

Reconeixement Visual

Codi: 43088
Crèdits: 6

Titulació	Tipus	Curs	Semestre
4314099 Visió per Computador / Computer Vision	OB	0	2

Professor/a de contacte

Nom: Joan Serrat Gual

Correu electrònic: Joan.Serrat@uab.cat

Equip docent

Jose Luis Gomez Zurita

Carles Ventura Royo

Luis Herranz Arribas

Lluis Gomez Bigorda

Utilització d'idiomes a l'assignatura

Llengua vehicular majoritària: anglès (eng)

Equip docent extern a la UAB

Adriana Romero

Michal Drozdal

Prerequisits

- Grau en enginyeria, matemàtiques, física o similar.
- Modul 2 "Machine learning for computer vision"

Objectius

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 al mòdul M2, 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. A més d'aquestes dues tasques visuals, aquest mòdul també aborda temes avançats d'aprenentatge profund com ara arquitectures per a la generació d'imatges (GANs i VAEs) i arquitectures de codificadors i decodificadors per a aplicacions multimodals.

Competències

- Assumir tasques de responsabilitat en la gestió de la informació i el coneixement.
- Comprendre, analitzar i sintetitzar els coneixements avançats que hi ha en l'àrea, així com proposar idees innovadores.
- Conceptualitzar alternatives de solucions complexes per a problemes de visió i crear prototips que demostrin la validesa del sistema proposat.
- Identificar els conceptes i aplicar les tècniques fonamentals més adequades per resoldre els problemes bàsics de la visió per computador.
- Planificar, desenvolupar, avaluar i gestionar solucions per a projectes en els diferents àmbits de la visió per computador.
- Que els estudiants sàpiguin aplicar els coneixements adquirits i la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.
- Que els estudiants tinguin les habilitats d'aprenentatge que els permetin continuar estudiant, en gran manera, amb treball autònom a autodirigit.
- Seleccionar les eines de programari i els conjunts d'entrenament més adequats per desenvolupar les solucions per als problemes de visió per computador.
- Tenir coneixements que aportin la base o l'oportunitat de ser originals en el desenvolupament o l'aplicació d'idees, sovint en un context de recerca.
- Treballar en equips multidisciplinaris.

Resultats d'aprenentatge

1. Assumir tasques de responsabilitat en la gestió de la informació i el coneixement.
2. Comprendre, analitzar i sintetitzar els coneixements avançats que hi ha en l'àrea, així com proposar idees innovadores.
3. Identificar els problemes bàsics que s'han de resoldre en reconeixement d'objectes i escenes, així com els algorismes específics.
4. Identificar les millors representacions que es puguin definir per resoldre problemes de reconeixement visual tant d'objectes com d'escenes
5. Planificar, desenvolupar, avaluar i gestionar una solució per a un problema concret de reconeixement visual.
6. Que els estudiants sàpiguin aplicar els coneixements adquirits i la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.
7. Que els estudiants tinguin les habilitats d'aprenentatge que els permetin continuar estudiant, en gran manera, amb treball autònom a autodirigit.
8. Seleccionar les tècniques apreses i entrenar-les per solucionar un projecte concret de reconeixement visual.
9. Tenir coneixements que aportin la base o l'oportunitat de ser originals en el desenvolupament o l'aplicació d'idees, sovint en un context de recerca.
10. Treballar en equips multidisciplinaris.

Continguts

1. Detecció d'objectes
2. Segmentació semàntica i d'instàncies
3. Aprenentatge per reforç

4. Arquitectures per a generació d'imatges
 - xarxes generatives adversàries (GAN)
 - autoencoders variacionals (VAE)
6. Aprenentatge de mètriques
7. Arquitectures codificador-decodificador per a aplicacions multimodals
8. GANs condicionals

Metodologia

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 mètodes bàsics de DL com ara xarxes de transmissió avançades (MLP) i xarxes neuronals convolutives (CNN). Inclou la comprensió de xarxes estàndard de detecció (RCNN, Fast RCNN, Faster RCNN, YOLO) i segmentació (FCN, SegNet, UNET). Els estudiants aprendran mitjançant una metodologia basada en un projecte utilitzant eines col·laboratives modernes en totes les etapes del desenvolupament del projecte.

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 dirigides: 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

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
classes de teoria	20	0,8	3, 4, 8
Tipus: Supervisades			

projecte	8	0,32	1, 2, 3, 4, 6, 7, 8, 9, 10
Tipus: Autònomes			
estudi, exercicis	112	4,48	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Avaluació

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. 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 3.0.

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. Tot això d'acord amb criteris específics com ara:

- Participació i discussió a les sessions i treball en grup (avaluacions entre 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.

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
examen	0.4	3	0,12	1, 2, 3, 4, 6, 9
projecte	0.55	7	0,28	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Bibliografia

Referències genèriques :

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

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.