



**Politecnico
di Torino**

Politecnico di Torino

Laurea Magistrale in Ingegneria Biomedica

A.a. 2020/2021

Dicembre 2021

**Sviluppo e generalizzazione di un
modello U-Net per il beamforming e la
segmentazione simultanea di
immagini ad ultrasuoni**

Relatrici:

Prof. Kristen Meiburger

Ing. Silvia Seoni

Candidato:

Nicola Casali

SOMMARIO

INTRODUZIONE	7
CAPITOLO 1.....	9
IMMAGINI AD ULTRASUONI	9
1.1 I PRINCIPI FISICI DEGLI ULTRASUONI	11
1.1.1 Generalità	11
1.1.1.1 L'equazione delle onde.....	11
1.1.1.2 L'impedenza acustica	14
1.1.1.3 Discontinuità in un mezzo	15
1.1.1.4 Attenuazione.....	17
1.2 LA GENERAZIONE DEL CAMPO DI PRESSIONE	19
1.2.1 La fisica del campo di pressione per onde continue	20
1.2.1.1 Caso di trasduttore reale formato da più elementi piezoelettrici	24
1.2.1.2 L'apodizzazione	27
1.3 IL BEAMFORMING.....	29
1.3.1 Composizione di un ecografo	29
1.3.1.1 Controlli principali di un ecografo	29
1.3.1.2 Diagramma a blocchi di un ecografo.....	31
1.3.2 Il beamforming per il filtraggio spaziale del segnale	32
1.3.2.2 Il beamforming in ultrasonografia.....	34
1.3.2.3 Algoritmi di beamforming alternativi	38
1.4 IL PROCESSAMENTO DEL SEGNALE ECOGRAFICO	40
1.4.1 Demodulazione IQ	40
1.4.1.1 Down-mixing.....	41
1.4.1.2 Filtraggio passa-basso	42
1.4.1.3 Decimazione	43
1.4.2 Dal segnale alla formazione dell'immagine	44
CAPITOLO 2.....	45
DEEP LEARNING: DAI FONDAMENTI ALLE APPLICAZIONI IN MEDICINA	45
2.1 IL DEEP LEARNING	47
2.1.1 Panoramica sui metodi di deep learning	49

2.1.1.1 Reti neurali artificiali (ANN)	50
2.1.1.2 Reti neurali convoluzionali (CNN)	52
2.1.1.3 Reti neurali ricorrenti (RNN)	55
2.1.1.4 U-Net	56
2.1.1.5 Rete generativa avversaria (GAN)	57
2.2 APPLICAZIONI DEL DEEP LEARNING IN MEDICINA	58
2.2.1 <i>Il deep learning nella genomica</i>	59
2.2.2 <i>Il deep learning nello studio dei biosegnali</i>	60
2.2.3 <i>Il deep learning nel bioimaging e medical imaging</i>	61
2.3 APPLICAZIONI DEL DEEP LEARNING IN ULTRASONOGRAFIA	64
2.3.1 <i>Computer-aided diagnosis (CAD)</i>	64
2.3.2 <i>Il deep learning per il processamento del segnale ecografico</i>	65
2.3.2.1 L'utilizzo delle reti neurali nel beamforming.....	66
CAPITOLO 3.....	69
MATERIALI E METODI.....	69
3.1 METODI.....	70
3.1.1 <i>Formulazione del problema</i>	70
3.1.2 <i>Architettura della rete</i>	71
3.1.2.1 Struttura dell'encoder	71
3.1.2.2 Struttura del decoder	73
3.1.3 <i>Preparazione dei dati</i>	73
3.1.3.1 Immagini con dynamic range ottimizzato	74
3.1.4 <i>Impostazioni di allenamento della rete</i>	75
3.1.5 <i>Metriche di valutazione dei risultati</i>	76
3.2 COSTRUZIONE DEL DATASET.....	78
3.2.1 <i>Simulazione delle immagini</i>	78
3.2.1.1 Dataset con cisti circolari anecogene.....	78
3.2.1.2 Introduzione di cisti iperecogene circolari	79
3.2.1.3 Introduzione di cisti doppie circolari.....	80
3.2.1.4 Introduzione di cisti di forma irregolare.....	81
3.2.1.5 Introduzione della simulazione del coefficiente di attenuazione dei tessuti.....	83
3.2.2 <i>Immagini sperimentali</i>	84
3.3 ALLENAMENTO DELLA RETE	84
CAPITOLO 4.....	87
RISULTATI E DISCUSSIONE	87

4.1 RISULTATI DEGLI ALLENAMENTI	87
4.1.1 Dataset con cisti anecogene	87
4.1.2 Dataset con cisti anecogene e iperecogene.....	90
4.1.2.1 Immagini simulate.....	90
4.1.2.2 Immagini sperimentali.....	91
4.1.3 Dataset con cisti anecogene, iperecogene e doppie	92
4.1.3.1 Immagini simulate.....	92
4.1.3.2 Immagini sperimentali.....	93
4.1.4 Dataset con cisti irregolari	94
4.1.4.1 Immagini simulate.....	94
4.1.4.2 Immagini sperimentali.....	97
4.1.5 Dataset con cisti irregolari e simulazione del coefficiente di attenuazione.....	100
4.1.5.1 Immagini simulate.....	101
4.1.5.2 Immagini sperimentali.....	104
4.2 DISCUSSIONE	108
CONCLUSIONI	111
BIBLIOGRAFIA.....	113

INTRODUZIONE

Le immagini ad ultrasuoni sono oggi ampiamente utilizzate come supporto alla diagnosi nel campo clinico. Questa modalità di imaging è stata riconosciuta negli anni come una delle più versatili, grazie ai costi contenuti e all'alta portabilità, specie se comparata con la CT e MRI. Tuttavia, la qualità delle immagini risulta spesso compromessa dalla presenza di rumore e artefatti, rendendo la diagnosi più difficoltosa al radiologo e dipendente dalle sue abilità ed esperienza.

Le acquisizioni in singola onda piana ottenute con l'algoritmo di beamforming Delay-And-Sum sono notoriamente poco dispendiose computazionalmente e garantiscono quindi alti *frame-rate*; tuttavia, l'alta risoluzione temporale viene pagata con la bassa qualità dell'immagine prodotta. Le soluzioni più innovative in questo campo hanno esplorato l'utilizzo del deep learning per la ricostruzione di immagini ecografiche, adoperando in particolar modo reti neurali convoluzionali (CNN).

In questa ottica, precedenti studi hanno mostrato la capacità delle reti DNN di estrarre informazioni direttamente dai dati grezzi a radiofrequenza, che permettano la formazione dell'immagine B-mode. Più nel dettaglio, è stato mostrato l'utilizzo di un modello U-Net con un encoder e due decoder, che consente di ricostruire l'immagine B-mode e contemporaneamente ottenere la maschera di segmentazione della struttura presente nell'immagine, fornendo in input alla rete solamente il dato a radiofrequenza grezzo. L'approccio permette di eliminare completamente le operazioni di beamforming classico, restituendo un'immagine di qualità superiore rimuovendo il rumore acustico presente all'interno degli oggetti anecoici.

A tal proposito, il lavoro di tesi si è posto l'obiettivo di ottenere un miglioramento dei risultati disponibili in letteratura, verificando gli effetti dovuti all'utilizzo di dataset più eterogeneo, che contenesse immagini con un numero variabile di cisti di differente ecogenicità e di forma irregolare, simulate grazie al pacchetto Field II su Matlab 2019b. Le performance sono state comparate testando tre differenti modelli per l'encoder dell'U-Net (VGG-13, VGG-16 e VGG-19), al fine di valutare eventuali risposte benefiche della rete utilizzando encoder con un numero maggiore di layer convoluzionali.

Inoltre, le immagini B-mode sono state generate con un valore di dynamic range tradizionale (60 dB) e con dynamic range automatico, specifico della singola immagine e stimato in modo tale da ottimizzare le caratteristiche dell'immagine.

L'attenzione maggiore si è focalizzata sui risultati conseguibili dalle predizioni su un test set di immagini sperimentali, comprendente 24 immagini phantom e 21 immagini *in-vivo* di carotide, acquisite mediante l'ecografo Verasonics® Vantage™ 128.

Il capitolo uno è dedicato alla trattazione dei principi dell'ultrasonografia classica, passando da brevi cenni sulla fisica degli ultrasuoni all'illustrazione delle tecniche odierne di beamforming, con particolare attenzione alla catena di elaborazione del segnale ecografico. Il capitolo due contiene nozioni essenziali per la comprensione dei moderni modelli di deep learning, osservando come questi hanno contribuito a rivoluzionare il mondo della medicina e in particolare dell'imaging biomedicale, terminando con l'esposizione dei modi con cui il deep learning si è inserito nell'ultrasonografia e dei metodi di beamforming.

Il lavoro sperimentale è oggetto dei capitoli tre e quattro, dove sono elencati materiali e metodi del progetto e infine i risultati ottenuti, rispettivamente.

CAPITOLO 1

IMMAGINI AD ULTRASUONI

L'importanza delle tecniche ultrasonografiche in medicina si deve sostanzialmente a tre fattori:

- l'innocuità per il paziente: grazie all'assenza di radiazioni ionizzanti, è noto come l'uso degli ultrasuoni a scopo di imaging sia completamente innocuo per il paziente, consentendo quindi misure prolungate nel tempo (monitoraggio) anche in pazienti a rischio, come le donne in gravidanza;
- l'ottima risoluzione temporale fornita, che consente l'uso dell'ultrasonografia anche per la documentazione di fenomeni molto rapidi, quali la contrazione cardiaca o l'apertura/chiusura delle valvole cardiache;
- lo sfruttamento di principi fisici differenti da quelli di altre tecniche di imaging come la radiografia, la RMN o la scintigrafia. L'ultrasonografia si basa infatti sulla riflessione e diffusione generate da interfacce acustiche, fornendo informazioni complementari a quelle ricavabili con le altre tecniche. Inoltre, un ecografo non richiede locali appositamente dedicati come nel caso della TC o RMN, tanto che per la sua eccellente portabilità viene definito da alcuni medici come un "moderno stetoscopio" [1], [2].

Gli ultrasuoni possono inoltre essere adoperati non solo per la produzione di immagini morfologiche, ma anche per applicazioni terapeutiche. In particolare, l'uso di fasci US focalizzati ad alta intensità ($100-10000 \text{ W/cm}^2$) permette la litotrissia di calcoli renali, la facoemulsificazione di cataratte o il trattamento di alcuni tumori mediante una procedura medica denominata *high-intensity focused US surgery* [1]. Lo sfruttamento dell'effetto Doppler consente inoltre agli apparecchi di stimare la velocità del sangue nei vasi, permettendo così indirettamente di valutare il grado di severità di una stenosi.

In Tab.1.1 e Tab.1.2 vengono riassunte le principali applicazioni biomedicali e mediche degli US.

<p><i>Medical diagnostic applications of US</i></p> <p>Diagnostic imaging</p> <p>Flow detection and quantification</p> <p>Tissue characterization</p> <p>Guide for interventional procedures</p> <p>Noninvasive hemodynamic</p> <p>Targeted microbubbles and US imaging for the assessment and tracking of biomarkers <i>in vivo</i> (imaging of angiogenesis, inflammation, and thrombus)</p> <p>Acoustic spectrography</p> <p>Therapy (focused, lithotripsy, phacoemulsification)</p> <p><i>Medical therapeutic applications of US</i></p> <p>Fracture healing</p> <p>Sonophoresis (increasing uptake of drugs by sonifying the site at which the drugs are applied)</p> <p>Sonoporation (manipulation of living cells to facilitate the transfer of drugs and genes across the cell membrane)</p> <p>Sonothrombolysis (plasminogen activator enhancement in stroke)</p> <p>Cataract surgery</p> <p>Extracorporeal shock wave lithotripsy</p> <p>High-intensity focused US (focused US surgery in cancer treatment)</p> <p><i>Nonmedical application of US</i></p> <p>Removal of biological contamination in water</p> <p>Removal of chemical contamination in water</p> <p>Air cleaning</p> <p>Acoustophoresis</p> <p>Removal of contamination in soil wastes</p> <p>Treatment of sewage sludge</p> <p>Crystallization and sonochemistry</p> <p>Food technology (degassing, drying, extraction, filtration, homogenization, meat tenderization, sterilization, and freezing)</p>	<p><i>Abdominal US</i></p> <p>Evaluation of abdominal organs</p> <p>Evaluation of peritoneal fluid</p> <p><i>Cardiology</i></p> <p>Cardiac anatomy</p> <p>Evaluation of cardiac function</p> <p>Hemodynamic evaluation (e.g., stroke volume, and pressures)</p> <p>Heart valve assessment</p> <p><i>Vascular</i></p> <p>Arterial and venous anatomy</p> <p>Arterial and venous hemodynamic</p> <p>Deep vein thrombosis evaluation</p> <p>Carotid for stroke assessment</p> <p><i>Obstetrics</i></p> <p>Routine fetal development</p> <p>Fetal location (intrauterine or ectopic)</p> <p>Multiple pregnancy</p> <p>Check for physical abnormalities and viability</p> <p>Determine sex</p>	<p><i>Urogenital</i></p> <p>Uterus, bladder, annexes, scrotum</p> <p>Endovaginal sonography</p> <p><i>Transrectal US</i></p> <p>Prostate</p> <p>Rectal cancer staging</p> <p><i>Ophthalmology</i></p> <p>Measure the cornea-retinal distance</p> <p>Image the parts of the eye without inspective access</p> <p><i>Thoracic US</i></p> <p>Pleural effusion</p> <p>Pneumothorax</p> <p>Interstitial diseases</p> <p>Consolidations</p> <p>Cortical tumors</p> <p>Diaphragmatic dynamics</p>
---	---	--

Tabella 1.1: Applicazioni biomedicali degli ultrasuoni.

Tabella 1.2: Applicazioni mediche degli ultrasuoni.

L'ecografia viene impiegata per lo studio di numerosi organi, essendo in grado di rilevare le alterazioni strutturali dovute alle più disparate condizioni patologiche. In particolare, può evidenziare noduli e calcoli di diversa natura che abbiano dimensioni di almeno 5-10 mm. Inoltre, dall'analisi dell'ecogenicità delle diverse strutture presenti nell'immagine, è possibile caratterizzare i tessuti differenziando le lesioni solide da quelle con infiltrazioni liquide.

Come si vedrà più nel dettaglio nei capitoli seguenti riguardanti la fisica degli ultrasuoni, il limite principale è costituito dal fatto che l'ecografia non è indicata per lo studio di alcuni distretti particolari circondati da aria e tessuto osseo (quali i polmoni e la teca cranica), per via della forte attenuazione che ne impedisce la penetrazione.

Da ultimo, come già accennato nelle righe precedenti, l'esame ecografico può essere condotto senza limiti di tempo e di indicazioni cliniche per via delle basse potenze adoperate durante l'imaging (10-100 mW/cm² in media). Un possibile rischio, che tuttavia si può manifestare solamente attraverso l'uso di potenze più elevate, è rappresentato dalla cavitazione del sangue, ossia la formazione di bolle gassose che potrebbero causare embolia nel paziente. L'effetto dell'energia pressoria è inoltre mitigato dall'intervallo di tempo relativamente lungo tra due impulsi [1].

1.1 I PRINCIPI FISICI DEGLI ULTRASUONI

1.1.1 Generalità [2]

1.1.1.1 L'equazione delle onde

Gli ultrasuoni possono essere definiti come onde sonore oltre la banda di percezione dell'*homo sapiens*, ossia a frequenze maggiori di 16-20 kHz. Sono costituiti da onde elastiche, longitudinali, che alternano fasi di compressione e di rarefazione, trasferendo energia meccanica e richiedendo dunque un mezzo per propagarsi, a differenza delle onde elettromagnetiche. Un generico mezzo può essere considerato come un insieme di particelle normalmente in stato di riposo, che quando vengono attraversate da un'onda longitudinale oscillano attorno alla loro posizione di equilibrio propagando l'onda (Fig.1.1). Questo insieme di particelle può anche essere immaginato come un insieme di masse connesse tra di loro da molle; perciò, in assenza di smorzamento, il loro moto viene governato dai valori di rigidità delle molle e dalle masse delle particelle.

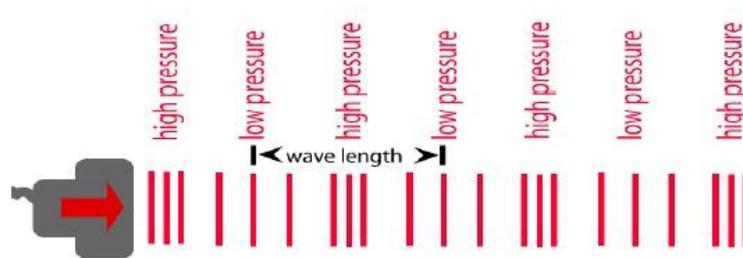


Figura 1.1: Rappresentazione della propagazione di un'onda longitudinale.

Supponendo che il moto sia sinusoidale, è possibile tracciare il grafico dello spostamento della particella $u(x)$ in funzione della pressione $p(x)$, la quale è proporzionale alla densità $\rho(x)$.

Le particelle vengono perturbate dalla loro posizione di equilibrio di una certa ampiezza (u) con una certa velocità (v). Questo cambiamento corrisponde ad una variazione nella pressione locale (p). Il semiciclo positivo dell'onda viene detto “di compressione”, mentre quello negativo” di “rarefazione” (Fig.1.2).

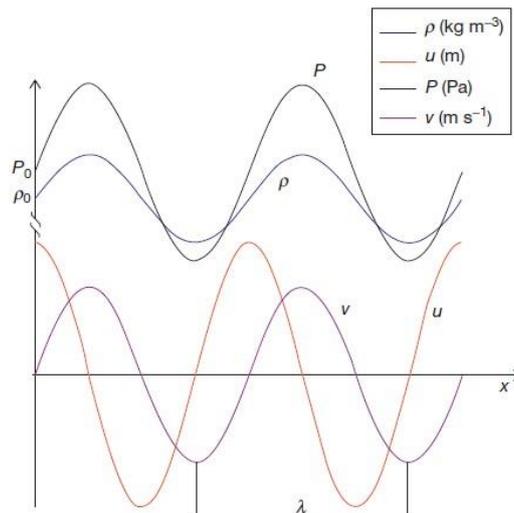


Figura 1.2: I grafici indicano la densità, lo spostamento, la velocità di una particella e la pressione in funzione della coordinata x .

Definendo con λ la lunghezza d'onda, cioè la distanza tra due successive onde di compressione o rarefazione, con c la velocità di propagazione (o velocità di fase) del suono nel mezzo considerato e con f la frequenza dell'onda, risulta:

$$c = \lambda \cdot f \quad (1.1)$$

Se la velocità di propagazione è indipendente dal valore della frequenza, allora si denota c con c_0 e il mezzo viene definito come “non dispersivo” (questa è un'approssimazione valida solo considerando dei mezzi ideali). Data l'esigenza in diagnostica medica di ottenere risoluzioni spaziali dell'ordine del millimetro o inferiori, le frequenze utilizzate spaziano solitamente da 1 MHz a 20 MHz.

L'onda può propagarsi solo se il mezzo è comprimibile, cioè se il volume del mezzo varia localmente con la pressione. Il modulo di comprimibilità di un mezzo viene definito come:

$$B = -V_0 \frac{\Delta p}{\Delta V} \quad (1.2)$$

Dove V_0 è un volume di controllo alla pressione p_0 . Si può dimostrare che la variazione dello spostamento Δu della particella lungo la direzione di propagazione è proporzionale alla pressione acustica p :

$$p = -B \frac{\partial u}{\partial x} \quad (1.3)$$

L'equazione del moto di Eulero è:

$$\rho \frac{dv}{dt} = -\nabla p$$

Nel caso di un'onda che si propaga lungo la direzione x e nel caso di piccoli segnali, l'equazione di Eulero diventa:

$$\rho \frac{\partial v}{\partial t} = -\frac{\partial p}{\partial x} \quad (1.4)$$

Occorre specificare che nel caso di piccoli segnali la densità che compare nell'Eq.1.4 può essere approssimata con la densità all'equilibrio (cioè in assenza di perturbazione) ρ_0 . Sostituendo ora l'Eq.1.3 nell'Eq.1.4 si ottiene:

$$\rho_0 \frac{\partial^2 u}{\partial t^2} = B \frac{\partial^2 u}{\partial x^2} \quad (1.5)$$

In questo modo è possibile ottenere l'equazione lineare della propagazione di un'onda acustica. Ricordando che la funzione spostamento contenuta nell'Eq.1.5 può essere espressa come:

$$u(x, t) = u_{max} \cos(\omega t - kx) \quad (1.6)$$

Sostituendo l'Eq.1.6 in Eq.1.5 e svolgendo le derivate, risulta che la velocità di fase dell'onda è ottenibile come la radice quadrata del rapporto tra il modulo di compressione e la densità del mezzo all'equilibrio:

$$c_0 = \sqrt{\frac{B}{\rho_0}} \quad (1.7)$$

In Tab.1.3 sono riportati i valori della velocità di propagazione e della densità di alcuni tessuti biologici.

	$c_0 (m s^{-1})$	$\rho_0 (kg m^{-3})$	$Z = c_0 \rho_0 (Rayls)$
Air (0 °C)	330	1.2	396
Water (20 °C)	1480	1000	1.48×10^6
Blood	1585	1060	1.68×10^6
Muscle	1600	1080	1.71×10^6
Fat	1440	930	1.34×10^6
Liver	1580	1050	1.66×10^6
Breast	1510	1020	1.54×10^6
Bone	4000	1900	7.60×10^6

Tabella 1.3: Valori di densità, velocità di propagazione e impedenza acustica per diversi tessuti e organi a temperatura fissata (37° C nel caso dei tessuti biologici).

Come visibile dalla tabella, nei mezzi biologici la velocità di propagazione differisce poco da un tessuto all'altro, risultando quindi in prima approssimazione uniforme all'interno del corpo umano ad eccezione dell'osso. Convenzionalmente per il tessuto molle si assume un valore medio, indipendente dalla frequenza dell'ultrasuono, di 1540 m/s. La velocità nei tessuti biologici è sempre superiore a quella nell'acqua, ad eccezione del grasso e del parenchima polmonare. La tecnica ultrasonografica è resa possibile proprio grazie al valore limitato della velocità di propagazione delle onde acustiche, rendendo così possibile la misurazione dei tempi di propagazione.

1.1.1.2 L'impedenza acustica

L'impedenza acustica (o caratteristica) è uno dei parametri più importanti in ultrasonografia. Viene definita come il rapporto tra la pressione esercitata dall'onda acustica su una particella del mezzo attraversato e la velocità di spostamento della particella:

$$Z = \frac{p}{v} \quad (1.8)$$

Sostituendo la relazione che lega la pressione e lo spostamento (Eq.1.3) nell'Eq.1.8 si ha:

$$Z = \frac{-B \frac{\partial u}{\partial x}}{\frac{\partial u}{\partial t}}$$

Avendo considerato che $\frac{du}{dt} \approx \frac{\partial u}{\partial t}$ nel caso di piccole variazioni di pressione (ossia si trascura la velocità delle particelle del mezzo rispetto alla velocità con cui si propaga l'onda). Le derivate possono essere svolte conoscendo la funzione spostamento (Eq.1.6). Semplificando il risultato si ottiene:

$$Z = \rho_0 c_0 \quad (1.9)$$

L'unità di misura adottata è il Rayl ($\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$). Nel caso di onde piane e in assenza di attenuazione, la pressione e la velocità sono in fase, per cui l'impedenza acustica è reale ed è data dal prodotto tra la velocità di fase e la densità del mezzo all'equilibrio (si veda l'ultima colonna della Tab.1.3). Nel caso di un mezzo ideale non dispersivo la velocità di fase non varia con la frequenza e di conseguenza anche l'impedenza acustica ne è indipendente. Si noti come il valore di impedenza acustica può anche assumere valore negativo, nel caso in cui l'onda si propaghi in verso opposto rispetto alla convenzione adottata.

1.1.1.3 Discontinuità in un mezzo

L'impedenza acustica appena definita è la grandezza fondamentale che consente il funzionamento della tecnica ultrasonografica e la possibilità di diversificare i vari tessuti nell'immagine. Un'onda che viaggia lungo una direzione e colpisce una discontinuità viene in parte riflessa e in parte rifratta; l'onda riflessa si propaga in verso opposto rispetto a quella incidente e con l'ampiezza modulata da un termine chiamato "fattore di riflessione".

Usando la notazione complessa si può esplicitare l'espressione dell'onda di pressione incidente, riflessa e trasmessa, rispettivamente:

$$\begin{aligned} p_I &= \exp\left(i\omega\left(t - \frac{1}{c_1}x\right)\right) \\ p_R &= R \exp\left(i\omega\left(t + \frac{1}{c_1}x\right)\right) \\ p_T &= T \exp\left(i\omega\left(t - \frac{1}{c_2}x\right)\right) \end{aligned} \quad (1.10)$$

Dove con R e T si indicano i coefficienti di riflessione e trasmissione. Considerando che l'onda incidente si propaga lungo la direzione dell'asse x per valori crescenti della coordinata spaziale, l'onda riflessa si propaga in verso opposto e perciò ha segno opposto nell'argomento dell'esponenziale.

Se nell'Eq.1.10 si pone $x=0$ (ossia si valuta cosa accade in corrispondenza del piano di separazione dei due mezzi, considerato idealmente di spessore infinitesimo) è facile verificare che:

$$1 + R = T \quad (1.11)$$

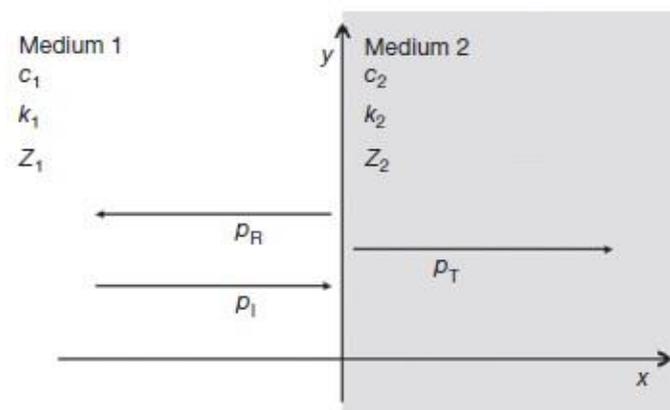


Figura 1.3: Due mezzi con impedenza acustica differente. Le frecce indicano l'onda di pressione incidente, riflessa e trasmessa.

I valori dei coefficienti sono:

$$R = \frac{Z_2 - Z_1}{Z_1 + Z_2} \quad (1.12)$$

$$T = \frac{2Z_2}{Z_1 + Z_2}$$

Se i due mezzi hanno impedenza acustica uguale, ne consegue che l'onda viene completamente trasmessa ($T=1$). Invece, nel caso in cui uno dei due mezzi abbia impedenza acustica molto superiore rispetto all'altra, come nel caso dell'interfaccia muscolo-osso, quasi tutta l'onda incidente viene riflessa ($|R| \approx 1$). Questa considerazione è fondamentale nel caso di interfacciamento delle onde sonore con l'aria, che ha impedenza acustica pressoché nulla e perciò l'onda incidente viene completamente riflessa. Da qui l'esigenza di utilizzare gel speciali tra trasduttore e tessuto per poter adattare l'impedenza acustica all'interfaccia di separazione.

Elevando i coefficienti R e T al quadrato si ottiene la frazione di energia posseduta dall'onda riflessa e da quella trasmessa rispetto a quella incidente.

Nel caso in cui l'onda incidente fosse inclinata di un certo angolo θ_i , una parte dell'onda verrebbe riflessa dello stesso angolo e una parte trasmessa con un angolo θ_t generalmente diverso da θ_i . Gli angoli sono ricavabili grazie all'applicazione della legge di Snell:

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{c_1}{c_2}$$

Dove c_1 e c_2 sono le velocità di propagazione nei due mezzi.

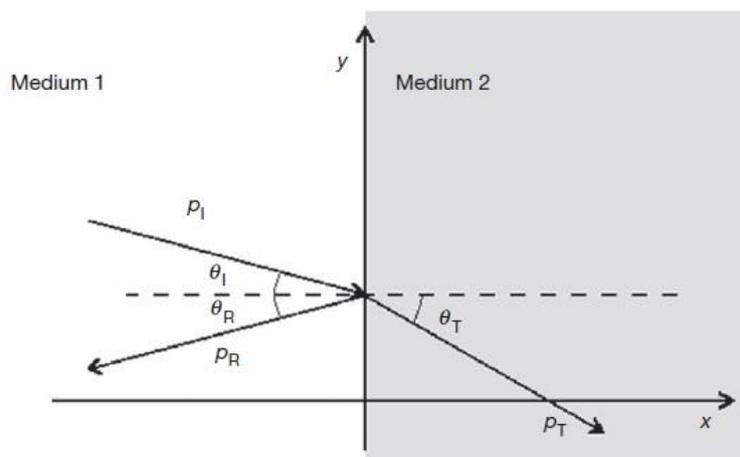


Figura 1.4: La figura mostra il caso in cui l'onda incidente non sia perpendicolare al piano di separazione.

$$R = \frac{Z_2 \cos \theta_i - Z_1 \cos \theta_t}{Z_1 \cos \theta_t + Z_2 \cos \theta_i}$$

$$T = \frac{2Z_2 \cos \theta_i}{Z_1 \cos \theta_t + Z_2 \cos \theta_i} \quad (1.13)$$

1.1.1.4 Attenuazione

Un'altra considerazione di carattere generale sugli ultrasuoni riguarda la descrizione dell'attenuazione dell'onda di pressione. L'intensità di un'onda che si propaga in un mezzo decresce esponenzialmente con la distanza a causa dell'assorbimento di energia e dei fenomeni di scattering. Per un'onda di pressione che si propaga lungo la direzione x , queste perdite possono essere modellate adoperando l'equazione di propagazione dell'onda piana (monocromatica), espressa in notazione complessa e modulata con una funzione esponenziale:

$$p(x, t) = p_0(x) e^{i(\omega t - kx)} \quad (1.14)$$

Dove:

$$p_0(x) = p_0(0) e^{-\alpha x} \quad (1.15)$$

Nel caso dei tessuti reali (viscoelastici) si osserva che il coefficiente di attenuazione dell'onda è dato dalla somma di due contributi principali:

$$\alpha = \alpha_{\text{assorbimento}} + \alpha_{\text{scattering}}$$

In particolare, l'assorbimento di energia avviene quando parte dell'energia meccanica dell'onda di pressione viene convertita in altre forme di energia (principalmente calore) durante l'attraversamento di un mezzo fisico (per via degli attriti meccanici causati dalle vibrazioni locali delle molecole che compongono la materia). Lo scattering avviene nel caso in cui il mezzo attraversato non sia omogeneo, in quanto parte dell'energia dell'onda acustica viene reindirizzata in direzioni differenti dalla direzione di propagazione originale. In particolare, strutture di dimensioni comparabili alla lunghezza dell'onda incidente (come i globuli rossi) possono riflettere parte dell'energia in tutte le direzioni.

In Eq.1.15 il coefficiente di attenuazione è espresso in α , ma solitamente in letteratura viene utilizzato il valore α_{dB} in $\text{dB} \cdot \text{cm}^{-1}$, ricavabile dal valore iniziale di pressione:

$$20 \log_{10} \left(\frac{p_0(0)}{p_0(x)} \right) = 20 \log_{10} e^{\alpha x} = 8.686 \cdot \alpha \cdot x = \alpha_{dB} \cdot x$$

Sia l'assorbimento sia lo scattering dipendono dalla lunghezza d'onda del fascio ad ultrasuoni, ossia dalla sua frequenza, per questo i sistemi di imaging sono dotati di dispositivi che consentono di aumentare l'amplificazione dell'onda di ritorno in base alla profondità (Time Gain Compensation). Questo implica che l'uso di elevate frequenze comporta una migliore risoluzione spaziale dell'immagine, al prezzo di una forte attenuazione che consente l'imaging solamente di strutture superficiali; al contrario, con basse frequenze si possono indagare distretti corporei più profondi al costo di una risoluzione spaziale peggiore.

Generalmente il coefficiente di attenuazione può essere espresso mediante funzioni dipendenti dalla frequenza elevata ad un certo coefficiente γ :

$$\alpha(f) = \alpha_0 + \alpha_1 |f|^\gamma \quad (1.16)$$

In cui solitamente α_0 è nullo, mentre l'esponente γ dipende dal tessuto, nonostante in genere si consideri una dipendenza lineare dalla frequenza ($\gamma = 1$), ottenendo quindi un'espressione semplificata:

$$\alpha_{dB} = \alpha_1 f$$

Esprimendo f in MHz e α_1 in $\text{dB}\cdot\text{cm}^{-1}\cdot\text{MHz}^{-1}$ [3].

Nel tessuto molle si può grossolanamente stimare il coefficiente di attenuazione con il valore di $1 \text{ dB}\cdot\text{cm}^{-1}\cdot\text{MHz}^{-1}$. L'osso rappresenta una barriera insormontabile per gli ultrasuoni ($8 \text{ dB}\cdot\text{cm}^{-1}\cdot\text{MHz}^{-1}$), mentre il polmone attenua molto di più per via dell'aria presente negli alveoli e delle riflessioni multiple che avvengono all'interno ($30 \text{ dB}\cdot\text{cm}^{-1}\cdot\text{MHz}^{-1}$). In Tab.1.4 vengono riportati alcuni valori del coefficiente di attenuazione in alcuni tessuti di interesse

Tessuto	Assorbimento (dB/MHz cm)
Aria	-
Grasso	0.5
Muscolo	2
Fegato	0.7
Cervello	1
Osso compatto	4-10
Acqua a 20° C	0.002

Tabella 1.4: Valori di attenuazione per diversi mezzi biologici.

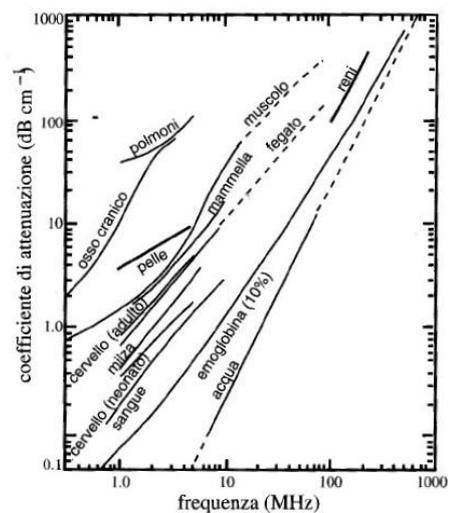


Figura 1.5: Curve di attenuazione di alcuni tessuti biologici.

Per via del limite di profondità imposto dal valore della frequenza dell'onda, si adoperano generalmente valori al di sotto di 10 MHz. In cardiologia, per esempio, si usano frequenze tra 2.5 e 5 MHz, per cui aventi lunghezze d'onda comprese tra 0.6-0.3 mm e penetrazione massima di 12-6 cm, rispettivamente [1].

1.2 LA GENERAZIONE DEL CAMPO DI PRESSIONE

Nelle sezioni precedenti sono stati riassunti i principi fondamentali su cui si basa la tecnica ultrasonografica. Per poter comprendere appieno il processo di formazione dell'immagine, è necessario discorrere della maniera con cui il trasduttore genera un campo di pressione. Questo aspetto consente di comprendere le possibilità e i limiti del sistema di imaging, specialmente in termini di risoluzione spaziale. Infatti, il modo con cui il trasduttore genera la forma d'onda si ripercuote sui parametri di qualità dell'immagine, che spesso sono in conflitto tra loro.

Alcuni dei parametri più importanti da tenere in considerazione sono:

- **risoluzione spaziale:** trattasi della più piccola distanza per la quale due riflettori differenti possono essere distinti nell'immagine. Può essere assiale (lungo la direzione di propagazione del fascio), laterale, o di elevazione (lungo il piano ortogonale alla direzione di propagazione);
- **risoluzione temporale:** rappresenta l'intervallo di tempo tra due immagini consecutive, espressa solitamente in Hz (immagini al secondo);
- **contrasto:** indica la capacità di distinguere visivamente due oggetti differenti (come due diversi tessuti) in un'immagine. Si esprime generalmente in dB ed è una misura relativa tra due valori di intensità;
- **profondità di penetrazione:** è la profondità massima per cui si mantiene un livello sufficientemente elevato di rapporto segnale-rumore (SNR). Si esprime generalmente in cm;
- **apertura dell'array:** rappresenta la grandezza fisica della superficie coperta dagli elementi piezoelettrici. Dipende dal numero di elementi, dalla loro grandezza e da come sono distribuiti. Si esprime in cm^2 ;
- **campo di vista** (spesso indicata con *field of view* o FOV): l'area o il volume dell'immagine ottenuta. Si esprime in cm^2 oppure in cm^3 [4].

Questi parametri sono strettamente collegati tra di loro, nonostante ciascuno venga definito in maniera diversa. Ad esempio, si può ottenere una risoluzione spaziale superiore o un FOV maggiore a scapito di una diminuzione della risoluzione temporale; per ottenere una profondità di penetrazione maggiore occorre diminuire la frequenza di trasmissione, peggiorando in tal modo la risoluzione spaziale; allargando il fascio trasmesso dal trasduttore si può visualizzare un'area più ampia, tuttavia al costo di generare un campo di pressione di ampiezza minore e ottenendo dunque dei valori di SNR inferiori.

1.2.1 La fisica del campo di pressione per onde continue [2], [3], [5]

La diffrazione è il fenomeno che descrive le onde generate dal trasduttore. Il fisico olandese Christiaan Huygens (1629-1695) definì la diffrazione come un processo di interferenza tra onde sferiche, generate da radiatori infinitesimi posti sulla superficie di un'apertura di dimensioni comparabili alla lunghezza d'onda dell'onda incidente.

