



ISTI Technical Reports

Analisi di immagini tomografiche ad alta risoluzione attraverso reti neurali convoluzionali per lo studio delle interstiziopatie polmonari

Rossana Buongiorno, ISTI-CNR, Pisa, Italy

Sara Colantonio, ISTI-CNR, Pisa, Italy

Danila Germanese, ISTI-CNR, Pisa, Italy



Analisi di immagini tomografiche ad alta risoluzione attraverso reti neurali convoluzionali per lo studio delle interstiziopatie polmonari
Buongiorno R.; Colantonio S.; Germanese D.
ISTI-TR-2020/007

Le interstiziopatie polmonari (Interstitial Lung Disease, ILD) sono patologie croniche che causano la cicatrizzazione del parenchima polmonare e dell'interstizio alveolare e la compromissione della funzionalità respiratoria. Dal momento che sono più di 200 le patologie raggruppate nella categoria delle ILD, una precisa identificazione è fondamentale per individuare la terapia migliore e formulare una prognosi. L'esame radiologico di riferimento è la tomografia computerizzata del torace ad alta risoluzione (High Resolution Computed Tomography, HRCT) e rappresenta un passaggio cruciale nel processo di diagnosi; nell'analizzare le immagini, infatti, il radiologo deve stabilire se vi è Usual Interstitial Pneumoniae (UIP), ovvero presenza di pattern istopatologici tipici della malattia, e valutarne l'estensione, correlata con la gravità delle alterazioni fisiologiche. Tuttavia, l'incidenza rara delle interstiziopatie fa sì che non tutti i radiologi abbiano un grado di esperienza adatto a individuare visivamente l'anomalia. Inoltre, la malattia si diffonde lungo tutti i polmoni e la segmentazione manuale risulta faticosa. Nel tentativo di rimediare alla variabilità intra- ed inter-osservatore, sono state sviluppate tecniche per il riconoscimento automatico dei pattern UIP; vi sono approcci basati sull'analisi dell'istogramma e della texture dell'immagine ma, dal momento che i classificatori sono stati addestrati su label definite da operatori clinici diversi, presentano comunque un bias che è causa di identificazioni errate, o mancate, dei pattern. Il deep learning, invece, si distingue dalle tecniche tradizionali perché fornisce strumenti che imparano autonomamente a classificare i dati. L'obiettivo del lavoro è stato, quindi, progettare e sviluppare la UIP-net, una rete neurale convoluzionale ad-hoc per la segmentazione automatica dei pattern UIP in immagini HRCT di pazienti con Fibrosi Idiopatica Polmonare (IPF), che è una sotto-categoria delle ILD.

Keywords: Image Processing, Convolutional Neural Network, Medical Imaging, Segmentation, Deep Learning, Artificial Intelligence.

Citation

Buongiorno R.; Colantonio S.; Germanese D. *Analisi di immagini tomografiche ad alta risoluzione attraverso reti neurali convoluzionali per lo studio delle interstiziopatie polmonari*. ISTI Technical Reports 2020/007. DOI: 10.32079/ISTI-TR-2020/007.

Introduzione

Le interstiziopatie polmonari (*Interstitial Lung Disease*, ILD) sono patologie croniche che causano la cicatrizzazione del parenchima polmonare e dell'interstizio alveolare e la compromissione della funzionalità respiratoria. Dal momento che sono più di 200 le patologie raggruppate nella categoria delle ILD, una precisa identificazione è fondamentale per individuare la terapia migliore e formulare una prognosi. L'esame radiologico di riferimento è la tomografia computerizzata del torace ad alta risoluzione (*High Resolution Computed Tomography*, HRCT) e rappresenta un passaggio cruciale nel processo di diagnosi; nell'analizzare le immagini, infatti, il radiologo deve stabilire se vi è *Usual Interstitial Pneumoniae* (UIP), ovvero presenza di pattern istopatologici tipici della malattia, e valutarne l'estensione, correlata con la gravità delle alterazioni fisiologiche. Tuttavia, l'incidenza rara delle interstiziopatie fa sì che non tutti i radiologi abbiano un grado di esperienza adatto a individuare visivamente l'anomalia. Inoltre, la malattia si diffonde lungo tutti i polmoni e la segmentazione manuale risulta faticosa.

Nel tentativo di rimediare alla variabilità intra- ed inter-osservatore, sono state sviluppate tecniche per il riconoscimento automatico dei pattern UIP; vi sono approcci basati sull'analisi dell'istogramma e della texture dell'immagine ma, dal momento che i classificatori sono stati addestrati su label definite da operatori clinici diversi, presentano comunque un bias che è causa di identificazioni errate, o mancate, dei pattern.

