

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
FACULTAD DE INGENIERIA ELECTRICA ELECTRONICA INFORMATICA Y
MECANICA

ESCUELA PROFESIONAL DE INGENIERIA INFORMATICA Y DE SISTEMAS



TESIS

**ADAPTACION DE TÉCNICAS DE APRENDIZAJE PROFUNDO
PARA LA RESTAURACIÓN DIGITAL VELOZ DE PINTURAS SIN
REGISTROS FOTOGRÁFICOS**

PRESENTADO POR:

Br. GUADALUPE VERENISE CAVERO MANZANARES

Br. CATHERINE LIZBED LOAYZA CONDOR

**PARA OPTAR AL TÍTULO PROFESIONAL DE
INGENIERO INFORMÁTICO Y DE SISTEMAS**

ASESOR:

Dr. LAURO ENCISO RODAS

CUSCO – PERÚ

2024

INFORME DE ORIGINALIDAD

(Aprobado por Resolución Nro.CU-303-2020-UNSAAC)

El que suscribe, Asesor del trabajo de investigación/tesis titulada: ADAPTACION DE TECNICAS DE APRENDIZAJE PROFUNDO PARA LA RESTAURACION DIGITAL VELOZ DE PINTURAS SIN REGISTROS FOTOGRAFICOS

presentado por: GUADALUPE VERENISE CAÑERO MANZANARES con DNI Nro.: 72424072 presentado por: CATHERINE LIZBETH LOAYZA CONDOR con DNI Nro.: 48130305 para optar el título profesional/grado académico de INGENIERO INFORMATICO Y DE SISTEMAS

Informo que el trabajo de investigación ha sido sometido a revisión por 3 veces, mediante el Software Antiplagio, conforme al Art. 6° del **Reglamento para Uso de Sistema Antiplagio de la UNSAAC** y de la evaluación de originalidad se tiene un porcentaje de 3%.

Evaluación y acciones del reporte de coincidencia para trabajos de investigación conducentes a grado académico o título profesional, tesis

Porcentaje	Evaluación y Acciones	Marque con una (X)
Del 1 al 10%	No se considera plagio.	X
Del 11 al 30 %	Devolver al usuario para las correcciones.	
Mayor a 31%	El responsable de la revisión del documento emite un informe al inmediato jerárquico, quien a su vez eleva el informe a la autoridad académica para que tome las acciones correspondientes. Sin perjuicio de las sanciones administrativas que correspondan de acuerdo a Ley.	

Por tanto, en mi condición de asesor, firmo el presente informe en señal de conformidad y adjunto la primera página del reporte del Sistema Antiplagio.

Cusco, 07 de ENERO de 2025



Firma

Post firma Laura Enciso Rodas

Nro. de DNI 23853228

ORCID del Asesor 0000-000162660838

Se adjunta:

1. Reporte generado por el Sistema Antiplagio.
2. Enlace del Reporte Generado por el Sistema Antiplagio: oid: 27259:419454622/

Verenice Lisbed Catherine Loayza, 4-Turnitin-TESIS-ADAPTACION-DE-TECNICAS-DE-APRENDIZ...

 Universidad Nacional San Antonio Abad del Cusco

Detalles del documento

Identificador de la entrega

trn:oid:::27259:419454622

Fecha de entrega

6 ene 2025, 6:59 a.m. GMT-5

Fecha de descarga

6 ene 2025, 9:59 p.m. GMT-5

Nombre de archivo

4-Turnitin-TESIS-ADAPTACION-DE-TECNICAS-DE-APRENDIZAJE-PROFUNDO-PARA-LA-RESTAURACI....pdf

Tamaño de archivo

1.6 MB

85 Páginas

25,304 Palabras

151,293 Caracteres




3% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- ▶ Bibliografía
- ▶ Texto citado
- ▶ Coincidencias menores (menos de 15 palabras)

Fuentes principales

- 1%  Fuentes de Internet
- 0%  Publicaciones
- 3%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Dedicatoria

A nuestros padres, cuyo amor, sacrificio y dedicación nos han guiado hasta este momento. Su ejemplo nos ha inspirado a alcanzar nuestras metas. A ustedes, nuestra eterna gratitud y amor.

Agradecimientos

Queremos expresar nuestro agradecimiento al Doctor Lauro Enciso Rodas por su contribución y guía durante este proceso. Su dedicación y conocimiento fueron una inspiración constante para nosotras.

Resumen

La tarea de la restauración digital de imágenes (Inpainting) supone una tarea fundamental dentro del campo de la visión computacional. En este trabajo se proponen adaptaciones a técnicas de aprendizaje profundo para la restauración digital de imágenes teniendo como objetivo mejorar la velocidad de ejecución. Luego de considerar un conjunto de técnicas, a partir de los criterios de ejecutabilidad, adaptabilidad, uniformidad y comparabilidad se seleccionan LaMa y EdgeConnect para su posterior adaptación. En la experimentación, se utiliza una base de datos de pinturas clásicas, cada una de las cuales es sometida a una distorsión uniforme y subsiguiente restauración con las técnicas seleccionadas y todas sus adaptaciones propuestas. Finalmente, de acuerdo con el análisis experimental a partir de los datos estadísticos obtenidos, la razón entre la diferencia de tiempo y la diferencia de calidad es de al menos 0.06 (6%), con lo cual hay una mejora del tiempo con relación a la pérdida de calidad. Esta mejora es constante en las técnicas finales seleccionadas, contribuyendo al avance de las aplicaciones de visión por computadora en restauración digital y su posterior uso en la restauración tradicional de pinturas sin registro fotográfico.

Palabras clave: *Restauración, Redes Neuronales Convolucionales, Inpainting, Deep Learning, Aprendizaje Automático, Tiempo de ejecución, Visión computacional, LaMa, EdgeConnect, Adaptación de técnicas*

Abstract

Digital image restoration (inpainting) is a fundamental task in computer vision. In this work we propose adaptations to deep learning techniques for digital image restoration to improve execution speed. After considering a set of techniques, based on the criteria of executability, adaptability, uniformity, and comparability, LaMa and EdgeConnect are selected for further adaptation. In the experimentation, a database of classical paintings is used, each of which is subjected to uniform distortion and subsequent restoration with the selected techniques and all their proposed adaptations. Finally, according to the experimental analysis from the statistical data obtained, the ratio between the time difference and the quality difference is at least 0.06 (6%), so there is an improvement in time about the loss of quality. This improvement is constant in the final techniques selected, contributing to the advancement of computer vision applications in digital restoration and its subsequent use in the traditional restoration of paintings without photographic records.

Keywords: *Restoration, Convolutional Neural Networks, Inpainting, Deep Learning, Machine Learning, Runtime, Computational Vision, LaMa, EdgeConnect, Adaptive of techniques*

Índice

Dedicatoria	2
Agradecimientos	3
Resumen	4
Abstract.....	5
Índice	6
Índice de Figuras	9
Índice de Tablas.....	10
Introducción.....	11
Aspectos Generales	13
1.1 Problema de investigación	13
1.1.1 Descripción del problema	13
1.1.2 Identificación del problema.....	14
1.1.3 Problemas específicos	14
1.2 Antecedentes	14
1.2.1 Context Encoders: Aprendizaje de características para restauración digital	14
1.2.2 MSNP: Restauración digital de alta resolución usando Síntesis de Parches Neuronales Multiescala.....	15
1.2.3 GLIC: Completación global y local de imágenes consistentemente.....	15
1.2.4 Shift-Net: Restauración digital de imágenes a través de Deep Features.....	15
1.2.5 GICA: Restauración digital de imágenes con atención contextual	16
1.2.6 FFGC: Restauración digital Free-Form con una convolución cerrada	16
1.3 Objetivos	17
1.3.1 Objetivo general	17
1.3.2 Objetivos específicos	17
1.4 Alcances y Limitaciones	18
1.5 Metodología	19
1.6 Justificación	20
Marco Teórico	22
2.1 Restauración de Pinturas.....	22
2.1.1 ¿Por qué es necesario restaurar una pintura?	23
2.1.2 Principios de la restauración	23
2.1.3 ¿Cómo se lleva a cabo la restauración de pinturas?	23

2.1.3.1. Examen preliminar y diagnóstico	24
2.1.3.2. Documentación	24
2.1.3.3. Reparación del soporte	24
2.1.3.3. Limpieza	24
2.1.3.4. Consolidación	24
2.1.3.5. Retención de la intervención mínima	24
2.1.3.6. Reintegración cromática	25
2.1.3.7. Barnizado final	25
2.2 Inteligencia Artificial	25
2.2.1 ¿Qué es la Inteligencia Artificial?	25
Motores de búsqueda	27
Sistemas de recomendación	28
Asistentes virtuales	29
2.2.2 Subáreas de la investigación en Inteligencia Artificial	32
2.3 Aprendizaje Automático	33
2.3.1 Aprendizaje no supervisado	37
2.3.2 Aprendizaje supervisado	40
2.3.2.1 Regresión	42
2.3.2.2 Clasificación	44
2.3.3 Aprendizaje semi-supervisado	46
2.3.4 Retos del aprendizaje automático	48
2.4 Aprendizaje profundo	50
2.4.1 Redes Neuronales Convolucionales	50
2.4.2 Autoencoders	52
2.4.3 BrushNet (Ju et al., 2024)	53
2.4.4 MAE: Aprendizaje de características previas y mejora de la atención para la restauración de imágenes (Cao Chenjieand Dong, 2022)	55
2.4.5 LaMa: Restauración de Máscaras Grandes Resistente a la Resolución con Convoluciones de Fourier (Suvorov et al., 2021)	57
2.4.6 EdgeConnect: Relleno Generativo de Imágenes con Aprendizaje Adversarial de Bordes (Nazeri et al., 2019)	58
Desarrollo del Proyecto	61
3.1. Análisis para la selección	61
3.1.1. BrushNet (Ju et al., 2024)	61

3.1.2. MAE: Aprendizaje de características previas y mejora de la atención para la restauración de imágenes (Cao Chenjieand Dong, 2022)	62
3.1.3. LaMa: Restauración de Máscaras Grandes Resistente a la Resolución con Convoluciones de Fourier (Suvorov et al., 2021)	64
3.1.4. EdgeConnect: Relleno Generativo de Imágenes con Aprendizaje Adversarial de Bordes (Nazeri et al., 2019)	66
3.2. Selección de técnicas en base a las características de interés	66
3.3. Propuesta de adaptación.....	68
3.3.1. Adaptaciones propuestas para LaMa	68
3.3.2. Adaptaciones propuestas para EdgeConnect	69
3.3.3. Proceso de ejecución y experimentación	70
3.4. Resultados experimentales.....	71
3.4.1. Evaluación el error de la restauración - calidad	72
3.5. Análisis y discusión de resultados	74
3.5.1. Interpretación de resultados con Edge Connect.....	75
3.5.2. Interpretación de resultados con LaMa.....	77
3.5.3. Resultados obtenidos en pintura de caballete real.....	80
- Con Edge Connect.....	80
- Con LaMa.....	80
Conclusiones.....	81
Recomendaciones	82
Bibliografía.....	83

Índice de Figuras

Figura 1 Metodología propuesta.....	20
Figura 2 Estudio de una pintura con microscopio USB	22
Figura 3 Comparación de arquitecturas inpainting anteriores y BrushNet.	54
Figura 4 Proceso de la restauración hecha por EdgeConnect.	59
Figura 5 Representación visual de la arquitectura de BrushNet.	62
Figura 6 Diagrama de la arquitectura de MAE.	63
Figura 7 Ejemplo de los resultados de MAE.....	63
Figura 8 Resultados de LaMa Fourier en comparación con otras arquitecturas.	64
Figura 9 Representación de la arquitectura de LaMa Fourier.	65
Figura 10 Resultados de la comparación de LaMa Fourier con otras técnicas de restauración digital de imágenes.	65
Figura 11 Diagrama de la arquitectura de EdgeConnect.....	66
Figura 12 Restauración de ejemplo realizado con la máscara aplicada.	72
Figura 13 Gráfico de porcentaje de ganancia de tiempo de procesamiento comparado con la pérdida de calidad - Edge Connect	76
Figura 14 Gráfico de error de restauración con Edge Connect	76
Figura 15 Gráfico de diferencia de tiempo (D.T.) con Edge Connect	77
Figura 16 Gráfico de porcentaje de ganancia de tiempo de procesamiento comparado con la pérdida de calidad - LaMa	77
Figura 17 Gráfico de error de restauración con LaMa	78
Figura 18 Gráfico de diferencia de tiempo (D.T.) con LaMa	78
Figura 19 Resultado visual de restauración con Edge Connect sin y con adaptaciones ...	80
Figura 20 Resultado visual de restauración con LaMa sin y con adaptaciones	80

Índice de Tablas

Tabla 1 Características de interés de las técnicas consideradas para su modificación.....	68
Tabla 2 Ejemplo del formato de la información obtenida de los experimentos hechos en LaMa con la modificación 1.....	73
Tabla 3 Comparaciones de error y tiempo junto con estadísticas de promedio, desviación estándar, mediana, mínimo y máximo.....	74
Tabla 4 Comparación porcentual de la diferencia de tiempo con respecto al error.	75

Introducción

La restauración profesional tradicional de pinturas de caballete comienza con una evaluación detallada del daño, proceso que combina herramientas avanzadas como luz ultravioleta, infrarroja y radiografías, junto con la experiencia y sensibilidad del restaurador. Según la Teoría de la Restauración de Cesare Brandi, este análisis técnico debe respetar la autenticidad histórica y estética de la obra, asegurando que las intervenciones sean reversibles y no comprometan su significado cultural (Brandi, 1995). Esta tarea se complica en ausencia de registros fotográficos de calidad, lo que obliga a recurrir a archivos históricos, estudios comparativos y colaboraciones interdisciplinarias para reconstruir el estado original de la obra. Identificar daños visibles e imperceptibles, como craqueladuras, decoloración o intervenciones inapropiadas, requiere un enfoque tanto técnico como interpretativo, permitiendo diseñar un plan de intervención que respete el valor histórico de la pintura y asegure su preservación para futuras generaciones (ICOM, 2020).

Gracias a los avances tecnológicos, los restauradores tienen acceso a nuevas herramientas digitales que complementan el proceso tradicional. La restauración digital de imágenes, conocida como inpainting, se ha convertido en una técnica clave en el campo de la visión computacional, reconstruir permitiendo áreas ausentes o dañadas en una imagen con el objetivo de recuperar su coherencia visual (Elharrouss et al., 2020). En la conservación del patrimonio cultural, su importancia es notable, especialmente en obras pictóricas que carecen de registros previos. Sin embargo, los métodos actuales enfrentan limitaciones relacionadas con la eficiencia temporal y los elevados requisitos computacionales, lo que restringe su uso en escenarios que exigen resultados rápidos y de alta calidad (Wang et al., 2021).

En los últimos años, el aprendizaje profundo ha surgido como el enfoque predominante para abordar los retos de la restauración digital. Estas técnicas, representando el estado del arte, han permitido generar reconstrucciones visualmente impactantes mediante redes neuronales convolucionales y generativas. Sin embargo, su implementación suele ser costosa en términos de tiempo de ejecución, lo que motiva la necesidad de explorar alternativas más eficientes que equilibren calidad y velocidad (Zeng et al., 2023).

¿Cómo se pueden adaptar técnicas avanzadas de aprendizaje profundo para la restauración digital de pinturas sin registros fotográficos, mejorando el tiempo de ejecución sin comprometer significativamente la calidad de los resultados?

La presente investigación, titulada "Adaptación de técnicas de aprendizaje profundo para la restauración digital veloz de pinturas sin registros fotográficos", se enmarca en este contexto. Su objetivo es adaptar técnicas avanzadas de inpainting para mejorar el tiempo de ejecución sin pérdidas significativas de calidad. La metodología utilizada incluye una revisión exhaustiva de las últimas técnicas publicadas sobre restauración digital, de las cuales se seleccionan diez que fueron evaluadas en base a los criterios propuestos para la investigación, seleccionando dos técnicas que cumplen satisfactoriamente para su posterior adaptación y experimentación (Zeng et otros, 2023).

Esta investigación busca contribuir tanto al avance académico en visión computacional como a la preservación del patrimonio cultural, proporcionando herramientas más eficientes y accesibles.

Aspectos Generales

1.1 Problema de investigación

1.1.1 Descripción del problema

La restauración digital de pinturas sin registros fotográficos representa un desafío significativo dentro del ámbito de la preservación del patrimonio cultural. Este proceso es crucial para reconstruir áreas dañadas o ausentes en obras artísticas, proporcionando una base visual que guíe las intervenciones tradicionales. Sin embargo, cuando no existen registros fotográficos confiables, los restauradores enfrentan limitaciones que dificultan la planificación de intervenciones y la preservación de los valores históricos de las pinturas.

