



UNIVERSITÀ DEGLI STUDI DI CATANIA
DIPARTIMENTO DI MATEMATICA E INFORMATICA
CORSO DI LAUREA MAGISTRALE IN INFORMATICA

Riccardo Raciti

Studio Comparativo di Approcci
di Segmentazione dell'MRI Cerebrale
per Applicazioni Neurologiche

RELAZIONE PROGETTO FINALE

Relatore: Prof. Sebastiano Battiato
Correlatore: Prof. Daniele Ravì

Anno Accademico 2023 - 2024

Abstract

Il costante sviluppo delle tecnologie di imaging medico ha rivoluzionato la diagnosi e il monitoraggio delle condizioni mediche. La risonanza magnetica (MRI) si è affermata come una delle tecniche più cruciali, offrendo dettagli anatomici e funzionali di straordinaria precisione. Tuttavia, l'interpretazione accurata di queste immagini richiede un impegno considerevole da parte di specialisti quali neurologi e radiologi.

L'avvento dell'intelligenza artificiale, in particolare del deep learning, ha aperto nuove prospettive nell'analisi delle immagini mediche. Questa tesi si inserisce in questo contesto, con l'obiettivo di esplorare e valutare l'efficacia di tecniche tradizionali di segmentazione e di approcci di deep learning sviluppati appositamente per la segmentazione delle strutture anatomiche nelle immagini MRI.

L'analisi dettagliata di queste diverse metodologie, considerando metriche quantitative e qualitative insieme ai tempi di elaborazione, offre una panoramica completa delle capacità di ciascuna architettura, identificando i rispettivi vantaggi e limitazioni.

Questa ricerca rappresenta un passo avanti nello sviluppo degli strumenti di analisi medica basati sull'elaborazione di immagini. L'obiettivo non è solo migliorare la precisione e l'efficienza delle diagnosi basate sulle MRI, ma anche rendere queste metodologie più accessibili e affidabili per un'ampia gamma di operatori sanitari.

La completa comprensione delle performance delle diverse architetture nell'analisi delle immagini MRI contribuisce in modo rilevante all'evoluzione dell'intero campo dell'elaborazione delle immagini mediche. Questi progressi rappresentano un'opportunità per affinare e migliorare le pratiche diagnostiche, aprendo la strada a metodologie più sofisticate e precise nella fornitura e interpretazione delle informazioni fondamentali per la pratica medica.

Indice

1	Introduzione	5
1.1	Medical Imaging	6
1.2	Dataset	7
1.2.1	Formati delle Scansioni	8
1.2.2	Pre-processing	9
1.2.2.1	Normalizzazione Lineare	9
1.2.2.2	Resize	10
1.3	MRI del Cervello	12
1.3.1	Principi fondamentali della MRI	13
1.3.2	Utilizzo e applicazioni della MRI	13
1.3.3	T1	14
1.3.3.1	Principi fondamentali della sequenza T1 . . .	14
1.3.3.2	Caratteristiche e applicazioni della sequenza T1	14
1.3.3.3	Ruolo clinico e utilizzo in contesti di ricerca .	15
1.3.4	T2	16
1.3.4.1	Principi fondamentali della sequenza T2 . . .	16
1.3.4.2	Caratteristiche e applicazioni della sequenza T2	16
1.3.4.3	Ruolo clinico e utilizzo in contesti di ricerca .	17
1.3.5	FLAIR	18
1.3.5.1	Principi fondamentali della sequenza FLAIR .	19
1.3.5.2	Caratteristiche e applicazioni della sequenza FLAIR:	19
1.3.5.3	Ruolo clinico e utilizzo in contesti di ricerca .	19
1.3.6	PET	20
1.3.6.1	Principi fondamentali della PET	20
1.3.6.2	Applicazioni e utilizzo in neuroscienze	20
1.3.6.3	Ruolo clinico e utilizzo in contesti di ricerca .	21

2 Descrizione Modelli	22
2.1 FAST	22
2.1.1 Tecnologie Utilizzate	23
2.1.1.1 Hidden Markov Random Field Model	23
2.1.1.2 Expectation-Maximization Algorithm	24
2.1.2 Vantaggi e Svantaggi	25
2.1.2.1 Vantaggi	25
2.1.2.2 Svantaggi	25
2.2 SAM	26
2.2.1 Tecnologie Utilizzate	26
2.2.1.1 Image Encoder	26
2.2.1.2 Prompt Encoder	29
2.2.1.3 Mask Decoder	31
2.2.2 Modalità d'utilizzo	32
2.3 SynthSeg	33
2.3.1 Tecnologie Utilizzate	33
2.3.1.1 Modelli Generativi	33
2.3.1.2 Rete di Segmentazione	33
2.3.1.3 Inferenza	35
2.4 UniverSeg	37
2.4.1 Tecnologie Utilizzate	37
2.4.1.1 CrossBlock	37
2.4.1.2 Rete	38
3 Risultati	39
3.1 Risultati Qualitativi	40
3.1.1 FAST	41
3.1.1.1 Risultati	41
3.1.1.2 Valutazioni	41
3.1.2 SAM	43
3.1.2.1 Risultati	44
3.1.2.2 Valutazioni	45
3.1.3 SynthSeg	46
3.1.3.1 Risultati	46
3.1.3.2 Valutazioni	46
3.1.4 UniverSeg	47
3.1.4.1 Risultati	48
3.1.4.2 Valutazioni	48
3.2 Tempi Esecuzione	49
3.2.1 FAST	50
3.2.1.1 Parametri	50

<i>INDICE</i>	4
3.2.1.2 Input	52
3.2.1.3 Valutazione	52
3.2.2 SAM	53
3.2.2.1 Parametri	53
3.2.2.2 Input	54
3.2.2.3 Valutazioni	54
3.2.3 SynthSeg	55
3.2.3.1 Parametri	55
3.2.3.2 Input	58
3.2.3.3 Valutazioni	58
3.2.4 UniverSeg	59
3.2.4.1 Parametri	59
3.2.4.2 Input	59
3.2.4.3 Valutazioni	59
3.2.5 Tabella Riassuntiva	60
Conclusione	61
Bibliografia	63

Capitolo 1

Introduzione

Il cervello umano subisce cambiamenti strutturali nel corso della vita, inclusi processi di atrofia che possono essere indicatori di condizioni neurodegenerative come la malattia di Alzheimer, la sclerosi multipla e altre patologie. L'atrofia cerebrale può manifestarsi attraverso la riduzione del volume di specifiche regioni cerebrali, fornendo importanti indizi sullo sviluppo e la progressione di queste condizioni patologiche.

La segmentazione precisa delle strutture cerebrali nelle immagini risonanza magnetica *MRI* è fondamentale per valutare e quantificare l'atrofia cerebrale. Le tecniche di imaging, in particolare le sequenze pesate nel rilassamento T1, rilassamento T2 e Fluid Attenuated Inversion Recovery (FLAIR), forniscono una rappresentazione dettagliata delle diverse regioni del cervello, consentendo l'analisi delle variazioni nel volume e nella morfologia dei tessuti cerebrali. Questo studio si concentra sull'utilizzo di quattro differenti architetture di segmentazione FAST, SynthSeg, UniverSeg e SAM per l'analisi dell'atrofia cerebrale nelle immagini MRI del cervello. L'obiettivo principale è valutare l'efficacia di queste architetture nella segmentazione automatica delle strutture cerebrali coinvolte nell'atrofia, come l'ippocampo, il giro temporale, la sostanza bianca, e altre regioni rilevanti.

Le tecniche di segmentazione automatica proposte dalle suddette architetture offrono la possibilità di quantificare le variazioni nel volume delle strutture cerebrali nel corso del tempo, permettendo una valutazione più accurata e precisa delle condizioni neurodegenerative e dell'atrofia cerebrale.

L'analisi comparativa di queste architetture si basa su una serie di metriche e la comparazione visiva tra i differenti risultati, al fine di determinarne l'accuratezza e l'affidabilità nella segmentazione delle strutture cerebrali. Questo studio mira a identificare la migliore architettura per la segmentazione dell'atrofia cerebrale nelle MRI, offrendo un contributo significativo nel campo della diagnosi precoce e del monitoraggio delle condizioni neurodegenerative.

basate sull’imaging cerebrale.

1.1 Medical Imaging

Il campo dell’imaging medico, in particolare l’analisi delle immagini ottenute tramite risonanza magnetica (MRI), svolge un ruolo cruciale nella diagnosi, nella pianificazione del trattamento e nel monitoraggio delle condizioni mediche. Le immagini MRI forniscono dettagli anatomici e funzionali *in vivo* di varie parti del corpo umano, consentendo la visualizzazione di tessuti molli, organi, vasi sanguigni e anomalie patologiche con una risoluzione elevata e senza l’impiego di radiazioni dannose.

La segmentazione delle strutture anatomiche nelle immagini MRI è un passo fondamentale per comprendere la morfologia e la distribuzione spaziale dei tessuti, facilitando la diagnosi e la valutazione delle condizioni mediche. Tuttavia, l’analisi manuale delle immagini richiede tempo e può essere soggetta a errori umani, sottolineando l’importanza di sviluppare metodi automatizzati e precisi per la segmentazione.

Negli ultimi anni, diverse architetture e algoritmi basati sull’apprendimento automatico e sulle reti neurali convoluzionali sono stati proposti per migliorare l’accuratezza e l’efficienza della segmentazione delle MRI. Questi approcci sfruttano tecniche di deep learning per l’identificazione e l’isolamento delle regioni di interesse all’interno delle immagini, consentendo una segmentazione più accurata e rapida rispetto ai metodi convenzionali.

Le architetture FAST, SynthSeg, UniverSeg e SAM rappresentano alcune delle proposte più recenti e promettenti per la segmentazione delle MRI. Ciascuna di queste architetture presenta peculiarità, vantaggi e limitazioni uniche, che devono essere valutate attentamente per determinarne l’efficacia nell’ambito specifico dell’analisi delle immagini mediche.

Questo contesto richiede una valutazione accurata delle prestazioni di queste architetture, considerando metriche come precisione, recall, dice coefficient e tempi di esecuzione. La comprensione approfondita delle prestazioni di ciascuna architettura è essenziale per identificare la più adatta alle esigenze della segmentazione delle strutture anatomiche nelle MRI, consentendo un progresso significativo nell’automatizzazione e nella precisione delle analisi diagnostiche basate sull’imaging medico.

1.2 Dataset

L'Archivio Immagini e Dati (IDA) [1] è gestito dal Laboratorio di Neuro Imaging (LONI) presso l'Istituto di Neuroimaging e Informatica Mark e Mary Stevens dell'USC. Il laboratorio ha iniziato a gestire i dati di neuroimaging per studi di ricerca multicentrici alla fine degli anni '90, il che ha portato allo sviluppo dell'IDA. Da allora, l'IDA è diventato un punto di riferimento per molte collaborazioni multisito. LONI si impegna ad essere un leader globale nella gestione dei dati in neuroscienze e nelle soluzioni informatiche che facilitano la conservazione, l'esplorazione e la condivisione dei dati.

L'Archivio Immagini e Dati fornisce strumenti e risorse per de-identificare, integrare, cercare, visualizzare e condividere una vasta gamma di dati sulle neuroscienze e facilita le collaborazioni tra scienziati di tutto il mondo. Gli investigatori degli studi mantengono il pieno controllo sui dati memorizzati nell'IDA, beneficiando al contempo di un'infrastruttura robusta e affidabile che protegge e conserva i dati di ricerca per massimizzare l'investimento nella raccolta dati.

All'interno di IDA è stato utilizzato il dataset Alzheimer's Disease Neuroimaging Initiative (ADNI) [2] il quale è un progetto di ricerca di vasta portata che si concentra sulla malattia di Alzheimer e coinvolge un consorzio di numerosi istituti di ricerca, tra cui centri universitari, ospedali e organizzazioni non profit. L'ADNI è stato lanciato negli Stati Uniti nel 2004 con l'obiettivo di migliorare la comprensione della malattia di Alzheimer attraverso la raccolta, l'analisi e la condivisione di dati clinici, di imaging e biologici.

Il dataset ADNI include una vasta gamma di informazioni, come dati clinici, dati genetici, immagini cerebrali (ad esempio, risonanza magnetica strutturale e funzionale, PET), dati biologici e campioni di sangue, liquido cerebrospinale e altri biomarcatori.

1.2.1 Formati delle Scansioni

Il formato DICOM è uno standard internazionale utilizzato per immagazzinare, visualizzare e trasmettere immagini mediche, come le immagini ottenute da apparecchiature di diagnostica per immagini, ad esempio risonanze magnetiche (MRI), tomografie computerizzate (CT), ecografie e altri dispositivi medici.

I file con estensione ".dcm" contengono dati dell'immagine e **metadati** associati, quali informazioni sulla struttura anatomica, i parametri di acquisizione, i dettagli del paziente, il dispositivo utilizzato per la scansione e altre informazioni pertinenti. Il formato DICOM è strutturato per garantire la compatibilità e l'interoperabilità tra diverse apparecchiature mediche e software, consentendo agli operatori sanitari di visualizzare, condividere e archiviare facilmente immagini mediche da svariate fonti.

Poiché il formato DICOM è ampiamente accettato e utilizzato nell'ambito della diagnostica per immagini mediche, i file con estensione ".dcm" sono riconosciuti da numerosi software di visualizzazione e elaborazione delle immagini mediche, offrendo la possibilità di analizzare dettagliatamente immagini provenienti da diverse fonti e strumenti diagnostici.

Il formato ".nii.gz" è una combinazione di due formati: NIfTI e GZIP.

Il formato *NIfTI* (.nii) è un formato standard per il salvataggio di dati di imaging cerebrale in neuroscienze e medicina. Il file con estensione ".nii" contiene informazioni dettagliate relative a immagini acquisite tramite risonanza magnetica (MRI), tomografia computerizzata (CT) e altri metodi di imaging. Questi file possono includere dati anatomici e funzionali e solitamente conservano informazioni come le dimensioni dell'immagine, la risoluzione, i parametri di acquisizione e i dati stessi dell'immagine.

GZIP (.gz) è un formato di compressione dei file che riduce le dimensioni dei file mediante l'algoritmo di compressione GZIP. L'estensione ".gz" indica che il file è stato compresso tramite questo metodo. Un file con estensione ".nii.gz" è quindi un file di imaging medico che contiene dati strutturati secondo lo standard NIfTI ed è stato compresso utilizzando l'algoritmo GZIP per ridurne le dimensioni. La compressione aiuta a risparmiare spazio di archiviazione e facilita la trasmissione o lo scambio di questi file.

