POLITECNICO DI TORINO

Laurea MAGISTRALE in Ingegneria Informatica

Tesi di laurea MAGISTRALE

Deep Learning per la previsione spaziale degli aftershock: un approccio end-to-end basato su dati InSAR

Supervisors

Prof. Paolo GARZA

Prof. Daniele REGE CAMBRIN

Candidate

Marco PERNO

LUGLIO 2026

Abstract

Questa tesi propone una pipeline end-to-end per la predizione della distribuzione spaziale degli aftershock a partire da dati interferometrici SAR (InSAR). Il lavoro copre l'intera catena di elaborazione: integrazione di cataloghi sismici globali, declustering mainshock–aftershock, selezione automatica dei frame LiCSAR, acquisizione degli interferogrammi e costruzione di un dataset geofisico spazialmente coerente, con normalizzazione basata sul raggio di Gardner come unità di scala adattiva all'epicentro.

Il modello proposto è composto principalmente da una rete neurale convoluzionale U-Net a scala locale con encoder ResNet. Vengono confrontate diverse configurazioni di input: il modello cardine utilizza spostamento InSAR e mappa di coerenza; configurazioni alternative sostituiscono o integrano questi canali con mappe di foreshock e con la rappresentazione delle faglie attive del dataset GEM. La supervisione avviene tramite mappe binarie su griglia 256×256, in cui ogni epicentro è rappresentato come un disco di raggio pari all'incertezza nominale di localizzazione dei cataloghi globali ($\sim$2 km), così da evitare una supervisione puntuale eccessivamente rigida. La funzione di perdita è formulata nell'ambito dei processi di intensità di Poisson, scelta motivata dall'estremo sbilanciamento tra pixel positivi e negativi che caratterizza questo tipo di mappe. I risultati sono valutati rispetto a una baseline ETAS isotropa.

I risultati mostrano che il segnale InSAR cosismico contiene informazione predittiva sulla distribuzione degli aftershock, ma richiede contesto sismologico per esprimere il proprio potenziale: in assenza di metadati scalari la rete fatica a calibrare l'ampiezza delle predizioni attraverso eventi di magnitudo e baseline temporale molto diversi. L'integrazione di magnitudo, intervallo temporale dall'evento principale, qualità della coerenza e conteggio dei foreshock tramite un meccanismo di condizionamento risolve questa limitazione in modo marcato, producendo guadagni sostanziali rispetto alla baseline statistica isotropa. Le configurazioni basate su mappe di densità dei foreshock e faglie attive raggiungono le prestazioni più elevate, con un vantaggio particolarmente netto in presenza di sismicità precursoria; nella condizione operativamente frequente di assenza di foreshock, il segnale interferometrico si conferma una sorgente informativa robusta. La tesi discute inoltre i limiti derivanti dall'incompletezza dei cataloghi sismici globali e le implicazioni operative per sistemi di previsione a breve termine del rischio sismico.

Acknowledgements

Ringrazio il professor Paolo Garza e Daniele Rege Cambrin per i suggerimenti e l'aiuto tecnico forniti durante il lavoro che ha dato vita a questa tesi. Ringrazio la mia famiglia e i miei amici per l'aiuto e il supporto che mi hanno dato durante gli anni di studio. Ringrazio HPC@POLITO per aver fornito le risorse computazionali necessarie a questo progetto.

Table of Contents

List of Tables VI

List of Figures VII

Acronyms IX

1 Introduzione 1
- 1.1 Motivazione e problema di ricerca . . . 1
- 1.2 InSAR come sorgente di informazione cosismica . . . 2
- 1.3 Apprendimento automatico per la previsione spaziale degli aftershock 2
- 1.4 Contributo di questa tesi . . . 3
- 1.5 Struttura della tesi . . . 5

2 Contesto scientifico e tecnologico 6
- 2.1 Sismologia delle sequenze di aftershock . . . 6
  - 2.1.1 Leggi empiriche della sismicità . . . 6
  - 2.1.2 Finestre spazio-temporali di Gardner–Knopoff . . . 7
  - 2.1.3 Il modello ETAS e la baseline spaziale isotropa . . . 8
  - 2.1.4 Trasferimento di stress di Coulomb . . . 9
  - 2.1.5 Stima di densità kernel per la sismicità spaziale . . . 10
- 2.2 Interferometria SAR per la sismologia . . . 12
  - 2.2.1 Principi dell'InSAR . . . 12
  - 2.2.2 Coerenza interferometrica . . . 14
  - 2.2.3 LiCSAR e il portale COMET-LiCS . . . 15
- 2.3 Apprendimento profondo per la segmentazione spaziale . . . 16
  - 2.3.1 Deep learning . . . 16
  - 2.3.2 Reti neurali convoluzionali . . . 17
  - 2.3.3 Residual Networks (ResNet) . . . 18
  - 2.3.4 U-Net per la segmentazione pixel-wise . . . 19
  - 2.3.5 Condizionamento su metadati scalari: FiLM . . . 20
  - 2.3.6 Funzione di perdita di Poisson . . . 21

3 Metodologia 23
3.1 Formulazione del problema . . . 23
3.2 Acquisizione del catalogo sismico . . . 24
3.3 Declustering e identificazione dei mainshock . . . 25
3.4 Selezione dei frame LiCSAR e acquisizione degli interferogrammi . . . 26
3.5 Costruzione del dataset . . . 27
3.5.1 Normalizzazione spaziale e estrazione delle patch . . . 27
3.5.2 Configurazioni di input . . . 27
3.5.3 Metadati sismologici . . . 29
3.5.4 Definizione del target binario . . . 29
3.5.5 Augmentazione dei dati . . . 30
3.6 Suddivisione del dataset . . . 31
3.7 Architettura del modello . . . 31
3.8 Encoder: backbone ResNet-18 preaddestrato . . . 33
3.9 Decoder con porte di attenzione . . . 34
3.9.1 Porta di attenzione (Attention Gate) . . . 35
3.9.2 Blocco residuo convoluzionale con GroupNorm . . . 36
3.10 Blocco di fusione e condizionamento FiLM . . . 37
3.11 Testa di output . . . 38
3.12 Funzione di perdita . . . 38
3.13 Procedura di addestramento . . . 39
3.13.1 Ottimizzatore e scheduling del learning rate . . . 39
3.13.2 Tecniche di regolarizzazione e stabilità . . . 39
3.13.3 Early stopping e checkpointing . . . 40
3.13.4 Riepilogo degli iperparametri . . . 40
3.13.5 Inizializzazione dei pesi . . . 40

4 Risultati sperimentali 42
4.1 Protocollo di valutazione . . . 42
4.2 Risultati globali sul test set . . . 43
4.3 Confronto con la baseline ETAS . . . 45
4.4 Effetto dei metadati sismologici . . . 47
4.5 Analisi per sottogruppi . . . 48
4.5.1 Numero di aftershock . . . 48
4.5.2 Numero di foreshock . . . 49
4.5.3 Magnitudo del mainshock . . . 50
4.5.4 Qualità della coerenza interferometrica . . . 50
4.5.5 Gap temporale InSAR–mainshock . . . 51
4.6 Discussione . . . 51

5 Conclusioni e Lavori Futuri 54
5.1 Conclusioni 54
5.2 Lavori Futuri 57
5.2.1 Miglioramento e ampliamento del dataset 57
5.2.2 Sviluppi architetturali 57
5.2.3 Validazione e deployment operativo 58
5.3 Riflessione finale 59

Bibliography 60

List of Tables

3.1 Iperparametri di addestramento. 41

4.1 Riepilogo delle configurazioni valutate. M = metadati FiLM; T = split temporale; S = split spaziale. 43
4.2 Metriche medie per campione sul test set (155 campioni). La riga ETAS riporta i valori della baseline valutata sullo stesso insieme. . . 44
4.3 Confronto con la baseline ETAS sulle metriche *pooled* e di verosimiglianza. LLR: log-likelihood ratio (positivo = meglio di ETAS). IG: information gain in bit per pixel. ΔAUC e ΔF1 calcolati rispetto alla corrispondente ETAS dello stesso split. 46
4.4 Guadagno assoluto introdotto dai metadati FiLM, calcolato come (con metadati) − (senza metadati). Metriche *pooled* e *per campione*. 47
4.5 AUC e F1 medie per campione stratificate per numero di aftershock. Solo configurazioni con split temporale. Sono esclusi i campioni senza aftershock (AUC non definita). 48
4.6 AUC e F1 medie per campione stratificate per numero di foreshock. Solo configurazioni con split temporale. 49
4.7 AUC media per campione stratificata per fascia di magnitudo. Split temporale. Tra parentesi il numero di campioni. 50
4.8 AUC e F1 medie per campione stratificate per quartile di coerenza. Split temporale. 50
4.9 AUC e F1 medie per campione stratificate per gap temporale InSAR–mainshock. Split temporale. 51

List of Figures

2.1 Cluster di terremoti dopo il mainshock. 8
2.2 Distribuzione di probabilità degli aftershock secondo il modello ETAS. 9
2.3 Campo di stress di Coulomb dopo il mainshock. 10
2.4 Esempio di mappa KDE costruita a partire da epicentri di foreshock. 12
2.5 Interferogramma InSAR del terremoto di Ridgecrest 2019. Fonte: [4] 14
2.6 Coherence del interferogramma del terremoto di Ridgecrest 2019. Fonte: [4] 15
2.7 Esempio di rete neurale a più strati e backpropagation. Fonte: [14]. 17
2.8 Funzionamento interno di una rete neurale convoluzionale. Fonte: [14]. 18
2.9 Architettura della U-Net. Fonte: [8] 20

3.1 Esempio di target binario dove la circonferenza rappresenta la finestra spaziale del mainshock definita dalla formula di Gardner–Knopoff. . 24
3.2 Architettura del modello proposto. 32
3.3 Architettura U-Net. 33
3.4 Architettura del singolo UpBlock. 34
3.5 Architettura del gate di attenzione nel decoder. 35
3.6 Architettura del blocco convoluzionale residuo con GroupNorm. . . 36
3.7 Architettura del blocco di fusione e condizionamento FiLM. 37

4.1 Previsioni del modello InSAR-T per il terremoto 2024-04-18 in Turchia. 45
4.2 Previsioni del modello InSAR-T con threshold. In nero TN, in rosso FP, in verde TP. 45
4.3 Previsioni del modello FF-T. 45

Acronyms

AI
artificial intelligence

ML
machine learning

DL
deep learning

CNN
Convolutional Neural Network

InSAR
Interferometric Synthetic Aperture Radar

NLL
Negative log-likelihood

KDE
Kernel Density Estimation

ETAS
Epidemic Type Aftershock Sequence

Chapter 1

Introduzione

Panoramica

In questo capitolo vengono presentati il problema di ricerca, la motivazione che lo ha originato e la struttura della tesi.

1.1 Motivazione e problema di ricerca

I terremoti rappresentano una delle minacce naturali più distruttive, con conseguenze che spaziano dalla perdita di vite umane al danneggiamento delle infrastrutture e all'impatto economico su intere regioni. Un terremoto di magnitudo elevata — il *mainshock* — è quasi invariabilmente seguito da una sequenza di eventi minori, gli *aftershock*, che si protraggono per giorni, mesi o addirittura anni dopo la rottura principale [1]. Sebbene gli aftershock siano individualmente meno potenti del mainshock, il loro effetto cumulativo aggrava i danni a strutture già compromesse e ostacola gravemente le operazioni di soccorso nelle ore e nei giorni immediatamente successivi all'evento principale. Una previsione affidabile di *dove* è più probabile che si verifichino gli aftershock è quindi di rilevanza operativa diretta: orienta le priorità di ricerca e soccorso, guida le ispezioni strutturali e supporta l'allocazione delle risorse durante la finestra post-sismica critica.

Nonostante decenni di ricerca, la *distribuzione spaziale* degli aftershock rimane difficilmente prevedibile. Il decadimento temporale è ben caratterizzato dalla legge empirica di Omori e dalla sua generalizzazione, ma il problema spaziale è intrinsecamente più complesso: gli aftershock non si distribuiscono isotropicamente attorno all'epicentro del mainshock, bensì riflettono la geometria del trasferimento cosismico di stress, la rete locale di faglie e lo stato pre-esistente della crosta. I modelli statistici classici, tra cui il modello ETAS (*Epidemic-Type Aftershock Sequence*) [1], descrivono con alta fedeltà la produttività e il clustering temporale

degli aftershock, ma adottano kernel spaziali isotropi che non sono in grado di riprodurre pattern direzionali o controllati dalla struttura delle faglie. L'approccio basato sullo stress di Coulomb è stato proposto come criterio spaziale fisicamente motivato [2], ma la sua capacità predittiva dipende fortemente dall'orientazione assunta delle faglie riceventi, un parametro raramente noto con precisione al momento della risposta all'emergenza.

1.2 InSAR come sorgente di informazione cosismica

L'Interferometria SAR (InSAR) da satellite offre una prospettiva osservativa complementare. A partire dalla prima misurazione geodetica della deformazione cosismica tramite InSAR da parte di Massonnet et al. [3] per il terremoto di Landers del 1992 (M_{w} 7.3), questa tecnica è divenuta lo strumento geodetico spaziale di riferimento per la mappatura degli spostamenti superficiali con precisione da centimetrica a millimetrica, su ampie aree e in qualsiasi condizione meteorologica. Un interferogramma cosismico codifica lo spostamento integrato lungo la linea di vista tra un'acquisizione precedente e una successiva all'evento, tracciando direttamente l'impronta delle variazioni di stress impresse alla crosta. Questo pattern di deformazione contiene informazioni sulla geometria della faglia, sull'estensione della rottura e sulla distribuzione dello slip — esattamente le grandezze fisiche che governano la localizzazione degli aftershock.

La disponibilità operativa dei dati Sentinel-1, processati automaticamente dal sistema LiCSAR [4] e distribuiti gratuitamente attraverso il portale COMET-LiCS, ha reso per la prima volta accessibili dataset InSAR cosismici a scala globale. LiCSAR produce interferogrammi geocodificati in fase avvolta e svolta e mappe di coerenza per tutte le principali regioni sismicamente attive, con un tempo di rivisita tipico di 6–12 giorni, consentendo la generazione quasi in tempo reale di interferogrammi per eventi di $M_{\mathrm{w}} \geq 5$.

1.3 Apprendimento automatico per la previsione spaziale degli aftershock

La disponibilità di ampi cataloghi sismici, combinata con i progressi nel *deep learning* per il riconoscimento di pattern spaziali, ha aperto nuove strade per la previsione *data-driven* degli aftershock. DeVries et al. [2] hanno dimostrato che una rete neurale addestrata su oltre 131.000 coppie mainshock–aftershock può superare i criteri classici basati sullo stress di Coulomb nella previsione della localizzazione degli aftershock (AUC 0.849 contro 0.583), e che il criterio appreso è fisicamente

interpretabile in termini di invarianti del tensore di stress.

Tuttavia, tale risultato è stato oggetto di revisione critica: Mignan and Broccardo [5] hanno dimostrato che una semplice regressione logistica a un neurone ottiene prestazioni equivalenti alla rete profonda di DeVries et al., suggerendo che il contributo determinante risieda nella scelta delle feature di stress piuttosto che nella complessità dell'architettura. Entrambi gli approcci condividono inoltre un limite fondamentale: si basano su campi di stress sintetici derivati da modelli di faglia ideali, disponibili solo a posteriori e sensibili a ipotesi geometriche difficilmente verificabili in fase di emergenza.

Questo apre la questione centrale affrontata in questa tesi: è possibile sostituire i campi di stress sintetici con *dati osservativi diretti*, disponibili operativamente nelle ore successive al mainshock? Tra questi, i più informativi sono il campo di deformazione cosismica misurato da InSAR — che integra implicitamente la geometria della faglia e la distribuzione dello slip — e la distribuzione spaziale dei foreshock, che traccia le strutture sismogeniche attivate prima della rottura principale. Nessuno dei due richiede ipotesi di modello: entrambi sono osservabili direttamente dal catalogo sismico e dai dati satellitari.

1.4 Contributo di questa tesi

Questa tesi affronta il seguente problema di ricerca:

> *È possibile addestrare una rete neurale convoluzionale a prevedere la distribuzione spaziale degli aftershock di un mainshock di* M_{w} *o* $M_{\mathrm{ww}} \geq 5.0$ *utilizzando esclusivamente dati osservativi — interferogramma InSAR, mappa di coerenza, metadati sismologici ed eventualmente distribuzione spaziale dei foreshock e faglie attive — senza ricorrere a modelli di stress sintetici?*

Per rispondere a questa domanda viene sviluppata una pipeline end-to-end completa, dall'acquisizione del catalogo sismico globale alla valutazione del modello addestrato, con le seguenti scelte progettuali fondamentali.

Declustering e selezione dei mainshock. I candidati mainshock sono identificati con un approccio cronologico: ogni evento di M_{w} o $M_{\mathrm{ww}} \geq 5.0$ che non ricade nella finestra spazio-temporale di un mainshock precedente viene accettato come nuovo mainshock. La distanza spaziale tra eventi è calcolata con la formula di Haversine, che tiene conto della curvatura terrestre. Il raggio spaziale segue la relazione empirica di Gardner & Knopoff $R = 10^{0.1238\,M+0.983}$ km [6], mentre la finestra temporale per gli aftershock è calibrata empiricamente tra 120 e 365 giorni in funzione della magnitudo, così da bilanciare completezza del catalogo e purezza

del clustering. La finestra per i foreshock è mantenuta conservativa (massimo 30 giorni).

Scala spaziale adattiva. Ogni campione è centrato sull'epicentro del mainshock e normalizzato spazialmente utilizzando il raggio di Gardner [6] come unità di scala adattiva alla magnitudo, producendo un sistema di coordinate in cui tutti gli eventi del dataset sono comparabili indipendentemente dalla dimensione del terremoto.

Definizione del target. Il target di previsione è una mappa binaria su griglia 256×256 pixel allineata all'estensione di Gardner. Ogni aftershock verificatosi dopo la data di fine dell'interferogramma viene marcato con un disco di raggio pari all'incertezza nominale di localizzazione del catalogo ($\sim$2 km). Questa formulazione è robusta all'incompletezza del catalogo e non richiede assunzioni sulla forma della distribuzione spaziale degli aftershock.

Configurazioni di input a confronto. Vengono confrontate principalemente tre configurazioni di input, con l'obiettivo di isolare il contributo informativo di ciascuna sorgente di dati: (*i*) solo InSAR e coerenza; (*ii*) mappa di densità kernel dei foreshock e mappa delle faglie attive del dataset GEM [7]; (*iii*) combinazione delle sorgenti precedenti con metadati sismologici — magnitudo, baseline temporale InSAR, coerenza media e conteggio dei foreshock — iniettati tramite strati FiLM (*Feature-wise Linear Modulation*). La valutazione principale adotta uno split temporale cronologico, che garantisce che il modello non osservi eventi futuri durante l'addestramento; viene inoltre esplorata una variante con split spaziale, in cui la separazione avviene per celle geografiche mantenendo l'ordinamento cronologico all'interno di ciascuna partizione.

Architettura e addestramento. Il modello è una U-Net [8] con encoder ResNet [9] e un relativo ramo di decodifica. L'integrazione dei metadati sismologici avviene attraverso un meccanismo di condizionamento FiLM (Feature-wise Linear Modulation), applicato prima degli strati convoluzionali finali necessari per adattare l'output alla dimensionalità del target. L'addestramento minimizza una funzione di perdita basata sulla log-verosimiglianza negativa di Poisson, motivata dall'estremo squilibrio tra pixel positivi e negativi su una griglia 256×256. La baseline statistica di confronto è un kernel spaziale ETAS isotropo [1] calibrato sul training set.

1.5 Struttura della tesi

Il resto della tesi è organizzato come segue. Il Capitolo 2 presenta il quadro scientifico di riferimento: i fondamenti sismologici delle sequenze di aftershock, i principi dell'InSAR e i principali lavori precedenti sulla previsione sismica *data-driven*. Il Capitolo 3 descrive la pipeline di dati: acquisizione del catalogo, declustering mainshock–aftershock, selezione dei frame LiCSAR, download degli interferogrammi, costruzione del dataset, l'architettura del modello, la funzione di perdita e la procedura di addestramento. Il Capitolo 4 riporta i risultati sperimentali, inclusi studi di ablazione sulle configurazioni di input e sui metadati, il confronto con la baseline ETAS. Il Capitolo 5 conclude la tesi, discute i limiti, i casi di fallimento e le direzioni per il lavoro futuro.

