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

Aceves Soley, Gerard; Lumbreras Ruíz, Felipe, dir. Identificació de fonts d'emissió de metà d'aigües residuals amb visió per computador i observacions de Sentinel-2. 2021. (958 Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/248528>

under the terms of the  license

Identificació de fonts d'emissió de metà d'aigües residuals amb visió per computador i observacions de Sentinel-2

Gerard Aceves Soley

Resum—Les tecnologies basades en aprenentatge profund s'han desenvolupat a gran velocitat en els últims anys i alhora proporcionen eines de molta efectivitat. En aquest treball s'aplicaran aquestes tecnologies per a la detecció de fonts del gas metà, un dels gasos més contaminants a l'atmosfera, alhora que fàcilment localitzable en les depuradores, on es centrarà el treball. Per a la localització de depuradores s'ha utilitzat una xarxa neuronal del estat de l'art de la segmentació, la U-Net, i s'ha entrenat amb un dataset creat mitjançant imatges del satèl·lit de l'agència europea de l'espai Sentinel-2. Al mateix temps, la U-Net és modificada per a tenir multi input/output, és a dir que a part de la segmentació també volem que faci de regressor de la capacitat d'aigua tractada de la depuradora, passant-li el nombre d'habitants pròxims com a input addicional. També es comprova l'eficàcia d'aplicar super-resolució abans de fer la segmentació d'objectes, que no ha resultat particularment eficaç. Finalment els resultats són satisfactoris per a la detecció de depuradores i estimació del volum d'aigua tractada.

Paraules clau— Super-Resolució, Segmentació, Regressió, Aprenentatge Profund, Aprenentatge Automàtic, Imatges de satèl·lit, Xarxa Neuronal, Arquitectures de múltiple entrada/sortida

Abstract— Deep learning-based technologies have developed rapidly in recent years and at the same time provide highly effective tools. In this project, these technologies will be applied to detect sources of methane gas, one of the most polluting gases in the atmosphere, while being easily located in waste water treatment plants, which this work will focus. In order to detect these wastewater treatment plants, a state-of-the-art neural network, the U-Net, has been trained with a dataset created using imagery provided by the Sentinel-2 mission, from the European Space Agency. At the same time, the U-Net is modified to have multi input / output, since apart from the segmentation we also want it to act as a regressor of the treated water capacity of the treatment plant, passing the number of nearby inhabitants as additional input. The effectiveness of applying super-resolution before doing object sequencing is also tested, which has not been particularly effective. Finally, the results are satisfactory for the detection of treatment plants and estimation of the volume of treated water.

Index Terms— Super-Resolution, Segmentation, Regression, Deep learning, Machine learning, Remote sensing, Neural network, Multi input/output architectures



1 INTRODUCCIÓ

Un dels reptes més importants del segle XXI és el canvi climàtic. En el present ja en notem alguns efectes i a mig i llarg termini pot afectar radicalment la vida al planeta. L'emissió de gasos contaminants és un dels principals causants de l'efecte hivernacle que provoca un escalfament global del planeta, un d'aquests gasos és el Metà (CH₄).

El metà és un gas d'efecte hivernacle que no té tanta presència a l'atmosfera com el CO₂, però és molt més potent a nivell de contribució a l'efecte hivernacle segons l'informe del Intergovernmental Panel on Climate Change (IPCC)[2]. Segons la base de dades d'emissions per a la investigació global de l'atmosfera[1], un estudi que realitza un inventari mundial d'emissions, el sector del tractament d'aigües residuals va representar entre el 5 i el 20% del metà emès durant l'any 2012.

Els inventaris d'emissions actuals manquen de representativitat espacial ja que es basen en l'ús de la densitat de població per representar les emissions de les plantes de tractament d'aigües residuals.

En aquest treball es pretén desenvolupar una metodologia basada en Deep Learning (DL) per a la identificació i localització de fonts d'emissions de metà a Espanya, amb un enfocament a les depuradores. Per a l'entrenament dels models s'utilitzaran imatges de la European Space Agency (ESA) de la missió Sentinel-2[4] i es compararan amb imatges del Instituto Geográfico Nacional.

També com a objectiu que la metodologia que es pugui replicar a altres àrees geogràfiques on la identificació de fonts d'emissió sigui més escassa i així millorar els inventaris de dades d'emissions actuals.

- E-mail de contacte: gerard.aceves@e-campus.uab.cat
- Menció realitzada: Computació
- Tutor d'empresa: Carlos Alberto Gómez González
- Tutor acadèmic: Felipe Lumbreras Ruiz
- Curs 2020/21

2 OBJECTIUS

L'objectiu d'aquest treball és la identificació i localització de fonts d'emissions de metà a Espanya (depuradores) mitjançant algoritmes de Deep Learning, també com a objectiu es vol regressionar el volum d'aigua tractada en les depuradores. A continuació els objectius desglossats:

2.1 Creació dels datasets

Inicialment abans d'entrenar els models cal tenir preparat el data set, que consta de diferents parts:

2.1.1 Verificació de les coordenades

En aquest cas es disposa d'un llistat de coordenades que representen les ubicacions de les depuradores, aquestes coordenades no necessàriament representen la ubicació exacte, sinó que una aproximació. Per tant, s'ha d'inspeccionar visualment, al principi, per tenir unes coordenades fiables.

2.1.2 Extracció d'imatges

Posteriorment, es pretén extreure imatges de dues fonts, com a mínim una de Sentinel -2, que té resolució a 10m. També es vol obtenir imatges d'una altre d'una font de més resolució com les dades que ofereix el Instituto Geográfico Nacional (IGN). Per comparar diferències de resolucions i també avaluar la segmentació d'objectes després de la super-resolució.