En este contexto, las técnicas de aprendizaje profundo han demostrado ser efectivas para generar reconstrucciones visualmente convincentes mediante algoritmos avanzados de inpainting. No obstante, estas técnicas presentan serias limitaciones que comprometen su aplicabilidad práctica. Entre los problemas más destacados se encuentran los altos tiempos de ejecución, que ralentizan significativamente el proceso, y los elevados requerimientos computacionales, que restringen su implementación en dispositivos con recursos limitados. Además, muchas de estas técnicas priorizan la calidad visual sin optimizar la eficiencia temporal ni el uso de recursos.

Estas deficiencias dificultan la adopción de estas herramientas en escenarios reales que demandan rapidez y eficiencia, limitando su potencial para democratizar la restauración digital en contextos profesionales y cotidianos. Ante esta situación, surge la necesidad de explorar cómo adaptar y optimizar las técnicas de aprendizaje profundo existentes para acelerar el proceso de restauración sin comprometer la calidad de los resultados obtenidos.

Por lo tanto, el problema principal de esta investigación consiste en determinar cómo implementar adaptaciones en técnicas avanzadas de aprendizaje profundo que permitan mejorar la velocidad de ejecución, mantener la calidad visual y reducir los costos computacionales. Resolver este problema contribuirá a desarrollar herramientas prácticas y accesibles que potencien la labor de los restauradores profesionales y permitan la conservación eficiente de obras culturales que carecen de referencias visuales.

1.1.2 Identificación del problema

¿La adaptación de técnicas de aprendizaje profundo permite acelerar el proceso de restauración digital de pinturas que no cuentan con registros fotográficos sin pérdidas significativas de calidad?

1.1.3 Problemas específicos

- ¿Cuáles son las técnicas de aprendizaje profundo que serán adaptadas para acelerar el proceso de restauración digital?
- ¿Qué adaptaciones se pueden proponer a las técnicas de aprendizaje profundo seleccionadas para acelerar su tiempo de ejecución sin mucha pérdida de calidad?
- ¿En qué medida las técnicas adaptadas aceleran el proceso de restauración digital y cuál es la mejor obtenida?

1.2 Antecedentes

Presentamos los antecedentes revisados que se utilizaron como base inicial de la investigación, así como los antecedentes directamente relevantes para el proyecto.

1.2.1 Context Encoders: Aprendizaje de características para restauración digital

En el artículo de (Pathak et al., 2016), titulado “*Context Encoders: Feature Learning by Inpainting*”, se presenta un algoritmo de aprendizaje no supervisado de características visuales, diseñado para la predicción de píxeles basada en su contexto. El objetivo principal de esta investigación fue desarrollar **Context Encoders**, una red neuronal convolucional capaz de generar contextos de una región arbitraria de una imagen condicionada a su entorno, permitiendo comprender el contenido completo de la imagen y generar hipótesis plausibles para las partes faltantes. La metodología incluyó experimentos con una pérdida de reconstrucción estándar por píxeles y una combinación de pérdida de reconstrucción y pérdida adversarial, donde esta última demostró producir resultados significativamente más nítidos y efectivos al manejar la diversidad en las salidas. Como conclusión, se encontró que los Context Encoders no solo capturan la apariencia, sino también la semántica de las estructuras visuales, siendo efectivos para tareas de clasificación, detección y segmentación. Además, los codificadores de contexto muestran ser valiosos para el preentrenamiento en aplicaciones de visión por computadora y útiles en tareas de restauración de imágenes, tanto de forma independiente como complementaria a otros métodos.

1.2.2 MSNP: Restauración digital de alta resolución usando Síntesis de Parches Neuronales Multiescala

En el artículo de (Yang et al., 2016), titulado “*High-Resolution Image Inpainting Usando Multi-Scale Neural Patch Synthesis*”, se propone un enfoque innovador para abordar las limitaciones de los métodos previos en aprendizaje profundo, los cuales, aunque efectivos para capturar características de alto nivel, enfrentaban restricciones de memoria y dificultades de entrenamiento, especialmente en imágenes de mayor resolución. El objetivo principal fue desarrollar un método basado en la síntesis de parches neuronales de múltiples escalas que optimiza tanto el contenido como las restricciones de textura de la imagen. Esta técnica no solo conserva las estructuras contextuales, sino que también mejora los detalles de alta frecuencia adaptando los parches según las correlaciones de características intermedias de redes neuronales profundas. La metodología fue evaluada utilizando los conjuntos de datos ImageNet y Paris StreetView , logrando una calidad de estado del arte en la reconstrucción digital de imágenes. Los resultados demostraron que este enfoque supera a los métodos anteriores, produciendo imágenes más nítidas y coherentes, particularmente en contextos de alta resolución.

1.2.3 GLIC: Completación global y local de imágenes consistentemente

En el trabajo de (Iizuka et al., 2017) titulado “*GLIC: Global and Local Consistent Image Completion*”, se presenta un enfoque novedoso para la restauración de imágenes digitales que asegura la consistencia tanto a nivel global como local. El objetivo principal fue diseñar una red neuronal convolucional completamente conectada capaz de completar imágenes de resoluciones arbitrarias, rellenando regiones faltantes de cualquier forma. Para entrenar esta red y garantizar la coherencia de los resultados, los autores implementaron discriminadores de contexto global y local: el discriminador global evalúa la coherencia de la imagen completa, mientras que el local se enfoca en las regiones generadas. La red fue entrenada para engañar a ambos discriminadores, generando imágenes indistinguibles de las reales en términos de consistencia y detalles. Este enfoque demuestra ser capaz de completar escenarios diversos, produciendo fragmentos nuevos que no aparecen en otras partes de la imagen, lo que permite la generación de detalles únicos y específicos, como rostros o estructuras complejas, de manera natural y precisa.

1.2.4 Shift-Net: Restauración digital de imágenes a través de Deep Features

En el trabajo de (Yan et al., 2018), titulado “*Shift-Net: Image Inpainting via Deep Feature Rearrangement*”, se introduce una capa especial de conexión llamada **Shift-Net**,

diseñada para ser integrada a la arquitectura **U-Net** con el fin de rellenar regiones faltantes en imágenes, manteniendo estructuras nítidas y texturas detalladas. El objetivo de la investigación fue desarrollar una red neuronal que desplazara características del codificador de una región conocida para predecir las partes faltantes. Para mejorar la precisión de las predicciones, se introduce una pérdida de guía en el decodificador, minimizando la diferencia entre las características del decodificador y las del codificador de la región de referencia. La metodología consistió en entrenar el modelo mediante un algoritmo de aprendizaje de extremo a extremo en los conjuntos de datos Paris StreetView y Places, demostrando que la Shift-Net genera imágenes más nítidas y detalladas, superando los métodos anteriores en términos de calidad visual. Como conclusión, se concluye que Shift-Net es eficaz para generar resultados más coherentes y plausibles en la restauración de imágenes digitales, especialmente en el contexto de regiones faltantes complejas.

1.2.5 GICA: Restauración digital de imágenes con atención contextual

En el trabajo de (Yu et al., 2018b), titulado “*Generative Image Completion with Contextual Attention*”, los autores proponen un enfoque innovador para la restauración digital de imágenes utilizando atención contextual. El objetivo principal fue mejorar la integración de las regiones circundantes durante la restauración, superando los problemas de distorsión de estructuras y texturas comunes en enfoques anteriores. La metodología incluyó una red neuronal convolucional totalmente integrada que podía procesar imágenes con múltiples agujeros de tamaño y ubicación variables. Se entrenó el modelo con la atención a las características circundantes, lo que mejoró la coherencia y detalle en las predicciones. A través de experimentos realizados en **CelebA**, **ImageNet** y **Places2**, se demostró que este enfoque produce reconstrucciones de mayor calidad que los métodos existentes. En conclusión, el modelo permitió una mejor integración de las características de la imagen circundante, destacando por su versatilidad y eficacia en la restauración digital.

1.2.6 FFGC: Restauración digital Free-Form con una convolución cerrada

En otro trabajo de, (Yu et al., 2018a) titulado “*Free-Form Image Completion with Gated Convolutions*”, presenta un innovador sistema generativo para completar imágenes utilizando máscaras y guías de forma libre. El objetivo de la investigación fue abordar la limitación de las convoluciones estándar que trataron todos los píxeles como válidos, introduciendo convoluciones cerradas que permitieron la selección dinámica de características. Se empleó un mecanismo de **GAN basado en parches** denominado **SN-**

PatchGAN para mejorar la generación de parches en imágenes con múltiples formas de máscaras. Los resultados mostraron que el sistema genera imágenes restauradas de alta calidad, con mayor flexibilidad en comparación con los métodos anteriores. Los experimentos realizados demostraron que este sistema optimiza la restauración automática de imágenes y facilita tareas de edición, como la eliminación de objetos o la modificación del diseño de la imagen, ampliando las posibilidades creativas en el manejo de imágenes. Como conclusión, se destacó que la SN-PatchGAN es eficiente, estable y adecuada para una restauración digital flexible, superando los métodos anteriores en rapidez y calidad.

Del grupo de antecedentes, se extrajo el conocimiento necesario, el cual se expone en su mayoría en el marco teórico, como por ejemplo, El desarrollo de técnicas avanzadas para la restauración digital de imágenes ha impulsado métodos basados en redes neuronales que integran contextos globales y locales, síntesis de parches y mecanismos de atención contextual. Desde los Context Encoders de Pathak et al. (2016), que introducen la predicción de píxeles basada en el entorno, hasta el enfoque multiescala de Yang et al. (2016) y las innovaciones como Shift-Net (Yan et al., 2018) y GLIC (Iizuka et al., 2017), se ha logrado una generación más nítida y semánticamente coherente de regiones faltantes. Avances como GICA (Yu et al., 2018b) y FFGC (Yu et al., 2018a) han destacado por mejorar la consistencia al integrar características de la imagen circundante y el manejo de máscaras de forma libre, respectivamente. Estas técnicas han demostrado eficacia en tareas como eliminación de objetos y edición creativa, ampliando significativamente el alcance y calidad de la restauración digital de imágenes. Estos antecedentes proporcionan los fundamentos y avances clave que guían el desarrollo y la comprensión de las redes neuronales convolucionales en el contexto de esta investigación.

1.3 Objetivos

1.3.1 Objetivo general

Adaptación de técnicas de aprendizaje profundo para la restauración digital veloz de pinturas sin registros fotográficos.

1.3.2 Objetivos específicos

- Determinar las técnicas de aprendizaje profundo que serán adaptadas para acelerar la restauración digital.

- Proponer adaptaciones a las técnicas de aprendizaje profundo seleccionadas para acelerar la restauración digital.
- Determinar en qué medida mejora la rapidez del proceso de restauración digital cuál es la calidad obtenida.

1.4 Alcances y Limitaciones

Con el fin de desarrollar una variante de una técnica para la restauración digital de imágenes, es conveniente limitar el trabajo a lo siguiente.

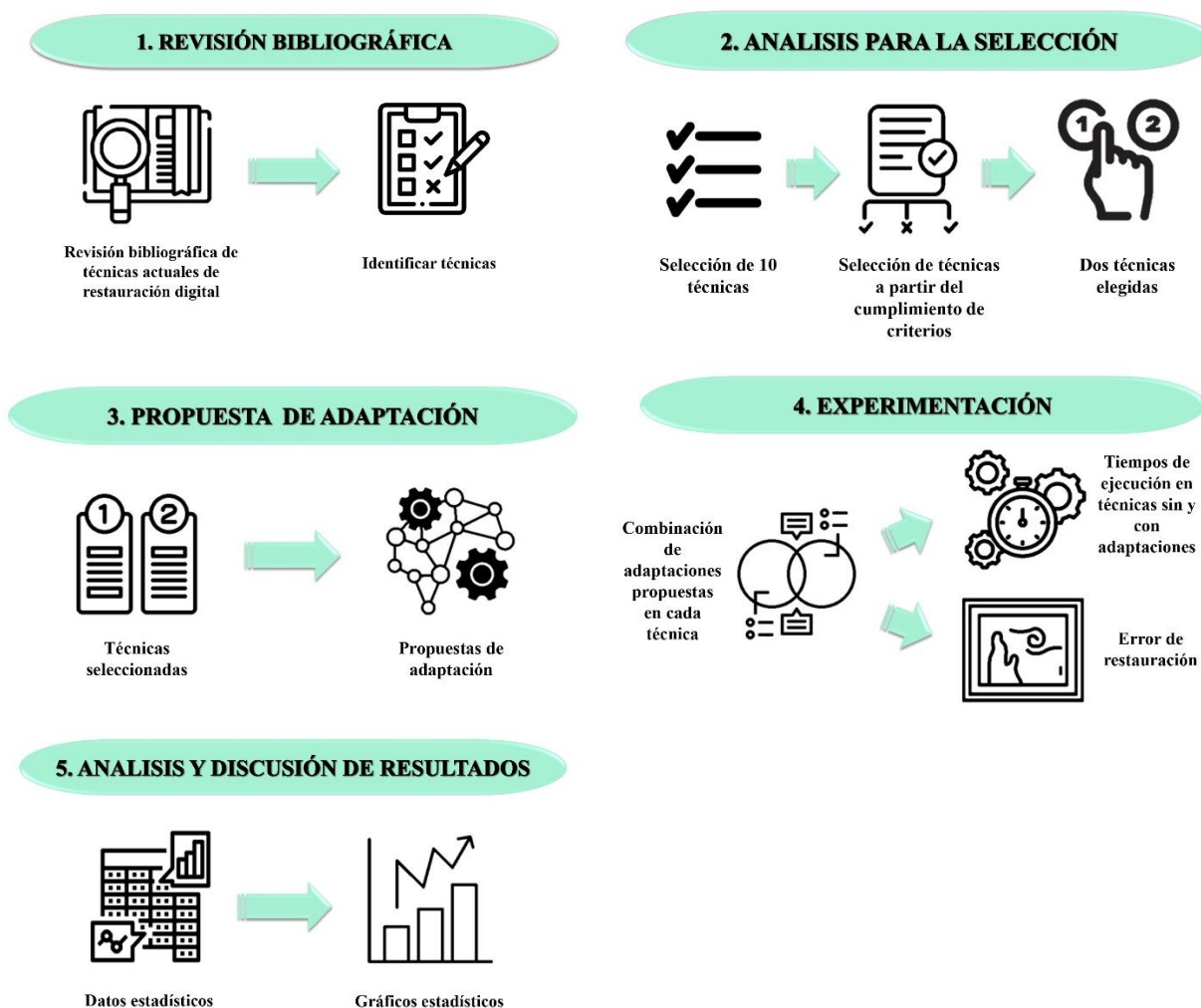
- Debido a la gran complejidad de las técnicas de restauración presentes en el estado del arte, se explorarán únicamente los conceptos teóricos básicos en los que se fundamentan. Esto se debe a que, una vez implementada la modificación propuesta, no es necesario controlar el entrenamiento de los modelos de inteligencia artificial, sino únicamente la fase de utilización de estos. La capacidad de aplicar estas adaptaciones sin necesidad de intervenir en el proceso de entrenamiento hace que sea más eficiente y práctico enfocarse en la implementación directa.
- Se llevarán a cabo experimentos únicamente con las técnicas cuyo código fuente esté disponible al público. Además, solo se considerarán aquellas técnicas que cuenten con una documentación suficientemente detallada para ser ejecutadas en una computadora convencional. Este criterio de selección asegura la reproducibilidad de los experimentos y la accesibilidad de las herramientas utilizadas, facilitando una evaluación más transparente y objetiva de los resultados obtenidos. No obstante, se mencionarán todas las técnicas consideradas durante la fase de realización del proyecto, especificando cuáles fueron finalmente seleccionadas y cuáles no formaron parte de la implementación. De esta manera, se proporciona una visión completa del proceso de selección y evaluación de las técnicas de restauración, subrayando tanto las metodologías adoptadas como las descartadas, y justificando las decisiones tomadas en función de su viabilidad práctica y documental.
- No se tomarán en cuenta algoritmos heurísticos que rellenan información basándose únicamente en reglas bien definidas como usar el color mayoritario alrededor de los píxeles faltantes. El enfoque utilizado será principalmente los algoritmos de aprendizaje automático, en particular, de deep learning debido a su calidad superior frente a estos otros algoritmos. Otro efecto secundario de esto es su posible fácil modificación posterior.

