

WAVELET-DRIVEN GENERATIVE MODELS

Luigi Sigillo, Eleonora Grassucci, Danilo Comminiello, Aurelio Uncini

Dipartimento di Ingegneria dell'Informazione, Elettronica e Telecomunicazioni (DIET)
"Sapienza" Università di Roma
Via Eudossiana 18, 00184 Roma

Parole chiave: modelli generativi, diffusion model, trasformate Wavelet.

I recenti progressi nei modelli generativi hanno influenzato in modo significativo diversi ambiti applicativi, mettendo in particolare evidenza i Large Language Models (LLM) per la generazione di contenuti testuali e i Diffusion Models (DM) per la generazione di contenuti multimediali, tra cui immagini, audio e video. Nonostante le loro capacità all'avanguardia, i Diffusion Models affrontano spesso notevoli sfide computazionali, in particolare in termini di tempo di generazione, che rimane considerevolmente elevato rispetto a framework alternativi come GAN o VAE.

Per affrontare queste limitazioni computazionali, la ricerca recente si è orientata verso l'integrazione di tecniche classiche di signal processing all'interno di architetture generative profonde. Particolarmente promettenti in questo contesto sono le trasformate Wavelet, che permettono ai Diffusion Models di manipolare esplicitamente le componenti a bassa e alta frequenza dei dati in ingresso, ottimizzando così sia l'efficienza computazionale che la qualità della generazione. Incorporando moduli basati su Wavelet, i processi diffusivi possono operare direttamente all'interno delle sottobande wavelet, migliorando significativamente l'efficienza dell'apprendimento del modello e riducendo sia i tempi di training che quelli di sampling [1, 3].

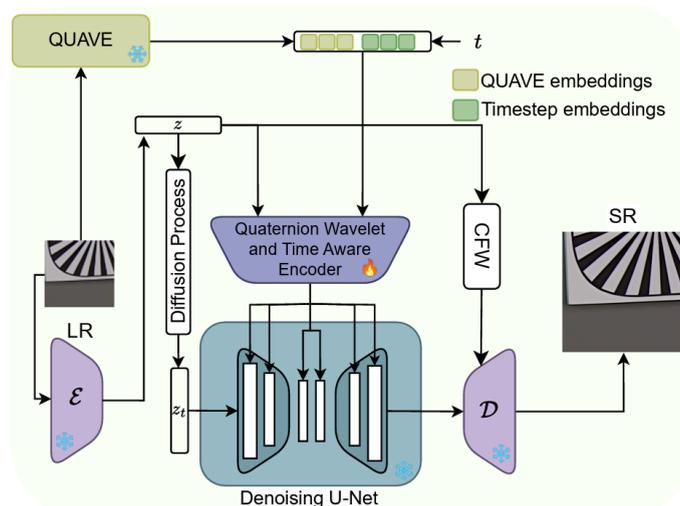


Figura 1: Diffusion model con moduli di estrazione delle sottobande Wavelet (QUAVE [2]) per migliorare il task di super resolution.

In particolare, l'uso di Wavelet Diffusion Models nei task di super-risoluzione delle immagini ha dimostrato un'efficacia notevole. Le sottobande wavelet ad alta frequenza forniscono informazioni dettagliate fondamentali, migliorando la qualità della ricostruzione delle immagini con risoluzione migliore. Il processo di diffusione inizia da un rumore gaussiano combinato con le sottobande ottenute dalle wavelet, per poi applicare una trasformata wavelet

inversa che consente di generare l'immagine ad alta risoluzione finale. Questo metodo riduce significativamente la dimensionalità dell'immagine grazie alle proprietà delle wavelet, diminuendo così il carico computazionale e la memoria GPU necessaria. Inoltre, questo approccio basato su wavelet gestisce direttamente i dettagli critici dell'immagine e converge più rapidamente rispetto ai modelli di diffusione tradizionali, che operano tipicamente in domini pixel o latenti senza decomposizione in frequenza. La letteratura recente ha ulteriormente sottolineato tali vantaggi esplorando processi diffusivi strutturati che sfruttano metodi gerarchici di signal processing per ottenere una generazione di immagini più veloce e scalabile.

Nel complesso, equipaggiare i Diffusion Models con trasformate Wavelet non solo migliora la qualità generativa, particolarmente importante per i task sensibili ai dettagli visivi, ma incrementa notevolmente anche l'efficienza computazionale. Questa combinazione rende i metodi generativi avanzati più accessibili anche in contesti con risorse computazionali limitate, colmando il divario tra performance e praticità.

Bibliografia

- [1] L. Sigillo, C. Bianchi, A. Uncini, D. Comminiello, "Quaternion Wavelet-Conditioned Diffusion Models for Image Super-Resolution", IJCNN 2025.
- [2] L. Sigillo, E. Grassucci, A. Uncini, D. Comminiello, "Generalizing medical image representations via quaternion wavelet networks", Neurocomputing, 2025.
- [3] Phung, Hao et al. "Wavelet Diffusion Models are fast and scalable Image Generators", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, 10199-10208.