

## Riassunto della tesi

**Titolo originale:** “Heterogeneous change detection on remote sensing data with self-supervised deep canonically correlated autoencoders”

**Titolo italiano:** “Rilevamento automatico di cambiamenti da dati eterogenei telerilevati con rete neurale profonda a massimizzazione di correlazione e non supervisionata”

La crescente disponibilità di immagini telerilevate, in particolare satellitari, rende possibile un loro utilizzo per lo studio della Terra, delle sue risorse e del loro sfruttamento, in modo da monitorare lo stato del nostro pianeta. Inoltre, possedere un maggior numero di immagini significa disporre di un campionamento più fine nel tempo, il che permette una maggiore accuratezza in tutte le applicazioni che seguono e tracciano l'evolversi dei cambiamenti, antropomorfi e non, che avvengono sulla Terra. In questa cornice si inquadra perfettamente il presente lavoro di tesi, il cui scopo è lo sviluppo di una nuova metodologia per il rilevamento automatico dei cambiamenti al suolo a partire da coppie di immagini telerilevate, siano esse ottiche o radar.

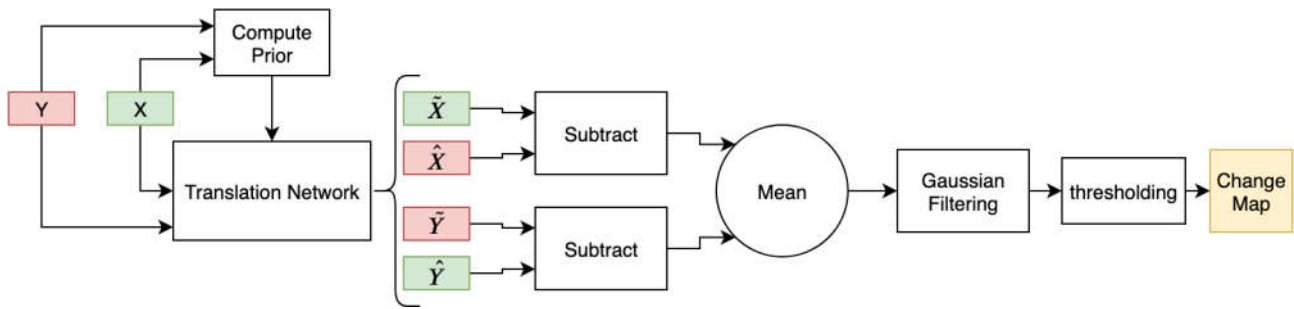
Il presente lavoro affronta il rilevamento automatico di cambiamenti a partire da immagini acquisite tramite sensori eterogenei, ovvero è in grado di utilizzare diversi tipi di acquisizioni del medesimo luogo per estrapolare l'informazione cercata. Questa metodologia, dunque, è particolarmente utile in caso di disastri naturali. In questi contesti è necessario utilizzare un'acquisizione immediata dell'area di indagine, ed effettuare il prima possibile una valutazione dell'evento e delle sue conseguenze in termini di estensione ed intensità. È quindi necessario disporre di un sistema atto a comparare immagini multimodali, poiché la frequenza di rivisitazione dei satelliti può risultare bassa e non permettere l'acquisizione dell'immagine con la modalità desiderata.

Benché incentrato sulle immagini multimodali, il presente studio può essere applicato anche a dati acquisiti con singola modalità. Inoltre, per quanto riguarda l'approccio bitemporale, ovvero la comparazione di due sole immagini alla volta, esso è facilmente estendibile e generalizzabile ad uno multi-temporale.

Entrando nel merito, la tesi si articola nei seguenti capitoli: Introduzione al telerilevamento, basi teoriche della metodologia, metodologia proposta, esperimenti e risultati ed infine le conclusioni.

La prima parte presenta un'introduzione generale al telerilevamento satellitare, mettendone in luce i vari tipi e i sensori utilizzati (passivi-attivi, pancromatici-multispettrali, ottici-termici-radar), e fornendo una breve spiegazione dei fenomeni fisici che sottendono ad essa. Tratta poi l'elaborazione dei dati e si sofferma in particolare sul rilevamento dei cambiamenti. A questo riguardo, una breve rassegna della letteratura di riferimento evidenzia le differenze metodologiche di base per quanto concerne il rilevamento da dati omogenei (da una coppia di immagini a singola modalità) ed eterogenei (da immagini multimodali). In questa sede preme rilevare quanto sia difficile il compito di rilevare cambiamenti in modo non supervisionato da immagini eterogenee, poiché implica la comparazione di dati che per loro natura appartengono a domini differenti, con statistiche diverse tra loro, per quanto riguarda sia l'intensità sia il rumore.