Il campo generato dall'onda US dipende dalla superficie di apertura del trasduttore. Si consideri una superficie emittente di forma rettangolare con lunghezza L_x e larghezza L_y , estesa lungo il piano $z=0$, definita dalle coordinate $r_0 = (x_0, y_0, 0)$ e oscillante con velocità $v(r_0, t) = v(r_0)\cos(\omega t)$, in direzione normale alla superficie S (Fig.1.6). Si definisce con $v(r_0)$ e ω rispettivamente ampiezza e pulsazione della particella sulla superficie S. L'oscillazione della superficie emittente S genera un campo di pressione con ampiezza $p(r)$ e pulsazione ω nel punto $r = (x, y, z)$. Si assume inoltre che la propagazione dell'onda avvenga in un mezzo liquido non dispersivo.

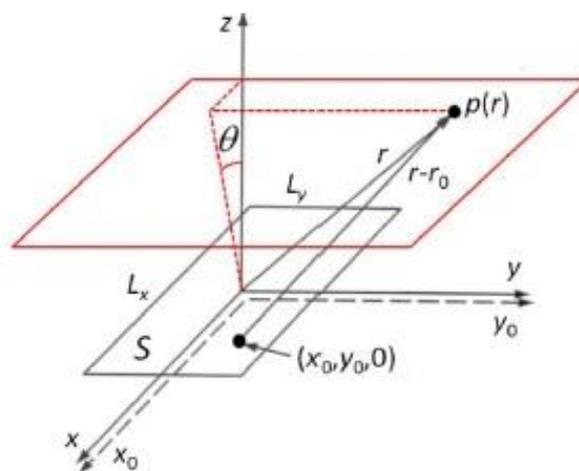


Figura 1.6: Lo schema geometrico utilizzato per descrivere una superficie emittente sul piano xy che irradia lungo z. Le coordinate dei punti appartenenti alla sorgente hanno il pedice 0.

La pressione $p(r, t)$ può essere calcolata mediante l'ausilio dell'integrale di Rayleigh-Sommerfeld, che si basa sul principio di Huygens-Fresnel:

$$p(r, t) = \frac{i\rho_0 c k}{2\pi} \iint_S \frac{e^{i[\omega t - k|r-r_0|]} v(r_0)}{|r-r_0|} dS = \frac{i\rho_0 c k v_0}{2\pi} \iint_S \frac{e^{i[\omega t - k|r-r_0|]} a(r_0)}{|r-r_0|} dS \quad (1.17)$$

Dalla relazione $\rho_0 c = \frac{p_0}{v_0}$ consegue:

$$p(r, t) = \frac{i p_0}{\lambda} \iint_S \frac{e^{i[\omega t - k|r-r_0|]} a(r_0)}{|r-r_0|} dS \quad (1.18)$$

Con:

- ρ_0 la densità del mezzo all'equilibrio;
- c velocità di propagazione dell'US nel mezzo;
- k numero d'onda;
- $\omega = 2\pi f$ è la velocità angolare;
- $v_0(r_0) = v_0 a(r_0)$ è l'ampiezza della velocità normale alla superficie della particella sulla superficie S , con $a(r_0)$ una funzione peso di v_0 (apodizzazione).

Assumendo ora che le dimensioni dell'elemento trasmittente siano molto più piccole della distanza a cui viene calcolata $p(r)$, ossia $r \gg r_0$, è possibile utilizzare l'approssimazione di Fresnel per semplificare il valore della distanza $|r - r_0|$:

$$|r - r_0| = \sqrt{z^2 + (x - x_0)^2 + (y - y_0)^2}$$

$$|r - r_0| \cong z \left[1 + \frac{1}{2} \left(\frac{x - x_0}{z} \right)^2 + \frac{1}{2} \left(\frac{y - y_0}{z} \right)^2 \right] \cong z$$

Quindi si approssima il valore della distanza $|r - r_0|$ con z . Sostituendo nell'Eq.1.18 si ottiene:

$$p(r, t) = \frac{i p_0 e^{i(\omega t - kz)} e^{-ik\left(\frac{x^2+y^2}{2z}\right)}}{z\lambda} \iint_S e^{-ik\left(\frac{x_0^2+y_0^2}{2z}\right)} a(x_0, y_0, 0) e^{ik\left(\frac{xx_0+yy_0}{z}\right)} dx_0 dy_0 \quad (1.19)$$

Nel caso di un trasduttore di forma rettangolare, come è di consuetudine nei dispositivi ecografici, allora si può utilizzare la funzione porta (*boxcar*) per descrivere la funzione di apodizzazione:

$$a(x_0, y_0, 0) = \prod_{L_x} x_0 \prod_{L_y} y_0 \quad (1.20)$$

$$a(x_0, y_0, 0) = a_x(x_0) a_y(y_0)$$

Dove $\prod_L(\cdot)$ è la funzione porta di larghezza L e ampiezza unitaria.

L'integrale di Eq.1.19 può essere scomposto nel prodotto di due integrali (uno lungo L_x e l'altro lungo L_y), perciò è possibile risolvere il problema considerando solamente una delle due direzioni. Il campo di pressione completo sarà poi ottenibile moltiplicando i contributi per le due dimensioni:

$$p(x, y, z) = p_x(x, z)p_y(y, z)$$

Trascurando il termine di fase $ie^{i(\omega t - kz)}$, l'integrale di Rayleigh-Sommerfeld per ciascuna direzione è:

$$p_x(x, z) = \Psi_{0x} e^{-ik\left(\frac{x^2}{2z}\right)} \int_{-\infty}^{\infty} e^{-ik\left(\frac{x_0^2}{2z}\right)} \prod_{L_x}(x_0) e^{ik\left(\frac{xx_0}{z}\right)} dx_0 \quad (1.21)$$

Con:

$$\Psi_{0x} = \sqrt{\frac{p_0}{z\lambda}}$$

Supponendo che il punto valutato in (x, z) sia abbastanza lontano dalla superficie del trasduttore (ossia valgono le condizioni di Fraunhofer: $\lambda z \gg x^2$; $\lambda z \gg x_0^2$), si può trascurare il termine che dipende da x_0^2 e x^2 :

$$p_x(x, z) = \Psi_{0x} \int_{-\infty}^{\infty} \prod_{L_x}(x_0) e^{i2\pi\left(\frac{x}{\lambda z}\right)x_0} dx_0 \quad (1.22)$$

Riconoscendo nell'Eq.1.22 la definizione dell'antitrasformata di Fourier nella variabile $\frac{x}{\lambda z}$, l'equazione diventa:

$$p_x(x, z) = L_x \Psi_{0x} \text{sinc}\left(\frac{xL_x}{\lambda z}\right)$$

Si rammenti che l'antitrasformata di una funzione porta è una funzione seno cardinale:

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

L'approssimazione di Fraunhofer consente di scrivere la coordinata x in funzione di z , come visibile dallo schema geometrico di Fig.1.7:

$$x = |r| \text{sen}\vartheta \cong z \text{sen}\vartheta$$

$$p_x(x, z) \cong L_x \Psi_{0x} \text{sinc} \left(\frac{L_x}{\lambda} \sin \vartheta \right) \quad (1.23)$$

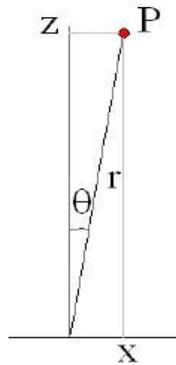


Figura 1.7: Schema geometrico per l'approssimazione di Fraunhofer.

Si introduce in questo modo una dipendenza esplicita dall'angolo di direzione del fascio. I valori della pressione sono nulli per valori tali per cui vale:

$$\sin \vartheta = \frac{\lambda}{L_x}$$

Questo risultato è di notevole importanza, in quanto fornisce un'espressione che consente di valutare la risoluzione laterale dell'immagine, determinata dalla posizione del primo zero della funzione che descrive il campo di pressione, ossia dall'ampiezza del lobo principale (Fig.1.8).

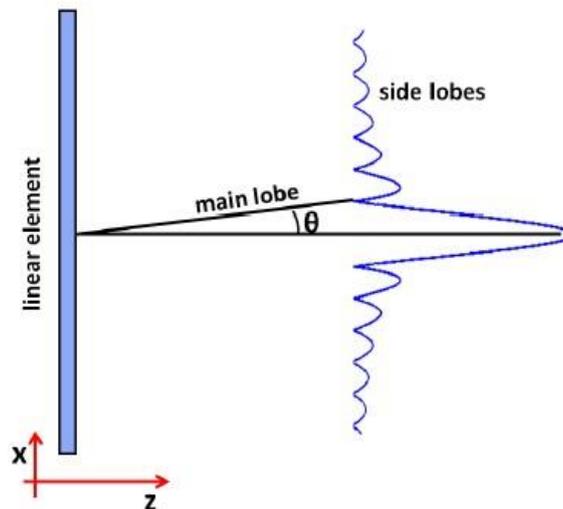


Figura 1.8: Rappresentazione schematica del campo di pressione in condizioni di campo lontano generato da un elemento lineare rettangolare.

Nella pratica, come misura per la risoluzione laterale, viene spesso utilizzata la larghezza a metà altezza (FWHM), ossia si considerano come estremi i punti a -6 dB dal valore massimo del campo di pressione, anziché i punti in cui questo si annulla. Si può dimostrare che questa vale:

$$FWHM = 1.206 \frac{\lambda z}{L_x} \quad (mm) \quad (1.24)$$

Da qui si può immediatamente osservare il motivo per cui ad una diminuzione della lunghezza d'onda (ossia una frequenza maggiore del segnale trasmesso dalla sonda) corrisponda una risoluzione laterale migliore, così come per grandi valori della lunghezza L_x . Tuttavia, le dimensioni della sonda non possono superare certi limiti spaziali, in particolare nel caso di applicazioni specifiche come l'imaging transtoracico ed endocavitario.

Nella Fig.1.8 si nota anche la presenza di lobi laterali oltre a quello principale, come deducibile dalla presenza della funzione *sinc* in Eq.1.23, che decrescono alla velocità di x^{-1} . La presenza dei lobi laterali rappresenta un effetto indesiderato durante una scansione ad ultrasuoni, poiché provocano la formazione di echi provenienti da direzioni diverse dalla linea di scansione, dove è presente il lobo principale.

Alla luce di queste considerazioni, alcune delle strategie principali adottate per il miglioramento della qualità dell'immagine sono: la focalizzazione del fascio, che riduce la larghezza del lobo principale; l'apodizzazione, che a seconda dei pesi utilizzati cerca di raggiungere un compromesso tra la velocità di decadimento e l'ampiezza dei lobi laterali e la larghezza del lobo principale.

1.2.1.1 Caso di trasduttore reale formato da più elementi piezoelettrici

In precedenza, si è discusso il caso in cui la sonda fosse formata da un unico elemento sensibile lineare di ampiezza L . Questo rappresenta un'approssimazione, poiché in realtà una sonda è costituita da un certo numero di elementi piezoelettrici (Fig.1.9) vicini tra loro. La distanza d tra i centri degli elementi adiacenti è costante ed è denominata *pitch*, mentre la lunghezza (*width*) di ciascun elemento rettangolare viene indicata con w .

Questo tipo di configurazione può essere descritta modificando l'equazione 1.20, precedentemente utilizzata per la descrizione di una sonda con un unico elemento sensibile lineare:

$$\prod_{L_x} [x_0] \rightarrow \prod_{L_x} [x_0] \left(\prod_w [x_0] * \sum_m \delta(x_0 - md) \right) \quad (1.25)$$

Dove:

- m rappresenta l'indice dell'elemento considerato;
- d è il pitch;
- $*$ è il simbolo di convoluzione;
- δ è la delta di Dirac;
- w è l'ampiezza dell'elemento.

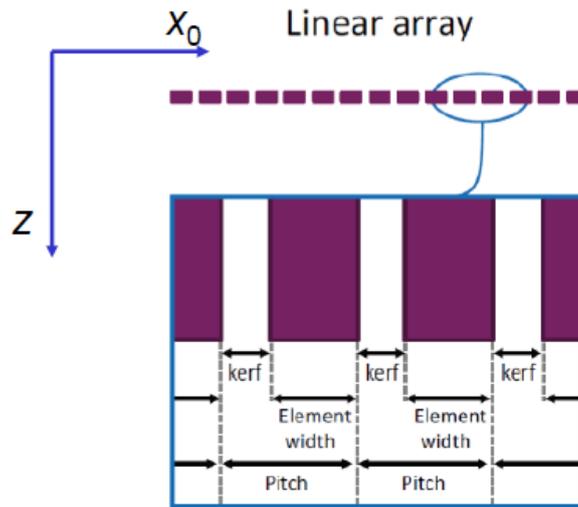


Figura 1.9: Nomenclatura di un array lineare per la generazione di ultrasuoni.

Dato che la convoluzione di una funzione con un treno di delta di Dirac rappresenta una ripetizione periodica della funzione stessa (teorema del campionamento), l'Eq.1.25 descrive la situazione della Fig.1.9.

Combinando l'Eq.1.22 con l'Eq.1.25 si ottiene:

$$p_x(x, z) = \Psi_{0x} \mathfrak{F}^{-1} \left\{ \prod_{L_x} [x_0] \left(\prod_w [x_0] * \sum_m \delta(x_0 - md) \right) \right\}$$

$$p_x(x, z) = \Psi_{0x} \mathfrak{F}^{-1} \left\{ \prod_w [x_0] \right\} \mathfrak{F}^{-1} \left\{ \prod_{L_x} [x_0] \sum_m \delta(x_0 - md) \right\} \quad (1.26)$$

Svolgendo le antitrasformate di Fourier:

$$p_x(x, z) = \Psi_{0x} \frac{L_x w}{d} \operatorname{sinc}\left(\frac{wx}{\lambda z}\right) \sum_m \operatorname{sinc}\left[\frac{L_x}{\lambda} \left(\frac{x}{z} - \frac{m\lambda}{d}\right)\right]$$

Utilizzando anche in questo caso l'approssimazione di Fraunhofer:

$$p_x(\vartheta, z) = \Psi_{0x} \frac{L_x w}{d} \operatorname{sinc}\left(\frac{w}{\lambda} \sin\vartheta\right) \sum_m \operatorname{sinc}\left[\frac{L_x}{\lambda} \left(\sin\vartheta - \frac{m\lambda}{d}\right)\right] \quad (1.27)$$

L'Eq.1.27 descrive il campo di pressione al variare dell'angolo azimutale (Fig.1.10). A causa del campionamento dell'elemento trasmettente, la situazione di Fig.1.10 differisce da quella di Fig.1.8 in quanto compare una ripetizione periodica del lobo principale. Intuitivamente si comprende che questo è da attribuirsi al treno di *sinc* involupate, che compaiono in Eq.1.27 per via del campionamento spaziale dell'array. I nuovi lobi vengono solitamente denominati "lobi di *grating*". I lobi di *grating* appaiono ad angoli di valore:

$$\vartheta = \arcsin\left(\frac{m\lambda}{d}\right) \quad (1.28)$$

Ne consegue che, scegliendo valori più bassi di frequenza e piccoli valori del pitch, è possibile traslare i lobi di *grating* verso angoli maggiori e quindi più lontani dalla direzione di insonazione. Dall'Eq.1.28 consegue che scegliendo i parametri in modo che $\frac{\lambda}{d} > 1$, i lobi di *grating* non compaiono nella zona insonata perché presenti ad $\vartheta > 90^\circ$.

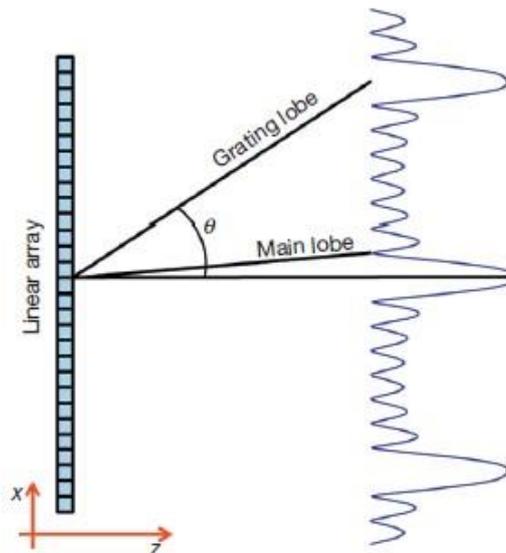


Figura 1.10: Rappresentazione del campo di pressione derivante dall'Eq.1.27. Si può notare la comparsa dei lobi di "grating" per via del campionamento spaziale degli elementi dell'array.

Lo specchietto di Fig.1.11 riassume i legami tra geometria dell'array e forma del campo di pressione generato.

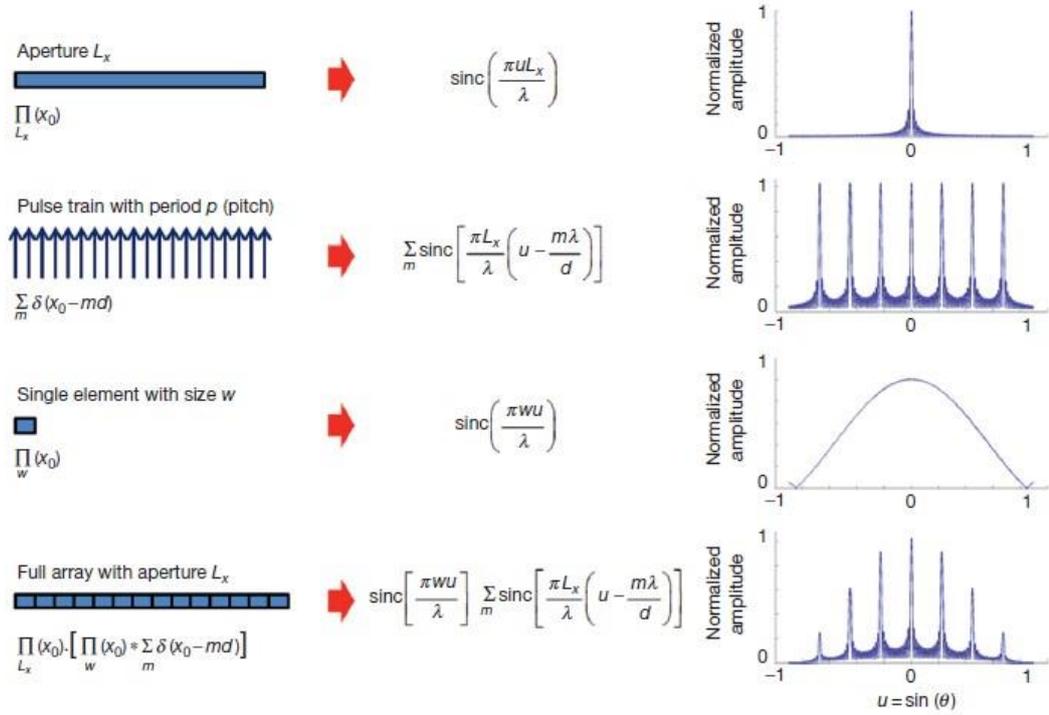


Figura 1.11: Derivazione del campo di ultrasuoni generato da un array lineare ed effetto del campionamento spaziale.

1.2.1.2 L'apodizzazione

Un ruolo fondamentale svolto dal *beamformer* è quello della generazione dei pesi da applicare in ogni istante agli elementi dell'array, sia in trasmissione sia in ricezione. Se si introduce nell'Eq.1.26 la funzione peso $a(x_0)$:

$$p_x(x, z) = \Psi_{0x} \mathfrak{F}^{-1} \left\{ \prod_w [x_0] \right\} \mathfrak{F}^{-1} \left\{ \prod_{L_x} (x_0) a(x_0) \sum_m \delta(x_0 - md) \right\} \quad (1.29)$$

Svolgendo le antitrasformate si ottiene:

$$p_x(\vartheta, z) = \Psi_{0x} \frac{w}{d} \text{sinc} \left(\frac{w}{\lambda} \sin \vartheta \right) \sum_m A_{L_x} \left(\frac{\sin \vartheta}{\lambda} - \frac{m}{d} \right) \quad (1.30)$$

Con A_{L_x} che rappresenta l'antitrasformata di Fourier di $a(x_0)$.

In Fig.1.12 vengono mostrate alcune delle finestre più comunemente utilizzate.

Si può osservare come la finestratura sia necessaria per ottenere un compromesso tra lobo principale (idealmente di larghezza infinitesima) e lobi secondari (idealmente di ampiezza nulla). Poiché i due obiettivi sono in contrasto tra loro, l'unità che gestisce la formazione dell'onda deve operare in modo da raggiungere un compromesso in base a quali parametri si vuole enfatizzare nell'immagine.

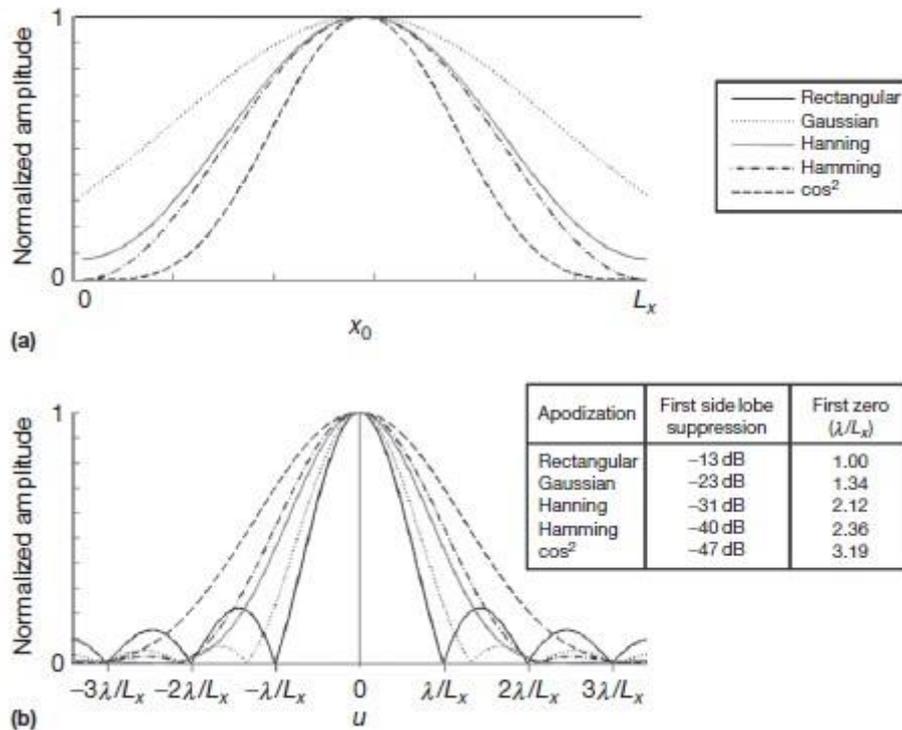


Figura 1.12: L'effetto dell'applicazione di diverse funzioni di apodizzazione: A) sul livello acustico normalizzato del segnale; B) disegnato in funzione di $u = \text{sen } \theta$.

1.3 IL BEAMFORMING

Dopo aver completato la trattazione sui principi fisici dell'ultrasonografia e del campo di pressione generato dal trasduttore, risulta utile esporre qualche cenno sulla composizione di un moderno ecografo e del relativo schema a blocchi, in modo da poter comprendere il ruolo posseduto dal beamformer e l'integrazione che ha con gli altri componenti.

1.3.1 Composizione di un ecografo [3]

La Fig.1.13 mostra le parti esterne che compongono un ecografo. Il display è montato su un apposito carrello per poterne consentire la mobilità. Sulla destra sono visibili alcuni tipi di trasduttori disponibili all'uso, connessi al sistema mediante appositi connettori. Al di sotto del display è presente il pannello di controllo, contenente la tastiera e le numerose leve e manopole per il controllo dell'intero sistema.

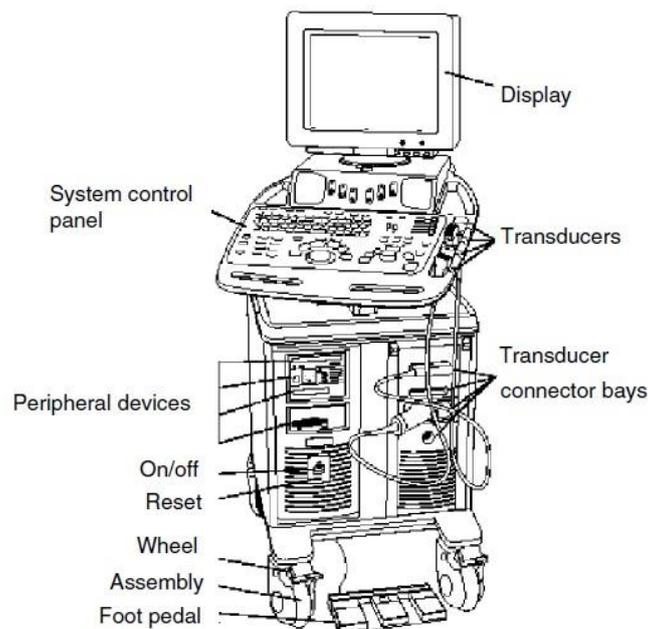


Figura 1.13: Visione d'insieme di un ecografo (Philips Medical Systems).

1.3.1.1 Controlli principali di un ecografo

Di seguito verranno mostrati alcuni dei controlli principali di un ecografo. Considerando il modo in cui sono organizzati e i loro nomi potrebbero variare a seconda del fabbricante, questi verranno presentati in maniera schematica e il più possibile generale.

Basandosi sulla Fig.1.14, i principali controlli presenti nel pannello del controllo del sistema possono essere riassunti in:

- **probe select:** questo switch consente di selezionare la sonda che si desidera adoperare, in quanto solitamente possono essere connessi da due a quattro trasduttori contemporaneamente;
- **mode select:** consente di selezionare il modo di scansione, ad esempio *B-Mode*, *color flow*, *Doppler*, *M-Mode*, etc.;
- **depth of scan control:** viene utilizzato per regolare il campo di vista (FOV);
- **focus:** consente di posizionare il punto focale del fascio a seconda della zona di interesse nell'immagine. Possono essere posizionati anche più punti focali, a discapito però del frame rate;
- **time gain compensation (TGC):** questa serie di controlli permette di compensare le perdite del segnale dovute all'attenuazione dell'onda. Generalmente sono dei controlli a scorrimento che consentono di modificare l'amplificazione a seconda della profondità da cui proviene il segnale. L'immagine viene suddivisa lungo l'asse z in diverse zone di profondità, ognuna delle quali controllata dalla propria leva del TGC;
- **transmit level control:** consente la regolazione dell'ampiezza dell'onda prodotta dal trasduttore. Il feedback per questo tipo di controllo è rappresentato dall'indice termico e dall'indice meccanico che appaiono sul display;

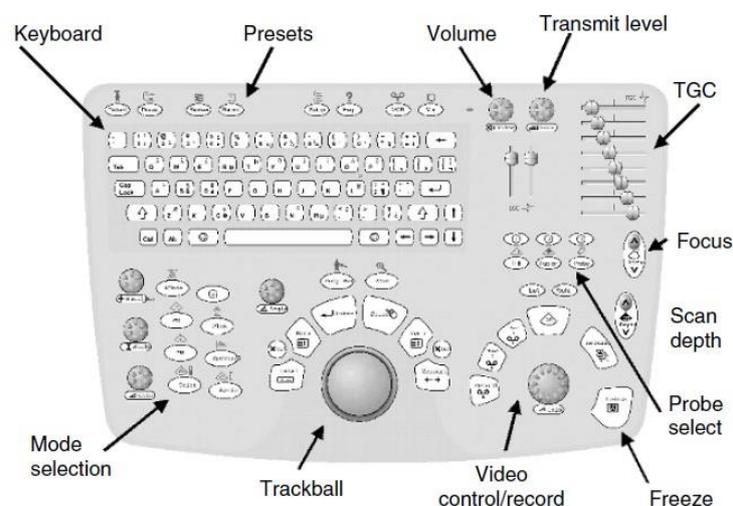


Figura 1.14: Tastiera di un ecografo (Philips Medical Systems).

- **display controls:** questi controlli consentono l'ottimizzazione dei parametri che formano l'immagine, includendo un controllo per la compressione logaritmica (ossia per modificare il *dynamic range*), la selezione di diverse curve per il *preprocessing* e *postprocessing*, la selezione della *colormap* e la possibilità di modificare la dimensione dell'immagine.

1.3.1.2 Diagramma a blocchi di un ecografo

La Fig.1.15 mostra lo schema a blocchi dei componenti interni dell'ecografo. Di seguito una rapida descrizione dei vari blocchi:

- **user interface:** la maggior parte dei blocchi mostrati sono sostanzialmente nascosti all'utente, il quale interagisce principalmente mediante la tastiera e il display, che si trovano raggruppati nello schema della Fig.1.15 sotto il nome di "interfaccia utente". Questa inoltre include anche interfacce di I/O per trasmettere e ricevere informazioni da e verso l'ecografo;
- **controller:** un tipico sistema possiede uno o più microprocessori o un PC che dirige le operazioni del sistema. Il sistema di controllo gestisce i segnali dei dispositivi di input, come la tastiera, ed esegue i comandi per controllare l'hardware. Dirige le impostazioni necessarie per la trasmissione e la ricezione delle forme d'onda, oltre al processamento del segnale, la visualizzazione dell'immagine e la stima in tempo reale del livello acustico emesso;
- **front end:** rappresenta l'interfacciamento tra il segnale che entra ed esce dal trasduttore. Il microprocessore trasmette il segnale che consente ai circuiti che regolano la trasmissione dell'onda di eccitare gli elementi piezoelettrici del trasduttore. In ricezione, raccoglie i segnali provenienti dal corpo indirizzandoli al TGC, per poi essere inviati all'unità di beamforming;
- **scanner (beamforming and signal processing):** questo blocco è fondamentale perché consente di creare le linee che compongono l'immagine applicando gli opportuni ritardi ai segnali ricevuti. Nel caso di unità di beamforming digitali, il segnale viene dapprima convertito mediante un convertitore A/D. Inoltre, il beamformer si occupa in fase di trasmissione di inviare gli impulsi agli elementi eccitabili;

- **back end:** questo blocco è necessario per svolgere le funzioni associate alla formazione dell'immagine e alla sua successiva visualizzazione. L'input di questo gruppo è un insieme di echi pulsati involuppati, provenienti dal processo di beamforming del segnale a radiofrequenza (RF). La formazione dell'immagine viene effettuata organizzando le linee di scansione e inviandole al *digital scan converter*, che si occupa di trasformare il segnale in un formato (*raster scan*) adatto ad essere visualizzato sul display.

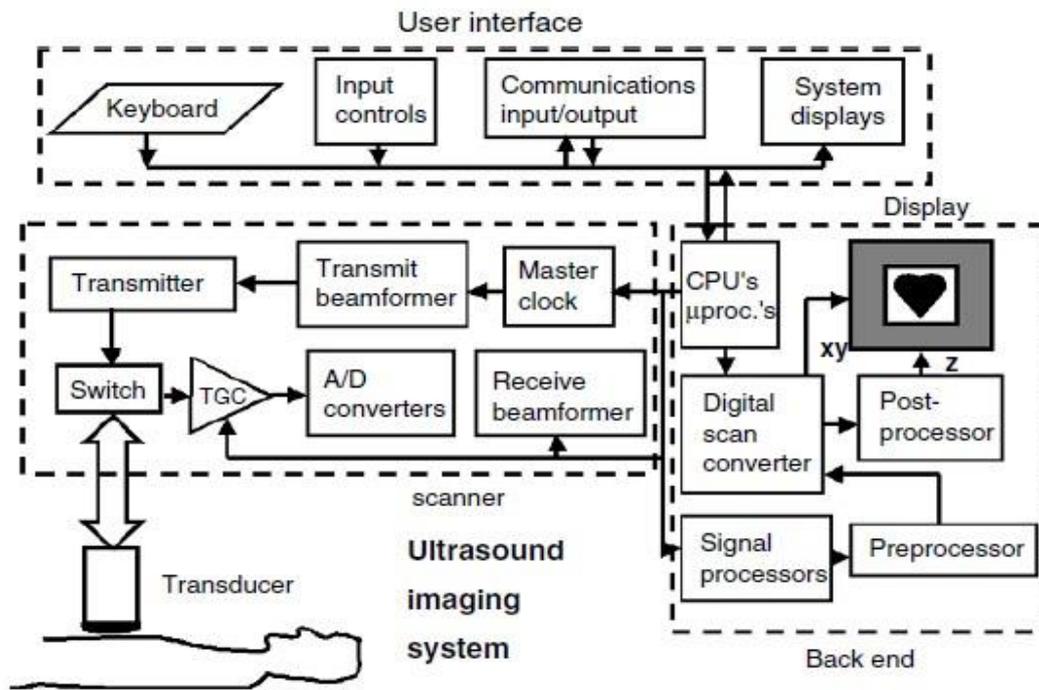


Figura 1.15: Diagramma a blocchi di un dispositivo ecografico.

1.3.2 Il beamforming per il filtraggio spaziale del segnale [6]

Il *beamforming* può essere definito come una tecnica di processamento del segnale, applicata in modo particolare nel mondo delle telecomunicazioni. Questa ha trovato larga applicazione nei RADAR e nei SONAR, oltre che in geofisica (per la mappatura della crosta terrestre e l'esplorazione di giacimenti petroliferi), astrofisica e naturalmente nell'imaging ad ultrasuoni, come riassunto nella Tab.5.

La tecnica viene adoperata, come già accennato, sia in fase di trasmissione del segnale (emissione di energia da un'antenna) sia in fase di ricezione (acquisizione di energia di un'antenna). In fase di ricezione il beamforming agisce come filtro spaziale. L'obiettivo è

Application	Description	References
RADAR	phased-array RADAR; air traffic control; synthetic aperture RADAR	Brookner [1985]; Haykin [1985]; Munson et al. [1983]
SONAR	source localization and classification	Knight et al. [1981]; Owsley [1985]
Communications	directional transmission and reception; sector broadcast in satellite communications	Mayhan [1976]; Compton [1978]; Adams et al. [1980]
Imaging	ultrasonic; optical; tomographic	Macovski [1983]; Pratt [1978]; Kak [1985]
Geophysical Exploration	earth crust mapping; oil exploration	Justice [1985]
Astrophysical Exploration	high resolution imaging of the universe	Readhead [1982]; Yen [1985]
Biomedical	fetal heart monitoring; tissue hyperthermia; hearing aids	Widrow et al. [1975]; Gee et al. [1984]; Peterson et al. [1987]

Tabella 5: Alcune applicazioni del beamforming per il filtraggio spaziale dei segnali.

quello di poter acquisire il segnale in arrivo da una direzione desiderata in presenza di più segnali sovrapposti e di rumore, che generano interferenza. Il beamformer opera in modo da poter separare più segnali che hanno un contenuto spettrale sovrapponibile (perciò non è possibile filtrare il segnale attenuando la banda in cui è presente l'interferenza) ma che originano da sorgenti distinte, isolando solamente quello proveniente dalla direzione di interesse, come mostrato nella Fig.1.16.

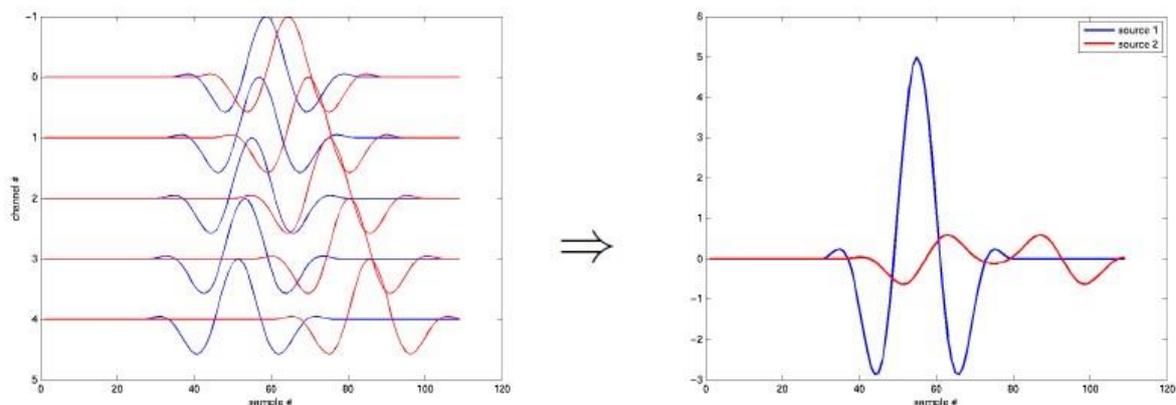


Figura 1.16: Utilizzo del beamforming come filtro spaziale. L'immagine mostra la ricezione di cinque segnali originati da due sorgenti differenti (blu e rosso). Conoscendo la posizione della sorgente di interesse e dei sensori di ricezione, è possibile calcolare il valore dei ritardi da applicare a ciascun canale, in modo che i contributi del segnale blu si sommino costruttivamente, mentre quelli del rosso distruttivamente.

I campioni dei segnali acquisiti dall'array di sensori vengono sommati coerentemente mediante l'applicazione di opportuni ritardi, calcolati in base alla direzione da cui proviene il segnale che si vuole isolare. Questo è facilmente comprensibile osservando il diagramma polare di radiazione (in letteratura noto come *beam pattern* o *radiation pattern*) di Fig.1.17, che confronta il guadagno di potenza (in ordinata) con il valore dell'angolo azimutale. Si nota come il

guadagno sia massimo nella direzione desiderata (il lobo principale), tuttavia sono presenti anche contributi minori provenienti da altre direzioni (lobi secondari) che si sommano incoerentemente (il guadagno di potenza per queste direzioni sarebbe nullo solo se i segnali, dopo l'applicazione dei ritardi, fossero esattamente in opposizione di fase tra di loro).

