Innovation and Technology Transfer

Reti Convoluzionali per la Segmentazione di Lesioni Cutanee in Immagini Dermoscopiche.†

A. Youssef,[a], D. D. Bloisi,[b] M. Muscio,[c] A. Pennisi,[d] D. Nardi,[a] A. Facchiano.[c]

1 Introduzione

Il melanoma è una delle più aggressive e letali forme di tumore. I decessi indotti dal melanoma costituiscono circa l'80% delle morti per cancro della pelle. La dermoscopia è uno dei più importanti strumenti a disposizione degli specialisti per la diagnosi precoce del melanoma. Si tratta di una tecnica non invasiva ed economicamente conveniente per il riconoscimento dei tumori della pelle utilizzando l'analisi di immagini. La dermoscopia si è dimostrata efficace nel ridurre il numero di diagnosi di tumore presunto da confermare istologicamente usando la biopsia.[1] Le immagini dermoscopiche sono ottenute combinando ingrandimenti ottici con luce polarizzata o immersione in liquido con illuminazione a basso angolo di incidenza. Tramite l'analisi dermoscopica è possibile estrarre informazioni sulla struttura dei bordi della lesione, sulla asimmetria e le irregolarità della lesione, sulla presenza di aree blu o bianche, reti pigmentate e globuli. Inoltre, si può tenere traccia della storia evolutiva della lesione per creare una diagnosi basata proprio sull'evoluzione nel tempo della lesione. Tuttavia, l'analisi automatica di immagini dermoscopiche è resa diffcile dalla grande varietà di forme, colori e dimensioni delle lesioni, i differenti tipi e texture di pelle, uniti alla possibile presenza di artefatti nelle immagini (per esempio, peli e bollicine di olio o aria).

2 Methods

L'approccio proposto si basa sull'uso di una rete neurale di tipo convoluzionale per la segmentazione a livello di singolo pixel. In particolare, si fa uso di una struttura di rete di tipo encoder/decoder,[2] , dove ogni encoder ha il suo decoder seguito da uno strato di classificazione. Si sottolinea il fatto che, nelle reti encoder/decoder, il decoder è sempre realizzato tenendo in considerazione l'architettura dell'encoder per poter produrre una mappa di feature con la stessa risoluzione dell'input. Così come avviene nella rete SegNet,[2]ogni strato nella rete di decodifica è connessa con il corrispondente strato di codifica e una politica max-pooling è utilizzata per poter trasferire i dati da uno strato all'altro. Le feature map prodotte dall'enconder sono utilizzate come input dagli straati convoluzionali del decoder. Tali strati eseguono operazioni ripetute di convoluzione sulle feature map per poter generare delle mappe dense. Al termine della fase di convoluzione, il classificatore soft-max produce immagini a 3 canali, con etichette relative alle classi "pelle", "lesione" e "sconosciuto". Le etichette vengono assegnate in base alla massima probabilità che un pixel appartenga ad una delle tre classi. In questo lavoro sono stati utilizzati un econder con 4 strati convoluzionali e un encoder basato sulla rete VGG16[3] avente 13 strati convoluzionali. Si evidenzia che gli strati convoluzionali, gli strati di pooling e i classificatori sia dell'encoder che del decoder sono stati addestrati ex-novo senza utilizzare tecniche di fine-tuning o learning transform. I problemi legati al possibile overfitting causati dal numero ridotto di immagini sono stati mitigati adottando tecniche di dropout in ognuno degli strati convoluzionali dell'encoder.[4]

Encoder a 13 strati convoluzionali. L'architettura di rete VGG16 permette di effetture object classification ed è usata come encoder (si veda la parte sinistra della Fig. 1).Tale rete consiste di 13 strati convoluzionali con un kernel di dimensione 3x3. Vengono usati, inoltre, un kernel di dimensione 2x2 per le operazioni di max-pooling, la batch normalization, e una funzione di attivazione di tipo ReLU.[5] Il decoder (mostrato nella parte destra della Fig. 1) è progettato con 13 strati convoluzionali, 5 strati di up-sampling e un classificatore di tipo soft-max pixel-wise.[2]

Encoder a 4 strati convoluzionali. La struttura della rete a 4 strati è mostrata in Fig. 2. L'architettura dell'encoder comprende 4 strati convoluzionali, di cui i primi due (denominati conv1 e conv2) aventi un kernel di dimensione 3x3 e 64 filtri, mentre gli ultimi due (denominati conv3 e conv4) presentano kernel da 7x7 con 32 filtri. L'output di ogni strato convoluzionale è rettificato utilizzando una funzione di attivazione di tipo ReLU. Gli strati conv1, conv2 e conv4 sono seguiti da un modulo che esegue operazioni di max-pooling senza overlapping con un kernel 2x2. Lo strato conv3 è seguito direttamente da conv4 per poter

[a] *Sapienza Università di Roma*
[b] *Università della Basilicata email: domenico.bloisi@unibas.it*
[c] *IDI-IRCCS*
[d] *Storelift*

† presentato a @ITIM 2019 - 19° Congresso Nazionale Associazione Italiana di Telematica ed Informatica Medica 11-12 Novembre 2019, Matera/Potenza.

