

POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Biomedica



TESI DI LAUREA MAGISTRALE

Modelli generativi basati su apprendimento profondo per la simulazione di immagini istopatologiche di tessuto prostatico

Relatori

Prof. Massimo Salvi

Ing. Francesco Marzola

Candidata

Ilaria Martini

Marzo 2023

Sommario

Il carcinoma prostatico è la neoplasia più comune tra gli uomini e rappresenta il 19,8% di tutti i tumori maschili. Nonostante sia molto diffuso, si tratta solitamente di una patologia a lenta evoluzione, asintomatica ed eterogenea. Di conseguenza, è fondamentale attuare delle politiche di screening volte a diagnosticare precocemente il tumore e stimare il corretto grado tumorale (Gleason Score). La diagnosi viene effettuata manualmente dal patologo attraverso l'esame istologico in H&E del campione asportato tramite agobiopsia. Poiché i principali criteri diagnostici si basano su aspetti puramente morfologici, l'assegnazione del Gleason Score è dispendiosa in termini di tempo e soffre di un'elevata variabilità inter- e intra-operatore. Per far fronte a tutte queste problematiche, il lavoro di tesi in questione mira allo sviluppo di un sistema automatico in grado di sintetizzare immagini istopatologiche di tessuto prostatico a partire da un ground truth. Lo scopo è quello di ideare un algoritmo di supporto che affianchi il patologo nel riconoscimento e nella classificazione delle lesioni tumorali, permettendo di velocizzare il processo diagnostico. Sono stati sviluppati due sistemi automatici basati su GAN condizionali: il primo consente la generazione di maschere di segmentazione, mentre il secondo la generazione semantica di immagini. Il dataset di partenza, fornito dal Vancouver Prostate Center, consiste in 333 WSI annotate manualmente da sei patologi esperti. Nella fase di pre-processing, le multiple annotazioni sono state unite in un unico ground truth label attraverso l'algoritmo STAPLE, in modo tale da ottenere una maschera altamente accurata e robusta. Le maschere binarie dei nuclei e dei bianchi ricavate mediante appositi algoritmi, sono servite per arricchire l'informazione semantica dei ground truth label e per creare le mappe d'istanza, in cui sono raffigurati i bordi dei nuclei e dei bianchi. Infine, la fase di estrazione delle patch ha consentito di selezionare immagini che contenessero almeno il 90% di una stessa classe. Per la sintesi delle maschere è stata implementata la StyleGAN2, addestrata su un dataset bilanciato. Affinché fosse possibile distinguere l'architettura del tessuto per differenti gradi patologici e confrontare le maschere reali e simulate, sono stati calcolati opportuni parametri morfologici e descrittori di tessitura. Si riportano come esempio alcuni parametri reali e generati. Numero nuclei: $289 \pm 100 / 258 \pm 82$ (B); $339 \pm 84 / 295 \pm 76$ (G3); $366 \pm 112 / 352 \pm 103$ (G4); $382 \pm 100 / 314 \pm 103$ (G5). Area nuclei: $69 \pm 11 / 67 \pm 12$ (B); $86 \pm 21 / 100 \pm 35$ (G3);

88±27/92±34 (G4); 84±28/105±22 (G5). Perimetro bianchi: 578±499/582±406 (B); 170±91/166±79 (G3); 152±99/167±96 (G4); 94±52/166±169 (G5). D'altra parte, per il task di generazione semantica, sono state implementate due varianti del modello Pix2Pix: SPADE e CLADE. Le immagini sintetiche sono state confrontate con le immagini originali attraverso cinque metriche di valutazione: PSNR, RMSE, SSIM, FSIM e VSI. Si raggiungono valori medi prossimi allo 0,25 e 0,90 per SSIM e VSI rispettivamente e superiori ai 16 dB per PSNR. In conclusione, i modelli permettono al medico di disporre di campioni di tessuto di classe nota e costituiscono dei metodi di data augmentation per altri algoritmi di segmentazione.

Indice

Elenco delle figure	6
Elenco delle tabelle	10
1 Introduzione	12
1.1 Anatomia e fisiologia della prostata	12
1.2 Iperplasia prostatica benigna	14
1.3 Carcinoma prostatico	15
1.3.1 Fattori di rischio	16
1.3.2 Diagnosi	17
1.3.3 Gleason Score	22
1.4 Finalità dello studio	25
2 Stato dell'arte	27
2.1 Generative Adversarial Network	27
2.2 Semantic image synthesis	31
2.2.1 Pix2Pix	31
2.2.2 Pix2PixHD	32
2.2.3 SPADE	34
2.2.4 CLADE	35
2.2.5 INADE	36
2.3 GAN class-conditional	38
2.3.1 StyleGAN	38
2.3.2 StyleGAN2	40
2.3.3 StyleGAN3	43
3 Metodi	47
3.1 Dataset di partenza	47
3.2 Pipeline	47
3.3 Pre-processing	48
3.3.1 Creazione dei ground truth label	50

3.3.2	Integrazione delle mappe semantiche e creazione delle mappe d'istanza	51
3.3.3	Estrazione delle patch	53
3.3.4	Costruzione dei dataset	55
3.4	Architetture GAN	57
3.5	Tuning dei parametri e training	58
3.5.1	StyleGAN2	58
3.5.2	SPADE e CLADE	60
3.6	Metodi di valutazione	60
3.6.1	Parametri morfologici	61
3.6.2	Feature di tessitura	61
3.6.3	Metriche di valutazione	64
4	Risultati	70
4.1	Task 1: Simulazione maschere	70
4.2	Task 2: Simulazione semantica	77
5	Conclusioni	83
	Bibliografia	86

Elenco delle figure

1.1	Zone della ghiandola prostatica. La prostata può essere suddivisa in quattro zone: anteriore (verde), centrale (viola), periferica (blu) e di transizione (rossa). Immagine tratta da [1].	13
1.2	Tumore della ghiandola anteriore con estensione extracapsulare. L'immagine di risonanza magnetica assiale pesata T2 (a) mostra un tumore ipointenso a contatto con la capsula prostatica (freccia). L'immagine assiale di livello leggermente superiore (b) rivela l'estensione extracapsulare del tumore (freccia). Immagine tratta da [2].	19
1.3	Ecografia transrettale in cui è presente un carcinoma ipoecogeno ben definito che invade la vescicola seminale (freccia). Immagine tratta da [3].	20
1.4	Evoluzione del sistema Gleason. A) Classificazione originale dei pattern del Gleason Score. B) Modifica proposta dall'International Society of Urological Pathology (ISUP) nel 2005. C) Modifica proposta dall'ISUP nel 2014. Immagine tratta da [4].	22
1.5	Patch di tessuto prostatico aventi diversi pattern istologici.	24
2.1	Architettura generica di una GAN. Immagine tratta da [5].	27
2.2	Grafico a torta che riporta la distribuzione delle applicazioni GAN nell'ambito dell'imaging medico entro il 2019. Immagine tratta da [6].	30
2.3	Confronto tra l'architettura encoder-decoder e U-Net del generatore. Immagine tratta da [7].	32
2.4	Architettura del generatore <i>coarse-to-fine</i> . Immagine tratta da [8]. . .	33
2.5	Tecnica di normalizzazione SPADE. Utilizza una rete di modulazione composta da due layer convoluzionali per modellare i parametri di denormalizzazione γ e β , in funzione della maschera semantica in ingresso. Immagine tratta da [9].	35
2.6	Struttura di un blocco residuo con SPADE (sinistra). Architettura del generatore SPADE, in cui ogni layer di normalizzazione utilizza la maschera di segmentazione per modulare le attivazioni del layer precedente (destra). Immagine tratta da [9].	36

ELENCO DELLE FIGURE

2.7	Normalizzazione adattata alle classi (a). CLADE utilizza un'efficiente operazione di campionamento guidato per campionare i parametri di modulazione adattati alla classe per ogni regione semantica. Variante con codifica posizionale intra-classe (b). La mappa della codifica posizionale viene calcolata a partire dalla mappa di segmentazione; d_0 e d_1 rappresentano la codifica posizionale lungo le dimensioni x e y . Immagine tratta da [10].	37
2.8	Normalizzazione INADE. Combina la distribuzione a livello semantico e il campionamento del rumore a livello di istanza. IGS indica il campionamento guidato dell'istanza. Immagine tratta da [11].	38
2.9	Architettura del generatore StyleGAN. Immagine tratta da [12].	39
2.10	Schematizzazione del processo di modulazione e demodulazione dei pesi associati a ciascun layer di convoluzione utilizzando il vettore di stile ottenuto dalla trasformazione affine.	41
2.11	Meccanismo ADA. In blu sono riportati i blocchi che caratterizzano operazioni di data augmentation, in arancione quelli che indicano la funzione di perdita, in verde quelli che vengono allenati all'iterazione di riferimento, mentre in grigio quelli temporaneamente bloccati (allenati all'iterazione successiva). G e D rappresentano il generatore e il discriminatore rispettivamente. Immagine tratta da [13].	42
2.12	Rappresentazione semplificata della modalità che permette di passare da un dominio discreto (Z) a uno continuo (z) e viceversa. Il passaggio da una mappa discreta a una continua è garantito attraverso la convoluzione con un kernel passa-basso, mentre il passaggio inverso è effettuato mediante la moltiplicazione puntuale di una griglia di Dirac. Immagine tratta da [14].	44
2.13	Architettura del generatore della StyleGAN3. Immagine tratta da [14].	45
3.1	Esempio di WSI (sinistra) e le relative annotazioni manuali fornite dai patologi (destra).	48
3.2	Pipeline del lavoro.	49
3.3	Esempio di WSI (sinistra), relativa annotazione manuale del patologo 2 (centro) e completamento della maschera (destra).	50
3.4	Esempio di WSI con le relative annotazioni manuali dei diversi patologi (a) e ground truth label ottenuto in seguito dell'applicazione dell'algoritmo STAPLE (b, in alto) e successiva integrazione del majority voting (b, in basso).	51
3.5	Esempio di immagine istologica originale, ottimizzata e normalizzata e relative maschere dei bianchi e dei nuclei.	52

3.6	Esempio di mappa semantica modificata e mappa d'istanza creata. Nella mappa semantica (a) si riporta la zona Gleason 4 (viola), i nuclei (arancione), i bianchi (giallo) e l'area non marcata (bianco); mentre nella mappa d'istanza (b) sono mostrati i bordi dei nuclei e dei bianchi. Le maschere d'esempio sono relative alla WSI della Figura 3.5a.	53
3.7	Patch di dimensione 512x512 estratte dalle stessa WSI, ma con differente scala d'ingrandimento.	54
3.8	Esempio di data augmentation su una patch di classe Gleason 5. . . .	56
3.9	Esempio di immagine del Test set (sinistra) appartenente alla classe Gleason 4, mappa d'istanza (centro) e mappa di segmentazione semantica (destra).	57
3.10	Andamento della FID durante il training della StyleGAN2.	60
4.1	Esempi di maschere di segmentazione simulate mediante StyleGAN2. Da sinistra verso destra si riportano campioni di classe Benigna, Gleason 3, Gleason 4 e Gleason 5.	71
4.2	Parametri morfologici dei nuclei. I grafici mostrano i valori medi e i relativi errori standard di area, perimetro, eccentricità e numerosità dei nuclei nelle maschere reali e generate.	72
4.3	Parametri morfologici dei bianchi. I grafici mostrano i valori medi e i relativi errori standard di area totale, perimetro, eccentricità e numerosità dei bianchi nelle maschere reali e generate.	73
4.4	Descrittori di tessitura del I ordine. I grafici mostrano i valori medi e i relativi errori standard di media, varianza, skewness, curtosi ed entropia delle maschere reali e generate per ciascuna classe.	75
4.5	Descrittori di tessitura del II ordine. I grafici mostrano i valori medi e i relativi errori standard di contrasto, dissimilarità, omogeneità, energia e correlazione delle maschere reali e generate per ciascuna classe.	76
4.6	Esempi di immagini istologiche simulate da SPADE e CLADE. Si riporta un esempio per ciascun grado della patologia, dall'alto verso il basso: Benigna, Gleason 3, Gleason 4 e Gleason 5. Da sinistra verso destra sono mostrate l'immagine reale, la mappa di segmentazione e le immagini generate dagli algoritmi SPADE e CLADE rispettivamente.	78
4.7	Metriche di valutazione della qualità delle immagini generate da CLADE e SPADE. Si riporta la media e la rispettiva deviazione standard sia sul Training set sia sul Test set.	80
4.8	Metriche di valutazione calcolate su ciascuna classe: Benigna, Gleason 3, Gleason 4 e Gleason 5. Si riporta la media e la rispettiva deviazione standard sul Test set ottenuta dai modelli CLADE e SPADE.	81

Elenco delle tabelle

1.1	Assegnazione del Grade Group in funzione del Gleason Score e stratificazione del rischio.	25
3.1	Numerosità delle patch estratte e selezionate per ciascuna classe d'interesse.	54
3.2	Struttura del dataset utilizzato per la generazione semantica di immagini istologiche.	58
3.3	Descrittori del primo ordine.	62
3.4	Descrittori del secondo ordine.	63

Capitolo 1

Introduzione

1.1 Anatomia e fisiologia della prostata

La prostata è una ghiandola esocrina dell'apparato genitale maschile situata nel compartimento sub-peritoneale e racchiusa dal diaframma pelvico e dalla cavità peritoneale. In particolare, è posizionata posteriormente alla sinfisi pubica, anteriormente al retto e inferiormente alla vescica urinaria. In un uomo adulto, in condizioni normali, la ghiandola ha dimensioni e forma assimilabili a una noce, con un peso compreso tra i 20 e i 30 g.

La prostata è costituita da tessuto ghiandolare e stroma, strettamente fuso all'interno di una pseudocapsula. Lo strato interno della capsula prostatica è composta da muscolo liscio rivestito esternamente da collagene.

I dotti eiaculatori attraversano obliquamente la ghiandola per entrare nell'uretra prostatica in corrispondenza del verumontano. Quest'ultimo, essendo posizionato a metà strada tra il collo della vescica e l'apice prostatico, suddivide l'uretra in prossimale e distale. La porzione prossimale è circondata da una cuffia di muscolo liscio nota come sfintere preprostatico. Tale cuffia si estende inferiormente fino a livello del verumontano e impedisce il passaggio dei dotti nella parte prossimale dell'uretra prostatica.

La ghiandola prostatica è composta da una base, un apice e da quattro superfici: anteriore, inferiore e infero-laterali. La base è adiacente al collo della vescica ed è attraversata dall'uretra nella porzione mediana vicino alla superficie anteriore. L'apice è invece localizzato sulla superficie superiore del diaframma urogenitale, a contatto con la superficie mediale dei muscoli elevatori anali. La superficie posteriore, triangolare e piatta, giace sulla parete anteriore del retto. Mentre, la superficie infero-laterale si unisce a quella anteriore e insieme poggiano sul fascio dei muscoli elevatori dell'ano al di sopra del diaframma urogenitale.

Inoltre, la prostata può essere suddivisa in quattro zone (Figura 1.1), che si diffe-

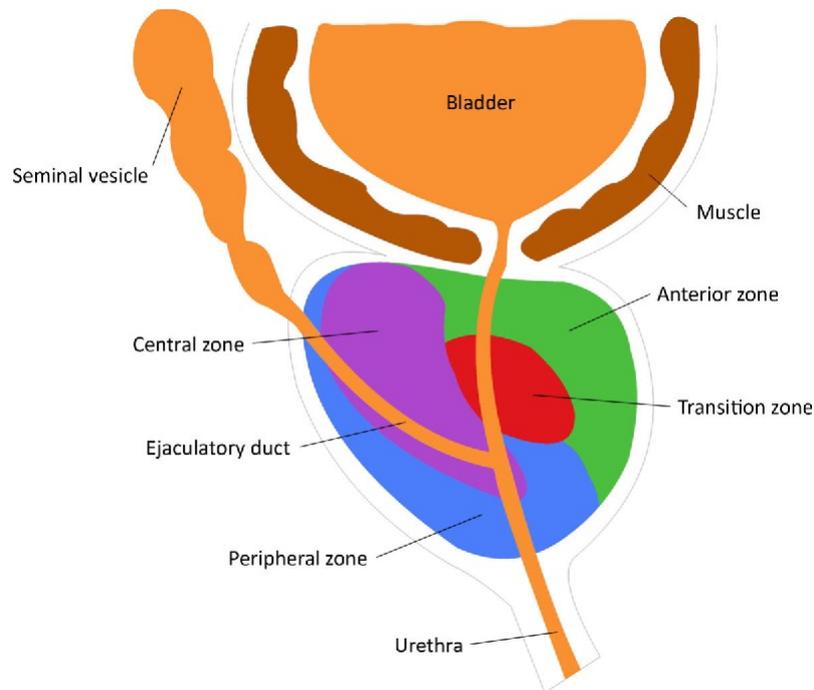


Figura 1.1: Zone della ghiandola prostatica. La prostata può essere suddivisa in quattro zone: anteriore (verde), centrale (viola), periferica (blu) e di transizione (rossa). Immagine tratta da [1].

renziano principalmente per le proprie origini embriologiche oltre alle caratteristiche anatomiche e istologiche. La zona periferica è maggiormente localizzata nella porzione dorsale della ghiandola, in prossimità della parete rettale, e circonda sia la zona centrale che l'uretra prostatica distale. Essa costituisce la principale componente ghiandolare della prostata, che comprende circa il 70% dell'intero volume prostatico. La zona periferica include dotti stretti che terminano in piccoli acini tondeggianti e permettono esclusivamente il drenaggio in direzione dell'uretra prostatica distale. Le cellule epiteliali acinari hanno bordi distinti, citoplasma chiaro e piccoli nuclei scuri posti basalmente, mentre lo stroma contiene muscolatura liscia intrecciata in modo casuale. La zona centrale circonda i dotti eiaculatori e l'uretra prossimale in posizione posterosuperiore, racchiudendo la zona di transizione. Si tratta di un'area conica avente l'apice posto alla confluenza dei dotti eiaculatori e dell'uretra prostatica in corrispondenza del verumontano. Costituisce la maggior parte del tessuto ghiandolare della base prostatica, andando a caratterizzare il 25% circa della ghiandola. Nella zona centrale, i dotti sono ampi e terminano in grandi acini irregolari con proiezioni stromali prominenti. Le cellule epiteliali acinari hanno bordi indistinti, citoplasma granuloso e grandi nuclei chiari localizzati a diversi livelli cellulari. Lo stroma invece contiene una muscolatura liscia compatta, disposta longitudinal-

mente. La zona di transizione invece circonda anteriormente e lateralmente l'uretra prossimale, costituendo solamente il 5% della ghiandola. Istologicamente, tale area è simile a quella periferica. Infine, lo stroma fibromuscolare anteriore (AFMS) è situato anteriormente all'uretra e si estende lungo la zona di transizione, formando la convessità della superficie esterna anteriore. Insieme all'uretra costituisce il principale elemento non ghiandolare. La metà apicale di quest'area è ricca di muscolo striato, che si fonde con la ghiandola e il muscolo del diaframma pelvico. Verso la base invece predominano le cellule muscolari lisce, che si fondono con le fibre del collo vescicale. La porzione distale dell'AFMS svolge un ruolo fondamentale nelle funzioni sfinteriche volontarie, mentre la porzione prossimale è centrale nelle funzioni sfinteriche involontarie.

Le differenti caratteristiche istologiche e anatomiche delle zone prostatiche comportano una diversa suscettibilità a eventi patologici. Oltre il 70% di tutti i tumori della prostata origina nella porzione periferica della ghiandola. Nella porzione antero-mediale transizionale dell'organo invece insorge il 20% delle neoplasie mentre la zona centrale è tipica dell'iperplasia prostatica benigna (IPB) ed è sede primaria solo del 5% delle neoplasie [15]. Sebbene la zona di transizione abbia origini embriologiche simili a quella periferica, la percentuale di cancro alla prostata che si sviluppa in tale area è nettamente inferiore. Tale discrepanza può essere spiegata dal fatto che lo stroma della zona di transizione ha una componente fibromuscolare maggiore. A tal proposito si ipotizza che l'iperplasia prostatica benigna sia una patologia dello stroma fibromuscolare, dal momento che insorge prevalentemente nella zona di transizione [16].

La funzione principale della prostata è quella di secernere il liquido prostatico, un liquido lattiginoso e alcalino che costituisce circa il 30% del volume dello sperma. Le secrezioni della prostata contengono diversi fattori, tra cui lo zinco, il citrato e il biomarcatore del carcinoma prostatico: l'antigene prostatico specifico (PSA). Il PSA è una glicoproteina secreta dalle cellule epiteliali della prostata in grado di liquefare il liquido seminale. Una porzione di PSA è presente nel circolo sanguigno ed elevati livelli di antigene nel sangue possono indicare la presenza di una condizione anomala della ghiandola prostatica, come ad esempio la presenza del carcinoma. Dal momento che il livello di PSA circolante è proporzionale alle dimensioni della prostata, la sua concentrazione nel circolo sanguigno può essere notevole anche in presenza di IPB [1].

1.2 Iperplasia prostatica benigna

L'iperplasia prostatica benigna è una patologia molto comune nel genere maschile, che colpisce il 5-10% degli uomini dopo i 40 anni di età e oltre l'80% dopo i 70 e 80 anni; tuttavia produce sintomi solo nella metà dei soggetti [17].

Si tratta di una patologia benigna reversibile confinata nella zona di transizione, caratterizzata dall'assenza di formazioni tumorali e infiltrazione dei tessuti. Si verifica un accumulo di diidrotestosterone (DHT) che inibisce la morte cellulare e promuove la proliferazione delle cellule prostatiche epiteliali e stromali, con conseguente ingrossamento della ghiandola. Può provocare sintomi del tratto urinario inferiore dovuti all'ostruzione dell'uscita della vescica, che vanno a interferire con le capacità urinarie del soggetto. Tra i sintomi ostruttivi si ha lo svuotamento incompleto della vescica, flusso urinario debole e intermittente e difficoltà a iniziare la minzione. Tuttavia, possono manifestarsi anche sintomi di tipo irritativo: frequenza nell'urinare, nicturia e urgenza minzionale. Quando si sviluppa l'ostruzione al deflusso urinario, i muscoli detrusori della vescica cercano di compensarlo con l'aumento di pressione, in modo tale da cercare di espellere l'urina, provocando però l'instabilità del muscolo e il peggioramento dei sintomi.

L'IPB sembra essere correlata all'età, agli androgeni, agli estrogeni e alla disfunzione detrusoriale del collo della vescica [18]. Generalmente, il trattamento iniziale per i pazienti affetti da IPB con sintomi del tratto urinario inferiore moderati o gravi prevede un approccio terapeutico. La terapia consiste sostanzialmente in due classi: gli antagonisti alfa-1 selettivi, che rilassano la muscolatura liscia della prostata e del collo vescicale per migliorare il flusso urinario, e gli inibitori della 5-alfa reduttasi di tipo II, che metabolizzano il testosterone in DHT riducendo le dimensioni della ghiandola mediante l'apoptosi prostatica. Questi farmaci sono indicati quando il peso della prostata è superiore a 40 g [19]. L'intervento chirurgico è invece riservato ai pazienti che non tollerano la terapia medica e con complicazioni correlate all'IPB, come la ritenzione urinaria refrattaria, i calcoli vescicali e l'insufficienza renale da ostruzione delle vie urinarie inferiori. La resezione transuretrale della prostata e la vaporizzazione laser prostatica sono i trattamenti chirurgici principali.

1.3 Carcinoma prostatico

Il carcinoma prostatico è la neoplasia maligna più diffusa negli uomini in Italia (40.500 nuovi casi stimati nel 2022) e rappresenta il 19,8% di tutti i tumori maschili. Il tasso di sopravvivenza a 5 anni dalla diagnosi è circa del 91% [20]. L'adenocarcinoma rappresenta circa l'80% di tutti i casi di cancro alla prostata e si sviluppa dalle cellule epiteliali secretorie della zona periferica della ghiandola.

I carcinomi prostatici sono per lo più multifocali ed eterogenei per grado e aspetto istologico. Di solito la neoplasia evolve lentamente, senza diffondersi al di fuori della ghiandola. Esistono tuttavia anche forme più aggressive, nelle quali le cellule neoplastiche invadono rapidamente tessuti e organi circostanti, come le vescicole seminali, la vescica urinaria e il retto. Nella fase iniziale il carcinoma è in genere asintomatico, ma con il progredire della malattia loco-regionale possono comparire alcuni sintomi,

tra cui la diminuzione della potenza del getto urinario e il dolore perineale. Nelle fasi più avanzate, essendo lo scheletro la prima sede di metastatizzazione, è caratteristico lo sviluppo di dolore osseo, localizzato principalmente a livello del rachide. Infatti, le cellule del cancro diffondono attraverso il flusso sanguigno e formano metastasi soprattutto a livello dei linfonodi e delle ossa. Di conseguenza, è fondamentale attuare delle politiche di screening volte a diagnosticare precocemente il carcinoma. Per valutare l'aggressività del tumore e la sua estensione locale e a distanza si utilizza il sistema di stadiazione TNM, che classifica i carcinomi in base all'analisi del tumore, dei linfonodi e delle metastasi, come illustrato in seguito:

- **T:** si riferisce alle dimensioni del tumore primario e al suo grado di invasione nei tessuti adiacenti. Si assegna un valore compreso tra 0 (assenza di tumore) e 4 (elevato grado di estensione e invasività).
- **N:** si utilizza per descrivere il coinvolgimento dei linfonodi nella regione tumorale. Si assegna un valore compreso tra 0 (assenza di diffusione linfonodale regionale) e 3 (elevata progressione distale).
- **M:** si utilizza per identificare la presenza di metastasi a distanza dal tumore primario (0 in caso di assenza, 1 in caso di evidenza).

Effettuare la stadiazione del carcinoma è sostanziale per stimare la prognosi, stabilire il protocollo di trattamento e organizzare il follow-up dei pazienti.

