



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

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

Tesi di Laurea

METODI DI APPRENDIMENTO AUTOMATICO  
PER LA FUSIONE DI IMMAGINI SAR  
DESPECKLED

MACHINE LEARNING METHODS FOR THE  
FUSION OF DESPECKLED SAR IMAGES

YOUNESS AHARRAM

Relatore: *Daniele Baracchi*

Correlatori: *Fabrizio Argenti, Tommaso Pecorella*

Anno Accademico 2024-2025



---

## CONTENTS

---

List of Figures	3
1 Introduzione	7
2 State of the Art	9
2.1 Origine fisica dello speckle . . . . .	10
2.2 Modello statistico . . . . .	11
2.3 Effetti e problematiche . . . . .	11
2.4 Tecniche di despeckling . . . . .	11
Bibliography	15



---

## LIST OF FIGURES

---

Figure 1	Impulsi radar inviati dal satellite SAR verso la Terra e riflessi indietro. . . . .	10
----------	--	----



*"Insert citation"*  
— *Insert citation's author*





---

## 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. Missioni e piattaforme di riferimento includono Sentinel-1 (ESA) [5], TerraSAR-X (DLR), COSMO-SkyMed (ASI) e RADARSAT. I satelliti SAR, grazie a questa loro capacità, trovano applicazione in molteplici contesti. In ambito geologico [4], 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. L'immagine generata dal satellite 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 difficile da interpretare ed analizzare. Il processo di riduzione dello speckle prende il nome di despeckling [10]. 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: alcuni si basano su filtri spaziali che analizzano i pixel vicini, altri usano tecniche più sofisticate come 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. Un primo approccio per ottenere ciò consiste nell'utilizzare 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. La fusione avviene tramite la media pesata dove i relativi pesi sono le mappe di qualità. Questo approccio però non sfrutta al massimo i punti di forza di ogni singola immagine despeckled portando ad un risultato finale non soddisfacente, in quanto la qualità del denoising viene stimata concentrandosi sul singolo pixel senza guardare i vicini. Un secondo approccio più efficiente è basato sull'attenzione. Invece di utilizzare mappe di qualità che determinano la bontà del denoising di un singolo pixel, si utilizzano meccanismi basati sulla self e cross attention. Questo consente di andare oltre la valutazione locale pixel per pixel, mettendo in relazione l'informazione proveniente da più immagini despeckled e valorizzando i dettagli complementari.

---

## STATE OF THE ART

---

Il funzionamento di questo tipo di satellite come spiegato nel sito della N.A.S.A. [3] 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 più 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. A seguito della cattura della scena di interesse, l'immagine ottenuta risulta disallineata, in quanto si genera un angolo  $\theta$  tra l'asse del satellite e la superficie terrestre. Per correggere questo effetto, si applica una tecnica di allineamento dell'immagine [6]. L'immagine generata dal satellite però presenta un rumore che prende il nome di speckle. Il fenomeno dello speckle è una caratteristica intrinseca delle immagini acquisite da sensori coerenti, come i radar ad apertura sintetica (SAR), i sistemi laser o gli interferometri ottici. Dal punto di vista fisico, lo speckle nasce dall'interferenza coerente tra le onde elettromagnetiche riflesse da molteplici scatterer presenti all'interno di una singola cella di risoluzione del sensore. Ciascuno di questi scatterer contribuisce con un segnale complesso avente ampiezza e fase proprie; la somma coerente di tali contributi produce una risultante la cui ampiezza varia casualmente nel tempo e nello spazio.

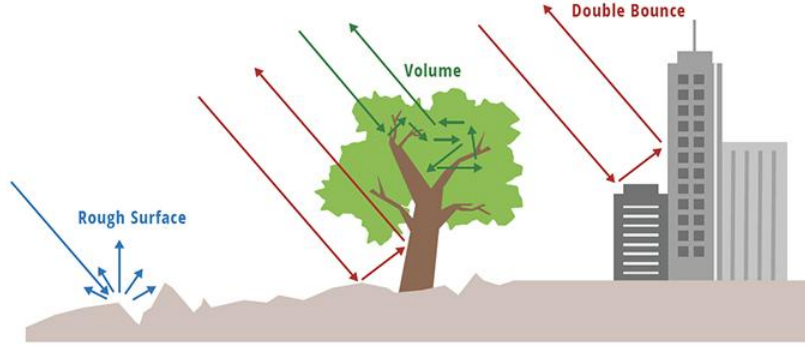


Figure 1: Impulsi radar inviati dal satellite SAR verso la Terra e riflessi indietro.

