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

Garrofé Urrutia, Marc; Benavente Vidal, Robert, dir. Detecció de piscines il·legals en imatges aèries utilitzant tècniques de deep learning. 2023. (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/280706>

under the terms of the  license

Detecció de piscines il·legals en imatges aèries utilitzant tècniques de deep learning

Marc Garrofé Urrutia

Juny 2023

Resum– Aquest projecte té com a objectiu detectar piscines il·legals mitjançant l'ús de tècniques de visió per computador i aprenentatge profund. La solució proposada implica l'extracció d'imatges satel·litals i capes multiespectrals, així com la utilització de dades cartogràfiques del registre del cadastre per determinar la legalitat de les construccions. Durant el desenvolupament del projecte, s'analitzen diverses arquitectures de xarxes neuronals com U-Net, FCN i DeepLab, i s'examinen múltiples combinacions de capes d'informació per aconseguir els millors resultats.

Paraules clau– Aprenentatge profund, visió per computador, U-Net, Fully Connected Networks, DeepLab, segmentació d'imatges, detecció de piscines, LIDAR, Near-infrared.

Abstract– This project is based on the detection of illegal swimming pools using deep learning and computer vision techniques. The proposed solution involves extracting satellite images and multispectral layers, along with cadastral map data, to determine the legal status of the construction. The project analyzes different neural network architectures, such as U-Net, FCN, and DeepLab, and compares multiple combinations of information layers.

Keywords– Deep Learning, Computer Vision, U-Net, Fully Connected Networks, DeepLab, Image segmentation, Pool detection, LIDAR, Near-infrared.



1 INTRODUCCIÓ

SEGONS un article de l'agost de 2022 del diari El País, el “Plan de Regularización Catastral” del Ministeri d'Hisenda realitzat entre els anys 2014 i 2018, ha permès la detecció de més de tres milions de construccions no regularitzades a tot l'estat espanyol. [1]

Pocs dies després d'aquest article, la BBC publicava la troballa per part de les autoritats fiscals franceses de més de 20.000 piscines no declarades gràcies a la col·laboració amb Google i la consultora Capgemini. Pel que fa als impostos, la revalorització de les propietats era de més de 10 milions d'euros; un increment mitjà de 200 euros anuals en concepte d'impostos per a cada piscina de 30 metres quadrats aproximadament. En la data de publicació, el projecte havia analitzat 9 departaments de França, però s'esperava expandir-ho a escala nacional. [2]

• E-mail de contacte: marc.garrofe@autonoma.cat

• Menció realitzada: Computació

• Treball tutoritzat per: Robert Benavente Vidal (Ciències de la Computació)

• Curs 2022/23

Tot i que la detecció de piscines en situació legal irregular representa un incentiu econòmic per a les administracions públiques, també hi ha un fort component mediambiental. Actualment, l'augment de les temperatures a escala mundial i la manca de precipitacions en el territori català provoca una situació de risc per a garantir l'abastiment d'aigua sense restriccions. Les autoritats públiques necessiten disposar d'un control més exhaustiu que mai sobre tots els elements de consum d'aigua per poder assegurar l'aprovisionament a la població.

Gràcies a una consulta a l'oficina provincial del cadastre de Barcelona, s'ha pogut identificar quines són les infraestructures de caràcter particular i residencial amb més dificultats a l'hora de legalitzar:

- Addició de plantes a l'habitatge.
- Ampliacions de soterranis, terrasses i porxos.
- Construcció de piscines.

La mateixa oficina cadastral consultada afirma que no disposa d'eines automàtiques per a realitzar la detecció, sinó que executen una tasca manual comparant les seves dades amb imatges de satèl·lits públiques en un procés que ells anomenen: “Cruce de datos”.

D'entre els tres tipus de construccions esmentades, aquest projecte pretén centrar-se en la detecció de les piscines perquè es considera un problema amb una complexitat que permet adequar-se a les limitacions temporals i de recursos que es disposa per dur a terme aquest TFG.

El processament d'imatges d'alta resolució de superfícies extenses pot ser, a escala computacional, molt complex. L'escalabilitat dels algorismes en visió per computador ha de ser un factor clau a considerar.

2 ESTAT DE L'ART

En observar una escena, el cervell humà és capaç de detectar de manera immediata tots els objectes que la componen. Avui en dia, l'extracció d'informació com ara detecció, segmentació i classificació d'objectes en imatges digitals o vídeos, ha esdevingut una necessitat primordial gràcies a l'alt nombre d'aplicacions que depenen d'aquesta tecnologia. Les tècniques tradicionals vistes fins fa uns anys en el camp de la visió per computador, són computacionalment cares i no estan a l'altura per poder resoldre els problemes moderns. La solució per aquest problema és l'ús del *Deep Learning*. Aquest és un tipus de *Machine Learning* que utilitza xarxes neuronals compostes de múltiples capes que, de manera progressiva, extreuen característiques d'alt nivell a partir de la informació inicial. [4]

El problema de la detecció de piscines en imatges aèries s'ha abordat de diferents maneres al llarg dels anys tot usant els recursos i els avanços tecnològics de cada època. Una de les primeres referències sobre la detecció de piscines es pot trobar en l'article de Rodríguez-Cuenca i Alonso [3]. Aquesta proposta combina tres fonts d'informació: imatges RGB, NIR (*Near InfraRed spectroscopy* o infraroges-properes) i LIDAR (*Light Detection and Ranging*) per a realitzar la segmentació de la imatge i assolir una *accuracy* del 99,86%.

Rodríguez-Cuenca i Alonso superposen les diferents capes per a extreure'n informació, com ara: l'índex NDVI (índex de vegetació de diferència normalitzada) i l'índex NDWI (índex d'aigua de diferència normalitzada). Aquests indicadors combinen les capes de color de l'espectre visible de la llum amb la capa infraroja (Equacions 1 i 2) per extreure la qualitat, quantitat i desenvolupament de la vegetació i també permeten capturar la característica senyal espectral que emeten els cossos d'aigua.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR} \quad (2)$$

on *NIR* és la capa d'espectre infraroig amb una longitud d'ona que oscil·la entre 690 i 850 nanòmetres (nm) segons les especificacions tècniques de l'ICGC [18]. *GREEN* representa la franja espectral corresponent al color verd i *RED* al color vermell.

Set anys després, en el treball de Coelho et al. [4], afegien tècniques de *Deep Learning* per a la detecció de les piscines i despleguen una xarxa RetinaNet amb imatges RGB on la majoria de les piscines són clarament distingibles i

visibles totalment. És a dir, quasi no tenen ombres ni vegetació frondosa en les seves proximitats i estan emplenades d'aigua. En les conclusions de l'article, els autors reporten problemes a l'hora de detectar piscines buides, i errors de classificació en pistes esportives i edificacions rodones de color blavós.

