

Multimodal Deep Learning Theories and Applications

Jin-Hwa Kim^{1,2} Jiyoung Lee¹

¹NAVER AI Lab ²Seoul National University

Fall 2023

Overview

This course is designed for graduate students interested in developing theories and applications of multimodal deep learning. The course will cover core concepts and fundamental theories that are applied in various enterprise applications. Assuming prior knowledge in machine learning and related topics, the course will explore multimodal deep learning from vision and language to audio and speech using deep neural networks and the techniques used in deep learning literature. If time permits, an emerging topic of neural graphics will also be discussed, focusing on studies on the NeRFs in an aspect of multi-view representation learning.

1 Target Audience

- Graduate students with prior knowledge in machine learning and related courses who are looking to specialize or broaden in multimodal deep learning.

2 Course Objectives

- Gain a deep understanding of multimodal deep learning theories and applications
- Explore various modalities such as vision, language, audio, and speech in deep learning
- Stay updated with emerging topics in neural graphics, specifically NeRFs

3 General Information

3.1 Grade

- Quiz (30%)
 - Five times at the beginning of class. Two to three weeks intervals depending on contents.
 - If you miss or fail to take a quiz, you will receive a zero point for that quiz.
- Project (proposal 30% + final 40% = 70%)
 - Develop a new research idea for multimodal deep learning.
 - Each team consists of *two* to *three* collaborators to prepare two (proposal and final) presentations.
 - Presentation must include the following contents: problem statement, data collection or preparation, model architecture, and theoretical and experimental analysis.
 - Rubrics for the assessments will be notified before the presentation dates.

*

4 Schedule

The following is a tentative schedule for the course:

Week	Topics	Note
1	Introduction	JH (Jin-Hwa Kim)
2	Multimodal tasks and datasets	JY (Jiyoung Lee)
3	Multimodal representation learning	JH
4	Multimodal reasoning	JY
5	Pre-training paradigm	JY
6	Project proposal	JH
7	Multimodal generation 1	Guest lecturer: Yunjey Choi
8	Multimodal generation 2	JY
9	Ethical AI	Guest lecturer: Hwaran Lee
10	Multimodal interpretability	JY
11	Neural graphics	JH
12	Advances in neural graphics	JH
13	Invited talk	Guest lecturer: Sangdoo Yun
14	Invited talk	Guest lecturer Eunwoo Song
15	Final presentations	JH
16	Final presentations	JY

Note: The course content and sequence may be adjusted based on the duration of the lecture and the interests of the students. Additional readings, assignments, and projects may also be included to enhance the learning experience.