

Informe Inicial TFG

Ús de models de difusió per alterar la realitat física de les escenes

Tutor: Ramon Baldrich Caselles
Alumna: Abril Piñol Chacon - 1604159

1 INTRODUCCIÓ

Aquest informe s'escriu amb la finalitat de presentar i definir el projecte de final de grau, proporcionant una visió detallada dels objectius establerts, l'estat de l'art i la metodologia prevista, així com la planificació del treball, organitzat setmanalment i per objectius. És a dir, en aquest document es descriu el punt de partida, en els futurs informes de progrés se'n farà el seguiment. En els darrers, s'hi exposaran els diferents reptes que s'han plantejat fins al moment i s'analitzaran tant l'evolució com els resultats aconseguits fins a la data, de la mateixa manera que el perquè de cada decisió presa.

El projecte es basa en un model d'intel·ligència artificial (IA) que ha experimentat un notable increment de popularitat i presència, especialment en les xarxes socials els darrers temps. Aquesta tendència s'ha fet visible en diverses ocasions, quan s'ha viralitzat contingut basat en imatges generades per IA. Els models en qüestió es coneixen com a models de generació d'imatges, i aquest projecte se centra específicament en els coneguts com a *Stable Diffusion* (SD).

2 OBJECTIUS

En aquest apartat, es plantegen els objectius que guiaran el desenvolupament i l'execució del projecte. Aquests objectius s'han establert

amb la finalitat de proporcionar una estructura clara i definida a l'hora d'afrontar les tasques necessàries per assolir l'objectiu final. A través d'aquests objectius específics, es busca comprendre a fons el funcionament dels models *Stable Diffusion* i aspectes directament relacionats amb aquests, com ho serien la *ControlNet* i diferents arquitectures específiques de tasques d'imatge restoration, entre altres. Es vol aplicar aquest coneixement per aconseguir resultats efectius i innovadors en la modificació d'imatges. Més endavant es detallen els objectius concrets de què es parla en una taula (*figura 1, annex*).

L'objectiu últim és replicar la tasca presentada a l'article "*Relighting from a Single Image: Datasets and Deep Intrinsic-based Architecture*"[1], substituint l'aproximació basada en les *Generative Adversarial Networks* (GANs) per models basats en *Stable Diffusion*.

3 ESTAT DE L'ART

En aquesta fase inicial, és important el context i una visió general del panorama actual dels models de generació d'imatges, especialment SD. És això el que es desenvolupa a continuació.

Dins dels models de generació d'imatges s'hi poden trobar els *Generative Adversarial Networks* (GANs)[10], a trets generals, aquests estan formats per dues xarxes: un generador i un discriminador. El generador crea imatges i

el discriminador intenta distingir si les imatges que li arriben han estat creades per aquest generador o no. Mitjançant l'entrenament adversari, el generador aprèn a produir imatges més realistes que puguin enganyar al discriminador.

Un altre tipus de models que permet crear imatges són els *Variational Autoencoders* (VAEs)[11], aquests consisteixen en un *encoder* que genera una representació de les imatges d'entrada en un espai latent, i un *decoder* que reconstrueix imatges a partir d'aquesta representació en l'espai latent. D'aquesta manera, mostrejant de l'espai latent el *decoder* pot generar imatges.

A diferència d'aquests, els models SD han mostrat millors resultats quan es tracta de crear imatges amb textures més complexes, detalls petits i cantonades afilades. Un dels trets responsables d'aquesta millora en els resultats és el procés iteratiu, aquest permet anar refinant la sortida del model i perfeccionar fins a l'últim detall.

El funcionament bàsic dels models de difusió consisteix a afegir un soroll conegut a les imatges d'entrenament per posteriorment fer el procés invers i aprendre a recuperar les imatges a partir del soroll. D'aquesta manera, un cop entrenats són capaços de crear imatges coherents partint de soroll.

En termes generals, es pot dividir la funcionalitat d'aquest model en dues categories: la creació d'imatges de zero i la modificació d'imatges ja existents. La creació d'imatges des de zero es pot condicionar utilitzant una *prompt* textual. Donat un fragment de text, aquest es codifica en el mateix espai latent i s'incorpora en el procés d'eliminació el soroll. En el cas de modificació d'imatges, no es parteix de només soroll, sinó que hi ha una imatge original que condiciona la sortida. El que es pretén és aplicar certes transformacions a aquesta imatge, però mantenint fidelitat a la imatge d'entrada, un exemple en seria l'eliminació de la pluja o de les ombres d'una imatge.

En aquest treball, l'objectiu és modificar imatges existents, concretament alterant les seves propietats físiques, però no sempre de la

mateixa manera. Per tant, a més, es vol fer servir una guia o instrucció específica, en el cas de la tasca de *relighting*, aquesta hauria d'indicar les condicions concretes d'il·luminació que s'esperen en el resultat.

