Aplicació de tècniques d'inpainting per a la reconstrucció d'imatges satèl·lit afectades per núvols

Marçal Muñoz Salat

Resum—Aquest projecte estudia l'aplicació de tècniques d'inpainting per eliminar núvols en imatges satel·litals, amb l'objectiu de restaurar de manera plausible informació perduda de l'escena. S'han avaluat diversos mètodes clàssics com TELEA, Navier-Stokes i FSR, i també arquitectures d'aprenentatge profund com U-Net i DeepFillv2, emprant el conjunt de dades CloudSen12. L'anàlisi inclou mètriques quantitatives com PSNR i SSIM, i també una avaluació qualitativa visual. Els resultats mostren que els mètodes basats en deep learning superen els clàssics, especialment en reconstruccions de grans zones ocultes. Per explorar el límit de rendiment dels models, s'han aplicat millores com l'ús de màscares separades i canals d'entrada addicionals, que han aportat una millora clara. Finalment, s'ha fet un darrer experiment amb augmentació de dades mitjançant rotacions aleatòries i inversions horitzontals i verticals, que ha aportat una millora addicional. En aquest context, també s'ha simulat el procés de restauració aplicant les màscares d'imatges ennuvolades sobre les seves versions sense núvols, permetent una avaluació més controlada.
Paraules clau—Inpainting, imatges satel·litals, visió per computador, núvols, aprenentatge profund, restauració d'imatges, CloudSen12.

Abstract— This project explores the application of inpainting techniques to remove clouds from satellite images, with the aim of plausibly restoring missing scene information. Several classical methods such as TELEA, Navier-Stokes and FSR have been evaluated, along with deep learning architectures like U-Net and DeepFillv2, using the CloudSen12 dataset. The analysis includes quantitative metrics such as PSNR and SSIM, as well as a qualitative visual evaluation. The results show that deep learning methods outperform classical ones, especially when reconstructing large occluded areas. To explore the performance limits of the models, improvements such as the use of separate masks and additional input channels were applied, leading to noticeable gains. Finally, a last experiment with data augmentation—using random rotations and horizontal and vertical flips—brought further improvements. In this context, the restoration process was also simulated by applying the cloud masks to the corresponding cloud-free images, enabling a more controlled evaluation.
Index Terms—Inpainting, satellite imagery, computer vision, clouds, deep learning, image restoration, CloudSen12.

1 Introducció

L'inpainting és una tècnica de processament d'imatges que permet reconstruir o completar zones d'una imatge que han estat danyades, ocultes o eliminades. Aquest procés té aplicacions en nombrosos àmbits, com la restauració de fotografies antigues [1], la reconstrucció d'imatges mèdiques [2] o la millora d'imatges comercials [3]. En el camp de la teledetecció —disciplina encarregada d'adquirir informació de la superfície terrestre mitjançant sensors remots— i la cartografia —disciplina que estudia la representació gràfica del territori i els fenòmens geoespacials—, l'inpainting té un paper fonamental en la restauració d'imatges satèl·lit i ortofotos [4], permetent tractar artefactes no desitjats causats per núvols, ombres o discontinuïtats generades durant la captació d'imatges.

La presència d'oclusions en imatges satel·litals és un problema recurrent que dificulta l'extracció d'informació acurada del terreny. Segons Karlsson et al. (2023) [5], que van analitzar 42 anys de dades AVHRR, la cobertura global de núvols es manté al voltant del 65 %, fet que indica que una gran part de la superfície terrestre està sovint oculta. Aquesta elevada presència compromet nombrosos estudis: en l'agricultura de precisió, per exemple, pot dificultar la detecció de cultius malmesos [6]; en la gestió de riscos naturals, interfereix en la identificació de zones inundades [7] o afectades per incendis; i en l'estudi del canvi climàtic, limita el seguiment continu del territori [8]. Tanmateix, en aquests contextos una reconstrucció errònia pot generar informació artificial i portar a conclusions contraproduents.

Aquest treball estudia l'aplicació de tècniques d'inpainting com a eina per mitigar problemàtiques més acotades i realistes derivades de la presència de núvols. Per exemple, un núvol que oculta parcialment un vehicle o una infraestructura pot impedir que un detector

- *E-mail de contacte: mmunozsalat@gmail.com*
- *Menció realitzada: Enginyeria de Computació*
- *Treball tutoritzat per: Daniel Ponsa Mussarra (Ciències de la Computació i Intel·ligència Artificial)*
- *Curs 2024/25*

d'objectes els identifiqui correctament. Si mitjançant inpainting s'aconsegueix una reconstrucció raonable de les àrees ocultes, es pot facilitar la detecció automatitzada d'aquests elements i millorar l'anàlisi posterior de les imatges.

2 Objectius

Aquest projecte té com a objectiu principal estudiar, implementar i avaluar diversos models d'inpainting per eliminar núvols i la projecció de les seves ombres en imatges satel·litals. Per assolir-ho, es defineixen els següents objectius específics seguint una seqüència ordenada i estructurada:

- Revisar l'estat de l'art. Analitzar les diferents tècniques d'inpainting existents, tant generals com específiques per a imatges satel·litals, així com identificar conjunts de dades públics adequats i seleccionar les mètriques més rellevants per avaluar la qualitat de les reconstruccions.
- Seleccionar i preparar un conjunt de dades d'imatges satel·litals amb i sense núvols. Aquest conjunt ha d'incloure també les màscares d'oclusions. L'ús d'un *dataset* adequat és imprescindible tant per entrenar els models com per mesurar el seu rendiment de manera objectiva.
- Aplicar models d'inpainting clàssics, no basats en l'aprenentatge computacional i avaluar-ne el rendiment per establir una referència on comparar propostes de l'estat de l'art.
- Implementar i entrenar models d'aprenentatge profund de l'estat de l'art per resoldre la problemàtica generada pels núvols en les imatges satel·litals, utilitzant el *dataset* preparat prèviament amb aquest propòsit, i avaluar-ne el rendiment mitjançant mètriques adequades per mesurar la qualitat de la reconstrucció, comparant els resultats amb els obtinguts pels models clàssics.

3 Metodologia

Per assolir els objectius del projecte, s'ha seguit una metodologia estructurada basada en un procés iteratiu que inclou l'estudi, la implementació i l'avaluació de diversos models d'inpainting. El primer pas ha consistit en preparar un conjunt d'imatges per entrenar i avaluar els models. Aquest conjunt s'ha dividit en tres parts: entrenament (train), validació (validation) i prova (test).

Per dur a terme l'experimentació, cada mètode considerat en aquest treball, ja sigui clàssic o basat en aprenentatge profund, s'ha estudiat seguint un procés estructurat en els següents passos:

1. Anàlisi del model: Estudiar el seu funcionament, requeriments i possibles avantatges i inconvenients respecte als anteriors.
2. Implementació del model en Python.
3. Entrenament i optimització (només per a mètodes basats en aprenentatge computacional): Ajust dels hiperparàmetres i millora del rendiment.
4. Avaluació: Anàlisi de la qualitat de la reconstrucció mitjançant mètriques quantitatives i qualitatives.
5. Anàlisi de resultats: Comparació amb els resultats obtinguts per altres mètodes i identificació de punts de millora.

Aquest enfocament iteratiu ha permès obtenir resultats parcials des de les primeres fases del projecte, fet que facilita la detecció primerenca de problemes i permet ajustar la metodologia de manera progressiva.

4 Planificació

Per assolir els objectius del projecte dins el calendari establert, s'ha planificat una seqüència de tasques estructurades en un diagrama de Gantt que es pot consultar a l'apèndix A1.

Fases del projecte

1. Revisar l'estat de l'art (del 17/02/2025 al 7/03/2025)
    - Cerca i lectura de publicacions científiques, informes i projectes relacionats.
    - Anàlisi de tècniques d'inpainting existents i seleccionar les més adequades per a imatges satel·litals.
    - Cerca i anàlisi de *datasets* d'imatges satel·litals.
2. Seleccionar i preparar un *dataset* (del 3/03/2025 al 14/03/2025)
    - Escullir i preparar el conjunt d'imatges per als experiments.
3. Estudiar, aplicar i avaluar models d'inpainting clàssics, no basats en l'aprenentatge computacional (del 17/03/2025 al 28/03/2025)
    - Aplicar els mètodes seleccionats sobre la partició de test.
4. Implementar, entrenar i avaluar de manera iterativa models basats en l'aprenentatge profund (del 31/03/2025 al 23/05/2025)
5. Redactar l'informe final (del 26/05/2025 al 06/06/2025)
6. Preparar els materials finals (del 09/06/2025 al 20/06/2025)
    - Fer la presentación.
    - Ordenar i revisar el dossier.
    - Fer el pòster.
7. Assajar la defensa del TFG (del 23/06/2025 al 03/07/2025)

5 Estat de l'art

La revisió de la primera part de l'estat de l'art s'ha estructurat en dues fases. Inicialment, s'han analitzat tècniques d'inpainting clàssiques, no basades en l'aprenentatge computacional, amb l'objectiu de comprendre els fonaments teòrics i pràctics d'aquest camp.

