



Elaborazione di Immagini Mediche

Corso di Laurea Magistrale in Ingegneria Biomedica

Polyp Segmentation Challenge

Ricciardelli Fabio (333858)

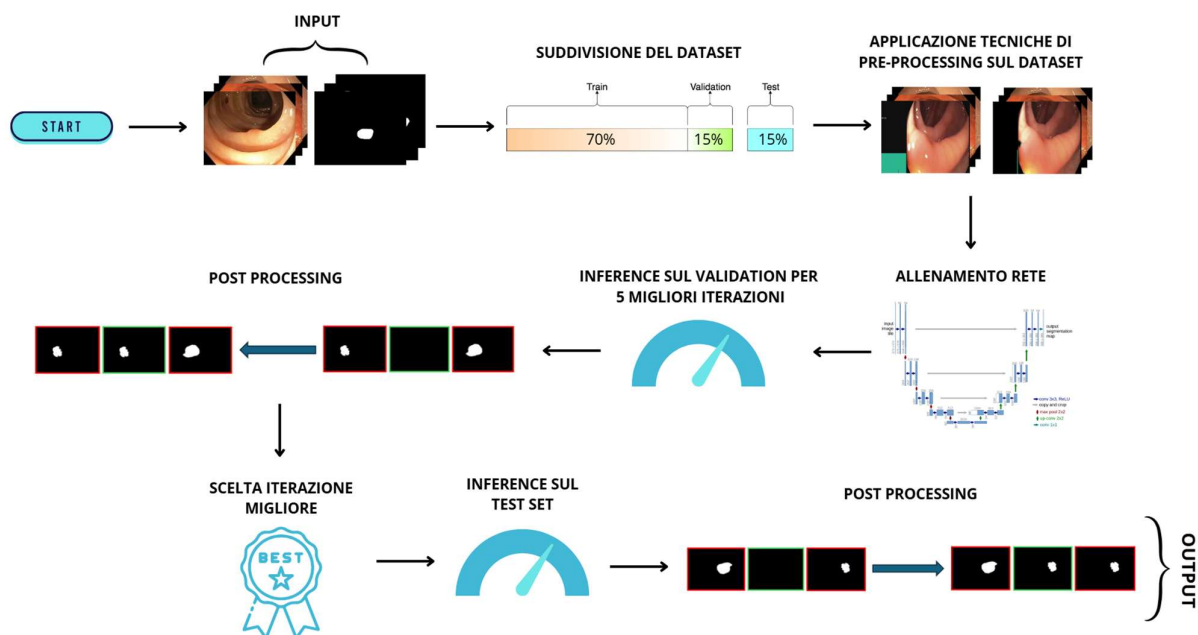
Spina Alessandro (324593)

Azzena Gabriele (319362)

1. INTRODUZIONE

La segmentazione dei polipi nel colon rappresenta un'importante sfida nel campo dell'elaborazione di immagini mediche, in quanto i polipi possono costituire lesioni precancerose e richiedono un'identificazione accurata durante la colonscopia [1]. Questo processo è cruciale per la prevenzione e la **diagnosi precoce** del cancro colorettale, che è una delle principali cause di morte oncologica a livello globale [2]. I polipi sono escrescenze anomale che si sviluppano sulla mucosa del colon e del retto. Sebbene molti polipi siano benigni, alcuni possono evolvere in adenomi, i quali rappresentano la fase iniziale della trasformazione in tumore maligno. Tuttavia, la rilevazione dei polipi può risultare complessa a causa della loro variabilità in termini di dimensioni, forma, colore e posizione all'interno del tratto intestinale. La segmentazione automatica dei polipi attraverso l'uso dell'intelligenza artificiale mira a supportare i medici nella localizzazione e classificazione delle lesioni, riducendo il rischio di diagnosi mancate, migliorando l'efficacia degli screening e riducendo l'incidenza dei tumori [3]. Tuttavia, la **variabilità dei dati**, dovuta a differenze nei dispositivi utilizzati dai diversi centri e nelle condizioni dei pazienti, rappresenta un ostacolo significativo per il task di segmentazione.

In questo contesto si colloca la *Polyp Segmentation Challenge*, che si pone l'obiettivo di sviluppare algoritmi capaci di segmentare accuratamente i polipi presenti nelle immagini di colonscopia. Una peculiarità della sfida è la necessità di garantire la coerenza tra frame consecutivi delle colonscopie. Per affrontare la sfida, uno degli approcci più utilizzati in letteratura per la segmentazione automatica è l'impiego di tecniche di *Deep Learning* basate su reti neurali convoluzionali. Tra queste, la rete **U-net** si distingue per la sua efficacia nella segmentazione semantica. L'obiettivo finale è progettare un sistema robusto e generalizzabile, capace di adattarsi alla variabilità dei dati clinici e di migliorare la qualità della diagnosi medica attraverso l'implementazione dei passaggi presentati in [Figura 1].

