UAB
Universitat Autònoma de Barcelona

Dipòsit digital de documents de la UAB

This is the **published version** of the bachelor thesis:

Auñon Ansio, Jordi. *Segmentació semàntica de parcel·les agrícoles en imatges aèries*. Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026 (Grau en Enginyeria de Dades)

This version is available at https://ddd.uab.cat/record/337360

under the terms of the CC BY-SA license.

Segmentació Semàntica de Parcel·les Agrícoles en Imatges Aèries

Jordi Auñon Ansio

Resum— Aquest treball proposa l'aplicació de tècniques de segmentació semàntica per a la identificació automàtica de parcel·les agrícoles sobre imatges aèries d'alta resolució del Plan Nacional de Ortofotografía Aérea (PNOA), emprant les dades del Sistema d'Informació Geogràfica de Parcel·les Agrícoles (SIGPAC) com a *ground truth*. L'àmbit geogràfic es centra en la província de Lleida. El treball segueix un flux experimental progressiu: construcció del *dataset*, entrenament d'un model *baseline* basat en U-Net amb diferents estratègies d'agrupament de classes, i exploració de l'arquitectura SegFormer basada en transformers. Finalment, s'analitza la separabilitat de les classes mitjançant *clustering* no supervisat sobre representacions extretes amb el model fundacional DINOv2, per validar el mapatge semàntic proposat i determinar el sostre raonable de rendiment assolible amb la configuració actual.

Paraules clau— segmentació semàntica, parcel·les agrícoles, PNOA, SIGPAC, U-Net, SegFormer, transformers, DINOv2, aprenentatge profund, *clustering*.

Abstract— This work proposes the application of semantic segmentation techniques for the automatic identification of agricultural parcels from high-resolution PNOA aerial imagery, using SIGPAC as ground truth. The geographic scope focuses on the Lleida region of Catalonia. The experimental workflow covers dataset construction, training of a U-Net baseline model under different class-grouping strategies, and exploration of the transformer-based SegFormer architecture. Finally, class separability is analysed through unsupervised clustering on features extracted with the DINOv2 foundation model, to validate the proposed semantic mapping and to assess the realistic performance ceiling achievable with the current configuration.

Keywords— semantic segmentation, agricultural parcels, PNOA, SIGPAC, U-Net, SegFormer, transformers, DINOv2, deep learning, clustering.

1. Introducció i Motivació

La classificació de parcel·les agrícoles a partir d'imatges aèries és un repte d'actualitat en la gestió del territori. Activitats com el control de declaracions agràries, la detecció de canvis d'ús del sòl o la planificació de recursos hídrics depenen d'aquesta classificació. Els processos manuals, realitzats per professionals del sector, resulten costosos i lents, i no escalen bé davant la necessitat de monitoratge continu a gran escala. Automatitzar aquesta tasca permetria realitzar aquestes activitats sobre territoris més extensos i amb una freqüència més alta, i una via per aconseguir-ho és l'aplicació de sistemes de segmentació semàntica.

La segmentació semàntica és la tasca de visió per computador que consisteix a assignar una categoria a cada píxel d'una imatge, de manera que píxels que pertanyen al mateix tipus d'objecte o classe reben la mateixa etiqueta. Aplicada a imatges aèries, permet obtenir una representació detallada de la distribució espacial dels cultius. Gràcies als avenços en xarxes neuronals convolucionals profundes, i més recentment en models basats en transformers, la qualitat d'aquesta tasca ha millorat molt en els darrers anys, obrint oportunitats reals per a la seva aplicació en teledetecció agrícola.

Aquest TFG proposa aplicar segmentació semàntica sobre imatges aèries d'alta resolució del Plan Nacional de Ortofotografía Aérea (PNOA) per identificar automàticament la tipologia de parcel·les agrícoles, usant el Sistema d'Informació Geogràfica de Parcel·les Agrícoles (SIGPAC) com a *ground truth*. El treball es focalitza en la província de Lleida, zona amb gran varietat i densitat de cultius (cereals, fruiters, vinya i olivera, entre d'altres), i pren com a referència metodologies ja validades en l'àmbit espanyol, en especial el treball de Pedrayes et al. [1].

El document s'estructura de la manera següent. La Secció 2 defineix els objectius del treball, i la Secció 3 revisa la bibliografia rellevant i estat de l'art. Les seccions 4 i 5 detallen la metodologia i la planificació del treball, respectivament. La Secció 6 recull tot el treball experimental: la construcció del *dataset*, els experiments amb el model *baseline* U-Net [10], l'avaluació de l'arquitectura SegFormer [5] basada en transformers, i l'anàlisi de separabilitat de classes mitjançant *clustering* sobre les característiques extretes amb

• E-mail de contacte: jordi.aunon@autonoma.cat
• Treball tutoritzat per: Daniel Ponsa (MSIAU)
• Curs 2025/26

DINOv2 [7]. Finalment, la Secció 7 recapitula el treball fet, les conclusions principals i les vies de continuació.

2. OBJECTIUS

L'objectiu principal d'aquest TFG és desenvolupar i avaluar un sistema de segmentació semàntica automàtica de parcel·les agrícoles sobre imatges aèries d'alta resolució.

Per assolir aquest objectiu, es desglossen els tres subobjectius següents:

1. Construir un *dataset* georeferenciat de qualitat. Obtenir i processar les fonts de dades per a la província de Lleida, generant automàticament parells imatge–màscara amb anotació semàntica fiable.

2. Definir un protocol d'avaluació rigorós i reproduïble. Establir les mètriques objectives, les particions del *dataset* amb separació geogràfica estricta, i el procediment experimental que permeti comparar models de manera quantitativa.

3. Entrenar, avaluar i comparar arquitectures de segmentació, i validació de decisions al disseny. En una primera fase, entrenar un model *baseline* consolidat en la tasca de segmentació semàntica i analitzar-ne els resultats per identificar-ne les limitacions. A partir d'aquest diagnòstic, avaluar un model basat en transformers sota les mateixes condicions experimentals, comparar el seu rendiment amb el *baseline* i investigar possibles millores en la *pipeline*. En una segona fase, motivada per les limitacions observades, validar amb tècniques no supervisades sobre representacions de models fundacionals si l'agrupació de categories SIGPAC proposada és coherent amb l'estructura de les dades o si n'existeix alguna alternativa millor intrínseca a les dades.

3. ESTAT DE L'ART

La segmentació automàtica de parcel·les agrícoles té un gran interès en l'àmbit de la teledetecció i la gestió del territori, i ha estat abordada amb estratègies que han evolucionat al ritme dels avenços metodològics en visió per computador.

El treball de Graesser i Ramankutty [3] és pioner a afrontar aquesta tasca a escala global, i proposa la detecció automàtica de camps agrícoles mitjançant Random Forests sobre imatges de 30 m/píxel. Aquest enfocament consolida els fonaments pre-*deep learning* del camp i demostra la viabilitat de l'automatització, però amb la limitació intrínseca de la resolució de les dades i la capacitat dels classificadors clàssics per modelar el context espacial.

Amb l'arribada de les tècniques d'aprenentatge profund a la teledetecció, els resultats milloren substancialment gràcies a xarxes convolucionals com U-Net i SegNet. Aung et al. [4] comparen aquestes dues arquitectures sobre cultius de Myanmar i destaquen que la qualitat del *ground truth* és tan crítica com la pròpia arquitectura, aspecte rellevant donades les possibles inconsistències que hi poden haver entre PNOA i SIGPAC.