A continuació, s'ha estudiat l'aplicació de tècniques

modernes basades en l'aprenentatge profund, que actualment s'han consolidat com l'enfocament dominant en aplicacions d'inpainting complexes.

5.1 Mètodes clàssics

Un treball pioner en aquest àmbit és el de Bertalmío et al. [9], que va introduir l'ús d'equacions diferencials parcials (EDPs), com el model de Navier-Stokes, per simular la propagació de la informació des de les vores cap a l'interior de les regions deteriorades. Aquest enfocament es basa en la continuació de línies isofòtiques —que uneixen punts d'igual intensitat— i resulta útil per a oclusions petites i textures simples. En paral·lel, Telea [10] va proposar un mètode més eficient basat en el Fast Marching Method, que propaga informació des del contorn utilitzant combinacions ponderades dels píxels veïns per preservar la coherència visual, sent especialment útil en formes estretes i contínues.

Ambdues tècniques estan implementades a OpenCV [11] i són les dues estratègies clàssiques més comunes. Recentment, s'hi han afegit nous mètodes al mòdul xphoto, com FSR_FAST i FSR_BEST (Fast Structure Reconstruction) [12], dissenyats per preservar millor l'estructura geomètrica dins les regions a completar. FSR_FAST prioritza la velocitat d'execució, mentre que FSR_BEST ofereix resultats més precisos a canvi d'un cost computacional més alt.

5.2 Mètodes basats en l'aprenentatge profund

Els mètodes d'inpainting basats en l'aprenentatge profund han experimentat un gran desenvolupament els darrers anys, oferint solucions capaces de reconstruir regions d'extensió i complexitat molt superiors a les tècniques clàssiques. Aquestes aproximacions generalment utilitzen xarxes neuronals convolucionals (CNN) o models generatius que aprenen a inferir contingut perdut a partir del context visual.

Una de les arquitectures més aplicada és la U-Net, inicialment desenvolupada per a segmentació semàntica d'imatges biomèdiques [13]. Gràcies als seus camins d'atenció ("skip connections") entre les capes d'encoding i decoding, permet conservar tant la informació global com els detalls espacials, fet que facilita la reconstrucció de regions ocultes amb coherència estructural i visual.

Altres models populars en inpainting són els basats en xarxes generatives adversarials (GAN), com els Context Encoders de Pathak et al. (2016) [14] que combinen funcions de pèrdua (loss) de reconstrucció i adversarials per generar resultats realistes i detallats. Aquest enfocament s'ha ampliat amb architectures com Globally and Locally Consistent Image Completion (GLCIC) Iizuka et al., 2017 [15], que afegeix discriminadors globals i locals per millorar la coherència a múltiples escales.

En l'àmbit més recent, models com DeepFillv2 [16] incorporen convolucions amb mecanismes de comportes (gating mechanism) i mecanismes d'atenció per gestionar oclusions irregulars i grans, permetent reconstruccions més naturals i detallades. Aquest tipus de models sovint es preentrenen en grans conjunts d'imatges i després es poden ajustar a casos específics per millorar la seva capacitat d'adaptació.

Altres aproximacions destacades inclouen arquitectures com LaMa (Suvorov et al., 2021) [17], que utilitza convolucions de Fourier ràpides i una pèrdua perceptual especialitzada per optimitzar el rendiment en màscares extenses i alta resolució, així com models que incorporen mecanismes d'atenció o xarxes de tipus transformador per capturar dependències globals en la imatge.

Aquestes tècniques basades en aprenentatge profund han establert l'estàndard en aplicacions d'inpainting complexes, especialment quan es requereix una reconstrucció realista en presència d'oclusions grans i amb estructura heterogènia, com és el cas de les imatges satel·litals afectades per núvols i ombres.

5.3 Conjunts de dades

En qualsevol sistema basat en aprenentatge profund, la qualitat i adequació del conjunt de dades és un factor determinant per a l'eficàcia del model. En el cas concret de l'inpainting, els *datasets* han de contenir imatges amb àrees degradades o ocloses, juntament amb les corresponents imatges originals, de manera que el model pugui aprendre a reconstruir les regions absents amb coherència visual i estructural.

En el camp de l'inpainting general —no específicament aplicat a imatges satel·litals— s'utilitzen conjunts com ImageNet [18], Places2 [19] o CelebA-HQ [20], formats per escenes naturals, objectes o rostres humans. Aquests *datasets* no contenen oclusions reals, i per tant s'hi apliquen màscares artificials per simular àrees tapades. Aquest enfocament, habitual en models com GLCIC [15], DeepFill [16] o LaMa [17], és útil per a tasques d'inpainting generatiu generalista.

Tanmateix, en el cas d'imatges satel·litals multiespectrals com les de Sentinel-2, es requereixen escenes reals afectades per núvols i ombres, així com versions netes equivalents, per tal d'entrenar models de manera supervisada. En aquest context, CloudSen12 [21] és un dels conjunts més complets. Està format per sèries d'imatges Sentinel-2 del mateix lloc en diferents dates, amb una imatge clara (0 % núvols) i quatre ennuvolades amb cobertures aproximada (12,5 %, 35 %, 55 % i 82,5 %). Tot i no estar dissenyat específicament per a inpainting, inclou màscares automàtiques generades per vuit mètodes i una col·lecció de màscares de referència anotades manualment. Això permet construir tríades imatge ennuvolada, màscara i referència clara.

També s'ha considerat All-Clear [22], amb imatges netes a les quals s'afegeixen núvols de manera artificial, cosa que en permet un control precís de l'oclusió. Les màscares es deriven de fonts com el Sentinel-2 Cloud

Probability i tècniques d'ombres de Google Earth Engine. Tot i això, la seva naturalesa sintètica i la manca de variabilitat temporal poden reduir-ne la capacitat de generalització.

Finalment, conjunts com 38-Cloud [23] i 95-Cloud [24], tot i incloure màscares manuals sobre imatges de Landsat 8, tenen un nombre limitat d'escenes i no proporcionen versions netes corresponents, la qual cosa dificulta el seu ús en entrenament supervisat.

5.4 Mètriques per a l'avaluació de resultats

L'avaluació objectiva dels resultats d'un model d'inpainting és essencial per validar-ne la qualitat i comparar-lo amb altres enfocaments. En aquest context, s'utilitzen habitualment mètriques quantitatives de similitud que comparen la sortida del model amb la imatge original no degradada. Aquestes mètriques han estat àmpliament adoptades tant en l'anàlisi de mètodes clàssics com en l'avaluació de models basats en aprenentatge profund, i es mantenen pràcticament invariables en la majoria de treballs recents.

No existeix una única mètrica capaç de reflectir amb exactitud tots els aspectes de qualitat perceptiva i estructural; sovint es combinen diversos indicadors per obtenir una valoració més completa. Entre les més utilitzades, destaquen el PSNR (Peak Signal-to-Noise Ratio), que mesura la relació entre la intensitat màxima d'una imatge i el soroll introduït en la reconstrucció. És una mètrica simple però efectiva per quantificar la distorsió global, tot i que no sempre reflecteix fidelment la qualitat perceptiva de la imatge reconstruïda. Per la seva banda, el SSIM (Structural Similarity Index) avalua la similitud estructural entre dues imatges, tenint en compte aspectes com la luminància, el contrast i l'estructura local. Aquesta mètrica proporciona una aproximació més alineada amb la percepció humana i és àmpliament emprada tant en estudis clàssics com en propostes recents.

La combinación d'aquestes mètriques permet obtenir una visió objectiva i multifactorial de la qualitat de reconstrucció, especialment rellevant en entorns complexos com l'inpainting satel·lital.

6 DESENVOLUPAMENT

6.1 Adaptació del conjunt de dades Cloudsen12

Entre els diversos *datasets* analitzats a l'Estat de l'Art, s'ha escollit el conjunt de dades CloudSen12 [21] com a base per al desenvolupament del projecte a causa de la seva estructura especialment adaptable per a tasques d'inpainting. En particular, aquest conjunt permet construir triades d'imatges corresponents a una mateixa ubicació i escenari temporal, formades per una imatge sense núvols, una altra amb núvols i la seva màscara corresponent a les zones ocloses. A partir d'aquest punt, ens referirem a aquestes imatges com a *clear* (imatge sense núvols), *cloudy* (imatge amb núvols) i *mask* (màscara d'oclusió), i al conjunt d'aquestes tres com una escena.

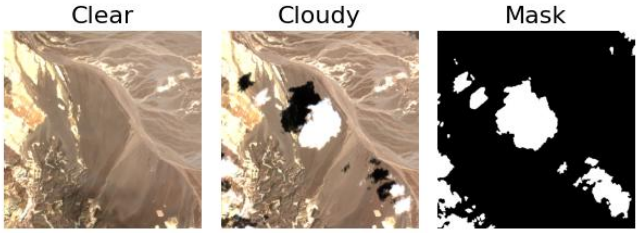

Figura 1. Exemple d'una escena, formada per *Clear*, *Cloudy* i *Mask*.

