

Generació de Retrats Robots i Identificació de Persones



Luis Lebron

Senior Machine Learning Engineer

Fundació i2CAT

01

Introducció

02

Objectius

03

Enfocament

04

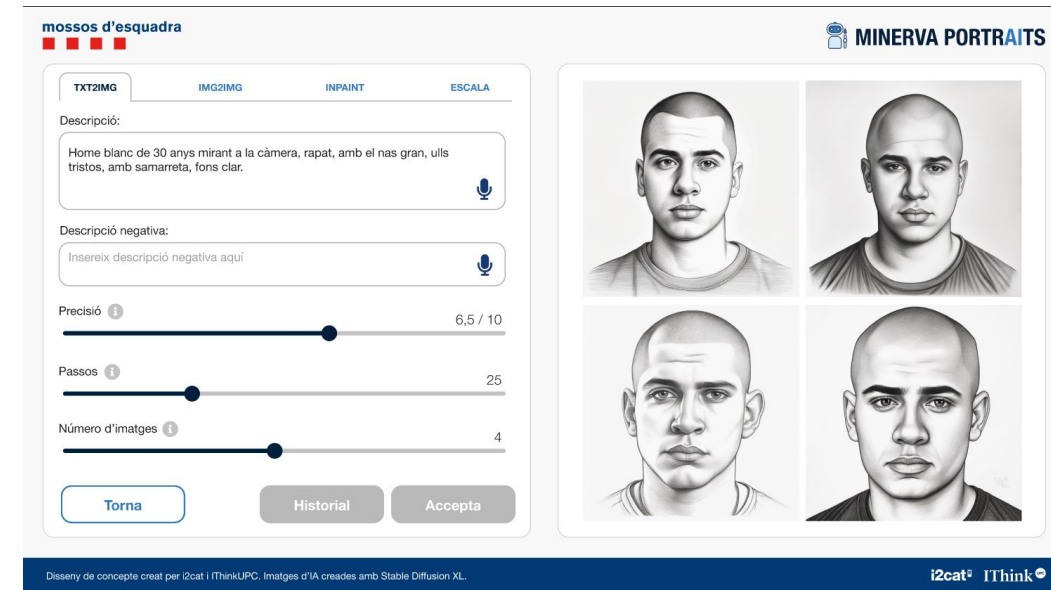
Resultats

05

Conclusions

Retrats Robot amb IA: Automatització dels processos

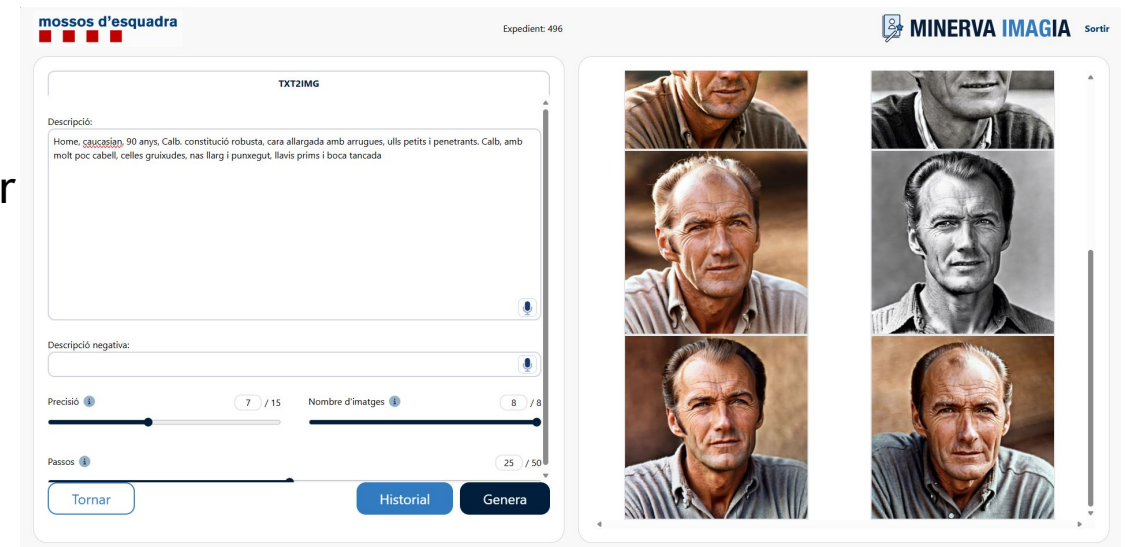
- Optimització de recursos: Eliminar el coll d'ampolla de la unitat centralitzada.
- Agilitat en la investigació: Eines precises per a l'ús policial
- Protecció de la víctima: Minimitzar el desgast psicològic i la revictimització.



Objectiu del Pilot

Desenvolupar una eina d'IA generativa integrada en el flux de treball de la policia científica, garantint un entorn segur i fluid.

