

Short report

L'esperimento consiste nella generazione di immagini a risonanza di una certa modalità (per esempio pesate in PD) corrispondenti alle immagini di un'altra modalità (per esempio pesate in T2) grazie ad un generator che è stato allenato appositamente tramite una GAN. L'esperimento punta a quantificare il contributo di ogni singolo componente della loss function proposta nel paper (1) e, in seconda battuta, a migliorare ulteriormente le performance del generator tramite il residual learning (2). Il punto (1) è eseguito nella traduzione T2 -> PD, mentre il punto (2) nella direzione inversa PD -> T2 per motivi che per ora tralasciamo e che rimandiamo alla relazione finale.

MAE, MSE e PSNR sono le metriche generali scelte per valutare la qualità della ricostruzione.

Preparazione del dataset

Abbiamo utilizzato parte delle MRI di oltre 500 pazienti raccolte nell'IXI brain dataset. Il volume del cervello di ogni soggetto è stato lavorato in modo da ottenere 32 grayscale slices trasversali di dimensione 224x224; è stata effettuata qui una prima semplificazione prendendo 1 slice ogni 3 (cioè 32 invece delle originali 96) per velocizzare tutte le successive fasi di training. I volumi così ottenuti sono stati trattati come matrici e quindi standardizzati.

250 volumi sono stati usati come train set, 35 come validation set per individuare i modelli migliori, 70 come test set per valutare la qualità delle immagini generate.

GAN architecture e training

Per la brevità di questo report vi rimandiamo alla descrizione nel paper. Specifichiamo solo che il generator, ricevendo in input una patch tridimensionale 32x32x32 (per esempio T2 pesata), genera una patch di uguali dimensioni; il discriminator (usato solo nel training) riceve in ingresso la stessa source image data al generator + l'output del generator stesso e calcola la probabilità che quest'ultimo sia reale anziché fittizio (synthetic). Naturalmente il discriminator viene anche allenato con coppie (source image, real target) autentiche, per poi essere in grado di individuare quelle sintetiche. L'architettura della rete è in definitiva una conditional GAN.

Poiché il training è patch-based, i volumi sono stati partizionati in cubi 32x32x32 per un totale di 49 cubi per paziente. Abbiamo adottato una mini batch di 7 cubi. Illustreremo la ricostruzione del volume originale nella relazione completa.

Esperimento 1: valutazione della loss function

La loss function del generator è la somma pesata di *Euclidian loss* (LG), *Adversarial loss* (LADV) e *Image Gradient Difference loss* (GDL), mentre il discriminator utilizza semplicemente la binary cross entropy. Le loro espressioni matematiche e i pesi si trovano nella sezione II. METHODS A del paper.

Euclidian loss (let's call this generator G1)

È la loss function tradizionale L2 che calcola la somma degli scarti quadratici punto per punto tra la ground-truth target image Y ed il volume generato G(X). L'utilizzo esclusivo dell'Euclidian loss non sfrutta la GAN architecture perché il discriminator non è chiamato in causa.

Euclidian loss + Adversarial loss (let's call this generator G2)

Il generator non deve solo minimizzare le differenze tra $G(X)$ e Y , ma deve anche ingannare il discriminator. Più precisamente $G(X)$ deve essere t.c. $D(G(X)) = 1$. Questo approccio richiede un training più lungo a causa del "braccio di ferro" tra generator e discriminator; i risultati da noi ottenuti dimostrano infatti che a parità di epoch di training le performance di G2 sono inferiori a G1.

Euclidian loss + Adversarial loss + Image Gradient Difference loss (let's call this generator G3)

L'Image Gradient Difference loss (GDL) minimizza la somma degli scarti quadratici punto per punto tra i gradienti di Y e $G(X)$ lungo l'asse verticale e l'asse orizzontale. Lo scopo è quello di preservare le regioni del cervello con forti gradienti, vale a dire i contorni tra le aree di intensità diverse (a cominciare dal bordo esterno del cervello rispetto allo sfondo nero). Le performance sono superiori a quelle di G1 e G2.