Aquesta organització facilita la formulació supervisada del problema d'inpainting, en què l'objectiu és reconstruir una imatge el més propera possible a la versió *clear*, utilitzant com a entrada la imatge *cloudy* i la seva *mask* corresponent.

Per adaptar CloudSen12 a les necessitats específiques del projecte, s'ha creat el subconjunt InpaintSen12, format exclusivament per escenes que compleixen un criteri de proximitat temporal entre les imatges *clear* i *cloudy*. S'ha fixat un llindar màxim de 30 dies de diferència entre captures, amb l'objectiu de minimitzar variacions reals en el terreny (com canvis estacionals, agrícoles o urbanístics) que podrien introduir soroll durant l'entrenament i l'avaluació. Aquest filtratge ha donat lloc a un total de 350 escenes vàlides per als experiments.

Pel que fa a les màscares, CloudSen12 proporciona tant versions generades automàticament com anotacions manuals. Després d'una anàlisi comparativa, s'ha optat per utilitzar les anotacions manuals, ja que presenten una delimitació més precisa de les àrees afectades. Aquestes màscares classifiquen cada píxel en quatre categories: àrees sense núvols, núvols translúcids, núvols densos i ombres. Per simplificar el problema i adaptar-lo als models d'inpainting, s'han binaritzat: es considera àrea a restaurar (valor 1) qualsevol píxel amb núvols densos o ombres, i àrea no afectada (valor 0) en la resta.

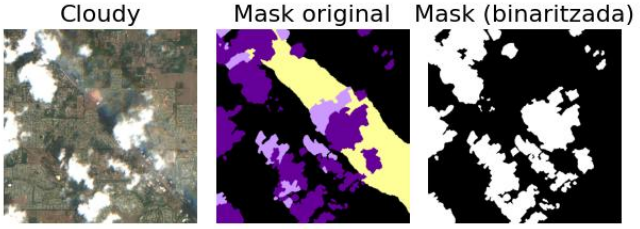

Figura 2. Exemple d'una *Mask* binaritzada.

Per garantir una avaluació rigorosa i representativa del rendiment dels models, s'ha dividit el conjunt InpaintSen12 en tres subconjunts: entrenament (252 escenes), validació (44 escenes) i prova (54 escenes). Aquesta partició s'ha realitzat manualment per assegurar una distribució equilibrada de contextos geogràfics i estructurals — incloent entorns urbans, zones agrícoles, vegetació natural i regions muntanyoses — a cadascun dels subconjunts. L'objectiu ha estat fomentar la capacitat de generalització dels models davant una gran varietat de situacions reals.

Les imatges es presenten en format TIFF de 17 canals,

incloent 12 bandes espectrals i diversos canals derivats per ampliar la informació disponible. No obstant això, en aquest projecte s'ha treballat exclusivament amb els canals RGB (vermell, verd i blau), per la seva compatibilitat amb arquitectures d'inpainting, la seva capacitat de visualització qualitativa i la reducció de càrrega computacional. Aquesta conversió es realitza dinàmicament durant la càrrega de dades, sense duplicar arxius en disc, preservant així l'estructura original i optimitzant l'ús d'espai.

6.2 Experiment amb mètodes clàssics

Amb el conjunt de dades InpaintSen12 ja preparat i estructurat, s'ha realitzat un primer experiment de restauració d'imatges emprant mètodes clàssics d'inpainting per establir una línea base de rendiment. S'han utilitzat algorismes clàssics disponibles a la llibreria OpenCV, concretament: Navier-Stokes (NS), TELEA, FSR_FAST i FSR_BEST.

Aquests algorismes s'han avaluat exclusivament en la partició de prova del conjunt de dades, donat que metodologicament, els mètodes basats en aprenentatge profund només es poden testejar en aquesta partició per evitar generar resultats esbiaixats.

Les imatges TIFF multiespectrals s'han convertit dinàmicament a format RGB seleccionant les bandes 4, 3 i 2 del satèl·lit Sentinel-2, corresponents als canals vermell, verd i blau respectivament. Aquest pas s'ha dut a terme utilitzant la llibreria rasterio, i s'ha completat amb una normalització a 8 bits per garantir la compatibilitat amb les funcions d'inpainting d'OpenCV. En el cas de les tècniques FSR, les imatges s'han reduït a un quart de la seva resolució original per tal d'adequar-les al format esperat pels algorismes, aplicant posteriorment una restauració a la mida original un cop completat el procés. Tots els resultats visuals s'han emmagatzemat en disc en format PNG per a la seva posterior avaluació quantitativa i qualitativa.

6.3 Implementació i entrenament del model U-Net

S'ha implementat la xarxa U-Net a partir del disseny original de Ronneberger et al. [9]. Tot i estar pensada inicialment per a segmentació semàntica, les seves característiques estructurals —com les connexions skip, que permeten preservar informació espacial, i la seva arquitectura simètrica encoder-decoder— la fan especialment adequada per a tasques d'inpainting.

Malgrat ser una arquitectura coneguda, el codi s'ha desenvolupat des de zero amb PyTorch Lightning, prenent com a guia un recurs audiovisual didàctic [25] per comprendre millor la seva lògica i garantir control sobre tot el procés. L'estructura segueix l'esquema clàssic en forma d'"U": un camí contractiu (encoder) amb dues convolucions 3×3 i activació ReLU, seguides d'un max pooling 2×2 per etapa; i un camí expansiu (decoder) amb up-convs 2×2 i connexions skip amb les sortides simètriques del camí descendent. Després de concatenar-les, s'apliquen dues convolucions addicionals per refinar la reconstrucció. La Figura 3 mostra l'esquema general de la U-Net original, tal com va ser proposada per a tasques de segmentació semàntica:

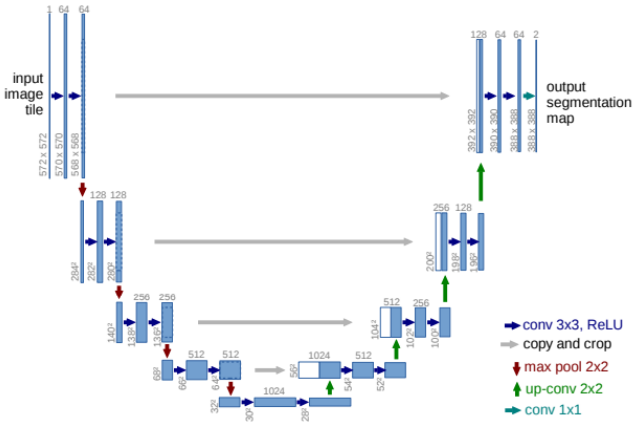

Figura 3. Arquitectura general de la U-Net. Font: [13]

S'han introduït ajustos per adaptar l'arquitectura a la tasca d'inpainting. En lloc de treballar amb imatges d'un sol canal, el model utilitza imatges RGB (3 canals) tant a l'entrada com a la sortida. També s'ha fixat una resolució de 256×256 píxels i s'han emprat 128 filtres base a la primera capa, en lloc dels 64 originals, per augmentar la capacitat de representació davant la complexitat de les escenes. Aquests canvis mantenen un bon equilibri entre expressivitat i eficiència computacional.

Per gestionar les dades, s'ha creat una classe pròpia (SatelliteInpaintingDataset) per carregar i preparar les imatges. Durant l'entrenament s'utilitzen exclusivament les particions d'entrenament i validació. El conjunt de prova es manté separat i s'utilitza únicament per a l'avaluació final. La validació serveix per prevenir el sobreajustament i identificar la millor configuració del model.

L'entrenament s'ha dut a terme amb PyTorch Lightning, usant l'optimitzador AdamW amb taxa d'aprenentatge inicial de 1e-4. S'ha aplicat un planificador ReduceLROnPlateau per reduir la taxa quan la millora s'estanca, i early stopping per evitar entrenaments innecessàriament llargs. Per monitorar i comparar experiments, s'ha integrat Weights & Biases (W&B), que permet enregistrar mètriques, visualitzar l'evolució de la pèrdua i gestionar múltiples configuracions de manera centralitzada.

Finalment, s'ha utilitzat una versió personalitzada de Masked L1 Loss, encapsulada dins la classe FlexibleCombinedLoss, que permet combinar diferents funcions de pèrdua amb pesos ajustables. Aquesta flexibilitat ha estat útil per adaptar el comportament del model segons la fase de desenvolupament i afinar la resposta a les característiques pròpies de l'inpainting.

6.4 Utilització i adaptació del model DeepFill v2

En aquest treball, també s'ha explorat l'ús d'una arqui-

tectura d'última generació desenvolupada específicament per a tasques d'inpainting: DeepFillv2. Aquesta proposta, disponible públicament mitjançant el repositori deepfillv2-pytorch [19], ha estat implementada per tercers i adaptada en aquest treball per aplicar-la al conjunt de dades InpaintSen12.