Finalment, en el mateix any de l'article anterior, l'equip de Lima et al. [5], implementa una solució que combina tècniques de *Deep Learning* amb el processament massiu de dades geogràfiques i informació pública sobre l'estat legal de les piscines. Lima et al., reporten una *accuracy* del 83% i finalitzen la proposta amb un programa interactiu que permet visualitzar les deteccions i mostrar dades sobre la parcel·la i l'estat legal de la construcció.

Tot i que les propostes obtenen bons resultats, encara hi ha problemes oberts, com ara la detecció de piscines buides, brutes o semiocloses. Amb la combinació de les tècniques de *Deep Learning* i l'ús de múltiples capes d'informació (RGB, Infraroig-proper i LIDAR) juntament amb índexs que permeten detectar característiques del terreny (NDVI, NDWI i c_3), es pretén poder millorar els resultats aconseguits fins ara en matèria de detecció de piscines.

3 OBJECTIUS

3.1 Objectius principals

L'objectiu principal del treball és poder identificar piscines de manera automàtica utilitzant tècniques de visió per computador i d'aprenentatge profund. Posteriorment, s'han de comparar les deteccions amb les dades públiques del registre del cadastre per determinar l'estat legal de la construcció.

Aquest objectiu principal es pot dividir en els següents subobjectius:

- Detectar piscines en imatges aèries d'alta resolució.
- Comparar les deteccions amb les dades del cadastre de manera automàtica per a obtenir la legalitat de la piscina.
- Utilitzar diferents estructures de xarxes neuronals: U-Net [6], FastFCN [7] i DeepLab [8] i comparar resultats.
- Treballar i donar solucions amb diferents tipus de dades, tals com: RGB, LIDAR, infraroges i geoespaci- als. Posteriorment, comparar els resultats entre les diferents fonts d'informació.

3.2 Objectius secundaris

Una vegada determinats els objectius més rellevants en l'apartat anterior, es proposa ampliar l'abast del projecte per tal d'assolir uns objectius complementaris. Aquests objectius secundaris són:

- Desenvolupar el codi de manera que permeti el processament massiu d'imatges d'alta resolució.
- Tenir en compte l'escalabilitat de la solució suggerida.
- Utilitzar *Transfer Learning* per escanejar paisatges no vistos en l'entrenament.

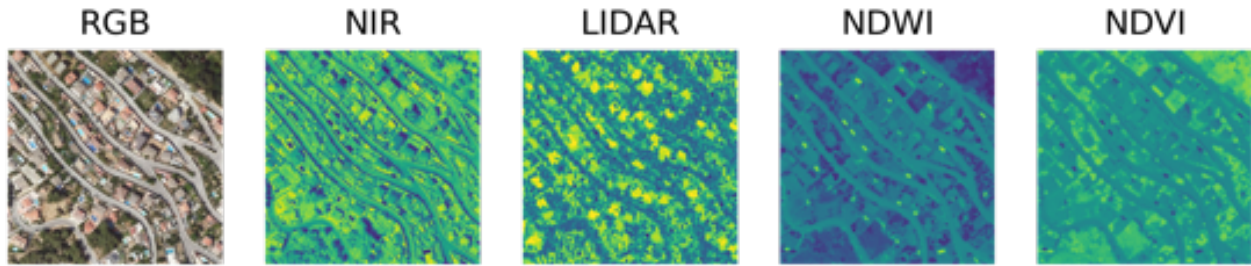


Fig. 1: Principals capes d'informació utilitzades per l'entrenament de les xarxes. Exceptuant la capa RGB, la resta d'elles són monocromàtiques. En aquesta figura es mostren les capes monocromàtiques amb un mapa de color (*cmap*) groc per facilitar la visualització.

- Implementar i comparar les funcions de *loss*: distància Jaccard i coeficient Sorenen-Dice.
- Exportar les deteccions a una plataforma que permeti visualitzar els resultats de manera gràfica.

4 METODOLOGIA

4.1 Planificació

Durant els mesos previs a l'inici oficial del TFG, s'ha pogut desenvolupar una versió preliminar que treballa amb imatges RGB i una xarxa U-Net. Això ha permès a escala personal quantificar el ritme i càrrega de treball suportable per tal de planificar de manera òptima les tasques i fites proposades en la Taula de planificació 5 a l'Annex.

4.2 Recursos

En el transcurs del projecte, es farà ús del GitHub per tenir un històric i control dels canvis en el codi a més de mantenir un seguiment de l'evolució del progrés. El *framework* utilitzat per treballar amb Keras en una targeta gràfica externa serà PlaidML [11]. S'ha optat per fer servir aquest compilador perquè permet treballar amb múltiples tipus de targetes gràfiques i sistemes operatius, incloent-hi MacOS. També s'ha disposat d'accés als recursos del Centre de Visió per Computador (CVC) i gràcies als seus servidors el temps d'entrenament s'ha reduït considerablement. Els recursos assignats constaven de dues targetes gràfiques Nvidia Titan Xp amb 3.840 nuclis CUDA cadascuna.

4.3 Capes d'informació

Les capes d'informació que proporciona l'Institut Cartogràfic i Geogràfic de Catalunya (ICGC) són les capes RGB, infraroges-properes i LIDAR. Aquestes, es complementaran amb la cartografia amb els índexs de vegetació i d'aigua (vegeu equacions 1 i 2 respectivament) utilitzades per Rodríguez-Cuenca i Alonso. Es poden observar les capes de les xarxes en la Figura 1.

Seguidament, es pretén millorar la precisió dels models a l'hora de detectar piscines parcialment o totalment cobertes per ombres i vegetació, afegint la banda c_3 (vegeu Equació 3) presentada per Ambrosio et. al. [20]. Aquest índex es basa en l'ús de la tercera banda de l'espai de color $c_1c_2c_3$ i les propietats invariants de les ombres en aquest espai.

$$c_3 = \arctan \left(\frac{BLUE}{\max \{RED, GREEN\}} \right) \quad (3)$$

on *RED*, *GREEN* i *BLUE* representen els components vermell, verd i blau respectivament de cada píxel en la imatge.

El conjunt de bandes d'entrada per les diferents xarxes neuronals es disposaria de la següent manera:

- RGB (3 canals de color)
- NIR (1 canal)
- LIDAR (1 canal)
- NDWI (1 canal) (Vegeu Equació 2)
- NDVI (1 canal) (Vegeu Equació 1)
- c_3 (1 canal) (Vegeu Equació 3 i Figura 2)

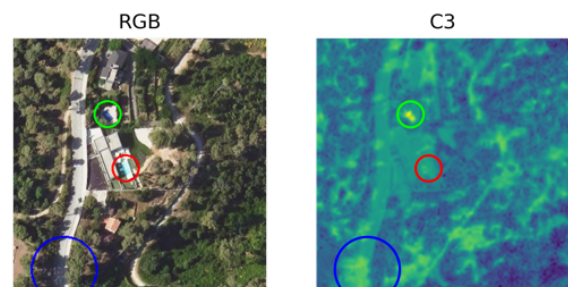


Fig. 2: Encerclada de color verd, es mostra una piscina coberta per l'ombra. S'observa que la seva intensitat és superior que la d'una piscina no tapada per una ombra (cercle de color vermell). Encerclada de color blau, es visualitza una àrea ombrívola a la carretera. En molts casos, aquestes últimes zones donen lloc a falsos positius.