L1 loss + Adversarial loss + Image Gradient Difference loss (let's call this generator G4)

È una nostra variante non presente nel paper. Abbiamo continuato il training del modello più performante di G3 (sul validation set) rimpiazzando l'Euclidian loss (L2) con il loss L1. Il motivo? A seguito della standardizzazione iniziale gli elementi delle celle nelle matrici risultavano già vicini a 0, ed il training ha ridotto ulteriormente gli scarti quadratici. Su differenze < 1 il loss L1, adottando il valore assoluto al posto del quadrato, permette di sfruttare per il training un margine di errore ancora superiore a quello ottenuto con L2. Si registrano miglioramenti in termini di MAE, MSE e PSNR già dopo 4 epoch di training.

T2 -> PD: DataFrames

<https://colab.research.google.com/drive/1cR4CY-nSuiFgNCeS38phEiTV8AmkQqEf>

Abbiamo riportato nel notebook gli score di MAE e PSNR ottenuti dai vari generatori epoch dopo epoch su ognuno dei 35 soggetti del validation set, oltre alla media e alla standard deviation di tali metriche sull'intero validation set. Per ogni generator (G1, G2, G3, G4) è stato scelto alla fine il modello corrispondente all'epoch in grado di segnare il PSNR medio più alto, il che ha corrisposto nella maggior parte dei casi al MAE medio più basso come si può vedere subito nel notebook successivo.

G2 è stato allenato fino alla 32esima epoch, quasi quanto G1 (34 epoch). G1 raggiunge il picco di avg PSNR alla 20esima epoch (26.657 ± 3.379) contro G2 alla 22esima (26.638 ± 3.35). Occorrerebbero più epoch a G2 per trarre beneficio dall'adversarial learning e, proprio per questo motivo, G3 è stato allenato fino alla 76esima ma con massimo avg PSNR registrato sempre alla 22esima (27.802 ± 3.294). In generale i tre generatori hanno mostrato un incremento del MAE e una diminuzione del PSNR dopo la 22esima epoch. Infine G4 supera G3 con avg PSNR pari a 29.765 ± 3.225 .

T2 -> PD: Plots

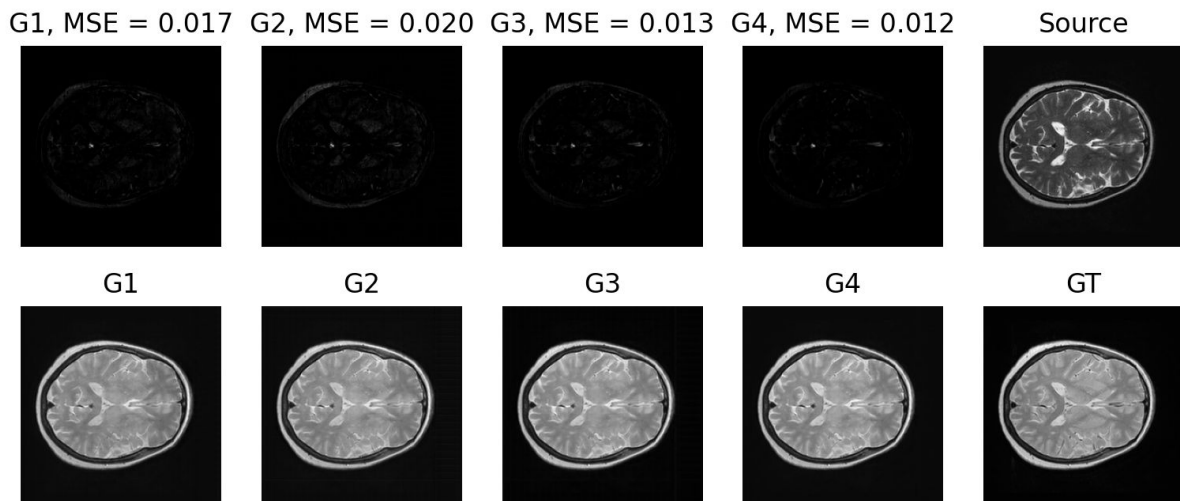
<https://colab.research.google.com/drive/1AsZ6Gd6wzXuHjyxsHQwZ1VM3njemC2UO>

Grafici cartesiani delle tabelle del notebook precedente. Si osservino in particolare gli andamenti dell'average MAE e dell'average PSNR di ogni generator sul validation set completo e su gruppi successivi di 5 soggetti alla volta.