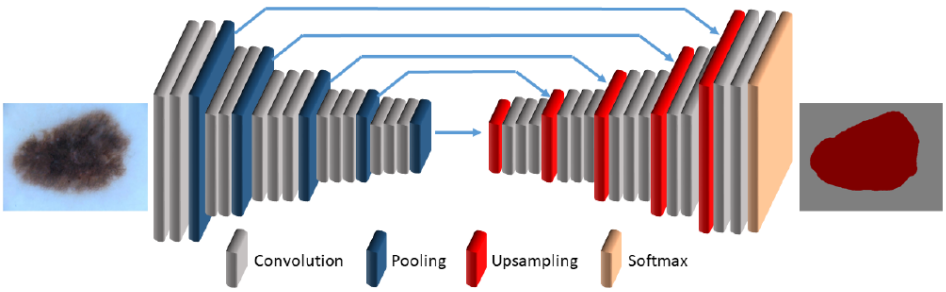

Fig. 1 Architettura della rete encoder-decoder a 13 strati convoluzionali. La rete permette di effetture una segmentazione dell'immagine al livello di singolo pixel.

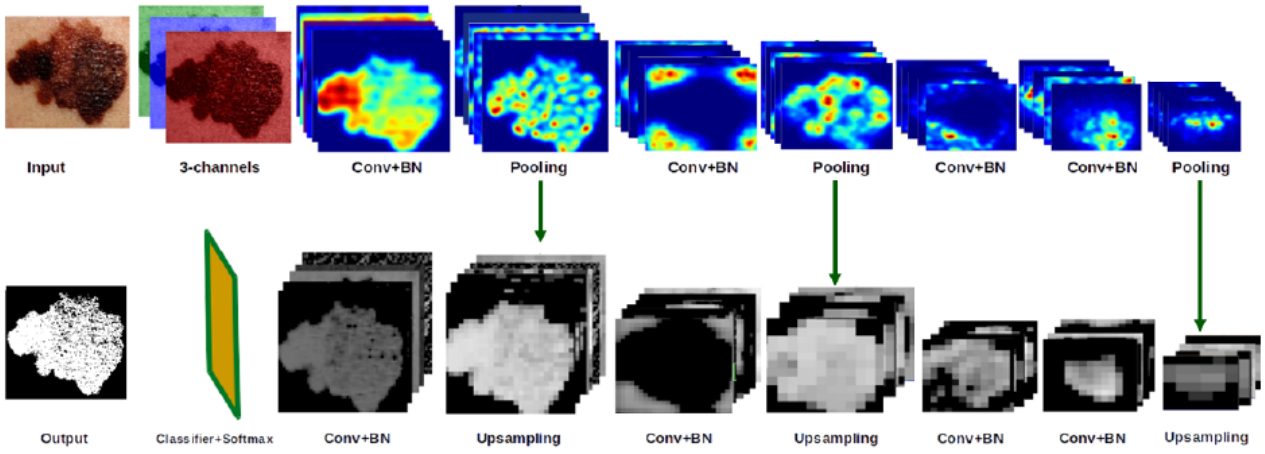

Fig. 2 Rete encoder-decoder a 4 strati. La figura mostra anche l'output di ogni singolo strato della rete.

apprendere i pesi delle operazioni di convoluzione. La tecnica del dropout è applicate ad ogni strato convoluzionale dell'encoder. Una procedura di upsampling viene effettuta usando gli indici provenineti da ogni operazione di pooling. Un classificatore soft-max pixel-wise viene usato per identificare una delle tre classi "pelle", "lesione" e "sconosciuto". La batch normalization è utilizzata dopo ogni strato convoluzionale e prima delle operazioni di pooling. La Fig. 2 mostra la capacità della rete di apprendere il valore dei pesi in base alla risposta dei neuroni per poter classificare ogni pixel dell'immagine in input come "pelle" o "lesione".

Bibliografia

1 L. Thomas, S. Puig, Dermoscopy, digital dermoscopy and other diagnostic tools in the early detection of melanoma and follow-up of high-risk skin cancer patients., Acta Dermato-Venereologica 97 (2017) 14–21.

2 A. Kendall, V. Badrinarayanan, R. Cipolla, Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding (2015). arXiv:1511.02680.

3 K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition (2014). arXiv:1409.1556.

4 N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research 15 (1) (2014) 1929–1958.

5 A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.