의료 이미지 인공지능관련 Foundation 모델의 연구 동향

홍원석 [1], 박인서 [2], 이현원 [3], 홍권 [3], 최현수 [3,*]
[1] 강원대학교, [2] 지오비전, [3] 서울과학기술대학교

4756hong@kangwon.ac.kr, inseo301@ziovision.co.kr, {lee.hyunwon999, ghdrnjs3, *choi.hyunsoo}@seoultech.ac.kr

Research trends in medical image artificial intelligence-related Foundation models

Hong Won-Seok[1], Park Inseo[2], Lee Hyun-Won[3], Hong Kwon[3], Choi Hyun-Soo[3,*]
[1]Kangwon National University, [2]ZIOVISION, [3]Seoul National University of Science and Technology

요 약

본 논문은 의료 분야에서 Foundation 모델들의 중요성을 강조하고, 프롬프트 기반 여러 Foundation 모델들에 대해 크게 텍스트 프롬프트모델, 시각적 프롬프트 모델로 분류하였고, 텍스트 프롬프트 모델은 대조적, 생성모델, 하이브리드 모델, 대화형 모델로 세분화하였고, 시각적 프롬프트 모델의 경우 적응성 모델과 일반화된 모델로 세분화하여 소개하였다.

Ⅰ. 서 론

기존의 딥 러닝은 과제특화적인 데이터, 레이블링 된 데이터에 의존하는 경향이 있다. 이런 단점들을 극복할 수 있는 모델이 Foundation 모델이다[1]. Foundation 모델 같은 경우에는 대규모의 데이터셋으로 사전 훈련된 딥러닝 모델로 미리 사전 학습이 완료된 후에 여러 Downstream task 에 적용될 수 있다 라는 장점이 있다. 이렇게 대규모 데이터셋으로 훈련된 Foundation 모델들은 딥러닝 분야에서 다양한 문제를 해결하기위한 범용적인 도구로써 주목을 받고 있고, 문맥적 추론, 일반화능력, 테스트시 프롬프트에 대한 기능을 용이하게 한다.

Foundation model 의 발전으로 의료 이미지 분야에서도 많은 발전이 있다. 의료데이터의 특성상 개인정보 문제, 데이터의 품질과 일관성, 많은 비용이 든다는 문제로 인해 좋은 질과 많은 양의 데이터를 수집하기 어렵다는 단점이 있다. 그러나 Foundation model 은 대규모의 데이터셋으로 사전 학습이 되어있어, 제한된 레이블 된 데이터만으로도 downstream task 에 적용될 수 있다. 최근 GPT[2]와 같은 모델들의 등장으로 대규모 언어모델(LLM)이 급속도로 발전하고 있다. 대규모 언어모델이 발전됨에 따라 CLIP[3]과 같은 시각-언어 모델의 발전이 촉진될 것으로 보이고, 의료 도메인에 접목하는 연구들이 등장하기 시작했다.

우리의 논문에서는 [1]에서 분류한 방식대로 의료 이미지 분야에서의 프롬프트 기반 Foundation 모델들을 크게 텍스트로 유도되는 특징 맵을 활용한 텍스트 프롬프트 모델(Textual Prompted Models, TPM)과 시각적 입력으로 유도될 수 있는 특징 맵을 이용한 시각적 프롬프트 모델(Visual Prompted Models, VPM)로 분류하고, 세부적으로 분류하여 조사한다. 또한 향후 의료 이미지 분야에서의 Foundation 모델의 연구 방향성에 대해 논한다.

Ⅱ. 본론

우리는 프롬프트 기반 Foundation 모델들에 대해 크게 텍스트 프롬프트 모델과 시각적 프롬프트 모델로 나눠서 분류하였고, 전체적인 구조는 그림 1 과 같다.