2.1.3 Etiquetatge

En els models de Deep Learning de *object segmentation* cal crear entrenar-los amb exemples, per a generar aquests exemples cal fer el *labeling*, unes anotacions que contenen una màscara de segmentació, és a dir una delimitació de l'objecte que es vol detectar.

2.2 Super-resolució

Al disposar de imatges de territori espanyol de més alta resolució, es pretén entrenar un model de super-resolució per a augmentar la resolució d'imatges de Sentinel 2, l'objectiu és fer una super-resolució de x4.

Augmentar la resolució de les imatges de Sentinel té com a objectiu millorar la segmentació i regressió que es vol fer a posterior.

2.3 Segmentació i regressió

Es vol crear un model que pugui alhora donar una resposta per a la detecció de les depuradores i que també pugui proporcionar un valor de la predicció de la capacitat d'aigua tractada de la planta.

2.3.1 Model de segmentació semàntica

Escollir un model de l'estat de l'art de segmentació i entrenar-lo amb el *dataset* que s'ha creat anteriorment. L'entrada del model és una imatge i la sortida la màscara binària que determina si es detecten depuradores.

2.3.2 Població pròxima a les depuradores.

Gràcies al data set públic [15] que proporciona un mapa amb la densitat de població en una quadrícula es pot extreure una dada rellevant, la població pròxima, per a poder predir el volum d'aigua.

2.3.3 Afegir regressor

Finalment es vol adaptar el model de segmentació i modificar-lo de tal manera que tingui dues entrades: una imatge i la població pròxima corresponent a les coordenades de la imatge.

El model ha de donar com a sortides la detecció de depuradores i la regressió d'un valor de capacitat d'aigua tractada en el cas de que hi hagi una depuradora.

2.4 Establir conclusions

Verificar la utilitat d'aplicar super-resolució, cal veure si els models entrenats amb el mateix set d'imatges augmentades de resolució són més eficaços que els models entrenats amb les imatges originals.

Poder determinar si els mètodes desenvolupats són útils i si proporcionen una fiabilitat a l'hora de detectar depuradores i fer-ne una estimació de la capacitat de volum d'aigua tractada.

3 ESTAT DE L'ART

3.1 Super-Resolució

Hi han moltes xarxes rellevants en la super-resolució, alguns exemples són la Residual Dense Network (RDN) [21] o la Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR)[17] que és la que es farà servir en aquest treball.

3.2 Segmentació semàntica

Hi han diferents models a tenir en compte, en el paper de Abdul Mueed Hafiz et al. [11] en fan una comparativa. Alguns dels models més populars són Masked R-CNN [12], YOLOACT [13] o la U-Net [6] entre d'altres per a tenir en consideració. En aquest treball es farà servir una modificació de la U-Net.

3.3 Regressió

Per a implementar regressió mitjançant *Deep Learning* també s'ha estudiat un paper que compara l'estat de l'art en aquest tema, Stephane Lathuilière et al. [10] analitzen i fan una comparativa de l'actualitat en aquest tema. Tot i que aquest cas és molt específic ja que es pretén implementar modificacions a una xarxa per a poder realitzar la regressió alhora que la segmentació

4 METODOLOGIA

El treball s'ha fet amb col·laboració amb el Barcelona Supercomputing Center (BSC), el tema del projecte ha estat escollit per al BSC també. El treball ha estat supervisat pel tutor d'empresa Carlos Alberto Gómez González i el tutor acadèmic Felipe Lumbreras Ruiz.

La metodologia a seguir ha consistit de treball autònom que es combinava amb reunions setmanals amb els dos tutors, on es proposaven els següents passos a fer i també resolien els dubtes o m'aconsellaven en problemes que havien sorgit.

També disposava d'accés al supercomputador POWER9 del BSC, fet que ha permès realitzar els entrenaments amb molta més rapidesa i que no haurien estat possibles amb l'ordinador que disposava.

5 PREPARACIÓ DE LES DADES

5.1 Sentinel - 2

La missió de l'ESA Copernicus Sentinel-2 [4] té com a objectiu controlar la variabilitat de les condicions de la superfície terrestre i donar suport al seguiment dels canvis de superfície de la Terra. Consta de dos satèl·lits en òrbita que proporcionen imatges d'una mateixa zona cada 5 dies aproximadament.

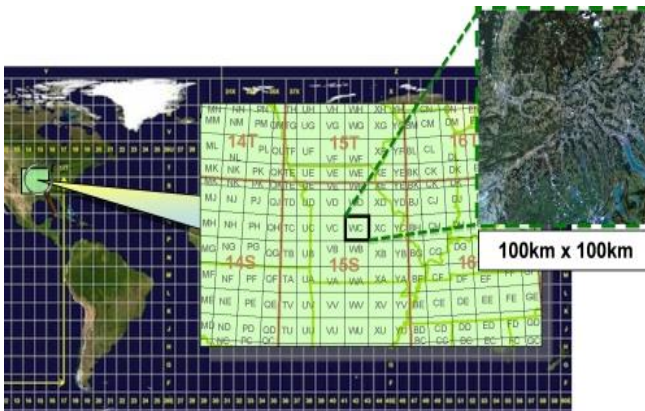


Fig. 1, distribució dels tiles de Sentinel-2

La missió de Sentinel-2 proporciona diferents tipus de productes, en aquest cas utilitzarem el Level-2A, que aporta imatges de *Bottom-Of-Atmosphere*, és a dir que generalment els núvols ja no es veuran. Totes les dades de la missió són públiques i es poden obtenir *on-line* gratuïtament.

