

**Reconocimiento Visual**

Código: 43088  
Créditos ECTS: 6

| Titulación                                      | Tipo | Curso | Semestre |
|---|------|-------|----------|
| 4314099 Visión por Computador / Computer Vision | OB   | 0     | 2        |

**Contacto**

Nombre: Maria Vanrell Martorell

Correo electrónico: maria.vanrell@uab.cat

**Equipo docente**

Joan Serrat Gual

Ernest Valveny Llobet

Dimosthenis Karatzas

Ruben Perez Tito

Andrés Mafla Delgado

Lluis Gomez Bigorda

**Uso de idiomas**

Lengua vehicular mayoritaria: inglés (eng)

**Equipo docente externo a la UAB**

Adriana Romero

David Vázquez

Michal Drozdal

Petia Radeva

**Prerequisitos**

- Grado en ingeniería, matemáticas, física o similar.
- Módulo 2 "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 el módulo M2, 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. Además de estas dos tareas visuales, este módulo también aborda temas avanzados de aprendizaje profundo, tales como arquitecturas para la generación de imágenes (GAN y VAE) más arquitecturas de codificador-decodificador para aplicaciones multimodales.

## Competencias

- Asumir tareas de responsabilidad en la gestión de la información y el conocimiento.
- Comprender, analizar y sintetizar los conocimientos avanzados que existen en el área, así como proponer ideas innovadoras.
- Conceptualizar alternativas de soluciones complejas a problemas de visión y crear prototipos que demuestren la validez del sistema propuesto.
- Identificar los conceptos y aplicar las técnicas fundamentales más adecuadas para la solución de los problemas básicos de la visión por computador.
- Planificar, desarrollar, evaluar y gestionar soluciones a proyectos en los diferentes ámbitos de la visión por computador.
- Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
- Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.
- Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.
- Seleccionar las herramientas software y los conjuntos de entrenamiento más adecuados para desarrollar las soluciones a los problemas de visión por computador.
- Trabajar en equipos multidisciplinares.

## Resultados de aprendizaje

1. Asumir tareas de responsabilidad en la gestión de la información y el conocimiento.
2. Comprender, analizar y sintetizar los conocimientos avanzados que existen en el área, así como proponer ideas innovadoras.
3. Identificar las mejores representaciones que se puedan definir para la resolución de problemas de reconocimiento visual tanto de objetos como de escenas.
4. Identificar los problemas básicos que se deben resolver en reconocimiento de objetos y escenas, así como los algoritmos específicos.
5. Planificar, desarrollar, evaluar y gestionar una solución a un problema concreto de reconocimiento visual.
6. Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
7. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.
8. Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.
9. Seleccionar las técnicas aprendidas y entrenarlas para solucionar un proyecto concreto de reconocimiento visual.
10. Trabajar en equipos multidisciplinares.

## Contenido

1. Detección de Objetos
2. Segmentación Semántica de imágenes e instancias
3. Aprendizaje de Métricas
4. Aprendizaje por Transferencia
5. Arquitecturas para la generación de imágenes: GANs y VAEs
6. Redes Neuronales Gráficas
7. Lenguaje y Visión

## Metodología

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 están utilizando diferentes marcos de programación de aprendizaje profundo (DL) (en la actualidad, PyTorch) y métodos básicos de DL como las redes de retroalimentación (MLP) y las redes neuronales convolucionales (CNN). Incluye la comprensión de redes estándar para detección (RCNN, Fast RCNN, Faster RCNN, YOLO) y segmentación (FCN, SegNet, UNET). Los estudiantes aprenderán a través de una metodología basada en proyectos utilizando herramientas modernas de colaboración en todas las etapas del desarrollo del proyecto.

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.

## Actividades

| Título              | Horas | ECTS | Resultados de aprendizaje     |
|---------------------|-------|------|-------------------------------|
| Tipo: Dirigidas     |       |      |                               |
| clases de teoría    | 20    | 0,8  | 4, 3, 9                       |
| Tipo: Supervisadas  |       |      |                               |
| proyecto            | 8     | 0,32 | 1, 2, 4, 3, 8, 7, 9, 6, 10    |
| Tipo: Autónomas     |       |      |                               |
| estudio, ejercicios | 112   | 4,48 | 1, 2, 4, 3, 5, 8, 7, 9, 6, 10 |

## Evaluación

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.

## Actividades de evaluación

| Título                    | Peso | Horas | ECTS | Resultados de aprendizaje     |
|---------------------------|------|-------|------|-------------------------------|
| asistencia a las sesiones | 0.05 | 0,5   | 0,02 | 1, 2, 7                       |
| examen                    | 0.4  | 2,5   | 0,1  | 1, 2, 4, 3, 8, 6              |
| proyecto                  | 0.55 | 7     | 0,28 | 1, 2, 4, 3, 5, 8, 7, 9, 6, 10 |

## Bibliografía

Referencias genéricas :

1. Deep Learning. Ian Goodfellow, Yoshua Bengio, Aaron Courville. MIT Press, 2016.
2. Neural networks and deep learning. Michael Nielsen. <http://neuralnetworksanddeeplearning.com>

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