1.5 Metodología

A continuación, se detalla la metodología que se siguió para la investigación.

1. Revisión bibliográfica. Se hizo un estudio de la bibliografía sobre las técnicas actuales de restauración digital de imágenes, para compilar una lista de técnicas potenciales para aplicarlas en la investigación. Este estudio se centró primordialmente en la revisión de técnicas de aprendizaje profundo (deep learning), que representan el estado del arte en la restauración digital de imágenes. Además, se revisaron técnicas adyacentes para identificar conceptos similares que podrían ser aplicados en futuras mejoras. Este último paso resultó crucial para identificar técnicas apropiadas susceptibles de adaptación, ampliando así el marco de opciones y potenciales innovaciones en el proceso de restauración digital.
2. Análisis para la selección. En esta etapa, se analizaron las características de cada técnica para determinar la viabilidad de su modificación, considerando las limitaciones mencionadas. Este análisis también involucró una evaluación de factores como la calidad, el tiempo de ejecución, y los requerimientos computacionales de cada técnica. El resultado de esta etapa fue la identificación y selección de dos técnicas de aprendizaje profundo que mostraron el mayor potencial para ser adaptadas a las necesidades específicas del proyecto. Estas técnicas se eligieron por su calidad en la restauración de imágenes y por su capacidad de optimizarse para mejorar su eficiencia en tiempo. Estas dos técnicas seleccionadas serán objeto de aplicación en la siguiente fase del proyecto.
3. Propuesta de adaptación. En esta etapa, se tomaron las técnicas seleccionadas en el paso anterior y se realizaron experimentos con adaptaciones en diversos parámetros característicos de cada técnica. Este enfoque permitió proponer una serie de ajustes específicos, que serán objeto de una comparación experimental detallada en la siguiente fase del proyecto.
4. Experimentación. Se consideró cada combinación de adaptaciones propuestas en la etapa anterior, poniendo un valor a los resultados mediante la métrica del número de píxeles correctamente clasificados. Esta consideración permitió cuantificar la eficacia de cada ajuste, facilitando una comparación precisa de su impacto en la calidad de la restauración digital en la siguiente etapa.
5. Análisis y discusión de resultados. En esta etapa se resumen los resultados de la experimentación realizada previamente.

Figura 1
Metodología propuesta



1.6 Justificación

Es común que un restaurador profesional se enfrente inicialmente a una larga lista de posibles restauraciones que debe atender, eligiendo de acuerdo a su conveniencia. En estos casos, uno de los factores decisivos es la cantidad de trabajo que cada restauración requerirá. Un aspecto crucial que facilitaría esta decisión es disponer de una visión general del aspecto original de las obras en la lista. Sin embargo, como se ha mencionado anteriormente, es frecuente la falta de registros fotográficos confiables para esta observación inicial.

La existencia de una técnica de inteligencia artificial capaz de realizar restauraciones digitales rápidas y con una calidad comparable a las mejores técnicas actuales representaría una herramienta invaluable para el trabajo diario de los restauradores profesionales. La implementación de tal herramienta permitiría a los restauradores obtener una estimación

preliminar precisa del estado original de una obra, facilitando así la planificación y priorización de las tareas de restauración.

La importancia de esta investigación futura radica en la creación de dicha herramienta, basada inicialmente en la selección y modificación de una técnica de aprendizaje profundo ya existente, seguida de su posterior simplificación controlada. Esta simplificación no solo buscaría mantener la alta calidad de las restauraciones, sino también mejorar la eficiencia y velocidad del proceso. Al reducir el tiempo y esfuerzo necesarios para las restauraciones preliminares, los restauradores podrían dedicar más recursos a intervenciones más detalladas y complejas, optimizando así el proceso de conservación y restauración del patrimonio artístico. El desarrollo de esta técnica avanzada de restauración digital podría transformar significativamente el campo de la restauración profesional, proporcionando un recurso eficiente y altamente efectivo para preservar obras de arte valiosas.

Marco Teórico

2.1 Restauración de Pinturas

La restauración de una pintura es un trabajo multidisciplinario que requiere la colaboración de un equipo diverso de expertos, entre ellos restauradores, historiadores, biólogos, químicos y fotógrafos. Para evaluar el estado de conservación de la obra y entender su evolución, se emplean diferentes tipos de análisis y métodos de investigación. El objetivo del restaurador es recopilar datos sobre las técnicas y materiales empleados en su creación, así como identificar posibles alteraciones en su soporte o estructura. (Instituto Andaluz del Patrimonio Histórico, s.f.)

Antes de llevar a cabo cualquier intervención, el restaurador debe tener un conocimiento detallado de la obra y su desarrollo a lo largo del tiempo. Para ello, se realizarán diversos estudios: algunos de carácter histórico, que permiten obtener información sobre la obra, su autor y el contexto temporal en el que fue creada, y otros de tipo analítico, enfocados en determinar los materiales utilizados, como el soporte, los pigmentos y barnices, así como las modificaciones o restauraciones previas. Toda esta información resulta esencial para evaluar el estado de conservación de la pieza y planificar el proceso de restauración adecuado. (Instituto Andaluz del Patrimonio Histórico, s.f.)

Figura 2

Estudio de una pintura con microscopio USB



Nota: Tomado de COREMANS, 2018

La restauración de pinturas es un proceso técnico y científico destinado para preservar, estabilizar y recuperar la integridad visual y estructural de una obra de arte. Este proceso busca garantizar que la pintura conserve su autenticidad y valor histórico, respetando siempre las características originales. La restauración no se limita a una reparación superficial, sino que implica un enfoque detallado que abarca la limpieza, consolidación, reintegración cromática y protección de la obra, con el objetivo de asegurar su conservación para las generaciones futuras (COREMANS, 2018)

2.1.1 ¿Por qué es necesario restaurar una pintura?

Con el paso del tiempo, toda obra de arte experimenta deterioro debido al envejecimiento de los materiales, a alteraciones físico-químicas, a las condiciones ambientales desfavorables a las que ha sido sometida, y en algunos casos, incluso al vandalismo. Estos son algunos de los factores principales que contribuyen al deterioro de las obras.

2.1.2 Principios de la restauración

Según Cesare Brandi, los principios fundamentales de la restauración se basan en el concepto de "valor histórico" de la obra de arte. Para Brandi (2005), la restauración no debe ser un proceso meramente de reparación, sino una intervención que respete la integridad histórica y estética de la obra.

El primer principio es la mínima intervención, lo que implica que cualquier acción tomada sobre la obra debe ser lo más discreta y reversible posible, para no alterar su esencia original. El segundo principio es la distinguir lo antiguo de lo nuevo, asegurando que las restauraciones sean identificables para no confundirlas con el trabajo original. Finalmente, Brandi sostiene que la unidad de la obra debe preservarse, no solo desde el punto de vista físico, sino también en su sentido simbólico y cultural (Brandi, 2005).

2.1.3 ¿Cómo se lleva a cabo la restauración de pinturas?

Según Ortiz (2012) La restauración de cuadros no sigue un método único, ya que depende de diversos factores como la época de creación de la obra y los daños que presenta. Cada caso requiere un enfoque particular, ya que incluso cuadros similares en técnica, temática o época pueden necesitar procedimientos de limpieza completamente diferentes.

En la restauración de pintura de caballete, el proceso de restauración se estructura en varias fases clave que garantizan la preservación y la integridad de la obra.

2.1.3.1. Examen preliminar y diagnóstico

El primer paso en el proceso es realizar un examen detallado de la pintura, utilizando métodos no invasivos para evaluar su estado de conservación. Esto incluye el análisis de la estructura del lienzo, la pintura, las capas de preparación y el soporte. Se identifican los problemas principales, como el envejecimiento de la capa pictórica, desgarros, pérdidas de pigmento o alteraciones en el barniz (Coremans, 1996).

2.1.3.2. Documentación

Una vez diagnosticados los problemas, se procede a la documentación meticulosa de todos los aspectos relevantes de la obra. Esto incluye fotografías detalladas, análisis de materiales y un registro exhaustivo de las condiciones previas y posteriores a la intervención. La documentación es fundamental para garantizar la trazabilidad del proceso de restauración (Coremans, 1996).

2.1.3.3. Reparación del soporte

En esta etapa se corrigen problemas estructurales del lienzo, como destensado, deformaciones, bordes desgastados o roturas. Además, se evalúa el estado del bastidor que sostiene la obra. Si está dañado, puede ser necesario sustituirlo, para lo cual se retira el lienzo del bastidor original y se monta en uno nuevo que lo mantenga correctamente tensado. (Ortiz, 2012)

2.1.3.3. Limpieza

La fase de limpieza es crucial para eliminar capas de suciedad, barnices oxidados y contaminantes sin dañar la pintura original. Esta etapa se realiza con mucha precaución, utilizando solventes adecuados y herramientas suaves para no alterar la superficie pictórica. Coremans destaca la importancia de evitar el uso de productos agresivos que puedan afectar la obra de manera irreversible (Coremans, 1996).

2.1.3.4. Consolidación

En casos de desprendimiento o flacidez del soporte pictórico, se recurre a técnicas de consolidación para fortalecer las capas de pintura. Esto implica el uso de resinas específicas que ayudan a estabilizar los pigmentos y restaurar la integridad estructural de la pintura (Coremans, 1996).

2.1.3.5. Retención de la intervención mínima

Coremans subraya la necesidad de que la intervención en la obra sea lo más mínima posible. Es decir, las restauraciones deben realizarse solo en áreas que realmente lo

necesiten, conservando la mayor cantidad de material original posible. Esto respeta la autenticidad y la integridad estética de la pintura (Coremans, 1996).

2.1.3.6. Reintegración cromática

Una vez realizadas las intervenciones estructurales, se lleva a cabo la reintegración cromática para restaurar las áreas que han sufrido pérdidas de pigmento. Para esto se emplean técnicas de retoque con colores compatibles que imitan la tonalidad original sin distorsionar la apariencia global de la obra (Coremans, 1996).

2.1.3.7. Barnizado final

Finalmente, se aplica una capa de barniz para proteger la pintura de factores externos como la humedad y la luz. Este barniz también ayuda a unificar el acabado visual de la pintura (Coremans, 1996).

2.2 Inteligencia Artificial

2.2.1 ¿Qué es la Inteligencia Artificial?

La Inteligencia Artificial es el desarrollo y estudio de métodos y programas que permiten a las máquinas mostrar inteligencia similar a la cognición humana. En esencia, la inteligencia artificial capacita a las computadoras para razonar, aprender y tomar decisiones basadas en datos, tareas que históricamente requerían inteligencia humana.

La inteligencia artificial permite a las máquinas percibir su entorno a través de sensores, cámaras y otras fuentes de entrada. Aprenden a partir de estos datos y adaptan su comportamiento en consecuencia (Panda et al., 2021).

La percepción es la capacidad de una computadora para recibir y procesar información sensorial de una manera similar a como los humanos perciben el mundo. Al igual que nuestros sentidos (vista, oído, tacto, gusto, olfato), la inteligencia artificial depende de sensores para recopilar datos de su entorno. Estos sensores incluyen visión artificial, donde las cámaras ópticas capturan información visual; audición computacional, donde los micrófonos recogen datos de audio; tacto artificial, donde los sensores táctiles proporcionan información sobre el contacto físico; olfato artificial, donde las narices electrónicas detectan olores; y gusto artificial, donde las lenguas electrónicas analizan las propiedades del gusto (Panda et al., 2021).

Una vez recopilados los datos sensoriales, los algoritmos de inteligencia artificial los procesan. Este proceso multietapa implica convertir los datos en bruto en un modelo de

lo que se está percibiendo. Por ejemplo, los algoritmos de visión por computadora analizan imágenes para reconocer objetos, rostros y escenas; los algoritmos de procesamiento de audio identifican sonidos, discursos y patrones; y los sensores táctiles ayudan a los robots a entender sus interacciones físicas con objetos (Panda et al., 2021).

La percepción es la primera etapa en los modelos sensoriales de la IA. Los algoritmos construyen una comprensión básica del mundo percibido. Esta comprensión sirve como base para un razonamiento y toma de decisiones más avanzados. La inteligencia artificial no necesita percibir exactamente como los humanos. En el diagnóstico médico, por ejemplo, los algoritmos de inteligencia artificial pueden proporcionar mejores análisis porque acceden a imágenes o datos más precisos que los que los humanos pueden percibir. El objetivo no es la imitación, sino la utilidad (Panda et al., 2021).

Los sistemas de inteligencia artificial toman acciones para maximizar sus posibilidades de alcanzar objetivos predefinidos. Ya sea jugando al ajedrez, conduciendo vehículos autónomos o generando arte creativo, la inteligencia artificial optimiza sus decisiones (Panda et al., 2021).

Cuando decimos que los sistemas de inteligencia artificial toman acciones para maximizar sus posibilidades de alcanzar objetivos predefinidos, estamos describiendo su comportamiento orientado a un propósito. A continuación, se explica cómo este principio se aplica en diversos dominios (Panda et al., 2021).

En la inteligencia artificial que juega al ajedrez, el objetivo es claro: ganar el juego. La inteligencia artificial evalúa los movimientos potenciales considerando factores como el valor de las piezas, la posición en el tablero y las consecuencias futuras. Optimiza sus decisiones seleccionando movimientos que lleven a resultados favorables, como capturar piezas del oponente, controlar el centro del tablero o preparar un jaque mate (Panda et al., 2021).

Los vehículos autónomos operan con el objetivo de una navegación segura y eficiente. Su toma de decisiones implica equilibrar múltiples objetivos: seguridad (evitar colisiones con peatones, otros vehículos y obstáculos), eficiencia (llegar al destino en el menor tiempo posible) y comodidad (proporcionar un viaje suave para los pasajeros). Los algoritmos de inteligencia artificial optimizan las acciones basándose en datos de sensores (lidar, cámaras, radar) y condiciones de tráfico en tiempo real (Panda et al., 2021).

La generación de arte impulsada por inteligencia artificial tiene como objetivo crear contenido novedoso y estéticamente agradable. Ya sea componiendo música, generando arte visual o escribiendo poesía, la inteligencia artificial optimiza sus elecciones para lograr variedad (explorando diversas posibilidades para evitar patrones repetitivos), armonía (equilibrando elementos como colores, notas y palabras para crear un conjunto coherente) y emoción (infundiendo al arte con emociones deseadas como alegría, melancolía o sorpresa). El éxito de la inteligencia artificial aquí no es ganar un juego, sino producir algo cautivador y único (Panda et al., 2021).

En los negocios, los sistemas de inteligencia artificial optimizan decisiones para lograr objetivos específicos. Esto incluye la maximización de ingresos mediante algoritmos de precios que ajustan los precios dinámicamente para maximizar las ganancias, la asignación de recursos a través de la optimización de la cadena de suministro que equilibra los niveles de inventario y los costos de transporte, y la satisfacción del cliente mediante chatbots que buscan resolver consultas de manera eficiente mientras mantienen una experiencia positiva para el usuario (Panda et al., 2021).

Algunas aplicaciones generales de la inteligencia artificial incluyen:

Motores de búsqueda

Google, por ejemplo, emplea algoritmos de inteligencia artificial para proporcionar resultados relevantes. Los motores de búsqueda web han evolucionado significativamente a lo largo de los años, y la inteligencia artificial juega un papel crucial en la mejora de la experiencia de búsqueda. Algunas maneras en que los motores de búsqueda utilizan la inteligencia artificial son:

La inteligencia artificial ayuda a los motores de búsqueda a indexar y comprender la vasta cantidad de contenido web. Procesa y analiza las páginas web, extrayendo información relevante de textos, imágenes y otros medios. Los motores de búsqueda utilizan algoritmos de inteligencia artificial para categorizar y organizar el contenido, facilitando la recuperación de resultados relevantes para las consultas de los usuarios (Russell & Norvig, 2016).

Esta también permite a los motores de búsqueda interpretar las consultas de búsqueda de manera más efectiva. En lugar de depender únicamente de palabras clave, los algoritmos de inteligencia artificial comprenden el lenguaje natural y el contexto. Al

comprender la intención del usuario, los motores de búsqueda pueden proporcionar resultados más precisos y contextualmente relevantes (Russell & Norvig, 2016).

La inteligencia artificial empareja las consultas de búsqueda con los resultados más relevantes y de alta calidad. Consideran factores como la relevancia, la autoridad y la actualidad. Esto asegura que los usuarios reciban la información más útil según su intención de búsqueda (Russell & Norvig, 2016).