4.4 Obtenció de les dades

L'ICGC permet descarregar tota mena de material cartogràfic i geogràfic a través de la seva pàgina web [12] i també gràcies a crides APIs [13], que faciliten l'automatització en l'extracció de grans àrees. Per dur a terme aquest projecte, primerament s'han descarregat ortoimatges a resolució 1:5.000 de diferents zones de la geografia catalana amb una alta densitat de piscines. En total, el conjunt de dades d'entrenament creat disposa de la següent estructura:

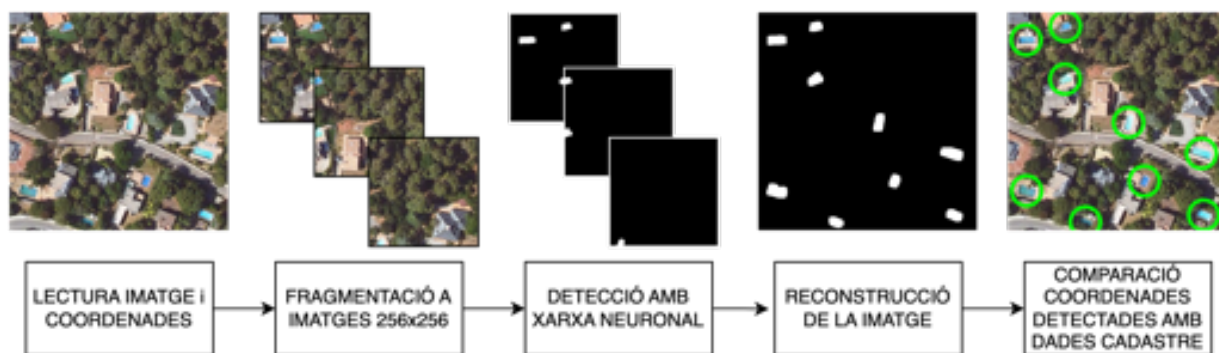


Fig. 3: Esquema sobre el funcionament de l'algorisme. Principalment dividit en 5 fases seqüencials on la sortida d'una etapa és l'entrada de la següent.

- 15 imatges de l'àrea metropolitana de Barcelona i urbanitzacions del litoral català.
- 60 quilòmetres quadrats de superfície.
- 8.270 piscines comptabilitzades.
- Imatges i fitxers RGB, NIR i LIDAR de resolució 4.000x4.000 píxels cadascuna.

El segon recurs utilitzat és la Seu Electrònica del Cadastre [14]. A través de la seva web, es poden descarregar de forma massiva dades cadastrals no protegides agrupades per municipi. Els arxius resultants de tipus *shapefile* [16], contenen les coordenades exactes de les construccions registrades i la seva forma espacial.

Per obtenir el *Ground Truth*, s'han generat imatges binàries amb les mateixes dimensions que la resta de fitxers de dades, on els píxels positius són aquells que contenen una piscina. Posteriorment, s'han validat i corregit de forma manual totes les piscines fent ús del programari d'Adobe Photoshop. Aproximadament el 20% de les piscines no estaven incloses en el registre oficial del cadastre o les seves coordenades presentaven errors de precisió.

4.5 Preprocessament i postprocessament

Les imatges aèries emprades tenen una resolució de 4000x4000 píxels (2 km x 2 km). Tal com es pot veure a petita escala en la Figura 3, l'algorisme fragmenta la imatge d'entrada en subimatges de la mida d'entrada de la xarxa neuronal i posteriorment, es combinen les segmentacions resultants per obtenir una imatge binària amb resolució 4000x4000 píxels.

Una característica per millorar la precisió en detecció de l'algorisme consisteix a superposar els fragments de les imatges que s'introdueixen a la xarxa i només utilitzar la zona central de la segmentació resultant. Aquesta solució augmenta considerablement el temps d'execució perquè s'incrementa el total de subimatges a processar, però augmenta de manera significativa les deteccions de piscines.

El fraccionament de les imatges ajuda a detectar les piscines, ja que, si una piscina es troba al límit de la imatge, es pot retallar i, com que no apareix completa, la xarxa disposa de menys informació per detectar els píxels contenidors de piscina.

En la Figura 4, l'algorisme ha generat per a una imatge de 768x768 píxels, un total de 25 mostres de 256x256 píxels (les finestres d'entrada corresponen als requadres vermells



Fig. 4: Mostreig d'imatge amb superposició. Els números de les imatges indiquen que els requadres de la imatge superior i inferior fan referència a la mateixa mostra. Els requadres vermells corresponen a l'àrea que analitzarà la xarxa neuronal. Els requadres blaus són les zones efectives reals que compondran la imatge binària final.

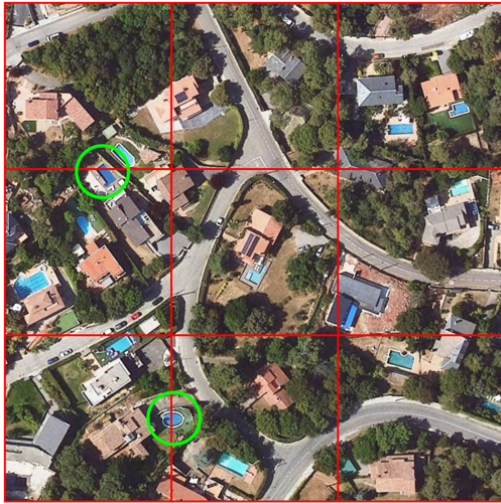


Fig. 5: Encerclades de color verd s'observen deteccions conflictives en cas de no aplicar solapament perquè es troben en el perímetre extern de la mostra.

de les Figures 4 i 5). Una vegada processades les 25 subimatges de la Figura 4, l'algorisme només utilitzarà la regió més cèntrica de cada finestra amb resolució 128x128 píxels (les regions finals corresponen als requadres blaus). Com s'observa en la Figura 5, la superposició és molt eficaç en aquelles piscines que, en cas de no aplicar la superposició, es trobarien en el límit dels fragments (piscines marcades en color verd) perquè la xarxa disposaria de poc context o àrea al voltant de la piscina per aplicar una correcta segmentació.

Cal afegir que l'algorisme amplia la regió de sortida (requadres de color blau de la Figura 4) en cas que aquestes regions es trobin en el perímetre extern de la imatge total d'entrada per tal d'emplenar la totalitat de la zona de segmentació. Aquestes mostres més externes del perímetre, no se superposen amb cap altra finestra (requadres de color vermell).