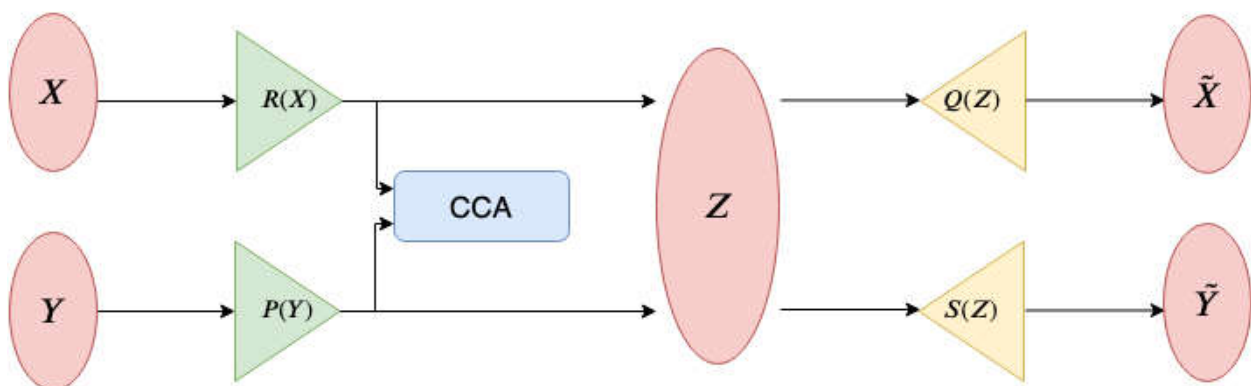
Seguono i due capitoli che si occupano di introdurre le nozioni teoriche fondamentali e la metodologia proposta, la quale prende spunto dalla letteratura scientifica, in particolare da *Luppino et al. 2020* (presenta la rete per la traslazione di dominio di immagini) e *Wang et al. 2015* (presenta un'architettura supervisionata neurale profonda a massimizzazione di correlazione). La metodologia proposta si articola come illustrato nell'immagine seguente:



X e Y sono le due immagini da comparare e i colori rappresentano l'appartenenza a domini differenti; la 'change map' è la mappa dei cambiamenti che vorremmo ottenere, ossia quella che indica i pixel rappresentanti porzioni di terreno cambiate tra un'acquisizione e l'altra. Il metodo si basa su una trasformazione che permetta di avere l'immagine X nel dominio di Y e l'immagine Y nel dominio di X. Per rendere più chiaro e senza perdita di generalità, possiamo supporre X come un'immagine ottica e Y come un'immagine radar (per esempio SAR): la capacità di traslare un'immagine nel dominio dell'altra ci permette di effettuare un confronto come se esse fossero acquisite da sensori omogenei. Dallo schema si evince che questa traslazione avviene per entrambe le immagini, e che si avranno due comparazioni e i due risultati verranno mediati.

L'operazione di traslazione è portata a termine tramite una rete neurale profonda, composta da due autoencoder, i quali, per essere allenati (fase di training della rete), hanno bisogno di esempi. Tuttavia, il nostro scopo era mantenere l'architettura non supervisionata, pertanto abbiamo sfruttato un metodo che consentisse l'uso di un'informazione a priori, estratta direttamente dalle due immagini mediante un procedimento basato su matrici di affinità. Esse sono costruite in modo da individuare pixel simili all'interno dell'immagine con un procedimento basato su distanze con kernel gaussiani (cfr. Eq. 4.1). Da queste due matrici è possibile ottenere un'informazione che abbiamo interpretato come la "probabilità" che un pixel sia cambiato tra un'acquisizione e l'altra. Essa è stata usata all'interno della rete per il training, che è avvenuto pesando di più i pixel a minor probabilità di cambiamento. In questo modo abbiamo allenato la rete a fare una traslazione di dominio per quelle parti di immagini di pixel non cambiati.

Per il suo funzionamento, questa rete ha bisogno che gli spazi di codice (code space) degli autoencoders appartengano allo stesso dominio, insieme Z nella figura sottostante. Ciò è stato assicurato sia da un'adeguata scelta delle funzioni di penalizzazione per l'allenamento della rete, sia dalla presenza di un blocco utile a far sì che i codificatori (encoders) massimizzino la correlazione tra le loro uscite. Il concetto è illustrato nella seguente immagine:



X e Y sono sempre la coppia di immagini (e anche i rispettivi domini),  $R(\cdot)$ ,  $P(\cdot)$  sono gli encoder,  $Q(\cdot)$ ,  $S(\cdot)$  i decoder; Z rappresenta invece il dominio comune nello spazio latente degli encoder. CCA è l'acronimo di "Canonical Correlation Analysis", il blocco che permette di calcolare la correlazione. Questa particolare

configurazione è chiamata in letteratura “Deep Canonical Correlated Autoencoders (DCCA)”, ma è la prima volta, per quanto ci è dato sapere, che viene utilizzata in maniera non supervisionata.

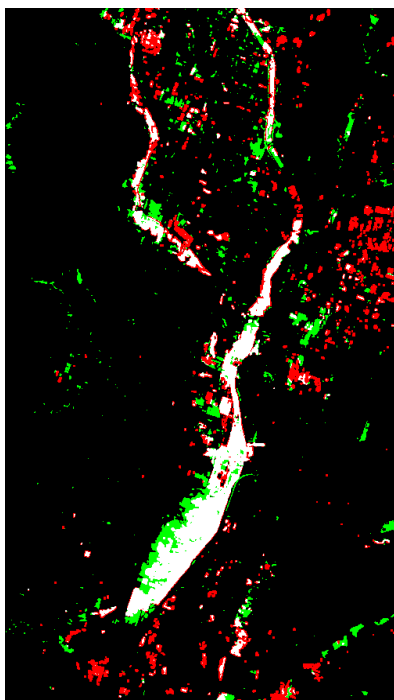
Per tutti i dettagli tecnici dell’architettura si rimanda al capitolo 4 della tesi.