Aquesta estratègia s'ha mostrat vàlida també en sistemes a gran escala. El treball de Pedrayes et al. [1], que és el punt de partida d'aquest TFG, avalua mètodes de segmentació semàntica per a la classificació d'usos del sòl sobre imatges PNOA emprant SIGPAC com a *ground truth*. Els resultats demostren que el *ground sampling distance* (GSD) és el factor més determinant: amb PNOA a 0,25 m/píxel, DeepLabV3+ classifica correctament el 78% dels píxels de cada classe i, alhora, el 75% dels píxels classificats en una classe realment hi pertanyen. DeepLabV3+ supera U-Net en imatges RGB, mentre que U-Net és preferible amb múltiples bandes espectrals.

Més recentment, treballs com FieldSeg [2] aborden aquesta tasca amb models fundacionals com SAM, sobre imatges de resolució mitjana a 10 m/píxel, obtenint un bon delineament en finques mitjanes i grans i un rendiment millorable en parcel·les petites, que es veuen més afectades per la baixa resolució de les imatges. Aquest treball il·lustra el potencial i les limitacions actuals dels models fundacionals en escenaris de resolució reduïda. En paral·lel, arquitectures basades en transformers com SegFormer [5] han mostrat grans resultats en segmentació de cobertes del sòl gràcies a la seva capacitat d'atenció global, motivant la seva exploració en aquest treball.

4. METODOLOGIA

A continuació es descriuen els aspectes metodològics fonamentals per garantir que els experiments siguin objectius, comparables i reproduïbles.

4.1 Dades i *ground truth*

Qualsevol sistema d'aprenentatge supervisat requereix un conjunt de dades amb anotacions fiables que actuïn com a *ground truth*. En aquest treball s'utilitzen dues fonts de dades públiques que es complementen: una proveïdora d'imatges aèries i una proveïdora d'anotacions semàntiques georeferenciades.

PNOA. [8] El Plan Nacional de Ortofotografía Aérea proporciona ortofotografies de cobertura nacional a 0,25 m/píxel. En aquest treball les imatges s'obtenen mitjançant el servei Web Map Service (WMS) de l'Instituto Geográfico Nacional (IGN) (`pnoa-ma`), descarregant diferents peces en format PNG directament per HTTP. L'alta resolució del PNOA és un dels factors que motiven la seva elecció com a font d'imatges, ja que treballs previs han demostrat que el GSD és un factor determinant en el rendiment de la segmentació, especialment en parcel·les petites com les que predominen a la zona d'estudi.

SIGPAC. [9] El Sistema d'Informació Geogràfica de Parcel·les Agrícoles és una base de dades geoespacial pública amb els polígons declarats pels agricultors i la seva tipologia d'ús. Els polígons es descarreguen per comarca des del portal de la Generalitat en format Shapefile. El fet que SIGPAC sigui una font oficial minimitza els errors d'etiquetatge manual (tot i que no evita possibles errors o imprecisions intrínsecs a les fonts de dades), i la seva georeferenciació permet generar les màscares d'anotació automàticament, sense intervenció manual.

No obstant, la coherència temporal entre les imatges aèries (PNOA) i les declaracions SIGPAC és un factor crític: imatges fotografiades en diferents parts de l'any poden mostrar un estat fenològic dels cultius inconsistent amb l'etiqueta SIGPAC assignada. Aquest aspecte es tracta explícitament en el disseny experimental.

4.2 Mètriques d'avaluació

Per mesurar el rendiment dels models de manera objectiva i comparable s'utilitzen cinc mètriques globals.

- **Overall Accuracy (OA):** proporció total de píxels classificats correctament respecte al total de píxels avaluats. És una mètrica global i intuïtiva, però poc informativa en *datasets* amb classes desbalancejades, ja que els resultats queden dominats per les classes majoritàries.
- **mIoU** (*mean Intersection over Union*): mètrica principal en segmentació semàntica. Per a cada classe mesura la proporció entre l'àrea de solapament entre la predicció i la màscara real i l'àrea de la seva unió, on un valor d'1 indica concordança perfecta i 0 cap solapament.
- **Coeficient Dice:** expressa de manera similar al mIoU la superposició entre predicció i realitat, però penalitza menys els errors en classes petites, cosa que el fa especialment informatiu en classes minoritàries.
- **Boundary F1:** avalua la precisió dels contorns entre parcel·les de manera independent de la classificació interior, aspecte rellevant per a aplicacions de gestió territorial on la delimitació exacta és important.

Per a l'anàlisi del rendiment per classe es reporta el **F1-score**, definit com la combinació entre la precisió (*precision*) i la sensibilitat (*recall*) dels píxels assignats a cada classe. És l'única mètrica que es desglossa per classe a les taules de resultats, i resulta especialment útil per detectar quines categories pateixen problemes.

4.3 Protocol experimental

Per garantir que les comparacions entre models siguin justes i coherents, s'adopten els principis següents. Cada model s'entrena i avalua sobre les mateixes particions, amb les mateixes mètriques i sota les mateixes condicions. Les corbes d'entrenament i validació es registren, i el millor model de cada experiment es selecciona en base a les mètriques calculades al conjunt de validació.

4.4 Gestió del treball

El treball proposat implica realitzar diferents experiments sobre el *dataset* construït. A l'hora de dur-los a terme, es planifica el treball experimental de manera incremental i iterativa: fins que un experiment no es realitza de manera completa no es comença el següent. Així, primer s'estableix una referència amb el model *baseline* i, a partir del diagnòstic dels seus resultats, es decideixen les arquitectures i millores a explorar, justificant cada decisió experimental amb evidència prèvia o hipòtesi formada a partir dels resultats previs.

El codi, les configuracions d'experiment i els resultats es gestionen en un repositori de *GitHub*. El seguiment i la visualització de totes les mètriques i corbes d'entrenament es realitza amb `Weights & Biases` (W&B), que permet comparar experiments de manera centralitzada i detectar sobreajustament/problemes en temps real.

5. PLANIFICACIÓ TEMPORAL

La durada total del TFG s'estima en 19 setmanes, amb una dedicació mitjana d'aproximadament 15 hores setmanals, el que representa 300 hores de treball total. La Taula 1 recull la distribució d'activitats i l'estimació d'hores per a cadascuna.

La planificació s'estructura en quatre fases diferenciades. La **fase inicial** (setmanes 1–5) cobreix la revisió bibliogràfica i la construcció del *dataset*: és la fase més crítica perquè un *pipeline* de dades defectuós invalidaria tots els experiments posteriors, i per això se li assigna una dedicació elevada (70 h). La **fase de *baseline*** (setmanes 5–8) inclou l'entrenament del model U-Net de referència sobre diferents estratègies d'agrupació de classes, així com la comparativa amb l'arquitectura SegFormer basada en transformers. La **fase d'exploració** (setmanes 8–12) està dedicada a la validació de les agrupacions de classes mitjançant tècniques de *clustering* no supervisat sobre representacions extretes amb DINOv2, i és la fase de major càrrega experimental (100 h). Finalment, la **fase de tancament** (setmanes 12–19) inclou l'anàlisi global dels resultats, la redacció de la memòria i la preparació de la defensa.

