

**City University of Hong Kong
Course Syllabus**

**offered by Department of Computer Science
with effect from Semester A 2022/23**

Part I Course Overview

Course Title: Vision and Language

Course Code: CS6187

Course Duration: One semester

Credit Units: 3

Level: P6

Medium of Instruction: English

Medium of Assessment: English

Prerequisites: CS5187 Vision and Image or
(Course Code and Title) CS5286 Algorithms and Techniques for Web Searching or
CS5487 Machine Learning: Principles and Practice

Precursors: Nil
(Course Code and Title)

Equivalent Courses: Nil
(Course Code and Title)

Exclusive Courses: Nil
(Course Code and Title)

Part II Course Details

1. 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.

2. Course Intended Learning Outcomes (CILOs)

(CILOs state what the student is expected to be able to do at the end of the course according to a given standard of performance.)

No.	CILOs	Weighting (if applicable)	Discovery-enriched curriculum related learning outcomes (please tick where appropriate)		
			A1	A2	A3
1.	Describe techniques in visual feature extraction, learning and representation.			✓	
2.	Explain the process of natural language processing, ontology construction and knowledge inferencing.			✓	
3.	Perform critical assessment on the effectiveness of rule-based, statistical-based and machine learning based translation models.		✓		
4.	Apply computer vision and natural language processing for innovative applications.		✓	✓	✓
		100%			

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 self-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.

3. Teaching and Learning Activities (TLAs)

(TLAs designed to facilitate students' achievement of the CILOs.)

TLA	Brief Description	CILO No.				Hours/week (if applicable)
		1	2	3	4	
Lecture	The lecture will focus on the introduction of computer vision, natural language processing, machine translation, and related real-world applications such as image/video captioning and robot dialog system.	✓	✓	✓		2 hours/ week
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 hour/ week
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.			✓		
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.				✓	
Final Exam	Final exam will include questions to assess the capability of students in performing and assessing the effectiveness of different algorithms for integration of vision and language.	✓	✓	✓		

4. Assessment Tasks/Activities (ATs)

(ATs are designed to assess how well the students achieve the CILOs.)

Assessment Tasks/Activities	CILO No.				Weighting	Remarks
	1	2	3	4		
Continuous Assessment: 70%						
Assignments	✓	✓	✓		30%	Expect to have at least two assignments.
Project				✓	40%	
Examination [^] : 30% (duration: 2 hours)						
Final Exam	✓	✓	✓		30%	
					100%	

[^] For a student to pass the course, at least 30% of the maximum mark for the examination must be obtained.

5. Assessment Rubrics

(Grading of student achievements is based on student performance in assessment tasks/activities with the following rubrics.)

Applicable to students admitted in Semester A 2022/23 and thereafter

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B)	Marginal (B-, C+, C)	Failure (F)
1. Assignment	The ability to implement and assess the effectiveness of different algorithms and techniques.	High	Significant	Moderate to Basic	Not even reaching marginal levels
2. Project	The ability and creativity in applying appropriate algorithms and techniques for real-world applications that require integration of vision and language.	High	Significant	Moderate to Basic	Not even reaching marginal levels
3. Examination	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.	High	Significant	Moderate to Basic	Not even reaching marginal levels

Applicable to students admitted before Semester A 2022/23

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B, B-)	Fair (C+, C, C-)	Marginal (D)	Failure (F)
1. Assignment	The ability to implement and assess the effectiveness of different algorithms and techniques.	High	Significant	Moderate	Basic	Not even reaching marginal level
2. Project	The ability and creativity in applying appropriate algorithms and techniques for real-world applications that require integration of vision and language.	High	Significant	Moderate	Basic	Not even reaching marginal level
3. Examination	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.	High	Significant	Moderate	Basic	Not even reaching marginal level

Part III Other Information (more details can be provided separately in the teaching plan)

1. Keyword Syllabus

(An indication of the key topics of the course.)

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.

2. Reading List

2.1 Compulsory Readings

(Compulsory readings can include books, book chapters, or journal/magazine articles. There are also collections of e-books, e-journals available from the CityU Library.)

2.2 Additional Readings

(Additional references for students to learn to expand their knowledge about the subject.)

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