

## Reconocimiento Visual

Código: 44777  
Créditos ECTS: 9

2024/2025

| Titulación                                      | Tipo | Curso |
|---|------|-------|
| 4318299 Visión por Computador / Computer Vision | OB   | 0     |

### Contacto

Nombre: María Isabel Vanrell Martorell

Correo electrónico: maria.vanrell@uab.cat

### Equipo docente

Joan Serrat Gual

Ernest Valveny Llobet

Petia Ivanova Radeva

Dimosthenis Karatzas

Joost van De Weijer

Lei Kang

German Barquero García

Julio Cezar Silveira Jacques-Junior

Jordi Casas Roma

Luis Gomez Bigorda

(Externo) David Vázquez

### Idiomas de los grupos

Puede consultar esta información al [final](#) del documento.

### Prerrequisitos

Grado en ingeniería, matemáticas, física o similar.

Asignatura C2: "Machine learning for computer vision"

### Objetivos y contextualización

Coordinador del Módulo: Dr. Joan Serrat Gual

En visión por computador, el reconocimiento visual corresponde a la tarea de explicar el contenido de una imagen en términos de "¿Qué es?" "¿Dónde está esto?". La respuesta a estas preguntas suele ser una

etiqueta de clase correspondiente al objeto o tipos de objeto en la imagen, un cuadro delimitador que contiene el objeto en cuestión o, en un nivel más fino, la región (píxeles) que es su contorno. Estas tareas se denominan, respectivamente, clasificación de imágenes, detección de objetos y segmentación semántica. Una pregunta es "dame objetos como este", que requiere aprender una métrica similar entre imágenes, incluso en el caso de que provengan de diferentes modalidades, como bocetos y fotografías, a través de las llamadas arquitecturas codificador-decodificador. El módulo VR cubre arquitecturas de redes neuronales que abordan estos cuatro tipos de tareas. Y, como complemento práctico, métodos para implementarlos.

Específicamente, en este módulo brindamos al estudiante una visión general de los últimos métodos basados en técnicas de aprendizaje profundo para resolver problemas de reconocimiento visual. El objetivo final es la comprensión de escenas complejas para construir sistemas viables para la comprensión automática de imágenes capaces de responder a la compleja pregunta de qué objetos y dónde están estos objetos en una escena compleja.

Habiendo abordado la tarea de clasificación en cursos anteriores, los estudiantes aprenderán una gran familia de arquitecturas exitosas de redes convolucionales profundas que han demostrado resolver las tareas visuales de detección, segmentación y reconocimiento. Adicionalmente, se abordan otros temas avanzados de aprendizaje profundo.

## Resultados de aprendizaje

1. CA02 (Competencia) Diseñar todos los componentes y su interconexión para un sistema completo de reconocimiento visual.
2. CA06 (Competencia) Conseguir los objetivos de un proyecto de visión realizado en equipo.
3. KA05 (Conocimiento) Identificar los problemas básicos que se deben solucionar en un problema de reconocimiento visual de una escena.
4. KA13 (Conocimiento) Proporcionar la mejor modelización de un problema de reconocimiento visuales, tales como la clasificación, la detección o la segmentación semántica.
5. SA05 (Habilidad) Resolver un problema de reconocimiento visual entrenando una arquitectura de red neuronal profunda y evaluar los resultados.
6. SA11 (Habilidad) Definir los mejores conjuntos de datos para entrenar arquitecturas de reconocimiento visual.
7. SA15 (Habilidad) Preparar un informe que describa, justifique e ilustre el desarrollo de un proyecto de visión.
8. SA17 (Habilidad) Preparar presentaciones orales que permitan debatir los resultados del desarrollo de un proyecto de visión.

## Contenido

1. Atención propia y Transformers
2. Detección de objetos
3. Segmentación Semántica de imágenes e instancias
4. Aprendizaje de Métricas
5. Aprendizaje por Transferencia: pre-entrenamiento, adaptación de dominio, no-supervisado, auto-supervisado.
6. Arquitecturas para la generación de imágenes: GAN y VAE
7. Aprendizaje por reforzamiento
8. Aprendizaje Continuo
9. Redes Neuronales Gráficas
10. Lenguaje y Visión
11. Aprendizaje profundo Multimodal
12. Modelos de difusión

- 13. Reconocimiento centrado en humanos
- 14. IA afectiva y fiable.