También se usa inteligencia artificial para evaluar continuamente la calidad del contenido para mejorar los resultados de búsqueda. Identifica páginas de baja calidad o desactualizadas y promueve fuentes autoritativas. Los motores de búsqueda utilizan la inteligencia artificial para mejorar la experiencia general de búsqueda refinando los algoritmos y ajustando los rankings (Russell & Norvig, 2016).

La inteligencia artificial generativa implica el uso de la inteligencia artificial para generar contenido, como imágenes, resúmenes e incluso respuestas. Algunos motores de búsqueda impulsados por inteligencia artificial ofrecen más que solo resultados basados en texto. Incorporan elementos visuales e información contextualmente relevante. Otros priorizan la privacidad del usuario. Evitan rastrear datos personales y proporcionan experiencias de búsqueda anónimas (Russell & Norvig, 2016).

Andi es otro ejemplo de motor de inteligencia artificial emergente que presenta la información de una manera novedosa. Utiliza la inteligencia artificial a lo largo de toda la página de resultados de búsqueda, no solo en la parte superior. Imágenes, resúmenes y opciones se ofrecen de manera contextual. Andi busca proporcionar respuestas confiables sin las alucinaciones comunes en algunos modelos de IA. Combina modelos de lenguaje de gran escala comerciales y de código abierto con gráficos de conocimiento y motores de búsqueda existentes (Russell & Norvig, 2016).

Sistemas de recomendación

La inteligencia artificial también juega un papel muy importante en la mejora de los sistemas de recomendación en muchas plataformas. Los algoritmos de inteligencia artificial analizan los datos de los usuarios, identificando patrones y preferencias. Consideran factores como el historial de navegación, el comportamiento de compra y las interacciones con el contenido. Estos sistemas de recomendación impulsados por inteligencia artificial predicen las preferencias del usuario basándose en el historial. Al comprender los gustos individuales, ofrecen sugerencias personalizadas (Russell & Norvig, 2016).

Las técnicas de inteligencia artificial manejan muchos y muy complicados conjuntos de datos muy rápidamente. Estas identifican correlaciones y patrones ocultos que los humanos podrían pasar por alto. Entre las técnicas de inteligencia artificial en sistemas de recomendación, encontramos el filtrado colaborativo, que analiza las interacciones entre usuarios y artículos y recomienda artículos basados en las preferencias de usuarios similares. Las técnicas incluyen el filtrado basado en usuarios y en artículos (Russell & Norvig, 2016).

Otra técnica muy usada es el filtrado basado en contenido, que considera los atributos de los artículos y los perfiles de los usuarios, recomendando artículos similares a aquellos en los que el usuario ha mostrado interés. Esta técnica usa la factorización de matrices para descomponer matrices de interacción usuario-artículo para extraer características latentes, popularizada por técnicas como SVD y MF. El aprendizaje profundo, mediante redes neuronales, captura patrones complejos y combina enfoques colaborativos y basados en contenido. Más adelante en la tesis se habla del aprendizaje profundo en detalle (Russell & Norvig, 2016).

Los modelos híbridos combinan múltiples técnicas para ofrecer recomendaciones más robustas. Estos sistemas de recomendación dan diversos beneficios, como la personalización, que mejora el compromiso del usuario, y la mayor satisfacción del usuario al recibir contenido relevante. Además, la inteligencia artificial maneja datos a gran escala de manera eficiente, lo que resulta en un procesamiento eficiente y aplicaciones diversas en comercio electrónico, servicios de streaming, redes sociales y más (Russell & Norvig, 2016).

Algunos desafíos y direcciones futuras para considerar son, por ejemplo, el problema del arranque en frío, que se refiere a recomendar para nuevos usuarios o artículos, y las consideraciones éticas, como asegurar la equidad y evitar sesgos, son áreas importantes de enfoque. Además, las recomendaciones contextuales, que incorporan información contextual, representan una dirección futura clave para mejorar aún más los sistemas de recomendación (Russell & Norvig, 2016).

Asistentes virtuales

Los asistentes virtuales representan una de las aplicaciones más visibles y ampliamente utilizadas de la inteligencia artificial en la vida cotidiana. Se han convertido en partes integrales de nuestras interacciones digitales, ofreciendo asistencia, realizando tareas y proporcionando información en una variedad de plataformas y dispositivos. Los asistentes virtuales son agentes de software diseñados para ayudar a los usuarios con

diversas tareas e investigaciones a través de interacciones en lenguaje natural. Utilizan tecnologías de inteligencia artificial para comprender la entrada del usuario, procesar solicitudes y proporcionar respuestas relevantes o realizar acciones. Estas técnicas pueden hacer cosas como programar citas y establecer recordatorios hasta responder preguntas y controlar dispositivos inteligentes en el hogar. Los asistentes virtuales han evolucionado hasta convertirse en herramientas muy útiles que mejoran la productividad (Russell & Norvig, 2016).

Algunas de las técnicas de inteligencia artificial que potencian los asistentes virtuales incluyen:

Procesamiento del Lenguaje Natural. El procesamiento de lenguaje natural permite a los asistentes virtuales comprender e interpretar la entrada de lenguaje humano. También se emplean técnicas como el análisis de sentimientos, el análisis de entidades nombradas y el análisis de texto para extraer el significado del texto o el habla (Russell & Norvig, 2016).

Reconocimiento de Voz. Esta técnica convierte el lenguaje hablado en texto, permitiendo a los asistentes virtuales procesar comandos de voz. Se utilizan técnicas avanzadas y modelos de aprendizaje automático para mejorar la calidad y también para adaptarse a diferentes acentos y patrones de habla (Russell & Norvig, 2016).

Aprendizaje Automático. Los modelos de aprendizaje automático permiten a los asistentes virtuales aprender de las interacciones con los usuarios y mejorar su rendimiento con el tiempo. Pueden reconocer patrones, personalizar respuestas y adaptarse a las preferencias del usuario en función de datos históricos. Más adelante se cubre más conceptos del aprendizaje automático (Russell & Norvig, 2016).

Grafos de Conocimiento. Los grafos de conocimiento organizan mucha información estructurada sobre entidades y sus relaciones. Esto permite a los asistentes virtuales proporcionar respuestas relevantes en el contexto. También facilitan la comprensión de consultas complejas y apoyan las interacciones conversacionales (Russell & Norvig, 2016).

Gestión de Diálogos. Las técnicas de gestión de diálogos controlan el flujo de conversación entre los usuarios y los asistentes virtuales. Esto asegura que las interacciones sean coherentes y contextualmente apropiadas. También se maneja el intercambio de turnos. También es posible gestionar las transiciones de estado y manejar errores o malentendidos (Russell & Norvig, 2016).

Algunas aplicaciones de los asistentes virtuales son:

Productividad Personal. Los asistentes virtuales ayudan a los usuarios a gestionar sus horarios, crear recordatorios, crear listas de tareas y organizar tareas. También simplifican las actividades cotidianas y mejoran la eficiencia (Russell & Norvig, 2016).

Recuperación de Información. Los asistentes virtuales brindan acceso rápido a información sobre muchos temas. Por ejemplo, pronósticos del tiempo, actualizaciones de noticias, puntuaciones deportivas y consultas de conocimiento general. También pueden utilizar APIs de búsqueda web y bases de datos de conocimiento para recuperar información más relevante (Russell & Norvig, 2016).

Control del Hogar Inteligente. Los asistentes virtuales se integran con dispositivos inteligentes en el hogar para permitir la automatización controlada por voz. Por ejemplo, se pueden controlar luces, termostatos, cámaras de seguridad y otros electrodomésticos conectados. Mejoran la comodidad y la accesibilidad en el entorno doméstico (Russell & Norvig, 2016).

Comercio Electrónico. Los asistentes virtuales ayudan a los usuarios con recomendaciones de productos, seguimiento de pedidos y consultas de atención al cliente en aplicaciones de comercio electrónico. También facilitan experiencias de compra fluidas y mejoran la participación del usuario (Russell & Norvig, 2016).

Servicio al Cliente. Los asistentes virtuales se despliegan en aplicaciones de servicio al cliente para manejar consultas rutinarias, proporcionar documentación de soporte y escalar problemas complejos a agentes humanos cuando sea necesario. Reducen los tiempos de espera y mejoran la eficiencia del servicio (Russell & Norvig, 2016).

Salud: Los asistentes virtuales se utilizan en entornos de atención médica para programar citas, proporcionar recordatorios de medicación, responder consultas médicas y brindar recomendaciones de salud personalizadas. Apoyan la participación del paciente y la adherencia a los planes de tratamiento (Russell & Norvig, 2016).

Los asistentes virtuales tienen profundas implicaciones para la sociedad, incluyendo:

Conveniencia y Accesibilidad: Los asistentes virtuales hacen que la tecnología sea más accesible para una gama más amplia de usuarios, incluidos aquellos con discapacidades

o habilidades técnicas limitadas. Reducen las barreras de acceso y capacitan a las personas para interactuar de manera más natural con los sistemas digitales (Russell & Norvig, 2016).

Transformación de la Fuerza Laboral: La adopción generalizada de asistentes virtuales está transformando la naturaleza del trabajo, especialmente en industrias como el servicio al cliente y el apoyo administrativo. Si bien automatizan tareas rutinarias, también crean nuevas oportunidades para que los trabajadores humanos se concentren en actividades de mayor valor que requieren creatividad y pensamiento crítico (Russell & Norvig, 2016).

Privacidad y Seguridad de Datos: Los asistentes virtuales plantean preocupaciones sobre la privacidad y la seguridad de los datos debido a la recopilación y procesamiento de información personal. Asegurar medidas robustas de protección de datos y transparencia en las prácticas de manejo de datos es esencial para mantener la confianza del usuario (Russell & Norvig, 2016).

Consideraciones Éticas: Los asistentes virtuales plantean preguntas éticas sobre sesgos, equidad y responsabilidad en la toma de decisiones de IA. Los desarrolladores deben abordar problemas como el sesgo algorítmico y las consecuencias no deseadas para garantizar que los asistentes virtuales satisfagan equitativamente las diversas necesidades de los usuarios (Russell & Norvig, 2016).

2.2.2 Subáreas de la investigación en Inteligencia Artificial

La investigación en inteligencia artificial abarca varios subcampos, cada uno con objetivos y herramientas específicas:

1. **Razonamiento:** Desarrollar capacidades de razonamiento lógico en los sistemas de Inteligencia Artificial.
2. **Representación del Conocimiento:** Representar la información de manera estructurada.
3. **Planificación:** Crear planes para lograr resultados deseados.
4. **Aprendizaje:** Permitir que las máquinas aprendan a partir de datos (por ejemplo, aprendizaje supervisado, no supervisado y por refuerzo).
5. **Procesamiento del Lenguaje Natural:** Enseñar a la inteligencia artificial a entender y generar lenguaje humano.
6. **Percepción:** Mejorar la capacidad de la inteligencia artificial para percibir su entorno (por ejemplo, visión por computadora).

7. Robótica: Integrar la inteligencia artificial con robots físicos.

A medida que la inteligencia artificial continúa evolucionando, debemos abordar preocupaciones éticas, garantizar la seguridad y encontrar un equilibrio entre la automatización y la toma de decisiones humanas. Las preocupaciones éticas incluyen temas de privacidad, sesgo y responsabilidad, que deben gestionarse cuidadosamente para prevenir daños y asegurar la equidad. Garantizar la seguridad implica desarrollar sistemas robustos y seguros que puedan resistir el uso indebido y los ataques adversarios. Además, equilibrar la automatización y la toma de decisiones humanas es crucial para mantener la supervisión y el control humano sobre los procesos críticos (Russell & Norvig, 2016).

El camino hacia el logro de una inteligencia general, donde la inteligencia artificial realice tareas a nivel humano, sigue siendo un desafío emocionante. Los investigadores se esfuerzan por desarrollar algoritmos que puedan entender, aprender y adaptarse de maneras similares a la cognición humana. Esta búsqueda implica superar importantes obstáculos técnicos y comprender las complejidades de la inteligencia humana (Russell & Norvig, 2016).

El impacto de la inteligencia artificial es omnipresente y profundo, influyendo en sectores tan diversos como la salud, las finanzas, la educación y el entretenimiento. En el ámbito de la salud, la inteligencia artificial ayuda a diagnosticar enfermedades con notable precisión, predecir resultados de pacientes y personalizar planes de tratamiento. En finanzas, mejora la detección de fraudes, el comercio algorítmico y el servicio al cliente a través de chatbots. En educación, la inteligencia artificial ofrece experiencias de aprendizaje personalizadas, automatiza tareas administrativas y proporciona información sobre el rendimiento estudiantil. En el entretenimiento, la inteligencia artificial contribuye a la creación de música, la generación de arte visual e incluso el desarrollo de guiones para películas (Russell & Norvig, 2016).

2.3 Aprendizaje Automático

La información de esta sección se obtuvo de (Burkov, 2020; Géron, 2022; Kelleher et al., 2020; Theobald, 2017).

El aprendizaje automático es una subdisciplina de la inteligencia artificial que implica el desarrollo de algoritmos y modelos estadísticos que permiten a las computadoras realizar tareas sin ser programadas explícitamente para ello. En lugar de depender de reglas

predefinidas, los sistemas de aprendizaje automático aprenden de los datos y mejoran su rendimiento con el tiempo (Burkov, 2020).

Los conceptos clave en el aprendizaje automático giran en torno a varios componentes críticos que, colectivamente, permiten a los sistemas aprender de los datos y tomar decisiones informadas (Burkov, 2020).

En primer lugar, los datos son la base del aprendizaje automático. Pueden categorizarse en datos estructurados, como las bases de datos, que están organizados y son fácilmente buscables, y datos no estructurados, como texto e imágenes, que no siguen un formato específico. La calidad y cantidad de los datos son fundamentales, ya que impactan significativamente en el rendimiento y precisión de los modelos de aprendizaje automático. Datos de alta calidad y en cantidad suficiente proporcionan una base robusta para el proceso de aprendizaje, permitiendo que los modelos generalicen mejor a datos nuevos y no vistos (Burkov, 2020).

Los algoritmos son los procedimientos o conjuntos de reglas que un sistema de aprendizaje automático sigue para aprender de los datos. Estos algoritmos pueden variar desde modelos simples de regresión lineal hasta complejas redes neuronales. La elección del algoritmo depende del tipo específico de tarea y la naturaleza de los datos. Por ejemplo, los árboles de decisión pueden usarse para tareas de clasificación, mientras que las redes neuronales convolucionales son adecuadas para el reconocimiento de imágenes (Burkov, 2020).

Los modelos son representaciones de lo que el sistema de aprendizaje automático ha aprendido de los datos. Capturan los patrones y relaciones identificados durante el proceso de aprendizaje. Una vez entrenado, un modelo puede hacer predicciones o tomar decisiones basadas en nuevos datos de entrada. Por ejemplo, un modelo entrenado podría predecir los precios de viviendas basándose en características como la ubicación, tamaño y comodidades. La capacidad de un modelo para generalizar desde los datos de entrenamiento a nuevos datos es crucial para su efectividad (Burkov, 2020).

El entrenamiento es el proceso de enseñar al modelo de aprendizaje automático utilizando un conjunto de datos. Durante el entrenamiento, el modelo ajusta iterativamente sus parámetros para minimizar errores y mejorar la precisión. Esto implica alimentar al modelo con datos de entrenamiento y utilizar técnicas de optimización para afinar los

parámetros del modelo. El objetivo es reducir la diferencia entre las predicciones del modelo y los resultados reales, a menudo medida mediante una función de pérdida (Burkov, 2020).

Las características son propiedades individuales medibles o características de los datos utilizados. En el contexto del aprendizaje automático, la ingeniería de características es el proceso de seleccionar y transformar datos brutos en características relevantes que representen mejor el problema subyacente a los modelos predictivos. Por ejemplo, en un conjunto de datos de viviendas, las características podrían incluir el número de habitaciones, la antigüedad de la casa y su proximidad a escuelas y centros comerciales. Una ingeniería de características efectiva puede mejorar significativamente el rendimiento del modelo (Burkov, 2020).

Finalmente, las etiquetas son las variables de salida o destino que el modelo intenta predecir. En el aprendizaje supervisado, cada pieza de datos de entrenamiento incluye una etiqueta, que sirve como la verdad fundamental que el modelo intenta aprender. Por ejemplo, en un sistema de detección de spam, los correos electrónicos se etiquetarían como 'spam' o 'no spam'. El modelo utiliza estas etiquetas durante el entrenamiento para entender la correspondencia entre las características de entrada y la salida objetivo. Un etiquetado preciso es esencial para construir modelos fiables, ya que las etiquetas incorrectas pueden llevar a un mal rendimiento del modelo (Burkov, 2020).