TAULA 1: PLANIFICACIÓ TEMPORAL DEL TFG AMB ESTIMACIÓ D'HORES

Set.	Activitat	Hores
Fase inicial (set. 1–5)		
1–2	Revisió bibliogràfica	25
2–4	Descàrrega PNOA i SIGPAC; *pipeline* del *dataset*	30
4–5	Anàlisi i caracterització del *dataset*	15
Fase de baseline (set. 5–8)		
5–7	U-Net amb diferents agrupacions de classes	40
7–8	Comparativa amb SegFormer	15
Fase d'exploració (set. 8–12)		
8–12	Validació de l'agrupació de classes amb *clustering* sobre DINOv2	100
Fase de tancament (set. 12–19)		
12–14	Anàlisi global dels resultats	10
14–17	Redacció de la memòria	50
17–19	Preparació presentació i defensa	15
	Total estimat	**300**

6. DESENVOLUPAMENT

En aquesta secció es presenta tot el treball experimental realitzat. Es comença per la construcció del *dataset* sobre el qual s'entrenen i avaluen tots els models/experiments (Secció 6.1), es continua amb els experiments del model *baseline* U-Net (Secció 6.2) i amb l'arquitectura SegFormer basada en transformers (Secció 6.3), i es tanca amb l'anàlisi de separabilitat de classes mitjançant *clustering* sobre les característiques extretes amb DINOv2 (Secció 6.4).

6.1 Construcció del *dataset*

El *dataset* s'ha generat automàticament a partir de les dues fonts de dades descrites a la Secció 4.1: les imatges PNOA i els polígons SIGPAC.

D'aquestes fonts de dades es volen recopilar imatges RGB aparellades amb la seva segmentació semàntica segons el registre a SIGPAC. Donada una regió geogràfica establerta, primer s'ha realitzat un mostreig aleatori dels polígons d'aquesta regió a SIGPAC. Per a cada polígon seleccionat,

s'ha triat un punt aleatori al seu interior que actua com a centre d'una finestra quadrada de 256×256 m sobre el territori. A continuació, s'ha consultat el servei WMS del IGN per descarregar la imatge PNOA corresponent a aquesta finestra, obtenint una imatge RGB de 1024×1024 píxels (amb un GSD de 0,25 m/píxel). Finalment, sobre aquesta mateixa finestra s'ha generat la màscara semàntica de manera automàtica, rasteritzant tots els polígons SIGPAC que la solapen i assignant a cada píxel la categoria d'ús del sòl declarada.

Per garantir la coherència temporal, el sistema consulta l'any de captura de cada imatge PNOA via WMS i utilitza el SIGPAC de l'any corresponent (2024 o 2025) per generar la màscara. La separació en particions es realitza per blocs espacials de 10×10 km, assignant cada bloc (d'on s'agafen els *tiles*) a train, val o test de manera aleatòria, assegurant que cap *patch* de test estigui geogràficament adjacent a patches d'entrenament.

El *dataset* final cobreix 13 comarques de la província de Lleida i conté un total de 3.250 patches/imatges, distribuïts tal com es recull a la Taula 2.

TAULA 2: CARACTERÍSTIQUES DEL *dataset* CONSTRUÏT

Característica	Valor
Comarques cobertes	13
Patches totals	3.250
Entrenament (train)	2.259
Validació (val)	574
Test	417
Mida del patch	256×256 m
Mida de la imatge	1024×1024 px
GSD original	0,25 m/píxel
Bandes espectrals	3 (RGB)
Format	PNG
Separació de particions	Blocs espacials 10×10 km

6.2 Model *baseline*: U-Net

A partir del *dataset* construït, es realitzen els primers experiments amb el model *baseline*. Aquests experiments estudien principalment l'impacte que té el criteri d'agrupament de classes sobre la qualitat de la segmentació. De tots els experiments realitzats, a continuació se'n presenta una selecció representativa centrada en els que han aportat informació més rellevant. En tots els casos s'utilitza augmentació de dades (rotacions, *flips* horitzontals i verticals, i variacions de brillantor i contrast), ja que els experiments preliminars van mostrar una millora consistent en totes les mètriques en activar-la.

Com a arquitectura *baseline* s'empra U-Net amb *encoder* ResNet-34 preentrenat sobre ImageNet i optimitzador Adam, per la seva àmplia adopció en segmentació semàntica i en problemes semblants al d'aquest treball [1, 4], així com per la seva senzillesa d'implementació i limitacions conegudes.

6.2.1 Disseny dels experiments

Experiment 1: 23 classes SIGPAC originals. Es parteix del cas més directe: entrenar el model sobre les 23 categories SIGPAC originals presents al *dataset*, sense cap agrupació. Aquest experiment serveix de punt de partida i permet identificar la principal dificultat d'aquest enfocament: el fort desbalanceig de classes, on les categories minoritàries no disposen de mostres suficients per ser apreses correctament. La taula completa de codis i descripcions de les categories SIGPAC es recull a l'Apèndix A.

Experiment 2: agrupament estadístic en classes majoritàries. Per mitigar el desbalanceig identificat a l'Experiment 1, es defineix un mapatge estadístic en què totes les categories amb menys del 4% de freqüència de píxels al *split* de *train* s'agrupen en una classe genèrica *Altres*. El resultat són 6 classes efectives: *Altres* (agrupació de CA, EP, AG, PS, ZU, IM, ED, VI, TH, FY, FS, OF, VO, FL, IV, VF, FV), PA, FO, TA, PR i OV. L'objectiu és obtenir un *dataset* menys desequilibrat i avaluar si això es tradueix en una millora de rendiment.

Experiment 3: agrupament semàntic per afinitat. A diferència de l'agrupament purament estadístic de l'Experiment 2, aquí les categories SIGPAC s'agrupen per similitud d'aparença espectral i estructural en imatges aèries RGB, amb l'objectiu de reduir la variabilitat intraclasse i la confusió entre categories visualment properes (per exemple, PA i FO, ambdues vegetació densa, o TA i PS, ambdues textures planes herbàcies). El mapatge defineix 7 categories:

- **HERBACI**: terres de conreu herbaci (TA), horts (TH) i pastures (PS). Textures planes i tons verds o ocres homogenis.
- **ARBORE**: fruiters (FY), oliveres (OV), fruita seca (FS) i cítrics. Estructura de copa individual sobre sòl visible, amb patrons regulars de plantació.
- **MIXT_VINYA**: vinya (VI) i totes les associacions que la inclouen (VF, FV, VO, OF, FL). La textura característica de la vinya en fileres la diferencia de les altres arbòries.
- **FORESTAL**: zones forestals (FO), pastures arbrades (PA) i pastures arbustives (PR). Cobertes de vegetació natural densa.
- **URBA**: edificacions (ED), vials (CA) i zones urbanes (ZU). Cobertes artificials.
- **AIGUA**: corrents i superfícies d'aigua (AG). Classe espectralment molt distintiva.
- **ALTRES**: improductiu (IM), elements del paisatge (EP) i altres cultius permanents. Les classes ZV i IV s'exclouen de l'entrenament, ja que són molt infreqüents i amb alta variabilitat visual.

6.2.2 Resultats

La Taula 3 recull les mètriques principals dels tres experiments sobre el conjunt de test.