Les imatges d'aquest producte de Sentinel-2 tenen diferents tipus de resolució, en el nostre cas per a aplicar mètodes de visió per computador ens interessa la màxima resolució per poder captar el detall. Les imatges amb més resolució són de 10 m (1 píxel = 10x10 m) i disposen 4 bandes,

els 3 canals de color RGB i la banda NIR (Near Infra-Red). Tot i que no s'utilitzarà la banda NIR és important a mencionar ja que en un futur seria una dada a utilitzar que podria millorar els resultats. A més, les imatges estan separades en forma de quadrícula en *tiles* de 100 x 100 km com es pot veure a la Fig. 1.

5.2 Creació del dataset

En la UWWT-DB [3], un llistat públic amb gran quantitat de dades relacionades amb les depuradores ofert per l'agència europea del medi ambient (EEA) es detallen les coordenades de les depuradores registrades a Espanya. En el nostre cas la regió delimitada. N'hi ha aproximadament unes >3000 depuradores llistades. En aquest treball es crearà un *dataset* inicialment de 500 plantes de tractament d'aigües.

Encara que hi ha les coordenades a la UWWT-DB cal inspeccionar visualment, ja que les ubicacions no són exactes en la majoria de casos. D'aquesta manera s'ha elaborat un llistat de coordenades en que les depuradores estaven confirmades en la ubicació.

En aquest cas s'ha decidit que per cada depuradora s'extraurà una imatge del Sentinel-2 de 1 km x 1 km que ve a ser 100 x 100 píxels tenint en compte la resolució de 10 m. Paral·lelament per poder comparar l'efecte en la diferència de resolució també s'han extret imatges del Instituto Geográfico Nacional (IGN) de les mateixes depuradores, tot i que la resolució d'aquestes pot arribar a ser de 1 m (1 píxel = 1x1 m), és a dir 10 vegades més resolució respecte a la de Sentinel-2.

Per a entrenar els models de *Object segmentation* cal tenir imatges de depuradores amb la màscara corresponent i el procés s'anomena *labeling*. Per al *labeling* de segmentació s'ha utilitzat VIA[14] i consta de manualment delimitar la regió de la imatge en que hi ha l'objecte que volem detectar per a totes les imatges de les que es disposa. En la figura 2 es pot veure una imatge de Sentinel-2 amb la màscara de segmentació anotada corresponentment.

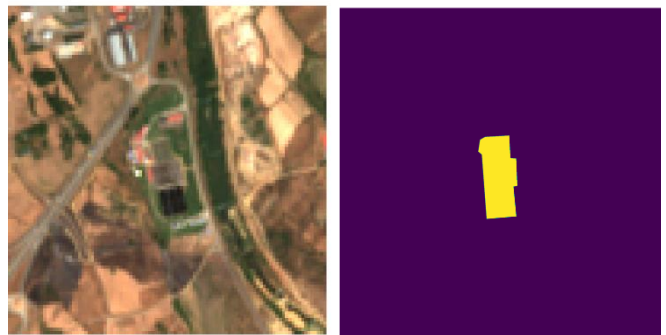


Fig. 2, exemple d'imatge del dataset de Sentinel-2 amb la màscara

6 SUPER-RESOLUCIÓ

La Super-Resolució permet augmentar la resolució d'una imatge mitjançant una xarxa convolucional de Deep Learning. L'objectiu és augmentar la resolució de les imatges de Sentinel-2 per a poder aplicar posteriorment els altres mètodes amb més èxit que tenint les imatges amb menys resolució. El factor de *scaling* que es vol aplicar és de x4.

Per a entrenar un model de super-resolució es necessita com a entrada una imatge a baixa resolució i com a sortida una imatge a alta resolució. Com que no disposem imatges de Sentinel-2 de més resolució es vol entrenar el model amb imatges del IGN, en el que les d'entrada tenen la mateixa resolució que Sentinel-2 i les de sortida tenen una resolució de x4.

Posteriorment es realitzarà inferència sobre les imatges de Sentinel-2 amb el model entrenat amb les imatges del IGN. Com que són les mateixes plantes i totes les imatges són de satèl·lit, és a dir amb la mateixa orientació serà possible realitzar aquesta transferència d'aprenentatge.

6.1 Model utilitzat

La xarxa utilitzada per a la super-resolució és de una estructura basada en la ResNet[18] però que utilitza els blocs residuals de la proposta Bee Lim et al. Anomenada EDSR[17] i que finalment llavors aplica una capa de Subpixel Convolution[16]. L'estructura de la xarxa utilitzada consta de múltiples blocs residuals, cada bloc residual conté dues capes de convolució separades per una ReLU com a funció d'activació, que finalment es sumen a l'entrada del bloc (*skip connection*). D'aquests blocs residuals s'hi ha 20 i són entre dues capes de convolució en les que es fa una *skip connection* també. Finalment hi ha el mòdul de Subpíxel-Convolution que és el que realitza l'augment de resolució, ja que reorganitza diferents canals i els intercala amb un mateix patró per a generar una imatge més gran. A la Fig. 3 podem veure l'estructura descrita amb més detall.

