Scuola di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea in Informatica

Tesi di Laurea

METODI DI APPRENDIMENTO AUTOMATICO PER LA PREDIZIONE DELLA QUALITÀ DI IMMAGINI SAR DESPECKLED

MACHINE LEARNING METHODS FOR THE PREDICTION OF DESPECKLED IMAGE QUALITY

YOUNESS AHARRAM

Relatore: *Daniele Baracchi*

Correlatori: *Fabrizio Argenti, Tommaso Pecorella*

Anno Accademico 2025-2026

CONTENTS

LIST OF FIGURES

"Insert citation"
— Insert citation's author

1

INTRODUZIONE

I satelliti SAR sono satelliti dotati di un radar ad apertura sintetica che permette loro di acquisire immagini della superficie terrestre indipendentemente dalle condizioni meteorologiche e dalla luce solare. I satelliti SAR, grazie a questa loro capacita, trovano applicazione in molteplici contesti disciplinari. Ambito geologico, sono impiegati per il monitoraggio del suolo e dei processi geomorfologici, consentendo la mappatura di foreste, deserti e aree soggette a trasformazioni ambientali. Inoltre, risultano particolarmente efficaci nell'analisi dei fenomeni di deforestazione attraverso il rilevamento dei cambiamenti nella copertura boschiva. Marittimo, permettono di localizzare navi anche in condizioni meteorologiche avverse e di rilevare sversamenti di petrolio o altre sostanze inquinanti. Infrastrutture e urbanistica, vengono utilizzati per misurare gli spostamenti del terreno e delle aree urbane, oltre che per il controllo di dighe, ponti e ferrovie, e per l'osservazione dello sviluppo delle città. Il funzionamento di questo tipo di satellite si basa sull'uso di onde radar che vengono inviate verso la Terra. Questi impulsi elettromagnetici rimbalzano sul terreno e sugli oggetti come edifici o vegetazione e tornano al satellite. Quest'ultimo analizzando il segnale di ritorno riesce ad ottenere informazioni sia sull'intensità del riflesso sia sul tempo impiegato dal segnale per tornare, dati fondamentali per ricostruire l'immagine del territorio. Il punto di forza del SAR è l'apertura sintetica. Poichè il satellite si muove lungo la sua orbita, i segnali raccolti in posizioni diverse vengono combinati insieme. Questo processo permette di simulare un'antenna molto più grande di quella reale, ottenendo così immagini ad altissima risoluzione, molto piu dettagliate di quelle che un radar di dimensioni fisiche limitate potrebbe generare da solo. In pratica, il movimento del satellite trasforma un radar relativamente piccolo in uno strumento potentissimo per osservare il pianeta. L'immagine così generata però presenta un particolare tipo di rumore. Quest'ultimo si forma quando un impulso radar colpisce il terreno, questo non riflette semplicemente un segnale uniforme. In

realtà, il segnale viene riflesso da moltissimi piccoli scatter presenti sulla superficie come foglie, rocce o edifici. Tutti questi ritorni interferiscono tra di loro, sommando le onde con fasi diverse. Il risultato di questa interferenza prende il nome di Speckle. Questo tipo di rumore non è un errore del satellite o del radar, ma una caratteristica intrinseca del tipo di misura e si presenta con un pattern granuloso che rende l'immagine difficle da interpretare ed analizzare. Il processo di riduzione dello speckle prende il nome di despeckling. Quest'ultimo cerca di smussare o filtrare il rumore granulare senza però perdere le informazioni reali presenti nell'immagine. In letteratura vi sono molteplici approcci: alcui si basano su filtri spaziali che analizzano i pixel vicini, altri usano tecniche più sofisticate come statistica multivarianza o metodi di deep learning. Ogni approccio ha i suoi punti di forza e le sue lacune sulla base del tipo di ambiente rappresentato nell'immagine. Lo scopo di questa tesi è cercare di unire i punti di forza di alcuni modelli in modo da ottenere l'immagine con il despeckling più accurato possibile. Per ottenere ciò si utlizza tecniche di machine learning per predire la qualità di un immagine denoised attraverso una mappa di qualità. Ad ogni modello, è associata una mappa che indica, pixel per pixel dove il modello ha funzionato meglio e dove invece peggio. Queste mappe sono essenziali nel processo di fusione delle immagini despeckled dei vari modelli in un unica immagine, in quanto determino i pesi della una media pesata.