TAULA 3: COMPARATIVA DE MÈTRIQUES DELS TRES EXPERIMENTS SOBRE EL CONJUNT DE TEST

Mètrica	E1 *23 cl.*	E2 *6 cl. estad.*	E3 *7 cl. sem.*
OA	0,676	0,734	0,871
mIoU macro	0,314	0,546	0,583
mIoU weighted	0,542	0,605	0,800
Dice macro	—	0,678	0,708
Boundary F1	0,194	0,202	0,238

Les taules 4, 5 i 6 detallen el F1 per classe per als experiments E1, E2 i E3 respectivament. Les taules no són directament comparables fila a fila, ja que els esquemes de classes són diferents per disseny.

TAULA 4: F1 PER CLASSE: E1 (23 CLASSES SIGPAC ORIGINALS)

Classe	E1	Classe	E1
background	0,003	VI	0,832
PA	0,319	OV	0,618
CA	0,576	TH	0,006
FO	0,610	FY	0,797
TA	0,898	FS	0,601
PR	0,721	OF	0,000
EP	0,239	VO	0,000
AG	0,493	FL	0,000
PS	0,520	IV	0,000
ZU	0,620	VF	0,000
IM	0,426	FV	0,000
ED	0,146		

TAULA 5: F1 PER CLASSE: E2 (MAPATGE ESTADÍSTIC)

Classe	E2
ALTRES	0,759
PA	0,235
FO	0,685
TA	0,918
PR	0,757
OV	0,713

TAULA 6: F1 PER CLASSE: E3 (MAPATGE SEMÀNTIC)

Classe	E3
HERBACI	0,892
ARBORE	0,841
MIXT_VINYA	0,792
FORESTAL	0,935
URBA	0,630
AIGUA	0,485
ALTRES	0,381

Les figures 1, 2 i 3 mostren els resultats qualitatius dels experiments E1, E2 i E3 respectivament.

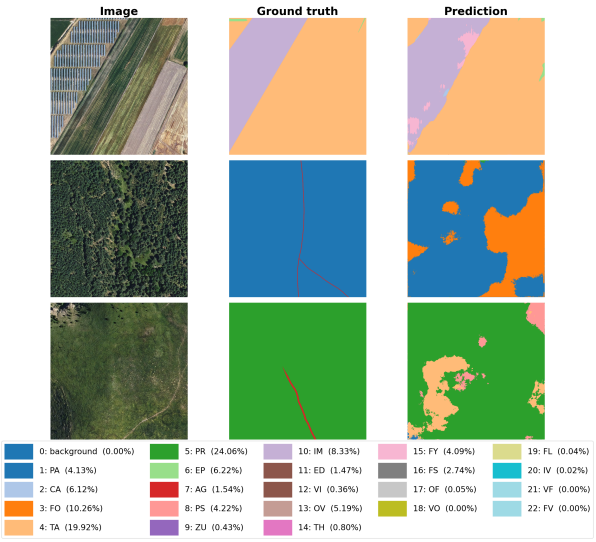


Fig. 1: Resultats qualitatius de l'Experiment 1 (U-Net, 23 classes originals).

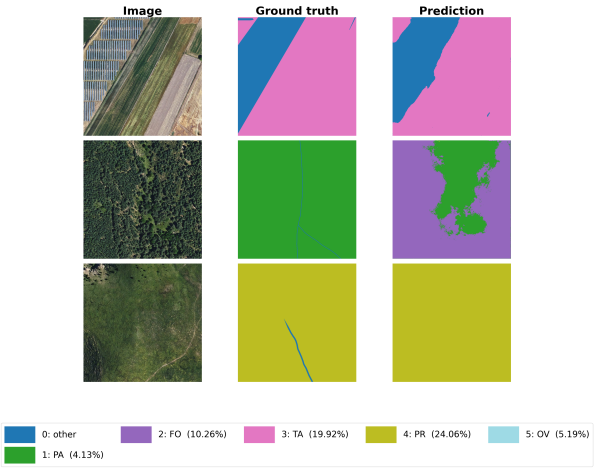


Fig. 2: Resultats qualitatius de l'Experiment 2 (U-Net, 6 classes, mapatge estadístic).

6.2.3 Anàlisi dels resultats

Impacte del mapatge de classes. El pas de 23 a 6 classes (E1→E2) suposa el factor de millora més determinant: el mIoU macro passa de 0,314 a 0,546. En E1, les categories minoritàries obtenen un F1 pràcticament nul i distorsionen l'aprenentatge de la resta. El salt a E3 (mIoU macro 0,583, OA 0,871) s'explica per la coherència espectral del nou agrupament: en unir PA, PR i FO sota FORESTAL desapareix la confusió que en E2 penalitzava greument PA (F1 = 0,235), i la classe assoleix F1 = 0,935. La contrapartida és ALTRES (F1 = 0,381) i AIGUA (F1 = 0,485), que esdevenen el nou punt feble. Tot i això, E3 s'adopta com a millor *baseline*.

Qualitat de les fronteres. El Boundary F1 es manté baix en tots tres experiments (0,194–0,238), indicant que U-Net no defineix contorns precisos amb les mides de peça i resolució actuals. Aquest fet motiva l'exploració d'arquitectures amb capacitat de modelar dependències globals, com SegFormer.

Lectures i hipòtesis. Més enllà dels números, els resultats apunten a una conclusió: el coll d'ampolla d'aquests experiments no és la capacitat del model sinó la definició de les classes i les propietats intrínseques de les dades. Aquesta observació és coherent amb la literatura, on Pedrayes et al. [1] ja reportaven aquesta conclusió i problema. A més, treballar amb només 3 bandes RGB imposa un límit de separabilitat inherent per a categories que es distingeixen sobretot per fenologia o textura subtil (PA, FO, PR), i cal tenir present que el propi SIGPAC pot contenir errors d'etiquetatge, que estableixen un sostre al rendiment assolible. D'aquí emergeixen dues hipòtesis que motiven els passos posteriors: (i) una arquitectura amb atenció global pot atenuar la limitació de fronteres i la confusió intraparcel·la (Secció 6.3), i (ii) la dificultat persistent en classes com ALTRES o AIGUA pot ser intrínseca a les dades més que al model, hipòtesi que cal contrastar amb una anàlisi no supervisada (Secció 6.4).

6.3 Arquitectura avançada: SegFormer

Com a primer pas més enllà del *baseline*, s'avalua l'arquitectura SegFormer [5], seleccionada per la seva capacitat d'atenció global mitjançant transformers. A diferència de les xarxes convolucionals com U-Net, que capturen context local, aquesta capacitat és especialment rellevant per a la tasca que volem fer, ja que una de les limitacions del *baseline* és precisament la qualitat de les fronteres de parcel·la i la confusió en classes dintre d'una mateixa parcel·la. El model s'entrena sobre el mateix esquema de 7 classes semàntiques que E3 i amb les mateixes particions, per tal que els resultats siguin comparables.

6.3.1 Resultats

La Taula 7 compara les mètriques globals de SegFormer amb les del millor model *baseline* (E3, U-Net) sobre el conjunt de test, i la Taula 8 en detalla el F1 per classe. Ambdós models s'avaluen sobre el mateix esquema de 7 classes semàntiques i les mateixes particions.