5 ARQUITECTURES DE XARXES NEURONALS

Per a la detecció de les infraestructures, s'ha optat per aplicar segmentació semàntica etiquetant tots els píxels de les imatges d'entrada segons si són píxels d'una piscina o no. Aquesta aproximació permet obtenir les superfícies de les piscines i classificar-les segons la seva àrea. Aquesta característica també és útil per descartar piscines inflables o temporals a causa de la seva superfície més reduïda.

5.1 U-Net

Per a la primera implementació del projecte s'ha optat per utilitzar una xarxa neuronal amb estructura de tipus U-Net [6] però modificada per tal de poder treballar amb imatges més reduïdes i així disminuir la complexitat computacional.

La xarxa U-Net és una arquitectura de xarxa neuronal emprada originalment en tasques de segmentació d'imatges biomèdiques. Es caracteritza per la seva estructura d'encoder-decoder, que permet capturar detalls locals mentre manté una visió contextual més ampla, proporcionant

resultats precisos en la segmentació de cèl·lules, òrgans o tumors en imatges mèdiques.

La xarxa resultant permet la segmentació d'imatges d'una resolució d'entrada de 256x256 amb profunditat variable segons el tipus de dada d'entrada. La xarxa manté les mateixes proporcions de capes i filtres que l'estructura original (vegeu diagrama d'arquitectura original en la Figura 6).

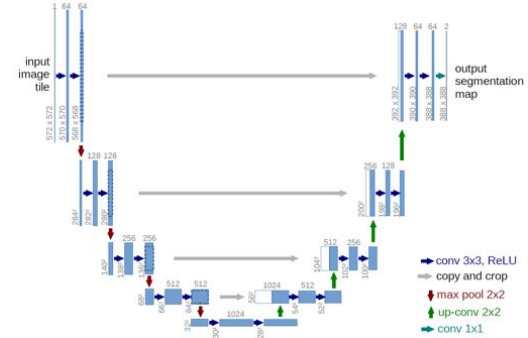


Fig. 6: Arquitectura de la xarxa U-Net presentada en l'article original. [6]

5.2 FCN

Les xarxes neuronals totalment connectades (*Fully Connected Networks*, FCN) són un tipus de xarxa neuronal artificial en què cada neurona en una capa està connectada a totes les neurones de la capa següent (vegeu Figura 7). Aquesta connexió completa permet que les dades flueixin lliurement a través de la xarxa, proporcionant una gran capacitat d'aprenentatge i modelatge de patrons, però no té en compte l'estructura espacial de les imatges. A diferència de la xarxa U-Net, l'arquitectura FCN no disposa de codificador i descodificador.

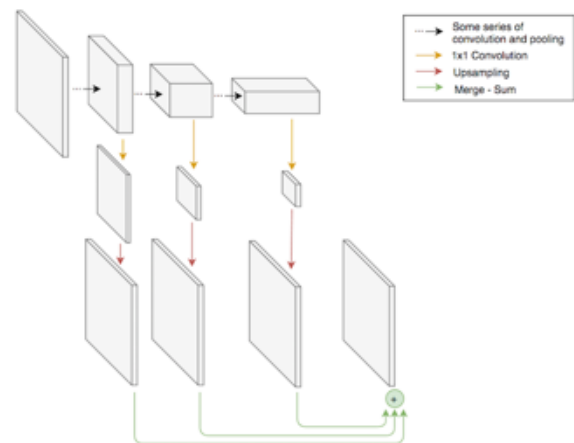


Fig. 7: Arquitectura FCN. [19]

5.3 DeepLab

A diferència de les xarxes U-Net i les xarxes totalment connectades (FCN), DeepLab implementa diverses tècniques clau per a millorar la precisió i la resolució en la segmentació. Una de les característiques principals de DeepLab és l'ús de dilatació espacial (o dilatació atrial) (vegeu dilatació dilatada en la Figura 9) en les capes convolucionals per augmentar el camp receptiu de les

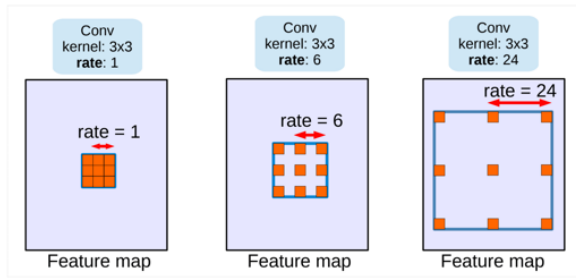


Fig. 9: Convolució atrial [10].

neurons sense augmentar significativament els paràmetres. Aquesta tècnica permet capturar informació contextual en diferents escales i millorar la precisió de la segmentació (vegeu arquitectura de la xarxa DeepLab en la Figura 8). [8, 9]

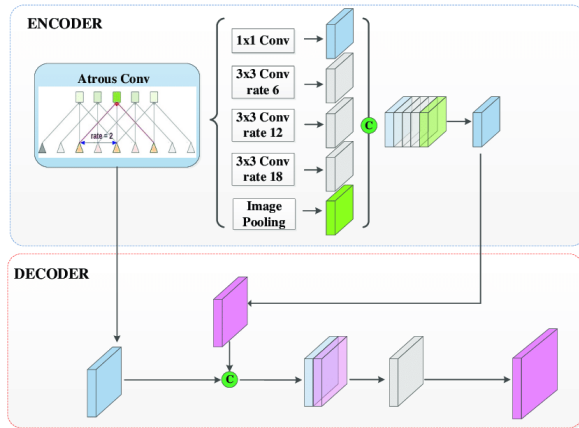


Fig. 8: Arquitectura DeepLab.

6 EXPERIMENTS

6.1 Dades d'entrenament i validació

Les dades de *test* són una mostra de 8 quilòmetres quadrats amb 1.675 piscines de la zona de Sant Quirze del Vallès i de Lliçà de Vall. Aquestes àrees no formen part de la zona d'entrenament. El conjunt de dades de *test* representa el 10% del total de les dades. El 90% restant s'utilitza completament per entrenar les xarxes (vegeu característiques de les dades d'entrenament en l'apartat 4.4). No es disposa de fase de validació a causa de la complexitat i cost de generar i etiquetar més imatges.

6.2 Mètriques de *loss*

Les mètriques de *loss* (o de pèrdua en català) s'utilitzen per avaluar com de bé la xarxa neuronal està aprenent durant el procés d'entrenament i són utilitzades per ajustar els pesos i les configuracions de la xarxa per aconseguir un millor rendiment.

Un dels problemes trobats a l'hora d'entrenar les xarxes és la manca de balanceig en les dades. El conjunt de píxels positius, aquells que pertanyen a una piscina, representen menys del 2% respecte al total de píxels.

Aquesta desigualtat no es pot corregir afegint més imatges de piscines perquè el conjunt de píxels positius en una imatge amb dues o tres piscines continua sent molt inferior respecte el total de píxels. Per poder afrontar aquest problema i evitar que la xarxa convergeixi en un punt on totes les classificacions són negatives, s'ha optat per utilitzar dues funcions de *loss* que s'adeqüen millor a la situació: la Distància de Jaccard (Equació 4) i el Coeficient de Sorensen-Dice (Equació 5).