1.2.2 Pre-processing

Nonostante la conversione delle immagini nel formato ”.nii.gz” le immagini ottenute necessitano di una fase di pre-processing. Ai modelli in esame verranno fornite delle *slice*.

Le slice in ambito di imaging medico, come la risonanza magnetica (MRI) o la tomografia computerizzata (CT), si riferiscono a sezioni tridimensionali dell’oggetto o del corpo analizzato. Queste sezioni sono ottenute attraverso l’acquisizione di immagini bidimensionali che mostrano la struttura interna dell’oggetto o del corpo da diverse prospettive.

Nel contesto dell’imaging medico, le slice rappresentano le sezioni sottili dell’area di interesse. Ad esempio, se si esegue una risonanza magnetica cerebrale, le slice sono le sezioni sottili del cervello ottenute durante la procedura di scansione. Queste sezioni sono visualizzate in successione per mostrare dettagli specifici o un’immagine tridimensionale dell’area di interesse.

Le slice possono essere acquisite lungo assi diversi (ad esempio, **sagittale**, **coronale** o **assiale** in riferimento al corpo umano) e vengono utilizzate per ottenere una visione dettagliata di varie strutture interne. Queste immagini a fette vengono poi analizzate e interpretate per fini diagnostici o di ricerca. Le slice sono fondamentali nell’interpretazione delle immagini di medical imaging, in quanto forniscono dettagli specifici e permettono di studiare la struttura e le caratteristiche di parti specifiche dell’organismo o dell’oggetto analizzato.

1.2.2.1 Normalizzazione Lineare

La normalizzazione lineare è un processo utilizzato per modificare la scala dei valori di un insieme di dati, rendendoli più adatti per un’analisi o una visualizzazione specifica. In generale, la normalizzazione lineare consiste nel trasformare i valori di un insieme di dati in un nuovo intervallo, solitamente compreso tra 0 e 1 o tra 0 e 255 in caso di immagini, mantenendo la relativa proporzione tra i valori. Nel nostro caso di studio verrà utilizzata per scalare le slice dal range 0/255 a 0/1.

Nel contesto dell’elaborazione di immagini, la normalizzazione lineare è comunemente utilizzata per regolare la scala dei valori di intensità dei pixel. Questo processo permette di ottimizzare la visualizzazione o l’analisi dell’immagine, poiché i valori possono variare notevolmente a seconda della tecnica di acquisizione o della modalità di imaging utilizzata.

La formula generale per la normalizzazione lineare è:

$$V_n = \frac{V_a - V_{min}}{V_{max} - V_{min}} \times Nuovo\ Intervallo \quad (1.1)$$

Dove:

- ” V_n ” è il valore normalizzato.
- ” V_a ” è il valore del dato originale.
- ” V_{min} ” è il valore minimo presente nel set di dati.
- ” V_{max} ” è il valore massimo presente nel set di dati.
- ”*Nuovo Intervallo*” è l’intervallo desiderato in cui si desidera riscalare i valori.

Questa trasformazione lineare assicura che tutti i valori siano ricalibrati proporzionalmente entro il nuovo intervallo definito, garantendo che i dati mantengano la relativa distribuzione e permettendo una migliore visualizzazione o analisi. Questa tecnica è spesso utilizzata in diversi campi, inclusi l’elaborazione di immagini mediche, la statistica, l’analisi dei dati e altri ambiti in cui è necessario regolare i valori dei dati per una migliore interpretazione o manipolazione.

Nel nostro caso d’utilizzo come detto precedentemente i nostri dati verranno riscalati nel range 0/1.

Quindi la formula diventa:

$$Slice = 255 - (255 \times \frac{V_{max} - V_a}{V_{max} - V_{min}}) \quad (1.2)$$

In questo modo tutti i pixel della slice hanno un valore compreso tra 0 e 1.

1.2.2.2 Resize

Il resize (ridimensionamento) si riferisce al cambiamento delle dimensioni di un’immagine, generalmente in termini di altezza e larghezza, per adattarla a una dimensione specifica desiderata. Questo processo può essere eseguito per adattare le immagini a dimensioni standard o specifiche richieste da un’applicazione o da un modello di machine learning. Il resize può essere effettuato in diversi modi, come l’interpolazione bilineare o l’interpolazione più vicina.

Il resize con padding e stride può essere combinato per gestire immagini di dimensioni diverse e regolare la loro grandezza per adattarle a un’analisi o

a un'applicazione specifica. Il padding viene utilizzato per aggiungere pixel intorno ai bordi dell'immagine prima di applicare il resize, garantendo che l'immagine risultante sia proporzionata e priva di distorsioni. Lo stride può essere utilizzato per controllare la quantità di sovrapposizione tra le trasformazioni o le operazioni applicate durante il processo di elaborazione dell'immagine.

Il padding, o riempimento, è l'aggiunta di pixel intorno ai bordi dell'immagine per estenderla e adattarla a una dimensione desiderata durante il processo di resize. Il padding può essere di diversi tipi, tra cui il padding zero (aggiunta di pixel con valore zero intorno ai bordi), il padding riflessivo (duplicando i pixel più esterni), o il padding costante (riempimento con un valore specifico).

Lo stride, o passo, è la distanza tra le posizioni in cui vengono applicate le trasformazioni (come il filtro in una convoluzione) su un'immagine durante l'elaborazione. Ad esempio, in una convoluzione, lo stride indica quante posizioni si sposta il filtro sull'immagine durante il processo di estrazione delle caratteristiche.

1.3 MRI del Cervello

La risonanza magnetica (MRI) è uno strumento fondamentale nell'ambito della neuroscienza in quanto consente di ottenere immagini dettagliate del cervello e del sistema nervoso [3]. Le immagini ottenute tramite MRI forniscono informazioni anatomiche e funzionali che sono cruciali per la comprensione della struttura cerebrale e l'identificazione di condizioni patologiche.

Nello specifico, le sequenze MRI pesate in T1, T2 e FLAIR offrono prospettive uniche e complementari. La sequenza T1 fornisce un'eccellente risoluzione spaziale e un notevole contrasto tra i tessuti, rendendo visibili strutture anatomiche specifiche come la corteccia cerebrale, i vasi sanguigni e le lesioni emorragiche. La sequenza T2, d'altro canto, evidenzia le differenze nella quantità di acqua nei tessuti, rendendo visibili infiammazioni, tumori, edemi e lesioni demielinizzanti. La sequenza FLAIR, una variante della sequenza T2, sopprime il segnale del liquido cerebrospinale, migliorando la visualizzazione delle lesioni vicino ai ventricoli cerebrali e consentendo una maggiore sensibilità nella rilevazione di lesioni sottili.

Oltre alle sequenze MRI, la neuroimaging può essere integrata con la tomografia ad emissione di positroni (PET). La PET fornisce informazioni funzionali rilevando le concentrazioni di radiotracers che riflettono l'attività metabolica o i cambiamenti neurochimici del cervello. Questa tecnica è cruciale per la comprensione dei processi fisiologici cerebrali, come il metabolismo del glucosio o il legame di neurotrasmettitori specifici, offrendo informazioni dettagliate sulla funzionalità cerebrale.

In sintesi, l'uso combinato di MRI con sequenze pesate in T1, T2, FLAIR e l'integrazione con la PET offre una panoramica completa e dettagliata della struttura e della funzione cerebrale. Questi strumenti di imaging giocano un ruolo fondamentale nella ricerca e nella diagnosi di condizioni neurologiche, contribuendo significativamente alla comprensione del cervello e delle sue patologie.

La risonanza magnetica (MRI), o risonanza magnetica nucleare (NMR), è una tecnica di medical imaging non invasiva e non dannosa che utilizza un forte campo magnetico e impulsi di radiofrequenza per generare immagini dettagliate del corpo umano. Questa tecnica permette la visualizzazione di tessuti molli, organi interni e strutture anatomiche in modo molto dettagliato e fornisce informazioni importanti sulla struttura e sulla composizione dei tessuti.

1.3.1 Principi fondamentali della MRI

La risonanza magnetica sfrutta il comportamento dei protoni presenti negli atomi di idrogeno nel corpo umano quando vengono esposti a un campo magnetico. Durante la procedura di MRI, i protoni si allineano con il campo magnetico e vengono successivamente eccitati dall'applicazione di impulsi di radiofrequenza. Quando questi protoni rilasciano l'energia assorbita, gli apparecchi MRI rilevano questo segnale e lo utilizzano per creare immagini dettagliate dei tessuti.

1.3.2 Utilizzo e applicazioni della MRI

Le MRI forniscono informazioni dettagliate sulla morfologia, la struttura e le patologie dei tessuti molli come il cervello, il midollo spinale, il cuore, i vasi sanguigni, i muscoli e le articolazioni. È ampiamente utilizzata per diagnosticare condizioni neurologiche, malattie cerebrovascolari, lesioni, tumori, anomalie congenite, nonché per lo studio delle patologie osteoarticolari e molte altre condizioni mediche.

Le sequenze MRI possono essere ponderate in diversi modi (T1, T2, FLAIR, ecc.), ciascuna offrendo informazioni specifiche sulla composizione e sulla struttura dei tessuti. La sequenza T1 offre un buon contrasto tra i tessuti molli, mentre la T2 è sensibile alla presenza di acqua, utile per rilevare infiammazioni, edemi e altre patologie. La sequenza FLAIR sopprime il segnale del liquido cerebrospinale, migliorando la visualizzazione delle lesioni adiacenti ai ventricoli cerebrali.

Inoltre, la MRI può essere utilizzata per monitorare la risposta ai trattamenti e per eseguire studi anatomici e funzionali [4, 5, 6], come l'attivazione cerebrale o lo studio della diffusione di fibre nervose attraverso le sequenze di risonanza magnetica a tensore di diffusione (DTI).

In conclusione, la risonanza magnetica è una tecnologia fondamentale nella diagnostica medica, offrendo immagini ad alta risoluzione e fornendo informazioni preziose per la diagnosi e il monitoraggio di una vasta gamma di condizioni mediche e neurologiche.

1.3.3 T1

La sequenza T1, abbreviazione di "tempo di rilassamento longitudinale", è un'importante tecnica di acquisizione di immagini utilizzata nella risonanza magnetica (MRI). La risonanza magnetica pesata in T1 è nota per la sua capacità di fornire immagini ad alta risoluzione con un'eccellente differenziazione dei tessuti.

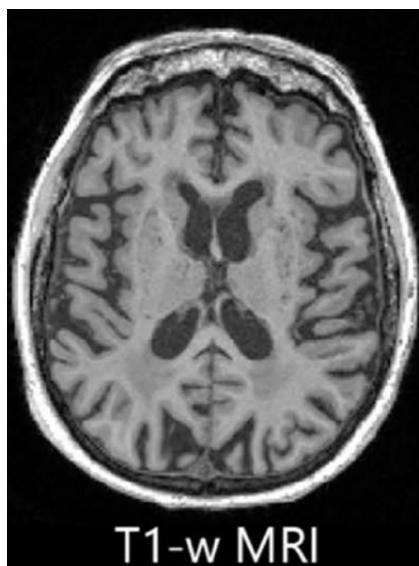


Figura 1.1: Risonanza Magnetica con peso T1.

1.3.3.1 Principi fondamentali della sequenza T1

La sequenza T1 sfrutta il tempo di rilassamento longitudinale dei protoni all'interno dei tessuti biologici. Quando un campo magnetico esterno viene applicato a una serie di atomi di idrogeno (protoni) allineati, questi protoni assorbiranno energia e si "ecciteranno". Quando questa energia viene rilasciata, i protoni ritornano al loro stato di equilibrio. Il tempo necessario affinché questi protoni ritornino al loro stato di equilibrio è chiamato tempo di rilassamento longitudinale (T1).

1.3.3.2 Caratteristiche e applicazioni della sequenza T1

Le immagini ottenute mediante la sequenza T1 forniscono un'eccellente risoluzione spaziale e un elevato contrasto tra i tessuti, evidenziando in particolare le differenze nella densità dei protoni all'interno dei vari tessuti. I tessuti ad alto contenuto lipidico, come il grasso, tendono a essere più luminosi nelle

immagini T1, mentre i tessuti ad alto contenuto di acqua, come la materia grigia, appaiono più scuri.

La sequenza T1 è utilizzata in diverse applicazioni diagnostiche, compresa l'analisi del cervello, della colonna vertebrale e di altre regioni anatomiche. È particolarmente utile per visualizzare strutture anatomiche come la corteccia cerebrale, i vasi sanguigni, il midollo spinale, le lesioni emorragiche e alcune patologie ossee. Inoltre, l'immagine T1 è essenziale nella valutazione di lesioni cerebrali, tumori, malattie demielinizzanti e nell'assistere nella pianificazione di interventi chirurgici.

1.3.3.3 Ruolo clinico e utilizzo in contesti di ricerca

Nel contesto clinico, le immagini T1 sono fondamentali per la valutazione di una vasta gamma di condizioni mediche, fornendo dettagli anatomici e aiutando i medici a formulare diagnosi precise. Inoltre, nel campo della ricerca, le immagini T1 sono utilizzate per studi longitudinali che valutano cambiamenti nel volume cerebrale nel corso del tempo, contribuendo alla comprensione delle malattie neurodegenerative e dello sviluppo cerebrale [7, 8, 9].

In sintesi, la sequenza T1 nella risonanza magnetica offre un'importante prospettiva nella diagnostica clinica, nella ricerca neuroscientifica e nell'analisi dettagliata delle strutture anatomiche, contribuendo in modo significativo alla comprensione delle condizioni patologiche e al monitoraggio delle variazioni anatomiche nel tempo.

1.3.4 T2

La sequenza T2 è un’importante tecnica di acquisizione di immagini utilizzata nella risonanza magnetica (MRI). La risonanza magnetica pesata in T2 è nota per la sua capacità di fornire immagini con un’elevato contrasto tra i tessuti molli, rivelando variazioni nella quantità di acqua e nell’organizzazione dei tessuti.

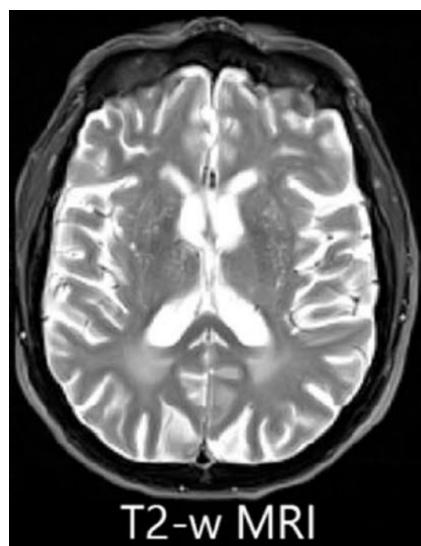


Figura 1.2: Risonanza Magnetica con peso T2.