Chapter 2

Contesto scientifico e tecnologico

In questo capitolo vengono presentati i fondamenti necessari alla comprensione del lavoro svolto. La trattazione è organizzata in tre blocchi tematici: la sismologia delle sequenze di aftershock e i modelli statistici e fisici classici per la loro previsione (Sezione 2.1); i principi dell'Interferometria SAR e la catena di produzione dei dati utilizzati (Sezione 2.2); le architetture di *deep learning* impiegate nel modello proposto (Sezione 2.3).

2.1 Sismologia delle sequenze di aftershock

2.1.1 Leggi empiriche della sismicità

La sismicità terrestre è governata da un insieme di leggi empiriche robuste, osservate indipendentemente in quasi tutte le regioni tettonicamente attive del globo.

Legge di Gutenberg–Richter. La relazione tra frequenza e magnitudo degli eventi sismici fu formulata da Gutenberg and Richter [10] su dati della California meridionale e successivamente verificata su scala globale. Essa afferma che il numero di terremoti N con magnitudo $\geq M$ segue una distribuzione a legge di potenza:

$$\log_{10} N(M) = a - b\,M, \tag{2.1}$$

dove a quantifica il livello di attività sismica di una regione e b descrive la distribuzione relativa di eventi grandi e piccoli. Il parametro b è prossimo a 1 nella maggior parte dei contesti tettonici, il che implica che per ogni evento di magnitudo M se ne verificano circa 10 di magnitudo $M - 1$. Questa invarianza di scala ha implicazioni dirette sul dataset costruito in questa tesi: la scelta di includere

tutti gli aftershock indipendentemente dalla magnitudo è coerente con la legge di Gutenberg–Richter, poiché il numero di eventi è dominato dai terremoti più piccoli.

Legge di Omori–Utsu. La frequenza degli aftershock decade nel tempo dopo il mainshock secondo una legge che fu osservata per la prima volta da Omori nel 1894 e formalizzata nella versione moderna da Utsu [11]:

$$n(t) = \frac{K}{(t+c)^p}, \tag{2.2}$$

dove $n(t)$ è il tasso di aftershock al tempo t dopo il mainshock, K è la produttività della sequenza, c è una costante che regola il comportamento nella fase iniziale e p è l'esponente di decadimento, tipicamente compreso tra 0.9 e 1.5. La legge di Omori–Utsu descrive con grande precisione il comportamento temporale delle sequenze, ma non fornisce alcuna informazione sulla distribuzione spaziale degli eventi.

2.1.2 Finestre spazio-temporali di Gardner–Knopoff

La definizione operativa di mainshock e aftershock richiede un criterio per separare gli eventi correlati da quelli indipendenti. Gardner and Knopoff [6] proposero finestre spaziali e temporali empiriche in funzione della magnitudo del mainshock. Il raggio spaziale è definito come:

$$R = 10^{0.1238\,M + 0.983} \quad \text{km}, \tag{2.3}$$

mentre la finestra temporale per gli aftershock dipende dalla magnitudo con una formula a tratti:

$$t(M) = \begin{cases} 10^{0.5409\,M - 0.547} & \text{se } M \geq 6.5, \\ 10^{0.032\,M + 2.7389} & \text{se } M < 6.5, \end{cases} \quad \text{giorni.} \tag{2.4}$$

Questa relazione è stata adottata come unità di scala adattiva in questa tesi: ogni campione del dataset è centrato sull'epicentro del mainshock e normalizzato spazialmente utilizzando R come riferimento, in modo che tutti gli eventi siano comparabili indipendentemente dalla magnitudo.

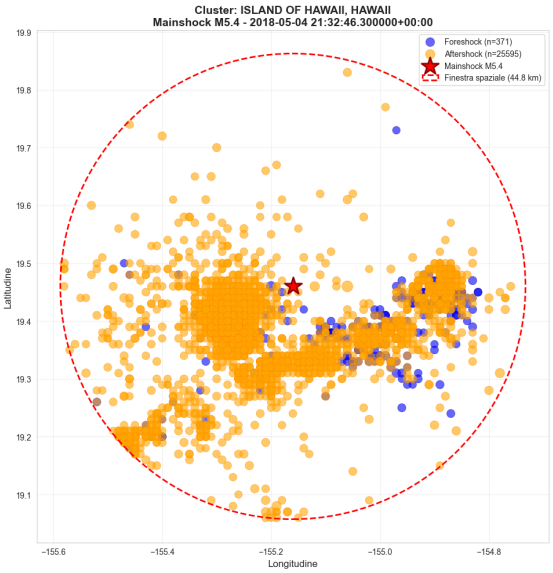


Figure 2.1: Cluster di terremoti dopo il mainshock.

2.1.3 Il modello ETAS e la baseline spaziale isotropa

Il modello ETAS (*Epidemic-Type Aftershock Sequence*), introdotto da Ogata [1], è il framework statistico di riferimento per la modellazione delle sequenze sismiche. Nella sua formulazione completa, il tasso di sismicità al tempo t è descritto da un processo di punto auto-eccitante in cui ogni evento può triggerarne altri:

$$\lambda(t) = \mu + \sum_{t_i<t} K\, e^{\alpha(M_i - M_0)} \left(1 + \frac{t - t_i}{c}\right)^{-p}, \tag{2.5}$$

dove μ è il tasso di background, K è la produttività, α controlla la dipendenza dalla magnitudo e i parametri c, p regolano il decadimento temporale alla Omori–Utsu.

L'estensione spaziale del modello adotta un kernel isotropo per la distribuzione degli aftershock attorno a ciascun evento:

$$K(r) = \frac{q-1}{\pi\,\sigma} \left(1 + \frac{r^2}{\sigma}\right)^{-q}, \qquad \sigma = d\, e^{\alpha M}, \tag{2.6}$$

dove r è la distanza dall'epicentro e i parametri d, α, q vengono calibrati sui dati.

In questa tesi il modello ETAS viene utilizzato in forma semplificata come **baseline spaziale isotropa**: si considera esclusivamente il kernel spaziale (2.6) centrato sull'epicentro del mainshock, con i parametri calibrati sul training set tramite massima verosimiglianza. La componente temporale e il meccanismo auto-eccitante non vengono impiegati, poiché il task affrontato è la previsione della distribuzione

spaziale degli aftershock nell'intera finestra temporale successiva all'evento, non la stima del tasso istantaneo di sismicità. Questa baseline rappresenta il riferimento statistico classico con il quale confrontare il modello proposto: essendo isotropa e dipendente solo dalla magnitudo del mainshock, cattura la distribuzione attesa in assenza di qualsiasi informazione osservativa sulla geometria della rottura.

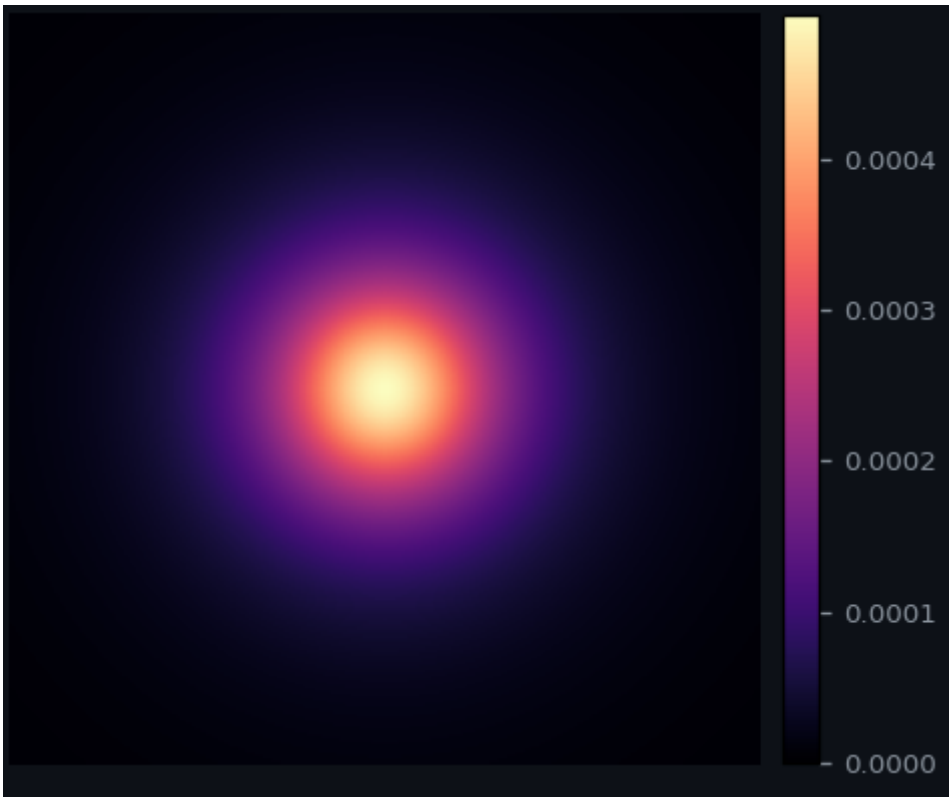


Figure 2.2: Distribuzione di probabilità degli aftershock secondo il modello ETAS.

2.1.4 Trasferimento di stress di Coulomb

La spiegazione fisica più consolidata per la localizzazione degli aftershock è il trasferimento di stress di Coulomb [12]. La variazione dello stress di Coulomb su una faglia ricevente è definita come:

$$\Delta\text{CFS} = \Delta\tau + \mu'\Delta\sigma_n, \tag{2.7}$$

dove $\Delta\tau$ è la variazione di sforzo di taglio nella direzione di slittamento della faglia, $\Delta\sigma_n$ è la variazione di sforzo normale (positivo per apertura) e μ' è il coefficiente di attrito apparente. Un valore $\Delta\text{CFS} > 0$ indica che la faglia è stata avvicinata alla rottura dal mainshock, e le regioni con ΔCFS positivo correlano statisticamente con la localizzazione degli aftershock.

Tuttavia, il calcolo di ΔCFS richiede la conoscenza dell'orientazione della faglia ricevente, del modello di slip del mainshock e dei parametri elastici della crosta: tutte quantità che nelle prime ore successive all'evento sono note solo approssimativamente o non ancora disponibili. DeVries et al. [2] hanno mostrato che un modello di deep learning addestrato su feature di stress può superare ΔCFS nella previsione degli aftershock, ma anch'esso si basa su campi di stress sintetici piuttosto che su osservazioni dirette. Questa tesi si propone di verificare se il campo di deformazione cosismica osservato da InSAR, che integra implicitamente le stesse informazioni fisiche, possa rappresentare un'alternativa operativamente accessibile.

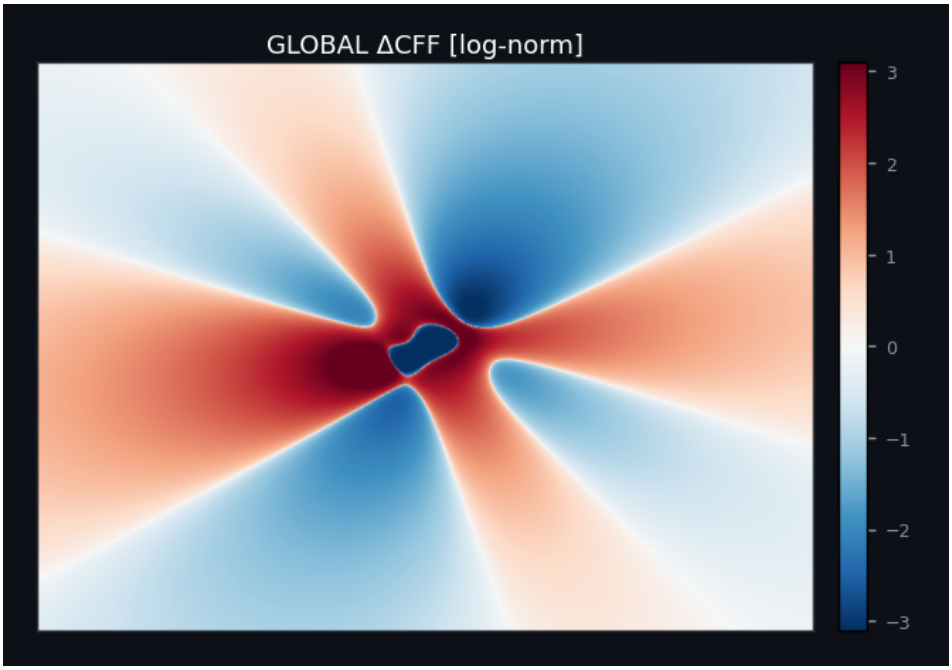


Figure 2.3: Campo di stress di Coulomb dopo il mainshock.

2.1.5 Stima di densità kernel per la sismicità spaziale

La stima di densità kernel (*Kernel Density Estimation*, KDE) è un metodo non parametrico per stimare la funzione di densità di probabilità di una variabile aleatoria continua a partire da un insieme discreto di osservazioni [13]. Applicata alla sismicità spaziale, la KDE trasforma un catalogo di epicentri puntiformi in una mappa continua di densità definita su una griglia regolare, rendendo confrontabili tra loro sequenze con numero di eventi molto diverso.

Formulazione base. Dato un insieme di n epicentri con coordinate $\{(x_i, y_i)\}_{i=1}^{n}$, la stima KDE nel punto $\mathbf{x}$ è:

$$\hat{f}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\sigma_i}(\mathbf{x} - \mathbf{x}_i)\,, \tag{2.8}$$

dove K_{σ_i} è il kernel associato all'i-esimo evento con parametro di banda σ_i. La scelta più comune in ambito geofisico è il kernel gaussiano isotropo:

$$K_\sigma(\mathbf{u}) = \frac{1}{2\pi\sigma^2} \exp\!\left(-\frac{\|\mathbf{u}\|^2}{2\sigma^2}\right), \tag{2.9}$$

dove $\|\mathbf{u}\|$ è la distanza fisica in chilometri tra il punto di valutazione e l'epicentro, e σ controlla il grado di smoothing spaziale.

Pesatura per magnitudo. La formulazione base assegna peso uguale a tutti gli eventi, indipendentemente dalla loro magnitudo. Nelle applicazioni sismologiche è tuttavia comune introdurre una pesatura che rifletta l'energia rilasciata: eventi di magnitudo maggiore contribuiscono con kernel di ampiezza maggiore, in accordo con la loro influenza sismogenica più estesa. La mappa pesata assume la forma:

$$\hat{f}(\mathbf{x}) = \frac{1}{W} \sum_{i=1}^{n} w_i\, K_{\sigma_i}(\mathbf{x} - \mathbf{x}_i)\,, \quad W = \sum_{i=1}^{n} w_i, \tag{2.10}$$

dove $w_i \geq 0$ è il peso associato all'i-esimo evento.

Banda adattiva. Il parametro di banda σ_i può essere fisso (uguale per tutti gli eventi) oppure adattivo, variando da evento a evento in funzione della magnitudo o della densità locale del catalogo. In sismologia, la banda adattiva è motivata fisicamente: la dimensione della zona di rottura cresce con la magnitudo, e pertanto l'incertezza spaziale associata a un evento grande è intrinsecamente maggiore di quella di un evento piccolo. Un approccio comune consiste nell'impostare σ_i come combinazione in quadratura dell'incertezza di localizzazione del catalogo σ_{loc} e di un termine di scala $\sigma_{\text{rup}}(M_i)$ dipendente dalla magnitudo:

$$\sigma_i = \sqrt{\sigma_{\text{loc}}^2 + \sigma_{\text{rup}}^2(M_i)}, \tag{2.11}$$

dove $\sigma_{\text{rup}}(M_i)$ cresce con la magnitudo secondo una relazione empirica calibrata sul raggio di rottura atteso.

Calcolo delle distanze fisiche. Poiché gli epicentri sono espressi in coordinate geografiche (latitudine e longitudine), il calcolo della distanza $\|\mathbf{u}\|$ richiede la conversione in un sistema metrico locale. Per aree di estensione limitata (dell'ordine di alcune decine di chilometri), la proiezione equirettangolare locale fornisce un'approssimazione accurata:

$$\Delta y_{\text{km}} = (\varphi - \varphi_i) \cdot 111.132, \tag{2.12}$$

$$\Delta x_{\text{km}} = (\lambda - \lambda_i) \cdot 111.132 \cdot \cos(\varphi_i), \tag{2.13}$$

dove φ e λ sono latitudine e longitudine del punto di valutazione, φ_i e λ_i quelle dell'epicentro, e il fattore $\cos(\varphi_i)$ corregge la contrazione metrica della longitudine alle latitudini più elevate. La distanza fisica è quindi

$$\|\mathbf{u}\| = \sqrt{\Delta x_{\text{km}}^2 + \Delta y_{\text{km}}^2}, \tag{2.14}$$

Nota che il fattore $\cos(\varphi_i)$ viene valutato alla latitudine *dell'epicentro* φ_i, non a quella del centro della griglia. Questa scelta è geometricamente più corretta: la longitudine si accorcia in funzione della latitudine del punto di riferimento, che in questo caso è l'epicentro da cui si misurano le distanze, non il centro della mappa.

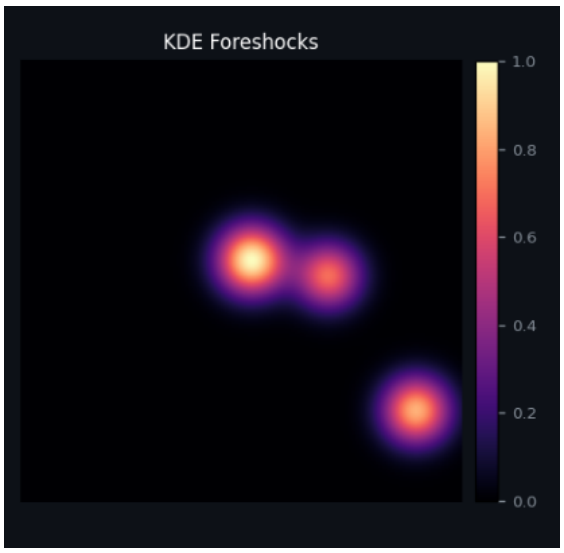


Figure 2.4: Esempio di mappa KDE costruita a partire da epicentri di foreshock.

2.2 Interferometria SAR per la sismologia

2.2.1 Principi dell'InSAR

Il Synthetic Aperture Radar (SAR) è un sensore a microonde che opera indipendentemente dalle condizioni di illuminazione solare e in presenza di nuvole. Il

satellite emette impulsi radar e registra la fase e l'ampiezza del segnale retrodiffuso dalla superficie terrestre. L'interferometria SAR (InSAR) sfrutta la coerenza di fase tra due acquisizioni dello stesso sensore per misurare variazioni di distanza superficie–satellite con precisione centimetrica o millimetrica.

L'interferogramma differenziale (*D-InSAR*) si ottiene calcolando la differenza di fase $\Delta\phi$ tra due immagini SAR acquisite prima e dopo un evento deformativo. La relazione tra la differenza di fase e lo spostamento lungo la linea di vista (*Line of Sight*, LOS) è:

$$\Delta\phi = \frac{4\pi}{\lambda} d_{\mathrm{LOS}} + \phi_{\mathrm{topo}} + \phi_{\mathrm{atm}} + \phi_n, \tag{2.15}$$

dove λ è la lunghezza d'onda del radar, d_{LOS} è la componente LOS del vettore di spostamento, ϕ_{topo} è la componente topografica (rimovibile con un DEM), ϕ_{atm} è il ritardo atmosferico e ϕ_n è il rumore. Per un terremoto superficiale di magnitudo elevata, il termine d_{LOS} è dominante e l'interferogramma cosismico fornisce una mappa diretta della deformazione del suolo associata all'evento.

La prima misurazione geodetica della deformazione cosismica tramite InSAR fu effettuata da Massonnet et al. [3] per il terremoto di Landers del 1992 (M_{w} 7.3), utilizzando dati ERS-1. Da allora, l'InSAR è diventato lo strumento standard per la caratterizzazione geodetica dei terremoti su scala globale.