TAULA 7: COMPARATIVA DE MÈTRIQUES: SEGFORMER VS. U-NET E3 SOBRE EL CONJUNT DE TEST

Mètrica	E3 (U-Net)	SegFormer
OA	0,871	0,868
mIoU macro	0,583	0,600
mIoU weighted	0,800	0,802
Dice macro	0,708	0,718
Boundary F1	0,238	0,235

TAULA 8: F1 PER CLASSE: SEGFORMER VS. U-NET E3

Classe	E3 (U-Net)	SegFormer
HERBACI	0,892	0,897
ARBORE	0,841	0,821
MIXT_VINYA	0,792	0,881
FORESTAL	0,935	0,935
URBA	0,630	0,656
AIGUA	0,485	0,498
ALTRES	0,381	0,338

La Fig. 3 mostra una comparació directa entre U-Net (E3) i SegFormer sobre una selecció de *tiles* del conjunt de test. La Fig. 4 presenta els 3 casos amb pitjor mIoU per al model SegFormer, incloent-hi la predicció de U-Net per als mateixos casos per facilitar la comparació.

6.3.2 Anàlisi dels resultats

Comparació global i guanys per classe. SegFormer millora lleugerament el *baseline* E3: el mIoU macro passa de 0,583 a 0,600 (+0,017) i el Dice macro de 0,708 a 0,718. L'OA i el mIoU weighted es mantenen pràcticament iguals, fet que indica que la millora es concentra en classes de menor representació més que no pas en una millora generalitzada. El guany més notable es produeix a MIXT_VINYA (+0,089, fins a F1 = 0,881), classe definida per una textura en fileres on l'atenció global pot ajudar a capturar el patró regular a escala superior a la del receptive field convolucional. URBA també millora lleugerament (+0,026).

Retrocessos i limitacions persistents. En contrapartida, ARBORE retrocedeix ($-0{,}020$) i ALTRES empitjora notablement ($-0{,}043$): els patrons locals i les classes amb alta variabilitat espectral semblen aprendre's millor amb convolucions que amb atenció global. AIGUA continua confonent-se amb FORESTAL, patró idèntic al *baseline*, i el Boundary F1 (0,235) és pràcticament idèntic al de E3 (0,238): el canvi d'arquitectura per si sol no millora els contorns ni resol les classes més difícils.

Lectures i hipòtesis. La lectura conjunta és que l'arquitectura no és la palanca principal en aquest problema. Aportar atenció global ajuda en patrons estructurats com la vinya, però no toca les limitacions que ja s'identificaven al *baseline*: la confusió en classes heterogènies (ALTRES), la dispersió d'AIGUA i la qualitat de les fronteres. Això reforça la hipòtesi formulada a la Secció 6.2.3: el sostre actual no és tant un problema de capacitat del model com de definició de classes i propietats intrínseques de les dades RGB. La validació d'aquesta hipòtesi és el que motiva l'anàlisi no supervisada amb DINOv2 de la Secció 6.4, que explora si l'agrupació semàntica proposada és coherent amb l'estructura natural de les representacions o si existeixen agrupaments alternatius més fonamentats.

6.4 Anàlisi de separabilitat de classes: *clustering* amb DINOv2

Els experiments supervisats anteriors han demostrat que algunes classes del mapatge semàntic (especialment ALTRES i AIGUA) presenten un rendiment baix independentment de l'arquitectura. Per comprendre si aquest problema té un origen en la definició de les classes o és degut a les limitacions dels models entrenats, es planteja un experiment de validació no supervisada: si les 7 classes estan espectralment ben diferenciades, un model de visió sense cap entrenament sobre el nostre *dataset* hauria de produir representacions on les classes formin agrupacions naturals i separades.

6.4.1 Disseny de l'experiment

S'utilitza DINOv2 [7] (variant `vits14`), un model fundacional de visió entrenat amb auto-supervisió sobre dades generals, sense cap adaptació al domini. Per a cada imatge del conjunt d'entrenament (2.275 imatges), el model extreu un vector de característiques per a cada *patch* de 14×14 píxels de la imatge centrada i retallada a 1008×1008 píxels, obtenint una representació de dimensió 384 per patch. L'etiqueta de cada patch s'assigna per majoria de píxels sobre la màscara SIGPAC corresponent. En total s'obtenen 254.583 patches vàlids, balancejats a 36.369 per classe (nombre total de patches de la classe menys repetida).

Sobre els vectors extrets s'aplica PCA per reduir la dimensió (conservant les components necessàries per explicar el 90% de la variància), seguida de t-SNE per a la projecció 2D. Complementàriament, s'executa KMeans [11] amb $k = 7$ sobre l'espai PCA per analitzar l'alineació entre els clústers obtinguts i les etiquetes reals. Aquesta alineació s'avalua observant, per a cada classe, com es distribueixen els seus *patches* entre els 7 clústers: si una classe està ben representada en l'espai de característiques, els seus *patches* haurien de concentrar-se majoritàriament en un o pocs clústers; en

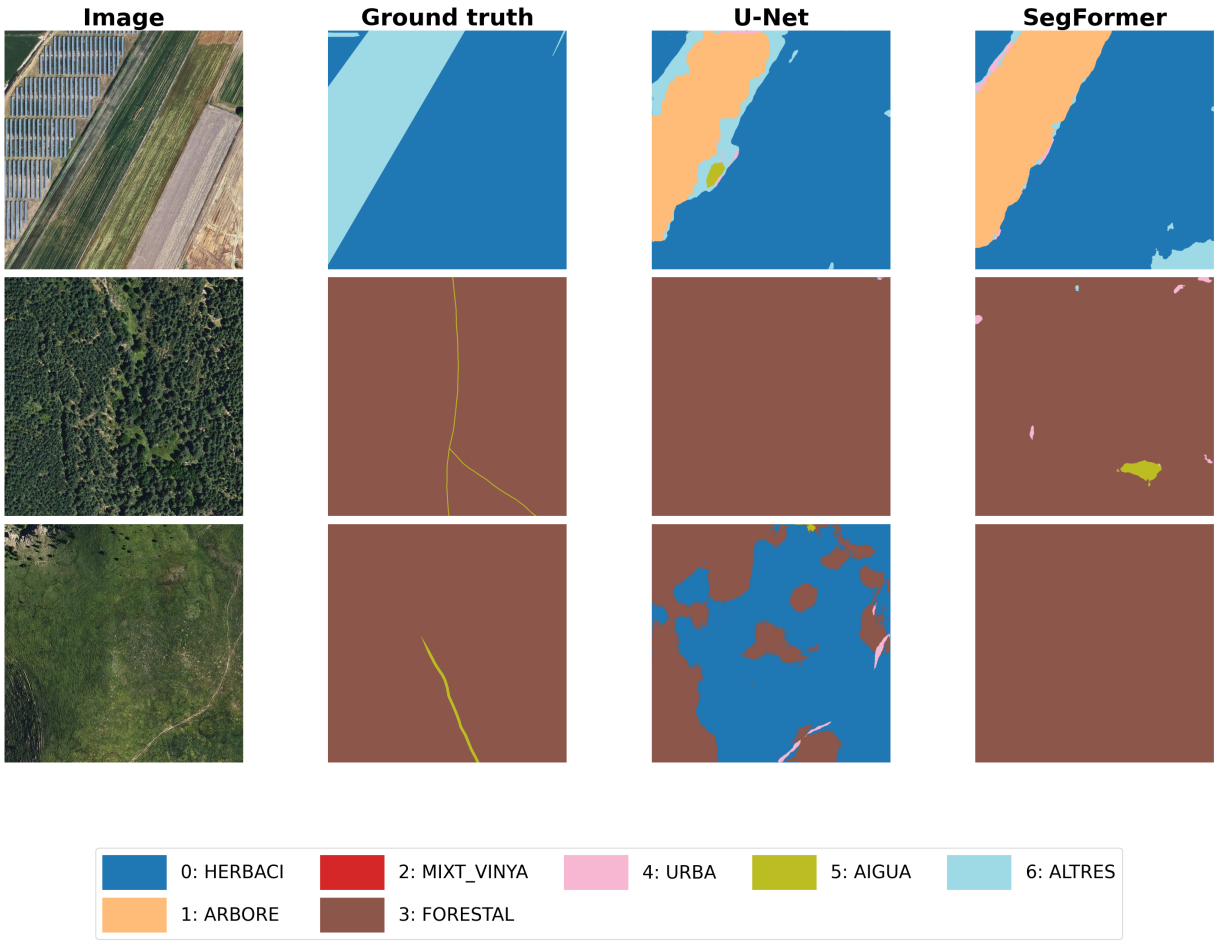


Fig. 3: Comparació qualitativa entre U-Net (E3) i SegFormer. Columnes: imatge aèria PNOA, *ground truth* SIGPAC, predicció U-Net i predicció SegFormer.