DeepFillv2 es basa en una arquitectura Generative Adversarial Network (GAN) dissenyada per a inpainting contextual. Els GANs consten de dues xarxes neuronals que s'entrenen de manera competitiva: un generador, encarregat de produir imatges completades a partir d'entrades degradades (amb màscares), i un discriminador, que intenta distingir entre imatges reals i les generades. Aquesta dinàmica de confrontació permet obtenir resultats realistes i coherents, especialment en les regions reconstruïdes.

En aquest treball s'ha utilitzat el model preentrenat amb el pes states_pt_places2.pth, que ha estat entrenat originalment sobre el *dataset* Places2, un conjunt massiu de prop de 2,5 milions d'imatges que cobreix una àmplia varietat d'escenes naturals i urbanes. Aquesta base d'entrenament proporciona al model una comprensió generalitzada de la composició d'escenes visuals, tot i que amb una semàntica diferent a la de les imatges satel·litals. A partir d'aquesta versió preentrenada, s'ha aplicat un procés de fine-tuning utilitzant els subconjunts d'entrenament i validació d'InpaintSen12. Aquest ajust fi permet adaptar les capacitats del model a la naturalesa específica del problema, mantenint l'avantatge d'un punt de partida robust i reduint el cost computacional associat a un entrenament des de zero.

7 RESULTATS

Per tal d'avaluar objectivament la qualitat de les imatges restaurades, s'han utilitzat dues de les mètriques més comunes en la literatura d'inpainting: PSNR (Peak Signal-to-Noise Ratio) i SSIM (Structural Similarity Index). Ambdues són mesures quantitatives que comparen la imatge restaurada amb la seva versió original no degradada, tot permetent quantificar el grau de similitud.

El PSNR mesura la distorsió global entre dues imatges en termes de soroll introduït. És una mètrica senzilla i àmpliament utilitzada per la seva interpretabilitat, expressada en decibels (dB), on valors més elevats indiquen una reconstrucció més precisa. No obstant això, pot no reflectir de manera fidel la qualitat perceptiva quan hi ha variacions locals o estructurals. Per això, s'ha complementat amb SSIM, que avalua la similitud estructural, tenint en compte paràmetres com la luminància, el contrast i l'estructura. Aquesta mètrica oscil·la entre 0 i 1, on valors propers a 1 indiquen una gran similitud visual i estructural entre la imatge generada i l'original.

Amb l'objectiu d'obtenir una anàlisi més acurada, les mètriques s'han calculat seguint dos enfocaments complementaris:

- Avaluació global: sobre la totalitat de la imatge.
- Avaluació focalitzada: sobre la regió a restaurar.

Aquesta doble avaluació permet identificar models que poden donar valors globals alts per mantenir intactes les zones no afectades, però que poden no oferir una restauració satisfactòria a la zona d'interès.

	Avaluació global		**Avaluació focalitzada**	
	PSNR (dB)	**SSIM**	**PSNR (dB)**	**SSIM**
TELEA	18.02	**0.6114**	16.01	0.6706
NS	18.37	0.6107	16.48	0.6762
FSR_FAST	18.12	0.5640	16.34	**0.6926**
FSR_BEST	17.76	0.5562	15.74	0.6761
U-NET	**19.59**	0.5624	18.13	0.6913
DEEPFILL	19.46	0.5833	**18.59**	0.6799

Taula 1. Resultats quantitatius dels diferents mètodes d'inpainting.

En termes de PSNR global (Taula 1), els mètodes basats en aprenentatge profund superen clarament els clàssics. U-Net assoleix un valor de 19.59 dB i DeepFill de 19.46 dB, per sobre dels valors que obtenen els mètodes clàssics, tots per sota dels 18.4 dB. Aquest resultat indica una menor distorsió global respecte a la imatge *clear* i reflecteix la capacitat d'aquests models de generar reconstruccions més precises, fins i tot en condicions de gran oclusió. Aquesta tendència es manté en l'avaluació focalitzada sobre la màscara, on tant U-Net com DeepFill tornen a destacar (18.13 dB i 18.59 dB respectivament), confirmant la seva superioritat també dins la regió reconstruïda.

Pel que fa al SSIM global, els millors resultats els obtenen els mètodes clàssics TELEA (0.6114) i Navier-Stokes (0.6107). Aquesta aparent superioritat, però, s'ha de matisar: aquests mètodes no modifiquen píxels fora de la màscara, fet que infla artificialment la similitud estructural global. En canvi, models com U-Net, DeepFill o fins i tot FSR introdueixen lleus modificacions fora de la màscara per millorar la coherència visual general, i això penalitza la seva puntuació SSIM. Per analitzar si aquestes modificacions afecten realment les mètriques, s'ha plantejat una nova avaluació global combinada. En aquest experiment, es genera una imatge on s'utilitza la reconstrucció dels models dins la màscara, mentre que la resta de píxels provenen de la imatge d'entrada *cloudy*. Això permet observar si modificar zones fora de la màscara influeix negativament en les mètriques quantitatives, especialment el SSIM.

	Avaluació global combinada	
	PSNR (dB)	**SSIM**
FSR_FAST	18.31	0.6150
FSR_BEST	17.99	0.6062
U-NET	19.07	0.5878
DEEPFILL	19.41	0.5820

Taula 2. Resultats quantitatius dels diferents mètodes d'inpainting combinant la imatge reconstruida per la regió de la màscara i la imatge *Cloudy* per la resta de la imatge.

El comportament del PSNR en aquesta avaluació combinada és desigual: els mètodes FSR mostren una lleu millora, mentre que U-Net i DeepFill presenten una petita reducció. Aquesta variació podria deure's al fet que, en modificar correctament regions fora de la màscara, aquestes, en ser substituïdes pels píxels de la imatge d'entrada, poden afectar la puntuació. Tot i això, les diferències de PSNR són mínimes. Pel que fa al SSIM, els resultats també varien segons el mètode: U-Net mostra una millora lleu, els mètodes FSR obetenen millores més significatives, mentre que DeepFill pateix una petita reducció. Aquestes diferències indiquen que, en alguns casos, limitar les modificacions a la màscara pot afavorir la coherència estructural percebuda pel SSIM, tot i que no sempre implica una millor valoració global.

7.1 Anàlisi del *dataset* i dificultat del problema

Tot i l'ús de tècniques d'inpainting avançades i un entrenament acurat, els resultats obtinguts no han assolit el nivell de qualitat esperat. Per tal d'entendre millor aquesta limitació, s'ha dut a terme una anàlisi complementària del conjunt de dades i de les condicions experimentals, amb l'objectiu de verificar si la dificultat del problema pot haver influït de manera substancial en els resultats.

7.1.1 Grau d'oclusió elevat

El percentatge mitjà de píxels oclosos per núvols en el conjunt de prova és del 41,91%, un valor notablement elevat en comparació amb altres estudis d'inpainting on habitualment les oclusions oscil·len entre el 10% i el 30%. Aquest alt grau d'oclusió implica que gairebé la meitat de cada imatge requereix ser reconstruïda, sovint en regions de gran extensió, fet que complica molt la tasca dels models, especialment en escenaris on la informació contextual disponible és limitada.

Aquest factor, per si sol, podria justificar una part important del rendiment suboptimal observat, ja que molts mètodes clàssics i basats en xarxes neuronals han estat optimitzats per a situacions on les zones a reconstruir són més petites, localitzades i envoltades d'informació visual suficient per inferir el contingut ocult.

7.1.2 Diferències entre imatges *Cloudy* i *Clear*

Una altra font rellevant de dificultat és la pròpia naturalesa del conjunt de dades, el qual està compost per imatges amb i sense núvols capturades en dies diferents. Tot i que s'ha aplicat un filtre temporal estricte que limita la diferència de captura a menys de 30 dies, és inevitable que hi hagi canvis reals en el terreny —com modificacions agrícoles, creixement de vegetació, canvis estacionals o transformacions urbanes— que introdueixen una distorsió addicional en les dades.

Aquest aspecte s'ha explorat quantitativament mitjançant dues comparacions de PSNR entre les imatges originals:

PSNR entre *Cloudy* i *Clear*: La informació que ens proporciona aquest càlcul indica des de quin punt estem partint. Amb una mitjana d'11,83 dB, aquesta mètrica mostra que les diferències estructurals i espectrals són molt significatives, especialment a mesura que augmenta la proporció de píxels ocults. Pot semblar un fet evident, però és remarcable que el PSNR disminueixi més de 15 dB. Aquest valor reflecteix la dificultat inherent de la tasca abans fins i tot de dur a terme cap reconstrucció.