- Mòdul PORTRAITS (Generació d'esbossos)
- Mòdul IMAGIA (Identificació i anàlisi)



Mòdul PORTRAITS (Generació d'esbossos)

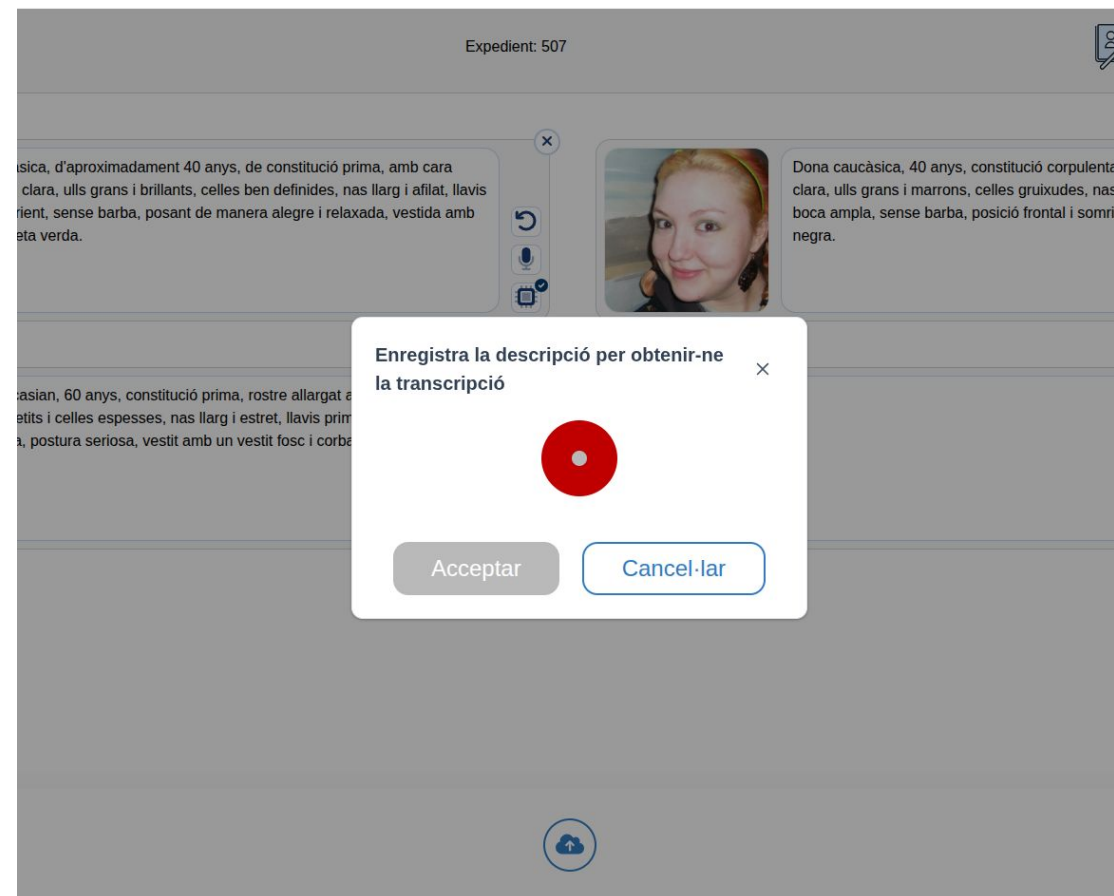
- Generació assistida mitjançant descripcions detallades (Text-to-Image).
- Refinament precís amb tècniques d'In-Painting per modificar trets específics (nas, ulls, etc.) de forma instantània.

Mòdul IMAGIA (Identificació i anàlisi)

- Recreació de subjectes a partir de poques fotografies d'entrenament.
- Funcions avançades d'envelliment/rejuveniment, canvis en condicions físiques, millora de la qualitat d'imatge i renderitzat 3D.