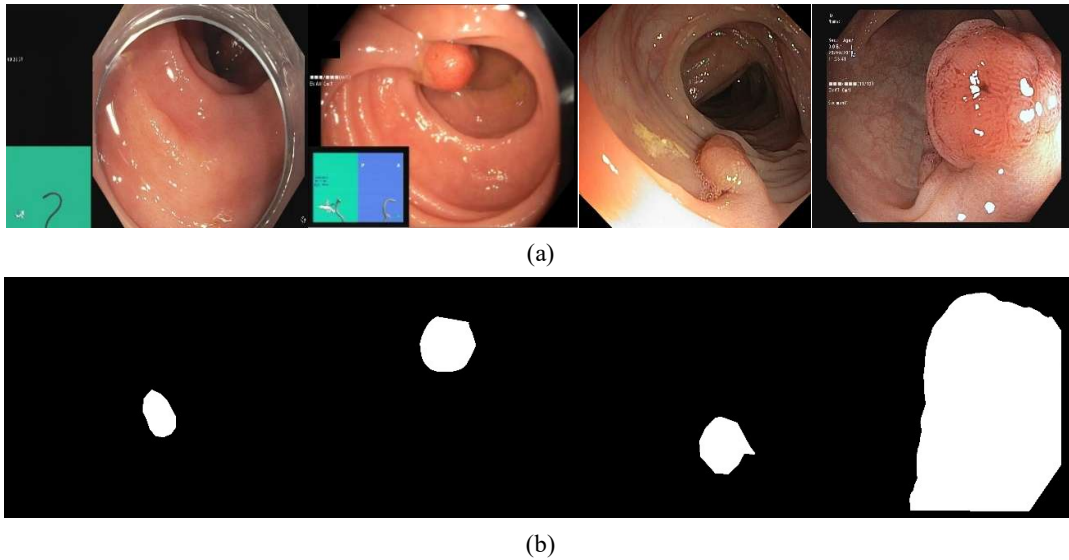


[Figura 1] Rappresentazione grafica della pipeline finale proposta.

2. MATERIALI E METODI

2.1 DataSet

Il dataset utilizzato per la challenge è denominato *PolypGen* e raccoglie dati provenienti da sei centri medici differenti, per un totale di oltre 300 pazienti. Sono disponibili 3762 tra immagini singole e sequenze video acquisite durante colonscopie. Le immagini sono fornite in formato JPG, con risoluzioni che variano da 384×288 a 1920×1080 pixel, rappresentando la natura **multicentrica** e **multistrumentale** del dataset. Inoltre, per ciascuna immagine, sono incluse maschere binarie in formato uint8 (0-255) che contengono le segmentazioni manuali dei polipi presenti nelle immagini. Le maschere sono state verificate da sei gastroenterologi esperti, garantendo un alto livello di affidabilità nelle annotazioni [4]. Nelle seguenti figure sono riportati esempi di immagini, appartenenti al dataset, e delle relative maschere manuali [Figura 2].



[Figura 2] Le immagini del dataset appartenenti a differenti centri (a) e le relative maschere (b)

2.2 Metriche per la valutazione dei risultati

Per la valutazione delle performance del sistema di segmentazione dei polipi, sono state adottate le seguenti metriche:

1. **Il coefficiente di similarità di Dice (DSC)** è utilizzato per misurare il grado di sovrapposizione tra la maschera manuale di riferimento e quella predetta automaticamente dall'algoritmo. Il valore del DSC varia tra 0, indicativo di nessuna corrispondenza tra le due maschere, e 1, che rappresenta una sovrapposizione perfetta.

$$DSC(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|}$$

Dove:

- X rappresenta i pixel della maschera manuale (*ground truth*).
- Y rappresenta i pixel della maschera predetta automaticamente.
- $|X \cap Y|$ indica il numero di pixel correttamente classificati come appartenenti alla lesione nella maschera predetta.

Per casi particolari in cui entrambe le maschere sono completamente vuote, ad esempio quando non sono presenti polipi, si otterrebbe una divisione indeterminata 0/0. In queste situazioni, si assegna un valore di Dice pari a 1, poiché l'algoritmo ha correttamente individuato l'assenza di lesioni.