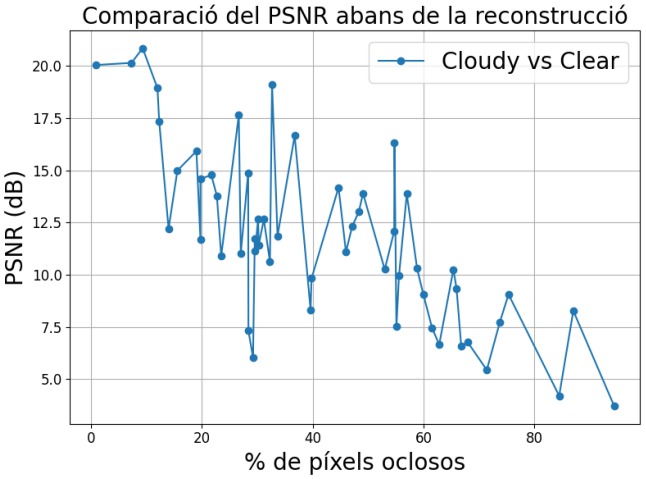

Figura 4. PSNR entre *Cloudy* i *Clear* en funció del percentatge de píxels de la màscara abans de la reconstrucción.

PSNR fora de la màscara: Aquest càlcul s'ha realitzat sobre els píxels no afectats per la màscara d'oclusió, és a dir, fora de la zona ennuvolada. Així, el PSNR reflecteix diferències no atribuïbles als núvols, sinó a canvis reals en l'escena deguts al pas del temps. La mitjana obtinguda és de 21,71 dB, un valor que mostra que les discrepàncies entre *Cloudy* i *Clear* són significatives fins i tot fora de les àrees amb núvols. Aquestes discrepàncies poden limitar el rendiment de l'entrenament, ja que el model ha d'aprendre a reconstruir informació en un context visual que pot no correspondre exactament a l'escena real de referència.

7.1.3 Relació entre percentatge de màscara i qualitat de reconstrucció

S'ha estudiat també la relació entre el percentatge de píxels tapats per la màscara i el rendiment dels models en termes de PSNR. La Figura 5 mostra que, en general, els valors de PSNR disminueixen a mesura que el percentatge d'oclusió augmenta. Aquest patró es manté coherent a través dels diferents mètodes analitzats, confirmant que la dificultat de la reconstrucció augmenta proporcionalment amb la mida de la zona oclosa.

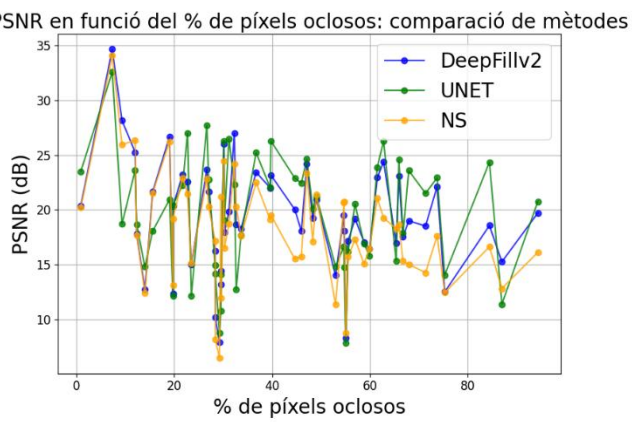

Figura 5. Comparació del PSNR en funció del percentatge de píxels de la màscara per a U-Net, DeepFill i NS (el millor mètode clàssic).

Aquest comportament reforça la hipòtesi que la dificultat del conjunt de dades, tant per la seva elevada oclusió com per les diferències temporals, representa un repte substancial que limita el rendiment absolut dels mètodes provats. No obstant això, també valida l'enfocament adoptat, ja que permet posar en relleu les fortaleses relatives entre mètodes sota condicions realistes i exigents.

7.2 Experiments complementaris i estratègies de millora

Un cop obtinguts els resultats principals, s'han explorat dues línies d'experiments addicionals amb l'objectiu de millorar el rendiment dels models i entendre millor les limitacions dels enfocaments anteriors.

7.2.1 Entrenament amb entrada multi-màscara (U-Net amb 6 canals)

En primer lloc, s'ha modificat l'esquema d'entrada de la U-Net per incloure informació més rica sobre les diferents categories d'oclusió. En lloc d'emprar una única màscara binària agregada, s'han concatenat tres canals corresponents a núvols densos, ombres i núvols translúcids (fins).

Aquest canvi augmenta el nombre de canals d'entrada de 3 (RGB) a 6, permetent que el model pugui diferenciar el tipus d'oclusió present a cada píxel. L'objectiu d'aquesta modificació és que la xarxa aprengui a reconstruir completament les zones de núvol dens, mentre que en àrees d'ombres o núvols translúcids pugui preservar o reutilitzar informació original de la imatge *Cloudy*.

Amb la voluntat d'explotar al màxim el potencial del model, també s'ha aplicat un procés d'augment de dades. Aquest augment consisteix en la generació de noves mostres a partir de les imatges originals mitjançant inversions (flips) horitzontals i verticals, així com cinc rotacions aleatòries. Per tal de preservar la coherència espacial, totes les transformacions han estat aplicades tenint en compte l'efecte mirall, garantint que les imatges resultants mantinguin una forma quadrada. Aquest procediment ha permès multiplicar per vuit el volum de dades disponibles durant l'entrenament, afavorint una millor generalització del model.

Els resultats obtinguts mostren una lleugera millora respecte a la versió estàndard de la U-Net:

	Avaluació global	
	PSNR (dB)	**SSIM**
U-NET 3 Canals	19.59	0.5624
U-NET 6 Canals	20.33	0.6081
U-NET 6 Canals augment de dades	**20.52**	**0.6306**

Taula 3. Resultats quantitatius de U-NET i les seves millores.

Aquestes millores indiquen que l'ús d'informació diferenciada sobre les oclusions pot ajudar la xarxa a gestionar millor la reconstrucció, especialment en contextos mixtos on conviuen diferents tipus de cobertures. El fet de proporcionar al model canals específics per a núvols densos, ombres i núvols translúcids li permet adaptar la seva estratègia de reconstrucció segons la naturalesa de l'oclusió, cosa que es reflecteix en l'increment de PSNR i SSIM.

A més, la incorporació d'augment de dades contribueix a millorar la robustesa del model davant variacions espacials i geomètriques, afavorint una millor generalització en entorns diversos. Això suggereix que, més enllà de la qualitat de les dades d'entrada, l'enriquiment del conjunt d'entrenament i la representació explícita de les característiques de l'oclusió són factors clau per millorar el rendiment en tasques de reconstrucció sota condicions de núvols.

7.2.2 Entrenament amb imatge *Clear* i *Mask* com a entrada

De manera anàloga a com es fa en molts algoritmes d'inpainting clàssic sobre imatges RGB, s'ha avaluat el rendiment dels models utilitzant la imatge *Clear* com a entrada en lloc de la imatge *Cloudy*, aplicant-hi artificialment la màscara d'oclusió corresponent. Aquesta configuració permet eliminar de l'experiment qualsevol variació deguda al pas del temps entre imatges, ja que l'entrada i el ground truth coincideixen exactament fora de la màscara.

L'objectiu és aïllar i analitzar estrictament la capacitat dels models per reconstruir les àrees ocultes, descartant el soroll afegit que suposen els canvis atmosfèrics o físics entre les imatges *Cloudy* i *Clear* reals.

Els resultats mostren una millora clara respecte a la configuració original, amb PSNR i SSIM notablement superiors. Això confirma que gran part de la dificultat inicial no rau en les capacitats dels models, sinó en les discrepàncies temporals i espectrals entre les imatges reals *Cloudy* i *Clear*:

	Avaluació global		Avaluació focalitzada	
	PSNR	**SSIM**	**PSNR**	**SSIM**
U-NET Clear	24.23 dB	0.6945	25.31 dB	0.8098
DEEPFILL Clear	**27.11 dB**	**0.8219**	**27.27 dB**	**0.8355**

Taula 4. Resultats quantitatius dels diferents mètodes d'inpainting amb input *Clear*.

Els resultats de DeepFillv2 destaquen especialment, amb un rendiment significativament superior. El model aprofita clarament el seu preentrenament sobre grans conjunts d'imatges naturals, aplicant aquest coneixement per generar reconstruccions més coherents i realistes, fins i tot en un domini tan diferent com el de les imatges satel·litals. Aquest comportament reforça el valor dels models preentrenats en escenaris on es disposa de dades específiques limitades per entrenar des de zero.

7.3 Avaluació qualitativa visual

Més enllà de les mètriques quantitatives, és fonamental analitzar el comportament visual dels diferents mètodes d'inpainting per tal de detectar artefactes, pèrdues de textura, incoherències estructurals o reconstruccions poc naturals que no sempre es reflecteixen en el PSNR o el SSIM. Per aquesta raó, s'han seleccionat sis escenes representatives i se n'han comparat els resultats visuals.