1.3.1 Fattori di rischio

L'eterogeneità clinica del carcinoma prostatico si riflette anche sull'eziologia della patologia, che risulta essere multifattoriale, risultato di una complessa interazione tra fattori biologici e stile di vita tra cui:

- **Età:** l'incidenza del tumore aumenta con l'età. Raramente si osservano casi in uomini con età inferiore a 40 anni, ma la possibilità di sviluppare il cancro alla prostata aumenta drammaticamente oltre i 55 anni [21].
- **Etnia:** l'incidenza è maggiore in uomini di origine Latina e Africana.
- **Storia familiare:** gli uomini con un padre o un fratello a cui è stato diagnosticato il cancro alla prostata hanno un rischio da due a tre volte maggiore di sviluppare la neoplasia. Il rischio è nove volte superiore per gli uomini con entrambi i familiari coinvolti [21].
- **Mutazioni genetiche:** per esempio mutazioni ereditarie dei geni BRCA1 o BRCA2 e la sindrome di Lynch.

- **Stile di vita:** obesità, eccessivo apporto di calcio e alimenti caseari, fumo.
- **Fattori metabolici:** l'aumento dei livelli di insulina, del fattore di crescita IGF-1 e della leptina [22].
- **Fattori ormonali:** elevate concentrazioni di testosterone [22].

Sono stati analizzati anche diversi fattori comportamentali protettivi che potrebbero essere coinvolti con la riduzione del rischio di insorgenza, tra i quali l'attività fisica e il consumo frequente di pomodori, alimenti a base di soia e pesce [22].

1.3.2 Diagnosi

La diagnosi precoce è una strategia chiave per ridurre significativamente il rischio di mortalità e migliorare la prognosi. Tuttavia, dal momento che si tratta di una neoplasia a lenta evoluzione e generalmente asintomatica negli stadi iniziali, la diagnosi precoce è molto complessa e si affida a esami di screening. Il 30% circa dei casi vengono diagnosticati quando la malattia si è già diffusa oltre la sede ghiandolare [23]. Dal 1993 l'obiettivo primario dell'ERSPC¹ è quello di valutare l'effetto dello screening regolare. Si è osservata una riduzione del rischio relativo del 27% a 13 anni di follow-up. D'altra parte, la riduzione del tasso di mortalità associata alle tecniche di screening è stata inferiore rispetto al rischio di sovradiagnosi e al conseguente sovratrattamento [24].

Esplorazione rettale

Si tratta di una procedura che consente all'urologo di valutare direttamente le dimensioni e la consistenza della prostata e rilevare la presenza di eventuali noduli o indurimenti sospetti della ghiandola prostatica. Dal momento che l'80% dei tumori della prostata è localizzato nella zona periferica della ghiandola, principalmente in posizione posteriore, l'esplorazione rettale risulta essere un approccio particolarmente agevole per la rilevazione quando il volume prostatico è superiore a 0,2 ml [25]. Gli svantaggi associati a tale metodo diagnostico sono correlati con le capacità del medico che effettua la diagnosi, dunque è solitamente affiancata ad altri test di screening.

Dosaggio del PSA

Il test consiste in un semplice prelievo di sangue che permette di misurare il livello dell'antigene prostatico specifico, poiché, come affermato in precedenza, tende ad

¹European Randomized Study of Screening for Prostate Cancer

aumentare enormemente nei pazienti che presentano una patologia prostatica. Sebbene l'80% dei tumori prostatici si manifesta inizialmente mediante livelli di PSA elevati nel circolo sanguigno, solo nel 25-30% dei casi si tratta di carcinoma [26]. I pazienti con un livello di PSA superiore a 4,0 ng/ml, una velocità del PSA maggiore di 0,75 ng/ml/anno o presenza di un nodulo prostatico riscontrato mediante palpazione devono sottoporsi a una biopsia prostatica per valutare la presenza del tumore [3, 24].

Nonostante il test del PSA sia un argomento controverso e oggetto di dibattito tra gli esperti, è ampiamente utilizzato per rilevare il carcinoma prostatico ed è uno strumento utile per valutare l'efficacia del trattamento o per monitorare l'evoluzione della malattia. Secondo l'ERSPC tale tecnica diagnostica ha aumentato il tasso di incidenza di 1,6 volte e ridotto la mortalità al 21% a 13 anni dal follow-up. La riduzione della mortalità sembra essere correlata alla durata dello screening [27].

Gli svantaggi associati al test del PSA includono una scarsa specificità, poiché, essendo l'antigene organo specifico, raggiunge elevati livelli anche in presenza di patologie prostatiche benigne e in determinate condizioni fisiologiche, come l'eiaculazione recente e l'attività fisica intensa. Questo si traduce in un trattamento aggressivo dovuto a una sovradiagnosi di tumore indolente e non letale. Inoltre, non esiste un valore soglia che possa garantire con certezza la presenza o l'assenza di tumore. A tal proposito, uno studio di prevenzione del tumore alla prostata ha dimostrato che il carcinoma può essere rilevato anche quando il livello di PSA è inferiore a 4 ng/ml [28]. Pertanto, l'interpretazione del risultato dell'esame deve essere sempre messa in relazione con l'età e con la storia clinica dell'individuo.

Risonanza magnetica

La risonanza magnetica multiparametrica (RMmp) prevede l'acquisizione di sequenze anatomiche T2-pesate, T1-pesate in studio di perfusione (DCE-MRI²) e pesate in diffusione (DWI³), fornendo una combinazione di informazioni anatomiche, biologiche e funzionali. Nello specifico, lo studio della diffusione molecolare dell'acqua consente di stabilire il grado di proliferazione della lesione, mentre lo studio della perfusione mediante contrasto permette di valutare la vascolarizzazione della ghiandola. Tale tecnica ha una risoluzione dei tessuti molli di gran lunga migliore rispetto all'ecografia ed è in grado di identificare aree ghiandolari sospette con un elevato grado di accuratezza e affidabilità. Si raggiunge una sensibilità complessiva dell'89% e una specificità del 73% per l'identificazione del cancro alla prostata [29].

Un risultato positivo della RMmp è talvolta sufficiente per diagnosticare il tumore della prostata senza necessariamente richiedere una conferma istologica. Oltre a

²Dynamic Contrast Enhanced Magnetic Resonance Imaging

³Diffusion Weighted Imaging

identificare e classificare noduli prostatici sospetti per aiutare la stadiazione e la localizzazione, permette di controllare l'estensione extracapsulare, valutare le vescicole seminali per un possibile coinvolgimento del tumore e determinare l'ingrossamento dei linfonodi regionali che potrebbero essere indice di malattia metastatica precoce.

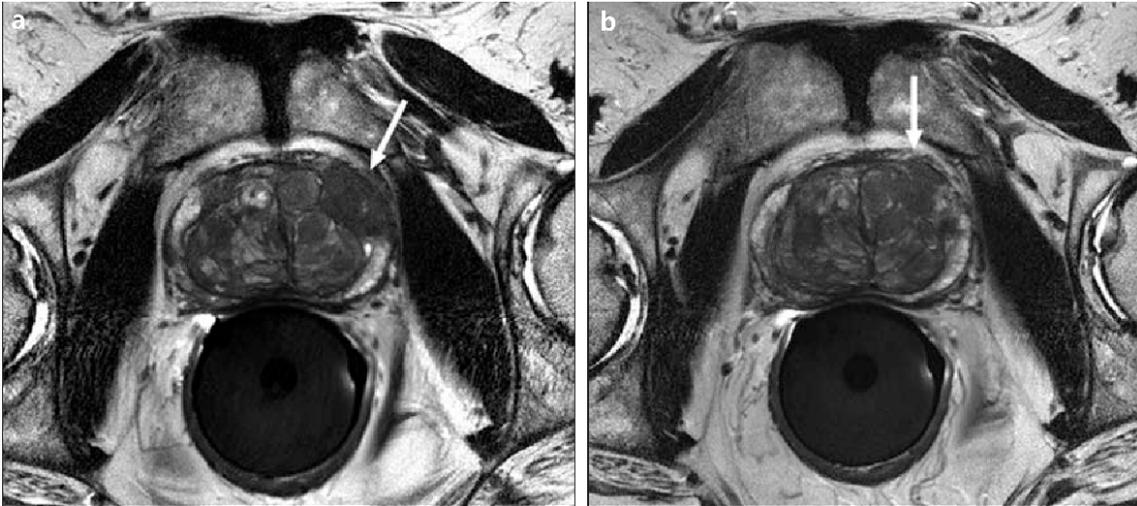


Figura 1.2: Tumore della ghiandola anteriore con estensione extracapsulare. L'immagine di risonanza magnetica assiale pesata T2 (a) mostra un tumore ipointenso a contatto con la capsula prostatica (freccia). L'immagine assiale di livello leggermente superiore (b) rivela l'estensione extracapsulare del tumore (freccia). Immagine tratta da [2].

Ecografia transrettale

L'ecografia transrettale è una tecnica diagnostica per immagini in grado di quantificare le dimensioni della ghiandola prostatica, che viene impiegata dall'urologo come guida real-time nell'esecuzione della biopsia. L'ecografia transrettale sfrutta la diversa ecogenicità del carcinoma, poiché un'area ipoecogena è indice di una regione potenzialmente sospetta. Tuttavia, non si può trascurare la possibile iperecogenicità di alcune forme di cancro alla prostata e che il 40% circa dei carcinomi mostrano un aspetto isoecogeno, anche se localizzati nella zona periferica, risultando occulti [15]. Si è cercato di porre rimedio a tale limite con lo sviluppo di metodiche quali la *Shear-wave elastography* (SWE), capace di individuare processi patologici tramite la loro alterata elasticità tissutale, e sonde ad alta frequenza, in grado di rilevare aree con anatomia duttale e densità cellulare alterati. Inoltre, l'ecografia tranrettale non è sempre in grado di differenziare i noduli proliferativi della zona di transizione dal cancro.

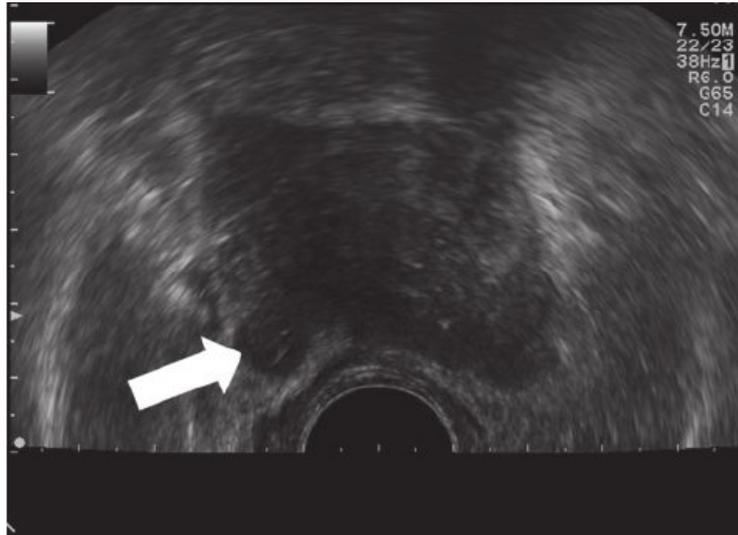


Figura 1.3: Ecografia transrettale in cui è presente un carcinoma ipoecogeno ben definito che invade la vescicola seminale (freccia). Immagine tratta da [3].

Agobiopsia ed esame istologico

L'agobiopsia è il gold standard per stabilire una diagnosi di adenocarcinoma prostatico. Pur essendo una procedura ampiamente utilizzata, la sensibilità e la specificità della biopsia sistematica sono bassi. I tassi di individuazione del carcinoma sono compresi tra il 20 e il 60%, mentre il tasso di falsi negativi è stimato intorno al 25-40% [30]. Per incrementare la qualità e l'accuratezza diagnostica della biopsia, ci si può affidare a strumenti diagnostici secondari. Esistono tre approcci metodologici differenti: (1) biopsia *in-bore* guidata da risonanza magnetica; (2) ispezione visiva della risonanza magnetica con conseguente biopsia cognitiva prostatica, ovvero guidata in real-time da ecografia transrettale; (3) biopsia *fusion* mirata di lesioni sospette, che è condotta mediante fusione dell'immagine di risonanza magnetica con l'ecografia transrettale utilizzando un software computerizzato. Quest'ultima procedura combina i vantaggi relativi alla risonanza magnetica nella localizzazione del tumore, e dell'ecografia nella valutazione in tempo reale, portabilità ed ergonomia. Nonostante il ridotto tasso di complicazioni, la biopsia prostatica è una procedura invasiva, pertanto si cerca di limitare il suo utilizzo a pazienti con un'elevata probabilità di ospitare un cancro clinicamente significativo. Attualmente l'indicazione all'esecuzione di biopsia prostatica per la diagnosi mediante esame istopatologico viene formulata sulla base del sospetto clinico, laboratoristico e/o esito positivo della RMmp. La biopsia prostatica eco-guidata viene eseguita mediante approccio transrettale o ransperineale. Le percentuali diagnostiche, in assenza di imaging preliminare con risonanza magnetica, sono comparabili tra le due metodiche; tuttavia

alcune evidenze suggeriscono di utilizzare preferibilmente l'approccio transperineale per il minor rischio di complicanze infettive.

Per stabilire la diagnosi istopatologica di cancro alla prostata è necessario l'esame al microscopio ottico di sezioni di tessuto colorate con ematossilina ed eosina. Tuttavia, prima di procedere con la colorazione dei campioni istologici asportati, è necessario trattarli opportunamente in modo tale da preservare la loro architettura interna. La prima operazione da eseguire è la fissazione del campione per prevenirne l'alterazione chimica e fisica. Dopo aver effettuato la deidratazione e la pulizia, si tratta il tessuto con la paraffina per fornire resistenza al campione stesso. Successivamente si seziona il tessuto con l'utilizzo di un microtomo rotante; lo spessore ottimale per poter osservare sia l'architettura tissutale sia la morfologia cellulare è $5\mu\text{m}$. Si esegue poi la colorazione del campione, utilizzando coloranti affini ai diversi costituenti del tessuto, e lo si racchiude con coperture di vetro trasparenti per proteggerlo da agenti esterni ed evitare la sua contaminazione [31]. I due coloranti sfruttano il diverso valore di pH dei componenti cellulari, causando un aumento di contrasto tra le diverse strutture. L'ematossilina è un colorante basico, che tende a legarsi a molecole acide, come gli acidi nucleici presenti a livello del nucleo, conferendogli il caratteristico colore blu-violaceo. Al contrario, l'eosina è un colorante acido, che si lega a molecole basiche, quali le proteine cellulari presenti a livello citoplasmatico, conferendogli la caratteristica colorazione rosacea. I vetrini istologici possono infine essere digitalizzati con l'ausilio di appositi scanner facilitandone la conservazione a lungo termine. L'uso di scanner differenti può imporre una variazione di colore del tessuto.

I principali criteri diagnostici per stabilire la presenza di cancro alla prostata sono la presenza di pattern di crescita infiltrativi, valutata a un ingrandimento medio-basso, l'assenza di cellule basali, l'ingrossamento nucleare e nucleoli prominenti, meglio apprezzabili a elevato ingrandimento [32].

Immunoistochimica

Sebbene la microscopia ottica sia il metodo cardine per la diagnosi del carcinoma prostatico, i casi complessi possono beneficiare di studi di immunoistochimica. Uno dei criteri principali su cui ci si basa per la diagnosi del carcinoma è l'assenza di cellule basali. Dal momento che la valutazione può essere ardua in specifiche porzioni di tessuto, soprattutto quando il numero di ghiandole atipiche è limitato, può essere necessario ricorrere all'uso di marcatori immunoistochimici.

L'uso di colorazioni immunoistochimiche specifiche per le cellule basali (KER 903, p63, CK5/6) può essere utile per stabilire una diagnosi di malignità, tuttavia è importante tenere presente che uno strato distinto di cellule basali non è sempre presente in piccole ghiandole benigne. Di conseguenza, l'assenza di cellule basali in un piccolo focus di ghiandole atipiche non è un criterio assoluto di malignità [32].

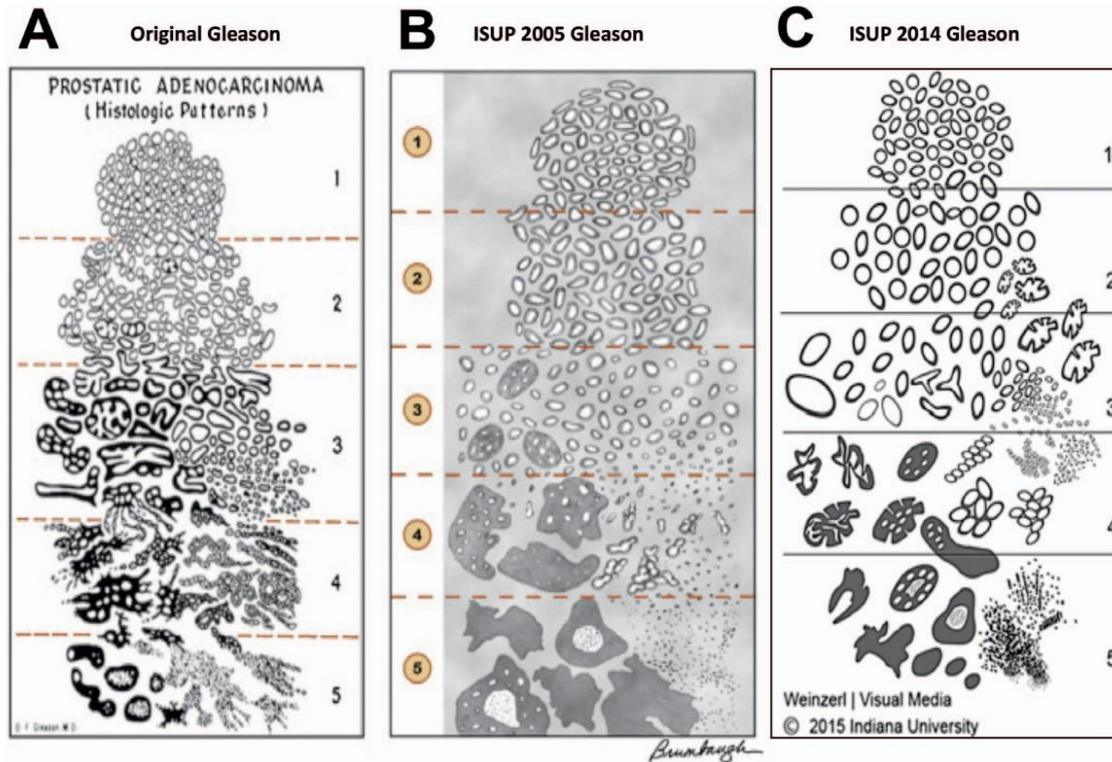


Figura 1.4: Evoluzione del sistema Gleason. A) Classificazione originale dei pattern del Gleason Score. B) Modifica proposta dall’International Society of Urological Pathology (ISUP) nel 2005. C) Modifica proposta dall’ISUP nel 2014. Immagine tratta da [4].

1.3.3 Gleason Score

L’adenocarcinoma prostatico presenta un’architettura ghiandolare anomala ed eterogenea, che viene classificata mediante il sistema Gleason. Ai diversi pattern istologici viene assegnato uno score compreso da 1 a 5 in base all’estensione della differenziazione ghiandolare e al modello di crescita del tumore nello stroma prostatico. Le caratteristiche strutturali del campione vengono valutate dall’anatomo-patologo a basso ingrandimento (da 4x a 10x) [33]. Le linee guida da seguire per l’assegnazione dello score sono state illustrate nel corso degli anni dalla ISUP⁴ (Figura 1.4). I cinque pattern vengono utilizzati per generare uno score istologico, che può variare da 2 a 10, sommando il grado primario e il grado secondario. Il cosiddetto pattern primario è quello che risulta predominante alla semplice ispezione visiva, mentre il pattern secondario corrisponde al secondo più rappresentato o al più alto dei meno

⁴International Society of Urological Pathology

rappresentati nella neoplasia. Se nel campione di tessuto è presente un solo pattern, si deve raddoppiare il grado primario per ottenere lo score finale.

I gradi terziari vengono gestiti in maniera differente a seconda del campione. Per le agobiopsie con pattern 3, 4 e 5 il punteggio viene assegnato in base al pattern predominante e il grado più alto. Al contrario, per i campioni di prostatectomia radicale, la presenza di un terzo pattern che sia inferiore al 5% del tumore e che sia di grado superiore al primario e al secondario, deve essere annotata nel referto piuttosto che essere incorporata nel punteggio. È chiaro che una componente terziaria di grado elevato che occupa il 5% del tumore nell'intera ghiandola, influenza lo stadio patologico e i tassi di progressione della neoplasia.

I carcinomi di basso grado sono raramente riscontrabili all'agobiopsia poichè sono localizzati prevalentemente anteriormente, nella zona di transizione, e tendono ad avere dimensioni ridotte. Dunque, un cancro con Gleason Score inferiore a 6 esiste, ma è raramente diagnosticato.

La classificazione Gleason è clinicamente un sistema non ottimale, dal momento che i pazienti con diagnosi di tumore prostatico con Gleason Score pari a 6 ritengono erroneamente che il loro carcinoma sia di rischio intermedio, nonostante sia il punteggio più basso utilizzato nelle agobiopsie prostatiche. Nel 2014 la ISUP ha introdotto il Gleason Grade Group per semplificare il sistema di gradazione del carcinoma prostatico e renderlo più comprensibile a medici e pazienti. Tra i vantaggi introdotti dai gruppi di grado, vi è l'allineamento della gradazione del carcinoma prostatico a quella di altri tumori, che hanno il grado più basso identificato con il gruppo 1, e la semplificazione del sistema. Inoltre, ha permesso una stratificazione del grado più accurata migliorando la distinzione tra neoplasie con Gleason Score 3+4 e 4+3, che in precedenza dividevano il medesimo score 7.

Le caratteristiche istologiche corrispondenti a ciascun gruppo sono le seguenti [34]:

- **Grade Group 1:** Solo ghiandole singole e ben distinte.
- **Grade Group 2:** Prevalenza di ghiandole regolari con una piccola componente di ghiandole irregolari, fuse e/o cribriformi.
- **Grade Group 3:** Prevalenza di ghiandole irregolari, fuse e/o cribriformi con una piccola componente di ghiandole regolari.
- **Grade Group 4:** Solo ghiandole irregolari, fuse e/o cribriformi (4+4). Prevalenza di ghiandole regolari con una componente minore tessuto privo di formazione ghiandolare (3+5). Prevalenza di tessuto privo di formazione ghiandolare con una componente minore di ghiandole regolari (5+3).
- **Grade Group 5:** Assenza di formazione ghiandolare o presenza di ghiandole necrotiche con o senza ghiandole irregolari, fuse e/o cribriformi.

Nella Figura 1.5 sono mostrate delle immagini istologiche di tessuto prostatico appartenenti al Vancouver Prostate Centre TMAs Dataset con diversa architettura e quindi diverso Gleason Score.

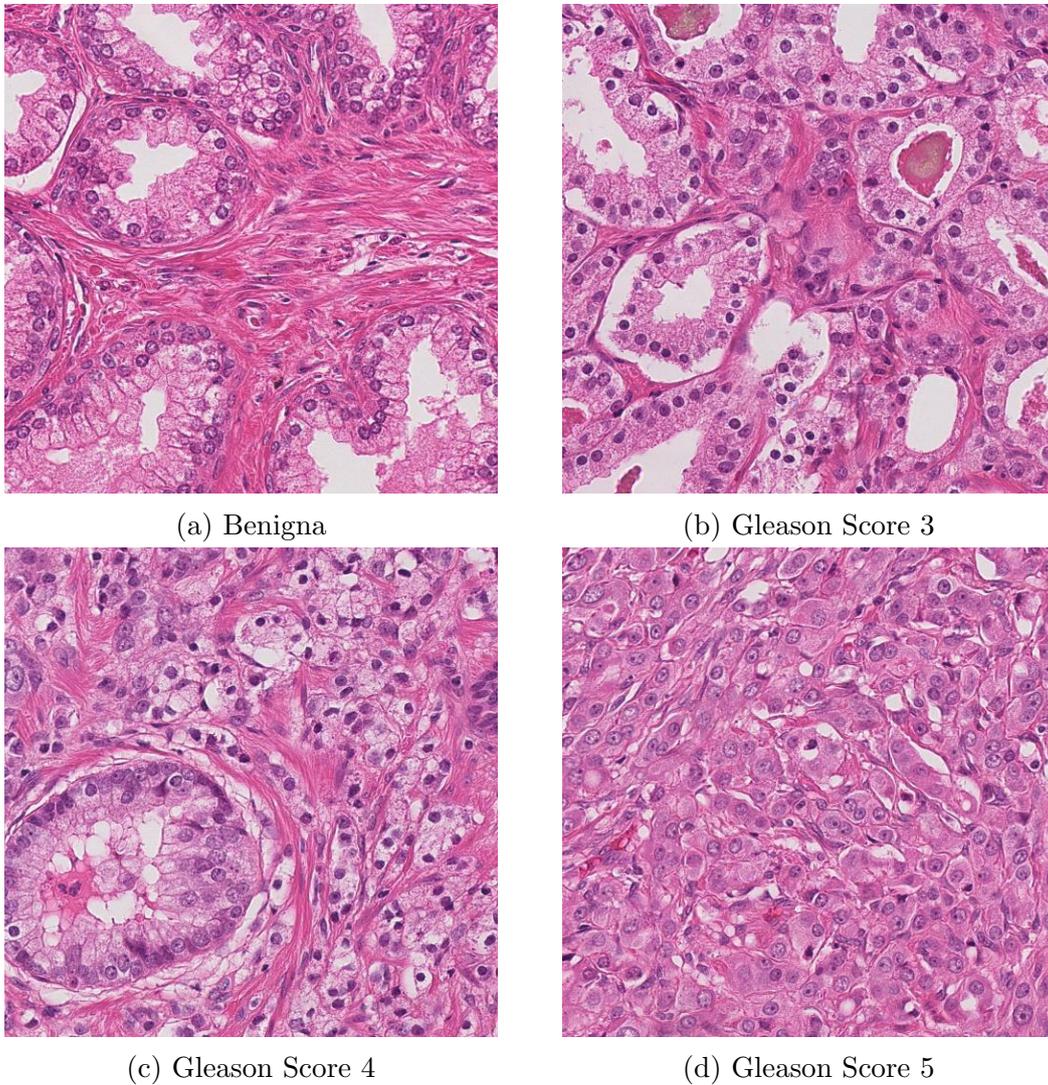


Figura 1.5: Patch di tessuto prostatico aventi diversi pattern istologici.

