Crowdsourced List

Note: These are the projects proposed by the community, and there is typically a mentor associated with each project. Please reach out to the mentor if you are interested in the project.

1. AI Model Evaluation using Synthetic Data for Radiology

Abstract: Diffusion models are commonly used with state-of-the-art results to generate synthetic data from text prompts. This data can be then further used for data augmentation and other tasks. In this project, we are interested to explore whether synthetic imaging data can be used for model evaluation and compare how the results of an AI model differ between real vs. synthetic data evaluation. A work that explored this idea on tabular data with interesting findings can be found here: https://arxiv.org/pdf/2310.16524.pdf

First, we will use a previously trained model for chest X-Ray generation from text prompts called RoentGen (https://arxiv.org/pdf/2211.12737.pdf). Then we will use the synthetic images for evaluation of AI models trained on different Chest X-Ray publicly available datasets. We will also explore whether generating subgroup-specific images, in terms of age, sex or race can provide meaningful X-Rays that can be used for fairness model evaluation.

Required skills: Python programming necessary, previous experience training and evaluating AI models on imaging data would be beneficial

Contact: Magdalini Paschali, paschali@stanford.edu

Preferred Advising Style: Regular weekly meetings

2. [TAKEN] Generating Affordance-Preserving and Efficient Collision Meshes for Physics Simulation

Abstract: In the context of robotics simulation and video games, to be able to physically simulate a 3D mesh object in a scene (e.g. have it collide with things, etc.), you can't use just use the mesh directly - collision meshes need to be volumes, and they need to be convex. Rather than just taking the convex hull of the object (which will overapproximate the collision mesh) one can approximately convert any arbitrary mesh into a set of convex shapes (a task known as approximate convex decomposition). A variety of ACD tools are available (Google V-HACD and CoACD and review their websites before reading any further). Both of these options can be used to generate reasonable approximations when their parameters are tuned correctly, but we have found it really hard to make them into a plug-and-play solution that works for any object. Instead, you have to tune their parameters for each object you want to use, which we have done for all 6k+ objects in the BEHAVIOR-1K dataset that we released by visually inspecting them. (e.g. the human expert picked the best mesh out of N for each object in the dataset, acting as a classification oracle) Now that we have all this labelled data (collision meshes generated with a number of different parameters on coacd and vhacd, and which of those options a human expert preferred) - I think we might be able to train a model that can look at a model (+ its metadata e.g. category, definition etc. if necessary) and output a good convex decomposition. This can first be in the shape of just learning which option to pick out of all of the options (e.g. we still pre-generate all the collision meshes and the model just picks one), which would be a simple classification task. As the input we can use a voxelized representation of the object, or multiple photos - this is the interesting research area. This is hopefully in-scope and achievable by a CS231N project team. I will provide the data in an easy-to-work-with format (.obj files for each candidate + a JSON file showing the selection). Interesting (and possibly publishable) follow-ups include whether it would be possible to generate the collision meshes directly from a learned solution.

Contact: Cem Gokmen, cgokmen@stanford.edu

Preferred Advising Style: Regular weekly meetings

3. Cortex-level Brain MRI Generation

Neuroimaging studies face challenges in drawing reliable conclusions based on current models, particularly those using deep learning techniques. Limited sample sizes and lack of diversity during training on MRI data sets lead to problems like biasing and overfitting. Generative AI holds promise in addressing these challenges as it can capture the distribution of real MRIs in high-dimensional space and then use that knowledge to produce high-resolution brain MRIs. This project aims to generate high-fidelity brain MRI scans that closely resemble real cortex structures, with the intention of aiding advancements in neuroimaging studies.

Contact: Wei Peng, wepeng@stanford.edu

Preferred Advising Style: Regular weekly meetings

4. Brain MRI foundation model

Design a foundation model for structure MRI, including tasks like Brain diagnosis, MRI registration, and brain tumor segmentation.

Contact: Wei Peng, wepeng@stanford.edu

Preferred Advising Style: Regular weekly meetings

5. [TAKEN] Combining adversarial training and unsupervised learning in vision foundation models to build more biologically plausible vision neural networks

In search for biologically realistic models of visual perception, prior research has demonstrated how adversarially trained self-supervised models achieve better neural productivity in the ventral visual stream. The intersection of adversarial and unsupervised learning (ACL, adversarial contrastive learning) in the context of visual modelling for better neural productivity has been less explored. One potential direction: Compare unsupervised vision models against adversarially trained unsupervised vision models via the Brainscore (https://www.brain-score.org) benchmark, and potentially build a new ACL model with high neural predictivity.

