

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

Scuola di Scienze  
Dipartimento di Fisica e Astronomia  
Corso di Laurea Magistrale in Fisica

# Applicazioni di Deep Learning per CAD Mammografico

Relatore:  
Prof. Renato Campanini

Presentata da:  
Federico Ridolfi

Correlatore:  
Dott.ssa Rosa Brancaccio

Anno Accademico 2015/2016

# Indice

<b>Introduzione</b>	<b>4</b>
<b>1 Il tumore alla mammella e nuove frontiere dei sistemi CAD</b>	<b>6</b>
1.1 Il tumore alla mammella . . . . .	6
1.2 Apprendimento Approfondito e Reti Neurali a Convoluzione . . . . .	10
1.2.1 Le Reti Neurali a Convoluzione . . . . .	13
1.2.2 CNN e CAD mammografico . . . . .	19
<b>2 Libreria e Database impiegati, post-processing delle immagini</b>	<b>22</b>
2.1 Caffe: Convolutional Architecture for Fast Feature Embedding . . . . .	22
2.2 Il database mini-MIAS . . . . .	25
2.2.1 Elaborazione delle immagini . . . . .	26
2.2.2 Data Augmentation . . . . .	27
<b>3 Architettura della rete e fase di Training</b>	<b>32</b>
3.1 La rete: CaffeNet_MAMMO . . . . .	32
3.2 Addestramento: il Solver del modello . . . . .	36
3.3 Addestramento con Fine Tuning . . . . .	38
<b>4 Classificazione di sezioni di immagine</b>	<b>40</b>
4.1 L'algoritmo di classificazione . . . . .	41

4.2	Risultati . . . . .	42
4.2.1	Classificazione con CaffeNet_MAMMO . . . . .	43
4.2.2	Classificazione con CaffeNet_MAMMO2 . . . . .	48
<b>5</b>	<b>Localizzazione della patologia in immagine completa</b>	<b>50</b>
5.1	L'algoritmo di localizzazione . . . . .	51
5.2	Risultati . . . . .	53
5.2.1	Localizzazione con CaffeNet_MAMMO . . . . .	54
5.3	Confronto con la letteratura . . . . .	60
	<b>Conclusioni</b>	<b>63</b>
	<b>Bibliografia</b>	<b>66</b>

E l'uomo creò le macchine.

A sua immagine e somiglianza Egli le  
creò.

---

A. e L. Wachowski, *Animatrix: Il  
Secondo Rinascimento*, 2003

# Introduzione

Il tumore alla mammella è una delle patologie oncologiche più diffuse nel mondo occidentale ed allo stato dell'arte attuale la più efficace protezione che si ha è la diagnosi precoce [1], che permette di intervenire per tempo prima che la patologia sviluppi metastasi e si diffonda nell'organismo. Per questo motivo sono da anni attivi a livello nazionale programmi di screening mammografico per tutte le donne di età superiore ai 50 anni (45 in alcune regioni tra cui il Lazio<sup>1</sup>). La tipica mammografia offre al medico che deve interpretarla un'immagine ricca di strutture, dovute al tessuto ghiandolare interno alla mammella e alla distribuzione spaziale del grasso che la costituisce. Questo fa sì che spesso l'immagine sia di difficile lettura e può pertanto capitare di marcare come negativi dei pazienti nei quali la patologia è ancora nei primi stadi.

Per limitare questo problema e accelerare l'analisi dei referti medici negli ultimi vent'anni si sono sviluppati numerosi software di Diagnosi Assistita Computerizzata (CAD) i quali hanno la funzione di assistere il radiologo nell'analisi dell'immagine, focalizzando la sua attenzione sulle zone nelle quali il programma ha identificato una potenziale patologia. Sebbene non possano sostituire l'esperienza dello specialista, tali algoritmi, eventualmente integrati direttamente all'interno dello scanner mammografico impiegato in sede di screening, hanno dato un sostanziale contributo nel soddisfare la richiesta sempre maggiore di mammografie di screening negli ultimi anni, aiutando di fatto a ridurre la mortalità per tumore alla mammella fino al calo del 35~40% [1] che possiamo

---

<sup>1</sup>Fonte AIRC

registrare ad oggi.

In questo lavoro verrà illustrato l'insieme delle procedure che ha portato alla realizzazione di un sistema di CAD mammografico basato sul *deep learning*, ossia la nuova generazione di reti neurali che a partire dalla fine degli anni 2000 ha iniziato a diffondersi anche per i software applicati in ambito medicale.

Nel primo capitolo saranno trattate le basi teoriche necessarie per comprendere questo lavoro, andando ad illustrare le caratteristiche del tumore alla mammella e delle reti neurali a convoluzione, analizzando lo stato dell'arte in merito; nel secondo capitolo sarà presentato *caffè*, il framework di lavoro impiegato, e sarà introdotto il database MiniMIAS utilizzato per addestrare la rete, unitamente alle procedure di processing a cui sono state sottoposte le immagini; nel terzo capitolo sarà discusso nel dettaglio il modello adottato e le particolarità della procedura di addestramento; nel quarto capitolo si esporranno i risultati relativi alla classificazione di porzioni ridotte delle immagini (*crop*); nel quinto capitolo si esporrà l'algoritmo di localizzazione e si discuteranno i risultati ottenuti alla luce dello stato dell'arte attualmente disponibile.

# Capitolo 1

## Il tumore alla mammella e nuove frontiere dei sistemi CAD

In questo capitolo andremo a fornire una sintetica trattazione delle caratteristiche della patologia in studio, il tumore alla mammella, e sulla tecnologia adoperata per trattare il problema, le reti neurali a convoluzione, sottolineando le differenze rispetto alle più conosciute e applicate reti neurali artificiali (ANN). Si osservi come la trattazione dell'aspetto medico non pretende in alcun modo di essere esaustiva in quanto una buona caratterizzazione eziologica del cancro mammario richiederebbe uno spazio superiore all'interesse della presente tesi. L'obiettivo sarà quindi di fornire alcuni dati e informazioni atti a comprendere la necessità di migliorare costantemente le capacità di screening a disposizione delle équipes mediche.

### 1.1 Il tumore alla mammella

Come accennato in introduzione, il tumore alla mammella rappresenta la principale causa di mortalità per tumore femminile (16.3% dei casi) [2], è maggiormente diffuso in popolazioni ad elevato tenore di vita e sviluppo sociale e può presentarsi sia in individui

di sesso femminile che di sesso maschile a causa della medesima composizione tissutale della mammella nei due generi, sebbene i casi di tumore alla mammella maschile coprano meno dell'1% dei casi totali.

La distribuzione d'età dei pazienti malati è sostanzialmente trascurabile fino ai 25 anni, quindi aumenta gradualmente, con un massimo intorno ai 70. In figura 1.1<sup>1</sup> notiamo come il numero di malati cala da tale età in poi in quanto rappresenta gli individui totali affetti da patologia. Risulta quindi più interessante studiare la linea continua, che rappresenta i casi patologici normalizzati al numero di donne totali aventi la stessa età, risultando quindi un dato statisticamente più significativo.

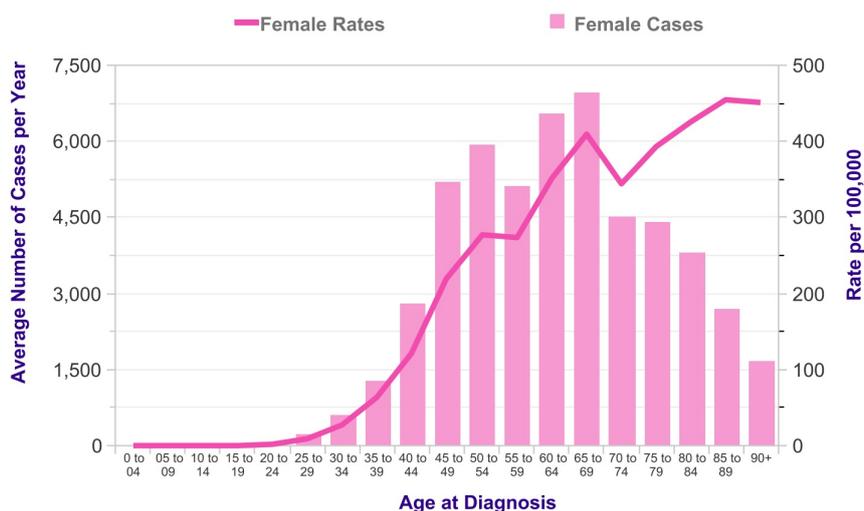


Figura 1.1: Casi (istogramma) e tassi (linea continua) di incidenza del tumore al seno in funzione dell'età di insorgenza.

Se si escludono le correlazioni con l'età e i relativi sbalzi ormonali indotti dalla menopausa, ad oggi non sono ancora chiare le cause del carcinoma mammario: sono presenti in letteratura studi che correlano tale patologia ad altri disturbi quali diabete, obesità e ipertensione [3], ma in molti casi si tratta di studi locali in aree dove queste patologie

<sup>1</sup>Tratta da Cancer Research UK, 2015

sono estremamente diffuse nella popolazione rendendo quindi difficile trarne conclusioni certe.

A causa di questa mancanza di informazioni l'unico metodo per dare una buona aspettativa di vita alle persone affette da tale patologia è una diagnosi precoce in modo da intervenire prima che il tumore inizi ad andare in metastasi; per dare un'idea, intervenire a seguito di una diagnosi precoce permette di dare al paziente un'aspettativa di vita di circa 20 anni, mentre nel caso in cui la diagnosi abbia luogo quando ormai il tumore è in uno stadio avanzato riduce l'aspettativa di vita a tempi dell'ordine del paio di anni. La tecnica più diffusa per lo screening è la mammografia, da eseguirsi per le donne ogni due anni una volta entrati nella fascia d'età a rischio. Tale pratica è divenuta lo standard in quanto non invasiva e, sebbene basata su radiazioni ionizzanti, offre un'immagine con una risoluzione non raggiungibile dall'ecografia, in grado di rivelare le microcalcificazioni di dimensione più ridotta, che possono avere dimensioni dell'ordine delle decine di micron. In figura 1.2 <sup>2</sup> possiamo vedere un esempio di mammografia completa di una singola mammella con le due proiezioni tipicamente adoperate, Cranio-Caudale e Medio-Laterale Obliqua.

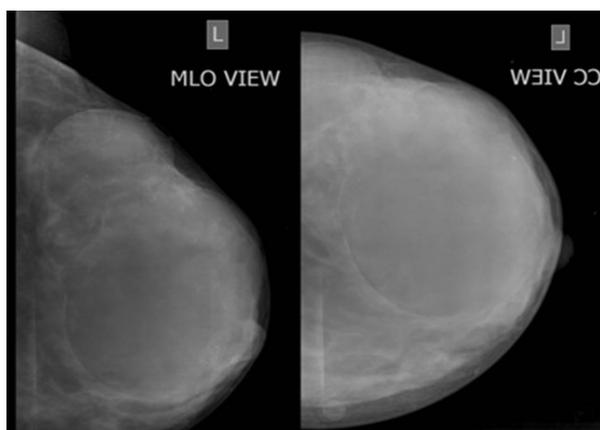


Figura 1.2: Tipica mammografia di screening non post-processata, sono visibili entrambe le proiezioni, MLO e CC.

---

<sup>2</sup>Tratta da Prasad Wavare, Vishal Bakare, Rakhi P. Puria, *Evaluation of Breast Masses Using Mammography and Sonography as First Line Investigations.*

Una tipica mammografia presenta una gran quantità di strutture, immerse all'interno di un campo circa uniforme dovuto al grasso presente in abbondanza nella mammella. All'interno dell'organo mammario sono immersi i tessuti ghiandolari, costituiti da un dotto primario che si ramifica in alcuni lobi ulteriormente frazionati fino a formare delle sezioni dette dotti terminali o lobuli, che terminano negli acini, all'interno dei quali avviene la secrezione del latte. Per dare un'idea dei numeri in gioco, la tipica mammella contiene 15-18 lobi, ognuno ramificato in 20-40 lobuli, da ognuno dei quali partono da 10 a 100 acini. Le zone lobulari sono molto importanti per la genesi del tumore e sono il sito di nascita di numerosi tumori invasivi, così come di calcificazioni che possono ostruirli [4].

Le calcificazioni risultano di estrema importanza per la diagnosi precoce in quanto si è osservato come, qualora si riuniscano in raggruppamenti (*cluster*), esse rappresentino un possibile precursore di masse tumorali essendo dovute a secrezioni di cellule mutate [4].

All'interno di una radiografia di screening è possibile identificare principalmente le suddette calcificazioni, anomalie strutturali o delle vere e proprie masse tumorali, circoscritte (ossia con un bordo e una forma ben definite) o spicolate (ossia con un nucleo centrale dal quale si estendono diverse ramificazioni di forma irregolare). Ognuna di queste anomalie può tuttavia essere benigna (e quindi non richiedere pratiche o interventi ma da tenere sotto controllo in caso di future mutazioni) o maligna, nel qual caso il radiologo deve richiedere ulteriori accertamenti eventualmente invasivi (e.g. una biopsia) per confermarne la natura e procedere con la sua rimozione, tipicamente per via chirurgica. Va specificato che in caso di diagnosi sufficientemente precoce si riesce a evitare una completa mastectomia, rimuovendo sezioni molto contenute di tessuto ed evitando quindi danni collaterali di carattere estetico e psicologico, altro motivo per il quale la prevenzione è caldamente consigliata.

## 1.2 Apprendimento Approfondito e Reti Neurali a Convoluzione

A partire dalla seconda metà degli anni '80, mentre le reti neurali iniziavano a diffondersi all'interno di istituti di ricerca e industria come applicazioni di frontiera per la classificazione, iniziava a diffondersi l'idea di andare oltre quanto ottenuto sinora cambiando il paradigma di apprendimento cui ispirarsi: iniziava a nascere il principio di cercare di insegnare al computer ad apprendere non più basandosi su una serie di osservazioni fornite dall'utente (le cosiddette caratteristiche, o *features*, tipiche del linguaggio delle reti neurali artificiali o *Artificial Neural Networks*) ma spingendolo ad apprendere le caratteristiche salienti di un problema durante l'addestramento stesso. Nasceva il *deep learning*, ma per quasi 30 anni gli studi in merito furono puramente teorici in quanto la potenza di calcolo disponibile all'epoca per chiunque non disponesse di accesso a un sistema di calcolo centralizzato ad alte prestazioni (*cluster High Performance Computing*, o HPC) era svariati ordini di grandezza al di sotto delle necessità degli algoritmi coinvolti, come visibile in figura 1.3<sup>3</sup>, nella quale si può osservare l'evoluzione della potenza in termini di operazioni in virgola mobile al secondo (*Floating Operations Per second*, *FLOPs*) del più potente computer al mondo (diamanti rossi), del 500esimo (triangoli gialli) e la potenza aggregata di tutti i 500 mainframe più potenti al mondo partecipanti alla TOP500 challenge<sup>4</sup>, (quadrati blu) anno per anno, con evidenziato il valore di 1 TeraFLOPs, ritenuto il valore di riferimento per addestrare una rete deep in tempi dell'ordine del giorno, ad oggi raggiungibile da qualunque scheda video di fascia medio-alta.

