ELECTRICAL & COMPUTER ENGINEERING
SUMMER
RESEARCH
INTERNSHIP
PROGRAM

FACULTY MENTOR

Nuno Vasconcelos

PROJECT TITLE

Language-Assisted Continual Visual Classification

PROJECT DESCRIPTION

Human brain can learn concepts continually without forgetting. This simple fact has attracted lots of attention from researchers. During the learning process of a human, two types of knowledge will occur: the low-level input directly from the visual system, typically modeled as the raw input image and the high-level understanding of an object from our brain; this is commonly represented as the text description or definition of this object. For a long period, the vision community has neglected the usage of text concepts during the learning of vision systems, while recently, the emergence of large-scale vision-text models (e.g., CLIP, BLIP) has aroused researcher's interest in combining vision and text together. In this project, we consider the usage of language models during the continual learning of a vision system. It is known that abstract concepts like task identity information are hard to be represented by a visual model, yet text models are born to learn these high level things. Moreover, we plan to organize the tasks based on some curriculum (for example, taxonomy) instead of random organization. In this case, the task-level concepts are realistic ones instead of artificial ones, which should give us a better continual vision learning system. The project aims for a top-tier conference publication.

Mentor: Zhiyuan Hu <z8hu@ucsd.edu>

This project can accommodate both remote and in-person students.

PREREQUISITES

- ➢ Experience with Python, Linux, and at least one popular deep learning framework such as PyTorch is an advantage.
- ➢ Stronger candidates will also have some knowledge in computer vision and continual learning.

FACULTY MENTOR

Nuno Vasconcelos

PROJECT TITLE

Customizing Radiation Cancer Treatment with Deep Learning

PROJECT DESCRIPTION

Brachytherapy is a treatment in which a radioactive source is used to deliver radiation internally to treat cancers such as cervical cancer. Currently, clinicians manually tune treatment parameters to customize the radiation to individual patient's anatomy. This process can take over an hour, which is problematic because patients are waiting in discomfort and often under sedation for this to occur. Deep learning can identify anatomical features that relate to ideal, customized radiation treatments by learning from past patient imaging and treatment data. In this project, we will generate new networks and inputs and/or modify existing networks to accurately predict radiation treatment parameters. The end goal is to automate the treatment customization process to ensure high-quality radiation treatments can be produced in a matter of minutes with a single button click. This project will involve working with a team of medical physicists (including Dr. Sandra Meyers), radiation oncologists, and electrical engineers and is a collaboration between the Vasconcelos and Meyers labs. The project aims for a top-tier conference or journal publication.

Mentors: Lance Moore <lcmoore@health.ucsd.edu>; Sandra Meyers <smmeyers@health.ucsd.edu>

This project can accommodate both remote and in-person students.

PREREQUISITES

- Experience with Python, Linux, and at least one popular deep learning framework such as PyTorch is an advantage.

FACULTY MENTOR

Nuno Vasconcelos

PROJECT TITLE

Accurate 3D-Hand Pose Estimation via Multi-Modal Fusion

PROJECT DESCRIPTION

3D hand pose estimation is a fundamental and challenging problem in vision. The goal is to recover the joint angles of the different finger and hand sections. This can be used, for example, in applications such as assistive remote surgery, training robots to grasp objects, or AR/VR problems that warrant the need for accurate 3D hand pose. However, accurate pose estimation using only RGB or depth cameras is impossible in occluded scenarios without the help of additional information. Such situations arise commonly since hands, fingers, and objects tend to occlude each other. To circumvent this, we propose to leverage the use of sensory information from a sensor attached to the human in conjunction with camera frames. In particular, we aim to explore different image/video and sensor fusion methods, including but not limited to early, mid, and late fusion methods. Further, the sensor information should be complementary to visual input that can provide predictions independent of the sensor inputs. This project involves developing an end-to-end pipeline from data collection, pre-processing, and designing the algorithm with a focus on transformer attention architectures. The project aims for a top-tier conference publication.

Mentor: Deepak Sridhar <desridha@ucsd.edu>

This project can accommodate both remote and in-person students.

