on deep learning and explainability for automatic report generation from medical images. ACM Computing Surveys (CSUR), 54(10s), 1-40.

Xuwen Wang, Yu Zhang, Zhen Guo, and Jiao Li. 2019. A Computational Framework Towards Medical Image Explanation. In Artificial Intelligence in Medicine: Knowledge Representation and Transparent and Explainable Systems: AIME 2019 International Workshops, KR4HC/ProHealth and TEAAM, Poznan, Poland, June 26–29, 2019. Springer-Verlag, Berlin, Heidelberg, 120–131.

Li, Feng, Hao Zhang, Yi-Fan Zhang, Shi Tong Liu, Jian Guo, Lionel Ming-shuan Ni, Pengchuan Zhang and Lei Zhang. Vision-Language Intelligence: Tasks, Representation Learning, and Large Models. https://doi.org/10.48550/arXiv.2203.01922

LI, Junnan, LI, Dongxu, XIONG, Caiming, et al. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In : International Conference on Machine Learning. PMLR, 2022. p. 12888-12900.

Sanjay Subramanian, Lucy Lu Wang, Sachin Mehta, Ben Bogin, Madeleine van Zuylen, Sravanthi Parasa, Sameer Singh, Matt Gardner, Hannaneh Hajishirzi. MedICaT: A Dataset of Medical Images, Captions, and Textual References,2020

O. Pelka, S. Koitka, J. Rückert, F. Nensa und C. M. Friedrich. Radiology Objects in COntext (ROCO): A Multimodal Image Dataset, Proceedings of the MICCAI Workshop on Large-scale Annotation of Biomedical data and Expert Label Synthesis (MICCAI LABELS 2018), Granada, Spain, September 16, 2018, Lecture Notes in Computer Science (LNCS) Volume 11043, Page 180-189.

Xuwen Wang, Zhen Guo, Chunyuan Xu, Lianglong Sun and Jiao Li. ImageSem Group at ImageCLEFmed Caption 2021 Task: Exploring the Clinical Significance of the Textual Descriptions Derived from Medical Images. CEUR Workshop Proceedings (CEUR-WS.org), CLEF 2021 Conference and Labs of the Evaluation Forum, September 21–24, 2021, Bucharest, Romania

Tongtong Liu, Fangxiang Feng, and Xiaojie Wang. 2021. Multi-stage Pre-training over Simplified Multimodal Pre-training Models. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 2556–2565, Online. Association for Computational Linguistics

Johnson AE, Pollard TJ, Berkowitz SJ, Greenbaum NR, Lungren MP, Deng CY, Mark RG, Horng S. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific Data. 2019;6.

Demner-Fushman D, Kohli MD, Rosenman MB, Shooshan SE, Rodriguez L, Antani S, Thoma GR, McDonald CJ. Preparing a collection of radiology examinations for distribution and retrieval. J Am Med Inform Assoc. 2016 Mar;23(2):304-10.

Xuwen Wang, Yu Zhang, Zhen Guo, Jiao Li. TMRGM: A Template-Based Multi-Attention Model for X-Ray Imaging Report Generation. Journal of Artificial Intelligence for Medical Sciences, Volume 2, Issue 1-2, June 2021, Pages 21 - 32

Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu and Ronald M. Summers, TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-Rays, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 9049-9058, doi: 10.1109/CVPR.2018.00943

Baoyu Jing, Pengtao Xie, Eric Xing, 2018. On the automatic generation of medical imaging reports, In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2577–2586, Melbourne, Australia. Association for Computational Linguistics.