DEEP LEARNING FOR COMPUTER VISION

PROF. VINEETH N BALASUBRAMANIAN Department of Computer Science and Engineering IIT Hyderabad

PRE-REQUISITES :

- Completion of a basic course in Machine Learning
- (Recommended, not mandatory) Completion of a course in Deep Learning, or exposure to topics in neural networks
- Knowledge of basics in probability, linear algebra, and calculus
- Experience of programming in Python

If you are unsure whether you meet the background requirements for the course, please look at Assignment 0 (both theory and programming). If you are comfortable solving/following these assignments, you are ready for the cours

INTENDED AUDIENCE : Senior undergraduate students, Post-graduate students, Industry professionals seeking to understand computer vision

INDUSTRIES APPLICABLE TO : Many organizations – national and international, industry and goverment – use computer vision in their products and services. This includes multinationals such as Google, Microsoft, Apple, Meta, Amazon, Netflix, Honeywell, etc; Indian companies such as Reliance Jio, Flipkart, TCS, Cognizant, L&T, etc; government organizations such as DRDO, Traffic Police Departments, etc; as well as start-ups such as Vehant, Netradyne, SigTuple, etc.

COURSE OUTLINE :

The automatic analysis and understanding of images and videos, a field called Computer Vision, occupies significant importance in applications including security, healthcare, entertainment, mobility, etc. The recent success of deep learning methods has revolutionized the field of computer vision, making new developments increasingly closer to deployment that benefits end users. This course will introduce the students briefly to traditional computer vision topics, before presenting deep learning methods for computer vision. This course delves into the fundamental concepts of neural networks, explores convolutional architectures, and covers the latest advancements in computer vision tasks such as image classification, object detection, segmentation, and generative modeling. Students will engage in hands-on programming assignments, as well as learn to implement and optimize deep learning models. By the end of the course, participants will be equipped with the skills and knowledge to contribute to the rapidly evolving field of computer vision, pushing the boundaries of what machines can perceive and understand.

The course assumes a basic background in machine learning, and may be most useful to students that have also completed basic introductory materials on deep learning.

ABOUT INSTRUCTOR :

Prof. Vineeth N Balasubramanian is an Associate Professor in the Department of Computer Science and Engineering at the Indian Institute of Technology, Hyderabad (IIT-H). He was also the Founding Head of the Department of Artificial Intelligence at IIT-H from 2019-22, and a Fulbright-Nehru Visiting Faculty at Carnegie Mellon University in 2022-23. His research interests include deep learning, machine learning, and computer vision. His research has resulted in over 160 peer-reviewed publications at various international venues, including top-tier venues such as ICML, CVPR, NeurIPS, ICCV, KDD, AAAI, and IEEE TPAMI, with Best Paper Awards at recent venues such as CODS-COMAD 2022, CVPR 2021 Workshop on Causality in Vision, etc. He served as a General Chair for ACML 2022, and serves as a Senior PC/Area Chair regularly for conferences such as CVPR, ICCV, AAAI, IJCAI and ECCV. He is a recipient of the Google Research Scholar Award (2021), NASSCOM AI Gamechanger Award (2022, both Winner and Runner-up), Teaching Excellence Award at IIT-H (2017 and 2021), Research Excellence Award at IIT-H (2022), among others. For more details, please see https://people.iith.ac.in/vineethnb/.

Course Plan:

Week 1: Introduction and Overview

- Course Introduction and Overview
- History (Optional)
- Image Formation (Optional)
- Image Representation
- Linear Filtering, Correlation, Convolution
- Code Walkthroughs

Week 2: Visual Features and Representations

- Edge Detection
- From Edges to Blobs and Corners
- Scale Space, Image Pyramids and Filter Banks
- SIFT and Variants
- Human Visual System (Optional)
- Code Walkthroughs

Week 3: Deep Learning Basics

- Neural Networks: A Review
- Feedforward Neural Networks and Backpropagation
- Gradient Descent and Variants
- Regularization in Neural Networks
- Improving Training of Neural Networks
- Code Walkthroughs

Week 4: Convolutional Neural Networks for Image Classification

- Convolutional Neural Networks: An Introduction
- Backpropagation in CNNs
- CNN Architecture for Image Classification
- Code Walkthroughs

Week 5: Beyond Basic CNNs: Architectures, Finetuning and Visualization

- Evolution of CNN Architectures: VGG, Inception, ResNets
- ResNet Variants, MobileNet, EfficientNet
- Finetuning CNNs
- Visualizing CNNs
- Code Walkthroughs

Week 6: CNNs for Object Detection and Segmentation

- CNNs for Object Detection: Two-stage Models
- CNNs for Object Detection: Single-stage Models
- CNNs for Segmentation
- Code Walkthroughs

Week 7: Recurrent Neural Networks and their use in Vision

- Recurrent Neural Networks: Introduction
- Backpropagation in RNNs
- LSTMs and GRUs
- Video Understanding using CNNs and RNNs
- Code Walkthroughs

Week 8: Attention Models and Transformers

- Attention in Vision Models: An Introduction
- Soft and Hard Attention: Image Captioning
- Self-Attention and Transformers
- Code Walkthroughs

Week 9: Vision Transformers and Applications

- From Transformers to Vision Transformers
- Transformers for Detection
- Transformers for Segmentation
- Code Walkthroughs

Week 10: Deep Generative Models: GANs and VAEs

- Deep Generative Models: An Introduction
- Generative Adversarial Networks
- GAN Hacks and Improvements
- Variational Autoencoders and Disentanglement
- Code Walkthroughs

Week 11: Deep Generative Models: Diffusion Models

- Introduction to Diffusion Models: DDPMs
- Classifier and Classifier-Free Diffusion Guidance
- Text-conditioned Diffusion Models
- Under the Hood: Sampling, Prediction Space, Noise Schedules, Architectures
- Code Walkthroughs

Week 12: Vision-Language Models and Recent Developments

- Self-Supervised Learning: SimCLR
- Contrastive Learning
- Vision-Language Models
- CLIP, BLIP, BLIP-2
- Code Walkthroughs
- Course Conclusion

Additional Material: Miscellaneous Advanced Topics (Additional, Optional)

- Applications and Case Studies
- Few-shot and Zero-shot Learning
- Adversarial Robustness
- Pruning and Model Compression
- Neural Architecture Search
- Recent Developments
    - From VLMs to MM-LLMS: LLaVA, Video ChatGPT, ChatGPT-4V, Gemini 1.5
    - Dall-E-1,2,3 + Imagen