UAB
Universitat Autònoma de Barcelona

Visual Recognition

Code: 44777
ECTS Credits: 9

2025/2026

Degree	Type	Year
Computer Vision	OB	1

Contact

Name: Maria Vanrell Martorell

Email: maria.vanrell@uab.cat

Teachers

Joan Serrat Gual

Ernest Valveny Llobet

Petia Ivanova Radeva

Dimosthenis Karatzas

Joost van De Weijer

Lei Kang

Carlos Boned Riera

German Barquero García

Julio Cezar Silveira Jacques-Junior

Jordi Casas Roma

Luis Gomez Bigorda

(External) David Vázquez

Teaching groups languages

Prerequisites

Degree in Engineering, Maths, Physics or similar.
Course 3: "Machine learning for computer vision"

Objectives and Contextualisation

Module Coordinator: Dr. Maria Vanrell / Dr. Julio Silveira

In Computer Vision, visual recognition corresponds to the task of explaining the content of an image in terms of "What is it?" "where is this?". The answer to these questions is usally a class label corresponding to the object or object types in the image, a tight bounding box containing the object in question, or, at a finer level, the region (pixels) that is its outline. These tasks are called, respectively, image classification, object detection and semantic segmentation. A question is "give me objects like this one", that requires learning a similar metric between images, even in the case come from different modalities, like sketches and photographs, through the so called encoder-decoder architectures. VR module covers neural network architectures addressing these four types of tasks. And, as a practical complement, methods to implement them.

Specifically, in this module we give to the student an overview of the latest methods based on deep learning techniques to solve visual recognition problems. The final aim is the understanding of complex scenes to build feasible systems for automatic image understanding able to answer the complex question of what objects and where are these objects in a complex scene.

Having addressed the task of classification in previous course, the students will learn a large family of successful architectures of deep convolutional networks that have been proved to solve the visual tasks of detection and segmentation and recognition. In addition to these two visual tasks, this module addresses also advanced topics in deep learning.

Learning Outcomes

1. CA02 (Competence) Design all the components and their interconnection for a complete system of visual recognition.
2. CA06 (Competence) Achieve the objectives of a project of vision carried out in a team.
3. KA05 (Knowledge) Identify the basic problems to be solved in a case of visual recognition of a scene.
4. KA13 (Knowledge) Provide the best modelling of a problem of visual recognition, such as classification, detection or semantic segmentation.
5. SA05 (Skill) Solve a problem of visual recognition training a deep neural network architecture and evaluate the results.
6. SA11 (Skill) Define the best data sets for training visual recognition architecture.
7. SA15 (Skill) Prepare a report that describes, justifies and illustrates the development of a project of vision.
8. SA17 (Skill) Prepare oral presentations that allow debate of the results of a project of vision.

Content

1. Self-attention and Transformers
2. Object detection
3. Semantic and instance segmentation
4. Transfer learning: pre-train, domain adaptation, un-supervised, self-supervised.
5. Metric learning
6. Generative models: GAN, VAE
7. Reinforcement Learning
8. Continual Learning
9. Graph Neural Networks
10. Vision and Language
11. Multimodal Deep Learning
12. Difusion Models
13. Human-centric recognition
14. Affective and trustworthy AI

Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
lectures	35	1.4	CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17, CA02
Type: Supervised			
project	10	0.4	CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17, CA02
Type: Autonomous			
homework, exercises	170	6.8	CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17, CA02

The learning methdology is based on the lectures and exercises but mainly on the project, which is developed during the whole module. It consists in solving some tasks for scene understanding applied to autonomous driving. The goal is to learn the basic concepts and techniques to build deep neural networks to detect, segment and recognize specific objects, focusing on images recorded by an on-board vehicle camera for autonomous driving.

The learning objectives are using different a deep learning (DL) programming frameworks (at present, PyTorch) and basic DL methods such as feed forward networks (MLP) and Convolutional Neural Networks (CNN). It includes the understanding of standard networks for detection (RCNN, Fast RCNN, Faster RCNN, YOLO) and segmentation (FCN, SegNet, UNET). The students will learn through a project based methodology using modern collaborative tools at all stages of the project development.

Students will acquire the skills for the tasks of designing, training, tuning and evaluating neural networks to solve the problem of automatic image understanding.

All this is done through three formats:

1. Supervised sessions: lectures where the instructors will explain general contents about the different topics. They will be used to solve the project and/or proposed exercises.
2. Directed sessions:
   1. Project Sessions, where the problems and goals of the projects will be presented and discussed, students will interact with the project coordinator about problems and ideas on solving the project. Additionally, the students give oral presentations about how have they solved the project and report results (approx once per week)
   2. Exam Session, where the students are evaluated individually. Knowledge achievements and problem-solving skills
4. Autonomous work :
   study and work with the materials derived from the lectures, plus solving some small practical exercises to better understand theoretical lectores that arenot directly involved in the project solution
   work in groups to solve the problems of the projects with deliverables: code, reports,oral presentations, exercises

Annotation: Within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Assessment

Continous Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
attendance	0.05	0.5	0.02	CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17
exam	0.4	2.5	0.1	CA02, KA05, KA13, SA05, SA11
project	0.55	7	0.28	CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17

The final marks for this module will be computed with the following formula:

Final Mark= 0.4 x Exam+ 0.55 x Project + 0.05 x Attendance

where

Exam:is the mark obtained in the Module Exam (must be >=3). This mark can be increased by getting extra points given by delivered exercises in specific lectures, but only if Exam Mark is greater than 3.

Attendance: is the mark derived from the control of attendance at lectures (minimum 70%).

Project:is the mark provided by the project coordinator based on the weekly follow-up of the project and deliverables (must be >=5). All accordingly with specific criteria such as:

- Participation in discussion sessions and in team work (inter-member evaluations)

- Delivery of mandatory and optional exercises.

- Code development (style, comments, etc.)

- Report (justification of the decisions in your project development)

- Presentation (Talk and demonstrations on your project)

Only those students that fail (Final Mark< 5.0) can do a retake exam.

Bibliography

Generic references :

1. Deep Learning. Ian Goodfellow, Yoshua Bengio, Aaron Courville. MIT Press, 2016
2. Dive into deep learning. Aston Zhang, Zack C. Lipton, Mu Li, Alex J. Smola. https://d2l.ai/

Most of the content is related to the state of the art in the different topics, so there are no books published but survey and research papers specific of each one that will be selected by the lecturers.

Software

Tools for Python programming with special attention to Computer Vision and Pythorch libraries

Groups and Languages

Please note that this information is provisional until 30 November 2025. You can check it through this link. To consult the language you will need to enter the CODE of the subject.

Name	Group	Language	Semester	Turn
(PLABm) Practical laboratories (master)	1	English	second semester	morning-mixed
(PLABm) Practical laboratories (master)	2	English	second semester	morning-mixed
(TEm) Theory (master)	1	English	second semester	morning-mixed