2

STATE OF THE ART

Negli ultimi trant'anni sono stati proposti numerosi metodi per la riduzione dello speckle nelle immagini SAR. I primi approcci sfruttano filti spaziali come Lee, Frost e Kuan. Questi operavano direttamente nel dominio dell'immagine, cioè sui pixel, sfruttando finestre locali per stimare statisticamente il rumore e ridurlo. Erano strumenti semplici, poco costosi dal punto di vista computazionale ed efficaci ma soffrivano di un limite strutturale. Per attenuare lo speckle tendevano a smussare anche i dettagli fini, specialmente lungo i bordi o nelle aree eterogenee. Con lo sviluppo della teoria delle trasformate multisensoriale negli anni Novanta , si passò ad un approccio diverso. Invece di agire direttamente sul'immagine, si inziò a trasformarla in un dominio in cui il segnale e il rumore potessero essere seprati. Nascono così i metodi basati su trasformata, come quelli che usano wavelet. Questi strumenti rappresentano un'evoluzione concettuale dei filtri spaziali, perchè superano alcune loro debolezze: riescono a distinguere meglio il rumore dalle strutture significative, ad adattarsi a diverse scale ed a preservare in maniera più accurata bordi, texture e linee sottili. Tuttavia, portano con sè una maggiore complessità computazionale e la possibilità di introdurre artefatti se non calibrati con attenzione. Infine dato che lo speckle è un rumore moltiplicativo e non semplicemente additivo, se non viene trasformato prima, la wavelet può non essere del tutto efficace. Alcuni di queste tipologie di filtri sono stati illustrati e confrontate nell'articolo *A Tutorial on Speckle Reduction in Synthetic Aperture Radar Images* [1]. Negli ultimi anni, l'attenzione si è spostata ancora più avanti verso i metodi non locali, come i filtri non local means o BM3D adattati per le immagini SAR. Qui l'idea è radicalmente diversa, ovvero non ci si limita più a gurdare in un introno locale del pixel, ma si cercano nel resto dell'immagine regioni simili e si usano queste corrispondenze per ridurre il rumore. In questo modo lo speckle viene attenuato in maniera molto efficace, mentre i dettagli strutturali si preservao quasi intatti. La qualità delle immagini risultanti è general-

mente superiore a quella ottenuta con filtri locali o multirisoluzione, ciò comporta però un costo computazionale elevato e la necessità di algoritmi sofisticati per gestire le similitudini tra regioni. Negli ultimi dieci anni si è aperta una nuova fase, spinta dall'esplosione del deep learning. Come riportato nell'articolo *Deep Learning for SAR Images Despeckling* [2], l'idea è che le reti neurali, in particolare convoluzionalil o basate su autoencoder, possano imparare direttamente dai dati le caratteristiche dello speckle e il modo migliore per ridurlo. Questo approccio non si basa più nell'assumere una ditribuzione statistica del rumore o una struttura matematica da preservare, ma si affida alla capacità della rete di apprendere automaticamante dalle coppie di immagini rumorose e pulite. I risultati hanno portato ad una qualità visiva migliore e un eccellente preservazione dei dettagli. D'altro canto, le reti neurali hanno bisogno di grandi quantità di dati ben calibrati per l'addestramento e possono soffrire di scarsa generalizzazione se applicate a scenari diversi da quelli su cui sono state addestrate oltre che ad un costo computazionale molto elevato. Le performance dei modelli di despeckling non è uniforme per tutti i tipi di scenari. La loro efficacia può variare in base alle caratteristiche statistiche del bioma come contesti di vegetazione, aree rocciose e urbane, poichè la distribuzione del rumore e le strutture da preservare differiscono sensibilmente. Un'immagine SAR potrebbe comprendere due o più tipi di biomi, ciò implica che utlizzando un unico modello di despeckling, indipendetemente da quale esso sia, l'immagine risultante avrà aree in cui è stata ripulita meglio e aree in cui è stata ripulita peggio a seconda di dove il modello per come è stato realizzato ha più facilità ad operare. L'idea da cui nasce questa tesi è quello di unire le caratteristiche migliori di determinate tecniche di despeckling, in modo tale che l'immagine risultante rispecchi il più possibile la realtà. Questo tipo di approccio non va a reinventare la ruota, cioè non punta a realizzare un nuovo modello con cui è possibile fare denoising, ma è mirato a sfruttare i punti di forza di modelli già esistenti.