Les tres primeres escenes presenten núvols de mida i opacitat moderada, i es podria esperar una reconstrucció raonablement bona. La quarta imatge representa un escenari complicat per la quantitat de petits detalls que conté. La cinquena il·lustra un cas amb canvis de color importants entre les imatges *Cloudy* i *Clear*, mentre que la sisena correspon a un cas límit, amb grans àrees ocloses i pràcticament cap context disponible.

A la Figura A2 (configuració original dels experiments), s'observa que els mètodes clàssics com TELEA i NS generen reconstruccions molt artificials, amb degradats poc naturals i patrons geomètrics repetitius. Les dues versions de FSR (FAST i BEST) mostren problemes de coloració —especialment en zones àmplies— amb l'aparició de tons verds o negres que no corresponen al context original. La U-Net de 3 canals millora notablement la integració visual, però pateix pèrdues de detall i artefactes a les vores. DeepFillv2 destaca per una reconstrucció més natural, amb textures més realistes i millor continuïtat espacial, tot i que sense arribar a l'excel·lència. En el cas límit (última fila), com era d'esperar, cap model aconsegueix un resultat convincent.

La Figura A3 mostra les variants del model U-Net. La millora entre la U-Net bàsica i la versió amb 6 canals és evident: s'aconsegueix una millor coherència amb el context, especialment en zones mixtes (amb ombres o núvols translúcids). L'augment de dades aporta un guany discret però estable en definició i estabilitat cromàtica. Tanmateix, en escenes amb detalls petits o estructures molt definides, es continuen observant inconsistències a les vores i zones lleugerament difuminades.

Finalment, la Figura A4 mostra els resultats quan l'entrada és una imatge *Clear* i *Mask*, de manera que es garanteix la coincidència fora de la zona oclosa. En aquest escenari, es neutralitzen els efectes dels canvis temporals en el terreny i es pot avaluar la capacitat de reconstrucció en condicions òptimes. S'hi observa una millora clara en tots els casos, especialment amb DeepFillv2, que produeix resultats notablement aproximats al Ground Truth tant en estructura com en textura. La U-Net també mostra un comportament més estable, amb menys artefactes i millor integració, tot i que continua sent limitada en definició.

8 Conclusions

Aquest treball ha abordat la problemàtica de la presència de núvols en imatges satel·litals mitjançant l'aplicació de tècniques d'inpainting, amb l'objectiu d'emplenar de manera coherent la informació visual de les zones ocloses. S'han explorat tant mètodes clàssics com arquitectures avançades d'aprenentatge profund, i s'ha avaluat el seu comportament mitjançant mètriques quantitatives (PSNR i SSIM) i una anàlisi qualitativa visual.

Els resultats obtinguts permeten extreure diverses conclusions rellevants:

Els mètodes clàssics com Navier-Stokes (NS) i TELEA són útils en escenaris molt locals, però generen reconstruccions artificials i poc naturals quan l'oclusió és extensa. El seu aparent bon rendiment en mètriques globals es deu, principalment, al fet que conserven intactes les regions no afectades, cosa que pot conduir a interpretacions esbiaixades si no s'utilitzen mètriques focalitzades sobre la màscara.

Els models basats en aprenentatge profund, com U-Net i especialment DeepFillv2, han demostrat una capacitat superior per reconstruir regions àmplies i complexes. Aquestes arquitectures són capaces de captar la semàntica contextual de l'escena i generar resultats visualment plausibles, tal com s'ha evidenciat tant en els valors de PSNR i SSIM com en l'anàlisi qualitativa.

El *dataset* InpaintSen12, tot i ser molt ric i útil per a entrenament supervisat, suposa un repte exigent: la mitjana de píxels oclosos per imatge (41,91 %) i les diferències temporals entre les imatges clares i ennuvolades dificulten la reconstrucció precisa, fins i tot per a models ben entrenats.

L'ús de màscares diferenciades (núvol dens, ombra, núvol translúcid) ha aportat una millora modesta però consistent en el model U-Net, especialment pel que fa a la integració visual, ja que el model pot captar millor el grau d'oclusió de cada regió.

L'experiment amb imatges *Clear* i *Mask* com a entrada ha evidenciat que la limitació principal no és tant l'arquitectura dels models com la qualitat i coherència de les dades. Amb aquesta configuració, tant U-Net com DeepFillv2 han assolit resultats molt més satisfactoris (PSNR > 24 dB i SSIM ≈ 0,83 en el cas de DeepFillv2), demostrant el potencial del sistema en condicions òptimes.

L'anàlisi qualitativa ha reforçat aquestes conclusions: DeepFillv2 genera reconstruccions més naturals i coherents, mentre que els mètodes clàssics o simplificats tenen dificultats especialment evidents en escenes complexes o amb grans zones ocloses.

En conjunt, aquest treball demostra que les tècniques d'inpainting tenen un potencial real per a la restauració d'imatges satel·litals afectades per núvols. No obstant això, també posa de manifest la importància de disposar de dades d'alta qualitat, coherents temporalment i amb informació contextual suficient. A mesura que es desen-

volupin nous *datasets* i arquitectures més robustes, serà possible abordar amb més garanties la reconstrucció precisa d'aquest tipus d'imatges, amb aplicacions clares en àmbits com la teledetecció.

De cara a futures investigacions, una millora rellevant seria aprofitar els canals addicionals que ofereix CloudSen12, com les bandes multiespectrals (per exemple, infraroig proper o SWIR), que proporcionen informació complementària sobre vegetació, humitat del sòl i tipus de cobertura. Aquests canals podrien guiar millor el procés de reconstrucció, especialment en zones amb visibilitat reduïda o amb escàs contrast en l'espectre RGB.

A més, es podria plantejar un entrenament en dues fases, on primer es faci un preentrenament massiu amb imatges *Clear* i *Mask* com a entrada, que presenten una correspondència perfecta amb el ground truth, i posteriorment es realitzi un ajust fi (fine-tuning) amb entrada *Cloudy* i *Mask* que tinguin una bona coherència temporal. Aquest enfocament permetria maximitzar l'aprenentatge estructural del model i, alhora, adaptar-lo a les condicions reals i més complexes del problema.

AGRAÏMENTS

Vull expressar el meu agraïment al meu tutor, Daniel Ponsa Mussarra, per la seva dedicació, orientació i suport constant al llarg del desenvolupament d'aquest treball. La seva experiència i els seus consells han estat fonamentals per enfocar i dur a terme aquest projecte amb rigor.

També voldria agrair al grup d'estudi i recerca MSIAU-Progress per generar un entorn col·laboratiu i enriquidor que m'ha permès aprendre i créixer en l'àmbit de la visió per computador.

FONTS D'INFORMACIÓ I BIBLIOGRAFIA

[1] Z. Wan, B. Zhang et al., "A Two-Stage Image Inpainting Technique for Old Photographs Based on Transfer Learning," Electronics, vol. 12, no. 15, Art. 3221, Jul. 25, 2023. doi:10.3390/electronics12153221

[2] M.-T. Tran, S.-H. Kim, H.-J. Yang, and G.-S. Lee, "Multi-Task Learning for Medical Image Inpainting Based on Organ Boundary Awareness," IEEE Access, vol. 11, pp. 61768–61781, 2023. doi: 10.1109/ACCESS.2023.3271276.

[3] Dotndot, "Mastering the Art of AI-Driven 3D Inpainting in Marketing Campaigns," *Dotndot*, [En línia]. Disponible a: https://dotndot.com/ai-driven-3d-inpainting-in-marketing-campaigns/. Accedit 7/3/2025

[4] X. Liu, "Inpainting Orthorectified Satellite Images," PRS Photogrammetry Remote Sensing, ETH Zurich, 2024.

[5] M. Karlsson, et al., "Global Cloudiness and Cloud Top Information from AVHRR Using 42 Years of Data (1979–2020)," *Remote Sensing*, vol. 15, no. 4, p. 904, Feb. 2023.

[6] M. L. Oliveira et al., "Impact of Cloud Cover on Monitoring Crops in Tropical Regions Using Remote Sensing Data," *Remote Sensing*, vol. 8, no. 3, p. 219, Mar. 2016.

[7] A. Smith, "These Satellites See Through the Clouds to Track Flooding," *Wired*, 2021. [En línia]. Disponible a: https://www.wired.com/story/these-satellites-see-through-the-clouds-to-track-flooding. Accedit: 28/5/2025

[8] European Environment Agency, "Use of Remote Sensing in Climate Change Adaptation," *Climate-ADAPT*, 2020. [En línia]. Disponible a: https://climate-adapt.eea.europa.eu/en/metadata/adaptation-options/use-of-remote-sensing-in-climate-change-adaptation Accedit: 28/5/2025

[9] M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in Proc. SIGGRAPH: ACM Special Interest Group on Computer Graphics and Interactive Techniques, New York, 2000, pp. 417–424.

[10] A. Telea, "An Image Inpainting Technique Based on the Fast Marching Method," J. Graph. Tools, vol. 9, no. 1, pp. 25–36, 2004. doi:10.1080/10867651.2004.10487596.

