ITCS 6010/8010
Advanced Computer Vision
Syllabus

Course Description

Computer Vision involves teaching machines to interpret and understand the visual world, namely images and videos. In this course, students will learn advanced computer vision algorithms applied in real-world scenarios. As a project-based curriculum, it offers hands-on experience by enabling students to implement industry-relevant computer vision algorithms. Additionally, the course will explore state-of-the-art approaches to visual representation learning.

Course Prerequisites

Unlike ITCS 6169, this course *does not* have any pre-requisite. However, the concepts of Machine Learning will be crucial in understanding the foundations of this course. If you have not mastered some concepts in ITCS 5156/3156, please get in touch with me or the TAs for additional resources/help. Also, the ability to program will be necessary. We will be programming in Python (OpenCV, NumPy, SciKit). For deep learning examples, we will use Pytorch.

Course & Instructor Information

Course Number ITCS 6010/8010
Course Title: Computer Vision
Credit Hours: 3
Online materials and video conferencing: Canvas and Zoom

Instructor: Srijan Das
Contact: sdas24 “at” charlotte.edu
Office Hours: Wednesday 4:00 - 5:00 PM
Location: Woodward 410E
College: College of Computing & Informatics
Department: Computer Science

TA: Vedha Sree Thummagunta (vthummag@charlotte.edu)
Meghna Neelayavalasa (mneelaya@charlotte.edu)
Navya Sri Kathi (nkathi@charlotte.edu)

TA Office Hours: Monday 3:00 PM - 4:00 PM
Wednesday 3:00 PM – 4:00 PM
Friday 3:00 PM - 4:00 PM

Instructional Method

Mode of delivery: in-person
Course type: Lecture

Instruction Method: All the lectures will be delivered in person in the classroom using PowerPoint presentations. There will be no online sessions available for the lectures. One lecture will feature a presentation by an external invited speaker. The lectures will follow a sequential order, adhering to the modules outlined in the course content. After completing each module, students will present research papers related to that particular module. At the end of the course, students are expected to complete a hands-on project and present it in class in pairs.

Course Objectives

Having successfully completed the first module of this course you will be able to:

- Implement basic Deep Neural Networks

Having successfully completed the next two modules of this course, you will be able to demonstrate knowledge and understanding of:

- Current approaches to basic computer vision
- Human and computer vision systems

Having successfully completed the last module of this course you will be able to:

- Develop and evaluate advanced solutions to problems in computer vision
- Demonstrate awareness of the current key research issues in computer vision
- Analyze and design a range of algorithms for computer vision

List of Topics to be covered

- **Module 1: Advanced Deep Neural Networks**
  - Basics of DNNs (Perceptrons)
  - Convolutional Neural Networks (CNNs)
  - Video Understanding

- Recurrent Neural Networks (RNNs)
- Image Captioning
- Attention Mechanism
- Vision Transformers

- **Module 2: Learning Mechanisms**
  - Self-supervised Learning
  - Semi-supervised Learning
  - Multi-tasking & Knowledge Distillation

- **Module 3: Vision-Language Models**
  - Vision-language Models
  - Large Vision Language Models

- **Module 4: Generative AI**
  - Generative Adversarial Networks (GANs)
  - Diffusion Models

- **Module 5**: **Robotic Vision**

Course content with calendar and due date –

	Lecture	**Date**			**Assignment due dates**	
1	Introduction (Basics of DNN and CNN)	*08/20*		Assignment 1	Week 4	
2	Video Understanding	*08/22*		Assignment 2	Week 8	
3	RNNs	*08/27*		Assignment 3	Week 11	
4	Image Captioning	*08/27*				
5	Attention mechanism	*08/29*			**Paper Presentation Days**	
6	Vision Transformers	*09/03, 09/05*	Module 1		*10/01*	ViTs
7	Self-supervised Learning	*09/10, 09/12*			*10/03*	SSLs
8	Semi-supervised Learning	*09/17*			*11/19*	VLMs
9	Multi-tasking and Knowledge Distillation	*09/19*	Module 2		*11/21*	Diffusion Models
10	Vision-Language Models	*10/08*				
11	Large Vision Language Models	*10/10*	Module 3		**Final Project Presentation**	
12	GANs	*10/24*	Module 4		11/26 and 12/3	

13	Diffusion Models	*11/5, 11/7*				
14	Robotic Vision	*11/12*	Module 5	Extra day	*09/24*	

Background reading materials

On Canvas, you will find a collection of basic Python programming examples presented in a Jupyter notebook format. Additionally, a set of slides covering fundamental mathematical concepts will be made available on Canvas.

