Relazione di stage - generazione di immagini con AI

Durante lo stage è stato svolto un lavoro di sperimentazione sulle potenzialità di **Stable Diffusion** e delle diverse piattaforme che lo supportano, con l'obiettivo di generare immagini e contenuti multimediali.
L'attività ha incluso sia un'analisi teorica dei modelli, sia una parte pratica di confronto tra piattaforme e la documentazione di esperimenti reali.

Funzionamento di Stable Diffusion

Stable Diffusion è un **modello di diffusione latente** che genera immagini trasformando rumore casuale in contenuti coerenti con un prompt. Il processo avviene in due fasi principali:
-**Encoding**: il testo viene compreso tramite un modello linguistico (CLIP) e tradotto in uno spazio latente;
-**Diffusione inversa**: a partire da rumore gaussiano, il modello applica progressivamente dei passi di denoising guidati dal prompt per generare l'immagine desiderata.

Parametri principali

Prompt: descrizione testuale dell'immagine desiderata.

Negative Prompt: specifica cosa escludere (es. *blurry, low quality*).

Steps: numero di iterazioni di raffinamento (più steps = maggiore dettaglio, ma tempi più lunghi).

CFG Scale: grado di aderenza al prompt (valori alti = maggiore fedeltà, valori bassi = più creatività).

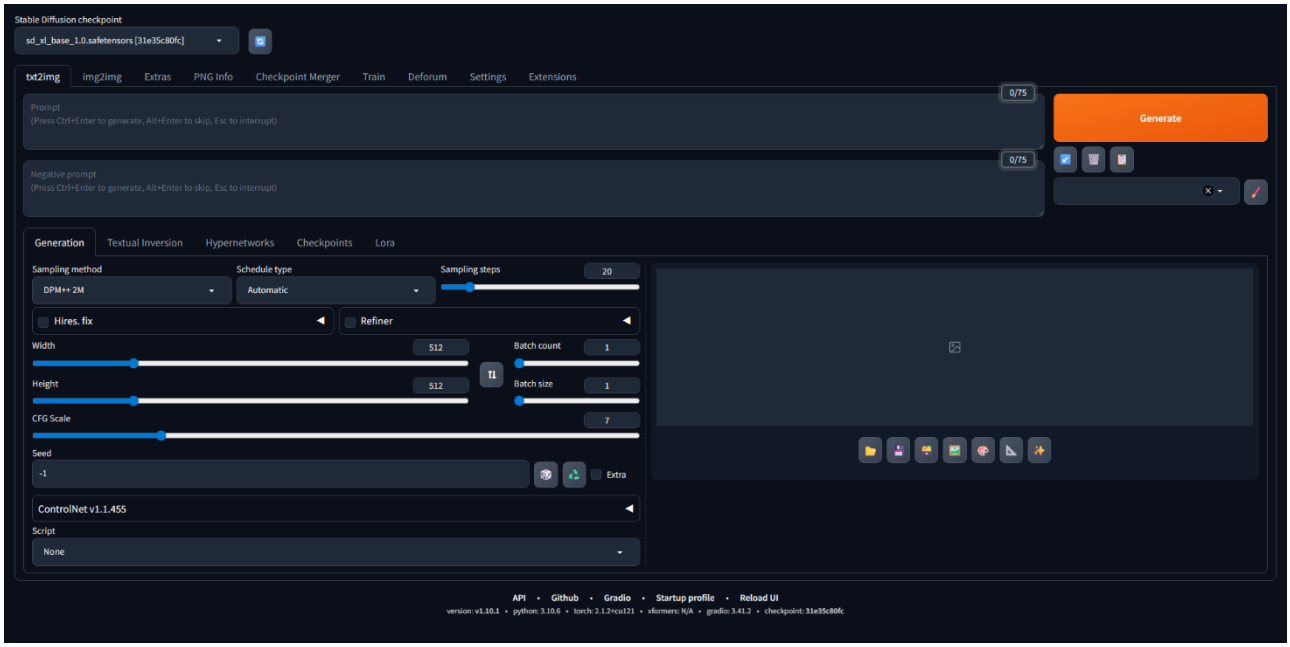
Seed: numero che controlla la casualità (stesso seed + stessi parametri = stesso output).

Resolution: dimensioni finali dell'immagine (es. 512x512, 768x512).

Strength (per Image-to-Image): grado di trasformazione rispetto all'immagine iniziale.

Frame Interpolation (per Video): parametri per generare continuità tra fotogrammi.