Le principali differenze tra le due tipologie di algoritmi di apprendimento risiede in poche ma sostanziali differenze:

- Nel deep learning tutti gli strati di neuroni "deep" applicano operazioni non lineari,

---

<sup>3</sup>Modificata da AI.Graphic - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=33540287>

<sup>4</sup>Per ulteriori informazioni <http://www.top500.org/> (visto il 09/11/2016)

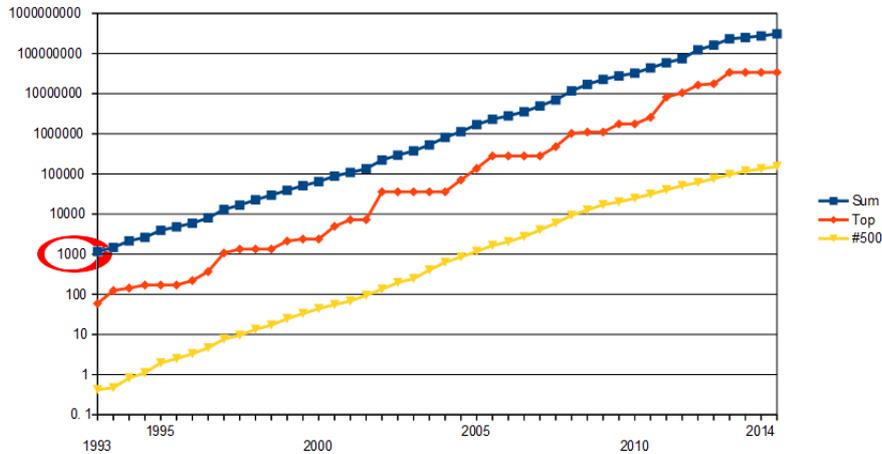


Figura 1.3: Evoluzione della potenza dei computer della TOP500, in ordinata potenza in GFLOPs, in ascissa anno di osservazione. È evidenziato il valore di 1 TeraFLOPs ritenuto potenza di riferimento per addestramento reti deep.

ottenendo quindi la possibilità di separare classi in situazioni dove una comune rete neurale troverebbe difficoltà notevoli;

- Le reti deep sono caratterizzate dal fatto di avere un numero di neuroni e collegamenti in numero molto più elevato rispetto alle ANN in quanto le decisioni sono dovute a risposte complessive di gruppi di neuroni, specialmente nelle reti a convoluzione impiegate in questo lavoro;
- Le reti deep possono non avere necessità di lavorare su features estratte a partire dai dati di input, ma sviluppano una serie di criteri (filtri) durante la fase di apprendimento, ottenendo performances paragonabili o addirittura superiori alle reti neurali classiche. Per realizzare questa operazione è necessario un numero di esempi per l'addestramento molto superiore rispetto a quanto necessario per una ANN tradizionale, e questa necessità rappresenta il principale problema nel realizzare un classificatore deep.

Alla luce di queste caratteristiche evidente che le reti ad apprendimento approfondito siano uno strumento estremamente potente in grado di lavorare su problemi difficilmente affrontabili dalle reti neurali tradizionali, tutto questo al prezzo di aumentare notevolmente il numero di parametri da configurare e quindi portando a addestramenti ingestibili da CPU multi-core, in quanto la loro limitata capacità di parallelizzazione porterebbe ad addestramenti con durate insostenibilmente lunghe. Per questo motivo le deep neural networks sono rimaste inutilizzate in ambito industriale e privato fino al 2007, anno in cui la NVIDIA con la serie 8xxx ha distribuito la prima versione di CUDA (*Compute Unified Device Architecture*), il proprio framework di calcolo distribuito su GPGPU (*General Purpose Graphic Processing Unit*, letteralmente "scheda grafica a scopo generico"), portando il calcolo parallelo su architettura many core dai mainframe (i centri di calcolo di università e istituti di ricerca privati) ai normali computer domestici, con costi inferiori ai mille euro<sup>5</sup>. Il principale vantaggio delle GPGPU sulle CPU sta nel fatto che tali dispositivi contengono un numero elevato (dell'ordine del migliaio) di *Shaders*, dei core estremamente veloci ma con un *Instruction Set* estremamente ridotto, contenente solo le istruzioni necessarie per eseguire i calcoli relativi a matrici e sistemi fisici (l'architettura NVidia GeForce nasce a partire dall'Ageia Physx, una PPU (*Physical Processing Unit*) commercializzata nel 2006 per alleggerire la CPU dal calcolo dei motori fisici dei videogiochi 3D), e risultano pertanto ideali per la parallelizzazione di un numero enorme di calcoli semplici, che è proprio il caso dell'addestramento di una rete deep. A causa del grande impiego di GPGPU nel calcolo scientifico NVidia ha in seguito sviluppato una linea di prodotti specificamente ingegnerizzata allo scopo dell'accelerazione del calcolo, NVidia Tesla, rivolta principalmente ad un utenza professionale e priva di velleità grafiche.

L'apprendimento profondo trova oggi applicazioni in numerosi campi, come ad esempio la crittografia [5], l'analisi del linguaggio naturale [6] e persino la simulazione del

---

<sup>5</sup><http://www.anandtech.com/show/2222>, visto il 30/10/16.

sogno<sup>6</sup>, e si ritiene che possano essere le reti in grado di superare il paradigma delle regole finite, che modella l'apprendimento da parte di un computer come limitato da un numero finito di regole che devono essere insegnate alla macchina prima di iniziarne il vero e proprio addestramento, arrivando ad una prima intelligenza artificiale a regole non finite, come sta avvenendo nelle IA deputate all'analisi del linguaggio naturale impiegate da Google, in grado di sviluppare spontaneamente un proprio set di regole crittografiche per evitare intercettazioni [7].

Inoltre una delle principali applicazioni del deep learning è la *computer vision*, ossia quell'insieme di procedure e algoritmi che permettono ad una macchina di estrarre informazioni a partire da un'immagine o un flusso video, specie per quanto concerne l'analisi e la classificazione di oggetti all'interno di scene complesse in tempo reale. A tale scopo i migliori risultati sono ottenuti dalle reti a convoluzione (*Convolutional Neural Networks*, CNN o ConvNet), impostesi come standard e dominanti in tutte le competizioni sull'argomento<sup>7</sup>. Per le loro performances nell'ambito della CV, le CNN saranno lo strumento adoperato in questo lavoro.

### 1.2.1 Le Reti Neurali a Convoluzione

Una CNN è un particolare tipo di rete basata sul perceptrone multistrato (MLP) ma ispirati alla corteccia visiva dei mammiferi. Tale regione del cervello elabora le immagini provenienti dal sistema visivo non solo grazie alla potenza di elaborazione delle singole unità funzionali, ma anche e soprattutto grazie alla loro disposizione e aggregazione [10]: le cellule sono raggruppate in gruppi, ognuno dei quali deputato all'elaborazione delle informazioni provenienti da uno specifico settore della scena individuata dall'occhio, agendo da filtro locale. Tale architettura risulta intuitivamente ottimale per la ricerca di forti correlazioni spaziali all'interno di scene complesse, tipica delle immagini naturali. A

---

<sup>6</sup>Google DeepDream, “<https://research.googleblog.com/2015/07/deepdream-code-example-for-visualizing.html>” (visto il 30/10/2016).

<sup>7</sup><http://image-net.org/challenges/LSVRC/2014/results> (visto 09/11/2016)

causa delle ottime performances della corteccia visiva nello svolgere tale compito, risulta naturale imitarne il comportamento all'interno di reti neurali applicate alla CV.

Sebbene le unità fondamentali di ogni strato della rete siano perceptron semplici e gli ultimi strati della rete, nei quali avviene la vera e propria classificazione, non siano altro che dei MLP completamente connessi, una CNN si differenzia da un MLP per delle sostanziali differenze di architettura negli strati convolutivi [8]:

**Connettività Sparsa:** Nelle CNN i neuroni appartenenti a diversi layers di uno strato convoluzionale sono collegati tra loro tramite un'architettura piramidale regolare; ogni neurone di un determinato strato riceve informazioni da uno specifico numero  $N$  di neuroni dello strato precedente tutti tra loro adiacenti, i quali inviano informazioni ad un massimo di  $N$  neuroni dello strato successivo, esattamente come avviene in un prodotto di convoluzione: lo strato di input svolge il ruolo della matrice su cui applicare l'operazione mentre i pesi che collegano i neuroni dello strato  $m-1$  a quelli dello strato  $m$  il filtro convolutivo, ed è proprio questa analogia a dare il nome alle CNN. Detta in altri termini, ogni unità è sensibile solo a variazioni e input provenienti da una specifica area di competenza, come visibile in figura 1.4. Questa architettura assicura che i pattern appresi dalle singole unità (filtri) diano forti risposte a input spazialmente contenuti. Una sufficiente profondità di tale struttura tuttavia permette ad informazioni locali di raggrupparsi gradualmente, portando alla creazione di filtri non lineari sempre più globali negli strati più prossimi all'output;

**Condivisione dei pesi:** Mentre la connettività sparsa caratterizza la gerarchizzazione dei diversi strati della rete, questa proprietà caratterizza i diversi neuroni appartenenti ad uno stesso strato; in una CNN ogni perceptrone appartenente a un dato strato ha i medesimi pesi e *bias* (soglie di attivazione) di ogni altro perceptrone suo pari, e l'insieme dei parametri di tutti i neuroni simili forma una *feature map* (v. figura 1.5), ossia l'insieme delle caratteristiche comuni a tutti i neuro-

ni presenti ad un determinato livello della rete. Questa proprietà è necessaria una volta impostata l'architettura piramidale di cui sopra per ottenere l'invarianza della risposta della rete per traslazioni, ossia la capacità di riconoscere un target indipendentemente dalla sua posizione all'interno della scena, una proprietà fondamentale per la generalizzazione di un'algoritmo di CV. Come effetto positivo collaterale, l'uso della feature map diminuisce i parametri liberi aumentando l'efficacia dell'apprendimento.

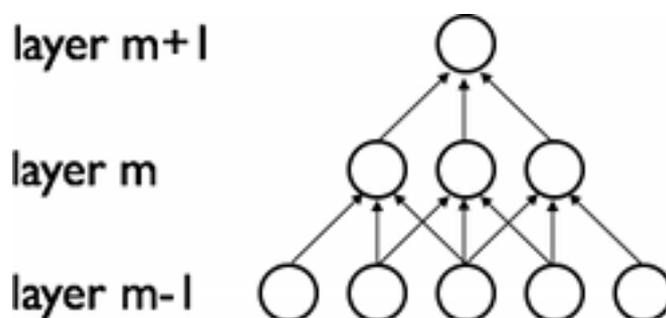


Figura 1.4: Esempio di connettività sparsa dei neuroni tra diversi strati convoluzionali: i neuroni di uno strato  $m$  sono connessi solo ad alcuni dei neuroni dello strato precedente  $m-1$  secondo principi di località.

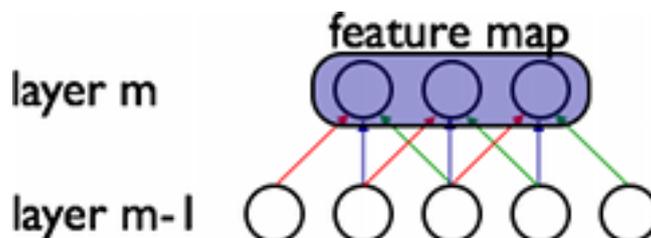


Figura 1.5: Esempio di Condivisione dei pesi (Feature Map) tra i neuroni tra diversi strati convoluzionali, a link di colore uguale corrisponde peso uguale.

Alla luce di queste caratteristiche, la funzione di un convolutional layer si può riassumere nel riconoscere determinate strutture, sempre più raffinate e nascoste man mano che il layer è collocato più in profondità all'interno della rete, e fornire una risposta quando il matching tra il filtro appreso e la sezione di immagine in studio è sufficientemente elevato.

Oltre a queste proprietà architetture la tipica CNN fa uso di strati ausiliari, detti di Aggregazione (*Pooling*) e Rettificazione (*ReLU*), atti a semplificare la comunicazione tra diversi strati convolutivi e diminuire ulteriormente il numero di parametri liberi [8]:

**Pooling:** Uno strato di Pooling ha il compito di aggregare i dati uscenti da uno strato convoluzionale riducendoli di numero, agendo anch'esso su gruppi spazialmente localizzati di neuroni di dimensioni pari al kernel del Pooling Layer. Nelle tipiche applicazioni delle CNN si parla di Max Pooling in quanto per ogni gruppo di neuroni viene selezionato solo il valore di massima intensità. Tipicamente le zone su cui viene applicato il pooling sono costruite in modo da essere parzialmente sovrapposte in modo da preservare l'informazione locale nonostante le soppressioni adoperate. Questa operazione ha due conseguenze, la prima (di ordine pratico) è che il numero di input nello strato successivo viene ridotto di un fattore paragonabile alla dimensione del kernel (pari se e solo se non si ha sovrapposizione tra le zone aggregate); il secondo è un rafforzamento della invarianza per traslazioni introdotta dalla condivisione dei pesi: il pooling infatti permette di riassorbire traslazioni di dimensioni dell'ordine del kernel adoperato nel passaggio dei dati tra uno strato e l'altro;

**ReLU:** La Rectified Linear Unit è uno strato di attivazione e in quanto tale agisce sui dati di ingresso in termini puntuali. La ReLU mantiene inalterati tutti gli input positivi e moltiplica per una costante (tipicamente 0) tutti gli input negativi. Viene spesso usata al posto di funzioni di attivazione più regolari, quali sigmoide o a tangente iperbolica, per la semplicità di implementazione e il quasi trascurabile carico computazionale introdotto da tale strato. Poiché inserite tra strati di convoluzione e pooling layers/strati completamente connessi, le ReLU aiutano a diminuire ulteriormente la quantità di calcoli operati negli strati bersaglio.

Questo elenco di caratteristiche permette quindi di capire che, sebbene le unità logi-

che operano la decisione siano collocate in strati con caratteristiche sostanzialmente uguali alle comuni ANN, le CNN introducono una quantità di operazioni volte a modificare la rappresentazione dell'immagine andando a creare dei filtri rappresentanti le sezioni di immagine più caratterizzanti e necessarie per la decisione. Questi filtri sostituiscono le features usate dalle vecchie generazioni di reti neurali e potenzialmente sono in grado di fornire una caratterizzazione degli oggetti in molti casi più efficace di un vettore di informazioni sull'immagine, la cui dimensionalità deve peraltro essere controllata per non soffrire la *maledizione della dimensionalità*, ossia quel fenomeno tipico del machine learning per cui superata una certa soglia del numero di features adottate a seconda dei casi si può avere una cattiva convergenza o nel caso opposto una macchina incapace di generalizzare a causa dell'eccessiva specificità dei parametri di classificazione. In figura 1.6<sup>8</sup> sono presenti gli output di alcuni strati di ImageNet, una CNN deputata al riconoscimento di immagine generica, allo scopo di spiegare con maggior chiarezza cosa siano i suddetti filtri: in figura 1.6a sono illustrati i filtri appresi dal primo strato convoluzionale di ImageNet, e come si può osservare essi sono fondamentalmente texture semplici estremamente generiche comunemente trovabili in immagini naturali (linee parallele a diverse orientazioni, colori fondamentali, pattern di intensità ecc.). In figura 1.6c si nota la risposta fornita dai diversi filtri appartenenti al primo strato convoluzionale alla sollecitazione da parte dell'immagine in figura 1.6b. Come si può osservare alcuni filtri non danno alcuna risposta, ma per quelli dove è presente un output si noti come esso ricalchi alcune caratteristiche dell'immagine originale: parti della silhouette del gatto, del prato, elementi della texture del pelo e così via. Man mano che si avanza tuttavia i filtri diventano sempre più numerosi e astratti in quanto agenti su rappresentazione sempre più distanti dall'immagine originale, e risulta quasi impossibile offrirne un'interpretazione, portando a risposte da parte degli strati più profondi come quella proposta in immagine 1.6d, relativa all'output del quinto strato convoluzionale, dove la risoluzione è degradata

---

<sup>8</sup>Tratta da <http://nbviewer.jupyter.org/github/BVLC/caffe/blob/master/examples/00-classification.ipynb> (visto il 16/11/16).

a un punto tale per cui l'occhio umano non riesce a estrarne più informazioni utili.