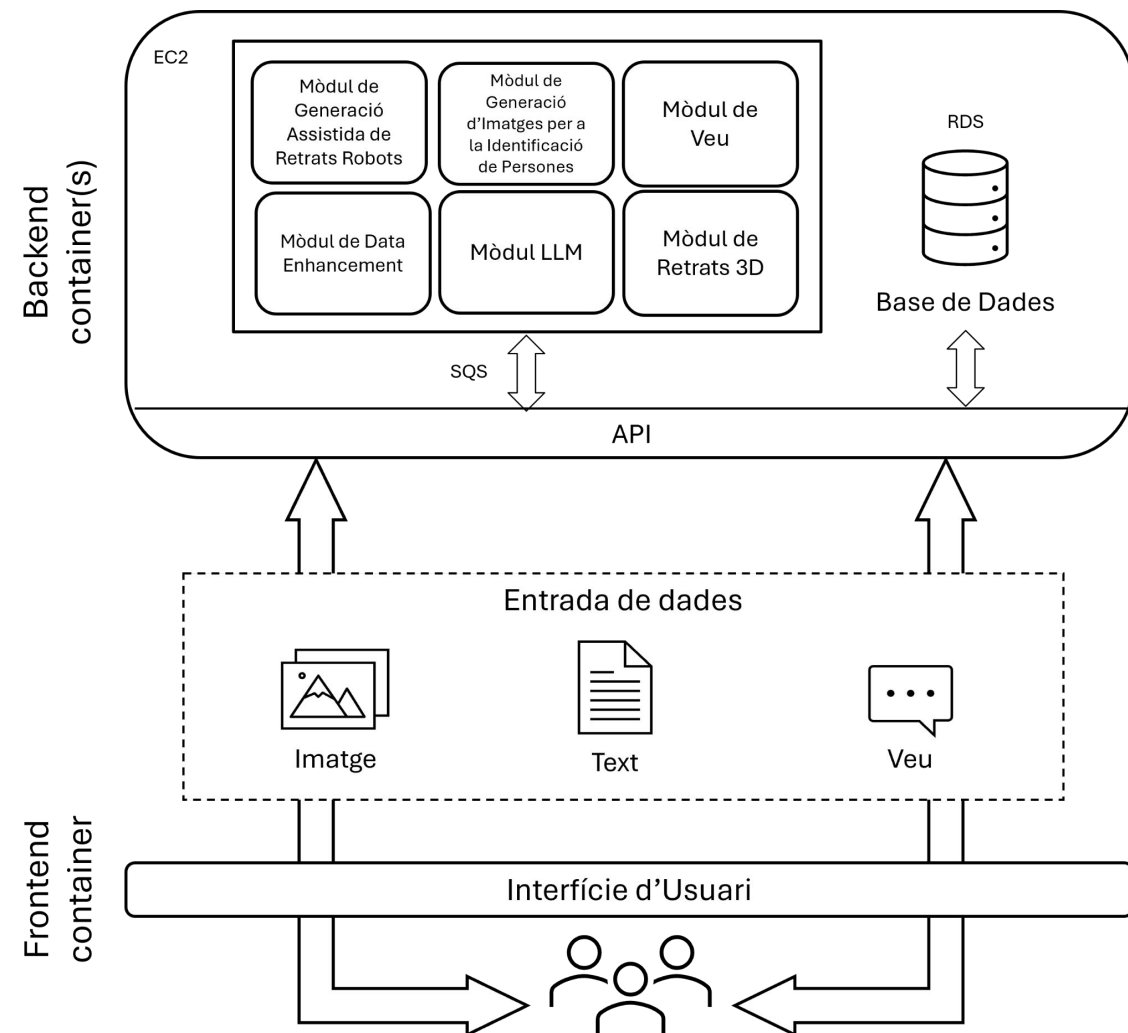
Assistent de veu multi-idioma

- Interfície transversal que permet una interacció natural i elimina les barreres tecnològiques per als investigadors.



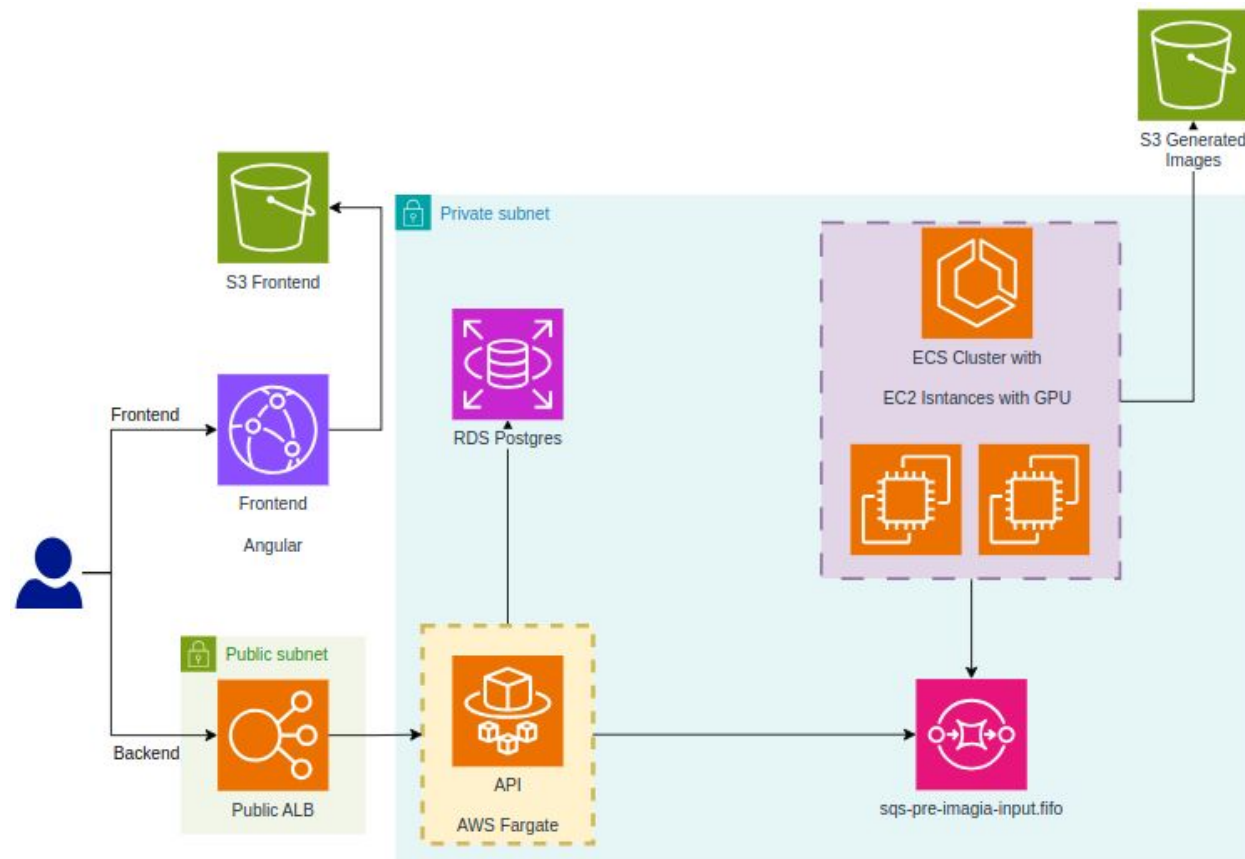
General

- Recepció i emmagatzematge
- Encuament
- Processament (Consumidor)
- Notificació i neteja