Primo test con WebUI (Automatic1111)

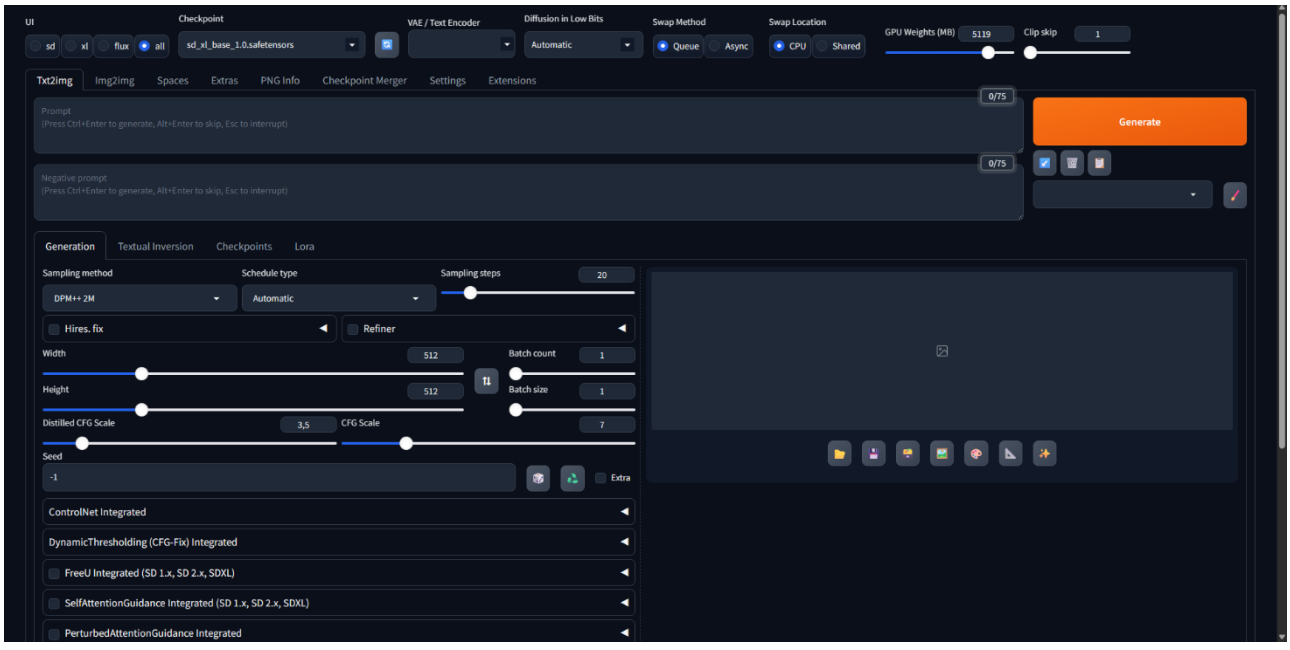


Pro: interfaccia semplice, molte estensioni disponibili, adatto a chi inizia.

Contro: pesante, poco ottimizzato, meno flessibile per workflow complessi.

Quando usarlo: per sperimentazioni veloci, senza troppi passaggi tecnici.