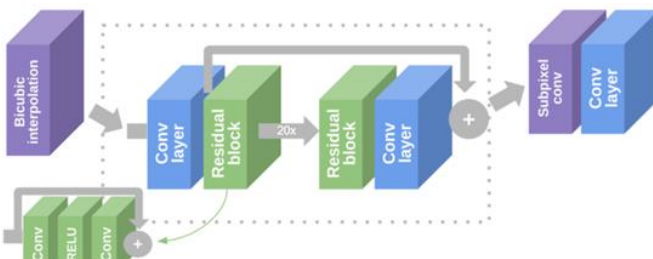


Fig. 3, arquitectura de la xarxa de super-resolució

6.2 Entrenaments

Per a entrenar el model de super-resolució s'ha utilitzat imatges del IGN de resolució 10 m i 2,5 m respectivament per als parells de baixa-alta resolució. Això significa que el factor d'augment és de x4. Tot i que les imatges del IGN

podrien arribar a tenir més resolució, creiem que aspirar a un x4 és realista si llavors es volen tenir resultats amb Sentinel-2. S'han utilitzat 20 capes de blocs residuals i s'ha entrenat 100 *epoch*. També s'ha realitzat *data augmentation* (flip vertical i horitzontal) perquè només disposem de 500 imatges.

Un cop entrenat el model es passen les imatges de Sentinel-2 per a que aquest faci una inferència. Com que Sentinel-2 no disposa de més resolució no podem entrenar un model en que la imatge d'alta resolució sigui de Sentinel-2 per tant utilitzarem la tècnica de *transfer learning*, ja que les imatges són similars perquè en tots casos imatges ortonormals, també de la mateixa resolució i dels mateixos llocs.

6.3 Resultats

Finalment, la inferència del model la podem veure en els exemples de les següents figures. En les figures 4 i 5 les imatges originals de S-2 i en les figures 6 i 7 les imatges de Sentinel-2 amb super-resolució de x4. Numèricament no es pot verificar la qualitat de la super-resolució ja que al no tenir imatges de més resolució de Sentinel-2 no es poden establir les mètriques habituals de super-resolució. Tot i així, visualment sembla correcte però tampoc aporten una gran millora de qualitat.



Fig. 4, imatge de Sentinel-2 nativa



Fig. 5, super-resolució resultant aplicada a la fig. 4



Fig. 6, imatge de Sentinel-2 nativa



Fig. 7, super-resolució resultant aplicada a la fig. 6

7 SEGMENTACIÓ I REGRESSIÓ

Aquesta fase té un objectiu doble, no només realitzar la detecció de les depuradores en imatges de satèl·lit, sinó que també predir un valor de capacitat d'aigua tractada de la planta.

La capacitat d'aigua tractada a una depuradora és un indicador del que es disposen dades, existents a la UWWT-DB [3]. Alhora, és una dada que està relacionada amb les emissions de metà ja que són molt directament proporcionals, així que és una bona aproximació.

7.1 U-Net i modificacions

U-Net és una de les xarxes que realitzen segmentació semàntica més utilitzada. En aquest cas, es volen introduir unes modificacions per tal de que pugui rebre dues entrades: una imatge del Sentinel-2 i un vector representant la població pròxima corresponent a les coordenades de la imatge per a diferents radis de distància.

L'estructura de la U-Net està dividida en l'encoder i el decoder que s'ajunten al *bridge*. El primer consisteix en una

convolució de 3×3 , i a continuació una activació de tipus ReLU, això dues vegades i posteriorment i una operació de *max-pooling* de 2×2 . Aquest bloc es repeteix cinc vegades, l'última ja és el *bridge* perquè just després comença el *decoder*. En aquest, es repeteix el mateix bloc de dues convolucions 3×3 , cadascuna seguida de una activació ReLU però amb la diferència que al final hi ha una operació de *up-convolution* de 2×2 . Després de cada bloc del codificador es realitza una còpia i es concatena al mateix nivell que al descodificador. Aquesta part de la xarxa és molt semblant a la de l'U-Net original, només canvien les mides d'entrada de les imatges. L'input1 és la imatge i l'output2 és la màscara binària que representa si hi ha trobat depuradores o no.

Per a adaptar la xarxa a dues entrades i dues sortides s'ha fet el següent; primer el vector que conté les quantitats de població segons els diferents radis és un vector de 6 posicions que va seguit d'una Dense de 64, aquesta va seguida d'una Dense de 32, posteriorment això es concatenarà amb el resultat del Flatten del *bridge* que segueix amb una Dense de 64, també seguida d'una Dense de 32. L'input2 és un vector de 6 posicions i l'output2 és un únic valor. A la figura 8 es pot veure l'arquitectura de la U-Net amb les modificacions afegides.

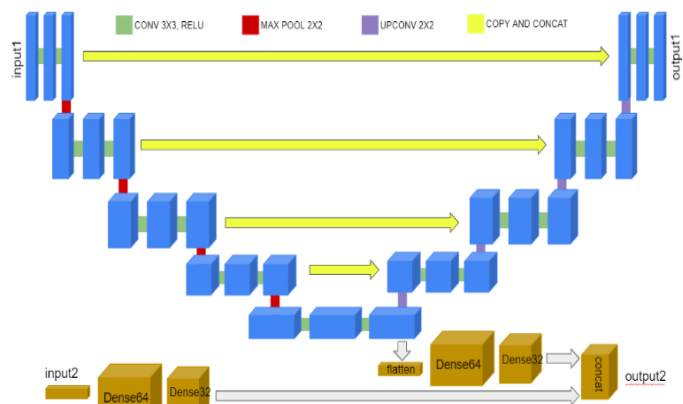


Fig. 8, arquitectura de la U-Net modificada

7.2 Dades disponibles

Les dades que disposem, a part de les imatges i les màscares binàries anotades, són la capacitat en volum d'aigua (metres cúbics) que tracta cada depuradora i un mapa de població proporcionat per la Global Human Settlement Layer[15] dades obertes al públic i gratuïtes proporcionades per la UE. Aquesta és una quadrícula que representa la densitat de població i que pot donar la població propera a la depuradora en un cert radi, ja que disposem de les coordenades de les plantes de tractament.