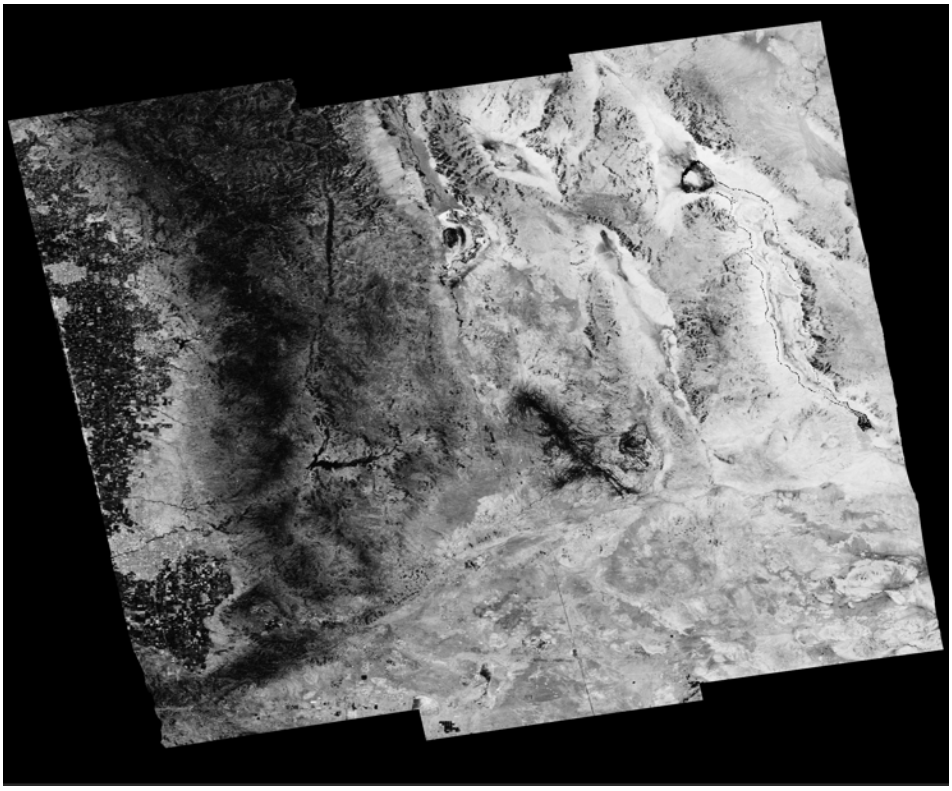


Figure 2.5: Interferogramma InSAR del terremoto di Ridgecrest 2019. Fonte: [4]

2.2.2 Coerenza interferometrica

La coerenza $\gamma \in [0, 1]$ quantifica la stabilità della fase dell'interferogramma:

$$\gamma = \frac{|\langle s_1 s_2^* \rangle|}{\sqrt{\langle |s_1|^2 \rangle \langle |s_2|^2 \rangle}}, \tag{2.16}$$

dove s_1 e s_2 sono i segnali SAR complessi delle due acquisizioni. Valori di coerenza prossimi a 1 indicano che la fase è stabile e lo spostamento misurabile; valori bassi indicano *decorrelazione*, dovuta a cambiamenti della superficie (vegetazione, neve, terremoti forti) o a un'eccessiva baseline temporale. Nel contesto di questa tesi la mappa di coerenza viene utilizzata come canale di input aggiuntivo, poiché fornisce informazioni sulla qualità locale della misura di spostamento e può indirettamente codificare la gravità della deformazione superficiale (zone di alta deformazione tendono a essere decorrelate).

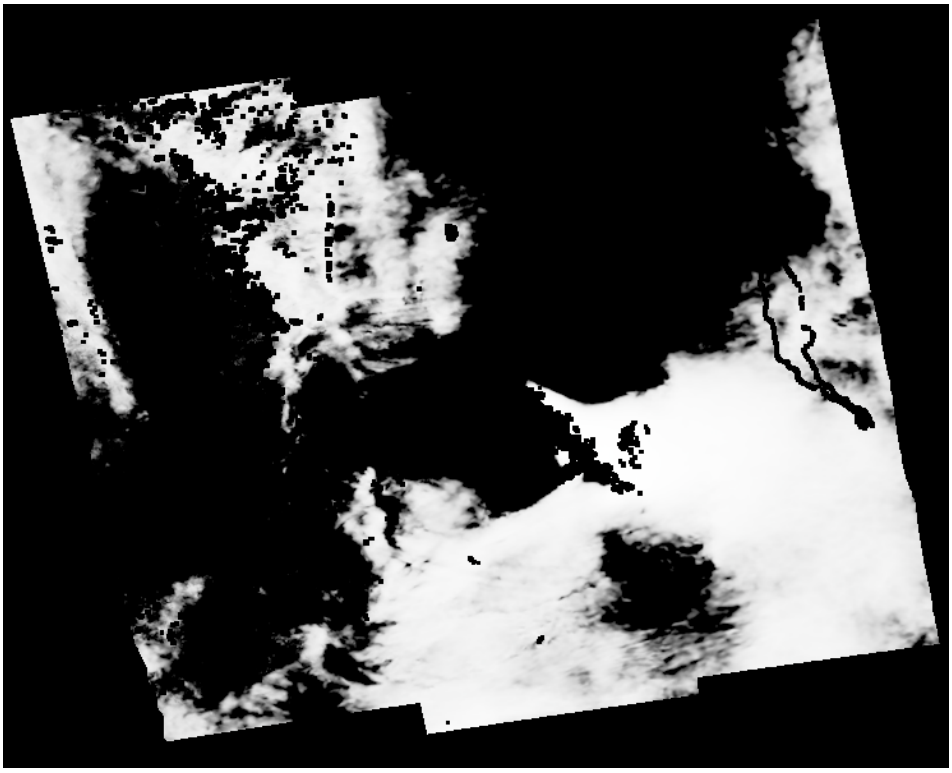


Figure 2.6: Coherence del interferogramma del terremoto di Ridgecrest 2019. Fonte: [4]

2.2.3 LiCSAR e il portale COMET-LiCS

Il sistema LiCSAR (*Looking Into Continents from Space with Synthetic Aperture Radar*) [4] è un processore automatico sviluppato dall'Università di Leeds e dal COMET (*Centre for the Observation and Modelling of Earthquakes, Volcanoes and Tectonics*) per la produzione sistematica di interferogrammi Sentinel-1 su scala globale.

Il sistema copre tutte le principali cinture sismiche del pianeta e produce automaticamente, per ciascuna coppia di acquisizioni compatibili:

- l'interferogramma in fase avvolta in radianti;
- l'interferogramma svolta (*unwrapped*), che rimuove l'ambiguità di 2π e fornisce lo spostamento LOS in centimetri;
- la mappa di coerenza (`.cc.tif`) associata.

Tutti i prodotti sono geocodificati su una griglia regolare a risoluzione 0.001°

(≈90 m), compatibile con la risoluzione dei dati Sentinel-1 in modalità IW (*Interferometric Wide Swath*). I dati sono accessibili liberamente attraverso il portale COMET-LiCS (`https://comet.nerc.ac.uk/COMET-LiCS-portal/`).

Il tempo di rivisita di Sentinel-1 varia tra 6 e 12 giorni secondo la latitudine e la disponibilità di acquisizioni congiunte delle due unità del satellite (Sentinel-1A e 1B). Questo intervallo è sufficientemente breve da consentire la generazione di interferogrammi cosismici nelle ore o nei giorni successivi a terremoti di $M_{\mathrm{w}} \geq 5$, rendendo il sistema operativamente utile per applicazioni di risposta rapida.

In questa tesi i dati LiCSAR vengono utilizzati come segue: per ciascun mainshock selezionato, viene identificato il frame LiCSAR che copre l'epicentro con la migliore geometria di acquisizione; viene quindi scaricato il miglior interferogramma disponibile, inteso come quello con la baseline temporale più breve compatibile con la copertura dell'evento, e la corrispondente mappa di coerenza.

2.3 Apprendimento profondo per la segmentazione spaziale

2.3.1 Deep learning

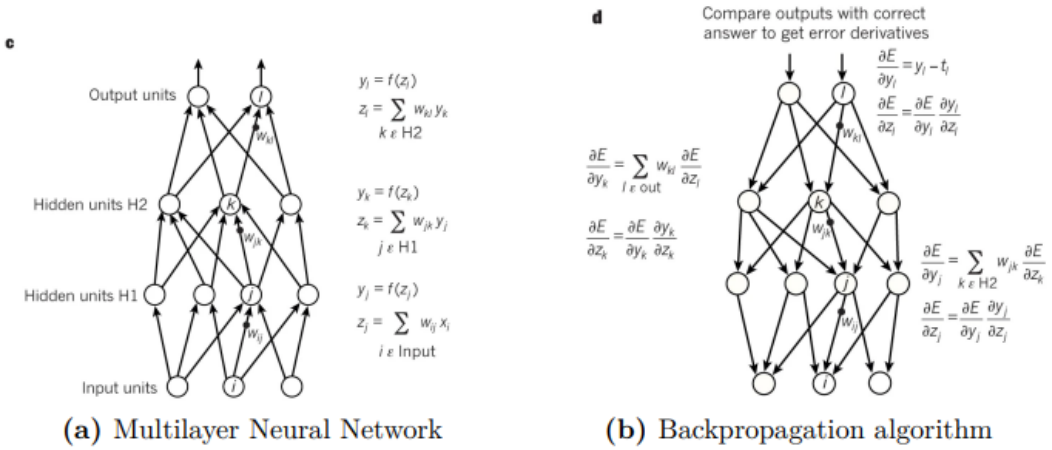
L'apprendimento profondo (*deep learning*, DL) è una branca dell'apprendimento automatico (*machine learning*, ML) fondata sull'utilizzo di reti neurali artificiali con molteplici strati di elaborazione [14]. L'idea portante è che una funzione di mappatura complessa — ad esempio quella che lega osservazioni geofisiche alla distribuzione spaziale degli aftershock — possa essere approssimata componendo una sequenza di trasformazioni più semplici, ciascuna applicata alla rappresentazione prodotta dallo strato precedente.

Ogni strato è costituito da unità computazionali elementari che applicano una trasformazione parametrica seguita da una non linearità; la composizione di molti di questi strati consente alla rete di apprendere funzioni arbitrariamente complesse [15], restando al contempo sensibile a variazioni rilevanti del segnale e robusta a fluttuazioni irrilevanti dell'input.

I parametri della rete vengono ottimizzati minimizzando una funzione di perdita su un insieme di esempi di addestramento tramite l'algoritmo di *backpropagation* [16], che calcola il gradiente della perdita rispetto a tutti i parametri applicando ricorsivamente la regola della catena lungo la sequenza degli strati. L'aggiornamento iterativo dei parametri avviene attraverso la discesa del gradiente stocastica (SGD) o sue varianti adattive come Adam [17], che modula il passo di aggiornamento in funzione della storia dei gradienti passati.

Il vantaggio fondamentale del deep learning rispetto ai metodi classici di machine learning è la capacità di apprendere automaticamente le rappresentazioni rilevanti

direttamente dai dati grezzi, senza richiedere l'estrazione manuale di feature (*feature engineering*) [14]. Questo è particolarmente prezioso in domini come la sismologia, dove le relazioni tra grandezze osservabili — deformazione cosismica, coerenza interferometrica, distribuzione dei foreshock — e la distribuzione spaziale degli aftershock sono complesse, non lineari e ancora oggetto di ricerca attiva. Tra le architetture di deep learning, le reti neurali convoluzionali (CNN) si sono affermate come il paradigma di riferimento per l'analisi di dati strutturati su griglia, quali immagini e mappe geofisiche, e vengono descritte in dettaglio nella sezione seguente.



(a) Multilayer Neural Network **(b)** Backpropagation algorithm

Figure 2.7: Esempio di rete neurale a più strati e backpropagation. Fonte: [14].

2.3.2 Reti neurali convoluzionali

Le reti neurali convoluzionali (CNN, *Convolutional Neural Networks*) [14] sono la classe di modelli di riferimento per l'elaborazione di dati strutturati su griglia, quali immagini e mappe geofisiche. A differenza delle reti fully-connected, le CNN sfruttano la stazionarietà spaziale dei pattern di interesse attraverso tre meccanismi fondamentali:

- **Condivisione dei pesi**: il medesimo filtro convoluzionale viene applicato in tutte le posizioni spaziali, riducendo drasticamente il numero di parametri rispetto a una rete densa e introducendo invarianza traslazionale.
- **Ricettività locale**: ogni neurone è connesso solo a una regione locale dell'input (*campo recettivo*), permettendo l'estrazione di feature locali come bordi, texture e strutture.
- **Pooling**: operazioni di aggregazione spaziale riducono progressivamente la risoluzione delle mappe di feature, aumentando il campo recettivo effettivo dei livelli successivi e introducendo robustezza a piccole traslazioni.

L'impilamento di più strati convoluzionali e di pooling produce una gerarchia di rappresentazioni: i livelli bassi catturano feature semplici (bordi, gradienti), mentre i livelli profondi combinano queste in strutture sempre più astratte e semanticamente significative.

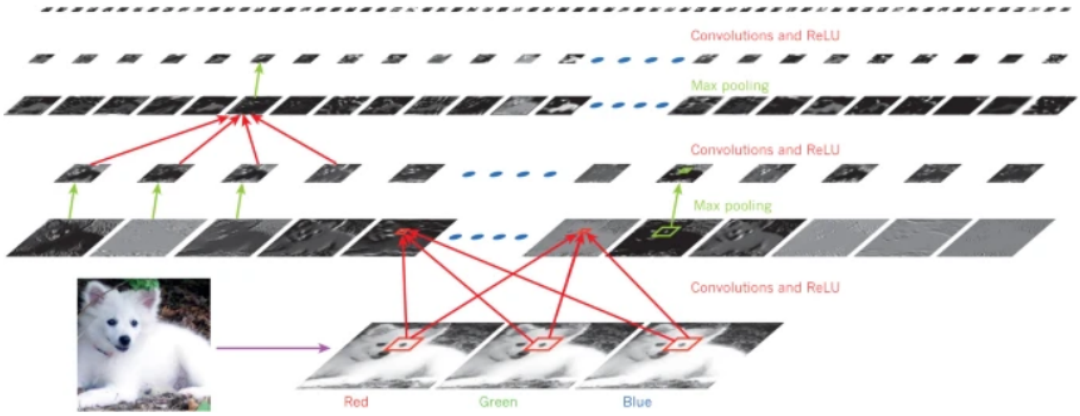


Figure 2.8: Funzionamento interno di una rete neurale convoluzionale. Fonte: [14].

2.3.3 Residual Networks (ResNet)

L'addestramento di reti molto profonde soffre tradizionalmente del problema della scomparsa del gradiente (*vanishing gradient*): il segnale di errore si attenua progressivamente durante la retropropagazione, rendendo inefficace l'aggiornamento dei pesi negli strati più vicini all'input. He et al. [9] hanno risolto questo problema introducendo le connessioni residuali (*skip connections*):

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}, \tag{2.17}$$

dove $\mathbf{x}$ è l'input del blocco, $\mathcal{F}$ è la funzione appresa (tipicamente due o tre strati convoluzionali con normalizzazione di batch e ReLU) e $\mathbf{y}$ è l'output. L'addizione diretta di $\mathbf{x}$ all'output del blocco garantisce che il gradiente possa fluire direttamente attraverso la connessione additiva, senza attraversare i parametri apprendibili, facilitando enormemente l'ottimizzazione di reti con decine o centinaia di strati.

Le architetture ResNet (ResNet-18, ResNet-34, ResNet-50, ecc.) hanno dominato i benchmark di classificazione di immagini dalla loro introduzione a CVPR 2016 e sono diventate l'encoder di riferimento per task di visione computazionale, inclusa la segmentazione semantica. In questa tesi ResNet viene utilizzato come encoder dell'architettura U-Net, sfruttando la sua capacità di estrarre rappresentazioni ricche dall'input multi-canale (InSAR, coerenza, mappa di faglie o foreshock).

2.3.4 U-Net per la segmentazione pixel-wise

La U-Net [8] è un'architettura encoder–decoder progettata per la segmentazione semantica pixel-wise, ossia per la predizione di un'etichetta o di un valore per ogni pixel dell'immagine di input. L'architettura è caratterizzata da una struttura simmetrica a "U":

- **Encoder (percorso di contrazione)**: una sequenza di blocchi convoluzionali con downsampling che comprimono progressivamente la risoluzione spaziale aumentando il numero di canali. A ogni livello, le mappe di feature catturano rappresentazioni sempre più astratte del contenuto dell'immagine.

- **Decoder (percorso di espansione)**: una sequenza speculare di blocchi con upsampling (o convoluzione trasposta) che ripristinano progressivamente la risoluzione originale. A ogni livello del decoder vengono concatenate le mappe di feature del corrispondente livello dell'encoder tramite *skip connections*.

- **Skip connections**: il collegamento diretto tra encoder e decoder a ogni scala permette al decoder di accedere sia alle feature semantiche di alto livello (dall'uscita dell'encoder) sia alle feature spaziali a alta risoluzione (dai livelli bassi dell'encoder), risolvendo il problema della perdita di localizzazione tipica delle architetture puramente encoder–decoder.

L'output della U-Net è una mappa della stessa dimensione spaziale dell'input, con un canale per classe nella segmentazione multi-classe, o un singolo canale con attivazione sigmoide nel caso binario.

Nel contesto di questa tesi, il problema di previsione degli aftershock è formulato esattamente come un task di segmentazione binaria pixel-wise: l'input è una mappa 256×256 con i canali geofisici, e l'output atteso è una mappa binaria della stessa dimensione in cui i pixel a 1 indicano le regioni con presenza di aftershock. La U-Net è quindi l'architettura naturale per questo task, e la sua capacità di sfruttare informazioni a scale diverse è particolarmente rilevante dato che gli aftershock possono presentare sia strutture lineari (lungo la traccia della faglia) sia distribuzioni più diffuse.

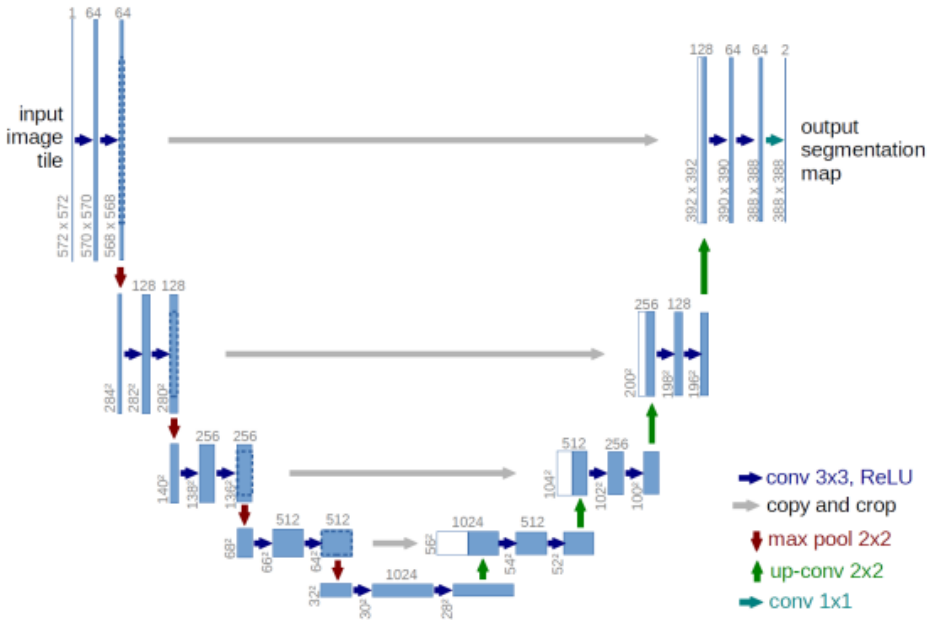


Figure 2.9: Architettura della U-Net. Fonte: [8]

2.3.5 Condizionamento su metadati scalari: FiLM

Un requisito specifico del modello proposto è la capacità di integrare informazioni scalari eterogenee — magnitudo del mainshock, baseline temporale dell'interferogramma, coerenza media e conteggio dei foreshock — con la rappresentazione spaziale estratta dalla U-Net. La concatenazione diretta di scalari alle mappe di feature non è una soluzione ottimale, poiché non consente al modello di modificare dinamicamente il proprio comportamento in funzione del contesto dell'evento.

