PhD. Statement of Purpose

Yuelin Xin
University of Leeds, University of California, Irvine
sc20yx2@leeds.ac.uk, yuelix1@uci.edu

1. Introduction & Motivation

Current transformer-based vision foundation methods have achieved great success in various basic tasks in computer vision, and we have not yet reached the performance bottleneck of these methods. While these methods are widely successful in natural images, they are not fully explored in the more challenging and specific domains. These domains highlight great obstacles that are ubiquitous in the field of computer vision – *How do we train models effectively with limited data? How do we prevent over-fitting in this case? How do we design models that can generalize well to unseen data?*

One of these domains is computer vision for medicine, which is my current research focus. Vision methods for medicine has long been plagued by the lack of large scale labeled data. This is especially true for 3D medical images, where the annotation process is extremely time consuming and requires expert knowledge. This lack of processed data leads to dissatisfactory training outcomes, and thus limits the effectiveness of vision methods for medicine. My research goal is to explore methods that can alleviate these problems, and I believe robust solutions to these problems can be beneficial to the field of computer vision as a whole. Advancing computer vision for medicine can also bring accessible medical care to more people than ever, giving second opinions to doctors and patients, and assisting rapid diagnosis in latent diseases such as breast and brain cancer.

2. Research Experience

During my undergrad years, I had the privilege to pursue research at 3 different institutions in 3 different countries. Meeting and working with these very different groups of people gave me an unique perspective that few others have. I experienced in firsthand how research is done in different parts of the world, and how different people produce unique insights for complex problems.

Object Tracking & Temporal Representation, SWJTU, China. Under the guidance of **Dr. Zhiguo Long** at SWJTU, I embarked on my first significant research project in object tracking and temporal representation in CNNs. Tasked with detecting minuscule production defects on electronic boards, a challenge exacerbated by scarce labeled data and a long-tail distribution of defects, we started by building the largest public dataset in this niche, comprising over 1,000 items and 2,800 annotations. We also designed a customized backbone network based on YOLOv5, enhancing detection accuracy while improving inference speed by 25%. Up until now, the work is still missing a part of the puzzle, CNN models like YOLO only detect objects in an image, in order to achieve real-time object tracking in video, we need to introduce temporal information into the detection process. We proposed a novel algorithm, 2SDS, that integrates with CNN-based models to process temporal information by performing temporal segmentation in parallel with the model. This plug-and-play technique enables real-time object tracking in video streams, and is able to process information in a fine temporal granularity, which is crucial for real-time applications. ***Our efforts were summarized into a publication, which I had the honor of writing the first draft and orally presenting at an IJCAI workshop in 2022.*** This work was pivotal in sharpening my research skills across the entire project lifecycle, from conceptualization to implementation and dissemination. This work

also underscored the importance of abundant data in training domain-specific methods, which is a challenge I have since been committed to addressing.

Pseudo Label Training & Medical Image Registration, UC Irvine, USA. I spent the summer of 2023 at UC Irvine, working with **Prof. Xiaohui Xie** and his group to find an alternative training methods for weakly-supervised and unsupervised image registration models. Weakly-supervised methods requires manually labeled data, while unsupervised methods rely on indirect heuristics to guide the training. This project centered on 3D medical images, confronting the persistent challenge of limited labeled data in medical imaging. We proposed On-the-Fly Guidance, a unified training framework employing pseudo labels to provide on-the-fly supervision for registration models. By integrating an optimizer in the training process, pseudo labels can be generate from model outputs, significantly refining the training process compared with that of unsupervised learning, without using any auxiliary label. Our method demonstrated significant and consistent performance gains across the tested models and datasets, outperforming the previous state-of-the-art. ***My role as the co-first author in our paper, submitted to CVPR 2024, was to design and implement the optimization stage of the framework, as well as conducting experiments and empirical analysis on the results. I also contributed to the majority of the writing and editing of the paper.*** This work is my first attempt at addressing challenges in model training for data-limited domains, and I am excited to continue exploring this topic in my future research.

Vision Backbone Model for Medical Image Analysis, Leeds & UC Irvine. The ongoing collaboration between the University of Leeds and UC Irvine, **led by myself**, and co-supervised by **Dr. Sharib Ali** and **Prof. Xiaohui Xie**, is a testament to the power of cross-institutional synergy. We are currently developing a new vision backbone model, Kernel Transformer, using a multiple and variable-granularity sliding kernel design, we are aiming to surpass the Swin Transformer's capabilities in fine-grained feature extraction, which is crucial for medical image analysis. The idea is to apply it to various downstream tasks such as medical image segmentation, object detection, etc. We are also exploring on various data augmentation methods to alleviate over-fitting problems presented by data-limited domains. This project is not only a continuation of my commitment to advancing computer vision for medicine but also an endeavor to contribute a pioneering method that could push the field forward. ***We are in the process of documenting our methods and results in a paper for ECCV 2024***, marking another significant progress in my research journey.

3. Goals & Plan

I plan to continue my research in computer vision for medicine, as well as computer vision as a whole. There are still numerous methods that can be explored to solve the problems mentioned in Sec. 1, and a solution to these problems is generic to the field of computer vision. More specifically, I plan to work on both domain-specific and natural images, and explore the possibility of transferring knowledge between these two domains.

4. Program Fit

I stumbled upon Dr. Jia-Bin Huang's group at UMD, and I was immediately drawn to the research they are doing. I was especially impressed by the work on various intriguing topics in natural images. Despite my current focus on medical images, I do plan to apply the techniques back to natural images, which is why I believe Dr. Huang's group is a great fit for me. And I cannot resist a group that has unlimited supply of snacks before deadlines, which seems funny but in fact is important in successful research projects. I also believe my diverse academic background can be a great addition to their group and the UMD community. Being an undergrad student, I realized that I may not have the most abundant research experience, but I'm confident in my ability to learn and adapt, and eventually thrive in a PhD program.