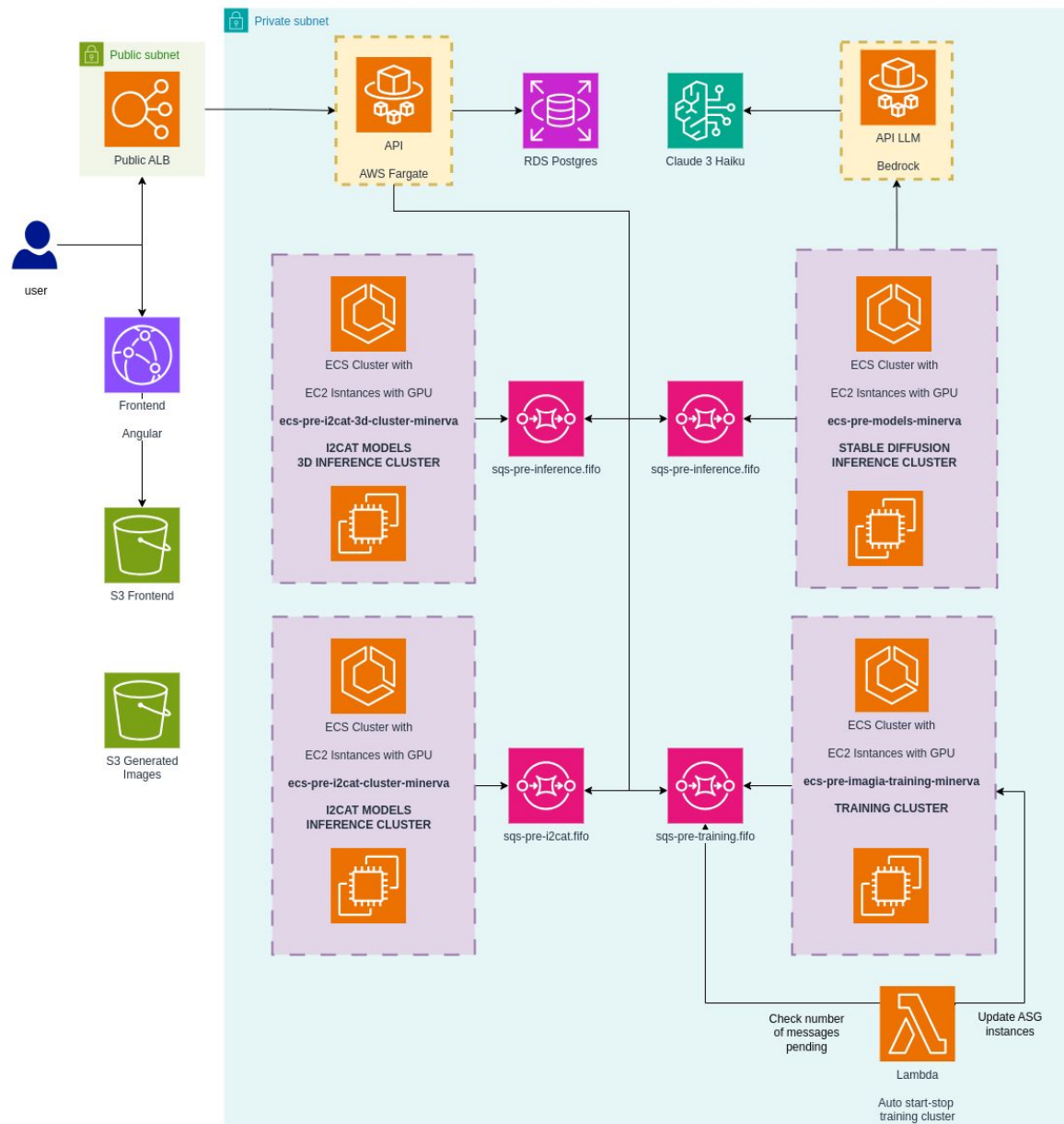
Portraits

- Inferència d'imatges amb Stable Diffusion XL
 - TXT2IMG
 - IMG2IMG
 - INPAINT
- Processament de Veu