[11] OpenCV. "Image Inpainting (Python)." [En línia]. Disponible a: https://docs.opencv.org/3.4/df/d3d/tutorial_py_inpainting.html Accedit: 28/3/2025

[12] OpenCV. "Image Inpainting (xphoto module)." [En línia]. Disponible a: https://docs.opencv.org/4.2.0/dc/d2f/tutorial_xphoto_inpainting.html Accedit: 28/3/2025

[13] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI, vol. 9351, pp. 234–241, 2015.*

[14] D. Pathak et al., "Context Encoders: Feature Learning by Inpainting," in Proc. CVPR, 2016.

[15] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and Locally Consistent Image Completion," ACM Trans. Graph., vol. 36, no. 4, pp. 1–14, 2017.

[16] Nipponjo, "deepfillv2-pytorch," GitHub repository, [En línia]. Disponible a: https://github.com/nipponjo/deepfillv2-pytorch Accedit: 28/5/2025.

[17] S. Matveev, A. Romanov, A. Bochkovskiy, and D. Vatolin, "LaMa: Resolution-robust Large Mask Inpainting with Fourier Convolutions," arXiv preprint arXiv:2109.07161, 2021.

[18] ImageNet, "Large Scale Visual Recognition Challenge," [En línia]. Disponible a: https://www.image-net.org/ Accedit: 28/3/2025.

[19] Places2, "Places2 Dataset – Scene Recognition with Deep Learning," Kaggle. [En línia]. Disponible a: https://www.kaggle.com/datasets/nickj26/places2-mit-dataset Accedit: 28/3/2025.

[20] CelebA-HQ, "High-Quality Version of CelebA Dataset," Papers with Code, [En línia]. Disponible a: https://paperswithcode.com/dataset/celeba-hq Accedit: 28/3/2025.

[21] X. Yang, Y. Wei, and L. Zhang, "Cloudsen12: A Benchmark Dataset for Cloud Removal in Optical Remote Sensing Images," Scientific Data, vol. 10, no. 88, 2023. doi:10.1038/s41597-022-01878-2. [En línia]. Disponible a: https://cloudsen12.github.io/ Accedit: 26/4/2025

[22] H. Zhou et al., "AllClear: A Comprehensive Dataset and Benchmark for Cloud Removal in Satellite Imagery," NeurIPS Datasets and Benchmarks Track, 2024. [En línia]. Disponible a: https://allclear.cs.cornell.edu/ Accedit: 12/3/2025

[23] Sorour, "38Cloud: Cloud Segmentation on Satellite Images," Kaggle Datasets. [En línia]. Disponible a: https://www.kaggle.com/datasets/sorour/38cloud-cloud-segmentation-in-satellite-images Accedit: 10/3/2025

[24] Sorour, "95Cloud: Cloud Segmentation on Satellite Images," Kaggle Datasets. [En línia]. Disponible a: https://www.kaggle.com/datasets/sorour/95cloud-cloud-segmentation-on-satellite-images Accedit: 10/3/2025

[25] D. F. M. E. K., "U-Net explained visually," YouTube, 2020. [En línia]. Disponible a: https://www.youtube.com/watch?v=dfMEK4bKjRE&t=13s Accedit: 5/5/2025

APÈNDIX

A1. DIAGRAMA DE GANTT

A continuació es mostra el diagrama de Gantt corresponent a la planificació del treball, on s'hi detallen les principals tasques i la seva distribució temporal al llarg del projecte.

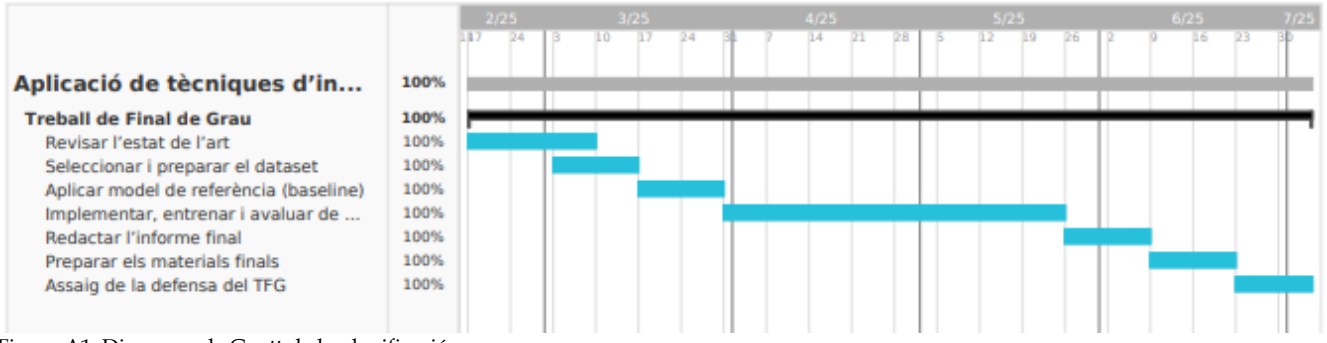

Figura A1: Diagrama de Gantt de la planificació

A2. RESULTATS VISUALS DELS EXPERIMENTS ORIGINALS

Comparació visual entre diferents mètodes d'inpainting, utilitzant com a entrada imatges ennuvolades (*Cloudy*). Cada fila mostra una escena diferent, i les columnes corresponen als següents elements:
D'esquerra a dreta:
Imatge original amb núvols, màscara binària, imatge + màscara, TELEA, NS, FSR_FAST, FSR_BEST, U-Net amb 3 canals, DeepFillv2 i Ground Truth.

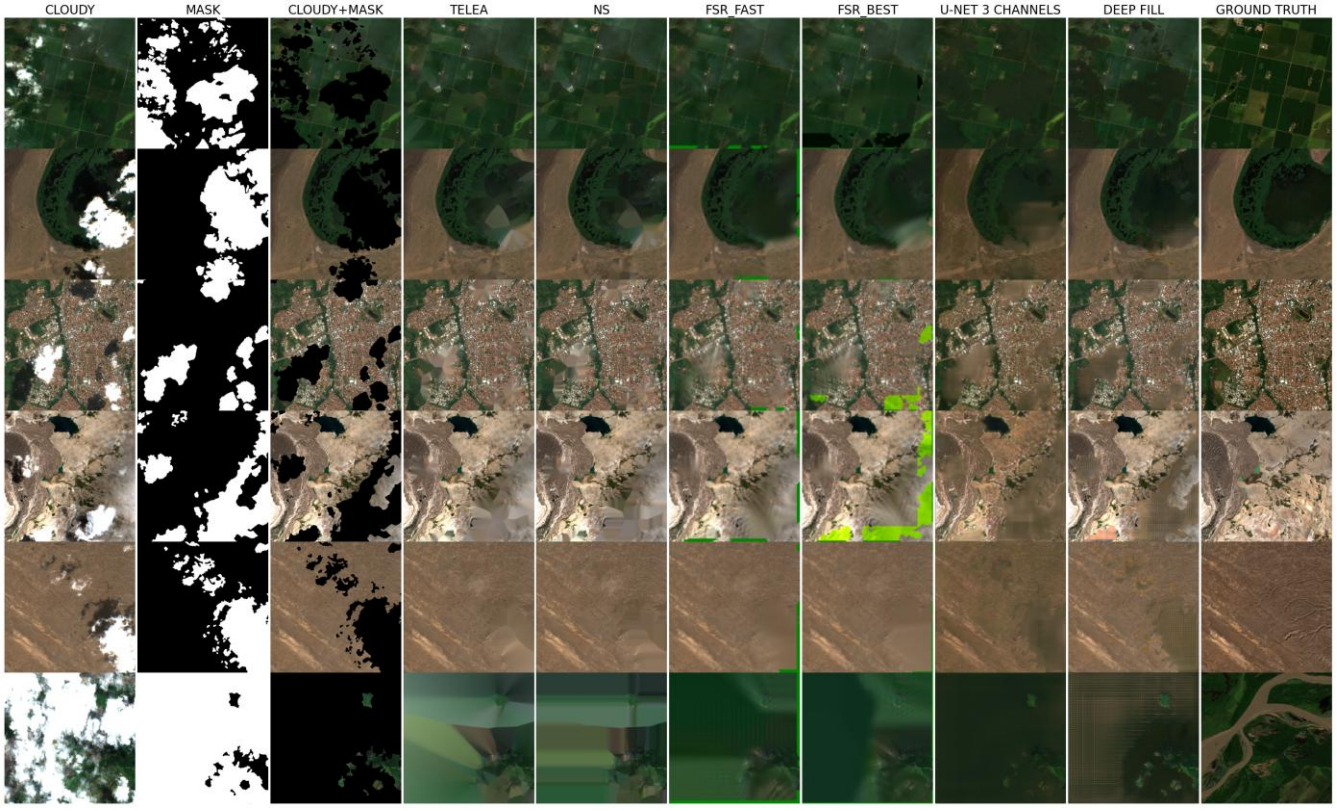

Figura A2: Comparació qualitativa dels resultats dels diferents mètodes d'inpainting, utilitzant imatges *Cloudy* com a entrada.