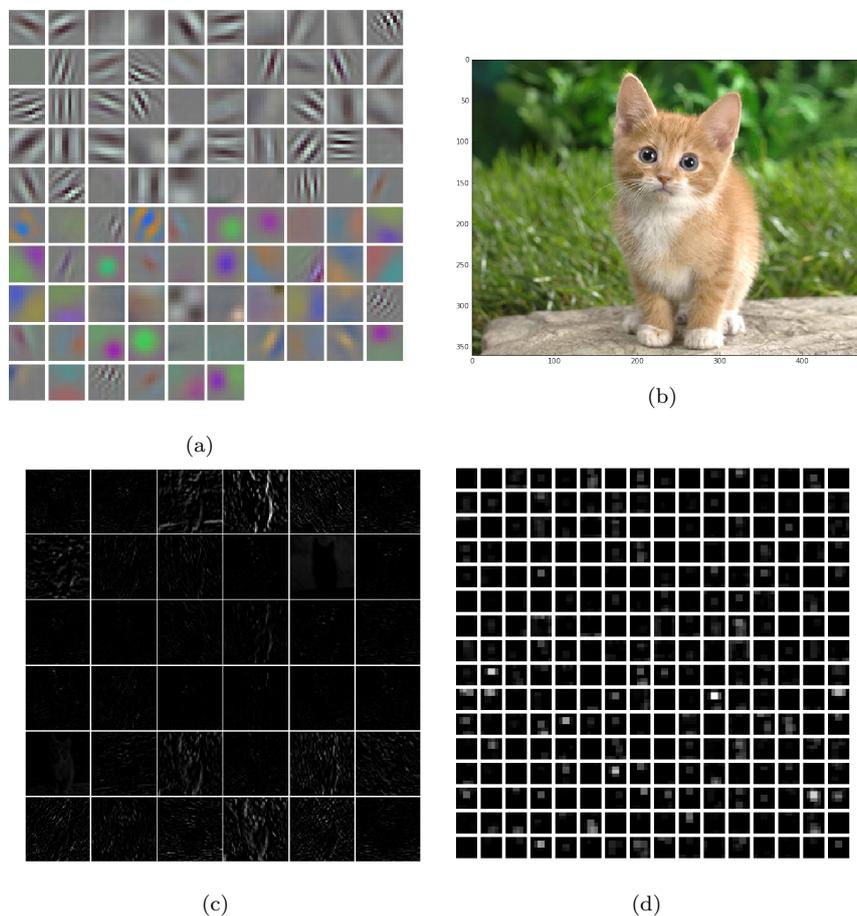


Figura 1.6: Filtri appresi dal primo strato convoluzionale di ImageNet (a) (si noti la distinzione tra filtri di intensità e filtri relativi a specifiche bande cromatiche), immagine di prova sottoposta alla rete (b), risposta (dopo Max Pooling per questioni di leggibilità) del primo strato convoluzionale (c) e del quinto (d).

La qualità di un classificatore binario (sia basato su CNN che su ANN) è stimata a partire dalle curve ROC (Caratteristica Ricevitore-operatore o Caratteristica Relativa di Operatività, in inglese *Receiver Operating Characteristic / Relative Operating Characteristic*) [9], curve ottenute graficando la sensibilità (rateo di vere classificazioni positive) in funzione del rateo di falsi positivi (1-specificità), i singoli punti sono ottenuti spostando la soglia di probabilità di appartenenza alla classe diagnostica a diversi valori. Il vantaggio

delle ROC è che permettono di avere una rappresentazione visiva delle performances del classificatore e di fissare la soglia di classificazione in funzione delle performances desiderate. Inoltre le ROC permettono di valutare quantitativamente la bontà del classificatore di discriminare le classi in oggetto tramite la valutazione dell'AUC (*Area Under Curve*), una stima di quanto la ROC presenti un andamento sopra o sotto-lineare rispetto alla bisettrice (che corrisponde a un classificatore completamente casuale e viene per questo detta in gergo *chance line*): una AUC minore di 0.5 implica che il classificatore ha pessime performance, man mano che il valore si sposta verso 1 (il massimo per costruzione) le performance sono sempre migliori. Un classificatore con  $AUC = 1$  è un classificatore che per qualunque valore di soglia non effettua mai misclassificazioni. Inoltre si può verificare la significatività dell'AUC calcolato tramite il test statistico di Wilcoxon-Mann-Whitney, o "statistica U" [17], un test non parametrico nato per confrontare le mediane di due popolazioni, dando ancora maggior importanza a tale valore.

### 1.2.2 CNN e CAD mammografico

Nonostante la tecnologia CUDA sia già in circolazione da diversi anni si è dovuto aspettare fino al 2015 per avere un primo lavoro che mettesse insieme pratiche di CAD mammografico e CNN [11] ottenendo risultati superiori alle *Support Vector Machine*, un particolare tipo di classificatori ANN non lineari tra i più avanzati disponibili. Nel loro lavoro, Arevalo e il suo team effettuano un confronto tra tre diverse soluzioni convoluzionali e tre metodiche basate su SVM. I tre metodi basati su CNN sono CNN2 e CNN3, reti create appositamente per lo scopo, composte da rispettivamente uno o due strati a convoluzione e con l'output su uno strato decisionale composto da MLP completamente connessi (strato a cui ci si riferisce solo come *fully-connected*), e DeCAF, una rete convoluzionale multistrato molto più complessa delle altre due, applicata per la classificazione di immagini generiche e vincitrice della ILSVRC del 2012 [12]. Le tre soluzioni basate su SVM differiscono nelle features adottate: in un caso si adopera l'istogramma

del gradiente orientato (HOG), nel secondo l'istogramma della divergenza del gradiente (HGD), nel terzo 17 features estratte manualmente da ogni immagine (*Hand Crafted Feats*, o *HCFeats*), tra cui i diversi momenti statistici dell'istogramma dell'immagine, area, perimetro, circolarità, elongazione, centro di massa e descrittori della texture dell'immagine. La valutazione delle performances dei classificatori viene svolta utilizzando le ROC descritte in precedenza. Tutte le reti sono fatte girare su GPGPU NVidia Tesla K40.

Il loro lavoro mette in luce come tutti i modelli convoluzionali superino i classificatori tradizionali sia in termini di ROC che di AUC, fintanto che non si inizino a combinare tra loro CNN e SVM, realizzando classificatori ibridi nei quali sono presenti almeno due stadi: nella prima fase una CNN elabora l'immagine senza effettuare la classificazione, fornendo un vettore di osservazioni dipendente dai filtri appresi, nel secondo strato tale vettore viene combinato con delle features più tradizionali (quali ad esempio le HCFeats, ma anche coefficienti wavelet o qualunque altro metodo caratterizzante si desideri) e dato in input ad una rete neurale non-deep, eventualmente reiterando a cascata questo procedimento più volte, combinando il meglio dei classificatori convoluzionali e di quelli features-based.

Andando a combinare i metodi con le features estratte a mano, tutti i modelli presentano performances comparabili, ad eccezione di CNN2 che, sebbene presenti un miglioramento rispetto alle performances ottenute senza combinare diversi metodi, rimane al di sotto degli altri classificatori di circa 5 punti percentuali. Questo è probabilmente da imputarsi alla semplicità della sua architettura che non riesce a trarre un così grande giovamento dall'accoppiamento del suo output con altre caratteristiche estratte dall'immagine. In figura 1.7 sono presenti i valori di AUC per tutti i metodi ivi proposti, sia impiegati in modo diretto che combinati con le HCFeats e impiegati per la realizzazione di un classificatore ibrido a due stadi.

**Table 2 – Summary of results in terms of AUC in the test set. Best results are shown in bold typeface and (\*) signals scores with no evidence of differences from the highest ( $\rho < 0.1$ ).**

Representation	Standalone	Combined with HCfeats
CNN3	<b>0.82 ± 0.03</b>	0.82 ± 0.03 (*)
CNN2	0.76 ± 0.05	0.78 ± 0.04
HGD	0.78 ± 0.04	<b>0.83 ± 0.04</b>
HOG	0.77 ± 0.03	0.81 ± 0.03 (*)
DeCAF	0.79 ± 0.05	0.82 ± 0.03 (*)
HCfeats	0.77 ± 0.02	–

Figura 1.7: Valori di AUC con relativo errore per i diversi metodi analizzati in [11] con e senza la combinazione con le HCFeats (tratto da [11]).

## Capitolo 2

# Libreria e Database impiegati, post-processing delle immagini

In questo capitolo andremo ad illustrare il database utilizzato per il training e la caratterizzazione della rete, mostrando inoltre le procedure di pre-elaborazione cui sono state sottoposte le immagini e le pratiche di data augmentation adottate per aumentare e migliorare la statistica a nostra disposizione. Sarà inoltre presentata "caffe", la libreria impiegata per realizzare, addestrare e manipolare la rete neurale che effettua buona parte del presente lavoro.

### 2.1 Caffè: Convolutional Architecture for Fast Feature Embedding

Caffè [16] è una libreria di deep learning sviluppata dal Berkeley Vision and Learning Center<sup>1</sup> open source disponibile dal 2014, nato a partire dalla libreria Decaf menzionata in precedenza, e rappresenta il cuore della parte computazionale di questo lavoro.

---

<sup>1</sup><http://bvlc.eecs.berkeley.edu/>

La libreria è scritta in C++ e CUDA, e viene fornita insieme a invocatori di funzioni (*wrapper*) Matlab e Python (il secondo dei quali compatibile con la popolare libreria numerica Numpy), rendendola ideale per lo sviluppo di prototipi grazie al supporto di linguaggi interpretati ma preservando i tempi di elaborazione grazie all'utilizzo di chiamate a routine precompilate. In figura 2.1 (tratta da [16]) è presente un confronto tra diverse librerie per il deep learning, dal quale si possono già intuire i punti di forza di tale implementazione.

Framework	License	Core language	Binding(s)	CPU	GPU	Open source	Training	Pretrained models	Development
Caffe	BSD	C++	Python, MATLAB	✓	✓	✓	✓	✓	distributed
cuda-convnet	unspecified	C++	Python		✓	✓	✓		discontinued
Decaf	BSD	Python		✓		✓	✓	✓	discontinued
OverFeat	unspecified	Lua	C++,Python	✓				✓	centralized
Theano/Pylearn2	BSD	Python		✓	✓	✓	✓		distributed
Torch7	BSD	Lua		✓	✓	✓	✓		distributed

Figura 2.1: Confronto di possibilità offerte al programmatore dalle più diffuse librerie per implementazione di reti deep.

Le caratteristiche salienti di caffe sono numerose e di notevole utilità per lo sviluppo di reti neurali a convoluzione. Le principali sono:

**Modularità:** Una rete in caffe è definita come una sequenza di layers, e la libreria si occupa di gestire buona parte dei parametri dello strato in modo da agevolare l'assemblaggio dell'architettura complessiva, inoltre viene fornita con un intero catalogo dei principali layers utilizzati per le reti a convoluzione, riducendo il compito dell'utente finale;

**Rappresentazione e Implementazione:** In tutte le librerie di machine learning, caffe inclusa, la rete può essere definita all'interno del codice. Tuttavia questa soluzione, sebbene efficiente, presenta alcuni problemi, tra i quali la difficoltà di definizione dei parametri e soprattutto l'impossibilità di condividere la rete in un formato indipendente dal codice. Alla luce di queste difficoltà, caffe offre la possibilità

di definire la rete in tutte le sue caratteristiche in un file esterno basato sullo standard Google Protobuf<sup>2</sup>, una piattaforma di serializzazione dei dati strutturati indipendente dal linguaggio utilizzato: la rete con tutte le sue caratteristiche viene dichiarata in un file di testo non formattato (tipicamente con estensione .prototxt) e quindi il file viene letto dal programma che si occuperà di gestire il flusso dei dati al suo interno, allocando dinamicamente la memoria necessaria a quella specifica rete;

**Supporto a GPU:** Caffe supporta la tecnologia NVidia CUDA con tutti i vantaggi prestazionali del caso. Inoltre al contrario di alcune possibili alternative, quali ad esempio TensorFlow e Theano, tutte le chiamate di funzione di caffe sono indipendenti dall'utilizzare GPU o CPU, rendendo il codice scritto su una workstation dotata di unità CUDA potenzialmente utilizzabile anche su sistemi privi di scheda video NVidia senza necessità di modificare il codice salvo l'aggiunta di una singola riga che selezioni quale dispositivo debba svolgere i calcoli. Questa caratteristica rende praticamente universale il codice scritto. Inoltre, sebbene non sia possibile ripartire un singolo processo (*job*) di caffe su più GPU, è possibile in caso di configurazioni a più GPU (standard NVidia SLI) lanciare diversi job in contemporanea, uno per scheda, utile per provare diverse configurazioni nel tempo di un singolo training;

**Possibilità di proseguire l'addestramento e disponibilità di modelli:** Caffe permette di riprendere l'addestramento in tempi successivi (utile su macchine a basse performances o alle quali non si abbia accesso in maniera continua), così come di addestrare solo alcuni strati della rete partendo da modelli pre-addestrati. A tale scopo sul sito del BVLC sono disponibili vari modelli già pronti all'uso, composti da prototxt della rete e file binario contenente i pesi dei vari collegamenti neuronali.

---

<sup>2</sup><https://developers.google.com/protocol-buffers/>

Tra i modelli proposti sono presenti reti che hanno vinto la ImageNet challenge del 2013, dimostrando quindi la bontà del framework per fini di riconoscimento delle immagini;

**Supporto database:** Caffe supporta nativamente i principali formati di database, quali `leveldb`<sup>3</sup> e `hdf5`<sup>4</sup>, ottimizzando in questo modo l'accesso ai dataset e raggiungendo velocità di picco nella lettura dei dati dell'ordine dei 150 MB/s, con un tempo di elaborazione dell'ordine dei 4 ms per ogni singola immagine in fase di training (dati relativi al set ILSVRC2012<sup>5</sup>).

Grazie a queste caratteristiche, caffe è una delle librerie di punta di machine learning, trovando applicazioni in corporazioni quali Facebook, Google e Flickr, oltre che in numerosi lavori di ricerca, tra cui il presente.

## 2.2 Il database mini-MIAS

Il MIAS ( Mammographic Image Analysis Society) è un database disponibile gratuitamente in rete curato dall'University College di Londra [13].

All'interno di questo database sono presenti 322 immagini mammografiche nativamente analogiche con una risoluzione di  $200 \mu m/\text{pixel}$  in formato `pgm` a 8 bit, ottenute da scansioni di immagini analogiche. A causa di ciò, esse presentano una qualità inferiore a quanto disponibile su un moderno mammografo digitale, il quale offre una risoluzione spaziale superiore ( $O(10 \mu m/\text{pixel})$ ) e una maggiore profondità di colore (12-14 bit), portando a una perdita di informazione per quanto riguarda calcificazioni isolate o cluster di piccole dimensioni e per regioni poco contrastate. Questa scelta è dovuta all'impossibilità di reperire un database di immagini nativamente digitali liberamente utilizzabile a fini

---

<sup>3</sup><http://leveldb.org/> (visto il 11/11/2016)

<sup>4</sup><https://www.hdfgroup.org/HDF5/> (visto il 11/11/2016)

<sup>5</sup>disponibile presso <http://image-net.org/challenges/LSVRC/2012/browse-synsets> (visto il 11/11/2016)

di ricerca e all'impossibilità di convertire le immagini del DDSM<sup>6</sup>, uno dei principali database di immagini analogiche digitalizzate ad alta risoluzione, dal formato proprietario LJPEG in cui è disponibile in uno dei formati supportati da caffè a causa dell'obsolescenza dei pacchetti software ivi reperibili. Sebbene queste caratteristiche limiteranno le potenzialità di generalizzazione del classificatore, a causa dell'impossibilità di trovare alternative valide e dei tempi ristretti si è accettato di lavorare con simili compromessi, fornendo comunque un metodo con buon potere predittivo (v. capp. 4 e 5).

Ogni immagine è estesa senza distorsioni, in questo modo tutto il database si presenta con una dimensione di 1024x1024 pixel. Inoltre le immagini sono relative alla sola proiezione Medio-Laterale Obliqua (MLO), a causa di restrizioni di tempo e di costi, e ottenute in condizioni di esposizione e collocazione del paziente il più uniforme possibile. I casi patologici costituiscono il 40% dei casi totali. Per ogni caso sono presenti due immagini: una relativa alla mammella sinistra (numerazione pari) e una alla destra (numerazione dispari). Tutte le mammografie provengono da diversi istituti medici inglesi e sono state ottenute nell'ambito della normale attività di screening.

Come detto in introduzione il database con cui abbiamo a che fare non è l'ideale per l'addestramento di un classificatore basato su tecniche di deep learning poichè le immagini sono poche e a risoluzione modesta. È pertanto necessario operare in modo pesante sulle immagini in modo da espandere il database di almeno un ordine di grandezza senza che diventi tuttavia ridondante, come vedremo nella sezione 2.2.2.

### 2.2.1 Elaborazione delle immagini

Per prima cosa si è osservato come, sebbene le immagini si presentino generalmente in un buono stato con poco rumore e/o artefatti, l'interno della mammella risulti poco contrastato, nascondendo buona parte delle strutture ghiandolari dietro una luminosità diffusa.

