Integració de Transformades Wavelet en la Superresolució d'Imatges de Satèl·lit

Rubén Martínez León

2 de juliol de 2024

Resum– El projecte tracta sobre la implementació i avaluació d'un model de superresolució d'imatges de satèl·lit i la integració de transformades Wavelet utilitzant una arquitectura basada en SRGAN (Superresolution Generative Adversarial Network). L'objectiu principal és modificar l'entrenament del generador per incorporar la transformada de Wavelet, donant lloc al model anomenat Wavelet-SRGAN. Es busca comparar la qualitat de les imatges generades amb el model SRGAN original per avaluar si els coeficients de Wavelet aporten millores significatives en termes de detall i nitidesa de les imatges. Aquest estudi té implicacions importants en àrees com la cartografia, la vigilància i la meteorologia, on la qualitat de les imatges de satèl·lit és crucial.

Paraules clau– Superresolució, xarxes generatives adversàries, SRGAN, transformada de Wavelet, imatges de satèl·lit.

Abstract– This project focuses on the implementation and evaluation of a super-resolution model for satellite images, and the integration of Wavelet transforms using the SRGAN architecture (Super-Resolution Generative Adversarial Network). The main objective is to modify the training of the generator to incorporate the Wavelet transform, resulting in a model called Wavelet-SRGAN. The goal is to compare the quality of the images generated with the original SRGAN model to assess whether Wavelet coefficients bring significant improvements in terms of image detail and sharpness. This study has important implications in areas such as cartography, surveillance, and meteorology, where the quality of satellite images is crucial.

Keywords– Super-resolution, generative adversarial networks, SRGAN, Wavelet transform, satellite images.

1 Introducció

La superresolució d'imatges és un tema d'investigació molt actiu en els últims anys, que tracta sobre la reconstrucció d'imatges d'alta qualitat a partir d'imatges de baixa qualitat. En el camp de la fotografia digital, la qualitat de les imatges es redueix considerablement quan s'intenta ampliar la resolució de tota una imatge o zona en específic. Aquest problema es va tractar de solucionar mitjançant algoritmes matemàtics que interpolaven les imatges per ampliar la resolució sense perdre qualitat, aquests algorismes són els que utilitzen els nostres ordinadors i telèfons mòbils en ampliar una imatge, però causen una perduda de detalls en els petits patrons, reducció de la nitidesa dels contorns i incapacitat de reconstruir parts no existents en la imatge original. Amb la implementació de xarxes neuronals i no simples interpolacions, es va veure que les xarxes podien aprendre una gran quantitat de característiques de les dades d'entrada, i fins i tot poder reconstruir zones no existents en la imatge original, gràcies a les dades que s'introduïen en la fase d'entrenament.

És per això que aquest treball se centra a abordar la millora de la qualitat de les imatges de satèl·lit mitjançant tècniques de superresolució. Aquestes imatges són capturades per satèl·lits que orbiten la Terra els quals estan equipats amb càmeres i diferents sensors. Solen tenir una perspectiva zenital, és a dir, una vista des de dalt, que permet tenir una visió àmplia a gran escala de grans àrees geogràfiques i els seus canvis al llarg del temps. Per tant, quan més gran sigui l'àrea de la imatge, més quantitat de terreny s'equipararà a un píxel i, per tant, menys detalls es podran distingir.

Les imatges de satèl·lit són una eina que s'utilitza en una gran varietat d'aplicacions com la cartografia, l'agricultura, la gestió de desastres, planificació urbana, monitorització de

• E-mail de contacte: ruben.mtez.leon@gmail.com
• Mención realitzada: Computació
• Treball tutoritzat per: Robert Benavente Vidal (Ciències de la Computació)
• Curs 2023/24

recursos i observació de la Terra. No obstant això, aquestes imatges solen tenir limitacions en la seva resolució a l'hora d'ampliar una zona en específic de la imatge, ja que un píxel pot representar una gran quantitat de terreny, i per tant pot afectar la seva utilitat i precisió en tots aquests contextos. Per tant cal tractar aquestes imatges per aconseguir una millora en la seva resolució i permetre que siguin mes aprofitables en tots el àmbits nombrats anteriorment, i la solució proposada es la utilització d'algorismes de superresolució. Aquests algoritmes poden ser claus per la detecció primerenca d'activitats il·legals com la desforestació, al detectar petits canvis en la cobertura forestal dins d'una imatge amb una gran quantitat de terreny boscos o per la identificació precisa d'àrees de cultius afectades per plagues o malalties, la qual cosa permet un millor ús de fertilitzants.

La transmissió de dades des dels satèl·lits cap a la Terra presenta reptes significatius, incloent-hi limitacions de cost i tècniques. El volum de dades generades per les càmeres i els sensors a bord dels satèl·lits pot ser massiu, i la transmissió d'aquestes dades cap a la Terra pot resultar costosa i tècnicament complexa. Com a resposta a aquests desafiaments, sovint s'opta per capturar imatges de baixa resolució durant els passatges dels satèl·lits sobre les àrees d'interès.

Aquesta estratègia, tot i que redueix els costos i la complexitat associats amb la transmissió de dades, comporta una pèrdua de detalls i resolució en les imatges capturades. Per superar aquesta limitació i obtenir imatges d'alta qualitat, és necessari aplicar tècniques de superresolució una vegada les imatges de baixa resolució s'han transmès a la Terra. Aquest procés permet millorar significativament la qualitat de les imatges sense la necessitat de transmetre grans volums de dades des de l'espai, facilitant-ne l'anàlisi i utilitat en una àmplia gamma d'aplicacions

Per a abordar aquest problema, s'ha realitzat una recerca exhaustiva sobre les tècniques de superresolució d'imatges i l'estat de l'art d'aquesta tecnologia. La cerca s'ha centrat en l'ús de xarxes neuronals convolucionals (CNN) a causa de la seva eficiència en tasques de processament d'imatges. S'han consultat estudis i documents tècnics relacionats amb la superresolució d'imatges i s'ha dut a terme una anàlisi i comparativa de diversos mètodes i arquitectures de models existent

2 Estat de l'art

En el context de les imatges digitals, la resolució es refereix a la capacitat de la imatge per representar detalls i informació visual. Aquesta resolució no depèn únicament del nombre absolut de píxels que componen l'imatge, sinó també de la densitat de píxels per unitat d'àrea, com són els píxels per polzada (PPI) o per centímetre (PPC) [1]. Per tant, una major quantitat de píxels no necessàriament es tradueix en una millor resolució; és la combinació de la quantitat de píxels i la seva densitat la que determina la qualitat visual i la capacitat per observar detalls en una imatge.

2.1 Interpolació de píxels

Quan intentem ampliar una imatge digital més enllà de la seva resolució original, el problema que existeix és que s'han de crear nous píxels que conviuran entre els píxels originals d'aquesta zona ampliada. El valor d'aquests nous píxels desconeguts es calcula mitjançant una sèrie d'operacions matemàtiques a partir de valors de píxels coneguts. És així, que programes de visualització o d'edició d'imatges utilitzen la interpolació de píxels existents per a crear nous píxels addicionals, això es coneix com a "escalat".