D'aquesta quadrícula representant la quantitat d'habitants per quilòmetre quadrat, en la figura 9 es pot veure una representació d'aquestes dades per la península ibèrica en que com més color significa més habitants per metre quadrat.

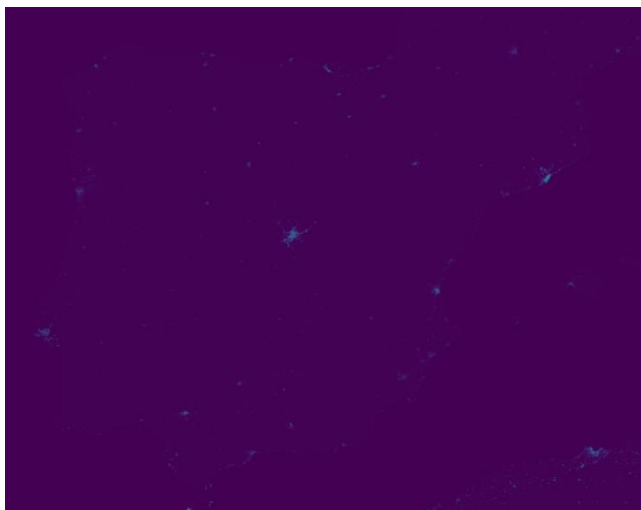


Fig. 9, densitat de població de la península ibèrica

Per a intentar relacionar i utilitzar com a regressor la dada de habitants en un cert radi al voltant de la planta s'han realitzat proves amb la correlació de Pearson, intentant relacionar capacitat de la planta amb nombre d'habitants en un cert radi. La millor relació és amb un radi de aproximadament 10 km al voltant, de aproximadament un coeficient de 0,7, això dona una correlació prou forta, la corba de correlacions es pot veure a la figura 10.

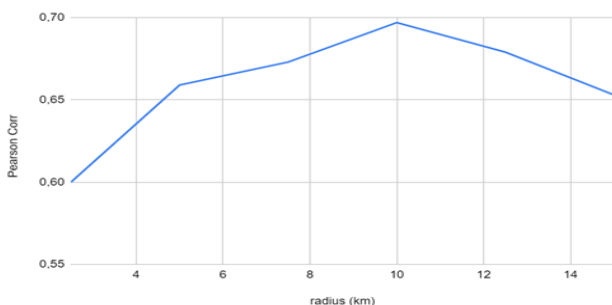


Fig. 10, gràfica de la correlació de Pearson de la capacitat d'aigua de la depuradora respecte a quantitat d'habitants segons radi

S'ha decidit que es passaran sis valors de població al voltant d'un cert radi, els valors són (2.5, 5, 7.5, 10, 12.5, 15) en km, ja que tot i que la correlació més forta és amb radi 10 km, per a casos de depuradores més grans també és bo passar-li àrees més grans. En l'apèndix A1 es pot veure els diferents radis amb la població per un exemple de depuradora.

Per a fer-nos una idea de les dades de població (utilitzarem les de habitants en un radi de 10 km) i capacitat d'aigua tractada en cada planta en la següent figura (fig. 11) podem veure un histograma en cada eix, on l'eix x és la població i l'eix y la capacitat de les depuradores i al centre el núvol de punts de les dues variables però agrupats en densitat, ja que sinó serien indistingibles. Per a fer el gràfic s'han obviat els casos on la població era superior a 25000 habitants ja que hi ha poblacions tan grans que provocaven una mala visualització del gràfic.

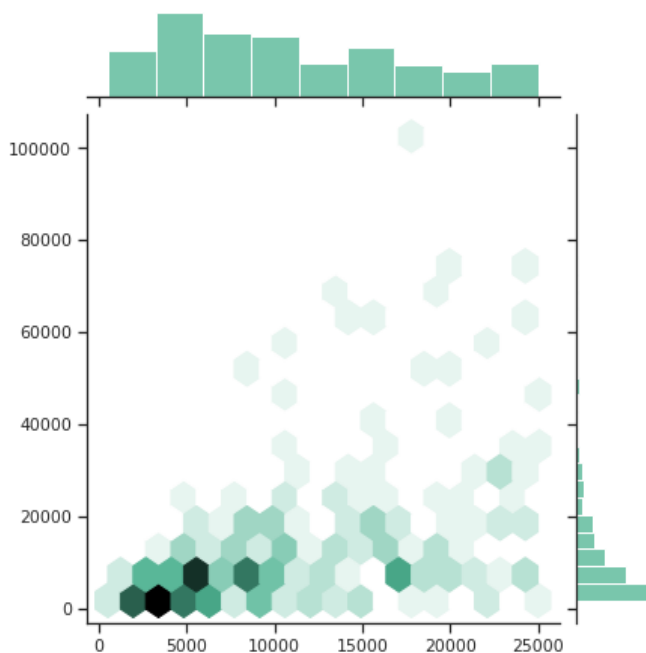


Fig. 11, histograma de població en radi de 10 km (eix x) i capacitat de la depuradora (eix y), al mig densitat dels exemples.

7.3 Entrenaments

Per a provar la U-Net modificada s'han provat amb dos sets. El primer és el de Sentinel-2 natiu sense aplicar super-resolució, que consta d'imatges de 100 x 100 píxels amb la màscara corresponent i llavors també el vector de poblacions i la capacitat de la planta. L'altre és el de Sentinel-2 un cop aplicat super-resolució en el que les imatges són de 400 x 400 píxels i la màscara corresponent així com el mateix vector de població i capacitat de la planta.