Il valore prognostico di questa classificazione è tale per cui nel libro blu della WHO⁵ 2016 si raccomanda di riportare sempre il Gleason Grade Group insieme al Gleason Score modificato secondo ISUP 2014 [15]. La Tabella 1.1 illustra la corrispondenza tra le due scale di classificazione.

⁵World Health Organization

Grade Group	Gleason Score	Rischio
I	≤ 6 (3+3)	Basso
II	7 (3+4)	Moderato favorevole
III	7 (4+3)	Moderato sfavorevole
IV	8 (4+4, 3+5, 5+3)	Alto
V	9 (4+5, 5+4) - 10 (5+5)	Alto

Tabella 1.1: Assegnazione del Grade Group in funzione del Gleason Score e stratificazione del rischio.

1.4 Finalità dello studio

L'assegnazione del Gleason Score non è solamente dispendiosa in termini di tempo, ma si affida all'esperienza del patologo e soffre di un'elevata variabilità inter-osservatore [35]. Per bypassare tali problematiche, il lavoro di tesi svolto mira allo sviluppo di un modello generativo per la sintesi di immagini istopatologiche prostatiche a partire da un ground truth. In questo modo, è l'immagine a essere generata a partire dall'annotazione manuale e non viceversa. I vantaggi legati all'implementazione di tale modello sono i seguenti:

- Superamento delle problematiche legate alla variabilità intra- e inter-operatore.
- Processi diagnostici più rapidi.
- Supporto decisionale all'anatomo-patologo.
- Minore mole di lavoro ai centri di anatomia patologica.

Si prospetta quindi che il sistema possa essere utilizzato come strumento di supporto da parte degli anatomo-patologi per una corretta ed efficiente diagnosi e stadiazione della neoplasia, che verranno effettuate paragonando i campioni con le immagini sintetiche di riferimento. Inoltre, è possibile sfruttare il sistema per scopi puramente didattici, tra cui la data augmentation, andando a integrare e arricchire il set di dati a disposizione con le immagini sintetiche in modo da poter sviluppare e validare altri algoritmi di Deep Learning.

Capitolo 2

Stato dell'arte

2.1 Generative Adversarial Network

Nel 2014 Goodfellow ha proposto un nuovo modello di rete neurale denominato Generative Adversarial Network (GAN) [36]. Una rete generativa avversaria è un framework di apprendimento automatico composto da due modelli convoluzionali: generativo e discriminativo. Il ruolo del generatore è quello di adattarsi il più possibile alla distribuzione di dati reali, producendo dei campioni sintetici che assomiglino ai campioni di addestramento a partire da una variabile casuale latente. Il discriminatore invece è un classificatore binario il cui scopo è distinguere i campioni generati da quelli reali (Figura 2.1).

Il processo di addestramento di una GAN consiste in un graduale aumento dell'efficienza del generatore e del discriminatore. Questo miglioramento avviene simultaneamente, partendo da caratteristiche di basso livello, fino ad acquisire la capacità di gestire dettagli sempre più fini. Le due reti vengono quindi fatte competere

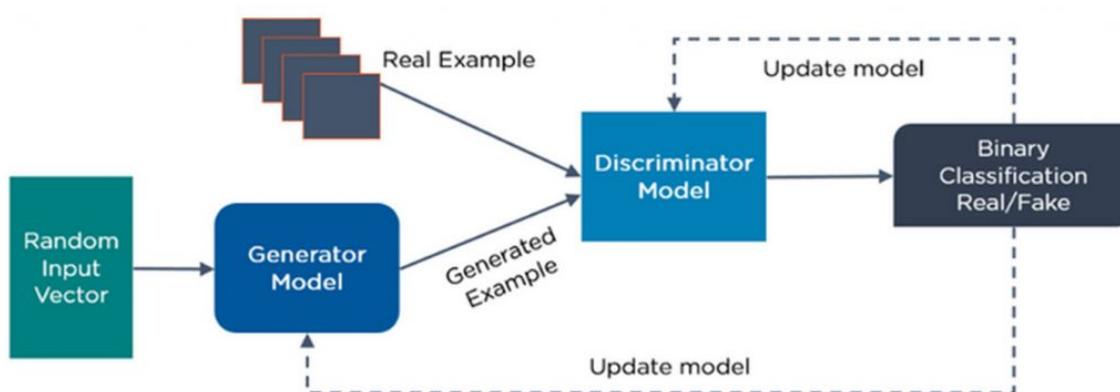


Figura 2.1: Architettura generica di una GAN. Immagine tratta da [5].

l'una contro l'altra finché non convergono e giungono al cosiddetto equilibrio di Nash, situazione in cui il generatore e il discriminatore raggiungono il massimo delle prestazioni e non sono più in grado di migliorare. Attraverso un meccanismo di *back-propagation* è possibile allenare alternativamente le reti. L'output di classificazione fornito dal modello discriminativo costituisce il segnale d'errore con cui viene addestrato il generatore, ma anche il discriminatore stesso. Il motivo per il quale l'apprendimento del generatore avviene attraverso l'interazione con il discriminatore, è legato al fatto che non ha accesso diretto alle immagini reali. La retropropagazione dell'errore avviene congelando i pesi associati a uno dei due modelli, durante l'addestramento del secondo, che avviene calcolando la loss e aggiornando i pesi della rete in questione. A questo punto si esegue lo stesso procedimento a reti invertite. L'obiettivo del discriminatore è massimizzare la probabilità che assegni le etichette corrette sia ai campioni di addestramento sia a quelli ottenuti in fase di generazione, mentre l'obiettivo del generatore è massimizzare la probabilità che il discriminatore commetta errori. In particolare, il ruolo del modello generativo è quello di trasformare una distribuzione casuale p_z (solitamente Gaussiana o uniforme) nella distribuzione di dati p_d , prendendo in input un vettore di rumore $z \sim p_z$ e mappandolo su un campione $x = G(z)$ derivante da p_g , che è la stima del generatore della vera distribuzione di dati p_d . L'apprendimento competitivo prende parte a un meccanismo denominato *min-max*, in cui il generatore tende a minimizzare la funzione obiettivo, mentre il discriminatore tende a massimizzarla. La funzione di perdita di una GAN standard è formulata nell'equazione seguente:

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_d} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))]. \quad (2.1)$$

La funzione può essere suddivisa in due parti:

- Loss del discriminatore, definita come:

$$\log D(x) + \log(1 - D(G(z))). \quad (2.2)$$

Il modello discriminativo tenta di massimizzare entrambi i termini, ovvero massimizzare la probabilità che il discriminatore classifichi correttamente sia i campioni reali sia i campioni prodotti dal generatore.

- Loss del generatore, definita come:

$$\log(1 - D(G(z))). \quad (2.3)$$

Il modello generativo tenta di minimizzare il termine sopra riportato, ovvero minimizzare la probabilità che il discriminatore etichetti come generata un'immagine realmente prodotta dal generatore.

Sebbene l'addestramento di una GAN possa essere considerato un problema di ottimizzazione basato su metodi del gradiente [37], non vi è alcuna garanzia di equilibrio tra l'allenamento delle due reti. Tra i problemi più comuni riscontrati durante l'addestramento di una GAN si ha:

- **Mancata convergenza.** Si verifica quando la rete non riesce a convergere verso un minimo globale. Ciò deriva da un addestramento instabile.
- **Vanificazione dei gradienti.** Quando il discriminatore diventa più bravo del generatore la sua perdita si azzerava, spingendo i gradienti retropropagati ad assumere valori sempre più piccoli, completamente inadatti a guidare il generatore verso la corretta distribuzione. In questo modo l'apprendimento del modello generativo si blocca verso soluzioni non ottimali. Questo problema è frequente quando si trattano immagini a elevata risoluzione, poiché i dettagli ad alta frequenza rendono difficile un adattamento ottimale del generatore ai dati reali [37]. Si può verificare anche nel caso contrario, quando il discriminatore è più debole e dunque la perdita non rappresenta nulla per il generatore.
- **Collasso della modalità.** Si verifica quando la distribuzione $p_g(x)$ appresa dal generatore si adatta solamente a una porzione di quella reale $p_d(x)$. Questo fenomeno limita la capacità di apprendimento del generatore, che produrrà dei campioni molto simili tra loro, mentre il discriminatore non sarà in grado di considerare la differenza tra i campioni generati.

Nel presente lavoro focalizzeremo la nostra attenzione su GAN condizionali, poiché l'intento è quello di controllare e indirizzare il processo di generazione dei dati, andando a condizionare il modello con informazioni aggiuntive. Oltre ad affinare la generazione dei dettagli presenti nell'immagine target, le GAN condizionali migliorano la stabilità dell'addestramento [6]. In questo modo, il generatore apprende distribuzioni condizionate di informazioni, in quanto è in grado di distinguerle dallo spazio latente complessivo; mentre il discriminatore non deve solamente valutare la somiglianza tra i campioni generati e quelli reali, ma considera anche la corrispondenza dell'immagine sintetica e l'etichetta fornita in input.

Le GAN condizionali sono comunemente impiegate nei cosiddetti compiti di *image-to-image translation*, in cui l'obiettivo è trasformare una rappresentazione di un'immagine in un'altra. Tuttavia, possono essere impiegate anche per la manipolazione di immagini, la generazione *text-to-image* e la generazione *class-conditional*. Esistono diverse architetture di GAN condizionali che si distinguono anche per come sono stati alleviati i problemi d'addestramento sopra descritti. Possono essere impiegate in molteplici settori, tra cui l'imaging medico, mediante applicazioni che spaziano tra la sintesi di immagini, la ricostruzione, la segmentazione, il rilevamento, la registrazione, il de-noising e la classificazione (Figura 2.2). Le diverse applicazioni

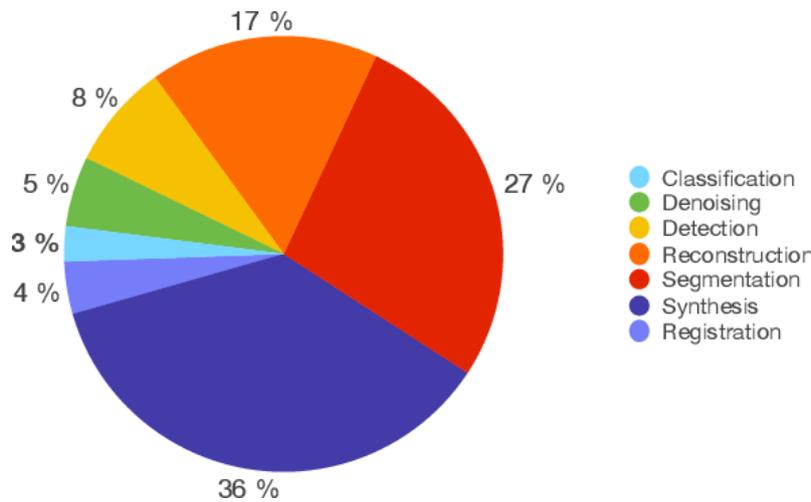


Figura 2.2: Grafico a torta che riporta la distribuzione delle applicazioni GAN nell'ambito dell'imaging medico entro il 2019. Immagine tratta da [6].

possono essere adoperate su differenti modalità di imaging: risonanza magnetica, tomografia computerizzata, radiografia, dermoscopia, ecografia e microscopia. Il successo dei metodi di Deep Learning supervisionati dipende fortemente dalla disponibilità di dati di addestramento etichettati. Nel campo dell'imaging medico, questo costituisce un problema fondamentale, poiché spesso vi è una grave mancanza di dati annotati. Tale scarsità deriva dal fatto che il processo di acquisizione è costoso e dispendioso in termini di tempo. Anche la mancanza di esperti che annotino le immagini costituisce un'ulteriore sfida per l'implementazione di questi algoritmi. Inoltre, la maggior parte dei set di dati disponibili soffre di un grave sbilanciamento tra le classi, a causa della rara natura di alcune patologie. Le GAN possono potenzialmente alleviare queste limitazioni, riducendo i problemi di privacy e generando delle immagini realistiche in grado di ridurre la disomogeneità dei campioni. Le GAN forniscono anche una possibile soluzione per ampliare il set di dati a disposizione, processo conosciuto come GAN augmentation, che può essere considerato un'alternativa o un complemento alla standard data augmentation. Il vantaggio di questo approccio rispetto a quello classico è basato sul fatto che una GAN è in grado di apprendere le diverse fonti di variazione presenti in un set di immagini. Fornisce quindi uno strumento efficiente per l'interpolazione della distribuzione dei dati di addestramento. Tuttavia, non è in grado di estrapolare oltre i suoi estremi senza l'aiuto della data augmentation standard. Perciò è fondamentale l'effetto sinergico tra i due metodi [38]. Nell'imaging istologico gli approcci basati su GAN possono essere molto utili per tenere conto di variazioni complesse, come quelle indotte dall'eterogeneità del tumore. Dunque, dato un sufficiente numero di dati di addestramento che illustrano questa eterogeneità, una GAN ben allenata

restituisce dei nuovi campioni “intermedi”. Nel presente lavoro si vuole tentare di sfruttare le immagini sintetiche anche per supportare l’anatomopatologo nel processo diagnostico. Le GAN sono quindi considerate degli strumenti potenti nell’estrarre caratteristiche semanticamente significative dalle immagini [6].

2.2 Semantic image synthesis

Una possibile applicazione di *image-to-image translation* è la sintesi semantica di immagini, in cui una GAN condizionale può essere addestrata a tradurre una maschera di segmentazione in un’immagine realistica che corrisponda al layout semantico fornito. Nel corso degli anni sono state sviluppate diverse varianti di GAN condizionali utilizzabili per questa tipologia di task. Oltre a descrivere i modelli impiegati nel lavoro in questione, è necessario illustrare le architetture GAN precedenti e comprendere il loro funzionamento.

2.2.1 Pix2Pix

La GAN Pix2Pix è stata proposta dai ricercatori della UC Berkeley nel 2017 [7]. Il modello è composto da un generatore con architettura basata su U-Net, simile all’architettura encoder-decoder, con l’inserimento di connessioni di salto tra ciascun layer i e $n-i$, dove n è il numero totale di layer (Figura 2.3). La mappa semantica fornita in input viene fatta passare attraverso una serie di layer che progressivamente effettuano un sottocampionamento dell’immagine, fino al raggiungimento di un livello di strozzatura. Da questo punto in poi il processo si inverte e la rappresentazione viene sottoposta a un sovracampionamento finché non si ottiene l’immagine finale con le dimensioni desiderate. L’aggiunta delle connessioni di salto si è dimostrata molto utile per la coerenza globale delle immagini sintetizzate [6]. Quando il codificatore esegue il sottocampionamento dell’immagine, l’output contiene le informazioni relative alle caratteristiche e alla classificazione di ciascuna classe, perdendo informazioni sulle caratteristiche di basso livello, come la disposizione spaziale intraclassa. Le connessioni di salto prevengono proprio la perdita delle caratteristiche di basso livello dell’immagine.

Il discriminatore è rappresentato da un classificatore PatchGAN convoluzionale, che è in grado di preservare i dettagli ad alta frequenza focalizzando la sua attenzione a livello di patch dell’immagine. Il discriminatore riceve in input la mappa semantica concatenata all’immagine corrispondente e classifica ogni patch dell’immagine come reale/generata. Successivamente, restituisce una matrice delle probabilità con le predizioni di ciascuna patch, che viene mediata per ottenere l’output finale. Un discriminatore PatchGAN consente di produrre immagini di alta qualità, sebbene sia caratterizzato da un numero inferiore di parametri. Inoltre, è veloce e può esse-

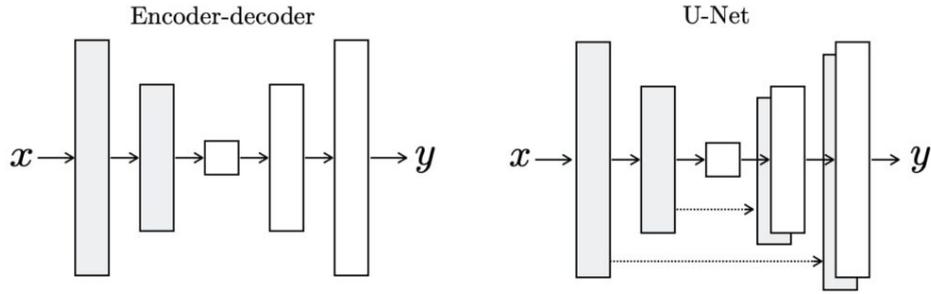


Figura 2.3: Confronto tra l'architettura encoder-decoder e U-Net del generatore. Immagine tratta da [7].

re applicato a immagini arbitrariamente grandi [7]. La funzione di perdita di una GAN condizionale viene combinata con la perdita di regolarizzazione L1, in modo che il generatore non solo inganni il discriminatore, ma produca immagini sintetiche il più simili possibile al ground truth. Dal punto di vista matematico, la funzione di perdita di una GAN condizionale integra al suo interno l'etichetta y ed è formulata come segue:

$$\mathcal{L}_{cGAN} = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))]. \quad (2.4)$$

La perdita di regolarizzazione è la seguente:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1]. \quad (2.5)$$

Dunque, la funzione obiettivo finale è:

$$\mathcal{L}_G = \operatorname{argmin}_G \operatorname{max}_D [\mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G)] \quad (2.6)$$

dove λ è un iperparametro introdotto per bilanciare le perdite.

2.2.2 Pix2PixHD

La GAN Pix2PixHD costituisce l'evoluzione della Pix2Pix, poiché l'obiettivo della sua implementazione è quello di risolvere i problemi che affliggono il modello precedente. La Pix2Pix soffre infatti di due problemi principali: la difficoltà di generare immagini ad alta risoluzione e la mancanza di dettagli e texture realistiche nelle immagini generate [8].

Il modello Pix2PixHD utilizza un generatore *coarse-to-fine*, un discriminatore multiscala e introduce un termine di perdita aggiuntivo. Per aumentare la risoluzione dell'immagine prodotta, il generatore è suddiviso in due reti: G1 e G2. Nello specifico, G1 costituisce la rete globale del generatore, mentre G2 corrisponde alla rete di

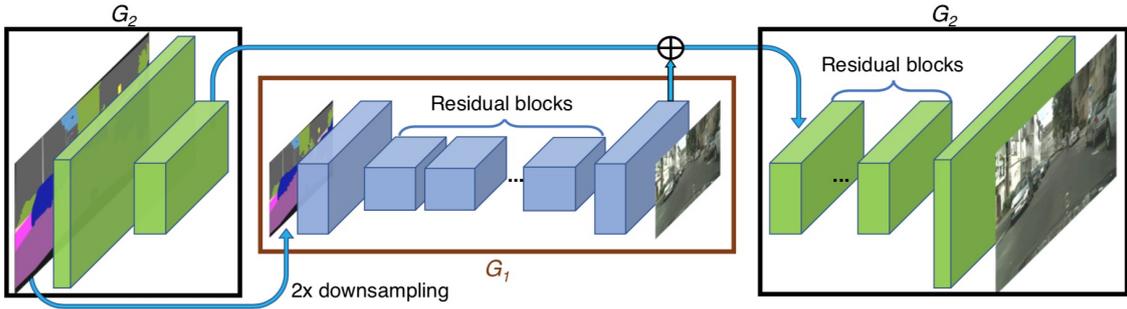


Figura 2.4: Architettura del generatore *coarse-to-fine*. Immagine tratta da [8].

miglioramento locale in grado di generare un'immagine con una risoluzione 4 volte superiore alla dimensione dell'immagine originale. Si addestra prima la rete residua G_1 su immagini a bassa risoluzione. Poi, a G_1 si aggiunge la rete residua G_2 e si addestrano le due reti congiuntamente su immagini ad alta risoluzione. In particolare, l'input dei blocchi residui di G_2 è la somma elemento per elemento della mappa di caratteristiche di G_2 e dell'ultima mappa di caratteristiche di G_1 (Figura 2.4). La sintesi di immagini ad alta risoluzione richiede un discriminatore ad ampio campo recettivo. Servirebbe una rete più profonda o kernel di convoluzione più grandi, che però comporterebbero un aumento delle capacità della rete e potrebbero causare overfitting. Dunque, si utilizza un discriminatore multiscala composto da tre discriminatori differenti aventi la stessa struttura di rete, ma che operano su scale di immagini diverse. Sia l'immagine reale sia l'immagine sintetizzata ad alta risoluzione vengono sottocampionate di un fattore 2 e 4 per creare una piramide di immagini a tre scale. In questo modo i discriminatori vengono addestrati a distinguere le immagini reali e sintetizzate su scale diverse. Il discriminatore che opera su scala maggiore ha un campo recettivo più grande e quindi è in grado di guidare il generatore a sintetizzare immagini globalmente coerenti. D'altra parte, il discriminatore alla scala minore incoraggia il generatore a produrre dettagli più fini.

La funzione di perdita della Pix2PixHD incorpora un termine aggiuntivo basato sul discriminatore, conosciuto come *feature matching loss*, in grado di stabilizzare l'addestramento. Invece di massimizzare direttamente l'output del discriminatore, si richiede al generatore di generare dati che corrispondano alle statistiche dei dati reali. Fondamentalmente, per ogni layer del discriminatore, si calcola l'errore assoluto tra il layer valutato sui dati reali e su dati sintetici. In questo modo, si vuole migliorare la corrispondenza tra le rappresentazioni intermedie dell'immagine reale e dell'immagine generata. La perdita è calcolata come segue:

$$\mathcal{L}_{FM}(G, D_k) = \mathbb{E}_{x,y} \sum_{i=1}^T \frac{1}{N_i} [\|D_k^{(i)}(x, y) - D_k^{(i)}(y, G(y))\|_1], \quad (2.7)$$

dove $D_k^{(i)}$ denota l' i -esimo layer del discriminatore D_k , T corrisponde al numero totale di layer e N_i è il numero di elementi in ogni layer.

Sia la Pix2Pix sia la Pix2PixHD forniscono direttamente in ingresso al generatore il layout semantico, che viene poi elaborato attraverso una serie di blocchi costituiti dai layer di convoluzione, normalizzazione e non linearità. Questa soluzione è spesso non ottimale, poiché i layer di normalizzazione tendono a eliminare importanti informazioni semantiche dalla mappa di segmentazione. A tal proposito, sono stati proposti alcuni metodi di normalizzazione condizionale che superano questa problematica.

2.2.3 SPADE

Nel 2019 è stato proposto un metodo di normalizzazione condizionale chiamato SPADE¹ [9], simile alla normalizzazione batch, in grado di modulare l'attivazione normalizzata in modo spazialmente adattativo e condizionato dalla maschera di segmentazione. Dunque, tutti i layer di normalizzazione vengono rimpiazzati con i blocchi SPADE, affinché l'informazione semantica possa essere propagata attraverso la rete, migliorando le prestazioni in termini di fedeltà visiva e allineamento spaziale. In SPADE la maschera di segmentazione viene prima proiettata in uno spazio di incorporazione e poi convoluta per ottenere i parametri di modulazione γ (scala) e β (bias). A differenza di altri metodi di normalizzazione condizionale, γ e β non sono vettori ma tensori con dimensioni spaziali. I parametri appresi risultano infatti spazialmente adattati e sono specifici per ciascuna classe semantica. Segue poi la fase di denormalizzazione, in cui γ e β vengono rispettivamente moltiplicati e sommati all'attivazione normalizzata elemento per elemento (Figura 2.5). Poiché i parametri di modulazione si adattano alla maschera di segmentazione in ingresso e codificano informazioni sufficienti sulla disposizione semantica, SPADE è un metodo appropriato alla sintesi semantica di immagini.

L'eliminazione del codificatore dal generatore, comunemente impiegato nelle architetture precedenti, è giustificata dal fatto che non è necessario fornire in ingresso al modello la maschera di segmentazione. Essa viene infatti fornita in input a ciascun layer intermedio del generatore. Il generatore SPADE è pertanto un decodificatore completamente convoluzionale, composto da una serie di blocchi ResNet e layer di sovracampionamento (Figura 2.6). Ogni blocco residuo opera a una scala differente, dunque si sottocampiona la maschera semantica per adattarla alla risoluzione spaziale. Il generatore riceve in input un vettore casuale con distribuzione normale e restituisce in output l'immagine sintetica. Il risultato è un modello generativo più semplice e leggero.

Per eseguire la sintesi multimodale, è possibile inserire un codificatore aggiuntivo in

¹SPatially-Adaptive (DE)normalization

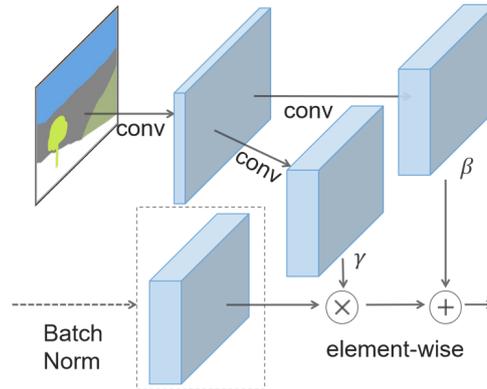


Figura 2.5: Tecnica di normalizzazione SPADE. Utilizza una rete di modulazione composta da due layer convoluzionali per modellare i parametri di denormalizzazione γ e β , in funzione della maschera semantica in ingresso. Immagine tratta da [9].

testa al generatore, che elabora l'immagine reale in un vettore casuale. In particolare, l'encoder è costituito da una serie di layer convoluzionali, di normalizzazione e di attivazione ReLu e produce il vettore media e varianza della distribuzione dell'immagine. I vettori casuali campionati da questa distribuzione sono forniti in input al generatore, agendo come guida di stile per garantire la diversità globale dei dati generati.

Nonostante la sua efficacia, la normalizzazione spazialmente adattativa introduce un significativo incremento di memoria e di calcolo, che può limitarne l'applicazione. Poiché la rete di modulazione a due layer è poco profonda, i parametri di denormalizzazione sono quasi invarianti all'interno di regioni appartenenti alla stessa classe semantica, soprattutto se le maschere di segmentazione fornite hanno una risoluzione elevata [10].