3

PRIMO APPROCCIO

Il primo approccio divideva il problema in due parti. La prima è relativa alla previsine della qualità tramite tecniche di machine learning di immagini denoised. La seconda parte invece riguarda la fusione delle immagini.

3.1 previsione della qualità: dataset

Come illustrato in Figura 1, il dataset impiegato è costituito da tre tipologie di immagini: clean, noisy e denoised. Queste immagini sono state realizzate tramite uno strumento ottico di un satellite per poi essere sporcate con speckle artificiale e su cui infine è stato fatto denoising. Questa scelta è stata fatta in quanto risulta difficile reperire un dataset contenente immagini SAR abbastanza grande da poter essere utlizzato per addestrare una rete neurale. Le immagini su cui viene fatto l'addestramento sono relative ad un singolo modello di despeckling ed ad un determinato look. Quest'ultimo è una metrica che indica l'intensità dello speckle artificiale, in quanto a più look. corrispondono più catture di quella che è la realtà di interesse e quindi si ha una maggior precisione e uno speckle ridotto. In fase di addestramento, le immagini noisy e denoised vengono concatenate in un unico tensore a due canali e utilizzate come input per la rete neurale. Le immagini clean, invece, assieme a quelle denoised, vengono impiegate per la generazione della mappa di qualità, che costituisce il riferimento necessario per il calcolo della funzione di perdita. Questo permette al modello di imparare a prevedere la qualità del denoising relative ad un dato modello di despeckling e ad una determinata intensità dello speckle.