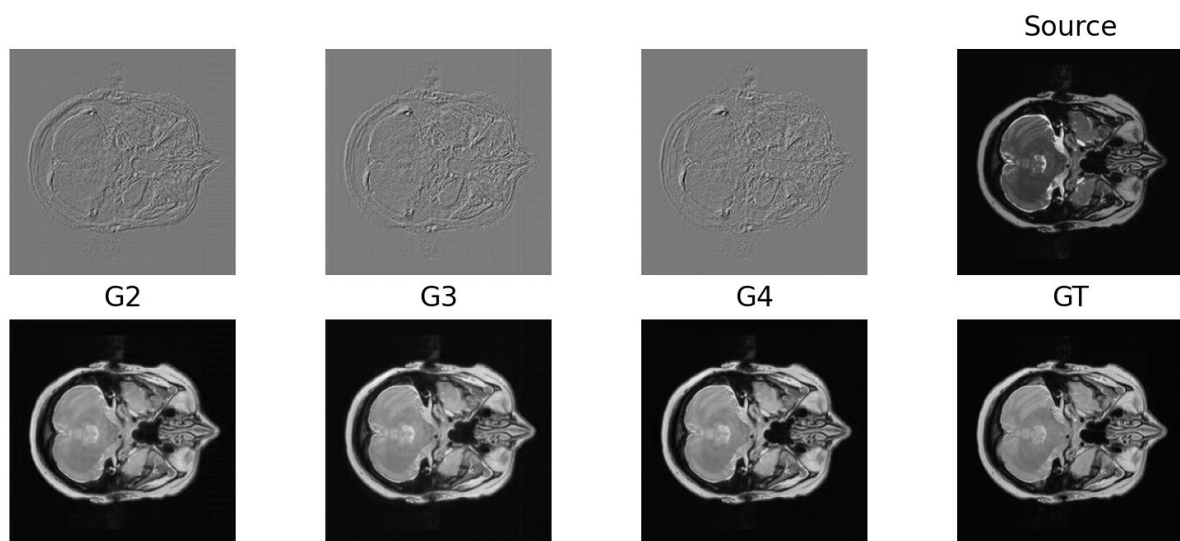
T2 -> PD: Confronto per immagini nel test set

<https://colab.research.google.com/drive/1nwGQFyoGwgHJfyhZx5XjTKG5NYzSmCHv>

Gli score di MAE e PSNR ottenuti sul test set di 70 soggetti confermano i “rapporti di forza” tra i generator evidenziati nel validation set. È questo il notebook in cui le immagini generate sono messe a confronto con il ground truth, prima in termini di MSE e poi in termini di image gradient difference. G3 mostra un calo importante del MSE (le differenze rispetto al ground truth sono visibilmente meno marcate se confrontate con G1 e G2).



Il MSE di G4 è ulteriormente più basso perché il loss L1, oltre a migliorare la resa del cervello stesso, riduce il rumore in fase di ricostruzione del volume completo come si può ben vedere in questa image gradient difference: le linee verticali e orizzontali in background ancora presenti in G3 scompaiono quasi del tutto in G4.



Esperimento 2: impatto del residual learning

Come annunciato all’inizio, la generazione sarà effettuata in verso opposto da PD a T2. Abbiamo confrontato le performance del modello L1 + LADV + GDL con e senza l’ausilio del residual learning, più precisamente con e senza la skip connection che somma pixel per pixel la source image iniziale all’output dell’ultimo dense layer del generator. L’intuizione alla base di questa input-output skip

connection è quella di accelerare il training fornendo subito un'immagine di partenza. Nella fase di training il generator imparerà a correggere epoch dopo epoch l'output del dense layer affinché, sommato all'input, generi un'immagine sempre più simile al real target T2. Si noti che questo approccio non sarebbe praticabile se le immagini PD e T2 non avessero alcuna somiglianza morfologica.