Secondo test con Forge/Torch

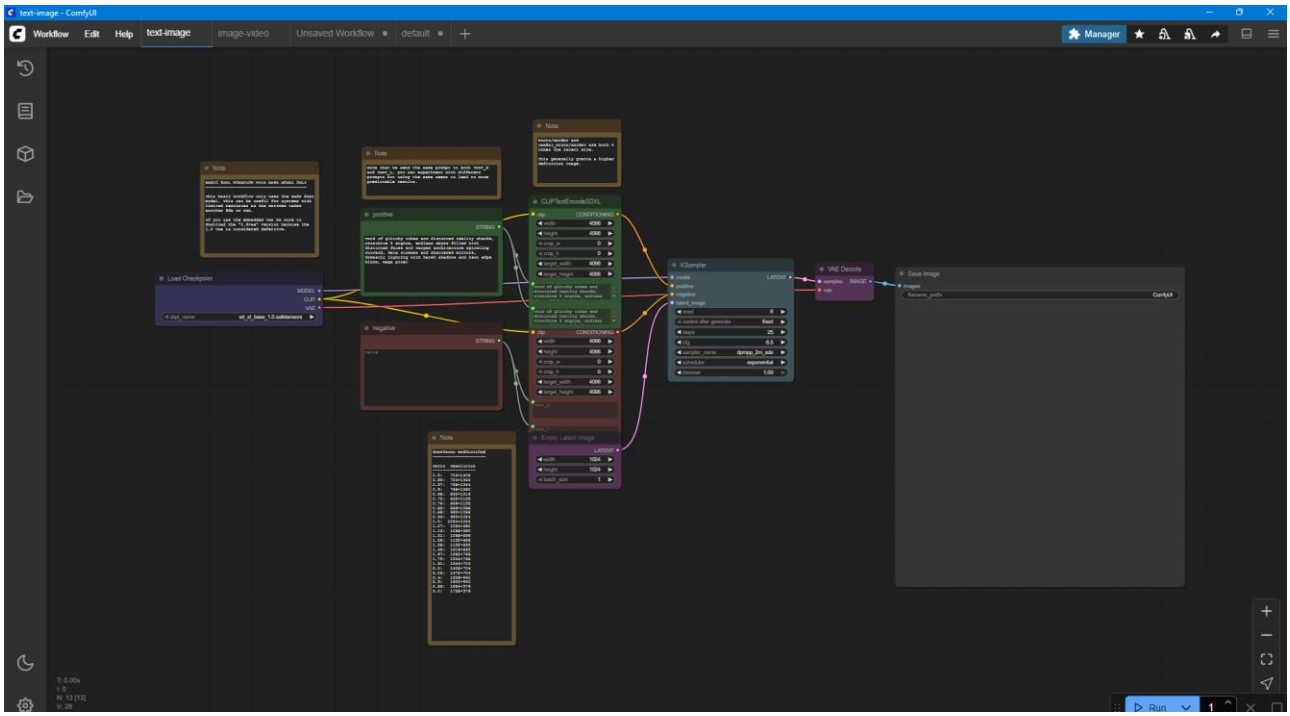


Pro: migliore gestione delle prestazioni, supporto a più modelli e plugin.

Contro: rimane legato a un approccio 'one-click' poco personalizzabile, workflow non modulari.

Quando usarlo: per chi vuole una WebUI più fluida ma senza complicarsi.

Test finale con ComfyUI



Pro: estremamente modulare (consente di costruire pipeline tramite nodi collegabili), prestazioni elevate con un migliore sfruttamento della GPU, maggiore stabilità, creazione di workflow riutilizzabili e condivisibili, altamente personalizzabile.

Contro: curva di apprendimento più alta all'inizio.

Quando usarlo: per progetti avanzati e workflow professionali.

Esempi con immagini

Prompt utilizzato: old house in the distance above a lake with an old man fishing

Prompt negativi utilizzabili: bad quality, worst quality, worst detail, sketch

Qualità: 1024x1024 **Steps**: 20-30

Automatic1111	*Forge/Torch*	*ComfyUI*
		

Tempi di generazione

Dai 5 ai 10 minuti	Circa 30-60 secondi	Meno di 30 secondi

Perfezionamento con ComfyUI

Una volta individuato ComfyUI come piattaforma principale, l'attenzione si è spostata sul perfezionamento dei risultati. Attraverso la creazione di workflow più strutturati e l'impiego dell'intelligenza artificiale per la generazione di prompt più mirati e dettagliati, è stato possibile migliorare sensibilmente la qualità delle immagini prodotte, rendendole più coerenti e vicine all'idea iniziale.

Esempio

Trasformazione prompt precedente: old house in the distance overlooking a still lake, old man fishing alone from a fragile wooden boat, distant forest reflecting on the surface, background filled with fog

Prompt negativo: Con questo workflow possiamo evitare di scriverlo, però si può usare per escludere specifiche cose non volute nelle immagini (oggetti, persone, colori…)

Passaggi per l'utilizzo

Scaricare e installare ComfyUI https://www.comfy.org/download ;

Scaricare il modello https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0/blob/main/sd_xl_base_1.0.safetensors e spostarlo nella cartella di ComfyUI (percorso scelto > ComfyUI > models > checkpoints);

Scaricare il workflow https://openart.ai/workflows/openart/basic-sdxl-workflow/P8VEtDSQGYf4pOugtnvO e trascinarlo semplicemente su ComfyUI;

Consiglio di impostare il seed su "increment" (in modo da generare immagini simili con lo stesso prompt senza cambiare nulla e tenere facilmente traccia dei seed usati) e di lasciare le altre impostazioni come sono (controllare di aver selezionato il modello giusto "sd_xl_base_1.0.safetensors" sul nodo viola iniziale);

Copiare il prompt (Rodolfo > Immagini > ChatGPT_prompt) su ChatGPT e chiedere parole chiave, come ad esempio "bellezza, oscurità" o "paura, intuizione" e così via;

Scegliere un prompt generato e copiarlo sul nodo "positive";

Cliccare "run" (CTRL + ENTER).

Conclusione

Il percorso ha permesso di confrontare diverse piattaforme e di individuare in ComfyUI la soluzione più adatta per la generazione avanzata di immagini. Attraverso workflow progressivamente più complessi è stato possibile raggiungere risultati di qualità professionale, comprendendo a fondo le potenzialità degli strumenti di AI generativa.