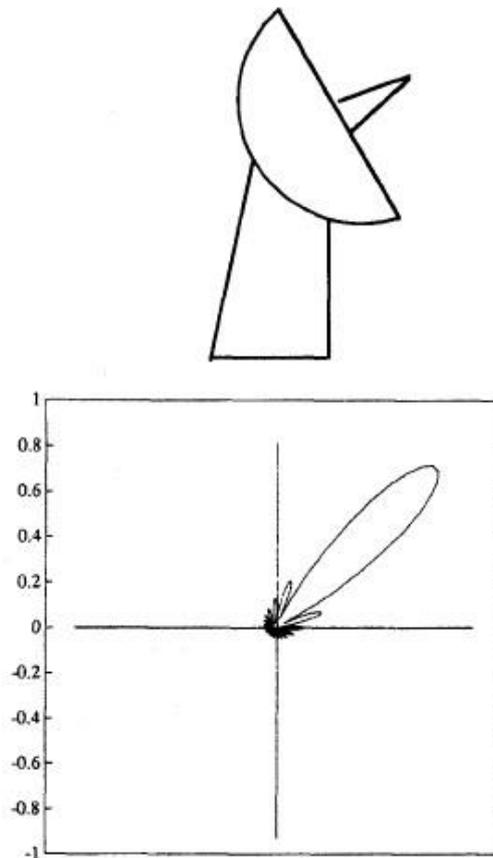


Figura 1.17: Digramma di radiazione polare di un'antenna. Grazie al beamforming è possibile isolare solamente il segnale proveniente dalla direzione di interesse, schematicamente rappresentata dalla posizione verso cui è orientata l'antenna.

1.3.2.2 Il beamforming in ultrasonografia

Come visto nell'introduzione alla sezione 1.2, la qualità dell'immagine ecografica, in termini di risoluzione e contrasto, è strettamente collegata alle proprietà dell'onda trasmessa. Si è visto come il compito del beamformer sia di generare un'onda con lobo principale stretto e con ampiezza dei lobi secondari minore possibile, i quali sfortunatamente sono due obiettivi in contrasto tra loro.

Il beamformer è l'unità incaricata di implementare tutte le tecniche atte alla focalizzazione e all'indirizzamento del fascio di ultrasuoni, generando i ritardi e i pesi da applicare al segnale

sia durante la trasmissione sia durante la ricezione [7]. Una progettazione accurata di questa unità è fondamentale per ottimizzare la forma dell'onda e di conseguenza anche le performance del sistema di imaging, che è essenziale per ottenere immagini di qualità che facilitino la diagnosi al medico. Per queste ragioni può essere considerato come uno dei blocchi più importanti e complessi progettualmente del sistema ecografico [8],[5]. Le sue funzioni si possono riassumere in quattro punti principali:

- gestire le temporizzazioni e i pesi che devono essere applicati in ogni istante ai segnali trasmessi dai trasduttori;
- generare i ritardi da applicare al segnale in ricezione;
- gestire i pesi e la somma degli echi di ritorno opportunamente ritardati;
- eventualmente effettuare ulteriori operazioni di processamento del segnale.

Nell'imaging ad ultrasuoni la tecnica di beamforming standard utilizzata per la ricostruzione dell'immagine è l'algoritmo Delay and Sum (DAS). La formula che restituisce il segnale in uscita dopo l'applicazione dell'algoritmo è:

$$y_{DAS}(t) = \sum_{i=1}^N \hat{s}_i(t - \tau_i) \quad (1.30)$$

Dove:

- $y_{DAS}(t)$ è il segnale in uscita dal beamformer;
- \hat{s}_i è il segnale ricevuto dal sensore i -esimo;
- τ_i è il ritardo applicato al segnale dell'elemento i -esimo;
- N è il numero di elementi attivi della sonda.

La Fig.1.18 mostra schematicamente le operazioni effettuate dall'algoritmo.

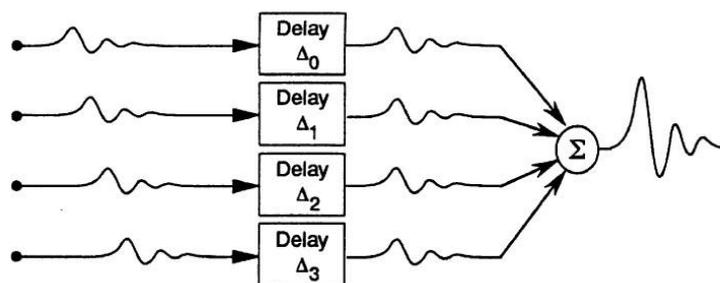


Figura 1.18: Le operazioni effettuate dall'algoritmo Delay and Sum.

Nel caso di un'immagine B-mode, l'algoritmo consente di assegnare un valore di intensità a ciascun pixel che compone l'immagine, ritardando i segnali ricevuti da ciascun elemento della sonda di un tempo ricavabile dalla conoscenza delle coordinate geometriche del pixel considerato.

Il calcolo dei ritardi avviene sfruttando qualche concetto basilare di geometria. Si può assumere che in fase di trasmissione l'onda sia piana, mentre gli echi di ritorno (generati dalle strutture riflettenti presenti nel mezzo scansionato) siano onde sferiche (Fig.1.20) [9],[10].

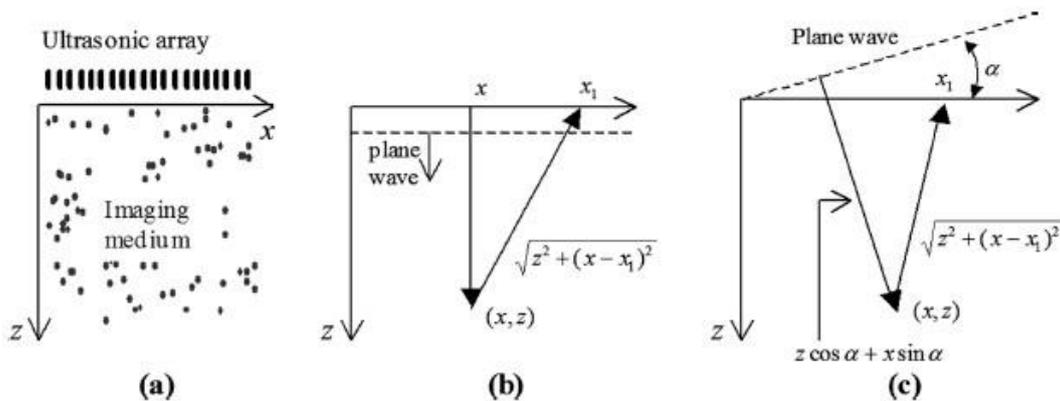


Figura 1.19 L'immagine mostra: A) la convenzione sugli assi adottata; B) il ritardo dell'eco di ritorno all'*i*-esimo elemento nel caso di trasmissione di un'onda piana; C) il calcolo del ritardo nel caso l'onda piana venga trasmessa inclinata di un angolo α .

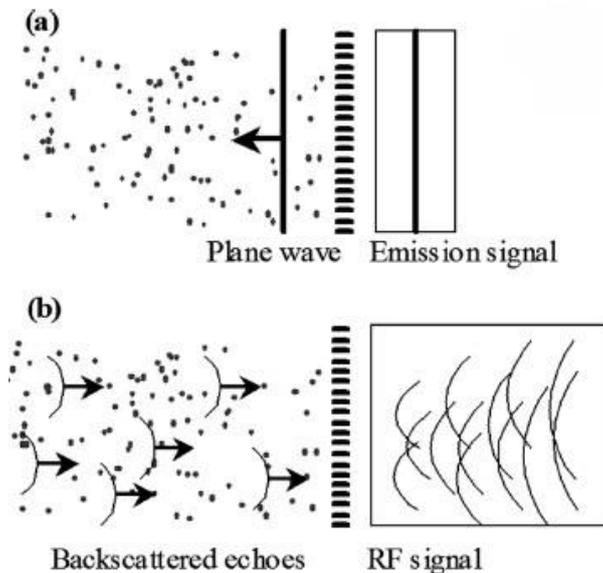


Figura 1.20: L'immagine mostra schematicamente a) la trasmissione di un'onda piana b) il segnale RF di ritorno generato dalle discontinuità nel mezzo e registrato dai trasduttori. Nota: in questo caso si mostra una tecnica differente (plane wave compounding) in cui si utilizzano tutti i trasduttori per generare e ricevere i segnali, mentre solitamente si utilizza una sub-apertura dell'array.

L'onda sferica dell'eco di ritorno giungerà naturalmente per prima al trasduttore posizionato alla stessa ascissa del punto considerato, mentre i ritardi da applicare ai segnali ricevuti dagli altri elementi sono calcolabili con il teorema di Pitagora (Fig.1.19):

$$\tau_i = \frac{\sqrt{z^2 + (x - x_i)^2} - z}{c} \quad (1.31)$$

Dove il punto di scansione ha coordinata (x, z) , x_i è la coordinata del n-esimo trasduttore della sonda e c la velocità di propagazione del suono nel mezzo.

In Fig.1.21 viene mostrato più nel dettaglio il diagramma a blocchi relativo al beamformer di ricezione. I dati in input sono campionati da un convertitore A/D, che indicativamente viene scelto con una frequenza di campionamento compresa tra 20 e 60 MHz, mentre la risoluzione è solitamente tra 8 e 14 bit.

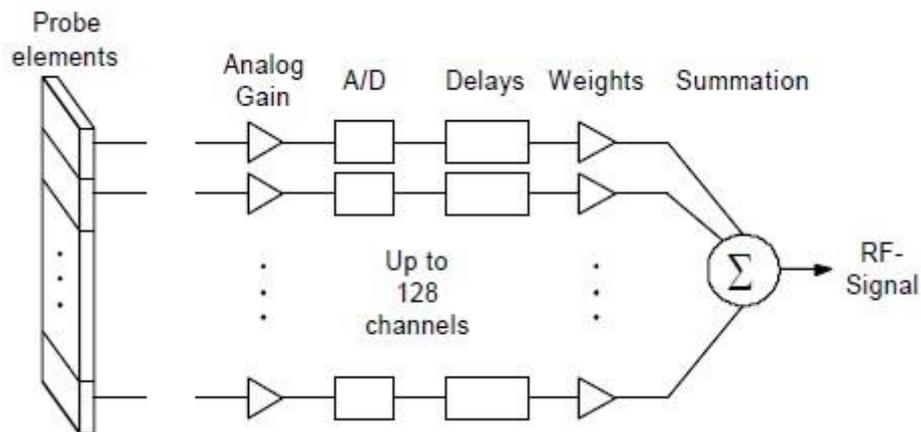


Figura 1.21: Schema a blocchi del beamformer in ricezione.

La fase di trasmissione è sostanzialmente inversa rispetto alle operazioni effettuate in ricezione. I ritardi in questa fase sono necessari per la focalizzazione e la sterzata del fascio e vengono generati da appositi contatori, andando ad eccitare gli elementi della sonda in momenti differenti. Un esempio di ritardi generati per la sterzata del fascio è mostrato in Fig.1.22.

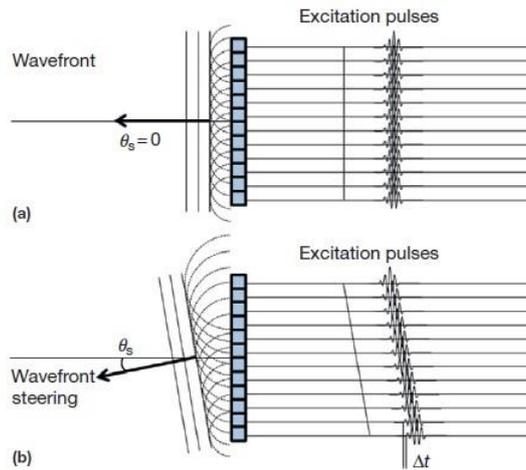


Figura 1.22: Applicazione dei ritardi per la sterzata dell'onda trasmessa.

I ritardi per la focalizzazione del fascio sono ricavati a partire dalle stesse considerazioni geometriche che hanno prodotto l'Eq.1.31, come visibile in Fig.1.23.

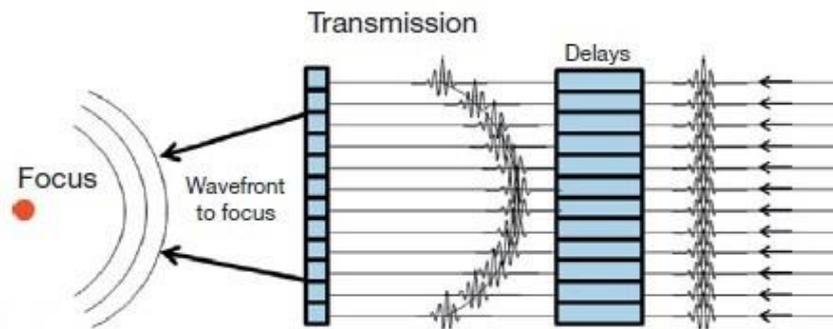


Figura 1.23: Applicazione di ritardi per la focalizzazione dell'onda in trasmissione.

1.3.2.3 Algoritmi di beamforming alternativi

La maggior parte dei sistemi di ultrasonografia adotta l'algoritmo standard Delay and Sum per la formazione dell'immagine B-mode, il cui limite consiste principalmente nella bassa risoluzione dell'immagine e la scarsa reiezione alle interferenze provenienti dalle direzioni non scansionate.

Un algoritmo proposto in [7] è il DMAS (Delay Multiply and Sum). Similmente al DAS, questo algoritmo focalizza gli echi applicando dei ritardi ai segnali ricevuti dagli elementi del trasduttore (o delle antenne nel caso di applicazioni non medicali). Tuttavia, la differenza consiste nel fatto che i segnali vengono poi accoppiati e moltiplicati tra loro, anziché

semplicemente sommati. Per via di quest'ultima operazione, l'algoritmo è non lineare. Il diagramma a blocchi dell'algoritmo è rappresentato in Fig.1.24.

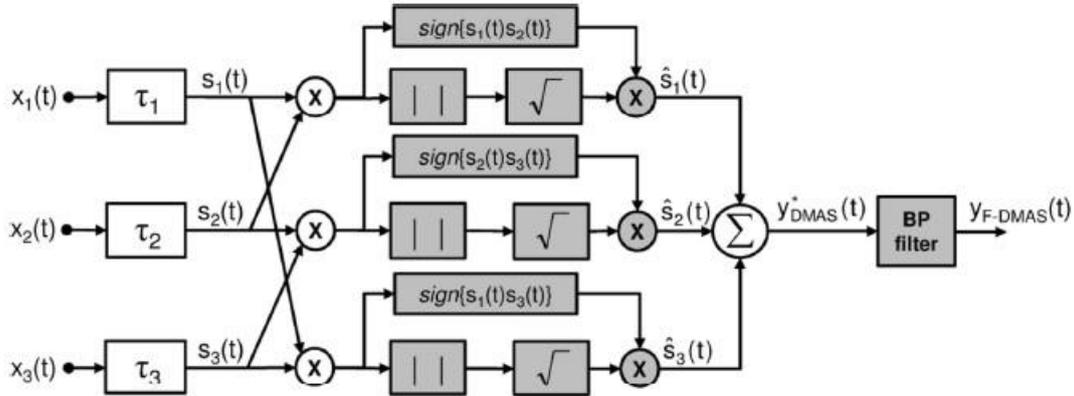


Figura 1.24 Digramma a blocchi dell'algoritmo DMAS.

Considerando N canali di ricezione, il numero di moltiplicazioni effettuate è dato dalle combinazioni di coppie possibili:

$$\binom{N}{2} = \frac{N^2 - N}{2}$$

Il segnale $y_{DMAS}(t)$ in uscita dal beamformer è ottenuto come:

$$y_{DMAS}^*(t) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \text{sign}(s_i(t)s_j(t)) \sqrt{|s_i(t)s_j(t)|} \quad (1.32)$$

Dove $s_i(t)$ è il segnale RF ritardato ricevuto dall' i -esimo elemento. Questa operazione può essere interpretata come la funzione di auto-correlazione degli elementi della sonda: ad ogni istante di tempo viene calcolata la funzione di cross-correlazione spaziale tra tutti i segnali ricevuti dagli elementi attivi. L'uso della radice quadrata del valore assoluto della coppia di segnali è necessario per poter ottenere un segnale in uscita che dimensionalmente sia in [Volt] anziché [Volt²]. La funzione *sign* viene adoperata per conservare il segno dopo l'operazione di radice quadrata.

Da notare come il segnale $y_{DMAS}(t)$ abbia valor medio nullo e lo spettro sia simile al segnale RF ricevuto da ciascun canale. Il segnale $y_{DMAS}^*(t)$, viceversa, possiede uno spettro in cui compaiono sia una componente a frequenza nulla, sia una componente posizionata in prossimità del doppio della frequenza di trasmissione. Perciò occorre filtrare il segnale $y_{DMAS}^*(t)$ con un filtro passa-banda, per poter attenuare la componente continua e le frequenze più elevate,

mantenendo quella a $2f_0$ inalterata. In definitiva, utilizzando l'algoritmo *filtered*-DMAS per il beamforming dei segnali in luogo del DAS, si ottengono miglioramenti nella risoluzione di contrasto, ovvero nell'abilità di rilevare oggetti con differente ecogenicità, al costo di un maggiore costo computazionale.

Un altro algoritmo di beamforming proposto in letteratura consiste nel pesare il segnale ottenuto dal DAS moltiplicandolo per il fattore di coerenza. Il fattore di coerenza fu inizialmente introdotto come un indicatore della qualità della focalizzazione dell'onda ricevuta [11], misurando come l'energia dell'ultrasuono si distribuisce tra il lobo primario e i lobi secondari [12]. Gli errori di focalizzazione dell'onda ricevuta sono dovuti principalmente alle variazioni della velocità del suono attraversando un mezzo non omogeneo, distorcendo l'ampiezza e la fase del segnale acustico [13]. Il fattore di coerenza viene definito come il rapporto tra la somma coerente e la somma incoerente dei segnali RF ricevuti dagli elementi attivi della sonda, dopo che sono stati applicati i ritardi di focalizzazione:

$$CF(t) = \frac{|\sum_{i=1}^N s_i(t)|^2}{N \sum_{i=1}^N s_i^2(t)} \quad (1.33)$$

Il fattore di coerenza è vicino a 1 quando la coerenza del segnale è alta, ossia quando la qualità della focalizzazione dell'onda è buona; viceversa assumerà valori bassi attorno allo 0 se la qualità è scarsa. Una volta calcolato, è possibile ottenere il segnale pesato dal fattore di coerenza:

$$y_{CF}(t) = CF(t)y_{DAS}(t)$$

1.4 IL PROCESSAMENTO DEL SEGNALE ECOGRAFICO

1.4.1 Demodulazione IQ [14]

Con riferimento alla Fig.1.21 si può notare come il segnale in uscita dal beamformer venga denominato "segnale a radiofrequenza" (*RF-signal*). Questo termine, adoperato in special modo nell'ingegneria delle telecomunicazioni, è stato adottato nell'industria che si occupa di apparecchi ad ultrasuoni per riferirsi al segnale non processato, in cui il contenuto in frequenza è intatto.

Il segnale RF ricevuto da un trasduttore ad ultrasuoni è un segnale passa-banda. La larghezza di banda relativa alla frequenza di trasmissione è solitamente attorno al 50-70 %. La larghezza

di banda viene definita come il range di frequenze che hanno contenuto con valore fino a -3 dB rispetto all'ampiezza della frequenza principale.

La Fig.1.25 mostra lo spettro di un segnale acquisito da una sonda ad ultrasuoni, con frequenza di trasmissione 2.5 MHz, larghezza di banda inferiore al 100% e frequenza di campionamento 20 MHz.

Per poter ridurre la quantità di dati ricevuti senza perdere le informazioni acquisite, è necessario l'utilizzo di una tecnica denominata "demodulazione in fase e in quadratura" (IQ).

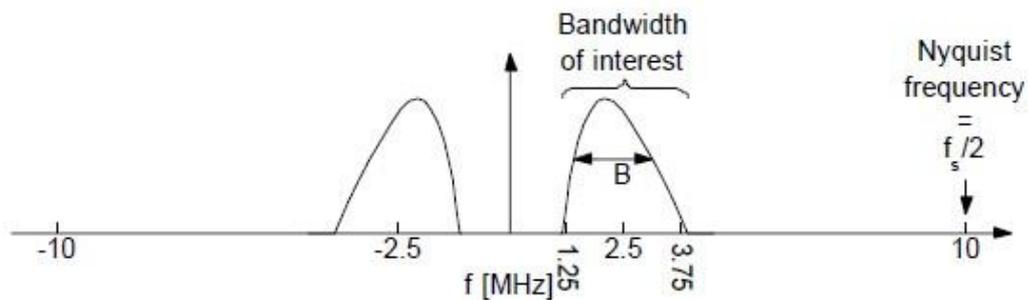


Figura 1.25 Spettro del segnale RF acquisito dal trasduttore.

La demodulazione IQ si compone di tre passi principali (Fig.1.26):

- down-mixing;
- filtraggio passa-basso;
- decimazione.

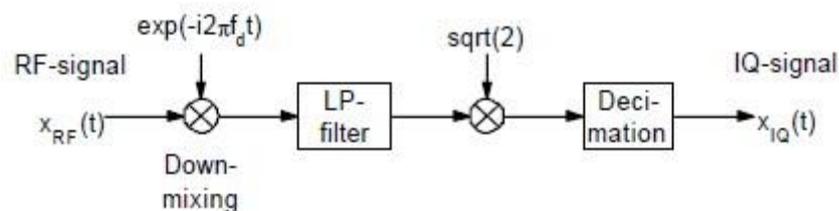


Figura 1.26 Demodulazione IQ del segnale RF.

1.4.1.1 Down-mixing

In questo primo passo, il segnale reale RF viene moltiplicato per un fasore, in modo da traslare lo spettro di una quantità pari alla frequenza di trasmissione:

$$x_{IQ}(t) = x_{RF}(t)e^{-i2\pi f_{demod}t} \quad (1.34)$$

Il segnale ottenuto è naturalmente complesso, perciò lo spettro non è simmetrico rispetto all'asse delle ordinate.

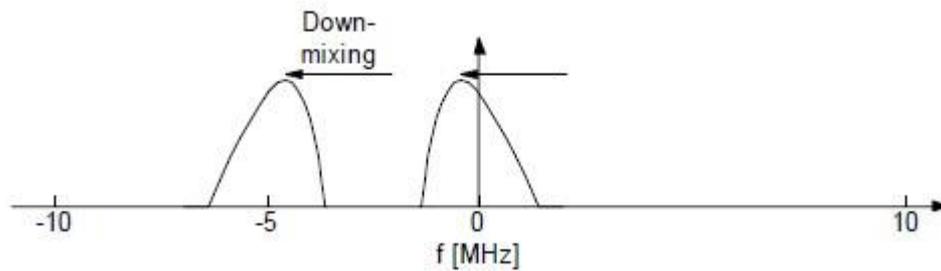


Figura 1.27 Down-mixing.

Considerando la relazione tra la funzione esponenziale complessa e le funzioni seno e coseno (equazione di Eulero):

$$e^{-i\omega t} = \cos(\omega t) - i \cdot \sin(\omega t)$$

Si può pensare l'Eq.1.34 come la combinazione del segnale RF con due sinusoidi che hanno differenza di fase tra loro di 90° (Fig.1.28).

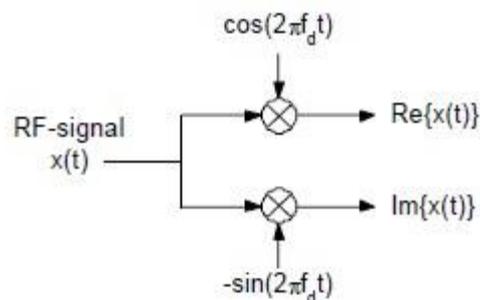


Figura 1.28 Combinazione del segnale RF con le funzioni seno e coseno.

1.4.1.2 Filtraggio passa-basso

Dopo l'operazione di down-mixing il segnale viene filtrato con un filtro passa-basso, in modo da rimuovere le frequenze negative dello spettro e il rumore presente al di fuori della banda di interesse (Fig.1.29). Nel caso preso in esame, la frequenza di taglio del filtro potrebbe essere scelta ragionevolmente con valore attorno a 1.5 MHz.

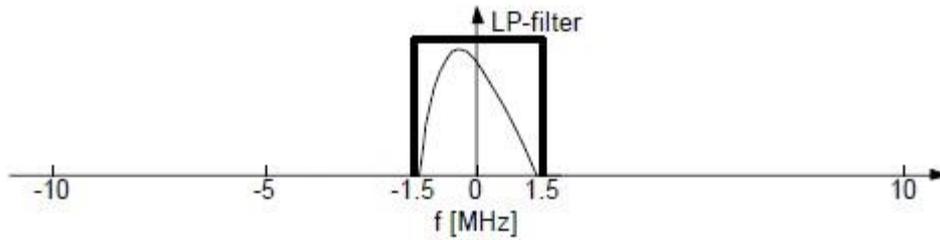


Figura 1.29 Filtraggio passa-basso.

Il filtro rimuove approssimativamente metà dell'energia del segnale, perciò per preservarne il contenuto energetico iniziale è necessario moltiplicare il segnale complesso per la radice di due.

1.4.1.3 Decimazione

Una volta che il segnale è stato filtrato, è possibile ridurre la frequenza di campionamento in modo da potere ridurre il quantitativo di dati memorizzato dal dispositivo. Il teorema di Nyquist afferma che la frequenza di campionamento deve essere pari ad almeno il doppio della frequenza di taglio del filtro. Alla luce di ciò, considerando che la nuova frequenza di campionamento deve essere un sottomultiplo intero di quella originale, una possibile scelta è 3.33 MHz, ossia un sesto del valore di partenza. Nella pratica viene mantenuto solamente un campione ogni sei, scartando i rimanenti, tuttavia preservando il contenuto originario di informazione contenuto nel segnale passa-banda (ciò significa che è possibile ricostruire il segnale RF iniziale a partire dal segnale demodulato IQ).

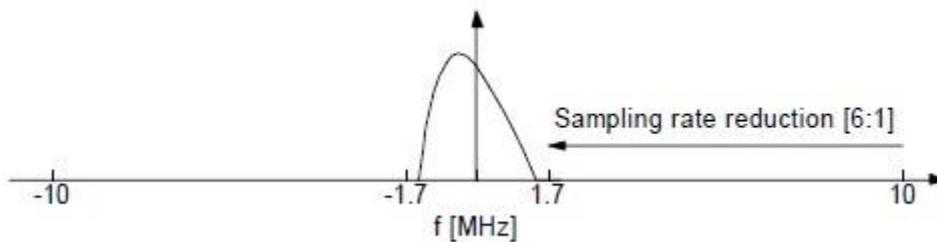


Figura 1.30 Decimazione.

Per ottenere l'involuppo del segnale si procede con il calcolo del modulo del segnale complesso demodulato IQ:

$$x_{envelope}(t) = \sqrt{I^2 + Q^2}$$

1.4.2 Dal segnale alla formazione dell'immagine [3], [15]

La rilevazione dell'involuppo del segnale rappresenta la prima delle operazioni del back-end. Da questo punto, l'informazione acquisita sotto forma di linee di scansione deve essere convertita in modo che alla fine della catena di processamento si possa visualizzare un'immagine sul monitor.

La fase di pre-processing avviene prima che i dati degli echi di ritorno vengano memorizzati, ed include alcune operazioni quali l'*edge-enhancement* e l'interpolazione dei pixel (ossia l'assegnazione di un valore di luminosità a pixel che sono al di fuori della linea di scansione) [15]. Il post-processing include invece le operazioni effettuate dopo la memorizzazione dei dati relativi agli echi di ritorno. Determina il modo in cui i dati appariranno sul display, assegnando un valore di luminosità a ciascun numero presente in memoria. Esistono diverse curve predefinite che consentono di modificare a quale valore di luminosità deve essere associato ciascun livello di grigio, permettendo di mettere in risalto anche delle piccole variazioni nelle intensità di due echi differenti. L'ultimo passo consiste nella conversione D/A dei dati presenti in memoria, in modo da ottenere in uscita dei valori analogici di tensione che potranno essere passati in input al display, determinando così la luminosità di ciascun pixel e visualizzando l'immagine.

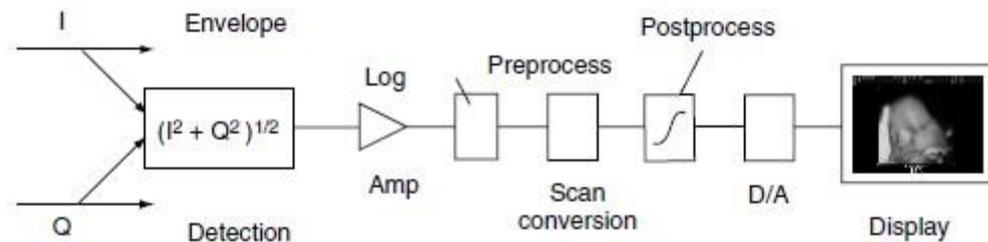


Figura 1.31: Schema a blocchi delle operazioni nel back-end.

CAPITOLO 2

DEEP LEARNING: DAI FONDAMENTI ALLE APPLICAZIONI IN MEDICINA

Negli ultimi anni, l'intelligenza artificiale è stata oggetto di numerose attenzioni e aspettative da parte dei media. Machine learning, deep learning e intelligenza artificiale sono parole che appaiono in un numero notevole e sempre crescente di pubblicazioni, anche in ambiti non propriamente affini a quello tecnologico.

L'intelligenza artificiale nasce attorno alla metà degli anni '50, grazie ad un gruppo di ricercatori informatici che tentava di rispondere al quesito relativo alla possibilità che i computer potessero essere in grado di “pensare”, una domanda le cui implicazioni vengono dibattute ancora oggi. Una possibile definizione di intelligenza artificiale potrebbe essere la seguente: “*il tentativo di automatizzare dei compiti e delle azioni normalmente svolti dagli esseri umani*” [16]. Secondo questa definizione, l'intelligenza artificiale comprenderebbe non solo il machine learning e il deep learning (Fig. 2.1), ma anche molti altri approcci che non utilizzano dei veri e propri processi di apprendimento. Infatti, inizialmente si credeva che un programma potesse simulare l'intelligenza umana mediante l'implementazione di un numero sufficientemente grande di regole, un metodo noto come intelligenza artificiale simbolica. L'adozione di questo paradigma fu dominante fino alla fine degli anni '80, ma fu poi superato

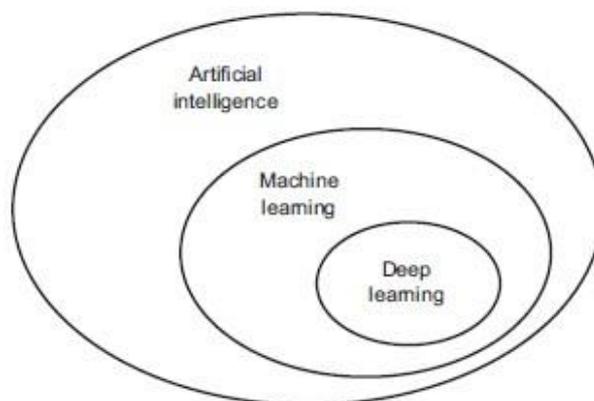


Figura 2.1: Intelligenza artificiale, machine learning e deep learning.

in quanto non consentiva di risolvere problemi più complessi in cui non vi fosse una soluzione al problema ben definita, come nel caso della classificazione di immagini o la traduzione linguistica. Per poter sopperire a queste mancanze, il machine learning nasce dall'ipotesi che un computer possa imparare da solo a risolvere uno specifico task, anziché essere programmato a priori tramite delle regole.

Infatti, in un problema classico di programmazione, la risposta viene data dal sistema in base ai dati e alle regole fornite in input, al contrario, nel machine learning, vengono restituite in uscita le regole che permettono la risoluzione del problema, fornendo in input i dati e le risposte previste (Fig.2.2). Perciò, più correttamente si può affermare che un sistema di machine learning non viene semplicemente programmato, ma allenato. Il modello modifica i propri parametri in base alla differenza tra gli output predetti dal sistema e gli output reali forniti a priori, utilizzando un criterio specifico per quantificare l'errore commesso dalla rete. Questo errore viene utilizzato come segnale di feedback per modificare il modo in cui l'algoritmo lavora tentando di minimizzare l'errore commesso.

Il deep learning è invece una particolare categoria di machine learning, che si basa sulla costruzione di reti neurali artificiali che mimano le funzioni e la struttura dell'encefalo umano. In pratica, il deep learning adopera un grande numero di strati nascosti (da qui l'aggettivo "profondo") che utilizzano operazioni non lineari per estrarre informazioni dai dati in ingresso, trasformando i dati su diversi livelli di astrazione.

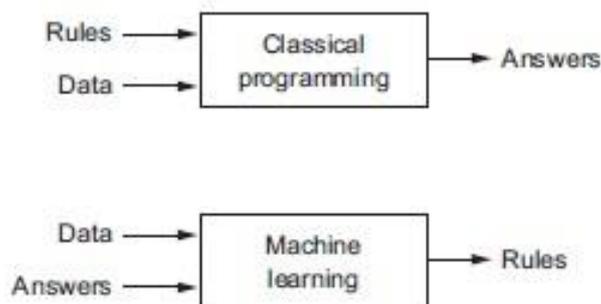


Figura 2.2: Il nuovo paradigma di programmazione del machine learning.

2.1 IL DEEP LEARNING

Un esempio classico di rete di deep learning è quella utilizzata per il riconoscimento di una cifra numerica all'interno di un'immagine. La rete trasforma l'immagine in diversi gradi di rappresentazione che si distaccano sempre più dall'immagine originale, mantenendo solamente l'informazione necessaria da restituire in uscita (Fig.2.3). L'informazione originaria è come se venisse passata all'interno di diversi filtri posti in serie e ne uscisse “purificata”, in modo da poter ottenere in uscita un valore che sia la risposta al problema (in questo caso il numero della cifra presente nell'immagine).

Il modo in cui il dato viene trasformato da un layer dipende dai valori dei pesi di quello strato, ossia da come è parametrizzato. Si può quindi interpretare il processo di apprendimento come la modifica dei pesi del layer, in modo tale che al termine del processo gli output restituiti dalla rete siano il più possibile simili ai dati target.

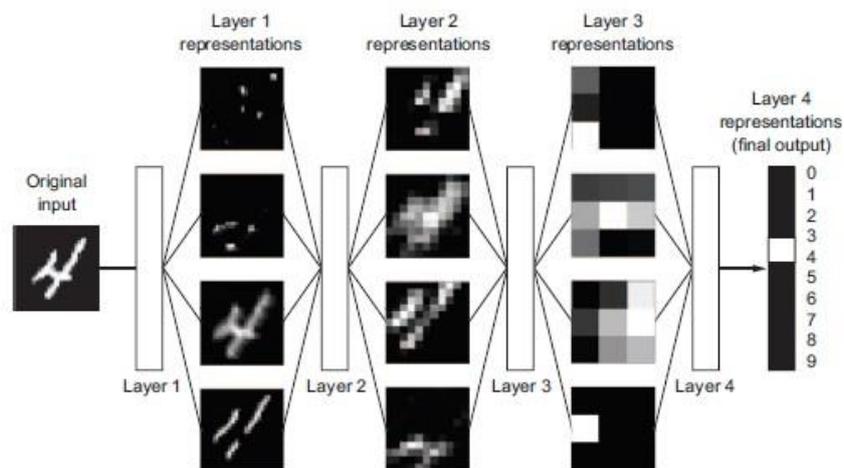


Figura 2.3: Rappresentazione del modo in cui i layer del modello rappresentano l'informazione dell'immagine in input.

Per poter modificare i pesi della rete è necessario poter monitorare quanto l'output predetto sia differente da quello reale. Questo è il compito della *loss function*, anche chiamata funzione obiettivo. La funzione ha il compito di implementare un criterio che misuri la distanza tra l'output predetto dalla rete e quello reale. Questo indice viene utilizzato come segnale di feedback per aggiustare i valori dei pesi dei diversi strati, in modo che il suo valore venga minimizzato. Le modifiche dei pesi avvengono tramite la funzione di ottimizzazione, che implementa l'algoritmo di *backpropagation* (Fig.2.4). L'algoritmo, per essere efficace, deve consentire di raggiungere un ottimo globale evitando di rimanere bloccato in minimi locali. Il

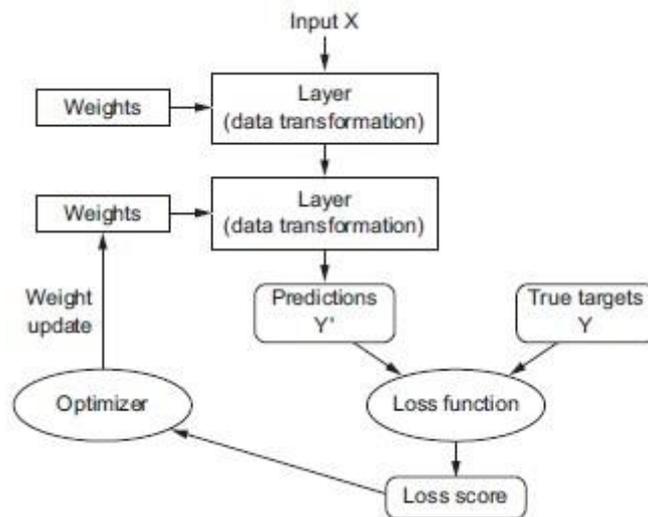


Figura 2.4: Il valore dato dalla loss function viene utilizzato come segnale di feedback per l'aggiustamento dei pesi.

learning rate è un parametro che consente di modificare la velocità con la quale l'algoritmo converge verso una soluzione ottima. Se il valore è troppo piccolo, l'allenamento potrebbe richiedere molte epoche prima di convergere, con l'ulteriore rischio che si fermi su un minimo locale delle soluzioni della loss function. Viceversa, se il valore è eccessivo, l'algoritmo potrebbe addirittura non convergere [17]. La Fig.2.5 mostra come l'algoritmo si muove sulle soluzioni della funzione obiettivo, per valori grandi e piccoli del learning rate.