Perez et al. [18] hanno proposto il meccanismo FiLM (*Feature-wise Linear Modulation*), che condiziona una mappa di feature tramite una trasformazione affine per-canale dipendente da un vettore di contesto esterno. Nella formulazione originale, dati i parametri γ_c e β_c prodotti da una rete generatrice in funzione del vettore di metadati $\mathbf{z}$, la trasformazione è:

$$\tilde{F}_c = \gamma_c(\mathbf{z})\, F_c + \beta_c(\mathbf{z}). \tag{2.18}$$

In questa tesi viene adottata una variante residuale:

$$\tilde{F}_c = F_c \cdot \Big(1 + \gamma_c(\mathbf{z})\Big) + \beta_c(\mathbf{z}). \tag{2.19}$$

Con questa formulazione, quando $\gamma_c \to 0$ la mappa di feature viene trasmessa invariata (il termine moltiplicativo tende a 1), mentre la formulazione originale (2.18) con $\gamma_c \to 0$ azzererebbe completamente il canale. Su dati geofisici con segnale

debole e distribuito in modo sparso — come le mappe di deformazione cosismica in prossimità di eventi di magnitudo moderata — questo comportamento residuale è preferibile: garantisce che i metadati scalari *modulino* le feature spaziali senza sopprimerle.

I parametri γ_c e β_c per tutti i C canali vengono prodotti congiuntamente da una rete fully-connected a due strati con attivazione ReLU applicata al vettore di metadati $\mathbf{z} \in \mathbb{R}^{d_m}$:

$$[\boldsymbol{\gamma};\, \boldsymbol{\beta}] = W_2\, \mathrm{ReLU}(W_1\, \mathbf{z}) \in \mathbb{R}^{2C}, \tag{2.20}$$

dove $W_1 \in \mathbb{R}^{d_h \times d_m}$ e $W_2 \in \mathbb{R}^{2C \times d_h}$ sono le matrici di peso della rete generatrice, e d_h è la dimensione del layer nascosto. Il vettore di output viene suddiviso in due blocchi di dimensione C, che forniscono rispettivamente i parametri di scaling e di shift per ciascun canale della mappa di feature.

2.3.6 Funzione di perdita di Poisson

La supervisione del modello avviene tramite una funzione di perdita basata sulla log-verosimiglianza negativa di un processo di punto spaziale di Poisson. Data una mappa di logit $\hat{\mathbf{y}} \in \mathbb{R}^{H \times W}$ e una mappa target binaria $\mathbf{k} \in \{0,1\}^{H \times W}$, l'intensità predetta per il pixel i è ottenuta tramite la funzione *softplus*:

$$\lambda_i = \log\!\left(1 + e^{\hat{y}_i}\right), \tag{2.21}$$

che garantisce $\lambda_i > 0$ senza imporre alcun limite superiore, a differenza della funzione sigmoide. Questa proprietà è fondamentale: l'intensità di un processo di Poisson è una grandezza non negativa ma non necessariamente limitata a 1, e vincolarne l'uscita a (0,1) introdurrebbe un'ipotesi non giustificata.

La funzione di perdita è la log-verosimiglianza negativa media del processo di Poisson su tutti i pixel della griglia:

$$\mathcal{L} = \frac{1}{HW} \sum_i \Big(\lambda_i - k_i\, \log(\lambda_i + \varepsilon) \Big), \tag{2.22}$$

dove ε è una costante di stabilità numerica.

Questa formulazione è motivata da tre considerazioni. In primo luogo, interpreta la mappa predetta come la mappa di intensità di un processo di punto spaziale di Poisson, che è il framework naturale per descrivere la distribuzione spaziale degli aftershock — gli stessi modelli statistici di riferimento come ETAS sono formulati in questo framework [1]. In secondo luogo, il termine $\sum_i \lambda_i$ nella loss penalizza *globalmente* la sovra-predizione dell'intensità totale: un modello che assegna alta intensità a molti pixel vedrà questo termine crescere anche in assenza di aftershock in quelle posizioni, inducendo la rete a essere selettiva nella localizzazione delle predizioni. In terzo luogo, la loss di Poisson tollera naturalmente lo squilibrio

estremo tra pixel positivi e negativi su una griglia 256×256, senza richiedere la specifica manuale di pesi per classe come nella Binary Cross-Entropy pesata.

Chapter 3

Metodologia

In questo capitolo viene descritta la pipeline end-to-end del sistema di apprendimento automatico, che integra la trasformazione dei cataloghi sismici grezzi e degli interferogrammi InSAR in un dataset strutturato, la definizione dell'architettura del modello, e la procedura di addestramento.

La trattazione segue l'intero flusso operativo del sistema: dalla definizione formale del problema e del dominio di apprendimento, al download e preprocessing dei dati, dal declustering delle sequenze sismiche alla costruzione delle mappe di input e di target, fino alla descrizione dell'architettura neurale, della funzione di perdita e della procedura di ottimizzazione.

Infine, vengono dettagliate le strategie di suddivisione dei dati in insiemi di addestramento, validazione e test, insieme alle metriche utilizzate per la valutazione delle prestazioni del modello.

3.1 Formulazione del problema

Il problema affrontato in questa tesi è formulato come un task di *segmentazione binaria pixel-wise*: dato un insieme di mappe geofisiche centrate sull'epicentro di un mainshock, il modello deve produrre una mappa della stessa risoluzione spaziale in cui ogni pixel indica se, nella regione corrispondente, si è verificato almeno un aftershock successivamente alla data di fine dell'interferogramma.

Formalmente, sia $\mathcal{X} \in \mathbb{R}^{C \times H \times W}$ la mappa di input con C canali e dimensioni spaziali $H \times W$, e sia $\mathbf{z} \in \mathbb{R}^{d_m}$ il vettore di metadati sismologici scalari associati all'evento. L'output atteso è una mappa binaria $\mathbf{y} \in \{0,1\}^{H \times W}$, in cui $y_{ij} = 1$ se nell'area corrispondente al pixel (i, j) si è verificato almeno un aftershock dopo la data di fine dell'interferogramma cosismico.

Questa formulazione è preferita a una stima di densità continua (ad esempio basata su KDE) per tre ragioni. In primo luogo, i cataloghi sismici globali sono

incompleti e caratterizzati da soglie di completezza variabili per regione e periodo: la mappa KDE degli aftershock è direttamente affetta da questa incompletezza, mentre il target binario — che si limita a marcare dove sono stati effettivamente registrati degli aftershock — è più robusto a eventi mancanti. In secondo luogo, la formulazione binaria è naturalmente compatibile con la funzione di perdita di Poisson descritta nel Capitolo 2.3, che interpreta i valori positivi come presenze di eventi in un processo di punto spaziale. In terzo luogo, le metriche di valutazione standard per la classificazione pixel-wise (AUC-ROC, F1) sono direttamente confrontabili con la baseline ETAS, facilitando la valutazione sperimentale.

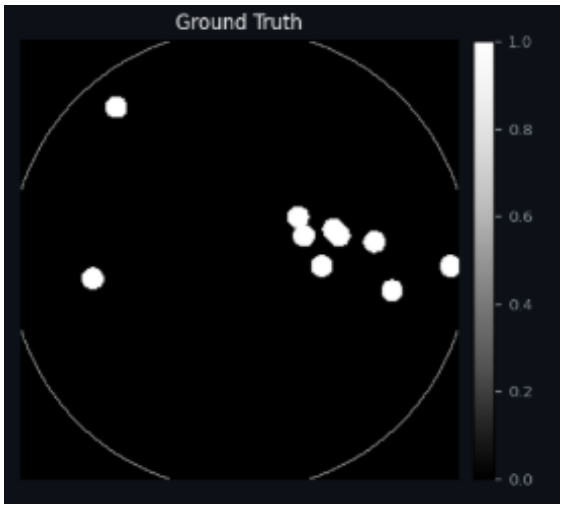


Figure 3.1: Esempio di target binario dove la circonferenza rappresenta la finestra spaziale del mainshock definita dalla formula di Gardner–Knopoff.

3.2 Acquisizione del catalogo sismico

Il catalogo sismico di riferimento è quello fornito dall'EMSC (*European Mediterranean Seismological Centre*) [19], che raccoglie dati sismici globali con copertura uniforme sulle principali cinture sismiche del pianeta. Il catalogo viene scaricato tramite le API pubbliche dell'EMSC e memorizzato in formato JSON per le elaborazioni successive.

La scelta dell'EMSC come sorgente primaria è motivata dalla buona copertura globale, dalla disponibilità di stime di magnitudo momento (M_w o M_{ww}) per gli eventi di interesse e dalla possibilità di accesso programmativo. Sono inclusi nel catalogo tutti gli eventi con magnitudo $M \geq 5.0$ registrati nel periodo di interesse, indipendentemente dalla profondità ipocentrale. La soglia $M \geq 5.0$ garantisce che ogni mainshock produca una deformazione cosismica sufficientemente ampia da essere rilevabile da Sentinel-1 e che la sequenza di aftershock abbia una numerosità

adeguata per la supervisione del modello.

Il download del catalogo EMSC utilizzato in questa tesi copre il periodo 2013-01-03–2025-12-31 (data di accesso 2026-03-15). Il declustering produce 6586 mainshock candidati; dopo i filtri di disponibilità InSAR e i vincoli di qualità descritti nelle sezioni successive, il dataset finale include 1033 mainshock, con date comprese tra il 2014-11-10 e il 2025-12-31.

3.3 Declustering e identificazione dei mainshock

Il declustering ha l'obiettivo di separare gli eventi indipendenti (mainshock) dagli eventi correlati (aftershock e foreshock). L'algoritmo adottato segue un approccio cronologico basato sulle finestre spazio-temporali di Gardner–Knopoff [6], già introdotte nella Sezione 2.1.2.

Logica di clustering. Gli eventi del catalogo vengono esaminati in ordine cronologico. Un evento è accettato come nuovo mainshock se e solo se la sua posizione non ricade nella finestra spazio-temporale di nessun mainshock precedentemente identificato. La distanza spaziale tra due eventi è calcolata mediante la formula di Haversine, che tiene conto della curvatura terrestre:

$$d = 2R_{\oplus} \arcsin\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cos\phi_2 \sin^2\left(\frac{\Delta\lambda}{2}\right)}, \tag{3.1}$$

dove $R_{\oplus} = 6371\,\text{km}$ è il raggio medio terrestre, ϕ_1 e ϕ_2 rappresentano le latitudini dei due eventi, mentre $\Delta\lambda$ è la differenza tra le loro longitudini.

Finestre spazio-temporali. Il raggio spaziale associato a ciascun mainshock segue la relazione empirica di Gardner–Knopoff:

$$R(M) = 10^{0.1238\,M+0.983} \qquad [\text{km}], \tag{3.2}$$

dove M rappresenta la magnitudo del mainshock.

La finestra temporale degli aftershock non è stata assunta direttamente dalla parametrizzazione classica di Gardner–Knopoff, ma calibrata empiricamente per ottenere un compromesso tra completezza delle sequenze e riduzione delle associazioni spurie. In particolare, la durata della finestra temporale è ottenuta mediante interpolazione lineare tra i seguenti punti di controllo:

$$(5.0{,}120), \quad (6.0{,}270), \quad (7.0{,}365), \quad (8.0{,}365),$$

dove il primo valore rappresenta la magnitudo e il secondo la durata della finestra in giorni.

Formalmente:

$$t_{\text{after}}(M) = \text{clip}(\text{interp}(M), 30, 365)\,, \tag{3.3}$$

dove $\text{interp}(M)$ indica l'interpolazione lineare dei punti precedenti e $\text{clip}(x, a, b)$ limita il valore di x all'intervallo $[a, b]$.

La finestra temporale dei foreshock è invece definita come:

$$t_{\text{before}}(M) = \min(30,\, \max(5,\, 0.1\, t_{\text{after}}(M)))\,, \tag{3.4}$$

così da mantenere una finestra conservativa, limitata a un massimo di 30 giorni precedenti il mainshock.

Separazione foreshock/aftershock rispetto alla data InSAR. Un aspetto critico del dataset è la separazione temporale tra eventi usati come contesto di input ed eventi usati come target. Piuttosto che usare la data del mainshock come spartiacque, si usa la data di fine dell'interferogramma cosismico selezionato (*InSAR end date*). In questo modo:

- gli eventi verificatisi prima della data di fine dell'interferogramma costituiscono il contesto osservativo disponibile al momento della previsione (foreshock e possibili repliche precoci);
- gli eventi verificatisi dopo tale data sono i target di previsione, ovvero gli aftershock che il modello deve localizzare.

Questa scelta garantisce la correttezza causale della pipeline: il modello non ha accesso ad alcuna informazione sugli aftershock futuri durante l'inferenza.

3.4 Selezione dei frame LiCSAR e acquisizione degli interferogrammi

Per ciascun mainshock identificato, si verifica la disponibilità di interferogrammi cosismici nel sistema LiCSAR [4]. I frame LiCSAR sono aree geografiche predefinite coperte sistematicamente da Sentinel-1; per ogni mainshock viene selezionato il frame che massimizza la copertura spaziale dell'epicentro e minimizza la distanza dal centro del frame al bordo dell'immagine, così da ridurre gli artefatti di bordo nei dati interferometrici.

Selezione del miglior interferogramma. Per ciascun frame selezionato, si scarica il miglior interferogramma disponibile, definito come quello con la baseline temporale più breve compatibile con la copertura dell'evento sismico. Una baseline

breve riduce il contributo dei fattori atmosferici e della decorrelazione temporale, aumentando la qualità della misura di spostamento.

I prodotti scaricati per ciascun campione sono:

- il file di fase svolta (`.unw.tif`), che fornisce lo spostamento lungo la linea di vista (LOS) in radianti su una griglia geocodificata a risoluzione 0.001ř (circa 90 m);
- la mappa di coerenza (`.cc.tif`), un valore nell'intervallo [0,255] che quantifica la stabilità della fase in ciascun pixel.

Entrambi i prodotti sono già geocodificati in coordinate geografiche (WGS-84) e non richiedono ulteriori operazioni di proiezione. Il gap temporale tra la data del mainshock e la data di fine dell'interferogramma viene registrato come metadato dell'evento (variabile `insar_mainshock_gap_days`) e utilizzato come feature di condizionamento del modello (Sezione 3.5.3).

3.5 Costruzione del dataset

3.5.1 Normalizzazione spaziale e estrazione delle patch

Poiché la dimensione caratteristica di una sequenza di aftershock scala con la magnitudo del mainshock, è necessaria una normalizzazione spaziale che renda comparabili eventi di diversa magnitudo. In questa tesi si adotta il raggio di Gardner come unità di scala adattiva: ogni campione è estratto come una patch quadrata di lato pari a $2R$ (dove R è il raggio di Gardner del mainshock, Equazione 3.2), centrata sull'epicentro.

La patch estratta viene ridimensionata a una griglia regolare di 256×256 pixel tramite interpolazione bilineare. La risoluzione al suolo effettiva dipende quindi dalla magnitudo dell'evento: per un mainshock di $M = 6.0$ il raggio di Gardner è circa $R \approx 53\,\text{km}$, e la patch copre $130 \times 130\,\text{km}^2$ con una risoluzione approssimativa di 500 m/pixel; per $M = 7.0$ il raggio è circa $R \approx 71\,\text{km}$ e la risoluzione scende a circa 1000 m/pixel. Nei casi in cui la patch supera il bordo dell'immagine InSAR, le regioni mancanti sono riempite con zeri.

Per ciascun campione vengono generati anche i sistemi di coordinate geografiche (ϕ, λ) per ogni pixel della griglia 256×256, allineati esattamente all'estensione fisica della patch. Questa griglia è utilizzata sia per la generazione delle mappe di target (Sezione 3.5.4) sia per la rasterizzazione della mappa delle faglie (Sezione 3.5.2).

3.5.2 Configurazioni di input

Vengono confrontate sistematicamente tre configurazioni di input, con l'obiettivo di isolare il contributo informativo di ciascuna sorgente di dati.

Configurazione (i): spostamento InSAR e coerenza

La configurazione di base utilizza due canali derivati direttamente dall'interferogramma LiCSAR:

- **Spostamento LOS** (canale 1): la fase svolta, proporzionale allo spostamento della superficie lungo la linea di vista del satellite. Il canale viene normalizzato con il metodo *percentile clipping*: i valori al di fuori dell'intervallo tra il 1° e il 99° percentile della distribuzione all'interno della patch vengono saturati, e il risultato viene scalato linearmente nell'intervallo $[0, 1]$. Questo approccio è più robusto della normalizzazione globale in presenza di outlier o di zone a bassa coerenza che producono artefatti di fase.

- **Coerenza interferometrica** (canale 2): la mappa di coerenza $\gamma \in [0, 255]$ viene riscala nell'intervallo $[0, 1]$ dividendo per 255. La coerenza fornisce informazioni sulla qualità locale della misura di spostamento.

Configurazione (ii): mappa dei foreshock e mappa delle faglie attive

Nella seconda configurazione i canali InSAR vengono sostituiti da due sorgenti di informazione complementari:

- **Mappa KDE dei foreshock** (canale 1): la distribuzione spaziale dei foreshock, rappresentata come mappa di densità kernel con pesatura quadratica per magnitudo e banda adattiva al raggio di rottura stimato. Questa mappa cattura le strutture sismogeniche attivate prima della rottura principale e può indicare le zone di massima suscettibilità sismica.

- **Mappa delle faglie attive GEM** (canale 2): una mappa binaria rasterizzata a partire dal dataset GEM Global Active Faults [7], che rappresenta le faglie attive note a scala globale. Le geometrie delle faglie vengono prima filtrate tramite un indice spaziale STR-tree sulla bounding box locale della patch e successivamente rasterizzate sulla griglia 256×256. I pixel attraversati da un segmento di faglia, eventualmente espansi mediante un piccolo spessore raster, assumono valore 1, mentre tutti gli altri pixel assumono valore 0.

Configurazione (iii): canali multipli con metadati FiLM

La terza configurazione combina ai canali InSAR (spostamento e coerenza) ed alle mappe dei foreshock e delle faglie i metadati sismologici scalari iniettati tramite strati FiLM (Sezione 2.3.5). Questa configurazione sfrutta la complementarità tra l'informazione spazialmente distribuita dei canali di immagine e l'informazione contestuale sui parametri dell'evento.

3.5.3 Metadati sismologici

Nel caso della terza configurazione (iii), il modello riceve un vettore di metadati scalari $\mathbf{z} \in \mathbb{R}^4$ che sintetizza le caratteristiche sismologiche e osservative dell'evento. I quattro componenti sono:

1. **Conteggio dei foreshock (log-normalizzato)**: $z_1 = \log(1+n_f)/\log(1+50)$, dove n_f è il numero di foreshock nella finestra temporale di 30 giorni precedenti il mainshock. La normalizzazione logaritmica comprime la distribuzione a coda pesante del numero di eventi.

2. **Magnitudo del mainshock (normalizzata)**: $z_2 = M/10$, ottenuta dividendo la magnitudo per il valore massimo atteso, così da mappare l'intervallo $[5, 10]$ nell'intervallo $[0.5, 1]$.

3. **Coerenza media della patch**: $z_3 = \bar{\gamma}$, calcolata come media della mappa di coerenza nella patch locale. Fornisce un indicatore globale della qualità dell'interferogramma e della distanza temporale tra l'acquisizione e l'evento.

4. **Gap temporale InSAR–mainshock (normalizzato)**: $z_4 = \min(\Delta t/30,\, 1)$, dove Δt è il numero di giorni tra la data del mainshock e la data di fine dell'interferogramma cosismico. Questo indicatore permette al modello di adattare la propria previsione in base alla freschezza dell'informazione InSAR rispetto all'evento sismico.

Tutti i componenti sono normalizzati nell'intervallo $[0, 1]$ prima dell'iniezione nel meccanismo FiLM, in modo da fornire input bilanciati alla rete generatrice dei parametri di modulazione.

3.5.4 Definizione del target binario

Il target di supervisione è una mappa binaria $\mathbf{k} \in \{0,1\}^{256\times256}$ allineata alla stessa griglia spaziale dell'input. La mappa viene costruita come segue: per ciascun aftershock verificatosi dopo la data di fine dell'interferogramma, tutti i pixel il cui centro dista meno di $r_{\text{unc}} = 2\,\text{km}$ dall'epicentro vengono impostati a 1. La scelta di $r_{\text{unc}} = 2\,\text{km}$ riflette l'incertezza nominale di localizzazione dei cataloghi globali (ISC, USGS).

Formalmente, per il pixel (i, j) con coordinate geografiche $(\phi_{ij}, \lambda_{ij})$:

$$k_{ij} = \mathbf{1}\left[\min_{s=1,\ldots,N_a} d(\phi_{ij}, \lambda_{ij},\, \phi_s, \lambda_s) \leq r_{\text{unc}}\right], \tag{3.5}$$

dove (ϕ_s, λ_s) sono le coordinate del s-esimo aftershock, N_a è il numero totale di aftershock nella finestra temporale e $d(\cdot,\cdot)$ è la distanza calcolata con la proiezione equirettangolare locale (Equazioni 2.12–2.13–2.14).