Contact: Jenny Xu, jennyxu6@stanford.edu

Preferred Advising Style: Need-based meetings

6. [TAKEN] Generalization of open-vocabulary 3D LiDAR-based unsupervised perception for self-driving

In 3D LiDAR-based perception for self-driving, supervised models achieve better performance than unsupervised models in terms of mAP on common datasets like Waymo, Lyft and NuScenes. However, supervised models have a fixed set of labels and have limited capacity for generalization. Previous unsupervised models mainly focused on detecting class-agnostic mobile objects, such as in MODEST (https://github.com/YurongYou/MODEST). The paper "Unsupervised 3D Perception with 2D Vision-Language Distillation for Autonomous Driving" (https://openaccess.thecvf.com/content/ICCV2023/papers/Najibi_Unsupervised_3D_Perception_with_2D_Vision-Language_Distillation_for_Autonomous_Driving_ICCV_2023_paper.pdf) enables 3D open-vocabulary detection of objects in the wild and was tested on the Waymo dataset. Due to the variability in sensor modality from various 3D LiDAR point-cloud datasets, this project aims to evaluate the generalizability of the aforementioned method (or another method) in a different dataset (such as Lyft or NuScenes).

Contact: Jenny Xu, jennyxu6@stanford.edu
Preferred Advising Style: Regular bi-weekly meetings

7. [TAKEN] Learning to Generate Interactive 3D Scenes for Robotics Simulation
In projects like BEHAVIOR-1K that can be used as simulation environments for robotics, one major issue is the difficulty of obtaining realistic, new scenes. Interactive scenes need to be designed by 3D artists and then processed with the right annotations by robotics researchers, making their creation extremely difficult and costly. There have been efforts to procedurally generate such scenes like ProcTHOR, but these solutions cause learned robot policies to perform poorly, because they generate scenes that are unrealistic and repetitive. This is because these solutions are built using rules that are far too simplistic to represent the full distribution of what a real-life indoor scene can look like. The goal of this project is to generate these scenes with a learned approach instead from large-scale data about houses available online. Today, there are many possibilities in this space: one can try to use diffusion models to pre-generate scene images and then populate those scenes with 3D objects that we already have. One can try to generate the objects / meshes too. One can apply an iterative process with existing vision language models to try to get them to populate the scene slowly. The possibilities are endless - you would be focusing on a particular subset of those possibilities that you find interesting, using the BEHAVIOR-1K simulation environment and its existing assets. This project does not directly involve any robotics work but it will involve getting yourself familiar with the BEHAVIOR-1K platform.
Contact: Cem Gokmen, cgokmen@stanford.edu
Preferred Advising Style: Regular weekly meetings

Inspiration List

Note: The list below should only serve as ideas for inspiration. You should target projects that you can find data, reasonable methods, existing base implementations, and/or pre-trained models to start with, and the projects should be doable in the timeframe of this course. When in doubt, submit your proposal and make necessary changes based on the feedback, or consult a project TA during office hours.

- Developing a tool that uses image recognition to identify and sort recyclable materials.
- Real-time tracking of wildlife for non-invasive ecological monitoring and research.
- Creating a diagnostic tool that uses image recognition to detect early signs of skin cancer.
- Developing a vision-based system to assist visually impaired individuals in navigating indoor environments.
- A tool for automated detection and blurring of sensitive information (e.g., faces, license plates) in public camera feeds.
- Automated analysis of aerial imagery for real estate and land use planning, including building detection and land cover classification.
- Explore using diffusion models for diverse and coherent story visualization from text.
- Creating a model to predict plant diseases using leaf imagery.
- Build a system for open-set fine-grained image retrieval using vision-language models.
- Implement a model that can generate 3D models of objects from 2D images to assist in virtual reality (VR) content creation.
- Development of a tool for analyzing athletes' performances by tracking movements and biomechanics in sports videos.
- Creating a system for real-time sign language interpretation using image recognition.

- Development of a tool for automated generation of fashion sketches from clothing images.
- Design a model for controllable and photorealistic sky replacement in outdoor images.
- Creating a model to predict traffic flow and congestion using real-time street view imagery.
- Creating a model to generate personalized fashion designs based on user preferences and trends.
- Investigate self-supervised learning techniques for video representation learning.
- Investigate and analyze the failure modes of multimodal language models on vision tasks.
- Automated detection of electrical or mechanical failures in industrial equipment through thermal imaging analysis.
- Create a model for generating photorealistic virtual try-on of clothing items.
- Design a model for controllable and photorealistic facial attribute editing.
- Create a model for photorealistic facial reenactment with audio-driven expressions.
- Develop a vision model to analyze drone footage for environmental monitoring, focusing on detecting changes in vegetation, water levels, and signs of pollution.
- Design a model for controllable and photorealistic portrait relighting.
- Investigate techniques for learning disentangled representations for controllable image synthesis.
- A system for automated comic book generation from movie scenes or video clips.
- Build a system for efficient and accurate hand tracking and reconstruction from monocular video.
- Real-time object removal and scene inpainting in live video streams for privacy protection.
- Create a model for generating diverse and realistic human motions from textual descriptions.
- Develop methods for unsupervised learning of disentangled shape and appearance for 3D objects.
- Build a system for interactive and controllable 3D avatar creation from a single image.
- Leverage robotic simulation to generate training data for 2D/3D point tracking.
- Investigate the analyze different pre-trained visual representations for visuomotor learning for robotics tasks.
- Build a system for interactive 3D modeling and editing using implicit neural representations.
- Build a system for photorealistic 3D reconstruction of indoor scenes from a single image.
- Design a model for generating diverse and realistic 3D room layouts from textual descriptions.