

**Reconeixement Visual**

Codi: 44777  
Crèdits: 9

**2024/2025**

Titulació	Tipus	Curs
4318299 Visió per Computador / Computer Vision	OB	0

**Professor/a de contacte**

Nom: Maria Isabel Vanrell Martorell

Correu electrònic: maria.vanrell@uab.cat

**Equip docent**

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

(Extern) David Vázquez

**Idiomes dels grups**

Podeu consultar aquesta informació al [final](#) del document.

**Prerequisits**

Grau en enginyeria, matemàtiques, física o similar.

Assignatura C3: "Machine learning for computer vision"

**Objectius**

Coordinador del mòdul: Dr. Joan Serrat Gual

En visió per computador, el reconeixement visual correspon a la tasca d'explicar el contingut d'una imatge en termes de "Què és?" "On és això?". La resposta a aquestes preguntes és habitualment una etiqueta de classe

corresponent als tipus d'objectes o objectes de la imatge, una caixa de delimitació estreta que conté l'objecte en qüestió o, a un nivell més fi, la regió (píxels) que és el seu esquema. Aquestes tasques s'anomenen, respectivament, classificació d'imatges, detecció d'objectes i segmentació semàntica. La pregunta és "doneu-me objectes com aquest", que requereix aprendre una mètrica similar entre imatges, fins i tot en el cas que provenen de diferents modalitats, com esbossos i fotografies, a través de les anomenades arquitectures de codificador-descodificador. El mòdul VR cobreix arquitectures de xarxes neuronals que aborden aquests quatre tipus de tasques. I, com a complement pràctic, mètodes per implementar-los.

En concret, en aquest mòdul oferim a l'alumne una visió general dels mètodes més recents basats en tècniques d'aprenentatge profund per resoldre problemes de reconeixement visual. L'objectiu final és comprendre escenes complexes per construir sistemes factibles per a la comprensió automàtica d'imatges capaços de respondre a la pregunta complexa de quins objectes i on es troben aquests objectes en una escena complexa.

Després d'haver abordat la tasca de classificació en cursos anteriors, els estudiants aprendran una gran família d'arquitectures d'èxit de xarxes profundes convolucionals que s'han demostrat per resoldre les tasques visuals de detecció i segmentació i reconeixement. Addicionalment, apart d'aquestes tasques visual el curs adreça altres temes avançat d'aprenentatge profund.

## Resultats d'aprenentatge

1. CA02 (Competència) Dissenyar tots els components i la seva interconnexió per a un sistema complet de reconeixement visual.
2. CA06 (Competència) Aconseguir els objectius d'un projecte de visió fet en equip.
3. KA05 (Coneixement) Identificar els problemes bàsics que cal solucionar en un problema de reconeixement visual d'una escena.
4. KA13 (Coneixement) Proporcionar la millor modelització d'un problema de reconeixement visuals, com ara la classificació, la detecció o la segmentació semàntica.
5. SA05 (Habilitat) Resoldre un problema de reconeixement visual entrenant una arquitectura de xarxa neuronal profunda i avaluar-ne els resultats.
6. SA11 (Habilitat) Definir els millors conjunts de dades per entrenar arquitectures de reconeixement visual.
7. SA15 (Habilitat) Preparar un informe que descriu, justifiqui i il·lustri el desenvolupament d'un projecte de visió.
8. SA17 (Habilitat) Preparar presentacions orals que permetin debatre els resultats del desenvolupament d'un projecte de visió.

## Continguts

1. Atenció pròpia i Transformers
2. Detecció d'Objectes
3. Segmentació Semàntica i d'instàncies
4. Aprenentatge per Transferència: pre-entrenament, adaptació de domini, no-supervisat, auto-supervisat.
5. Aprenentatge de mètriques
6. Arquitectures per a la generació d'imatges: GANs i VAEs
7. Aprenentatge per Reforçament
8. Aprenentatge Continu.
9. Xarxes Neuronals Gràfiques
10. Llenguatge i Visió
11. Aprenentatge profund multimodal
12. Models de difusió
13. Reconeixement centrat en humans

#### 14. IA afectiva i fiable.

### Activitats formatives i Metodologia

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
classes de teoria	35	1,4	CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17, CA02
Tipus: Supervisades			
projecte	10	0,4	CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17, CA02
Tipus: Autònomes			
estudi, exercicis	170	6,8	CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17, CA02

La metodologia d'aprenentatge es basa en les conferències i exercicis, però principalment en el projecte, que es desenvolupa durant tot el mòdul. Consisteix a resoldre algunes tasques de comprensió d'escenes aplicades a la conducció autònoma. L'objectiu és aprendre els conceptes i tècniques bàsiques per construir xarxes neuronals profundes per detectar, segmentar i reconèixer objectes específics, centrant-se en imatges enregistrades per una càmera de vehicles a bord per a la conducció autònoma.