2. **L'errore assoluto nella stima degli oggetti segmentati (AE)** calcola la differenza tra il numero di oggetti identificati dall'algoritmo nella maschera predetta e il numero di oggetti annotati manualmente nella maschera di riferimento.

$$Errore\ Assoluto = |N_p - N_g|$$

Dove:

- N_p è il numero di oggetti segmentati nella maschera predetta.
- N_g il numero di oggetti annotati nella maschera manuale.

3. **L'indice di stabilità temporale (TSI)** è stato introdotto per valutare la coerenza tra maschere predette in frame consecutivi di una sequenza video. Questa metrica è particolarmente importante per garantire che il sistema di segmentazione produca risultati consistenti, riducendo discontinuità o fluttuazioni non realistiche tra un frame e l'altro.

$$TSI = \frac{|Area_{t-1} - Area_t|}{Area_{t-1} + Area_t}$$

Dove:

- $Area_{t-1}$ è la maschera predetta precedente a quella considerata.
- $Area_t$ è la maschera predetta considerata.
- $|Area_{t-1} - Area_t|$ rappresenta il valore assoluto tra le aree di due frame consecutivi.
- $Area_{t-1} + Area_t$ indica la somma dei pixel segmentati nei due frame.

Un valore di TSI vicino a 0 indica un'ottima stabilità temporale, con variazioni minime tra i frame, mentre valori più vicini ad 1 segnalano una maggiore instabilità. È risultato opportuno adottare questa metrica in quanto i polipi, nei video colonscopici, sono strutture relativamente stabili e localizzate. Quindi, se la telecamera si sposta o ruota, i polipi rimangono generalmente visibili e occupano un'area simile tra frame consecutivi.

Queste metriche sono state calcolate per tutte le immagini che costituiscono le sequenze video.

2.3 Suddivisione del Dataset

Il dataset fornito è stato suddiviso in tre parti per garantire un'efficace valutazione del modello:

- il **train set** (70%), utilizzato per l'addestramento del modello;
- il **validation set** (15%), impiegato per ottimizzare i parametri e prevenire l'*overfitting*;
- il **test set** (15%), riservato per la valutazione finale delle performance su dati completamente nuovi.

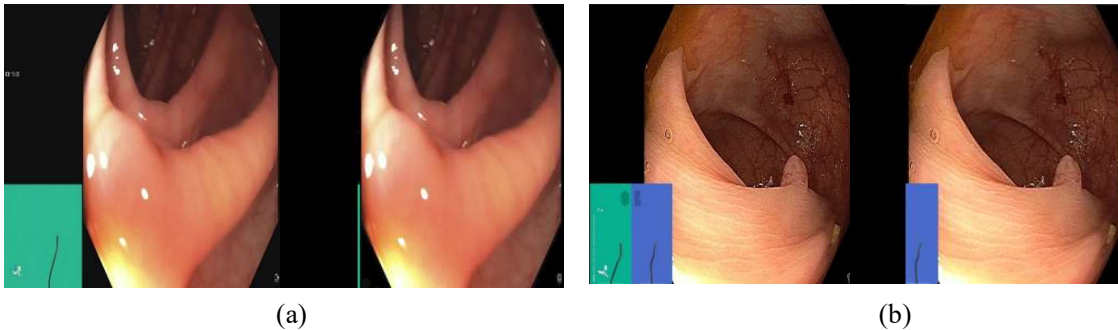
Tutte le immagini sono state ridimensionate a 256x256 per poterle dare in input alla rete.

2.4 Pre-processing

Di seguito vengono riportate le tecniche di pre-processing utilizzate ed implementate nell'algoritmo.

2.4.1 Regional Masking

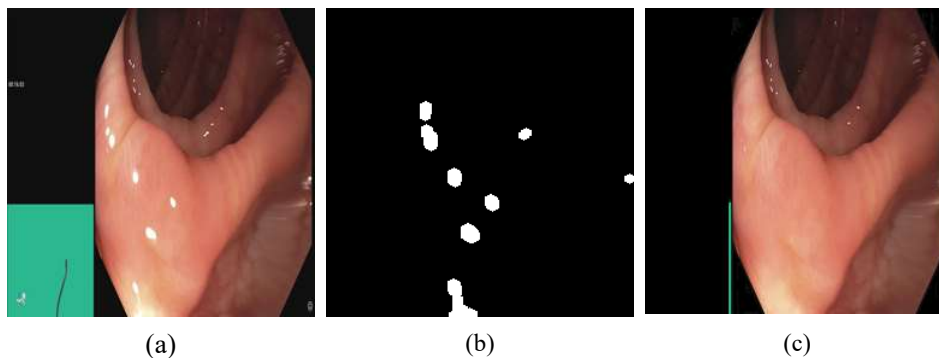
Il *regional masking* è stato adottato per **rimuovere dettagli non rilevanti** dalle immagini, come annotazioni e riquadri inseriti da alcuni centri sulle aree di interesse. Questo consente alla rete di concentrare l'apprendimento esclusivamente sulle regioni rilevanti, migliorando così la qualità della segmentazione e l'efficacia del modello. Per far questo le aree di disturbo sono state oscurate impostando i relativi valori di pixel a zero. Questo approccio è particolarmente utile per far fronte alla grande variabilità del dataset. Nelle seguenti immagini sono riportati degli esempi [Figura 3].