2.2.4 CLADE

A differenza di SPADE, il metodo di normalizzazione denominato CLADE² utilizza la maschera di segmentazione per modulare l'attivazione normalizzata in modo adattativo alla classe. In particolare, CLADE si adatta solamente alle diverse classi per mantenere la proprietà essenziale della consapevolezza semantica, indipendentemente dalla posizione spaziale, dalla forma semantica o dal layout complessivo della maschera. Tale tecnica è semplice da implementare e consente di raggiungere prestazioni comparabili a SPADE senza richiedere alcuna rete di modulazione aggiuntiva; pertanto i parametri coinvolti e il costo computazionale associato sono pressoché

²Class-Adaptive (DE)normalization

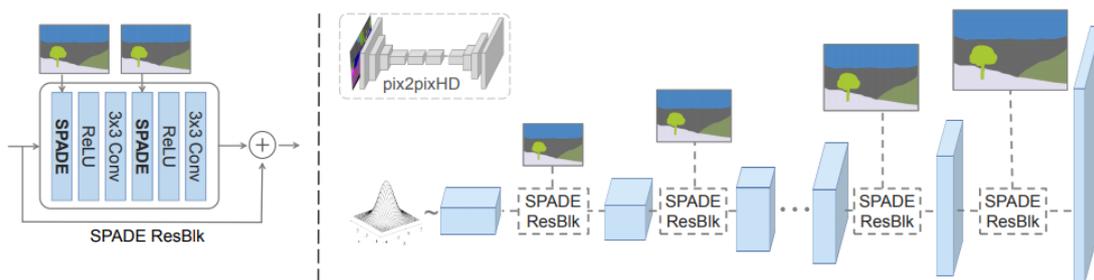


Figura 2.6: Struttura di un blocco residuo con SPADE (sinistra). Architettura del generatore SPADE, in cui ogni layer di normalizzazione utilizza la maschera di segmentazione per modulare le attivazioni del layer precedente (destra). Immagine tratta da [9].

trascurabili.

Poiché è stato osservato che l'adattamento spaziale apporta benefici alla sintesi semantica, è stata creata una variante di CLADE spazialmente adattativa denominata CLADE-ICPE. Essa utilizza una mappa di codifica posizionale che rappresenta la varianza spaziale intra-classe, definita come la distanza relativa normalizzata di ciascun pixel dal centro dell'oggetto semantico. Questa codifica posizionale viene poi integrata all'interno dei parametri di modulazione di CLADE, rendendoli spazialmente adattativi nelle regioni appartenenti alla stessa classe.

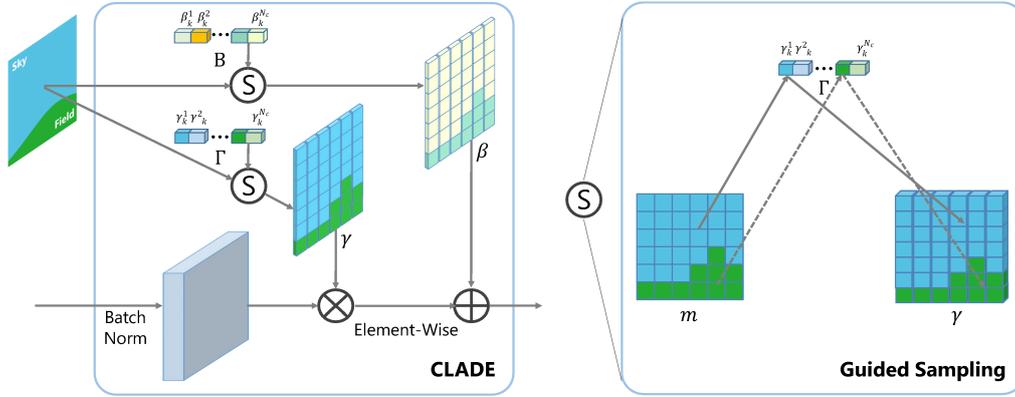
La Figura 2.7 riporta i diagrammi illustrativi delle due tecniche di normalizzazione CLADE e CLADE-IPCE.

2.2.5 INADE

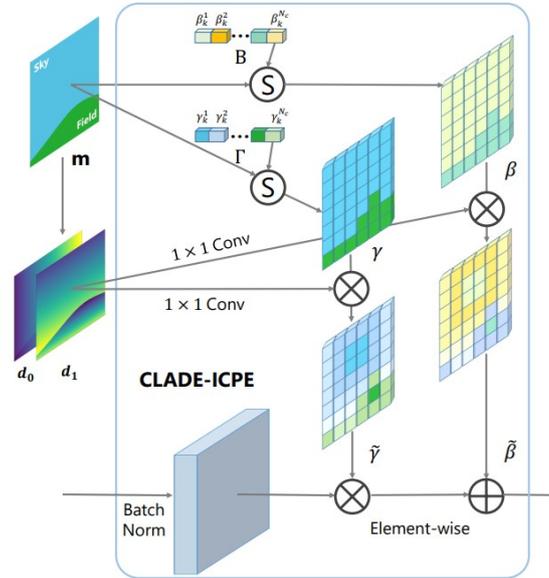
A causa della mancanza di condizionamento dell'istanza, le istanze con la stessa etichetta semantica tendono a convergere verso uno stile simile, danneggiando in modo significativo la diversità della generazione [11]. INADE³ permette di controllare la diversità semantica nelle immagini dal punto di vista delle distribuzioni di probabilità semantiche. Dunque, in aggiunta alla maschera di segmentazione, si utilizza la mappa di istanza, che differenzia le istanze di oggetti diversi che condividono la stessa etichetta semantica (Figura 2.8). La mappa semantica non è in grado di distinguere oggetti affiancati appartenenti alla stessa classe, quindi, utilizzando la mappa delle istanze, la separazione degli oggetti migliora.

L'idea è quella di utilizzare dei modelli di modulazione variazionali che consentano di modellare i parametri di modulazione discreti a livello semantico come distribu-

³Instance-Adaptive DEnormalization



(a) CLADE



(b) CLADE-ICPE

Figura 2.7: Normalizzazione adattata alle classi (a). CLADE utilizza un'efficiente operazione di campionamento guidato per campionare i parametri di modulazione adattati alla classe per ogni regione semantica. Variante con codifica posizionale intra-classe (b). La mappa della codifica posizionale viene calcolata a partire dalla mappa di segmentazione; d_0 e d_1 rappresentano la codifica posizionale lungo le dimensioni x e y . Immagine tratta da [10].

zioni di probabilità continue, in grado di incorporare le caratteristiche generali di ciascuna categoria semantica in modo adattivo alla classe. A partire dai modelli variazionali, si effettua un campionamento stocastico dei parametri di modulazione per raggiungere la diversità a livello di istanza. Inoltre, per facilitare l'addestramento supervisionato e il controllo dello stile dell'istanza in fase di test, si effettua la rimappatura preventiva del rumore. Durante questa fase si trasformano i campioni di rumore in un insieme di parametri di perturbazione lineare codificati a partire da riferimenti arbitrari, preservandone la distribuzione originale.

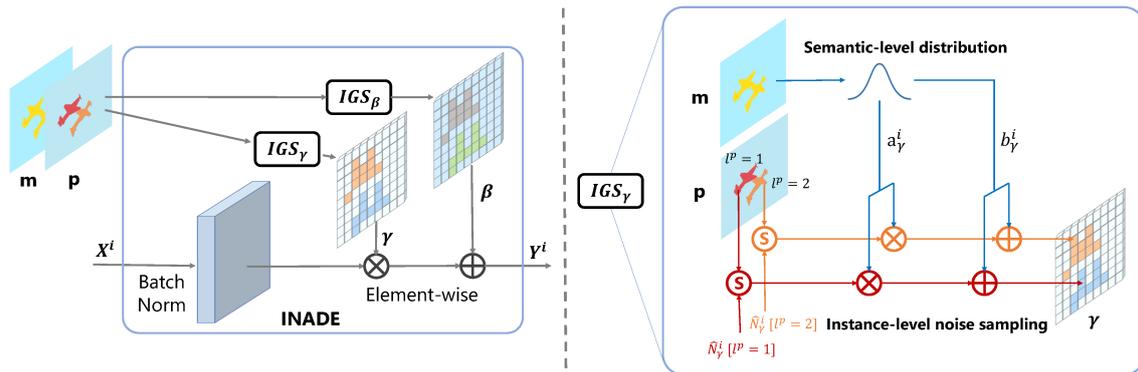


Figura 2.8: Normalizzazione INADE. Combina la distribuzione a livello semantico e il campionamento del rumore a livello di istanza. IGS indica il campionamento guidato dell'istanza. Immagine tratta da [11].

2.3 GAN class-conditional

Le GAN class-conditional facilitano il compito di apprendimento della distribuzione dei dati condizionando le etichette e suddividendo efficacemente i campioni.

Nel seguito verranno presentati una serie di modelli StyleGAN, progettati dai ricercatori della NVIDIA. Queste tipologie di GAN possono essere utilizzate sia in modo non condizionato sia condizionato, per controllare lo stile delle immagini generate. Le StyleGAN condizionali integrano lo spazio latente di ingresso con le condizioni di classe precedentemente corrette da rumore.

2.3.1 StyleGAN

La StyleGAN è un'evoluzione della PGGAN⁴, un modello che inizia il suo addestramento con immagini a bassa risoluzione e progressivamente aumenta la risoluzione

⁴Progressive Growing GAN

aggiungendo dei nuovi layer sia al generatore sia al discriminatore. La StyleGAN effettua importanti modifiche a livello del modello generativo (Figura 2.9).

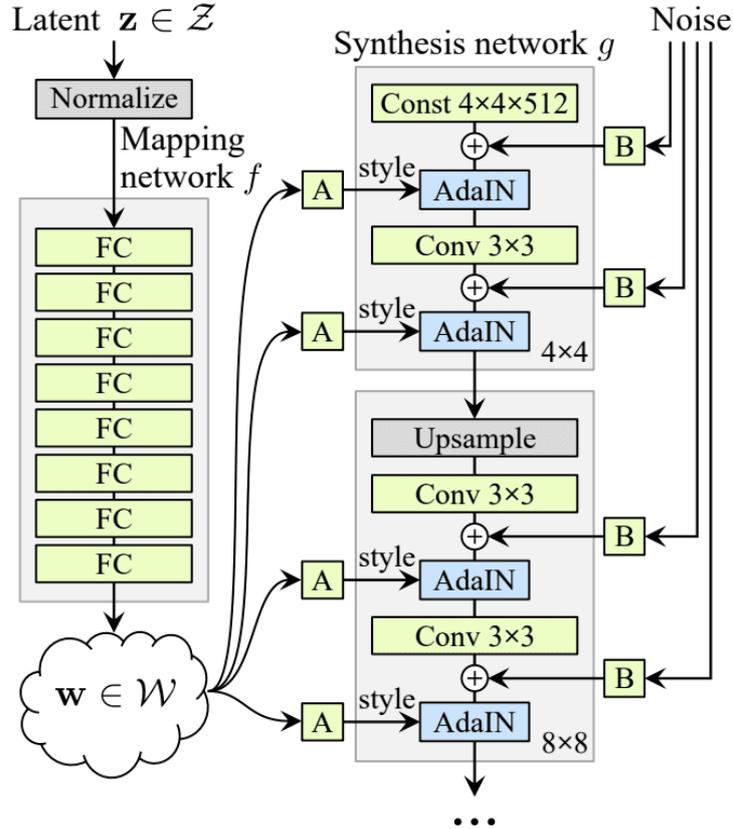


Figura 2.9: Architettura del generatore StyleGAN. Immagine tratta da [12].

In primo luogo, utilizza una rete di mappatura, che consiste in un *multi-layer perceptron*, il cui ruolo è quello di codificare il vettore latente z in ingresso in uno spazio latente intermedio W , affinché sia possibile controllare lo stile dell'immagine generata. Il codice latente w subisce una trasformazione affine A che lo converte in un vettore di stile. Tale vettore viene incorporato in ogni blocco del modello generativo a seguito del layer convoluzionale attraverso un'operazione di normalizzazione adattiva dell'istanza denominata AdaIN⁵. Di conseguenza, la mappa delle caratteristiche viene prima normalizzata in una Gaussiana standard e poi modulata dal vettore di stile mediante i parametri di scala e bias.

In secondo luogo, si introduce rumore come fonte di variazione stocastica in ogni blocco del generatore. Il rumore viene trasmesso a ciascuna mappa di caratteristiche utilizzando i fattori di scala B appresi per ciascuna feature, dunque viene aggiunto

⁵Adaptive Instance Normalization

all'output della convoluzione corrispondente. Queste variazioni stocastiche possono essere regolate senza influire sulla rappresentazione complessiva dell'immagine e sulle caratteristiche ad alto livello.

A differenza delle architetture tradizionali, in cui il vettore latente viene fornito in input al primo layer del generatore, nella StyleGAN la sintesi dell'immagine parte da un tensore costante appreso. Inoltre, è possibile effettuare la miscelazione di stili, dall'inglese *style mixing*. Entrando nello specifico, si fanno scorrere due codici latenti z_1 e z_2 attraverso la rete di mappatura e si ottengono i due vettori latenti intermedi w_1 e w_2 , corrispondenti a due stili differenti. w_1 si applica prima del punto di crossover, mentre w_2 dopo. Questo trucco di regolarizzazione impedisce alla rete di assumere che gli stili adiacenti siano correlati.

Il modello risultante, oltre a produrre immagini di alta qualità, è quindi in grado di controllare lo stile dell'immagine generata a diversi livelli di dettaglio attraverso la variazione dei vettori di stile e del rumore. Tuttavia, nei lavori precedenti, è stato rilevato che la maggior parte delle immagini generate dalla StyleGAN presentavano caratteristici artefatti simili a gocce d'acqua. Questo fenomeno si verificava anche nelle mappe di caratteristiche intermedie, diventando progressivamente più evidente a risoluzioni elevate [39]. È stato constatato che il problema sia riconducibile all'operazione AdaIN, poiché il generatore lascia passare intenzionalmente informazioni relative all'intensità del segnale mediante la normalizzazione dell'istanza. In questo modo, si crea un picco intenso e localizzato che domina le statistiche. Infatti, nel momento in cui viene rimosso lo step di normalizzazione, l'artefatto scompare completamente.

2.3.2 StyleGAN2

L'architettura StyleGAN2 apporta alcune modifiche riguardanti il modello generativo della StyleGAN. L'idea è quella di sostituire la normalizzazione dell'istanza con un'operazione di demodulazione applicata ai pesi associati a ciascun layer convoluzionale. Inizialmente i pesi vengono modulati utilizzando il vettore di stile, il quale contiene le informazioni caratteristiche di una determinata distribuzione di dati, che quindi verranno trasmesse alle operazioni successive. Matematicamente, il processo di modulazione è sintetizzato come segue:

$$w'_{ijk} = s_i \cdot w_{ijk}, \quad (2.8)$$

dove w e w' rappresentano i pesi originali e modulati rispettivamente, s_i è lo stile estratto dallo spazio latente attraverso la trasformazione affine, quindi la scala corrispondente all' i -esima mappa di caratteristiche di input, mentre j e k fanno riferimento agli indici spaziali delle mappe di caratteristiche di output.

Arrivati a questo punto, lo scopo della normalizzazione dell'istanza era essenzial-

mente quello di rimuovere l'effetto di stile dalle caratteristiche in uscita dalla convoluzione. Proprio questo entra in gioco la fase di demodulazione dei pesi, riassunta nella formula seguente:

$$w''_{ijk} = \frac{w'_{ijk}}{\sqrt{\sum_{i,k} w'_{ijk}{}^2 + \epsilon}}, \quad (2.9)$$

dove ϵ è una costante di stabilizzazione.

In conclusione, l'intero blocco di stile è composto da un singolo layer di convoluzione, i cui pesi sono regolati dal vettore di stile utilizzando le equazioni 2.8 e 2.9. Si riporta la schematizzazione del processo all'interno della Figura 2.10.

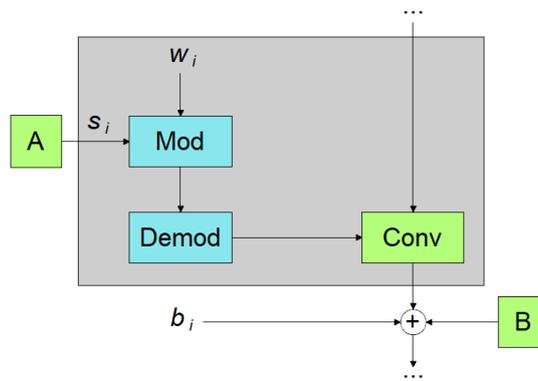


Figura 2.10: Schematizzazione del processo di modulazione e demodulazione dei pesi associati a ciascun layer di convoluzione utilizzando il vettore di stile ottenuto dalla trasformazione affine.

Uno dei fattori che frena l'utilizzo delle GAN è spesso la grande dimensione e qualità del set di dati richiesti per un opportuno addestramento. Se la quantità dei dati non è adeguata a una specifica situazione, è facile incorrere nel problema dell'overfitting: il discriminatore si adatta eccessivamente ai dati forniti in fase di allenamento, l'errore di retropropagazione risulta insignificante per il generatore e l'allenamento inizia a divergere. Tuttavia, la data augmentation come soluzione standard all'overfitting non è sufficiente nei casi in cui il dataset è particolarmente limitato e inoltre non è una via sempre percorribile. A tal proposito, il discriminatore utilizza un meccanismo denominato ADA⁶ (Figura 2.11), che stabilizza in modo significativo l'addestramento della GAN quando si utilizzano set di dati limitati.

Il meccanismo ADA consiste nell'applicare una trasformazione alle immagini generate con una probabilità $p \in [0, 1]$ prima che vengano giudicate dal discriminatore.

⁶Adaptive Discriminator Augmentation

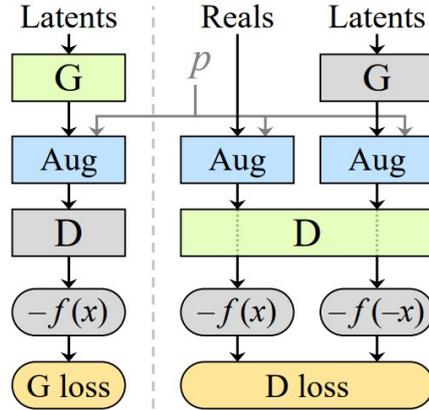


Figura 2.11: Meccanismo ADA. In blu sono riportati i blocchi che caratterizzano operazioni di data augmentation, in arancione quelli che indicano la funzione di perdita, in verde quelli che vengono allenati all'iterazione di riferimento, mentre in grigio quelli temporaneamente bloccati (allenati all'iterazione successiva). G e D rappresentano il generatore e il discriminatore rispettivamente. Immagine tratta da [13].

Anche per bassi valori di p è molto improbabile che il discriminatore riceva un'immagine generata "pulita", perché le operazioni di trasformazione sono molteplici all'interno della pipeline e sono posizionate dopo la fase di sintesi.

Le trasformazioni a cui le immagini sono sottoposte sono molte, tra le più efficaci *pixel blitting* (x-flip, rotazioni di 90° , traslazioni), trasformazioni geometriche e modifiche del colore. Il valore di intensità della trasformazione p è progettato affinché sia controllato dinamicamente sulla base del grado di overfitting dell'allenamento. Esistono sostanzialmente due vie per quantificare il livello di overfitting. Il metodo standard consiste nell'utilizzare un validation set e osservare il suo comportamento rispetto al training set. In caso di overfitting il validation set inizia a comportarsi in modo sempre più simile alle immagini generate. Lo svantaggio di questa tecnica è legato alla necessità di disporre di un set di validazione separato, quando i dati di addestramento sono limitati. Il secondo metodo si basa sulla seguente relazione:

$$r_t = E[\text{sign}(D_{train})] \quad (2.10)$$

Si va a calcolare $r_t \in [0, 1]$, osservando il segno dell'uscita del discriminatore. Idealmente, il valore 0 indica che l'immagine è sintetica, mentre 1 è associato all'immagine reale. Poiché le immagini del training set sono tutte reali, se r_t tende a 1 ci si trova in una condizione di overfitting, se invece tende a 0 non si verifica overfitting. Dunque, gli output del discriminatore per le immagini reali e generate divergono simmetricamente attorno allo zero mano a mano che la situazione peggiora.

2.3.3 StyleGAN3

La StyleGAN3 tenta di risolvere il problema della “texture sticking” che affligge le versioni precedenti, ovvero cerca di rendere le animazioni di transizione più naturali [14]. Tipicamente il processo di sintesi delle GAN dipende fortemente dalle coordinate assolute dei pixel. Questo si manifesta con la generazione di dettagli che sembrano essere incollati alle coordinate dell’immagine anziché alle superfici degli oggetti raffigurati. Questo fenomeno è dovuto alla presenza di aliasing nella rete generatrice, riconducibile all’applicazione puntuale delle non linearità e all’uso di filtri di sovracampionamento non ideali. L’obiettivo è quindi quello di eliminare tutte le fonti di riferimenti spaziali, affinché i dettagli delle immagini possano essere generati ugualmente bene indipendentemente dalle coordinate dei pixel. Per rimuovere questi riferimenti spaziali, la StyleGAN3 propone di rendere la rete completamente equivariante alla traslazione e opzionalmente alla rotazione dei sub-pixel in tutti i layer. Le reti neurali operano su mappe di caratteristiche campionate in modo discreto, per cui è necessario operare nel dominio continuo per sopprimere efficacemente il problema dell’aliasing. Pertanto, è fondamentale riprogettare il generatore della StyleGAN2 per ridurre l’applicazione di operazioni nel discreto e per rendere ogni layer del generatore equivariante rispetto al segnale continuo. Dunque, la caratteristica fondamentale che permette di distinguere la StyleGAN3 dalle versioni precedenti, è l’eseguire una trasformazione di non linearità, come la ReLU, non più a partire da una mappa di attivazioni nel discreto, ma da una mappa di attivazioni nel continuo. È fondamentale chiarire che in questo contesto per continuo si intende sovracampionamento. Per operare il passaggio da un dominio all’altro, e viceversa, gli autori della StyleGAN3 hanno pensato di sostituire il filtro bilineare di sovracampionamento con un’operazione di filtraggio passa-basso e una moltiplicazione puntuale. Nella Figura 2.12 è mostrata una rappresentazione di come si cambia il dominio di una mappa di attivazioni. La frequenza di taglio di questi filtri dipende dal livello di risoluzione in cui ci si trova. I filtraggi dei livelli più profondi del generatore hanno una frequenza di taglio minore, poiché operano sulle caratteristiche a basso livello. Mano a mano che si raggiungono i livelli finali del generatore, le frequenze di taglio dei filtri aumentano per permettere la generazione di immagini con dettagli sempre più fini.

Oltre all’aggiunta del filtraggio passa-basso alle operazioni di campionamento e intorno alle non linearità per evitare aliasing indesiderati, l’architettura del generatore della StyleGAN3 presenta le seguenti modifiche strutturali (Figura 2.13):

- Aggiunta del blocco EMA⁷ prima di ogni layer di convoluzione. Si tratta di un blocco di normalizzazione non allenabile che consiste nella divisione delle attivazioni per la media mobile esponenziale.

⁷Exponential Moving Average

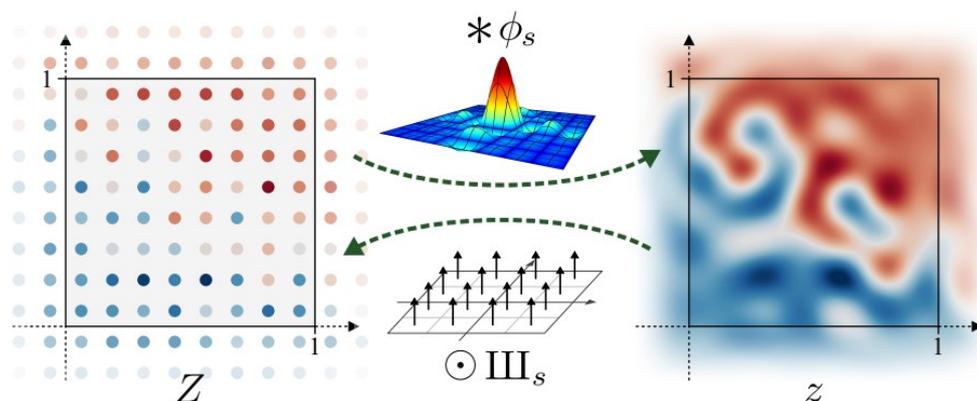


Figura 2.12: Rappresentazione semplificata della modalità che permette di passare da un dominio discreto (Z) a uno continuo (z) e viceversa. Il passaggio da una mappa discreta a una continua è garantito attraverso la convoluzione con un kernel passa-basso, mentre il passaggio inverso è effettuato mediante la moltiplicazione puntuale di una griglia di Dirac. Immagine tratta da [14].

- Sostituzione del tensore costante di input con la *Fourier feature*. Anche in questo caso si tratta di un blocco non allenabile che permette una mappatura profonda delle attivazioni che riceve in input, facilitando la continua traslazione e rotazione di z .
- Rimozione del rumore fornito in input agli strati intermedi del generatore, poiché si è rivelato una fonte di generazione di riferimenti posizionali.
- Riduzione della profondità della *mapping network* per semplificare la configurazione del generatore.

Il generatore della StyleGAN3 corrisponde a quello della StyleGAN2 in termini di FID, pur essendo leggermente più pesante dal punto di vista computazionale [14].