canvi, una distribució uniforme entre tots els clústers indica una baixa separabilitat. La tria d'un k fix es justifica pel fet que es vol contrastar directament amb les 7 classes del mapatge semàntic; alternatives sense necessitat d'especificar k, queden proposades com a línia futura.

6.4.2 Resultats

La Taula 9 mostra la distribució percentual de *patches* per classe i clúster, i la Fig. 5 la projecció t-SNE colorejada per etiqueta.

6.4.3 Anàlisi dels resultats

La distribució observada confirma que cap classe es concentra majoritàriament en un sol clúster: en mitjana, els *patches* de cada classe queden repartits entre diversos clústers, fet que indica una baixa separabilitat global. La projecció t-SNE ho il·lustra: no existeixen regions nítidament separades per classe, sinó una superposició generalitzada, especialment entre HERBACI, ARBORE i ALTRES.

Les classes amb millor separabilitat són URBA i FORESTAL: URBA concentra el 29% dels seus patches al C0 i el 47% al C5, mentre que FORESTAL agrupa el 46% al C1. En canvi, HERBACI i ARBORE es distribueixen de manera força uniforme per tots els clústers sense cap dominància clara. AIGUA és especialment problemàtica: tot i dominar C2 i C3, cap d'aquests clústers supera el 28% dels seus patches, indicant una dispersió molt elevada.

Aquests resultats suggereixen que la dificultat que els models supervisats troben amb certes classes no és exclusivament un problema d'arquitectura o de dades d'entrenament insuficients: l'agrupació de classes implica una variabilitat intraclasse elevada (almenys en el *dataset* inicial).

6.4.4 Experiment complementari: *clustering* sobre les classes originals

L'experiment de *clustering* anterior està basat sobre el mapatge semàntic de 7 classes definit manualment (E3). Per validar si aquest mapatge és òptim o si existeix un agrupament alternatiu suggerit per les pròpies dades, es replanteja l'experiment partint de les 22 categories SIGPAC originals (que apareixen dintre del conjunt de train) sense cap agrupació prèvia. L'objectiu és determinar si KMeans amb $k = 7$ sobre les característiques DINOv2 convergeix cap a un mapatge diferent al proposat, i si aquest mapatge podria ser útil per entrenar SegFormer.

Disseny de l'experiment. S'utilitza el mateix procediment que a l'experiment anterior: extracció de característiques DINOv2 (`vits14`) sobre els *patches* del conjunt d'entrenament, reducció de dimensió per PCA i posterior KMeans amb $k = 7$.

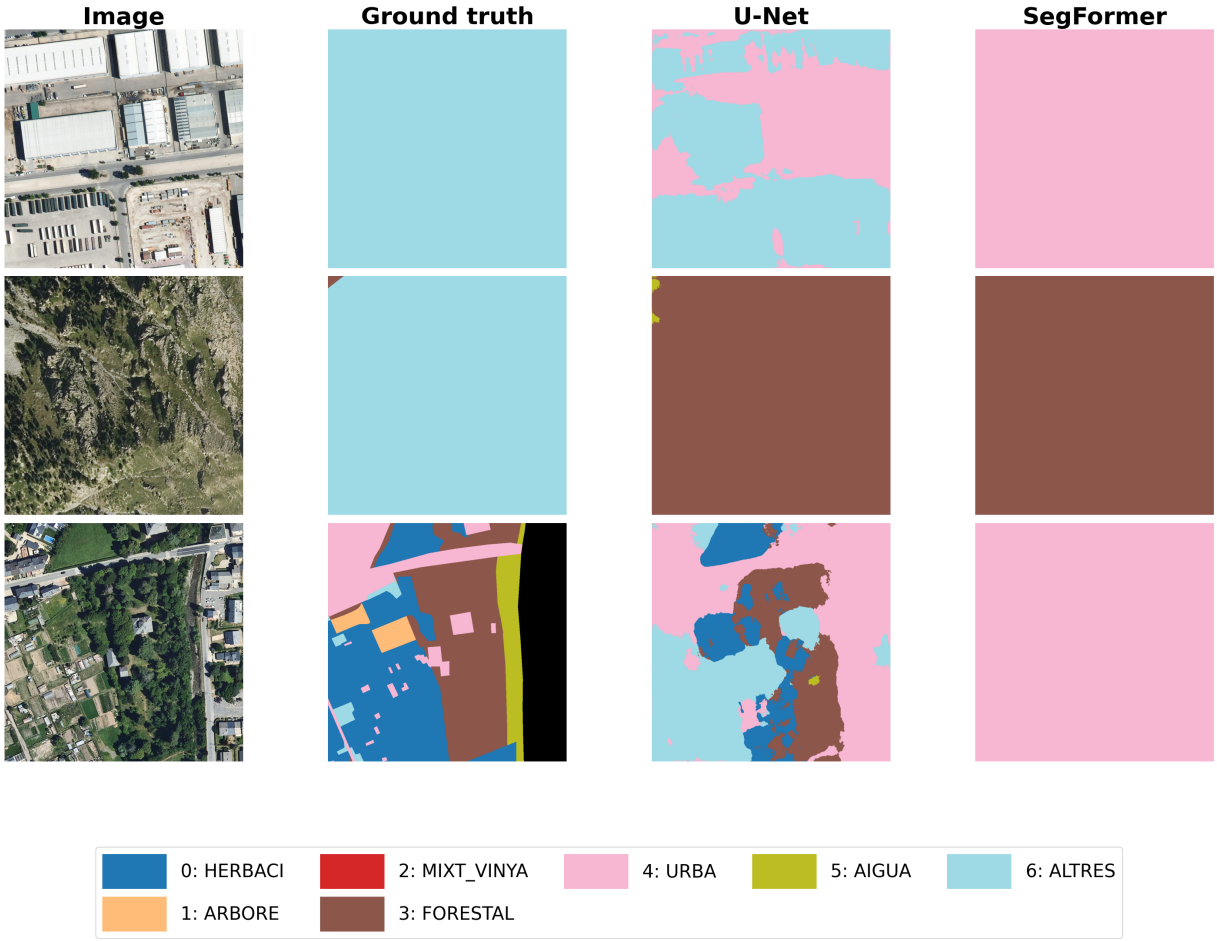


Fig. 4: Els 3 casos amb pitjor mIoU del conjunt de test per al model SegFormer. Columnes: imatge aèria PNOA, *ground truth* SIGPAC, predicció SegFormer i predicció U-Net (E3) per als mateixos casos.