[Figura 3] Un esempio in cui il regional masking elimina tutti gli oggetti non rilevanti (a) e un esempio in cui non li rimuove (b)

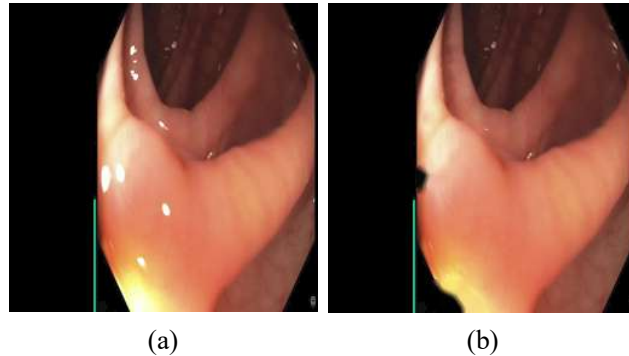
2.4.2 Inpainting

La tecnica di *inpainting* è stata implementata utilizzando il *fast marching method* implementato da *Alexandru Telea* [5]. Questa tecnica è stata selezionata per **ricostruire piccole regioni dell'immagine** in cui, per via dei **riflessi di luce**, i canali RGB saturano. Tali riflessi nell'immagine sono causati da un'illuminazione intensa, che risulta necessaria per permettere al medico di poter distinguere tutte le anomalie presenti sulla mucosa e di analizzare ogni minimo dettaglio dell'intestino. Per individuare gli artefatti causati dalla luce intensa è stata creata una maschera binaria in cui vengono posti ad 1 i pixel da correggere e a 0 quelli da mantenere invariati. Il metodo riempie le aree mancanti propagando i colori dei pixel noti verso quelli sconosciuti, basandosi sul gradiente dell'immagine, che indica sia la direzione che l'intensità delle variazioni nei valori dei pixel. Un esempio è riportato in [Figura 4].



[Figura 4] L'immagine originale (a), la maschera dei riflessi (b) e l'immagine corretta con il fast marching method (c)

Il metodo migliora visivamente tutte le immagini ma in alcune introduce anche delle piccole sbavature dovute al fatto che il difetto si trova vicino a regioni di bordo e perciò viene riempito con dei pixel neri [Figura 5].



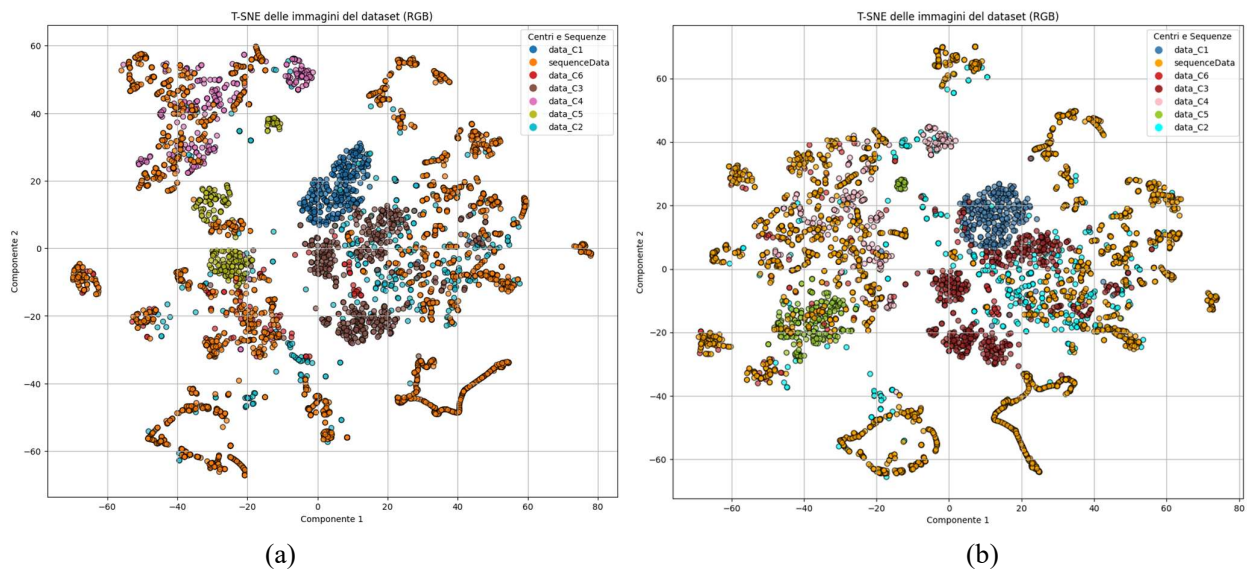
[Figura 5] Immagine originale (a) ed immagine corretta con dei piccoli difetti sul bordo (b)