PD -> T2: DataFrames

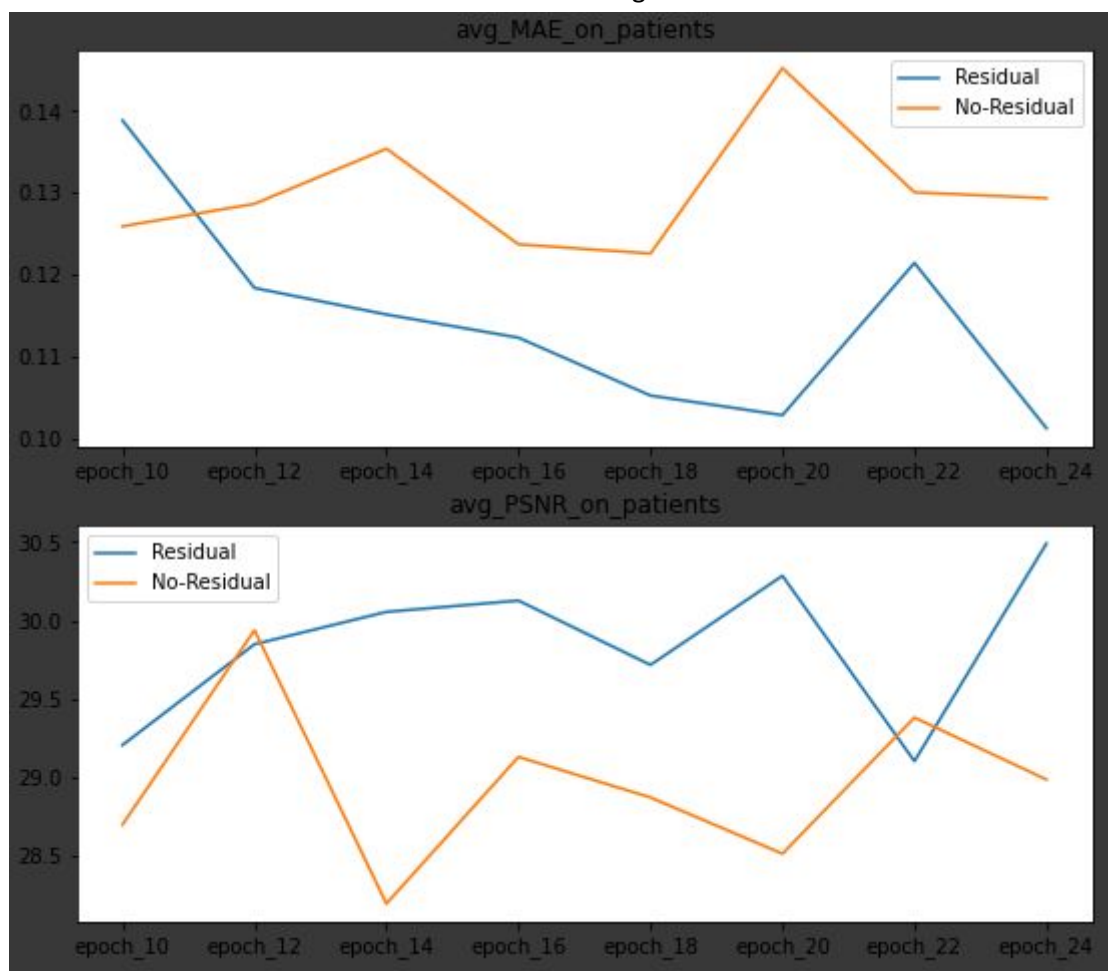
<https://colab.research.google.com/drive/1FU7Xj5neWYXImp1VoVrkeKdfud7snEtk>

I due generator sono stati allenati per 24 epoch con batch size = 49. Il generator senza skip-connection ottiene un avg PSNR pari 29.378 ± 2.627 alla 12esima epoch sul validation set, mentre il generator con il residual learning raggiunge 30.491 ± 2.336 alla 24esima.

PD -> T2: Plots

<https://colab.research.google.com/drive/1AsZ6Gd6wzXuHjyxsHQwZ1VM3njemC2UO>

Scorrere fino alla sezione dedicata al residual learning.



