Project summary

Tom Hope’s lab guided by Nir Mazor
Noam Delbari

Introduction

In a world where there is an ever-growing accessibility for medical imaging tools, their demand is growing as well. Since those tools depend on authorized technicians and radiologists which have an extensive training in managing the machines and providing the results, their time is extremely valuable. With the growing need of radiologist their workload may exceed their capabilities which could result in less focus for each patient and inaccurate prognosis. Considering this reality, there is a necessity for tools which will assist the radiologist and lower his workload which is the goal of this project. In this project I developed a demo for medical search engine web application that assists radiologists and physicians by providing pathological insights for their patient’s case using a specialized neural network pretrained on medical images named CLIP.

Literary review

While image search became useful and commonly used feature in search engines, widely used medical search engines such as Ovid and PubMed lack this option. Creating capable content-based image retrieval system is a complex problem, requires large amount of labeled data, which can be time-consuming and may require complex models. Developing medical CBIR systems is an uprising research field in the wake of powerful LLMs. Although there are not many articles providing information on entire CBIR systems, many delve into developing algorithms to accomplish the task of retrieving medical images based on an image query: One of the earliest ideas was to construct vector space based on image features derived from CNN. In the article Deep learning-based search engine for biomedical[i] the authors acquired feature vectors from trained CNN to each class in their medical images database. When a user entered an image query, they used the CNN to retrieve feature vector from the image and performed similarity search between each of the classes mean feature vectors to get the class of the image. Then, the algorithm performed similarity searched between the image’s feature vector to each image feature vector in the class in the database the retrieve the closest image. To measure their results, the authors compared the true class of the image being provided to the search engine and the retrieved image class, achieving 82.89% accuracy. Another article which used more recent approach named Text-guided visual representation learning for medical image retrieval systems[ii]. In this article the authors aimed to improve image retrieval in health care systems by using ‘dormant’ data from medical resources to improve image representation. Using CLIP architecture with modified pretrained in-domain image encoder (Clincal BERT), the authors performed contrastive learning between encoded images and their corresponding text encoded captions to gain additional sematic information on images during retrieval. To test their algorithm, the authors created custom data set from ROCO dataset by mapping images to labels by mapping CUI to semtypes in

UMLS library, by using only images with semtypes “Diagnostic procedure”, “body part” and “organ”. To annotate their test set, they’ve used doctors and measure the precision of the retrieval by getting the most K closes images. Comparing their results by using data from custom retrieval dataset using precision at 5 images they compared to general CLIP model: achieving 93.2% compared to 84% in general clip for “modality” and 65% compared to 50.8% for “organ”.

Medical CLIP search engine

In the web application a user will be presented with an easy to use and accessible UI, where he can use either enter text or image as an input to retrieve similar cases from the literature. Furthermore, description of the pathology will be presented with a link to the article. During user’s searches, he’ll be able to save favorite images for detailed inspection later. Additional usage of favorite results is for more accurate search results such in the case when user disliked one of the results, the algorithm will use favorite images to replace undesirable results. Disliking a result will cause the app to search for more appropriate cases based on the liked images. During the project I’ve used 2 datasets which used as a knowledge base from which the user retrieved similar cases. The main database which was used during the testing in the lab’s cluster is PMC Open Access Subset contains 1.65M image-text pairs from more than 3.4 million journal articles and preprints that are made available under license terms that allow reuse. During development process and local testing, I’ve used Radiology Objects in Context dataset consists of 79,789 images text pairs large-scale medical and multimodal imaging dataset. The listed images are from publications available on the PubMed Central Open Access. In order to be able retrieve and search efficiently from many thousands up to millions of images, data has to be encoded in such a manner that will allow to perform simillarity search fast during user search and store all the data on a local Database. For those purposes I’ve used BiomedCLIP. This model is a biomedical vision-language foundation model that is pretrained on PMC dataset with domain-specific adaptations tailored to biomedical vision-language processing[iii]. Foundational models are large-scale artificial intelligence models trained on vast and diverse datasets that can be adapted to perform a wide range of downstream tasks. CLIP model trained image and text encoders to embed (image, text) pairs in a shared space, optimizing for a high cosine similarity among positive pairs and a low similarity for negative pairs. Specifically, I’ve used its image encoder to create vectors which were stores in a local Sqlite DB. Since storing vectors is not a default functionality enabled by Sqlite, the encoded vectors has to be adapted to json format in order to store them as objects. During retrieval, I’ve searched for nearest neighbors at encoded space by cosine similarity. For the langugae modality, where a user enters a description or key words for a pathology I’ve used BM25 which is a ranking function used in information retrieval to estimate the relevance of documents to a given search query. It's a widely adopted algorithm in search engines and databases for ranking documents based on their content's similarity to the search terms provided by a user. For this part I’ve used Pyserini which is an open-source Python toolkit designed to facilitate information retrieval research and applications. It provides a simple Python interface to the Java-based Anserini IR toolkit, which is built on top of the widely-used Apache Lucene search library. Pyserini implements BM-25 retrieval model. As was described previously, datasets consits of image,text pairs together with an ID. I've constructed a .jsonl file consistsing of rows containing ID and image description. This file is called a collection used as an input to create an indexed dataset. When a user enters his query, it being split to individual terms. Each term is then being searched to provide potential IDs of images. Then, each document given a score based on BM-25 ranking algorithm to provide most relevant image, description pairs. Top 5 document scores are chosen, and then displayed to the user. BM-25 uses Inverse Document Frequency to evaluate how common a term is across all documens. Combined with term frequency for each document and its length the score is provided for the term, and finally summed for all terms in query to provide final score to each document.

Conclusion

During this project I've expanded my knowledge over new field of medical search engines, both theoretically discovering the field of vision-languange with understanding how CLIP model works and by searching relevant literature how search engine systems implemented and reading articles describing various algorithms implementing CBIR systems. On the other hand, I've gained experience developing demo for a search engine, expirimanting at both languange and vision modalities using various algorithms and NNs. Getting to know how to work with important libraries such as pyserini and hugginface. More over, I had the chance to work on a cloud and specifically on Lab's cluster using large comute to encode many GBs of data. In my opinion the field of medical search engine has great potential in the age of large datasets and new AI models taking one step closer to create reliable systems which professionals can use.

[i] Mishra, R., Tripathi, S.P. Deep learning based search engine for biomedical images using convolutional neural networks. Multimed Tools Appl 80, 15057–15065 (2021). https://doi.org/10.1007/s11042-020-10391-w

[ii] G. Sérieys, C. Kurtz, L. Fournier and F. Cloppet, "Text-guided visual representation learning for medical image retrieval systems," *2022 26th International Conference on Pattern Recognition (ICPR)*, Montreal, QC, Canada, 2022, pp. 593-598, doi: 10.1109/ICPR56361.2022.9956402.

[iii] BiomedCLIP: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs Sheng Zhang, Yanbo Xu, Naoto Usuyama, Hanwen Xu, Jaspreet Bagga, Robert Tinn, Sam Preston, Rajesh Rao, Mu Wei, Naveen Valluri, Cliff Wong, Andrea Tupini, Yu Wang, Matt Mazzola, Swadheen Shukla, Lars Liden, Jianfeng Gao, Matthew P. Lungren, Tristan Naumann, Sheng Wang, Hoifung Poon. arXiv:2303.00915