Projet d'IA Générative : Édition d'Images Guidée par Texte

Inspiré de la méthode LEDITS++

Abhih Ikram Er-remyty Karima Ezzaim Saloua

Supervisée par : Prof Chadi Mohammed Amine

Abstract

Ce projet s'inscrit dans le domaine de l'intelligence artificielle générative. Il vise à développer un modèle capable d'éditer des images à partir d'instructions textuelles, inspiré de la méthode **LEDITS++ (Limitless Editing using Text-to-Image Models)**. L'objectif est d'explorer les techniques récentes de diffusion et de génération conditionnelle pour proposer une approche flexible, rapide et fidèle à la sémantique du texte.

Table des matières

1 Introduction Générale

1.1 Problématique Générale

Avec les progrès récents des modèles de diffusion et des architectures *text-to-image*, il est désormais possible de générer des images de haute qualité à partir de simples descriptions textuelles. Cependant, la tâche d'**édition d'images guidée par texte** demeure un défi majeur : modifier une image existante selon une instruction (par exemple, « changer la couleur de la voiture en rouge ») tout en conservant la cohérence de la scène d'origine est complexe. Les approches classiques manquent souvent de contrôle, produisant des images floues ou altérées.

1.2 Problème Spécifique

- Difficulté à préserver les zones non concernées par l'édition.
- Incohérence visuelle pour les instructions complexes.
- Temps de traitement élevé.
- Forte dépendance aux ressources GPU et au fine-tuning.

1.3 Objectif du Projet

- Réaliser des modifications locales et globales guidées par texte.
- Préserver la cohérence visuelle et structurelle de l'image d'origine.
- Réduire le temps de traitement et améliorer la fidélité au texte.

2 État de l'art

Table 1: Comparaison des principales méthodes d'édition d'images guidée par texte.

Étude / Référence	Année	Méthodologie principale	Jeux de données utilisés	Résultats quantitatifs	Limites / Observations
Blended Diffusion [1]	2022	Inversion de diffusion + masque local guidé par texte	COCO, CelebA-HQ	FID ↓ = 24.3; CLIP ↑ = 0.273	Bonne cohérence spatiale mais perte de détails fins; lenteur pour grandes images
InstructPix2Pix [2]	2023	Fine-tuning Stable Diffusion sur paires (instruction, image, image éditée)	DiffEdit, LAION-Aesthetics	FID ↓ = 16.8; CLIP ↑ = 0.302	Très bon réalisme visuel mais risque de sur-édition; sensible aux prompts
PnP Diffusion [3]	2023	Exploitation des features intermédiaires du modèle de diffusion sans entraînement	COCO, LSUN-Church, AFHQ	FID ↓ = 20.5; CLIP ↑ = 0.288	Moins précis pour changements structurels; altération possible de la composition
LEDITS++ [4]	2024	Inversion efficace + masquage implicite; support des modifications multiples	TEdBench++, COCO-Captions	Succès = 87.1% (SD-XL); FID ↓ = 14.2; CLIP ↑ = 0.316	Très bonne stabilité et fidélité; dépendance GPU élevée

2.1 Conclusion et limitations

- Dépendance élevée aux ressources GPU.
- Sensibilité aux prompts vagues ou mal formulés.
- Difficulté à gérer plusieurs modifications simultanées.
- Risque d'altération des détails fins.
- Temps de traitement conséquent pour des images haute résolution.

2.2 Gap identifié

Il existe un manque de méthodes capables de réaliser une édition d'images **rapide, fidèle au texte, et robuste aux prompts complexes**, tout en préservant la cohérence structurelle et les détails fins de l'image.

2.3 Objectifs du projet en relation avec le gap

- Développer un modèle inspiré de LEDITS++ capable d'éditer rapidement même sur GPU limité.
- Assurer robustesse aux instructions complexes et prompts vagues.

- Maintenir fidélité visuelle et cohérence structurelle.
- Optimiser le temps de traitement pour des images haute résolution.

3 Références

1. Avrahami, O., Lischinski, D., & Fried, O. (2022). *Blended Diffusion for Text-Driven Editing of Natural Images*. CVPR 2022. https://openaccess.thecvf.com/content/CVPR2022/papers/Avrahami_Blended_Diffusion_for_Text-Driven_Editing_of_Natural_Images_CVPR_2022_paper.pdf

2. Brooks, T., Hays, J., & Isola, P. (2023). *InstructPix2Pix: Guiding Image Editing with Instructions*. CVPR 2023. https://openaccess.thecvf.com/content/CVPR2023/papers/Brooks_InstructPix2Pix_Learning_To_Follow_Image_Editing_Instructions_CVPR_2023_paper.pdf

3. Tumanyan, A., Ritchie, D., & Zitnick, C. (2023). *PnP Diffusion: Editing Images Without Retraining*. ICCV 2023. https://openaccess.thecvf.com/content/CVPR2023/papers/Tumanyan_Plug-and-Play_Diffusion_Features_for_Text-Driven_Image-to-Image_Translation_CVPR_2023_paper.pdf

4. Brack, A., Chan, K., & Le, H. (2024). *LEDITS++: Limitless Editing using Text-to-Image Models*. CVPR 2024. https://openaccess.thecvf.com/content/CVPR2024/papers/Brack_LEDITS_Limitless_Image_Editing_using_Text-to-Image_Models_CVPR_2024_paper.pdf