Per poder quantificar l'impacte real de l'ús d'aquestes dues mètriques proposades, es defineix la *Binary Cross-Entropy* (BCE) (vegeu Equació 6) com a mètrica de punt de partida.

$$Jaccard = 1 - \frac{|U \cap V|}{|U \cup V|} \quad (4)$$

$$Sorensen - Dice = 1 - \frac{2 * |U \cap V|}{|U| + |V|} \quad (5)$$

$$BCE = -\frac{1}{N} \sum_{i=1}^N (U \log(V) + (1 - U) \log(1 - V)) \quad (6)$$

on U representa la classificació del *Ground Truth* i V el resultat de la segmentació.

Les dues *loss* proposades quantifiquen la diferència entre el *Ground Truth* i la sortida de la xarxa comparant les àrees d'intersecció. Per això, resulta molt adient per a aquest problema perquè no té en compte la majoria dels píxels negatius de la imatge i permet corregir la manca de balanceig de les dades.

6.3 Test i anàlisi dels models

En l'avaluació dels resultats, les deteccions es quantifiquen a escala de piscina i no de píxel. En finalitzar l'execució de l'experiment, s'obté una matriu de confusió (vegeu les Taules 6, 7 i 8 en l'apèndix amb les execucions amb millor resultat). Una vegada extretes les matrius de confusió, es calculen les mètriques del model utilitzant les equacions 7, 8, 9 i 10.

$$Accuracy = \frac{TP}{TP + FN + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (10)$$

on TP són el total de piscines que el model ha classificat correctament, FN són les piscines no detectades i FP són les deteccions incorrectament classificades com a piscines.

6.4 Entrenament dels models

Els models no tenen valors de pesos predefinitos durant la inicialització, s'utilitza la configuració per defecte. Segons la documentació oficial de Tensorflow, aquesta inicialització s'anomena: *Glorot uniform*. A continuació, es detallen els paràmetres d'entrenament:

- Mètrica de *loss*: Distància Jaccard.
- 500 èpoques.
- Optimitzador: Adam (amb paràmetres per defecte del *framework* Tensorflow).
- *Learning Rate*: 0.001.
- Únicament s'emmagatzema l'època amb la millor puntuació de *loss*.

6.5 Resultats execucions

Seguidament, en les Taules 1, 2 i 3, s'adjunten totes les execucions per a les tres arquitectures proposades amb combinacions de capes d'entrada diferents.

6.5.1 U-Net

Com s'observa en la Figura 10 entrenant els models únicament amb imatge RGB, s'obtenen molts falsos positius amb estructures de color verd i no es detecten correctament piscines brutes. Les capes NIR i LIDAR afegeixen la informació necessària per descartar les edificacions que poden confondre's amb una piscina i també per classificar correctament piscines brutes de color verd.

Això es pot atribuir a la diferència d'energia en les capes infraroges de l'aigua respecte a altres materials i també a l'ús de l'altura de les capes de LIDAR.



Fig. 10: Comparativa resultats capes d'entrada Xarxa U-Net. Cercles verds indiquen una correcta detecció, cercles vermells incorrecta i en blau piscines no detectades.

TAULA 1: RESULTATS EXECUCIONS U-NET

RGB	NIR	LIDAR	NDWI	NDVI	c_3	Accuracy	Recall	Precision	F_1 score
✓	✗	✗	✗	✗	✗	80,82	87,04	91,87	89,39
✗	✓	✗	✗	✗	✗	66,37	71,52	90,21	79,79
✓	✓	✗	✗	✗	✗	85,22	88,47	95,86	92,02
✓	✓	✓	✗	✗	✗	84,98	87,16	97,13	91,87
✓	✓	✗	✓	✗	✗	84,37	86,08	97,69	91,52
✓	✗	✗	✓	✗	✓	84,36	86,32	97,37	91,51
✓	✓	✓	✓	✓	✗	85,34	87,48	97,21	92,09
✓	✓	✓	✓	✓	✓	84,84	86,92	97,26	91,80

6.5.2 FCN

La xarxa FCN obté uns resultats numèrics similars a la xarxa U-Net en les mètriques d'*accuracy* i *recall*. Quan s'afegeixen capes d'informació per complementar les imatges només RGB, els models són més robustos i precisos a l'hora de detectar piscines (vegeu Taula 2).

Amb les mateixes capes RGB, NIR i LIDAR i comparant respecte a la xarxa U-Net, l'arquitectura FCN és més propensa a etiquetar incorrectament les ombres sobre carreteres. Per contra, l'arquitectura FCN és capaç de classificar piscines parcialment closes i buides (o en construcció) millor que l'arquitectura U-Net (vegeu Figura 11).

TAULA 2: RESULTATS EXECUCIONS FCN

RGB	NIR	LIDAR	NDWI	NDVI	c_3	Accuracy	Recall	Precision	F_1 score
✓	✗	✗	✗	✗	✗	80,84	88,95	89,86	89,40
✗	✓	✗	✗	✗	✗	65,66	71,82	88,45	79,27
✓	✓	✗	✗	✗	✗	82,72	88,05	93,17	90,54
✓	✓	✓	✗	✗	✗	84,88	88,83	95,01	91,82
✓	✓	✗	✓	✗	✗	83,79	88,29	94,26	91,18
✓	✓	✓	✓	✗	✗	84,64	89,19	94,31	91,68
✓	✓	✓	✓	✓	✓	84,24	89,07	92,95	90,97

6.5.3 DeepLab

Finalment, l'arquitectura DeepLab és la xarxa més conservadora de les tres. Genera menys deteccions que la resta, però la seva precisió és major (vegeu Taula 3 de resultats). Si s'observa la Figura 11, les classificacions de la xarxa DeepLab són encertades, però no és capaç de detectar totes les piscines conflictives, és a dir, aquelles que són buides i ocloses.

TAULA 3: RESULTATS EXECUCIONS DEEPLAB

RGB	NIR	LIDAR	NDWI	NDVI	c_3	Accuracy	Recall	Precision	F_1 score
✓	✗	✗	✗	✗	✗	78,72	83,52	93,20	88,09
✗	✓	✗	✗	✗	✗	69,99	73,25	94,00	82,34
✓	✓	✗	✗	✗	✗	84,38	86,80	96,80	91,53
✓	✓	✓	✗	✗	✗	84,99	86,20	98,36	91,88
✓	✓	✓	✓	✗	✗	83,39	84,53	98,40	90,94
✓	✓	✓	✓	✓	✗	84,10	85,91	97,55	91,36
✓	✓	✓	✗	✗	✓	82,95	84,83	97,39	90,68

6.6 Comparativa mètriques de loss

Les mètriques de *loss* utilitzades per a l'entrenament de les xarxes es compararan amb la *loss* més comuna en classificació binària, coneguda com a *Binary Cross-Entropy*, per avaluar el seu rendiment.