1.3.4.1 Principi fondamentali della sequenza T2

La sequenza T2 sfrutta il tempo di rilassamento trasversale dei protoni all’interno dei tessuti biologici. Quando i protoni assorbono energia in un campo magnetico esterno, rilasciano questa energia nel tempo. Il tempo necessario affinché i protoni si disperdano e ritornino al loro stato di equilibrio è chiamato tempo di rilassamento trasversale (T2). I tessuti con maggiore contenuto di acqua o fluidi tendono ad avere tempi T2 più lunghi.

1.3.4.2 Caratteristiche e applicazioni della sequenza T2

Le immagini ottenute mediante la sequenza T2 mostrano una buona risoluzione spaziale e un elevato contrasto tra i tessuti. I tessuti con maggiore contenuto di acqua appaiono più luminosi, mentre i tessuti con meno acqua, come la materia bianca del cervello, possono apparire più scuri. Questa sequenza è particolarmente efficace nel rilevare ed evidenziare patologie che

causano un aumento di acqua o di fluidi, come ad esempio gli edemi, le infiammazioni o le lesioni.

Le immagini T2 sono comunemente utilizzate nella diagnostica del cervello, della colonna vertebrale e dei tessuti molli, permettendo la visualizzazione di lesioni, patologie infiammatorie, processi degenerativi, tumori, infarti cerebrali e anomalie vascolari. Inoltre, la sequenza T2 è fondamentale per valutare lo stato di salute del midollo spinale, la presenza di cisti, lesioni demielinizzanti e altre patologie neurologiche.

1.3.4.3 Ruolo clinico e utilizzo in contesti di ricerca

Le immagini T2 sono di grande importanza nella pratica clinica per la diagnosi e il monitoraggio delle condizioni patologiche, fornendo informazioni dettagliate sui tessuti e sulle lesioni. Inoltre, nella ricerca scientifica, la sequenza T2 è utilizzata per studiare la progressione di malattie neurodegenerative, la valutazione del cervello in diverse condizioni fisiologiche e patologiche e per approfondire la comprensione della fisiologia e della patologia cerebrale [10, 11, 12].

In conclusione, la sequenza T2 nella risonanza magnetica fornisce informazioni dettagliate sulle variazioni nei tessuti e sulle patologie, offrendo una prospettiva fondamentale sia nella pratica clinica che nella ricerca medica e neuroscientifica.

1.3.5 FLAIR

La sequenza FLAIR, Fluid Attenuated Inversion Recovery, è una variante della sequenza T2 che sopprime il segnale del liquido cerebrospinale (CSF), consentendo una migliore visualizzazione delle lesioni o delle anomalie dei tessuti vicini al liquido cerebrospinale. Questo metodo migliora il contrasto tra le lesioni e i tessuti normali, riducendo l'effetto di parziale volume del liquido cerebrospinale nelle immagini, consentendo una maggiore sensibilità nella rilevazione delle lesioni sottili o vicine ai ventricoli cerebrali.

La sequenza FLAIR è estremamente utile nella diagnosi di lesioni cerebrali, come le lesioni infiammatorie, le sclerosi multipla, le infarti cerebrali, le lesioni demielinizzanti e i tumori che altrimenti potrebbero essere mascherati dal segnale del liquido cerebrospinale nelle immagini convenzionali T2. La sequenza FLAIR, acronimo di Fluid Attenuated Inversion Recovery, è una tecnica di risonanza magnetica (MRI) che sfrutta l'inversione del segnale del fluido per migliorare la visualizzazione delle lesioni o delle patologie cerebrali nelle immagini. Questa sequenza è una variante delle sequenze pesate in T2, progettata per sopprimere il segnale del liquido cerebrospinale (CSF) al fine di migliorare la sensibilità nel rilevare le lesioni nei tessuti circostanti.

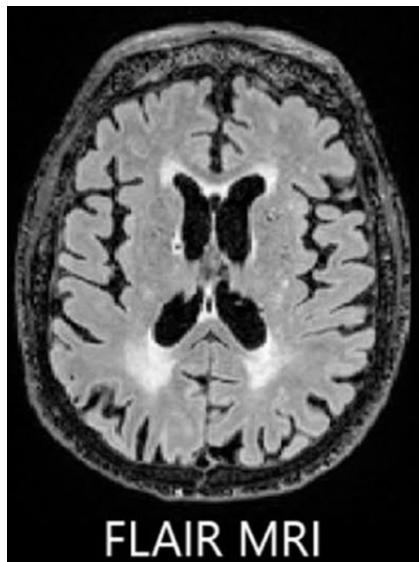


Figura 1.3: Risonanza Magnetica FLAIR.

1.3.5.1 Principi fondamentali della sequenza FLAIR

La sequenza FLAIR utilizza l'inversione del segnale del fluido, come il CSF, per sopprimirne il segnale nelle immagini. Ciò viene ottenuto utilizzando un impulso di inversione specifico per il CSF prima dell'acquisizione delle immagini. Questa tecnica permette di ridurre il segnale proveniente dal CSF, migliorando il contrasto tra le lesioni e i tessuti circostanti.

1.3.5.2 Caratteristiche e applicazioni della sequenza FLAIR:

Le immagini ottenute tramite la sequenza FLAIR forniscono un'eccellente visualizzazione delle lesioni o delle patologie che altrimenti potrebbero essere nascoste o mascherate dal segnale del liquido cerebrospinale nelle sequenze T2 convenzionali. Le lesioni o le anomalie che sono in prossimità dei ventricoli cerebrali o del CSF sono più facilmente identificabili e visualizzate con maggiore chiarezza nelle immagini FLAIR.

La sequenza FLAIR è ampiamente utilizzata nella diagnostica neuroradiologica per la valutazione delle lesioni cerebrali, incluse patologie come la sclerosi multipla, le infiammazioni, i tumori cerebrali, le lesioni demielinizzanti, gli infarti e altre condizioni patologiche che coinvolgono il sistema nervoso centrale. Inoltre, la FLAIR è fondamentale per la valutazione della materia bianca e per rilevare lesioni sottili o lesioni vicine ai ventricoli cerebrali.

1.3.5.3 Ruolo clinico e utilizzo in contesti di ricerca

La sequenza FLAIR è diventata uno strumento essenziale nella pratica clinica poiché fornisce una maggiore sensibilità nel rilevare e visualizzare lesioni o patologie cerebrali, migliorando la precisione diagnostica. È anche ampiamente utilizzata nella ricerca neuroscientifica per studiare la progressione di patologie cerebrali, valutare le variazioni nella materia bianca e studiare le condizioni neurodegenerative [13, 14, 15, 16, 17, 18].

In conclusione, la sequenza FLAIR nella risonanza magnetica fornisce una prospettiva chiara e dettagliata delle lesioni o delle patologie cerebrali, migliorando la sensibilità diagnostica e contribuendo notevolmente alla comprensione e alla valutazione delle condizioni patologiche del sistema nervoso centrale.

1.3.6 PET

La Tomografia ad Emissione di Positroni (PET) è una tecnica di imaging molecolare che fornisce informazioni sulla funzione e sull'attività biochimica del corpo umano, inclusi il cervello e altri organi. Questa tecnologia permette di visualizzare la distribuzione di molecole biologicamente rilevanti all'interno dell'organismo, consentendo di rilevare la presenza, la localizzazione e la quantità di sostanze radioattive introdotte nel corpo (radiotracers).

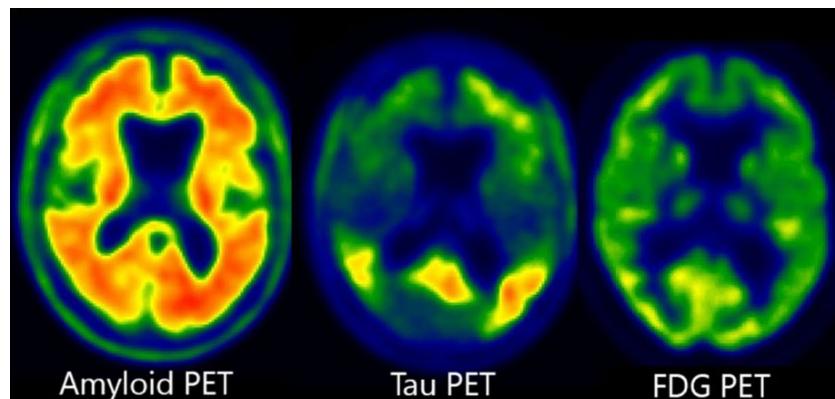


Figura 1.4: Immagine PET.

1.3.6.1 Principi fondamentali della PET

La PET coinvolge l'uso di radiotracers, molecole biologicamente attive mamate con isotopi radioattivi. Questi radiotracers emettono positroni, che a loro volta si scontrano con gli elettroni nel corpo producendo una coppia di fotoni gamma. Questi fotoni vengono rilevati da una serie di scanner esterni che convertono i dati raccolti in immagini tridimensionali che mostrano la distribuzione del radiotracer nel corpo.

1.3.6.2 Applicazioni e utilizzo in neuroscienze

Nell'ambito delle neuroscienze, la PET è estremamente utile per studiare l'attività cerebrale e le disfunzioni neurologiche. Una delle applicazioni più comuni è la valutazione del metabolismo cerebrale mediante l'uso del radiotracer FDG (Fluorodesossiglucosio). La PET-FDG rivela i livelli di consumo di glucosio nel cervello, mostrando aree con un maggior o minore metabolismo. Questa informazione è cruciale per la diagnosi e il monitoraggio di condizioni cerebrali, come tumori, demenza, epilessia, e per comprendere le basi neurobiologiche di varie patologie.

Altre applicazioni PET includono l'uso di radiotracers specifici per rilevare recettori neurotrasmettitori nel cervello, permettendo di studiare malattie come la malattia di Alzheimer, il morbo di Parkinson e la schizofrenia. Ad esempio, i radiotracer specifici per i recettori dopaminergici possono essere utilizzati per valutare la funzione dei neuroni dopaminergici nel morbo di Parkinson.

1.3.6.3 Ruolo clinico e utilizzo in contesti di ricerca

La PET gioca un ruolo fondamentale nella ricerca neuroscientifica, consentendo di studiare dettagliatamente l'attività cerebrale e di identificare cambiamenti chimici e funzionali. Nell'ambito clinico, la PET è utilizzata per valutare il trattamento e monitorare la progressione di malattie neurologiche, fornendo informazioni vitali per la diagnosi precoce e la valutazione del successo del trattamento [19, 20, 21].

In conclusione, la PET rappresenta uno strumento essenziale nella ricerca e nella pratica clinica delle neuroscienze, offrendo una prospettiva unica sulla funzione cerebrale e giocando un ruolo chiave nella diagnosi e nel monitoraggio di patologie neurologiche.

Capitolo 2

Descrizione Modelli

In questo capitolo, saranno presentati i modelli impiegati per la segmentazione cerebrale al fine di individuare regioni soggette ad atrofia nella ricerca.

2.1 FAST

La seguente sezione fornirà un'analisi dettagliata di un articolo chiave nel campo della segmentazione di immagini mediche, intitolato "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm" [22]. L'articolo presenta un approccio innovativo basato su un modello di campo casuale di Markov nascosto (hidden Markov random field, HMRF) e l'algoritmo di aspettativa-massimizzazione (expectation-maximization, EM) per la segmentazione di immagini risonanza magnetica (MR) del cervello.

2.1.1 Tecnologie Utilizzate

2.1.1.1 Hidden Markov Random Field Model

L'articolo propone l'utilizzo di un modello di campo casuale di Markov nascosto (HMRF) [23] per la segmentazione delle immagini cerebrali. Questo modello fornisce un approccio robusto per modellare le relazioni spaziali tra i pixel dell'immagine, considerando la struttura adiacente durante il processo di segmentazione. Di seguito vengono riportati i componenti principali dell'HMRF.

Campo Casuale di Markov (MRF): Iniziamo definendo il concetto di campo casuale di Markov (MRF), un modello statistico che descrive l'incertezza nelle variabili casuali distribuite su una griglia. Nell'MRF, le variabili casuali sono associate ai nodi di un grafo, con nodi adiacenti considerati dipendenti tra loro.

Variabili Nasoste e Struttura Adiacente: L'HMRF estende l'MRF introducendo il concetto di variabili nascoste. Queste variabili non sono osservabili direttamente ma giocano un ruolo significativo nel processo generativo del modello. L'HMRF modella anche la struttura adiacente, considerando le relazioni spaziali tra le variabili, un aspetto cruciale quando si tratta di dati bidimensionali come le immagini.

Applicazioni in Imaging Biomedico: Nel contesto dell'imaging biomedico, l'HMRF trova applicazione nella segmentazione di immagini, come ad esempio nell'analisi di immagini cerebrali. L'utilizzo di HMRF consente di catturare le correlazioni spaziali tra i pixel, migliorando la precisione della segmentazione.

Stima dei Parametri: L'ottimizzazione dei parametri in un HMRF può essere complessa. Spesso, vengono impiegati algoritmi di apprendimento, come l'aspettativa-massimizzazione (EM), per stimare accuratamente i parametri del modello.

Vantaggi e Limitazioni: Gli HMRF offrono una rappresentazione flessibile delle dipendenze tra le variabili, specialmente nelle applicazioni in cui le correlazioni spaziali sono rilevanti. Tuttavia, la complessità computazionale e la scelta dei parametri possono rappresentare sfide che richiedono particolari accorgimenti.

2.1.1.2 Expectation-Maximization Algorithm

L'algoritmo di aspettativa-massimizzazione (EM) [24] viene impiegato per ottimizzare i parametri del modello HMRF. Questo algoritmo iterativo è ampiamente utilizzato in problemi di stima dei parametri, consentendo una migliore adattabilità del modello ai dati dell'immagine.

L'Algoritmo di Aspettativa-Massimizzazione (EM) è un metodo iterativo utilizzato per stimare i parametri di un modello statistico quando ci sono variabili latenti o mancanti nei dati. Si applica a una vasta gamma di problemi e ha trovato ampio utilizzo in campi come l'apprendimento automatico, la visione artificiale e la statistica.

Fase di Aspettativa (E-step): Nella **prima fase dell'EM**, chiamata **fase di aspettativa (E-step)**, vengono calcolate le aspettative delle variabili latenti dati i dati osservati e i parametri correnti del modello. Questo coinvolge il calcolo delle probabilità condizionate o delle aspettative delle variabili nascoste.