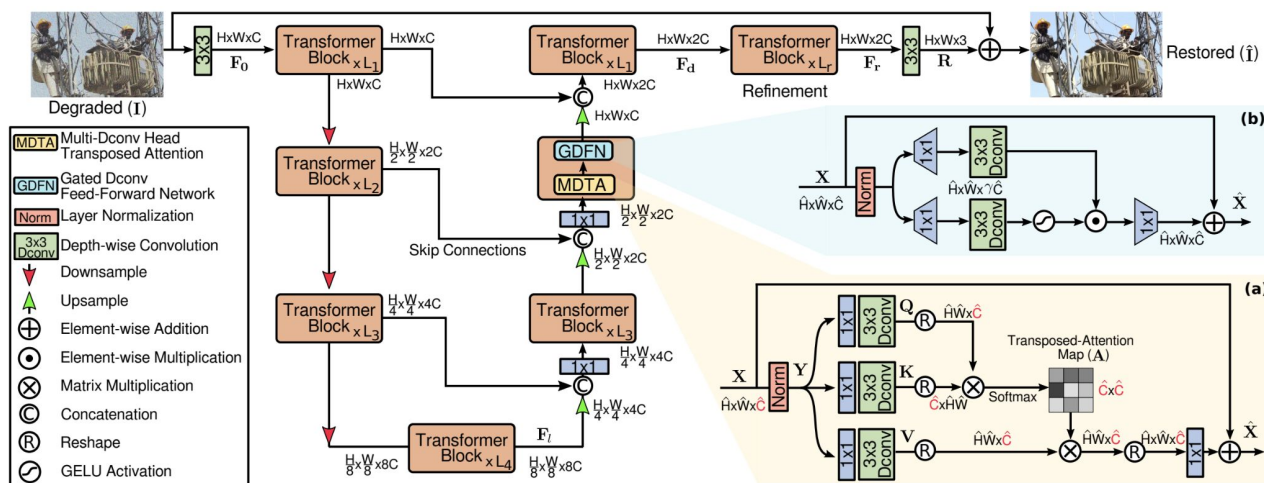
Imagia

- Data Enhancement
- Entrenament de Models (LoRA)
- Inferència amb Model Personalitzat
- Retrats 3D

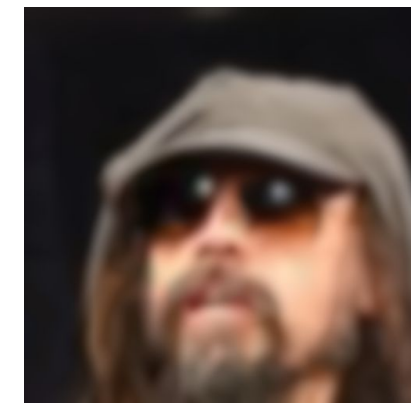


Com les imatges poden provenir de diverses fonts, la qualitat no és sempre l'òptima.

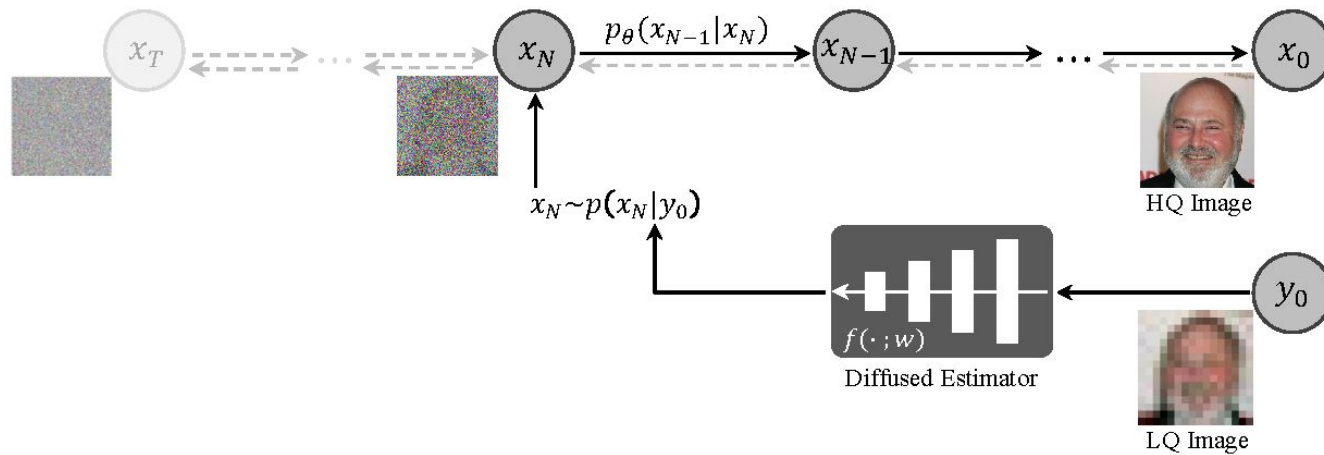
Per resoldre aquest problema, apliquem opcionalment tècniques de image enhancement per a solucionar possibles problemes de desenfocament o soroll.



Liang, Jingyun, et al. "Swinir: Image restoration using swin transformer." Proceedings of the IEEE/CVF international conference on computer vision. 2021.



Per a imatges més malmeses, apliquem un Diffusion Model que tenen la capacitat de complementar la imatge original amb informació addicional.



Yue, Zongsheng, and Chen Change Loy. "Difface: Blind face restoration with diffused error contraction." IEEE Transactions on Pattern Analysis and Machine Intelligence (2024).