PREREQUISITES

- Experience with Python, Linux, and at least one popular deep learning framework such as PyTorch is an advantage. Experience with 3D vision is an asset.

FACULTY MENTOR

Nuno Vasconcelos

PROJECT TITLE

Generalizable Neural Radiance Fields (NeRF) with Few Images

PROJECT DESCRIPTION

Neural Radiance Fields lean a continuous representation through multiview consistency, which can then be rendered from any viewpoint. Recently, NeRF-based representations have made significant progress in novel view synthesis and produce photo-realistic rendering results. However, NeRF optimization usually requires a large number of images to model accurate geometry and texture. It is observed that the rendering results decay fast as the number of image inputs decreases. In this project, we will investigate how to learn NeRF with fewer images. Training NeRF with fewer images would open the door for various applications in the real world. For example, one can possibly turn several images taken by your phone into an interactive 3D scene. The project aims for a top-tier conference publication.

Mentor: Jiteng Mu <jmu@ucsd.edu>

This project can accommodate both remote and in-person students.

PREREQUISITES

➢ Candidates are expected to be adept with deep learning (DL) and Python/PyTorch. Experience with 3D vision is an asset.

ELECTRICAL & COMPUTER ENGINEERING
SUMMER
RESEARCH
INTERNSHIP
PROGRAM

FACULTY MENTOR

Nuno Vasconcelos

PROJECT TITLE

Novel Benchmark for Human-level Long-form Video Understanding

PROJECT DESCRIPTION

Large vision-language foundation models (LVLMs), e.g., GPT-4V, BLIP-2, LLaVA), have revolutionized the area of multi-modal learning. Human-like reasoning on image data (e.g., visual question answering) has almost been achieved by LVLMs. However, applying LVLMs to video understanding, especially with a long temporal context, remains a challenging open problem. One limitation of current research is the lack of high-quality long-form video understanding benchmarks since existing ones are obsolete and unsuitable for probing the true capabilities of modern LVLMs in the open world. In this project, we aim to build from scratch a new comprehensive benchmark dataset to facilitate the research in this area. The project aims for a top-tier conference publication.

Mentors: Jiacheng Cheng <jicheng@ucsd.edu>, Yi Li <yil898@ucsd.edu>

This project can accommodate both remote and in-person students.

PREREQUISITES

- ➢ Candidates are expected to be adept with deep learning (DL) and Python/PyTorch.
- ➢ Experience with video deep learning and/or vision-language learning is preferred.
- ➢ Experience with data collection and annotation tools (e.g., Amazon Turk) is a plus.

ELECTRICAL & COMPUTER ENGINEERING
SUMMER
RESEARCH
INTERNSHIP
PROGRAM

FACULTY MENTOR

Nuno Vasconcelos

PROJECT TITLE

Ensemble Modeling for Text-to-Image Generation in Stable Diffusion

PROJECT DESCRIPTION

Text-to-image generation is a multifaceted task, demanding a delicate interplay between generation and reasoning. Existing end-to-end models, such as stable diffusion, often lack explicit differentiation between these components, leading to challenges in interpretability and generalization. In our pursuit of more nuanced solutions, we propose an alternative approach that involves leveraging specialized diffusion models. Each of these models contributes distinct traits to the collaborative process, with a Large Language Model (LLM) serving as the central controller. This task centers on ensemble modeling for text-to-image generation within the diffusion framework, exploring the potential of specialized models guided by the LLM. By adopting this strategy, we aim to enhance interpretability, overcome limitations in generalization, and achieve a more sophisticated balance between visual processing and reasoning. The ensemble, directed by the centralized LLM, promises a novel avenue for advancing the state-of-the-art in text-to-image generation.

Mentor: Alakh Desai <ahdesai@ucsd.edu>

This project can accommodate both remote and in-person students.

PREREQUISITES

- Candidates are expected to be adept with deep learning (DL) and Python/PyTorch.
- Stronger candidates will also have experience with running stable diffusion and some knowledge of diffusion models.