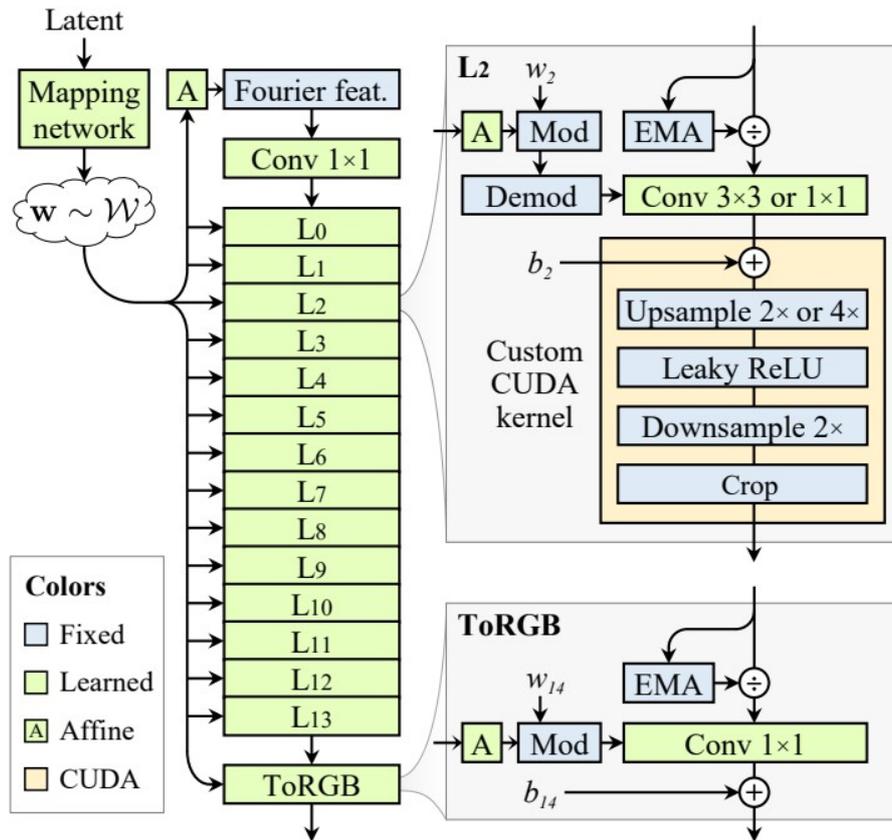


Figura 2.13: Architettura del generatore della StyleGAN3. Immagine tratta da [14].

Capitolo 3

Metodi

3.1 Dataset di partenza

Per lo svolgimento del lavoro di tesi è stato utilizzato il dataset fornito dal Vancouver Prostate Centre [40]. Esso comprende 6 tissue microarrays, ognuno dei quali è composto da una serie di campioni di tessuto, per un totale di 333 WSI¹. I campioni sono stati prelevati da 231 pazienti che hanno subito un intervento di prostatectomia radicale all'ospedale di Vancouver tra giugno del 1997 e giugno del 2011. Ciascun vetrino è stato colorato con ematossilina-eosina e successivamente digitalizzato mediante scanner SCN400 (Leica Microsystems) con una risoluzione di 40x. La dimensione delle immagini varia da 4608 a 5632. Sei patologi con in media 16 anni di esperienza nella classificazione del carcinoma prostatico hanno annotato le WSI, assegnando a ciascuna regione tissutale una delle seguenti 5 classi: Benigna, Gleason Score 3, Gleason Score 4, Gleason Score 5 e Altro. In particolare, quattro patologi hanno annotato tutti i campioni, mentre i restanti due hanno etichettato rispettivamente 191 e 92 immagini. Le annotazioni sono state realizzate in un arco temporale compreso tra dicembre del 2016 e ottobre del 2017. Un esempio di WSI con le relative annotazioni manuali è mostrato nella Figura 3.1.

3.2 Pipeline

L'obiettivo del progetto è la realizzazione di due modelli generativi automatici condizionali: il primo che permetta di simulare delle maschere di segmentazione semantica mediante l'implementazione di una rete GAN a condizione di classe, il secondo invece che consenta di sintetizzare delle immagini istopatologiche prostatiche a partire da un determinato ground truth, effettuando una sintesi semantica. Dunque, l'int-

¹Whole Slide Image

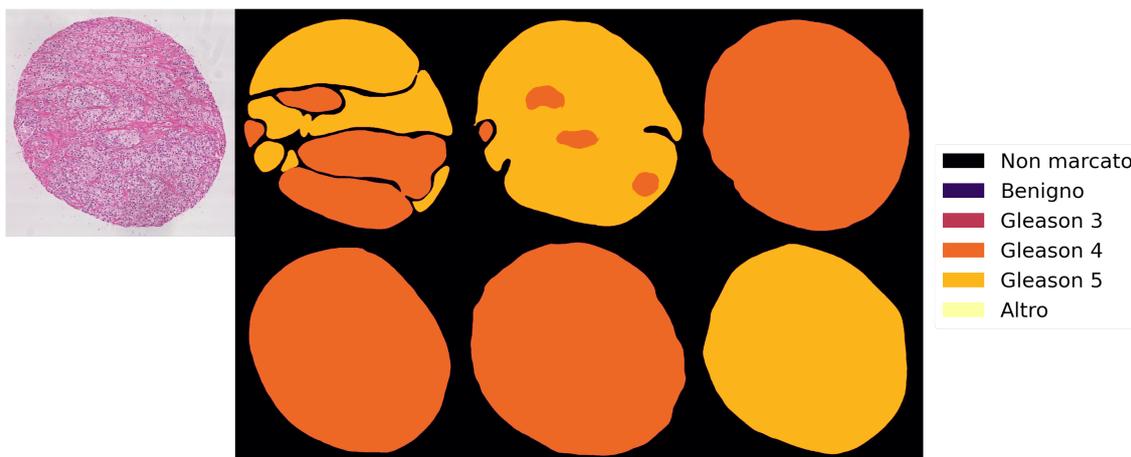


Figura 3.1: Esempio di WSI (sinistra) e le relative annotazioni manuali fornite dai patologi (destra).

ro lavoro di tesi prevede sostanzialmente due strade distinte e parallele, che a loro volta possono essere suddivise in tre fasi:

- **Creazione del dataset:** completamento manuale delle annotazioni e realizzazione di un unico ground truth label, pre-processing delle immagini istopatologiche, creazione delle maschere dei nuclei e dei bianchi, estrazione e selezione delle patch e divisione dell'intero set di dati in training set e test set.
- **Implementazione del modello:** scelta del modello e ottimizzazione degli iperparametri di training.
- **Valutazione dei risultati:** calcolo di opportuni parametri/metriche.

All'interno della Figura 3.2 è raffigurato il diagramma che illustra i principali stadi del lavoro di tesi.

3.3 Pre-processing

Prima di effettuare la simulazione, è stato inevitabile eseguire una fase di pre-processing del set di dati a disposizione. Di seguito verranno illustrate dettagliatamente le strategie utilizzate per la preparazione dei dataset che saranno impiegati nell'allenamento e validazione dei modelli.

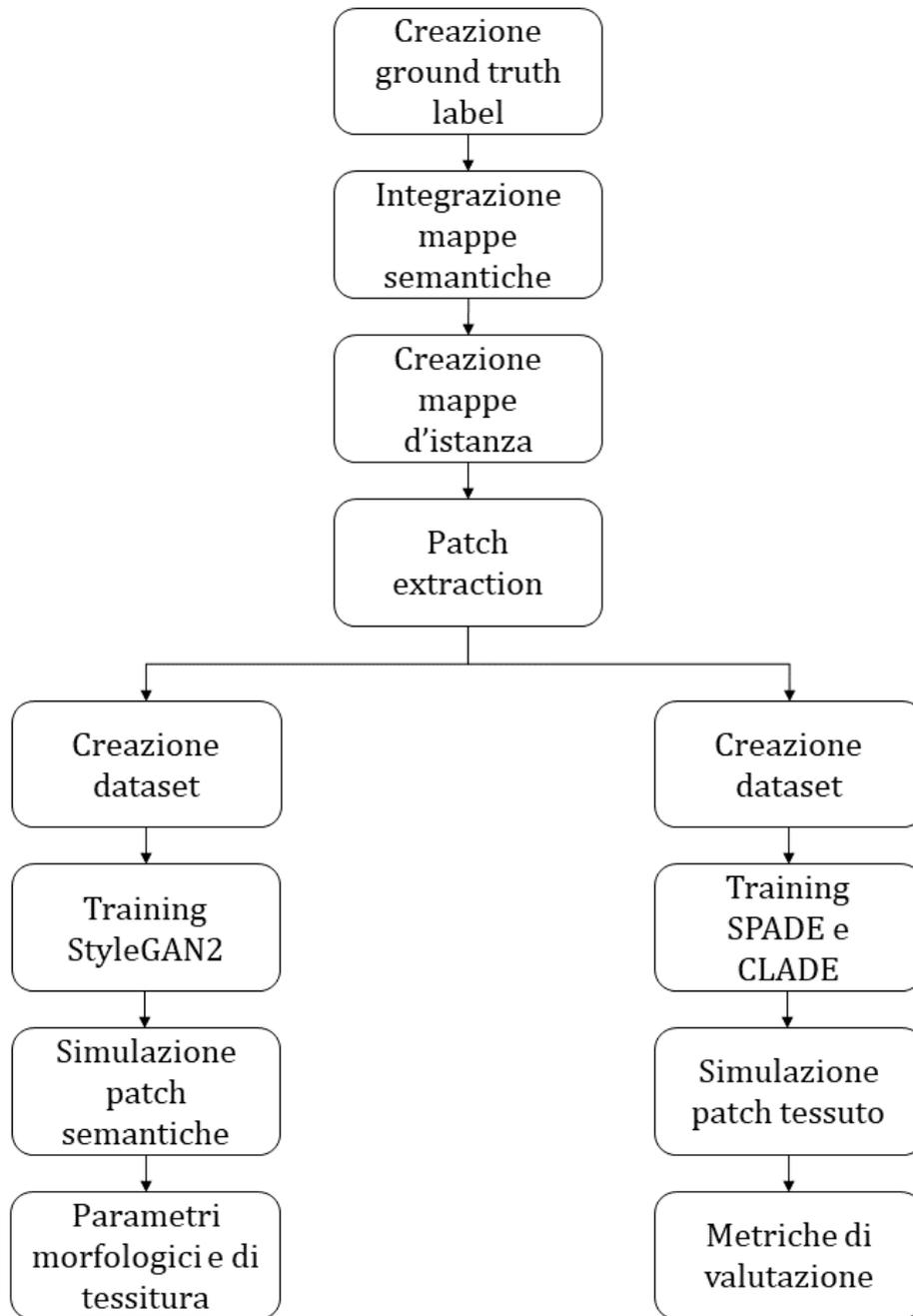


Figura 3.2: Pipeline del lavoro.

3.3.1 Creazione dei ground truth label

Attraverso l'ispezione visiva delle annotazioni manuali, è stato notato che alcune di esse risultavano incomplete, poiché presentavano delle regioni tissutali evidenziate con un contorno aperto e non pienamente marcate. Dunque, è stato necessario selezionare le maschere che mostravano tale problema, chiudere manualmente il contorno di marcatura e imporre all'area d'interesse l'etichetta assegnata dal patologo. Si riporta un esempio nella Figura 3.3.

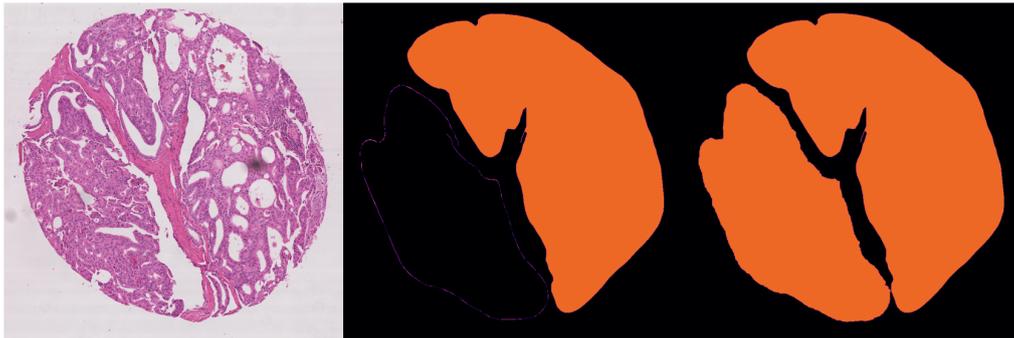


Figura 3.3: Esempio di WSI (sinistra), relativa annotazione manuale del patologo 2 (centro) e completamento della maschera (destra).

Dal momento che il dataset di partenza era costituito da annotazioni multiple, si è dovuta applicare una tecnica che consentisse di combinare le varie maschere semantiche di ciascun esperto per creare un ground truth label altamente accurato e robusto. La sua realizzazione è stata possibile implementando l'algoritmo STAPLE² [41]. Si tratta di un algoritmo di voto ponderato che considera una serie di segmentazioni e calcola sia una stima probabilistica della segmentazione “reale” sia una misura delle performance rappresentata da ciascuna segmentazione. La stima probabilistica della “reale” segmentazione è ottenuta stimando una combinazione ottimale delle segmentazioni, ponderando ciascuna di esse in base al livello di prestazione stimato e incorporando un modello per la distribuzione spaziale delle strutture segmentate, nonché i vincoli di omogeneità spaziale. Dunque, non tiene conto soltanto della deviazione sistematica causata dalla differenza delle annotazioni manuali, ma valuta anche la qualità delle annotazioni di ciascun esperto. L'output è quindi un'immagine costituita da pixel con valore numerico compreso tra 0 e 1, che rappresentano la probabilità che ciascun pixel appartenga a una specifica etichetta semantica. Tuttavia, nel presente lavoro l'algoritmo STAPLE è stato adattato a mappe multi-label. Se un pixel aveva la massima probabilità di appartenere a due classi differenti, allora non è stata attribuita alcuna etichetta.

²Simultaneous Truth And Performance Level Estimation

Inoltre, per ottenere dei ground truth label maggiormente sicuri, è stata applicata la tecnica di votazione a maggioranza (*majority voting* in inglese) solamente alle aree tissutali di dubbia classificazione per la maggior parte dei patologi. In questo modo tutte le zone non annotate dalla maggioranza degli operatori, a cui l'algoritmo STAPLE aveva attribuito una label, non sono state marcate. La Figura 3.4 illustra un esempio di ground truth label, ottenuto dall'unione delle annotazioni manuali disponibili.

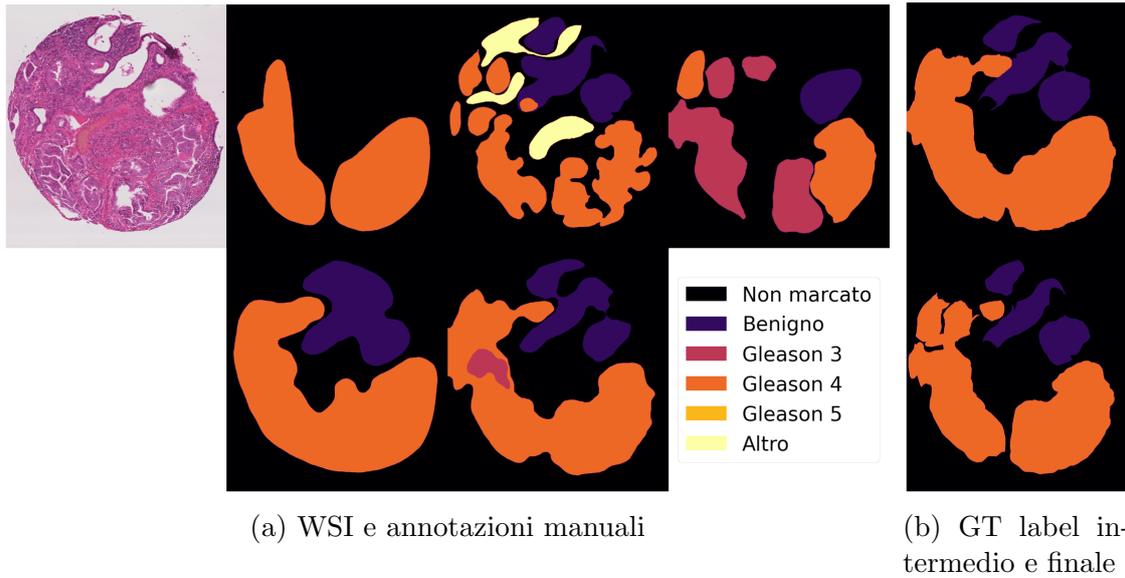


Figura 3.4: Esempio di WSI con le relative annotazioni manuali dei diversi patologi (a) e ground truth label ottenuto in seguito dell'applicazione dell'algoritmo STAPLE (b, in alto) e successiva integrazione del majority voting (b, in basso).

3.3.2 Integrazione delle mappe semantiche e creazione delle mappe d'istanza

Mediante appositi algoritmi di ottimizzazione, normalizzazione e segmentazione delle immagini, sono state prima ottenute le WSI ottimizzate e normalizzate e successivamente, da queste ultime, sono state ricavate le seguenti maschere:

- **Bianchi.** Si tratta di maschere binarie che identificano la zone dell'immagine appartenenti al vetrino, non incluse nello stroma e nei nuclei. Da queste maschere sono stati eliminati i piccoli artefatti.
- **Nuclei.** Si tratta di maschere in scala di grigi in cui si differenziano i nuclei (pixel a 255) e i rispettivi bordi (pixel a 127) dal resto dell'immagine.

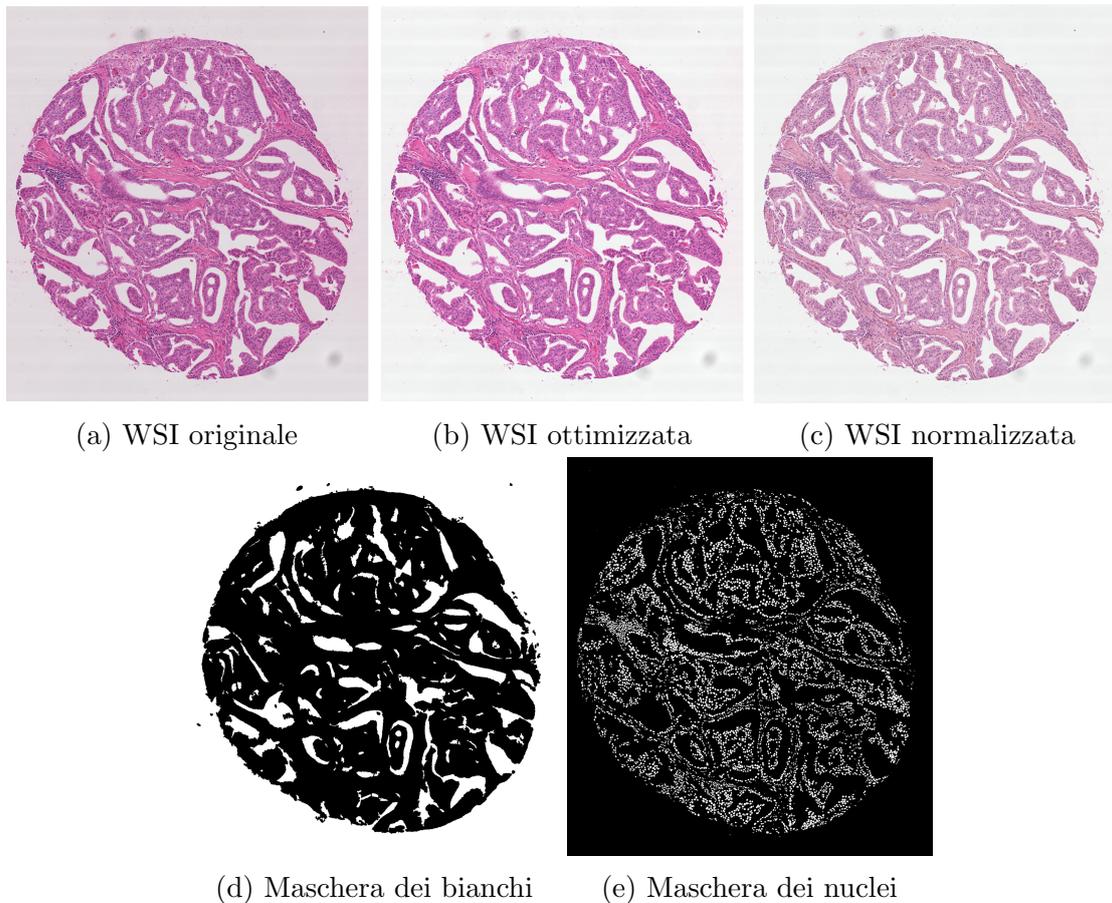


Figura 3.5: Esempio di immagine istologica originale, ottimizzata e normalizzata e relative maschere dei bianchi e dei nuclei.

Nella Figura 3.5 si riporta un esempio di WSI ottimizzata e normalizzata con le relative maschere dei bianchi e dei nuclei.

Tali maschere sono state utilizzate per due scopi: l'integrazione dei bianchi e dei nuclei all'interno delle mappe semantiche e la creazione delle mappe d'istanza. In primo luogo, sono state modificate le etichette dei ground truth label, andando a differenziare le regioni dell'immagine appartenenti allo sfondo dalle aree tissutali non marcate perché di dubbia classificazione. Quindi per ciascuna WSI, a partire dalla maschera dei bianchi, è stata ricavata la maschera dello sfondo, successivamente integrata all'interno della corrispondente mappa semantica. La conseguente ripartizione delle classi è stata modificata in questo modo: Sfondo, Benigna, Gleason 3, Gleason 4, Gleason 5, Non marcato/Altro. Affinché il processo di sintesi fosse maggiormente guidato nella localizzazione dei nuclei e dei bianchi, è stato pensato di arricchire l'informazione semantica contenuta nelle mappe di segmentazione aggiun-

gendo due nuove classi. Così facendo il numero di etichette semantiche è aumentato a otto: Sfondo, Benigno, Gleason 3, Gleason 4, Gleason 5, Nuclei, Bianchi, Non marcato/Altro.

In secondo luogo, sono state create le cosiddette mappe d'istanza, che sono sostanzialmente delle maschere binarie contenenti i contorni dei nuclei e dei bianchi. Esse verranno fornite in input alla GAN insieme alle mappe di segmentazione semantica solamente nel task di sintesi semantica di immagini istologiche, al fine di distinguere in maniera netta i nuclei e i bianchi dal resto dell'immagine.

La Figura 3.6 mostra un esempio di mappa semantica e mappa d'istanza.

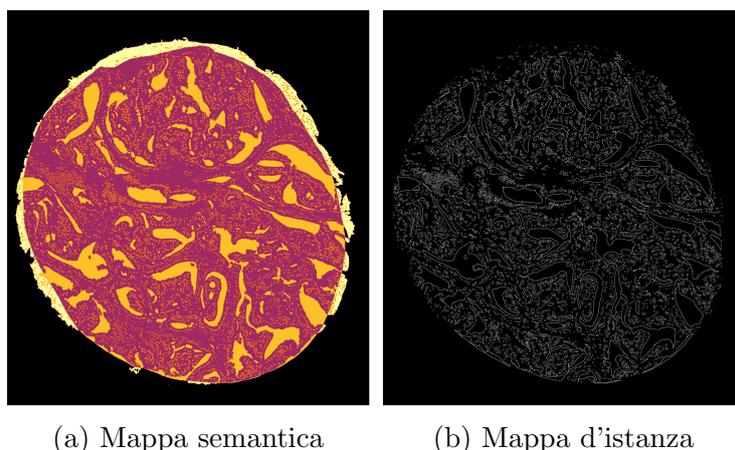


Figura 3.6: Esempio di mappa semantica modificata e mappa d'istanza creata. Nella mappa semantica (a) si riporta la zona Gleason 4 (viola), i nuclei (arancione), i bianchi (giallo) e l'area non marcata (bianco); mentre nella mappa d'istanza (b) sono mostrati i bordi dei nuclei e dei bianchi. Le maschere d'esempio sono relative alla WSI della Figura 3.5a.

Da questo punto della trattazione in poi si farà riferimento alle immagini istologiche ottimizzate, poiché saranno impiegate nella creazione del dataset per la simulazione semantica.

3.3.3 Estrazione delle patch

I parametri di estrazione delle patch giocano un ruolo fondamentale, poiché determinano il grado di informazione contenuto nelle diverse patch e il numero di elementi ottenuto. Le caratteristiche istologiche del tessuto possono essere significative a risoluzioni differenti, dunque è importante che la scala di dimensione impostata consenta di ricavare delle patch che contengano al proprio interno le caratteristiche architettoniche necessarie per la loro sintesi. Patch più grandi di bassa scala presentano delle caratteristiche istologiche meno localizzate, mentre patch più piccole di scala elevata

potrebbero non contenere sufficienti informazioni semantiche. Si riporta un esempio di patch di ugual dimensione ma appartenenti a una scala d'ingrandimento differente (Figura 3.7).

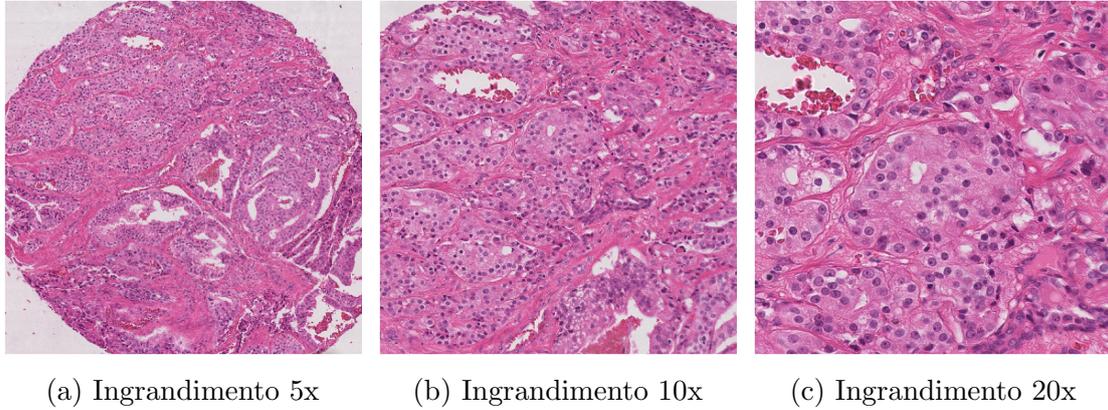


Figura 3.7: Patch di dimensione 512x512 estratte dalle stessa WSI, ma con differente scala d'ingrandimento.

Un altro parametro fondamentale è l'overlap, ovvero la sovrapposizione tra patch adiacenti, che regola in maniera diretta il numero di patch estratte da ciascuna immagine. Se l'overlap è troppo elevato, si ottiene un insieme di dati altamente ripetitivo e il tempo computazionale richiesto per l'estrazione risulta essere notevole. Al contrario, un overlap troppo basso, produce un numero insufficiente di esempi. Per bilanciare i possibili effetti è stato impostato un valore di overlap pari al 50% e una dimensione delle patch di 512x512. Prima di applicare la sliding window ed effettuare la divisione in patch, è stato scelto di sottocampionare i dati di un fattore 2, ottenendo immagini con ingrandimento a 20x.