Existeixen diferents tècniques d'interpolació de píxels que ajudaven a augmentar la resolució d'una imatge. La més senzilla és la interpolació proximal o interpolació del veí més proper, com es veu a la figura (1) aquesta augmenta la resolució de la imatge principal deixant forats negres entre píxels i es completen amb el valor del píxel més proper a aquest forat [2]. Per tant, el que s'està fent és un augment de la mida dels píxels, en conseqüència es perd qualitat a la imatge.

Amb la implementació de xarxes neuronals i no simples interpolacions, es va veure que podien aprendre una gran quantitat de característiques de les dades d'entrada, i fins i tot poder reconstruir zones no existents en la imatge original, gràcies a les dades que s'introduïen en la fase d'entrenament.

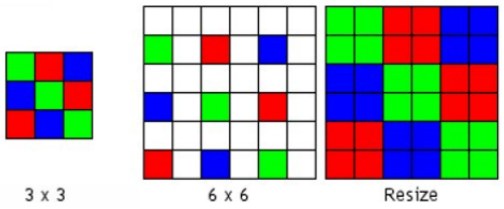

Fig. 1: Interpolació proximal [3]

L'estàndard utilitzat en programes de disseny d'imatges o senzillament que necessitin augmentar la mida d'aquestes, han sigut la interpolació bilineal i bicúbica [4]. Com es veu a la figura (2), aquestes tècniques calculen una interpolació entre els 8 o 16 píxels més propers als forats negres que es creen a l'augmentar la mida de la imatge, en el cas de la interpolació bilineal de 8 píxels [5], i en el cas de la bicúbica de 16 píxels, per tant, als veïns en un radi de 2 i 4.

Aquestes tècniques proporcionen més nitidesa que la interpolació proximal, i en conseqüència un augment del temps de processament.

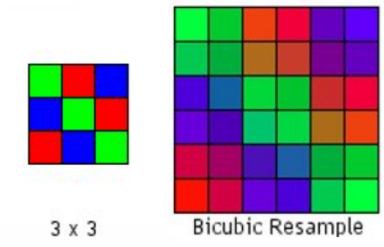

Fig. 2: Interpolació bicubica [6]

El problema principal dels processos explicats anteriorment és la reducció de la qualitat de la imatge, ja que s'estan afegint detalls que no són presents en la imatge original i l'ampliació excessiva d'una imatge pot portar a l'aparició d'efectes visuals adversos, com a vores pixelades, falta de nitidesa i pèrdua de detall. També un dels grans problemes d'aquests algoritmes es que no poden introduir res que no sigui visible en la imatge original.

Amb l'arribada d'intel·ligència artificial es va revolucionar la manera en què es millorava la qualitat d'imatges en baixa resolució, creant algoritmes i models de superresolució que milloraven les tècniques anteriors. Ja no es dependria només de la interpolació de píxels. Aquestes noves tècniques de superresolució aprofiten la potència de la intel·ligència artificial, en particular, de les xarxes neuronals convolucionals (CNN). Aquestes xarxes poden aprendre directament la relació entre les imatges de baixa resolució i alta resolució a partir de dades d'entrenament dels dos conjunts mencionats.

A través de l'aprenentatge profund, les CNN poden capturar característiques complexes de les imatges i generar detalls realistes en les imatges d'alta resolució, sense simplement interpolar els píxels [7]. Aquestes característiques es poden extreure mitjançant l'aplicació de filtres als inputs d'entrada de cada capa. Aquests filtres es realitzen mitjançant la finestra de kernel, i la distància de desplaçament d'aquesta finestra pels píxels de la imatge. Això permet obtenir resultats de superresolució més precisos i realistes.

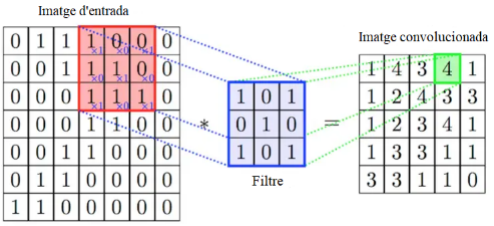

Fig. 3: Pas d'una imatge en una capa de convolució [8]

A la figura (3) es pot observar com és tractada una imatge d'entrada a una capa de convolució. A la imatge d'entrada se li aplica un filtre (matriu blava) a cada posició del kernel (matriu vermella), aquest kernel es va desplaçant per la imatge d'entrada horitzontalment i, per tant, en aplicar-se el filtre es va creant una nova imatge (matriu verda).

2.2 Models generatius

La SRCNN [9] va ser la primera xarxa neuronal convolucional presentada al 2014 per atacar el problema de la superresolució. Aquesta xarxa estava composta per nomes tres capes convolucionals, de 64 , 32, i 1 filtre respectivament, finestres kernel de 9x9 en la primera capa, 5x5 a la segona i 5x5 a la tercera, el que fa que actualment no es consideri una xarxa profunda. La imatge de baixa resolució d'entrada s'augmentava a les dimensions de sortida desitjades mitjançant la interpolació de píxels explicada anteriorment, i la imatge ja estava preparada per ser modificada pels diferents filtres de la xarxa.

Aquesta arquitectura va poder reconstruir detalls més nítids i textures més realistes comparades amb els algoritmes existents, i va augmentar significativament el PSNR (Peak Signal-to-Noise Ratio), la ràtio de relació entre la imatge de sortida resultant de la xarxa i la imatge real d'alta resolució en el set de dades de validació. Aquest avenç significatiu va establir un nou estàndard en la millor de qualitat d'imatges, demostrant el potencial de les xarxes convolucionals.

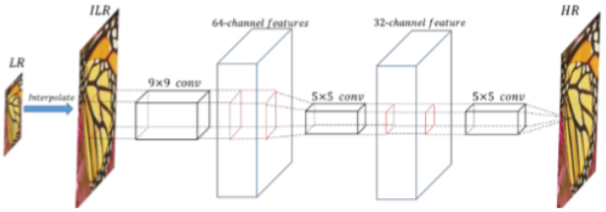

Fig. 4: Arquitectura SRCNN [9]

La utilització de blocs convolucionals d'iteracions es va utiltzar per primera vegada a RCAN (Residual Channel Attention Networks) [10]. Al contrari que SRCNN, aquesta xarxa sí que es considera profunda, ja que com bé s'explica a l'article publicat, a la seva versió més profunda arriba fins a 400 capes convolucionals. La clau de RCAN és la utilització en la seva arquitectura d'una gran quantitat de blocs constituïts per un conjunt de capes, el que facilita molt la seva implementació i mitjançant un bucle es poden aplicar més o menys blocs depenent de la profunditat desitjada. Aquesta tècnica es va utilitzar en tots els grans models posteriors, inclòs en el model realitzat en aquest projecte.