Fase di Massimizzazione (M-step): Nella **seconda fase**, chiamata **fase di massimizzazione (M-step)**, i parametri del modello vengono aggiornati per massimizzare la verosimiglianza dei dati, tenendo conto delle aspettative calcolate nella fase precedente. Questo comporta l'ottimizzazione dei parametri del modello rispetto ai dati osservati e alle stime delle variabili latenti. **Iterazioni:** Le **fasi di E-step e M-step** vengono eseguite iterativamente fino a raggiungere la convergenza, ossia quando i parametri del modello smettono di cambiare significativamente tra le iterazioni.

Applicazioni in Segmentazione di Immagini: L'EM trova applicazione in vari contesti, inclusa la **segmentazione di immagini**. In particolare, può essere utilizzato per ottimizzare i parametri di modelli probabilistici, come nel caso di modelli di segmentazione basati su **Hidden Markov Random Field (HMRF)** per immagini mediche.

Vantaggi e Limitazioni: L'EM è un potente strumento quando si tratta di dati mancanti o variabili latenti, ma può essere sensibile alle condizioni iniziali e potrebbe richiedere un numero significativo di iterazioni. È essenziale eseguire l'algoritmo con attenzione e considerare le specificità del problema in questione.

2.1.2 Vantaggi e Svantaggi

2.1.2.1 Vantaggi

Il modello HMRF e l'algoritmo EM offrono una robusta capacità di segmentazione, tenendo conto delle relazioni spaziali complesse presenti nelle immagini cerebrali. La combinazione di questi approcci consente un'adeguata rappresentazione delle strutture anatomiche e delle regioni di interesse.

2.1.2.2 Svantaggi

Tuttavia, è importante notare che l'efficacia di questo approccio può essere influenzata dalla qualità dei dati di input e dalla scelta appropriata dei parametri. Inoltre, il tempo di calcolo potrebbe essere una considerazione in contesti in cui la velocità è critica.

2.2 SAM

Nella sezione successiva, esploreremo un articolo che ha catturato notevole attenzione nella comunità scientifica di recente, intitolato "Segment Anything" [25], da cui il nome **SAM** *Segment Anything Model*. Il seguente articolo presenta un modello e un dataset innovativo per la segmentazione delle immagini, con un dataset che comprende oltre 1 miliardo di maschere su 11 milioni di immagini. SAM, sfruttando l'intelligenza artificiale, si distingue come uno strumento significativo in questo contesto.

Quando si affronta l'incertezza nell'identificazione dell'oggetto da segmentare, SAM dimostra la sua capacità di generare multiple maschere valide. Inoltre, il modello ha la straordinaria abilità di individuare e generare automaticamente maschere per tutti gli oggetti presenti in un'immagine. Sfruttando le incorporate dell'immagine precalcolate, SAM è in grado di fornire istantaneamente una maschera di segmentazione per qualsiasi prompt, consentendo un'interazione in tempo reale con il modello.

2.2.1 Tecnologie Utilizzate

L'architettura di SAM comprende tre componenti che lavorano insieme per restituire una maschera di segmentazione valida.

2.2.1.1 Image Encoder

Al livello più alto vi è un image encoder che genera le componenti uniche delle immagini e può essere applicato prima di presentare un'interrogazione al modello.

Per questo scopo vengono utilizzati **MAE** e **ViT**.

MAE Il **Masked Auto-Encoder** (MAE) è una variante di un autoencoder che sfrutta l'utilizzo di maschere durante il processo di encoding e decoding.

Esaminandolo più nel dettaglio il funzionamento: **Autoencoder di Base**:

- In un **autoencoder** tradizionale, si ha un **encoder** che riduce la dimensione dell'input, generando una rappresentazione compressa detta codice.
- Il **decoder**, poi, ricostruisce l'input originale da questa rappresentazione compressa.

Aggiunta di Maschere (Masking):

- Nel contesto del **MAE**, durante l'**encoding**, si applicano maschere all'input originale. Queste maschere determinano quali parti dell'input verranno ignorate o "mascherate".
- L'obiettivo è quello di insegnare al modello a concentrarsi su specifiche caratteristiche rilevanti, ignorando o "mascherando" le informazioni meno importanti durante il processo di encoding.

Processo di Decoding:

- Nel processo di **decoding**, il modello cerca di ricostruire l'input originale tenendo conto delle maschere applicate durante l'encoding.
- L'uso delle maschere durante il decoding contribuisce a generare una rappresentazione dell'input originale che tiene conto delle informazioni rilevanti identificate durante l'encoding.

Applicazioni e Vantaggi:

- L'approccio del **MAE** può essere particolarmente utile in scenari in cui si desidera dare più peso a specifiche regioni o caratteristiche dell'input.
- Può essere impiegato in varie applicazioni, compresa la segmentazione di immagini, in cui l'attenzione a determinate parti dell'immagine è cruciale.

Quindi lo scopo di **MAE** è progettato per fornire un meccanismo di mascheramento durante l'encoding per concentrare l'attenzione su aspetti specifici dei dati, contribuendo così a generare rappresentazioni più significative e rilevanti.

Vit Il **Vision Transformer** (ViT) è un’architettura di rete neurale profonda progettata per l’elaborazione di immagini. A differenza delle tradizionali reti neurali convoluzionali (CNN), che utilizzano layer convoluzionali per catturare pattern spaziali nelle immagini, il ViT si basa su un’architettura completamente trasformativa e sfrutta il meccanismo di autoattenzione.

Autoattenzione: Il cuore del ViT è l’utilizzo di meccanismi di **autoattenzione** (trasformatore) per catturare le relazioni tra le diverse parti di un’immagine. Questo permette al modello di attribuire pesi differenti a diverse regioni dell’immagine in base al contesto e alle relazioni semantiche.

Divisione in Patches: L’immagine di input viene suddivisa in piccoli blocchi rettangolari noti come **patches**. Ogni patch viene poi *appiattita* e trattata come una sequenza, che è la forma di input predefinita per il trasformatore.

Inclusione di Informazioni Globali: L’architettura ViT include informazioni globali attraverso un token speciale di contesto, che viene concatenato alle rappresentazioni delle patch. Questo consente al modello di catturare informazioni globali dell’intera immagine.

Struttura Trasformata: Il modello è composto da una serie di layer trasformati. Ogni layer ha una parte di autoattenzione e una parte di rete completamente connessa (feedforward). Gli output di ciascun layer vengono passati al successivo.

Classificazione di Immagini: Anche se è stato originariamente progettato per compiti di **classificazione di immagini**, il ViT può essere adattato per affrontare compiti di visione artificiale più complessi, come la segmentazione delle immagini o la generazione di immagini.

Scalabilità: Una delle forze principali del ViT è la sua **scalabilità**: funziona bene su immagini di diverse dimensioni senza richiedere modifiche all’architettura del modello.

Il *Vision Transformer* ha dimostrato risultati sorprendenti su benchmark di classificazione di immagini e ha contribuito a consolidare l’importanza delle architetture basate su trasformatori nell’ambito della visione artificiale.

2.2.1.2 Prompt Encoder

L'encoder di prompt codifica in tempo reale punti di sfondo, maschere, bounding boxes o testi in un vettore di incorporamento. La ricerca considera due insiemi di prompt: sparso (punti, boxes, testo) e denso (maschere).

Punti e boxes sono rappresentati da codifiche posizionali e sommati con incorporamenti appresi per ogni tipo di prompt. I prompt di testo in forma libera sono rappresentati con un encoder di testo pronto all'uso da CLIP. I prompt densi, come le maschere, sono incorporati con convoluzioni e sommati elemento per elemento con l'incorporamento dell'immagine.

Codifiche posizionali (positional encodings) Le **codifiche posizionali** sono un meccanismo cruciale nelle architetture di reti neurali, come il **Transformer**, per introdurre informazioni sulla posizione relativa di elementi in una sequenza. In contesti sequenziali, l'ordine degli elementi è fondamentale, ma le reti neurali non hanno una percezione intrinseca dell'ordine o della posizione.

La **strategia comune** per introdurre informazioni sulla posizione è attraverso l'utilizzo di funzioni sinusoidali e cosinusoidali. La codifica posizionale per l'elemento alla posizione pos e la dimensione i è data dalle formule:

$$\text{Pos_encoding}(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d}}\right)$$

$$\text{Pos_encoding}(pos, 2i + 1) = \cos\left(\frac{pos}{10000^{2i/d}}\right)$$

dove pos è la posizione e i è la dimensione della codifica posizionale. Queste codifiche vengono sommate o concatenate agli incorporamenti degli elementi nella sequenza, fornendo al modello informazioni sulla posizione relativa.

L'utilizzo del **grassetto** nel testo mira a evidenziare l'importanza di questa tecnica nella comprensione del contesto sequenziale per le reti neurali.

CLIP Acronimo di *Contrastive Language-Image Pre-training*, è un modello di intelligenza artificiale sviluppato da OpenAI. Progettato per comprendere e collegare in modo semantico il linguaggio naturale e le immagini, CLIP è addestrato su un vasto dataset di coppie di immagini e testi. Caratteristiche Chiave di CLIP:

- **Architettura Multimodale:** CLIP è un modello multimodale che può comprendere simultaneamente il linguaggio naturale e le immagini, eseguendo compiti di associazione tra testo e immagini.
- **Addestramento Contrastivo:** Il modello è addestrato attraverso un processo contrastivo, massimizzando la similarità tra rappresentazioni di coppie positive (testo e immagini semanticamente simili) e minimizzando la similarità con rappresentazioni di coppie negative.
- **Dataset Ampio:** CLIP viene addestrato su un ampio dataset, che include testi e immagini da una vasta gamma di contesti, contribuendo a creare una rappresentazione generale e flessibile.
- **Applicazioni Diverse:** Il modello può essere utilizzato in compiti come la classificazione di immagini, la ricerca di immagini e la generazione di testo basata su immagini.
- **Zero-shot Learning:** CLIP ha la capacità di eseguire compiti di zero-shot learning, generalizzando su nuove categorie senza ulteriore addestramento specifico.
- **OpenAI API:** L'accesso a CLIP è reso disponibile tramite l'API di OpenAI, consentendo agli sviluppatori di sfruttare le sue capacità multimodali nelle loro applicazioni.

2.2.1.3 Mask Decoder

Un **lightweight mask decoder** predice le maschere di segmentazione basandosi sugli incorporamenti generati sia dall’immagine che dagli encoder di prompt. Questo decodificatore mappa l’incorporamento dell’immagine, gli incorporamenti del prompt e un token di output in una maschera. Tutti gli incorporamenti vengono aggiornati dal blocco di decodifica, che utilizza autoattenzione e interattenzione tra prompt in entrambe le direzioni (dal prompt all’incorporamento dell’immagine e viceversa).

Le maschere vengono annotate e utilizzate per aggiornare i pesi del modello. Questo schema arricchisce il dataset e consente al modello di imparare e migliorare nel tempo, rendendolo efficiente e flessibile.

Il **mask decoder** è una componente fondamentale in un sistema di segmentazione che ha il compito di generare le maschere di segmentazione. La sua principale funzione è tradurre gli **incorporamenti** (embeddings) estratti dall’immagine e dai prompt in una rappresentazione dettagliata della segmentazione dell’immagine, ovvero le maschere che indicano le regioni di interesse. Caratteristiche chiave del **mask decoder**:

1. **Predizione delle Maschere:** Il mask decoder predice le maschere di segmentazione basandosi sugli incorporamenti estratti dall’immagine e dai prompt. Utilizza una rappresentazione ad alta dimensionalità per catturare i dettagli fini delle strutture nell’immagine.
2. **Utilizzo di Autoattenzione e Interattenzione:** La componente chiave del processo di decodifica è l’uso di autoattenzione e interattenzione. L’autoattenzione consente al decoder di concentrarsi su diverse parti dell’input, mentre l’interattenzione consente l’interscambio di informazioni tra gli incorporamenti del prompt e quelli dell’immagine in entrambe le direzioni.
3. **Aggiornamento degli Incorporamenti:** Gli incorporamenti dell’immagine e del prompt vengono continuamente aggiornati dal blocco di decodifica. Questo processo iterativo consente al modello di affinare progressivamente la sua rappresentazione delle maschere durante il processo di apprendimento.
4. **Addestramento con Annotazioni delle Maschere:** Le maschere generate dal decoder sono confrontate con le maschere annotate nel dataset. Le discrepanze tra le predizioni e le annotazioni vengono utilizzate per calcolare la perdita durante l’addestramento.

5. **Efficienza e Flessibilità:** L'approccio del mask decoder è progettato per essere leggero, consentendo una generazione efficiente delle maschere. La sua struttura bidirezionale e il processo iterativo di aggiornamento degli incorporamenti rendono il modello flessibile e adattabile a una varietà di contesti.
6. **Apprendimento Continuo:** L'uso delle maschere annotate per aggiornare i pesi del modello consente un apprendimento continuo. Il modello può migliorare la sua capacità di segmentazione man mano che viene esposto a nuovi dati e nuove annotazioni.

Il **mask decoder** svolge un ruolo critico nella generazione di maschere di segmentazione, incorporando informazioni dall'immagine e dal prompt e iterando per migliorare progressivamente la qualità delle predizioni. La sua architettura leggera e il processo di aggiornamento continuo lo rendono una componente chiave nei sistemi di segmentazione avanzati.

2.2.2 Modalità d'utilizzo

Il modello SAM mette a disposizione due modalità di utilizzo:

1. **Generazione Automatica:** In questa modalità, il modello prende in input l'immagine e genera tutte le maschere plausibili relative all'immagine.
2. **Generazione da Bounding Box:** In questa modalità, si richiede di disegnare una bounding box, e il modello calcolerà le maschere nell'area racchiusa dalla bounding box.

Nella fase di sperimentazione della seguente tesi è stata utilizzata la prima modalità esposta, su ogni slice.

2.3 SynthSeg

SynthSeg [26] rappresenta un software innovativo dedicato alla sintesi di immagini di risonanza magnetica (MRI) cerebrale con etichette di segmentazione. Questo strumento è stato sviluppato con l'obiettivo di affrontare le sfide legate alla disponibilità limitata di dataset di imaging cerebrale annotati, fornendo una soluzione efficiente per la generazione di dati sintetici.

La segmentazione cerebrale, ovvero l'identificazione accurata delle diverse strutture cerebrali in immagini MRI, è un compito critico per molte applicazioni mediche e di ricerca. Tuttavia, l'addestramento di modelli di apprendimento automatico per questa attività richiede un ampio volume di dati etichettati, e la raccolta di tali dataset può essere un compito oneroso e costoso.

