Deep Learning for Perception

RBE-474X/595-B01-ST

Syllabus

Welcome to **RBE474X/595-B01-ST: Deep Learning for Perception** course at WPI. This course is taught by Prof. Wei Xiao . This course is both an undergraduate (RBE474X) and a graduate course (RBE595-B01-ST) with the same course content but slightly different assignments.

All the class announcements will be made through Piazza (please sign up first if you have not yet). Please use Piazza to contact TAs. Please do NOT contact the TA's or the Instructor via email unless it's an emergency, and do NOT contact the TA's on any social media platform such as Facebook or WhatsApp (please respect their privacy) regarding course content. If you want to have a chat about our research, feel free to reach out after class.

All the student reports will be released publicly online to enable better learning experience for others. We will be also announce top three submissions for each announcement on the website. All the projects and homeworks are to be submitted using Canvas (474X) or Canvas (595). If you find any errors/typo in the course, please post on Piazza.

What is this course about?

This course is about learning the mathematical foundations of deep learning applied to images. Perception stacks in state-of-the-art robots are rapidly adapting the latest advancements in deep learning due to their efficacy and high accuracy. These deep learning-based methods are also accelerable using parallelized hardware such as GPUs that can enable low latency operations of complex tasks such as real-time scene segmentation. You will learn to formulate, develop and implement of deep learning solutions for common computer vision problems in the context of robot perception. The course will cover advanced and state-of-the-art topics such as sim2real, adversarial attacks on neural networks, vision transformers and diffusion models. Additional topics explored in this course include image formation, linear classifiers, neural networks and backpropagation, Convolutional Neural Networks (CNNs), CNN architectures, data generation for sim2real, black- and white-box attacks on neural networks as applied to build state-of-the-art robotic stack. You will gain knowledge about the considerations required to enable a robotic system with the state-of-the-art deep learning toolkit. A unique aspect of this course is that the course is designed to balance theory with applications through projects.

Team

Prof. Wei Xiao (Instructor), Shreyas Devdatta Khobragade (Grader/Teaching Assistant)

Pre-requisites

Programming proficiency (preferably in Python 3), experience with Linux and completion of P0 is a hard pre-requisite. We assume that the students are proficient with basic Linear Algebra (MA2071/2072), calculus (MA1024) and probability (MA2621/2631). Experience working on images programmatically is a big plus.

Class Timings and Location

- Timings: Tuesdays and Fridays 16:00PM to 17:50PM (10/20/2025 – 12/12/2025).
- Class Location: Washburn 323.

Office Hours

- Shreyas Devdatta Khobragade: Tuesday 2:00pm-3:00pm, UH243 Curtain Space.
- Prof. Wei Xiao: ONLY for serious issues with Appointment, UH287.

Software Environment

We will use Python 3 as the programming platform throughout this course along with packages from OpenCV, PyTorch, TensorFlow, Numpy, Scikit and Matplotlib, etc.

Assignments (codebase)

All projects are to be done in groups of TWO to THREE. However, we encourage you to discuss with your peers and not cheat. For clarifications, post a private post on Piazza. We are pro ChatGPT (and other LLMs) in this course as long as the chat prompts are included in the report and more than 30% assignment is not made by LLM prompts.

Assignment Name	Grade Percentage	Deadline
In-class quiz	5	
HW0: Alohomora **(Individual)**	10	10/31/2025
HW1: Nifty Neural Networks! **(Group of 2 or 3)**	15	11/14/2025
HW2: Dreaming Data! **(Group of 2 or 3)**	20	12/05/2025
Course Project: Learning for End-to-End Robot Policies: From Perception to Control **(Group of 2 or 3) – coding, report, and oral presentations**	50	12/12/2025

Late assignments are penalized for 15% per day after the due date. But life is unpredictable (Please let the instructor or TA know if there are special reasons)

Class Schedule

10.21 Class 1: Introduction, Logistics and Sensors

10.24 Class 2: Multi-Layer Perceptron and Backpropagation

10.28 Class 3: NN Tuning, Image Filtering and Convolutional Neural Networks

10.31 Class 4: Advanced CNN Architectures and Image Warping

11.4 No class

11.7 Class 5: Simulation for Data Generation and Sim2Real

11.11 Class 6: Object Detection and Segmentation

11.14 Class 7: Learned Depth: Monocular + Stereo

11.18 Class 8: Vision Transformers, Can We Trust Neural Networks?

11.21 Class 9: Single Pixel Attacks, Patch Based Attacks

11.25 Class 10: Generative Models: VAEs, GANs, Attacking GANs

11.28 No class

12.2 Class 11: Advanced Generative Models: Diffusion Models

12.5 Class 12: Advanced Generative Models++: Multi-modal Generative Deep Learning

12.9 Presentation 1 (Group-based)

12.12 Presentation 2 + Course Summary (Group-based)

Reference Books

All concepts will be covered in class lecture, and in the lecture notes. However, we also recommend the following books as good references:

- Deep Learning Book, MIT Press, Ian Goodfellow, Yoshua Bengio, and Aaron Courville, 2016.
- Computer Vision: Algorithms and Applications, Springer, Richard Szeliski, 2010.
- Digital Image Processing, Prentice Hall, Rafael Gonzalez, and Richard Woods, 2008.

Furthermore, refer to the slides from courses in the acknowledgement section.

Acknowledgements

This course is developed by adapting some of the best parts of each of the courses at multiple universities and these resources are linked below and you are encouraged to look at their content to learn from them as well. **The goal of this course is to be the best undergraduate deep learning course in the world.**

- University of Maryland, College Park's CMSC733
- University of Minnesota, Twin Cities' CSCI5563
- Stanford University's CS231n
- University of Maryland, College Park's CMSC764
- WPI, RBE 474X/595 2024