Questo effetto di interferenza, costruttiva o distruttiva, genera un pattern granulare nell'immagine osservata, noto appunto come speckle. Quest'ultimo degrada la qualità visiva e radiometrica dell'immagine, rendendo più complessa l'analisi e l'interpretazione dei dati. Le immagini ottenute da sistemi radar ad apertura sintetica (SAR) sono caratterizzate da un disturbo granulare denominato speckle. Questo fenomeno è intrinseco a tutti i sistemi di imaging coerente e rappresenta un rumore di natura moltiplicativa, dipendente dal segnale, che degrada l'aspetto visivo e le prestazioni dei processi automatici di analisi dell'immagine, come classificazione o rilevamento di cambiamenti.

## 2.1 ORIGINE FISICA DELLO SPECKLE

Il sensore SAR emette un'onda elettromagnetica e riceve il segnale retrodiffuso da una piccola area della scena, detta *cella di risoluzione*. Tale cella contiene numerosi scatterer elementari, ciascuno caratterizzato da una propria ampiezza  $A_i$  e fase  $\phi_i$ . Il segnale complesso ricevuto può essere modellato come somma coerente dei contributi di tutti gli scatterer presenti nella cella [1]:

$$z = \sum_{i=1}^N A_i e^{j\phi_i}. \quad (1)$$

Le fasi  $\phi_i$  variano casualmente a causa delle differenze di cammino ottico e, per un numero sufficientemente grande di scatterer indipendenti, il teorema del limite centrale implica che le componenti reale e immaginaria di  $z$  sono variabili gaussiane a media nulla e varianza  $\sigma^2/2$ . In tal caso, si parla di *speckle pienamente sviluppato* (*fully developed speckle*).

## 2.2 MODELLO STATISTICO

Il modulo del segnale complesso  $A = |z|$  segue una distribuzione di Rayleigh:

$$p_A(A) = \frac{A}{\sigma^2} e^{-A^2/(2\sigma^2)} \quad (2)$$

mentre l'intensità  $I = A^2$  segue una distribuzione esponenziale:

$$p_I(I) = \frac{1}{\sigma^2} e^{-I/\sigma^2} \quad (3)$$

Il valore medio dell'intensità è  $\mathbb{E}[I] = \sigma^2$ , che rappresenta la riflettività media del bersaglio o Radar Cross Section (RCS). Si ottiene così il cosiddetto modello moltiplicativo del rumore:

$$I = \mu u \quad (4)$$

dove  $\mu$  è la riflettività media (informazione utile) e  $u$  è una variabile casuale esponenziale a media unitaria che rappresenta il rumore speckle. La varianza di  $I$  è proporzionale al quadrato della media, quindi i pixel più luminosi sono affetti da un disturbo più forte. Per ridurre la varianza, si può effettuare una media su  $L$  osservazioni indipendenti (tecnica detta *multilooking*), ottenendo una distribuzione  $\mathcal{C}$  per l'intensità media  $I_L$ , con varianza ridotta a  $\sigma^2/L$ .

## 2.3 EFFETTI E PROBLEMATICHE

Lo speckle riduce il rapporto segnale rumore (SNR) e ostacola l'estrazione di informazioni affidabili da immagini SAR. Inoltre, poiché la sua intensità è proporzionale al segnale stesso, il rumore è più evidente nelle aree ad alta riflettività. La sua natura coerente e moltiplicativa richiede quindi modelli e tecniche di riduzione specifiche, diverse dai filtri per rumore additivo.

## 2.4 TECNICHE DI DESPECKLING

Negli ultimi trent'anni sono stati proposti numerosi metodi per la riduzione dello speckle nelle immagini SAR. I primi approcci sfruttano filtri spaziali