2.4.3 Min-Max Scaling

Per garantire una rappresentazione consistente dei valori dei pixel e migliorare la convergenza del modello durante l'allenamento, è stato adottato un processo di normalizzazione basato sulla tecnica *Min-Max Scaling* nell'intervallo **0-255**.

2.5 Analisi della Variabilità e Valutazione del pre-processing tramite t-SNE

È stata condotta un'analisi della **variabilità** del dataset utilizzando il t-SNE (*t-Distributed Stochastic Neighbor Embedding*), una tecnica di riduzione dimensionale non lineare. Questo approccio ha permesso di visualizzare la distribuzione dei dati nello spazio a bassa dimensionalità, fornendo una rappresentazione chiara della loro variabilità e delle possibili sovrapposizioni tra le diverse caratteristiche presenti nel dataset. Inizialmente, è stato applicato il t-SNE sulle immagini originali. Successivamente, utilizzando lo stesso piano di proiezione, è stata ripetuta l'analisi sulle immagini sottoposte a pre-processing, al fine di verificare la reale riduzione della variabilità [Figura 6].



[Figura 6] Il t-SNE prima del dataset originale (a) e del dataset dopo aver applicato il pre-processing (b)

Inoltre, attraverso il calcolo della varianza è stata fornita un'evidenza quantitativa della riduzione della variabilità dei dati che è fondamentale per ottenere prestazioni più robuste e affidabili dalla rete neurale, come riportato in Tabella 1.

DATASET	VARIANZA
NO PRE-PROCESSING	2113.36
PRE-PROCESSING	2001,88

Tabella 1: Analisi quantitativa della varianza prima e dopo il pre-processing

2.6 Architettura e parametri della rete

Il modello utilizzato è implementato attraverso il framework *MMSegmentation* e si basa su una rete *U-Net* pre-addestrata su *Cityscapes*. Questo approccio sfrutta le conoscenze acquisite durante il pre-training su *Cityscapes*, aiutando a migliorare sia la velocità di addestramento che la robustezza del modello. La rete è configurata per gestire immagini con dimensioni 256x256x3. Il numero di classi è 2, corrispondenti a "background" e "object". In seguito, sono state eseguite le prove per ottimizzare gli **iperparametri** della rete:

- **Batch Size:** Sono stati provati valori pari a 4, 8, 16 e 24.
- **Epoche di allenamento:** Sono stati testati valori pari a 20 e 45 epoche.
- **Funzione di Loss:** Sono state valutate tre funzioni di perdita: *Cross Entropy Loss*, *Dice Loss* e *Tversky Loss*.
- **Weight loss:** Sono stati provati delle combinazioni dei pesi di *loss* differenti, lasciando sempre il peso del *decode head* pari a 1 e variando quelli dell'*auxiliary head* con i valori 0.2, 0.4 e 0.6.

La configurazione finale ha previsto una *backbone Unet* per l'encoder, una *FCNHead* come decoder e *auxiliary head*.

2.7 Post-processing

2.7.1 Eliminazioni delle regioni isolate

Il processo di post-processing ha incluso l'eliminazione delle regioni connesse che occupano **meno del 0.6%** dei pixel totali dell'immagine. Questo passo è stato adottato per rimuovere piccole zone isolate vicine alla maschera principale, che non fanno parte di essa. Un esempio viene riportato in [Figura 7].