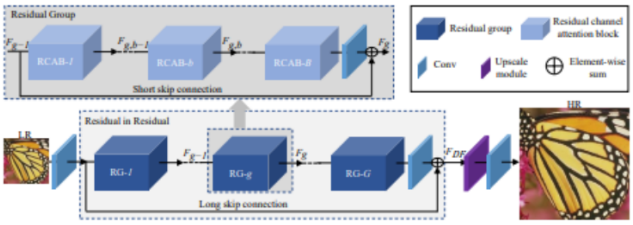

Fig. 5: Arquitectura RCAN [10]

La imatge original en baixa resolució s'introdueix en una primera capa convolucional i seguidament s'introduirà en el primer dels 6 Residual Groups (RG), on cadascun d'aquests està compost per un conjunt de 16 Residual Channel Attention Blocks (RCAB). Aquests estan formats per 4 capes convolucionals amb funcions Relu i normalitzacions, que afegeixen no linealitat al model. Aquests blocs fan que el model se centri en les característiques més importants de la imatge i les zones que s'han de modificar per millorar la qualitat de la imatge i eliminar el soroll. Finalment, les últimes dues capes del model són una capa d'escalat i una capa final de convolució.

Finalment, el model que s'utilitzarà com a punt de partida d'aquest treball és SRGAN (Superresolution Generative Adversarial Network) [11]. És un model que segueix la metodologia GAN, les quals estan compostes per dues xarxes neuronals que lluiten entre elles amb la finalitat d'aconseguir la imatge amb millor qualitat a partir de les dades d'entrada. Aquestes dues xarxes s'anomenen Discriminador i Generador [12].

Discriminador o Model Discriminatiu: Classifica la imatge com a real o falsa i estima la probabilitat que una mostra provingui des de les dades d'entrenament, aprèn quina és real i quina no.

Generador o Model Generatiu: És el que rep com a entrada imatges de baixa resolució, que pot ser una imatge real de baixa resolució o una imatge escalada de l'original, i intenta replicar la versió en alta resolució, creant unes noves

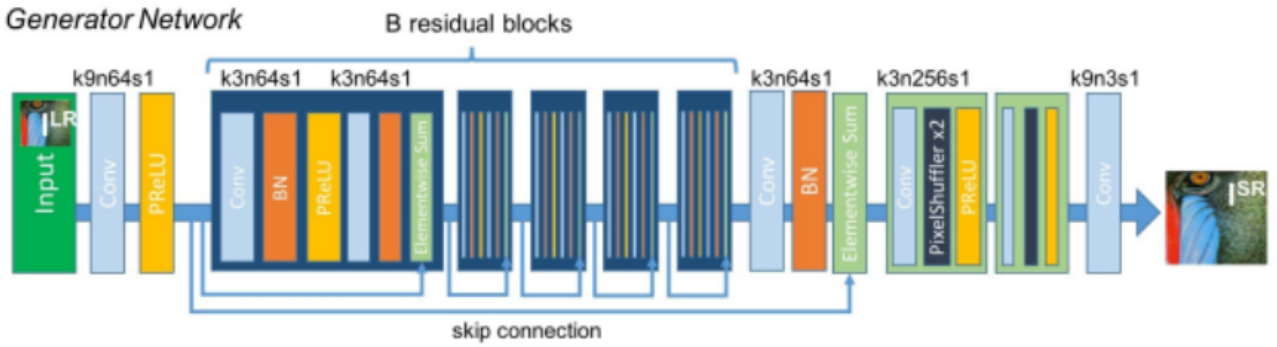

Fig. 6: Arquitectura Generador SRGAN [11]

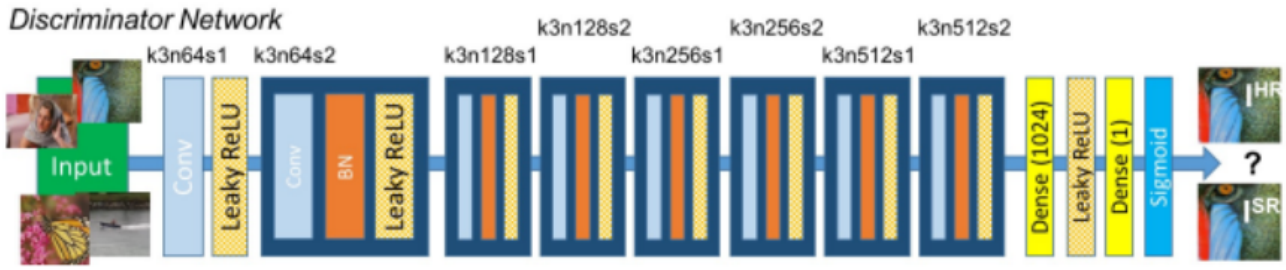

Fig. 7: Arquitectura Discriminador SRGAN [11]

imatges a fi d'enganyar el Discriminador. En alguns casos es rep un vector aleatori, això permet tenir més flexibilitat i poder tenir una gamma més extensa de característiques al passar pels filtres de les capes denses.

A la figura (6) es pot veure l'arquitectura del generador de SRGAN. Aquesta està formada per una capa convolucional amb una funció Relu posterior, amb un valor 9 de kernel, un desplaçament d'1 i un augment de 4 píxels a la vora de la imatge. Seguidament 7 Residual Blocks, on els primers 5 Residual Blocks estan formats per dues capes convolucionals amb un valor 3 de kernel i un desplaçament de 2, amb Batch Normalitation. Aquests blocs ajuden a prevenir el problema de desenfocament a l'hora d'aplicar la superresolució. El cinquè bloc només conté una capa convolucional amb una capa de Batch Normalitation, i l'últim bloc està destinat a reescalar la imatge.

El discriminador està format per una consecució de capes convolucionals seguides d'una Batch Normalitation amb funció Relu, la capa convolucional més profunda arriba aplica fins a 1024 filtres. És important destacar que la sortida del discriminador està constituïda per una funció Sigmoid, aquesta s'utilitza per comprimir el valor de la sortida entre 0 i 1.

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

0n 0 serà una classificació "falsa"de la imatge i un 1 indica la classificació "real".

3 Proposta d'Objectiu

L'objectiu principal d'aquest treball és desenvolupar un model funcional basat en l'arquitectura SRGAN, una metodologia de Generative Adversarial Network ja existent, y realitzar modificacions per obtenir millors resultats. Es modificarà la forma com s'entrena el generador, ja que una part s'entrenarà a partir dels 4 coeficients en la transformada de Wavelet [13] de les imatges, sent així una Wavelet-SRGAN. S'aprofundeix més en el mètode utilitzat en l'apartat 4 de Metodologia. L'objectiu és comparar els resultats obtinguts amb els del model SRGAN original per avaluar les diferències en la qualitat de les imatges generades, i veure si un entrenament d'una xarxa GAN es veu afectat positivament pels coeficients Wavelet. Un dels principals desafiaments dels models GAN és la intensitat computacional necessària per al seu entrenament. A més, es pretén que aquesta millora sigui específicament aplicable per a la millora de la qualitat d'imatges obtingudes per satèl·lits, una aplicació crítica en diverses àrees com la cartografia, la vigilància i la meteorologia. La prioritat dels objectius establerts ha estat dividida en objectius principals i objectius secundaris.

Objectius principals:

- Aconseguir detectar diferències significatives entre els dos mètodes, especialment en la nitidesa i el detall de les imatges superresoltes.
- Implementar canvis lògics i justificats en el model, assegurant una millora tangible en els aspectes de rendiment i eficiència.
- Avaluar exhaustivament el rendiment del model modificat que treballa en les transformades de Wavelet, en termes de qualitat d'imatge, temps de processament i recursos computacionals requerits, comparant-lo amb el model original.