Comprender y utilizar eficazmente estos conceptos clave es esencial para desarrollar sistemas robustos de aprendizaje automático que puedan analizar datos con precisión y hacer predicciones fiables.

El aprendizaje automático abarca varios tipos distintos, cada uno adaptado a tipos específicos de tareas y estructuras de datos, proporcionando una variedad de metodologías para extraer información significativa y hacer predicciones.

El aprendizaje supervisado es uno de los tipos más comunes de aprendizaje automático. Implica entrenar un modelo con un conjunto de datos que incluye tanto los datos de entrada como la salida correspondiente correcta, conocida como datos etiquetados. El objetivo principal es que el modelo aprenda una correspondencia de entradas a salidas, permitiéndole hacer predicciones precisas sobre datos nuevos y no vistos. El aprendizaje supervisado es particularmente efectivo en tareas donde hay datos históricos disponibles y se puede establecer una clara relación entre entrada y salida. Ejemplos comunes de aprendizaje supervisado incluyen la regresión, que predice resultados continuos (como los

precios de viviendas), y la clasificación, que predice resultados discretos (como si un correo electrónico es spam o no) (Burkov, 2020).

El aprendizaje no supervisado, por otro lado, trabaja con conjuntos de datos que no tienen salidas etiquetadas. El modelo debe encontrar de manera independiente patrones, estructuras y relaciones dentro de los datos. Este tipo de aprendizaje es útil en el análisis exploratorio de datos, donde el objetivo es descubrir patrones ocultos o estructuras intrínsecas en los datos. Las técnicas comúnmente utilizadas en el aprendizaje no supervisado incluyen la agrupación, que agrupa puntos de datos similares (como la segmentación de clientes en marketing), y la reducción de dimensionalidad, que reduce el número de características en un conjunto de datos mientras se preservan sus características esenciales (como el Análisis de Componentes Principales) (Burkov, 2020).

El aprendizaje semisupervisado cierra la brecha entre el aprendizaje supervisado y el no supervisado al aprovechar una combinación de datos etiquetados y no etiquetados. Típicamente, una pequeña porción del conjunto de datos está etiquetada, y una gran porción permanece sin etiquetar. El modelo utiliza los datos etiquetados para hacer predicciones iniciales, que luego se refinan y mejoran utilizando los datos no etiquetados. Este enfoque es particularmente útil cuando obtener una gran cantidad de datos etiquetados es costoso o lleva mucho tiempo, pero hay abundantes datos no etiquetados disponibles. El aprendizaje semisupervisado puede mejorar significativamente el proceso de aprendizaje al aprovechar al máximo los datos etiquetados limitados, mejorando la precisión y el rendimiento del modelo (Burkov, 2020).

El aprendizaje por refuerzo es un enfoque dinámico donde el modelo aprende interactuando con su entorno y recibiendo retroalimentación en forma de recompensas o penalizaciones basadas en sus acciones. El objetivo es desarrollar una estrategia, o política, que maximice la recompensa acumulativa a lo largo del tiempo. Este tipo de aprendizaje se utiliza a menudo en situaciones donde un agente debe tomar una secuencia de decisiones, como en la robótica, donde un robot aprende a navegar en un entorno, o en el juego, donde un algoritmo aprende a jugar un juego como el ajedrez o Go. Los algoritmos de aprendizaje por refuerzo están diseñados para manejar tareas complejas donde la acción óptima no es inmediatamente obvia y debe descubrirse mediante ensayo y error (Burkov, 2020).

Estos diversos tipos de aprendizaje automático proporcionan herramientas poderosas para analizar datos y hacer predicciones, cada uno adecuado para diferentes escenarios y desafíos.

2.3.1 Aprendizaje no supervisado

El aprendizaje no supervisado, una rama vital del aprendizaje automático, opera bajo la premisa de entrenar algoritmos con datos no etiquetados, es decir, sin categorías o etiquetas asociadas. Estos datos solo comprenden características de entrada, lo que obliga a los algoritmos a discernir de manera autónoma patrones o agrupaciones significativas sin guía externa. A menudo empleado en el análisis exploratorio de datos, el aprendizaje no supervisado sirve como herramienta para descubrir la estructura latente dentro de los conjuntos de datos. Al revelar patrones ocultos o agrupaciones, permite a los analistas obtener valiosas percepciones sobre la distribución inherente de los datos, ofreciendo una comprensión más profunda de sus características subyacentes (Géron, 2022).

Además, los algoritmos de aprendizaje no supervisado desempeñan un papel fundamental en las tareas de extracción de características y reducción de dimensionalidad. Al identificar características significativas o reducir la dimensionalidad del conjunto de datos, el aprendizaje no supervisado facilita la simplificación de estructuras de datos complejas. Este proceso no solo mejora la interpretabilidad, sino que también incrementa el rendimiento de las tareas subsiguientes al enfocarse en los aspectos más destacados de los datos (Géron, 2022).

La agrupación destaca como una aplicación prominente del aprendizaje no supervisado, donde los algoritmos agrupan puntos de datos similares en clústeres basados en similitudes inherentes. El objetivo principal de los algoritmos de agrupación es maximizar la similitud dentro del clúster mientras se minimiza la similitud entre clústeres, delineando así agrupaciones distintivas dentro del conjunto de datos. Esto permite a los analistas discernir agrupaciones o asociaciones naturales presentes en los datos, ayudando en tareas de segmentación y reconocimiento de patrones (Géron, 2022).

Asimismo, el aprendizaje no supervisado encuentra utilidad en la detección de anomalías, donde los algoritmos identifican patrones irregulares o inesperados que se desvían de la norma. Las anomalías, que van desde errores y valores atípicos hasta fenómenos novedosos, representan puntos de datos que merecen una investigación más profunda. Al detectar tales anomalías, el aprendizaje no supervisado mejora la robustez y la

fiabilidad de los procesos de análisis de datos, facilitando la identificación de aberraciones y posibles áreas de preocupación dentro de los conjuntos de datos (Géron, 2022).

Los algoritmos de aprendizaje no supervisado abarcan una amplia gama de técnicas destinadas a extraer información valiosa y patrones de datos no etiquetados. Los algoritmos de agrupación forman la base del aprendizaje no supervisado, con la tarea de agrupar puntos de datos similares en clústeres basados en su proximidad en el espacio de características. Los métodos de agrupación populares incluyen k-means, que particiona los datos en k clústeres optimizando iterativamente los centroides del clúster, la agrupación jerárquica, que construye una jerarquía en forma de árbol de clústeres, y algoritmos basados en densidad como DBSCAN, que identifican regiones densas de puntos de datos. Estos algoritmos son ampliamente utilizados en diversos dominios, incluyendo la segmentación de clientes, la segmentación de imágenes y la detección de anomalías (Géron, 2022).

Las técnicas de reducción de dimensionalidad ofrecen otra herramienta poderosa en el arsenal del aprendizaje no supervisado, enfocándose en reducir el número de características en los datos mientras se preserva la mayor cantidad posible de información. El Análisis de Componentes Principales es una técnica ampliamente utilizada que identifica ejes ortogonales que representan los componentes principales de variación en los datos, facilitando la reducción de dimensionalidad sin una pérdida significativa de información. Otros métodos como t-SNE (t-distributed Stochastic Neighbor Embedding) y los autoencoders proporcionan enfoques alternativos para capturar la estructura subyacente de datos de alta dimensionalidad, particularmente útiles en tareas de visualización y compresión de datos (Géron, 2022).

Los algoritmos de aprendizaje de reglas de asociación exploran la revelación de relaciones entre variables dentro del conjunto de datos, descubriendo patrones frecuentes o asociaciones que existen entre los elementos de datos. El algoritmo Apriori, un ejemplo clásico en este contexto, identifica conjuntos de elementos frecuentes en datos de transacciones, permitiendo el análisis de cestas de mercado y la extracción de información procesable sobre el comportamiento del consumidor. El aprendizaje de reglas de asociación encuentra aplicaciones en campos diversos, incluyendo el comercio minorista, la salud y los sistemas de recomendación, donde comprender las asociaciones subyacentes es crucial para la toma de decisiones y la modelización predictiva (Géron, 2022).

Los modelos generativos representan otra área fascinante del aprendizaje no supervisado, enfocándose en aprender la distribución de probabilidad subyacente de los datos. Estos modelos pueden generar nuevas muestras que se asemejan a los datos de entrenamiento, ofreciendo capacidades valiosas para la aumentación de datos, la síntesis y la detección de anomalías. Los Modelos de Mezcla Gaussianos, por ejemplo, representan datos como una mezcla de varias distribuciones gaussianas, mientras que los Autoencoders Variacionales y las Redes Generativas Adversarias proporcionan técnicas avanzadas para generar muestras de datos realistas. Los modelos generativos encuentran aplicaciones en la generación de imágenes, la generación de textos y la aumentación de datos, contribuyendo a avances en áreas como la visión por computadora, el procesamiento del lenguaje natural y el arte generativo (Géron, 2022).

El aprendizaje no supervisado abarca una rica variedad de algoritmos, cada uno ofreciendo capacidades únicas para extraer percepciones y patrones de datos no etiquetados. Desde la agrupación y la reducción de dimensionalidad hasta el aprendizaje de reglas de asociación y la modelización generativa, estas técnicas desempeñan un papel crucial en la exploración de datos, el descubrimiento de conocimiento y la toma de decisiones en una amplia gama de dominios y aplicaciones (Géron, 2022).

Las aplicaciones del aprendizaje no supervisado abarcan diversos dominios, reflejando su versatilidad para descubrir patrones ocultos y percepciones de datos no etiquetados, ofreciendo contribuciones invaluableles a los procesos de toma de decisiones y al descubrimiento de conocimientos en diversas industrias.

La segmentación de mercados representa una aplicación fundamental, donde los algoritmos de aprendizaje no supervisado disciernen grupos distintos de clientes que exhiben comportamientos o preferencias de compra similares. Al segmentar a los clientes en clústeres significativos, las empresas pueden adaptar estrategias de marketing, ofertas de productos y campañas promocionales para dirigirse efectivamente a segmentos específicos. Este enfoque mejora la satisfacción del cliente, maximiza el retorno de la inversión en marketing y fomenta relaciones duraderas con los clientes (Géron, 2022).

En la agrupación de imágenes, las técnicas de aprendizaje no supervisado son fundamentales para organizar grandes colecciones de imágenes en grupos cohesivos basados en similitudes visuales. Esto facilita la recuperación eficiente de imágenes, la organización de contenido y el etiquetado automatizado, agilizando así los sistemas de

gestión de contenido y mejorando las experiencias de usuario en aplicaciones que van desde bases de datos de imágenes y plataformas de redes sociales hasta sitios web de comercio electrónico (Géron, 2022).

La detección de anomalías, otra aplicación destacada, implica identificar patrones anómalos o valores atípicos en los datos que se desvían significativamente de la norma. En dominios como la ciberseguridad, las finanzas y la manufactura, los algoritmos de aprendizaje no supervisado desempeñan un papel crucial en la detección de actividades inusuales en el tráfico de red, las transacciones financieras o los procesos de producción. La detección temprana de anomalías permite intervenciones oportunas para mitigar riesgos, prevenir fraudes, optimizar la eficiencia operativa y garantizar la calidad y seguridad del producto (Géron, 2022).

El modelado de temas surge como una herramienta poderosa para extraer temas subyacentes de grandes corpora de texto, permitiendo una comprensión y organización exhaustiva de los datos textuales. Al identificar automáticamente temas o materias recurrentes dentro de los documentos, el aprendizaje no supervisado facilita tareas de análisis de contenido, recuperación de información y resumen de documentos. Esto tiene profundas implicaciones en diversos dominios, incluyendo la investigación académica, los sistemas de recomendación de contenido y el análisis de sentimientos en el monitoreo de redes sociales (Géron, 2022).

Además, el análisis del comportamiento del cliente aprovecha el aprendizaje no supervisado para analizar patrones de navegación web, datos de clics o historiales de transacciones, descubriendo patrones y tendencias intrincados en el comportamiento del usuario. Al identificar comportamientos, preferencias e intereses comunes de los usuarios, las empresas pueden mejorar la usabilidad del sitio web, personalizar las comunicaciones de marketing y optimizar las estrategias de participación del cliente. Esto, a su vez, fomenta la satisfacción, lealtad y retención del cliente, impulsando el crecimiento de ingresos y la competitividad en el mercado (Géron, 2022).

2.3.2 Aprendizaje supervisado

El aprendizaje supervisado es muy importante dentro del aprendizaje automático. En este, los algoritmos se nutren de datos etiquetados para adquirir la capacidad de realizar predicciones precisas. En este contexto, el conjunto de datos de entrenamiento incluye tanto los datos de entrada como las salidas correspondientes, lo que permite al algoritmo crear

una relación entre ambos y, así, realizar predicciones acertadas sobre nuevos datos no vistos (Kelleher et al., 2020).

Dentro del contexto del aprendizaje supervisado, diversos conceptos clave sientan las bases para la construcción de modelos predictivos efectivos. Por ejemplo, la existencia de datos etiquetados resulta muy importante, donde cada punto de datos contiene características de entrada y etiquetas de salida correspondientes. Este proceso de etiquetado permite al algoritmo comprender la relación entre las entradas y las salidas, capacitándolo para realizar predicciones sobre nuevos datos no vistos. Por ejemplo, en una tarea de diagnóstico médico, los datos de cada paciente podrían incluir diversas métricas de salud como características, con el diagnóstico correspondiente como etiqueta (Kelleher et al., 2020).

Durante el proceso de entrenamiento, el algoritmo aprende de estos datos etiquetados ajustando iterativamente sus parámetros internos. De esta forma, el algoritmo minimiza la diferencia entre sus predicciones y las etiquetas verdaderas. Esta adaptación se logra típicamente mediante algoritmos de optimización. Estos algoritmos buscan minimizar una función de pérdida seleccionada, que cuantifica la discrepancia entre los valores predichos y reales. La elección de la función de pérdida depende de la naturaleza de la tarea. Para tareas de regresión, se utiliza comúnmente el Error Cuadrático Medio. Para las tareas de clasificación, se suelen emplear otras técnicas como la Pérdida de Entropía Cruzada (Kelleher et al., 2020).

Luego del entrenamiento, el rendimiento del modelo se evalúa a través de la evaluación del modelo utilizando un conjunto de datos separado. Este conjunto se denomina conjunto de validación o prueba. Este paso de evaluación es crucial para evaluar qué tan bien se generaliza el modelo a nuevos datos no vistos. Considerando más allá de los ejemplos en los que fue entrenado. Evaluar el modelo en un conjunto de datos distinto ayuda a identificar cualquier problema de sobreajuste o subajuste. Esto también proporciona información sobre su rendimiento en el mundo real (Kelleher et al., 2020).

Además, se emplean diversas técnicas para mejorar la efectividad de los modelos de aprendizaje supervisado. La ingeniería de características implica seleccionar o transformar características de entrada. Esto se usa para mejorar el rendimiento del modelo y capturar información relevante de los datos. Además, se utilizan métodos de regularización como la

regularización L1 y L2 que previenen el sobreajuste penalizando modelos excesivamente complejos (Kelleher et al., 2020).

El aprendizaje supervisado en el aprendizaje automático se basa en datos etiquetados, un proceso de entrenamiento que ajusta los parámetros del modelo para minimizar errores de predicción y evaluación del modelo. Luego se evalúa el rendimiento de la generalización. Estos conceptos constituyen la base para construir modelos predictivos precisos en diversos campos y aplicaciones (Kelleher et al., 2020).

El aprendizaje supervisado abarca dos tipos principales de problemas: regresión y clasificación. En la regresión, el objetivo es predecir una variable de salida continua basada en características de entrada. Por ejemplo, en la predicción de precios de viviendas, se utilizan características como tamaño, ubicación y número de habitaciones para predecir el precio. Por otro lado, la clasificación implica predecir una variable de salida discreta. Ejemplos incluyen clasificar correos electrónicos como spam o no spam o reconocer dígitos escritos a mano. Estos dos tipos de tareas de aprendizaje supervisado constituyen la base para una amplia gama de aplicaciones en diversos dominios, proporcionando percepciones y predicciones valiosas a partir de datos etiquetados (Kelleher et al., 2020).

2.3.2.1 Regresión

La regresión abarca varios aspectos fundamentales para construir modelos predictivos precisos. Una característica central de las tareas de regresión es la presencia de una variable de salida continua. Esto implica que el valor predicho puede caer dentro de un rango infinito. Esta característica es evidente en varios escenarios del mundo real, como predecir precios de viviendas, donde el precio predicho puede ser cualquier número real dentro de un rango especificado (Kelleher et al., 2020).

