CS6187: VISION AND LANGUAGE

Effective Term
Semester A 2025/26

Part I Course Overview

Course Title
Vision and Language

Subject Code
CS - Computer Science
Course Number
6187

Academic Unit
Computer Science (CS)

College/School
College of Computing (CC)

Course Duration
One Semester

Credit Units
3

Level
P5, P6 - Postgraduate Degree

Medium of Instruction
English

Medium of Assessment
English

Prerequisites
CS5187 Vision and Image or
CS5286 Algorithms and Techniques for Web Searching or
CS5487 Machine Learning: Principles and Practice

Precursors
Nil

Equivalent Courses
Nil

Exclusive Courses
Nil

Part II Course Details

Abstract

This course introduces algorithms and techniques for integration of computer vision and natural language processing for innovative applications, such as robot dialog system, image/video captioning, cross-media multimedia search and question-answering. The course will discuss the latest technologies in bridging the gap between vision and language, with topics ranging from machine translation, feature extraction and learning, to design of deep neural network architecture.

Course Intended Learning Outcomes (CILOs)

	CILOs	Weighting (if app.)	DEC-A1	DEC-A2	DEC-A3
1	Describe techniques in visual feature extraction, learning and representation.			x	
2	Explain the process of natural language processing, ontology construction and knowledge inferencing.			x	
3	Perform critical assessment on the effectiveness of rule-based, statistical-based and machine learning based translation models.		x		
4	Apply computer vision and natural language processing for innovative applications.		x	x	x

A1: Attitude
Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability
Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to real-life problems.

A3: Accomplishments
Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

Learning and Teaching Activities (LTAs)

	LTAs	Brief Description	CILO No.	Hours/week (if applicable)
1	Lecture	Students will engage on topics about the introduction of computer vision natural language processing machine translation, and related real-world applications such as image/video captioning and robot dialog system.	1, 2, 3	2 hours/ week

2	Tutorial	Students will work on a different problem set each week during the tutorial sessions, through which they can discover the differences and commonalities in vision, speech and text, and the solutions to bridge the gap between different domains. They will also be invited to present their solutions, and the class will be encouraged to provide comments.	1, 2, 3	1 hour/ week
3	Assignment	The students will implement basic vision and language algorithms for processing different modalities, and translation models in modelling or learning the embedded relationship between different modalities. In this way, students can observe the characteristics and perform critical assessment of different machine translation models.	3	
4	Project	The students will create a new system design or solution to solve a real-world problem. The students will apply the principles they have learnt from the course for innovative application.	4	

Assessment Tasks / Activities (ATs)

	ATs	CILO No.	Weighting (%)	Remarks ("-" for nil entry)	Allow Use of GenAI?
1	Assignments	1, 2, 3	30	Expect to have at least two assignments.	Yes
2	Project	4	40	-	Yes

Continuous Assessment (%)

70

Examination (%)

30

Examination Duration (Hours)
2

Minimum Examination Passing Requirement (%)
30

Additional Information for ATs
For a student to pass the course, at least 30% of the maximum mark for the examination must be obtained.

Assessment Rubrics (AR)

Assessment Task
Assignment (for students admitted before Semester A 2022/23 and in Semester A 2024/25 & thereafter)

Criterion
The ability to implement and assess the effectiveness of different algorithms and techniques.

Excellent
(A+, A, A-) High

Good
(B+, B, B-) Significant

Fair
(C+, C, C-) Moderate

Marginal
(D) Basic

Failure
(F) Not even reaching marginal level

Assessment Task
Project (for students admitted before Semester A 2022/23 and in Semester A 2024/25 & thereafter)

Criterion
The ability and creativity in applying appropriate algorithms and techniques for real-world applications that require integration of vision and language.

Excellent
(A+, A, A-) High

Good
(B+, B, B-) Significant

Fair
(C+, C, C-) Moderate

Marginal
(D) Basic

Failure
(F) Not even reaching marginal level

Assessment Task
Examination (for students admitted before Semester A 2022/23 and in Semester A 2024/25 & thereafter)

Criterion
The extent to which the students can understand the differences and commonalities between different modalities and ability to perform appropriate algorithms and techniques for integration of them, and evaluate their performances.

Excellent
(A+, A, A-) High

Good
(B+, B, B-) Significant

Fair
(C+, C, C-) Moderate

Marginal
(D) Basic

Failure
(F) Not even reaching marginal level

Assessment Task
Assignment (for students admitted from Semester A 2022/23 to Summer Term 2024)

Criterion
The ability to implement and assess the effectiveness of different algorithms and techniques.

Excellent
(A+, A, A-) High

Good
(B+, B) Significant

Marginal
(B-, C+, C) Moderate to Basic

Failure
(F) Not even reaching marginal levels

Assessment Task
Project (for students admitted from Semester A 2022/23 to Summer Term 2024)

Criterion
The ability and creativity in applying appropriate algorithms and techniques for real-world applications that require integration of vision and language.

Excellent
(A+, A, A-) High

Good
(B+, B) Significant

Marginal

(B-, C+, C) Moderate to Basic

Failure

(F) Not even reaching marginal levels

Assessment Task

Examination (for students admitted from Semester A 2022/23 to Summer Term 2024)

Criterion

The extent to which the students can understand the differences and commonalities between different modalities and ability to perform appropriate algorithms and techniques for integration of them, and evaluate their performances.

Excellent

(A+, A, A-) High

Good

(B+, B) Significant

Marginal

(B-, C+, C) Moderate to Basic

Failure

(F) Not even reaching marginal levels

Part III Other Information

Keyword Syllabus

Visual understanding, natural language processing, speech transcription, knowledge graph representation and reasoning, canonical correlation analysis, feature extraction and learning, deep neural network, image / video tagging and captioning, zero-shot learning, question-answering.

Reading List

Compulsory Readings

	Title
1	Nil

Additional Readings

	Title
1	Yoshua Bengio, Ian Goodfellow, Aoran Courville, Deep Leaning MIT Press, 2014.
2	Richard Szeliski, Computer Vision: Algorithms and Applications Springer; 2011.
3	Daniel Jurafsky, James H. Martin, Speech and Language Processing Prentice Hall, 2009.
4	Philipp Koehn, Statistical Machine Translation Cambridge University Press, 2010.