Per aquesta prova, totes les xarxes neuronals s'han entrenat en una arquitectura U-Net i amb els mateixos paràmetres definits en la secció 6.4. A més, les capes d'entrada per a les tres xarxes seran: RGB, NIR i LIDAR.

TAULA 4: RESULTATS EXECUCIONS MÈTRQUES LOSS

<i>loss</i>	Accuracy	Recall	Precision	F_1 score
Distància de Jaccard	84,98	87,16	93,13	90,05
<i>Binary Cross-Entropy</i>	75,25	84,23	87,58	85,87
Coefficient de Sorensen-Dice	84,22	85,73	97,95	91,43

Analitzant numèricament la Taula 4, s'observa que les dues mètriques de *loss* proposades (Distància de Jaccard i Coeficient de Sorensen-Dice) augmenten en 9 punts l'*accuracy* respecte a la *loss* sovint utilitzada de *Binary Cross-Entropy*.

En la Figura 12, observem que la distància de Jaccard és capaç de detectar piscines closes, però també és una mica més sensible a classificar incorrectament objectes de forma rodona com si fossin piscines buides.

Per contra, el coeficient de Sorensen-Dice no genera tantes deteccions, però la majoria de deteccions que realitza, les encerta. Per tant, la seva precisió és major que la resta de *losses*.

Finalment, *Binary Cross-Entropy* no convergeix de la mateixa manera que les altres dues mètriques. És més propensa a classificar incorrectament estructures semblants a piscines buides i també ombres de vegetació sobre carreteres. Podem atribuir aquest fet a la incapacitat d'aquesta *loss* de treballar amb dades imbalancejades, ja que el conjunt de piscines buides i closes és menor respecte a les piscines emplenades d'aigua.

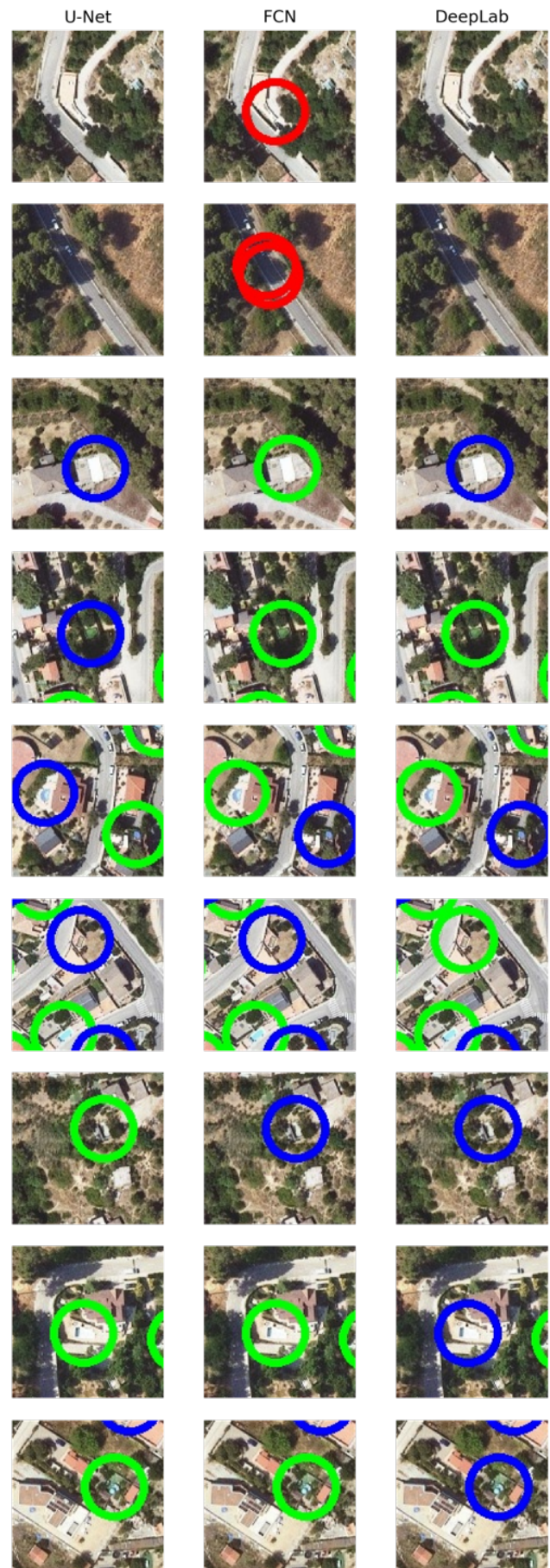


Fig. 11: Comparativa resultats xarxa U-Net, FCN i DeepLab. Totes elles entrenades amb les capes RGB, NIR i LIDAR. Cercles verds indiquen una correcta detecció, cercles vermells incorrecta i en blau piscines no detectades.

7 VISUALITZACIÓ DINÀMICA DELS RESULTATS

Com s'ha vist al llarg del treball (vegeu Figura 3), l'algorisme detecta les piscines de la imatge d'entrada i, següidament, converteix les coordenades píxel a coordenades geogràfiques per a comparar les deteccions amb el registre del cadastre.

Aquestes classificacions sobre l'estat legal de la piscina es complementen amb informació relativa a la identificació de la propietat, tals com: el codi de referència cadastral i la direcció postal.

Finalment, s'ha optat per mostrar els resultats i deteccions de l'algorisme en fitxers CSV a través de la plataforma de Carto. Carto és una eina que permet l'anàlisi i visualització de dades geospacial en el núvol. [17] Aquest recurs permet superposar les coordenades de les deteccions que genera l'algorisme amb mapes dinàmics juntament amb estadístiques, com ara: el percentatge de piscines no regulades donada una àrea (vegeu Figura 13 a l'Apèndix) o la referència cadastral associada a la parcel·la de la piscina (vegeu Figura 14 a l'Apèndix).

Els resultats de mostra es poden consultar a través del següent enllaç: <https://bit.ly/43L2s1S>. Aquests estaran disponibles fins quatre mesos després de la presentació del treball.

8 CONCLUSIONS

En aquest projecte, s'han estudiat tres mètodes diferents de segmentació d'imatges per a la detecció de piscines en imatges multispectrals. També s'ha determinat l'estat legal de la piscina comparant les deteccions respecte al registre del cadastre de manera automàtica.

Totes les arquitectures de xarxes neuronals analitzades han obtingut uns resultats molt similars entre elles. Es conclou que l'arquitectura no és un factor clau a l'hora de segmentar les imatges per satèl·lit.

El factor més significatiu i que millora considerablement les deteccions són les combinacions de capes d'entrada, juntament amb escollir una mètrica de *loss* que compensi la manca de balanceig en les dades. Les imatges RGB tradicionals i les infraroges per separat no són tan eficaces, però quan aquestes dues s'ajunten i es formen combinacions multispectrals, l'increment d'*accuracy* augmenta entre el 7% i el 31%.

