EE/CS/CNS 148b - Large Language and Vision Models

TIME: TuTh 10:30-12:00
PLACE: Chen 100
Instructors: Gkioxari, Perona, Achille, Paolini

What is this class about

Large language and vision models are transforming the way we process and generate text and images. Models, such as GPT-3, trained on massive amounts of text and image data have achieved human-like levels of performance on a wide range of language tasks. This has the potential to transform many human activities including teaching, industry and science. Understanding how these models work and how they can be utilized can lead to new breakthroughs in artificial intelligence and natural language processing. Studying large language models can also provide insights into human communication, and contribute to our understanding of the complex relationship between language, pictures, thought, and intelligence.

The class has three goals:

- Offering an in-depth introduction to LLVMs to Caltech students
- Exploring the application of LLVMs to science
- Developing teaching material for hands-on exploration and learning

Prerequisites

- Intermediate Python programming, including some experience with PyTorch
- Foundations of machine learning incl. unsupervised learning
- Foundations of deep learning
- Linear algebra
- Probability and statistics

Projects

The class will include both regular lectures and hands-on learning. Sudents will be organized into groups of five (give or take) and work on projects lead by an experienced graduate student.

Syllabus (may change slightly)

Intro to neural networks and deep learning. Optimization for deep learning. Language modeling: word embeddings, seq2seq, transformers. Vision modeling: resnets and transformers. Discriminative modeling: classification and detection (Mask R-CNN, DETR), segmentation (UNets), motion (RAFT, PIP), depth estimation. Self-supervised learning, contrastive learning, SimCLR. Generative models (GANs, VAEs, AR, Diffusions). Alignment (RL, ChatGPT). Unified language and vision models (CLIP, BLIP), latent diffusion models (DALLE), Flamingo.