Pels entrenaments s'ha fet servir en tots el casos 20 *epochs*, 0,0001 *learning rate*, amb un lleuger *decay* a partir de la 10 *epoch* i finalment seleccionem el model amb menys *validation loss*. També cal mencionar que s'estandarditzen abans d'entrenar els valors numèrics de població i capacitat.

Del total de depuradores, 500, s'han establert un 10% pel test, 10% de validació i 80% de *training*, és a dir un total de 50 de test, 50 de validació i 400 per entrenament. Només en les mostres d'entrenament s'ha fet *data-augmentation* ja que disposen de poques dades, *flip* horitzontal, *flip* vertical, i les tres rotacions possibles de 90 graus. També s'han realitzat els entrenaments a dues particions per tal de poder fer la mitjana dels resultats.

Els paràmetres que s'avaluaran en els experiments són la quantitat de filtres en cada nivell de la U-Net, i el *loss weight* que consisteix en la ponderació a la *loss* total de cada sortida, és a dir la segmentació té una *loss* i la regressió una altre i llavors es multipliquen per un paràmetre cadascuna que conforma la *loss* total de la xarxa.

7.4 Resultats

En la següent taula (taula 1) podem veure els experiments realitzats, amb les especificacions de cada experiment.

Nom	Loss_weights	Filtres
exp1	1, 1	[32, 64, 128, 256, 512]
exp2	5, 1	[32, 64, 128, 256, 512]
exp3	10, 1	[32, 64, 128, 256, 512]
exp4	1, 5	[32, 64, 128, 256, 512]
exp5	10, 1	[64, 128, 256, 512, 1024]
exp6	10, 1	[16, 32, 64, 128, 256]

Taula 1

En la següent taula (taula 2) podem veure la mitjana de les diferents mètriques per a les dues particions de les dades per a cada experiment per les imatges de Sentinel-2 de resolució nativa. Cal notar que IoU (Intersection over Union), *precision* i *recall* avaluen la segmentació i el *mean absolute error* (mae) avalua la part de regressió.

Exp	IoU	precicison	recall	mae
exp1	0,371	0,749	0,461	0,465
exp2	0,381	0,759	0,454	0,485
exp3	0,387	0,722	0,506	0,470
exp4	0,345	0,694	0,448	0,542
exp5	0,380	0,714	0,482	0,516
exp6	0,348	0,756	0,450	0,462

Taula 2

Primer de tot, comparant els resultats en que diferien els *loss_weights*, és a dir exp1, exp2, exp3 i exp4, podem dir que no canvia gaire per els tres primers, excepte que donant-li més pes a la *loss* de regressió si que els resultats són pitjors (inclús el mae). En tot cas el que dona IoU més elevat és amb la combinació de *loss_weights* de [10, 1]. Llavors per a comparar les diferents quantitats de filtres serien l'exp3, exp5 i exp6 i tampoc es poden apreciar millores especialment significatives quan s'augmenten el nombre de filtres de la xarxa de la mateixa manera que al reduir-los empitjora, per això escollir la menor quantitat de filtres resulta en una estructura més lleugera.

A continuació, en la següent taula (taula 3) podem veure la mitjana de les diferents mètriques per a les dues particions de les dades segons cada experiment per les imatges de Sentinel-2 un cop aplicada la super-resolució x4. Similar al que passa amb els experiments de Sentinel-2 a resolució

nativa, si comparem l'efecte del canvi de *loss_weights*, donar més pes a la regressió no té efecte en el mae, en canvi, si que empitjora les mètriques de la part de segmentació. La millor combinació de *loss_weights* podríem dir que experimentalment és la [10,1]. Per altra banda si analitzem l'efecte de la quantitat de filtres sembla que no tenen pràcticament efecte augmentar-los o disminuir-los, per tant segurament seria preferible la menor quantitat ja que la xarxa és més senzilla i requerirà menys espai i menys temps d'entrenament.

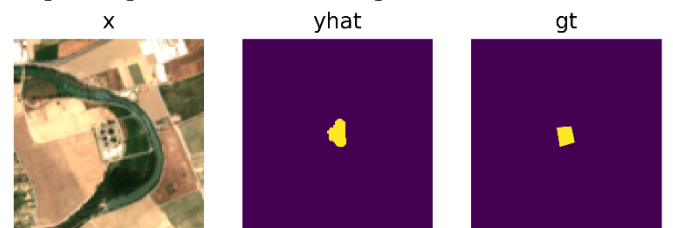
Exp	IoU	precicison	recall	mae
exp1	0,286	0,714	0,461	0,473
exp2	0,339	0,710	0,491	0,454
exp3	0,349	0,754	0,514	0,450
exp4	0,241	0,755	0,399	0,451
exp5	0,376	0,777	0,508	0,501
exp6	0,365	0,764	0,497	0,479

Taula 3

Resumint, analitzant entre experiments que comparin *loss_weights* donar-li més pes a la *loss* de regressió afecta negativament, sobretot a la IoU i no millora el mae. I la quantitat de filtres no sembla afectar als resultats. També si comparem els resultats de Sentinel-2 amb resolució nativa i Sentinel-2 amb super-resolució x4, es pot dir que no aporta cap millora la super-resolució, empitjora una mica les mètriques de segmentació i millora lleugerament la regressió.

En la següent figura (fig. 12) hi ha un exemple de la inferència que retorna la xarxa sobre dades de test. Més exemples a l'apèndix A2.