A3. Resultats Visuals de U-Net i les seves millores

Aquest apartat mostra els resultats visuals obtinguts amb les diferents variants del model U-Net. S'hi inclouen millores com l'ús de 6 canals d'entrada (amb màscares separades) i l'augment de dades durant l'entrenament.
D'esquerra a dreta:
Imatge original amb núvols, màscara binària, imatge + màscara, U-Net amb 3 canals, U-Net amb 6 canals, U-Net amb 6 canals amb augment de dades i Ground Truth.

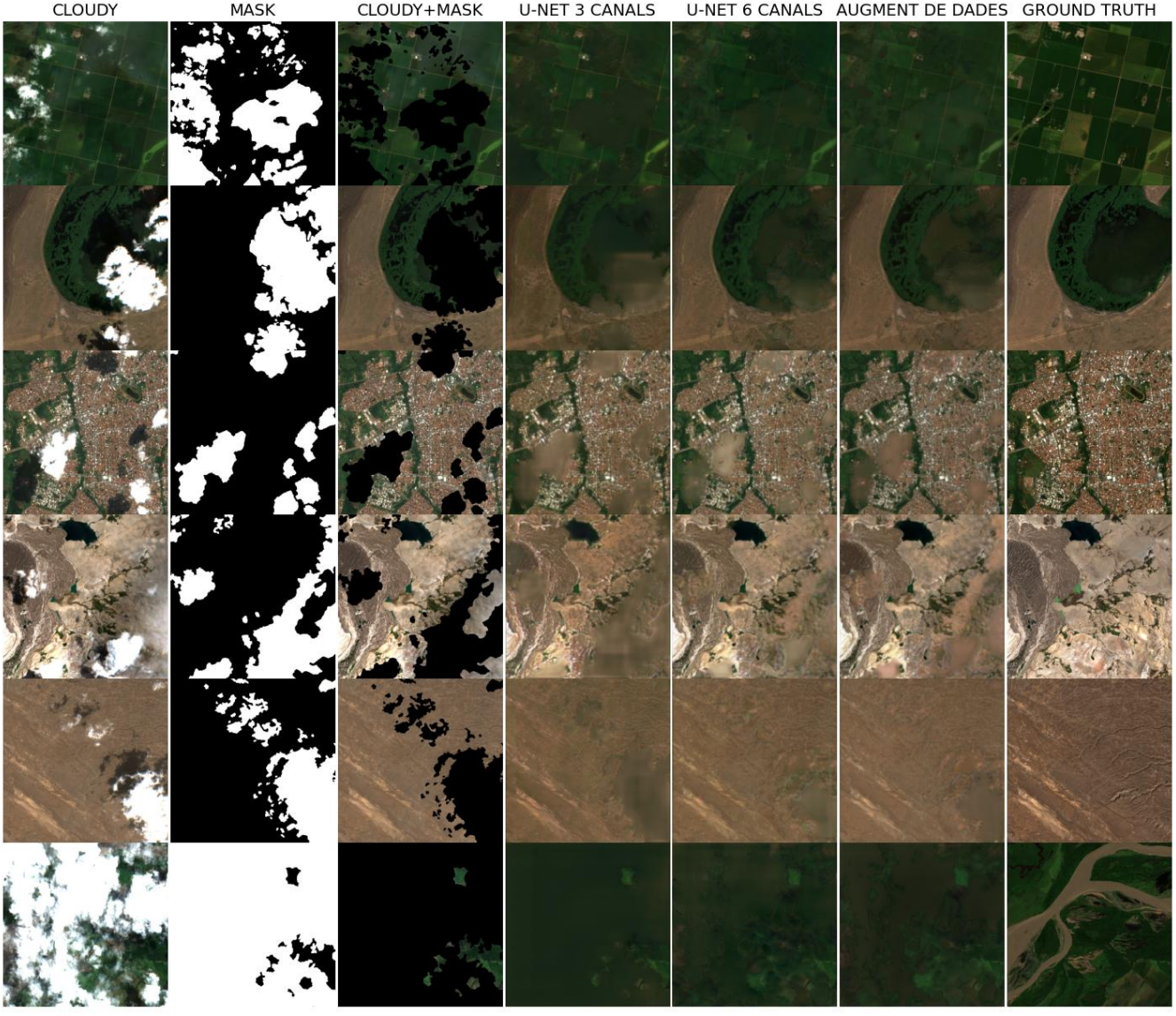

Figura A3: Comparació visual de les variants del model U-Net.

A4. RESULTATS VISUALS AMB INPUT *CLEAR* I *MASK*

Aquest experiment mostra els resultats obtinguts quan es disposa d'una imatge *Clear* i la màscara corresponent com a entrada.
D'esquerra a dreta:
Màscara, entrada *Clear* + *Mask*, reconstrucció amb U-Net, reconstrucció amb DeepFillv2 i Ground Truth.

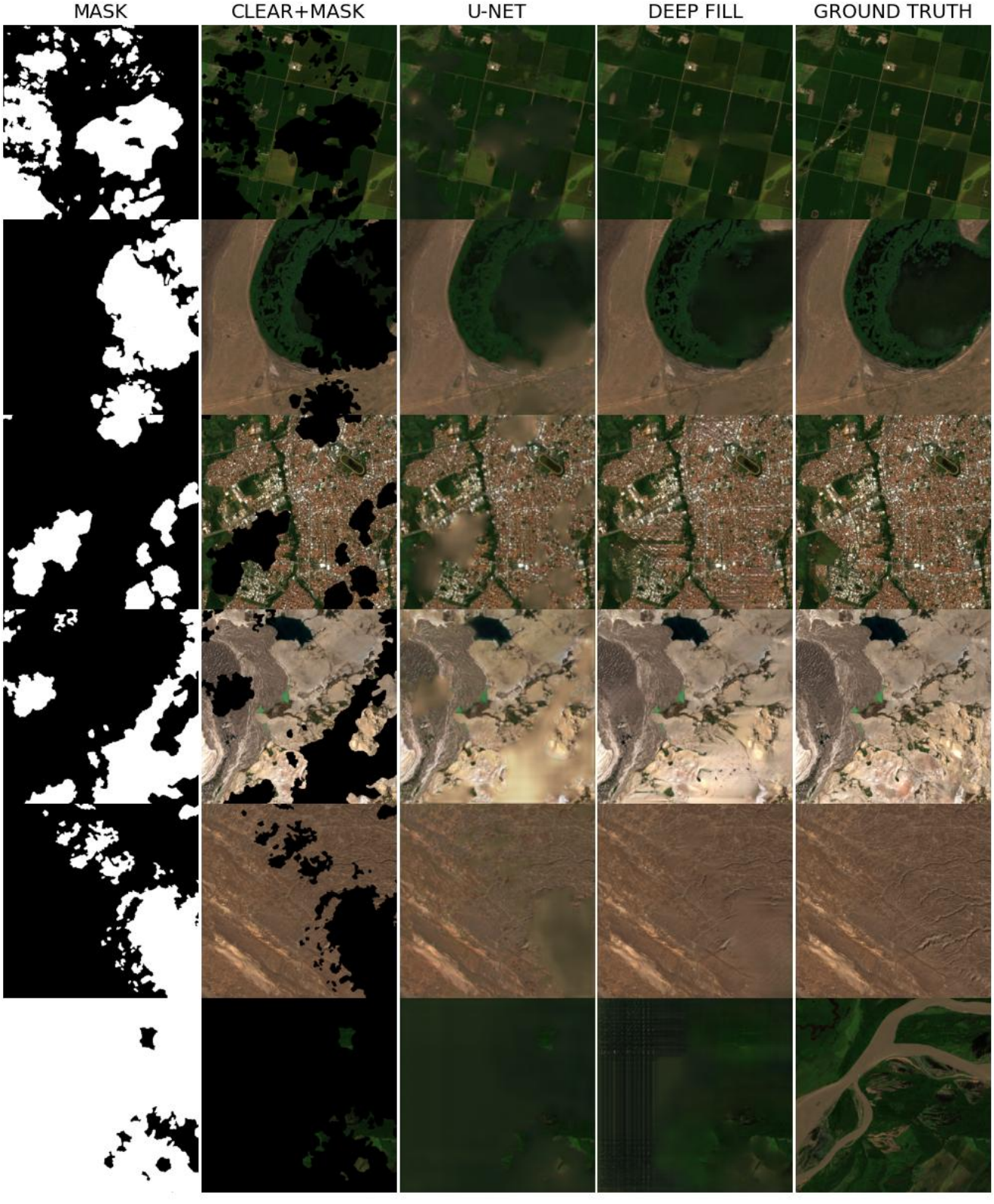

Figura A4: Comparació visual dels resultats dels models U-Net i DeepFillv2 quan l'entrada és una imatge *Clear* i *Mask*.