Com a possibles vies de continuació d'aquest projecte, es podria explorar la segmentació i detecció d'altres infraestructures o problemàtiques que s'experimenten en el territori català, com ara: dins del context de la sequera es podria detectar el rec desmesurat i il·legal de jardins o també les zones més propenses a criar mosquits. També es proposa augmentar i posteriorment analitzar nous canals d'entrada gràcies al gran nombre de sensors que cada vegada més satèl·lits incorporen.



Fig. 12: Comparativa resultats de les mètriques de *loss* utilitzant xarxa U-Net i capes RGB, NIR i LIDAR. Cercles verds indiquen una correcta detecció, cercles vermells incorrecta i en blau piscines no detectades.

REFERÈNCIES

- [1] J. Bueno del Almo. (2022). “A la caza de las piscinas y ampliaciones de casas ilegales”. El País. Recuperat de: <https://elpais.com/economia/negocios/2022-08-20/a-la-caza-de-las-piscinas-y-ampliaciones-de-casas-ilegales.html>
- [2] BBC News. (2022). “Undeclared pools in France uncovered by AI technology”. Recuperat de: <https://www.bbc.com/news/world-europe-62717599>
- [3] B. Rodríguez-Cuenca i M.C. Alonso. (2014) “Semi-Automatic Detection of Swimming Pools from Aerial High-Resolution Images and LIDAR Data”. Remote Sens, n.6, 2628-2646.
- [4] C. Coelho, M. Fernanda P. Costa, L.L. Ferrás i A.J. Soares. (2021). “Development of a Machine Learning model and a user interface to detect illegal swimming pools”. SYMCOMP 2021, 5th International Conference on Numerical and Symbolic Computation: Developments and Applications.
- [5] B. Lima, L. Ferreira i J Martinho. (2021). “Helping to detect legal swimming pools with Deep Learning and Data Visualization”. Procedia Computer Science, n.181, 1058-1065.
- [6] O. Ronneberger, P. Fischer i T. Brox. (2015). “U-Net: Convolutional Networks for Biomedical Image Segmentation”. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 234-241.
- [7] H. Wu, J. Zhang, K. Huang, K. Liang i Y. Yu. (2019). “FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation”. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3217-3226.
- [8] L. Chen, G. Papandreou, I Kokkinos, K Murphy i A. Yuille. (2016). “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs”. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), n.40, 834-848.
- [9] L. Chen, Yukun Zhu, G. Papandreou, F. Schroff. (2018). “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation”. European Conference on Computer Vision (ECCV), 801-818.
- [10] L. Chen, G. Papandreou, F. Schroff i H. Adam. (2017). “Rethinking Atrous Convolution for Semantic Image Segmentation”.
- [11] PlaidML. Data d'accés: Novembre 2022. Recuperat de: <https://plaidml.github.io/plaidml/>
- [12] Institut Cartogràfic i Geològic de Catalunya. “Visor de descàrregues”. Data d'accés: Novembre 2022. Recuperat de: <http://www.icc.cat/appdownloads/>
- [13] Institut Cartogràfic i Geològic de Catalunya. “WMS Ortofoto”. Data d'accés: Febrer 2023. Recuperat de: <https://www.icgc.cat/ca/Administracio-i-empresa/Serveis/Geoinformacio-en-linia-Geoserveis/WMS-i-tessel·les-Cartografia-de-referencia/WMS-Ortofotos>
- [14] Sede Electrónica del Catastro. “Difusión de datos catastrales”. Data accés: Novembre 2022. Recuperat de: <https://www.sedecatastro.gob.es/Accesos/SECacDescargaDatos.aspx>
- [15] Institut Cartogràfic i Geològic de Catalunya. “Referencial Topogràfic Territorial”. Data d'accés: Novembre 2022. Recuperat de: <https://www.icgc.cat/Descarregues/Referencial-Topografic-Territorial>
- [16] ArcGIS Desktop. “Qué es un shapefile?”. Data accés: Abril 2023. Recuperat de: <https://desktop.arcgis.com/es/arcmap/latest/manage-data/shapefiles/what-is-a-shapefile.htm>
- [17] Carto. Data accés: Març 2023. Recuperat de: <http://carto.com/>
- [18] Institut Cartogràfic i Geològic de Catalunya. “ESPECIFICACIONES TÉCNICAS de l'Ortofoto infraroja de Catalunya 1:5 000 (OI- 5M) v2.0”. Data d'accés: Abril 2023. Recuperat de: <https://www.icgc.cat/Administracio-i-empresa/Mediana-tural/Imatges-aeries-i-de-satel·lit/Ortofoto-infraroja>
- [19] Lewis Fishgold i Rob Emanuele. “Deep Learning for Semantic Segmentation of Aerial and Satellite Imagery”. Data d'accés: Maig 2023. Recuperat de: <https://www.azavea.com/blog/2017/05/30/deep-learning-on-aerial-imagery/>
- [20] G. Ambrosio, V. Arévalo, J. González i J. Valdés. (2006). “Detección de sombras en imágenes de satélite de alta resolución Quickbird”. Recuperat de: <http://mapir.isa.uma.es/varevalo/drafts/ambrosio2006dsi.pdf>

APÈNDIX

A continuació, es mostren captures de pantalla de la plataforma interactiva Carto. El recurs estarà temporalment disponible en línia a través del següent enllaç: <https://bit.ly/43L2s1S>. Els punts verds de les imatges 13 i 14 fan referència a piscines inscrites en el registre del cadastre. Totes les piscines que per coordenades no es troben registrades en el cadastre, s'etiqueten com a il·legals i es mostren amb un cercle vermell.