La utilización de características de entrada es central en la modelización de regresión. Estas características (también conocidas como variables independientes), se usan para hacer predicciones sobre la variable de salida. Estas características pueden abarcar una amplia cantidad de tipos de datos, incluyendo datos numéricos, categóricos o incluso temporales. Esto proporciona información valiosa para el proceso de predicción. Por ejemplo, al predecir precios de viviendas, características como el área en pies cuadrados, la ubicación y el número de dormitorios sirven como variables de entrada que influyen en el precio predicho (Kelleher et al., 2020).

Los modelos de regresión dependen de capturar la relación entre las características de entrada y la variable de salida. Esta relación se aprende a partir de los datos de entrenamiento. Esto permite al modelo generalizar y hacer predicciones precisas sobre nuevos datos no vistos. Se emplean diversas técnicas matemáticas, como la regresión lineal, la regresión polinomial o métodos más avanzados. Estos incluyen la regresión de vectores de soporte y las redes neuronales, para capturar y modelar esta relación de manera efectiva (Kelleher et al., 2020).

Evaluar el rendimiento de un modelo de regresión es muy importante para evaluar sus capacidades predictivas. Muchas métricas se utilizan comúnmente para este propósito. Algunas son el Error Cuadrático Medio, el Error Cuadrático Medio Raíz, el Error Absoluto Medio y el coeficiente de determinación R-cuadrado. Estas métricas cuantifican el grado en que las predicciones del modelo coinciden con los valores reales en el conjunto de datos. Esto proporciona información sobre su precisión y confiabilidad (Kelleher et al., 2020).

También, el análisis de regresión permite identificar tendencias y patrones dentro de los datos. Esto facilita percepciones valiosas sobre la dinámica subyacente y las relaciones entre variables. Esta comprensión es particularmente importante en campos como la economía, las finanzas y la ciencia ambiental, donde predecir y comprender resultados continuos es de suma importancia (Kelleher et al., 2020).

Los algoritmos de regresión abarcan muchas técnicas utilizadas para modelar la relación entre características de entrada y una variable de salida continua. Uno de los métodos más simples y ampliamente empleados es la Regresión Lineal. Esta asume que existe una relación lineal entre las características de entrada y la variable de salida. Luego busca encontrar la mejor línea recta que minimiza el error de predicción. Esto la hace adecuada para tareas donde la relación es aproximadamente lineal (Kelleher et al., 2020).

La Regresión polinomial extiende la regresión lineal al juntar términos polinomiales, como cuadráticos o cúbicos. Estos capturan relaciones no lineales entre las características de entrada y la variable de salida. Esta nueva técnica permite modelar relaciones más complejas, lo que la hace útil en escenarios donde la suposición de que las relaciones son lineales no es suficiente (Kelleher et al., 2020).

Las técnicas de regularización, como la Regresión Ridge y la Regresión Lasso, ofrecen soluciones a los problemas comunes de sobreajuste en modelos de regresión que se mencionó antes. La regresión Ridge, o regularización L2, agrega un término de penalización

a la función objetivo de regresión lineal. Esto reduce la sensibilidad del modelo a los datos de entrenamiento. Por otro lado, la regresión Lasso, o regularización L1, tiende a producir modelos dispersos al reducir algunos coeficientes a cero, facilitando la selección de características en conjuntos de datos de alta dimensión (Kelleher et al., 2020).

La Regresión ElasticNet combina las penalizaciones de la regresión Ridge y Lasso. Esto ofrece un enfoque equilibrado que aborda tanto el sobreajuste como la selección de características. Esto la hace particularmente útil al tratar conjuntos de datos que contienen múltiples características correlacionadas (Kelleher et al., 2020).

La Regresión de Vectores de Soporte extiende los principios de las máquinas de vectores de soporte a problemas de regresión. Esta técnica busca encontrar un hiperplano que se ajuste mejor a los datos dentro de un margen de tolerancia especificado. Esto la hace efectiva en espacios de alta dimensión donde las técnicas de regresión tradicionales pueden tener dificultades (Kelleher et al., 2020).

Cada uno de estos algoritmos de regresión tiene ventajas y desventajas. Es necesario analizar cuál de estos es adecuado para cada conjunto de datos y tarea objetivo para el modelado. En el fondo, las técnicas que se consideran en el proyecto son técnicas de regresión, ya que usan redes neuronales convolucionales (Kelleher et al., 2020).

2.3.2.2 Clasificación

En las tareas de clasificación dentro del aprendizaje supervisado, la variable de salida asume una naturaleza finita. Esto significa que los posibles valores que la variable de salida puede asumir se puede contar. Esta característica es muy importante en escenarios como la clasificación de correos electrónicos (spam o no spam), la identificación de especies basada en características de plantas o el reconocimiento de dígitos escritos a mano (Kelleher et al., 2020).

En este caso, las características de entrada sirven como base para predecir las etiquetas de clase para los puntos de datos. Estas características pueden abarcar una variedad de tipos, incluidos numéricos, categóricos o una combinación de ambos, ofreciendo información valiosa para el proceso de clasificación (Kelleher et al., 2020).

Un muy importante de los modelos de clasificación implica aprender los límites de decisión dentro del espacio de características. Esto puede ayudar a separar efectivamente diferentes clases. Al discernir los límites de decisión óptimos, el modelo puede clasificar

con precisión nuevos puntos de datos no vistos. Este objetivo busca maximizar la separación entre clases, lo que asegura un rendimiento de clasificación más resistente en diversos conjuntos de datos (Kelleher et al., 2020).

Por otro lado, la evaluación de los modelos de clasificación requiere la aplicación de varias métricas para medir el rendimiento. Métricas como la precisión, la sensibilidad, la puntuación F1 y el área bajo la curva Característica de Operación del Receptor se utilizan comúnmente. Estas métricas proporcionan percepciones sobre el grado en que las predicciones del modelo coinciden con las etiquetas de clase reales. Esto facilita las evaluaciones informadas de la eficacia de la clasificación y ayuda en la refinación del modelo (Kelleher et al., 2020).

Los algoritmos de clasificación ofrecen diversas estrategias para categorizar puntos de datos en clases o categorías distintas. Estos algoritmos varían en complejidad, enfoque y principios conceptuales. Para su correcto uso, es necesario elegir el método más adecuado según las características del conjunto de datos y los requisitos específicos de la tarea de clasificación (Kelleher et al., 2020).

La Regresión Logística es un algoritmo de clasificación binaria ampliamente utilizado. Esta técnica modela la probabilidad de que un punto de datos pertenezca a una cierta clase utilizando la función logística. Es particularmente efectiva cuando la relación entre las características de entrada y el resultado binario puede representarse por un límite de decisión lineal. La regresión logística no solo proporciona predicciones, sino que también estima la probabilidad de pertenencia a cada clase, lo que la hace valiosa en escenarios donde entender la incertidumbre es esencial (Kelleher et al., 2020).

Los Árboles de Decisión son otra técnica de clasificación. Estos ofrecen un enfoque versátil e intuitivo para la clasificación, construyendo una estructura similar a un árbol donde cada nodo interno representa una decisión basada en una característica. Cada nodo hoja corresponde a una etiqueta de clase. Los árboles de decisión dividen recursivamente el espacio de características en regiones. Esto permite una interpretación y visualización sencilla de los procesos de toma de decisiones. Aunque los árboles de decisión son propensos al sobreajuste, utilizar métodos en conjunto como Random Forest hacen que este riesgo disminuya. Esto se logra agregando predicciones de múltiples árboles de decisión. Esto, a su vez, mejora la precisión y resistencia de la clasificación del modelo (Kelleher et al., 2020).

Las Máquinas de Vectores de Soporte son clasificadores muy potentes que buscan identificar el hiperplano óptimo que separa diferentes clases en el espacio de características. Esta técnica busca maximizar el margen entre clases, lo que las hace efectivas en escenarios con clases bien separadas y límites de decisión complejos. Además, esta técnica puede manejar datos de alta dimensión de manera eficiente. Esto la hace adecuada para aplicaciones como la clasificación de texto y el reconocimiento de imágenes (Kelleher et al., 2020).

Otro algoritmo de aprendizaje basado en instancias utilizado para la clasificación es k-vecinos más cercanos. Este clasifica un punto de datos identificando sus k vecinos más cercanos en el espacio de características y asignando la etiqueta de clase mayoritaria entre ellos. Aunque este algoritmo es simple y fácil de implementar, puede sufrir de ineficiencia computacional y sensibilidad a la elección del número k. Sin embargo, sigue siendo una opción popular por su simplicidad y flexibilidad (Kelleher et al., 2020).

Naive Bayes es otro algoritmo de clasificación probabilístico basado en el teorema de Bayes. Este asume que las características son independientes condicionalmente dada la etiqueta de clase. A pesar de su suposición simplista, Naive Bayes a menudo funciona sorprendentemente bien en la práctica, especialmente en tareas de clasificación de texto y detección de spam. Es computacionalmente eficiente y resistente a características irrelevantes, lo que lo hace muy adecuado para conjuntos de datos a gran escala con espacios de características de alta dimensión (Kelleher et al., 2020).

Existen muchos otros algoritmos de clasificación que ofrecen un conjunto de herramientas muy amplio para resolver muchas de tareas de clasificación. Cada uno, al igual que en el caso de regresión, tiene ventajas y desventajas (Kelleher et al., 2020).

2.3.3 Aprendizaje semi-supervisado

El aprendizaje semi-supervisado es una técnica un poco distinta dentro del espectro de paradigmas de aprendizaje automático. Esta técnica combina metodologías de aprendizaje supervisado y no supervisado. A diferencia del aprendizaje supervisado, donde los algoritmos dependen únicamente de datos etiquetados, o el aprendizaje no supervisado que opera exclusivamente con datos no etiquetados, el aprendizaje semi-supervisado aprovecha una combinación de datos etiquetados y no etiquetados para entrenar modelos (Theobald, 2017).

La integración de datos etiquetados y no etiquetados distingue al aprendizaje semi-supervisado como un enfoque híbrido. Esto ofrece ventajas distintas sobre las técnicas de aprendizaje automático supervisadas y no supervisadas. Al incorporar datos etiquetados, los algoritmos de aprendizaje semi-supervisado pueden aprovechar la información contenida en instancias etiquetadas para guiar el proceso de aprendizaje. Esto facilita la extracción de patrones y relaciones significativas del conjunto más grande de datos no etiquetados (Theobald, 2017).

En términos prácticos, los escenarios de aprendizaje semi-supervisado a menudo surgen en situaciones donde adquirir datos etiquetados no es muy conveniente o consume mucho tiempo o es poco práctico pero los datos no etiquetados son abundantes o fácilmente disponibles. Esto ocurre con frecuencia en problemas como el procesamiento del lenguaje natural o el reconocimiento de imágenes y el diagnóstico médico. En estos problemas, obtener datos etiquetados puede requerir la anotación de expertos o esfuerzos de etiquetado manual. Por lo tanto, el aprendizaje semi-supervisado presenta una solución pragmática para maximizar la utilidad de los recursos de datos disponibles mientras se minimiza la carga de los requisitos de etiquetado (Theobald, 2017).

Además, los algoritmos de aprendizaje semi-supervisado muestran resistencia a datos etiquetados con ruido o incompletos. Esto se debe a que la integración de instancias no etiquetadas puede ayudar a mitigar el impacto de errores o inconsistencias en el etiquetado. Esta robustez mejora las capacidades de generalización de los modelos semi-supervisados y les permite manejar de manera efectiva escenarios de datos del mundo real que se caracterizan por un etiquetado imperfecto o datos etiquetados dispersos (Theobald, 2017).

En el aprendizaje semi-supervisado, los algoritmos aprovechan las fortalezas complementarias de datos etiquetados, que comprenden pares de entrada-salida, y datos no etiquetados, que solo contienen características de entrada. Esta relación capacita a los algoritmos para extraer ideas de las instancias etiquetadas. También aprovechan simultáneamente la abundancia de datos no etiquetados para discernir patrones y estructuras subyacentes dentro del conjunto de datos (Theobald, 2017).

Una característica distintiva del aprendizaje semi-supervisado es su hábil utilización de los amplios conjuntos de datos no etiquetados prevalentes en los conjuntos de datos del mundo real. Esta incorporación poderosa de instancias no etiquetadas durante el

entrenamiento permite a los algoritmos mejorar sus capacidades predictivas. Esto es así principalmente en escenarios donde los datos etiquetados son dispersos o prohibitivamente caros de adquirir. Por ejemplo, en el reconocimiento de voz, el aprendizaje semi-supervisado puede aprovechar grandes cantidades de datos de audio no etiquetados junto con muestras etiquetadas limitadas para mejorar la precisión y robustez del modelo. Además, el aprendizaje semi-supervisado representa una unión importante de las metodologías de aprendizaje supervisado y no supervisado. Mientras aprovechan los datos etiquetados para guiar tareas predictivas, estos algoritmos también exploran el territorio inexplorado de los datos no etiquetados para descubrir estructuras y relaciones latentes. Por ejemplo, en el procesamiento de imágenes médicas, los algoritmos de aprendizaje semi-supervisado pueden combinar imágenes etiquetadas con una multitud de exploraciones no etiquetadas para identificar patrones sutiles que indican la progresión de la enfermedad o la respuesta al tratamiento (Theobald, 2017).

Hay muchos enfoques que crean el panorama del aprendizaje semi-supervisado, ofreciendo soluciones personalizadas para diferentes escenarios de datos y objetivos de aprendizaje. Los algoritmos de autoentrenamiento mejoran iterativamente el rendimiento del modelo mediante la pseudoetiquetación de instancias no etiquetadas basándose en predicciones actuales del modelo. Esto expande gradualmente el conjunto de datos etiquetados. Los algoritmos de co-entrenamiento, por otro lado, aprovechan múltiples vistas o representaciones de los datos para mejorar la resistencia y precisión del modelo. Por ejemplo, en el procesamiento del lenguaje natural, los algoritmos de co-entrenamiento pueden utilizar texto y metadatos como vistas distintas de los datos para mejorar el análisis de sentimientos o las tareas de modelado de temas. Además, las variantes semi-supervisadas de los algoritmos de aprendizaje supervisado tradicionales como las máquinas de vectores de soporte semi-supervisadas o los modelos de aprendizaje profundo semi-supervisados adaptan técnicas establecidas para incorporar datos no etiquetados. Esto ofrece una escalabilidad y rendimiento mejorados en aplicaciones a gran escala (Theobald, 2017).

Aunque en este proyecto no se abordan formas de aprendizaje no supervisadas o semi-supervisadas, el aprendizaje semi-supervisado tiene un potencial transformador de aprovechar tanto datos etiquetados como no etiquetados en los esfuerzos de aprendizaje automático.

2.3.4 Retos del aprendizaje automático

El aprendizaje automático enfrenta varios desafíos significativos que pueden afectar la efectividad, confiabilidad y consideraciones éticas de los modelos. Uno de los principales desafíos es la calidad y cantidad de datos. El rendimiento de un modelo depende en gran medida de tener muchos datos y de alta calidad. Datos insuficientes o de mala calidad pueden llevar a modelos inexactos que no generalizan bien a nuevos datos no vistos. Además, el preprocesamiento y limpieza de datos son pasos críticos para garantizar que los datos alimentados en los modelos sean confiables y relevantes. En (Kaddour et al., 2022) se puede ver una compilación de los problemas abiertos del aprendizaje automático reciente. A continuación, damos un resumen de algunos de estos (Theobald, 2017).

El sobreajuste y el subajuste, mencionados anteriormente, también son problemas críticos en el aprendizaje automático. El sobreajuste ocurre cuando un modelo aprende demasiado bien los datos de entrenamiento, incluido el ruido y las fluctuaciones menores, lo que conduce a un rendimiento deficiente en nuevos datos no vistos. Esto suele ser el resultado de un modelo demasiado complejo en relación con la cantidad de datos de entrenamiento. Por el contrario, el subajuste ocurre cuando un modelo es demasiado simplista para capturar los patrones subyacentes en los datos, lo que resulta en un rendimiento deficiente tanto en los datos de entrenamiento como en los nuevos datos. Equilibrar la complejidad del modelo y la cantidad de datos de entrenamiento es crucial para evitar estos problemas.