La rappresentazione a disco (piuttosto che a singolo pixel) serve a evitare una supervisione puntuale eccessivamente rigida: le posizioni degli epicentri nei cataloghi globali hanno un'incertezza intrinseca, e penalizzare il modello per predizioni corrette ma leggermente spostate rispetto alla posizione catalogata introdurrebbe un bias indesiderato. Questa scelta è coerente con l'interpretazione Poissoniana della funzione di perdita, in cui k_{ij} rappresenta la presenza osservata di almeno un evento nell'area del pixel.

Incompletezza del catalogo e impatto sulla supervisione. Un limite intrinseco di questa formulazione è l'incompletezza dei cataloghi sismici globali utilizzati per la costruzione del target. La magnitudo di completezza dei cataloghi EMSC e ISC varia regionalmente e può superare $M_c = 3.0$ in aree a bassa densità di stazioni, implicando che una frazione non trascurabile degli aftershock reali non sia registrata. Questo si riflette nella distribuzione fortemente asimmetrica del numero di aftershock per campione: la mediana è pari a 9 eventi, mentre la media raggiunge 129, indicando che pochi eventi con sequenze molto ricche dominano la statistica aggregata. In pratica, il modello viene penalizzato per predizioni spazialmente corrette in zone dove gli aftershock si sono verificati ma non sono stati catalogati, introducendo un rumore sistematico nel segnale di supervisione che non è rimovibile con il dataset attuale. Le metriche di valutazione riportate nel Capitolo 4 vanno quindi interpretate come stime per difetto della capacità discriminativa reale del modello.

3.5.5 Augmentazione dei dati

Per ridurre il rischio di *overfitting* e aumentare la variabilità del dataset, vengono applicati in fase di addestramento i seguenti trasformate di augmentazione, applicate in modo consistente a tutti i canali di input e alla mappa di target:

- **Flip casuale**: riflessione orizzontale e verticale, ciascuna con probabilità $p = 0.5$. Questa augmentazione è fisicamente motivata dall'assenza di una direzione privilegiata nella distribuzione degli aftershock rispetto all'asse della faglia, almeno in media sul dataset.
- **Rotazione casuale di 90°**: rotazione di un multiplo di 90° estratto uniformemente da $\{0ř, 90ř, 180ř, 270ř\}$, applicata con probabilità $p = 0.5$. Come per il flip, la rotazione discreta a multipli di 90° non altera la struttura geometrica della patch.
- **Rumore gaussiano**: perturbazione additiva $\mathcal{N}(0, \sigma^2)$ con $\sigma = 0.01$ applicata al canale di spostamento InSAR dopo la normalizzazione. Modella le fluttuazioni di fase residue non rimosse durante il processamento LiCSAR.

- **Scaling di intensità**: moltiplicazione del canale di spostamento per un fattore scalare estratto uniformemente dall'intervallo $[0.9, 1.1]$. Modella le incertezze sulla stima dell'ampiezza della deformazione dovute alla geometria di acquisizione e ai ritardi atmosferici residui.

In fase di validazione e test non viene applicata alcuna augmentazione, così da garantire la riproducibilità delle metriche di valutazione.

3.6 Suddivisione del dataset

Il dataset viene suddiviso in tre insiemi disgiunti: addestramento (*train*), validazione (*val*) e test, con proporzioni 70%, 15% e 15% rispettivamente. Il dataset finale contiene 1033 campioni (uno per mainshock), che con tali proporzioni producono 723/155/155 campioni in train/val/test.

Split temporale (configurazione principale). Nella configurazione principale, la suddivisione segue l'ordinamento cronologico dei mainshock: tutti i campioni vengono ordinati per data di origine del mainshock, e i primi 70% vengono assegnati al training set, il successivo 15% al validation set, e il restante 15% al test set. Questa scelta garantisce la correttezza causale della valutazione: il modello viene addestrato esclusivamente su eventi storicamente precedenti agli eventi di test, simulando fedelmente uno scenario operativo in cui il modello deve generalizzare a sequenze future. Lo split temporale è la scelta naturale per sistemi di previsione sismica, dove i cataloghi crescono nel tempo e non è possibile assumere scambiabilità degli esempi.

Split spaziale (configurazione alternativa). Come configurazione alternativa, viene implementato uno split spaziale in cui i campioni vengono raggruppati per celle geografiche di $20ř \times 20ř$ e i gruppi vengono assegnati ai tre insiemi rispettando le proporzioni target. Lo split spaziale verifica la capacità del modello di generalizzare a regioni sismiche non viste durante l'addestramento, separando la valutazione dalla dipendenza geografica. All'interno di ciascun insieme l'ordine dei campioni è mantenuto cronologico. Entrambe le modalità di split vengono esplorate negli esperimenti del Capitolo 4.

3.7 Architettura del modello

Il modello proposto è un'architettura encoder–decoder basata su U-Net con backbone ResNet-18, estesa con porte di attenzione nel decoder, un blocco di fusione

convoluzionale e un meccanismo FiLM per l'iniezione dei metadati sismologici. La Figura 3.2 illustra il percorso di elaborazione.

Dato l'input $\mathcal{X} \in \mathbb{R}^{C\times 256\times 256}$ (con C canali geofisici) e il vettore di metadati $\mathbf{z} \in \mathbb{R}^4$, il modello produce: una mappa di logit ad alta risoluzione $\hat{\mathbf{y}} \in \mathbb{R}^{1\times 256\times 256}$, usata come output principale per la funzione di perdita e l'inferenza.

Il percorso principale si articola in cinque fasi sequenziali: encoder ResNet-18 $\rightarrow$ decoder $\rightarrow$ condizionamento FiLM $\rightarrow$ testa di output.

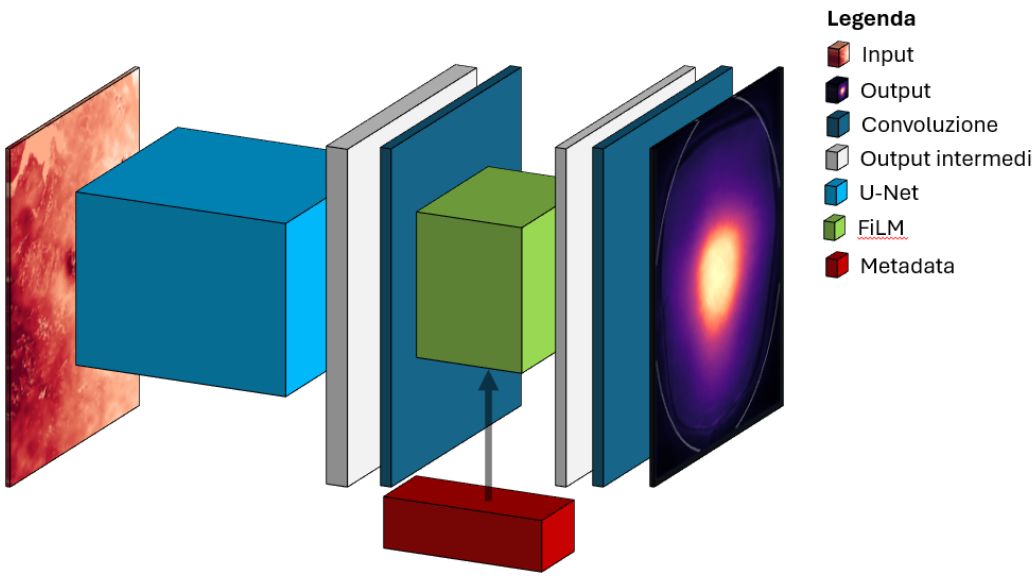


Figure 3.2: Architettura del modello proposto.

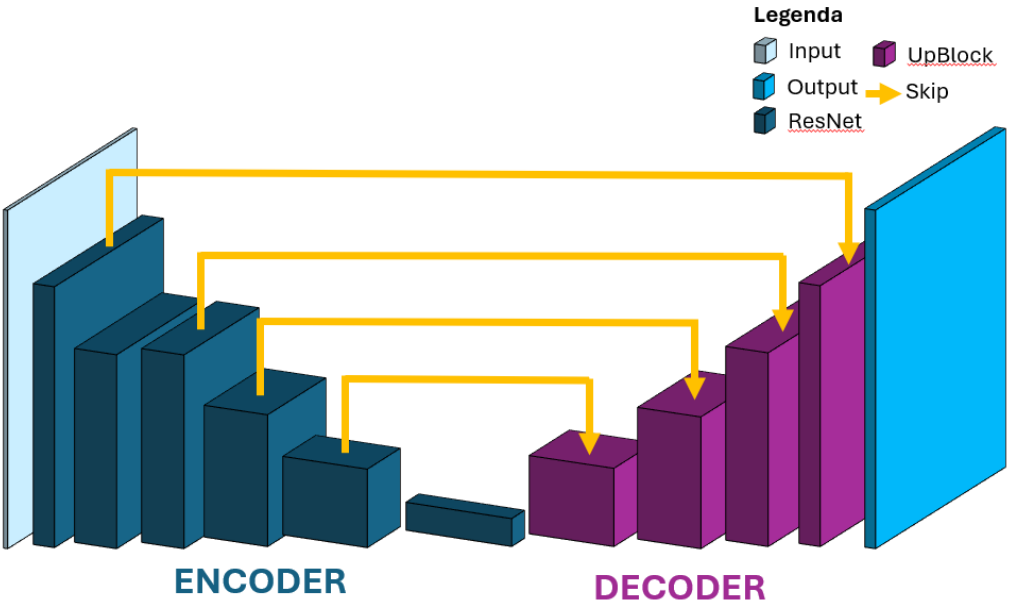


Figure 3.3: Architettura U-Net.

3.8 Encoder: backbone ResNet-18 preaddestrato

L'encoder è implementato con l'architettura ResNet-18 [9], inizializzata con i pesi preaddestrati su ImageNet. Per ciascuna configurazione di input (Sezione 3.5.2), il numero di canali in ingresso C differisce dai 3 canali RGB su cui ResNet-18 è stato preaddestrato. Il primo strato convoluzionale viene quindi sostituito con un nuovo Conv2d$(C, 64, 7 \times 7, \text{stride} = 2, \text{padding} = 3)$ e i suoi pesi sono inizializzati mediando i pesi del canale preaddestrato originale lungo la dimensione del canale:

$$W_{\text{new}}[:, c, :, :] = \frac{1}{3} \sum_{k=1}^{3} W_{\text{pretrained}}[:, k, :, :] \quad \forall\, c = 1, \ldots, C. \tag{3.6}$$

Questo approccio trasferisce la struttura dei filtri preaddestrati — che catturano bordi e gradienti orientati — a tutti i canali di input, indipendentemente dalla loro interpretazione fisica, e risulta preferibile a un'inizializzazione casuale in termini di stabilità nelle prime epoche.

Il backbone produce cinque insiemi di feature maps a risoluzioni progressivamente dimezzate rispetto all'input 256×256:

- **Skip 1** (dopo Conv1 + BN + ReLU): $\mathbb{R}^{B \times 64 \times 128 \times 128}$, risoluzione 1/2;
- **Skip 2** (dopo Layer1): $\mathbb{R}^{B \times 64 \times 64 \times 64}$, risoluzione 1/4;

- **Skip 3** (dopo Layer2): $\mathbb{R}^{B\times128\times32\times32}$, risoluzione 1/8;
- **Skip 4** (dopo Layer3): $\mathbb{R}^{B\times256\times16\times16}$, risoluzione 1/16;
- **Bottleneck** (dopo Layer4): $\mathbb{R}^{B\times512\times8\times8}$, risoluzione 1/32.

I quattro skip vengono passati al decoder; il bottleneck costituisce il punto di ingresso della fase di espansione.

3.9 Decoder con porte di attenzione

Il decoder è strutturato come una sequenza di quattro *UpBlock*, ciascuno corrispondente a un livello di skip dell'encoder, e da un blocco convoluzionale per l'espansione alla dimensione originale. Ogni UpBlock esegue la seguente sequenza di operazioni:

1. **Upsampling bilineare** $\times2$ delle feature map del livello precedente del decoder;
2. **Porta di attenzione** (Attention Gate) sul corrispondente skip dell'encoder;
3. **Concatenazione** delle feature upsamplate con lo skip modulato;
4. **Blocco residuo convoluzionale** con GroupNorm.

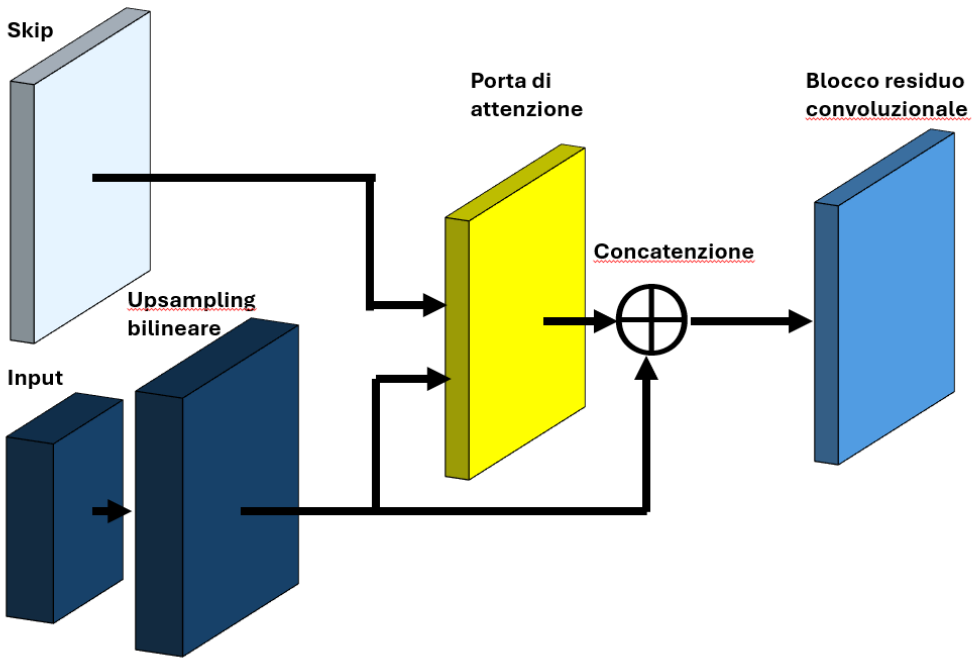


Figure 3.4: Architettura del singolo UpBlock.

3.9.1 Porta di attenzione (Attention Gate)

La porta di attenzione segue la formulazione additiva di Oktay et al. [20]. Date le feature dell'encoder al livello l (signal $\mathbf{s}$, risoluzione fine) e le feature del decoder al livello successivo (gate $\mathbf{g}$, risoluzione più grossolana), la mappa di attenzione scalare è:

$$\alpha = \sigma_2\Big(\boldsymbol{\psi}^\top \mathrm{ReLU}(W_s\mathbf{s} + W_g\mathbf{g} + b_g) + b_\psi\Big), \tag{3.7}$$

dove σ_2 è la funzione sigmoide, W_s e W_g sono convoluzione 1×1 con $d_{\text{inter}} = \max(C_s/2, 8)$ canali nascosti, e $\boldsymbol{\psi}$ è una convoluzione 1×1 che collassa a un singolo canale. Se $\mathbf{g}$ ha una risoluzione spaziale inferiore a $\mathbf{s}$, viene prima portato alla stessa risoluzione tramite interpolazione bilineare. L'output è $\tilde{\mathbf{s}} = \alpha \odot \mathbf{s}$, con $\alpha \in [0,1]$ che seleziona spazialmente le regioni del feature map dell'encoder più rilevanti per il livello corrente del decoder.

L'uso delle porte di attenzione è particolarmente motivato nel contesto della previsione degli aftershock: le regioni di interesse (zone di faglia, pattern di deformazione) occupano tipicamente una frazione ridotta della patch, e la selezione adattiva dei feature encoder aiuta il decoder a concentrare la propria capacità su tali strutture localizzate.

Nella figura 3.5 è illustrata l'architettura del gate di attenzione nel decoder. Con il termine *gate* si intende la feature map del decoder ottenuta dopo l'upsampling, mentre con *skip* si indica la corrispondente feature map proveniente dall'encoder.

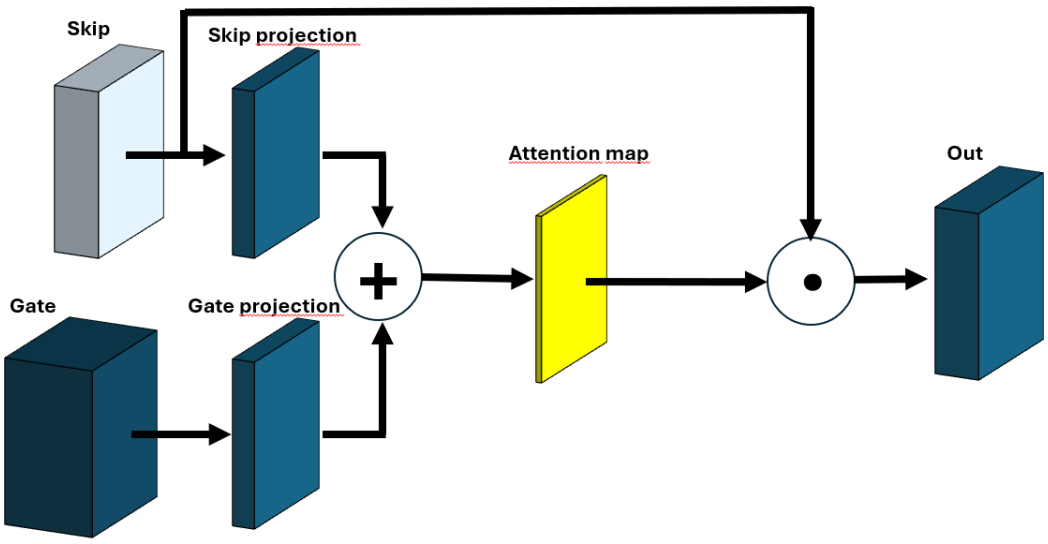


Figure 3.5: Architettura del gate di attenzione nel decoder.

3.9.2 Blocco residuo convoluzionale con GroupNorm

A differenza della formulazione originale della U-Net, che utilizza la normalizzazione di batch (Batch Normalization), il blocco convoluzionale del decoder impiega la *Group Normalization* (GroupNorm), con un numero di gruppi automaticamente adattato al numero di canali (default 8 gruppi). La scelta è motivata dal batch size ridotto ($B = 4$) imposto dai vincoli di memoria GPU: con batch size piccoli, la normalizzazione di batch soffre di stime instabili della media e varianza, mentre GroupNorm è indipendente dalla dimensione del batch.

Il blocco residuo ha la struttura:

$$\mathbf{h} = \mathrm{ReLU}\Big(\mathrm{GN}(\mathrm{Conv}_{3\times3}(\mathrm{Dropout2d}(\mathrm{ReLU}(\mathrm{GN}(\mathrm{Conv}_{3\times3}(\mathbf{x})))))) + \mathrm{shortcut}(\mathbf{x})\Big), \tag{3.8}$$

dove lo shortcut è una convoluzione 1×1 senza normalizzazione quando i canali di input e output differiscono, e una connessione identità altrimenti. Il dropout spaziale (`Dropout2d`) con $p = 0.1$ viene applicato all'interno del blocco come regolarizzazione. Nella Figura 3.6 è mostrata l'architettura del blocco residuo convoluzionale con GroupNorm. La feature map *input* è ottenuta concatenando le feature del decoder dopo l'operazione di upsampling con l'output del gate di attenzione applicato alla corrispondente skip connection dell'encoder; la feature map *output* rappresenta invece l'uscita finale del blocco residuo, e anche dell'UpBlock.

L'uscita del decoder è una mappa di feature $\mathbf{F}_{\mathrm{dec}} \in \mathbb{R}^{B\times64\times256\times256}$ con $d_{\mathrm{dec}} = 64$ canali e risoluzione piena 256×256.