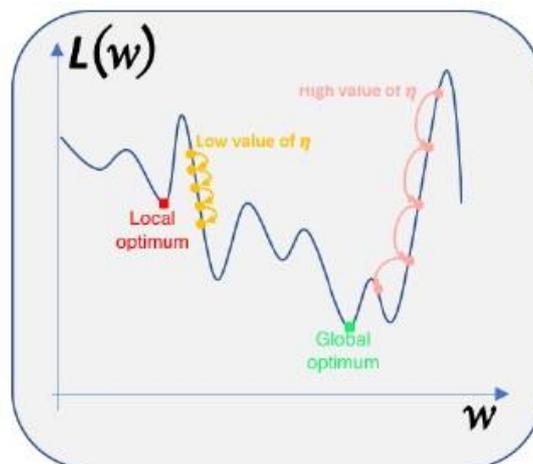


Figura 2.5: La figura mostra il modo in cui la variazione del learning rate modifica il comportamento dell'algoritmo. Valori troppo piccoli potrebbero causare lunghi tempi di convergenza e il rischio di fermarsi in un ottimo locale. Viceversa, valori troppo elevati potrebbero causare la non convergenza dell'algoritmo.

I pesi vengono inizialmente assegnati casualmente, perciò all'inizio il valore restituito dalla loss function sarà molto alto. La funzione di loss viene minimizzata a mano a mano che vengono

processati i campioni del dataset (solitamente decine di iterazioni su migliaia di campioni). Una rete con la funzione di loss minimizzata restituisce degli output il più possibile simili ai target.

2.1.1 Panoramica sui metodi di deep learning [18]

I metodi di machine learning vengono generalmente suddivisi in algoritmi di apprendimento supervisionati e non supervisionati.

Negli algoritmi supervisionati si utilizza un dataset composto da dati di input x associati ai corrispondenti target y . Questa y può essere di tipologie differenti, a seconda del tipo di problema che si intende risolvere. Ad esempio, in un problema di classificazione potrebbe trattarsi di un valore scalare che rappresenta una particolare classe di appartenenza (si pensi ad un caso in cui si vuole classificare una lesione come maligna o benigna mediante una rappresentazione binaria), oppure può essere un vettore di variabili continue nel caso di un problema di regressione. Nell'allenamento supervisionato il fine è quello di trovare i parametri del modello che minimizzano la differenza tra i dati predetti dalla rete e i dati target, come già visto nel paragrafo precedente. Utilizzando questi parametri è poi possibile effettuare predizioni su dati dei quali la classe di appartenenza è sconosciuta.

Nel caso dell'allenamento non supervisionato non sono noti i valori y dei dati in ingresso, perciò la rete opera in maniera tale da riconoscere dei pattern nascosti nei dati che consentano di raggrupparli. Un esempio di algoritmi tradizionali di apprendimento non supervisionato sono i metodi di clustering e l'analisi delle componenti principali.

La Fig.2.6 offre una panoramica sui modelli di deep learning maggiormente utilizzati in medicina.

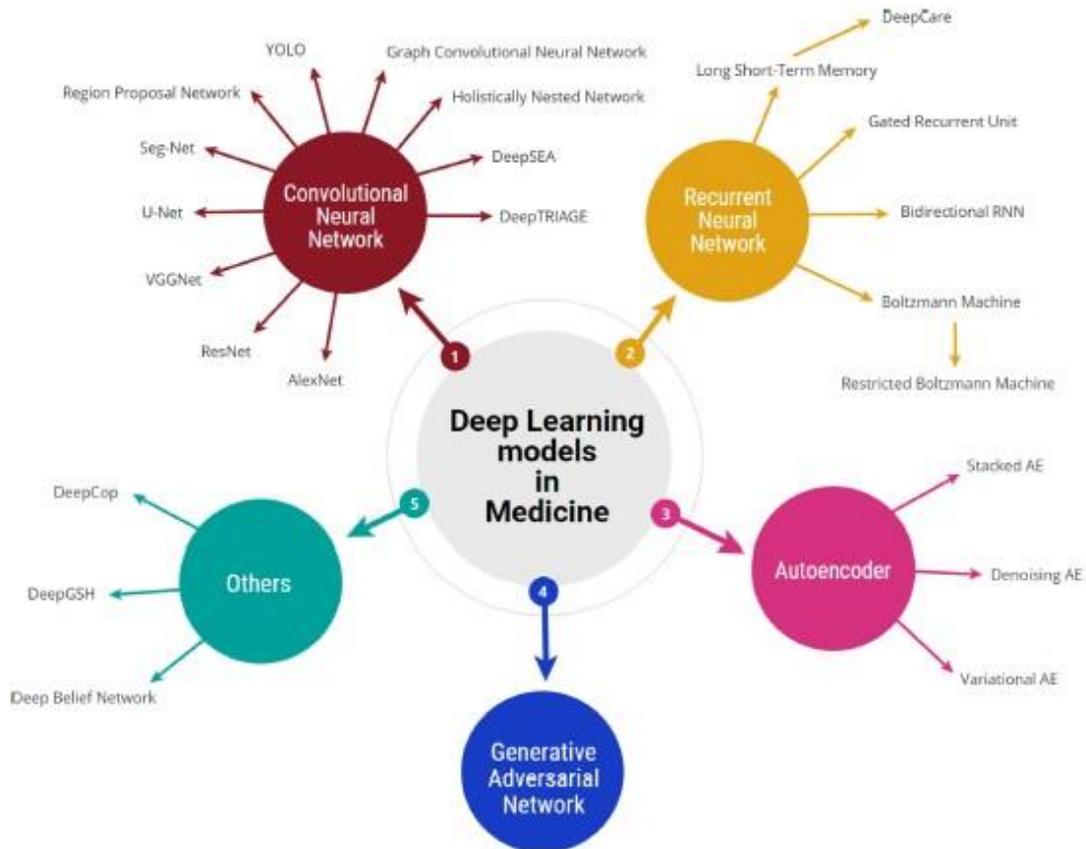


Figura 2.6: Schema riassuntivo dei modelli di deep learning principalmente adoperati in medicina.

2.1.1.1 Reti neurali artificiali (ANN)

Le reti neurali sono una tipologia di modello computazionale alla base della maggior parte dei metodi di deep learning.

Una rete neurale è composta da un certo numero di “neuroni” artificiali, ciascuno caratterizzato da un set di parametri e da una funzione di attivazione. Ogni neurone, similmente a quanto avviene nel mondo biologico, riceve in input le uscite degli altri neuroni per mezzo di connessioni pesate dai valori del vettore peso w , che viene modificato durante il corso della fase di apprendimento:

$$u = \sum_i w_i x_i$$

La somma pesata viene poi passata attraverso una funzione di attivazione $f(u)$, ottenendo l'uscita del neurone che può essere così inviata ad altri neuroni oppure direttamente all'uscita della rete.

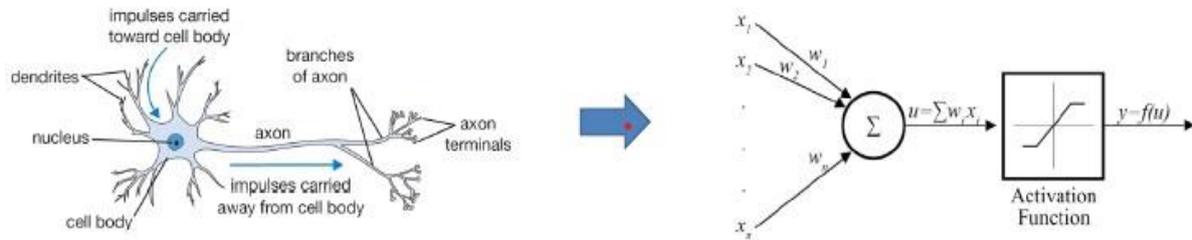


Figura 2.7: La figura mostra il parallelismo tra un neurone "biologico" e il modello del neurone artificiale di McCulloch e Pitts.

Si parla di *multi-layer perceptron* (MLP) quando la rete è costituita da: uno strato di input che riceve le informazioni dall'esterno; uno o più strati nascosti che elaborano le informazioni; un layer di output che restituisce all'utente il risultato del processo di allenamento. Quando tra l'input e l'output vengono interposti un certo numero di strati nascosti (tipicamente sei o più) si parla più propriamente di *deep neural network* (DNN). La differenza viene mostrata in Fig.2.8.

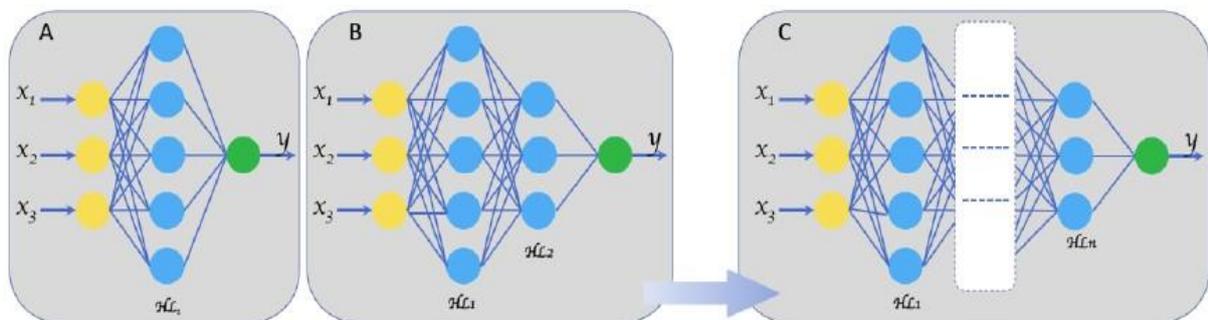


Figura 2.8: Rappresentazione di una rete neurale con: A) un layer nascosto; B) due layer nascosti; C) un multiplice numero di layer nascosti (DNN).

La funzione di attivazione dipende dall'architettura adoperata e dal numero di layer. Nel caso di architetture con pochi layer nascosti sono utilizzate di norma la funzione sigmoide o la tangente iperbolica, mentre nel caso di DNN una delle funzioni di attivazione più utilizzate è la *rectified linear unit* (ReLU), ossia:

$$f(z) = \max(z, 0)$$

La funzione di attivazione del layer di uscita deve essere scelta con cautela, a seconda del tipo di problema che si sta cercando di risolvere. Nel caso di una classificazione binaria (per esempio

per classificare un'immagine medica come benigna o maligna) si potrebbe usare una funzione sigmoide che mappa i valori di uscita della rete tra zero e uno in modo continuo (naturalmente in seguito si deve utilizzare un criterio per poter assegnare solamente uno dei due valori); nel caso di un problema in cui l'input può essere assegnato ad un certo numero di classi (come il riconoscimento di una cifra numerica in un'immagine), una soluzione potrebbe essere l'utilizzo della funzione *softmax*.

2.1.1.2 Reti neurali convoluzionali (CNN)

Le reti neurali convoluzionali sono risultate essere tra le più efficaci nella risoluzione di problemi legati all'analisi di immagini. Il lavoro su questo tipo di reti comincia alla fine degli anni '80, ispirandosi al concetto di rappresentazione gerarchica dei neuroni in biologia, scoperto da Hubel e Wiesel nel 1962. I due neurofisiologi avevano osservato che nella corteccia visiva del gatto erano presenti neuroni che rispondevano individualmente a porzioni limitate del campo visivo. Questa regione del campo visivo che influenza l'eccitazione del singolo neurone è denominata come il suo campo recettivo. Identificarono due tipi di cellule visive: le cellule semplici, che vengono eccitate quando nel campo recettivo vengono riconosciute delle figure orientate in una direzione particolare, e le cellule complesse, che hanno un campo recettivo più esteso. Combinando queste due tipologie di cellule, i ricercatori formularono un modello per la risoluzione di task di *pattern recognition* [19].

L'importanza del concetto di campo recettivo formulato da Hubel e Wiesel venne notata da Kunihiko Fukushima, che nel 1980 propose una rete neurale per il riconoscimento di pattern all'interno di immagini, la quale si basava principalmente su due tipi di layer: *convolutional layers* e *downsampling layers* [20]. Le unità dei *convolutional layers* hanno un campo recettivo che è associato ad una porzione del layer precedente. I *downsampling layers*, nell'idea di Fukushima, dovevano avere il ruolo di generalizzare il riconoscimento di pattern complessi, in modo che eventuali rotazioni, traslazioni o piccoli cambiamenti fossero ininfluenti sul processo di riconoscimento. Il "neocognitrone", come venne battezzata dal suo autore, può essere considerato il primo esempio di CNN (Fig.2.9).

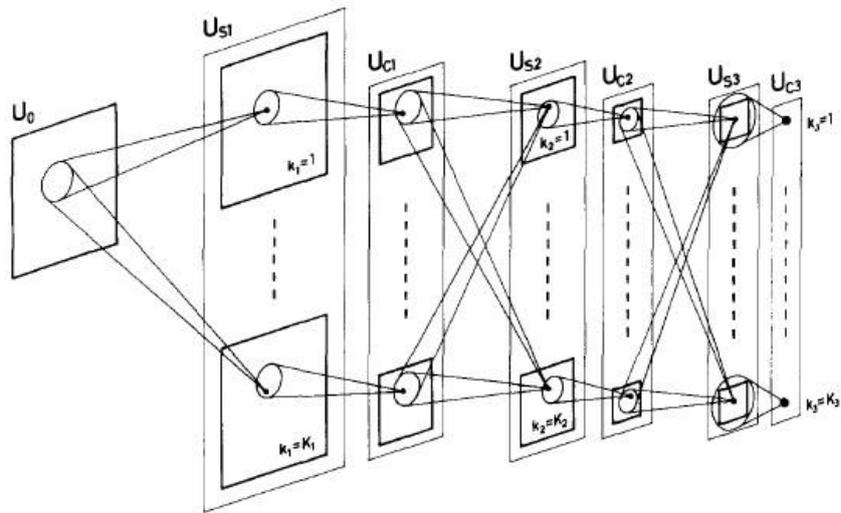


Figura 2.9: L'architettura del neocognitrone. Si noti come le unità di un layer convoluzionale sono connesse al campo recettivo del layer precedente.

In seguito, il lavoro sulle CNN continuò, trovando applicazione nell'analisi di immagini mediche già nel 1995. Nel 1998 *LeNet* viene utilizzata con successo per il riconoscimento automatico di immagini rappresentanti cifre scritte a mano. Il grande salto di qualità ci fu però nel 2012, con la vittoria a grande margine di *AlexNet* nella competizione *ImageNet*, che riversò l'interesse della comunità accademica su questa tipologia di reti. Da allora si sono sviluppate architetture sempre più "profonde", divenendo le reti di elezione per la risoluzione di problemi nell'ambito della computer vision [18].

La differenza principale tra CNN e MLPs consiste nel fatto che le prime sono in grado di riconoscere dei pattern locali presenti nell'immagine, ossia che coinvolgono solo un sottoinsieme di pixel, rendendo la rete in grado di riconoscere quel particolare motivo anche se si presenta in punti diversi dell'immagine. Viceversa, i MLPs riconoscono dei pattern globali, ossia che coinvolgono tutti i pixel dell'input. Le reti convoluzionali inoltre imparano a riconoscere i pattern suddividendoli secondo una gerarchia spaziale: i primi layer convoluzionali sono sensibili al riconoscimento di piccoli particolari di un pattern visivo, come ad esempio i contorni di un dettaglio, i layer successivi invece riconoscono sequenze più grandi che includono i dettagli riconosciuti dagli strati precedenti [16]. La rete infine sarà quindi in grado di associare all'input la classificazione corrispondente, quale per esempio la classe "gatto" ad un'immagine (Fig.2.10).

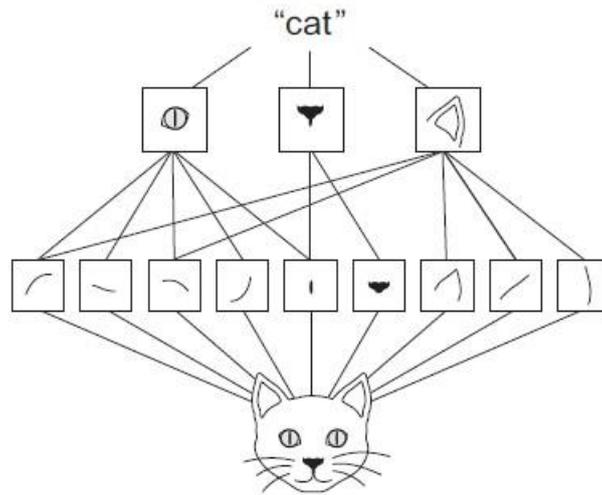


Figura 2.10: La figura mostra come una CNN organizza l'informazione gerarchicamente attraverso una serie di layer convoluzionali. I dettagli più minuti, come i contorni, vengono combinati per formare occhi, naso e orecchio. A loro volta, questi saranno combinati per poter associare all'input il concetto più alto livello di "gatto".

L'architettura di una CNN moderna è strutturata attraverso una serie di blocchi (Fig.2.11). Il primo tipo di operazione, che dà il nome alla rete, è naturalmente la convoluzione (Fig.2.12). Questa consiste nello scorrimento di un filtro (o *kernel*) al tensore di input, che matematicamente esegue l'operazione di convoluzione discreta. Ciascun filtro applicato va supposto come un campo recettivo dell'unità in uscita, per utilizzare l'analogia biologica di Hubel e Wiesel, ossia risponde alla presenza di un particolare pattern presente nell'input. I risultati delle somme pesate sono poi trasformati da una funzione di attivazione non lineare, come ad esempio ReLU.

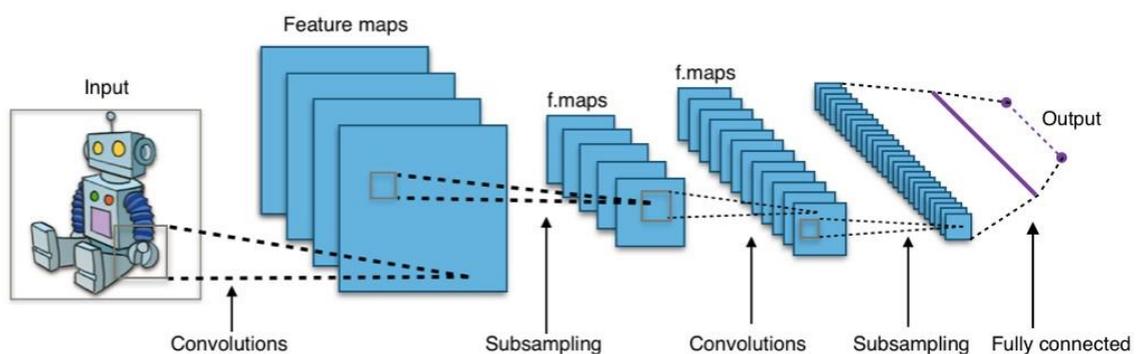


Figura 2.11: Architettura tipica di una CNN.

Le operazioni di convoluzione sono generalmente seguite da un layer di tipo *downsampling* o *pooling*, il quale ha un meccanismo di funzionamento piuttosto simile alla convoluzione (Fig.2.13). Anche in questo caso si utilizza un filtro che scorre sul tensore di input, tuttavia invece di calcolare la convoluzione discreta viene utilizzato un criterio, quale la media, la norma

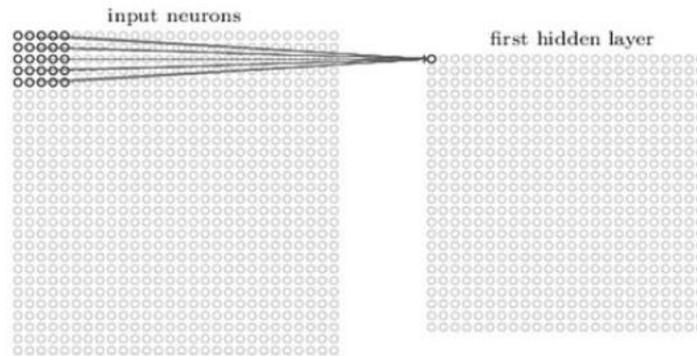


Figura 2.12: L'operazione di convoluzione.

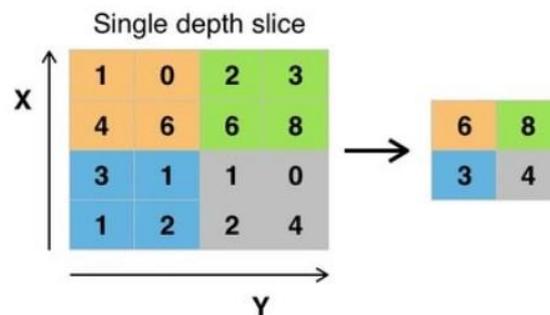


Figura 2.13 L'operazione di max-pooling.

L2 o il valore massimo, da applicare alla regione coperta dal filtro. In questo modo viene diminuita la dimensione del tensore di input.

2.1.1.3 Reti neurali ricorrenti (RNN) [16], [18]

Le reti *feedforward* viste precedentemente si caratterizzano per il fatto di non possedere memoria. Ogni input viene processato indipendentemente dalla rete senza che venga mantenuto uno stato. In questa tipologia di reti, per poter processare una sequenza o una serie temporale di dati, è necessario convertirla in un unico vettore da mostrare in input. Viceversa, una *recurrent neural network* (RNN) processa una sequenza di dati un elemento alla volta, mantenendo nell'unità nascosta un "vettore stato" che contiene l'informazione relativa agli ingressi precedenti. La RNN viene comunque resettata quando si procede a processare una nuova sequenza indipendente dalla precedente. Dal punto di vista pratico, una RNN possiede un loop interno che consente di memorizzare l'uscita al tempo *i*-esimo.

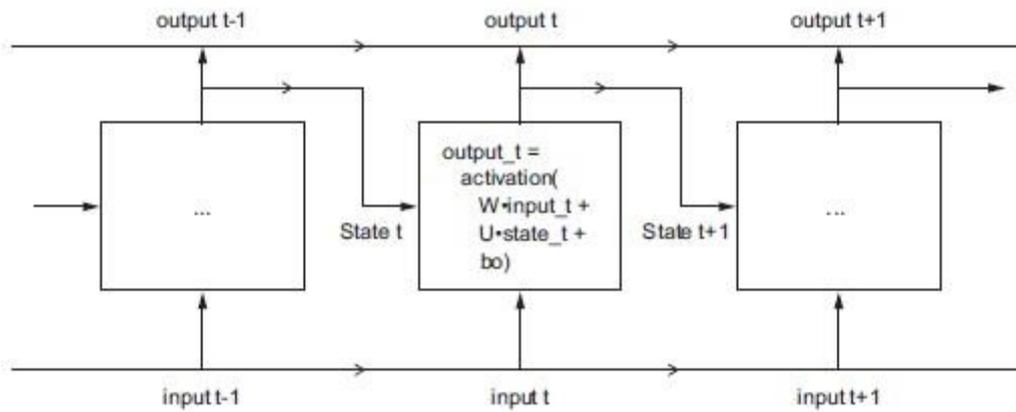


Figura 2.14: Una RNN rappresentata in funzione del tempo.

2.1.1.4 U-Net [21]

L'U-Net venne sviluppata appositamente per la risoluzione di problemi relativi al processamento di immagini biomediche, in particolar modo per ottenere delle maschere di segmentazione. La Fig.2.15 mostra la rappresentazione dell'architettura della rete, il cui nome si deve alla particolare forma ad "U" con cui sono disposti i layer.

L'architettura consiste in uno stadio di contrazione, che possiede essenzialmente gli stessi layer di una tipica CNN, e uno stadio di espansione.

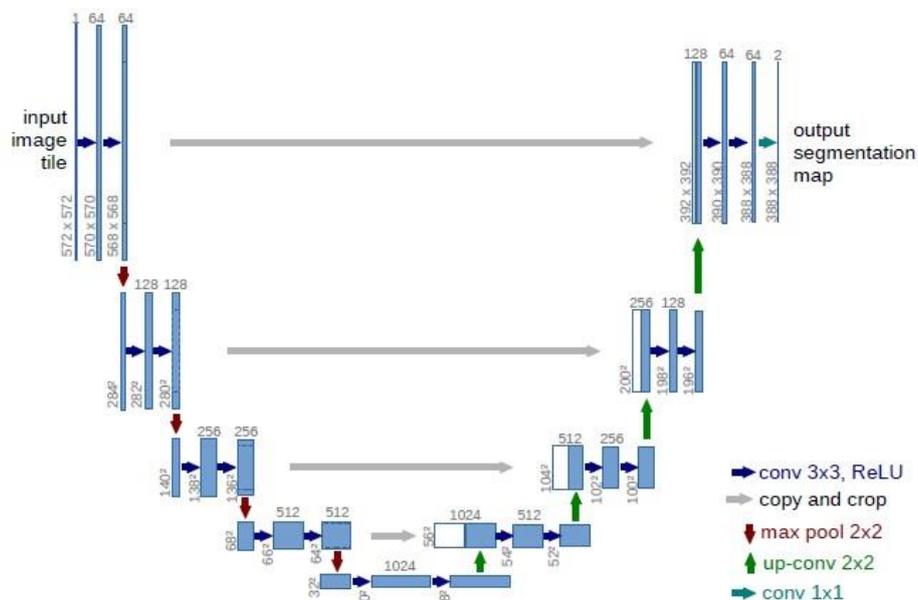


Figura 2.15: Rappresentazione dell'architettura di un U-Net

Il primo stadio consiste essenzialmente nell'applicazione ripetuta di convoluzioni 3x3, seguite da una funzione ReLU e da un layer *max-pooling* 2x2, che consente di dimezzare le dimensioni di input. Al termine di ogni blocco il tensore di input risulterà dimezzato lungo l'altezza e la larghezza, mentre la profondità (*feature channels*) è raddoppiata. Lo stadio di espansione è invece costituito da una serie di layer *upsampling* seguiti da convoluzioni 2x2 (operazione complessivamente indicata in Fig.2.15 come *up-conv*), che dimezza il numero di *feature channels* del tensore. A questa operazione seguono la concatenazione con la *feature map* corrispondente appartenente allo stadio di contrazione, due convoluzioni 3x3 e una attivazione ReLU.

2.1.1.5 Rete generativa avversaria (GAN)

Le reti generative avversarie sono modelli, introdotti nel 2014 da Goodfellow [22], che consentono di generare delle immagini che siano statisticamente indistinguibili dalle immagini reali. Gli strati nascosti di una GAN sono composti essenzialmente da due parti: un generatore, che sfrutta la distribuzione di probabilità dei dati in ingresso per creare una nuova immagine (falsa); un discriminatore, che ha il compito di riconoscere le immagini false e distinguerle da quelle reali (Fig.2.15) [16], [23]. Il generatore ha il compito di ingannare il discriminatore, generando immagini via via sempre più simili ed indistinguibili da quelle reali, al punto tale che il discriminatore non è in grado di distinguere quella simulata da quella vera. Sia il generatore sia il discriminatore vengono allenati simultaneamente, con i pesi del generatore che vengono aggiornati in modo che le immagini create siano riconosciute come reali dal discriminatore, mentre i pesi del discriminatore vengono ottimizzati per discriminare

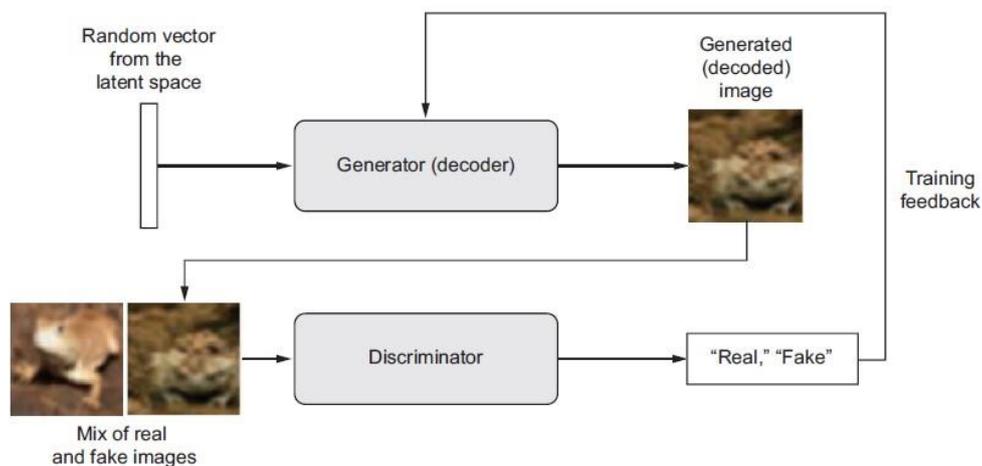


Figura 2.15: Schema rappresentante una GAN. Il generatore trasforma un punto casuale dello spazio latente in un'immagine, mentre il discriminatore ha il compito di distinguere le immagini reali da quelle fasulle.

accuratamente le immagini false da quelle reali [24]. Le GAN sono oggetto di ricerca per via delle ampie applicazioni che hanno nelle immagini mediche [17]. In particolare, possono essere utilizzate per l'aumento della qualità delle immagini, come nel caso della riduzione di artefatti in immagini CT.

2.2 APPLICAZIONI DEL DEEP LEARNING IN MEDICINA

I dati biomedici presenti in medicina sono caratterizzati dalla loro eterogeneità e notevole complessità, cresciuta negli anni per via dei miglioramenti scientifici e tecnologici che hanno permesso di avere a disposizione un grande numero di tipologie di analisi mediche, parametri biologici e tecniche di imaging. Il machine learning si inserisce in questo contesto come opportunità e strumento per il raggiungimento di svariati obiettivi: il supporto di medici e biologi nell'analisi e nello sfruttamento di big data biomedici; la riduzione del rischio di errori medici nella diagnosi; il miglioramento di protocolli di prognosi data una particolare diagnosi [17].

Il deep learning, in particolare, grazie all'architettura che si sviluppa su più layer, consente di ottenere risultati ragguardevoli in termini di accuratezza e qualità, in quanto è in grado di ottenere un alto livello di astrazione lavorando con grandi quantità di campioni [23]. La Fig.2.16 mostra il crescente interesse nell'uso del deep learning in ambito biomedico, espresso in termini di numero di pubblicazioni.

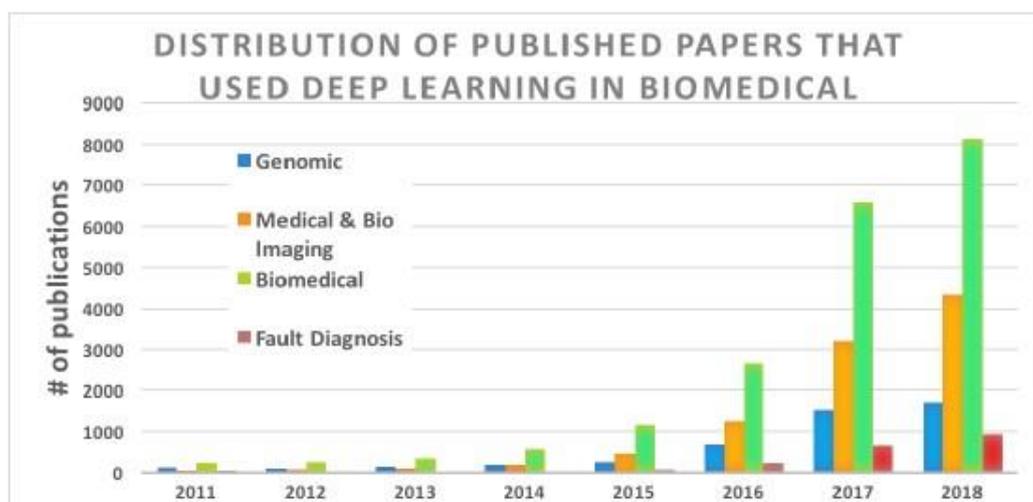


Figura 2.16: Distribuzione negli anni dei paper che trattano dell'uso del deep learning in campo biomedico.

Una possibile suddivisione degli ambiti biomedici in cui trova applicazione il deep learning è:

- genomica e proteomica: lo studio del sequenziamento del DNA/RNA e delle interazioni tra le proteine;
- biosegnali: segnali termici, meccanici ed elettrici prodotti dal corpo umano (EEG, EMG, ECG ecc.);
- *bioimaging*: lo studio delle cellule biologiche e dei tessuti mediante l'analisi delle immagini istopatologiche e immunoistochimiche;
- *medical imaging*: lo studio degli organi umani mediante raggi-X, MRI, CT, US ecc.;
- *public and medical health management* (PmHM): lo studio e l'analisi di big data medici al fine di sviluppare e migliorare le decisioni relative ai servizi e alle prestazioni sanitarie.

Di seguito si esporranno alcuni cenni su come il deep learning abbia contribuito alla progressione delle conoscenze di alcuni di questi campi, con particolare attenzione all'utilizzo nelle immagini mediche.

2.2.1 Il deep learning nella genomica [17]

Questo campo di ricerca racchiude tutti gli studi che spaziano dal sequenziamento del genoma all'espressione genica, dalla predizione della struttura delle proteine alle interazioni tra proteine e farmaci. Si tratta di un terreno molto fertile per l'utilizzo delle reti neurali artificiali, il cui uso sta crescendo rapidamente. Il fine di questi studi è quello di comprendere i processi biologici ad un livello molecolare, in modo tale da poter predire e prevenire malattie attraverso cure personalizzate per il paziente.

Lo studio del genoma attraverso tecniche di intelligenza artificiale è volto alla risoluzione di problemi che concernono: le interazioni tra DNA e proteine; la predizione della regolazione dell'espressione genica; il sequenziamento del genoma.

Le interazioni DNA-proteina ricoprono un ruolo fondamentale nei processi cellulari di trascrizione, traduzione, riparazione e replicazione. L'interazione chimica tra una proteina e un filamento di DNA o RNA è chiamata *protein binding*. In questo caso la predizione dei siti di legame (*protein binding prediction* o PBP) può essere ottenuta mediante l'uso di CNN, utilizzando in input la matrice posizione-frequenza ed ottenendo in output un valore (detto

binding score) che indica l'affinità di legame tra la proteina e un ligando. La predizione del sito di legame è un fattore chiave per l'ideazione e lo sviluppo di nuovi farmaci.

L'espressione genica è un processo biologico che si compone di tre fasi: la trascrizione, il processamento del filamento di RNA e la traduzione in proteina. Durante lo splicing della molecola di pre-mRNA, le sequenze di esoni possono essere riarrangiate in maniera differente per via di un processo denominato "splicing alternativo". Le tecniche di DL, in questo caso, possono fornire dei modelli in grado di prevedere il risultato dello splicing alternativo nei diversi contesti cellulari. Oltre a ciò, possono essere sviluppati modelli che predicono il livello di espressione di un gene, restituendo un valore binario in uscita (che indica una bassa o alta espressione).

Il sequenziamento del DNA è un processo attraverso il quale si determina l'ordine preciso dei nucleotidi in una molecola di DNA. In particolare, le CNN si sono rivelate utili per l'analisi dell'attività funzionale del DNA.

2.2.2 Il deep learning nello studio dei biosegnali [23]

Con il termine "biosegnali" si denota l'insieme dei segnali termici, meccanici ed elettrici prodotti dal corpo umano. La loro acquisizione viene utilizzata dallo specialista per la rilevazione di eventuali patologie e per migliorare l'accuratezza della diagnosi. Nonostante siano fonte di grandi informazioni sullo stato di salute del soggetto, la loro analisi risulta spesso difficoltosa per la presenza di rumore e per la non stazionarietà di alcuni di essi. Il DL si è rivelato promettente nell'estrapolazione di informazioni a partire da questo tipo di dati complessi e rumorosi.

Nel caso del segnale elettrocardiografico (ECG), il DL trova applicazione nell'analisi di alcuni tracciati che possono riflettere la presenza di alcune patologie, come ad esempio aritmie cardiache e l'analisi delle apnee notturne e la relazione con i livelli di glucosio. Nel caso del riconoscimento della fibrillazione atriale e dell'identificazione di altre aritmie cardiache, le CNN ed RNN sono le reti più comuni utilizzate in letteratura. Le CNN sono state adoperate anche per la classificazione dell'apnea notturna di tipo ostruttivo e l'ipopnea.

L'EEG è un segnale, derivante dall'attività elettrica dell'encefalo, sfruttato per la diagnosi e il trattamento di disturbi di origine neurologica e la conseguente riabilitazione. Il segnale è particolarmente arduo da interpretare, per via del rumore e dell'alta variabilità. Anche in questo

caso il DL si è rivelato utile per superare alcune di queste difficoltà. In letteratura è possibile trovare modelli per implementare un algoritmo che individui automaticamente la fase del sonno del soggetto, estraendo le caratteristiche dal segnale riducendo i livelli di rumore. L'RNN si è rivelata una scelta popolare nello studio del segnale di soggetti epilettici, a volte combinata in un modello ibrido con una CNN.

2.2.3 Il deep learning nel bioimaging e medical imaging

Il gradino successivo allo studio a livello molecolare del DNA e delle proteine è rappresentato dalla citologia e dell'istologia, ossia rispettivamente lo studio delle cellule e dei tessuti. Questo è il campo di ricerca principale dove più di tutti trova applicazione il DL (si veda Fig.2.16), che in letteratura va sotto il nome di *biological imaging*. Con *medical imaging* si intende più propriamente il processo di acquisizione di immagini di organi umani mediante tecniche diagnostiche quali MRI, CT, PET, US ecc. [17]

L'accuratezza della diagnosi basata su immagini mediche dipende strettamente sia dalla qualità del sistema di acquisizione, sia dall'interpretazione dell'immagine stessa da parte dello specialista, che può essere pesantemente influenzata dall'esperienza e dalla valutazione soggettiva dello stesso, oltre banalmente a motivi di stress e fatica [25]. L'uso di algoritmi basati sul deep learning ha quindi il compito di facilitare l'interpretazione e l'analisi delle immagini, soprattutto perché al medico potrebbero sfuggire dettagli chiave per una diagnosi corretta.