La nuova architettura proposta è stata validata con due dataset differenti:

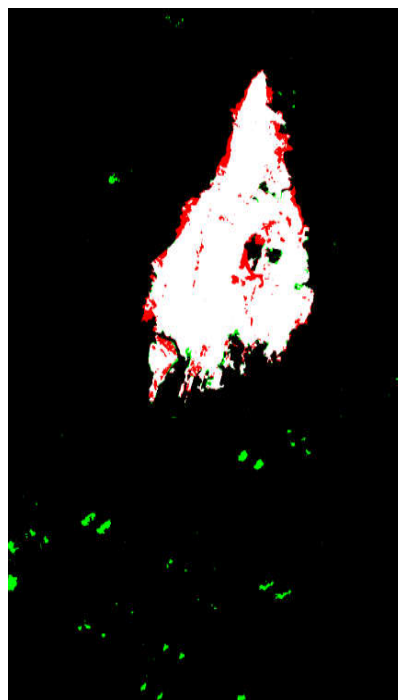
- Il primo rappresenta tre contee della California che sono state vittima di un’alluvione; delle due immagini a nostra disposizione, la prima acquisita dal Landsat 8 (RGB+TIRS), la seconda invece da Sentinel-1A è un’immagine puramente radar con 3 canali (VV, VH, rapporto tra VV e VH). Ai dati è stata applicata una trasformazione logaritmica per poter operare sull’immagine SAR.
- Il secondo è invece composto da due immagini acquisite con modalità più simili tra loro, perché entrambe hanno canali RGB e TIRS, anche se differenti tra loro, e una delle due acquisizioni ha anche un canale pancromatico. Le immagini riportano cambiamenti dovuti ad un incendio nello stato del Texas.

Gli esperimenti sono stati condotti cento volte, in quanto l’allenamento della rete neurale non avviene in modo deterministico. Questo ha permesso di ottenere dei risultati e di valutare l’accuratezza degli stessi. Il metodo è stato inoltre confrontato con metodologie similari adottate a riferimento come lo stato dell’arte e il nostro metodo, in entrambi i dataset, è stato in grado di ottenere risultati migliori.

I risultati sono riportati in funzione della metrica scelta, la  $K$  di Cohen. Essa è stata scelta poiché molto usata in letteratura per questa famiglia di metodi. Inoltre perché è simile alla misura di accuratezza, ma differisce da essa in quanto tiene conto della possibilità che una quantità di pixel possa essere classificata correttamente anche operando in maniera casuale. Inoltre, dato che il nostro problema riguarda la classificazione binaria, abbiamo considerato anche gli errori e la loro tipologia. Nei risultati riportati di seguito, i pixel bianchi (cambiamento) e neri (non-cambiamento) sono i pixel correttamente classificati, i verdi (falsi positivi) e i rossi (falsi negativi) sono gli errori divisi per tipo:



California dataset, Kappa = 0.51.



Texas dataset, kappa = 0.90.

Texas Results			California Results		
	<i>Mean <math>\pm \sigma</math></i>	<i>Max</i>		<i>Mean <math>\pm \sigma</math></i>	<i>Max</i>
<i>Cohen's kappa</i>	0.83 $\pm$ 0.12	0.91	<i>Cohen's kappa</i>	0.45 $\pm$ 0.04	0.51
<i>Overall Accuracy</i>	0.97 $\pm$ 0.02	0.98	<i>Overall Accuracy</i>	0.91 $\pm$ 0.01	0.93
<i>Time</i>	18 [min] $\pm$ 3 [s]	18.4[ <i>min</i> ]	<i>Time</i>	9 [min] $\pm$ 3 [s]	9 .17[ <i>min</i> ]

Immagini e tabelle mostrano i risultati e i tempi richiesti per ottenerli. Le metriche scelte sono state appunto la  $K$  di Cohen e l'accuratezza totale. Il tempo invece è stato misurato a partire da una rete totalmente 'vuota' fino alla fine del processo di allenamento e di valutazione dell'immagine. I tempi ovviamente cambiano in base alle dimensioni in pixel delle immagini.

In conclusione, la tesi ha affrontato il rilevamento automatico di cambiamenti non supervisionato da immagini eterogenee, e ha proposto una nuova architettura allo scopo. I risultati sono molto buoni, tuttavia prospettive di miglioramento possono essere individuate nella formulazione dell'informazione a priori, e possibilmente nell'integrazione di un sistema di apprendimento incrementale, che permetta di risparmiare tempo di calcolo ed eviti l'allenamento della rete per ogni coppia di immagini presentata.

*Tutti i riferimenti e le citazioni qui presenti rimandano direttamente alla tesi e alla sua bibliografia.*