SynthSeg si propone di superare questa limitazione offrendo la possibilità di generare immagini MRI cerebrali sintetiche, accompagnate da etichette di segmentazione precise. Questo software è progettato per facilitare lo sviluppo e la validazione di algoritmi di segmentazione cerebrale basati sull'apprendimento automatico, consentendo ai ricercatori di affinare e valutare le proprie metodologie in un contesto controllato.

2.3.1 Tecnologie Utilizzate

2.3.1.1 Modelli Generativi

La metodologia di SynthSeg si basa su un modello generativo dal quale vengono estratte scansioni sintetiche per addestrare una rete di segmentazione. Un aspetto cruciale di questa procedura è che le immagini di addestramento sono generate dinamicamente con parametri completamente randomizzati. Ciò implica che la rete viene esposta a varie combinazioni di contrasto, risoluzione, morfologia, artefatti e rumore in ciascun mini-batch. La generazione dinamica delle immagini di addestramento, con una variazione continua dei parametri, contribuisce a migliorare la robustezza della rete di segmentazione.

Questa metodologia consente a SynthSeg di fornire un set di dati di addestramento vario e ricco di variazioni, contribuendo così a ottimizzare le prestazioni della rete di segmentazione nell'affrontare una vasta gamma di condizioni di imaging.

2.3.1.2 Rete di Segmentazione

Nell'implementazione, viene impiegata un'architettura UNet tridimensionale. Questa architettura consiste in cinque livelli, ciascuno separato da uno

strato di normalizzazione del batch insieme a un'operazione di max-pooling (percorso contrattivo) o di upsampling (percorso espansivo). Ogni livello comprende due strati convoluzionali con kernel $3 \times 3 \times 3$. Ogni strato convoluzionale è associato a un'attivazione Exponential Linear Unit, eccetto l'ultimo, che utilizza un softmax. Mentre il primo livello conta 24 mappe di caratteristiche, questo numero viene raddoppiato dopo ogni max-pooling e dimezzato dopo ogni upsampling. Seguendo l'architettura UNet, vengono utilizzate connessioni skip attraverso i percorsi contrattivi ed espansivi. Si noti che l'architettura di rete non è l'oggetto principale di questo lavoro: sebbene sia impiegato un UNet (la rete più diffusa per immagini mediche), in linea di principio potrebbe essere sostituita con qualsiasi altra architettura di segmentazione.

Unet L'architettura U-Net [27] è una rete neurale convoluzionale (CNN) ampiamente utilizzata per compiti di segmentazione delle immagini. Proposta nel 2015 da Olaf Ronneberger, Philipp Fischer e Thomas Brox, l'U-Net ha dimostrato grande efficacia, in particolare nelle applicazioni mediche, dove la segmentazione accurata delle strutture anatomiche è fondamentale.

- **Struttura a U:** L'architettura prende il nome dalla sua forma a "U". Comprende un percorso contrattivo (downsampling) a sinistra della "U" e un percorso espansivo (upsampling) a destra, consentendo di catturare dettagli a diverse scale spaziali.
- **Percorso Contrattivo:** Attraverso strati convoluzionali e di max-pooling, l'immagine di input viene ridotta progressivamente di dimensioni. Questo percorso è progettato per catturare caratteristiche di livello superiore e ridurre la dimensione spaziale.
- **Percorso Espansivo:** L'immagine viene ricostruita attraverso operazioni di upsampling e concatenazione con le caratteristiche del percorso contrattivo corrispondente tramite connessioni skip. Ciò aiuta a mantenere dettagli importanti durante l'espansione.
- **Connessioni Skip:** Le connessioni skip collegano i corrispondenti strati contrattivi ed espansivi, preservando informazioni dettagliate durante il downsampling e riutilizzandole durante l'upsampling.
- **Funzione di Attivazione:** Di solito, l'U-Net utilizza la funzione di attivazione ReLU per gli strati convoluzionali, tranne l'ultimo strato del percorso espansivo, che spesso utilizza la funzione di attivazione Softmax per problemi di classificazione multclasse.

- **Applicazioni Mediche:** Grazie alla sua capacità di gestire immagini mediche e mantenere dettagli durante la segmentazione, l'U-Net è diventata un riferimento per la segmentazione di organi e lesioni nelle immagini mediche.

2.3.1.3 Inferenza

Durante il test, l'input viene campionato nuovamente a una risoluzione rHR utilizzando l'interpolazione trilineare (in modo che l'output della CNN sia a risoluzione HR), e le sue intensità vengono ridimensionate tra 0 e 1 con normalizzazione min-max (utilizzando i percentili 1° e 99°). Le scansioni preprocessate vengono quindi alimentate alla rete per ottenere mappe di predizione morbide per ciascuna etichetta. Nella pratica, viene eseguita anche un'aggiustamento durante il test, che migliora leggermente i risultati sul set di convalida. In particolare, vengono segmentate due versioni di ciascuna scansione di test: quella originale e una versione ribaltata da destra a sinistra. Le predizioni morbide dell'input ribaltato vengono quindi ribaltate nuovamente nello spazio originale (assicurando che le etichette da destra a sinistra finiscano sul lato corretto) e vengono mediate con le predizioni della scansione originale. Una volta eseguita l'aumentazione durante il test, le segmentazioni finali sono ottenute mantenendo il componente connesso più grande per ciascuna etichetta.

rHR La sigla "rHR" sta per "risoluzione ad alta risoluzione" (high resolution). Nel contesto dell'algoritmo descritto, rHR rappresenta una risoluzione a cui l'input dell'immagine viene campionato durante il test. L'operazione di campionamento viene eseguita utilizzando l'interpolazione trilineare. L'obiettivo di questa operazione è ottenere un'immagine con una risoluzione più elevata, in modo che l'output della rete neurale convoluzionale (CNN) sia a una risoluzione ad alta fedeltà.

Durante il test, l'immagine di input viene quindi riscalata o campionata a questa risoluzione specifica. Questo processo può migliorare la precisione e i dettagli delle previsioni della rete, consentendo una rappresentazione più dettagliata delle caratteristiche dell'immagine e delle strutture da segmentare.

L'utilizzo della risoluzione ad alta risoluzione (rHR) nel processo di test contribuisce a garantire che le previsioni della rete siano più dettagliate e accurate, migliorando complessivamente le performance della segmentazione, specialmente quando si tratta di strutture più piccole o dettagliate nell'immagine medica.

Interpolazione trilineare L’interpolazione trilineare è una tecnica utilizzata per stimare i valori intermedi tra i punti di dati noti in uno spazio tridimensionale. Questa tecnica è spesso utilizzata in contesti in cui è necessario ottenere valori intermedi su una griglia tridimensionale regolare. Nel contesto dell’immagine medica e della segmentazione, l’interpolazione trilineare è comunemente utilizzata per campionare i valori dei pixel o voxel tra le immagini acquisite.

Il termine ”voxel” è una contrazione di ”volume element” (elemento di volume) ed è utilizzato per descrivere i pixel tridimensionali, ovvero gli elementi di volume in uno spazio tridimensionale.

Ecco come funziona l’interpolazione trilineare:

1. **Definizione della Griglia:** Avendo una griglia tridimensionale di dati, dove ogni punto della griglia ha un valore associato. In questo caso, la griglia potrebbe rappresentare i voxel in uno spazio tridimensionale di un’immagine medica.
2. **Posizione del Punto da Interpolare:** Se si ha un punto in uno spazio tridimensionale che non coincide con i punti della griglia, ma si vuole stimarne il valore, si utilizza l’interpolazione trilineare. Questo punto può trovarsi in una posizione non allineata con gli assi della griglia.
3. **Interpolazione:** L’interpolazione trilineare coinvolge tre passaggi principali lungo ciascun asse della griglia (x , y , z). Per stimare il valore del punto desiderato, si effettuano le seguenti interpolazioni:
 - **Interpolazione lungo l’asse x :** Viene eseguita un’interpolazione lineare tra i valori noti sui due lati più vicini lungo l’asse x .
 - **Interpolazione lungo l’asse y :** Viene eseguita un’interpolazione lineare tra i valori ottenuti nella fase precedente lungo l’asse y .
 - **Interpolazione lungo l’asse z :** Infine, viene eseguita un’interpolazione lineare tra i valori ottenuti lungo l’asse z .

Questo processo di interpolazione consente di stimare il valore del punto desiderato in modo continuo e regolare all’interno dello spazio tridimensionale della griglia. Nell’ambito della segmentazione medica, l’interpolazione trilineare viene spesso utilizzata durante operazioni di ridimensionamento o campionamento di immagini per garantire una rappresentazione continua e dettagliata delle strutture anatomiche.

2.4 UniverSeg

UniverSeg [28], ”Universal Medical Image Segmentation”, è un nuovo metodo per la segmentazione delle immagini mediche che può generalizzare a compiti non visti, che coinvolge nuove anatomia, modalità di immagine o etichette, senza bisogno di ulteriori addestramenti. Questa è una soluzione al problema della formazione del modello che richiede tempo o della messa a punto in genere richiesta dai ricercatori di fronte a un nuovo compito di segmentazione. UniverSeg utilizza un meccanismo Cross-Block per creare mappe di segmentazione accurate data un’immagine di query e un insieme di coppie immagine-etichetta che definiscono la nuova attività. Per garantire la sua capacità di generalizzare, UniverSeg è stato addestrato su MegaMedical, una vasta collezione di 53 set di dati di segmentazione medica ad accesso aperto con oltre 22.000 scansioni. È stato trovato che il metodo supera diversi altri approcci su compiti inediti, fornendo informazioni utili sul funzionamento del sistema proposto.

2.4.1 Tecnologie Utilizzate

L’implementazione del modello avviene mediante una rete neurale completamente convoluzionale. Vengono utilizzati due blocchi di costruzione: lo strato di cross-convoluzione e il modulo CrossBlock.

2.4.1.1 CrossBlock

Per trasferire informazioni tra l’insieme di supporto e l’immagine di interrogazione, è stato introdotto uno strato di cross-convoluzione che interagisce con una mappa di caratteristiche dell’immagine di interrogazione u e un insieme di supporto $V = \{v_i\}_{i=1}^n$:

$$\text{CrossConv}(u, V; \theta_z) = \{z_i\}_{i=1}^n, \\ \text{per } z_i = \text{Conv}(u || v_i; \theta_z),$$

Dove, $||$ rappresenta l’operazione di concatenazione lungo la dimensione delle caratteristiche, e $\text{Conv}(x; \theta_z)$ denota uno strato convoluzionale con parametri apprendibili θ_z . Grazie al riutilizzo dei pesi θ_z , le operazioni di cross-convoluzione risultano invarianti rispetto alla permutazione di V . Da questo strato, è stato progettato un blocco di costruzione di livello superiore che produce versioni aggiornate delle rappresentazioni dell’immagine di interrogazione u e dell’insieme di supporto V ad ogni passo nella rete:

$$\text{CrossBlock}(u, V; \theta_z, \theta_v) = (u', V'),$$

dove:

$$\begin{aligned} z_i &= \mathcal{A}(\text{CrossConv}(u, v_i; \theta_z)), \quad \text{per } i = 1, 2, \dots, n \\ u' &= \frac{1}{n} \sum_{i=1}^n z_i \\ v'_i &= \mathcal{A}(\text{Conv}(z_i; \theta_v)), \quad \text{per } i = 1, 2, \dots, n \end{aligned}$$

dove $\mathcal{A}(x)$ è una funzione di attivazione non lineare. Questa strategia consente alle rappresentazioni di ciascuna voce dell’insieme di supporto e della query di interagire con le altre attraverso la loro rappresentazione media e facilita l’uso di insiemi di supporto di dimensioni variabili.

2.4.1.2 Rete

Per integrare informazioni su scale spaziali diverse, vengono composti i moduli CrossBlock in una struttura encoder-decoder con connessioni residue, analogamente alla popolare architettura UNet. La rete riceve in input l’immagine di interrogazione x_t e l’insieme di supporto ($S_t = \{(x_{ti}, y_{it})\}_{i=1}^n$) di coppie immagine e mappa di etichette, ciascuna concatenate lungo la dimensione dei canali, e produce in output la mappa di predizione della segmentazione \hat{y}_t . Ogni livello nel percorso dell’encoder consiste in un blocco CrossBlock seguito da un’operazione di downsampling spaziale delle rappresentazioni sia dell’immagine di interrogazione che dell’insieme di supporto. Ogni livello nel percorso espansivo consiste nell’upsampling di entrambe le rappresentazioni, il cui livello di risoluzione spaziale viene raddoppiato, concatenandole con la rappresentazione di dimensioni equivalenti nel percorso dell’encoding, seguito da un CrossBlock. Viene eseguita una singola convoluzione 1x1 per mappare la rappresentazione finale dell’interrogazione in una previsione.

Capitolo 3

Risultati

Durante il corso di questa indagine, sono stati esaminati e valutati i modelli utilizzando un set di dati composto dalle risonanze magnetiche pesate in T2 di cinque pazienti, ciascuno sottoposto a due MRI effettuate in tempi differenziati. Ciascuna MRI è stata suddivisa in 256 slice. Poiché approcci come SAM e UniverSeg operano su immagini 2D, l'elaborazione è considerata completa solo quando applicata a ciascuna slice. Per modelli come FAST e SynthSeg, che elaborano immagini 3D, il processo è considerato terminato al raggiungimento dell'output. Questo capitolo esplorerà in dettaglio i risultati derivati dall'analisi di tali dati.

3.1 Risultati Qualitativi

In questa sezione, verranno valutati i risultati qualitativi ottenuti dai diversi modelli, utilizzando un criterio soggettivo umano per la valutazione.

Ricordiamo che l'obiettivo di questa tesi è confrontare i risultati ottenuti da modelli sviluppati per la segmentazione, utilizzando i modelli per la segmentazione del tessuto cerebrale, mirando a ottenere le maschere corrispondenti alle diverse sezioni del cervello.

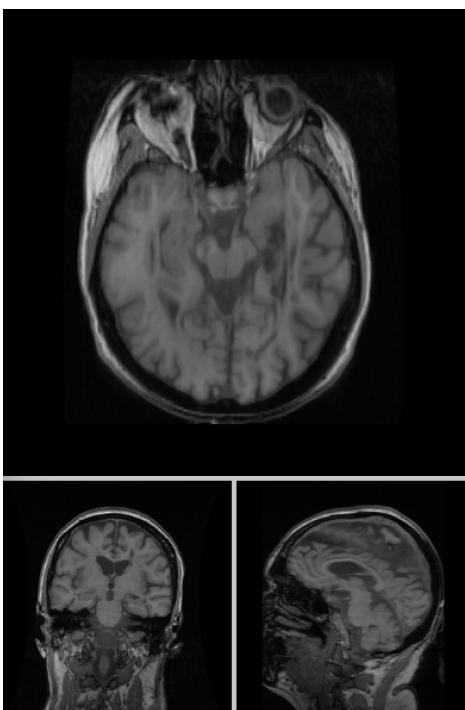


Figura 3.1: Risonanza magnetica.