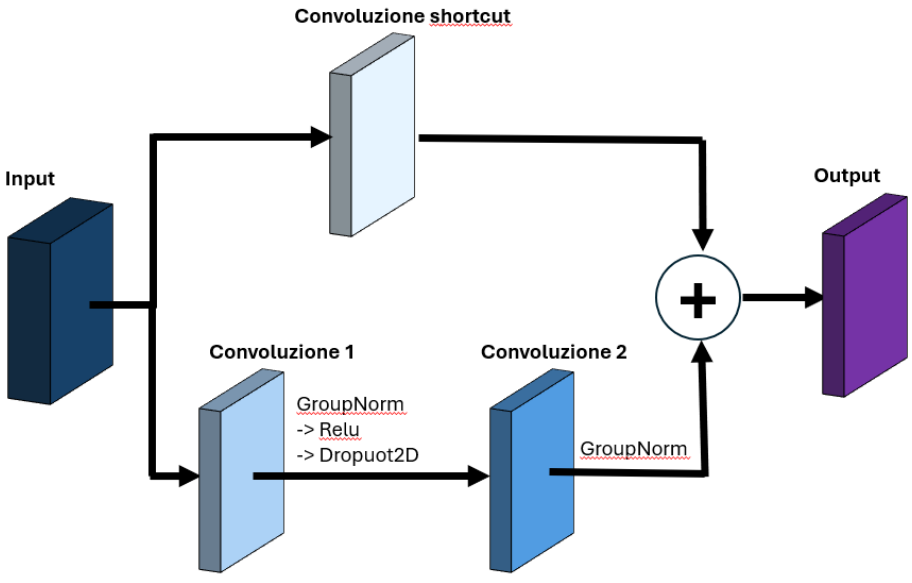


Figure 3.6: Architettura del blocco convoluzionale residuo con GroupNorm.

3.10 Blocco di fusione e condizionamento FiLM

Le feature map del decoder vengono elaborate da un blocco di fusione che ne riduce la dimensionalità prima dell'iniezione dei metadati:

$$\mathbf{F}_{\text{fus}} = \text{ReLU}(\text{BN}(W_f * \mathbf{F}_{\text{dec}})) \in \mathbb{R}^{B\times 32\times 256\times 256}, \tag{3.9}$$

dove W_f è un filtro convoluzionale 3×3 con $d_{\text{dec}}/2 = 32$ canali di output. Questo strato riduce la dimensione del canale e aumenta il campo recettivo prima del condizionamento scalare.

Il condizionamento FiLM è implementato nella variante residuale descritta nella Sezione 2.3.5. La rete generatrice dei parametri è un percettrone a due strati con attivazione ReLU:

$$[\boldsymbol{\gamma};\, \boldsymbol{\beta}] = W_2\, \text{ReLU}(W_1\, \mathbf{z}) \in \mathbb{R}^{2\times 32}, \tag{3.10}$$

con $W_1 \in \mathbb{R}^{32\times 4}$ e $W_2 \in \mathbb{R}^{64\times 32}$ (hidden dim $d_h = 32$, output $2 \times C = 2 \times 32 = 64$). La modulazione residuale è:

$$\tilde{F}_c = F_{\text{fus},c} \cdot (1 + \gamma_c(\mathbf{z})) + \beta_c(\mathbf{z}), \quad c = 1, \ldots, 32. \tag{3.11}$$

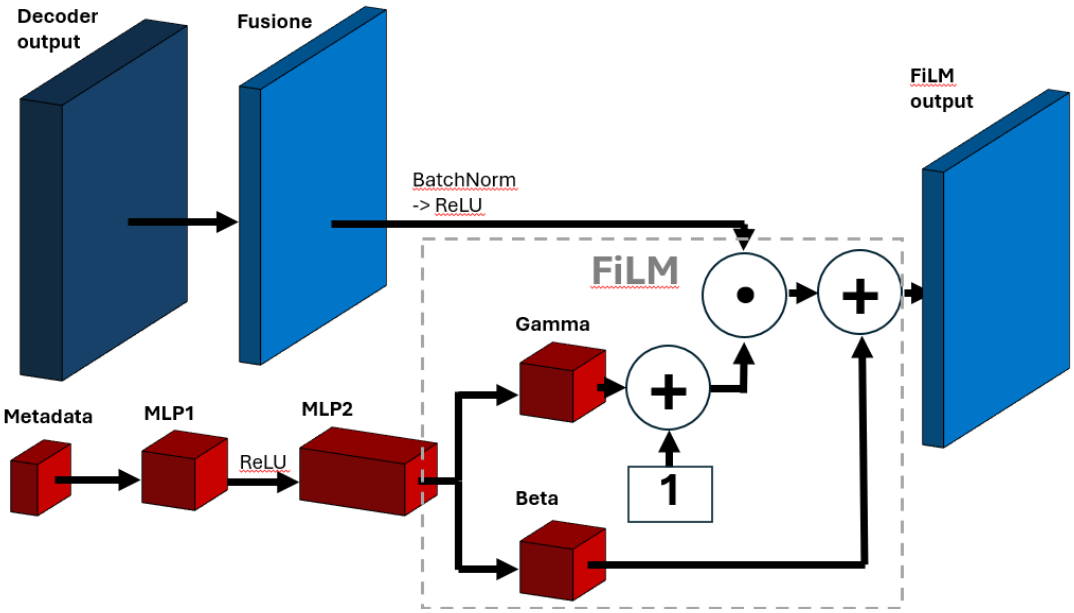


Figure 3.7: Architettura del blocco di fusione e condizionamento FiLM.

Inizializzazione della rete generatrice. L'ultimo strato lineare della rete generatrice (W_2, bias $\mathbf{b}_2$) viene inizializzato a zero: $W_2 = \mathbf{0}$, $\mathbf{b}_2 = \mathbf{0}$. Di conseguenza, all'inizio dell'addestramento $\gamma_c = 0$ e $\beta_c = 0$ per ogni canale, e le feature

map attraversano il blocco FiLM inalterate (moltiplicatore unitario, shift nullo). Questo garantisce che le prime epoche siano stabili come quelle di un modello senza condizionamento, lasciando alla rete la libertà di apprendere gradualmente l'influenza dei metadati.

3.11 Testa di output

La testa di output trasforma le feature map modulate in una mappa di logit scalare ad alta risoluzione:

$$\hat{\mathbf{y}} = W_{\text{out},2} * \text{Dropout2d}\Big(\text{ReLU}\Big(W_{\text{out},1} * \tilde{\mathbf{F}}\Big)\Big) \in \mathbb{R}^{B\times 1\times 256\times 256}, \tag{3.12}$$

dove $W_{\text{out},1}$ è un filtro 3×3 con 32 canali di output, $W_{\text{out},2}$ è una convoluzione 1×1 che collassa al singolo canale di output, e Dropout2d è applicato con $p = 0.1$.

L'output della rete è una mappa di *logit* grezzi, senza alcuna funzione di attivazione finale. La trasformazione in intensità di Poisson avviene nella funzione di perdita tramite *softplus* (Sezione 3.12), coerentemente con la formulazione descritta nella Sezione 2.3. Durante l'inferenza, la mappa di intensità predetta è $\hat{\boldsymbol{\lambda}} = \log(1 + e^{\hat{\mathbf{y}}})$. La conversione in mappa binaria avviene tramite sogliatura su $\hat{\boldsymbol{\lambda}}$: la soglia ottimale τ^* è determinata massimizzando l'indice di Youden $J = \text{TPR} - \text{FPR}$ sui valori della curva ROC calcolata sul validation set, ovvero $\tau^* = \arg\max_\tau \Big(\text{TPR}(\tau) - \text{FPR}(\tau)\Big)$. La stessa soglia è poi applicata a tutti i campioni del test set senza ulteriori adattamenti.

3.12 Funzione di perdita

La supervisione del modello avviene esclusivamente tramite la log-verosimiglianza negativa del processo di Poisson, già formalmente introdotta nella Sezione 2.3. Qui si descrive l'implementazione concreta e le motivazioni della scelta rispetto alle alternative considerate durante lo sviluppo.

Data la mappa di logit $\hat{\mathbf{y}} \in \mathbb{R}^{H\times W}$ prodotta dalla rete e la mappa target binaria $\mathbf{k} \in \{0,1\}^{H\times W}$, l'intensità predetta per pixel è:

$$\lambda_i = \log(1 + e^{\hat{y}_i}) = \text{softplus}(\hat{y}_i), \tag{3.13}$$

e la funzione di perdita è:

$$\mathcal{L}_{\text{Poisson}} = \frac{1}{HW} \sum_{i=1}^{HW} \Big(\lambda_i - k_i \, \log(\lambda_i + \varepsilon)\Big), \tag{3.14}$$

con $\varepsilon = 10^{-8}$ costante di stabilità numerica.

Confronto con le alternative. Durante lo sviluppo del progetto sono state valutate diverse funzioni di perdita, incluse Binary Cross-Entropy pesata (BCE) e una loss focale. La BCE pesata richiedeva la specificazione manuale del peso per la classe positiva $w_+ = 28.82$ (rapporto approssimato tra pixel negativi e positivi su una griglia 256×256 con sequenze di magnitudo moderata), e si è rivelata sensibile all'incompletezza del catalogo: campioni con pochi aftershock catalogati producevano gradienti dominati dai falsi negativi piuttosto che dai pattern spaziali. La perdita di Poisson, non richiedendo la specifica di pesi, si è dimostrata più stabile su dataset con densità di aftershock variabile tra eventi, in accordo con la sua motivazione probabilistica come framework naturale per i processi di punto spaziale.

3.13 Procedura di addestramento

3.13.1 Ottimizzatore e scheduling del learning rate

Il modello viene ottimizzato con AdamW [21], una variante di Adam con decoupled weight decay, con i seguenti iperparametri:

- learning rate iniziale: $\eta_0 = 10^{-4}$;
- weight decay: $\lambda_{\text{wd}} = 10^{-5}$;
- parametri di momento: $\beta_1 = 0.9$, $\beta_2 = 0.999$ (default PyTorch).

Lo scheduling del learning rate segue la strategia *warmup coseno* (Sezione 2.3), implementata dalla classe `WarmupCosineScheduler`:

$$\eta_t = \begin{cases} \eta_0 \cdot \dfrac{t}{T_{\text{warm}}} & t < T_{\text{warm}}, \\ \eta_{\min} + \dfrac{\eta_0 - \eta_{\min}}{2}\left(1 + \cos\!\left(\pi \dfrac{t - T_{\text{warm}}}{T - T_{\text{warm}}}\right)\right) & t \geq T_{\text{warm}}, \end{cases} \tag{3.15}$$

con $T_{\text{warm}} = 5$ epoche di warmup lineare, $T = 100$ epoche totali e $\eta_{\min} = 10^{-7}$. Il warmup lineare evita l'instabilità numerica nelle prime epoche, quando i parametri non preaddestrati (decoder, testa di output, rete FiLM) sono ancora lontani da una buona inizializzazione e la loss può presentare gradienti di grande ampiezza.

3.13.2 Tecniche di regolarizzazione e stabilità

Gradient clipping. Prima di ogni aggiornamento dei pesi, la norma globale del gradiente viene calcolata su tutti i parametri del modello e limitata a $\|\nabla\|_2 \leq 1.0$:

$$\nabla \leftarrow \nabla \cdot \min\!\left(1,\ \frac{1.0}{\|\nabla\|_2}\right). \tag{3.16}$$

Il gradient clipping è particolarmente importante nelle prime epoche, quando il ramo FiLM — con pesi inizialmente nulli — può produrre gradienti concentrati sui parametri della rete generatrice.

Precisione mista (AMP). L'addestramento utilizza la precisione mista automatica (*Automatic Mixed Precision*, AMP) tramite `torch.amp.GradScaler`. I forward pass e il calcolo della loss avvengono in `float16` per ridurre l'utilizzo della memoria GPU e aumentare il throughput computazionale, mentre i parametri del modello e l'aggiornamento restano in `float32`. Il gradient scaler scala dinamicamente il loss per prevenire l'underflow nella rappresentazione in mezza precisione, e de-scala prima del clipping così da garantire che la soglia di 1.0 sia applicata ai gradienti effettivi in float32.

Dropout spaziale. Il dropout spaziale (`Dropout2d`, $p = 0.1$) è applicato internamente ai blocchi residuali del decoder e alla testa di output. A differenza del dropout scalare standard, `Dropout2d` azzera interi canali per campione, promuovendo la ridondanza tra canali e riducendo la co-adattazione delle feature maps.

3.13.3 Early stopping e checkpointing

L'addestramento si arresta anticipatamente se la loss di validazione non migliora di almeno $\delta_{\min} = 10^{-5}$ per $P = 30$ epoche consecutive (*patience*). Ad ogni miglioramento del minimo di validazione, il checkpoint corrente viene salvato come `checkpoint_best.pt`. Vengono inoltre salvati checkpoint periodici ogni 5 epoche, mantenendo gli ultimi 3. Al termine dell'addestramento, il checkpoint migliore viene ricaricato per la valutazione sul test set.

3.13.4 Riepilogo degli iperparametri

La Tabella 3.1 riassume gli iperparametri di addestramento adottati in tutti gli esperimenti riportati nel Capitolo 4.

3.13.5 Inizializzazione dei pesi

I pesi del backbone ResNet-18 sono inizializzati con i pesi preaddestrati su ImageNet, fatta eccezione per il primo strato convoluzionale inizializzato come descritto nell'Equazione 3.6. I pesi di tutti gli altri moduli non preaddestrati (decoder, blocco di fusione, testa di output) sono inizializzati con la tecnica di He et al. [22] (*Kaiming normal initialization*), appropriata per strati seguiti da attivazioni ReLU:

$$w \sim \mathcal{N}\left(0,\, \sqrt{\frac{2}{n_{\text{in}}}}\right), \tag{3.17}$$

Table 3.1: Iperparametri di addestramento.

Iperparametro	**Simbolo**	**Valore**
Ottimizzatore	—	AdamW
Learning rate iniziale	η_0	10^{-4}
Weight decay	λ_{wd}	10^{-5}
Epoche di warmup	T_{warm}	5
Epoche totali (max)	T	100
Learning rate minimo	$\eta_{\min}$	10^{-7}
Batch size	B	4
Gradient clipping	$\|\nabla\|_2$	1.0
Dropout (Dropout2d)	p	0.1
Patience early stopping	P	30
Min delta early stopping	$\delta_{\min}$	10^{-5}
Seed	—	42
Precisione	—	mista (AMP)

dove n_{in} è il numero di ingressi del neurone. I bias sono inizializzati a zero. L'ultimo strato lineare della rete generatrice FiLM è inizializzato a zero, come descritto nella Sezione 3.10.

Chapter 4

Risultati sperimentali

4.1 Protocollo di valutazione

Il set di test è composto da 155 campioni, selezionati come il 15% cronologicamente più recente del dataset (split temporale) oppure come il 15% geograficamente più isolato (split spaziale), a seconda della configurazione. Le metriche di valutazione sono organizzate su due livelli complementari.

Metriche per campione (macro). Per ogni campione del test set, le mappe di intensità predette vengono convertite in mappe binarie tramite sogliatura a un valore di threshold ottimizzato sul validation set tramite ricerca sul log-spazio. Le metriche risultanti — F1, precisione, recall e IoU — vengono poi mediate sui campioni. L'AUC-ROC per campione viene calcolata pixel-wise all'interno della finestra di Gardner e aggregata come media dei valori per campione (metrica *macro*).

Metriche aggregate (*pooled*). I pixel di tutti i campioni del test set vengono concatenati in un unico vettore prima del calcolo delle metriche. L'AUC-ROC *pooled* è quindi influenzata principalmente dagli eventi con maggiore densità di aftershock, che contribuiscono con più pixel positivi al pool complessivo, ed è pertanto più sensibile alla capacità discriminativa del modello sugli eventi più energetici.

Confronto con la baseline ETAS. La baseline statistica di riferimento è il kernel spaziale isotropo ETAS (Sezione 2.1.3), calibrato sul training set. Il confronto viene condotto tramite tre quantità:

- **Information gain** (IG, bit): differenza di log-verosimiglianza media per pixel tra il modello proposto e la baseline ETAS. Un valore positivo indica che il

modello assegna probabilità più alte nelle zone dove gli aftershock si verificano effettivamente.

- **Log-likelihood ratio** (LLR): differenza di log-verosimiglianza totale sul test set; un valore positivo indica superiorità rispetto a ETAS.
- **ΔAUC** e **ΔF1**: differenze rispetto all'ETAS nelle metriche AUC-ROC *pooled* e F1 micro, calcolate allo stesso threshold ottimale.

Le sette configurazioni valutate sono riepilogate nella Tabella 4.1. Le configurazioni (i)–(iii) descritte in Sezione 3.5.2 corrispondono rispettivamente a InSAR, FF e ALL; le varianti +M aggiungono i metadati FiLM e le sigle T/S indicano lo split temporale o spaziale.

Table 4.1: Riepilogo delle configurazioni valutate. M = metadati FiLM; T = split temporale; S = split spaziale.

Sigla	Canali di input	Metadati	Split
InSAR-T	Spostamento LOS + coerenza	No	Temporale
InSAR-T+M	Spostamento LOS + coerenza	Sì	Temporale
InSAR-S	Spostamento LOS + coerenza	No	Spaziale
InSAR-S+M	Spostamento LOS + coerenza	Sì	Spaziale
FF-T	KDE foreshock + faglie GEM	No	Temporale
FF-T+M	KDE foreshock + faglie GEM	Sì	Temporale
ALL-T	Tutti e quattro i canali	Sì	Temporale

Significatività statistica. Per quantificare l'affidabilità delle differenze osservate tra configurazioni, le metriche AUC per campione vengono confrontate tramite test di Wilcoxon a ranghi con segno appaiato [23], non-parametrico e appropriato per distribuzioni asimmetriche come quelle dell'AUC su sequenze con pochi aftershock catalogati. I confronti sono limitati alle coppie metodologicamente rilevanti ($k = 7$); la soglia di significatività è corretta per confronti multipli con il metodo di Bonferroni ($\alpha_{\text{corr}} = 0.05/7 \approx 0.007$). L'effect size è riportato come $r = |Z|/\sqrt{n}$, con $|r| < 0.1$ trascurabile, 0.1–0.3 piccolo, 0.3–0.5 medio. I 21 campioni senza aftershock nel test set vengono esclusi dal calcolo dell'AUC per campione ($n = 134$ campioni validi su 155).

4.2 Risultati globali sul test set

La Tabella 4.2 riporta le metriche medie per campione sull'intero test set.

Table 4.2: Metriche medie per campione sul test set (155 campioni). La riga ETAS riporta i valori della baseline valutata sullo stesso insieme.

Configurazione	**AUC**	**F1**	**Precisione**	**Recall**	**IoU**
ETAS (temporale)	0.647	0.095	0.065	0.368	0.054
InSAR-T	0.621	0.093	0.067	0.404	0.054
InSAR-T+M	0.613	0.080	0.050	0.340	0.047
FF-T	**0.663**	**0.102**	**0.072**	**0.408**	**0.061**
FF-T+M	0.663	0.093	0.061	0.344	0.055
ALL-T	0.652	0.091	0.058	0.356	0.054
ETAS (spaziale)	0.683	0.069	0.044	0.349	0.040
InSAR-S	0.642	0.063	0.040	0.382	0.036
InSAR-S+M	**0.666**	0.050	0.030	0.383	0.028

A livello macro, InSAR-T si posiziona al di sotto della baseline ETAS sia per AUC (0.621 vs 0.647) che per F1 (0.093 vs 0.095). Questo risultato è in larga parte attribuibile all'alta varianza della metrica AUC per campione: nei campioni con pochi aftershock catalogati anche pochi eventi mancanti possono alterare significativamente il ranking tra pixel positivi e negativi. Il vantaggio di FF-T rispetto a InSAR-T (ΔAUC $= +0.043$) è invece statisticamente significativo (Wilcoxon, $p = 0.002$, $r = 0.25$), così come quello di FF-T+M rispetto a InSAR-T+M (ΔAUC $= +0.050$, $p < 0.001$, $r = 0.33$). Il calo di AUC macro introdotto dai metadati FiLM in entrambe le configurazioni (-0.008 per InSAR, ≈ 0 per FF) non raggiunge invece la significatività statistica (Wilcoxon, $p = 0.239$ e $p = 0.542$ rispettivamente), e va interpretato come rumore di stima piuttosto che come effetto sistematico. Come si vedrà nella Sezione 4.3, i metadati producono un impatto molto diverso — e marcatamente positivo — sulle metriche *pooled*.

Nello split spaziale, InSAR-S+M supera InSAR-S per AUC (0.666 vs 0.642) e mostra recall simile ad altre configurazioni (0.383), a indicare una buona copertura delle zone positive, sebbene entrambe le configurazioni restino al di sotto della rispettiva baseline ETAS (0.683). Il contesto geografico non visto durante l'addestramento rappresenta pertanto una sfida aggiuntiva per il modello basato su InSAR, indipendentemente dalla presenza dei metadati.