Objectius secundaris:

- Documentar exhaustivament tot el procés de modificació i entrenament del model, així com els resultats obtinguts, per facilitar futures investigacions i aplicacions en el camp.

En resum, aquest treball busca no només crear i modificar un model GAN existent per entrenar el seu generador a partir d'imatges Wavelet, sinó també comparar els resultats de manera visual i numèrica amb el model original per determinar les diferències en la qualitat i l'eficiència computacional. A més de concloure conclusions a partir dels resultats obtinguts.

4 METODOLOGIA

La metodologia que es seguirà es basa en l'aplicació de tècniques de xarxes convolucionals, específicament s'utilitza l'arquitectura SRGAN prèviament desenvolupada en el camp de la superrresolució d'imatges. En aquest treball, es fa una modificació del model SRGAN per incorporar la descomposició de coeficients de la Transformada Wavelet [13] a les imatges d'entrada del generador. La Transformada Wavelet s'utilitza per descompossar una imatge en una sèrie de quatre coeficients Wavelet, que representen la informació de la imatge en diferents escales i orientacions d'ones.

- Coeficient d'Aproximació (Low-Low): Representa la part de baixa freqüència de la imatge.
- Coeficient de Detall Horitzontal (Low-High): Representa la informació de detall horitzontal de la imatge.
- Coeficient de Detall Vertical (High-Low): Representa la informació de detall vertical de la imatge.
- Coeficient de detall diagonal (High-High): Representa la informació de detall diagonal de la imatge.

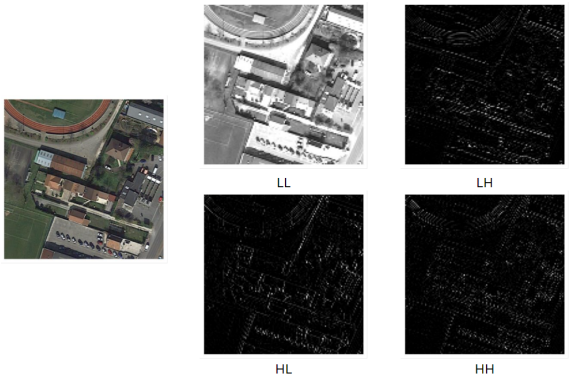

Fig. 8: Coeficientes Wavelet de la imatge P000-3 del dataser DOTA

Aquest procés s'implementa en les imatges de baixa resolució introduïdes al generador. Cada coeficient es processa de manera independent a través d'un camí de 20 WaveletBlocks. Això permet que el model aprofundeixi en els detalls i característiques particulars de cada component de la imatge en diferents escales i orientacions. Aquests WaveletBlocks estan formats per tres capes convolucionals de 32 filtres cadascuna amb una funció ReLU per cada capa. Una vegada processats aquests coeficients, es combinen amb la imatge original processada, mitjançant una ponderació de 0.2 per a cada coeficient.

$$out = n * p + LL * p + LH * p + HL * p + HH * p \quad (2)$$

On n es la imatge normal processada i out la imatge generada . S'assegura que la informació rellevant i característiques de cada component contribueixi adequadament a la imatge final.

A la figura (10) podem observar els 4 coeficients Wavelet una vegada han sigut processats pels WaveletBlocks del generador en contrast amb l'imatge original de baixa resolució, s'han extret les característiques i detalls de cadascun.

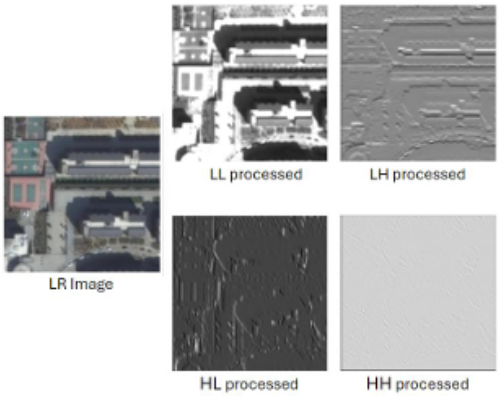

Fig. 10: Coeficients processats [Propia]

A la figura (11) podem observar un exemple de com afecta la incorporació de les característiques processades dels coeficients Wavelet a la sortida del Generador. A l'esquerra una imatge processada pel model original, a la dreta la mateixa imatge amb les característiques dels coeficients Wavelet incorporades, es pot observar que es dona una major importància a les característiques i detalls verticals.

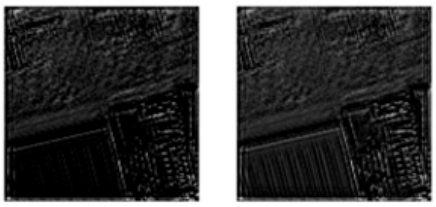

Fig. 11: Comparació de imatges procesades pel generador [Propia]

El model SRGAN original i el model modificat s'entrenen utilitzant un conjunt de dades estàndard en la comunitat de recerca, el dataset DOTA [15] amb més de 1000 imatges d'alta resolució, que formen el conjunt d'entrenament, de test i de validació. Un cop s'obtinguin resultats, aquests es compararan numèricament amb el model SRGAN original. La mètrica principal utilitzada serà el PSNR. Es quantifica la qualitat de la imatge superrresolta en termes de fidelitat a la imatge original, quantificant la diferència entre els valors dels píxels de les dues imatges.

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \quad (3)$$

Aquí, MAX és el valor màxim possible d'un píxel a la imatge (per exemple, 255 per a imatges de 8 bits). MSE es el Mean Squarred Error, calculat com:

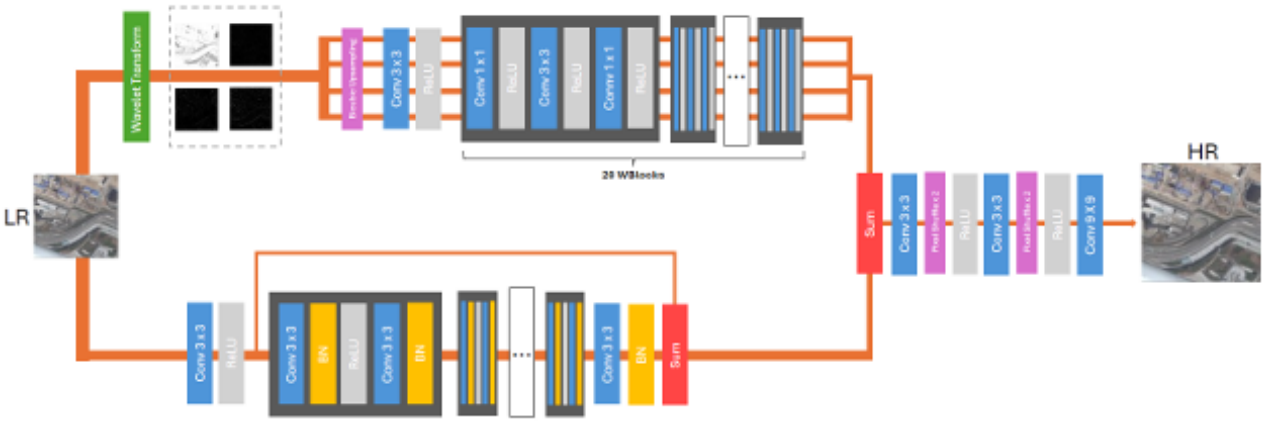

Fig. 9: Arquitectura Generador WaveletSRGAN [Propia]

$$\text{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j) - K(i,j)]^2 \qquad (4)$$

On *I(i,j)* i *K(i,j)* son els valors de la intensitat dels pixels en las posicions *(i,j)*, de la imatge original i la imatge generada respectivament.