- Python programming:
  - Python for Everybody
  - Python lecture
- Probability and statistics:
  - Basic probability theory (pp. 12-19) in Pattern Recognition and Machine Learning.
- Linear Algebra:
  - Chapter 2 in DL textbook on Linear Algebra.
  - Chapter 2 on Linear Algebra in Mathematics for Machine Learning.
  - Inderjit Dhillon's Linear Algebra Background
  - Gilbert Strang's Introduction to Linear Algebra
  - Petersen et al.'s The Matrix Cookbook
  - Mike Brookes' Matrix Reference Manual
- Calculus:
  - Basic properties for derivatives, integrals, exponentials, and logarithms.
  - Chapter 4.3 in DL textbook on Numerical Computation.
  - Gilbert Strang's Calculus texbook.

Textbooks

Foundations of Computer Vision by Antonio Torralba, Phillip Isola and William T. Freeman
Computer Vision: Algorithms and Applications by Richard Szeliski
Computer Vision: A Modern Approach by David Forsyth and Jean Ponce

Video Materials

First Principles of Computer Vision – by Shree Nayar
CS231n: Convolutional Neural Networks for Visual Recognition – by Fei Fei Li, Andrej Karpathy, Justin Johnson
Introduction to Computer Vision – Free course at Udacity (also offered at Georgia Tech as CS 6476)
Video Understanding: How to model Time? – by Srijan Das [video]
Tips to attend attention! – by Srijan Das [video]

Assessment

Course Grades

- **3 Assignments** (15%)
- **Group Paper Presentation + Paper Reports** (25%): team presentation of a recent paper
- **Project** (40%): team project with presentation and reports
- **Class participation + Class Quiz** (20%)

Paper Presentation Options		
10/01	CvT: Introducing Convolutions to Vision Transformers	https://arxiv.org/abs/2103.15808
	Swin Transformer: Hierarchical Vision Transformer using Shifted Windows	https://arxiv.org/abs/2103.14030
	Transformer in Transformer	https://proceedings.neurips.cc/paper/2021/file/854d9fca60b4bd07f9bb215d59ef5561-Paper.pdf
	A ConvNet for the 2020s	https://arxiv.org/pdf/2201.03545
10/03	DINOv2: Learning Robust Visual Features without Supervision	https://arxiv.org/abs/2304.07193
	Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture	https://arxiv.org/abs/2301.08243
	VideoMAE V2: Scaling Video Masked Autoencoders with Dual Masking	https://arxiv.org/abs/2303.16727
	OmniMAE: Single Model Masked Pretraining on Images and Videos	https://arxiv.org/abs/2206.08356

11/19	InternVideo: General Video Foundation Models via Generative and Discriminative Learning	https://arxiv.org/pdf/2212.03191
	IMAGEBIND: One Embedding Space To Bind Them All	https://openaccess.thecvf.com/content/CVPR2023/papers/Girdhar_ImageBind_One_Embedding_Space_To_Bind_Them_All_CVPR_2023_paper.pdf
	VideoPrism: A Foundational Visual Encoder for Video Understanding	https://arxiv.org/pdf/2402.13217v1
	MiniGPT-4: Enhancing Vision-Language Understanding with Advanced Large Language Models	https://arxiv.org/pdf/2304.10592
11/21	High-Resolution Image Synthesis with Latent Diffusion Models	https://arxiv.org/pdf/2112.10752
	Residual Denoising Diffusion Models	https://arxiv.org/abs/2308.13712
	Object-Centric Diffusion for Efficient Video Editing	https://arxiv.org/abs/2401.05735
	Latte: Latent Diffusion Transformer for Video Generation	https://arxiv.org/abs/2401.03048v1

Grading Scale

- 90% and above: A
- 80% to 89.99%: B
- 70% to 79.99%: C
- Below 70%: U

The instructor has the authority to adjust the grading scale at any point during the course. Such modifications are likely to occur when the class average falls below a specific threshold.