## Actividades formativas y Metodología

| Título              | Horas | ECTS | Resultados de aprendizaje                            |
|---------------------|-------|------|--|
| Tipo: Dirigidas     |       |      |  |
| clases de teoría    | 35    | 1,4  | CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17, CA02 |
| Tipo: Supervisadas  |       |      |  |
| proyecto            | 10    | 0,4  | CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17, CA02 |
| Tipo: Autónomas     |       |      |  |
| estudio, ejercicios | 170   | 6,8  | CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17, CA02 |

La metodología de aprendizaje se basa en las conferencias y ejercicios, pero principalmente en el proyecto, que se desarrolla durante todo el módulo. Consiste en resolver algunas tareas para la comprensión de la escena aplicada a la conducción autónoma. El objetivo es aprender los conceptos y técnicas básicos para construir redes neuronales profundas para detectar, segmentar y reconocer objetos específicos, centrándose en las imágenes grabadas por una cámara de vehículo a bordo para la conducción autónoma.

Los objetivos de aprendizaje utilizan diferentes marcos de programación de aprendizaje profundo (DL) (en la actualidad, PyTorch) y la construcción de bloques/capas de diferentes arquitecturas (MLP, CNN, Transformers visuales). Incluye la comprensión de redes estándar para detección (RCNN, Fast RCNN, Faster RCNN, YOLO) y segmentación (FCN, SegNet, UNET) y más aplicaciones.

Los estudiantes adquirirán las habilidades para las tareas de diseño, capacitación, ajuste y evaluación de redes neuronales para resolver el problema de la comprensión automática de imágenes.

Todo esto se realiza a través de tres formatos:

- Sesiones supervisadas: conferencias donde los instructores explicarán contenidos generales sobre los diferentes temas. Se utilizarán para resolver el proyecto y / o los ejercicios propuestos.
- Sesiones dirigidas:
  - Sesiones de proyectos, donde se presentarán y debatirán los problemas y objetivos de los proyectos, los estudiantes interactuarán con el coordinador del proyecto sobre problemas e ideas para resolver el proyecto. Además, los estudiantes realizan presentaciones orales sobre cómo han resuelto el proyecto e informan los resultados (aproximadamente una vez por semana)
  - Sesión de examen, donde los estudiantes son evaluados individualmente. Logros de conocimiento y habilidades para resolver problemas.
- Trabajo autónomo:
  - Estudiar y trabajar con los materiales derivados de las conferencias, además de resolver algunos pequeños ejercicios prácticos para comprender mejor a los lectores teóricos que no están directamente involucrados en la solución del proyecto.
  - Trabajar en grupos para resolver los problemas de los proyectos con entregables: código, informes, presentaciones orales, ejercicios

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

## Evaluación

### Actividades de evaluación continuada

| Título                    | Peso | Horas | ECTS | Resultados de aprendizaje                      |
|---------------------------|------|-------|------|--|
| asistencia a las sesiones | 0.05 | 0,5   | 0,02 | CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17 |
| examen                    | 0.4  | 2,5   | 0,1  | CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17 |
| proyecto                  | 0.55 | 7     | 0,28 | CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17 |

La nota final se calculará mediante la siguiente fórmula :

$$\text{Nota final} = 0.4 \times \text{Examen} + 0.55 \times \text{Proyecto} + 0.05 \times \text{Asistencia}$$

donde

Examen: es la nota obtenida del examen (debe ser  $\geq 3$ ). Puede ser incrementada con puntos extra correspondientes a los ejercicios propuestos en las clases de algunos temas, pero sólo si la nota de examen es como mínimo de 3.

Asistencia: nota derivada del control de asistencia a las clases (mínimo 70%)

Proyecto: nota otorgada por coordinador del proyecto basada en el seguimiento que hace semanalmente y en las entregas del proyecto (debe ser  $\geq 5$ ). Todo ello de acuerdo con criterios específicos como :

- Participación y discusión en las sesiones y trabajo en grupo (evaluaciones entre pares)
- Entregas de partes obligatorias y opcionales
- Código desarrollado (estilo, comentarios, etc.)
- Informe escrito (justificación de las decisiones de desarrollo)
- Presentación oral y demostración

Sólo los estudiantes que han suspendido (nota final  $< 5.0$ ) podrán hacer el examen de recuperación.

## Bibliografía

Referencias genéricas :

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/>

La mayoría de los contenidos están relacionados con el estado del arte en los diferentes temas, así que no existen libros publicados sino artículos de revisión (surveys) y de investigación de cada uno de los temas, que serán seleccionados por los profesores.

## Software

Entorno de programación en Python con especial atención a las librerías de visión por computador y Pythorch

## Lista de idiomas

| Nombre                                    | Grupo | Idioma | Semestre             | Turno        |
|---|-------|--------|----------------------|--------------|
| (PLABm) Prácticas de laboratorio (máster) | 1     | Inglés | segundo cuatrimestre | mañana-mixto |
| (PLABm) Prácticas de laboratorio (máster) | 2     | Inglés | segundo cuatrimestre | mañana-mixto |
| (TEm) Teoría (máster)                     | 1     | Inglés | segundo cuatrimestre | mañana-mixto |