In letteratura è possibile trovare una possibile suddivisione dei compiti svolti dalle reti di DL applicate alle immagini cliniche [18]:

- **classificazione:** in input solitamente possono essere presenti una o più immagini appartenenti allo stesso esame, lo scopo della rete è di restituire in uscita un valore che indichi se è presente o no una determinata patologia. La classificazione può anche riguardare oggetti che fanno parte di una piccola porzione dell'immagine, con l'obiettivo di distinguerli tra due o più possibili classi di appartenenza (per esempio, la classificazione di noduli in un'immagine CT del torace);
- **identificazione:** processo necessario per localizzare (nel dominio dello spazio o del tempo) un oggetto anatomico, quale un organo o una regione di interesse. Questa può prevedere anche l'identificazione di piccoli oggetti (come lesioni) in un'immagine. La

fase di identificazione è necessaria e deve essere eseguita prima della classificazione dell'oggetto;

- **segmentazione:** la segmentazione è la materia di più comune interesse negli articoli riguardanti l'applicazione del DL nelle immagini mediche. La segmentazione, ossia la suddivisione dell'immagine in regioni che sono omogenee rispetto ad una o più caratteristiche dell'immagine, occorre per effettuare analisi quantitative relative a parametri legati al volume o alla forma, in particolare nell'analisi di immagini cardiache e cerebrali. Per questo compito, la rete più famosa e adoperata è l'U-Net, trattata nella sezione 2.1.1.4;
- **generazione e miglioramento dell'immagine:** in letteratura sono stati proposti diversi algoritmi per il miglioramento della qualità dell'immagine, per la rimozione di alcuni elementi di disturbo, normalizzazione dell'intensità e riduzione del rumore. Un esempio è dato dall'eliminazione della presenza delle costole in radiografie del torace.

Una visione di insieme sulla quantità di articoli pubblicati in merito all'uso del DL nelle immagini mediche è data dalla Fig.2.20. Da notare l'importanza che assume la segmentazione tra i task inerenti alle immagini mediche.

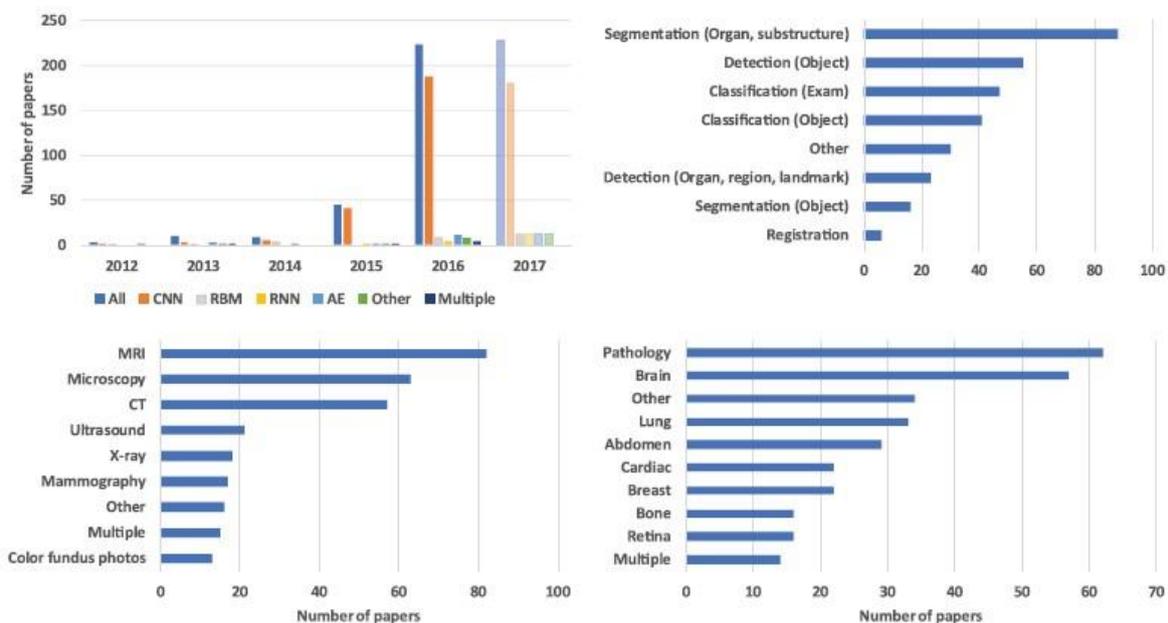


Figura 2.17: Panoramica sul numero di articoli pubblicati ordinati per: A) tipologia di rete utilizzata; B) obiettivo della rete; C) modalità di imaging; D) distretto di applicazione.

Come visibile dalla Fig.2.17, è evidente la preferenza dei ricercatori verso l'utilizzo di reti neurali convoluzionali. Nonostante al momento le CNN siano le architetture in grado di

raggiungere le migliori performance nella maggior parte delle competizioni legate all'analisi di immagini mediche, non bisogna pensare che la topologia della rete sia il fattore determinante nell'ottenimento di buoni risultati [18]. Molti gruppi di ricerca hanno ottenuto risultati molto differenti nelle varie challenge proposte, nonostante utilizzassero lo stesso tipo di rete. Due variabili importanti sono rappresentate infatti dalle tecniche di pre-processing (per poter generalizzare i dati in ingresso alla rete) e di *data augmentation* (per rendere la rete più robusta), il cui utilizzo in maniera corretta può sensibilmente aumentare le prestazioni della rete. L'ottimizzazione degli iper-parametri (come il learning rate e il dropout rate) può portare la rete a risultati migliori, sebbene non esistano linee guida chiare che possano aiutare nella scelta dei parametri, ma solamente approcci empirici e basati sull'intuizione del ricercatore stesso.

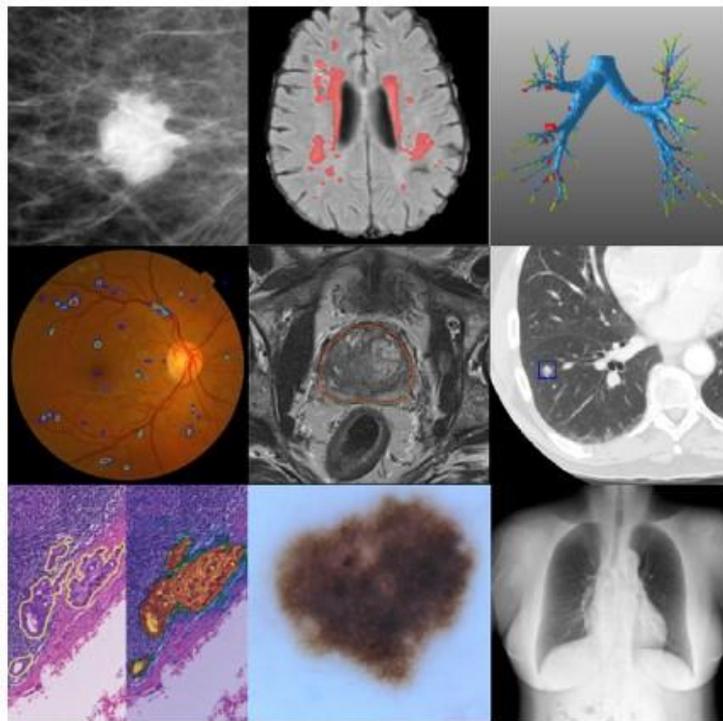


Figura 2.18: Un collage di applicazioni del DL alle immagini mediche in cui sono stati raggiunti risultati allo stato dell'arte. In ordine dall'alto a sinistra verso in basso a destra: A) classificazione di lesioni in mammografia; B) segmentazione di lesioni nell'encefalo; C) rilevazioni di ostruzioni nelle vie aeree; D) classificazione di retinopatia diabetica; E) segmentazione della prostata; F) classificazioni di noduli in CT torace; G) rilevazione di metastasi nei linfonodi dovute a carcinoma mammario; H) classificazione di lesioni della pelle; I) rimozione delle ossa in radiografia del torace.

2.3 APPLICAZIONI DEL DEEP LEARNING IN ULTRASONOGRAFIA

Come già si è visto nel corso del capitolo uno, l'ultrasonografia rappresenta uno dei metodi più flessibili di diagnostica per immagini tra quelli attualmente disponibili. Questo è dovuto principalmente all'eccellente portabilità del sistema di indagine, ai costi relativamente bassi se confrontati con gli apparecchi necessari ad altre metodiche e all'assenza di radiazioni ionizzanti. Naturalmente questi vantaggi sono in contrasto con la qualità dell'immagine ultrasonografica, spesso inficiata dalla presenza di rumore e per la risoluzione spaziale non ottimale. [26]. I task che coinvolgono l'imaging ad ultrasuoni sono principalmente quelli già discussi nella sezione precedente, come la classificazione, l'identificazione di lesioni e la segmentazione, riguardando anche la misurazione di parametri biometrici, la diagnosi assistita dal computer e, più recentemente, anche video-chirurgia e terapia (Fig.2.22).

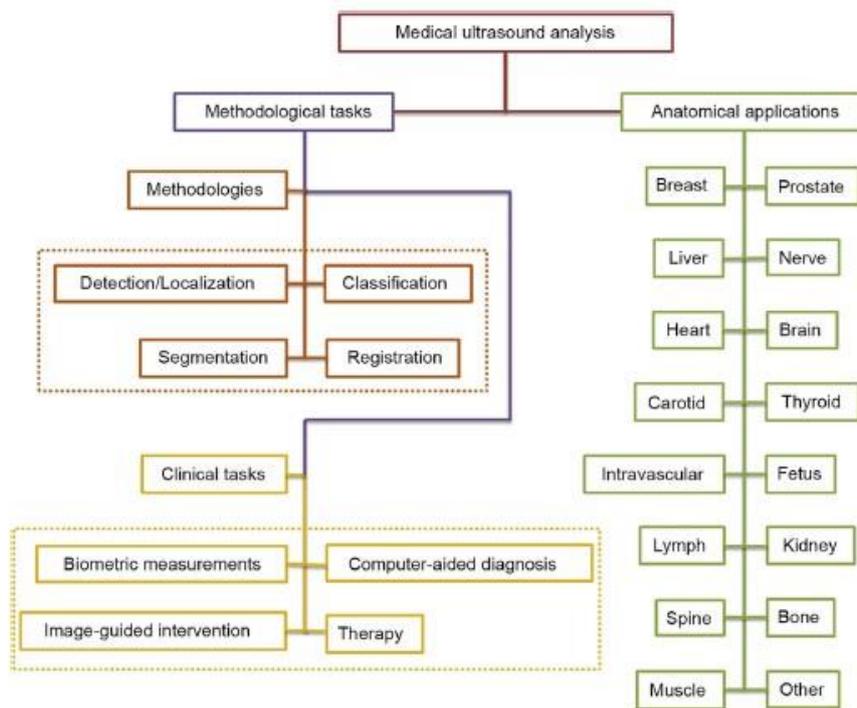


Figura 2.22: Analisi condotte mediante ultrasonografia.

2.3.1 Computer-aided diagnosis (CAD) [27]

Nell'ultrasonografia la diagnosi è fortemente dipendente dall'esperienza dell'operatore e dal tipo di dispositivo utilizzato. Risulta quindi evidente come sia necessario affiancare un sistema che possa fornire metodi di analisi automatica dell'immagine, che siano accurati, intelligenti e automatici. Questo consente di ottenere una "seconda opinione" che possa assistere il radiologo sulla diagnosi. Questo strumento di assistenza viene denominato *computer-aided diagnosis*

(CAD), ed è uno dei campi di ricerca più attivi nella radiologia diagnostica. Inizialmente sono stati proposti molti sistemi CAD con algoritmi basati sul machine learning classico, più recentemente invece si è fatto strada l'utilizzo di modelli di deep learning, basati in gran parte su reti neurali convoluzionali.

Le indagini cliniche per cui la ricerca su strumenti CAD è particolarmente attiva riguardano principalmente la classificazione e l'identificazione di lesioni a carico di seno, tiroide e fegato.

- Secondo le statistiche del *Centers for Disease Control and Prevention*, il cancro al seno è divenuto il più comune tra le neoplasie e la seconda causa di morte tra le donne nel mondo. Sebbene la mammografia sia il riferimento per lo screening e la diagnosi, l'ecografia al seno gioca comunque un ruolo fondamentale nella pratica clinica. Attraverso l'uso di CAD ad alta sensibilità (il più possibile vicina al 100% per evitare i falsi negativi) si può ridurre il rischio di effettuare delle biopsie non necessarie.
- I noduli tiroidei sono lesioni estremamente comuni che si stima siano presenti nel 50% della popolazione. L'unico procedimento non chirurgico per la diagnosi della lesione è la biopsia mediante ago aspirato, dalle cui analisi emerge che il 90% delle lesioni analizzate è benigna. Per questa ragione si rende necessario un metodo alternativo non invasivo per la classificazione delle lesioni, in modo da evitare delle biopsie non necessarie.
- Nel caso delle patologie a carico del fegato, l'ecografia è la modalità d'elezione per comprendere lo stato di salute dell'organo, in quanto è possibile avere una visione generale sull'aspetto del fegato e del flusso sanguigno attraverso la vena porta. Per la diagnosi delle alterazioni epatiche, anche in questo caso, la biopsia è la tecnica *gold standard*, che però può essere fonte di errori di campionamento e di complicazioni per il paziente.

2.3.2 Il deep learning per il processamento del segnale ecografico [28]

L'uso del deep learning in ultrasonografia non si limita solamente a strumento per l'analisi delle immagini in fase di post-processing, ma può anche inserirsi in uno o più blocchi della catena di elaborazione del segnale. Il fine principale è quello di ottenere nelle immagini formate un miglioramento del contrasto e della risoluzione, la soppressione del *clutter* (ossia la ricezione di echi indesiderati) e il miglioramento della stima spettrale (nella modalità Doppler).

Il capitolo uno ha mostrato come la semplicità e bassa complessità computazionale dell'algoritmo delay-and-sum lo abbia reso il metodo di beamforming standard in ecografia. Tuttavia, la semplicità di questo sistema di filtraggio spaziale è resa possibile al prezzo di alcuni fattori che causano il deterioramento della qualità dell'immagine. In particolare, assumendo una velocità costante del suono nel mezzo, si ottengono aberrazioni che alterano il corretto allineamento degli echi provenienti da un oggetto riflettente. Anche i pesi (l'apodizzazione) sono determinati a priori, mentre idealmente richiederebbero un'ottimizzazione tra la larghezza del lobo principale (risoluzione dell'immagine) e l'ampiezza dei lobi laterali (dispersione di potenza). Per questo sono stati proposti diversi algoritmi che sfruttano le reti neurali, alcuni che sostituiscono l'intero processo di beamforming, altri solamente alcuni anelli della catena.

2.3.2.1 L'utilizzo delle reti neurali nel beamforming

Per superare ad alcuni dei limiti del DAS, sono presenti in letteratura alcuni algoritmi "adattivi", che a seconda del segnale di input ricevuto, calcolano i pesi ottimali da moltiplicare a ciascun canale, in modo da minimizzare la varianza del segnale ricevuto dopo l'apodizzazione, mantenendo comunque guadagno unitario nella direzione desiderata. In questo modo è possibile ridurre il segnale indesiderato ricevuto da altre direzioni. Uno di questi approcci va sotto il nome di *minimum variance beamformer* (MV). Sebbene abbia mostrato di poter significativamente migliorare la qualità dell'immagine, il peso computazionale elevato ne impedisce una realizzazione applicabile in real-time all'ecografo. Per questo, invece di sostituire tutto il processo di beamforming, si utilizza una rete neurale per la predizione dei pesi ottimi da applicare al segnale focalizzato, cioè a cui sono stati preliminarmente applicati i ritardi. Il vettore peso ottimo, di dimensione pari al numero di elementi dell'array, viene predetto dalla rete per ciascun pixel dell'immagine, ricevendo in input il segnale a radiofrequenza a cui sono stati applicati i ritardi [29], [30].

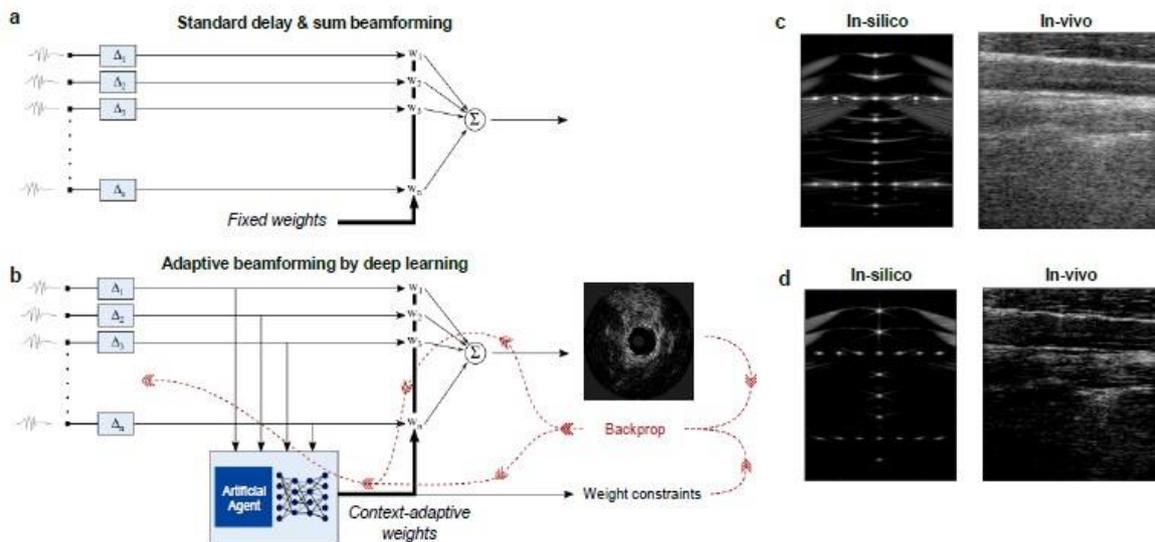


Figura 2.23: La figura mostra: A) beamforming con algoritmo DAS; B) beamforming adattivo usando il deep learning; C, D) confronto tra immagini ottenute con entrambi i metodi.

Sono stati proposti per il miglioramento dei parametri dell'immagine anche altri metodi di beamforming che utilizzano le reti neurali. In [31], [32] viene utilizzata una *deep neural network fully connected* per processare i dati nel dominio di Fourier. I dati dei canali (a cui sono già stati applicati i ritardi) vengono dapprima trasformati nel dominio della frequenza utilizzando la *short-time Fourier transform* (STFT), poi processati dalla rete neurale e infine trasformati nuovamente nel dominio del tempo e sommati per ottenere il segnale RF associato ad una particolare posizione assiale (Fig. 2.24). La rete neurale viene allenata appositamente per eliminare il segnale indesiderato proveniente dalle zone al di fuori dell'area di insonificazione (*off-axis scattering*).

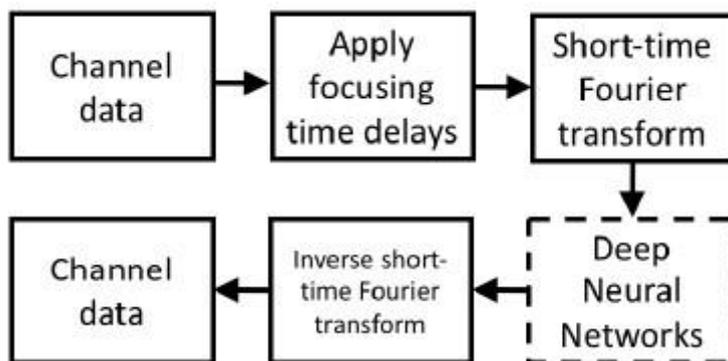


Figura 2.24: Flow chart del metodo proposto in [31]

Un'ulteriore classe di proposte riguarda l'utilizzo del deep learning per ottenere delle immagini di qualità superiore a partire da immagini ottenute con la trasmissione di una singola onda piana. I target della rete possono essere immagini di qualità superiore ottenute con acquisizione in

synthetic aperture [33], considerata la modalità di acquisizione gold standard, oppure immagini ottenute come *compounding* di più onde piane [34].

Un'altra classe di lavori mira invece al miglioramento del frame rate di acquisizione, utilizzando i modelli di deep learning per produrre immagini di alta qualità a partire dal segnale RF sottocampionato [35]–[37].

La classe di lavori di maggiore interesse, in ottica del presente elaborato, prevede invece l'utilizzo diretto del segnale RF grezzo, ossia prima dell'applicazione dei ritardi, da utilizzare con input alla rete neurale. In [38] viene utilizzata una GAN per l'estrazione diretta delle informazioni dai dati RF, generando contemporaneamente in uscita l'immagine e la maschera di segmentazione di una ciste presente nell'immagine. Questo studio presenta una certa rilevanza, in quanto dimostra che una rete neurale può estrarre un quantitativo sufficiente di informazioni dal segnale RF grezzo per poter compiere questo tipo di task.

CAPITOLO 3

MATERIALI E METODI

Come mostrato nel capitolo precedente, il deep learning può essere sfruttato per lo svolgimento di alcuni task atti al miglioramento della qualità dell'immagine. In [38]–[42] è stato mostrato come sia possibile estrarre informazione direttamente dai dati RF e IQ del segnale ecografico, in modo che la DNN possa generare in uscita sia l'immagine B-mode, sia la maschera di segmentazione. Questo tipo di implementazione costituisce una trasformazione end-to-end, in cui il processo di beamforming tradizionale viene sostituito completamente dalla rete neurale, e non solamente supportato dalla rete come nei metodi [29]–[32] visti in 3.1.2.1, in cui vengono modificati unicamente alcuni step intermedi. Questo tipo di approccio nasce da alcuni limiti posti dall'ultrasonografia: la presenza di speckle e clutter può complicare l'interpretazione dell'immagine, in particolar modo se si stanno svolgendo delle operazioni che richiedono l'uso di algoritmi automatici; il rumore può rendere difficoltosa la segmentazione dell'immagine poiché confonde gli algoritmi di filtraggio e di *thresholding* automatico, costringendo all'utilizzo di algoritmi più complessi e computazionalmente dispendiosi; la segmentazione viene generalmente effettuata dopo la formazione dell'immagine, incrementando la complessità computazionale delle operazioni.

L'utilizzo di una trasformazione end-to-end che sostituisca il processo tradizionale di beamforming porterebbe quindi vantaggi anche sul fronte della velocità di formazione dell'immagine: infatti, per ottenere immagini di qualità elevata è necessario trasmettere un elevato numero di onde piane con angolazioni diverse, che vengono poi mediate tra loro. Questo, tuttavia, riduce drasticamente il frame-rate; perciò, la rete di deep learning potrebbe rappresentare una soluzione nel fornire un'immagine di qualità superiore anche utilizzando l'acquisizione a singola onda piana.

Il lavoro effettuato si pone l'obiettivo di estendere quanto già ottenuto in [41]. Nel corso del capitolo si procederà descrivendo l'architettura della rete, la preparazione dei dati in ingresso, le impostazioni di allenamento della rete e i diversi tipi di dataset adoperati.

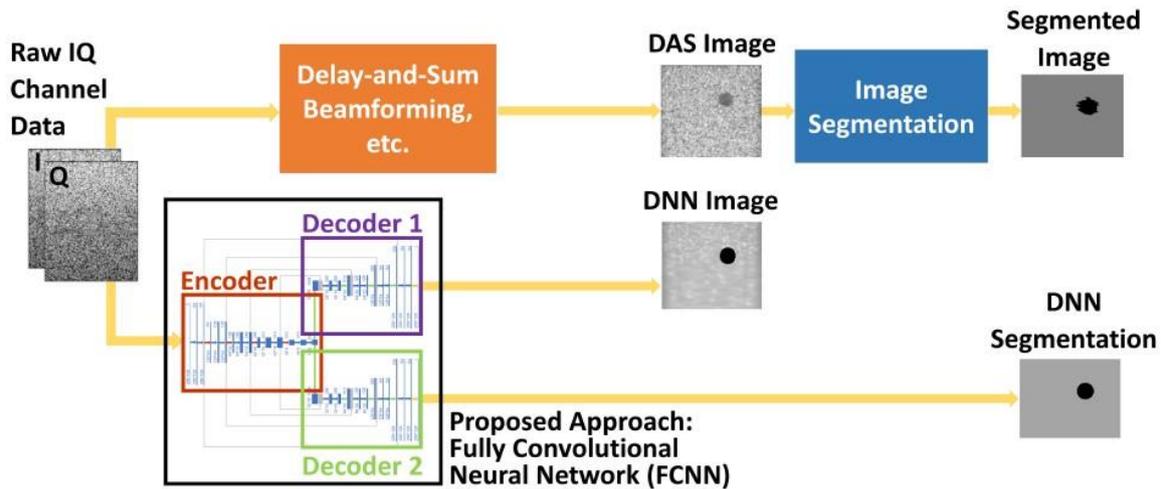


Figura 3.1: Illustrazione del metodo proposto confrontato con il metodo tradizionale. A partire dal segnale grezzo vengono predetti in uscita dalla FCNN (con un encoder e due decoder) sia l'immagine B-mode sia la maschera di segmentazione. Nel caso tradizionale, invece, si utilizza il beamforming DAS, seguito da envelope-detection, log-compression e filtraggio.

3.1 METODI

3.1.1 Formulazione del problema

La rete neurale può essere adattata per avere in input sia i dati radiofrequenza (RF), sia i dati in fase e in quadratura (IQ). I dati RF sono rappresentabili mediante un tensore tridimensionale di dimensione $d \times w \times q$, in cui d rappresenta la lunghezza del segnale, w è la larghezza dell'immagine, pari al numero di trasduttori della sonda e q è il numero di canali, pari ad uno nel caso del segnale RF. La situazione è analoga per il tensore IQ, tuttavia essendo il segnale complesso (per via della trasformazione di Hilbert) si avranno due canali (uno per la componente in fase e l'altro per quella in quadratura). Chiamando I_d il tensore d'ingresso, l'obiettivo è di ottenere l'immagine B-mode predetta " D " e la maschera di segmentazione " S_p ", entrambe di dimensioni $d \times w$. La rete (*fully convolutional neural network*) viene allenata in modo tale da apprendere i parametri ϑ che minimizzano la differenza tra l'input e il riferimento y , costituito dall'immagine generata mediante il metodo standard Delay and Sum (DAS) e dalla maschera di segmentazione " S_t ". Per il primo target, sono state testate due diverse tipologie di immagini: l'immagine post-beamforming (B-mode) oppure l'immagine immagine rinforzata " E " (*enhanced*), generata dalla sovrapposizione dell'immagine DAS con la maschera di segmentazione, ossia ponendo a zero tutti i pixel appartenenti alla ciste (Fig.3.2). Dal punto di vista concettuale, questo consente di sensibilizzare la rete a enfatizzare l'anecogenicità della ciste, differenziando in modo più efficiente il tessuto esterno dall'interno della struttura. Lo

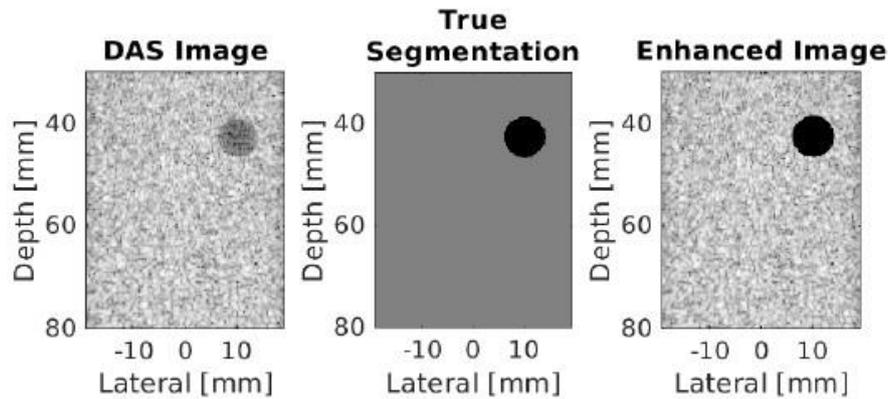


Figura 3.2: L'immagine simulata DAS e la maschera di segmentazione. Il target che la rete deve riprodurre è l'immagine DAS sovrapposta alla maschera di segmentazione, in modo che l'oggetto anecoico venga messo in evidenza.

scopo della rete è infatti non quello di replicare l'immagine DAS originale, in cui la ciste intrinsecamente anecogena è sovrapposta a rumore acustico, ma di fornire un'immagine di qualità superiore.

3.1.2 Architettura della rete

La rete si basa sull'U-Net [21], costituita in questa particolare applicazione da un encoder e due decoder, in modo tale che in uscita si abbia sia l'immagine sia la maschera di segmentazione (Fig.3.3).

3.1.2.1 Struttura dell'encoder

L'encoder si basa sulla struttura della rete convoluzionale Visual Geometry Group 13 (VGG-13) [43], in cui viene aggiunto, dopo ogni operazione di convoluzione, un layer di batch-normalization, per rendere la rete più stabile e velocizzare il tempo di convergenza. L'encoder riceve l'input (di dimensione 256x128) e applica ripetutamente una serie di due layer convoluzionali 3x3 (con padding=1 e stride=1), seguiti dalla funzione di attivazione ReLU e dal layer di max-pooling 2x2 (con padding=0 e stride=2). Dopo ognuno di questi step (ripetuti per cinque volte), le dimensioni spaziali dell'input vengono dimezzate, mentre il numero di *feature channel* raddoppia. Le prove sono state condotte anche variando la struttura

dell'encoder, utilizzando VGG-16 e VGG-19, che differiscono per il numero di layer convoluzionali presenti (Fig.3.4).

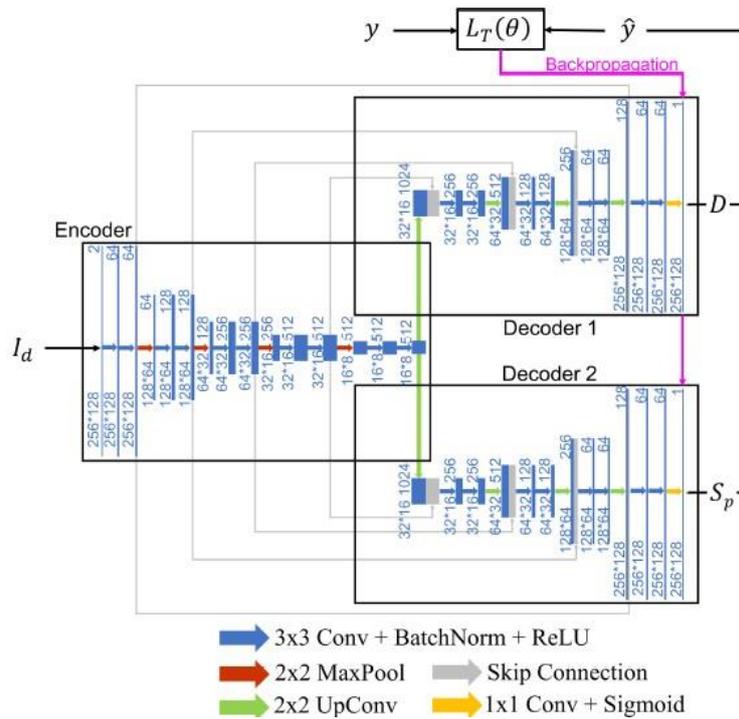


Figura 3.3: Architettura della rete.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512 conv3-512

Figura 3.4: Confronto tra VGG-13 (B), VGG-16 (D) e VGG-19 (E).

3.1.2.2 Struttura del decoder

L'encoder viene seguito da due decoder identici e speculari, che producono in uscita l'immagine e la maschera di segmentazione. Ogni blocco del decoder (ripetuto quattro volte) è costituito da: un layer di up-sampling che opera una convoluzione 2x2 (padding=0 e stride=2), la concatenazione del layer con la feature map corrispondente dell'encoder, una doppia convoluzione 3x3 (padding=1 e stride=1) e la funzione di attivazione ReLU. Dopo ogni blocco le dimensioni spaziali raddoppiano mentre la dimensione del feature channel dimezza. L'ultimo layer è costituito per entrambi da una funzione sigmoide, che restringe i valori possibili in uscita tra 0 e 1.

3.1.3 Preparazione dei dati

Una volta simulate le immagini che comporranno il dataset di training (come si vedrà nelle sezioni successive), occorre pre-processare i dati. I dati RF vengono acquisiti dalla sonda in numero pari ai trasduttori. La lunghezza del segnale, in numero di campioni, dipende dalla frequenza di campionamento impostata e dalla profondità della zona insonata. Il segnale viene sottocampionato mediante interpolazione lineare, in modo che il tensore di ingresso alla rete abbia dimensione fissata (256x128 pixel in questo caso), ossia si fa corrispondere ad ogni campione temporale una precisa posizione spaziale, determinata assumendo che la velocità del suono nel mezzo sia fissata. I valori di ampiezza del segnale vengono quindi normalizzati tra -1 ed 1, dividendo ciascun campione per il valore massimo assoluto. In questo modo si ottiene il tensore di ingresso indicato con I_d (il procedimento è identico anche per il tensore dei dati IQ, con la differenza che possiede due canali anziché uno solo).

L'immagine target (denominata con E) viene ricavata a partire dal segnale RF grezzo mediante i passaggi già descritti nel capitolo uno: beamforming con algoritmo DAS, demodulazione IQ, ricampionamento e log-compression, ottenendo così I_{dB} . In questo modo i valori da assegnare a ciascun pixel dell'immagine sono in dB, ristretti tra -60 dB e 0 (nel caso si utilizzi un dynamic range standard). Questi devono tuttavia essere normalizzati tra 0 e 1 per l'ottimizzazione delle performance della rete, secondo la formula:

$$I_n = \frac{I_{dB} + 60}{60} \quad (3.1)$$

Come già accennato, per ottenere un'immagine target di qualità superiore, I_n viene sovrapposta alla maschera di segmentazione, in modo tale che i pixel appartenenti alla ciste anecogena abbiano valore nullo.

3.1.3.1 Immagini con dynamic range ottimizzato

Lo scopo del lavoro è di verificare gli effetti della generalizzazione del dataset, non solo introducendo immagini con forme non circolari o di differente ecogenicità, ma anche verificando un'eventuale risposta benefica della rete se si utilizzano immagini nell'allenamento in cui il dynamic range (dR) viene calcolato automaticamente. Le prime prove sono state effettuate allenando la rete solamente con immagini in cui il dynamic range fosse standard (60 dB), mentre nelle successive si sono confrontati i risultati tra allenamenti con immagini a dynamic range standard e immagini in cui venisse stimato automaticamente. Infatti, attraverso l'istogramma delle luminosità dell'immagine è possibile stimare un valore automatico di dynamic range che ottimizza la visualizzazione dell'immagine, che potrebbe essere influenzata dalle diverse posizioni ed ecogenicità delle strutture. Si utilizza l'algoritmo proposto in [44] per il calcolo automatico del valore ottimale dR; in seguito, i valori dei pixel dopo l'operazione di log-compression vengono compressi nell'intervallo $[-dR \ 0]$, e infine normalizzati tra 0 e 1 mediante la formula:

$$I_n = \frac{I_{dB} + dR}{dR} \quad (3.2)$$

Dal punto di vista della notazione, l'immagine B-mode con dynamic range standard verrà identificata con B-mode₆₀ (oppure E₆₀ nel caso in cui sia rinforzata con la maschera di segmentazione), viceversa per le immagini a dynamic range automatico si adopererà la sigla B-mode_{auto} (oppure E_{auto}).

In questo modo l'obiettivo è di disporre di una DNN che oltre a fornire contemporaneamente immagini di qualità superiore e la relativa maschera di segmentazione, possa anche predire un'immagine con dynamic range ottimale.

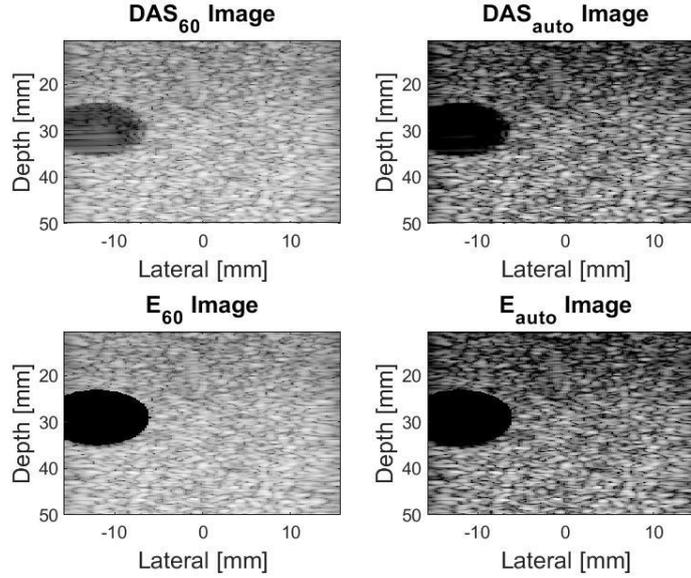


Figura 3.5: Confronto tra le diverse immagini usate negli allenamenti. A) Con $B\text{-mode}_{60}$ si denota l'immagine B-mode con dynamic range standard; B) con $B\text{-mode}_{auto}$ si denota l'immagine B-mode con dynamic range automatico (-40 dB nel caso mostrato); C) con E_{60} si denota l'immagine "enhanced" con dynamic range standard; D) con E_{auto} si denota l'immagine "enhanced" in cui viene adoperato il dynamic range automatico.