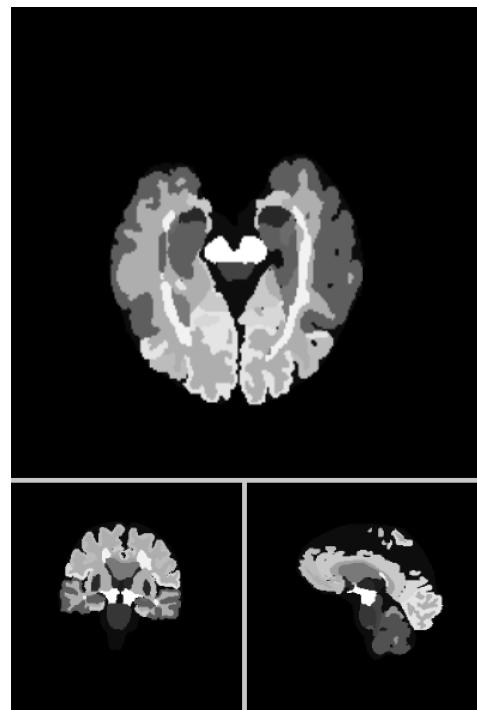


Figura 3.2: Maschera dell'MRI.

Nella Fig. 3.1 è possibile osservare l'immagine di una risonanza magnetica, mentre nella Fig. 3.2 sono visualizzate le maschere estratte dalla corrispondente risonanza magnetica.

3.1.1 FAST

3.1.1.1 Risultati

Come descritto in precedenza FAST è stato utilizzando impostando il valore delle etichette dei tessuti su 5, ciò vuol dire che come risultato ci apsetteremo 5 immagini contenenti le segmentazioni del cervello, più altre immagini contenente altre informazioni estratte da FAST.

Nella Fig. 3.3 sono presenti i risultati dei 5 Layer prodotti da fast.

Come si può vedere nel Layer 0(3.3a) e nel Layer 1(3.3b) FAST estrapola la parte interna del cranio, andando ad escludere la porzione non inherente del cervello.

Nel Layer 2(3.3c) il modello segmenta correttamente il cranio e lo spazio che intercorre tra il cranio e il cervello.

I risultati più interessanti li troviamo nelle immagini relative al terzo(3.3d) e quarto(3.3e) Layer, dove FAST identifica e differisce i vari strati e tipologie di tessuti.

Negli esempi riportati vengono prese sole delle slice di riferimento, ma FAST come detto in precedenza produce un output tridimensionale.

3.1.1.2 Valutazioni

I risultati mostrano che FAST svolge un lavoro eccellente nella segmentazione delle diverse sezioni del cervello, nonostante sia un modello proposto molti anni fa. Tuttavia, è evidente la difficoltà del modello nel distinguere le varie aree interne delle sezioni, come la corteccia o la differenziazione degli emisferi.

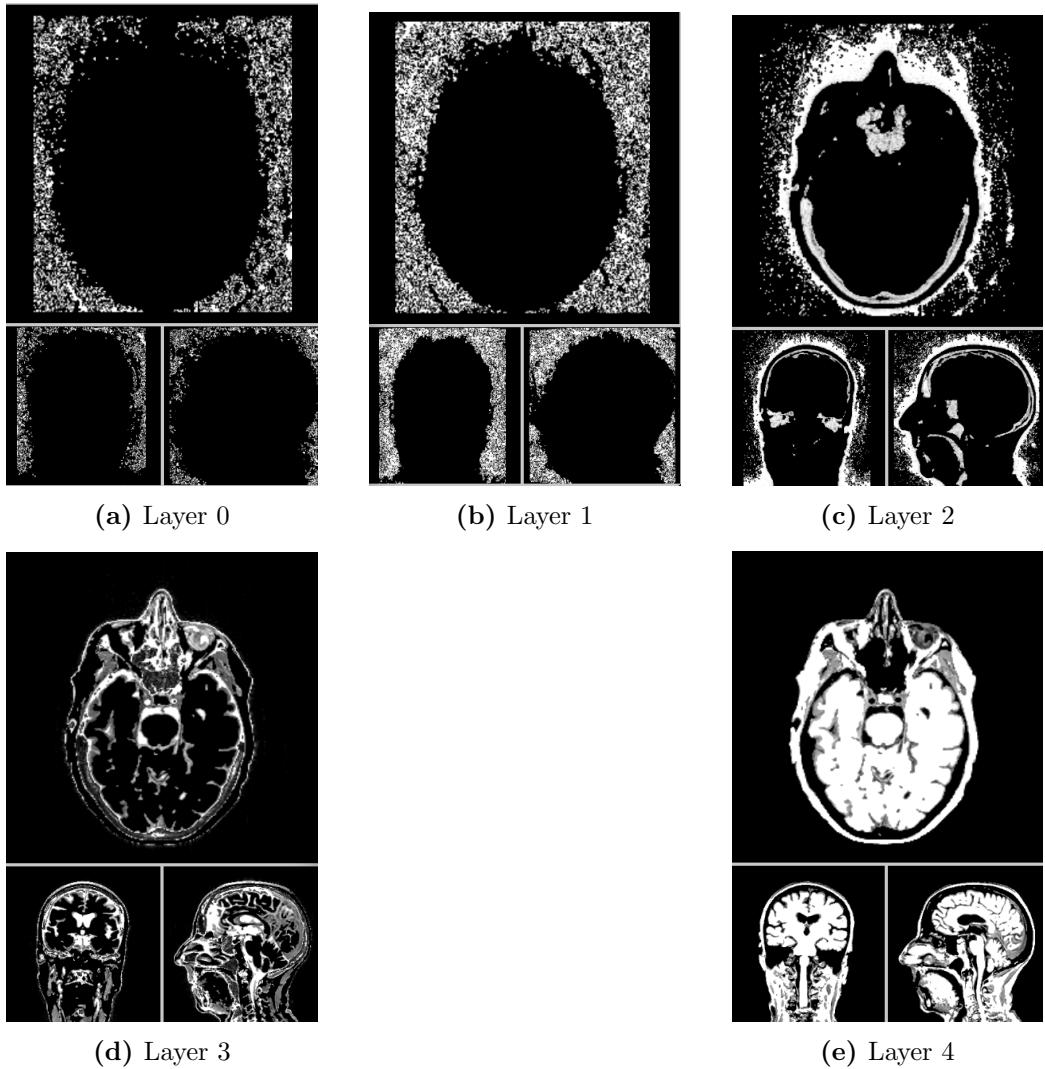


Figura 3.3: Risultati esecuzione FAST.

3.1.2 SAM

Per ottenere delle immagini tridimensionali dai risultati di SAM, che fornisce come output le maschere che trova in ogni slice sono stati svolti i seguenti passaggi:

1. Per ciascuna slice, le maschere sono state combinate per ottenere una maschera unica in cui l'intensità del colore indica la presenza del pixel in più maschere.
2. Ogni maschera complessiva ottenuta è stata trattata come un tensore bidimensionale. Questi tensori bidimensionali sono stati successivamente concatenati lungo l'asse appropriato per creare un tensore tridimensionale che rappresenta le maschere lungo l'intero asse.

3.1.2.1 Risultati

Nell'immagine seguente vengono mostrato come esempio le maschere ottenute da una sclice centrale di una risonanza magnetica

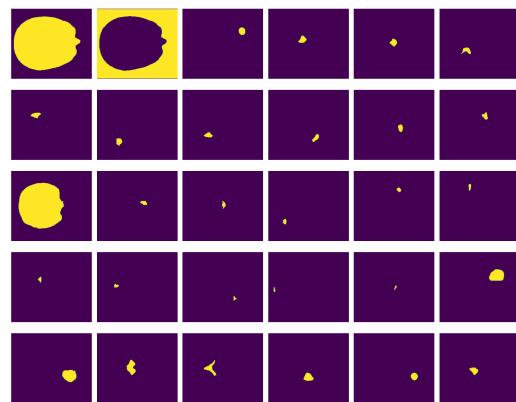


Figura 3.4: Maschere SAM Slice Centrale.

Una volta che tutte le maschere vengono unificate in una sola maschera il risultato sarà il seguente

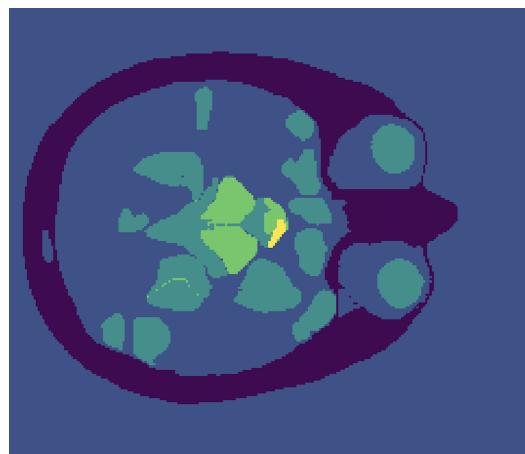


Figura 3.5: Maschera unificata.

3.1.2.2 Valutazioni

SAM non è stato in grado di ottenere le maschere delle sezioni del cervello, come evidenziato nella Figura 3.5, dove sono assenti molte informazioni rilevanti. È importante notare che il risultato non è ottenuto automaticamente, ma richiede un lavoro di post-processing per raggiungere il risultato descritto. Queste operazioni aggiuntive aumentano il tempo di esecuzione complessivo per il calcolo delle maschere.

3.1.3 SynthSeg

SynthSeg prende in input la risonanza magnetica, la normalizza e restituisce come output sia la maschera complessiva del cervello che la segmentazione dettagliata delle diverse regioni cerebrali.

3.1.3.1 Risultati

Come evidenziato nella Figura 3.7, SynthSeg riesce a segmentare correttamente tutte le aree del cervello e a rilevare con precisione i differenti tipi di tessuti cerebrali.

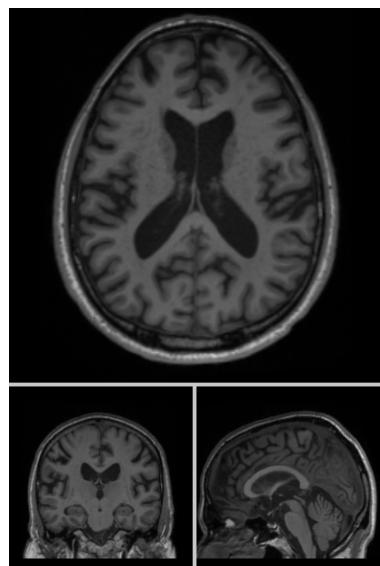


Figura 3.6: MRI Normalizzata.

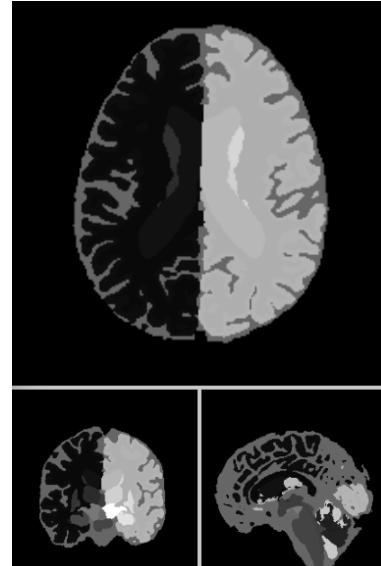


Figura 3.7: Segmentazione.

3.1.3.2 Valutazioni

SynthSeg dimostra la capacità di differenziare i vari tipi di tessuti cerebrali e di rilevare con notevole precisione le strutture anatomiche, inclusi i solchi e i lobi cerebrali.

3.1.4 UniverSeg

Come menzionato nella sezione dedicata, UniverSeg accetta un pool di immagini e relative etichette come input per eseguire l'inferenza.

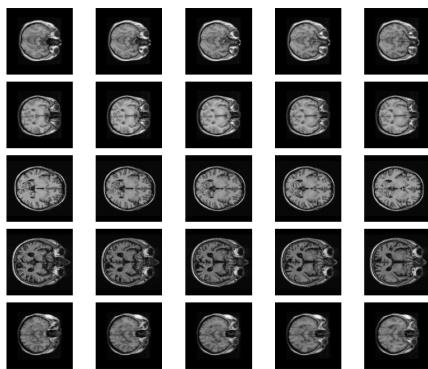


Figura 3.8: Sampler.

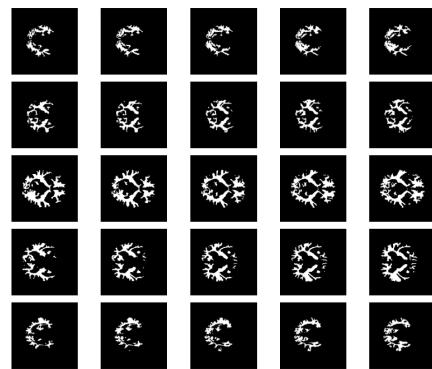


Figura 3.9: Label.

Figura 3.10: Pool di esempio.

3.1.4.1 Risultati

Poiché i test possono essere eseguiti solo sulle slice centrali, si è scelto di utilizzare UniverSeg per identificare, in base a una specifica etichetta, il risultato ottenuto su tali slice.

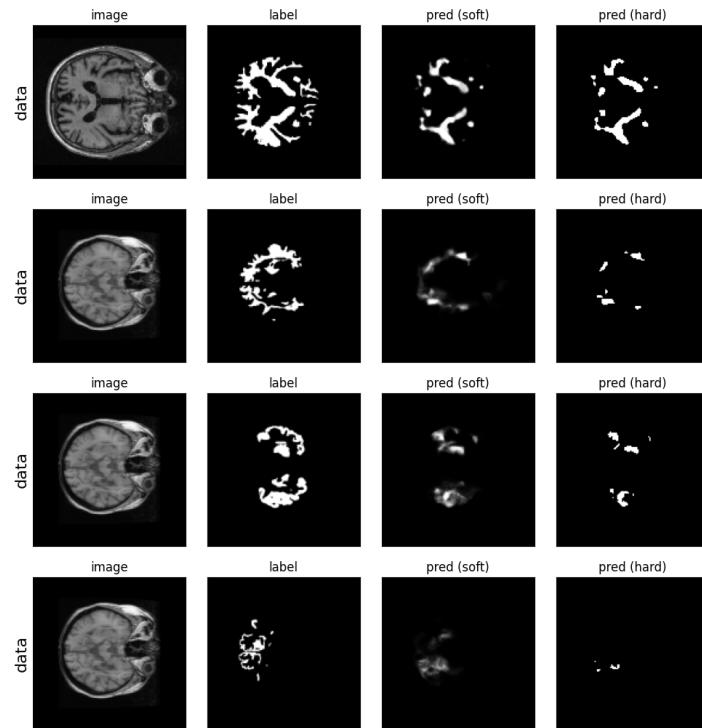


Figura 3.11: Risultati UniverSeg.