Il deep learning, invece, si distingue dalle tecniche tradizionali perché fornisce strumenti che imparano autonomamente a classificare i dati. L'obiettivo del lavoro è stato, quindi, progettare e sviluppare la **UIP-net**, una rete neurale convoluzionale ad-hoc per la segmentazione automatica dei pattern UIP in immagini HRCT di pazienti con Fibrosi Idiopatica Polmonare (IPF), che è una sotto-categoria delle ILD.

Metodi

La UIP-net ha una struttura *Encoder-Decoder*, adatta ad analizzare immagini risolte limitando la perdita di informazione dovuta all'elaborazione dei dati da parte di strati consecutivi della rete e ottimizzando le risorse computazionali a fronte dell'elevato contenuto informativo da processare. Infatti, l'esame HRCT, rispetto alle comuni metodiche di imaging, fornisce immagini a elevata risoluzione spaziale e profondità del pixel, codificato con 12 bit. La rete Encoder-Decoder è stata progettata per la segmentazione della cartilagine di ginocchio[1]; dopo aver individuato le differenze sostanziali tra le due applicazioni, sono state effettuate delle

modifiche alla versione originale. Infatti, quest'ultima classifica i pixels sulla base dell'*intensità* dei livelli di grigio; la segmentazione dei pattern UIP nei campi polmonari, invece, si basa sulla texture, quindi sulla *differenza* delle intensità dei livelli di grigio di pixels adiacenti.

L'architettura finale dell'UIP-net si compone di uno strato di input che riceve un'immagine di dimensioni 492×492 , tre strati di codifica e tre strati di decodifica convolutivi, ad ognuno dei quali segue uno strato di attivazione ReLU, a eccezione dell'ultimo in cui si applica una funzione di attivazione tangente iperbolica. Gli strati di codifica riducono la dimensione delle immagini in input, eseguendo un'operazione di convoluzione con un filtro 5×5 , nel primo strato, e 3×3 negli strati successivi. Gli strati di decodifica, invece, sovra-campionano i dati ingrandendo le immagini progressivamente. Gli output degli strati di codifica e decodifica sono delle *features maps*, ovvero le risposte delle ROI delle immagini ispezionate ai filtri applicati. Ad ogni step di codifica, il numero di features maps raddoppia; al contrario, si dimezza durante la decodifica. Infine, l'ultimo strato è convolutivo e assembla tutte le features maps in un'immagine con le stesse dimensioni dell'input.

Dopo averne definito l'architettura, sono stati scelti l'entropia incrociata binaria come funzione di costo, l'*Adaptive Moment Estimation* come funzione di ottimizzazione e l'indice di Dice come metrica di accuratezza per l'addestramento della rete.

Il dataset per l'addestramento e il test della UIP-net contava 20 HRCT, con circa 250 slices ognuna, di pazienti affetti da IPF, messe a disposizione dall'Unità Operativa Radiodiagnostica 2 dell'AOUP di Cisanello. Da queste, è stata ottenuta la *ground truth* attraverso *Computer-Aided Lung Informatics for Pathology Evaluation and Ratings* (CALIPER), software di riferimento per la quantificazione dei pattern UIP in pazienti affetti da ILD. CALIPER si basa sull'analisi della texture dell'immagine, è stato ottenuto dall'elaborazione di una quantità consistente di dati e validato su test di funzionalità respiratoria[2]; elaborando le intere scansioni volumetriche, fornisce in output le singole slices segmentate, sottoforma di immagini RGB, per la differenziazione dei pattern UIP. Dalle immagini RGB sono state ottenute delle maschere binarie che avessero pixels uguali a 1 se appartenenti ad uno dei pattern UIP e pixels uguali a 0 se appartenenti al tessuto sano.

Dopo aver ottenuto la ground truth, le immagini sono state pre-processate per abbattere il tempo computazionale necessario per l'addestramento, riducendo sia le dimensioni del campo visivo dell'immagine, che il numero di pixels da analizzare. Per questo, è stato implementato un algoritmo di clustering non esclusivo, il *fuzzy C-mean*, che associasse a ogni pixel dell'immagine il grado di appartenenza a due classi: campi polmonari e fondo o regione addominale. L'algoritmo, implementato in Matlab®, ha fornito i centri dei due clusters e il grado di appar-

tenenza dei pixels a ognuno di loro; a partire da questi, è stata ottenuta una maschera binaria con valori dei pixels a 1 se il grado di appartenenza al fondo o ai polmoni superava il 70%, 0 altrimenti. In seguito, con delle operazioni morfologiche e aritmetiche, sono stati eliminati i pixel rumorosi del fondo e dell'addome. Dopo averla ottenuta, l'immagine a livelli di grigio contenente solo i pixel appartenenti ai polmoni è stata adattata alle dimensioni dello strato di input della rete; quindi, da un'immagine 512×512 , è stata ottenuta un'immagine 492×492 , attraverso un crop centrato nel centro dell'immagine, eseguito anche sulla ground truth, affinché il matching delle due matrici fosse perfetto.