Per a entrenar els models de StableDiffusionXL amb LoRa cal generar captions per lo que amb Claude generem descripcions similars a la que demanem als usuaris.

mossos d'esquadra Expedient: 496 **MINERVA IMAGIA** Sortir

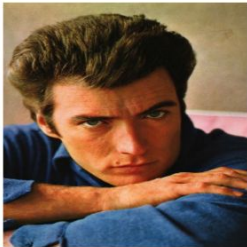
Historial de l'expedient de treball

IMAGIA



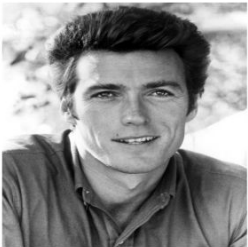
n, 60 anys, constitució allargada amb arrugues, tez petits i penetrants, celles arg i punxegut, llavis primis, barba espessa, postura t, vestit amb barret de

E3 IMAGIA



Descripció
Home, caucasià, 35 anys, fit, rostre allargat, pell clara, ulls grans i blaus, celles gruixudes, nas llarg i recte, llavis primis, sense barba, mirada directa, camisa a

E4 IMAGIA



Descripció
Home, caucasià, 35 anys, constituït, rostre allargat amb arrugues, tez clara, ulls grans i brillants, celles gruixudes, nas llarg i recte, llavis primis i boca somrient, sense barba, postura relaxada, vestit amb camisa

E5 IMAGIA



Descripció
Home, caucasià, 35 anys, constitució atlètica, rostre oval, pell clara, ulls blaus, celles gruixudes, nas recte i allargat, llavis primis, sense barba, postura relaxada,

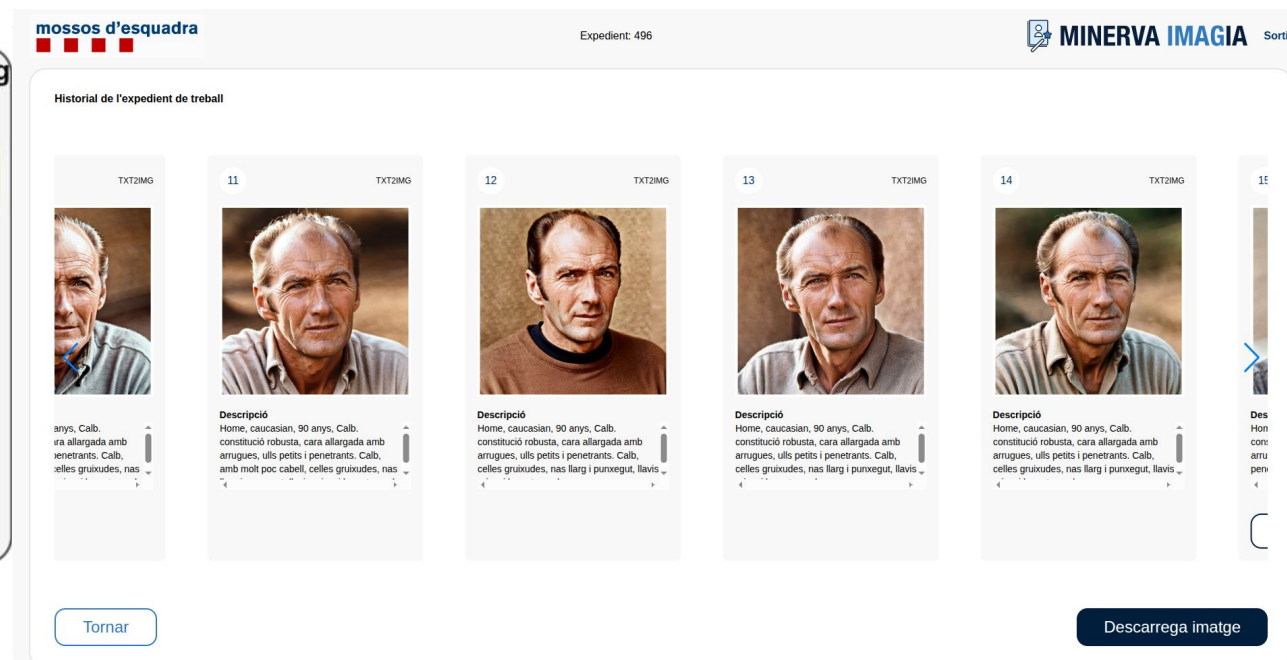
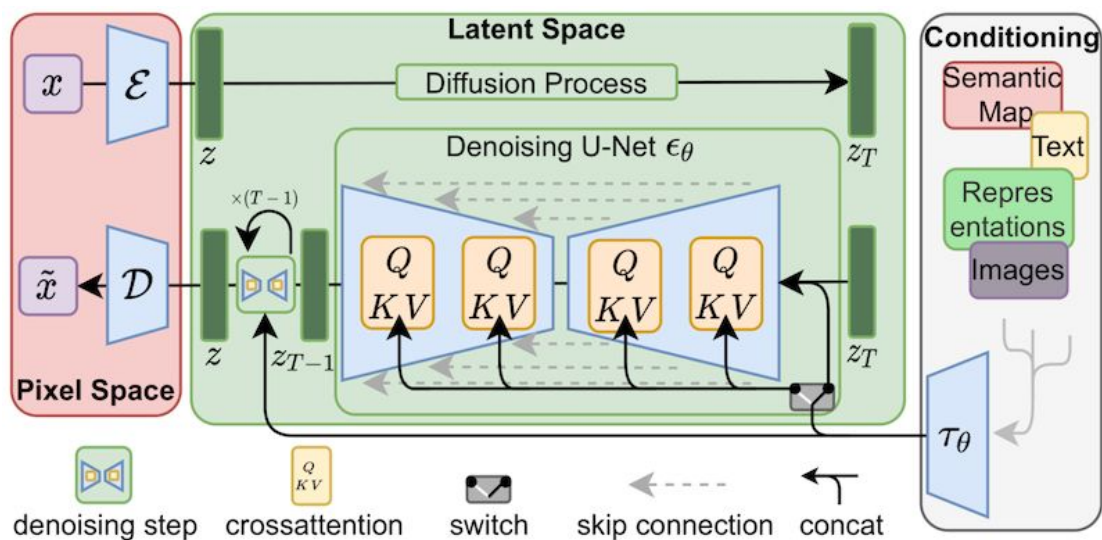
E6 IMAGIA