També s'utilitza la mètrica SSIM (Índex de Similitud Estructural, Structural Similarity Index), que és utilitzada per comparar la similitud entre dues imatges, enfocant-se en la percepció visual de l'estructura i la qualitat. A diferència de mètriques com el PSNR, que mesuren diferències absolutes en els valors dels píxels, el SSIM considera canvis en la luminància, el contrast i l'estructura, cosa que s'alinea més estretament amb la percepció humana.

$$\text{SSIM}(I,K) = \frac{(2\mu_I\mu_K + C_1)(2\sigma_{IK} + C_2)}{(\mu_I^2 + \mu_K^2 + C_1)(\sigma_I^2 + \sigma_K^2 + C_2)} \qquad (5)$$

On μ_I i μ_K són les mitjanes dels valors dels píxels de les finestres locals (regió específica de la imatge seleccionada per ser analitzada o processada de manera independent de la resta) de I i K. σ_I^2 y σ_K^2 són les variàncies dels valors dels píxels en les finestres locals. σ_{IK} és la covariància dels valors dels píxels entre les finestres locals de I i K. C_1 i C_2 són constants per estabilitzar la divisió quan els denominadors són petits.

A més de la comparació numèrica, també s'analitzarà visualment la qualitat de les imatges generades per cada model. Aquest anàlisi qualitatiu permetrà avaluar aspectes com la nitidesa, la coherència dels detalls i la percepció dels colors, oferint una perspectiva més completa del rendiment del model en diferents escenaris. Finalment, es realitzaran proves de rendiment per avaluar l'eficiència computacional del model modificat, considerant factors com el temps d'entrenament i d'inferència. Aquesta informació complementària proporcionarà una comprensió més profunda dels avantatges i limitacions de la proposta.

5 Experimentació

En aquesta primera fase de l'experimentació, s'ha dut a terme el tractament de les dades del dataset DOTA perquè puguin ser processats pel model, així com el desenvolupament i entrenament del model SRGAN, tal com es descriu en el propi document, amb la finalitat d'observar les mètriques i resultats obtinguts del model original mitjançant l'entrenament d'imatges de satèl·lit. Aquest experiment s'ha realitzat en una màquina amb sistema operatiu Linux-5.4.0, equipada amb una GPU NVIDIA GeForce RTX 2080 Ti amb 12GB de memòria VRAM. Els programes utilitzats han estat Visual Studio Code com a entorn de desenvolupament, amb Python com a llenguatge principal de programació. Les biblioteques més rellevants empleades han estat NumPy, OpenCV i PyTorch, aquesta última sent una llibreria d'aprenentatge profund àmpliament utilitzada en la comunitat científica.

5.1 Tractament de dades

El Dataset DOTA és el conjunt d'imatges que s'utilitza per realitzar l'entrenament i el test de la xarxa SRGAN. El tractament d'aquestes imatges és crucial per introduir al model dades de qualitat, les quals ajuden a extreure millors resultats. Aquest està constituït per imatges de satèl·lit recopilades de Google Earth, dividit en un conjunt de Train amb 1411 imatges amb un total de 9.9 GB de memòria, un conjunt de Test de 928 imatges amb un total de 6.8 GB de memòria, i un conjunt de validació de 458 imatges amb 3.26GB de memòria, el que fa un total de 2800 imatges. Aquestes són les imatges d'alta resolució, i a partir d'aquestes es crearan les imatges de baixa resolució. En aquest tractament s'ha realitzat un filtre per GSD (Ground Sample Distance). Aquest filtre ens ajuda a distingir les altures a les quals s'han realitzat totes les imatges del dataset, per unificar un valor estàndard de GSD on només s'utilitzen les imatges dins d'un rang de 0,1 metres i 0,15 metres, i per tant, imatges que s'han realitzat en altures similars. Aquest valor escollit ens indica que cada píxel de les imatges representa de 10 a 15 centímetres reals. Com es veu a l'histograma següent, escollim el rang amb més quantitat d'imatges al dataset, l'eix y representa el nombre d'imatges, i l'eix x representa el valor de GSD en metres.

Dins d'aquest grup escollit, hi ha imatges amb les vores defectuoses o directament negres i imperfectes. Aquestes imatges passen per un filtre i són eliminades.

Seguidament aquestes imatges seleccionades tenen mides diferents, i per tant a partir d'aquestes es creen dife-

rents imatges amb resolució de 1024 x 1024 per crear un estàndard d'imatges d'alta resolució. Amb aquest conjunt d'imatges d'alta resolució de 1024 x 1024 ens ajuden a crear les imatges de baixa resolució les quals es poden obtenir amb un reescalat a 256 x 256 i una interpolació bicubica. Aquest tractament es realitza per als tres conjunts de dades, train, test i validació. Amb totes aquestes fases, el conjunt d'imatges final que s'introduirà al model per ser entrenat es de 1113 imatges d'entrenament (70 %), 294 imatges per la validació (15%), i 294 imatges de test (15%). El que fa un total de 1611 imatges entre els tres conjunts.

A la figura (12) Podem observar un exemple visual del retall de la imatge P0347 d'alta resolució del dataset DOTA que ha passat pel filtre GSD i el filtre d'imperfeccions. Aquesta imatge es iterada i retalla per obtenir el màxim conjunt d'imatges possibles de 1024 x 1024. En aquest cas s'obtenen tres imatges d'alta resolució.

Fig. 12: Retall d'imatges originals

5.2 Entrenament dels models

Per a dur a terme l'entrenament del model original, es va implementar l'arquitectura descrita a l'article de SRGAN. Aquest procés va començar amb la creació dels arxius .h5 a partir de les 1113 imatges d'entrenament finals i 294 imatges de validació seleccionades. Els arxius .h5 són arxius de model de Keras que contenen tant l'arquitectura com els pesos del model, i per tant permeten la reutilització i distribució de models de manera eficient. Dins del procés d'entrenament, es va aplicar la funció de pèrdua tal com es detallava a l'article. Aquesta funció esta constituïda per:

- Adversarial Loss: Utilitzada per a entrenar la xarxa generativa per a generar imatges que siguin indistingibles de les imatges d'alta resolució reals. On G: És el generador de la GAN, que pren z_i com a entrada i genera una imatge artificial. D: És el discriminador de la GAN, que pren com a entrada una imatg i avalua la probabilitat que aquesta imatge sigui una mostra real en lloc d'una generada per G. $D(G(z_i))$: Representa la sortida del discriminador, és la probabilitat que $G(z_i)$ (la imatge generada) sigui classificada com una mostra real per D.