TAULA 9: DISTRIBUCIÓ DE *patches* PER CLASSE I CLÚSTER, EN PERCENTATGE PER CLASSE (7 CLASSES SEMÀNTIQUES, DINOV2 + KMEANS)

Classe	**C0**	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**
HERBACI	0,7	5,7	14,8	19,1	27,5	26,6	5,6
ARBORE	0,6	1,8	10,9	11,6	25,4	27,1	22,6
MIXT_VINYA	0,4	1,5	15,6	18,2	40,0	16,8	7,4
FORESTAL	0,1	45,7	11,9	12,9	0,2	5,5	23,6
URBA	29,0	3,0	3,5	2,9	1,3	47,1	13,3
AIGUA	1,4	16,5	16,7	28,4	0,6	18,1	18,3
ALTRES	20,1	10,6	6,2	4,8	2,5	35,3	20,5
Dominant	URBA	FORESTAL	AIGUA	AIGUA	MIXT_VINYA	URBA	FORESTAL

En aquest cas s'empren les 22 classes amb tots els vectors extrets possibles del conjunt d'entrenament, obtenint un total d'$n = 11.710.656$ patches vàlids.

Anàlisi, resultats i conclusions. La Taula 10 mostra la distribució percentual de *patches* per classe i clúster. Els resultats revelen que el *clustering* no supervisat no convergeix cap a un mapatge alternativament útil en 7 classes. Tres dels set clústers queden dominats per FO i dos per PR, cosa que indica que DINOv2 no diferencia de manera natural les categories de vegetació natural (PA, FO, PR) en grups semànticament coherents. En conseqüència, no és possible convertir aquest experiment en un mapatge *data-driven* que pugui substituir o millorar el mapatge semàntic de E3.

L'anàlisi per classe confirma aquest patró: les classes espectralment més distintives concentren la majoria dels seus patches al clúster 4 (ED 89,5%, IV 89,6%, CA 77,4%, ZU 72,9%, TA 60,4%), mentre que les categories de vegetació natural (PA, FO, PR) apareixen distribuïdes de manera quasi uniforme per tots els clústers, evidenciant que la seva superposició en l'espai de característiques és una propietat intrínseca de les dades en RGB i no un artefacte del mapatge manual.

Cal remarcar, però, que el plantejament d'aquest experiment té una limitació: el *clustering* opera sobre vectors de característiques extrets de *patches* de 14×14 píxels, sense tenir directament en compte el context global de la parcel·la. Una alternativa potser més adequada consistiria a agregar els

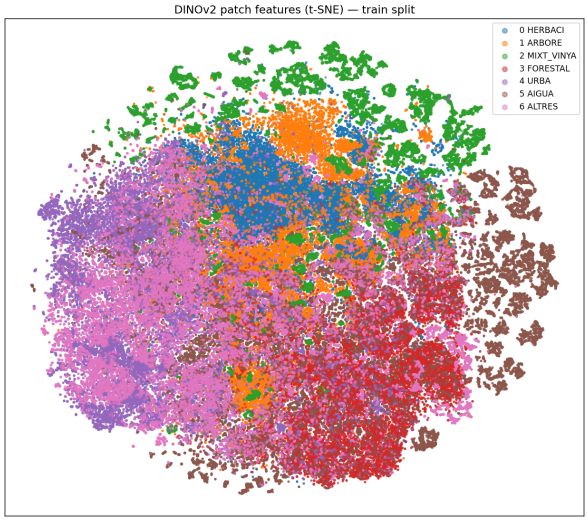


Fig. 5: Projecció t-SNE de les característiques DINOv2 sobre el conjunt d'entrenament, colorejada per etiqueta SIGPAC.

vectors de tots els *patches* que cauen dins d'una mateixa parcel·la SIGPAC en una representació única (per exemple, mitjançant *pooling* o concatenació) i aplicar el *clustering* sobre aquestes representacions de parcel·la completa. La hipòtesi és que, amb el context integral d'una parcel·la, instàncies d'una mateixa classe presentarien una major similitud en l'espai de característiques, i el *clustering* podria revelar agrupaments més coherents amb les categories SIGPAC. Aquest enfocament queda proposat com a línia de treball futura.

7. Conclusions i Treball Futur

Aquest treball ha abordat la segmentació semàntica de parcel·les agrícoles sobre imatges PNOA d'alta resolució, usant SIGPAC com a *ground truth*, i ha permès construir una base experimental sòlida sobre la qual fonamentar decisions posteriors.

Pel que fa als **models de segmentació**, els experiments amb el *baseline* U-Net han mostrat que la definició de les classes és un factor tant o més determinant que l'arquitectura: el pas a un mapatge semàntic de 7 classes (E3), agrupant categories per similitud espectral i estructural en lloc de per freqüència, ha estat la decisió que ha aportat el salt qualitatiu més gran. La transició a SegFormer ha confirmat que l'atenció global dels transformers aporta una petita millora addicional, especialment en classes amb patrons estructurals regulars com la vinya, però també ha mostrat que el canvi d'arquitectura no resol per si sol les limitacions persistents: la qualitat de les fronteres de parcel·la i el rendiment baix en classes heterogènies (ALTRES) o poc representades (AIGUA).

L'anàlisi de **separabilitat amb DINOv2** ha aportat una perspectiva complementària. Els experiments de *clustering* no supervisat, tant sobre les 7 classes semàntiques com sobre les 22 categories SIGPAC originals, han evidenciat que part de la confusió no és un problema exclusivament d'arquitectura, sinó una propietat intrínseca de les dades en RGB: categories com PA, FO i PR es solapen de manera natural en l'espai de característiques, i no emergeix cap agrupament *data-driven*. Aquest resultat valida *a posteriori* la decisió de E3 i alhora delimita el sostre raonable de rendiment assolible amb la configuració actual.

A partir d'aquestes observacions, es plantegen dues **línies de continuació**. La primera consisteix a repetir el *clustering* no a nivell de *patch*, sinó agregant les característiques DINOv2 a nivell de parcel·la SIGPAC sencera i aplicant algorismes més avançats que KMeans. Amb el context complet de la parcel·la, instàncies d'una mateixa tipologia presentarien una similitud més alta, i si el *clustering* convergís cap a un mapatge ben definit, aquesta agrupació, validada de manera no supervisada, podria emprar-se com a esquema d'etiquetes per entrenar models de segmentació amb el suport d'una estructura emergent de les dades.

La segona línia proposa replantejar l'àmbit geogràfic del *dataset*: el conjunt actual pren 13 comarques de Lleida, fet que introdueix una variabilitat intraclasse molt alta. Centrar el *dataset* en un territori més acotat i homogeni permetria que els models aprenguessin patrons més consistents per classe, i ajudaria a discernir fins a quin punt les limitacions observades són atribuïbles a l'heterogeneïtat del territori més que a les arquitectures.

En conjunt, el treball deixa establert un *pipeline* reproduïble, un mapatge semàntic validat i un diagnòstic clar de les limitacions actuals, conformant una base sobre la qual les futures iteracions poden avançar amb fonaments sòlids.