come Lee, Frost e Kuan [9]. 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 sull'immagine, si iniziò a trasformarla in un dominio in cui il segnale e il rumore potessero essere separati. Nascono così i metodi basati su trasformata, come quelli che usano wavelet [2]. 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 sé 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 [1]. Negli ultimi anni, l'attenzione si è spostata verso metodi non locali, come il filtro Non-Local Means e l'algoritmo SARBM3D, adattati specificamente per il despeckling di immagini SAR. Qui l'idea è radicalmente diversa, ovvero non ci si limita più a guardare in un intorno 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 preservano quasi intatti. La qualità delle immagini risultanti è generalmente 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 [7]. L'idea è che le reti neurali, in particolare convoluzionali 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 distribuzione statistica del rumore o una struttura matematica da preservare, ma si affida alla capacità della rete di apprendere autonomamente 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 utilizzando un unico modello di despeckling, indipendentemente 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 di interesse. Questo tipo di approccio non va a reinventare la ruota, cioè non punta a realizzare un nuovo modello con cui è possibile fare despeckling, ma è mirato a sfruttare i punti di forza di modelli già esistenti. Inizialmente è stata usata una tecnica naive che prevede un architettura Unet per addestrare tanti modelli quante sono le tecniche di despeckling della quale si vuole imparare a prevedere la qualità del denoising generando così mappe di qualità che determinano quanto dell'immagine denoised di un modello prendere in relazione alla bontà del denoising. Tuttavia, questa tecnica non riesce a sfruttare appieno le potenzialità dei singoli modelli, poichè la stima della qualità è effettuata a livello locale e si concentra sul singolo pixel, trascurando così informazioni contestuali di più ampia scala. Inoltre, la combinazione pesata pixel-per-pixel non permette di cogliere la complementarità delle caratteristiche tra i diversi metodi a livello di patch, limitando la capacità del sistema di valorizzare i punti di forza specifici di ciascun modello. L'approccio proposto in [8], invece, supera tali limiti introducendo meccanismi di attenzione incrociata, che consentono una fusione più coerente e informata tra le diverse rappresentazioni. Gli approcci basati su questa tecnica propongono esattamente questa linea di azione: invece di pesare singoli pixel, si estraggono rappresentazioni (feature) dai diversi output despeckled e si usa un modulo di attenzione per selezionare, a livello di feature e di contesto, quali informazioni preservare e quali attenuare. Questo approccio permette di superare i limiti dell'approccio pixel per pixel, in quanto una fusione basata su patch consente di catturare informazioni contestuali e di valorizzare le relazioni strutturali presenti nell'immagine. Operando su blocchi di feature e non su singoli pixel isolati, il modello è in grado di preservare meglio i bordi e le discontinuità, la coerenza

spaziale della patch riduce il rischio di smussare i contorni netti, tipico delle fusioni locali.



---

## BIBLIOGRAPHY

---

- [1] Fabrizio Argenti, Alessandro Lapini, Tiziano Bianchi, and Luciano Alparone. A tutorial on speckle reduction in synthetic aperture radar images. *IEEE Geoscience and Remote Sensing Magazine*, 1(3):6–35, 2013.
- [2] Fabrizio Argenti and Gionatan Torricelli. Speckle suppression in ultrasonic images based on undecimated wavelets. *EURASIP Journal on Advances in Signal Processing*, (5):379638, 2003.
- [3] NASA Earthdata. Synthetic aperture radar (sar), 2025.
- [4] G. Esposito, I. Marchesini, A. C. Mondini, P. Reichenbach, M. Rossi, and S. Sterlacchini. A spaceborne sar-based procedure to support the detection of landslides. *Natural Hazards and Earth System Sciences*, 20(9):2379–2395, 2020.
- [5] European Space Agency (ESA). Sentinel-1: Radar vision for copernicus, 2025.
- [6] Lloyd Haydn Hughes, Diego Marcos, Sylvain Lobry, Devis Tuia, and Michael Schmitt. A deep learning framework for matching of sar and optical imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169:166–179, 2020.
- [7] Lattari, Leon, Asaro, Rucci, Prati, and Matteucci. Deep learning for sar image despeckling. *remote sensing*, 1(1):20, 2019.
- [8] Hui Li and Xiao-Jun Wu. CrossFuse: A Novel Cross Attention Mechanism based Infrared and Visible Image Fusion Approach. *Information Fusion*, 103:102147, 2024.
- [9] Sanjjushri Varshini R, Rohith Mahadevan, Bagiya Lakshmi S, Mathivanan Periasamy, Raja CSP Raman, and Lokesh M. Speckle noise analysis for synthetic aperture radar (sar) space data, 2024.
- [10] Yongjian Yu and S.T. Acton. Speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing*, 11(11):1260–1270, 2002.