3.1.4 Impostazioni di allenamento della rete

Considerando che la rete deve produrre due uscite contemporaneamente, la loss function si basa sulla somma di due contributi: il primo che misura la distanza tra l'immagine predetta D e il target E , il secondo che quantifica la differenza tra la maschera predetta S_p e quella reale S_t . I pesi θ della rete vengono aggiornati dopo un certo numero di campioni dati in ingresso, che vengono posti in ingresso alla rete concatenati tra loro in un unico tensore (il numero di campioni è specificato dalla grandezza del mini-batch).

Per la quantificazione dell'errore di predizione dell'immagine si adopera la norma L1, definita come:

$$L1Loss(\theta) = \frac{1}{n} \sum_{i=1}^n \frac{\|D_i(I_d; \theta) - E_i\|_1}{N} \quad (3.3)$$

Dove:

- $\| \cdot \|_1$ è la norma L1;
- D_i è l'i-esima immagine predetta vettorizzata;
- E_i è l'i-esima immagine target vettorizzata;

- N è il numero totale di pixel dell'immagine;
- n è la dimensione del mini-batch.

La seconda loss function si basa sul coefficiente di similarità Dice, calcolato tra la maschera di segmentazione vera e predetta:

$$DSCLoss(\theta) = \frac{1}{n} \sum_{i=1}^n 1 - 2 \frac{|S_{p,i}(I_d; \theta) \cap S_{t,i}|}{\underbrace{|S_{p,i}(I_d; \theta)| + |S_{t,i}|}_{DSC}} \quad (3.4)$$

Il coefficiente di similarità Dice (si veda la sezione 3.1.5) ha valore 1 nel caso in cui le due immagini siano identiche, 0 se tutti i pixel sono sbagliati. Per cui, dato che la funzione di loss è un valore da minimizzare, la DSCLoss è data da 1-DSC.

Anche in questo caso le maschere di segmentazione predette e reali sono vettorizzate. La maschera reale S_t è binaria, con i valori dei pixel che possono assumere solo 0 o 1. I valori della maschera predetta S_p possono assumere qualsiasi valore nell'intervallo tra 0 e 1, dove 0 indica che la rete identifica un particolare pixel come tessuto con confidenza al 100%, viceversa con 1 identifica il pixel come appartenente alla ciste con confidenza al 100%. Durante la fase di allenamento è necessario mantenere i valori della maschera predetta tra 0 e 1 per poter aggiornare i pesi, nonostante la maschera di segmentazione target sia binaria. Al contrario, in fase di validazione, i valori predetti vengono trasformati in 0 o 1 mediante l'uso di una soglia.

La funzione di loss totale è la somma delle due descritte:

$$L_T(\theta) = DSCLoss(\theta) + L1Loss(\theta) \quad (3.5)$$

3.1.5 Metriche di valutazione dei risultati

Si descrivono ora le metriche adoperate per valutare la qualità dei risultati predetti della rete.

- 1) **Dice Similarity Coefficient (DSC):** il DSC [45] quantifica il grado di sovrapposizione tra due maschere di segmentazione:

$$DSC(S_p, S_t) = 2 \frac{|S_p \cap S_t|}{|S_p| + |S_t|} \quad (3.6)$$

Nel caso in cui la maschera DNN predetta sia esattamente uguale alla maschera reale, il DSC avrà valore pari a 1. Dato che i valori predetti dalla rete sono compresi tra 0 e 1,

prima della visualizzazione e della valutazione del DSC è necessario sogliare i valori dei pixel in modo che quelli >0.5 siano posti a 1 e quelli <0.5 a 0.

- 2) **Contrast (CR):** il contrasto viene utilizzato per dare una misura quantitativa delle differenze tra due regioni. Nel caso in esame, si utilizza per valutare la differenza tra una regione interna alla ciste e quella esterna. Viene definito come:

$$Contrast = 20 \log_{10} \left(\frac{S_i}{S_o} \right) \quad (3.7)$$

Dove S_i è la media dei valori dei pixel interni alla regione di interesse (ROI); S_o la media dei valori esterni. La ROI viene selezionata automaticamente nel caso delle immagini simulate, in particolare la ROI interna è una regione circolare di raggio 2 mm posta vicino al bordo della ciste, la ROI esterna è geometricamente identica ma posta al di fuori della ciste, alla stessa profondità della ROI interna. Nel caso delle immagini reali (phantom e carotide), le ROI vengono selezionate manualmente.

- 3) **Signal To Noise Ratio (SNR):** il rapporto segnale rumore della regione esterna alla ciste quantifica la qualità con cui la rete riproduce lo speckle:

$$SNR = \frac{S_o}{\sigma_o} \quad (3.8)$$

Dove S_o è la media dei valori dei pixel contenuti nella ROI esterna alla ciste, mentre σ_o la loro deviazione standard.

- 4) **Generalized Contrast-to-Noise Ratio (gCNR):** il gCNR [46] è un indice introdotto recentemente in letteratura, per fornire una stima più accurata della rilevabilità della lesione.

$$gCNR = 1 - \sum_{x=0}^1 \min_x \{p_i(x), p_o(x)\} \quad (3.9)$$

Con $p_i(x)$ funzione di massa di probabilità dei pixel interni alla ROI, $p_o(x)$ di quelli esterni.

- 5) **Peak Signal-to-Noise Ratio (PSNR):** il PSNR [47], [48] quantifica la similarità tra l'immagine DNN predetta D e l'immagine E reale (l'immagine DAS rinforzata con la maschera di segmentazione). Viene definita come il rapporto tra il valore massimo che può assumere un pixel in E (ossia 1) e il *mean square error* tra D ed E .

$$PSNR(D, E) = 10 \log_{10} \left(\frac{MAX_E^2}{MSE} \right)$$

$$PSNR(D, E) = 10 \log_{10} \left(\frac{1}{\frac{\|D - E\|_2^2}{N}} \right) \quad (3.10)$$

Con D ed E in forma vettorizzata, $\| \cdot \|_2$ è la norma L2 e N è il numero di pixel totali dell'immagine.

3.2 COSTRUZIONE DEL DATASET

Le differenti prove condotte sono state condotte consecutivamente, adoperando differenti tipologie di dataset per il training della rete. Inizialmente, come *proof-of-concept*, si è utilizzato un dataset interamente costituito da cisti anecogene circolari di differenti dimensioni, passando quindi ad allargare il dataset introducendo dapprima cisti iperecogene e multiple (ossia immagini che presentavano contemporaneamente due cisti di differente ecogenicità), poi cisti irregolari costruite come sovrapposizione di diverse ellissi. Si è studiato da ultimo l'effetto che produceva nella rete DNN l'introduzione di immagini che venivano generate simulando anche l'effetto prodotto dal coefficiente di attenuazione dei tessuti.

3.2.1 Simulazione delle immagini

3.2.1.1 Dataset con cisti circolari anecogene

Per la prima prova effettuata si costruisce un dataset contenente solamente immagini di cisti anecoiche circolari. Per la simulazione si utilizza il pacchetto Field II su Matlab 2019b. Il dataset consiste in 19978 immagini, ripartite in 16384 nel training set, 1697 nel validation set e 1717 nel test set. Nella simulazione vengono generati e salvati: i dati RF, i dati IQ, l'immagine B-mode DAS (D), l'immagine enhanced (E) e la maschera di segmentazione (Fig.3.6). Le immagini hanno dynamic range standard (60 dB).

Le cisti sono contenute all'interno di un volume cuboidale, di larghezza laterale 32 mm, profondità assiale tra 10 e 40 mm e spessore 1 mm. I parametri delle cisti sono illustrati in Tab.3.2, dove si osserva anche entro quale range possibile poteva essere scelta casualmente la posizione del centro della ciste e il raggio. La velocità del suono nel mezzo è costante ed

Parameter	Value
Element number	128
Element width	0.215 mm
Kerf	0.030 mm
Pitch	0.245 mm
Aperture	31.36 mm
Transmit frequency	5 MHz
Sampling frequency	100 MHz

Tabella 3.1: Parametri di simulazione del trasduttore.

Parameter	Range	Increment
Radius (r)	2-10 mm	1 mm
Lateral position of cyst center (x)	-16–16 mm	1 mm
Axial position of cyst center (z)	20-40 mm	1 mm

Tabella 3.2: Parametri di simulazione delle cisti.

impostata pari a 1540 m/s. All'interno del phantom viene selezionato un numero di *scatterer* pari a 10000, per poter ottenere uno speckle completamente sviluppato.

I parametri della simulazione relativi alle impostazioni del trasduttore sono mostrati in Tab.3.1.

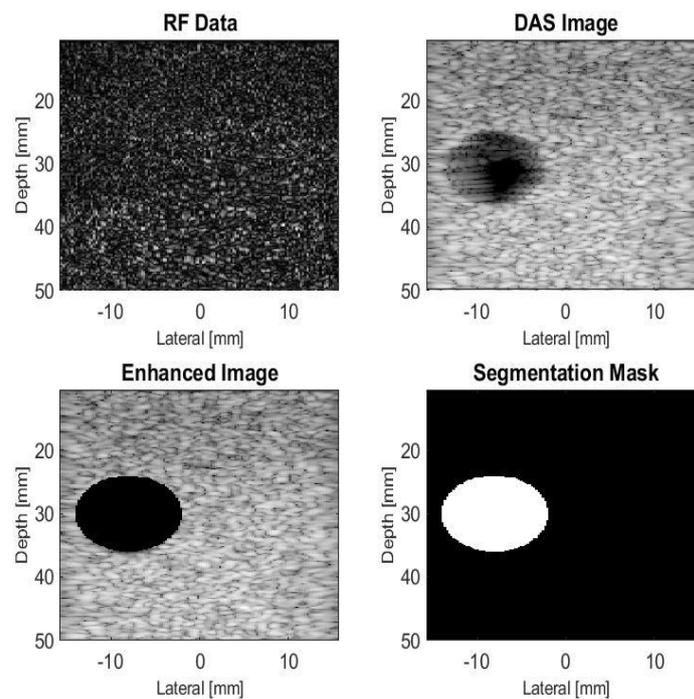


Figura 3.6: Le immagini generate su Field II per la costruzione del dataset.

3.2.1.2 Introduzione di cisti iperecogene circolari

Le immagini utilizzate per questa prova sono analoghe a quella precedente, con i parametri identici a quelli di Tab.3.1 e Tab 3.2. La differenza consiste nell'aggiunta al dataset di immagini

con cisti iperecogene (si simula la ciste in modo che i valori di intensità interni siano 10 volte maggiori di quelli esterni, ossia con contrasto di +20 dB). Si ripete lo stesso ragionamento effettuato per le immagini rinforzate nel caso delle cisti anecoiche: siccome i valori della ciste interna sono corrotti da rumore acustico, si costruisce un'immagine rinforzata "enhanced", in cui i valori dei pixel appartenenti all'interno della ciste vengono posti a 1. Le immagini enhanced rappresentano le immagini target con cui viene allenata la rete. Complessivamente il dataset si compone di: un training set di 8192 immagini, ripartiti in 4096 con target anecoico e 4096 iperecoico; un validation set e un test set contenenti entrambi 816 immagini, ripartite egualmente tra target anecogeni e iperecogeni (Fig.3.7).

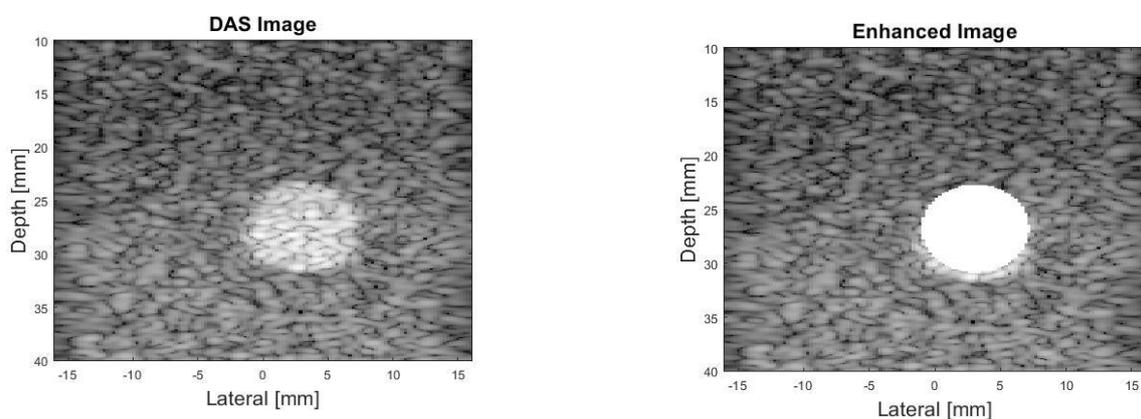


Figura 3.7: Immagine di ciste iperecogena.

3.2.1.3 Introduzione di cisti doppie circolari

In questa prova il dataset viene ulteriormente ampliato aggiungendo, oltre alle immagini con cisti anecogene e iperecogene, anche immagini in cui fosse presente contemporaneamente una ciste anecogena e una iperecogena, per sensibilizzare la rete a riconoscere immagini in cui fossero presenti anche più oggetti (Fig.3.8).

Il dataset si compone di: un training set di 6144 immagini, ripartite in egual misura tra le tre classi (cisti singole anecoiche; cisti singole iperecoiche; doppia ciste); un validation set di 1224 immagini; un test set di 1224 immagini (sempre ripartite tra le tre classi in egual misura). In questa prova si effettua anche data augmentation per incrementare il numero di immagini disponibili, che vengono ottenute capovolgendo attorno all'asse $x=0$ le immagini disponibili e raddoppiando così il contenuto del training set (che quindi ammonta in definitiva a 12288

immagini). In questo modo si simula che le nuove immagini vengano acquisite mediante la rotazione di 180° del trasduttore.

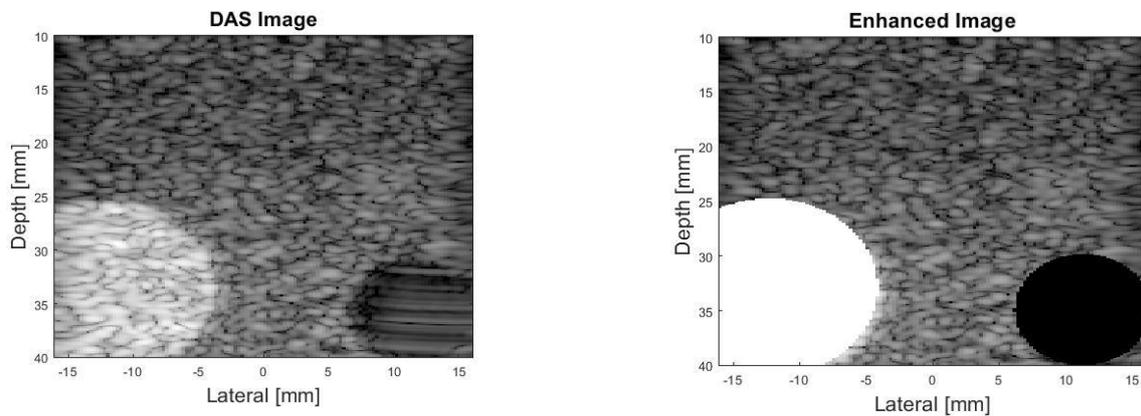


Figura 3.8: Immagine con due cisti: una anecogena e una iperecogena.

3.2.1.4 Introduzione di cisti di forma irregolare

La quarta prova consiste nell'utilizzare un dataset più generale rispetto ai precedenti, introducendo nuove immagini con cisti dalla forma irregolare. Le cisti vengono generate sempre tramite Field II, mediante un'apposita modifica dello script già adoperato in precedenza. Il programma, invece di modellare la ciste come una circonferenza, utilizza una sovrapposizione di più ellissi, delle quali la prima viene generata utilizzando i parametri di Tab.3.3 e scegliendo i valori dai range casualmente. Grazie all'equazione completa dell'ellisse traslata, è possibile generare la figura a partire dal valore del semiasse maggiore, dell'eccentricità e dell'angolo di rotazione. La seconda ellisse viene sovrapposta alla prima, scegliendo il nuovo centro casualmente tra i punti appartenenti alla prima ellisse, modificando casualmente il semiasse maggiore di massimo ± 2 mm e scegliendo un nuovo valore per l'eccentricità e l'angolo di rotazione. Per aumentare la complessità della forma e la variabilità delle immagini, si imposta che una terza ellisse possa essere generata e sovrapposta alla struttura con probabilità pari al 70%.

Si introducono anche cisti ipoecogene, ossia in cui l'intensità dei punti interni sia un valore compreso tra 0.2 e 0.5 volte rispetto ai punti esterni (che si traduce in un contrasto tra -14 e -6 dB, rispettivamente). Per questo tipo di cisti, l'immagine "enhanced" è identica all'immagine

Parameter	Range	Increment
Semi-major axis (a)	2.5-9 mm	0.5 mm
Eccentricity (e)	0.3-0.85	0.05
Angle of rotation (ϑ)	0-179°	1°
Lateral position of cyst center (x)	-16-16 mm	1 mm
Axial position of cyst center (z)	20-50 mm	1 mm

Tab.3.3: Parametri di simulazione delle cisti irregolari.

DAS originale, ossia non si utilizza una maschera di rinforzo come nel caso delle cisti anecogene e iperecogene. In questo modo la rete viene sensibilizzata a distinguere tra una ciste realmente anecogena ed una ipoecogena, permettendo di discernere tra i due casi nel caso non fosse possibile per via della presenza di clutter all'interno della ciste (Fig.3.9).

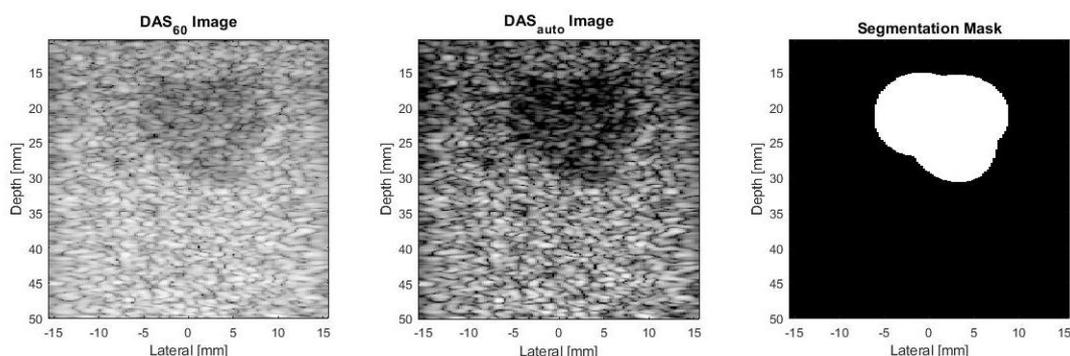


Figura 3.9: Esempio di simulazione di immagine di ciste irregolare ipoecogena con: A) dynamic range 60 dB; B) dynamic range regolato automaticamente; C) maschera di segmentazione.

In questa prova si utilizzano anche le immagini con dynamic range automatico, descritte in 3.1.3.1. Queste verranno utilizzate come immagini target in fase di allenamento per verificare se sia effettivamente un miglioramento rispetto all'utilizzo di immagini con dynamic range standard. Complessivamente il dataset si compone di: un training set di 6560 immagini; un validation set di 816 immagini; un test set di 816 immagini. Ciascun set contiene in parti uguali le seguenti categorie di cisti:

- Cisti anecogene singole;
- Cisti ipoecogene singole;

- Cisti iperecogene singole;
- Cisti doppie (una anecogena e una iperecogena).

Per la composizione del dataset si riutilizzano anche le immagini circolari generate precedentemente, in modo che per ogni categoria si abbia metà delle immagini con cisti circolari e metà con cisti a forma di grappolo di ellissi.

Anche in questo caso si utilizza la tecnica di data augmentation già descritta precedentemente, ottenendo così un training set composto da 13120 immagini.

3.2.1.5 Introduzione della simulazione del coefficiente di attenuazione dei tessuti

Il dataset utilizzato in questa prova utilizza cisti con forme e parametri identici a quelli già descritti in 3.2.1.4, con la differenza che le immagini vengono simulate anche introducendo un coefficiente di attenuazione dei tessuti pari a $1.5 \text{ dB}\cdot\text{cm}^{-1}\text{MHz}^{-1}$, come visibile dalla Fig.3.10. Inoltre, considerando i risultati delle prove precedenti, si utilizza la maschera di rinforzo solamente per le immagini in cui sono presenti cisti anecogene e non più anche per le iperecogene (come in Fig.3.6).

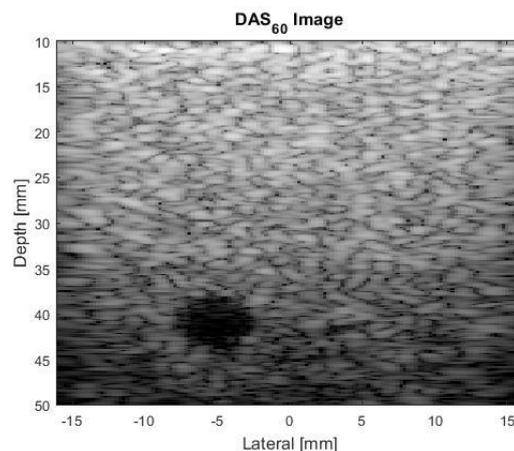


Figura 3.10: Esempio di simulazione di immagine di ciste con attenuazione introdotta dai tessuti.

Complessivamente il dataset si compone di: un training set di 6560 immagini; un validation set di 816 immagini; un test set di 816 immagini. Con l'uso di data augmentation il training set ammonta a 13120 immagini.

3.2.2 Immagini sperimentali

Le immagini sperimentali utilizzate in fase di test delle diverse reti allenate consistono in 24 immagini phantom (modello 040GSE, CIRS Inc, USA) e 21 immagini *in-vivo* di carotide umana, acquisite su due volontari. Per l'acquisizione *plane wave* si utilizza un ecografo Verasonics Vantage 128 system (L11-5v linear array). Il phantom contiene oggetti target anecoici, iperecoici e ipoecoici, localizzati ad una profondità compresa tra 5 e 60 mm (Fig.3.11).

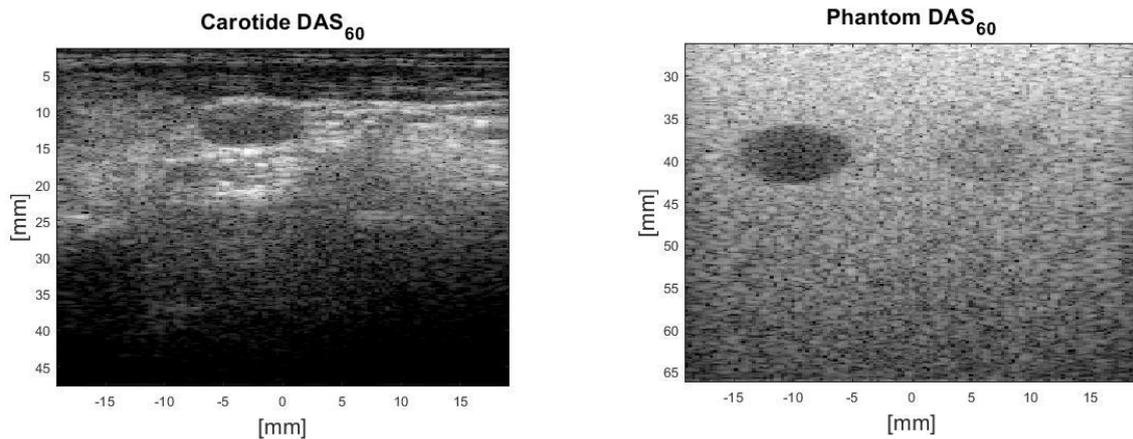


Figura 3.11: L'immagine mostra due esempi di immagini reali: A) carotide; B) phantom con target anecoico ed ipoecoico.

3.3 ALLENAMENTO DELLA RETE

Per le prime prove effettuate l'allenamento viene condotto sia utilizzando i dati RF in ingresso, sia i dati IQ. Gli iperparametri vengono modificati leggermente di prova in prova, in base ai risultati ottenuti dalle prove precedenti. Nella prova del dataset 3.2.1.3 si utilizza anche un criterio di *early stopping*, in cui l'allenamento viene interrotto nel caso la loss calcolata sul validation set non dovesse migliorare per quattro epoche consecutive, utilizzando 50 epoche come limite superiore.

Per gli allenamenti con il dataset di sezione 3.2.1.4 e 3.2.1.5, gli allenamenti vengono effettuati utilizzando in input i dati IQ. Gli iperparametri utilizzati sono riassunti in Tab.3.4.

Nel caso delle prove effettuate con i dataset 3.2.1.4 e 3.2.1.5, si confrontano i risultati utilizzando diversi tipi di immagini target, osservando eventuali differenze tra l'allenamento con target senza maschera di rinforzo a dynamic range standard e automatico (B-mode₆₀ e B-mode_{auto}), e in seguito utilizzando immagini con maschera di rinforzo (E₆₀ ed E_{auto}). La rete viene allenata variando anche l'encoder dell'U-Net, sostituendo alla struttura VGG-13 gli

encoder basati sulle reti convoluzionali VGG-16 e VGG-19 (descritti in 3.1.2.2). Le prove condotte con questi due dataset sono descritte in Tab.3.5.

Dataset	Number of epochs	Batch size	Learning rate
Section 3.2.1.1	25	16	10^{-5}
Section 3.2.1.2	35	16	10^{-5}
Section 3.2.1.3	36 (RF) / 45 (IQ)	16	10^{-5}
Section 3.2.1.4	50	16	$5 \cdot 10^{-5}$
Section 3.2.1.5	50	8	$5 \cdot 10^{-5}$

Tabella 3.4: Selezione degli iperparametri.

Dataset section 3.2.1.4		Dataset section 3.2.1.5	
Encoder	Target image	Encoder	Target image
VGG-13	B-mode ₆₀	VGG-13	B-mode ₆₀
	E ₆₀		E ₆₀
	B-mode _{auto}		B-mode _{auto}
VGG-16	E _{auto}	VGG-16	E _{auto}
	B-mode _{auto}		B-mode ₆₀
	E ₆₀		E ₆₀
VGG-19	B-mode _{auto}	VGG-19	B-mode _{auto}
	E ₆₀		E ₆₀
	E _{auto}		E _{auto}

Tabella 3.5: Prove condotte utilizzando: A) il dataset descritto in 3.2.1.4; B) il dataset descritto in 3.2.1.5.

CAPITOLO 4

RISULTATI E DISCUSSIONE

4.1 RISULTATI DEGLI ALLENAMENTI

4.1.1 Dataset con cisti anecogene

La curva di allenamento della rete con dati IQ in ingresso, che raffigura il valore della loss function in funzione del numero di epoche, viene mostrata in Fig.4.1.

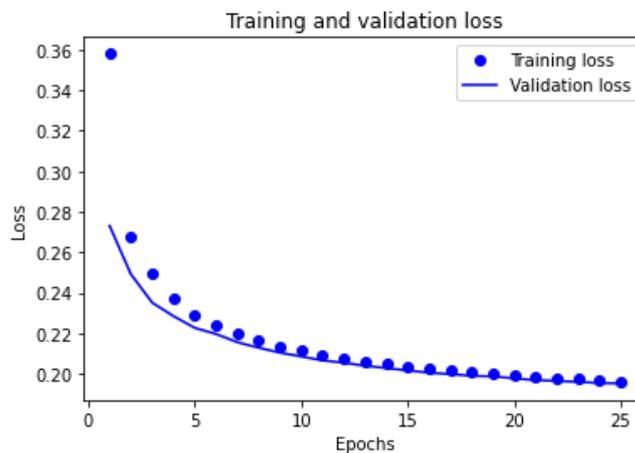


Figura 4.1: Curva di allenamento della rete allenata con dati IQ in ingresso.

I risultati relativi alle metriche di valutazione vengono riportati in Tab.4.1, mentre un esempio di immagine simulata predetta contenuta nel test set viene mostrato in Fig.4.2. I numeri contenuti in tabella sono valori medi \pm una deviazione standard.

Metrics	DAS	DNN (RF)	DNN (IQ)
CR (dB)	$-10,1 \pm 4,20$	$-54,18 \pm 8,55$	$-57,79 \pm 10,51$
SNR	$7,09 \pm 0,87$	$8,41 \pm 0,97$	$9,79 \pm 1,39$
gCNR	$0,94 \pm 0,09$	$0,99 \pm 0,01$	$0,99 \pm 0,01$
PSNR (dB)	/	$26,52 \pm 0,32$	$24,68 \pm 0,38$

DSC	/	$0,99 \pm 0,01$	$0,99 \pm 0,01$
-----	---	-----------------	-----------------

Tabella 4.1: Confronto tra i valori medi delle metriche di valutazione sulle immagini simulate DAS e le immagini predette DNN allenate con i dati RF e IQ.

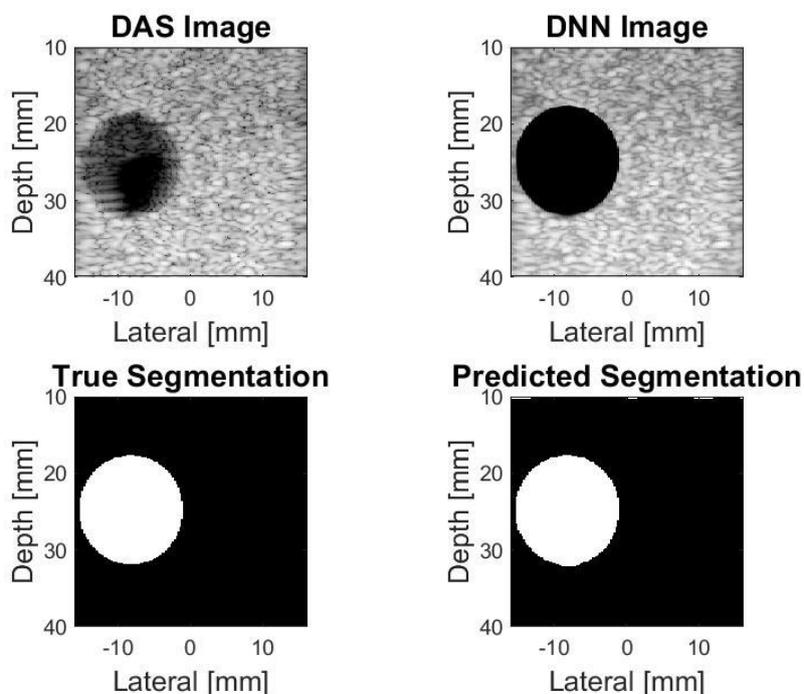


Figura 4.2: Confronto tra immagine simulata del test set e immagine predetta dalla rete (prima riga). Nella seconda riga il confronto viene effettuato tra la maschera di segmentazione reale e quella predetta.

Come visibile dal confronto tra le immagini predette e reali, la rete è in grado di predire un'immagine di qualità superiore eliminando il rumore acustico all'interno della ciste anecogena, testimoniato dal notevole aumento del valore del contrasto (da $-10,1 \pm 4,20$ dB a $-57,79 \pm 10,51$) Grande accuratezza è presente anche nella predizione della maschera di segmentazione, con un DSC medio pari a $0,99 \pm 0,01$. Non si osserva una sostanziale differenza nell'utilizzo dei dati RF oppure IQ in input alla rete.

Si effettua anche un ulteriore test sulla rete per verificare se essa possa predire anche la presenza di due cisti anecogene, sebbene sia stata allenata esclusivamente su immagini con cisti singole. Si utilizza un piccolo test set composto da 30 immagini contenenti due cisti anecogene, con raggi compresi tra 2 e 6 mm. Si osserva dalla Fig.4.3 che la rete è in grado di ricostruire e segmentare correttamente anche questa tipologia di immagini.

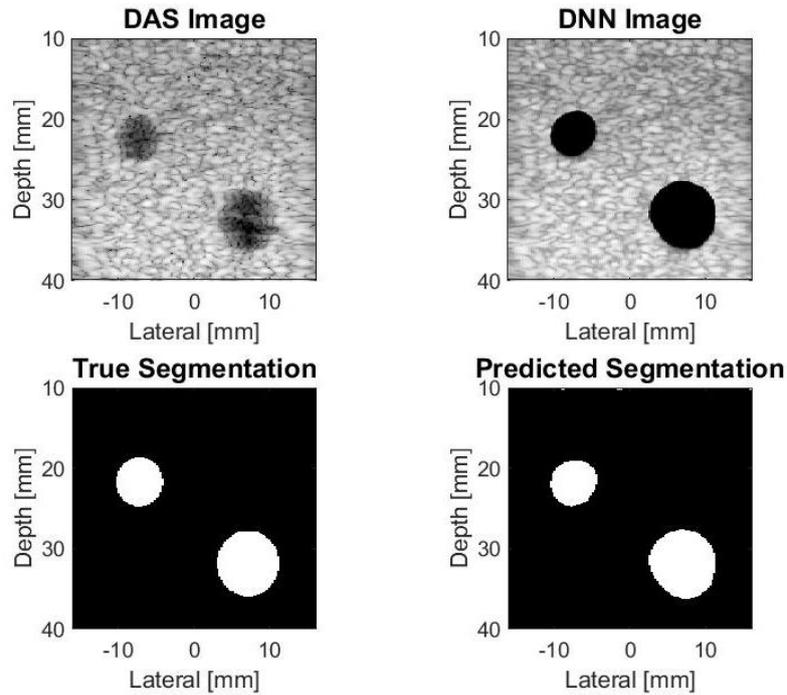


Figura 4.3: Predizione di un'immagine con due cisti anecogene.

La rete viene testata anche su immagini con cisti iperecogene, non presenti nel dataset. Dato che la rete è addestrata per ricostruire e segmentare immagini con cisti anecogene, in questo caso il modello predice un'immagine errata, riconoscendo il tessuto circostante alla ciste come

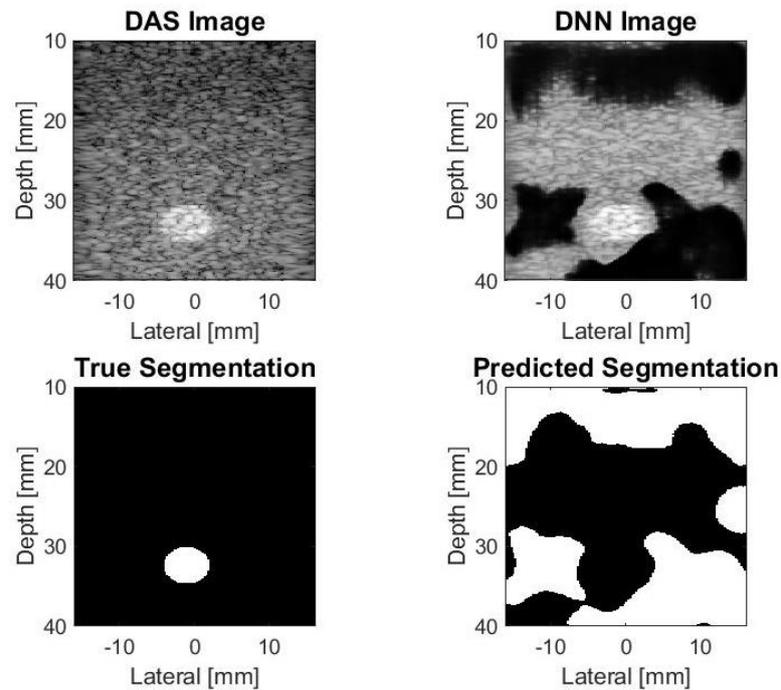


Figura 4.4: Predizione di un'immagine con ciste iperecogene. La rete riconosce erroneamente il tessuto circostante alla ciste come anecogeno e quindi da segmentare.

la porzione anecogena da segmentare (Fig.4.4). Questo giustifica l'introduzione di cisti iperecogene nel dataset della prova successiva.

4.1.2 Dataset con cisti anecogene e iperecogene

4.1.2.1 Immagini simulate

I risultati dei confronti tra le metriche valutate sulle immagini simulate sono visibili in Tab.4.2, mentre in Fig.4.5 viene mostrato un esempio di immagine con ciste iperecogene predetta dalla rete.

Metrics	Simulated anechoic			Simulated hyperechoic		
	DAS	DNN (RF)	DNN (IQ)	DAS	DNN (RF)	DNN (IQ)
CR (dB)	$-9,97 \pm 0,49$	$-88,11 \pm 20,71$	$-77,11 \pm 16,17$	$4,84 \pm 0,70$	$6,89 \pm 0,68$	$6,85 \pm 0,73$
SNR	$7,15 \pm 0,86$	$13,68 \pm 3,03$	$10,06 \pm 1,41$	$4,57 \pm 0,68$	$17,1 \pm 3,03$	$12,21 \pm 2,31$
gCNR	$0,93 \pm 0,10$	$0,99 \pm 0,01$	$0,99 \pm 0,01$	$0,92 \pm 0,05$	$0,99 \pm 0,01$	$0,99 \pm 0,01$
PSNR (dB)	/	$21,96 \pm 0,38$	$23,25 \pm 0,43$	/	$21,66 \pm 0,56$	$22,59 \pm 0,62$
DSC	/	$0,99 \pm 0,01$	$0,99 \pm 0,01$	/	$0,96 \pm 0,03$	$0,96 \pm 0,03$

Tabella 4.2: Confronto tra le metriche di valutazione per le immagini simulate.