---

<sup>6</sup><http://marathon.csee.usf.edu/Mammography/Database.html> (visto il 14/11/16)

Analizzando lo stato dell'arte presente nella letteratura sui sistemi CAD mammografici, il miglior sistema per il pretrattamento delle immagini mammografiche risulta essere la Recursive Mean Separate Histogram Equalization (RMSHE) [14], una procedura di equalizzazione nella quale l'istogramma viene diviso in due sotto-istogrammi (rispettivamente sopra e sotto l'intensità media dell'immagine) e riequalizzato, reiterando il processo sui sotto-istogrammi finchè non avvengano più variazioni rilevanti [15]. Il principale vantaggio di questa tecnica è la preservazione della luminosità generale dell'immagine, dando quindi all'utente la percezione di avere a che fare con l'immagine originale o, nel caso di un sistema automatizzato, senza cambiare il range di livelli di grigio coinvolti in modo sostanziale, aumentando però il contrasto e quindi la percezione dei dettagli. L'implementazione di questo algoritmo è stata svolta in ambiente matlab. In figura 2.2 possiamo vedere l'immagine relativa al caso #5 prima e dopo l'applicazione dell'RMSHE assieme ai relativi istogrammi. Come si può osservare, l'immagine elaborata presenta un aspetto più definito e appaiono numerose strutture che prima erano nascoste dall'uniformità del tessuto mammario; questo si riflette in un istogramma più uniforme dell'immagine processata (si osservi in particolare la riduzione del picco per  $x \sim 125$ ) rispetto all'originale e quindi in un miglior contrasto.

Oltre a queste operazioni, in fase di classificazione si procede a sottoporre ogni immagine alla sottrazione dell'intensità media e alla soppressione dei livelli di grigio più bassi. L'insieme di tutte queste procedure di post processing dell'immagine permette una buona classificazione.

## 2.2.2 Data Augmentation

Una volta eseguita la procedura di miglioramento del contrasto delle immagini, si è affrontato il problema del numero di immagini presenti nel database: la sua ridotta dimensione è a malapena sufficiente per addestrare un classificatore basato su ANN, quindi decisamente sottodimensionato per uno basato su CNN. Per questo motivo in

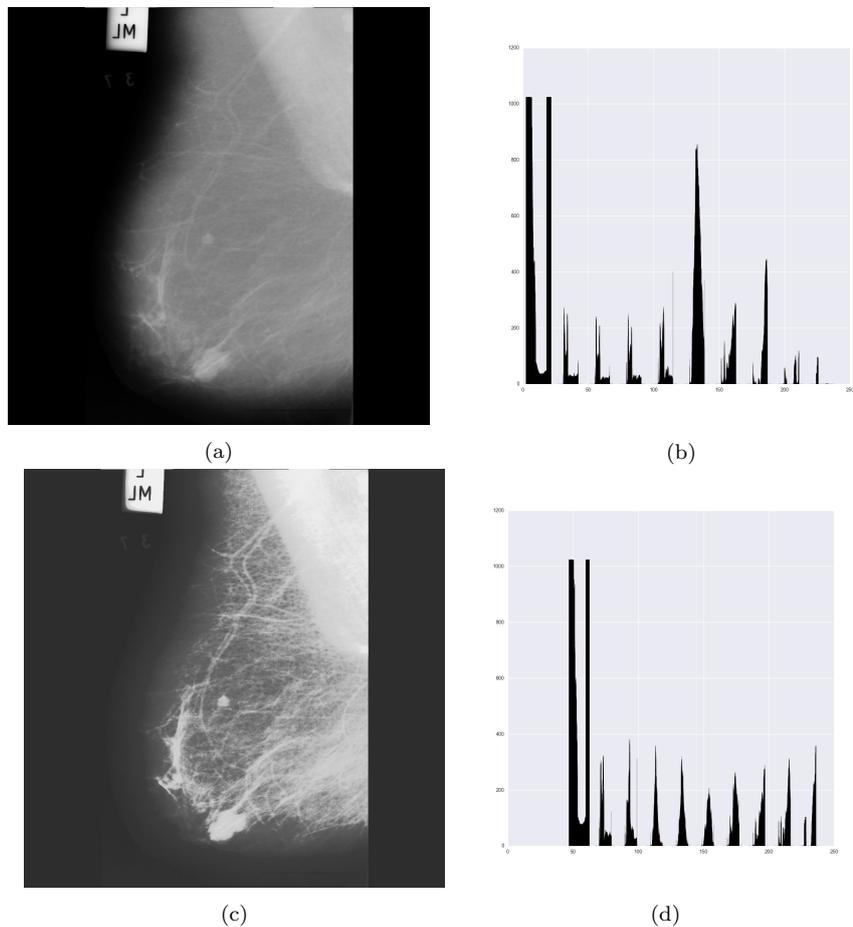


Figura 2.2: Mammografia relativa al caso #5 prima dell'elaborazione (a), e dopo l'applicazione dell'RMSHE (c). In (b) e (d) i rispettivi istogrammi.

questa sezione descriveremo la procedura di estrazione dei campioni di immagine (*crop*), allo scopo di lavorare con un gran numero di immagini al prezzo di limitarsi a lavorare solo con piccole porzioni delle mammografie originali.

Per questioni di ottimizzazione dell'addestramento, è necessario fissare una volta per tutte la dimensione delle *crop*. A causa della risoluzione delle mammografie (1024x1024 pixel), si è scelto di usare *crop* con rapporto di dimensione 1:1. La dimensione del lato delle *crop* è stata fissata grazie a delle considerazioni svolte sui metadati delle immagini: per le immagini relative a mammografie contenenti lesioni vengono riportate il tipo di

lesione, le coordinate (con origine degli assi posta in basso a sinistra, una convenzione discordante con quelle adottate dalle librerie di manipolazione immagini) del centroide della lesione e il raggio in pixel della circonferenza circoscrivente la lesione. Si può notare come il raggio medio sia di 54 pixel, e tutte le lesioni tranne tre abbiano un raggio inferiore ai 64 pixel. Alla luce di questi dati, il valore di 64 è stato adottato come lato delle crop; inoltre tale dimensione è sottomultipla di 1024, proprietà che semplificherà alcuni passaggi all'interno dell'algoritmo di localizzazione della patologia. Il fatto di sottocampionare le (poche) patologie più estese è stato ritenuto un compromesso più che accettabile in quanto il software qui discusso intende svolgere da supporto al lavoro del medico specialista, andando a mettere in luce quelle patologie che per dimensioni, forma o scarsità di contrasto non siano facilmente identificabili dall'occhio umano. Le patologie più estese, grazie alle loro dimensioni abbondanti, non dovrebbero rientrare in questa categoria e pertanto dovrebbero essere rilevate dall'utilizzatore finale. Si noti inoltre che in questo modo molte crop conterranno oltre alla patologia una buona regione di tessuto circostante (nel seguito *contesto*), che è spesso utilizzato dagli stessi specialisti per discernere la reale natura di una potenziale lesione.

Selezionata la dimensione ottimale delle crop, viene effettuata la lettura automatizzata (*parsing*) della documentazione a corredo del database, identificando per ogni immagine contenente tessuto patologico le coordinate del centroide della lesione. A partire da queste coordinate vengono quindi estratte 22 crop per ogni mammografia, nel seguente modo:

- Una crop della mammografia priva di trasformazioni geometriche, centrata sul centroide della lesione (nel caso di immagine patologica) o nel punto centrale dell'immagine (nel caso di immagine di mammella sana);
- Una crop estratta da una versione della mammografia originaria riflessa rispetto all'asse verticale, centrata sulla nuova posizione del centroide della lesione (nel caso

di immagine patologica) o nel punto centrale dell'immagine (nel caso di immagine di mammella sana);

- Dieci crop ottenute ruotando la mammografia di un angolo compreso tra 0 e  $2\pi$  e distribuito in modo uniforme, al fine di aumentare i dati (le crop così estratte hanno il centro uguale a meno di rotazioni ma diverso contenuto nelle zone più prossime agli angoli) e migliorare l'invarianza per rotazioni delle performances del classificatore, proprietà utile soprattutto per il riconoscimento di cluster di microcalcificazioni presenti nelle immagini e lesioni spicolate. Tali crop vengono estratte sia dall'immagine originaria che dall'immagine riflessa da cui si è estratta in precedenza una crop.

Al fine di addestrare la rete le crop ottenute vengono convertite da pgm a bmp in quanto il formato in cui il database viene distribuito non è supportato da caffe. Prima di procedere con l'addestramento, le crop, come da convenzione in un lavoro di machine learning, vengono divise in tre gruppi:

- Un gruppo di addestramento, o *training*, composto dalle crop estratte dalle prime 193 immagini;
- Un gruppo di test, usato da caffe stesso per controllare lo stato dell'addestramento, composto dalle crop estratte dalle immagini da 194 a 291;
- Un gruppo di validazione, usato per costruire le ROC e valutare le performances del classificatore, composto dalle restanti crop e mai sottoposte alla rete prima dei test finali e a eventuali lesioni secondarie presenti nelle immagini usate per i due set precedenti (per ogni immagine di test e addestramento viene adoperata solo la prima lesione elencata nella documentazione dei dati in modo da aumentare i casi patologici nel set di validazione, altrimenti troppo esigui).

Si noti come al termine di questa procedura si è passati da 322 a 7084 immagini le quali, sebbene parzialmente correlate, permettono un buon addestramento della rete, che sarà discusso nel prossimo capitolo.

# Capitolo 3

## Architettura della rete e fase di Training

In questo capitolo saranno descritte l'architettura della CNN realizzata per affrontare il problema della classificazione di crop e per la localizzazione di eventuali patologie, descrivendo i layers adottati, e alcuni aspetti salienti della fase di training. Il secondo punto sarà di particolare interesse in quanto esibirà due diverse procedure adottate: addestrare una rete dall'inizio e specializzare una rete generica preesistente e preaddestrata.

### 3.1 La rete: CaffeNet\_MAMMO

A causa della notevole complessità delle CNN, crearne una da zero è un lavoro che da solo può richiedere anni grazie anche all'enorme numero di variabili in gioco e alla minor solidità delle basi matematiche alle spalle di questi modelli. Per questo motivo non è inusuale utilizzare reti preesistenti nate per scopi simili e applicarle ad un nuovo problema, adattando i parametri necessari. Questa strada è stata seguita anche in questo lavoro, pertanto si descriverà la rete, della quale sono stati cambiati alcuni parametri numerici oltre che il numero di classi possibili, da 1000 a 2.

CaffeNet\_MAMMO è una CNN sviluppata a partire da CaffeNet, una rete specializzata nella classificazione di immagini naturali che ha dato risultati superiori ai vincitori della PASCAL\_VOC Challenge dal 2007 al 2012 e della ILSVRC del 2012, in grado di discriminare tra 1000 possibili classi di appartenenza. Grazie alla sua potenza essa è impiegata in numerosi contesti, anche sensibilmente diversi dallo scopo originario, ed è stata scelta per questo lavoro, alla luce sia di precedenti sue applicazioni in campo medicale, che della sua notevole versatilità grazie al grande numero di strati convoluzionali coinvolti.

La rete, illustrata in figura 3.1, è così composta: nel primo strato legge le immagini, caricandole in memoria in pacchetti di dati (batch) che saranno analizzati in parallelo. Ogni crop in questa fase viene ritagliata ulteriormente in modo casuale in un quadrato di lato 60 in modo da simulare lo scenario (comune nelle applicazioni reali) di patologie non totalmente inquadrare in una crop, e le viene sottratto il valore di intensità medio dell'intero database per eliminare offset comuni all'intero set. I batch sono generati prendendo le crop in ordine casuale e attribuendo ad ogni crop un'etichetta: 0 nel caso in cui si riferisca a tessuto sano, 1 nel caso in cui si riferisca a tessuto malato. Questa fase viene distinta a seconda del caso in cui i pesi della rete varino (TRAIN), nel qual caso i batch sono creati a partire dal gruppo di addestramento e contengono 128 elementi cadauno, o nel caso in cui i pesi siano statici e si stiano valutando le prestazioni della rete (TEST), situazione in cui i batch sono presi dal gruppo di test e hanno dimensione pari a 64.

A questo punto si susseguono due blocchi identici composti da quattro layers cadauno:

- Un layer convolutivo, nel primo blocco lo spostamento della finestra (*stride*) è inferiore alla dimensione del kernel, causando una sovrapposizione tra le varie windows nel primo passaggio. La sovrapposizione è voluta per migliorare il riconoscimento nel caso in cui non si abbia perfetto allineamento tra margine della lesione e finestra di convoluzione;

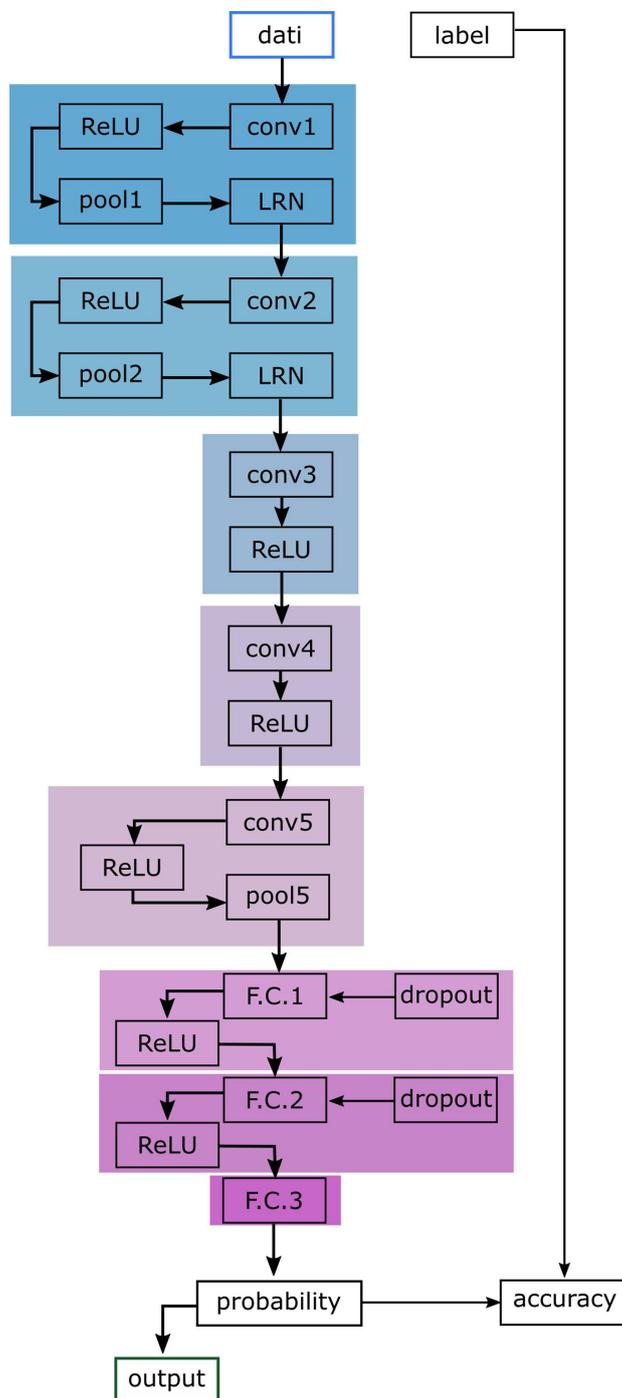


Figura 3.1: Schema della CNN impiegata per l'addestramento del classificatore. Sono riquadrati gli insiemi di layers costituenti un unico blocco logico (strato di elaborazione + strati ausiliari). Su campo bianco sono indicati gli strati di I/O e di confronto tra risultati prodotti e classificazioni attese.

- Una ReLU agente sul layer convolutivo;
- Un Pooling Layer che filtra l'uscita della ReLU, con kernel e stride uguali nei due blocchi;
- Un layer di normalizzazione della risposta locale (LRN), all'interno del quale ogni valore in uscita dal pooling layer viene moltiplicato per un termine definito dall'intorno dell'input stesso, portando a un fattore correttivo come quello in eq. 3.1:

$$\frac{1}{(1 + (\alpha/n) \sum_i x_i^2)^\beta} \quad (3.1)$$

ove  $n$  è il numero di pixel contenuti nell'intorno,  $x_i$  è l'intensità dell' $i$ -esimo pixel,  $\alpha$  e  $\beta$  due costanti scelte empiricamente. Entrambi i blocchi usano gli stessi parametri numerici.