Descripció
Home, caucasià, 35 anys, constituït, rostre allargat amb arrugues, tez clara, ulls grans i brillants, celles gruixudes, nas llarg i recte, llavis primis i boca somrient, sense barba, postura relaxada, vestit amb camisa

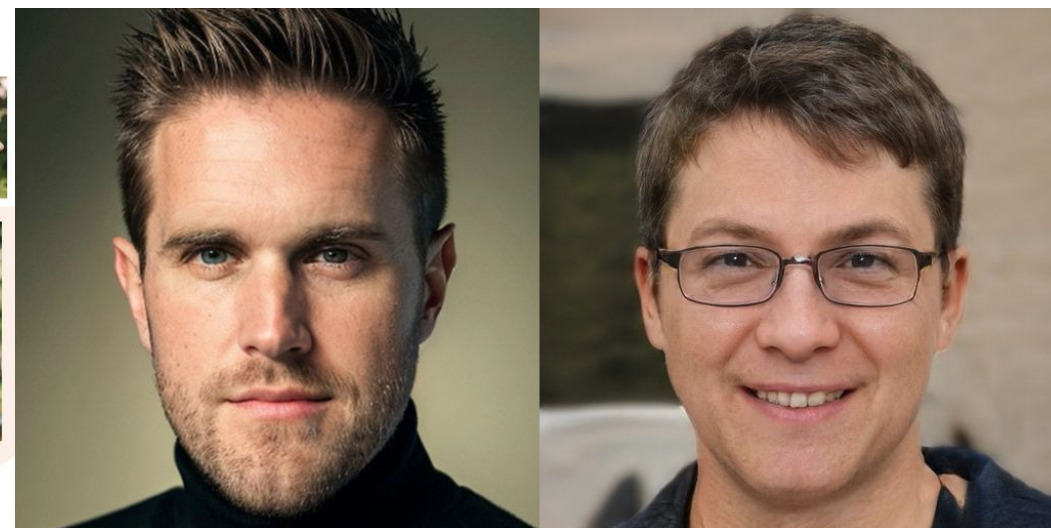
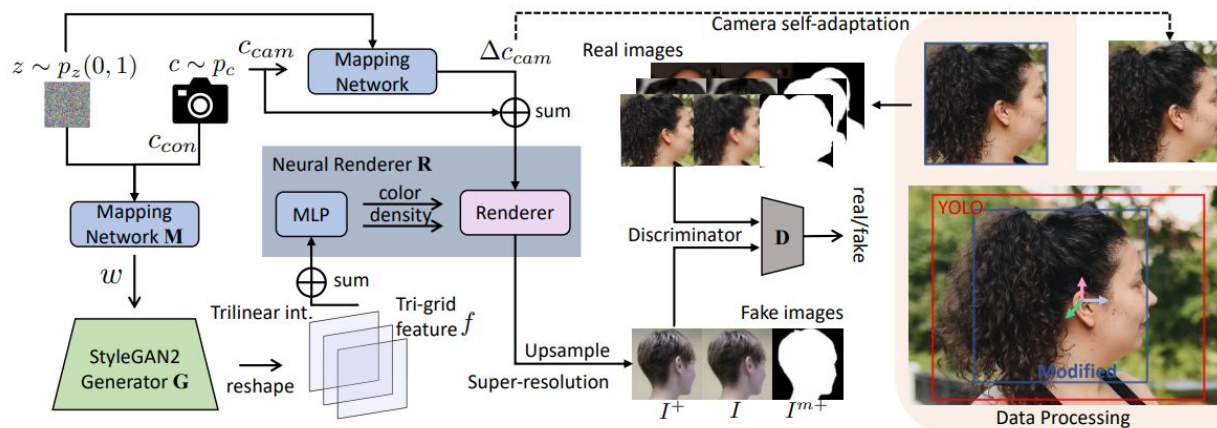
Tornar Descarrega imatge

Amb les imatges originals o millorades i la descripció, s'adapta un model de StableDiffusionXL2 amb LoRa1 per a generar un nou conjunt d'imatges alterades de la persona, e.g. més vella, amb un altre tallat de cabells, etc.



Hu, Edward J., et al. "Lora: Low-rank adaptation of large language models." ICLR 1.2 (2022): 3.

Per enriquir la visualització de les imatges generades, utilitzem un model per traduir una imatge frontal en una representació 3D del cap.



An, Sizhe, et al. "Panohead: Geometry-aware 3d full-head synthesis in 360deg." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.