4 METODOLOGIA

La metodologia adoptada per aquest treball està basada en el plantejament *Agile*, aquest és un mètode seqüencial i d'adaptació continua. És cert que inicialment s'estableix un pla de treball concret, no obstant això, aquest es revisa sovint i s'adapta a les necessitats i característiques reals del projecte. És per aquest motiu que és important una comunicació freqüent i eficient que permeti traçar el millor camí per avançar de cara a l'objectiu final.

En aquest treball, les característiques esmentades pel que fa a la metodologia es veuen reflectides en una planificació setmanal, que es pot veure en detall a la següent secció (*secció: 4 Planificació*), i per objectius, ja descrits en una de les seccions anteriors (*secció: 2 Objectius*). Tot això s'acompanya de reunions setmanals amb el tutor on es discuteixen els resultats i/o problemes que puguin sorgir i es determina la millor manera de procedir. A més a més, s'avalua si s'han aconseguit els objectius establerts i es determina la millor manera de procedir. Aquestes trobades són clau per evitar l'estancament del projecte i optimitzar el procés de treball, ja que permeten identificar i abordar els problemes de manera ràpida, minimitzant l'impacte que poden tenir a llarg termini.

En resum, la metodologia adoptada combina una planificació flexible i adaptativa amb una comunicació i col·laboració constants, així com una avaluació periòdica del progrés, per garantir un desenvolupament eficient i satisfactori del projecte. Aquest enfocament proporciona una estructura coherent i a la vegada permet una resposta àgil als canvis i les necessitats que sorgeixen durant el procés de treball.

Fins ara es parla de la metodologia de desenvolupament establerta, però en una fase

posterior, s'explorà amb més detall la metodologia tècnica, que inclourà una descripció exhaustiva de les eines utilitzades en el codi, com ara llibreries específiques i altres recursos rellevants per a l'execució del projecte. Aquesta anàlisi tècnica permetrà una comprensió més profunda de les tecnologies emprades que contribueixen a la implementació i execució dels objectius.

5 PLANIFICACIÓ

En aquesta secció es descriu minuciosament el pla de treball establert d'acord amb la metodologia descrita prèviament, aquest ens permet tenir una perspectiva general de l'evolució prevista del projecte al llarg del temps. Per una visió més clara es mostra en format tabular (*figura 2*), aquesta taula conté quatre columnes on s'especifiquen les setmanes, juntament amb les seves dates corresponents, els identificadors prèviament assignats als objectius previstos a assolir en aquests períodes i finalment una pinzellada dels assoliments reals, que es detallaran en més profunditat en els informes de progrés. En aquest punt del projecte només es té informació de la darrera columna fins a la cinquena setmana que és el temps transcorregut fins al moment. Addicionalment, les files en negre destaquen les dates on hi ha entregues previstes.

6 REFERÈNCIES

- [1] Yixiong Yang, Hassan Ahmed Sial, Ramon Baldrich, Maria Vanrell (Juliol 2023) "Relighting from a Single Image: Datasets and Deep Intrinsic-based Architecture".
- [2] Steins (2023) "Stable diffusion clearly explained!", *Medium*. Disponible a: <https://medium.com/@steinsfu/stable-diffusion-clearly-explained-ed008044e07e#0b5c> (Accés: 05 Febrer 2024).
- [3] Jamil, U. (2023) "Coding stable diffusion from scratch in Pytorch", *YouTube*. Disponible a: https://www.youtube.com/watch?v=ZBKpAp_6TGI&t=2185s (Accés: 08 Febrer 2024).
- [4] Adaloglou, N. and Karagiannakos, S. (2022) "How diffusion models work: The math from scratch, *AI Summer*." Disponible a: <https://theaisummer.com/diffusion-models/> (Accés: 07 Febrer 2024).
- [5] Lin, X. *et al.* (2023) "Diffbir: Towards blind image restoration with generative diffusion prior", *arXiv.org*. Disponible a: <https://arxiv.org/abs/2308.15070> (Accés: 13 Febrer 2024).
- [6] Luo, Z. *et al.* (2023) "Refusion: Enabling large-size realistic image restoration with latent-space diffusion models", *arXiv.org*. Disponible a: <https://arxiv.org/abs/2304.08291> (Accés: 15 Febrer 2024).
- [7] Lin, X. *et al.* (2023) "XPixGroup/Diffbir: Official Codes of diffbir: Towards blind image restoration with generative diffusion prior", *GitHub*. Disponible a: <https://github.com/XPixGroup/DiffBIR/tree/main> (Accés: 20 Febrer 2024).
- [8] Luo, Z. *et al.* (2023) "ALGOLZW/Image-Restoration-SDE: Image Restoration with mean-reverting stochastic differential equations, *ICML 2023. winning solution of the NTIRE 2023 image shadow removal challenge*.", *GitHub*. Disponible a: <https://github.com/Algolzw/image-restoration-sde?tab=readme-ov-file> (Accés: 21 Febrer 2024).
- [9] Rombach, R. *et al.* (2021) "COMPMVIS/stable-diffusion: A latent text-to-image diffusion model", *GitHub*. Disponible a: <https://github.com/CompVis/stable-diffusion> (Accés: 09 Març 2024).
- [10] Goyal, S. (2019) "Gans-a brief introduction to generative Adversarial Networks, *Medium*." Disponible a: <https://medium.com/analytics-vidhya/gans-a-brief-introduction-to-generative-adversarial-networks-f06216c7200e> (Accés: 18 Febrer 2024).
- [11] Zhu, D. (2020) "Generate images using variational Autoencoder (VAE), *Medium*." Disponible a: <https://medium.com/@judyys10/generate-images-using-variational-autoencoder-vae-4d429d9bdb5> (Accés: 18 Febrer 2024).

