ECE 499 Directed Study / Independent Research Form

ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY OF MICHIGAN

Electrical and Computer Engineering
Undergraduate Advising Office
3415 EECS Bldg., eceadvising@umich.edu, 734.763.2305

SEE DIRECTIONS ON BACK

Once completed, this form should be returned
to the ECE Undergraduate Advising Office via email at eceadvising@umich.edu

Student Name: Jiaming Zeng	
UMID: 08866142	Unique Name: zjiaming
Date: 1/19/2022	
Major (e.g., CE, EE): Data Science, LSA	
Faculty Mentor Name: Prof.Atul Prakash	
EECS 499 Section Number: 137	Credit Hours: 4
Project Title: Adversarial Attack Detection and Model Improvement Based on Semantic Similarity	

You must attach a 1-2 paragraph description of your project. You will need to discuss this thoroughly with your faculty director prior to completing this form. After the description of the project, answer the following questions:

1. How will you be evaluated? (Report, oral presentation, exam?)
2. Will materials from other classes you have taken be used in this project? If so, list the classes and materials/topics that will be used.
3. How often will you meet with your faculty director?
4. How will the completion of your project be determined (what are the deliverables)?

NOTE: The prerequisite for this course is senior standing (85 CTP or above). If you do not meet the prerequisite you will need to complete a permission form in the EECS advising office or register for EECS 399.

Jiaming Zeng 1/19/2021
Student Date

I certify that the above information is correct, and that this directed study course replaces an equivalent number of credits of a regularly offered 400-level EECS course in terms of content and learning experience.

Faculty Director Date

How to set up Directed Research / Independent Study
(EECS 499)

---

1. Fill out this form completely.

2. Attach a brief description of the independent study project.

3. Have the form signed by the professor with whom you will be working (this is your faculty mentor).

4. Submit an **electronic copy** of the documents to the Undergraduate Advising Office at eceadvising@umich.edu.

5. Register for the appropriate section* of EECS 499.

*You must register for the correct section of EECS 499. Each faculty member in the EECS department has a section number assigned to him or her. A list of these section numbers is posted on the bulletin board in the EECS Undergraduate Advising Office and on the web at:

https://drive.google.com/drive/folders/15uH3kHQ9Ofmtzbt-HVkSs_8gvE_y2M4O

A maximum of 4 credits of Independent Study (including EECS 499 or other independent study work from other departments) may be applied toward Flexible Technical Electives. Anything beyond 4 credits will be applied toward General Electives.

Paid employment and/or financially-compensated project work is not eligible for EECS 499 credit.

Project Proposal

Jiaming Zeng

This project would be based on Professor Prakash's research on physical-world attacks on deep learning models. While deep learning models are capable of classifying images and applied to real-world applications, it is vulnerable to adversarial attacks from small-magnitude perturbations added to the input data. Input would become mislabeled by the deep learning models with the small perturbation and the model would fail. In order to detect an adversarial attack and improve the accuracy of the final result of deep learning models, this project proposes the idea of pre-filtering the input with extra data analysis methods based on semantic similarity. It also has to be precise for filtering all the images which would be mislabeled by the original model.

In this project, machine learning and statistical analyzation techniques I learned from STATS 415 and STATS 406 might be applied to generate a better filter. Every Wednesday, Professor Prakash, Ryan Feng, a PhD Student and I would meet over zoom or in person. In the end of the semester, there would be a report describing the result of the filter and whether it would work. The deliverables in the report would be experimental results and summarization from different semantic measures from at least one dataset including GTSRB, and the evaluation of their effectiveness on adaptive adversarial attacks. The source code will be provided under the MIT open source license.