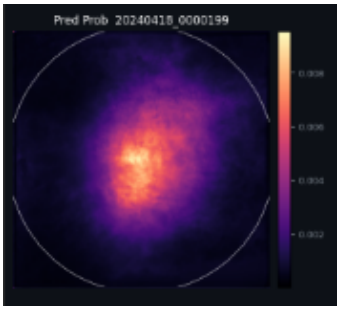


Figure 4.1: Previsioni del modello InSAR-T per il terremoto 2024-04-18 in Turchia.

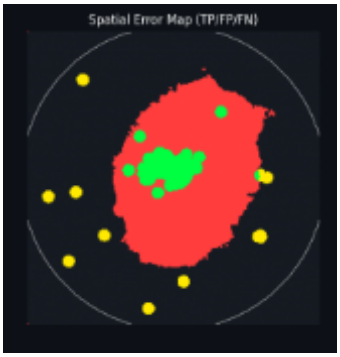


Figure 4.2: Previsioni del modello InSAR-T con threshold. In nero TN, in rosso FP, in verde TP.

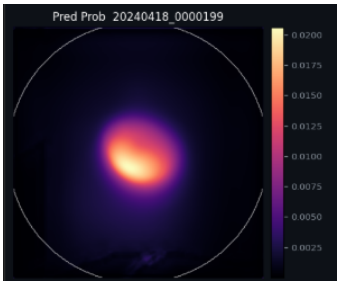


Figure 4.3: Previsioni del modello FF-T.

4.3 Confronto con la baseline ETAS

La Tabella 4.3 riporta le metriche di confronto diretto con la baseline ETAS, calcolate aggregando i pixel di tutti i campioni del test set (metriche *pooled*)

insieme al log-likelihood ratio e all'information gain.

Table 4.3: Confronto con la baseline ETAS sulle metriche *pooled* e di verosimiglianza. LLR: log-likelihood ratio (positivo = meglio di ETAS). IG: information gain in bit per pixel. ΔAUC e ΔF1 calcolati rispetto alla corrispondente ETAS dello stesso split.

Config.	$\mathbf{AUC_{pool}}$	$\mathbf{F1_{micro}}$	**LLR**	**IG (bit)**	$\mathbf{\Delta AUC_{pool}}$	$\mathbf{\Delta F1_{micro}}$
ETAS (T)	0.615	0.119	—	—	—	—
InSAR-T	0.589	0.117	+222	+0.023	−0.027	−0.002
InSAR-T+M	0.797	0.199	+413	−0.024	+0.181	+0.080
FF-T	0.674	0.151	+578	+0.162	+0.059	+0.032
FF-T+M	**0.818**	**0.225**	**+750**	+0.156	**+0.203**	**+0.106**
ALL-T	**0.818**	0.219	+721	+0.123	+0.202	+0.100
ETAS (S)	0.700	0.080	—	—	—	—
InSAR-S	0.653	0.071	−81	−0.094	−0.048	−0.009
InSAR-S+M	0.755	0.090	−24	+0.040	+0.055	+0.010

InSAR senza metadati non supera ETAS. Sia InSAR-T che InSAR-S producono AUC *pooled* e F1 micro inferiori alla rispettiva baseline ETAS. L'information gain è positivo ma modesto per InSAR-T (+0.023 bit) e negativo per InSAR-S (−0.094 bit): in assenza di informazioni contestuali sull'evento sismico, la rete apprende feature di texture specifiche all'evento che faticano a generalizzare. Il problema non è l'assenza di segnale informativo — il LLR positivo (+222) conferma che la verosimiglianza complessiva è superiore a ETAS — ma una calibrazione sistematicamente insufficiente dell'ampiezza delle predizioni.

I metadati FiLM risolvono il problema di calibrazione InSAR. L'iniezione dei quattro scalari sismologici trasforma radicalmente il comportamento del modello: InSAR-T+M porta l'AUC *pooled* da 0.589 a 0.797 (+0.208), l'F1 micro da 0.117 a 0.199, e il LLR quasi raddoppia (da +222 a +413). L'apparente contraddizione con l'IG lievemente negativo (−0.024 bit) si spiega considerando le diverse sensibilità dei due indicatori. Per un campione s con n_s aftershock e probabilità normalizzate p_i, p_i^{ETAS} sui pixel positivi $\mathcal{P}_s$:

$$\mathrm{IG}_s = \frac{1}{n_s} \sum_{i \in \mathcal{P}_s} \log_2 \frac{p_i}{p_i^{\mathrm{ETAS}}}, \qquad \mathrm{LLR}_s = \sum_{i \in \mathcal{P}_s} \log \frac{p_i}{p_i^{\mathrm{ETAS}}} = n_s \ln 2 \cdot \mathrm{IG}_s.$$

L'IG aggregato è la media di IG_s a peso uniforme sui campioni, mentre l'LLR aggregato è la media di LLR_s, proporzionale a n_s, ed è quindi dominato dagli eventi con molte repliche. Nei campioni ricchi di aftershock il modello condizionato

concentra l'intensità su zone ristrette e diagnostiche, ottenendo LLR_s elevati; nei campioni poveri penalizza i pixel positivi periferici, abbassando IG_s sotto lo zero. Il LLR aggregato positivo riflette dunque un guadagno reale di verosimiglianza sugli eventi che più contano operativamente, non un artefatto di aggregazione. Questa relazione considera solo i pixel positivi e probabilita normalizzate ed e quindi una semplificazione a fini illustrativi. Nella loss Poisson (Eq. 3.14) il termine $\sum_i \lambda_i$ agisce su tutti i pixel (positivi e negativi) e il LLR completo include anche questa penalizzazione della sovrapredizione.

Il contributo del canale InSAR nella fusione multimodale. Le configurazioni FF-T+M e ALL-T raggiungono entrambe AUC *pooled* pari a 0.818, con LLR rispettivamente di +750 e +721, e una differenza di F1 micro di soli 0.006 punti (0.225 vs 0.219). A livello di metrica macro, ALL-T mostra invece un'AUC leggermente inferiore a FF-T+M ($\Delta = -0.012$); questa differenza risulta marginalmente significativa senza correzione per confronti multipli (Wilcoxon, $p = 0.023$), ma non supera la soglia di Bonferroni ($p > \alpha_{\mathrm{corr}} = 0.007$), e va quindi interpretata con cautela. Nel complesso, una volta disponibili foreshock e metadati sismologici, l'aggiunta del canale interferometrico non produce un guadagno robusto sulle metriche globali. Il contributo del canale InSAR appare selettivo piuttosto che sistematico, e la sua utilità è condizionata alla qualità e all'ampiezza del segnale di deformazione disponibile.

4.4 Effetto dei metadati sismologici

Per isolare il contributo dei metadati scalari (magnitudo, gap InSAR–mainshock, coerenza media, conteggio dei foreshock), vengono confrontate le coppie di configurazioni con e senza condizionamento FiLM. La Tabella 4.4 riassume i guadagni assoluti.

Table 4.4: Guadagno assoluto introdotto dai metadati FiLM, calcolato come (con metadati) − (senza metadati). Metriche *pooled* e *per campione*.

Coppia	$\mathbf{\Delta AUC_{pool}}$	$\mathbf{\Delta F1_{micro}}$	$\mathbf{\Delta AUC_{macro}}$	$\mathbf{\Delta F1_{macro}}$
InSAR-T+M vs InSAR-T	+0.208	+0.082	−0.008	−0.013
FF-T+M vs FF-T	+0.144	+0.074	+0.000	−0.009
InSAR-S+M vs InSAR-S	+0.103	+0.019	+0.024	−0.013

L'effetto dei metadati è positivo sulle metriche *pooled* in tutte e tre le coppie, con guadagni compresi tra +0.103 (split spaziale) e +0.208 (InSAR temporale). Il guadagno maggiore si osserva proprio per la coppia InSAR: è la configurazione

più dipendente dal contesto dell'evento, e sono la magnitudo e il gap temporale i metadati che ne sbloccano il potenziale discriminativo.

La metrica *macro* racconta una storia diversa: l'F1 macro peggiora in tutte le coppie e l'AUC macro di InSAR-T+M cala lievemente (-0.008). Tuttavia, i test statistici indicano che queste variazioni non sono significative (Wilcoxon, $p = 0.239$ per la coppia InSAR e $p = 0.542$ per la coppia FF, entrambi $\gg \alpha_{\text{corr}}$): si tratta di fluttuazioni di stima, non di un effetto sistematico dei metadati sulle sequenze povere di aftershock. L'interpretazione causale — i metadati orientano il modello verso eventi ricchi penalizzando le sequenze povere — rimane plausibile sul piano qualitativo, ma non trova conferma quantitativa a livello macro.

4.5 Analisi per sottogruppi

L'analisi stratificata che segue mira a identificare le condizioni in cui il segnale InSAR risulta più o meno efficace, esaminando cinque variabili di stratificazione: numero di aftershock, numero di foreshock, magnitudo del mainshock, qualità della coerenza e gap temporale InSAR–mainshock.

4.5.1 Numero di aftershock

Table 4.5: AUC e F1 medie per campione stratificate per numero di aftershock. Solo configurazioni con split temporale. Sono esclusi i campioni senza aftershock (AUC non definita).

	1–2 aftershock		**10–50 aftershock**		**50+ aftershock**	
Config.	AUC	F1	AUC	F1	AUC	F1
InSAR-T	0.626	0.005	0.612	0.071	0.628	0.258
InSAR-T+M	0.616	0.001	0.613	0.054	0.638	0.238
FF-T	**0.689**	0.003	**0.669**	**0.081**	0.687	0.289
FF-T+M	0.672	0.001	0.669	0.061	**0.696**	0.279
ALL-T	0.643	0.001	0.659	0.065	0.683	**0.265**

Per i campioni con 1–2 aftershock, tutte le AUC si comprimono nell'intervallo 0.62–0.69 e le F1 sono sostanzialmente nulle: localizzare correttamente uno o due eventi su una griglia 256×256 è intrinsecamente difficile per qualsiasi configurazione. In questa fascia, InSAR-T (AUC 0.626, F1 0.005) è comparabile alle alternative basate su foreshock, confermando che il segnale di deformazione è già informativo anche con pochissimi punti di riferimento positivi.

Nella fascia 10–50 aftershock, le differenze tra configurazioni diventano più marcate. InSAR-T mantiene una AUC (0.612) e una F1 (0.071) competitive,

sebbene inferiori alle migliori configurazioni. Per gli eventi con oltre 50 aftershock, InSAR-T+M migliora rispetto alla versione senza metadati su AUC (0.638 vs 0.628), ma non su F1 (0.258 vs 0.238 per InSAR-T): su eventi ricchi di repliche, il condizionamento FiLM scala correttamente l'intensità delle predizioni. La configurazione ALL-T, che include il canale InSAR insieme ai foreshock e ai metadati, ottiene la seconda AUC più alta (0.683) e una F1 superiore a InSAR-T (0.265 vs 0.258), ma inferiore a FF-T (0.289) e FF-T+M (0.279). Questo suggerisce che l'interferogramma contribuisce positivamente quando le altre sorgenti informative sono già presenti, senza però rappresentare il top performer in questa fascia.

4.5.2 Numero di foreshock

Table 4.6: AUC e F1 medie per campione stratificate per numero di foreshock. Solo configurazioni con split temporale.

	0 foreshock *(n=65)*		**1–2** *(n=39)*		**3–5** *(n=22)*		**6+** *(n=29)*	
Config.	AUC	F1	AUC	F1	AUC	F1	AUC	F1
InSAR-T	0.601	**0.026**	0.652	0.066	0.638	0.151	0.599	0.238
InSAR-T+M	0.582	0.011	0.639	0.052	0.636	0.121	0.614	0.243
FF-T	0.613	0.017	**0.657**	0.056	0.701	**0.154**	0.727	**0.314**
FF-T+M	**0.615**	0.007	0.646	0.059	**0.702**	0.147	**0.737**	0.288
ALL-T	0.609	0.006	0.648	**0.068**	0.673	0.139	0.712	0.276

Questa stratificazione è particolarmente rilevante per valutare la robustezza del modello InSAR nelle condizioni in cui i foreshock non sono disponibili — che rappresentano il 42% del test set ($n = 65$). In assenza di precursori, InSAR-T ottiene AUC 0.601 e F1 0.026, risultati superiori a quelli di InSAR-T+M (0.582 e 0.011): il condizionamento FiLM non riesce a compensare l'assenza del segnale di foreshock e introduce invece un lieve bias. Le configurazioni FF, prive di quel canale informativo, convergono verso valori simili a InSAR-T, con il divario di AUC ridotto a soli 0.01–0.02 punti.

Al crescere del numero di foreshock, il vantaggio delle configurazioni FF si amplifica progressivamente, mentre InSAR-T e InSAR-T+M non mostrano un trend monotono in termini di AUC — coerentemente col fatto che questi modelli non utilizzano direttamente tale informazione — sebbene la f1 aumenti e sia competitiva con gli altri modelli. Nella fascia 6+ foreshock, InSAR-T registra la F1 più bassa (0.238 vs 0.314 di FF-T), a conferma che in presenza di una sismicità precursore ben definita l'interferogramma, da solo, non riesce a catturare la direzionalità della struttura già attivata. In quest'ultimo scenario, la configurazione ALL-T (AUC

0.712, F1 0.276) indica che il canale InSAR non porta un ulteriore contributo predittivo.

4.5.3 Magnitudo del mainshock

Table 4.7: AUC media per campione stratificata per fascia di magnitudo. Split temporale. Tra parentesi il numero di campioni.

Config.	[**5.0**,**5.5**) *(101)*	[**5.5**,**6.0**) *(35)*	[**6.0**,**6.5**) *(13)*	[**6.5**, $+\infty$) *(6)*
InSAR-T	0.617	0.613	0.669	0.624
InSAR-T+M	0.608	0.621	0.654	0.552
FF-T	**0.670**	0.651	0.733	0.459
FF-T+M	0.666	0.642	**0.755**	0.516
ALL-T	0.650	**0.635**	0.731	**0.594**

Le differenze tra le configurazioni rimangono relativamente contenute in tutte le fasce di magnitudo, con AUC generalmente comprese tra 0.60 e 0.75. Nella fascia dominante $[5.0, 5.5)$, che rappresenta circa il 65% dei campioni del test set, le configurazioni che includono l'informazione sui foreshock mostrano prestazioni mediamente superiori rispetto ai modelli basati esclusivamente sull'InSAR, suggerendo che la sismicità precursoria fornisca un segnale spaziale più stabile per eventi di moderata magnitudo.

4.5.4 Qualità della coerenza interferometrica

La qualità dell'interferogramma è stratificata in quattro quartili della coerenza media della patch, da Q1 (bassa coerenza) a Q4 (alta coerenza).

Table 4.8: AUC e F1 medie per campione stratificate per quartile di coerenza. Split temporale.

	Q1 (bassa)		**Q2**		**Q3**		**Q4 (alta)**	
Config.	AUC	F1	AUC	F1	AUC	F1	AUC	F1
InSAR-T	0.644	**0.064**	0.613	**0.071**	0.642	0.092	0.584	0.147
InSAR-T+M	0.610	0.034	0.586	0.063	0.660	0.079	0.596	0.146
FF-T	0.668	0.057	0.617	0.063	**0.683**	**0.094**	**0.684**	**0.193**
FF-T+M	**0.676**	0.059	**0.627**	0.060	0.672	0.073	0.676	0.178
ALL-T	0.657	0.045	0.618	0.065	0.671	0.082	0.659	0.172

Per InSAR-T, l'AUC e la F1 mostrano andamenti qualitativamente diversi

al crescere della coerenza. L'AUC si mantiene sostanzialmente stabile attraverso i quartili, mentre la F1 cresce monotonicamente (0.064 → 0.147). Questa dissociazione riflette la natura complementare dei due indicatori: l'AUC misura la capacità del modello di ordinare correttamente i pixel per probabilità di aftershock, indipendentemente dalla soglia di classificazione, mentre la F1 è sensibile anche alla densità di aftershock nei campioni esaminati. Nei quartili ad alta coerenza il rumore di fase è ridotto, e il modello produce distribuzioni di punteggio più polarizzate; ciò favorisce l'identificazione di una soglia che cattura una quota maggiore di veri positivi, innalzando la F1 pur in assenza di un corrispondente miglioramento nel ranking globale. In altri termini, il modello riconosce *che* si verificheranno aftershock, ma incontra difficoltà nel localizzare *dove*.

FF-T, pur non utilizzando il dato interferometrico, si conferma la configurazione più robusta attraverso i quartili, raggiungendo i valori massimi di AUC (0.684) e F1 (0.193) in Q4.

4.5.5 Gap temporale InSAR–mainshock

Table 4.9: AUC e F1 medie per campione stratificate per gap temporale InSAR–mainshock. Split temporale.

	0–5 gg *(n=66)*		**5–10 gg** *(n=53)*		**10–20 gg** *(n=28)*		**20+ gg** *(n=8)*	
Config.	AUC	F1	AUC	F1	AUC	F1	AUC	F1
InSAR-T	0.626	0.106	0.631	0.094	0.629	0.084	0.417	0.019
InSAR-T+M	0.624	0.100	0.609	0.082	0.599	0.054	0.599	0.000
FF-T	**0.676**	**0.111**	0.657	0.100	**0.656**	**0.092**	0.620	**0.067**
FF-T+M	0.667	0.101	**0.675**	**0.104**	0.644	0.063	0.605	0.051
ALL-T	0.664	0.104	0.659	0.096	0.612	0.064	**0.627**	0.053

Nei gap brevi (0–5 giorni), che rappresentano lo scenario operativo più rilevante, InSAR-T ottiene AUC 0.626 e F1 0.106 — valori sostanzialmente stabili anche nelle fasce 5–10 gg (0.631 / 0.094) e 10–20 gg (0.629 / 0.084). Questo andamento indica che, finché il gap rimane entro i 20 giorni, la qualità del segnale di deformazione non degrada significativamente.

4.6 Discussione

I risultati sperimentali permettono di delineare un quadro coerente sul valore informativo del segnale InSAR per la previsione spaziale degli aftershock e sulle condizioni che ne determinano l'efficacia.

Il segnale InSAR è informativo ma richiede contesto scalare. Nella configurazione senza metadati, l'interferogramma cosismico non supera la baseline ETAS sulle metriche *pooled*, pur mantenendo un LLR positivo (+222). Il nodo è la calibrazione: senza metadati, la rete non riesce a normalizzare l'ampiezza delle predizioni rispetto alle caratteristiche fisiche dell'evento. L'aggiunta dei soli quattro scalari FiLM risolve questo problema in modo marcato, portando l'AUC *pooled* da 0.589 a 0.797. Il segnale InSAR è quindi necessario ma non sufficiente: il suo potenziale discriminativo dipende dalla disponibilità del contesto sismologico dell'evento.

Condizioni di utilizzo del segnale InSAR. L'analisi per sottogruppi suggerisce che il contributo del segnale InSAR dipenda fortemente dalle caratteristiche dell'evento e dalla qualità dei dati disponibili. In generale, le prestazioni tendono a migliorare per eventi di magnitudo più elevata, in cui la deformazione superficiale risulta più evidente e strutturata. Analogamente, gap temporali contenuti tra il mainshock e l'acquisizione InSAR favoriscono una migliore leggibilità del segnale cosismico, mentre intervalli più lunghi possono introdurre contributi post-sismici che rendono la localizzazione degli aftershock meno stabile.

La stratificazione per qualità della coerenza mostra inoltre che il legame tra qualità interferometrica e prestazioni non è monotono, suggerendo che la sola coerenza media non sia sufficiente a descrivere la quantità di informazione utile contenuta nell'interferogramma. Infine, il modello basato esclusivamente su InSAR risulta relativamente più competitivo nei campioni privi di foreshock, dove manca una sorgente informativa sismologica diretta sulla geometria della sequenza.

Fusione multimodale come paradigma operativo. ALL-T e FF-T+M raggiungono entrambe AUC *pooled* 0.818 (ΔAUC +0.202 rispetto a ETAS), con una differenza di F1 micro di soli 0.006 punti. Il contributo del canale InSAR nella fusione non è quindi sistematico a livello aggregato, ma emerge selettivamente nelle condizioni in cui il segnale di deformazione è più diagnostico. In queste condizioni, la morfologia dell'interferogramma aggiunge informazione che né i foreshock né i metadati da soli riescono a catturare, giustificando l'inclusione del canale InSAR in un sistema operativo completo.