L'avaluació exhaustiva dels models s'ha quantificat tenint en compte els temps de resposta en inferència/entrenament i aplicant mètriques de qualitat visual rigoroses per als models de Data Enhancement.

- En el procés de generació iterativa d'imatges amb Stable Diffusion XL s'ha comprovat els següents temps:
 - Generació d'1 imatge: ~11 segons.
 - Generació d'un panell de 4 imatges: ~44 segons.
 - Entrenament complet d'un model LoRApersonalitzat: Entre 4 i 6 hores.

Pel que fa a la qualitat perceptual i al rendiment tècnic dels models de Data Enhancement, els resultats han evidenciat el següent:

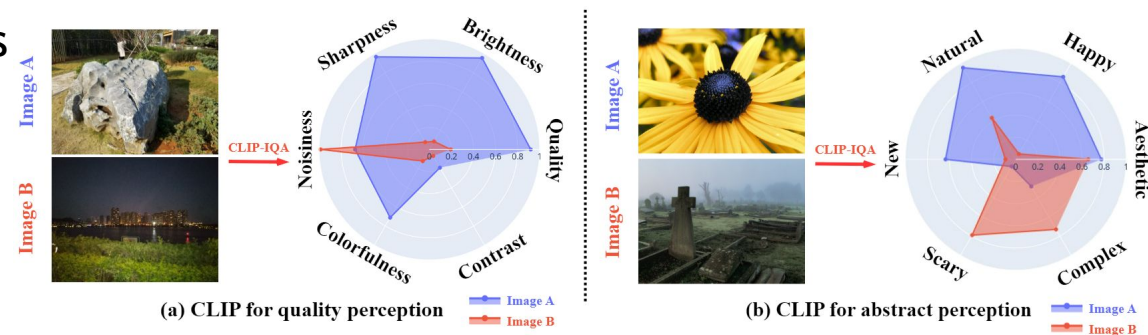
Model	# Params	VRAM	CLIP-IQA↑	SSIM↑	Temps (per imatge)
SwinIR	16 M	8GB	19.02	0.71	4 seg
DiffFace	176 M	8GB	56.41	0.67	14 seg

Taula 1: Resultats utilitzan la base de dades de CelebA per a calcular el rendiment dels models de Data Enhancement en un GPU A6000.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

on:

- μ_x i μ_y són la luminància mitjana de les dues imatges que s'estan comparant.
- σ_x^2 i σ_y^2 són les variàncies de les intensitats dels píxels a les dues imatges.
- σ_{xy} és la covariància de les intensitats dels píxels entre les dues imatges.
- C_1 i C_2 són constants afegides per prevenir la inestabilitat quan els denominadors estan prop de zero.



DifFace (Risc d'Al·lucinacions)

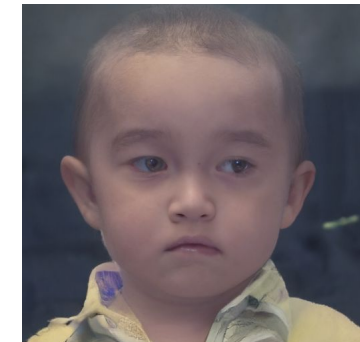
- Podria inventar i introduir de manera artificial artefactes i detalls facials minúsculs que no estaven presents a la fotografia original, alterant potencialment elements identificatius clau.

Panohead(3D Angles cecs)

- El model té una forta tendència a generar detalls incongruents a les zones o parts del rostre i cap que romanien amagats a la fotografia original, com per exemple la forma posterior de les orelles o la disposició exacta del pentinat al clatell

Stable Diffusion XL (LoRA i perfils sensibles)

- S'ha observat que si a l'eina se li apliquen més de 60 passos de generació, la imatge resultant perd l'estètica funcional d'esbós i la fidelitat cau cap a una "sobre-definició" innecessària.



Conclusions

- **Viabilitat del Model:**

Èxit en la integració de models d'IA generativa multimodal en l'entorn policial, demostrant ser una eina pràctica i altament intuïtiva.

- **Infraestructura Escalable:**

L'arquitectura sobre AWS garanteix alta disponibilitat i la capacitat d'estendre el sistema a múltiples comissaries sense risc de col·lapse.

- **Precisió amb IMAGIA:**

Fidelitat en l'adaptació de rostres mitjançant entrenaments LoRA sobre espais latents (òptim amb fotografies inicials tipus "carnet").

- **Eficiència Operativa:**

Reducció dràstica en els temps de confecció dels retrats, agilitzant la investigació policial.

- **Impacte Humanitari:**

Millora directa en el confort psicològic dels testimonis i víctimes en reduir la durada i l'estrès de les sessions.

Possibles millores

- **Preprocessament Automàtic:**

Integració d'IA per a la detecció facial i l'enquadrament intel·ligent, eliminant la preparació manual d'imatges.

- **Optimització per a imatges CCTV:**

Millora de l'encoder per processar amb precisió fotografies forenses de baixa qualitat (soroll, mala il·luminació o angles complexos).

- **Correcció de biaixos:**

Refinament del model per garantir la fidelitat i representació exacta de totes les ètnies i gèneres, eliminant limitacions dels conjunts de dades.

Moltes gràcies!

Luis Lebron Casas
*Senior Machine Learning
Engineer*
luis.lebron@i2cat.net