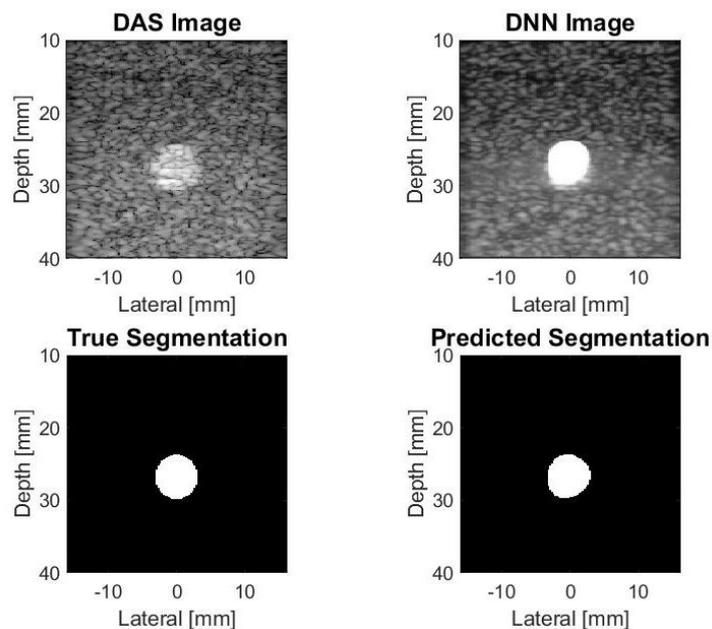


Figura 4.5: Predizione di un'immagine con ciste iperecogene.

Anche in questo caso si registrano notevoli miglioramenti rispetto all'utilizzo del beamforming classico, in termini di contrasto e di gCNR sia per le immagini con cisti anecoiche ($-77,11 \pm 16,17$ dB vs $9,97 \pm 0,49$ dB e $0,99 \pm 0,01$ vs $0,93 \pm 0,10$), sia iperecoiche ($6,85 \pm 0,73$ dB vs $4,84 \pm 0,70$ dB e $0,99 \pm 0,01$ vs $0,92 \pm 0,05$). Anche i valori di DSC risultano molto elevati, in particolare modo per le maschere di segmentazione di cisti anecoiche ($0,99 \pm 0,01$).

4.1.2.2 Immagini sperimentali

Si valuta la rete sulle immagini sperimentali descritte in 3.2.2, ossia 21 immagini acquisite su un phantom contenente target di differente ecogenicità (Tab.4.3), e 18 immagini di carotide umana acquisite su due volontari (Tab.4.4).

Metrics	Phantom images					
	Hypo-Anechoic targets			Hyperechoic targets		
	DAS	DNN (RF)	DNN (IQ)	DAS	DNN (RF)	DNN (IQ)
CR (dB)	-3.76 ± 2.62	-1.24 ± 1.62	6.62 ± 12.01	2.43 ± 0.46	2.85 ± 0.76	0.83 ± 0.42
SNR	6.94 ± 0.79	15.16 ± 3.16	1.24 ± 0.55	6.37 ± 0.82	13.56 ± 2.43	11.08 ± 2.46
gCNR	0.69 ± 0.26	0.51 ± 0.33	0.21 ± 0.34	0.76 ± 0.04	0.88 ± 0.12	0.45 ± 0.11
PSNR (dB)	/	14.96 ± 0.45	9.40 ± 0.44	/	16.27 ± 0.60	12.49 ± 0.45
DSC	/	0.00 ± 0.00	0.23 ± 0.07	/	0.29 ± 0.20	0.00 ± 0.00

Tabella 4.3: Confronto tra le metriche di valutazione per le immagini del phantom, suddivise tra target ipo-anechoici e target iperecoici.

Metrics	Carotid images		
	DAS	DNN (RF)	DNN (IQ)
CR (dB)	-4.80 ± 2.40	-3.62 ± 3.23	-3.52 ± 1.67
SNR	4.20 ± 0.94	9.42 ± 7.81	5.04 ± 1.02
gCNR	0.67 ± 0.19	0.69 ± 0.26	0.73 ± 0.21
PSNR (dB)	/	15.52 ± 1.14	16.47 ± 0.34
DSC	/	0.019 ± 0.024	0.131 ± 0.059

Tabella 4.4: Confronto tra le metriche di valutazione per le immagini in-vivo di carotide.

Si può notare come le immagini phantom predette dalla rete non siano soddisfacenti in termini di qualità rispetto a quelle DAS. La rete ricostruisce erroneamente le immagini con target anecoici, in particolar modo utilizzando i dati IQ in ingresso. Gli unici risultati accettabili in termini di contrasto (2.85 ± 0.76 dB vs 2.43 ± 0.46 dB) e gCNR (0.88 ± 0.12 vs 0.76 ± 0.04) si ottengono per i target iperecoici, adoperando i dati RF in input. Tutte le reti non riescono a fornire delle maschere di segmentazione corrette, come testimoniato dai bassi valori di DSC.

Nel caso delle immagini *in-vivo* di carotide, la qualità risulta leggermente inferiore rispetto alle immagini DAS, con eccezione per i valori di gCNR (0.73 ± 0.21 vs 0.67 ± 0.19). Anche in questo caso la rete non riesce a segmentare l'immagine in maniera corretta.

4.1.3 Dataset con cisti anecogene, iperecogene e doppie

4.1.3.1 Immagini simulate

Nella training set di questa prova vengono introdotte immagini con doppia ciste, in quanto la rete allenata con il dataset precedente non era in grado di riconoscere immagini in cui fossero presenti contemporaneamente due cisti di ecogenicità differente. I risultati dei confronti tra le metriche valutate sulle immagini simulate sono visibili in Tab.4.5, mentre in Fig.4.6 viene mostrato un esempio di immagine con doppia ciste predetta dalla DNN.

Metrics	Simulated anechoic			Simulated hyperechoic		
	DAS	DNN (RF)	DNN (IQ)	DAS	DNN (RF)	DNN (IQ)
CR (dB)	$-9,97 \pm 0,49$	$-88,71 \pm 17,99$	$-80,13 \pm 15,29$	$4,84 \pm 0,70$	$7,05 \pm 0,80$	$6,92 \pm 0,82$
SNR	$7,15 \pm 0,86$	$8,54 \pm 1,07$	$8,14 \pm 0,99$	$4,57 \pm 0,68$	$8,21 \pm 1,56$	$6,58 \pm 0,86$
gCNR	$0,93 \pm 0,10$	$0,99 \pm 0,01$	$0,99 \pm 0,01$	$0,92 \pm 0,05$	$0,99 \pm 0,01$	$0,99 \pm 0,01$
PSNR (dB)	/	$25,16 \pm 0,50$	$26,89 \pm 0,56$	/	$24,59 \pm 0,78$	$25,75 \pm 1,17$
DSC	/	$0,99 \pm 0,01$	$0,99 \pm 0,01$	/	$0,98 \pm 0,03$	$0,98 \pm 0,02$

Tabella 4.5: Confronto tra le metriche di valutazione per le immagini simulate.

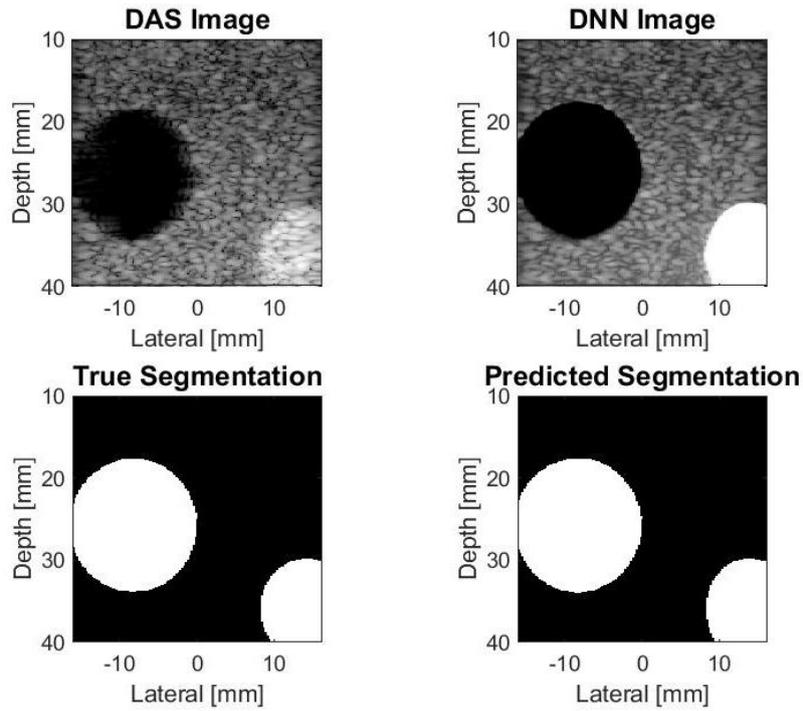


Figura 4.6: Predizione di un'immagine con due cisti di differente ecogenicità.

Anche in questo caso, come già visto con la prova descritta in 4.1.2.1, la rete ottiene buoni risultati con il test set composto da immagini simulate, con differenze non significative tra l'utilizzo in input dei dati RF o IQ.

4.1.3.2 Immagini sperimentali

Metrics	Phantom images					
	Hypo-Anechoic phantom targets			Hyperechoic phantom targets		
	DAS	DNN (RF)	DNN (IQ)	DAS	DNN (RF)	DNN (IQ)
CR (dB)	-3.76 ± 2.62	-4.55 ± 7.54	-27.12 ± 11.98	2.43 ± 0.46	1.34 ± 0.42	31.30 ± 7.66
SNR	6.94 ± 0.79	7.14 ± 1.02	1.44 ± 0.39	6.37 ± 0.82	9.66 ± 1.47	1.64 ± 0.21
gCNR	0.69 ± 0.26	0.54 ± 0.28	0.82 ± 0.22	0.76 ± 0.04	0.56 ± 0.15	0.86 ± 0.07
PSNR (dB)	/	16.00 ± 0.52	7.49 ± 0.25	/	15.75 ± 0.61	8.09 ± 0.35
DSC	/	0.00 ± 0.01	0.08 ± 0.06	/	0.03 ± 0.04	0.01 ± 0.01

Tabella 4.6: Confronto tra le metriche di valutazione per le immagini del phantom.

Metrics	Carotid images		
	DAS	DNN (RF)	DNN (IQ)
CR (dB)	-4.80 ± 2.40	-4.97 ± 6.60	-6.31 ± 7.09
SNR	4.20 ± 0.94	15.15 ± 25.15	3.64 ± 1.71
gCNR	0.67 ± 0.19	0.70 ± 0.19	0.71 ± 0.18
PSNR (dB)	/	14.51 ± 1.07	12.98 ± 0.61
DSC	/	0.18 ± 0.14	0.25 ± 0.13

Tabella 4.7: Confronto tra le metriche di valutazione per le immagini in-vivo di carotide.

I risultati relativi alle immagini phantom non sono soddisfacenti neanche in questa prova (Tab.4.6). Gli alti valori di contrasto nelle prove IQ sono dovuti in realtà a ricostruzioni incorrette dell'immagine (testimoniato dai bassi valori di PSNR), con il tessuto circostante alla ciste che viene predetto erroneamente come anecogeno.

I risultati ottenuti con le immagini *in-vivo* di carotide (Tab.4.7) sono piuttosto simili a quelli della prova precedente (Tab.4.4), senza registrare notevoli miglioramenti nella predizione dell'immagine.

4.1.4 Dataset con cisti irregolari

Il dataset adoperato in questa prova è quello descritto in 3.2.1.4, in cui vengono introdotte cisti dalla forma irregolare e cisti ipoecogene. Come esposto in 3.3, a partire da questa prova si utilizzano tre differenti encoder per l'allenamento dell'U-Net, oltre all'impiego di immagini target con dynamic range automatico ($B\text{-mode}_{\text{auto}}$ e E_{auto}) che verranno confrontate con i modelli allenati con immagini a dynamic range standard ($B\text{-mode}_{60}$ e E_{60}).

4.1.4.1 Immagini simulate

I risultati dei confronti tra le metriche valutate sulle immagini simulate sono visibili in Tab.4.8, Tab.4.9 e Tab.4.10, in cui vengono riportati i risultati ottenuti con le reti ad encoder VGG-13, VGG-16 e VGG-19, rispettivamente. Sono stati divisi i risultati per ciascuna classe di ciste

(iperecogene, ipoecogene, ipoecoiche), in modo tale che i valori fossero confrontabili. Alcuni esempi di immagini predette dalla rete E_{60} e E_{auto} sono mostrati in Fig.4.7.

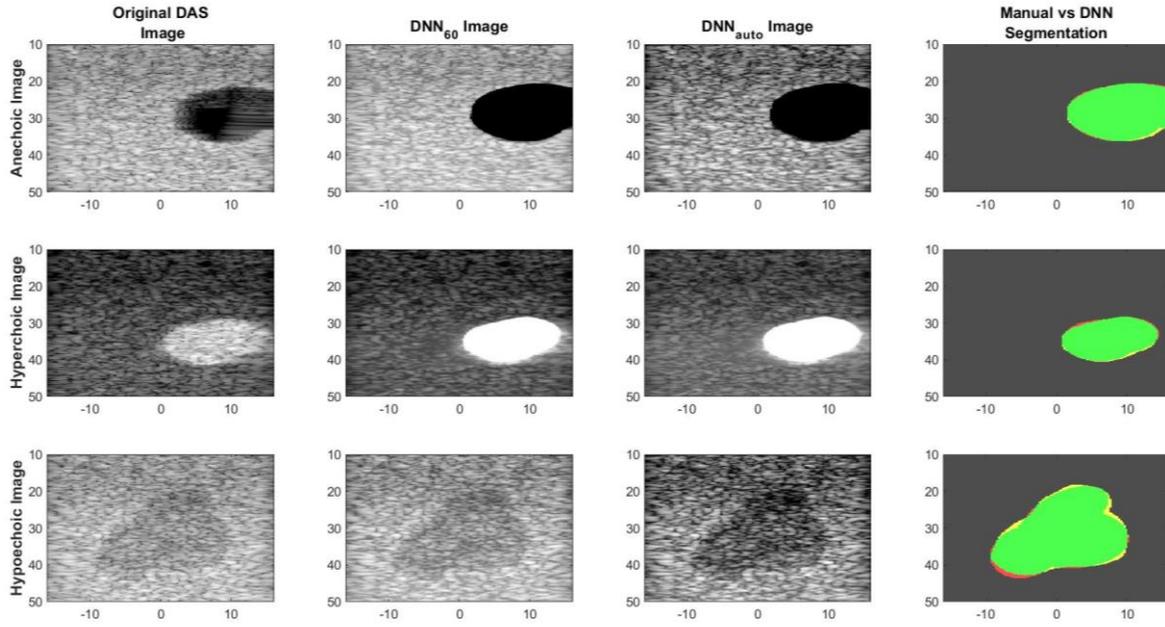


Figura 4.7: Esempi di immagini con ciste anecogena, iperecogena e ipoecogena (prima colonna) predette dalle reti allenate con immagini E_{60} (seconda colonna) e E_{auto} (terza colonna), con encoder VGG-13.

Metrics	VGG-13								
	Simulated anechoic			Simulated hyperechoic			Simulated hypoechoic		
	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀
CR (dB)	-11.04 ± 3.97	-10.97 ± 4.04	-80.52 ± 16.23	6.11 ± 1.45	6.20 ± 1.26	8.25 ± 1.32	-2.49 ± 1.18	-2.61 ± 1.30	-2.37 ± 1.09
SNR	6.79 ± 0.81	8.19 ± 1.50	8.02 ± 1.13	3.87 ± 0.82	6.60 ± 1.39	7.26 ± 1.50	6.84 ± 0.78	8.36 ± 1.46	8.08 ± 1.09
gCNR	0.95 ± 0.06	0.98 ± 0.04	0.99 ± 0.01	0.93 ± 0.06	0.98 ± 0.04	0.99 ± 0.01	0.62 ± 0.15	0.68 ± 0.17	0.65 ± 0.17
PSNR (dB)	/	25.16 ± 0.61	24.34 ± 1.09	/	24.80 ± 0.94	24.37 ± 1.16	/	25.47 ± 0.52	25.64 ± 1.22
DSC	/	0.97 ± 0.02	0.97 ± 0.02	/	0.97 ± 0.01	0.97 ± 0.01	/	0.95 ± 0.05	0.95 ± 0.04
Metrics	Simulated anechoic			Simulated hyperechoic			Simulated hypoechoic		
	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}

CR (dB)	-31.86± 13.71	-30.39 ± 10.97	-78.51 ± 13.04	5.24 ± 1.14	5.00 ± 1.02	7.47 ± 1.21	-7.33 ± 4.74	-7.35 ± 4.94	-11.61 ± 16.91
SNR	3.53 ± 0.83	4.06 ± 0.77	3.95 ± 0.77	4.71 ± 1.12	8.13 ± 2.46	8.39 ± 2.45	3.09 ± 0.52	3.58 ± 0.65	3.54 ± 0.64
gCNR	0.95 ± 0.07	0.98 ± 0.06	0.99 ± 0.04	0.93 ± 0.05	0.96 ± 0.05	0.99 ± 0.01	0.64 ± 0.15	0.66 ± 0.16	0.67 ± 0.17
PSNR (dB)	/	22.90 ± 1.22	22.49 ± 1.33	/	25.69 ± 1.39	24.84 ± 1.66	/	22.96 ± 0.71	22.92 ± 0.68
DSC	/	0.97 ± 0.02	0.97 ± 0.02	/	0.97 ± 0.02	0.97 ± 0.02	/	0.95 ± 0.04	0.95 ± 0.04

Tabella 4.8: Confronto tra le metriche di valutazione per le immagini simulate, predette dalla rete con encoder VGG-13. Si confrontano le immagini ottenute con beamforming classico a dynamic range standard e automatico (DAS_{60} e DAS_{auto}) con le reti allenate con immagini rinforzate (E_{60} e E_{auto}) e non rinforzate ($B-mode_{60}$ e $B-mode_{auto}$).

Metrics	VGG-16								
	Simulated anechoic			Simulated hyperechoic			Simulated hypoechoic		
	DAS_{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS_{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS_{auto}	DNN B-mode _{auto}	DNN E _{auto}
CR (dB)	-31.86± 13.71	-30.86 ± 9.82	-80.91 ± 9.35	5.24 ± 1.14	5.07 ± 1.11	7.31 ± 1.09	-7.33 ± 4.74	-7.34 ± 5.20	-10.90 ± 15.55
SNR	3.53 ± 0.83	4.03 ± 0.73	3.97 ± 0.71	4.71 ± 1.12	7.87 ± 2.11	8.38 ± 1.86	3.09 ± 0.52	3.56 ± 0.58	3.57 ± 0.56
gCNR	0.95 ± 0.07	0.98 ± 0.04	0.99 ± 0.03	0.93 ± 0.05	0.96 ± 0.05	0.99 ± 0.01	0.64 ± 0.15	0.66 ± 0.16	0.67 ± 0.18
PSNR (dB)	/	23.29± 1.43	22.70 ± 1.43	/	25.89 ± 1.53	25.21 ± 1.58	/	23.49 ± 0.78	23.23 ± 0.75
DSC	/	0.97 ± 0.02	0.97 ± 0.02	/	0.97 ± 0.02	0.97 ± 0.03	/	0.95 ± 0.04	0.95 ± 0.04

Tabella 4.9: Confronto tra le metriche di valutazione per le immagini simulate, predette dalla rete con encoder VGG-16.

Metrics	VGG-19								
	Simulated anechoic			Simulated hyperechoic			Simulated hypoechoic		
	DAS_{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS_{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS_{auto}	DNN B-mode _{auto}	DNN E _{auto}
CR (dB)	-31.86± 13.71	-30.87 ± 11.11	-95.10 ± 12.50	5.24 ± 1.14	4.98 ± 1.08	7.34 ± 1.20	-7.33 ± 4.74	-7.14 ± 4.78	-13.31 ± 20.35

SNR	3.53 ± 0.83	4.09 ± 0.82	4.02 ± 0.84	4.71 ± 1.12	8.29 ± 2.41	8.26 ± 2.21	3.09 ± 0.52	3.56 ± 0.61	3.45 ± 0.60
gCNR	0.95 ± 0.07	0.98 ± 0.06	0.99 ± 0.03	0.93 ± 0.05	0.96 ± 0.057	0.99 ± 0.01	0.64 ± 0.15	0.65 ± 0.16	0.68 ± 0.17
PSNR (dB)	/	23.60 ± 1.13	23.57 ± 1.28	/	26.43 ± 0.92	26.02 ± 1.15	/	23.61 ± 0.59	24.00 ± 0.72
DSC	/	0.97 ± 0.03	0.97 ± 0.02	/	0.97 ± 0.03	0.97 ± 0.02	/	0.94 ± 0.08	0.95 ± 0.04

Tabella 4.10: Confronto tra le metriche di valutazione per le immagini simulate, predette dalla rete con encoder VGG-19.

Dal confronto tra i risultati ottenuti non emergono differenze significative tra i tre differenti encoder nella predizione di immagini simulate. Anche l'utilizzo di immagini target a dynamic range standard o automatico non porta a differenze tangibili nella qualità dell'immagine predetta, come visibile dal confronto dei valori di PSNR e DSC delle colonne relative a DNN E₆₀ e DNN E_{auto}.

4.1.4.2 Immagini sperimentali

I risultati dei confronti tra le metriche valutate sulle immagini phantom sono visibili in Tab.4.11, Tab.4.12 e Tab.4.13, in cui vengono riportati i risultati ottenuti con le reti ad encoder VGG-13, VGG-16 e VGG-19, rispettivamente. I risultati relativi alle immagini di carotide sono mostrati in Tab.4.14 e Tab.4.15. Un esempio di immagini sperimentali predette dalla rete è esposto in Fig.4.8.

Metrics	VGG-13					
	Hypo-Anechoic phantom targets			Hyperechoic phantom targets		
	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀
CR (dB)	-3.76 ± 2.62	-15.54 ± 17.02	-19.29 ± 15.51	2.43 ± 0.46	2.31 ± 0.75	1.92 ± 0.87
SNR	6.94 ± 0.79	3.95 ± 0.32	4.45 ± 0.49	6.37 ± 0.82	5.07 ± 0.56	6.06 ± 0.56
gCNR	0.69 ± 0.26	0.71 ± 0.29	0.78 ± 0.24	0.76 ± 0.04	0.53 ± 0.10	0.54 ± 0.19
PSNR (dB)	/	9.13 ± 0.30	10.05 ± 0.36	/	10.17 ± 0.34	11.43 ± 0.38
DSC	/	0.15 ± 0.04	0.22 ± 0.05	/	0.00 ± 0.00	0.00 ± 0.00
Metrics	Hypo-Anechoic phantom targets			Hyperechoic phantom targets		

	DAS_{auto}	DNN B-mode_{auto}	DNN E_{auto}	DAS_{auto}	DNN B-mode_{auto}	DNN E_{auto}
CR (dB)	-5.40 ± 3.96	-17.69 ± 19.35	-30.01 ± 19.78	2.88 ± 0.51	2.04 ± 0.45	12.34 ± 27.16
SNR	5.35 ± 0.75	4.14 ± 0.55	0.88 ± 0.49	5.25 ± 0.83	4.87 ± 0.73	0.53 ± 0.28
gCNR	0.70 ± 0.25	0.75 ± 0.25	0.16 ± 0.31	0.77 ± 0.046	0.47 ± 0.08	0.47 ± 0.33
PSNR (dB)	/	11.48 ± 0.36	8.31 ± 0.35	/	11.94 ± 0.34	8.93 ± 0.44
DSC	/	0.27 ± 0.05	0.12 ± 0.03	/	0.01 ± 0.02	0.05 ± 0.01

Tabella 4.11: Confronto tra le metriche di valutazione per le immagini del phantom, predette dalla rete con encoder VGG-13.

Metrics	VGG-16					
	Hypo-Anechoic phantom targets			Hyperechoic phantom targets		
	DAS_{auto}	DNN B-mode_{auto}	DNN E_{auto}	DAS_{auto}	DNN B-mode_{auto}	DNN E_{auto}
CR (dB)	-5.40 ± 3.96	-17.11 ± 15.37	-16.34 ± 8.28	2.88 ± 0.51	3.29 ± 1.34	30.13 ± 5.77
SNR	5.35 ± 0.75	2.73 ± 0.69	0.76 ± 0.28	5.25 ± 0.83	3.31 ± 0.65	0.47 ± 0.16
gCNR	0.70 ± 0.25	0.74 ± 0.25	0.01 ± 0.01	0.77 ± 0.046	0.55 ± 0.10	0.72 ± 0.29
PSNR (dB)	/	10.63 ± 0.33	8.49 ± 0.30	/	11.37 ± 0.36	8.67 ± 0.32
DSC	/	0.14 ± 0.03	0.14 ± 0.03	/	0.01 ± 0.01	0.02 ± 0.02

Tabella 4.12: Confronto tra le metriche di valutazione per le immagini del phantom, predette dalla rete con encoder VGG-16.

Metrics	VGG-19					
	Hypo-Anechoic phantom targets			Hyperechoic phantom targets		
	DAS_{auto}	DNN B-mode_{auto}	DNN E_{auto}	DAS_{auto}	DNN B-mode_{auto}	DNN E_{auto}
CR (dB)	-5.40 ± 3.96	-11.66 ± 9.34	-0.96 ± 2.16	2.88 ± 0.51	2.09 ± 0.89	3.81 ± 8.39
SNR	5.35 ± 0.75	4.38 ± 1.74	1.02 ± 0.25	5.25 ± 0.83	5.61 ± 0.98	0.67 ± 0.18
gCNR	0.70 ± 0.25	0.76 ± 0.25	0.00 ± 0.00	0.77 ± 0.05	0.52 ± 0.13	0.04 ± 0.06
PSNR (dB)	/	10.18 ± 0.28	7.47 ± 0.39	/	11.08 ± 0.34	7.98 ± 0.38
DSC	/	0.14 ± 0.03	0.13 ± 0.03	/	0.01 ± 0.01	0.08 ± 0.02

Tabella 4.13: Confronto tra le metriche di valutazione per le immagini del phantom, predette dalla rete con encoder VGG-19.

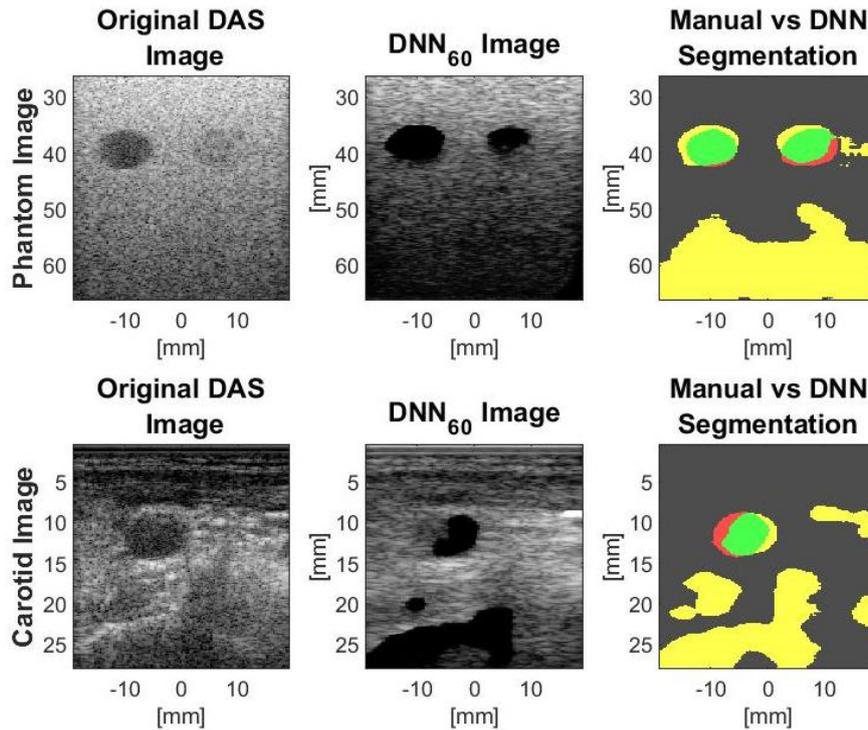


Figura 4.8: In figura vengono riportate un'immagine di phantom (prima riga) e un'immagine di carotide (seconda riga) predette dalla rete allenata con target E_{60} ed encoder VGG-13. Le porzioni in giallo visibili nel confronto tra le maschere (terza colonna) sono quelle over-segmentate dalla rete DNN.

Nel caso delle immagini phantom si osserva come non si ottengano miglioramenti apprezzabili variando la struttura dell'encoder. In particolare, la problematica principale risulta essere l'attenuazione del segnale proveniente dagli strati insonificati più profondi, che viene riconosciuto dalla rete come appartenente ad una struttura anecogena. Questi strati vengono dunque riconosciuti come strutture da segmentare, come visibile nell'immagine phantom di Fig.4.8, conducendo a bassi valori di DSC. Dato che queste porzioni vengono riconosciute erroneamente come anecogene, anche l'immagine predetta presenta zone in cui i valori dei pixel vengono posti vicino allo zero, portando a bassi valori di PSNR. Questo vale in modo particolare per le reti E_{60} e E_{auto} , allenate con la maschera rinforzata, che in questo caso non si rivela essere un vantaggio per la formazione di un'immagine di qualità superiore.

Metrics	Carotid images					
	VGG-13					
	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}
CR (dB)	-4.80 ± 2.40	-3.85 ± 3.00	-17.61 ± 22.28	-3.30 ± 1.75	-3.24 ± 3.10	-3.54 ± 2.53
SNR	4.20 ± 0.94	4.96 ± 0.92	4.97 ± 1.01	5.63 ± 1.12	4.83 ± 0.77	4.49 ± 0.87

gCNR	0.67 ± 0.19	0.57 ± 0.24	0.72 ± 0.26	0.67 ± 0.19	0.53 ± 0.22	0.52 ± 0.21
PSNR (dB)	/	16.50 ± 0.46	15.49 ± 0.27	/	16.26 ± 0.77	14.70 ± 0.79
DSC	/	0.18 ± 0.079	0.20 ± 0.06	/	0.26 ± 0.12	0.18 ± 0.05

Tabella 4.14: Confronto tra le metriche di valutazione per le immagini di carotide, predette dalla rete con encoder VGG-13.

Metrics	Carotid images				
	DAS _{auto}	VGG-16		VGG-19	
		DNN B-mode _{auto}	DNN E _{auto}	DNN B-mode _{auto}	DNN E _{auto}
CR (dB)	-3.30 ± 1.75	-2.92 ± 2.59	-12.35 ± 11.68	-5.16 ± 3.51	-32.71 ± 19.99
SNR	5.63 ± 1.12	4.69 ± 0.75	3.28 ± 1.64	4.47 ± 1.11	3.38 ± 1.86
gCNR	0.67 ± 0.19	0.49 ± 0.20	0.66 ± 0.21	0.64 ± 0.22	0.89 ± 0.15
PSNR (dB)	/	16.20 ± 0.72	12.24 ± 1.13	14.85 ± 0.67	11.48 ± 0.70
DSC	/	0.20 ± 0.07	0.17 ± 0.06	0.16 ± 0.09	0.19 ± 0.10

Tabella 4.15: Confronto tra le metriche di valutazione per le immagini di carotide, predette dalla rete con encoder VGG-16 e VGG-19.

Osservando le Tab.4.14 e Tab.4.15 relative ai risultati sulle immagini di carotide, si nota come l'architettura con encoder VGG-19 consenta di ottenere valori di gCNR e contrasto superiori rispetto all'immagine DAS_{auto} (0.89 ± 0.15 vs 0.67 ± 0.19 e -32.71 ± 19.99 dB vs -3.54 ± 2.53 dB). I valori DSC risultano bassi anche in questo caso, soprattutto per l'errata segmentazione delle porzioni profonde soggette ad attenuazione, di cui si è accennato nel precedente paragrafo e il cui effetto è visibile nell'immagine di carotide in Fig.4.8.

4.1.5 Dataset con cisti irregolari e simulazione del coefficiente di attenuazione

In questa prova il dataset adoperato è quello descritto in 3.2.1.5, in cui le immagini vengono simulate anche tenendo in considerazione il coefficiente di attenuazione introdotta dai tessuti. Questo viene introdotto per tentare di ottenere migliori risultati nella predizione di immagini sperimentali. Infatti, in 4.1.4.2 si è osservato come l'attenuazione presente veniva scambiata dalla rete per un oggetto anecoico presente nell'immagine, segmentando porzioni extra di tessuto nella maschera di segmentazione (Fig.4.8). Inoltre, sempre nell'ottica di migliorare le predizioni di immagini sperimentali, le maschere di rinforzo vengono applicate solo alle immagini con cisti anecoiche e non più iperecoiche.

4.1.5.1 Immagini simulate

I risultati sono visibili in Tab.4.16, Tab.4.17 e Tab.4.18, suddivisi tra gli encoder VGG-13, VGG-16 e VGG-19, rispettivamente. Alcuni esempi di immagini predette dalla rete sono mostrati in Fig.4.9.

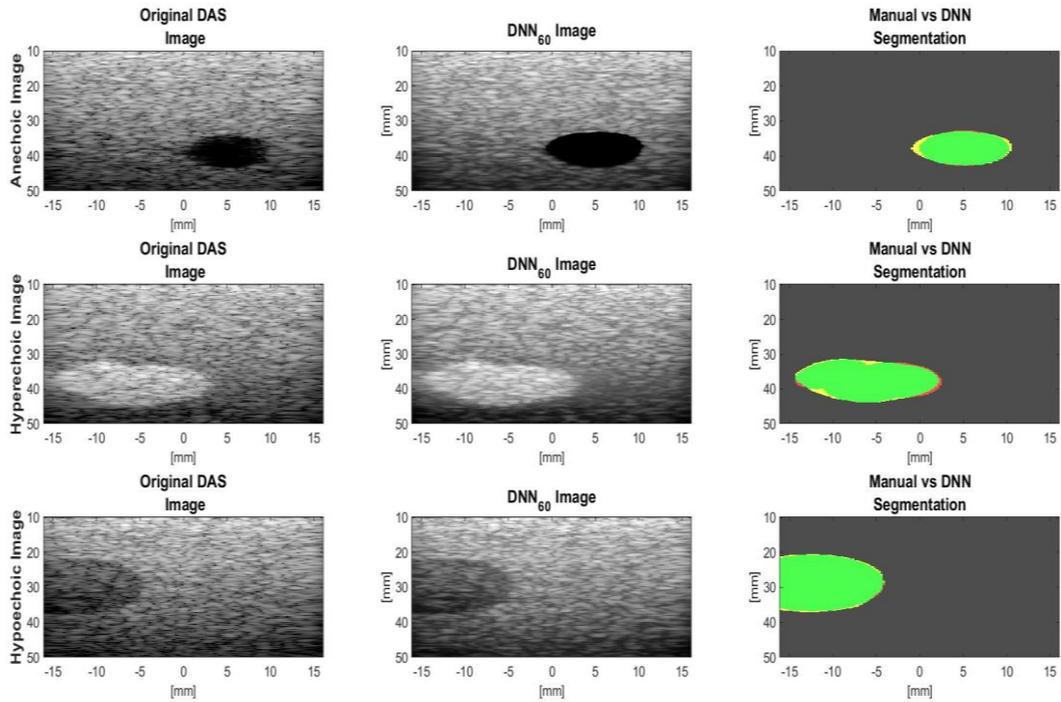


Figura 4.9: Esempi di immagini con ciste anecogena, iperecogena e ipoecogena (prima colonna) predette dalle reti allenate con immagini E_{60} (seconda colonna) con encoder VGG-13.

Metrics	VGG-13								
	Simulated anechoic			Simulated hyperechoic			Simulated hypoechoic		
	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀
CR (dB)	-16.70 ± 7.45	-15.67 ± 8.78	-92.70 ± 21.30	5.38 ± 1.04	5.00 ± 0.78	5.32 ± 0.94	-4.74 ± 3.01	-4.73 ± 4.674	-6.25 ± 7.83
SNR	4.37 ± 1.43	6.13 ± 1.34	5.73 ± 1.45	4.03 ± 0.87	9.06 ± 1.58	8.27 ± 1.36	4.27 ± 1.39	5.94 ± 1.42	5.54 ± 1.43
gCNR	0.90 ± 0.10	0.95 ± 0.10	0.99 ± 0.04	0.90 ± 0.05	0.98 ± 0.05	0.98 ± 0.04	0.60 ± 0.14	0.66 ± 0.19	0.70 ± 0.19
PSNR (dB)	/	22.95 ± 0.54	24.13 ± 0.77	/	22.29 ± 0.99	23.77 ± 0.89	/	22.77 ± 0.49	24.35 ± 0.75

DSC	/	0.82 ± 0.15	0.96 ± 0.04	/	0.85 ± 0.12	0.97 ± 0.02	/	0.80 ± 0.17	0.96 ± 0.03
Metrics	Simulated anechoic			Simulated hyperechoic			Simulated hypoechoic		
	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}
CR (dB)	-17.72 ± 9.14	-15.18 ± 5.50	-80.70 ± 17.22	6.52 ± 2.62	5.78 ± 1.77	5.91 ± 2.13	-5.71 ± 4.80	-4.17 ± 2.65	-7.59 ± 10.60
SNR	4.43 ± 1.69	6.80 ± 1.43	5.77 ± 1.77	3.72 ± 1.76	9.02 ± 3.20	7.45 ± 2.67	4.03 ± 1.37	6.29 ± 1.13	5.10 ± 1.54
gCNR	0.90 ± 0.09	0.97 ± 0.07	0.99 ± 0.05	0.90 ± 0.05	0.99 ± 0.03	0.97 ± 0.042	0.61 ± 0.14	0.71 ± 0.20	0.69 ± 0.19
PSNR (dB)	/	22.51 ± 0.83	24.00 ± 1.03	/	21.83 ± 1.23	23.49 ± 1.49	/	21.97 ± 0.84	24.38 ± 0.58
DSC	/	0.92 ± 0.09	0.95 ± 0.09	/	0.96 ± 0.03	0.97 ± 0.03	/	0.83 ± 0.20	0.95 ± 0.05

Tabella 4.16: Confronto tra le metriche di valutazione per le immagini simulate, predette dalla rete con encoder VGG-13.