$$L_{adv} = \frac{1}{N}\sum_{i=1}^{N}(1 - D(G(z_i))) \tag{6}$$

- Perception Loss: Incorpora una mesura de distància perceptiva per a millorar la qualitat visual de les imatges generades. On $\phi(I_{\text{real}_i})$ són les característiques extretes de la imatge generada per G i de la imatge real I_{real_i}, respectivament. $\|\cdot\|_2^2$ és la norma euclidiana al quadrat, que representa la distància euclidiana al quadrat entre dos vectors.

$$L_{\text{percep}} = \frac{1}{N}\sum_{i=1}^{N}\|(G(z_i)) - (I_{\text{real}_i})\|_2^2 \tag{7}$$

- Image Loss: S'encarrega de minimitzar la diferència entre les imatges generades i les imatges d'alta resolució originals.

$$L_{\text{img}} = \frac{1}{N}\sum_{i=1}^{N}\|G(z_i) - I_{\text{real}_i}\|_2^2 \tag{8}$$

La funció final es:

$$L_{\text{total}} = L_{\text{img}} + 0.001 \cdot L_{\text{adv}} + 0.006 \cdot L_{\text{percep}} \tag{9}$$

La pèrdua del discriminador és la següent:

$$L_{\text{D}} = \log D(I_{\text{real}}) + \log\left(1 - D(G(z_i))\right) \tag{10}$$

Un cop s'havia establert la funció de pèrdua, es van configurar els paràmetres del model d'acord amb les recomanacions de l'article i les necessitats del projecte. Per exemple, es va establir un factor de reescalat de 4 per a millorar la resolució de les imatges generades, ja que les imatges de baixa qualitat que tenim per entrenar tenen 4 vegades menys píxels que les de alta qualitat. Es va triar el optimitzador Adam per la seva eficàcia en la convergència de l'entrenament amb $\beta = 0.9$.

Una vegada avaluats els resultats obtinguts amb els models entrenats durant 20 èpoques, es va determinar que era necessari augmentar el nombre d'èpoques a 40 per a millorar el rendiment i maximitzar el procés d'entrenament. Aquest ajust va duplicar el temps necessari per a l'entrenament, estenent-lo a 8 hores. Es van dur a terme 5 iteracions d'entrenament amb el model original durant 40 èpoques cadascuna. Les dades resultants d'aquestes iteracions van ser comparats, i es va seleccionar el model més robust i amb millors mètriques PSNR i SSIM per a les imatges de prova del conjut de validació.

Posteriorment, es va procedir a implementar els canvis detallats en la secció 4, (Metodologia), al model original. Es van realitzar 5 iteracions d'entrenament amb aquest nou model, WaveletSRGAN, utilitzant exactament les mateixes dades d'entrenament. Aquest nou model té una càrrega computacional mes gran i arriba a trigar unes 14 hores en realitzar l'entrenament, un 30% més que el model original. De la mateixa manera que amb el model SRGAN original, es va seleccionar el model amb les millors estadístiques per al conjunt de validació DOTA. La figura (13) mostra la comparació de mètriques per a cada època d'entrenament d'aquest models seleccionats.

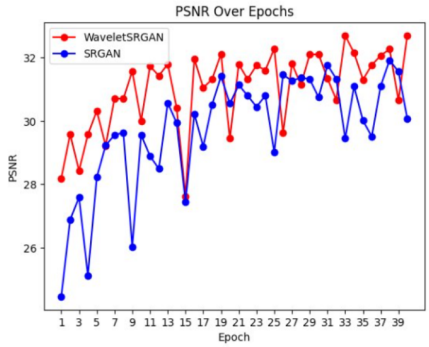

Fig. 13: PSNR del conjunt de validació durant l'entrenament de SRGAN i WSRGAN

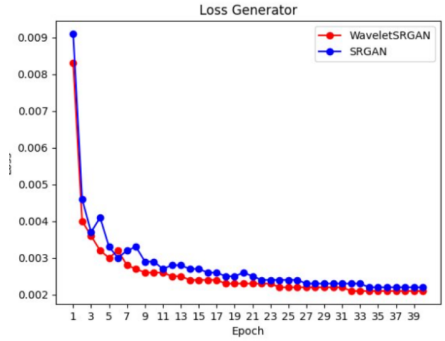

Fig. 14: Perdua durant l'entrenament de SRGAN i WSRGAN

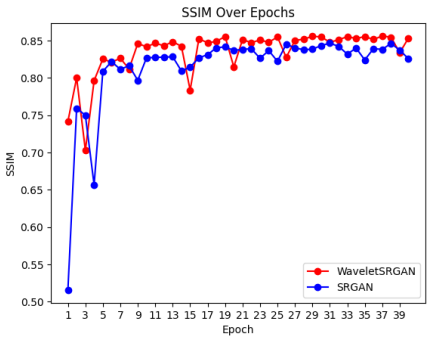

Fig. 15: SSIM del conjunt de validació durant l'entrenament de SRGAN i WSRGAN

A la figura (13), la corba vermella que representa a WaveletSRGAN, manté consistentment un PSNR més alt en comparació amb la corba blava de SRGAN, la qual cosa indica una millor similitud d'imatge amb la imatge d'alta resolució. Encara que tots dos models mostren fluctuacions, WaveletSRGAN mostra una tendència ascendent més consistent, aconseguint valors màxims de PSNR més alts. D'altra banda, a la figura (14) de la pèrdua del generador, totes dues corbes mostren una disminució ràpida en les primeres èpoques, indicant un aprenentatge inicial eficient. Després d'aquestes primeres èpoques, on SRGAN mostra algunes fluctuacions, la pèrdua s'estabilitza, però WaveletSRGAN tendeix a tenir valors lleugerament menors de pèrdua en comparació amb SRGAN, suggerint un rendiment superior. En conclusió, WaveletSRGAN a priori ofereix una millor similitud d'imatge amb la imatge d'alta qualitat original i mostra una menor pèrdua del generador en comparació amb SRGAN, la qual cosa ho fa més robust i efectiu per a tasques de superresolució d'imatges.

Model	Nombre de paràmetres
SRGAN	192,011
WSRGAN	1,104,399

TAULA 1: NOMBRE DE PARÀMETRES DE SRGAN I WSRGAN.

5.3 Resultats de l'entrenament

En aquesta secció, es presentaran i analitzaran els resultats obtinguts del procés d'entrenament del model SRGAN i WaveletSRGAN. Es realitzarà una comparació visual entre algunes imatges d'alta resolució del conjunt de proves i les imatges respectives generades pels models. A més, es mostraran les imatges d'entrada utilitzades durant el procés de generació, la qual cosa permetrà una comprensió més completa del rendiment del model i la seva capacitat per a produir imatges d'alta qualitat a partir d'imatges de baixa resolució.

Model	PSNR	SSIM
SRGAN	29.6238	0.8212
WSRGAN	30.6486	0.8427

TAULA 2: MÈTRIQUES MITJANES DE SRGAN I WSRGAN PER A LES 294 IMATGES DE TEST

Model	Temps
SRGAN	8 hores
WSRGAN	14 hores

TAULA 3: TEMPS D'ENTRENAMENT DE SRGAN I WSRGAN.