Garantizar la imparcialidad en los modelos de aprendizaje automático es otro desafío importante. Los modelos pueden perpetuar involuntariamente o incluso exacerbar los sesgos existentes presentes en los datos, lo que lleva a resultados injustos o discriminatorios. Aunque los algoritmos funcionan exactamente como se espera de ellos, esto puede tener serias implicaciones éticas y legales, especialmente en aplicaciones sensibles como la contratación, los préstamos o la aplicación de la ley. Se emplean técnicas como la detección de sesgos, las restricciones de equidad y conjuntos de datos de entrenamiento diversos para mitigar estos riesgos (Theobald, 2017).

Los recursos computacionales representan otro desafío importante. Esto es así especialmente para tareas de aprendizaje automático que involucran conjuntos de datos grandes y modelos complejos como el aprendizaje profundo, como es el caso de este proyecto. Estas tareas requieren una potencia computacional, memoria y almacenamiento sustanciales, que pueden ser costosos y requieren muchos recursos. Los avances en

hardware, como las GPU y las UTP, así como los marcos de computación distribuida, han aliviado algunos de estos problemas. Sin embargo, las limitaciones de recursos económicos aún pueden limitar la viabilidad y escalabilidad de ciertos proyectos de aprendizaje automático (Theobald, 2017).

Además, otro desafío es la interpretabilidad y explicabilidad de los modelos de aprendizaje automático. Muchos modelos avanzados, como las redes neuronales profundas, funcionan como cajas negras. Estas tienen mecanismos internos complejos que son difíciles de entender. Esta falta de transparencia puede ser problemática, especialmente en aplicaciones donde comprender el razonamiento detrás de las decisiones de un modelo es muy importante. Un ejemplo es el cuidado de la salud o las finanzas. Los esfuerzos para desarrollar modelos más interpretables y herramientas para la explicación de modelos son áreas de investigación en curso (Theobald, 2017).

2.4 Aprendizaje profundo

2.4.1 Redes Neuronales Convolucionales

Una red neuronal convolucional es un tipo de red neuronal profunda diseñada específicamente para procesar y analizar datos que tienen una estructura de cuadrícula, como imágenes. Este tipo de red ha demostrado ser particularmente eficaz en tareas de visión por computadora, como la clasificación de imágenes, la detección de objetos, el reconocimiento facial, entre otras (Theobald, 2017).

Un componente importante de las redes neuronales convolucionales son las capas convolucionales. Estas son el corazón de una red neuronal convolucional. Estas capas aplican una operación matemática llamada convolución a la imagen de entrada. En este proceso, se utilizan filtros que son matrices de valores, para recorrer la imagen y detectar patrones locales, como bordes, texturas o formas. A medida que el filtro se desplaza por la imagen, se generan mapas de características que representan la presencia de ciertas características detectadas en diferentes ubicaciones de la imagen. Cada filtro está diseñado para detectar un patrón específico, y la red puede aprender múltiples filtros para identificar diferentes características en las imágenes (Theobald, 2017).

Este componente es solo una parte de uno más grande llamado capas de agrupación, también conocidas como capas de submuestreo o capas de pooling. Estas se utilizan para reducir la dimensionalidad de los mapas de características, es decir, para disminuir el tamaño de las representaciones sin perder información importante. Una técnica común es el

max pooling, que divide el mapa de características en pequeñas regiones y selecciona el valor máximo dentro de cada región. Esto no solo reduce el tamaño de los datos, sino que también proporciona invariancia a pequeñas traslaciones y distorsiones en la imagen, lo que ayuda a la red a reconocer objetos en diferentes posiciones (Theobald, 2017).

Al igual que en las redes neuronales clásicas, también en las redes neuronales convolucionales se pueden encontrar capas completamente conectadas. Al final de la red, las capas completamente conectadas toman las características extraídas y reducidas por las capas anteriores y las utilizan para realizar la clasificación final. En estas capas, cada neurona está conectada a todas las neuronas de la capa anterior, lo que permite que la red combine todas las características extraídas para decidir la etiqueta o clase de la imagen. Este proceso es similar al de las redes neuronales tradicionales, donde las conexiones densas permiten aprender combinaciones complejas de características (Theobald, 2017).

La función de activación usada en estas redes es normalmente ReLU, que introduce no linealidad en la red. Esto es crucial porque permite que la red neuronal modele relaciones más complejas y aprenda patrones más sofisticados en los datos. Sin la no linealidad, la red sería esencialmente una combinación lineal de las capas anteriores y no podría capturar la complejidad de los datos visuales (Theobald, 2017).

Cuando se ingresa una imagen en una red neuronal convolucional, la imagen pasa a través de varias capas convolucionales donde los filtros, previamente aprendidos, detectan características importantes como bordes, esquinas, y texturas. Las capas de agrupación luego reducen la dimensionalidad de estas representaciones, manteniendo solo la información más relevante. Finalmente, las capas completamente conectadas toman estas características procesadas y las utilizan para determinar la clase de la imagen, como identificar si en la imagen hay un perro, un gato, o cualquier otro objeto que la red haya sido entrenada para reconocer (Theobald, 2017).

Las CNNs ofrecen varias ventajas importantes en comparación con otros tipos de redes neuronales. Por ejemplo, la extracción automática de características, que a diferencia de los métodos tradicionales de visión por computadora, que requieren que las características sean definidas manualmente por los ingenieros, las redes neuronales convolucionales aprenden automáticamente las características más relevantes para la tarea en cuestión, como la clasificación de imágenes. Otro ejemplo es la invariancia a traslaciones y distorsiones, ya que gracias a las capas de pooling y la estructura de la red, las redes

neuronales convolucionales pueden reconocer objetos en una imagen incluso si estos están en diferentes posiciones, o si la imagen ha sido ligeramente alterada, como rotada o escalada (Theobald, 2017).

2.4.2 Autoencoders

Los autoencoders son un tipo de red neuronal que se utiliza para aprender representaciones compactas y eficientes de los datos. Estos modelos son particularmente valiosos en tareas como la reducción de dimensionalidad, la eliminación de ruido y la generación de nuevos datos que sean similares a los originales. Lo que hace a los autoencoders especialmente útiles es su capacidad para captar la estructura interna de los datos sin la necesidad de etiquetarlos o clasificarlos, lo cual los distingue de otras redes neuronales que suelen estar más orientadas hacia tareas supervisadas (Theobald, 2017).

En términos de estructura, un autoencoder se compone de dos partes principales: el codificador y el decodificador. El codificador es responsable de comprimir los datos de entrada en una representación más pequeña y de menor dimensión, conocida como el espacio latente o código latente. Este proceso se logra a través de una serie de capas, que generalmente son densas, y que reducen progresivamente la dimensionalidad de los datos de entrada. La finalidad del codificador es aprender a capturar las características más relevantes de los datos mientras descarta la información redundante o menos importante (Theobald, 2017).

Por otro lado, el decodificador toma esta representación comprimida y trata de reconstruir los datos originales a partir del código latente. Es decir, el decodificador tiene la tarea de transformar el código latente de vuelta a su forma original o a una forma lo más cercana posible a los datos originales. Si el autoencoder ha sido entrenado correctamente, la salida del decodificador debería ser muy similar a los datos de entrada, aunque estos hayan sido comprimidos en el espacio latente (Theobald, 2017).

El proceso de entrenamiento de un autoencoder es peculiar porque, a diferencia de otros tipos de redes, utiliza los propios datos de entrada como la "verdad" que debe predecir. Es decir, la red se entrena para que la salida, tras pasar por el codificador y el decodificador, se parezca lo más posible a la entrada original. Para medir el rendimiento de la red en esta tarea, se emplea una función de pérdida, como el error cuadrático medio, que calcula la diferencia entre la entrada original y la salida reconstruida. El objetivo durante el

entrenamiento es minimizar esta pérdida, lo que se logra ajustando los pesos de la red de manera que las reconstrucciones sean cada vez más precisas (Theobald, 2017).

Existen varios tipos de autoencoders, cada uno con características y aplicaciones específicas. Los autoencoders clásicos son los modelos estándar que utilizan una red completamente conectada tanto en el codificador como en el decodificador. Los variational autoencoders son una variante que, además de codificar los datos de entrada en un espacio latente, también modelan la distribución probabilística de estos datos. Esto les permite generar nuevas muestras de datos al muestrear del espacio latente, lo que es útil en tareas de generación de imágenes y otros tipos de datos. Los denoising autoencoders están diseñados específicamente para eliminar el ruido de los datos. Durante el entrenamiento, se les proporciona una versión ruidosa de los datos de entrada y se les pide que reconstruyan la versión original limpia, lo que los convierte en herramientas eficaces para la limpieza de datos. Por último, los sparse autoencoders introducen una restricción de dispersión en el espacio latente, lo que significa que solo unas pocas unidades activas representan la entrada. Esta característica es útil para capturar características importantes de manera más eficiente y con menos redundancia (Theobald, 2017).

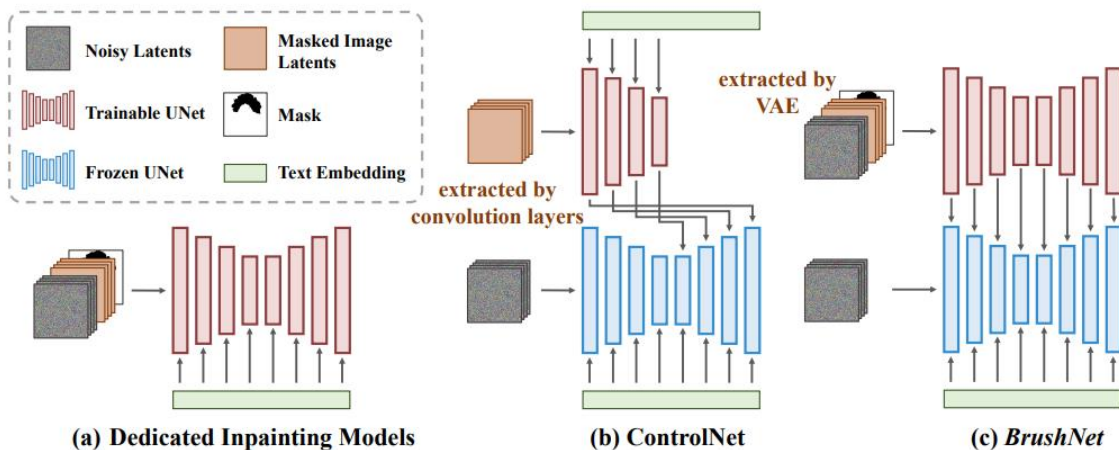
Los autoencoders tienen una amplia gama de aplicaciones en diversos campos, además de ser fundamentales en una de las técnicas que se consideran en este proyecto. En la reducción de dimensionalidad, pueden usarse de manera similar al análisis de componentes principales para reducir la dimensionalidad de los datos manteniendo las características más importantes en un espacio de menor dimensión. También son efectivos en la eliminación de ruido, donde los denoising autoencoders limpian los datos ruidosos, como imágenes o señales contaminadas. En cuanto a la compresión de datos, los autoencoders pueden aprender a comprimir datos de alta dimensión en un formato más compacto, lo cual es útil en aplicaciones de almacenamiento y transmisión de información. Además, los variational autoencoders pueden generar nuevas muestras de datos similares a los originales, lo que resulta valioso en la creación de imágenes, música, texto, y más. Otro uso importante de los autoencoders es el análisis de anomalías; al entrenar un autoencoder con datos normales, las diferencias significativas en la reconstrucción de nuevos datos pueden indicar anomalías o irregularidades, lo cual es útil en la detección de fraudes, defectos en productos, o patrones inusuales en datos médicos (Theobald, 2017).

2.4.3 *BrushNet (Ju et al., 2024)*

Por hacer mención del contexto histórico, recientemente, los modelos de difusión han mostrado un rendimiento impresionante en la generación de imágenes, lo que ha llevado a investigadores a explorar pipelines basados en difusión para la restauración de imágenes de alta calidad alineadas con indicaciones de texto. Los métodos actuales de restauración guiados por texto y basados en difusión se dividen en dos categorías. La primera es la modificación de la estrategia de muestreo, que modifica el proceso de eliminación de ruido al muestrear las regiones enmascaradas de un modelo de difusión preentrenado, mientras que las áreas no enmascaradas se copian y pegan directamente de la imagen original. Aunque este método puede aplicarse a diversos modelos de difusión, a menudo resulta en restauraciones incoherente debido al conocimiento perceptivo limitado de los bordes de las máscaras y el contexto de las regiones circundantes. La segunda categoría comprende los modelos de restauración dedicados, que ajustan un modelo de restauración de imágenes especialmente diseñado al expandir la dimensión del canal de entrada de los modelos de difusión base para incorporar la imagen dañada y la máscara. Aunque estos modelos pueden generar resultados más satisfactorios, su arquitectura puede no ser la más adecuada para la restauración basada en difusión debido a la influencia de la incrustación de texto y la carga adicional sobre el marco UNet (Ju et al., 2024).

Figura 3

Comparación de arquitecturas inpainting anteriores y BrushNet.



Nota: Tomado de Ju et al., 2024.

Para abordar estos problemas, el reciente trabajo de (Ju et al., 2024) propone BrushNet, que introduce una rama adicional en el marco de difusión para una arquitectura de restauración más adecuada. Las características clave de BrushNet incluyen el uso de un codificador VAE para el procesamiento de imágenes enmascaradas, la adopción de un

enfoque jerárquico para incorporar gradualmente las características de UNet preentrenadas y la eliminación de la atención cruzada del texto en UNet para asegurar información de imagen pura en la rama adicional. Además, BrushNet propone una estrategia de mezcla difuminada para lograr una mejor consistencia y control sobre las regiones no enmascaradas. La efectividad del sistema se evalúa en dos tipos de tareas de restauración: máscaras de pincel aleatorio y máscaras basadas en segmentación, utilizando benchmarks como EditBench y los recién introducidos BrushBench y BrushData. Los resultados indican que BrushNet logra un rendimiento de vanguardia en múltiples métricas, incluyendo la calidad de la imagen, la preservación de la región enmascarada y la alineación del texto. Estos resultados la hacen el uso de la red adecuada para los propósitos del proyecto debido a la gran precisión alcanzada. Sin embargo, al limitarse al uso de los parámetros obtenidos durante el preentrenamiento, la operación de feed-forward aplicada a la red no puede modificarse de manera flexible para reducir el tiempo de ejecución. Además, las operaciones del codificador VAE y las demás etapas previas que se utilizan como entrada a BrushNet son estrictamente necesarias como parte integral del modelo, y no pueden ser alteradas sin comprometer su funcionalidad (Ju et al., 2024).

2.4.4 MAE: Aprendizaje de características previas y mejora de la atención para la restauración de imágenes (Cao Chenjieand Dong, 2022)

Este trabajo (Cao Chenjieand Dong, 2022) reconoce que, en imágenes de alta resolución, reparar imágenes dañadas por máscaras arbitrarias con resultados razonables es un desafío. Esto se debe a que el modelo de restauración necesita comprender la información semántica de las imágenes enmascaradas, lo que exige prioridades impulsadas por datos y capacidades de modelo suficientes. Además, se observan los siguientes problemas al momento de realizar esta tarea

(i) Capacidades limitadas para modelar buenas prioridades. Muchos trabajos pioneros han intentado introducir información previa a los modelos de restauración. Algunos trabajos proponen modelos de múltiples etapas, que reparan varias informaciones auxiliares e imágenes dañadas secuencialmente para mejorar la restauración de imágenes. Estos métodos aprenden prioridades en campos específicos, como estructuras o semántica, con buena interpretabilidad visual en lugar de características con prioridades más informativas. Otros métodos utilizan pérdidas auxiliares para introducir información previa adicional sin extender capacidades de modelo suficientes. Las funciones de pérdida complejas causan una sintonización de hiperparámetros sofisticada y un entrenamiento de

restauración más difícil. Además, algunos métodos de restauración basados en transformadores dependen en gran medida de imágenes de baja resolución generadas por el muestreo iterativo que consume mucho tiempo, y luego las aumentan con redes neuronales convolucionales. Algunos otros modelos aprenden prioridades latentes globales con Generative Adversarial Networks, que solo pueden resolver escenas simples con un solo objeto. El modelo propuesto por este artículo incorpora características previas efectivas del aprendizaje de representaciones basado en transformadores para mejorar el restauración, lo que hace que logre resultados superiores sin sobreajustar los resultados del transformador (Cao Chenjieand Dong, 2022).