	Numerosità
Benigna	2557
Gleason 3	9305
Gleason 4	15442
Gleason 5	527
Totale	35340

Tabella 3.1: Numerosità delle patch estratte e selezionate per ciascuna classe d'interesse.

Poiché il presente lavoro è finalizzato alla sintesi di immagini istopatologiche di tessuto prostatico aventi un determinato grado di patologia, è stata applicata una specifica condizione di estrazione. Sono state selezionate solamente le patch che contenessero almeno il 90% di una stessa classe tra quelle d’interesse, ovvero “Benigna”, “Gleason 3”, “Gleason 4” e “Gleason 5”.

Nella Tabella 3.1 si riporta la numerosità delle patch ottenute per ciascuna classe.

3.3.4 Costruzione dei dataset

Da questo punto in poi del lavoro, si procede parallelamente con la simulazione di mappe di segmentazione mediante l’implementazione di un modello generativo a condizione di classe da un lato e la generazione semantica di immagini istopatologiche dall’altro. I dataset utilizzati per i due task di simulazione sono differenti, poiché adattati al compito e al tipo di architettura scelta. Nel primo caso, la scelta di operare con le maschere di segmentazione piuttosto che con le immagini istologiche è dovuta alla semplificazione del task di generazione e del modello implementato, poiché si lavora con immagini in scala di grigi. Inoltre, il sistema dovrebbe imparare a focalizzarsi esclusivamente sul contenuto semantico associato a ciascun grado patologico, senza considerare altri aspetti di minor interesse, come lo staining del campione.

Task 1

Per la generazione di mappe di segmentazione semantica è stato utilizzato un dataset costituito esclusivamente dal training set. Esso era composto da un totale di 8000 mappe di segmentazione in scala di grigi, 2000 per ciascuna classe. Poiché le immagini appartenenti al gruppo “Gleason 5” erano solamente 527, è stata effettuata la data augmentation. In particolare, a tutte le maschere sono state applicate le seguenti trasformazioni: rotazione di 90°, 180°, 270°, flip orizzontale e flip verticale (Figura 3.8). Successivamente, è stato selezionato in modo casuale un numero sufficiente di maschere aumentate, affinché si raggiungesse la soglia richiesta.

È fondamentale precisare che per semplificare il task di generazione, è stata effettuata una conversione del valore dei pixel associati a ciascuna label, affinché venisse sfruttata l’intera dinamica dell’immagine. Perciò, le diverse aree semantiche sono state categorizzate in modo tale che la distanza tra i pixel appartenenti a etichette differenti fosse massimizzata, massimizzando di conseguenza il contrasto tra le strutture istologiche. A ciascuna label sono stati associati i seguenti valori:

- Sfondo \rightarrow 0
- Benigno \rightarrow 35
- Gleason 3 \rightarrow 72

- Gleason 4 → 108
- Gleason 5 → 145
- Nuclei → 181
- Bianchi → 218
- Altro/Non marcato → 255

Inoltre, dal momento che il task era di tipo condizionale, all'interno del dataset è stato inserito un file *json* che riportava il nome e la classe di appartenenza di ogni mappa di segmentazione: 0 per la classe Benigna, 1 per la classe Gleason 3, 2 per la classe Gleason 4 e 3 per la classe Gleason 5.

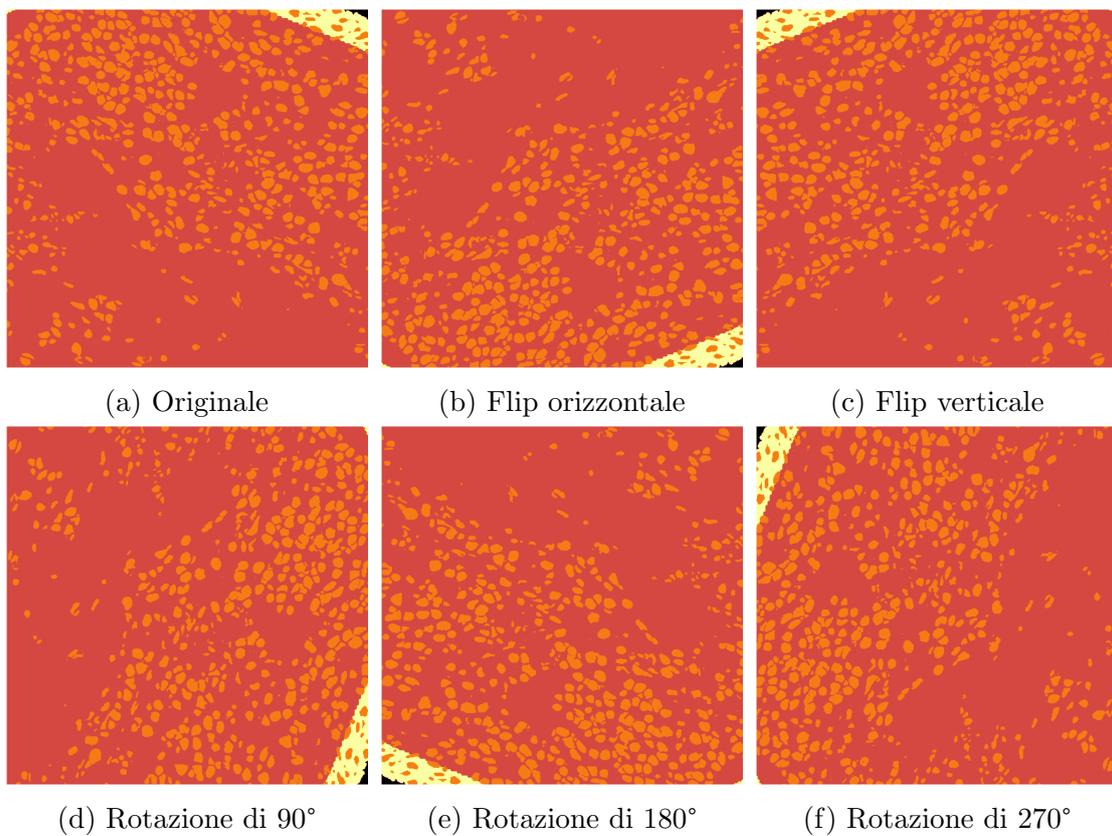


Figura 3.8: Esempio di data augmentation su una patch di classe Gleason 5.

Task 2

Differentemente da quanto fatto in precedenza, il dataset utilizzato per il task di generazione semantica è stato suddiviso in Training set e Test set. La distribuzione

delle immagini e delle relative mappe semantiche e mappe d'istanza nei due gruppi è stata casuale, affinché il training set fosse comprendesse il 70% dei dati a disposizione. Nella Figura 3.9 è mostrato un esempio di patch con le relative mappe d'istanza e semantica, che appartengono al dataset creato.

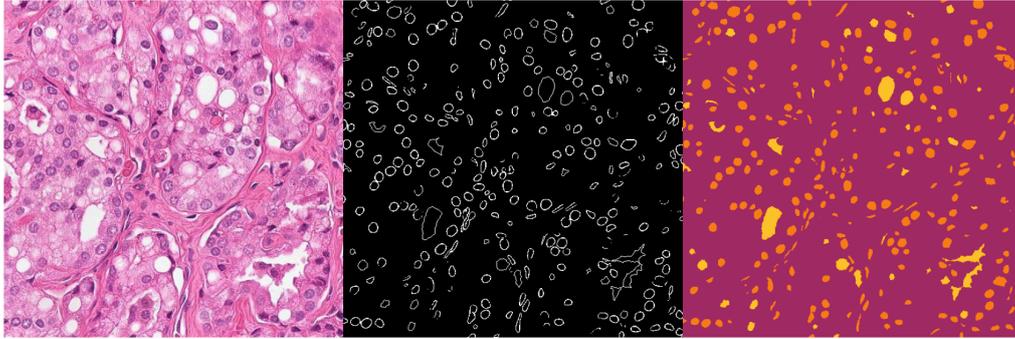


Figura 3.9: Esempio di immagine del Test set (sinistra) appartenente alla classe Gleason 4, mappa d'istanza (centro) e mappa di segmentazione semantica (destra).

A ciascun pixel delle maschere di segmentazione semantica è stato attribuito uno specifico valore in base alla label di appartenenza:

- Sfondo \rightarrow 0
- Benigno \rightarrow 1
- Gleason 3 \rightarrow 2
- Gleason 4 \rightarrow 3
- Gleason 5 \rightarrow 4
- Nuclei \rightarrow 5
- Bianchi \rightarrow 6
- Altro/Non marcato \rightarrow 7

La struttura del dataset impiegato nella generazione semantica e la distribuzione delle immagini nelle varie classi sono mostrate all'interno della Tabella 3.2.

3.4 Architetture GAN

Questo punto della trattazione è dedicato alla scelta dell'architettura GAN, che possa essere promettente nell'adattarsi al tipo di dati a disposizione.

Per quanto riguarda la fase di simulazione di mappe di segmentazione semantica,

	Training set	Test set
Benigno	1787	770
Gleason 3	6503	2802
Gleason 4	10816	4626
Gleason 5	376	151
Totale	19482	8349

Tabella 3.2: Struttura del dataset utilizzato per la generazione semantica di immagini istologiche.

si è scelto di implementare una rete GAN che effettuasse una generazione a condizione di classe. Tra le tre possibili versioni di StyleGAN presentate all'interno del Capitolo 2 è stata scelta la StyleGAN2, poiché computazionalmente più leggera rispetto alla StyleGAN3 e più prestante della StyleGAN classica. D'altra parte, per la sintesi di immagini istopatologiche si è tentato l'utilizzo di tre modelli di generazione semantica che si basano sull'architettura pix2pix con generatore SPADE e discriminatore PatchGAN multiscala illustrati anch'essi nel capitolo precedente. La differenza sostanziale tra questi modelli ricade sulla tecnica di normalizzazione implementata all'interno dei vari blocchi che compongono il generatore: SPADE, CLADE e INADE. Dopo il processo di prova, l'architettura INADE è stata scartata, poiché produceva immagini sintetiche di qualità nettamente inferiore e non sempre realistiche rispetto agli altri due modelli.

3.5 Tuning dei parametri e training

A partire da un modello è possibile apportare diverse modifiche sia a livello strutturale sia a livello degli iperparametri, i quali costituiscono tutte le variabili della rete che possono essere inizializzate e ottimizzate prima della fase di allenamento. In questo lavoro sono state effettuate numerose sperimentazioni per trovare la combinazione di iperparametri migliore per entrambi i task di generazione.

3.5.1 StyleGAN2

I parametri di allenamento utilizzati per la StyleGAN2 sono correlati con il numero di GPU e la risoluzione delle immagini, pari a 512x512, ma certamente dipendono anche dalla natura dei dati che si sta trattando. Dunque, a partire dalle impostazioni

di training consigliate dagli autori [42] sono state effettuate diverse prove, principalmente legate a uno dei principali iperparametri: γ . Esso costituisce il mezzo di regolarizzazione dei pesi durante la fase di addestramento. Un valore di γ elevato predilige la qualità della generazione, mentre un basso valore la diversificazione dei campioni. Poiché è stato difficile ottenere un addestramento stabile che non sfociasse nel collasso della modalità, è stato deciso di settare γ a un valore estremo elevato, pari a 50, per consentire alla rete di compiere un allenamento stabile. Dopo 350 epoche, il valore di γ è stato diminuito significativamente a 2 per le restanti 150 epoche, affinché fosse garantita la variabilità intra-classe dei campioni. Nonostante questa configurazione producesse maschere migliori, la diversificazione dei prodotti non è stata propriamente ottenuta. Tuttavia, in fase di inference si è riusciti a raggiungere dei buoni risultati.

Dal momento che, generalmente, lavorare con immagini istologiche prostatiche annotate presuppone una situazione di scarsa numerosità dei dati iniziali, durante l'allenamento sono stati abilitati tutti i meccanismi di StyleGAN2 che scongiurassero l'overfitting. Perciò, è stato utilizzato sia il meccanismo ADA descritto nel Capitolo 2 sia l'aumento dei dati mediante l'impostazione *mirror*, che permetteva di raddoppiare il dataset di partenza effettuando il flip orizzontale di ciascuna immagine. In questo modo, la StyleGAN2 è stata addestrata su un dataset di 16000 campioni.

La dimensione di batch è stata mantenuta a 8, mentre i learning rate del generatore e del discriminatore sono stati settati a 0,001 per le prime 350 epoche e sono stati poi ridotti a 0,0001 per le restanti 150 epoche. Per quanto riguarda l'algoritmo di ottimizzazione si utilizza ADAM³ con $\beta_1 = 0, \beta_2 = 0.99$.

Disponendo di un dataset costituito da immagini in scala di grigi e di risoluzione non particolarmente elevata, si è pensato di alleggerire il modello, accelerando l'addestramento della rete.

Per quantificare le prestazioni della GAN è stato valutato il grado di avanzamento dell'allenamento mediante la metrica FID⁴. Essa è in grado di stimare la somiglianza delle immagini sintetiche rispetto a quelle originali, calcolando la distanza tra i vettori delle feature delle immagini reali e quelli delle immagini generate. Un valore di FID prossimo allo zero è indice di perfetta somiglianza tra i due gruppi di immagini. Il grafico riportato all'interno della Figura 3.10 mostra l'andamento della FID al variare delle epoche durante l'addestramento della StyleGAN2. La FID è stata calcolata a intervalli regolari (ogni 50 epoche). Come si può osservare, nella fase iniziale, il miglioramento della FID è maggiore e quindi la convergenza delle immagini simulate rispetto a quelle originali è più rapida. Nella parte finale del training invece la FID descrive un plateau, raggiungendo il minimo a 49,8.

³ADaptive Moment estimation

⁴Fréchet Inception Distance

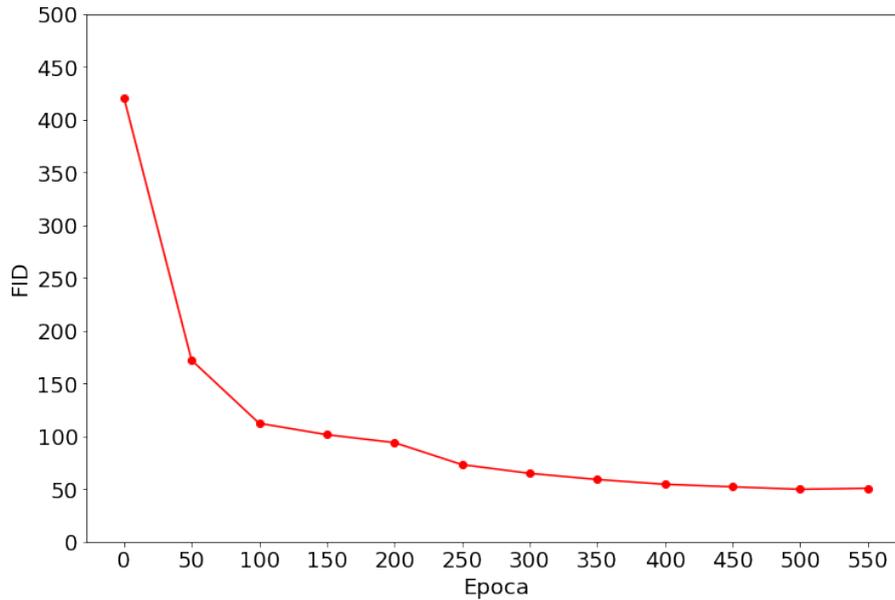


Figura 3.10: Andamento della FID durante il training della StyleGAN2.

3.5.2 SPADE e CLADE

Per quanto concerne la generazione semantica, gli iperparametri di allenamento impostati per SPADE e CLADE sono i seguenti. Anche in questo caso è stato utilizzato l'algoritmo di ottimizzazione ADAM con $\beta_1 = 0$, $\beta_2 = 0.99$ e learning rate iniziale pari a 0.0002. La dimensione di batch è stata settata a 1 e l'addestramento è stato effettuato su 100 epoche; durante le prime 80 epoche il learning rate è mantenuto costante, mentre nelle restanti 20 epoche viene fatto decadere linearmente a zero.

3.6 Metodi di valutazione

Per quantificare numericamente la bontà dei risultati di generazione sono stati utilizzati differenti parametri e metriche di validazione, che saranno discussi dettagliatamente nei seguenti paragrafi. Specificatamente, la valutazione della sintesi delle maschere di segmentazione avviene mediante l'uso di parametri morfologici e descrittori di tessitura. Mentre, per il task di generazione semantica sono state confrontate le immagini sintetiche con quelle reali, sia per il Training set sia per il Test set, utilizzando specifiche metriche di qualità.

3.6.1 Parametri morfologici

I parametri morfologici possono essere molto utili per confrontare immagini istologiche, poiché permettono di comprendere l'architettura del tessuto. Nel presente lavoro, è stato deciso di ricavare alcuni parametri morfologici caratteristici delle immagini reali, che permettano di distinguere il grado della patologia. In particolare, sono stati calcolati: numerosità, area, perimetro ed eccentricità, sia dei nuclei sia dei bianchi. Quest'ultimo è una misura della deviazione della forma di un oggetto dalla forma circolare ideale. Dunque, un valore di eccentricità prossimo allo zero è sinonimo di un oggetto con forma circolare, mentre i nuclei e i bianchi di forma allungata o irregolare possiedono un valore di eccentricità più elevato.

Prima di stimare i parametri morfologici, le maschere di segmentazione generate sono state categorizzate in modo analogo a quelle originali, poiché la StyleGAN2 non è in grado di garantire la corretta suddivisione delle etichette. Successivamente, per ciascuna maschera di segmentazione sintetica vengono calcolati i parametri elencati in precedenza. In questo modo, è stato possibile confrontare le distribuzioni dei valori ottenute per una determinata classe dai dati di riferimento e dai campioni simulati.

3.6.2 Feature di tessitura

L'analisi della tessitura permette di descrivere la natura del tessuto attraverso una serie di informazioni relative al valore assoluto dei pixel e alla loro distribuzione all'interno dell'immagine. Per estrarre le caratteristiche si utilizzano dei descrittori statistici, indipendenti dalla modalità di imaging, che non forniscono alcuna informazione a livello morfologico. I descrittori del primo ordine sono delle funzioni matematiche che estraggono feature a partire dai singoli valori dei pixel delle immagini, perciò si basano sull'istogramma delle luminosità; quelli del secondo ordine o superiore invece considerano la dipendenza che intercorre tra due o più pixel dell'immagine e si basano su specifiche matrici, a partire dalle quali vengono calcolati i descrittori.

Prima di presentare nel dettaglio i descrittori impiegati nel presente progetto, è fondamentale dire che le maschere di segmentazione semantica reali e simulate sono state elaborate, affinché i parametri di tessitura calcolati potessero essere confrontabili tra le varie classi. Dunque, è stato deciso di settare provvisoriamente le label in questo modo:

- Grado della patologia $\rightarrow 0$
- Nuclei $\rightarrow 127$
- Bianchi $\rightarrow 255$

Descrittori del primo ordine

Come anticipato, i descrittori del primo ordine dipendono esclusivamente dal singolo valore del pixel e sono indipendenti dalla relazione spaziale tra i pixel dell'immagine. Questi parametri si ricavano dall'istogramma delle luminosità, che fornisce informazioni sulla distribuzione dei livelli di grigio all'interno della maschera di segmentazione. Si tratta quindi di una rappresentazione grafica bidimensionale, in cui sull'asse delle ascisse si riportano i valori di grigio, mentre sull'asse delle ordinate il numero di volte che compare tale valore nei pixel dell'immagine. I descrittori del primo ordine che sono stati adoperati in questo lavoro sono la media, la varianza, la skewness, la curtosi e l'entropia dell'istogramma, le cui formule matematiche sono descritte nella Tabella 3.3.

Feature I ordine	Formula
Media	$\mu = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M I(i, j)$
Varianza	$\sigma^2 = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (I(i, j) - \mu)^2$
Skewness	$\frac{1}{NM} \frac{1}{\sigma^3} \sum_{i=1}^N \sum_{j=1}^M (I(i, j) - \mu)^3$
Curtosi	$\frac{1}{NM} \frac{1}{\sigma^4} \sum_{i=1}^N \sum_{j=1}^M (I(i, j) - \mu)^4$
Entropia	$-\sum_{i=1}^N \sum_{j=1}^M I(i, j) \log_2 I(i, j)$

Tabella 3.3: Descrittori del primo ordine.

In particolare, la media rappresenta il valore medio di intensità dei pixel dell'immagine. La varianza esprime la dispersione dell'intensità dei pixel dell'immagine; un valore elevato indica che i pixel sono molto distribuiti, mentre un valore basso che i pixel sono vicini all'intensità media. La curtosi è una misura di quanto la distribuzione dei valori di intensità dei pixel si discosta dalla distribuzione normale; un valore positivo implica una distribuzione "appuntita" e concentrata intorno alla media, mentre una curtosi negativa denota una distribuzione "piatta" e meno concentrata. La skewness è una misura dell'asimmetria della distribuzione dei valori di intensità dei pixel rispetto alla media; un valore positivo è indice di una distribuzione "spostata" verso destra rispetto alla media, mentre una skewness negativa fa riferimento a una distribuzione "spostata" verso sinistra. L'entropia quantifica la caoticità della distribuzione dei livelli di grigio dell'immagine, dunque è una misura di complessità del tessuto che valuta la sua varietà e disorganizzazione; una distribuzione uniforme avrà un'entropia elevata, mentre una distribuzione con un solo valore di intensità avrà un'entropia bassa.

Descrittori di Haralick e Galloway

I descrittori di Haralick e Galloway, conosciuti come descrittori del secondo ordine, consentono di valutare la dipendenza spaziale tra i pixel dell'immagine. Per calcolarli è necessario costruire la matrice delle co-occorrenze, o GLCM⁵, cioè una matrice quadrata le cui righe e colonne rappresentano tutti i livelli di grigio che possono assumere i pixel dell'immagine. Ogni cella della matrice contiene il numero di volte in cui un pixel con un determinato tono di grigio è vicino a un altro pixel di un altro colore, dunque riporta quante volte occorre l'accoppiamento di due pixel nell'immagine. Proprio per la sua conformazione, la matrice delle co-occorrenze è nota come istogramma di secondo livello, dal momento che costituisce una rappresentazione bidimensionale dei livelli di grigio dell'immagine.

Attraverso la GLCM è dunque possibile valutare la presenza di pattern e l'omogeneità delle maschere di segmentazione semantica. La matrice delle co-occorrenze è stata calcolata per quattro angoli di inclinazione differenti (0°, 45°, 90°, 135°). I descrittori del secondo ordine ottenuti da ciascuna delle quattro matrici sono stati mediati. Le cinque feature calcolate sono le seguenti: il contrasto, la correlazione, l'omogeneità, la dissimilarità e l'energia, le cui formule matematiche sono riportate all'interno della Tabella 3.4.

Feature II ordine	Formula
Contrasto	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - j)^2 P(i, j)$
Correlazione	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{(i - \mu_i)(j - \mu_j) P(i, j)}{\sigma_i \sigma_j}$
Omogeneità	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i, j)}{1 + (i - j)^2}$
Dissimilarità	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i - j P(i, j)$
Energia	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j)^2$

Tabella 3.4: Descrittori del secondo ordine.

Entrando nel dettaglio, il contrasto misura il grado di fluttuazione locale: maggiore è il valore, maggiore è la variazione di intensità tra i pixel dell'immagine. La correlazione è un indice del parallelismo tra i livelli di grigio dell'immagine, il cui valore è compreso nell'intervallo 0-1; un valore prossimo all'unità indica una maggiore dipendenza lineare dei livelli di grigio dell'immagine. L'omogeneità quantifica l'affinità dei livelli di grigio in co-occorrenza, ovvero la presenza di pixel con intensità simile vicini tra loro. La dissimilarità misura la differenza di intensità tra i pixel della GLCM, considerando anche la distanza tra i pixel. Infine, l'energia fornisce

⁵Gray Level Co-occurrence Matrix

informazioni sulla distribuzione energetica e quindi sull'uniformità dell'immagine, il cui valore è compreso nell'intervallo 0-1; un valore prossimo all'unità esprime una bassa variazione dell'immagine rispetto all'intensità.

3.6.3 Metriche di valutazione

Le metriche impiegate nel presente lavoro per la valutazione della qualità delle immagini istologiche sono sostanzialmente cinque e saranno discusse nei paragrafi sottostanti.

Peak Signal-to-Noise Ratio (PSNR)

Il rapporto segnale-rumore valuta la qualità di un'immagine compressa rispetto all'originale, confrontando i valori dei pixel delle due immagini. È definito come il rapporto tra la massima potenza del segnale e la potenza del rumore di corruzione che influisce sulla fedeltà della rappresentazione. Il segnale è considerato come il dato originale, mentre il rumore è l'errore prodotto dalla compressione o distorsione [43]. Il PSNR è una metrica di riferimento non lineare, ma logaritmica, solitamente espressa in decibel. Per calcolare il PSNR, è necessario calcolare prima l'errore quadratico medio o MSE⁶:

$$MSE = \frac{1}{NM} \sum_0^{M-1} \sum_0^{N-1} \|f(i, j) - g(i, j)\|^2 \quad (3.1)$$

$$PSNR = 20 \log_{10} \left(\frac{MAX_f}{\sqrt{MSE}} \right) \quad (3.2)$$

dove f, g rappresentano l'immagine originale e generata rispettivamente, N, M corrispondono all'altezza e alla larghezza dell'immagine rispettivamente, MAX_f corrisponde al massimo valore possibile assunto dai pixel.

Più basso è l'MSE, minore è l'errore e più alto è il risultato del PSNR. Un punteggio PSNR elevato corrisponde a una migliore ricostruzione dell'immagine degradata rispetto all'immagine di riferimento, il che significa che anche l'algoritmo utilizzato per la ricostruzione è migliore [44]. Il rapporto segnale-rumore è quindi comunemente utilizzato nei compiti di image-to-image traslation poiché permette di quantificare la perdita di qualità dell'immagine sintetica rispetto all'immagine reale di riferimento. Maggiore è la somiglianza dell'immagine generata con quella reale, maggiore è il valore del PSNR.