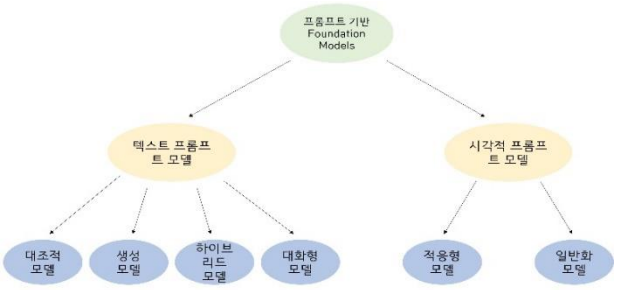

그림 1. 프롬프트 기반 Foundation 모델의 트리구조 시각화

A.텍스트 프롬프트 모델

텍스트 프롬프트 모델에서는 대조적 텍스트 프롬프트 모델, 생성 모델, 하이브리드 텍스트 프롬프트 모델, 대화형 텍스트 프롬프트 모델 네 가지로 나눠서 설명한다[1].

대조학습이란 특징 공간 내에서 비슷한 샘플들을 서로 가깝게, 다른 샘플들은 멀어지도록 학습하는 방법이다. 대조적 모델에는 Tiu, Ekin, et al 가 제안한 CHeXzero[4] 라는 모델이 있다. 이 모델은 자기지도학습을 사용한 모델이다. 이 모델은 레이블이 달리지 않은 병에 대해서도 예측을 할 수 있다. 이를 가능하게 하기 위해 이미지와 텍스트 쌍을 사용하여 자기지도학습을 수행하여 제로 샷 멀티 레이블 분류를 가능하게 하는 표현을 학습한다고 한다. 따라서 이 방법은 레이블 데이터를 사용한 fine-tuning 을 하지 않고, 테스트 시에만 레이블이 필요하다. 이 방법은 흉부 X-선 분류에서 우수한 성능을 보인다고 한다. 이 외에도 MedCLIP[5]과 같은 모델들도 존재한다.

생성모델은 텍스트 프롬프트를 기반으로 현실적인 의료 이미지를 생성하기위해 고안되었다. 그러나 의료 이미지가 아닌 의학적 질문에 의사 수준의 답변을 생성해주는 모델도 존재한다. Singhal et al.가 제안한 Med-PaLM 2[6] 모델은 1066 개의 소비자 의학 질문에 대한 쌍 별 비교 순위에서 의사들은 9 개의 축중 8 개 축에서 의사들이 만든 답변보다 Med-PaLM 2 의 답변을 선호한다고 한다. 즉, 의료 분야에

대한 질문에 대답하는데 있어서 매우 높은 수준의 성능을 보인다. 이를 위해 Med-PaLM 2 는 기본 LLM 의 개선과 의료 분야에 대한 fine-tuning 및 앙상블 정제 접근법을 포함한 프롬프팅 전략을 활용하여 성능을 높인다.

하이브리드 모델은 생성 모델 방법론과 대조적 방법론을 모두 사용한 형태의 모델이다. 하이브리드 모델의 예시로는 Singhal, Karan, et al.가 제안한 MedBLIP[7]이 있다. MedBLIP 은 LLM 을 사용하여 VLP(Vision-Language Pre-training)를 부트스트랩하는 새로운 CAD(Computer-Aided Diagnosis, CAD)이다. 그리고 3D 의료 이미지와 2D 로 사전 학습된 이미지 인코어와 LLM 사이의 간극을 MedQFormer 모듈을 설계함으로써 줄인다.

대화형 모델은 의료 전문가들이 모델에게 의료 이미지에 대한 설명을 요구하는 것과 같이 의료 전문가들과 모델 간 상호작용이 가능하게 한다. 대화형 텍스트 프롬프트 모델 중에 LLM 을 의료 이미지 CAD 네트워크에 통합하는 새로운 방법을 제안한 논문이 있다[8]. 이것은 LLM 을 통해 의료 분야에 대한 지식과 추론을 가지고, 의료 이미지 CAD 모델의 시각 이해능력과 결합하여 기존 CAD 시스템에 비해 환자가 더 이해하기 쉬은 시스템을 만들었다. 이 외에도 DeID-GPT[9]같은 모델도 존재한다.

B. 시각적 프롬프트 모델

시각적 프롬프트 모델의 경우 적응성 모델과 일반화된 모델로 나누어서 설명한다[1]. 적응성 모델은 기존 분할 모델에 대한 적응과 수정을 탐구하여 의료 이미지 작업에 대한 특수성과 성능을 향상시킨다. Ma, Jun, et al.은 100 만개 이상의 의료 이미지-마스크 쌍으로 구성된 데이터셋을 사용하여 의료 이미지 segmentation 분야의 foundation model 인 MedSAM 을 제안했다[10]. MedSAM 은 다양한 이미지 모달리티에 사용할 수 있으며, 특히 종양에 대한 과제에서 뛰어난 성능을 보인다. 이후 MedSAM 을 활용한 MedLSAM[11] 같은것들이 등장하였다.