Course Policies

- **My (instructor's) commitment**: You can expect me to be courteous, respectful, and punctual; be well organized and prepared for class; answer questions clearly and in a non-negative fashion; be available during posted student drop-in / office hours or notify you beforehand if I am unable to keep them.
- **Your commitment**: The best way to learn and master the concepts covered in this course is to engage with the course material and then engage in discussions and activities with your fellow students and your instructional team (i.e., instructor/TAs).
- **Course material**: All course material will be available on Canvas. Please check Canvas periodically. Course materials, including presentations, tests, exams, outlines, and similar materials, are protected by copyright. You are free to take notes and make copies of

course materials for your own educational use. However, you may not, nor must you allow others to reproduce or distribute lecture notes and course materials publicly without my express written consent. This includes providing materials to commercial course material suppliers such as CourseHero, Chegg and other similar services. If you publicly distribute or display or help others publicly distribute or display copies or modified copies of the course material, you may be in violation of University Policy 406 (https://legal.uncc.edu/policies/up406), The Code of Student Responsibility.

- **Course announcements and communication**: I will use Canvas for all announcements and other course related communications. Please make sure to check Canvas on a regular basis to ensure that you do not miss any important information. You should also use Canvas to ask/answer questions pertaining to the class. Please use email communication only for personal/grade related issues.
- **Exercises/Assignments**: You will work on numerous exercises/assignments throughout the semester. These exercises are designed to help you apply and / or get hands-on experience with course concepts. You are encouraged to discuss the questions with your lightweight team, but you must individually submit the exercise/assignment on Canvas (e.g., for questions that require a written response, you may discuss the idea in your team, but must write the final answer in your own words; for exercises/assignments that require you to write a program, you may discuss the algorithm/steps with your team, but you must write the final program on your own). Most exercises/assignments that have auto-graded components allow two attempts, giving you an opportunity to correct mistakes you may make on auto-graded questions during your first attempt. All exercises/assignments have firm due dates, with a grace period where you can still submit, but will incur a small penalty. Having firm due dates will help you stay on track with course material and will help us give you timely feedback on your submissions.
- **Grading**: Teaching assistants will be responsible for grading in-class activities, tests and assignments. If you believe that an error has been made in grading, please bring it to the TA. If you are unable to resolve the issue with the TA, please reach out to me.
- **Engagement**: As mentioned earlier, this class is a collaborative space. A portion of your overall course grade will be determined based on how you engage with me, your TAs and with your lightweight team members. Details about how points will be awarded/deducted will be posted explicitly on Canvas.
- **Missed course work/tests**: If you have any medical emergencies, care-giver concerns or other documented reasons for being unable to submit exercises/assignments on time, being unable to attend a test, etc., please reach out to me (ahead of time whenever possible) and we can work together to find a suitable solution.
- **Academic Integrity**: You are expected to follow the UNC Charlotte Code of Student (http://legal.uncc.edu/policies/up-407) Academic Integrity (http://legal.uncc.edu/policies/up-407) for all activities, exercises, assignments and tests in this course. This includes following all the written / verbal instructions given to you by the instructional team and guidelines for authorized collaboration where relevant. I will have to report all cases of academic misconduct to the Dean of Students' Office.

- **Disability services**: Many students have visible or invisible disabilities. UNC Charlotte is committed to access to education and offers accommodations that allow all students to achieve their full potential. If you have a disability and need academic accommodations, please send me your accommodation letter as early as possible. You are encouraged to meet with me to discuss the accommodations outlined in your letter. For more information on accommodations, contact the Office of Disability Services (https://ds.uncc.edu/) at 704-687-0040 or disability@uncc.edu.
- **Non-discrimination:** No student will be discriminated against in this class based on age, race, nationality, religion, sexual orientation, gender identity/expression, veteran's status, country of origin, or group affiliation. You are expected to behave in a respectful manner towards others in the class, both in virtual and face-to-face settings. Continuous or repeated disrespectful behavior will be considered to be creating a hostile environment, which constitutes a violation to the University Policy 406 (https://legal.uncc.edu/policies/up-406), Code of Student Responsibility. In such a case, you will be referred to the Office of Student Conduct or the Title IX Office. Based on such referral, the Director or designee will determine whether a Formal Charge(s) shall be pursued and whether the Formal Charge(s) constitutes a Minor Violation or a Serious Violation, based on your prior record or facts and circumstances related to the case.
- **Syllabus Revisions**: I may modify standards and requirements set forth in this syllabus if a need arises. I will notify you of any such changes by announcement on Canvas and revision to this syllabus page.

Acknowledgement

This course has been partially inspired by the Computer Vision course by Michael S. Ryoo (Salesforce AI/Stony Brook University). The background reading materials are adapted from Razvan Bunesco's Machine Learning course. This syllabus structure is adapted from ITSC3146 by Harini Ramaprasad (UNC Charlotte).