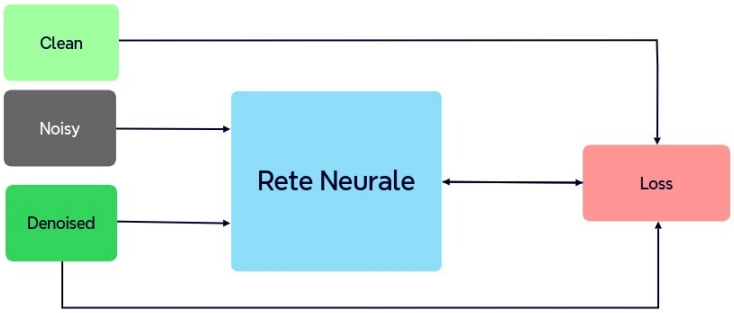

Figure 1: Mock della struttura logica per la previsione della qualità

3.2 PREVISIONE DELLA QUALITÀ: ARCHITETTURA RETE NEURALE

Come Architettura della rete neurale è stata utilizzata la Unet. Questo perchè risulta particolarmente adatto per la generazione di una mappa di qualità che valuti l'affidabilità di un'immagine sottoposta a denoising. Il compito di questa rete è assegnare a ciascun pixel un valore continuo compreso nell'intervallo $[0, 1]$ che ne rappresenti la qualità locale. Questo output per la sua natura richiede un'architettura in grado di produrre una mappa di densità della stessa risoluzione spaziale dell'input, preservando la localizzazione precise delle informazioni. L'architettura è composta da due bracci principali. Il braccio sinistro (encoder) opera una contrazione progressiva, applicando strati convoluzionali e di pooling per estrarre features astratte e gerarchiche dall'input. Questo processo permette alla rete di comprendere il contenuto semantico dell'immagine e la natura del rumore, apprendendo a riconoscere pattern complessi come texture, bordi e regioni omogenee. Il braccio destro (decoder) svolge il compito simmetrico di espansione, ricostruendo gradualmente la risoluzione spaziale attraverso operazioni di up-sampling e convoluzioni. È questa struttura ad "U" che permette di trasformare le feature astatte e globali apprese dall'encoder in una mappa dettagliata. L'elemento più innovativo dell'architettura U-Net è la presenza di connessioni skip che collegano i layer dell'encoder ai layer corrispondenti nel decoder. Questi collegamenti trasferiscono le mappe di feature ad alta risoluzione dalla fase di contrazione a quella di espansione. La loro presenza, è rilevante per la corretta generazione della mappa di qualità. Senza di esse, il decoder avrebbe perso l'informazione spaziale riguardante l'esatta localizzazione di dettagli critici come bordi sottili o piccole texture. Poiché sono proprio

queste aree ad essere più suscettibili ad artefatti da oversmoothing o errata ricostruzione da parte degli algoritmi di denoising, le skip connections forniscono al decoder il contesto spaziale necessario per produrre una valutazione di qualità estremamente precisa e localizzata.Il modello deve superare la semplice stima di una differenza assoluta tra l'immagine denoised e un ground truth ideale. Deve invece sviluppare una percezione semantica dell'errore, ossia la capacità di valutare la gravità di una discrepanza in base al contesto visivo in cui appare. Un errore di grande entità in una regione uniforme (es. un cielo sereno) è percepibilmente molto più fastidioso di un errore di pari entità in una regione altamente texturizzata (es. del fogliame), dove può essere naturalmente mascherato. Analogamente, un piccolo errore di disallineamento su un bordo netto può essere molto visibile e quindi criticabile. L'encoder della U-Net, stratificando features da semplici a complesse, apprende automaticamente questa gerarchia di contenuti visivi. Il modello può dunque fondere la misura dell'errore con la comprensione del significato dell'area in cui esso si manifesta, producendo una mappa di qualità che pesa intelligentemente l'importanza visiva di ogni discrepanza.

3.3 FUSIONE DEI MODELLI

La fusione delle immagini denoised [I] prodotte tramite i rispettivi modelli di despeckling vengono fusi attraverso una media pesata sfuttando le mappa di qualità [M] come pesi per le singole immagini ovvero:

$$\frac{\sum_{i=1}^{n} I_i M_i}{\sum_{i=1}^{n} M_i} \tag{1}$$

Non è una fusione "cieca" (es. una media semplice). È un processo che si comporta in modo diverso per ogni singolo pixel dell'immagine. Se il Modello A ha una qualità stimata molto alta in una regione e il Modello B molto bassa, la fusione privilegerà quasi esclusivamente il Modello A in quell'area. Mentre alcuni sono eccellenti nel preservare i bordi netti ma possono lasciare del rumore residuo nelle aree lisce, altri sono ottimi nell' eliminare il rumore dalle superfici lisce ma tendono a sfocare (oversmooth) i bordi e le texture. La fusione pesata permette di prendere il meglio di ogni modello usando l'output dell'algoritmo migliore per una data caratteristica dell'immagine, scartando virtualmente i suoi contributi peggiori.

BIBLIOGRAPHY

[1] Argenti, Lapin, Bianchi, and Alparone. A tutorial on speckle reduction in synthetic aperture radar images. *IEEE Xplore*, 1(2):6–35, 2013.

[2] Lattari, Leon, Asaro, Rucci, Prati, and Matteucci. Deep learning for sar image despeckling. *remote sensing*, 1(1):20, 2019.