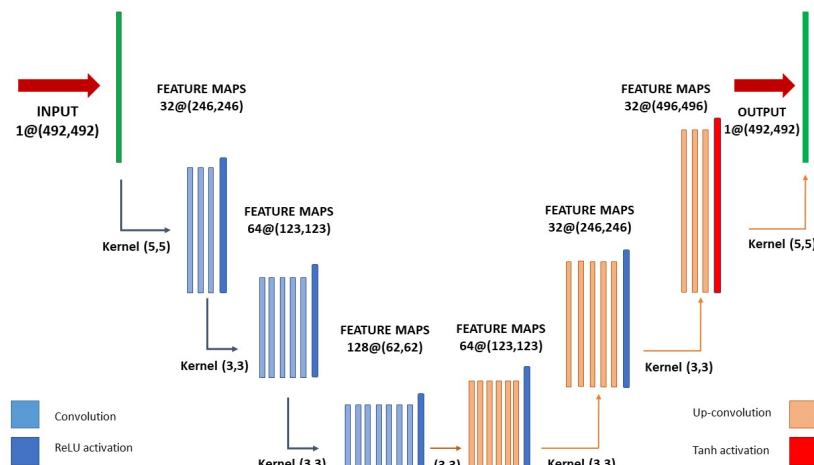


Fig 1. Architettura della UIP-net.

Risultati

L'addestramento della UIP-net è stato eseguito in Python 3.7 attraverso Keras, che utilizza come back-end la libreria TensorFlow. Sono state utilizzate 13 delle 20 TC ma, dal momento che lo strato di ingresso della UIP-net riceve immagini 2-D, il training set intero contava 3200 immagini. Per individuare la strategia di addestramento migliore per limitare il fenomeno dell'*overfitting*, sono stati eseguiti 20 addestramenti: variando la numerosità del dataset mantenendo fisso il numero di epoche, sono stati forniti alla rete 200, 400, 800, 1600 e 3200 esempi; variando il numero di epoche mantenendo fisso il numero di esempi, la rete è stata addestrata per 50, 100, 150, 200 epoche. Tutti gli addestramenti sono stati eseguiti con la dimensione del batch costante e pari a 10. Il modello migliore è risultato quello addestrato con 800 campioni per 50 epoche, in quanto la funzione di perdita e la metrica di accuratezza, in termini di indice di Dice, calcolati sul training e validation set si assestavano pressappoco sugli stessi valori a ogni epoca. Il validation set è stato composto dal 30% dei dati totali di training, mescolati a ogni epoca. Una volta scelto il numero di campioni e di epoche coi quali addestrare la rete,

è stata eseguita una *k-fold cross validation*, imponendo k uguale a 5. Quindi, il training set è stato diviso in 5 gruppi di uguale numerosità e per 5 volte la rete è stata addestrata su 640 immagini e testata su 160. Il valor medio dell'indice di Dice, calcolato per ogni split della convalida incrociata, è $76,1\% \pm 4,45\%$.

Una volta addestrato il modello con convalida incrociata, è stato eseguito il test su 7 delle 20 TC, con un totale di circa 1800 slices, attraverso un'analisi quantitativa e un'analisi qualitativa. L'analisi quantitativa, eseguita in Matlab®, è consistita nel calcolo di quattro indici di confronto con la ground truth: indici di Dice, BF-score, sensibilità e specificità; l'analisi qualitativa, invece, si è basata sul giudizio attribuito da un radiologo esperto che si è rivelato complessivamente positivo: la UIP-net risulta, spesso, più sensitiva. Tuttavia, rispetto a CALIPER, la UIP-net classifica erroneamente i vasi sanguigni e le vie aeree come appartenenti ai polmoni o al parenchima polmonare e include, spesso, anche organi esterni come l'intestino. Per far fronte alla problematica, attraverso il software 3D-Slicer, che implementa un algoritmo di segmentazione automatica per l'isolamento dei campi polmonari, dei vasi sanguigni e delle vie aeree, sono state ottenute delle maschere binarie, con le quali sono stati annullati i pixels erroneamente segmentati dalla UIP-net. Gli indici di qualità della performance sono stati calcolati sulle maschere prima e dopo le operazioni di correzioni in post-processing; in entrambi i casi, il BF-score e la sensibilità hanno presentato i valori massimi, arrivando rispettivamente a 83,73% e 84,6%, senza post-processing e 96,7% e 85,9%, con il post-processing.