⁶Mean Square Error

Root Mean Square Error (RMSE)

L'errore quadratico medio è una misura utilizzata frequentemente per valutare le differenze tra il valore predetto da uno stimatore e il valore effettivo [43]. L'RMSE deriva dall'MSE e dunque valuta l'entità dell'errore. L'RMSE assume sempre valori positivi e un valore nullo indica un perfetto adattamento dei dati. Nel nostro caso, un basso valore di RMSE è indice di buona somiglianza tra l'immagine generata e il riferimento. La formula per il calcolo dell'RMSE è la seguente:

$$RMSE = \sqrt{MSE} = \frac{1}{NM} \sum_0^{M-1} \sum_0^{N-1} \|f(i, j) - g(i, j)\|^2 \quad (3.3)$$

Mean Structural Similarity Index (SSIM)

A differenza dei parametri riportati in precedenza, che stimano gli errori assoluti, l'indice di somiglianza strutturale è un modello basato sulla percezione, che confronta la luminanza, il contrasto e la struttura dell'immagine originale e di quella degradata. In particolare, considera il degrado dell'immagine come una variazione dell'informazione strutturale. Il termine informazione strutturale fa riferimento ai pixel fortemente interdipendenti, i quali contengono importanti informazioni sulla struttura degli oggetti. L'SSIM incorpora anche importanti fenomeni percettivi, tra cui il mascheramento della luminanza e il mascheramento del contrasto. Il primo è un termine in cui la distorsione di un'immagine è meno visibile nei bordi dell'immagine. D'altra parte, il secondo è un termine in cui le distorsioni tendono a essere meno visibili nella texture dell'immagine [43].

Dunque, l'SSIM consente di stimare la qualità percepita sia di immagini sia di video; in questo caso, si utilizza per quantificare la somiglianza tra l'immagine originale e l'immagine sintetica. L'SSIM è misurato su una scala da 0 a 1, dove più il punteggio è vicino a 1, più l'immagine generata è simile all'immagine di riferimento. Matematicamente, l'indice di somiglianza strutturale può essere espresso attraverso questi tre termini:

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (3.4)$$

dove l è la luminanza, c è il contrasto e s è la struttura, mentre α , β e γ sono delle costanti positive. A loro volta, la luminanza, il contrasto e la struttura possono essere espresse separatamente come segue:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (3.5)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (3.6)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \quad (3.7)$$

dove μ_x e μ_y sono le medie locali, σ_x e σ_y le deviazioni standard e σ_{xy} la cross-covarianza per le immagini x e y sequenzialmente. Nel caso in cui $\alpha = \beta = \gamma = 1$, l'indice può essere semplificato come segue utilizzando le Equazioni 3.5-3.7:

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3.8)$$

Visual Saliency-induced Index (VSI)

La VSI si basa sul presupposto che la mappa VS⁷ di un'immagine sia correlata con la sua qualità percettiva. Infatti, i valori VS variano con la qualità dell'immagine e sono influenzati dalla presenza di eventuali distorsioni sovrasoglia. I modelli VS si basano fondamentalmente sulle feature di basso livello dell'immagine. Tuttavia, le mappe VS non sono in grado di caratterizzare un'eventuale perdita di contrasto tra due immagini. Dunque, per compensare questa mancanza di sensibilità al contrasto, si utilizza una feature aggiuntiva: la GM⁸. Esistono diversi operatori per calcolare il gradiente di un'immagine che possono essere espressi mediante maschere di convoluzione; i tre più utilizzati sono l'operatore di Sobel, di Prewitt e di Scharr. La VS e la GM sono quindi complementari e riflettono aspetti differenti del sistema visivo umano nella valutazione della qualità locale di un'immagine. Inoltre, anche le variazioni di saturazione del colore non ben sono caratterizzate, perciò si utilizza la crominanza. La VS ha un duplice ruolo nel calcolo della VSI: da un lato viene utilizzata come feature nel calcolo della mappa di similarità locale, dall'altro viene impiegata come funzione di ponderazione per riflettere l'importanza delle regioni locali e ottenere un singolo punteggio di qualità. Gli autori hanno constatato che la VSI sia in grado di fornire risultati statisticamente significativi migliori in termini di accuratezza della predizione rispetto ai metodi proposti in precedenza, pur mantenendo una complessità computazionale moderata [45].

Il calcolo della VSI tra l'immagine reale e l'immagine generata è scomposto in tre componenti:

$$S_{VS} = \frac{2VS_1VS_2 + C_1}{VS_1^2 + VS_2^2 + C_1} \quad (3.9)$$

$$S_G = \frac{2G_1G_2 + C_2}{G_1^2 + G_2^2 + C_2} \quad (3.10)$$

$$S_C = \frac{2M_1M_2 + C_3}{M_1^2 + M_2^2 + C_3} \cdot \frac{2N_1N_2 + C_3}{N_1^2 + N_2^2 + C_3} \quad (3.11)$$

⁷Visual Saliency

⁸Gradient Magnitude

dove VS_1 e VS_2 rappresentano le mappe VS estratte dall'immagine reale e generata, G_1 e G_2 sono i valori di gradiente dell'immagine reale e generata, M_1 , N_1 , M_2 e N_2 sono i due canali di crominanza dell'immagine reale e generata rispettivamente, mentre C_1 , C_2 e C_3 sono tre costanti positive.

Le Equazioni 3.9-3.11 sono unite per ottenere l'indice di similarità locale:

$$S(x) = [S_{VS}(x)] \cdot [S_G(x)]^\alpha \cdot [S_C(x)]^\beta \quad (3.12)$$

dove α e β sono due parametri utilizzati per regolare l'importanza relativa delle tre feature. Per semplicità si imposta $\alpha = \beta = 1$.

Infine, è possibile ottenere la metrica VSI tra le due immagini, il cui valore è compreso tra 0 e 1:

$$VSI = \frac{\sum_{x \in \Omega} S(x) \cdot VS_m(X)}{\sum_{x \in \Omega} VS_m(x)} \quad (3.13)$$

dove Ω fa riferimento all'intero dominio spaziale e $VS_m(x) = \max(VS_1(x), VS_2(x))$.

Feature Similarity Index Matrix (FSIM)

L'indice di similarità delle caratteristiche mappa le feature e misura la somiglianza tra due immagini. La metrica si basa sul fatto che il sistema visivo umano comprende un'immagine principalmente in base alle sue feature di basso livello. In particolare, è emerso che le caratteristiche visibilmente distinguibili coincidono con i punti a elevata congruenza di fase. Dunque si impiega la PC⁹ come feature primaria per valutare l'importanza di una struttura locale ed estrarre caratteristiche altamente informative. La PC è invariante rispetto al contrasto, sebbene l'informazione relativa al contrasto influisca sulla percezione della qualità di un'immagine. Perciò, anche in questo caso, si utilizza la GM dell'immagine come feature secondaria. Perciò, analogamente a quanto detto in precedenza, la PC e la GM sono complementari e riflettono aspetti differenti del sistema visivo umano nella valutazione della qualità locale di un'immagine. Dopo aver calcolato la mappa di somiglianza locale, la PC viene nuovamente usata come funzione di ponderazione per ricavare un singolo punteggio di similarità [46]. Il calcolo della FSIM tra l'immagine reale e l'immagine simulata è suddiviso in due componenti, una basata su PC (3.14) e una su GM (3.15):

$$S_{PC} = \frac{2PC_1PC_2 + T_1}{PC_1^2 + PC_2^2 + T_1} \quad (3.14)$$

$$S_G = \frac{2G_1G_2 + T_2}{G_1^2 + G_2^2 + T_2} \quad (3.15)$$

dove PC_1 e PC_2 rappresentano i valori di congruenza di fase dell'immagine reale e generata, G_1 e G_2 sono i valori di gradiente dell'immagine reale e generata, T_1 e

⁹Phase Congruency

T_2 sono due costanti positive che dipendono rispettivamente dal range dinamico dei valori di PC e GM.

Le Equazioni (3.14) e (3.15) sono combinate per calcolare l'indice di similarità in ogni posizione x :

$$S_L(x) = [S_{PC}(x)]^\alpha \cdot [S_G(x)]^\beta \quad (3.16)$$

dove α e β sono due parametri utilizzati per regolare l'importanza relativa di PC e GM. Per semplicità si imposta $\alpha = \beta = 1$.

Successivamente è possibile ottenere la FSIM tra le due immagini, il cui valore è compreso tra 0 e 1:

$$FSIM = \frac{\sum_{x \in \Omega} S_L(x) \cdot PC_m(X)}{\sum_{x \in \Omega} PC_m(x)} \quad (3.17)$$

dove Ω fa riferimento all'intero dominio spaziale e $PC_m(x) = \max(PC_1(x), PC_2(x))$

Capitolo 4

Risultati

All'interno del presente capitolo verranno illustrati i risultati ottenuti dai modelli generativi implementati.

4.1 Task 1: Simulazione maschere

Sebbene la fase di addestramento della StyleGAN2 sia stata difficoltosa, in fase di inference si è riusciti a simulare 5000 maschere di segmentazione per ciascuno dei quattro gradi patologici. In questo modo è stato possibile valutare i campioni sintetizzati e metterli a confronto con quelli reali, sia dal punto di vista morfologico sia mediante l'analisi della tessitura. La comparazione dell'informazione semantica è avvenuta sostanzialmente in due modi: intra-classe, paragonando i parametri relativi alle maschere reali e generate del medesimo grado patologico, e inter-classe, verificando eventuali differenze al variare della malignità. Prima di commentare i risultati ottenuti, si mostrano alcuni campioni simulati dal modello generativo e successivamente categorizzati (Figura 4.1).

In prima battuta, esaminando i parametri morfologici fondamentali, quali area, perimetro, eccentricità e numerosità dei nuclei e dei bianchi, si riscontra una discreta corrispondenza tra le maschere reali e le maschere generate per ciascuna classe. A supporto di quanto affermato, le Figure 4.2 e 4.3 illustrano i valori medi e i relativi errori standard dei quattro parametri morfologici calcolati sui nuclei e bianchi rispettivamente. In particolare, per i bianchi, si sceglie di riportare il grafico relativo all'area totale dell'immagine occupata dai pixel etichettati come "Bianchi" poiché di più facile interpretazione rispetto a quello dell'area.

Entrando più nel dettaglio, si osserva che gli andamenti dei valori medi dei parametri morfologici ottenuti dalle maschere simulate sono analoghi a quelli delle maschere originali al variare della classe, salvo qualche eccezione.

A cominciare dai nuclei, l'area stimata dalle maschere generate appartenenti alle

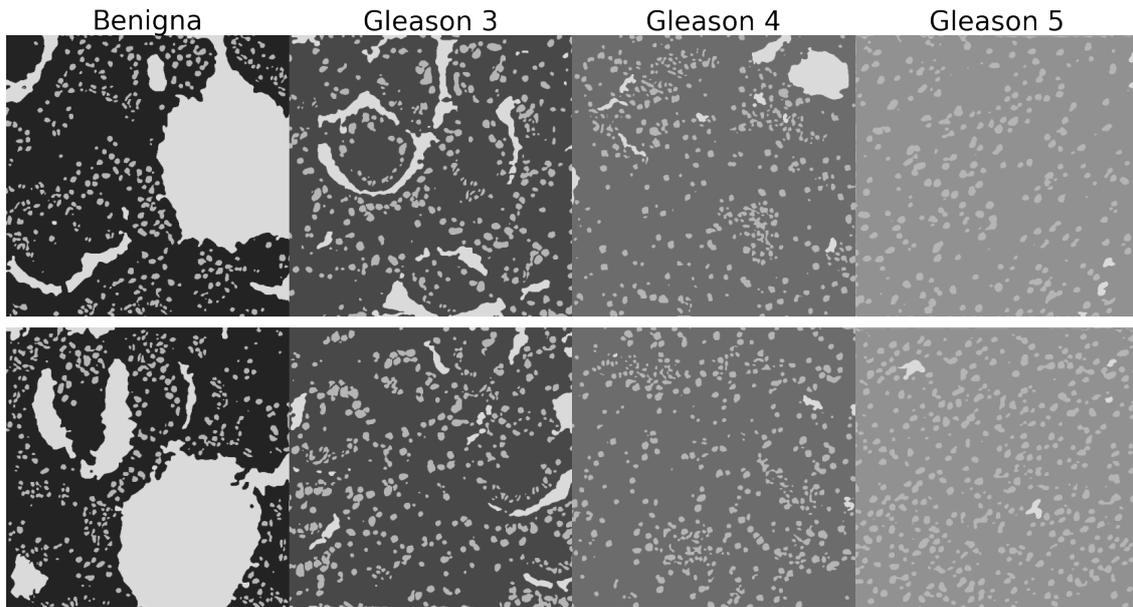


Figura 4.1: Esempi di maschere di segmentazione simulate mediante StyleGAN2. Da sinistra verso destra si riportano campioni di classe Benigna, Gleason 3, Gleason 4 e Gleason 5.

classi Benigna e Gleason 4 è coerente con il rispettivo valore medio riscontrato dalle maschere reali, i cui errori percentuali risultano essere circa dell'1,47% e 5,53% rispettivamente. Non si verifica lo stesso invece per le classi Gleason 3 e Gleason 5, in cui i valori medi di area sono superiori e di conseguenza anche gli errori percentuali (16,2% e 25,4% circa rispettivamente). Per il perimetro si riscontrano dei valori medi paragonabili tra le distribuzioni reale e simulata, a esclusione della classe Gleason 5, per la quale si osserva un valore medio sovrastimato (errore percentuale del 13,8% circa). Al contrario, l'eccentricità dei nuclei subisce una sottostima generale, ciò nonostante l'andamento dei valori medi al variare del grado di malignità rispecchia quello di riferimento. L'errore percentuale è inferiore allo 0,50% per la classe Benigna, si aggira intorno al 5,60% e 3,94% per le classi Gleason 3 e Gleason 4 rispettivamente, fino ad aumentare al 13,8% per la classe Gleason 5. Anche il parametro di numerosità è sottostimato dal modello, soprattutto per la classe Gleason 5, caratterizzata da un errore percentuale del 17,7% circa. Inoltre ci si aspetterebbe che il numero di nuclei, seppur sottostimato, segua un trend crescente all'aumentare del grado della patologia. Dal momento che si riscontra una generale sovrastima dell'area e sottostima della numerosità, si può constatare che la StyleGAN2 tende a fondere i nuclei vicini in un unico oggetto.

In merito alla morfologia dei bianchi, i grafici mostrano che gli andamenti dei valori

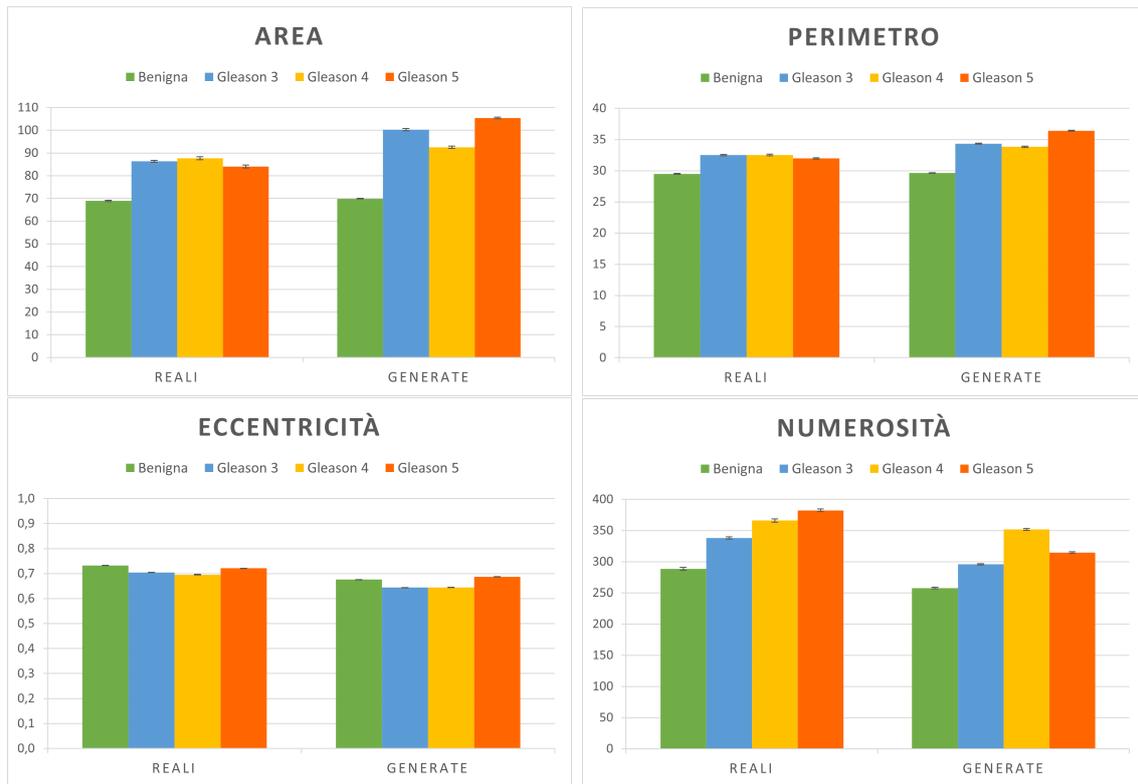


Figura 4.2: Parametri morfologici dei nuclei. I grafici mostrano i valori medi e i relativi errori standard di area, perimetro, eccentricità e numerosità dei nuclei nelle maschere reali e generate.

medi ottenuti dai campioni simulati in funzione della classe aderiscono a quelli di riferimento, a eccezione del perimetro. Quest'ultimo assume valori coerenti a quelli reali per le classi Benigna e Gleason 3, i cui errori percentuali si aggirano intorno allo 0,61% e 2,69% rispettivamente. Diversamente accade per i gradi più severi, in cui non si verifica un decremento del valore medio, ma un allineamento alla classe Gleason 3. Questo comporta un errore percentuale significativo per la classe Gleason 5 del 76,8% circa. Invece, il trend descritto dai parametri di area totale, eccentricità, numerosità dei bianchi delle maschere simulate si adatta maggiormente a quello di riferimento. In particolare, l'area totale media associata a ciascuna classe subisce una sovrastima; infatti si ottengono i seguenti errori percentuali: 31,6% per la classe Benigna, 9,65% per la classe Gleason 3, 29,3% per la classe Gleason 4 e 72,4% per la classe Gleason 5. L'eccentricità viene leggermente sovrastimata per la classe Gleason 5 (0,45%), mentre per i restanti gradi patologici si osserva una sottostima, inferiore al 3%. Esaminando le maschere reali e generate appartenenti alla classe

4.1. TASK 1: SIMULAZIONE MASCHERE

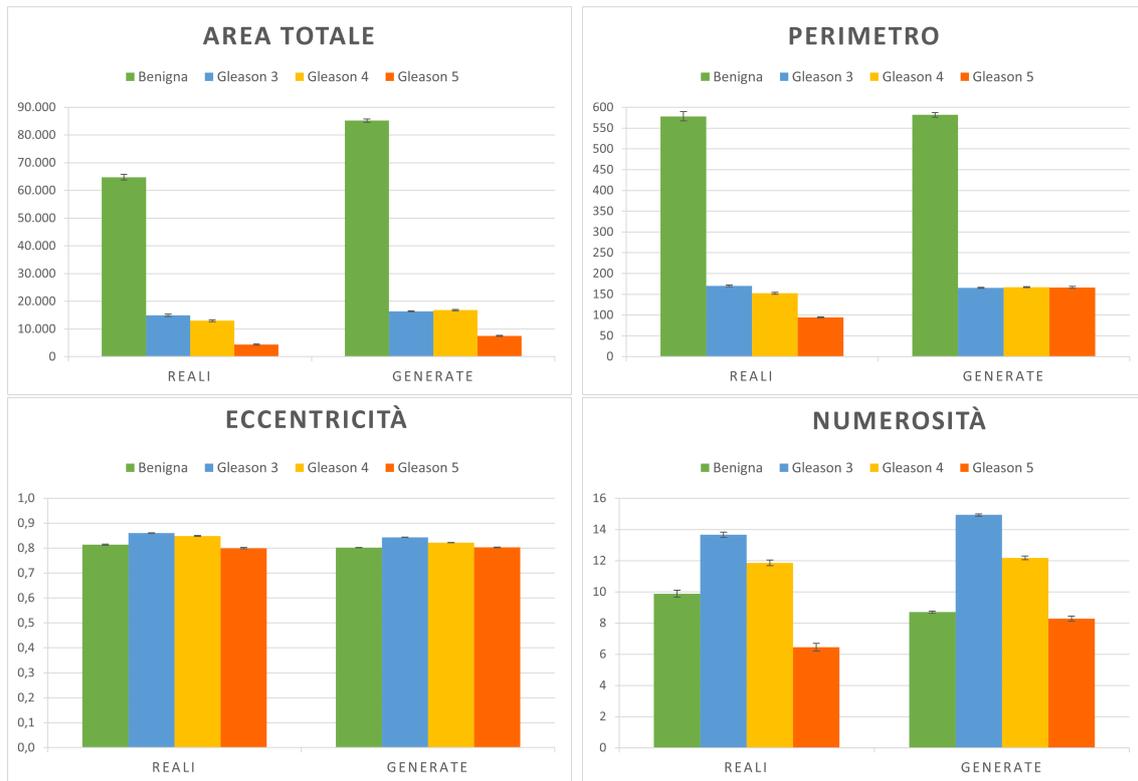


Figura 4.3: Parametri morfologici dei bianchi. I grafici mostrano i valori medi e i relativi errori standard di area totale, perimetro, eccentricità e numerosità dei bianchi nelle maschere reali e generate.

Gleason 5 e i rispettivi parametri morfologici dei bianchi, si può affermare che la StyleGAN2 tende a simulare maschere con un contenuto di bianchi superiore, quando il 45% dei campioni di addestramento non presenta alcuna zona bianca. Inoltre, si riscontra una buona capacità di generazione dei bianchi da parte del modello per le altre classi, nelle quali la dimensione e l'estensione di tali zone risulta essere un fattore fondamentale per la garantire la corretta distinzione della lesione tumorale. È possibile concludere che complessivamente i grafici relativi ai parametri morfologici ricavati dalle maschere simulate rispecchiano l'andamento di quelli reali, a eccezione della classe Gleason 5, per la quale il modello incontra delle difficoltà. Esaminando i descrittori di tessitura è possibile valutare sia il valore assoluto dei singoli pixel sia la loro distribuzione all'interno della maschera di segmentazione. Anche in questo caso si mettono a confronto i campioni di addestramento con quelli simulati, suddividendoli per classe di appartenenza. Si riportano i valori medi delle feature di tessitura del I e II ordine rispettivamente all'interno delle Figure 4.4 e

4.5. In generale, dai grafici riportati si osserva che l'andamento dei valori medi di ciascun descrittore ottenuto dalle maschere simulate rispecchia il trend descritto dal rispettivo riferimento all'aumentare del grado patologico.

Per quanto concerne i descrittori del primo ordine, l'entropia riveste particolare importanza nella caratterizzazione dell'informazione semantica di un'immagine istologica, poiché può essere utilizzata per descrivere la complessità della texture o per identificare la presenza di zone di tessuto con pattern particolari. Dai risultati ottenuti, si osserva che il valore medio di entropia diminuisce all'aumentare del grado della patologia. Si ricorda che un valore di entropia maggiore è sinonimo di una distribuzione di intensità dei pixel disomogenea e disordinata, mentre un valore minore è indice di una distribuzione uniforme e ordinata. Dunque, si può affermare che le maschere di segmentazione di classe Benigna presentano un contenuto semantico più complesso ed eterogeneo, caratterizzato dalla presenza di nuclei e bianchi. Al contrario, le maschere di tessuto di grado più elevato, come quelle di classe Gleason 5, possiedono un valore di entropia minore dal momento che costituite quasi esclusivamente da nuclei e quindi da pixel di ugual intensità. Gli errori percentuali calcolati tra i valori medi di riferimento e quelli relativi ai campioni simulati sono i seguenti: 10,3% per la classe Benigna, 4,55% per la classe Gleason 3, 8,87% per la classe Gleason 4 e 12,2% per la classe Gleason 5. I restanti descrittori statistici del primo ordine restituiscono informazioni più comprensibili. I grafici relativi alla media e alla varianza mostrano un andamento analogo a quello dell'entropia. Come atteso, minore è l'aggressività del tumore e maggiore è il valore medio calcolato di intensità media dei pixel e varianza. Invece, la skewness relativa alla classe Benigna appare più che dimezzata rispetto ai valori medi ottenuti dalle altre classi, poiché i tessuti tumorali di grado maggiore tendono a essere più uniformi nella distribuzione di intensità dei pixel. La curtosi descrive una tendenza simile alla skewness, in cui il valore medio calcolato dalla classe Benigna è significativamente inferiore. Questo denota la presenza di zone disomogenee di tessuto con valori di intensità dei pixel differente. Sia per la skewness sia per la curtosi, l'errore percentuale calcolato dalla classe Benigna risulta essere pari al 42% circa, mentre per le classi Gleason non supera il 13%.

Tra i descrittori del secondo ordine è essenziale valutare l'energia e la correlazione. I grafici raffigurano una corrispondenza inversamente proporzionale dei due parametri al variare del grado patologico. Nello specifico, si osserva un incremento del valore medio di energia e un decremento del valore medio dell'indice di correlazione all'aumentare del Gleason Score. Il trend crescente associato all'energia è spiegato dal fatto che generalmente i tessuti neoplastici presentano una maggiore eterogeneità in termini di texture e struttura. Mentre, un valore medio minore del parametro di correlazione per la classe Gleason 5 è dovuto a una maggiore variazione di intensità dei pixel tra le diverse regioni della maschera e a una minore coerenza spaziale dei

4.1. TASK 1: SIMULAZIONE MASCHERE



Figura 4.4: Descrittori di tessitura del I ordine. I grafici mostrano i valori medi e i relativi errori standard di media, varianza, skewness, curtosi ed entropia delle maschere reali e generate per ciascuna classe.