Capacitat predita: 22084 m3 Capacitat real: 30000 m3



Capacitat predita: 261721 m3 Capacitat real: 767283 m3

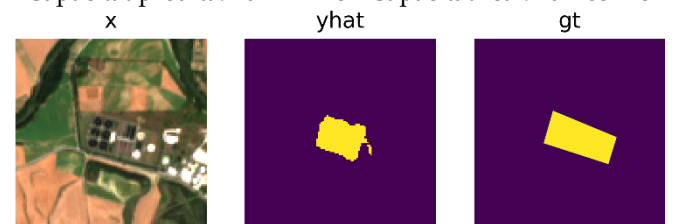


Fig. 12, exemples de resultats de la xarxa de segmentació i regressió

8. CONCLUSIONS

En aquest projecte s'ha implementat una xarxa de super-resolució i una de segmentació i regressió combinades i s'han avaluat els resultats de la segona xarxa comprovant tant amb imatges a resolució nativa i les imatges augmentades de resolució.

Els resultats dels experiments realitzats conclouen que els *loss_weights* han de donar-li més pes a la segmentació i que la quantitat de filtres no sembla afectar als resultats.

Els resultats de la super-resolució no han aportat millores significatives, en algun cas lleugerament han empitjorat les mètriques de segmentació i en algun cas milloren les de regressió.

Tot i així, cal remarcar que al no tenir imatges de Sentinel-2 de alta resolució, és un repte intentar aplicar super-resolució ja que no es pot entrenar una xarxa amb parells de imatges baixa-alta resolució del mateix tipus, per això s'ha hagut de realitzar *transfer learning* i inferir les imatges de Sentinel-2 d'alta resolució a partir d'un model entrenat amb imatges del Instituto Geográfico Nacional. És a dir, queda la incògnita de què passaria amb un model de super-resolució més eficaç.

També, el *dataset* té una mida més aviat reduïda, i encara que aplicant *data-augmentation* falten dades d'entrenament, a que en models de deep learning se solen utilitzar molts més exemples. Tot i que les mètriques i resultats a nivell visual són bons, podrien ser millors. Possiblement, la causa principal és la mida petita del data set que es disposa i d'altra banda que la resolució de les imatges de Sentinel-2 no és gaire detallada (per exemple, una petita depuradora en que la piscina fa 4m quadrats seran literalment 4 píxels) i és un repte intentar-hi fer segmentació semàntica.

Durant aquest treball hem pogut aprendre sobre dos dels paradigmes de la visió per computador més rellevants, la super-resolució i també la segmentació. Alhora que també hem après a construir xarxes neuronals, tant de segmentació com de super-resolució, des de zero, entrenar-les, fer-li *fine-tuning* i inclús modificar-les.

També hem pogut utilitzar unes instal·lacions com dels de Barcelona Supercomputing Center, tenir contacte amb investigadors del centre i veure el funcionament, l'organització i interacció d'unes màquines tan potents.

AGRAÏMENTS

Primer de tot als tutors. Tant el d'empresa, Carlos Alberto Gómez González com l'acadèmic, Felipe Lumbreras Ruiz, per a haver realitzat un molt bon seguiment i haver ajudat en moltes casos i aportar idees quan no sabia per on continuar.

També a la institució del Barcelona Supercomputing

Center per haver-me acceptat com a aprenent i també posar a disposició uns recursos computacionals de últimes prestacions amb els quals s'han pogut realitzar els experiments amb molta més rapidesa.

BIBLIOGRAFIA

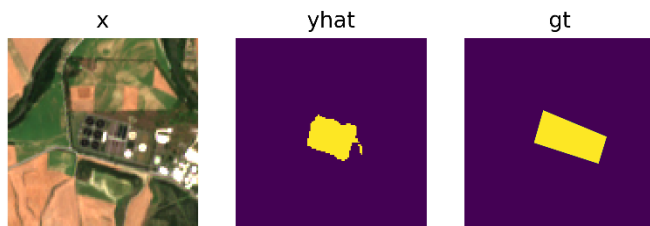
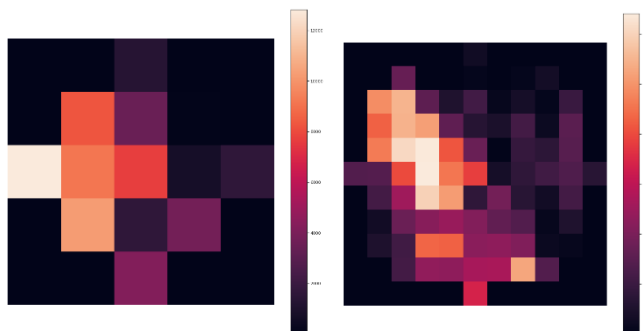
- [1] Greet Janssens-Maenhout, Monica Crippa, Diego Guizzardi, Marilena Muntean, Edwin Schaaf, Frank Dentener, Peter Bergamaschi, Valerio Pagliari, Jos G. J. Olivier, Jeroen A. H. W. Peters, John A. van Aardenne, Suvi Monni, Ulrike Doering, A. M. Roxana Petrescu, Efisio Solazzo, and Gabriel D. Oreggioni. "EDGAR v4.3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970–2012", 2019, <https://essd.copernicus.org/articles/11/959/2019/>, accedit juny 2021
- [2] Intergovernmental Panel on Climate Change (IPCC), 2019, <https://www.ipcc.ch/report/2019-refinement-to-the-2006-ipcc-guidelines-for-national-greenhouse-gas-inventories/>, accedit juny 2021
- [3] Urban Waste Water Treatment DataBase (UWWT-DB), <https://www.eea.europa.eu/data-and-maps/data/waterbase-uwwt-d-urban-waste-water-treatment-directive-z>, accedit juny 2021
- [4] European Space Agency (ESA) Sentinel-2 mission <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>, accedit juny 2021
- [5] LabelImg <https://github.com/tzutalin/labelImg>, accedit juny 2021
- [6] Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", 2015, <https://arxiv.org/abs/1505.04597>, accedit juny 2021
- [7] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", 2016, <https://arxiv.org/abs/1506.01497>, accedit juny 2021
- [8] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár. "Focal Loss for Dense Object Detection", 2017, <https://arxiv.org/abs/1708.02002>, accedit juny 2021
- [9] Licheng Jiao, Fan Zhang, Fang Liu, Shuyuan Yang, Lingling Li, Zhixi Feng, Rong Qu. "A Survey of Deep Learning-based Object Detection", 2019, <https://arxiv.org/abs/1907.09408>, accedit juny 2021
- [10] Stephane Lathuilière, Pablo Mesejo, Xavier Alameda-Pineda, Member IEEE, and Radu Horaud. "A Comprehensive Analysis of Deep Regression", 2020, <https://arxiv.org/abs/1803.08450>, accedit juny 2021
- [11] Abdul Mueed Hafiz, Ghulam Mohiuddin Bhat. "A Survey on Instance Segmentation: State of the art", 2020, <https://arxiv.org/abs/2007.00047>, accedit juny 2021
- [12] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick. "Mask R-CNN", 2017, <https://arxiv.org/abs/1703.06870>, accedit juny 2021
- [13] Daniel Bolya, Chong Zhou, Fanyi Xiao, Yong Jae Lee. "YOLACT: Real-time Instance Segmentation", 2019, <https://arxiv.org/abs/1904.02689>, accedit juny 2021
- [14] Abhishek Dutta, Andrew Zisserman, VGG Image