Limiti della supervisione da catalogo globale. I valori assoluti di F1 restano modesti per tutte le configurazioni. Una causa strutturale è l'incompletezza dei cataloghi sismici globali: in alcune regioni la magnitudo di completezza EMSC supera $M_c = 3$, e la distribuzione del numero di aftershock per campione è fortemente asimmetrica — mediana 9, media 129 — a indicare che la qualità della supervisione varia considerevolmente tra eventi. Il modello viene penalizzato per predizioni spazialmente corrette in zone dove le repliche si sono verificate ma non sono state

catalogate, un bias non rimovibile con il dataset attuale che rappresenta il limite fondamentale della formulazione in termini di supervisione binaria da catalogo globale.

Split temporale vs spaziale. Il confronto tra InSAR-T e InSAR-S mostra che lo split spaziale produce AUC macro superiore (0.642 vs 0.621), a indicare una capacità di generalizzazione geografica discreta per gli eventi di piccola magnitudo. Tuttavia, l'AUC *pooled* di InSAR-S (0.653) resta al di sotto della baseline ETAS spaziale (0.700), segnalando che la discriminazione degli eventi più energetici si deteriora su regioni non viste durante l'addestramento. Lo split temporale rimane la modalità di valutazione più fedele allo scenario operativo reale, in cui il modello è addestrato su eventi storici e applicato a sequenze future tendenzialmente nella stessa area geografica.

Chapter 5

Conclusioni e Lavori Futuri

Panoramica

In questo capitolo vengono sintetizzati i principali contributi della tesi, discussi i limiti strutturali della pipeline proposta e delineate le direzioni di ricerca che appaiono più promettenti per estendere e consolidare i risultati ottenuti.

5.1 Conclusioni

Questa tesi ha affrontato il problema della previsione spaziale degli aftershock a partire da dati osservativi — interferogrammi InSAR, mappe di coerenza, distribuzioni spaziali dei foreshock e faglie attive — senza ricorrere a modelli di stress sintetici. L'obiettivo dichiarato era verificare se una rete neurale convoluzionale addestrata su dati disponibili operativamente nelle ore successive al mainshock potesse superare la baseline statistica isotropa ETAS nella discriminazione spaziale delle repliche.

Pipeline end-to-end. Il primo contributo rilevante è la realizzazione di una pipeline completa e riproducibile che copre l'intera catena di elaborazione: dall'acquisizione del catalogo EMSC tramite API pubblica, al declustering cronologico basato sulle finestre di Gardner–Knopoff, alla selezione automatica dei frame LiCSAR e al download degli interferogrammi Sentinel-1, fino alla costruzione di un dataset geofisicamente coerente su griglia 256×256 pixel. La normalizzazione spaziale adattiva basata sul raggio di Gardner come unità di scala ha reso confrontabili tra loro sequenze di magnitudo molto diversa, consentendo un addestramento unificato su un catalogo globale eterogeneo. La separazione causale tra contesto di input e target di previsione — imperniata sulla data di fine dell'interferogramma piuttosto

che sulla data del mainshock — garantisce che il modello non osservi alcuna informazione sulle repliche future durante l'inferenza, condizione indispensabile per la correttezza operativa della valutazione.

Il segnale InSAR è informativo ma non autosufficiente. I risultati sperimentali mostrano con coerenza che il canale di spostamento InSAR contiene informazione predittiva sulla distribuzione degli aftershock: il log-likelihood ratio di InSAR-T rispetto alla baseline ETAS è positivo (+222), confermando che la verosimiglianza globale supera quella del kernel isotropo. Tuttavia, nella configurazione senza metadati, l'AUC pooled (0.589) non raggiunge il livello della baseline ETAS (0.615), segnalando un problema di calibrazione sistematica: la rete apprende rappresentazioni spazialmente informative ma non riesce a normalizzare l'ampiezza delle predizioni attraverso eventi di magnitudo e baseline temporale molto diversi.

I metadati FiLM risolvono il problema di calibrazione. L'iniezione di quattro scalari sismologici — magnitudo del mainshock, gap temporale InSAR–mainshock, coerenza media della patch e conteggio dei foreshock — tramite il meccanismo di condizionamento FiLM trasforma radicalmente il comportamento del modello. InSAR-T+M porta l'AUC pooled da 0.589 a 0.797 (+0.208), con un guadagno di F1 micro di +0.082 e un LLR quasi raddoppiato (da +222 a +413) rispetto alla baseline ETAS. Questo risultato indica che la relazione tra il pattern di deformazione InSAR e la localizzazione degli aftershock è significativamente dipendente dal contesto dell'evento: la stessa impronta cosismica ha implicazioni predittive diverse a seconda della magnitudo, della freschezza dell'interferogramma e della qualità della coerenza. Il meccanismo residuale FiLM adottato — $\tilde{F}_c = F_c \cdot (1 + \gamma_c(\mathbf{z})) + \beta_c(\mathbf{z})$ — si è dimostrato particolarmente adatto a dati geofisici con segnale debole e sparso, garantendo che i metadati scalari modulino le feature spaziali senza sopprimerle nelle epoche iniziali.

Foreshock e faglie come sorgenti complementari. Le configurazioni basate sulla mappa KDE dei foreshock e sulle faglie attive GEM (FF-T, FF-T+M) raggiungono i valori più elevati di AUC pooled tra le configurazioni con split temporale: FF-T+M ottiene 0.818 con un ΔAUC di +0.203 rispetto a ETAS e un LLR di +750. Il vantaggio delle configurazioni FF è particolarmente marcato quando i foreshock sono numerosi: nella fascia 6+ precursori, FF-T registra AUC macro 0.727 contro 0.599 di InSAR-T, a conferma che la distribuzione spaziale dei precursori traccia in modo efficace le strutture sismogeniche già attivate. Nella condizione operativamente rilevante di assenza di foreshock — che rappresenta il 42% del test set ($n = 65$) — InSAR-T ottiene invece AUC macro 0.601, comparabile a quella delle configurazioni FF (0.613–0.615), dimostrando che il segnale interferometrico è una sorgente di informazione robusta quando i precursori non sono disponibili.

La fusione multimodale non produce guadagni sistematici aggregati. ALL-T e FF-T+M raggiungono entrambe AUC pooled 0.818 con una differenza di F1 micro di soli 0.006 punti (0.219 vs 0.225). A livello aggregato, l'aggiunta del canale InSAR alla configurazione FF-T+M non migliora le metriche pooled in misura rilevante. L'analisi per sottogruppi rivela tuttavia che il contributo del canale interferometrico è selettivo: nella fascia di magnitudo $[6.0, 6.5)$ ALL-T ottiene AUC macro 0.731 contro 0.733 di FF-T, indicando una complementarità sostanziale; analogamente, nei campioni con gap breve (0–5 giorni) e alta coerenza (Q4), InSAR-T produce le F1 più elevate della propria categoria (0.106 e 0.147 rispettivamente). Questi risultati suggeriscono che in un sistema operativo la scelta dei canali di input dovrebbe essere guidata dalle caratteristiche dell'evento piuttosto che da una selezione fissa a priori.

Generalizzazione geografica e temporale. Il confronto tra split temporale e spaziale mostra che il modello basato su InSAR generalizza con discreta efficacia su eventi futuri nella stessa regione (scenario temporale), mentre incontra difficoltà maggiori su regioni non viste durante l'addestramento (scenario spaziale). InSAR-S produce AUC pooled 0.653, al di sotto della baseline ETAS spaziale (0.700), con un information gain negativo (-0.094 bit); InSAR-S+M recupera parzialmente (0.755, $\Delta\text{AUC} + 0.055$), ma rimane inferiore alle corrispondenti configurazioni temporali. Lo split temporale rimane la modalità di valutazione più aderente allo scenario operativo reale, nel quale il modello è addestrato su eventi storici e applicato a sequenze future nella stessa area geografica.

Limiti assoluti e incompletezza del catalogo. I valori assoluti di F1 restano modesti per tutte le configurazioni — il massimo osservato è F1 micro 0.225 per FF-T+M in regime pooled — con una causa strutturale identificabile nell'incompletezza dei cataloghi sismici globali. La magnitudo di completezza EMSC varia regionalmente e può superare $M_c = 3$ in aree a bassa densità di stazioni. La distribuzione del numero di aftershock per campione è fortemente asimmetrica: mediana 9 eventi, media 129, a indicare che pochi eventi con sequenze molto ricche dominano la statistica aggregata. Le metriche riportate vanno pertanto interpretate come stime per difetto della capacità discriminativa reale del modello: la supervisione binaria penalizza predizioni spazialmente corrette in zone dove le repliche si sono verificate ma non sono state catalogate, introducendo un rumore sistematico non rimovibile con il dataset attuale.

5.2 Lavori Futuri

I risultati ottenuti aprono numerose direzioni di sviluppo, che possono essere organizzate in tre livelli tematici: qualità dei dati, architettura del modello e validazione operativa.

5.2.1 Miglioramento e ampliamento del dataset

Integrazione di cataloghi regionali ad alta completezza. La limitazione più impattante sui valori assoluti di F1 è l'incompletezza del catalogo EMSC nelle regioni a bassa densità di stazioni sismiche. Una direzione naturale è l'integrazione di cataloghi regionali ad alta completezza che offrono soglie di completezza inferiori. La costruzione di un dataset ibrido, con campioni selezionati prioritariamente dalle regioni coperte da cataloghi densi, ridurrebbe il rumore di supervisione e consentirebbe una valutazione più accurata della capacità discriminativa reale del modello.

Fusione di orbite SAR ascendenti e discendenti. Ogni interferogramma LiCSAR misura lo spostamento lungo la sola linea di vista del satellite (LOS), che combina in modo ambiguo le componenti orizzontale e verticale del campo di deformazione. La disponibilità di acquisizioni sia ascendenti che discendenti per lo stesso evento — ormai sistematica per Sentinel-1 nelle aree con buona copertura — consentirebbe la decomposizione del campo di spostamento nelle componenti Est–Ovest e verticale, fornendo una rappresentazione fisicamente più completa della deformazione cosismica. L'architettura U-Net proposta è direttamente estendibile a un input a quattro canali (spostamento e coerenza per ciascuna orbita), senza modifiche strutturali all'encoder.

5.2.2 Sviluppi architetturali

Calibrazione delle probabilità. I risultati mostrano che buone prestazioni in termini di AUC non implicano necessariamente una buona calibrazione delle probabilità predette. In alcuni casi il modello riesce infatti a ordinare correttamente le aree più probabili, pur assegnando probabilità poco calibrate in valore assoluto. Tecniche di calibrazione post-hoc o funzioni di loss specifiche per la calibrazione potrebbero quindi migliorare ulteriormente la qualità probabilistica delle predizioni senza modificare significativamente la capacità discriminativa del modello.

Architetture alternative basate su transformer. L'architettura U-Net con encoder ResNet adottata in questa tesi sfrutta l'induzione di località propria delle reti convoluzionali, che si è rivelata adeguata per catturare i pattern spaziali del

segnale cosismico a scala locale. Una direzione esplorativa consiste nel valutare architetture basate su transformer in grado di modellare dipendenze a lungo raggio attraverso meccanismi di self-attention e potrebbero risultare vantaggiose per catturare relazioni spaziali estese lungo la traccia della faglia, non facilmente accessibili a un campo recettivo convoluzionale finito. Va tuttavia osservato che i transformer richiedono generalmente dataset di addestramento di dimensioni maggiori rispetto a quelli disponibili in questo lavoro per esprimere il loro pieno potenziale, e che il vantaggio rispetto a architetture ibride convolution–attention — come quella già proposta con le porte di attenzione nel decoder — non è garantito a priori su dati geofisici sparsi. Il loro utilizzo rimane pertanto una direzione da esplorare in concomitanza con un ampliamento del dataset descritto nella sezione precedente.

Estensione alla previsione temporale. La formulazione attuale produce una mappa binaria per l'intera finestra temporale successiva alla data di fine dell'interferogramma, senza fornire informazioni sul ritmo temporale delle repliche. Un'estensione naturale è la previsione congiunta della distribuzione spaziale e dell'evoluzione temporale della sequenza, formulata come stima della mappa di intensità del processo ETAS nella finestra $[t_0, t_0 + \Delta t]$ per valori diversi di Δt. Questo richiederebbe l'introduzione di un condizionamento sul parametro temporale — ad esempio tramite FiLM o codifiche sinusoidali — e una funzione di perdita che tenga conto della produttività attesa della sequenza in funzione della magnitudo secondo la legge di Omori–Utsu.

5.2.3 Validazione e deployment operativo

Integrazione in un sistema di risposta rapida. La pipeline proposta è progettata per essere operativamente accessibile: i dati LiCSAR sono distribuiti gratuitamente e con latenza di pochi giorni, e il modello addestrato può essere eseguito su GPU consumer in tempo reale. Un passo naturale verso il deployment è l'integrazione in un sistema automatico di risposta rapida, in cui il rilevamento di un mainshock $M \geq 5.0$ da parte di un catalogo in tempo reale (EMSC, USGS/ANSS) avvia automaticamente la selezione del frame LiCSAR, il download dell'interferogramma cosismico non appena disponibile e l'esecuzione della previsione, producendo una mappa di intensità degli aftershock pubblicabile nelle ore successive all'evento. Tale sistema richiederebbe la gestione della latenza di acquisizione Sentinel-1 (tipicamente 6–12 giorni per il primo interferogramma), che potrebbe essere parzialmente compensata da un modello di predizione pre-InSAR — basato sui soli foreshock e metadati — attivato immediatamente dopo il mainshock e aggiornato con il canale interferometrico non appena disponibile.

Interpretabilità e analisi delle feature apprese. Un aspetto non affrontato in questa tesi è l'interpretabilità delle rappresentazioni interne del modello. Tecniche di spiegabilità post-hoc applicate all'encoder ResNet-18, permetterebbero di identificare quali pattern del segnale InSAR guidano maggiormente le predizioni ad alta intensità. Tali analisi non hanno solo valore scientifico, ma sono necessarie per guadagnare la fiducia degli operatori di protezione civile che dovrebbero utilizzare il sistema in contesti di emergenza reale.

5.3 Riflessione finale

Il risultato complessivo di questa tesi è duplice. Da un lato, viene dimostrato che i dati InSAR cosismici contengono informazione predittiva sulla distribuzione spaziale degli aftershock, estraibile da una rete convoluzionale senza la mediazione di modelli fisici di stress, a condizione che il contesto sismologico dell'evento sia fornito tramite un meccanismo di condizionamento scalare come FiLM. Dall'altro, viene mostrato che le sorgenti di informazione complementari — distribuzione spaziale dei foreshock e faglie attive — non sono intercambiabili con il segnale interferometrico, ma si posizionano come alternative migliori nelle regioni a bassa qualità InSAR e come complementi efficaci in quelle ad alta deformazione cosismica.

Il limite fondamentale non è architetturale, ma di supervisione: i cataloghi sismici globali non registrano la totalità delle repliche, e la valutazione delle prestazioni è inevitabilmente affetta da questo rumore sistematico. Il passo più impattante per il progresso del campo è probabilmente l'integrazione di cataloghi regionali ad alta completezza che renderebbe il segnale di supervisione significativamente più pulito e consentirebbe una valutazione equa della capacità discriminativa reale dei modelli proposti.

La pipeline end-to-end sviluppata in questa tesi costituisce un'infrastruttura riproducibile e modulare su cui innestare questi sviluppi futuri, con l'obiettivo di avvicinare la previsione spaziale degli aftershock a uno standard operativo utile per la gestione del rischio sismico post-evento.

Bibliography

[1] Y. Ogata. «Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes». In: *Journal of the American Statistical Association* 83.401 (Mar. 1988), pp. 9–27. DOI: 10.1080/01621459.1988.10478560 (cit. on pp. 1, 4, 8, 21).

[2] Phoebe M. R. DeVries, Fernanda Viégas, Martin Wattenberg, and Brendan J. Meade. «Deep learning of aftershock patterns following large earthquakes». In: *Nature* 560.7718 (2018), pp. 632–634. DOI: 10.1038/s41586-018-0438-y (cit. on pp. 2, 10).

[3] D. Massonnet, M. Rossi, C. Carmona, F. Adragna, G. Peltzer, K. Feigl, and T. Rabaute. «The Displacement Field of the Landers Earthquake Mapped by Radar Interferometry». In: *Nature* 364.6433 (1993), pp. 138–142. DOI: 10.1038/364138a0 (cit. on pp. 2, 13).

[4] M. Lazecký et al. «LiCSAR: An Automatic InSAR Tool for Measuring and Monitoring Tectonic and Volcanic Activity». In: *Remote Sensing* 12.15 (July 2020), p. 2430. DOI: 10.3390/rs12152430 (cit. on pp. 2, 14, 15, 26).

[5] A. Mignan and M. Broccardo. «One neuron linearizes and solves the deep learning of aftershock patterns». In: *Nature* 574 (Oct. 2019), E1–E3. DOI: 10.1038/s41586-019-1582-8 (cit. on p. 3).

[6] J. K. Gardner and L. Knopoff. «Is the Sequence of Earthquakes in Southern California, with Aftershocks Removed, Poissonian?» In: *Bulletin of the Seismological Society of America* 64.5 (1974), pp. 1363–1367 (cit. on pp. 3, 4, 7, 25).

[7] R. Styron and M. Pagani. «The GEM Global Active Faults Database». In: *Earthquake Spectra* 36.1_suppl (Oct. 2020), pp. 160–180. DOI: 10.1177/8755293020944182 (cit. on pp. 4, 28).

[8] O. Ronneberger, P. Fischer, and T. Brox. «U-Net: Convolutional Networks for Biomedical Image Segmentation». In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Vol. 9351. Lecture Notes in

Computer Science. Springer, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28 (cit. on pp. 4, 19, 20).

[9] K. He, X. Zhang, S. Ren, and J. Sun. «Deep Residual Learning for Image Recognition». In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016, pp. 770–778 (cit. on pp. 4, 18, 33).

[10] B. Gutenberg and C. F. Richter. «Frequency of Earthquakes in California». In: *Bulletin of the Seismological Society of America* 34.4 (1944), pp. 185–188 (cit. on p. 6).

[11] T. Utsu. «A Statistical Study on the Occurrence of Aftershocks». In: *Geophysical Magazine* 30 (1961), pp. 521–605 (cit. on p. 7).

[12] G. C. P. King, R. S. Stein, and J. Lin. «Static Stress Changes and the Triggering of Earthquakes». In: *Bulletin of the Seismological Society of America* 84.3 (June 1994), pp. 935–953. DOI: 10.1785/BSSA0840030935 (cit. on p. 9).

[13] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall, 1986. DOI: 10.1007/978-1-4899-3324-9 (cit. on p. 10).

[14] Y. LeCun, Y. Bengio, and G. Hinton. «Deep Learning». In: *Nature* 521.7553 (May 2015), pp. 436–444. DOI: 10.1038/nature14539 (cit. on pp. 16–18).

[15] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL: https://www.deeplearningbook.org (cit. on p. 16).

[16] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. «Learning representations by back-propagating errors». In: *Nature* 323 (1986), pp. 533–536 (cit. on p. 16).

[17] Diederik P. Kingma and Jimmy Ba. «Adam: A Method for Stochastic Optimization». In: *ICLR* (2015) (cit. on p. 16).

[18] E. Perez, F. Strub, H. de Vries, V. Dumoulin, and A. Courville. «FiLM: Visual Reasoning with a General Conditioning Layer». In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. Feb. 2018, pp. 3942–3951. DOI: 10.1609/aaai.v32i1.11671 (cit. on p. 20).

[19] European Mediterranean Seismological Centre (EMSC). *EMSC Earthquake Catalog Web Service*. https://www.seismicportal.eu. Accessed: 2026-03-15. 2024 (cit. on p. 24).

[20] Ozan Oktay et al. «Attention U-Net: Learning Where to Look for the Pancreas». In: *Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL)*. 2018. URL: https://arxiv.org/abs/1804.03999 (cit. on p. 35).

[21] Ilya Loshchilov and Frank Hutter. «Decoupled Weight Decay Regularization». In: *ICLR* (2019) (cit. on p. 39).

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. «Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification». In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2015, pp. 1026–1034. DOI: 10.1109/ICCV.2015.123 (cit. on p. 40).

[23] Frank Wilcoxon. «Individual Comparisons by Ranking Methods». In: *Biometrics Bulletin* 1.6 (1945), pp. 80–83 (cit. on p. 43).