Figura 4.5: Descrittori di tessitura del II ordine. I grafici mostrano i valori medi e i relativi errori standard di contrasto, dissimilarità, omogeneità, energia e correlazione delle maschere reali e generate per ciascuna classe.

valori assunti dai pixel. Sia per l'energia sia per la correlazione, gli errori percentuali ricavati da ciascuna classe sono inferiori al 5%. Inoltre, è risaputo che il contrasto misura la differenza media di intensità tra pixel adiacenti della maschera. Dunque, come ci si aspettava, il valore medio di contrasto diminuisce all'aumentare del Gleason Score. Anche in questo caso gli errori percentuali sono estremamente bassi e non superano il 4,4%. Similmente al contrasto, la dissimilarità considera anche la distanza tra i pixel. Non si raggiunge un valore medio maggiore in corrispondenza delle maschere di classe Benigna, bensì simile alle classi Gleason; quanto ottenuto è indice del fatto che la variazione di intensità dei pixel tra le diverse regioni delle maschere è molto simile per i differenti gradi tumorali. Analogamente a quanto detto in precedenza, l'errore percentuale ottenuto da ciascuna classe è inferiore al 6%. Infine, l'omogeneità del tessuto, quantifica la coerenza spaziale dei valori dei pixel nell'immagine. Tuttavia, operando con maschere di segmentazione piuttosto che con immagini istologiche, non si osserva una variazione dei valori medi al variare del grado patologico.

4.2 Task 2: Simulazione semantica

Dal punto di vista puramente qualitativo, paragonando le immagini sintetiche con quelle originali, si riscontra una buona coerenza semantica. Si ricorda che lo scopo del presente lavoro è proprio quello di simulare l'architettura del tessuto prostatico per diversi gradi della patologia, ponendo particolare attenzione al contenuto semantico più che alla colorazione delle differenti componenti strutturali. Detto ciò, è possibile constatare che sia il modello generativo SPADE sia CLADE riescono a localizzare correttamente i bianchi e i nuclei del tessuto, restituendo immagini istologiche con sembianze realistiche. D'altra parte, entrambi gli algoritmi hanno difficoltà nella generazione dei globuli rossi e di alcune porzioni di stroma.

In alcune immagini simulate mediante SPADE si riscontra la presenza di un artefatto, che si manifesta in forma di pattern geometrico ripetitivo sulla superficie dell'intero campione. Tuttavia, SPADE sembrerebbe maggiormente in grado di esaltare il contrasto delle diverse strutture tissutali, focalizzandosi sulla loro colorazione; CLADE invece tende a omogeneizzare le differenti componenti stromali, in particolare nelle immagini appartenenti alla classe Gleason 5. Si riportano degli esempi illustrativi all'interno della Figura 4.6, in cui si confronta ciascuna immagine reale e la relativa mappa di segmentazione semantica con le immagini simulate dai due modelli implementati.

Come descritto all'interno del Capitolo 3.6, per quantificare la qualità delle immagini istologiche sintetiche rispetto ai campioni reali, sono state utilizzate cinque metriche di valutazione differenti: PSNR, RMSE, SSIM, VSI e FSIM. Nella Figura 4.7 sono riportati i valori stimati dei parametri sia sul Training set sia sul Test Set.

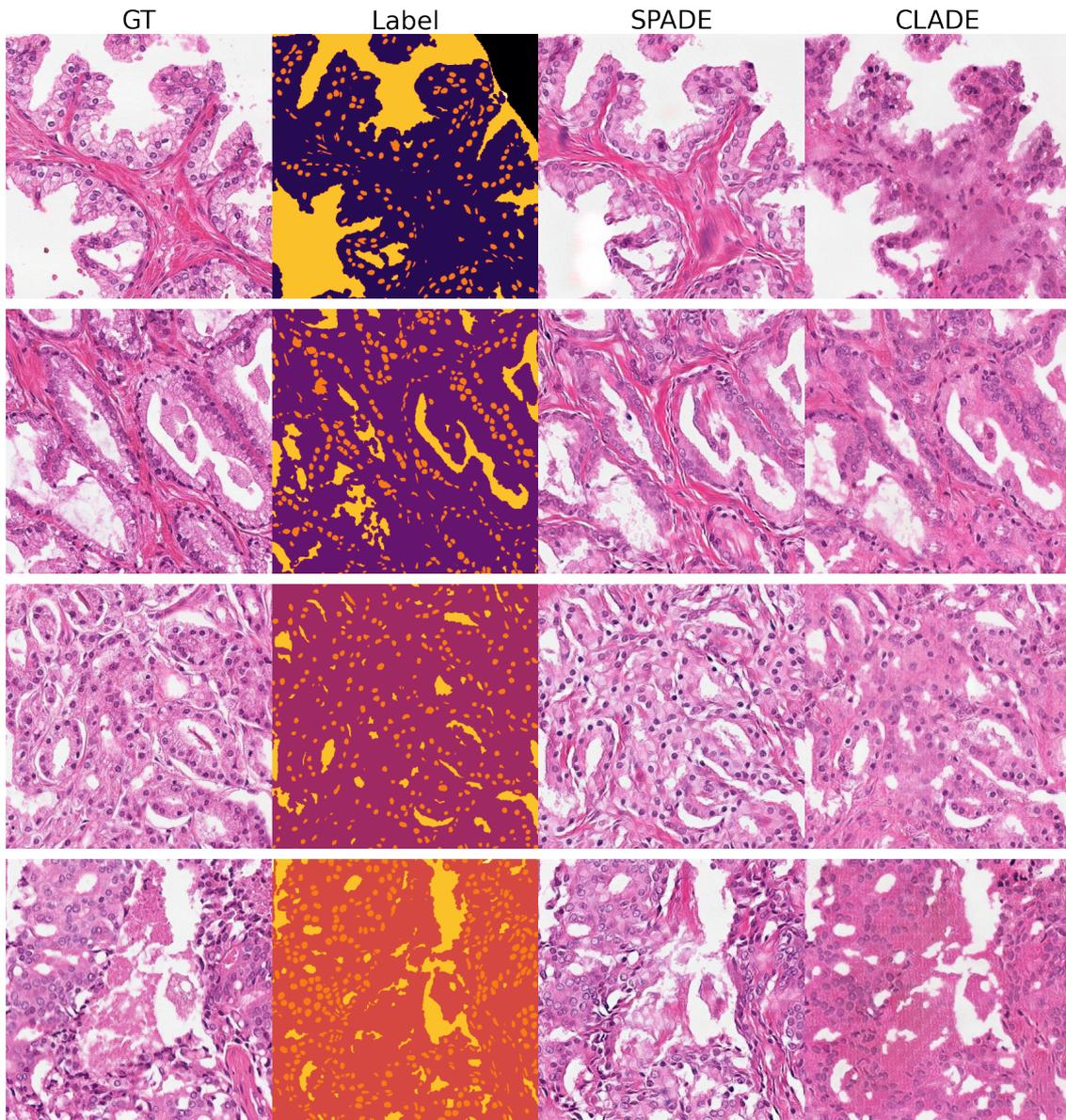


Figura 4.6: Esempi di immagini istologiche simulate da SPADE e CLADE. Si riporta un esempio per ciascun grado della patologia, dall'alto verso il basso: Benigna, Gleason 3, Gleason 4 e Gleason 5. Da sinistra verso destra sono mostrate l'immagine reale, la mappa di segmentazione e le immagini generate dagli algoritmi SPADE e CLADE rispettivamente.

Dai grafici mostrati si può affermare che i valori delle metriche calcolate sulle intere immagini siano in linea con quanto ci aspettassimo, vista l'eterogeneità della patologia prostatica. In particolare, la differenza percentuale tra i valori ottenuti sul Training set e sul Test set risulta essere sempre inferiore all'1%, indice del fatto che entrambi i modelli siano in grado di adattarsi al tipo di dato fornito. Anche la differenza percentuale tra i due modelli generativi implementati risulta essere inferiore al 5% per ciascuna metrica, a eccezione dell'RMSE per cui si aggira intorno al 9,5%.

Analizzando dettagliatamente i risultati si osserva un valore di PSNR medio complessivo di circa 16,1 e 16,9 dB per SPADE e CLADE rispettivamente, arrivando a raggiungere anche valori prossimi a 18 dB. Per quanto riguarda l'indice di similarità si ottengono valori medi che si aggirano intorno a 0,245 per SPADE e 0,251 per CLADE, giungendo a picchi più elevati in immagini ad alto contenuto di zone bianche. Invece, i valori medi di FSIM e VSI risultano essere nettamente superiori a quelli di SSIM: il primo è circa pari allo 0,734 e allo 0,753, mentre il secondo eguaglia lo 0,892 e lo 0,904 per SPADE e CLADE rispettivamente. Questi risultati possono essere dovuti al fatto che i due indici tentano di correlare maggiormente la qualità dell'immagine al sistema visivo umano, ponendo l'accento sulle feature a basso livello dell'immagine.

Inoltre, sono stati calcolati gli stessi indici suddividendo le immagini istologiche nelle varie classi d'appartenenza. Così facendo, è stato possibile confrontare i valori medi di ciascuna metrica per differenti gradi della patologia, in modo tale da verificare se i modelli fossero più facilitati nella simulazione di alcune classi rispetto ad altre. Come ci si aspettava, la classe Benigna ha raggiunto risultati significativamente superiori, probabilmente a causa della morfologia e dell'architettura del tessuto, il quale include molteplici zone bianche. Nello specifico, si ottengono valori di PSNR di 17,5 dB e 18,0 dB, SSIM di 0,390 e 0,397, RMSE di 34,2 e 32,3, FSIM di 0,759 e 0,775 e VSI di 0,915 e 0,923 per SPADE e CLADE rispettivamente. Al contrario, per le classi Gleason 3 e Gleason 4 i valori ottenuti risultano essere in linea coi valori medi complessivi. La classe Gleason 5 si comporta diversamente, poiché associata a indici peggiori rispetto alle medie globali. In particolare, si osservano valori di SSIM di 0,199 e 0,207, FSIM di 0,713 e 0,693 e VSI di 0,877 e 0,872 per SPADE e CLADE rispettivamente. Questa differenza potrebbe essere associata anche alla complessità e scarsità dei campioni di grado Gleason 5 all'interno del dataset. Nella Figura 4.8 si riportano i valori medi e le rispettive deviazioni standard delle metriche di valutazione per ciascuna classe. Esaminando i grafici, si nota un allineamento delle metriche tra i due modelli generativi, i cui andamenti seguono un trend decrescente all'aumentare del grado della patologia a eccezione dell'indice RMSE che segue un trend crescente.



Figura 4.7: Metriche di valutazione della qualità delle immagini generate da CLADE e SPADE. Si riporta la media e la rispettiva deviazione standard sia sul Training set sia sul Test set.

4.2. TASK 2: SIMULAZIONE SEMANTICA



Figura 4.8: Metriche di valutazione calcolate su ciascuna classe: Benigna, Gleason 3, Gleason 4 e Gleason 5. Si riporta la media e la rispettiva deviazione standard sul Test set ottenuta dai modelli CLADE e SPADE.

Capitolo 5

Conclusioni

Il lavoro di tesi sviluppato tenta di affrontare uno dei problemi fondamentali ancora aperti inerente all'imaging medico: la difficoltà di eseguire una diagnosi precoce accurata e stimare il corretto grado tumorale di una determinata patologia. Per tentare di alleviare tali problematiche si è pensato di creare un tool di intelligenza artificiale in grado di fornire dei campioni sintetici che possano supportare l'anatomopatologo nel processo diagnostico. Ripercorrendo quanto descritto in precedenza, sono stati implementati due modelli generativi basati su apprendimento profondo per la simulazione di immagini istopatologiche di tessuto prostatico a partire da un ground truth. Il primo utilizza una rete GAN a condizione di classe, mentre il secondo effettua una sintesi semantica delle immagini.

A seguito dei risultati ottenuti, si possono effettuare le seguenti considerazioni. La StyleGAN2 si è rivelata un attuabile sistema per la simulazione di immagini, che consente di generare una vasta gamma di campioni annotati. Durante la fase di training del modello si sono riscontrate delle difficoltà nel tuning degli iperparametri, affinché si conseguisse un allenamento stabile che non sfociasse nella problematica del collasso della modalità. Tuttavia, in fase di inference si è riusciti a garantire una migliore variabilità dei campioni. Buona parte delle maschere di segmentazione simulate possiede parametri di tessitura che rientrano all'interno della distribuzione dei valori di riferimento, mentre si ottengono risultati meno confortanti dal punto di vista morfologico. Nonostante ciò, si riscontra una buona coerenza tra i valori medi dei parametri ottenuti dalle maschere reali e generate al variare del grado patologico, salvo qualche eccezione per la classe Gleason 5. Dal momento che all'interno del presente lavoro la StyleGAN2 è stata addestrata su maschere di segmentazione di tessuto prostatico, è ragionevole ristrutturare il modello su un dataset costituito da immagini istologiche. Questo permetterebbe di valutare la capacità del sistema di simulare zone tissutali differenti ma appartenenti alla stessa label semantica ed esaminare i descrittori di tessitura in maniera più coerente.

Per quanto riguarda la generazione semantica, è possibile affermare che i modelli

generativi SPADE e CLADE consentono di raggiungere risultati simili in termini di metriche di valutazione, nonostante le immagini istologiche simulate siano differenti sia dal punto di vista dell'informazione semantica sia in merito alla colorazione. SPADE riesce a simulare in maniera più realistica le differenti strutture tissutali, soprattutto per la classe Gleason 5, e restituisce immagini a elevato contrasto. Ciò nonostante, parte dei campioni presenta un pattern geometrico ripetitivo sovrapposto che interferisce con la qualità dell'immagine. Questo aspetto dimostra che valutare la qualità di un'immagine in termini quantitativi è una procedura particolarmente difficoltosa, che non costituisce un mezzo alternativo alla percezione visiva del sistema umano. Nel complesso, CLADE permette di raggiungere una qualità di generazione paragonabile a quella di SPADE con una migliore efficienza, coinvolgendo un minor numero di parametri aggiuntivi e comportando un costo computazionale inferiore.

In entrambi i task la classe Gleason 5 risulta più difficile da simulare. Questo potrebbe essere correlato con la sua eterogeneità ma soprattutto con il fatto che è sotto rappresentata. Dunque, è fondamentale disporre di un dataset di partenza numeroso e bilanciato, in cui anche le classi patologiche più rare siano ben rappresentate.

A questo punto, possiamo sostenere che il presente lavoro di tesi costituisce un buon punto di partenza per riuscire a fornire all'anatomopatologo un sistema di supporto decisionale che lo affianchi nel riconoscimento e nella classificazione della neoplasia. L'idea è proprio quella di permettere al medico effettuare una diagnosi per confronto che sia il più oggettiva possibile, predisponendo di una vasta gamma di campioni di riferimento di classe nota. Questo consentirebbe di velocizzare l'intero iter diagnostico, limitare i problemi relativi alla variabilità inter- e intra-operatore nell'assegnazione del Gleason Score e di conseguenza ridurre la mole di lavoro nei centri di anatomia patologica. In aggiunta, i modelli implementati costituiscono dei possibili metodi di data augmentation per altri algoritmi di segmentazione, che vanno a integrare il dataset di partenza con i campioni sintetici. In questo modo, si dispone di dati già annotati in grado di ridurre la disomogeneità dei campioni appartenenti a classi differenti e apprendere le diverse fonti di variazione della patologia.

Per quanto riguarda i possibili sviluppi futuri, è certamente possibile estendere il presente lavoro a immagini istologiche derivanti da altri tessuti od organi umani, ma anche adattarlo a ulteriori modalità di imaging. Inoltre, sarebbe interessante “dosare” il grado della patologia a partire dallo stesso contenuto semantico per simulare la degradazione progressiva del tessuto e fare studi sul comportamento e l'evoluzione del tumore.

Bibliografia

- [1] M-C Franz, Pascale Anderle, Marc Bürzle, Y Suzuki, MR Freeman, MA Hedges, and Gergely Kovacs. Zinc transporters in prostate cancer. *Molecular aspects of medicine*, 34(2):735–741, 2013.
- [2] Fatma Nur Soylu, Scott Eggener, and Aytekin Oto. Local staging of prostate cancer with MRI. *Diagnostic and Interventional Radiology*, 18(4):365–73, 2012.
- [3] CJ Harvey, J Pilcher, Jonathan Richenberg, Uday Patel, and Ferdinand Frauscher. Applications of transrectal ultrasound in prostate cancer. *The British journal of radiology*, 85:S3–S17, 07 2012.
- [4] Gladell P Paner, Jatin Gandhi, Bonnie Choy, and Mahul B Amin. Essential updates in grading, morphotyping, reporting, and staging of prostate carcinoma for general surgical pathologists. *Archives of pathology & laboratory medicine*, 143(5):550–564, 2019.
- [5] Reti neurali generative (gan). <https://humanativaspa.it/reti-neurali-generative-gan/>, 2018.
- [6] Salome Kazemina, Christoph Baur, Arjan Kuijper, Bram van Ginneken, Nassir Navab, Shadi Albarqouni, and Anirban Mukhopadhyay. Gans for medical image analysis. *Artificial Intelligence in Medicine*, 109:101938, 2020.
- [7] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pages 5967–5976, 2017.
- [8] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8798–8807, 2018.
- [9] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *2019 IEEE/CVF*

- Conference on Computer Vision and Pattern Recognition*, pages 2332–2341, 2019.
- [10] Zhentao Tan, Dongdong Chen, Qi Chu, Menglei Chai, Jing Liao, Mingming He, Lu Yuan, Gang Hua, and Nenghai Yu. Semantic image synthesis via efficient class-adaptive normalization. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):4852–4866, 2022.
- [11] Zhentao Tan, Menglei Chai, Dongdong Chen, Jing Liao, Qi Chu, Bin Liu, Gang Hua, and Nenghai Yu. Diverse semantic image synthesis via probability distribution modeling. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7958–7967, 2021.
- [12] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4396–4405, 2019.
- [13] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12104–12114. Curran Associates, Inc., 2020.
- [14] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 852–863. Curran Associates, Inc., 2021.
- [15] AIOM. Linee guida: Carcinoma della prostata. <https://www.aiom.it/linee-guida-aiom-2021-carcinoma-della-prostata/>, 2021.
- [16] Christine H Lee, Oluyemi Akin-Olugbade, and Alexander Kirschenbaum. Overview of prostate anatomy, histology, and pathology. *Endocrinology and metabolism clinics of North America*, 40(3):565–575, 2011.
- [17] Iperplasia prostatica benigna. <https://www.humanitas.it/malattie/iperplasia-prostatica-benigna>.
- [18] David Rakel. Benign prostatic hyperplasia. In *Integrative Medicine (Second Edition)*, pages 637–643. W.B. Saunders, 2007.
- [19] David Yeo and Graeme Steele. Benign prostatic hypertrophy. In *Decision Making in Medicine (Third Edition)*, pages 564–565. Mosby, 2010.

- [20] I numeri del cancro in italia. https://www.aiom.it/wp-content/uploads/2022/12/2022_AIOM_NDC-web.pdf, 2022.
- [21] Claire H Pernar, Ericka M Ebot, Kathryn M Wilson, and Lorelei A Mucci. The epidemiology of prostate cancer. *Cold Spring Harbor perspectives in medicine*, 8(12):a030361, 2018.
- [22] Michael F Leitzmann and Sabine Rohrmann. Risk factors for the onset of prostatic cancer: age, location, and behavioral correlates. *Clinical epidemiology*, 4(1):1–11, 2012.
- [23] Tumore della prostata. <https://www.humanitas.it/malattie/tumore-della-prostata/>.
- [24] CJ Das, A Razik, S Sharma, and S Verma. Prostate biopsy: when and how to perform. *Clinical Radiology*, 74(11):853–864, 2019.
- [25] Shikha Sharma, Julia Zapatero-Rodriguez, and Richard O’Kennedy. Prostate cancer diagnostics: Clinical challenges and the ongoing need for disruptive and effective diagnostic tools. *Biotechnology Advances*, 35(2):135–149, 2017.
- [26] Stephen W Leslie, Taylor L Soon-Sutton, Anu R I, Hussain Sajjad, and Larry E Siref. Prostate cancer. <https://www.ncbi.nlm.nih.gov/books/NBK470550/>, 2022.
- [27] Jonas Hugosson, Monique J Roobol, Marianne Månsson, Teuvo LJ Tammela, Marco Zappa, Vera Nelen, Maciej Kwiatkowski, Marcos Lujan, Sigrid V Carlsson, Kirsi M Talala, et al. A 16-yr follow-up of the european randomized study of screening for prostate cancer. *European Urology*, 76(1):43–51, 2019.
- [28] Jean-Luc Descotes. Diagnosis of prostate cancer. *Asian Journal of Urology*, 6(2):129–136, 2019. Prostate Cancer : On the Road of Progress.
- [29] Mark S Litwin and Hung-Jui Tan. The diagnosis and treatment of prostate cancer: A review. *JAMA*, 317(24):2532–2542, 06 2017.
- [30] Nicolai Hübner, Shahrokh Shariat, and Mesut Remzi. Prostate biopsy: guidelines and evidence. *Current Opinion in Urology*, 28(4):354–359, 07 2018.
- [31] Massimo Salvi, U Rajendra Acharya, Filippo Molinari, and Kristen M Meiburger. The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. *Computers in Biology and Medicine*, 128:104–129, 2021.

- [32] Cristina Magi-Galluzzi. Prostate cancer: diagnostic criteria and role of immunohistochemistry. *Modern Pathology*, 31:12–21, 2018.
- [33] Peter A. Humphrey. Histopathology of prostate cancer. *Cold Spring Harbor perspectives in medicine*, 7:10, 2017.
- [34] Jonathan I Epstein, Michael J Zelefsky, Daniel D Sjoberg, Joel B Nelson, Lars Egevad, Cristina Magi-Galluzzi, Andrew J Vickers, Anil V Parwani, Victor E Reuter, Samson W Fine, et al. A contemporary prostate cancer grading system: A validated alternative to the gleason score. *European Urology*, 69(3):428–435, 2016.
- [35] Yi-hong Zhang, Jing Zhang, Yang Song, Chaomin Shen, and Guang Yang. Gleason score prediction using deep learning in tissue microarray image. *arXiv preprint arXiv:2005.04886*, 2020.
- [36] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.
- [37] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, 58:101552, 2019.
- [38] Yves-Rémi Van Eycke, Adrien Foucart, and Christine Decaestecker. Strategies to reduce the expert supervision required for deep learning-based segmentation of histopathological images. *Frontiers in medicine*, 6:222, 2019.
- [39] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8107–8116, 2020.
- [40] Guy Nir, Davood Karimi, S Larry Goldenberg, Ladan Fazli, Brian F Skinnider, Peyman Tavassoli, Dmitry Turbin, Carlos F Villamil, Gang Wang, Darby JS Thompson, et al. Comparison of artificial intelligence techniques to evaluate performance of a classifier for automatic grading of prostate cancer from digitized histopathologic images. *JAMA Network Open*, 2(3):e190442–e190442, 2019.
- [41] Simon K Warfield, Kelly H Zou, and William M Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE transactions on medical imaging*, 23(7):903–921, 2004.

BIBLIOGRAFIA

- [42] Training configurations. <https://github.com/NVlabs/stylegan3/blob/main/docs/configs.md>, 2021.
- [43] Umme Sara, Morium Akter, and Mohammad Shorif Uddin. Image quality assessment through FSIM, SSIM, MSE and PSNR - A comparative study. *Journal of Computer and Communications*, 7(3):8–18, 2019.
- [44] Full-reference quality metrics: VMAF, PSNR and SSIM. <https://audiovideotestlab.com/blog/full-reference-quality-metrics-vmaf-psnr-and-ssim/>, 2022.
- [45] Lin Zhang, Ying Shen, and Hongyu Li. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23(10):4270–4281, 2014.
- [46] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011.

Ringraziamenti

Desidero ringraziare tutte le persone che hanno arricchito il mio percorso universitario, seppur in maniera differente. In primo luogo, ringrazio il Professore Massimo Salvi per avermi dato l'opportunità di realizzare questo progetto di tesi e per l'interesse che ha suscitato in me questa tematica. Ringrazio anche l'Ing. Francesco Marzola per la disponibilità, i consigli ricevuti e il costante supporto durante l'intero lavoro.

Un ringraziamento sincero ai compagni di corso conosciuti in questi anni, con i quali ho passato le mie giornate e condiviso i progetti di laboratorio. In particolare, vorrei ringraziare tre amiche: Chiara per la sua caparbieta, Eleonora per la sua generosità e Rosi per la sua bontà (ma soprattutto per essere indecisa quanto me).

Ringrazio tutti i miei amici che mi sono stati accanto e non mi hanno quasi mai fatta sentire sola. Chi c'è sempre stato e chi ho conosciuto meglio in questi ultimi anni. Chi ha vissuto a Torino e chi nonostante la lontananza ha saputo mantenere viva la nostra amicizia. Chi mi ha fatta sorridere nei momenti bui e chi ha imparato a volermi bene veramente. Tra loro ci tengo a menzionare le mie coinquiline, Chiara e Sofie, senza le quali questo percorso non sarebbe stato lo stesso, e tutto il gruppo Aula studio per aver reso le sessioni di studio meno pesanti.

Il ringraziamento più grande va alla mia famiglia, a cominciare da mamma e papà. Grazie per il continuo sostegno e aiuto, senza di voi niente sarebbe stato possibile. Grazie per l'affetto che mi avete sempre donato e per avermi concesso tutto senza chiedermi niente in cambio. Grazie per i consigli, gli insegnamenti e le passioni trasmesse. Ringrazio mio fratello Lorenzo, per il quale ci sarò sempre; so che i nostri litigi e punzecchiamenti inutili nascondono tanto bene e stima. Infine, ringrazio i miei nonni, Miranda e Romano, per aver creduto in me in qualsiasi circostanza e per avermi amata incondizionatamente. Chi mi conosce sa quanto ho desiderato avervi qui con me oggi, ma purtroppo le cose cambiano e non vanno sempre come vogliamo noi. Ho capito che bisogna imparare a essere grati per ciò che si ha e ricordare i bei momenti passati, senza rimpiangere quello che non si può più avere indietro.