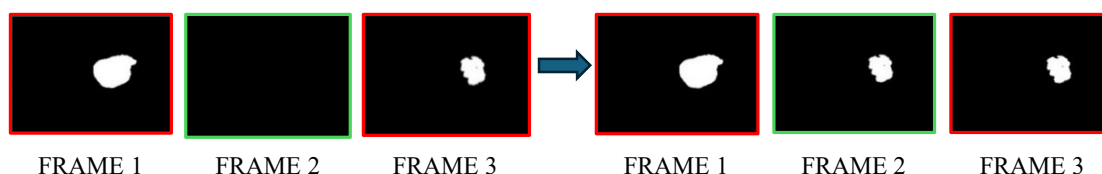
[Figura 7] La maschera predetta senza l'eliminazione delle regioni isolate, sulla sinistra, e la stessa maschera dopo l'eliminazione sulla destra

2.7.2 Correzione della coerenza temporale

Poiché la rete elabora ogni frame singolarmente, mentre la colonscopia è un video, è stato applicato un post-processing per migliorare la coerenza delle predizioni tra frame consecutivi. In particolare, viene inserito un polipo nel frame corrente se presente nei frame precedente e successivo attraverso i seguenti passaggi:

1. Si seleziona il **frame intermedio** in cui la rete non ha segmentato il polipo.
2. Viene effettuata un'**intersezione** tra i due frame adiacenti, per individuare le parti della maschera in comune.
3. La nuova regione segmentata viene inserita nel frame intermedio per garantire una **transizione più uniforme** e realistica nella sequenza.

Questa strategia si basa sull'assunto che la presenza di polipi in frame consecutivi sia altamente probabile in contesti di colonscopia. Tale strategia, non si applica invece a situazioni in cui il polipo è assente per più di un frame consecutivo, poiché ciò potrebbe rappresentare una vera discontinuità della lesione o un cambiamento nel campo visivo della colonscopia. Un esempio è riportato in [Figura 8].



[Figura 8] I tre frame considerati prima dell'aggiunta del polipo, a sinistra, e dopo l'aggiunta a destra

3. RISULTATI E DISCUSSIONE

In questa sezione sono riportati i risultati ottenuti applicando le tecniche di pre-processing e modificando i parametri dell'architettura della rete. La rete allenata solo con le immagini ridimensionate è stata utilizzata come punto di partenza per tutte le successive analisi. Dopo ogni fase di allenamento, sono state selezionate le **5 migliori iterazioni** e valutate sul Validation Set. L'iterazione con le performance migliori è stata riportata nelle tabelle seguenti ed è servita come riferimento per prendere tutte le decisioni. Infine, la migliore iterazione tra tutti i casi analizzati è stata utilizzata per eseguire l'*inference* sul Test Set. Le metriche sono presentate come media più o meno deviazione standard.

La Tabella 2 riporta i risultati derivanti dall'applicazione delle diverse tecniche di pre-processing. Con l'aggiunta dell'*inpainting* al *regional masking*, si osserva una diminuzione del DSC rispetto al caso con solo il *regional masking*. Questa operazione è però necessaria per applicare correttamente il *min-max scaling*, che porta ad una riduzione di variabilità dei dati, un aumento del dice e una riduzione del TSI. Per l'AE, invece, si nota un peggioramento che si cerca di risolvere con il post-processing.

MODALITÀ UTILIZZATA	DSC	ABSOLUTE ERROR	TSI
NO PRE-PROCESSING	0.5202 ± 0.4291	2.0035 ± 4.2588	0.2529 ± 0.3606
REGIONAL MASKING	0.5390 ± 0.4473	1.8081 ± 4.1799	0.2342 ± 0.3552
REGIONAL MASKING + INPAINTING	0.5287 ± 0.4259	1.6549 ± 4.2980	0.2579 ± 0.3719
REGIONAL MASKING + INPAINTING + MIN-MAX SCALING	0.5391 ± 0.4525	2.0423 ± 4.4296	0.2034 ± 0.3483

Tabella 2: Confronto delle performance ottenute combinando differenti tecniche di pre-processing.

Le diverse scelte degli iperparametri della rete sono state effettuate modificandone uno alla volta, mantenendo invariati gli altri. La Tabella 3 riporta i risultati derivanti dalla modifica della funzione di Loss, addestrando la rete per 20 epoche con batch size pari a 8 e valori di *weight loss* fissati rispettivamente a 1 per il *decode head* ed a 0.4 per l'*auxiliary head*. Si nota come la miglior funzione di loss è quella di Dice.

LOSS FUNCTION	DSC	ABSOLUTE ERROR	TSI
CROSS ENTROPY	0.4723 \pm 0.4419	1.8873 \pm 4.4252	0.2914 \pm 0.3830
DICE	0.5391 \pm 0.4525	2.0423 \pm 4.4296	0.2034 \pm 0.3483
TVERSKY (alpha=0.3 e beta=0.7)	0.1834 \pm 0.2722	2.3715 \pm 3.6830	0.1537 \pm 0.2347

Tabella 3: Confronto delle performance in funzione della Loss utilizzata.

La Tabella 4 illustra i risultati derivanti dalla modifica dei valori di *weight loss*. La combinazione di *weight loss* pari a 1 per il *decode head* ed a 0.2 per l'*auxiliary head* porta ad un aumento del DSC e dell'AE a fronte di un peggioramento del TSI.

