Project Proposal

2021320071_허천윤

Introduction

With the rapid development of deep learning models, particularly in natural language processing (NLP) and computer vision, pre-trained foundation models like CLIP and LLaMA have demonstrated exceptional performance across various tasks. However, these models are often not fully optimized for specific domain tasks, and fine-tuning them to achieve better performance in specialized applications is a growing challenge. In this project, we aim to apply and fine-tune pre-trained foundation models to improve domain-specific tasks such as text-image matching in medical datasets.

Related Works: Recent works such as CLIP have showcased strong generalization across tasks, yet their performance in medical imaging and text analysis is less explored. Our work builds upon these models, aiming to overcome the challenges of domain adaptation.

Challenges: Fine-tuning large pre-trained models on domain-specific data presents computational challenges and risks overfitting, given the typically limited size of domain datasets.

Problem Definition: We aim to improve text-image matching performance in a medical dataset by leveraging pre-trained models, focusing on overcoming domain adaptation issues and generalization limitations.

Datasets

We will use publicly available medical image datasets like MedMNIST and MIMIC-CXR for the text-image matching task. These datasets are chosen for their well-documented utility in medical research and their alignment with our task of interest. Additionally, we will explore other domain-specific datasets as required.

Goals

1. Fine-tune pre-trained models (e.g., CLIP) on medical datasets for text-image matching tasks.
2. Evaluate the performance improvement compared to state-of-the-art methods.
3. Develop a generalizable framework for domain-specific model fine-tuning.

Brief Tentative Schedule

1. Week 1-2: Dataset preparation and preprocessing.
2. Week 3-4: Fine-tuning pre-trained models (CLIP) on the datasets.
3. Week 5: Performance evaluation and comparison with existing state-of-the-art models.
4. Week 6: Report writing and submission of results.

Roles

This project is done individually, with the focus on both dataset preparation and model fine-tuning. Code and results will be shared through GitHub for transparency.

Comparison with SOTA

We will compare the results of our fine-tuned model with the original CLIP model, as well as other SOTA methods in medical text-image matching. We aim to achieve a performance improvement of at least 5% in accuracy over existing methods.