Seguono due blocchi composti da strato convolutivo e ReLU, infine l'ultimo blocco contenente architettura deep è composto da uno strato convolutivo, una ReLU e un layer di Pooling.

Gli ultimi tre blocchi sono, al contrario dei precedenti, privi di unità convolutive, e sono quelli nei quali avviene la classificazione vera e propria dopo i cambi di rappresentazione introdotti dalle precedenti elaborazioni: in ognuno di essi troviamo uno strato completamente connesso, all'interno del quale ogni neurone esegue semplicemente un prodotto scalare tra tutti i dati in ingresso e i corrispondenti vettori peso, diversi per ogni percettore (e abbandonando quindi le features map usate nei layers convolutivi), e una ReLU per diminuire il carico computazionale. Per ridurre il rischio di overfitting, ossia la degenerazione di una rete "troppo" addestrata e quindi estremamente performante sui gruppi di addestramento e test ma incapace di generalizzare, sui primi due blocchi è attivo il Dropout dei pesi: in ogni ciclo di addestramento, solo una frazione di collegamenti sinaptici scelti a caso tra tutti i possibili è attiva e viene modificata.

Chiude il tutto uno strato di verifica dove viene valutata l'accuratezza della previsione della rete confrontando la risposta della rete all'etichetta della crop, usando come funzione costo per verificare lo stato dell'apprendimento una curva logistica valutata in funzione dell'errore di classificazione commesso (eventualmente multilogistica nel caso di classificatore a più di due classi).

L'algoritmo di addestramento corregge i pesi (inizialmente assegnati casualmente da una distribuzione gaussiana diversa per ogni strato) tramite back propagation e salva i risultati in un file binario in formato proprietario con estensione *.caffemodel*.

La rete usata per la classificazione vera e propria (*deploy*) è identica ma priva di valutazione dell'accuratezza e dei layers di input, in quanto i dati vengono direttamente inseriti dallo script python che si occupa di lanciare la classificazione e di interpretare i risultati.

## 3.2 Addestramento: il Solver del modello

In caffe, oltre al file di definizione della rete (sia di addestramento che di deploy), è possibile specificare tutti i parametri dell'addestramento in un file esterno, anch'esso scritto secondo lo standard Google Protobuf, detto in gergo *Solver*. Anche questo file nasce dal principio di condivisione alla base di caffe, rendendo tutti i parametri più propri dell'addestramento facilmente accessibili e modificabili indipendentemente dal proprio linguaggio di elezione e dal file contenente i parametri architetturali della rete.

In particolare il compito del solver è di orchestrare le fasi di addestramento, istanziando la rete, scegliendo il modello di apprendimento, ottimizzandola ad ogni iterazione cambiando il verso in cui i dati fluiscono (dall'inizio alla fine per classificazione e valutazione, dalla fine all'inizio per computo dei gradienti dei pesi e addestramento), chiamando le fasi di test e occasionalmente salvando lo stato dei pesi per realizzare un backup.

Tra i diversi tipi di modelli di apprendimento offerti da caffe, in questo lavoro si è

scelto di adottare quello basato sulla discesa stocastica del gradiente (SGD) in quanto poco oneroso in termini di prestazioni e, una volta opportunamente impostato, si rivela una scelta ottimale per evitare di incorrere in minimi locali del gradiente nei quali i pesi possano bloccarsi portando a un addestramento di bassa qualità (se non praticamente assente) della rete. In un modello SGD ad ogni iterazione (analisi di un batch completo) i pesi  $W_t$  sono aggiornati secondo la coppia di equazioni 3.2 e 3.3.

$$V_{t+1} = \mu V_t - \alpha \nabla L(W_t) \quad (3.2)$$

$$W_{t+1} = W_t + V_{t+1} \quad (3.3)$$

Ove

$V_t$  rappresenta la variazione dei pesi al tempo  $t$ ,

$\nabla L(W_t)$  il gradiente della funzione costo calcolata rispetto alla configurazione dei pesi al tempo  $t$ ,

$\alpha$  e  $\mu$  rappresentano rispettivamente il learning rate e il momento, che fanno da costanti di proporzionalità che regolano cosa influenzi maggiormente l'apprendimento della rete tra la variazione della funzione costo e l'ultima correzione adottata.

Mentre il momento una volta impostato rimane tale e quale per tutto l'addestramento, il learning rate può variare durante l'addestramento secondo diverse politiche: per addestrare CaffeNet\_MAMMO la politica di aggiornamento adottata è la *stepwise*: ogni  $N$  iterazioni (*step*) il learning rate viene moltiplicato per un termine minore di 1 in modo che l'importanza della funzione errore nel tempo si riduca, accelerando le prime fasi dell'addestramento e rallentando le ultime in modo da fare correzioni più raffinate verso la fine del training.

Oltre ai suddetti parametri è possibile modulare l'intera correzione dei pesi moltiplicando quanto ottenuto dallo SGD per un fattore di degradazione, o *Weight Decay*, al fine di controllare la rapidità di addestramento a prescindere dal valore del gradiente e

dalla percentuale di completamento del training.

Tutte le procedure di addestramento e produzione di risultati sono state eseguite su una macchina messa a disposizione dal gruppo XRAY Tomography del DIFA, dotata di processore Intel Core i7-6700 (4 cores @ 3.40 GHz, 8 MB cache), scheda video NVidia GeForce GTX 750 (512 CUDA cores @ 1085 MHz, 1024 MB VRAM GDDR5) e 32GB di memoria RAM DDR3.

### 3.3 Addestramento con Fine Tuning

Come abbiamo visto in precedenza, l'architettura della rete è presa da un altro progetto, CaffeNet, adattata alle necessità del problema in studio e quindi addestrata da zero. Tuttavia, date le ottime performances di CaffeNet nel riconoscimento di immagini generiche, risulta interessante mettere alla prova la sua capacità di generalizzare andando ad applicarla su un problema nuovo: grazie al suo allenamento su una casistica estremamente differenziata, infatti, si suppone che i filtri che ha già appreso al suo interno siano in grado di riconoscere anche le strutture caratteristiche di un carcinoma mammario; l'unica cosa che bisogna insegnare alla rete è quindi l'attribuzione di un significato a tali lesioni.

Uno dei motivi per cui è stata scelta caffe tra le varie alternative è proprio la possibilità di eseguire questo tipo di operazione: all'interno del *Model Zoo* del BLVC sono infatti presenti gli stessi pesi dei link di CaffeNet con cui tale rete ha vinto le principali competizioni di CV, e grazie alle funzioni della libreria è possibile procedere ad un nuovo addestramento di CaffeNet\_MAMMO scegliendo quali layers ereditare da CaffeNet e quali invece imparare ex novo modificando poche righe del prototxt della rete. Nei prossimi capitoli ci si riferirà a questa rete come CaffeNet\_MAMMO2.

Risulta quindi intuitivo come tutti gli strati convoluzionali siano stati ereditati in toto dalla rete preaddestrata. A causa della natura a colori delle immagini per cui CaffeNet

è stata sviluppata tutte le immagini di addestramento, test e validazione sono state convertite da PGM monocromatici a BMP a tre canali: il canale rosso porta l'informazione vera e propria mentre i canali verde e blu sono costituiti da immagini completamente nere in modo da essere ignorati per la costruzione dei pesi della rete, riducendo il carico computazionale ma rispettando i requisiti di input del classificatore preaddestrato. Un altro possibile atteggiamento sarebbe stato di modificare il primo strato convoluzionale, riaddestrandolo in modo da accettare immagini monocromatiche, ma questo sarebbe andato contro l'obiettivo teorico-speculativo del finetuning: riaddestrandolo, lo strato avrebbe appreso nuovi filtri e quindi qualunque considerazione sulla generalizzazione del sistema visivo appreso per la ILSVRC challenge sarebbe stata priva di senso. Il metodo adottato invece fa sì che l'apparato visivo non venga alterato, mentre le uniche modifiche rimangono nell'ambito del sistema decisionale.

A questo punto la rete è in grado di discriminare gli oggetti allo stesso modo della rete madre. Per insegnarle a dare un'etichetta agli output coerente con il nuovo problema è quindi necessario agire sugli strati decisionali, ossia i MLP costituenti gli ultimi tre strati della rete: oltre a ridurre la rosa delle possibilità da 1000 a 2, infatti, i perceptroni devono capire di nuovo quali input producano quale output. Come solver si è adoperato lo stesso impiegato per l'addestramento da zero della rete.

L'obiettivo di questo secondo addestramento è duplice: in primo luogo risulta interessante realizzare un confronto tra una rete nata per uno scopo specifico ma addestrata su un piccolo database e una rete nata per uno scopo generico ma addestrata su decine di migliaia di immagini naturali presentanti una varietà irripetibile di forme e texture apprendibili, permettendo di fare ipotesi sul miglior modello di apprendimento per queste reti di ispirazione biologica. In secondo luogo, la possibilità di riaddestrare reti per i più differenti problemi è sempre interessante per effettuare considerazioni sulla possibilità di generalizzare il lavoro su altre immagini diagnostiche, sia inerenti diversi organi e tessuti coinvolti, sia relative a diverse tecniche diagnostiche.

## Capitolo 4

# Classificazione di sezioni di immagine

In questo capitolo sarà illustrato l'algoritmo di classificazione impiegato per caratterizzare le performances della rete in entrambe le sue versioni. Saranno inoltre valutate le sue prestazioni mediante il metodo delle ROC ed eseguito un confronto con lavori presenti in letteratura, tenendo conto della limitatezza del dataset impiegato. Gli oggetti della classificazione in questa parte di lavoro saranno le crop provenienti dal gruppo di validazione presentato in sezione 2.2.2 in quanto, a causa del problema in oggetto e dell'approccio adottato, risulta di scarso significato tentare di classificare intere immagini in quanto le lesioni occupano mediamente porzioni dell'immagine dell'ordine del 3 per mille (assumendo un'estensione pari a un quadrato di lato pari al lato medio), pertanto trascurabili nell'ambito di una scansione dell'intera immagine. Per ulteriori dettagli sull'analisi di immagini intere si rimanda al cap. 5.

## 4.1 L'algoritmo di classificazione

L'algoritmo di classificazione è scritto in linguaggio Python. Tale linguaggio è stato adottato per le sue qualità di immediatezza e semplicità nel manipolare dati multidimensionali e immagini e per la sua capacità di poter lanciare codice precompilato C++ in parallelo a comandi interpretati, rendendo possibile la creazione di un prototipo agile e modulare (grazie alla parte interpretata) ma senza rinunciare all'efficienza del codice (deputando le parti computazionalmente più gravose a funzioni compilate). Sebbene anche MATLAB offra queste caratteristiche (oltre che la compatibilità con caffè), si è preferito Python in quanto open-source.

Nella parte preliminare, lo script carica la rete utilizzando il modello definito all'interno di un protobuf simile a quello usato per la rete d'addestramento ma con un layer di input tarato per una singola immagine caricata direttamente dalla memoria e non dall'hard disk, e i pesi dei link ottenuti dall'ultima iterazione dell'algoritmo. Viene quindi definito un oggetto proprio del framework di caffè, un *transformer*. Il transformer si occuperà in seguito di applicare tutte le trasformazioni sia di preprocessing che di ritaglio e riscaldamento dell'immagine, che viene caricata in memoria come un array Numpy, svolgendo quindi funzioni simili al layer di input della rete di training. Sebbene i preprocessing non siano di utilità ai nostri fini in quanto tutte le immagini provengono dalla stessa fonte e hanno già tutte subito il medesimo trattamento, come descritto nel capitolo 2, tale oggetto riveste un ruolo chiave qualora si vogliano classificare immagini provenienti da fonti differenti o ottenute in condizioni di esposizione o contrasto differenti, in modo da fornire al classificatore oggetti il più omogenei possibile. In particolare il transformer adoperato in questo lavoro:

- riorganizza la rappresentazione in memoria dell'immagine in modo che le prime due dimensioni siano le coordinate x e y e i canali siano la dimensione più esterna;<sup>1</sup>

---

<sup>1</sup>Indipendentemente dal fatto che le immagini siano a colori o in scala di grigio, caffè ragiona sempre in termini di immagini a colori in quanto casistica più generale.

- Rinormalizza le intensità dei pixel dell'immagine ad una scala arbitraria ( $0 \rightarrow 255$ ), utile nel caso si usino immagini provenienti da diversi scanner con diversa profondità di colore;
- Nel caso di immagini rgb cambia l'ordine dei canali da BGR (standard Python) a RGB;
- Sottrae l'intensità media dai pixel dell'immagine, in modo da eliminare eventuali offsets dovuti a calibrazioni diverse degli strumenti;
- Riorganizza nuovamente la rappresentazione in memoria dell'immagine in modo da farla combaciare con la forma del primo layer convoluzionale.

A questo punto, ogni crop viene caricata in memoria, trasformata e sottoposta alla rete, la quale procede con la classificazione fornendo nell'ultimo strato un array di due elementi, uno per possibile classe, ognuno contenente la probabilità che la crop vi appartenga. Come ulteriore misura per filtrare i risultati può essere applicata una soglia minima sulla probabilità di una crop di appartenere alla classe patologica: valori bassi di tale soglia faranno sì che tutte le crop relative a tessuto malato siano identificate come tali, ma che lo siano anche numerose crop sane con un conseguente calo della specificità. Alzando la soglia si aumenta la specificità al prezzo però di un calo della sensibilità, pertanto è fondamentale trovare un valore che offra un buon compromesso. Tramite analisi ROC è possibile identificare in modo quantitativo tale compromesso una volta decisa la sensibilità o la specificità desiderata.

## 4.2 Risultati

Le performances del classificatore sono di seguito riportate e interpretate. Oltre che con considerazioni di carattere generale/intuitivo, la bontà di classificazione sarà valutata anche sfruttando le ROC. Per generare i punti necessari alla loro valutazione sono state

eseguite numerose istanze del classificatore analizzando tutte le crop di validazione. Tra una esecuzione e la successiva è stata variata la soglia di probabilità al di sopra della quale si considera patologica un'immagine, partendo da un valore estremamente basso (0.01) fino ad una probabilità del 100%, con incrementi dell'1% per iterazione. A causa della rapidità di variazione della sensitività ad alti valori, l'intervallo relativo a soglie comprese tra 0.99 e 1 è stato ulteriormente suddiviso in 10 sottointervalli in modo da avere un buon ricoprimento della ROC.

#### 4.2.1 Classificazione con CaffeNet\_MAMMO

Sono riportati in figura 4.1 gli andamenti di sensitività, ovvero il rateo di risposte corrette del classificatore sottoposto a immagini patologiche (veri positivi nel seguito), e specificità, ossia la capacità del classificatore di fornire come risposta la classe patologica solo in presenza di immagini ad essa appartenente, assimilabile al tasso di veri negativi<sup>2</sup>, in funzione della soglia di riconoscimento e in figura 4.2 la ROC del nostro classificatore.

Come si può osservare, al crescere del valore di soglia la sensitività cala (ossia sempre più immagini patologiche vengono ignorate dall'algoritmo). Di contro, l'aumento di tale valore aumenta la specificità del classificatore, ossia la sua capacità di non misclassificare porzioni di tessuto sano. Nel caso in oggetto si osserva come per un grande intervallo ([0:0.9]) dei valori di soglia le due curve sono disgiunte, evidenziando come per valori bassi e medi di soglia il classificatore abbia un'ottima sensitività ma una specificità mediocre. È necessario utilizzare valori molto elevati di soglia per abbassare i falsi positivi, collocando la soglia nella regione dove si intersecano, perdendo un buon 15% della sensitività ma guadagnando un aumento di specificità dell'ordine del 30%<sup>3</sup>. Alla luce di queste considerazioni quindi si può stimare 0.994 come un possibile optimum

---

<sup>2</sup>Per la costruzione delle ROC è necessario usare il complemento ad uno della specificità, ossia il rateo di falsi negativi, che per brevità sarà chiamato *aspecificità*.

