Wai Tak Lau (Michael)

Personal Statement: Computer Science PhD at Columbia University

ChatGPT told me that the determinant of a diagonal matrix is the sum of the diagonal elements, but it should be the product instead. [1]Although trivial in this case, large language models hallucinating pose a serious problem when used in domains such as healthcare. One of the central problems in modern deep learning is that it fits to observational distribution and does not consider the cause and effect between different variables. These problems have motivated my **research interest**: 1) Learn causal representations of high-dimensional and multi-modal data, that allows models to perform well on unseen data and when few labels are available. These representations can be reused to build new models iteratively ; 2) Using the learnt causal representations for developing machine learning models in healthcare, applying to diseases classification, biomarker discovery and clinical deployment.

I discovered my passion for research much later than my peers. When applying for masters program during senior year undergrad, my goal was to go into industry as I had yet to gain in depth knowledge and problems that I am passionate about. However, this changed as I dove deeper into different areas in machine learning and discovered many interesting and important problems. This led me to work in the AI4VS lab, led by **Prof. Kaveri Thakoor**, along with other amazing collaborators. Over the process, I grew to enjoy research and gained valuable experience. I have keen interest and am now looking to further develop my skills at Columbia University.

[1] I asked ChatGPT to generate math questions and it failed to provide the correct answer.

Research at AI4VS Lab Data scarcity is one of the main problems in machine learning in healthcare, involving not only gathering a large amount of training data but also obtaining high-quality labeled data for training supervised algorithms.

I am excited to address this problem by using data of different modalities, similar to how text and images can be combined. At AI4VS lab, we used eye tracking data to address this problem. Eye tracking has long been used to understand the subject's cognitive process when examining stimuli, and the relationship was proposed in the Eye-Mind hypothesis [1] [2]. Eye tracking data reflects the regions and order of importance as clinicians examine the reports, and they can be used as surrogate information that acts as labels in a situation where ground truth is not available. Most of the current approaches when using eye tracking data with images ignore the temporal aspect of the data, therefore the challenge is to create a reasonable encoding to capture both spatial and temporal relationships. To address this problem, I designed a new way to encode eye tracking information as words, inspired by natural language processing. A transformer encoder based model is then used to train with a multitask objective to learn meaningful embeddings where eye tracking on similar images are close to each other. Our goal is to use these embeddings to generate pseudo labels to train a model that uses optical coherence tomography (OCT) reports of the retina to predict Glaucoma (a neurodegenerative disease that leads to blindness). Although we need more data to support this conclusion, the results have shown that eye tracking data aids downstream tasks like Glaucoma classification. More surprisingly, eye tracking data alone achieves relatively high accuracy in classifying Glaucoma. During this process, I learnt how to maneuver in the complex search space of possible hypothesis and convert a hypothesis into concrete problems that we can solve. I presented this work at Columbia

Center of AI Technology Symposium: AI & Healthcare and I am working to submit this work to as the **first author**.

In the context of healthcare, I am interested in developing and applying techniques in causal representation learning to learn interpretable representations that can aid understanding new biomarkers, interpretability and be used for downstream applications such as segmentation or even hypothesis generation. As a first step, we can work towards a joint embedding space between images and eye tracking, and even text from clinicians inspired by CLIP [4].

Other Projects and Experiences One interesting application that arises from GANs is performing image translation by generating samples from the conditional distribution $P(x \mid y)$, where y is the input image and x is the translated image. However, this translation might not be accurate due to lack of conditioning on set of covariates z. For example z could be which city we would like to translate input into in Cityscape dataset [1], since different cities might have different architecture style. To overcome this challenge, we can not only generate from $P(x \mid y, z)$ but also take into account the causal relationships, where each variable is modeled by a neural network.

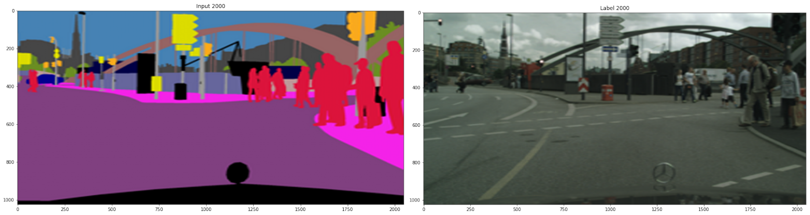

Figure 1: Example inputs and outputs from CityScape Dataset [1]. The dataset is collected from different cities. Left is the input, and right is output.

As I embark on my research journey, I have realized in many ways research is similar to entrepreneurship as they both require extreme ownership. During my undergraduate studies, I ventured into the entrepreneurial world through the Alchemy program at University of Illinois at Urbana-Champaign (UIUC), lead by **Professor Sanjay Patel**. Our goal was to build a minimum viable product for predicting realtime audio quality in the hopes that this information could be helpful for improving audio delivery and ultimately video as well. There were two main challenges: finding high quality training dataset that contains audio degradations and the real-time constraint. To solve the first problem, we decided to build a synthetic dataset by collecting our own data using Amazon Mechanical Turk. I helped design and build the survey as well as the different simulators to generate the synthetic audio dataset that consists 30 hours of audio. This gave me the chance to dive into audio quality metrics, compression codecs and audio quality degradations. To address the second problem, I also suggested the two stage features extractor and regression model to accommodate the real time needs. Ultimately, the model performed competitively and I learnt how to pivot around a obstacles creatively.

At Columbia My long term goal is to become a researcher in using machine learning to enhance healthcare. I am especially excited to work with **Professor Kaveri Thakoor** and **Professor Richard Zemel**. I specifically want to work in the AI4VS Lab, as the lab and my interests aligns in robust and interpretable machine learning in healthcare using data from different sources and different modalities. Professor Richard Zemel's interests in few-shot learning and continual learning aligns with my interest learn useful representation that allows model to perform well when few labels are available and build models iteratively. I appreciate the close connection between the medical school and the computer science department, making it the ideal environment for me to transition from theory to application. With the excellent faculty, students and collaborative environment, Columbia is the ideal place for me to continue my academic journey.

References

1. Just, M. A., and Carpenter, P. A. (1976b). Eye fixations and cognitive processes. Psychol. Rev. 87, 329–354. doi: 10.1016/0010-0285(76)90015-3

2. Rayner K. Eye movements in reading and information processing: 20 years of research. Psychol Bull. 1998 Nov;124(3):372-422. doi: 10.1037/0033-2909.124.3.372. PMID: 9849112.

3. M. Cordts, M. Omran, S. Ramos, *et al.*, *The cityscapes dataset for semantic urban scene understanding*, 2016. DOI: 10.48550/ARXIV.1604.01685. [Online]. Available: https: // arxiv.org/abs/1604.01685.

4. Radford, A. et al. (2021) ‘Learning Transferable Visual Models From Natural Language Supervision’, arXiv [cs.CV]. Available at: http://arxiv.org/abs/2103.00020.