Les taules presentades detallen els resultats obtinguts de l'entrenament dels models SRGAN i WaveletSRGAN. La Taula 2 mostra que el model WaveletSRGAN supera al SRGAN en les mètriques PSNR i SSIM. Cal tenir en compte que el PSNR és una mesura logarítmica, la qual cosa significa que petites diferències absolutes en el PSNR poden reflectir diferències significatives en la percepció visual de la qualitat de la imatge. La Taula 3 reflecteix el temps d'entrenament, suggereix un major cost computacional per a WSRGAN a causa de la seva arquitectura més complexa.

Les imatges generades pels dos models presenten diferències substancials en la qualitat dels detalls i en les mètriques quantitatives utilitzades per avaluar el seu rendiment.

La capacitat de WSRGAN de descompondre la imatge en freqüències diferentes és crucial per millorar la reconstrucció en zones d'alta freqüència de la imatge, on hi ha transicions brusques i detalls fins, com ara línies i textures complexes. Gràcies als coeficients que consideren les característiques verticals, horitzontals i diagonals, WSRGAN

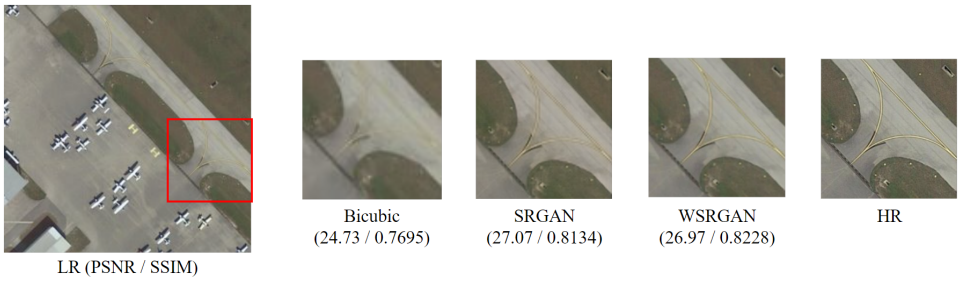

Fig. 16: Comparació resultats P0179-11.png

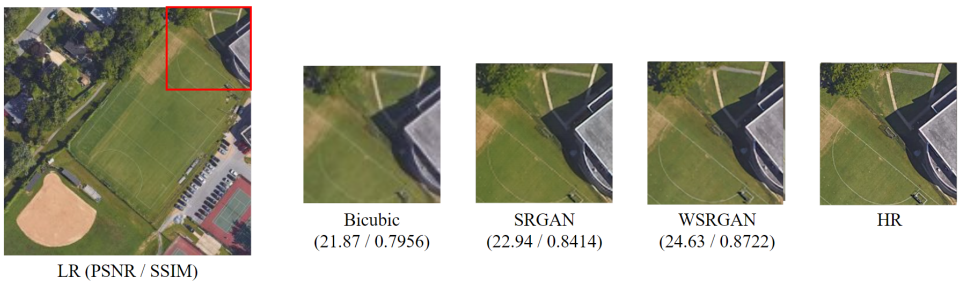

Fig. 17: Comparació resultats P0665-0.png

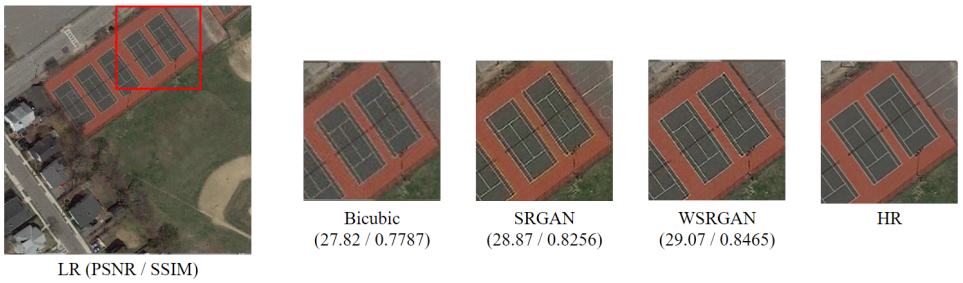

Fig. 18: Comparació resultats P0352-0.png

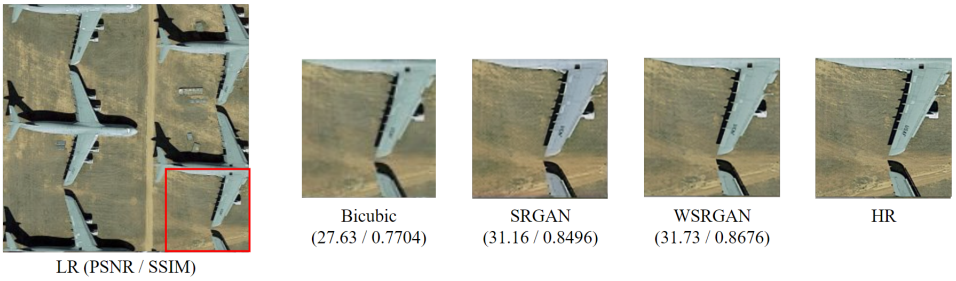

Fig. 19: Comparació resultats P1398-11.png

pot capturar millor aquestes estructures, resultat en una major nitidesa i precisió en aquestes àrees. Per exemple, en la primera imatge d'exemple, les línies de l'aeroport són més definides amb WSRGAN que amb SRGAN.

En contrast, en les zones de baixa freqüència de la imatge, on els canvis de color i textura són més suaus i menys abruptes, les diferències entre SRGAN i WSRGAN són molt menys notòries. Aquestes àrees no requereixen el mateix nivell de precisió en la reconstrucció de detalls fins, per la qual cosa tots dos models ofereixen un rendiment similar. Per exemple, a la figura (16), tant SRGAN com WSRGAN produeixen resultats pràcticament indistingibles per a l'ull humà en zones de baixa freqüència.

6 Conclusions

En aquest treball, hem implementat i avaluat un model de superresolució per a imatges de satèl·lit, anomenat Wavelet-SRGAN, que combina la xarxa generativa adversària SRGAN amb la transformada de Wavelet. L'objectiu ha estat millorar la qualitat de les imatges en comparació amb el model SRGAN estàndard, augmentant detall i nitidesa gràcies a l'ús de coeficients de Wavelet.

Durant el desenvolupament, es van adaptar els paràmetres del model SRGAN per incorporar la transformada de Wavelet en l'entrenament del generador. Això va permetre generar imatges amb una qualitat visual superior, demostrant avantatges significatius en la reconstrucció de detalls respecte al model original. La comparativa amb

SRGAN va mostrar millores en la percepció visual i en mètriques com PSNR i SSIM.

Els resultats subratllen el potencial de les transformades Wavelet en zones d'alta freqüència, millorant el rendiment per a aplicacions com cartografia, vigilància i meteorologia, on la qualitat de les imatges de satèl·lit és crucial.