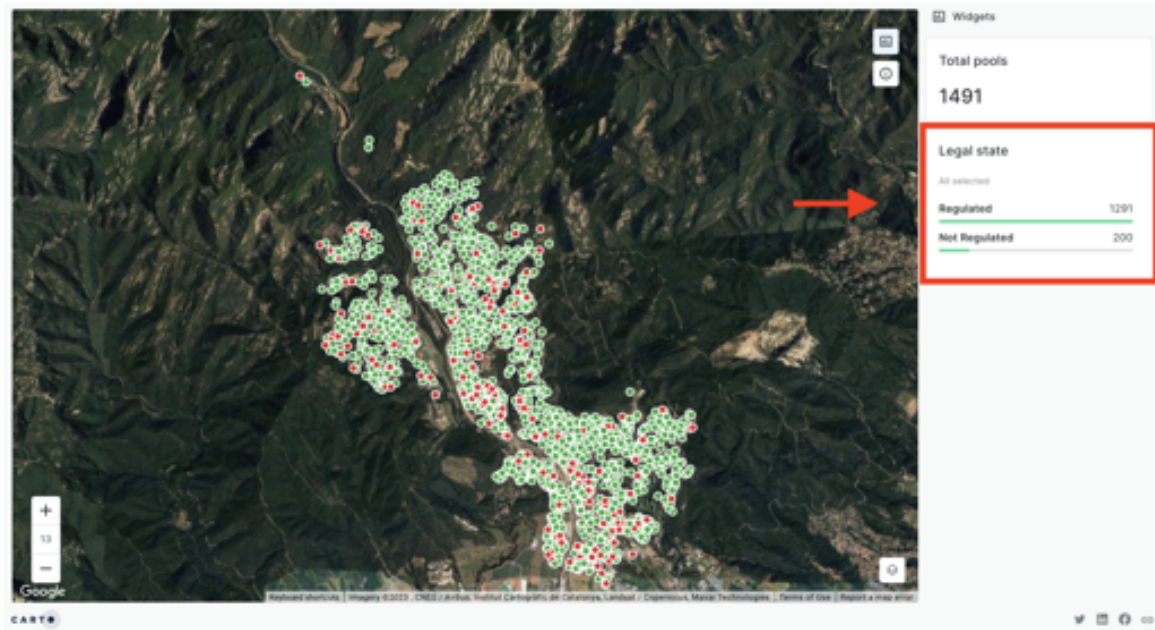


Fig. 13: Captura plataforma Carto amb mapa interactiu i estadístiques.

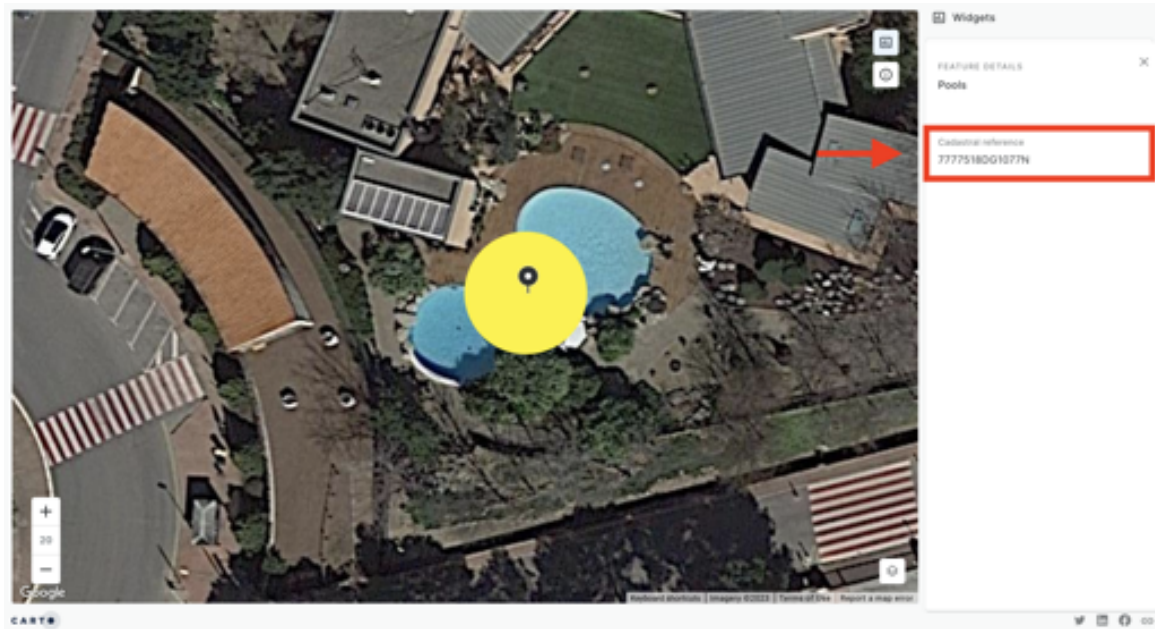


Fig. 14: Captura plataforma Carto amb mapa interactiu i referència cadastral de la parcel·la seleccionada.

TAULA 6: CONFUSION MATRIX XARXA U-NET I CAPES RGB, NIR I LIDAR

		GT	
		Piscina	No Piscina
Model	Piscina	1460	43
	No Piscina	215	∅

TAULA 7: CONFUSION MATRIX XARXA FCN I CAPES RGB, NIR I LIDAR

		GT	
		Piscina	No Piscina
Model	Piscina	1488	78
	No Piscina	187	∅

TAULA 8: CONFUSION MATRIX XARXA DEEPLAB I CAPES RGB, NIR I LIDAR

		GT	
		Piscina	No Piscina
Model	Piscina	1444	24
	No Piscina	231	∅

TAULA 5: PLANIFICACIÓ

ITERACIÓ	SETMANA	DATA INICI	TASQUES
1a	Setmana 1	20/02/2023	Reunió inicial Recopilar documentació tècnica Definir objectius
	Setmana 2	27/02/2023	Familiarització arquitectura U-Net Implementació U-Net
	Setmana 3	06/03/2023	Determinar mètriques de qualitat Codificar ETL per accedir a les dades Cadastre Implementar primera versió funcional Redacció informe inicial
2a	Setmana 4	13/03/2023	Generar dataset amb dades LIDAR Generar dataset amb dades infraroges Integrar noves dades en el codi actual
	Setmana 5	20/03/2023	Sol·licitar accés a recursos del CVC i realitzar migració de dades Buscar documentació d'arquitectura FastFCN Familiarització arquitectura FastFCN
	Setmana 6	27/03/2023	Desenvolupar model FastFCN
	Setmana 7	03/04/2023	Realitzar entrenament de la xarxa FastFCN Quantificar nous resultats respecte arquitectures prèvies
	Setmana 8	10/04/2023	Desenvolupar codi per a l'exportació de les deteccions Manteniment general del codi garantint l'escalabilitat
	Setmana 9	17/04/2023	Entrevista 2a iteració Redacció informe de progrés I
3a	Setmana 10	24/04/2023	Afegir canvis sobre correcció de l'informe previ Familiarització arquitectura DeepLab
	Setmana 11	01/05/2023	Buscar documentació d'arquitectura DeepLab Desenvolupar model DeepLab
	Setmana 12	08/05/2023	Realitzar entrenament de la xarxa DeepLab Quantificar nous resultats respecte arquitectures prèvies
	Setmana 13	15/05/2023	Avaluar els models en zones amb paisatge diferent respecte les dades de <i>training</i> Buscar documentació sobre mètriques de <i>loss</i>
	Setmana 14	22/05/2023	Extreure conclusions provisionals Entrevista 3a iteració Redacció informe de progrés II
4a	Setmana 15	29/05/2023	Realitzar experiments amb diferents <i>loss</i> Definir estructura i apartats de l'informe final
	Setmana 16	05/06/2023	Recopilació de dades per l'informe final Transcripció de resultats i generació de gràfics
	Setmana 17	12/06/2023	Entrevista 4a iteració Redactar informe final
5a	Setmana 18	19/06/2023	Preparar contingut gràfic presentació Reunió 5a iteració
	Setmana 19	26/06/2023	Realitzar correccions pertinents a la presentació Elaborar dossier final
	Setmana 19	03/07/2023	Maquetar pòster Assajar presentació