3.1.4.2 Valutazioni

Come evidenziato dai risultati di UniverSeg, nonostante vengano fornite immagini di campionamento con le rispettive etichette, il modello non è in grado di generare una maschera accurata quando gli viene presentato un nuovo dato. Inoltre, la limitazione nell'uso del modello solo sulle slice centrali aumenta l'inadeguatezza di questo approccio per ottenere maschere da un'intera risonanza magnetica, soprattutto considerando che l'esecuzione su immagini bidimensionali richiederebbe un campionamento adeguato per ogni slice e per ogni asse, risultando dispendioso e poco pratico.

3.2 Tempi Esecuzione

La valutazione delle prestazioni delle varie tecnologie proposte si basa su un'analisi cruciale: l'efficienza dei modelli. In questo contesto, l'efficienza si concentra principalmente sul tempo di calcolo, rappresentando un elemento chiave per valutare l'efficacia e la praticità delle soluzioni proposte.

Efficienza come tempo di calcolo L'efficienza in termini di tempo di calcolo si riferisce alla capacità di un modello informatico di eseguire le sue operazioni o completare un compito nel minor tempo possibile, utilizzando le risorse di calcolo disponibili in modo efficiente. Questa metrica è particolarmente critica in scenari in cui la velocità di esecuzione è una considerazione importante, come nel caso di applicazioni in tempo reale, elaborazioni di grandi volumi di dati o esecuzioni su risorse limitate.

Per misurare l'efficienza temporale di un modello, è comune considerare il tempo di calcolo necessario per eseguire operazioni specifiche, come l'addestramento del modello o l'inferenza su nuovi dati, come nel seguente caso studio.

L'obiettivo è ridurre al minimo il tempo richiesto per ottenere risultati accettabili, senza compromettere eccessivamente la qualità del modello.

3.2.1 FAST

In questa tesi è stato utilizzato FAST di **FSL** [29, 30, 31], in cui la versione utilizzabile di FAST sostituisce le versione FAST v4.0 e FAST v3.

3.2.1.1 Parametri

FAST permette di utilizzare molti valori e opzioni per l'esecuzione della segmentazione, i comandi sono mostrati nella tabella 3.1.

L'esecuzione di FAST avviene nella CPU e non nella GPU, ed inoltre non sempre è possibile utilizzare un'interfaccia grafica, come quella mostrato di seguito.

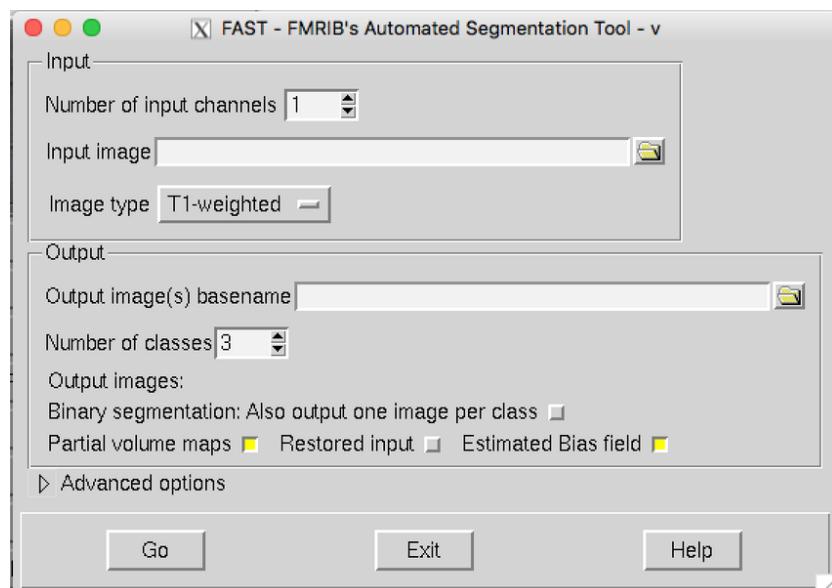


Figura 3.12: Interfaccia grafica FAST.

Per questo motivo nel caso in cui non si possa utilizzare l'interfaccia grafica bisognerà utilizzare il terminale e utilizzare un comando del seguente tipo

```
1 !fast <PPath_Image>
```

Nelle fasi di sperimentazioni sono stati utilizzati diversi parametri, ne segue un esempio dove viene modificato il numero di classi di tessuto da ricercare.

```
1 !fast -n 5 <PPath_Image>
```

Tabella 3.1: Descrizione dei parametri di FAST

Opzione	Descrizione
-n, -class	Numero di classi di tipo tessuto; default=3
-I, -iter	Numero di iterazioni del ciclo principale durante la rimozione del campo di bias; default=4
-l, -lowpass	Estensione di smoothing del campo di bias (FWHM) in mm; default=20
-t, -type	Tipo di immagine 1=T1, 2=T2, 3=PD; default=T1
-f, -fHard	Smusatezza spaziale iniziale della segmentazione (durante la stima del campo di bias); default=0.02
-g, -segments	Produce un'immagine binaria separata per ogni tipo di tessuto
-a <standard2input.mat>	Inizializza usando prior; è necessario fornire una trasformazione FLIRT
-A <prior1><prior2><prior3>	Immagini priori alternative
-nopve	Disabilita la stima del volume parziale (PVE)
-b	Producere il campo di bias stimato
-B	Producere l'immagine corretta dal bias
-N, -nobias	Non rimuovere il campo di bias
-S, -channels	Numero di immagini di input (canali); default 1
-o, -out	Nome di base per l'output
-P, -Prior	Usa prior in tutto; è necessario impostare anche l'opzione -a
-W, -init	Numero di iterazioni per l'inizializzazione della segmentazione; default=15
-R, -mixel	Liscezza spaziale per il tipo di mixel; default=0.3
-O, -fixed	Numero di iterazioni del ciclo principale dopo la rimozione del campo di bias; default=4
-H, -Hyper	Liscezza spaziale della segmentazione; default=0.1
-v, -verbose	Attiva i messaggi diagnostici
-h, -help	Visualizza questo messaggio
-s, manualseg <filename>	Nome del file contenente le intensità per la segmentazione manuale
-p	Producere mappe di probabilità individuali

3.2.1.2 Input

FAST richiede un’immagine 3D nel formato *nii.gz* come input, poiché questo formato conserva tutti i metadati necessari della risonanza magnetica. Il programma scorrerà automaticamente tutte le slice in ogni asse dell’immagine di input e restituirà i risultati nella stessa posizione dell’immagine originale.

3.2.1.3 Valutazione

A causa dell’utilizzo della CPU per i calcoli e della necessità di processare tutte le slice dell’MRI, questo modello risulta essere caratterizzato da una notevole lentezza nell’ottenimento dei risultati. Il calcolo delle maschere è stato eseguito sia con il numero di classi di tessuto da identificare (n) impostato a 3 (valore di default) che a 5.

In media, l’esecuzione della ricerca delle maschere con n pari a tre richiede circa **35 minuti**. Durante l’esecuzione delle stesse MRI con n impostato a 5, il tempo di esecuzione aumenta di circa venti minuti, portando il totale dell’esecuzione a circa **55 minuti** per ogni risonanza magnetica.

3.2.2 SAM

Nell'utilizzo di SAM è stato scelto di utilizzare la modalità di generazione automatica delle maschere.

3.2.2.1 Parametri

Sono stati svolti test con i parametri di default del generatore di maschere automatico ***”SamAutomaticMaskGenerator”*** e test con i parametri modificati per ottenere risultati qualitativamente superiori.

Tabella 3.2: Parametri di SamAutomaticMaskGenerator

Parametro	Spiegazione
model	Modello utilizzato per la generazione delle maschere.
points_per_side	Numero di punti su ciascun lato dell'immagine considerati durante la generazione della maschera.
pred_iou_thresh	Soglia di IoU predetta, al di sopra della quale una regione è inclusa nella maschera.
stability_score_thresh	Soglia di stabilità del punteggio, al di sopra della quale una regione è considerata stabile e inclusa nella maschera.
crop_n_layers	Numero di strati (layers) da ritagliare durante la generazione della maschera.
crop_n_points_downscale_factor	Fattore di riduzione per il numero di punti considerati durante il ritaglio.
min_mask_region_area	Area minima della regione di maschera. Regioni più piccole sono soggette a post-processing.

Di seguito viene mostrato come sono stati impostati i parametri personalizzati per la fase di test.

```

1 mask_generator = SamAutomaticMaskGenerator(
2     model=sam,
3     points_per_side=50,
4     pred_iou_thresh=0.86,
5     stability_score_thresh=0.92,
6     crop_n_layers=1,
7     crop_n_points_downscale_factor=2,
8     min_mask_region_area=100,
9 )

```

3.2.2.2 Input

SAM, in fase di input, non elabora l'intera risonanza magnetica (MRI), ma è in grado di operare solamente su immagini bidimensionali (2D). Questa caratteristica ha portato alla decisione di limitare le valutazioni all'asse assiale delle risonanze magnetiche. Di conseguenza, è stato fornito al modello ciascun piano (slice) dell'asse assiale, e le maschere risultanti sono state salvate.

3.2.2.3 Valutazioni

Va notato che i tempi di esecuzione menzionati sono riferiti esclusivamente alla versione personalizzata di SAM, e non tengono in considerazione i tempi della versione predefinita.

Nella tabella seguente vengono riportate le misurazioni:

Tabella 3.3: Valori dei tempi di esecuzione per pazienti

Paziente	Tempo (secondi)
Paziente 1	5894.87
Paziente 2	7916.63
Paziente 3	3823.30
Paziente 4	6175.85
Paziente 5	4713.63
Media Pazienti	5704.86

È importante sottolineare che, quando si menziona il termine "pazienti", si fa riferimento all'elaborazione di entrambe le risonanze magnetiche, totalizzando così 512 slice complessive. Considerando questa prospettiva, possiamo affermare che, in media, ogni risonanza magnetica richiede circa **44** minuti di elaborazione, ma questo si applica a un solo asse.

È fondamentale notare che le risonanze magnetiche sono composte da tre assi, pertanto, l'eventuale elaborazione dell'intera risonanza magnetica richiederebbe più di **2 ore** per ciascuna MRI.

3.2.3 SynthSeg

Per la successiva valutazione dei risultati, è cruciale tenere presente che SynthSeg è un modello specificamente progettato per la sintesi di immagini di risonanza magnetica (MRI) cerebrale, includendo etichette di segmentazione.

3.2.3.1 Parametri

Il modello può essere eseguito mediante una semplice linea di comando

```
1 python ./scripts/commands/SynthSeg_predict.py --i <input> --o  
    <output> [--parc --robust --ct --vol <vol> --qc <qc> --  
    post <post> --resample <resample>]
```

Dove i parametri sono spiegati nella tabella 3.4 (dove sono presenti tutti i parametri della versione iniziale) e nella tabella 3.5 (contenente dei parametri opzionali aggiunti nelle versioni successive).

Tabella 3.4: Parametri di SynthSeg

Parametro	Spiegazione
<input>	Percorso per una scansione da segmentare, o per una cartella. Può anche essere il percorso di un file di testo, dove ogni riga è il percorso di un'immagine da segmentare.
<output>	Percorso in cui verranno salvate le segmentazioni di output. Deve essere dello stesso tipo di <input> (ovvero, il percorso di un file, una cartella o un file di testo dove ogni riga è il percorso di una segmentazione di output).
--parc	(Opzionale) Per eseguire la parcellazione corticale oltre alla segmentazione dell'intero cervello.
--robust	(Opzionale) Per utilizzare la variante per una maggiore robustezza (ad esempio, nell'analisi di dati clinici con ampi spazi tra le slice). Può essere più lento rispetto all'altro modello.
--ct	(Opzionale) Utilizzare su scansioni CT nella scala di Hounsfield. Clip degli intensità a [0, 80].
<vol>	(Opzionale) Percorso per un file CSV in cui saranno salvati i volumi (in mm ³) di tutte le regioni segmentate per tutte le scansioni.
<qc>	(Opzionale) Percorso per un file CSV in cui saranno salvati i punteggi QC.
<post>	(Opzionale) Percorso in cui verranno salvate le probabilità soft come mappe di probabilità.
<resample>	(Opzionale) Le segmentazioni SynthSeg sono sempre date a una risoluzione isotropica di 1 mm. Utilizzare questo flag per salvare le immagini risample (stessi requisiti di formato di <output>).

Tabella 3.5: Parametri aggiuntivi di SynthSeg

Parametro	Spiegazione
--cpu	(Opzionale) Per forzare l'esecuzione del codice sulla CPU, anche se è disponibile una GPU.
--threads	(Opzionale) Numero di threads da utilizzare da Tensorflow.
--crop	(Opzionale) Per ritagliare le immagini di input a una data forma prima della segmentazione.
--fast	(Opzionale) Per disabilitare alcune operazioni per una previsione più veloce (meno accurata). Non si applica quando è usato il flag --robust.
--v1	(Opzionale) Per eseguire la prima versione di SynthSeg (SynthSeg 1.0).

3.2.3.2 Input

SynthSeg, in fase di input, è in grado di elaborare l'intera risonanza magnetica (MRI) e restituire come output il risultato complessivo delle maschere.

3.2.3.3 Valutazioni

SynthSeg, essendo stato appositamente progettato per l'elaborazione di risonanze magnetiche, presenta ottimi risultati in termini di tempi di esecuzione, sia quando utilizza la CPU che quando sfrutta la GPU.

Nel caso dell'utilizzo della CPU, il tempo di esecuzione medio per ciascuna risonanza magnetica è di circa **100 secondi**, mentre con l'ausilio della GPU, il tempo di esecuzione si riduce di oltre la metà, raggiungendo una media di **40 secondi**.

3.2.4 UniverSeg

UniverSeg è un modello che non richiede una fase di addestramento e, come presentato dagli sviluppatori, promette di offrire prestazioni efficienti su una vasta gamma di compiti.