Aquest projecte obre la porta a futures millores. Seria interessant explorar altres transformades o tècniques de preprocessament per millorar la reconstrucció de detalls. A més, l'optimització del procés d'entrenament, incloent l'ajust de hiperparàmetres i estratègies avançades de regularització, que podria reduir el temps d'entrenament i millorar el rendiment en escenaris més variats i complexos, o la utilització de dades sintètiques. També es podria considerar la integració del model amb sistemes de captura d'imatges en temps real per a aplicacions de monitorització contínua, adaptant-lo per treballar amb dades en streaming. La seva aplicació a altres tipus de dades visuals, com imatges de drons o sensors aeris, podria estendre l'ús de la superresolució més enllà de les imatges de satèl·lit tradicionals.

AGRAÏMENTS

En primer lloc, al meu tutor, el Professor Robert Benavente Vidal, per la seva constant orientació, paciència, i consells al llarg de tot el procés. La seva experiència i dedicació han estat essencials per a la consecució d'aquest treball.

També vull agrair a la Universitat Autonoma de Barcelona, que amb els seus recursos i suport acadèmic m'ha brindat l'oportunitat de desenvolupar les meves habilitats i coneixements en aquest camp. Els recursos bibliogràfics, les infraestructures i l'entorn de treball que ofereix han estat fonamentals per al desenvolupament d'aquest projecte.

REFERÈNCIES

[1] C. Christian Gamdek, "Confusion about the image resolution in digital photography," Litmind, accés febrer de 2024. [En línea]. Disponible: https://litmind.com/confusion_sobre_la_resolucion_de_la_imagen_en_fotografia_digital

[2] Matthew Giassa, "Image Processing - Nearest Neighbour Interpolation — GIASSA.NET," GIASSA.NET — Engineering, DIY, and Everything Else, accés febrer de 2024. [En línea]. Disponible: https://www.giassa.net/?page_id=207

[3] Igor Yefremov, "RESAMPLING," HobbyMaker, accés març de 2024. [En línea]. Disponible: https://hobbymaker.narod.ru/English/Articles/resampling_eng.htm

[4] Wikimedia Contributors, "Bicubic interpolation," Wikipedia, accés febrer de 2024. [En línea]. Disponible: https://en.wikipedia.org/wiki/Bicubic_interpolation

[5] J. Nett, "Bilinear Interpolation," Homepage — Portland State University, accés febrer de 2024. [En línea]. Disponible: https://web.pdx.edu/ jduh/courses/geog493f09/Students/W6_Bilinear%20Interpolation.pdf

[6] Liz Sanderson, "Modify Raster Cell Size by Resampling,"FME Support Center, accés abril de 2024. [En línea]. Disponible: https://support.safe.com/hc/en-us/articles/25407477353741-Modify-Raster-Cell-Size-by-Resampling

[7] "Convolutional Neural Networks for Visual Recognition," cs231n, accés febrer de 2024. [En línea]. Disponible: http://cs231n.github.io/convolutional-networks/

[8] "Tobacco Plant Detection in RGB Aerial Images," Research Gate, accés març de 2024. [En línea]. Disponible: https://www.researchgate.net/publication/339578955/figure/fig4/AS:863732028669952@1582941171319/Calculation-process-of-a-filter-in-the-convolutional-layer-The-first-matrix-is-the.ppm

[9] Chao Dong, Chen Change, Loy Kaiming He, and Xiaoou Tang, "Image Super-Resolution Using Deep Convolutional Networks," arXiv.org, accés el 21 de abril de 2024. [En línea]. Disponible: https://arxiv.org/abs/1501.00092

[10] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," arXiv.org, accés abril de 2024. [En línea]. Disponible: https://arxiv.org/abs/1807.02758

[11] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," arXiv.org, accés març de 2024. [En línea]. Disponible: https://arxiv.org/abs/1609.04802

[12] M. Del Pra, "Generative Adversarial Networks," Medium, accés març de 2024. [En línea]. Disponible: https://medium.com/@marcodelpra/generative-adversarial-networks-dba10e1b4424

[13] Youssef Abarca, Daniel Calle, Mathews Castillo, and José Guayllas, "Transformada Wavelet aplicada en imágenes,"accés març de 2024. Disponible: https://www.monografias.com/trabajos105/vtransformada-wavelet-aplicada-imagenes/vtransformada-wavelet-aplicada-imagenes

[14] kmlahan, "Super-Resolution of Satellite Images using Deep Learning," GitHub, accés abril de 2024. [En línea]. Disponible: https://github.com/kmalhan/SatelliteSR

[15] "DOTA," captainWHU, accés el 21 de abril de 2024. [En línea]. Disponible: https://captain-whu.github.io/DOTA/dataset.html

APÈNDIX

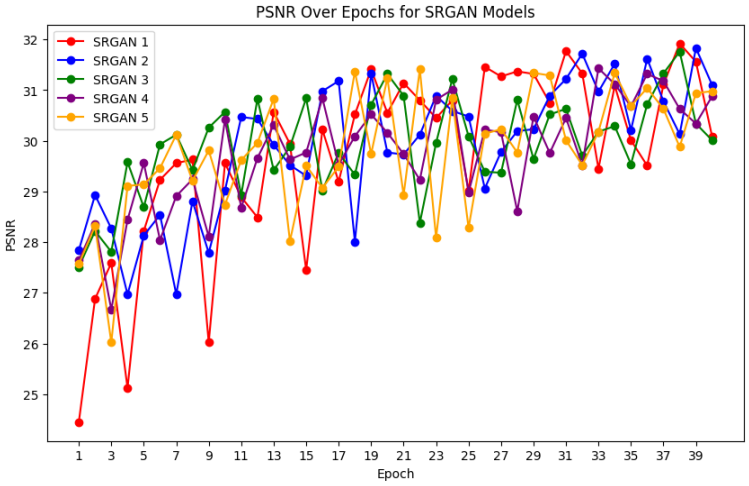

Fig. 20: PSNR de conjunt de validació pels 5 entrenaments realitzats amb SRGAN

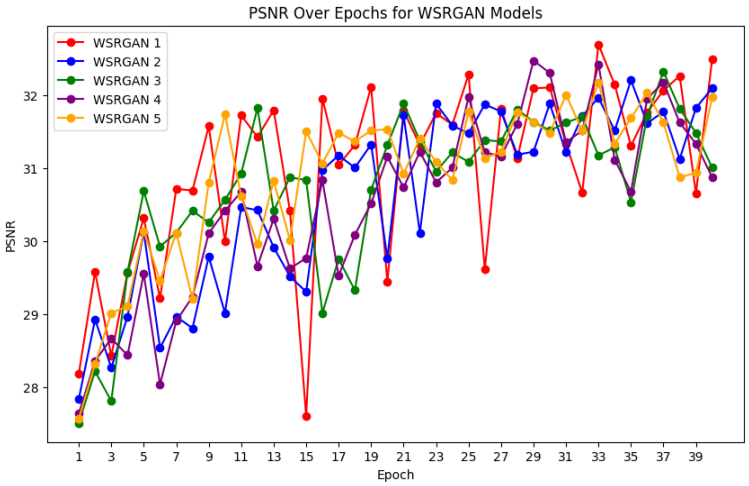

Fig. 21: PSNR de conjunt de validació pels 5 entrenaments realitzats amb WSRGAN