WEIGHT LOSS	DSC	ABSOLUTE ERROR	TSI
DECODE HEAD: 1 AUXILIARY HEAD: 0.2	0.5492 \pm 0.4344	1.8028 \pm 4.0758	0.2244 \pm 0.3479
DECODE HEAD: 1 AUXILIARY HEAD: 0.4	0.5391 \pm 0.4525	2.0423 \pm 4.4296	0.2034 \pm 0.3483
DECODE HEAD: 1 AUXILIARY HEAD: 0.6	0.4963 \pm 0.4459	1.6673 \pm 4.1427	0.2605 \pm 0.3809

Tabella 4: Confronto delle performance al variare dei valori di weight loss.

Nella Tabella 5 vengono riportati i risultati ottenuti modificando il *batch size* utilizzato durante l'addestramento mantenendo invariato il numero di epoche. Il *batch size* di 8 risulta il più performante.

BATCH SIZE	DSC	ABSOLUTE ERROR	TSI
4	0.4654 \pm 0.4369	1.8151 \pm 3.9907	0.3408 \pm 0.4067
8	0.5492 \pm 0.4344	1.8028 \pm 4.0758	0.2244 \pm 0.3479
16	0.5185 \pm 0.4330	1.7095 \pm 4.1936	0.2787 \pm 0.3842
24	0.5367 \pm 0.4303	1.7764 \pm 4.2548	0.3086 \pm 0.3865

Tabella 5: Confronto delle performance al variare del batch size.

La Tabella 6 sintetizza i risultati ottenuti modificando il numero di epoche di addestramento. Utilizzando 45 epoche invece di 20 si ha un peggioramento di tutte le metriche.

NUMERO EPOCHES	DSC	ABSOLUTE ERROR	TSI
20	0.5492 \pm 0.4344	1.8028 \pm 4.0758	0.2244 \pm 0.3479
45	0.5020 \pm 0.4287	3.0493 \pm 4.7339	0.3697 \pm 0.3951

Tabella 6: Confronto delle performance al variare del numero di epoche.

Infine, nella Tabella 7 sono riportati i risultati derivanti dall'applicazione di diverse tecniche di post-processing sulle maschere generate. Con l'aggiunta della correzione della coerenza temporale nelle maschere predette si osserva una minima diminuzione del DSC accompagnata da una diminuzione della sua deviazione standard e un miglioramento dell'AE e del TSI. Per questa ragione si è preferito mantenere questo post-processing.

MODALITÀ UTILIZZATA	DSC	ABSOLUTE ERROR	TSI
NO POST-PROCESSING	0.5492 ± 0.4344	1.8028 ± 4.0758	0.2244 ± 0.3479
ELIMINAZIONE DELLE REGIONI ISOLATE	0.5926 ± 0.4301	1.3539 ± 4.3105	0.1777 ± 0.3250
ELIMINAZIONE DELLE REGIONI ISOLATE + CORREZIONE DELLA COERENZA TEMPORALE	0.5925 ± 0.4275	1.3451 ± 4.2971	0.1628 ± 0.2977

Tabella 7: Confronto delle performance in base al tipo di post-processing applicato.

La Tabella 8 evidenzia l'impatto del post-processing sul validation e test set. Si evidenzia che l'aggiunta del post-processing sul test set peggiora il DSC ma migliora l'AE e il TSI. Inoltre, il DSC nel test risulta superiore rispetto a quello del validation set, una differenza attribuibile alla natura casuale della suddivisione dei dati e all'elevata variabilità del dataset.

POST-PROCESSING	DSC	ABSOLUTE ERROR	TSI
TEST SET			
NO	0.6788 ± 0.3661	1.2007 ± 2.2294	0.1945 ± 0.3197
SI	0.6557 ± 0.3965	0.5115 ± 1.5874	0.1391 ± 0.2757
VALIDATION SET			
NO	0.5492 ± 0.4344	1.8028 ± 4.0758	0.2244 ± 0.3479
SI	0.5925 ± 0.4275	1.3451 ± 4.2971	0.1628 ± 0.2977