<sup>3</sup>rispetto ai valori relativi a una soglia pari a 0.01

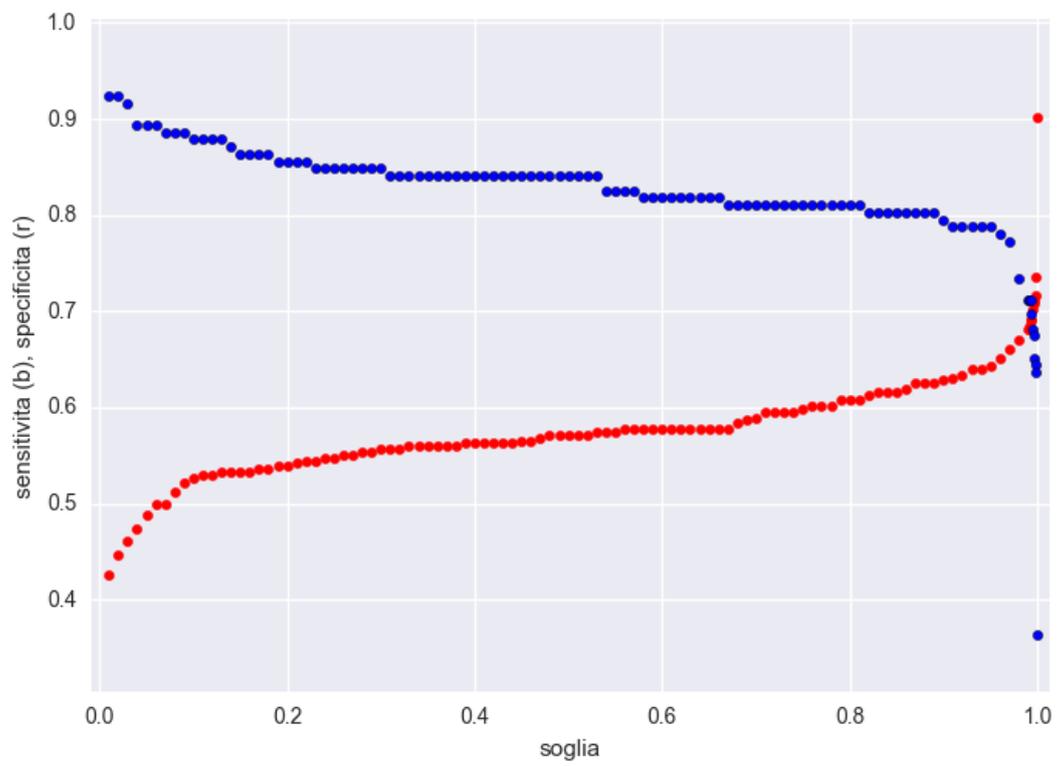


Figura 4.1: Andamento di sensibilità e specificità in funzione della soglia di detezione.

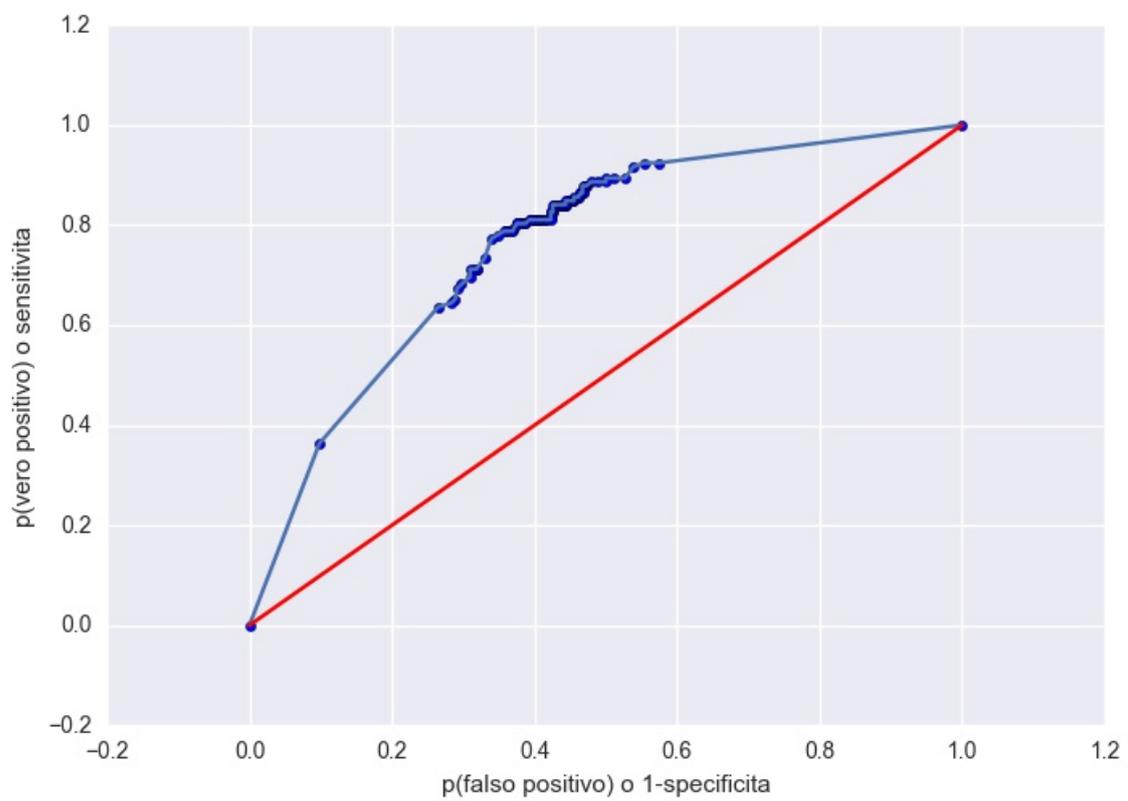


Figura 4.2: ROC del nostro sistema (linea e punti blu) e *chance line* (linea rossa).

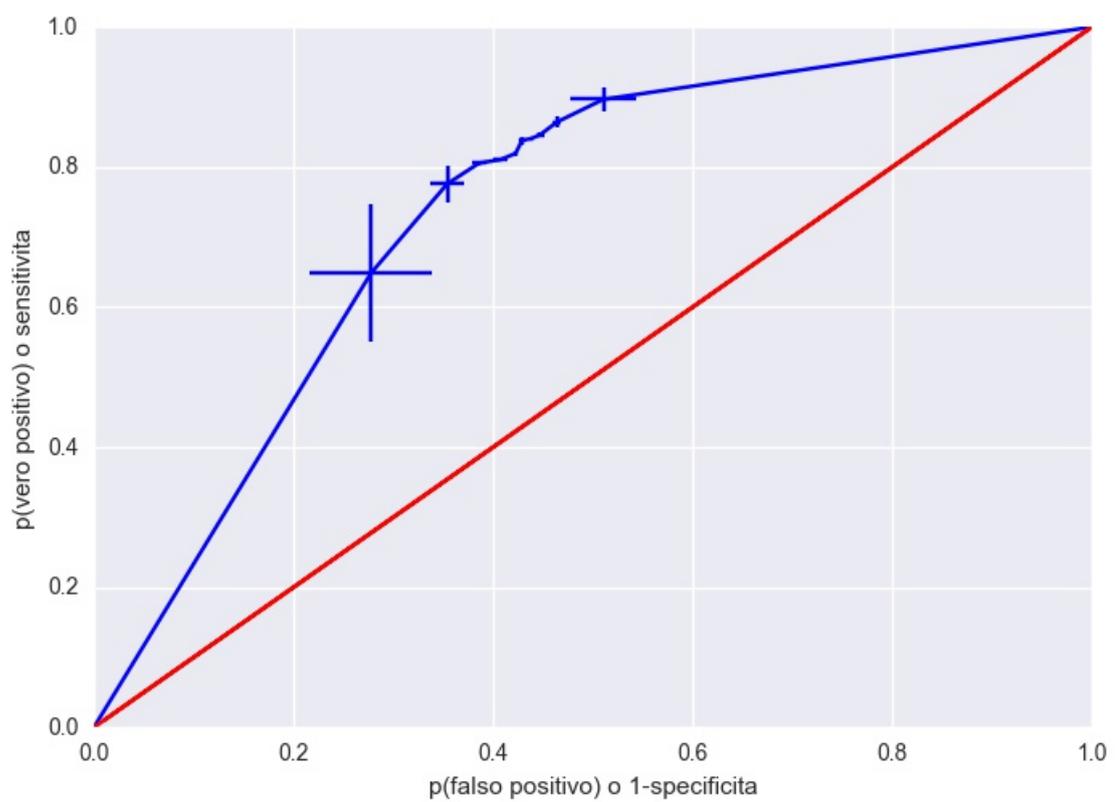


Figura 4.3: ROC mediata del nostro sistema con errori delle misure (linea blu) e *chance line* (linea rossa).

per il valore di soglia, identificando un buon compromesso tra specificità e sensibilità, entrambe pari a 0.70 per detta soglia.

La curva ROC mostra un andamento crescente sopralineare come atteso, seppur con un raggruppamento molto poco omogeneo dei punti, riflettendo la situazione osservabile nel primo grafico: a bassi valori di probabilità di falso positivo (1-specificità) corrisponde un singolo punto, relativo alla soglia massima di 1, mentre la maggior parte dei punti, inclusi quelli che corrispondono a soglie con differenze dell'ordine dell'1 per mille, si trovano nella zona centrale del grafico. Per conferire dignità statistica ai dati, grazie anche al notevole numero di punti disponibili, si è scelto di aggregare i dati, calcolando l'area sotto la ROC (AUC) non sulla curva riportata in precedenza ma su quella presente in figura 4.3, nella quale ogni punto corrisponde alla media di 11 punti della curva originale (fig. 4.2), mediando sia sulla sensibilità che sull'aspecificità.

Adoperando dette medie è possibile stimare per ogni misura un'incertezza e quindi fornire una AUC maggiormente consistente. A causa di una zona a sensibilità costante per valori di soglia compresi tra 0.31 e 0.53, non è possibile clusterizzare i valori senza ottenere almeno un insieme a varianza nulla (a meno di escludere dei punti dal calcolo), ma questa irregolarità non si ritiene abbia effetti troppo rilevanti nella stima dell'AUC e del suo errore. Con i dati ottenuti si è stimata una AUC pari a

$$AUC_{CaffeNet\_Mammo} = 0.68 \pm 0.08.$$

Per verificarne la rilevanza statistica si è calcolata la statistica U di questo valore in modo da stimare quantitativamente quanto la ROC differisca in modo significativo dalla chance line (corrispondente a un classificatore totalmente casuale [9], studiando lo scostamento della AUC dal valore atteso di 0.5 nell'ipotesi nulla che non vi siano differenze statisticamente significative tra le mediane delle due popolazioni, ossia nel nostro caso

tra la ROC e detta bisettrice, secondo la relazione

$$z = \frac{AUC - 0.5}{\sigma_{AUC}}$$

ottenendo un valore di  $z$  pari a 2.43, superiore al valore critico di 1.97 della statistica di Wilcoxon-Mann-Whitney corrispondente ai 10 gradi di libertà disponibili, confermando in modo inequivocabile la buona qualità della classificazione ottenuta.

Per un confronto con quanto attualmente disponibile in materia si rimanda alla sezione 5.3, dove sarà ritratto un quadro conclusivo confrontando il presente lavoro con la letteratura sull'argomento.

Se confrontato con lo stato dell'arte dei classificatori puramente convolutivi [11], il classificatore proposto in questo lavoro presenta performances comparabili con le AUC di tutti i metodi standalone ivi proposti ad eccezione del CNN3 (tra i migliori classificatori a singolo stato basato su CNN attualmente disponibile), il che rappresenta un risultato estremamente positivo se si considera la scarsità di immagini di addestramento inizialmente disponibili, evidenziando la bontà del metodo qui proposto. Si noti tuttavia come il classificatore qui realizzato sia puramente convolutivo, e le sue performances rimangono nettamente inferiori rispetto a classificatori più avanzati che tengono conto oltre che dell'immagine in sé anche di features appositamente costruite.

## 4.2.2 Classificazione con CaffeNet\_MAMMO2

Nonostante le attese, in questo caso specifico il fine tuning della rete non ha dato risultati incoraggianti: i filtri appresi per il riconoscimento di immagini generiche infatti si sono rivelati totalmente inadatti se applicati alle immagini mammografiche impiegate, facendo sì che per ogni immagine la rete fosse incapace di proporre classificazioni di alcun tipo: l'output proposto dalla rete infatti è per ogni crop pari a  $[0.5, 0.5]$  indipendentemente da quali e quanti strati convoluzionali si ereditassero dal modello base, ottenendo cambi

di prestazioni solo quando ormai praticamente l'intera rete stava venendo riaddestrata (mantenendo dalla configurazione base solo i pooling layers). Tuttavia questa situazione non è di alcun interesse in quanto il grosso del sistema visivo di una CNN risiede nei layer a convoluzione. Riaddestrandoli da capo essi hanno assunto le stesse caratteristiche di CaffeNet\_MAMMO, portando in pratica ad una rete identica.

Si può quindi concludere che il modello originale sia troppo sofisticato per poterne prendere in prestito alcuni strati senza portare a effetti devastanti sulle performances di classificazione, specie nel caso di un cambio di destinazione d'uso così rilevante<sup>4</sup>.

---

<sup>4</sup>A causa di questo insuccesso, nel capitolo 5 non si tratterà delle performances di CaffeNet\_MAMMO2 come localizzatore in quanto l'algoritmo di proposta delle ROI è sostanzialmente una variazione di quello di classificazione e non darebbe risultati significativi

## Capitolo 5

# Localizzazione della patologia in immagine completa

In questo capitolo si discuterà dell'algoritmo di localizzazione della patologia, appoggiato fortemente sull'algoritmo di classificazione, e se ne discuteranno le performances. L'obiettivo di questo software è di fornire al medico specialista un numero ristretto di siti di indagine (o di cluster di siti) relativo a potenziali collocazioni della lesione, senza però sostituirlo: come affermato in precedenza (cap. 2) ci si aspetta che molte delle lesioni più estese non siano identificabili a causa del sottocampionamento cui sono andate incontro, pertanto l'identificazione di tali masse è sicuramente affidata allo specialista. Tuttavia questa caratteristica non viene sentita come una limitazione in quanto va ricordato che i software di CAD nascono dall'esigenza di aiutare a focalizzare l'attenzione su quei dettagli scarsamente visibili che possano sfuggire con elevata probabilità.

A causa dell'alto rateo di falsi positivi ottenuto, i comuni metodi quantitativi per la valutazione della bontà della localizzazione non risultano applicabili. Si è pertanto preferito dare un'idea qualitativa delle performances del localizzatore, illustrando i risultati ottenuti su alcune immagini esemplificative.

## 5.1 L'algoritmo di localizzazione

In questa sezione analizzeremo l'algoritmo di localizzazione. Per valutare se una sezione possa essere o meno interessante viene impiegato il classificatore descritto al capitolo precedente.

Caricata la rete e predisposto il transformer, il programma richiede all'utente la soglia che si intende impiegare. Tale possibilità permette all'utilizzatore di modificare il numero e il potenziale significato delle proposte del software: soglie più basse forniscono più punti, aumentando la probabilità di localizzare la patologia ma ottenendo anche un gran numero di suggerimenti errati; soglie più alte al contrario riducono il numero di falsi positivi, al prezzo però di rischiare di non ricevere l'informazione desiderata. Per questo motivo si invita l'utilizzatore a utilizzare un valore suggerito, calcolato a partire dalle ROC esposte nel capitolo precedente, che rappresenta l'optimum di default, e di modificare questo parametro solo per ulteriori indagini. Il programma controlla inoltre che la soglia di classificazione non assuma valori al di fuori del range  $(0:1]$  in modo da prevenire esecuzioni del programma prive di rilevanza diagnostica.

Impostato il valore desiderato, l'immagine viene caricata in memoria e tramite una sogliaatura che esclude tutti i livelli di grigio inferiori a 1.5 volte il livello di grigio minimo (valore determinato euristicamente) vengono eliminati tutti i pixel appartenenti allo sfondo e al bordo della mammella, all'interno del quale non essendo presente tessuto ghiandolare non può essere presente un tumore agli stadi iniziali. Viene calcolata l'intensità media e minima dei pixel non nulli di quest'ultima immagine.

Si definisce quindi la dimensione della finestra di scansione: tale numero rappresenta l'indeterminazione con cui si voglia conoscere la posizione della patologia, di default pari a 64, e a partire da questo valore sono calcolati una serie di parametri accessori che serviranno a definire gli estremi dei loops impiegati in seguito. Un parametro importante è lo stride della finestra di scorrimento, pari a  $1/4$  della dimensione della stessa, scelto in modo da aumentare la probabilità di ottenere una buona corrispondenza tra la regione

che sta venendo esaminata e quella in cui la patologia è veramente collocata. Questo fa sì che patologie molto piccole possano essere caratterizzate da una clusterizzazione di regioni proposte, agevolandone l'identificazione.