ANNEX

ID	Objectiu	Descripció
0	Entendre	Guanyar una comprensió exhaustiva del funcionament dels models de difusió i en concret del codi escollit.
1	Primer entrenament	Entrenar el model per dur a terme la tasca <i>deshadow</i> , és a dir, donada una imatge eliminar les ombres que hi apareguin. El dataset previst per aquest entrenament és l'SRD format per 3960 imatges i les mètriques a recollir: PSNR, SSIM, LPIPS i MSE. En futurs informes es profunditzarà més en aquests detalls.
2	Integració de condicionants externs (<i>prompts</i>).	Incorporar un condicionant extern a la tasca, per començar en format de text <i>prompt</i> i posteriorment, utilitzant directament una imatge com a condicionant. Pel que fa al <i>dataset</i> i les mètriques, el primer està per determinar i les mètriques es mantenen en totes les tasques, amb la possibilitat d'incloure'n de noves sempre que sigui necessari.
3	Entrenament i anàlisi de resultats en la tasca final.	Entrenar el model per la tasca de <i>Relighting</i> , condicionada per imatge, i analitzar de manera detallada els resultats obtinguts. Aquest entrenament està pensat pels datasets ISR i RSR, proposats per Yixiong Yang et al. [1].

Figura 1. Taula d'objectius.

Setmana	Data/es	ID Objectiu/s	Assoliments
1	05/02 - 11/02	0	Familiarització amb conceptes bàsics de SD [2], [3], [4] i l'article de relighting [1].
2	12/02 - 18/02	0	Millor comprensió dels models SD. Comprensió del context actual de SD i concretament dels models SD de restauració d'imatges [5], [6].
3	19/02 - 25/02	0	Consolidació de coneixements base en SD. Proves dels codis més rellevants i afins als projectes trobats. [7], [8], [9] Tria del codi més adequat de cara a la tasca final.
4	26/02 - 03/02	0	Comprensió a trets generals del codi escollit. Primeres proves del funcionament del codi. Detecció i millora dels primers errors trobats.
5	04/03 - 10/03	0, 1	Millor comprensió del procés d'entrenament en el codi. Mètriques a recollir escollides. Wandb en funcionament per recollir aquestes mètriques.
10/03/2024		Informe Inicial	

Setmana	Data/es	ID Objectiu/s	Assoliments
6	11/03 - 17/03	0, 1	-
7	18/03 - 24/03	0, 1	-
8	25/03 - 31/03	0, 1	-
9	01/04 - 07/04	0, 1	-
10	08/04 - 14/04	0, 1	-
11	15/04 - 21/04	0, 1	-
21/04/2024		Informe de Progrés I	
12	22/04 - 28/04	0, 2	-
13	29/04 - 05/05	0, 2	-
14	06/05 - 12/05	0, 2	-
15	13/05 - 19/05	0, 2	-
16	20/05 - 26/05	0, 2	-
26/05/2024		Informe de Progrés II	
17	27/05 - 02/06	0, 3	-
18	03/06 - 09/06	0, 3	-
19	10/06 - 16/06	0, 3	-
16/06/2024		Proposta Informe Final	
20	17/06 - 23/06	0, 3	-
21	24/06 - 30/06	0, 3	-
30/06/2024		Proposta de Presentació	
30/06/2024		Dossier i Informe Final	
01/07/2024		Pòster	
03/07/2024-10/07/2024		Defensa TFG	

Figura 2. Taula de planificació del projecte.