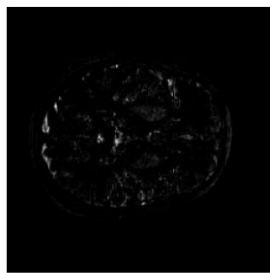
Per larga parte del training il residual learning permette di ottenere valori medi di MAE e di PSNR rispettivamente più bassi e più alti del modello standard. Inoltre gli andamenti del Residual model lasciano sperare in futuri miglioramenti oltre la 24esima epoch.

PD -> T2: Confronto per immagini nel test set

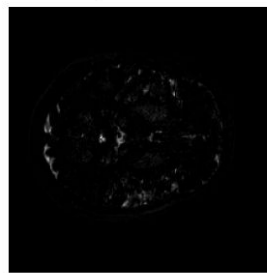
<https://colab.research.google.com/drive/1nwlivEiso1Y5KpST2EIGr8UjVe2QwXMI>

L'effetto positivo del residual learning è confermato anche dai dati nel test set.

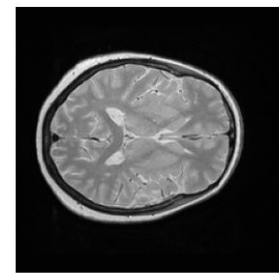
No Residual, MSE = 0.040 Residual, MSE = 0.028



No Residual

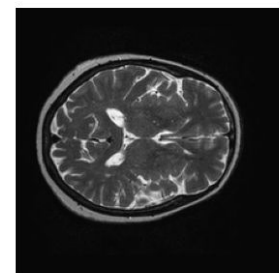
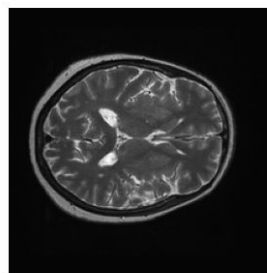
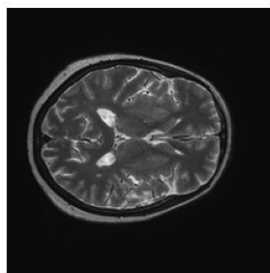


Residual

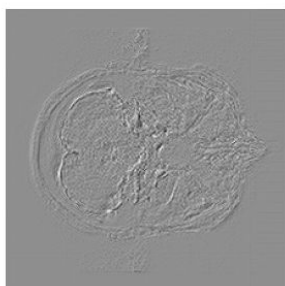


Source

GT



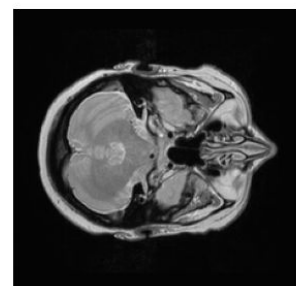
L'applicazione della skip-connection ha anche il vantaggio di appiattire le differenze di gradiente rispetto al ground truth (specialmente lungo i bordi esterni del cervello), e di eliminare totalmente il rumore di ricostruzione.



No Residual

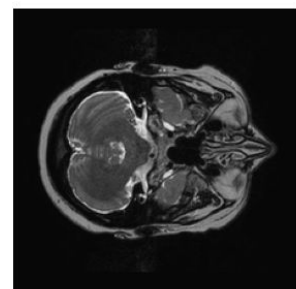
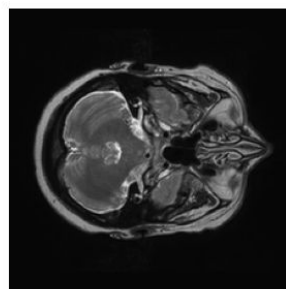
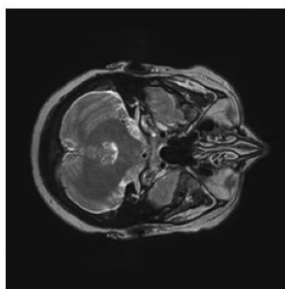


Residual



Source

GT



Cartella di Google Drive con il codice completo

<https://drive.google.com/drive/folders/128in-Ss7Dt3zHcF4leLOVX7oNlliSBY4?usp=sharing>