- Annotator, 2019, <https://arxiv.org/abs/1904.10699> , accedit juny 2021
- [15] GHSL - Global Human Settlement Layer, <https://ghsl.jrc.ec.europa.eu/download.php?ds=pop>, accedit juny 2021
- [16] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, Zehan Wang, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network", 2016, <https://arxiv.org/abs/1609.05158>, accedit juny 2021
- [17] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, Kyoung Mu Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution", 2017, <https://arxiv.org/abs/1707.02921>, accedit juny 2021
- [18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun , "Deep Residual Learning for Image Recognition", 2015, <https://arxiv.org/abs/1512.03385> , accedit juny 2021
- [19] TensorFlow Object Detection API, https://github.com/tensorflow/models/tree/master/research/object_detection , accedit juny 2021
- [20] Coco dataset, <https://cocodataset.org/#home> , accedit juny 2021
- [21] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, Yun Fu, "Residual Dense Network for Image Super-Resolution", 2018, <https://arxiv.org/abs/1802.08797> , accedit juny 2021

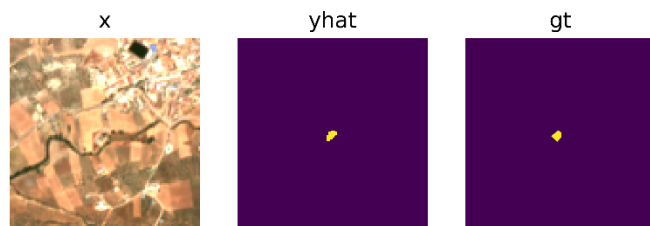
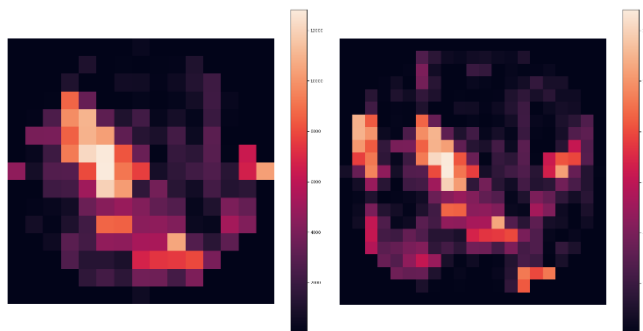
APÈNDIX

pred cap: 261721.47843348814, gt cap: 767283.0

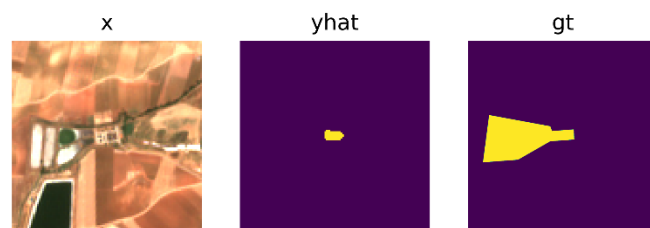
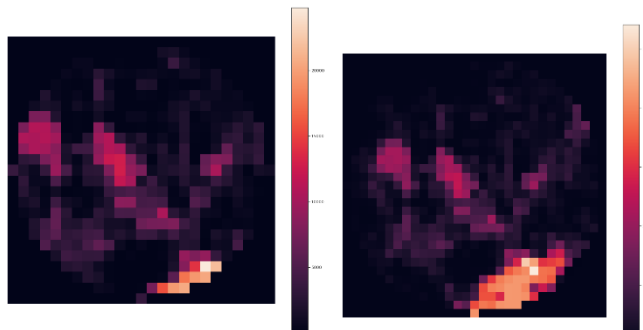
A1. SECCIÓ D'APÈNDIX



pred cap: 48545.542932009244, gt cap: 6000.0



pred cap: 42786.66830050481, gt cap: 28560.0



A2. Secció d'apèndix

pred cap: 262685.5288813551, gt cap: 330000.0

pred cap: 676362.6936371123, gt cap: 1353600.0



pred cap: 22084.184109669062, gt cap: 30000.0