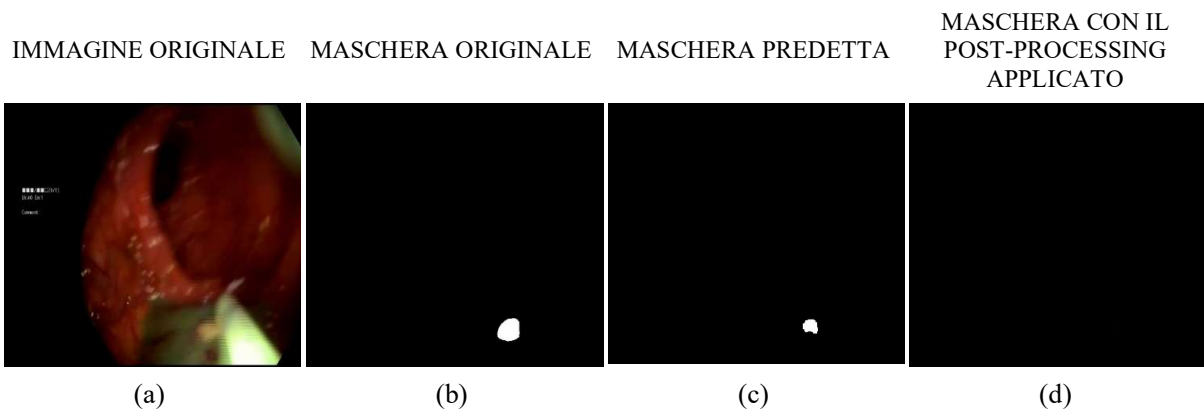
Tabella 8: Effetto del post-processing sulle performance nel validation e test set.

4. LIMITAZIONI E SVILUPPI FUTURI

La principale limitazione riscontrata è dovuta all'impossibilità di utilizzare il metodo implementato per applicazioni mediche **real time** come la colonscopia. Infatti, per la predizione delle maschere finali è richiesto un tempo legato alla generazione della maschera da parte dell'algoritmo e inoltre, per applicare l'aggiunta dei polipi come post-processing, vi è la necessità di ottenere prima tutte le maschere.

Un'ulteriore limitazione riguarda il dataset utilizzato per sviluppare il sistema, che include immagini provenienti da un **numero limitato** di centri e nazioni. In futuro, sarebbe interessante testare l'algoritmo su immagini provenienti da centri non rappresentati nel dataset e, se necessario, arricchire la varietà delle immagini utilizzate per l'addestramento, al fine di migliorare la capacità di generalizzazione del modello.

L'ultima limitazione riguarda la **scelta delle soglie** per l'eliminazione delle piccole regioni connesse durante il post-processing. A causa delle diverse dimensioni delle immagini, può accadere che la soglia risulti troppo elevata, portando alla cancellazione completa di polipi molto piccoli. Per affrontare questo problema, è stata implementata una strategia automatica che adatta la soglia in proporzione alle dimensioni dell'immagine. Sebbene questa soluzione abbia risolto il problema legato alla variabilità delle dimensioni, resta la difficoltà di gestire polipi talmente piccoli da occupare un numero di pixel inferiore alla soglia impostata. Questo limite è stato riscontrato in alcuni frame del test set che presentavano un polipo molto piccolo, di cui un esempio è riportato in [Figura 9].



[Figura 9] L'immagine originale (a), la maschera originale (b), la maschera predetta (c) e maschera predetta con il post-processing applicato (d)

5. CONCLUSIONI

L'obiettivo di questo lavoro è stato lo sviluppo di un sistema di segmentazione dei polipi basato su un algoritmo di deep learning. Questo algoritmo rappresenta un supporto fondamentale per il gastroenterologo, che quotidianamente esegue numerosi esami. La stanchezza accumulata nel corso della giornata può aumentare il rischio di errori, con potenziali implicazioni sulla salute del paziente. In questo contesto, l'algoritmo non solo funge da seconda opinione, ma aiuta anche a ridurre la probabilità di errori, migliorando l'affidabilità dell'esame.

In conclusione, il sistema proposto ha il potenziale di ottimizzare l'esame di colonscopia, con vantaggi economici derivanti dalla diagnosi più precisa e dalla possibilità di interventi tempestivi nella rimozione di polipi che potrebbero essere sospetti.

BIBLIOGRAFIA E SITOGRAFIA

- [1] <https://www.humanitas.it/news/polipi-intestinali-i-sintomi-e-le-cure/>
- [2] <https://www.airc.it/news/i-numeri-del-cancro-fotografia-dal-mondo>
- [3] <https://www.humanitas.it/news/polipi-intestinali-diagnosi-tipologie/>
- [4] Ali, S., Jha, D., Ghatwary, N. *et al.* A multi-centre polyp detection and segmentation dataset for generalisability assessment. *Sci Data* 10, 75 (2023). <https://doi.org/10.1038/s41597-023-01981-y>
- [5] Telea, Alexandru. (2004). "An Image Inpainting Technique Based on the Fast Marching Method." *Journal of Graphics Tools*, 9(1), 25–36. DOI: 10.1080/10867651.2004.10487596