Impostati tutti i parametri, parte il cuore dell'algoritmo: la finestra di scansione viene collocata al vertice superiore sinistro ed estrae una crop dell'immagine, se la sua intensità media è inferiore all'intensità minima calcolata in precedenza la crop viene ignorata in quanto relativa a sfondo e/o bordo, altrimenti l'immagine viene mantenuta, riscalata a 64x64 nel caso in cui non abbia già queste dimensioni e, a causa di limitazioni di caffè, salvata su disco su un file temporaneo che viene immediatamente ricaricata con il loader della libreria<sup>1</sup>. Su questa versione viene applicato il transformer e la sottrazione dell'intensità media e infine avviene una classificazione: qualora la probabilità di appartenenza alla classe delle immagini patologiche sia superiore alla soglia di classificazione, il software memorizza le coordinate in un array, altrimenti la finestra di scansione viene spostata al punto successivo e il procedimento si ripete. L'utente è informato in tempo reale dello stato di avanzamento dell'elaborazione e di quando sia stata identificata una patch patologica, fornendo la possibilità di conservare un'immagine della stessa per future elaborazioni.

Conclusa la scansione dell'immagine, le coordinate identificate vengono salvate su file di testo e viene archiviata una versione dell'immagine con sovrapposte le regioni patologiche proposte dalla rete.

Per caratterizzare le performances dell'algoritmo è stato utilizzato l'intero database mini-MIAS: sebbene buona parte delle immagini contengano le crop utilizzate per l'addestramento infatti si è ritenuto che l'utilizzo di crop non allineate alla patologia in modo accurato come quelle usate per il training, unitamente al fatto che la stragrande maggio-

---

<sup>1</sup>Questa operazione di salvataggio e riapertura è estremamente inefficiente in termini di tempi di calcolo ma necessaria in quanto il loader di caffè carica le immagini come array sui quali intervenire è estremamente diverso a causa della loro forma non conforme agli standard del linguaggio Python. Tuttavia visti i tempi complessivamente ridotti di esecuzione dell'algoritmo (20 secondi per immagine completa, pochi ms per crop) non si è ritenuto necessario intervenire per risolvere questo problema

ranza di tali immagini non sono mai state effettivamente viste dal classificatore, faccia sì che non si debba temere in modo eccessivo l'overfitting (v. sez. 3.1) e permette inoltre di avere una serie di osservazioni inerenti alla massima varietà di lesioni e collocazioni, contesti inclusi.

## 5.2 Risultati

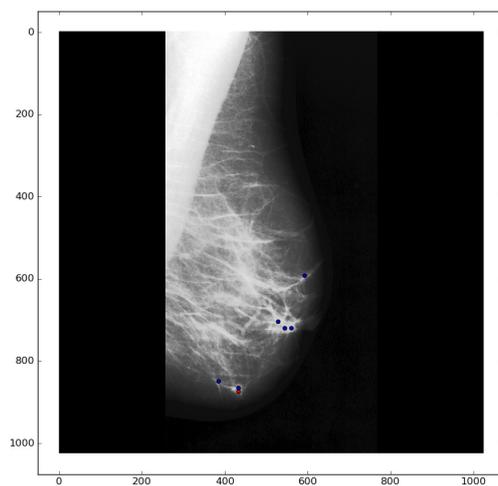
In questa sezione saranno illustrati i risultati ottenuti commentando l'andamento generale della localizzazione e illustrando, sia per immagini sane che per immagini patologiche, il miglior risultato, il peggiore e un'immagine con performances medie. A causa del numero non trascurabile di punti identificati in ogni immagine risulta privo di senso fornire i tipici stimatori delle performances di un localizzatore (come ad esempio la distanza della previsione dall'effettiva lesione o il tasso di falsi positivi per immagine), ma si ricorda ancora una volta come l'obiettivo del software sia di fornire un aiuto alla diagnosi, pertanto la valutazione sarà effettuata in termini di utilità dei punti evidenziati nel ridurre il tempo di lettura dell'immagine da parte dell'osservatore umano.

Le immagini seguenti presentano la mammografia originale migliorata tramite RM-HSE (v. sez. 2.2.1) con sovrainpressi in blu i centri delle lesioni predette dall'algoritmo (ogni ROI predetta è di forma quadrata con un lato di 64 pixel, la distanza minima tra due centri è pari a 16 pixel) e in rosso come riferimento il reale centri della lesione (per immagini sane il punto rosso è collocato nell'angolo superiore sinistro in modo tale che guardando tale angolo sia possibile immediatamente valutare la reale natura (*ground truth*) dell'immagine: se il punto è ivi presente l'immagine si riferisce a una mammella sana, se è assente significa che in base alla documentazione a corredo è presente almeno una lesione).

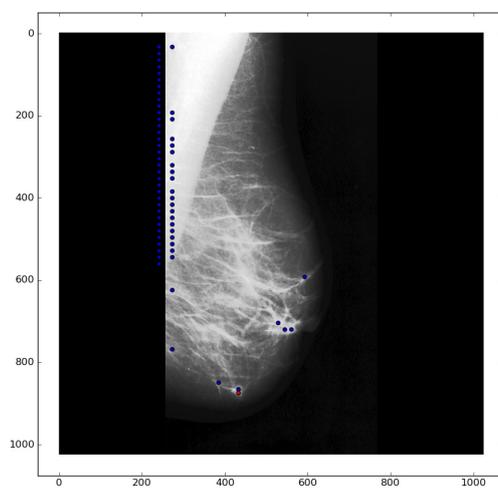
### 5.2.1 Localizzazione con CaffeNet\_MAMMO

Poichè l'algoritmo di localizzazione si basa su quello di classificazione è importante notare come la soglia adottata ne influenzi pesantemente le performances: soglie troppo basse aumentano il rateo di falsi positivi, marcando più zone del necessario, mentre soglie troppo alte aumentano i falsi negativi, nascondendo quindi lesioni. Per questo motivo, sebbene nel capitolo precedente si sia identificata 0.994 come una soglia ottimale, si è scelto di adoperare una soglia pari a 0.99 (corrispondente a una sensibilità pari a 0.72 e una specificità pari a 0.68) a causa della particolare applicazione dell'algoritmo: in contesto medico infatti è preferibile evitare un falso negativo piuttosto che un falso positivo. Tuttavia non è possibile neanche abbassare troppo la soglia per evitare di avere un'immagine risultante satura di spot di localizzazione e quindi sostanzialmente priva di utilità per una CAD.

Mediante l'algoritmo di localizzazione evidenzia circa una decina di punti per immagine all'interno della mammella, esibendo un raggruppamento degli stessi in prossimità delle lesioni come è lecito attendersi. Tuttavia in molti casi la densità di questi centri di proposta non è sufficiente per inserire un criterio di soglia sopprimendo quindi i punti isolati, lasciando allo specialista un'indeterminazione su 4 ~ 5 zone di interesse da interpretare. Sebbene nel caso ideale e nei prodotti commerciali si dovrebbe ottenere una singola ROI proposta per lesione, il numero ottenuto risulta un compromesso accettabile. A causa del ridotto dataset adoperato si osserva come l'algoritmo ha sviluppato una deformazione cognitiva: sono presenti infatti nel set di addestramento 5 immagini che presentano la lesione sul confine tra mammella e sfondo nero. Questo ha fatto sì che il software tenda a proporre come regioni patologiche tutti i punti iperintense prossimi ad una regione di intensità nulla, placche di identificazione incluse. Sebbene le placche siano un problema trascurabile, la misclassificazione dei bordi dal lato del torace del paziente rappresenta un problema rilevante che pone davanti al programmatore due possibili vie:



(a)



(b)

Figura 5.1: Immagine del caso #80 con overlay di localizzazione con (a) e senza (b) soppressione delle proposte di bordo.

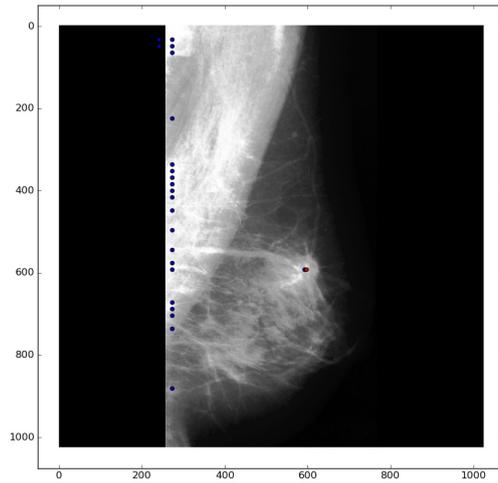
**Sopprimere i bordi:** A causa del pattern estremamente regolare dei bordi risulta semplice implementare un criterio di edge detection e soglia combinati che escluda le regioni dei bordi dalla localizzazione, così come classificarli e a posteriori eliminare le colonne con un numero eccessivo di punti (sempre corrispondenti ai bordi della mammella) (figura 5.1a);

**Mantenere i bordi:** A causa della presenza di tessuto ghiandolare anche nella regione più vicina al torace del paziente è possibile che essa contenga lesioni, pertanto un approccio più cauto può essere di mantenere i punti identificati senza criteri di soppressione (figura 5.1b).

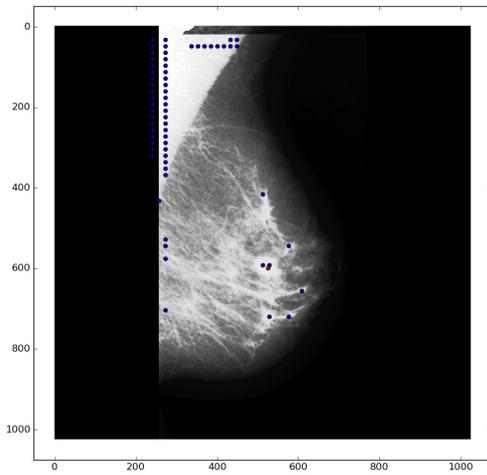
In questo lavoro si è adottata la scelta più cauta, evitando di perdere potere predittivo al prezzo di complicare la lettura dell'immagine da parte del radiologo. A causa di questa potenziale abbondanza di punti si è scelto di usare come marcatore della ROI il solo centro e non i suoi bordi, in modo da non compromettere la leggibilità complessiva dell'immagine. Si noti come questo bias sia eliminabile utilizzando un database più ampio, nel quale le lesioni sui bordi rappresentino una percentuale negligibile e quindi non sufficiente ad insegnare un criterio errato al classificatore. In caso di raggruppamento di regioni proposte, nella trattazione che segue tutti i punti sufficientemente vicini saranno contati come un'unica proposta.

Come da attesa le lesioni più estese vengono completamente ignorate dall'algoritmo, ma le più piccole sono quasi sempre identificate. Contando anche le lesioni di maggiore estensione, sono localizzate in modo corretto 77 immagini su 111 mammografie patologiche disponibili, con una sensibilità della classificazione pari al 69%, confrontabile con la sensibilità del classificatore dovuta alla soglia adottata, il che rappresenta un risultato che conferma la non casualità della proposta delle ROI.

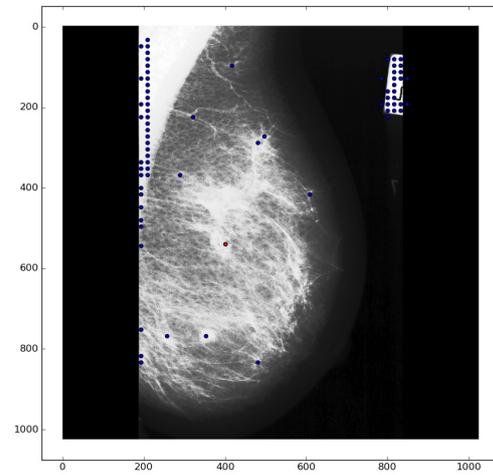
Per quanto concerne le immagini patologiche, il miglior risultato si ha per l'immagine #264 (proveniente dal gruppo di test), nella quale è presente un'unica lesione maligna su tessuto ghiandolare non grasso. In questo caso, come visibile in figura 5.2a, ad eccezione



(a)



(b)



(c)

Figura 5.2: Immagini relative ai casi #264 (a), #10 (b) e #256 (c), tutti patologici, con overlay di localizzazione (blu = predizione, rosso = ground truth).

delle proposte di bordo è presente un'unico centroide proposto praticamente coincidente con il centroide reale.

Il risultato medio può essere osservato in figura 5.2b, relativa all'immagine #10 (unica lesione circolare benigna su tessuto grasso, gruppo di addestramento), nel quale la lesione è identificata da una coppia di ROI proposte ma sono presenti altri 5 punti corrispondenti a falsi positivi.

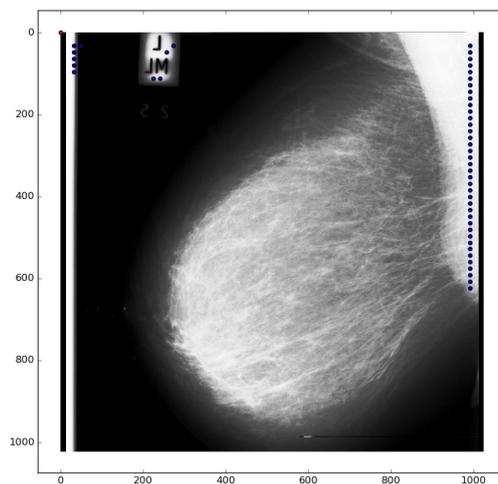
Il peggior risultato si ha per il caso #256 (calcificazioni maligne su tessuto ghiandolare), nella quale si hanno 8 ROI proposte ma tutte distanti dal luogo ove la patologia è situata, visibile in figura 5.2c.

Per quanto concerne le immagini non patologiche il miglior risultato si ha per il caso #135 (immagine con in prevalenza tessuto grasso), riportata in figura 5.3a, in cui ad eccezione dei falsi positivi di bordo non si hanno regioni proposte all'interno della mammella.

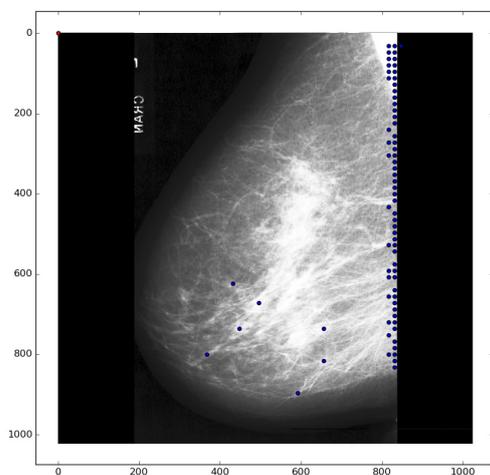
Nella maggior parte dei casi in un'immagine sana si ha una proposta di 5-8 regioni proposte (sempre senza contare i falsi positivi di bordo), come visibile in figura 5.3b (caso #173, tessuto prevalentemente grasso). È interessante notare come il numero di tasso di regioni proposte in un'immagine sana sia mediamente molto più basso ( $8 \text{ regioni proposte} / 3600 \text{ possibili regioni} = 0.002$ ) dell'aspecificità dovuta alla soglia impiegata (0.32).

Il peggior risultato si ha in situazioni analoghe a quella proposta in figura 5.3c (caso #129, tessuto denso-ghiandolare), nel quale il numero di regioni proposto è così elevato da obbligare il radiologo ad analizzare l'intera mammografia senza che sia effettivamente presente una sezione patologica al suo interno. Il numero di regioni proposte in questo caso diventa confrontabile con l'aspecificità del classificatore impiegato.