Els objectius d'aprenentatge utilitzen diferents marcs de programació d'aprenentatge profund (DL) (actualment, PyTorch) i la construcció de blocs/capes de diferents arquitectures (MLP, CNN, Transformers visuals). Això inclou la comprensió de xarxes de detecció (RCNN, Fast RCNN, Faster RCNN, YOLO), segmentació (FCN, SegNet, UNET) i més aplicacions.

Els estudiants adquiriran les habilitats per a les tasques de dissenyar, entrenar, afinar i avaluar xarxes neuronals per resoldre el problema de comprensió automàtica de la imatge.

Tot això es realitza mitjançant tres formats:

- Sessions supervisades: conferències on els instructors explicaran continguts generals sobre els diferents temes. S'utilitzaran per resoldre el projecte i / o exercicis proposats.
- Sessions dirigides:
  - Sessions de projectes, on es presentaran i discuteixen els problemes i objectius dels projectes, els estudiants interactuaran amb el coordinador del projecte sobre problemes i idees sobre la resolució del projecte. A més, els estudiants fan exposicions orals sobre com han resolt el projecte i informen dels resultats (aproximadament una vegada per setmana)
  - Sessió d'exàmens, on els estudiants s'avaluen individualment. Assoliments de coneixements i habilitats de resolució de problemes
- Treball autònom:
  - estudiar i treballar amb els materials derivats de les conferències, a més de resoldre alguns petits exercicis pràctics per entendre millor els lectors teòrics que no participen directament en la solució del projecte

- treballar en grup per resoldre els problemes dels projectes amb lliuraments: codi, informes, presentacions orals, exercicis

Nota: es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, per a la complementació per part de l'alumnat de les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura/mòdul.

## Avaluació

### Activitats d'avaluació continuada

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
assistència a les sessions	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
projecte	0.55	7	0,28	CA02, CA06, KA05, KA13, SA05, SA11, SA15, SA17

La nota final es calcularà mitjançant la següent fórmula :

$$\text{Nota final} = 0.4 \times \text{Examen} + 0.55 \times \text{Projecte} + 0.05 \times \text{Assistència}$$

on

Examen: és la nota obtinguda de l'examen (ha de ser  $\geq 3$ ). Pot ésser incrementada per punts extra corresponents als exercicis proposats en les classes d'alguns temes, però només si la nota de l'examen és com a mínim 3.

Assistència: nota derivada del control d'assistència a les classes (mínim 70%)

Projecte: nota atorgada pel coordinador del projecte basada en el seguiment que fa setmanalment i en els lliuraments del projecte (ha de ser  $\geq 5$ ). Tot això d'acord amb criteris específics com ara:

- Participació i discussió a les sessions i treball en grup (avaluacions intre pars)
- Lliurament de parts obligatòries i opcionals
- Codi desenvolupat (estil, comentaris, etc.)
- Informe escrit (justificació de les decisions de desenvolupament)
- Presentació oral i demostració

Només els estudiants que han suspès (Nota final  $< 5.0$ ) poden fer l'examen de recuperació.

## Bibliografia

Referències genèriques :

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 majoria dels continguts està relacionat amb l'estat de l'art en els diferents temes així que no hi ha llibres publicats sinó articles de revisió (surveys) i de recerca de cada un dels temes, que seran seleccionats pels professors.

## Programari

Entorn de programació en Python amb especial atenció a les llibreries de visió per computador i Pythorch

## Llista d'idiomes

Nom	Grup	Idioma	Semestre	Torn
(PLABm) Pràctiques de laboratori (màster)	1	Anglès	segon quadrimestre	matí-mixt
(PLABm) Pràctiques de laboratori (màster)	2	Anglès	segon quadrimestre	matí-mixt
(TEm) Teoria (màster)	1	Anglès	segon quadrimestre	matí-mixt