

Architettura del Sistema di Captioning SVG

March 29, 2025

Introduzione

Questo documento descrive in dettaglio le architetture implementate per il sistema di generazione di descrizioni da file SVG. La versione corrente elimina ogni riferimento al doppio percorso (raster+vettoriale) mantenendo solo l'approccio vettoriale.

1 Architettura Decoder-Only

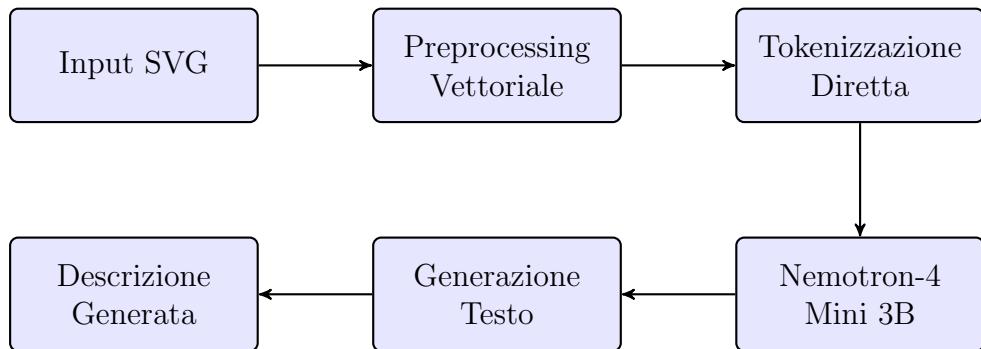


Figure 1: Flusso completo dell'architettura decoder-only

Dettagli Implementativi

- **Preprocessing Vettoriale:** Estrazione degli elementi vettoriali (path, forme geometriche) con relative proprietà
- **Attention Meccanism:** Self-attention nello spazio vettoriale con dimensioni chiave 512
- **Hardware Requirements:** 16GB GPU RAM per batch size 8

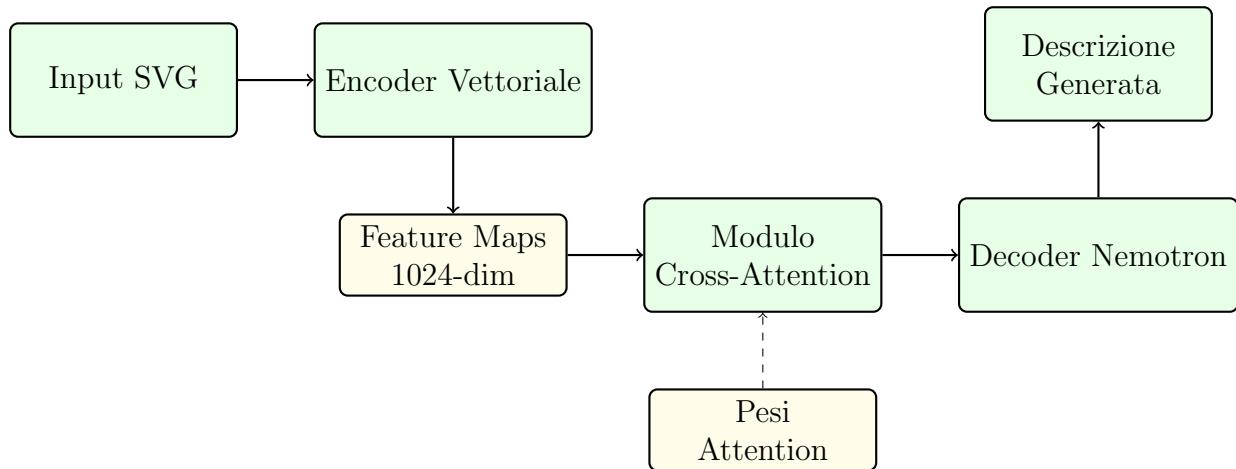


Figure 2: Architettura completa encoder-decoder con cross-attention

2 Architettura Encoder-Decoder Completa

Specifiche Tecniche

| Componente | Configurazione | Note |
|-----------------|--------------------------|-------------------------|
| Encoder | 12 layer, 768 hidden dim | Addestrato su 1M SVG |
| Cross-Attention | 8 heads | Temperatura 0.1 |
| Decoder | Nemotron-4 3B | Congelato per 50 epoche |