



UNIVERSITÀ
DEGLI STUDI
DI TERAMO



Principali modelli di diffusione per la generazione di immagini

Prof. ssa Romina Eramo

Università degli Studi di Teramo

Dipartimento di Scienze della Comunicazione

reramo@unite.it

Introduzione

- » I modelli di diffusione sono una classe di algoritmi di intelligenza artificiale che generano immagini sintetiche partendo da rumore casuale.
 - Utilizzano una tecnica di addestramento inverso in cui imparano a "rimuovere il rumore" fino a creare un'immagine chiara basata su un input testuale.
- » Come funzionano:
 - Si parte da un'immagine completamente rumorosa.
 - Il modello applica una serie di passaggi per ridurre il rumore gradualmente, utilizzando le informazioni fornite dal prompt testuale.
 - Il risultato finale è un'immagine che rispecchia il testo fornito.

I Principali Modelli di Diffusione

DALL-E (OpenAI)

Piattaforma ufficiale OpenAI: <https://platform.openai.com/dalle> (offre dei crediti gratuiti), in alternativa usare ChatGPT (<https://chatgpt.com>)

Caratteristiche principali:

- » Generazione di immagini da descrizioni testuali.
- » Capacità di creare composizioni creative e dettagliate.
- » Supporta la modifica di immagini esistenti (inpainting).

Punti di forza:

- » Creatività elevata e capacità di creare immagini surreali.
- » Ampia comprensione dei dettagli nei prompt.

Limiti:

- » Non sempre coerente in dettagli complessi.
- » Richiede specificità nei prompt per ottenere risultati precisi.

I Principali Modelli di Diffusione

MidJourney

Utilizzato principalmente tramite Discord (<https://discord.com/>)

Caratteristiche principali:

- » Ottimizzato per la creazione di immagini artistiche e stilizzate.

Punti di forza:

- » Eccezionale per stili visivi unici e creatività artistica.
- » Ottima resa per illustrazioni, paesaggi fantastici e concetti astratti.

Limiti:

- » Meno adatto a immagini fotorealistiche rispetto ad altri modelli.
- » Richiede un po' di pratica per sfruttare al massimo i suoi strumenti.

I Principali Modelli di Diffusione

Stable Diffusion

Utilizzabile tramite, ad esempio:

- DreamStudio (versione base gratuita - <https://dreamstudio.ai/>)
- OpenArt (credits gratuiti - <https://openart.ai/>)
- Leonardo.Ai (credits gratuiti – <https://leonardo.ai/>)

Caratteristiche principali:

- » Modello open-source per la generazione di immagini.
- » Altamente personalizzabile e integrabile in altre applicazioni.

Punti di forza:

- » Flessibilità e accesso alla comunità open-source.
- » Adatto a scenari personalizzati e specifici.
- » Supporta la generazione di immagini fotorealistiche e stilizzate.

Limiti:

- » Può richiedere conoscenze tecniche per personalizzazioni avanzate
- » Meno intuitivo rispetto a DALL-E o MidJourney

I Principali Modelli di Diffusione

Gemini (Google DeepMind)

Al momento, Google Gemini non offre strumenti gratuiti per la generazione di immagini.

Caratteristiche principali:

- » Integrazione di capacità multimodali avanzate.
- » Generazione di immagini e video (testo-immagine e testo-video).

Punti di forza:

- » Capacità di comprendere contesti complessi nei prompt.
- » Supporto per input e output multimodali.

Limiti:

- » Tecnologia emergente, ancora in fase di affinamento.
- » Accessibilità limitata rispetto agli altri modelli.

Generazione di Contenuti Video (Text-to-Video)

Evoluzione del Text-to-Video

- » I modelli di diffusione vengono adattati per creare video partendo da descrizioni testuali.

Espansione verso nuove frontiere:

- » Text-to-Video: Generazione di video a partire da descrizioni testuali.
- » Image-to-Video: Creazione di animazioni coerenti basate su immagini fornite.
- » Video-to-Video: Modifica e trasformazione creativa di video esistenti.

Generazione di Contenuti Video (Text-to-Video)

Principali modelli emergenti:

- » RunwayML Gen-2:
 - Co-creato da uno dei fondatori di Stable Diffusion.
 - Leader nel campo del Text-to-Video, capace di produrre risultati sempre più utilizzabili.
 - Focus su video brevi e animazioni generate da testo o immagini.
- » Stable Video Diffusion:
 - Lanciato a novembre 2023, permette di trasformare il testo in clip video brevi o di animare immagini esistenti.
- » Stable Diffusion Turbo:
 - Ottimizzato per la generazione di immagini in tempo quasi reale, con potenziali applicazioni per video frame-by-frame.

Generazione di Contenuti Video (Text-to-Video)

- » ...
- » Sora (OpenAI):
 - Rilasciato a febbraio 2024, dimostra che OpenAI sta investendo nella tecnologia Text-to-Video.
 - Integra avanzate capacità multimodali per espandere l'interazione tra testo, immagini e video.
- » Google Phenaki: Progetta sequenze video lunghe e coerenti.
- » Meta Make-A-Video: Integra capacità di generazione video con animazioni fluide.

Applicazioni del Text-to-Video

- » Marketing: Creazione di brevi spot pubblicitari generati automaticamente.
- » Intrattenimento: Generazione di clip per giochi o scenari cinematografici.
- » Educazione: Creazione di contenuti didattici interattivi.

Generazione di Contenuti Video (Text-to-Video)

Modello	Punti di forza	Limiti	Applicazioni ideali
DALL-E	Creatività, modifica immagini (inpainting)	Meno fotorealistico in dettagli complessi	Creazione artistica e concettuale
MidJourney	Stile artistico unico, grande flessibilità visiva	Limitato per il fotorealismo	Illustrazioni e design creativo
Stable Diffusion	Open-source, personalizzabile	Richiede conoscenze tecniche avanzate	Progetti personalizzati e complessi
Gemini	Multimodalità, integrazione avanzata	Accessibilità limitata	Progetti di ricerca e contenuti multimodali
Text-to-Video	Generazione di video da testo	Tecnologia emergente, limitata nella durata	Brevi video per marketing e creatività

Generazione di Contenuti Video (Text-to-Video)

Alcuni strumenti gratuiti:

- RunwayML (Gen-2) <https://runwayml.com>, la versione gratuita ha un limite di generazioni mensili)
- Pika Labs, accessibile tramite server Discord ufficiale
- Kaiber AI, la versione gratuita è limitata nel numero di video generabili.

Confronto tra modelli

Input:

Un corgi in cima alla Porta di Brandeburgo

DALL-E 3



Midjourney v6



Stable Diffusion XL