일반화된 모델의 경우 X-Ray, MRI 같은 영상 이미지부터 환자의 유전데이터를 포함한 여러 데이터 모달리티를 처리하도록 설계되어 Foundation 모델의 능력을 향상시킨다. Singhal, Karan, et al.가 제안한 Med-PaLM M[12]은 의료 영상, 유전자등 다양한 의료 모달리티를 처리할 수 있는 멀티모달 생물의학 인공지능 시스템이다. 이는 ViT 와 LLM 을 융합했고, 새로 구성된 MultiMedBench 데이터셋에서 fine-tuning 한다. MEd-PaLM M 은 매우 뛰어난 제로 샷 성능을 보이며 사전훈련 없이 가슴 X-Ray 에서 결핵을 매우 잘 식별한다. 게다가 사람평가에서는 방사선 전문의의 성능에 필적한다고 한다.

Ⅲ. 결론

우리는 [1]을 정리하여, 의료 이미지 분야에서 프롬프트 기반 Foundation 모델들에 대해 텍스트 프롬프트 모델과 시각적 프롬프트 모델로 분류하고, 그 안에서 세분화하여 모델을 분류하였다. 특히 텍스트 프롬프트 모델에서 생성모델의 Med-PaLM 2[6]는 이미 의학적 질문에 대해 전문가 수준의 답변을 만든다. 이런 의료분야에서 프롬프트 기반 모델들의 기술적 진보는 앞으로 의료분야에 있어서 중대한 이점을 가져다 줄 것으로 기대된다. 그러나 아무리 Foundation 모델이 대규모 데이터셋으로 사전학습을 할지라도, 모든 데이터를 학습하기에는 불가능에 가깝다. 그러므로 훈련 과정에서 인종이나 민족같이 미처 신경 쓰지 못한 편향이 존재할 수 있다[1]. 이런 편향들 뿐 아니라 여러 외압에 강인한 모델들에 대해 탐구하는 것이 앞으로의 과제이다.

ACKNOWLEDGMENT

본 논문은 한국과학기술연구원 기본사업 (2E32260 and 2E32264) 및 한국연구재단 이공계기초연구사업 (NRF-2022R1F1A1076454)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] Azad, Bobby, et al. "Foundational models in medical imaging: A comprehensive survey and future vision." arXiv preprint arXiv:2310.18689 (2023).

[2] Brown, Tom, et al. "Language models are few-shot learners." Advances in neural information processing systems 33 (2020): 1877-1901.

[3] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." International conference on machine learning. PMLR, 2021.

[4] Tiu, Ekin, et al. "Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning." Nature Biomedical Engineering 6.12 (2022): 1399-1406.

[5] Wang, Zifeng, et al. "Medclip: Contrastive learning from unpaired medical images and text." arXiv preprint arXiv:2210.10163 (2022).

[6] Singhal, Karan, et al. "Towards expert-level medical question answering with large language models." arXiv preprint arXiv:2305.09617 (2023).

[7] Chen, Qiuhui, et al. "MedBLIP: Bootstrapping Language-Image Pre-training from 3D Medical Images and Texts." arXiv preprint arXiv:2305.10799 (2023).

[8] Wang, Sheng, et al. "Chatcad: Interactive computer-aided diagnosis on medical image using large language models." arXiv preprint arXiv:2302.07257 (2023).

[9] Liu, Zheng liang, et al. "Deid-gpt: Zero-shot medical text de-identification by gpt-4." arXiv preprint arXiv:2303.11032 (2023).

[10] Ma, Jun, et al. "Segment anything in medical images." arXiv preprint arXiv:2304.12306 (2023).

[11] Lei, Wenhui, et al. "MedLSAM: Localize and Segment Anything Model for 3D Medical Images." arXiv preprint arXiv:2306.14752 (2023).

[12] Tu, Tao, et al. "Towards generalist biomedical AI." arXiv preprint arXiv:2307.14334 (2023).