Ús d'eines d'intel·ligència artificial

D'acord amb la normativa de l'assignatura, durant l'elaboració d'aquest treball s'han fet servir eines d'intel·ligència artificial generativa (concretament Claude, d'Anthropic, i ChatGPT, d'OpenAI) com a suport puntual en aspectes formals i tècnics: revisió de redacció, estil i correcció ortogràfica de la memòria, i assistència en dubtes concrets d'implementació o de configuració durant la fase experimental. L'ús ha estat limitat i sempre supervisat. El plantejament metodològic, el disseny i l'execució dels experiments, l'anàlisi i la interpretació dels resultats, la formulació d'hipòtesis i conclusions, i l'estructura del document han estat fets completament per l'estudiant. Qualsevol contingut generat amb el suport d'aquestes eines ha estat revisat i adaptat per l'estudiant abans d'incorporar-lo al treball.

Agraïments

Vull agrair al meu tutor, l'orientació i el suport al llarg de tot el treball. També a la meva família i a la meva parella, per l'acompanyament i la paciència durant el grau.

Referències

[1] O. D. Pedrayes, D. G. Lema, D. F. García, R. Usamentiaga i À. Alonso, "Evaluation of Semantic Segmentation Methods for Land Use with Spectral Imaging Using Sentinel-2 and PNOA Imagery," *Remote Sensing*, vol. 13, núm. 12, p. 2292, 2021.

[2] "FieldSeg: A scalable agricultural field extraction framework based on the Segment Anything Model and 10-m Sentinel-2 imagery," *Computers and Electronics in Agriculture*, 2025.

[3] J. Graesser i N. Ramankutty, "Detection of cropland field par-

TAULA 10: DISTRIBUCIÓ DE *patches* PER CLASSE I CLÚSTER, EN PERCENTATGE PER CLASSE (22 CLASSES ORIGINALS A TRAIN, DINOV2 + KMEANS)

Classe	**C0**	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**
background	6,1	12,5	11,8	17,5	19,8	19,0	13,3
PA	18,4	13,5	12,9	20,3	2,3	20,6	12,1
CA	1,0	2,3	1,7	4,0	77,4	10,9	2,7
FO	24,4	14,5	14,3	12,6	2,2	19,5	12,6
TA	0,0	14,1	9,1	2,2	60,4	2,8	11,4
PR	3,8	12,2	11,2	34,3	8,4	19,0	11,1
EP	0,1	2,4	1,6	4,2	57,0	31,7	3,0
AG	3,1	18,0	17,9	10,9	20,9	15,0	14,1
PS	2,2	11,2	10,1	17,2	32,9	15,4	11,1
ZU	0,2	4,5	3,2	1,1	72,9	12,6	5,5
IM	0,6	4,4	4,6	11,4	59,0	13,4	6,5
ED	0,0	2,1	1,0	0,5	89,5	4,1	2,9
VI	0,0	14,3	16,1	1,1	53,7	1,0	13,8
OV	0,0	6,8	5,4	3,8	50,0	25,9	8,0
TH	0,0	7,1	2,2	0,2	64,3	9,5	16,7
FY	0,0	12,0	9,2	1,8	53,6	12,2	11,2
FS	0,0	9,7	6,7	1,6	51,9	21,6	8,6
Classes molt minoritàries ($n < 5.000$ patches)							
OF	0,0	7,8	2,8	10,8	39,8	30,9	7,8
VO	0,0	0,0	0,0	0,0	88,0	0,0	12,0
FL	0,0	12,3	5,4	0,3	55,3	23,8	2,9
IV	0,0	0,0	0,0	0,0	89,6	0,0	10,4
VF	0,0	5,6	0,0	0,0	87,4	0,0	6,6
Dominant	FO	FO	FO	PR	TA	PR	PR

cels from Landsat imagery," *Remote Sensing of Environment*, vol. 201, pp. 165–180, 2017.

[4] S. N. Aung et al., "Comparative analysis of U-Net, SegNet and FCN for crop segmentation," 2023.

[5] E. Xie et al., "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," *NeurIPS 2021*.

[6] A. Kirillov et al., "Segment Anything," *ICCV 2023*, pp. 4015–4026.

[7] M. Oquab et al., "DINOv2: Learning Robust Visual Features without Supervision," *Transactions on Machine Learning Research*, 2024.

[8] Instituto Geográfico Nacional, *Plan Nacional de Ortofotografía Aèrea (PNOA)*. `https://pnoa.ign.es/`

[9] Generalitat de Catalunya, Dept. d'Agricultura, *SIGPAC — Descàrregues per comarca .shp*. `https://agricultura.gencat.cat`

[10] O. Ronneberger, P. Fischer i T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *MICCAI 2015*, pp. 234–241.

[11] L. Mahon i M. Lapata, "K*-Means: A Parameter-free Clustering Algorithm," *arXiv preprint* arXiv:2505.11904, 2025. `https://arxiv.org/abs/2505.11904`

APÈNDIX

A. CATEGORIES SIGPAC

La Taula 11 recull les categories d'ús del sòl definides per SIGPAC i utilitzades com a *ground truth* al llarg del treball, amb el seu codi i descripció. S'hi inclou també la categoria *background*, que correspon als píxels sense cap ús definit a SIGPAC.

TAULA 11: CATEGORIES SIGPAC: DESCRIPCIÓ I CODI

Descripció	Codi
Altres cultius permanents	OP
Associació vinya – fruiter	VF
Associació fruita seca – olivera	FL
Associació fruita seca – vinya	FV
Associació olivera – fruiter	OF
Associació vinya – olivera	VO
Cítric	CI
Corrents i superfícies d'aigua	AG
Edificació	ED
Element del paisatge	EP
Forestal	FO
Fruiter	FY
Fruita seca	FS
Hivernacle	IV
Horta	TH
Improductiu	IM
Olivera	OV
Pastiu	PS
Pastura arbrada	PA
Pastura arbustiva	PR
Terra campa	TA
Vial	CA
Vinya	VI
Zona censurada	ZV
Zona urbana	ZU
Sense ús definit a SIGPAC	*background*

B. DISTRIBUCIÓ DE CLASSES AL *dataset*

La Taula 12 recull el percentatge de parcel·les que correspon a cada categoria SIGPAC dins del *dataset* construït. Aquesta distribució evidencia el fort desbalanceig de classes al qual es fa referència al llarg del treball: només cinc categories (PR, TA, FO, IM i EP) acumulen aproximadament el 70% de les parcel·les, mentre que diverses categories minoritàries presenten una representació pràcticament nul·la.

TAULA 12: DISTRIBUCIÓ DE PARCEL·LES PER CATEGORIA SIGPAC AL *dataset*

Codi	Descripció	% parcel·les
PR	Pastura arbustiva	24,06
TA	Terra campa	19,92
FO	Forestal	10,26
IM	Improductiu	8,33
EP	Element del paisatge	6,22
CA	Vial	6,12
OV	Olivera	5,19
PS	Pastiu	4,22
PA	Pastura arbrada	4,13
FY	Fruiter	4,09
FS	Fruita seca	2,74
AG	Corrents/superfícies aigua	1,54
ED	Edificació	1,47
TH	Horta	0,80
ZU	Zona urbana	0,43
VI	Vinya	0,36
OF	Associació olivera–fruiter	0,05
FL	Associació f. seca–olivera	0,04
IV	Hivernacle	0,02
VO	Associació vinya–olivera	0,00
VF	Associació vinya–fruiter	0,00
FV	Associació f. seca–vinya	0,00
background	Sense ús definit a SIGPAC	0,00