3.2.4.1 Parametri

UniverSeg non presenta parametri specifici; l'unico flag disponibile è *pre-trained*, che deve essere impostato su true per indicare l'utilizzo del modello preaddestrato.

```
1 model = universeg(pretrained=True)
```

3.2.4.2 Input

Il modello UniverSeg accetta come input un insieme di immagini insieme alle rispettive maschere di target. Successivamente, esegue l'inferenza su queste immagini per generare i risultati desiderati.

```
1 prediction = model(
2     target_image,
3     support_images,
4     support_labels,
5 )
```

3.2.4.3 Valutazioni

Grazie alla rapida inferenza eseguita durante la fase di runtime, UniverSeg fornisce risultati in **pochi secondi**. Tuttavia, è importante notare che, nella versione attuale del modello, l'unico input accettato è la slice di mezzo, poiché il modello è stato addestrato esclusivamente su tali slice. Pertanto, nonostante la velocità di ottenimento dei risultati, l'output riguarda solo una singola slice.

3.2.5 Tabella Riassuntiva

Di seguito è presentata una tabella che riporta le considerazioni e le valutazioni emerse dalle varie fasi di sperimentazione condotte nel contesto della presente ricerca. Questa tabella fornisce un quadro dettagliato delle osservazioni raccolte durante il processo sperimentale, insieme alle relative valutazioni, al fine di illustrare in modo chiaro e accurato i risultati ottenuti e le implicazioni derivanti dalla ricerca condotta. Nella seguente tabella sono elencati i nomi dei modelli utilizzati nello studio, insieme al tempo di esecuzione dedicato al processo di segmentazione per l'intera risonanza magnetica (MRI). È importante sottolineare che per i casi in cui è stato impiegato il metodo SAM, il tempo di esecuzione riportato rappresenta il valore medio calcolato su un singolo asse moltiplicato per tre, al fine di considerare l'intero volume tridimensionale dell'MRI. Inoltre, per il modello UniverSeg, che è in grado di elaborare solo la slice centrale, il tempo di esecuzione è stato moltiplicato per 256×3 , al fine di riflettere l'analisi dell'intera immagine. Il 'tipo di input' si riferisce al formato dell'immagine accettato dal modello, che può essere 2D o 3D. Si è valutato se il risultato della segmentazione generale è accettabile, con particolare attenzione alla capacità di cogliere i dettagli delle macroaree. Inoltre, è stata valutata l'efficacia del modello nella rilevazione accurata delle strutture anatomiche, come solchi e lobi cerebrali. La colonna 'Deep Learning', che indica se il modello utilizza tecnologie di deep learning o meno. Questi criteri di valutazione forniscono una panoramica completa delle prestazioni dei modelli considerati nel contesto della ricerca condotta.

Tabella 3.6: Risultati Qualitativi Segmentazione

Modello	Segmentazione Semantica	Qualitativamente Attendibile
<i>FAST</i>	No	Si
<i>SAM</i>	No	Parzialmente
<i>SynthSeg</i>	Si	Si
<i>UniverSeg</i>	No	No

Tabella 3.7: Risultati Tempi Esecuzione

Modello	Tempo Esecuzione	Tipo Input	Deep Learning
<i>FAST</i>	Medio	3D	No
<i>SAM</i>	Alto	2D	Si
<i>SynthSeg</i>	Basso	3D	Si
<i>UniverSeg</i>	Alto	2D	Si

Conclusione

L’obiettivo di questa tesi è stato quello di valutare l’efficacia di tecniche tradizionali di segmentazione e di approcci di deep learning sviluppati appositamente per la segmentazione delle strutture anatomiche nelle immagini MRI. L’analisi dettagliata di queste diverse metodologie, considerando metriche quantitative e qualitative insieme ai tempi di elaborazione, offre una panoramica completa delle capacità di ciascuna architettura, identificando i rispettivi vantaggi e limitazioni. L’analisi dei risultati ottenuti da diversi modelli di segmentazione, quali FAST, SAM, SynthSeg, e UniverSeg, offre una panoramica dettagliata delle capacità e delle limitazioni di ciascun approccio nell’ambito della segmentazione cerebrale. FAST dimostra un’eccellente capacità di segmentazione delle diverse sezioni del cervello, evidenziando tuttavia difficoltà nella distinzione delle varie aree interne, come la corteccia o la differenziazione degli emisferi. L’utilizzo di parametri specifici, come l’impostazione del valore delle etichette dei tessuti su 5, contribuisce a ottenere risultati migliori. SAM, pur presentando una metodologia di output basata su maschere in ogni slice, mostra una limitata capacità nella segmentazione delle sezioni cerebrali, richiedendo un significativo lavoro di post-processing per ottenere risultati significativi. La mancanza di informazioni rilevanti nelle maschere evidenzia la necessità di miglioramenti. SynthSeg si distingue per la sua capacità di segmentare accuratamente tutte le aree del cervello, rilevando con precisione i diversi tipi di tessuti cerebrali, inclusi solchi e lobi. La sua capacità di differenziare i vari tipi di tessuti e rilevare strutture anatomiche lo posiziona come un modello avanzato nella segmentazione cerebrale. UniverSeg, sebbene richieda un pool di immagini e relative etichette per l’inferenza, presenta limitazioni nell’ottenere maschere accurate quando si affrontano nuovi dati. L’utilizzo esclusivo sulle slice centrali aumenta ulteriormente l’inadeguatezza dell’approccio, risultando poco pratico su immagini tridimensionali e richiedendo un campionamento adeguato per ogni slice e asse. In generale, i risultati indicano che, nonostante alcuni modelli presentino limitazioni, la segmentazione cerebrale è un campo in evoluzione, con alcuni modelli come SynthSeg che dimostrano un notevole potenziale. Tut-

tavia, l'efficacia di ciascun modello dipende dall'applicazione specifica e dalla qualità dei dati utilizzati. L'efficienza dei modelli di segmentazione cerebrale è stata valutata principalmente in base al tempo di calcolo, considerando questa metrica come cruciale per valutare l'efficacia e la praticità delle soluzioni proposte. FAST, pur offrendo risultati accettabili nella segmentazione cerebrale, risulta caratterizzato da un tempo di esecuzione significativo. L'utilizzo della CPU per i calcoli e la necessità di processare tutte le slice dell'MRI contribuiscono alla lentezza dell'ottenimento dei risultati. Nonostante alcuni parametri possano essere regolati per migliorare le prestazioni, l'approccio di FAST risulta meno efficiente rispetto ad altri modelli. SAM, con la sua modalità di generazione automatica delle maschere, mostra un miglioramento delle prestazioni rispetto a FAST, ma presenta limitazioni nell'elaborazione di immagini bidimensionali e nell'uso esclusivo sulle slice centrali. L'efficienza di SAM è influenzata dalla necessità di post-processing e dalla mancanza di considerazione dei tempi della versione predefinita. SynthSeg si distingue per la sua eccellente efficienza temporale, soprattutto quando sfrutta la GPU. La capacità di elaborare l'intera risonanza magnetica e restituire risultati di alta qualità in tempi ridotti lo posiziona come un modello altamente efficiente per la segmentazione cerebrale. UniverSeg, sebbene offra una rapida inferenza durante la fase di runtime, presenta limitazioni nell'accettare solo slice di mezzo come input. La velocità di ottenimento dei risultati è indubbiamente un punto forte, ma l'inadeguatezza nell'uso su intere risonanze magnetiche e su più assi limita la sua efficacia pratica. In generale, la scelta del modello più adatto dipenderà dall'applicazione specifica e dai requisiti temporali. SynthSeg si distingue come il modello più efficiente in termini di tempo di calcolo e qualità dei risultati, mentre altri modelli mostrano vantaggi e limitazioni che devono essere considerati in base alle esigenze specifiche dell'utente.

Bibliografia

- [1] IDA. Abide project, Anno. Informatics & Data Analytics Core, University of Southern California.
- [2] Alzheimer's Disease Neuroimaging Initiative (ADNI). <https://adni.loni.usc.edu>. Laboratory of Neuro Imaging, University of Southern California.
- [3] Christopher G. Schwarz. Uses of human mr and pet imaging in research of neurodegenerative brain diseases. *Neurotherapeutics*, 18(2):661–672, 2021.
- [4] Benjamin Billot, Daniel Moyer, Neel Dey, Malte Hoffmann, Esra Abaci Turk, Borjan Gagoski, Ellen Grant, and Polina Golland. Se(3)-equivariant and noise-invariant 3d motion tracking in medical images, 2023.
- [5] Oscar Dabrowski, Jean-Luc Falcone, Antoine Klauser, Julien Songeon, Michel Kocher, Bastien Chopard, François Lazeyras, and Sébastien Courvoisier. Sismik for brain mri: Deep-learning-based motion estimation and model-based motion correction in k-space, 2023.
- [6] Alexander Frotscher, Jaivardhan Kapoor, Thomas Wolfers, and Christian F. Baumgartner. Unsupervised anomaly detection using aggregated normative diffusion, 2023.
- [7] Henrik Skibbe, Michal Byra, Akiya Watakabe, Tetsuo Yamamori, and Marco Reisert. Patchmorph: A stochastic deep learning approach for unsupervised 3d brain image registration with small patches, 2023.
- [8] Bethany Little, Carly Flowers, Andrew Blamire, Peter Thelwall, John-Paul Taylor, Peter Gallagher, David Andrew Cousins, and Yujiang Wang. Multivariate brain-cognition associations in euthymic bipolar disorder, 2023.

- [9] Teresa Lemainque, Nicola Pridöhl, Shuo Zhang, Marc Huppertz, Manuel Post, Can Yüksel, Masami Yoneyama, Andreas Prescher, Christiane Kuhl, Daniel Truhn, and Sven Nebelung. Time-efficient combined morphologic and quantitative joint mri based on clinical image contrasts – an exploratory in-situ study of standardized cartilage defects, 2023.
- [10] Navodini Wijethilake, Steve Connor, Anna Oviedova, Rebecca Burger, Tom Vercauteren, and Jonathan Shapey. A clinical guideline driven automated linear feature extraction for vestibular schwannoma, 2023.
- [11] Marta Moraschi, Silvia Tommasin, Laura Maugeri, Mauro DiNuzzo, Julien Cohen-Adad, Marco Masullo, Fabio Mangini, Lorenzo Giovannelli, Daniele Mascali, Tommaso Gili, Valerio Pisani, Ugo Nocentini, Federico Giove, and Michela Fratini. Influence of scanning plane on human spinal cord functional magnetic resonance echo planar imaging, 2024.
- [12] Christiane Posselt, Mehmet Yigit Avci, Mehmet Yigitsoy, Patrick Schünke, Christoph Kolbitsch, Tobias Schäffter, and Stefanie Remmele. Simulation of acquisition shifts in t2 flair mr images to stress test ai segmentation networks, 2023.
- [13] Giulia Baldini, Melanie Schmidt, Charlotte Zäske, and Liliana L. Caldeira. Mri scan synthesis methods based on clustering and pix2pix, 2023.
- [14] I-Jieh Liu, Ci-Siang Lin, Fu-En Yang, and Yu-Chiang Frank Wang. Language-guided transformer for federated multi-label classification, 2023.
- [15] Zhuoyao Xin, Christopher Wu, Dong Liu, Chunming Gu, Jia Guo, and Jun Hua. Enhancing ct image synthesis from multi-modal mri data based on a multi-task neural network framework, 2023.
- [16] Wei-Chun Kevin Tsai, Yi-Chien Liu, Ming-Chun Yu, Chia-Ju Chou, Sui-Hing Yan, Yang-Teng Fan, Yan-Hsiang Huang, Yen-Ling Chiu, Yi-Fang Chuang, Ran-Zan Wang, and Yao-Chia Shih. Using resnet to utilize 4-class t2-flair slice classification based on the cholinergic pathways hyperintensities scale for pathological aging, 2023.
- [17] Eberechi Wogu, Patrick Filima, Bradley Caron, Daniel Levitas, Peer Herholz, Catherine Leal, Mohammed F. Mehboob, Soichi Hayashi, Simisola Akintoye, George Ogoh, Tawe Godwin, Damian Eke, and Franco Pestilli. A labeled clinical-mri dataset of nigerian brains, 2023.

- [18] Long Chen, Liben Chen, Binfeng Xu, Wenxin Zhang, and Narges Razavian. Automatic detection of alzheimer’s disease with multi-modal fusion of clinical mri scans, 2023.
- [19] Fumio Hashimoto and Kibo Ote. Reconu-net: a direct pet image reconstruction using u-net architecture with back projection-induced skip connection, 2023.
- [20] Matthias Hadlich, Zdravko Marinov, Moon Kim, Enrico Nasca, Jens Kleesiek, and Rainer Stiefelhagen. Sliding window fastedit: A framework for lesion annotation in whole-body pet images, 2023.
- [21] Huidong Xie, Weijie Gan, Bo Zhou, Xiongchao Chen, Qiong Liu, Xueqi Guo, Liang Guo, Hongyu An, Ulugbek S. Kamilov, Ge Wang, and Chi Liu. Ddpet-3d: Dose-aware diffusion model for 3d ultra low-dose pet imaging, 2023.
- [22] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm. *IEEE Trans Med Imag*, 20(1):45–57, 2001.
- [23] S. Z. Li, S. Zhu, and T. F. Chan. Segmentation of brain mr images through a hidden markov random field model and the expectation–maximization algorithm. *IEEE Transactions on Medical Imaging*, 23(2):130–137, 2005.
- [24] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics, 2008.
- [25] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arXiv:2304.02643*, 2023.
- [26] Benjamin Billot, Douglas N. Greve, Oula Puonti, Axel Thielscher, Koen Van Leemput, Bruce Fischl, Adrian V. Dalca, and Juan Eugenio Iglesias. Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. *Medical Image Analysis*, 86:102789, May 2023.
- [27] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical*

- Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
- [28] Victor Ion Butoi, Jose Javier Gonzalez Ortiz, Tianyu Ma, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. Universeg: Universal medical image segmentation, 2023.
 - [29] M.W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C.F. Beckmann, M. Jenkinson, and S.M. Smith. Bayesian analysis of neuroimaging data in fsl. *NeuroImage*, 45(Suppl 1):S173–86, 2009.
 - [30] S.M. Smith, M. Jenkinson, M.W. Woolrich, C.F. Beckmann, T.E.J. Behrens, H. Johansen-Berg, P.R. Bannister, M. De Luca, I. Dronjak, D.E. Flitney, R. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J.M. Brady, and P.M. Matthews. Advances in functional and structural mr image analysis and implementation as fsl. *NeuroImage*, 23(Suppl 1):208–19, 2004.
 - [31] M. Jenkinson, C.F. Beckmann, T.E. Behrens, M.W. Woolrich, and S.M. Smith. Fsl. *NeuroImage*, 62:782–90, 2012.