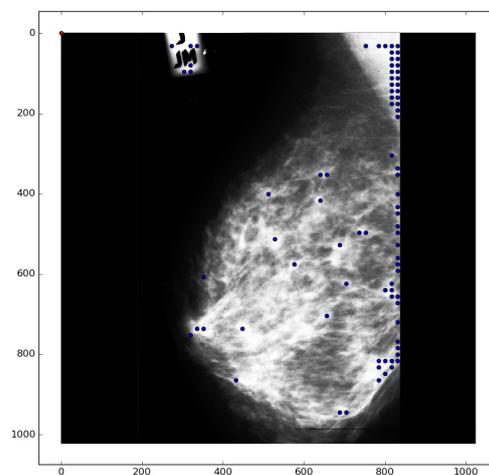
Non si è individuata alcuna correlazione tra le performances del localizzatore e caratteristiche della lesione quali sua natura (maligna o benigna) o forma della stessa, tuttavia si nota come le immagini dotate di maggior contrasto interno tendano a fornire



(a)



(b)



(c)

Figura 5.3: Immagini relative ai casi #135 (a), #173 (b) e #129 (c), tutti sani, con overlay di localizzazione (blu = predizione).

maggiori misclassificazioni rispetto a quelle con variazioni più morbide, portando a maggiori proposte in regioni con tessuto denso-ghiandolare le quali, esibendo più strutture maggiormente contrastate rispetto ai tessuti grassi o grasso-ghiandolari, possono trarre maggiormente in inganno l'algoritmo di classificazione.

### 5.3 Confronto con la letteratura

In questa sezione andremo a confrontare i risultati ottenuti sia in termini di classificazione che di localizzazione con quanto presente nella letteratura sull'argomento.

Per fare ciò è necessaria fare un'importante distinzione necessaria a collocare nella corretta fascia di riferimento il metodo qui proposto. I classificatori moderni si dividono principalmente in tre famiglie:

**Classificatori a singolo stadio:** sono algoritmi che eseguono la classificazione in una singola fase, utilizzando una sola rete per proporre una categoria di appartenenza, sia partendo da immagini complete (es. classificatori basati su CNN) che da features estratte dall'immagine, in modo automatizzato o manuale (es. SVM). CaffeNet\_MAMMO è una rete la cui implementazione come classificatore ricade in questa categoria;

**Classificatori ibridi:** sono algoritmi che eseguono la classificazione in due fasi distinte. Nella prima fase una CNN estrae le features (in termini di risposta dei filtri dopo un certo numero di cambi di rappresentazione, pari al numero di blocchi logici della rete) dall'immagine completa. Quindi tali valori sono accoppiati con caratteristiche più intuitive come le HCFeats[11] e il nuovo vettore di features ibrido viene inserito in una SVM o in un altro classificatore non deep per determinare la classe di appartenenza dell'immagine;

**Classificatori a più fasi:** sono gli algoritmi di classificazione più raffinati, composti da un numero di reti superiore a due e posti in cascata. Ad ogni fase (che può essere sia normale che deep) vengono accoppiati i risultati della fase precedente e delle features dell'immagine. L'unico classificatore che assegna una proposta di etichetta è quello presente nello strato più esterno. A causa della loro grande complessità essi richiedono tempi notevoli di sviluppo ma sono in grado di fornire i migliori risultati. Un esempio di questa categoria è il Galileo<sup>2</sup>, sviluppato dal 2001 al 2006 dall'Università di Bologna e l'R2 ImageChecker di Hologic<sup>3</sup>.

Risulta intuitivo comprendere come all'aumentare della complessità del classificatore aumentino le sue performances, sia in termini di classificazione che di localizzazione.

Per quanto concerne la classe più evoluta, Galileo riesce a raggiungere su microcalcificazioni una sensibilità del 95% (85% su masse), mentre il prodotto Hologic raggiunge il 97% di sensibilità su calcificazioni e l'88% di sensibilità su masse. Per quanto riguarda la localizzazione, Galileo offre tassi di falsi positivi per immagine pari a 0.2 per calcificazioni e 0.3 per masse, mentre il prodotto Hologic si ferma a 0.64 per calcificazioni e 1 per masse.

I classificatori ibridi e a singolo stadio sono analizzati nel dettaglio in [11], e esibiscono performance sensibilmente inferiori, con sensibilità massime prossime all'80% [19] nel caso degli ibridi e del 75% nel caso di singolo stadio [18], con ratei di falsi positivi per immagine prossimi a 1 per gli ibridi e ancora superiori per i metodi a singolo stadio.

In questo scenario, il metodo qui proposto si inserisce nella fascia di eccellenza dei classificatori a singolo stadio basati su CNN, offrendo una buona sensibilità e specificità, prossime al 70%, se comparato a quanto presente in letteratura, ma offre performances notevolmente inferiori se confrontato con i metodi più evoluti, soprattutto in fase di loca-

---

<sup>2</sup><http://www.bioretics.com/?project=galileo-cad> (visto il 14/11/16).

<sup>3</sup><http://www.hologic.com/products/imaging/mammography/image-analytics>  
[https://www.hologic.com/sites/default/files/product-files/ImageChecker%208.7%20Data%20Sheet%20\(DS-00054\)%20English%2012-08.pdf](https://www.hologic.com/sites/default/files/product-files/ImageChecker%208.7%20Data%20Sheet%20(DS-00054)%20English%2012-08.pdf)  
(visti il 15/11/2016)

lizzazione, dove il tasso di falsi positivi per l'immagine si assesta intorno a 5 fp/immagine (escludendo i falsi positivi di bordo, altrimenti supera i 10 fp/immagine).

Nonostante ciò, il risultato ottenuto è da intendersi positivo in quanto il metodo qui proposto si basa su un'architettura deep addestrata a partire da un database insufficientemente popoloso per una rete del genere. L'aver realizzato un sistema che riesca a posizionarsi in una posizione di rilievo tra gli altri appartenenti alla stessa famiglia risulta quindi uno dei punti di forza del presente lavoro che potrà in tempi successivi essere incluso in un classificatore ibrido o a più fasi con la speranza di aumentarne ulteriormente le incoraggianti performances registrate.

# Conclusioni

È stato illustrato il processo che ha portato allo sviluppo di un sistema di CAD mammografico basato su un classificatore, costruito su una rete neurale a convoluzione. Il punto di forza del metodo qui proposto è la possibilità di addestrare una rete ad apprendimento approfondito a partire da un dataset estremamente ridotto (322 immagini) rispetto agli standard del deep learning, richiedenti alcune migliaia di immagini di addestramento.

L'impiego di tecniche di post-processing ampiamente collaudate in ambito CAD, quali l'RMSHE, e un'oculata data augmentation hanno infatti permesso di realizzare un database sintetico di dimensioni molto maggiori dell'originale, con un aumento delle immagini di più di 20 volte, al prezzo di lavorare su porzioni di immagine anziché su immagini complete.

L'impiego di una rete di affermata qualità nel riconoscimento di immagini naturali è stato determinante, sopperendo così con la potenza della rete a ciò che il dataset non era in grado di fornire a livello di varietà.

Il classificatore proposto mostra un'efficienza in termini di AUC sorprendentemente elevata se si pensa alla scarsità di dati disponibili, con un valore pari a  $0.68 \pm 0.08$ , statisticamente molto differente dall'ipotesi nulla di classificatore puramente casuale. Tale valore risulta paragonabile con la maggior parte delle soluzioni adottate in sistemi per CAD mammografico [11].

Inoltre si è realizzato un prototipo di un localizzatore basato su CNN. Questo software, che rappresenta il cuore di un sistema CAD, riesce nella maggior parte dei casi

a fornire un numero sufficientemente contenuto di ROI contenenti potenziali patologie al medico specialista, riducendo i tempi di analisi delle mammografie e le possibilità di errate diagnosi in modo considerevole. Il localizzatore tuttavia presenta un bias cognitivo che marca come ROI tutti i bordi dell'immagine e delle placche. Questo difetto è dovuto alla presenza nel set di training di numerose immagini contenenti lesioni in tali aree, portando la rete all'errata convinzione che il trovarsi sul bordo sia una delle caratteristiche ricorrenti dei tumori al seno, e pertanto eliminabile utilizzando un database di addestramento dove tali casistiche siano una minoranza trascurabile. Sebbene facilmente eliminabile, tale difetto è stato mantenuto nella versione del localizzatore qui presentata, per evitare di ignorare potenziali patologie realmente collocate in tali aree, adottando in generale un principio di prudenza.

Infine, si è tentato l'addestramento della rete partendo da un modello pronto all'uso per il riconoscimento di immagini generiche (CaffeNet), caratterizzato da filtri estremamente raffinati in grado di discriminare fino a 1000 diverse categorie di soggetti dell'immagine. L'idea di tale riaddestramento era di sfruttare suddetti filtri, ritenendo che potessero rivelarsi adatti all'analisi di immagini mammografiche seppur eventualmente con risultati inferiori, mettendo però in luce la possibilità di generalizzazione offerte dal modello. Tuttavia tale tentativo non ha fornito i risultati attesi, portando ad una classificazione delle immagini totalmente arbitraria e priva di significato. L'interpretazione che si propone è che sebbene la rete alla base di CaffeNet sia la stessa che permette al classificatore proposto di funzionare, la sua versione addestrata, sebbene ideata proprio per riconoscimento di immagini generiche, sia paradossalmente estremamente specifica per quello scopo. Per permettere alla rete di generalizzare è necessario procedere ad un riaddestramento che coinvolga un numero di strati così elevato da in pratica costituire un addestramento ex novo, risultando quindi scarsamente interessante alla luce di quanto già affrontato nel resto della presente tesi.

Alla luce dei risultati ottenuti, il metodo illustrato risulta estremamente interessante

per la realizzazione di sistemi CAD addestrati su dataset ridotti quali quelli costruibili a partire dalle cartelle cliniche offerte dalla maggior parte dei laboratori di analisi specializzati in mammografie, che offra però prestazioni comparabili con sistemi realizzati addestrando la rete sottostante su dati aggregati di dimensioni ben superiori a quello qui impiegato. Il presente lavoro, sebbene distante da più raffinate soluzioni commerciali in termini di performances, può trovare future espansioni sia nel miglioramento delle sue prestazioni, addestrando la rete su database maggiormente consistenti, che nella realizzazione della parte convoluzionale di un sistema di classificazione ibrido basato sia su CNN che su ANN, naturale evoluzione dei classificatori CNN puri [18] grazie alle performances estremamente promettenti di CaffeNet\_MAMMO confrontate con altri classificatori convoluzionali a singolo stato usati in ambito mammografico [11].

# Bibliografia

- [1] Peter C Gøtzsche, Karsten Juhl Jørgensen, *Screening for breast cancer with mammography*, Cochrane Breast Cancer Group 6, 2013;
- [2] AA. VV., *World Cancer Report*, International Agency for Research on Cancer, 2008
- [3] F. De Waard, E. A. Baanders-Van Halewijn e J. Huizinga, *The Bimodal Age Distribution Of Patients With Mammary Carcinoma*, disponibile presso PubMed (visto 09/11/2016), 1963;
- [4] Robin Smithuis and Ruud Pijnappel, *Breast - Calcifications Differential Diagnosis*, Radiology department, Rijnland Hospital, Leiderdorp and Martini Ziekenhuis, 2008;
- [5] Yoonseop Kang, Saehoon Kim, Seungjin Choi, *Deep Learning to Hash with Multiple Representations*, IEEE XPLORE, 2012;
- [6] Ronan Collobert, Jason Weston, *A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning*, ICML, 2008;
- [7] Martìn Abadi, David G. Andersen, *LEARNING TO PROTECT COMMUNICATIONS WITH ADVERSARIAL NEURAL CRYPTOGRAPHY*, arXiv, 2016 (visto il 09/11/2016);
- [8] AA.VV., *Convolutional Neural Networks (LeNet) – DeepLearning 0.1 documentation*, DeepLearning 0.1, LISA Lab., 2010;

- [9] Ezio Bottarelli, Stefano Parodi, *Un approccio per la valutazione della validità dei test diagnostici: le curve R.O.C. (Receiver Operating Characteristic)*, Ann. Fac. Medic. Vet. di Parma Vol. XXIII, pag. 49-68, 2003;
- [10] Hubel, D. and Wiesel, T., *Receptive fields and functional architecture of monkey striate cortex* Journal of Physiology (London), 195, pag. 215–243, 1968;
- [11] Arevalo, Gonzàles, Ramos-Pollàn, Oliveira, Guevara Lopez, *Representation learning for mammography mass lesion classification with convolutional neural networks*, Computer Methods And Programs in Biomedicine, 2015;
- [12] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell, *Decaf: A deep convolutional activation feature for generic visual recognition*, arXiv, 2013 (visto il 09/11/2016);
- [13] J Suckling et al, *The Mammographic Image Analysis Society Digital Mammogram Database*, Excerpta Medica, International Congress Series 1069 pag. 375-378, 1994;
- [14] K. Akila, L.S. Jayashree, A. Vasuki, *Mammographic Image Enhancement Using Indirect Contrast Enhancement Techniques – A Comparative Study*, Procedia Computer Science Volume 47, Pages 255-261, 2015;
- [15] Soong-Der Chen, Abd. Rahman Ramli, *Contrast Enhancement using Recursive Mean-Separate Histogram Equalization for Scalable Brightness Preservation*, IEEE XPLORE, 2003;
- [16] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, Trevor Darrell, *Caffe: Convolutional Architecture for Fast Feature Embedding* , Open Source Software Competition, 2014;

- [17] H. B. Mann, D. R. Whitney, *On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other*, Ann. Math. Statist. Volume 18, Number 1, pag. 50-60, 1947;
- [18] Ketan Sharma, Bobbin Preet, *Classification of mammogram images by using CNN classifier*, IEE XPLORE, 2016 (visto il 11/11/2016);
- [19] Wener Borges Sampaioa, Edgar Moraes Diniza, Aristófanés Corrêa Silvaa, Anselmo Cardoso de Paivaa, Marcelo Gattassb, *Detection of masses in mammogram images using CNN, geostatistic functions and SVM*, Computers in Biology and Medicine, 41, 2011;

# Elenco delle figure

1.1	Casi (istogramma) e tassi (linea continua) di incidenza del tumore al seno in funzione dell'età di insorgenza. . . . .	7
1.2	Tipica mammografia di screening non post-processata, sono visibili entrambe le proiezioni, MLO e CC. . . . .	8
1.3	Evoluzione della potenza dei computer della TOP500, in ordinata potenza in GFLOPs, in ascissa anno di osservazione. É evidenziato il valore di 1 TeraFLOPs ritenuto potenza di riferimento per addestramento reti deep. . . . .	11
1.4	Esempio di connettività sparsa dei neuroni tra diversi strati convoluzionali: i neuroni di uno strato $m$ sono connessi solo ad alcuni dei neuroni dello strato precedente $m-1$ secondo principi di località. . . . .	15
1.5	Esempio di Condivisione dei pesi (Feature Map) tra i neuroni tra diversi strati convoluzionali, a link di colore uguale corrisponde peso uguale. . . . .	15
1.6	Filtri appresi dal primo strato convoluzionale di ImageNet (a) (si noti la distinzione tra filtri di intensità e filtri relativi a specifiche bande cromatiche), immagine di prova sottoposta alla rete (b), risposta (dopo Max Pooling per questioni di leggibilità) del primo strato convoluzionale (c) e del quinto (d). . . . .	18
1.7	Valori di AUC con relativo errore per i diversi metodi analizzati in [11] con e senza la combinazione con le HCFeats (tratto da [11]). . . . .	21

2.1	Confronto di possibilità offerte al programmatore dalle più diffuse librerie per implementazione di reti deep. . . . .	23
2.2	Mammografia relativa al caso #5 prima dell'elaborazione (a), e dopo l'applicazione dell'RMSHE (c). In (b) e (d) i rispettivi istogrammi. . . .	28
3.1	Schema della CNN impiegata per l'addestramento del classificatore. Sono riquadrati gli insiemi di layers costituenti un unico blocco logico (strato di elaborazione + strati ausiliari). Su campo bianco sono indicati gli strati di I/O e di confronto tra risultati prodotti e classificazioni attese. . . . .	34
4.1	Andamento di sensibilità e specificità in funzione della soglia di detezione.	44
4.2	ROC del nostro sistema (linea e punti blu) e <i>chance line</i> (linea rossa). . .	45
4.3	ROC mediata del nostro sistema con errori delle misure (linea blu) e <i>chance line</i> (linea rossa). . . . .	46
5.1	Immagine del caso #80 con overlay di localizzazione con (a) e senza (b) soppressione delle proposte di bordo. . . . .	55
5.2	Immagini relative ai casi #264 (a), #10 (b) e #256 (c), tutti patologici, con overlay di localizzazione (blu = predizione, rosso = ground truth). . .	57
5.3	Immagini relative ai casi #135 (a), #173 (b) e #129 (c), tutti sani, con overlay di localizzazione (blu = predizione). . . . .	59