Metrics	VGG-16								
	Simulated anechoic			Simulated hyperechoic			Simulated hypoechoic		
	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀
CR (dB)	-16.70 ± 7.45	-17.87 ± 8.96	-92.75 ± 23.46	5.38 ± 1.04	5.03 ± 0.78	5.16 ± 0.79	-4.74 ± 3.01	-5.08 ± 3.86	-6.31 ± 8.92
SNR	4.37 ± 1.43	5.73 ± 1.53	5.71 ± 1.49	4.03 ± 0.87	8.53 ± 1.41	8.64 ± 1.51	4.27 ± 1.39	5.56 ± 1.51	5.51 ± 1.51
gCNR	0.90 ± 0.10	0.97 ± 0.08	0.99 ± 0.05	0.90 ± 0.05	0.98 ± 0.05	0.98 ± 0.05	0.60 ± 0.14	0.70 ± 0.18	0.69 ± 0.19
PSNR (dB)	/	24.69 ± 0.65	24.49 ± 0.76	/	23.87 ± 0.86	23.75 ± 1.15	/	24.56 ± 0.58	24.65 ± 0.63
DSC	/	0.96 ± 0.07	0.97 ± 0.03	/	0.97 ± 0.02	0.97 ± 0.02	/	0.96 ± 0.04	0.95 ± 0.06
Metrics	Simulated anechoic			Simulated hyperechoic			Simulated hypoechoic		
	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}
CR (dB)	-17.72 ± 9.14	-19.78 ± 11.60	-74.49 ± 19.03	6.52 ± 2.62	5.85 ± 2.04	6.38 ± 2.41	-5.71 ± 4.80	-6.40 ± 5.79	-7.74 ± 9.39

SNR	4.43 ± 1.69	5.48 ± 1.64	5.56 ± 1.83	3.72 ± 1.76	7.36 ± 2.58	7.04 ± 2.81	4.03 ± 1.37	4.85 ± 1.49	4.91 ± 1.67
gCNR	0.90 ± 0.09	0.96 ± 0.08	0.99 ± 0.05	0.90 ± 0.05	0.97 ± 0.04	0.98 ± 0.04	0.61 ± 0.14	0.67 ± 0.18	0.71 ± 0.19
PSNR (dB)	/	24.48 ± 1.21	24.13 ± 1.17	/	23.74 ± 1.57	23.55 ± 1.45	/	24.76 ± 0.55	24.56 ± 0.47
DSC	/	0.96 ± 0.07	0.96 ± 0.06	/	0.97 ± 0.02	0.97 ± 0.02	/	0.96 ± 0.04	0.96 ± 0.04

Tabella 4.17: Confronto tra le metriche di valutazione per le immagini simulate, predette dalla rete con encoder VGG-16.

Metrics	VGG-19								
	Simulated anechoic			Simulated hyperechoic			Simulated hypoechoic		
	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀
CR (dB)	-16.70 ± 7.45	-17.64 ± 8.89	-82.08 ± 19.28	5.38 ± 1.04	5.23 ± 0.73	5.23 ± 0.77	-4.74 ± 3.01	-5.15 ± 3.86	-6.09 ± 7.24
SNR	4.37 ± 1.43	5.79 ± 1.53	6.05 ± 1.49	4.03 ± 0.87	8.72 ± 1.52	8.82 ± 1.47	4.27 ± 1.39	5.62 ± 1.53	5.84 ± 1.50
gCNR	0.90 ± 0.10	0.96 ± 0.08	0.99 ± 0.04	0.90 ± 0.05	0.98 ± 0.05	0.99 ± 0.04	0.60 ± 0.14	0.70 ± 0.18	0.72 ± 0.18
PSNR (dB)	/	24.65 ± 0.60	23.95 ± 0.69	/	23.95 ± 0.79	23.68 ± 0.53	/	24.53 ± 0.50	24.06 ± 0.48
DSC	/	0.96 ± 0.06	0.96 ± 0.03	/	0.97 ± 0.02	0.97 ± 0.02	/	0.96 ± 0.03	0.95 ± 0.06
Metrics	Simulated anechoic			Simulated hyperechoic			Simulated hypoechoic		
	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}
CR (dB)	-17.72 ± 9.14	-20.04 ± 11.60	-77.74 ± 26.83	6.52 ± 2.62	6.46 ± 2.41	6.39 ± 2.50	-5.71 ± 4.80	-6.97 ± 6.32	-7.11 ± 6.64
SNR	4.43 ± 1.69	5.55 ± 1.84	5.63 ± 1.68	3.72 ± 1.76	7.21 ± 3.07	7.35 ± 3.11	4.03 ± 1.37	4.91 ± 1.69	5.07 ± 1.55
gCNR	0.90 ± 0.09	0.96 ± 0.09	0.99 ± 0.06	0.90 ± 0.05	0.98 ± 0.04	0.98 ± 0.04	0.61 ± 0.14	0.70 ± 0.18	0.72 ± 0.18
PSNR (dB)	/	24.68 ± 0.93	24.42 ± 0.95	/	23.89 ± 1.18	23.93 ± 1.22	/	24.67 ± 0.42	24.61 ± 0.59

DSC	/	0.96 ± 0.05	0.96 ± 0.04	/	0.97 ± 0.02	0.97 ± 0.02	/	0.95 ± 0.03	0.96 ± 0.06
-----	---	-------------	-------------	---	-------------	-------------	---	-------------	-------------

Tabella 4.18: Confronto tra le metriche di valutazione per le immagini simulate, predette dalla rete con encoder VGG-19.

Come visibile in Tab.4.16, anche in questa prova l'uso della maschera di rinforzo per le immagini anecoiche consente di ottenere migliori risultati rispetto alle immagini DAS₆₀, in termini di contrasto (-82.08 ± 19.28 dB vs -16.70 ± 7.45 dB) e gCNR (0.99 ± 0.04 vs 0.90 ± 0.10) con encoder VGG-13 (miglioramenti presenti anche con encoder VGG-16 e VGG-19).

Per le cisti ipoeoiche e iperecoiche, i parametri ottenuti dalle reti DNN B-mode e DNN E sono molto simili alle immagini originali DAS, in quanto questa tipologia di cisti non adopera la maschera di rinforzo in fase di allenamento. Come in precedenza, non si osservano sostanziali differenze nei risultati ottenuti con i tre differenti encoder sul test set di immagini simulate, così come per l'uso di target con dR standard o automatico.

4.1.5.2 Immagini sperimentali

I risultati dei confronti tra le metriche valutate sulle immagini phantom sono visibili in Tab.4.19, Tab.4.20 e Tab.4.21.

Metrics	VGG-13					
	Hypo-Anechoic phantom targets			Hyperechoic phantom targets		
	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀
CR (dB)	-3.76 ± 2.62	-3.65 ± 2.13	-7.41 ± 11.35	2.43 ± 0.46	2.09 ± 0.31	3.17 ± 1.11
SNR	6.94 ± 0.79	7.59 ± 1.07	5.84 ± 0.81	6.37 ± 0.82	10.55 ± 1.92	7.34 ± 0.75
gCNR	0.69 ± 0.26	0.68 ± 0.25	0.64 ± 0.27	0.76 ± 0.04	0.81 ± 0.05	0.53 ± 0.18
PSNR (dB)	/	14.39 ± 0.38	13.59 ± 0.36	/	15.14 ± 0.47	14.51 ± 0.44
DSC	/	0.14 ± 0.12	0.22 ± 0.23	/	0.24 ± 0.06	0.68 ± 0.17
Metrics	Hypo-Anechoic phantom targets			Hyperechoic phantom targets		
	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}
CR (dB)	-5.40 ± 3.96	-7.43 ± 6.17	-2.50 ± 1.90	2.88 ± 0.51	1.58 ± 0.23	1.36 ± 0.30

SNR	5.35 ± 0.75	7.16 ± 0.81	7.85 ± 0.87	5.25 ± 0.83	11.74 ± 2.52	9.54 ± 0.94
gCNR	0.70 ± 0.25	0.75 ± 0.25	0.57 ± 0.31	0.77 ± 0.05	0.59 ± 0.23	0.40 ± 0.18
PSNR (dB)	/	17.15 ± 0.32	16.97 ± 0.33	/	18.12 ± 0.27	17.91 ± 0.23
DSC	/	0.19 ± 0.21	0.11 ± 0.07	/	0.62 ± 0.12	0.25 ± 0.03

Tabella 4.19: Confronto tra le metriche di valutazione per le immagini del phantom, predette dalla rete con encoder VGG-13.

Metrics	VGG-16					
	Hypo-Anechoic phantom targets			Hyperechoic phantom targets		
	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀
CR (dB)	-3.76 ± 2.62	-5.13 ± 2.12	-12.42 ± 19.52	2.43 ± 0.46	2.73 ± 0.30	2.75 ± 0.30
SNR	6.94 ± 0.79	7.78 ± 0.95	7.22 ± 0.69	6.37 ± 0.82	8.61 ± 1.30	8.87 ± 0.66
gCNR	0.69 ± 0.26	0.63 ± 0.28	0.70 ± 0.29	0.76 ± 0.04	0.55 ± 0.20	0.69 ± 0.13
PSNR (dB)	/	14.78 ± 0.41	14.96 ± 0.43	/	15.20 ± 0.56	15.90 ± 0.61
DSC	/	0.24 ± 0.16	0.43 ± 0.26	/	0.26 ± 0.14	0.79 ± 0.13
Metrics	Hypo-Anechoic phantom targets			Hyperechoic phantom targets		
	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}
	CR (dB)	-5.40 ± 3.96	-5.78 ± 4.39	-8.13 ± 8.53	2.88 ± 0.51	1.91 ± 0.13
SNR	5.35 ± 0.75	6.58 ± 1.02	9.68 ± 1.96	5.25 ± 0.83	7.84 ± 0.78	11.49 ± 1.28
gCNR	0.70 ± 0.25	0.63 ± 0.28	0.66 ± 0.31	0.77 ± 0.05	0.47 ± 0.24	0.63 ± 0.12
PSNR (dB)	/	16.94 ± 0.32	16.83 ± 0.48	/	17.63 ± 0.19	17.84 ± 0.24
DSC	/	0.15 ± 0.13	0.10 ± 0.15	/	0.45 ± 0.10	0.29 ± 0.26

Tabella 4.20: Confronto tra le metriche di valutazione per le immagini del phantom, predette dalla rete con encoder VGG-16.

Metrics	VGG-19					
	Hypo-Anechoic phantom targets			Hyperechoic phantom targets		
	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀
CR (dB)	-3.76 ± 2.62	-3.96 ± 2.35	-14.66 ± 15.78	2.43 ± 0.46	2.64 ± 0.25	2.24 ± 0.22
SNR	6.94 ± 0.79	7.81 ± 0.56	8.11 ± 0.67	6.37 ± 0.82	7.84 ± 1.21	8.48 ± 1.31

gCNR	0.69 ± 0.26	0.72 ± 0.24	0.69 ± 0.33	0.76 ± 0.04	0.70 ± 0.11	0.67 ± 0.11
PSNR (dB)	/	15.81 ± 0.43	15.55 ± 0.55	/	16.46 ± 0.52	15.74 ± 0.48
DSC	/	0.60 ± 0.11	0.35 ± 0.17	/	0.86 ± 0.02	0.78 ± 0.11
Metrics	Hypo-Anechoic phantom targets			Hyperechoic phantom targets		
	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}
CR (dB)	-5.40 ± 3.96	-3.49 ± 2.25	-31.72 ± 33.03	2.88 ± 0.51	1.84 ± 0.19	1.55 ± 0.24
SNR	5.35 ± 0.75	8.55 ± 0.62	9.40 ± 0.77	5.25 ± 0.83	9.54 ± 1.10	10.74 ± 1.88
gCNR	0.70 ± 0.25	0.71 ± 0.28	0.71 ± 0.32	0.77 ± 0.05	0.60 ± 0.13	0.65 ± 0.14
PSNR (dB)	/	17.72 ± 0.11	17.18 ± 0.80	/	18.21 ± 0.23	18.16 ± 0.43
DSC	/	0.52 ± 0.14	0.54 ± 0.18	/	0.79 ± 0.12	0.76 ± 0.14

Tabella 4.21: Confronto tra le metriche di valutazione per le immagini del phantom, predette dalla rete con encoder VGG-19.

Al contrario delle prove precedenti, in questo caso si osserva un netto miglioramento dei parametri relativi alle predizioni delle immagini phantom. In particolare, i risultati migliori si raggiungono con l'utilizzo dell'encoder VGG-19 e target E_{auto}, specialmente per il DSC delle immagini con cisti ipo-anechoiche (0.54 ± 0.18 vs 0.11 ± 0.07 con encoder VGG-13) e iperechoiche (0.76 ± 0.14 vs 0.25 ± 0.03 con encoder VGG-13). In Fig.4.10 si riportano due immagini phantom e le rispettive predizioni della rete DNN.

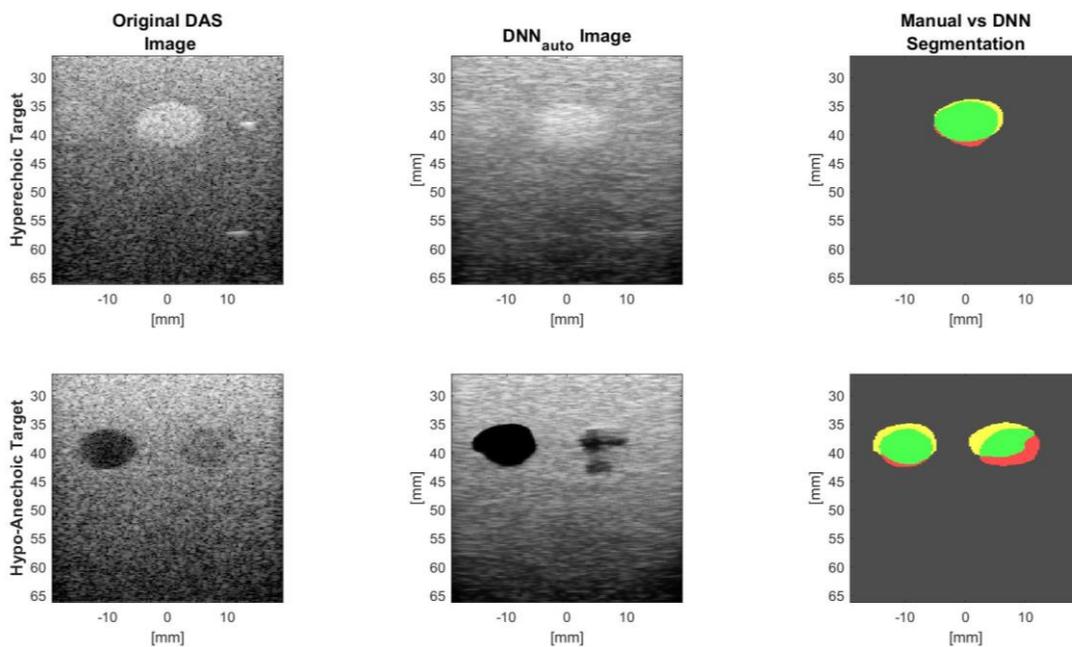


Figura 4.10: In figura vengono riportati due esempi di immagini phantom predetti dalla rete E_{auto} con encoder VGG-19.

I risultati dei confronti tra le metriche valutate sulle immagini *in-vivo* di carotide sono visibili in Tab.4.22.

Metrics	Carotid images					
	VGG-13					
	DAS ₆₀	DNN B-mode ₆₀	DNN E ₆₀	DAS _{auto}	DNN B-mode _{auto}	DNN E _{auto}
CR (dB)	-4.43 ± 1.65	-2.53 ± 2.07	-3.60 ± 2.49	-2.15 ± 0.83	-2.72 ± 1.84	-3.68 ± 9.00
SNR	4.18 ± 0.55	6.28 ± 1.34	5.41 ± 1.06	7.64 ± 0.94	8.95 ± 2.02	8.28 ± 1.08
gCNR	0.66 ± 0.13	0.71 ± 0.27	0.71 ± 0.26	0.65 ± 0.14	0.73 ± 0.27	0.61 ± 0.32
PSNR (dB)	/	18.57 ± 0.52	18.53 ± 0.58	/	17.36 ± 0.61	15.40 ± 0.77
DSC	/	0.03 ± 0.02	0.02 ± 0.02	/	0.09 ± 0.10	0.01 ± 0.01
Metrics	VGG-16					
CR (dB)	-4.43 ± 1.65	-3.29 ± 1.39	-4.52 ± 3.32	-2.15 ± 0.83	-1.82 ± 0.80	-2.63 ± 1.88
SNR	4.18 ± 0.55	6.26 ± 0.73	6.05 ± 1.13	7.64 ± 0.94	8.68 ± 0.92	8.40 ± 1.31
gCNR	0.66 ± 0.13	0.73 ± 0.23	0.75 ± 0.27	0.65 ± 0.14	0.68 ± 0.28	0.67 ± 0.28
PSNR (dB)	/	18.83 ± 0.59	18.23 ± 0.85	/	15.28 ± 0.95	14.54 ± 0.92
DSC	/	0.01 ± 0.01	0.02 ± 0.02	/	0.01 ± 0.01	0.02 ± 0.02
Metrics	VGG-19					
CR (dB)	-4.43 ± 1.65	-3.52 ± 1.60	-5.47 ± 2.84	-2.15 ± 0.83	-1.70 ± 1.41	-15.53 ± 14.31
SNR	4.18 ± 0.55	5.97 ± 0.81	6.28 ± 1.12	7.64 ± 0.94	7.95 ± 1.71	9.77 ± 1.77
gCNR	0.66 ± 0.13	0.74 ± 0.23	0.79 ± 0.20	0.65 ± 0.14	0.69 ± 0.27	0.85 ± 0.24
PSNR (dB)	/	18.28 ± 0.90	19.01 ± 0.67	/	16.30 ± 0.64	18.46 ± 0.79
DSC	/	0.07 ± 0.08	0.03 ± 0.03	/	0.06 ± 0.04	0.07 ± 0.08

Tabella 4.22: Confronto tra le metriche di valutazione per le immagini di carotide.

Le immagini di carotide si confermano per la DNN come le più complesse da predire. La rete presenta difficoltà nel segmentare correttamente la carotide, come testimoniato dai bassi valori DSC. Nonostante ciò, si registrano miglioramenti nell'utilizzo della rete con encoder VGG-19 e target E_{auto} (Fig.4.11), con gCNR superiore rispetto all'immagine DAS_{auto} originale (0.85 ± 0.24 vs 0.65 ± 0.14) e PSNR superiore rispetto all'allenamento con encoder VGG-13 (18.46 ± 0.79 dB vs 15.40 ± 0.77 dB).

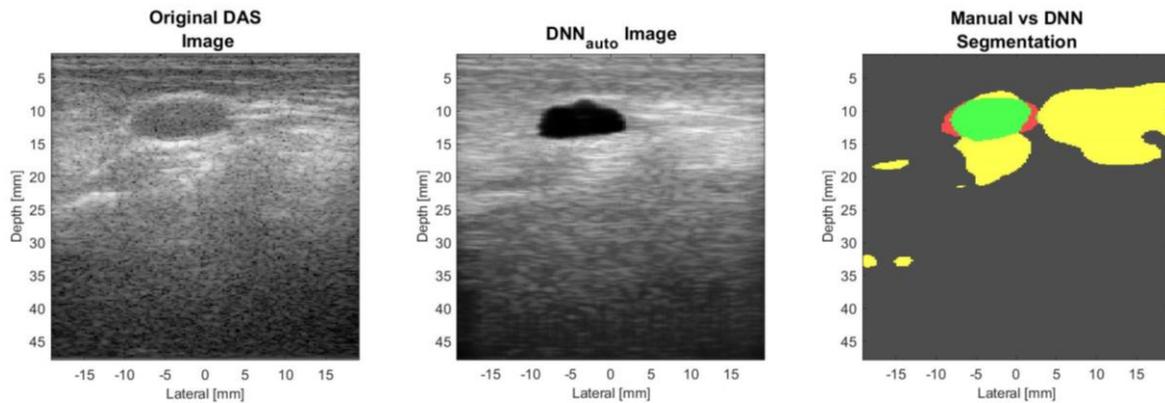


Figura 4.11: In figura viene riportato un'immagine di carotide predetta dalla rete E_{auto} con encoder VGG-19.

4.2 DISCUSSIONE

I risultati di questo studio mostrano come le reti DNN siano in grado di formare e segmentare delle immagini ad ultrasuoni che contengono strutture con diversa forma ed ecogenicità. Le reti non sono solo in grado di riconoscere cisti anecogene circolari dal tessuto circostante dell'immagine [41], ma riescono a generalizzare la segmentazione distinguendo anche più di una ciste all'interno della stessa immagine e con diversa ecogenicità.

Il vantaggio principale dell'approccio presentato consiste nella possibilità di utilizzare immagini simulate con trasmissione a singola onda piana per l'allenamento della rete, effettuando poi predizioni su immagini sperimentali e producendo immagini B-mode con contrasto e gCNR superiori (nel caso di target anecogeni). A questo si aggiunge il vantaggio di ottenere la maschera di segmentazione in parallelo, diminuendo la complessità computazionale rispetto ad approcci in cui si utilizzano algoritmi di segmentazione a valle della formazione dell'immagine. Tuttavia, l'utilizzo del dataset contenente solamente target con cisti anecogene circolari non consente l'identificazione di target iperecoici, generando maschere di

segmentazione errate (Fig.4.4), in quanto la rete identifica il segnale del tessuto circostante come proveniente da target anecoici.

L'utilizzo di dataset contenente anche cisti doppie anecogene e iperecogene (sezione 4.1.3), consente di riconoscere nelle immagini anche la presenza simultanea di più cisti con differente ecogenicità. Tuttavia, questo ampliamento del dataset non permette di ottenere notevoli miglioramenti nelle predizioni delle immagini sperimentali phantom e *in-vivo*.

L'attenzione principale si è rivolta dunque all'utilizzo dei dataset presentati in 4.1.4 e 4.1.5, dalle cui prove si è dimostrata la possibilità di ottenere il beamforming e la segmentazione di immagini ecografiche contenenti strutture irregolari e di differente ecogenicità (sono state incluse cisti anecogene, iperecogene, ipoecogene e doppie). Nel dataset della prova in 4.1.4 l'immagine di rinforzo è stata adoperata per le cisti anecogene e iperecogene, mentre nel caso delle cisti ipoecogene il target coincideva con l'immagine B-mode DAS originale (senza maschera di rinforzo). Questo ha consentito alla rete di produrre immagini con strutture ipoecogene di qualità comparabile al beamforming tradizionale, e al contempo immagini con contrasto e gCNR migliorato per i target iper/anecogeni, enfatizzando le differenze di ecogenicità presenti nelle strutture. Tuttavia, l'utilizzo della maschera di rinforzo anche per le strutture iperecogene non si rivelato essere un vantaggio, in particolare per la predizione delle immagini *in-vivo* di carotide, in cui la rete enfatizzava incorrettamente anche le strutture iperecogene del tessuto endoteliale circostante. Per questa motivazione, nella prova 4.1.5 le maschere di rinforzo vengono adoperate solamente per le cisti anecoiche, utilizzando le immagini B-mode DAS originali per le cisti ipo/iperecoiche.

L'uso di immagini nell'allenamento in cui venisse simulato anche l'effetto del coefficiente di attenuazione introdotto dai tessuti ha consentito di ottenere migliori maschere di segmentazione per il test set sperimentale. La Fig.4.8 mostra come l'attenuazione presente negli strati profondi delle immagini reali venga riconosciuta dalla rete come segnale proveniente da una struttura anecogena, nel caso in cui la DNN venga allenata solamente con immagini simulate senza attenuazione. Questo si ripercuote sulla qualità dell'immagine B-mode predetta e sul DSC della relativa maschera di segmentazione.

L'uso in fase di allenamento di target con dR automatico [44] rappresenta una novità rispetto allo studio precedente, che ha consentito di ottenere miglioramenti per il contrasto e DSC delle immagini phantom con target ipo/anecoici (prova 4.1.5). L'allenamento con target a dR automatico permette di allenare una rete che possa predire immagini visivamente simili tra di

loro, in quanto il dynamic range potrebbe risultare alterato dalle differenti posizioni ed ecogenicità delle cisti. Nel caso delle immagini sperimentali (specialmente quelle *in-vivo*), l'uso di una rete con dR automatico appiattisce le notevoli differenze nell'intensità del segnale alle diverse profondità provocate dall'attenuazione dei tessuti, predicendo con più facilità l'immagine B-mode e la maschera di segmentazione rispetto ad una rete allenata solo con target a dR standard.

I risultati sul test set di immagini simulate non si sono rivelati essere significativamente differenti variando la struttura dell'encoder (VGG-13, VGG-16 e VGG-19), tuttavia si sono registrati notevoli miglioramenti sulle immagini phantom predette dalla rete con encoder VGG-19 e immagini target E_{auto} . La rete produce maschere di segmentazione con DSC più elevati rispetto all'allenamento con VGG-13 e VGG-16, oltre a immagini con migliore contrasto (per i target anecoici) e SNR. Le immagini *in-vivo* di carotide si sono rivelate per la DNN come le più complesse da segmentare, nonostante l'utilizzo del dataset di prova 4.1.5, encoder VGG-19 e target E_{auto} abbia consentito di ottenere immagini con contrasto e gCNR della carotide migliorato (visibile in Fig.4.11). I DSC medi sono risultati tuttavia molto bassi, con il miglior risultato raggiunto dalla maschera di Fig.4.11 (DSC=0.31).

Risulta quindi di interesse un eventuale studio futuro in cui nel training set vengano incluse anche immagini sperimentali (phantom e carotide) assieme a quelle simulate, verificando se un dataset così eterogeneo possa sopperire alle difficoltà della rete nella predizione di immagini *in-vivo*.

Indubbiamente i risultati sono condizionati dall'uso di un training set relativamente piccolo rispetto a quello presentato in [41] (6560 vs. 22230), che ha consentito di ridurre i tempi di simulazione delle immagini e il costo computazionale dell'allenamento, ottenendo comunque ottimi risultati su test set di immagini simulate.

Gli incoraggianti risultati ci spingono a continuare lo studio sulla generalizzazione delle reti per il beamforming. Futuri studi potrebbero riguardare un'analisi degli effetti prodotti da dataset più ampi, che includano strutture con maggiore variabilità sia in forma sia in ecogenicità, oltre all'uso di dataset costituiti dall'unione di immagini con e senza attenuazione. Alle considerazioni sull'ampliamento del dataset adoperato, si presenta anche la necessità di investigare l'effetto dell'uso di altre architetture DNN, sia per l'encoder, sia per il decoder.

CONCLUSIONI

Nel presente elaborato si è osservato quali siano le numerose possibilità offerte oggi dal deep learning, spaziando per le innumerevoli applicazioni in medicina e in particolare nell'imaging biomedicale. Si è visto come la ricerca sia particolarmente fertile nelle applicazioni del deep learning in ultrasonografia, sia per quanto riguarda le potenzialità offerte da strumenti di *computer-aided diagnosis*, sia per un utilizzo più radicale delle reti neurali nella catena di elaborazione del segnale ecografico. In particolare, l'interesse verso l'uso del deep learning nel beamforming ecografico è cresciuto enormemente negli ultimi anni, spinto dalla potenzialità di poter migliorare la qualità dell'immagine prodotta e facilitando in tal modo la diagnosi da parte del radiologo.

Il lavoro presentato ha dimostrato la possibilità di produrre un'immagine B-mode e la relativa maschera di segmentazione a partire dal segnale a radiofrequenza grezzo di una scansione ecografica, avvalendosi di una U-Net dotata di un encoder e di due decoder. L'estensione del dataset con strutture di diversa ecogenicità e dalle forme irregolari ha consentito di allenare una rete che potesse ottenere migliori risultati nella predizione di immagini sperimentali (phantom e carotide *in-vivo*), rispetto ai risultati relativi a reti allenate solamente con cisti anecogene circolari. Il confronto tra i differenti tipi di encoder e i risultati incoraggianti sulle immagini sperimentali ottenuti adoperando una struttura VGG-19 lasciano spazio per la sperimentazione futura di altre architetture. In particolar modo, si rivela incoraggiante l'utilizzo innovativo in fase di allenamento di immagini target con dynamic range regolato automaticamente, il quale garantisce maggiori possibilità di generalizzazione della rete.

In conclusione, la possibilità di trasferire la conoscenza acquisita da immagini simulate per predire immagini phantom reali e *in-vivo* apre ampie opportunità sul prosieguo della ricerca dell'uso del deep learning nel beamforming ecografico, trovando applicazione nella risoluzione di problemi legati ai processi decisionali automatizzati.

BIBLIOGRAFIA

- [1] G. Coppini, S. Diciotti, and G. Valli, *Bioimmagini*. Pàtron, 2012.
- [2] M. Demi and D. Panetta, *Comprehensive Biomedical Physics: X-Ray and Ultrasound Imaging*. New York, NY: Elsevier, 2014.
- [3] T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out*. New York, NY: Elsevier, 2004.
- [4] L. Demi, "Practical guide to ultrasound beam forming: Beam pattern and image reconstruction analysis," *Applied Sciences (Switzerland)*, vol. 8, no. 9, Sep. 2018, doi: 10.3390/app8091544.
- [5] K. E. Thomenius, "Recent Trends in Beamformation in Medical Ultrasound," 2005.
- [6] B. van Veen and K. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.
- [7] G. Matrone, A. S. Savoia, G. Caliano, and G. Magenes, "The delay multiply and sum beamforming algorithm in ultrasound B-mode medical imaging," *IEEE Transactions on Medical Imaging*, vol. 34, no. 4, pp. 940–949, Apr. 2015, doi: 10.1109/TMI.2014.2371235.
- [8] K. E. Thomenius, "Evolution of Ultrasound Beamformers," in *Proc. IEEE Int. Ultrason. Symp*, 1996, pp. 1615–1622.
- [9] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 3, pp. 489–506, Mar. 2009, doi: 10.1109/TUFFC.2009.1067.
- [10] S. Holm, "Digital Beamforming in Ultrasound Imaging," 1994. [Online]. Available: <https://www.researchgate.net/publication/242391444>
- [11] G. Matrone, A. Ramalli, J. D'Hooge, P. Tortoli, and G. Magenes, "A Comparison of Coherence-Based Beamforming Techniques in High-Frame-Rate Ultrasound Imaging with Multi-Line Transmission," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 2, pp. 329–340, Feb. 2020, doi: 10.1109/TUFFC.2019.2945365.

- [12] K. W. Hollman, K. W. Rigby, and M. O'donnell, "Coherence Factor of Speckle from a Multi-Row Probe," in *IEEE Ultrasonics Symposium*, Oct. 1999, pp. 1257–1260.
- [13] P. C. Li and M. L. Li, "Adaptive imaging using the generalized coherence factor," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 50, no. 2, pp. 128–141, Feb. 2003, doi: 10.1109/TUFFC.2003.1182117.
- [14] J. Kirkhorn, "Introduction to IQ-demodulation of RF-data," 1999.
- [15] F. W. Kremkau, *Sonography Principles and Instruments*, 9th ed. Elsevier.
- [16] F. Chollet, *Deep Learning with Python*, 2nd ed. Shelter Island, NY: Manning Publications, 2017.
- [17] R. Zemouri, N. Zerhouni, and D. Racoceanu, "Deep learning in the biomedical applications: Recent and future status," *Applied Sciences (Switzerland)*, vol. 9, no. 8, Apr. 2019, doi: 10.3390/app9081526.
- [18] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.
- [19] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *J. Physiol*, vol. 195, pp. 215–243, 1968.
- [20] K. Fukushima, "Biological Cybernetics Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position," *Biol. Cybernetics*, vol. 36, pp. 193–202, 1980.
- [21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, May 2015, pp. 234–241. [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [22] I. J. Goodfellow *et al.*, "Generative Adversarial Nets," *Advances in neural information processing systems*, vol. 27, Jun. 2014, [Online]. Available: <http://arxiv.org/abs/1406.2661>
- [23] F. Piccialli, V. di Somma, F. Giampaolo, S. Cuomo, and G. Fortino, "A survey on deep learning in medicine: Why, how and when?," *Information Fusion*, vol. 66, pp. 111–137, Feb. 2021, doi: 10.1016/j.inffus.2020.09.006.
- [24] Stuart Russell and Peter Norvig, *Artificial Intelligence-A Modern Approach*, 4th ed.

- [25] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5. Institute of Electrical and Electronics Engineers Inc., pp. 1153–1159, May 01, 2016. doi: 10.1109/TMI.2016.2553401.
- [26] S. Liu *et al.*, "Deep Learning in Medical Ultrasound Analysis: A Review," *Engineering*, vol. 5, no. 2, pp. 261–275, Apr. 2019, doi: 10.1016/j.eng.2018.11.020.
- [27] Z. Akkus *et al.*, "A Survey of Deep-Learning Applications in Ultrasound: Artificial Intelligence-Powered Ultrasound for Improving Clinical Workflow," *Journal of the American College of Radiology*, vol. 16, no. 9, pp. 1318–1328, Sep. 2019, doi: 10.1016/j.jacr.2019.06.004.
- [28] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep Learning in Ultrasound Imaging," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 11–29, Jan. 2020, doi: 10.1109/JPROC.2019.2932116.
- [29] B. Luijten *et al.*, "Deep Learning for Fast Adaptive Beamforming," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 1333–1337.
- [30] B. Luijten *et al.*, "Adaptive Ultrasound Beamforming Using Deep Learning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 3967–3978, Dec. 2020, doi: 10.1109/TMI.2020.3008537.
- [31] A. C. Luchies and B. C. Byram, "Deep Neural Networks for Ultrasound Beamforming," *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 2010–2021, Sep. 2018, doi: 10.1109/TMI.2018.2809641.
- [32] A. C. Luchies and B. C. Byram, "Training improvements for ultrasound beamforming with deep neural networks," *Physics in Medicine and Biology*, vol. 64, no. 4, Feb. 2019, doi: 10.1088/1361-6560/aafd50.
- [33] D. Perdios, A. Besson, F. Martinez, M. Vonlanthen, M. Arditi, and J.-P. Thiran, "On Problem Formulation, Efficient Modeling and Deep Neural Networks for High-Quality Ultrasound Imaging," in *53rd Annual Conference on Information Sciences and Systems (CISS)*, 2019, pp. 1–4.
- [34] M. Gasse, F. Millioz, E. Roux, D. Garcia, H. Liebgott, and D. Friboulet, "High-quality plane wave compounding using convolutional neural networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 10, pp. 1637–1639, Oct. 2017, doi: 10.1109/TUFFC.2017.2736890.

- [35] Y. H. Yoon and J. C. Ye, "Deep Learning for Accelerated Ultrasound Imaging," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Sep. 2018, vol. 2018-April, pp. 6673–6676. doi: 10.1109/ICASSP.2018.8462304.
- [36] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, "Efficient B-Mode Ultrasound Image Reconstruction From Sub-Sampled RF Data Using Deep Learning," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 325–336, Feb. 2019, doi: 10.1109/TMI.2018.2864821.
- [37] S. Khan, J. Huh, and J. C. Ye, "Deep Learning-based Universal Beamformer for Ultrasound Imaging," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Apr. 2019, pp. 619–627. [Online]. Available: <http://arxiv.org/abs/1904.02843>
- [38] Nair A, Tran T, Reiter A, and Lediju Bell M, "A Generative Adversarial Neural Network for Beamforming Ultrasound".
- [39] A. A. Nair, T. D. Tran, A. Reiter, and M. A. Lediju Bell, "A Deep Learning Based Alternative to Beamforming Ultrasound Images," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Sep. 2018, vol. 2018-April, pp. 3359–3363. doi: 10.1109/ICASSP.2018.8461575.
- [40] Nair A, Tran T, Reiter A, and Lediju Bell M, "One-Step Deep Learning Approach to Ultrasound Image Formation and Image Segmentation with a Fully Convolutional Neural Network," p. 2644.
- [41] A. A. Nair, K. N. Washington, T. D. Tran, A. Reiter, and M. A. Lediju Bell, "Deep Learning to Obtain Simultaneous Image and Segmentation Outputs from a Single Input of Raw Ultrasound Channel Data," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2493–2509, Dec. 2020, doi: 10.1109/TUFFC.2020.2993779.
- [42] Nair A, Rajugopal Gubbi M, Duy Tran T, Reiter A, and Lediju Bell M, "A Fully Convolutional Neural Network for Beamforming Ultrasound Images; A Fully Convolutional Neural Network for Beamforming Ultrasound Images," 2018.
- [43] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014, [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [44] K. M. Meiburger, S. Seoni, and G. Matrone, "Automatic dynamic range estimation for ultrasound image visualization and processing," in *IEEE International Ultrasonics Symposium, IUS*, Sep. 2020, vol. 2020-September. doi: 10.1109/IUS46767.2020.9251470.

- [45] K. H. Zou *et al.*, "Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index," *Acad Radiol*, vol. 11, no. 2, pp. 178–189, 2004, [Online]. Available: <http://www.slicer.org>
- [46] A. Rodriguez-Molares *et al.*, "The Generalized Contrast-to-Noise Ratio: A Formal Definition for Lesion Detectability," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 4, pp. 745–759, Apr. 2020, doi: 10.1109/TUFFC.2019.2956855.
- [47] S. Yao, W. Lin, E. Ong, and Z. Lu, "Contrast Signal-to-Noise Ratio For Image Quality Assessment," in *International Conference on Image Processing, ICIP, 2005*, pp. 397–400.
- [48] J. Dinu, R. Ganesan, A. A. Kebede, and B. Veerasamy, "Performance Analysis and Comparison of Medical Image Compression Techniques," in *Int. Conf. Control Instrum. Commun. Comput. Technol. ICCICCT, 2016*, pp. 738–745.