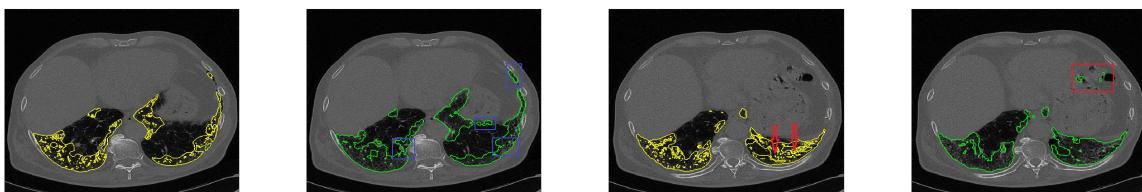


Fig 2. da sinistra a destra: segmentazione di CALIPER con i pattern mancanti, segmentazione UIP-net più sensitiva, segmentazione di CALIPER coi vasi sanguigni correttamente segmentati, intestino erroneamente segmentato dalla UIP-net.

Discussione

Le analisi quantitative e qualitative dei risultati hanno messo in luce diversi aspetti; innanzitutto, dal confronto tra l'indice di Dice calcolato durante la convalida incrociata in fase di training e quello calcolato sulle segmentazioni dell'insieme di test, è emerso che il valore del primo supera quello del secondo, perché le reti neurali predicono meglio i dati coi quali viene eseguito l'addestramento. In seguito, è emerso che l'indice di Dice non rispecchia il punteggio

visivo; infatti, è adatto per valutare le segmentazioni di regioni compatte ma non di malattie sparse come le interstiziopatie. Al contrario, il BF-score si è rivelato il più coerente con la valutazione qualitativa e presenta un valore sempre maggiore rispetto all'indice di Dice; inoltre, la specificità è sbilanciata a causa dei molti pixel nulli nelle maschere binarie delle regioni apicali dei polmoni (la malattia si concentra nei campi medi-inferiori).

Dai risultati sulle segmentazioni corrette in post-processing, si conclude che tutti gli indici aumentano; sebbene quello di Dice resti penalizzante, il BF-score arriva ad un valore massimo del 96,7%.

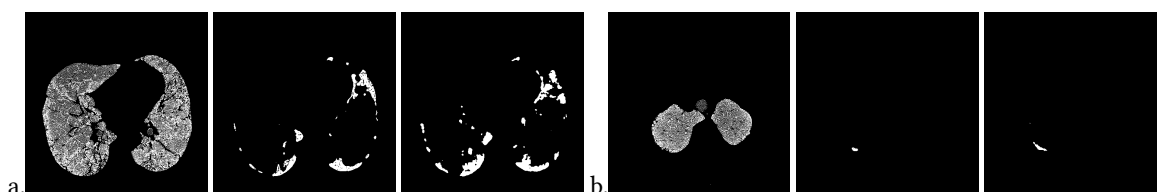


Fig 3. a): disomogeneità della malattia; b): immagine con pochi pixel segmentati ed elevata specificità. A sinistra: immagine originale; al centro: ground truth; a destra: segmentazione della rete.

Conclusioni

L'obiettivo del lavoro svolto era di indagare la capacità del deep learning di segmentare autonomamente le interstiziopatie polmonari in immagini HRCT. I risultati numerici ottenuti sugli indici di prestazione della UIP-net hanno dimostrato che le reti convoluzionali possono superare i limiti delle tecniche allo stato dell'arte, primo fra tutti il bias dovuto al giudizio degli operatori clinici di cui necessitano per classificare le immagini.

Ci sono dei limiti, però, che offrono spunti a sviluppi futuri. Innanzitutto, la numerosità del dataset: aggiungere campioni al test set può conferire maggior robustezza ai risultati della UIP-net; in seguito, addestrare la UIP-net su un insieme di immagini ricostruite con diversi kernel spaziali e acquisite con macchinari diversi potrebbe aumentare la riproducibilità dei risultati; in questo lavoro, infatti, la UIP-net è stata addestrata con immagini ricostruite attraverso un solo kernel e acquisite da un solo scanner. Infine, aumentando il numero di campioni del training set, l'architettura della UIP-net potrebbe essere arricchita, così da riuscire a rilevare molte più caratteristiche e a ottenere una differenziazione più fine dei pattern UIP.

Inoltre, data l'affinità dei pattern nelle due patologie, la UIP-net potrebbe essere prima testata, poi migliorata, per segmentare anche le immagini HRCT di pazienti affetti da Covid-19.

Riferimenti bibliografici

- [1] Fang Liu e al. «Deep Convolutional Neural Network and 3D Deformable Approach for Tissue Segmentation in Musculoskeletal Magnetic Resonance Imaging.» In: *Magnetic Resonance in Medicine* 79 (2018), pp. 2379–2391.
- [2] JJacob e al. «Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures.» In: *Eur Respir* (2017).