LICDIA
Laboratorio de Investigación en Ciencia de Datos & Inteligencia Artificial

Línea de I+D - Gobernanza y transformación digital

JORNADAS 2026 DE CIENCIA, TECNOLOGÍA Y EXTENSIÓN
DEPARTAMENTO DE CIENCIAS BÁSICAS

Restauración de Fachadas Patrimoniales mediante Modelos Generativos de Difusión y Transformers Multimodales

Leandro Marina[1] - Malek Camilo[1] - Juan Manuel Fernández[1,2]

[1] LICDIA, Departamento de Ciencias Básicas, Universidad Nacional de Luján - [2] ITBA
leandro.marina88@gmail.com - mcamilo@unlu.edu.ar - jmfernandez@{unlu, itba}.edu.ar

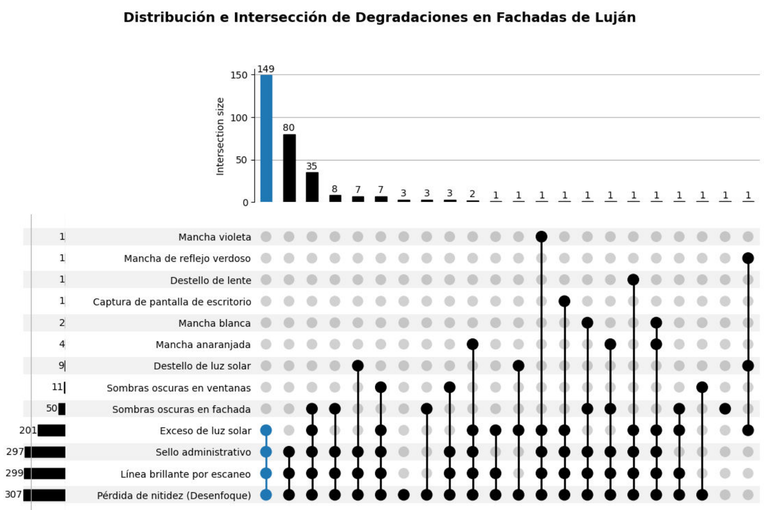
Los Transformers Multimodales con flujo rectificado (SD3 y SD3.5) y desacople temporal superan a Pix2Pix y SDXL, eliminando sellos y manchas del archivo patrimonial de Luján sin deformar la geometría arquitectónica.

1 Problema

Las fotografías escaneadas del relevamiento patrimonial de Luján presentan deterioros compuestos: desenfoque, sellos administrativos, líneas de escaneo y manchas. Los modelos generativos pueden reconstruir texturas dañadas, pero corren el riesgo de inventar detalles arquitectónicos inexistentes, un error documental inaceptable para registros históricos.

2 Degradaciones

Relevamos 309 fotos reales de Luján (13 degradaciones) y sintetizamos 1380 pares (512×512 px) sobre CMP Facades.

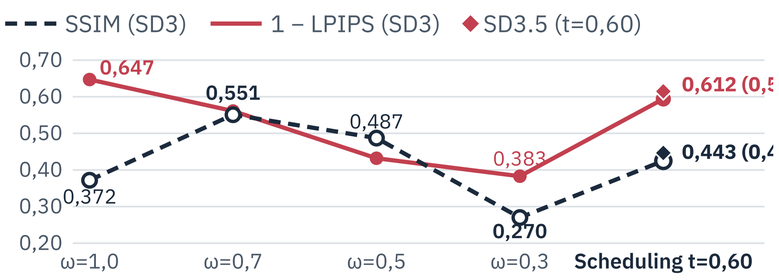


3 Comparativa inter-técnica

Modelo	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓	VRAM (GB)	Latencia (s)
Entrada degradada	28,39 ± 0,74	0,944 ± 0,008	0,046 ± 0,007	12,99	—	—
Pix2Pix (cGAN) 200 épocas	27,83 ± 0,28	0,917 ± 0,005	0,108 ± 0,005	16,41	0,89	0,014
SDXL UNet + ControlNet 200 épocas	16,97 ± 0,24	0,463 ± 0,002	0,578 ± 0,004	144,29	19,04	8,25
SD3 MMDiT + ControlNet 150 épocas, t_{corte} = 0,60	**13,42** ± 0,22	**0,425** ± 0,007	**0,406** ± 0,004	**92,42**	**15,66**	**12,16**
SD3.5 MMDiT + ControlNet 125 épocas, t_{corte} = 0,60	**13,70** ± 0,20	**0,443** ± 0,009	**0,388** ± 0,005	**92,86**	**18,49**	**14,28**

Evaluación sobre conjunto de prueba (n = 144, 512×512 px) en FP32 (BS = 16) en Clementina XXI.

4 Impacto del desacople temporal



Escalas estáticas altas copian los defectos; escalas bajas (ω = 0,3) deforman la fachada (SSIM 0,270). Desactivar ControlNet en el 40% final (t_{corte} = 0,60) erradica sellos y manchas preservando fidelidad perceptual (1 – LPIPS: 0,594 en SD3 y 0,612 en SD3.5).

5 Evidencia visual en Luján

Antes (Fachada degradada) | **Después (Fachada restaurada)**



Ilustración del esquema de restauración: remoción de mancha anaranjada y veladura superficial conservando íntegra la geometría arquitectónica original.

6 Conclusiones

GEOMETRÍA
MMDiT preserva la arquitectura
Flujo rectificado en 16 canales mantiene plomos, vanos y cornisas sin distorsiones.

CONTROL
El scheduling evita copiar defectos
Desactivar ControlNet en el 40% final borra sellos y manchas sin alucinar.

CÓMPUTO
Compromiso entre calidad y latencia
Pix2Pix infiere en 0,014 s (0,89 GB); SD3/SD3.5 exigen 15,66–18,49 GB pero superan a SDXL (LPIPS 0,388 vs. 0,578).

Sigue: incorporar condicionamiento geométrico mediante mapas de normales, destilar pasos de inferencia para uso interactivo y realizar validación a ciegas con arquitectos especialistas en patrimonio.

Referencias
1. Isola et al. Image-to-Image Translation with cGANs. CVPR, 2017.
2. Rombach et al. High-Resolution Image Synthesis with LDMs. CVPR, 2022.
3. Podell et al. SDXL: Improving Latent Diffusion Models. ICLR, 2024.
4. Esser et al. Scaling Rectified Flow Transformers. ICML, 2024.
5. Meng et al. SDEdit: Guided Image Synthesis with SDEs. ICLR, 2022.

Agradecimientos
Agradecemos al archivo patrimonial del Municipio de Luján y al Sistema Nacional de Computación de Alto Desempeño (SNCAD) por el acceso a Clementina XXI (proyecto IPAC PCI-146).

Repositorio

Líneas de I+D - Investigación aplicada en ciencia de datos e IA, con fuerte impronta en los problemas regionales

Modelos de lenguaje abiertos | IA en educación | Diálogo y RAG | **Gobernanza y transformación digital** | Blockchain y credenciales | Imágenes y datos geoespaciales

licdia.unlu.edu.ar