(ii) Prioridades informativas para casos de alta resolución. La restauración de imágenes de alta resolución disfruta de más implicaciones prácticas con productos electrónicos avanzados e imágenes de alta calidad en el mundo real. Algunas investigaciones se dedican a facilitar la restauración de imágenes de alta resolución con campos receptivos más grandes, transferencia de atención para el residuo de alta frecuencia y upsampling en dos etapas. Desafortunadamente, estos métodos aún tienden a copiar texturas existentes sin sentido en lugar de comprender realmente la semántica de las imágenes enmascaradas de alta resolución sin entrenar directamente con datos de alta resolución costosos. Nuestro método aprovecha la codificación posicional continua para aumentar las características previas para un rendimiento de restauración superior en imágenes de alta resolución (Cao Chenjieand Dong, 2022).

(iii) Falta de discusiones sobre el aprendizaje de representaciones para restauración. Recientemente, los modelos de lenguaje preentrenados auto-supervisados han logrado un gran éxito en los campos de Procesamiento de Lenguaje Natural. La idea de enmascarar y predecir y las arquitecturas basadas en transformadores también han sido bien exploradas en tareas de visión. Pero estos transformadores de visión solo consideran el aprendizaje de representaciones para tareas de clasificación. Los autores mencionan que, hasta donde alcanza su conocimiento, nadie ha explorado la aplicación de modelos de visión preentrenados auto-supervisados para tareas generativas, y mucho menos para la restauración de imágenes. En este trabajo se presentan discusiones sobre el aprendizaje de representaciones basado en el preentrenamiento para la restauración de imágenes de imágenes en este documento (Cao Chenjieand Dong, 2022).

Para abordar estos problemas, esta técnica propone guiar la restauración de imágenes con un modelo de preentrenamiento eficiente de AutoCodificador Enmascarado (MAE), que se llama Restauración Mejorada con Características y Atención Previas. Específicamente, un MAE se preentrena primero con la tarea de predicción visual enmascarada. Los autores reemplazan algunas máscaras aleatorias con máscaras grandes y contiguas para hacer que el MAE sea más adecuado para la tarea posterior. Luego, las características del decodificador MAE se agregan a la CNN de restauración para la restauración guiado por prioridades de imágenes. Además, se destaca que las relaciones de atención de MAE entre regiones enmascaradas y no enmascaradas son compatibles con el aprendizaje de restauración de CNN. Por lo tanto, se utilizan convoluciones grupales para agregar características de CNN con puntuaciones de atención de MAE, lo que puede mejorar mucho el rendimiento de restauración (Cao Chenjieand Dong, 2022).

2.4.5 LaMa: Restauración de Máscaras Grandes Resistente a la Resolución con Convoluciones de Fourier (Suvorov et al., 2021)

El rellenado realista de partes faltantes requiere tanto entender la estructura a gran escala de las imágenes naturales como realizar síntesis de imágenes. Este tema ha sido estudiado en la era previa al aprendizaje profundo (Bertalmio et al., 2003; Criminisi et al., 2003; Hays & Efros, 2007) . La práctica habitual es entrenar sistemas de restauración en un gran conjunto de datos generado automáticamente, creado mediante el enmascaramiento aleatorio de imágenes reales. Es común usar modelos complicados de dos etapas con predicciones intermedias, como imágenes suavizadas, bordes y mapas de segmentación (H. Liu et al., 2020; Yi et al., 2020; Zeng et al., 2020) . En el artículo de (Suvorov et al., 2021) , se logra resultados del estado del arte con una red de una sola etapa. Un campo receptivo efectivo grande es esencial para entender la estructura global de una imagen y, por lo tanto, resolver el problema de la restauración. Además, en el caso de una máscara grande, un campo receptivo aún mayor pero limitado puede no ser suficiente para acceder a la información necesaria para generar una restauración de calidad. Se observa que las arquitecturas convolucionales populares podrían carecer de un campo receptivo efectivo lo suficientemente grande.

También se interviene cuidadosamente en cada componente del sistema para aliviar el problema y desbloquear el potencial de la solución de una sola etapa. Específicamente:

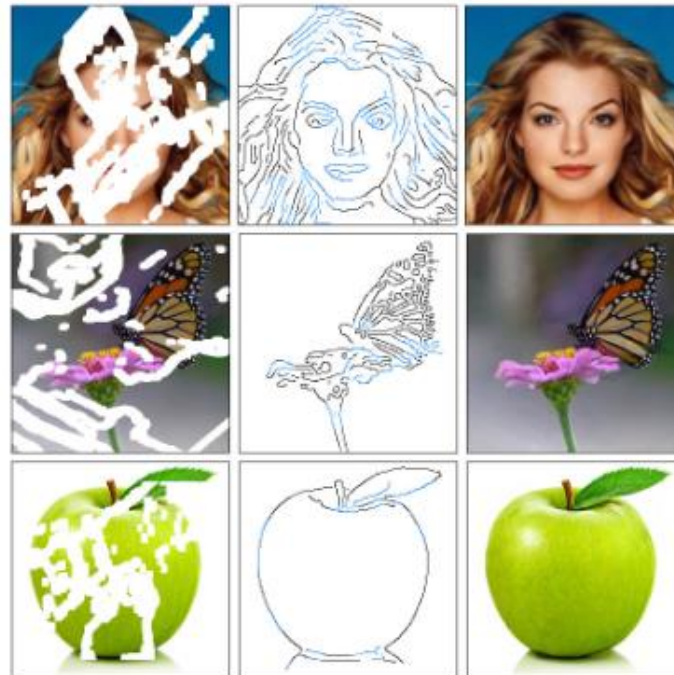
- i) Se propone una red de restauración basada en convoluciones rápidas de Fourier (FFCs) desarrolladas recientemente. Las FFCs permiten un campo receptivo que cubre toda la imagen incluso en las capas tempranas de la red. El trabajo muestra que esta propiedad de las FFCs mejora tanto la calidad perceptual como la eficiencia de los parámetros de la red. Un efecto interesante es que el sesgo inductivo de las FFC permite que la red se generalice a resoluciones altas que nunca se ven durante el entrenamiento. Este hallazgo aporta beneficios prácticos significativos, ya que se necesitan menos datos de entrenamiento y cálculos.
- ii) Se propone el uso de la pérdida perceptual basada en una red de segmentación semántica con un campo receptivo alto. Esto se basa en la observación de que el campo receptivo insuficiente afecta no solo a la red de restauración, sino también a la pérdida perceptual. La pérdida promueve la consistencia de las estructuras y formas globales.
- iii) Se introduce una estrategia agresiva para el entrenamiento de generación de máscaras, para desbloquear el potencial de un alto campo receptivo de los dos primeros componentes (Suvorov et al., 2021).

2.4.6 EdgeConnect: Relleno Generativo de Imágenes con Aprendizaje Adversarial de Bordes (Nazeri et al., 2019)

El trabajo presentado en este artículo (Nazeri et al., 2019) está motivado por la observación de que muchas técnicas existentes de restauración de imágenes generan regiones excesivamente suavizadas o borrosas, fallando en reproducir detalles finos. Por este motivo, los autores dividen la restauración de imágenes en un proceso de dos etapas: generación de bordes y completación de imagen, como se muestra en la figura 3.

Figura 4

Proceso de la restauración hecha por EdgeConnect.



Nota: Tomado de Nazeri et al., 2019.

La generación de bordes se enfoca únicamente en recrear bordes en las regiones faltantes. La red de completación de imagen utiliza los bordes recreados y estima las intensidades de los píxeles RGB de las regiones faltantes. Ambas etapas siguen un marco adversarial para asegurar que los bordes recreados y las intensidades de los píxeles RGB sean visualmente consistentes. Ambas redes incorporan pérdidas basadas en características profundas para garantizar resultados perceptualmente realistas. Al igual que la mayoría de los problemas de visión computacional, la restauración de imágenes precede el uso generalizado de técnicas de aprendizaje profundo. Sin embargo, estos métodos, según los autores, hacen un mal trabajo al reconstruir detalles complejos que pueden ser locales a la región faltante (Nazeri et al., 2019).

Más recientemente, los enfoques de aprendizaje profundo han encontrado un éxito notable en la tarea de la restauración de imágenes. Estos esquemas rellenan los píxeles faltantes utilizando distribuciones de datos aprendidas. Son capaces de generar estructuras coherentes en las regiones faltantes, algo que era casi imposible para las técnicas tradicionales antes mencionadas. Si bien estos enfoques pueden generar regiones faltantes con estructuras significativas, las regiones generadas a menudo son borrosas o carecen de

ciertos artefactos, lo que sugiere que estos métodos tienen dificultades para reconstruir información de alta frecuencia con precisión (Nazeri et al., 2019).

Dado que la estructura de la imagen está bien representada en su máscara de bordes, los autores muestran que es posible generar resultados superiores condicionando una red de restauración de imágenes en los bordes de las regiones faltantes. Como no se tiene acceso a los bordes en las regiones faltantes, en su lugar, se entrena un generador de bordes que reconstruye bordes en estas áreas. El enfoque de "primero líneas, luego color" está parcialmente inspirado en una comparación directa de cómo trabajan los artistas. Como dice Betty Edwards, "En el dibujo lineal, las líneas no solo delimitan y definen espacios y formas; también juegan un papel vital en la composición", destacando la importancia de los bocetos desde un punto de vista artístico (Edwards, 1997).

Desarrollo del Proyecto

3.1. Análisis para la selección

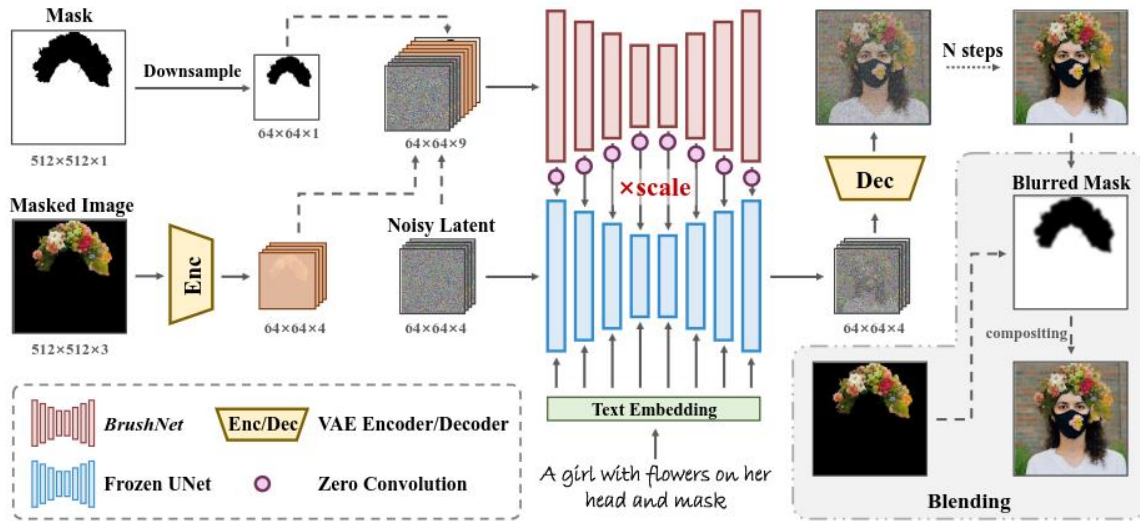
En esta sección se exponen algunas de las técnicas que se evaluaron como posibles candidatas para la adaptación propuesta en este trabajo. En términos generales, los enfoques tradicionales para la restauración de imágenes pueden dividirse en dos grupos: los basados en difusión y los basados en parches. Los métodos basados en difusión propagan datos del fondo en la región faltante siguiendo un proceso difusivo típicamente modelado utilizando operadores diferenciales. Ejemplos de este tipo pueden ser encontrados en (Ballester et al., 2001; Bertalmio et al., 2000; Esedoglu & Shen, 2002; D. Liu et al., 2007). Por otro lado, los métodos basados en parches rellenan las regiones faltantes con parches de una colección de imágenes fuente que maximizan la similitud de los parches. Algunos ejemplos son expuestos en (Darabi et al., 2012; Huang et al., 2014). Una de las técnicas consideradas para este proyecto, EdgeConnect, no pertenece a ninguno de estos tipos. Los detalles de esta técnica se dan más adelante.

Es importante destacar que también se consideraron las técnicas mencionadas en la sección de Antecedentes, dado que inicialmente fueron propuestas en el plan de tesis. Sin embargo, lamentablemente, ninguna de ellas demostró ser adecuada para una posterior adaptación, debido a una variedad de razones. Las siguientes técnicas, por otro lado, mostraron potencial para su adaptación posterior. Cada una de ellas logró ser ejecutada siguiendo las instrucciones disponibles en el repositorio del código fuente, o con ayuda externa. Las descripciones proporcionadas de cada técnica fueron obtenidas de los artículos científicos publicados al presentarlas.

3.1.1. *BrushNet (Ju et al., 2024)*

Para recapitular BrushNet, brevemente explicado en el marco teórico. A continuación, se presenta un esquema visual de la arquitectura del modelo.

Figura 5
Representación visual de la arquitectura de BrushNet.



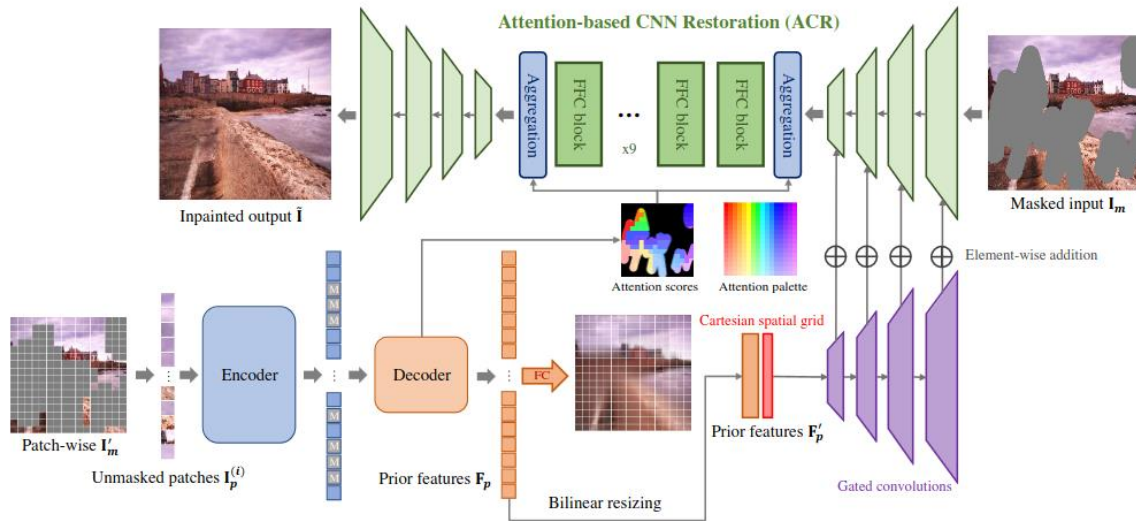
Nota: Tomado de (Ju et al., 2024)

El modelo genera una imagen restaurada basada en la máscara de entrada y la imagen enmascarada. Inicialmente, la máscara se reduce de tamaño para que coincida con el espacio latente, y la imagen enmascarada se introduce en el codificador VAE para armonizar la distribución del espacio latente. Posteriormente, el espacio latente ruidoso, el espacio latente de la imagen enmascarada y la máscara reducida se concatenan y se utilizan como entrada para BrushNet. Las características extraídas de BrushNet se incorporan gradualmente a las capas preentrenadas de UNet, siguiendo un bloque de convolución cero. Tras el proceso de eliminación de ruido, la imagen generada y la imagen enmascarada original se mezclan utilizando una máscara difuminada. La única instancia donde la modificación es viable es en el procesamiento feed-forward de BrushNet, alterando las dimensiones de la capa o el número de neuronas. No obstante, esta flexibilidad no está respaldada por Torch, la plataforma en la que se encuentra implementado el modelo. Cualquier ajuste en las dimensiones resulta en un error de tiempo de ejecución.

3.1.2. MAE: Aprendizaje de características previas y mejora de la atención para la restauración de imágenes (Cao Chenjie and Dong, 2022)

El modelo puede extenderse efectivamente a la restauración de alta resolución con una pequeña puesta a punto de características MAE redimensionadas bilinealmente y la cuadrícula espacial cartesiana. Una visión general del modelo se aprecia en la figura 5.

Figura 6
Diagrama de la arquitectura de MAE.



Nota: Tomado de Cao Chenjie and Dong, 2022

Los resultados obtenidos mediante esta técnica son sumamente prometedores, ya que logran restaurar con una notable precisión la máscara aplicada en un tiempo considerablemente reducido, tal como se evidencia a continuación.

Figura 7
Ejemplo de los resultados de MAE.



Nota: Tomado de Cao Chenjie and Dong, 2022

Una de las limitaciones de este modelo, reconocida por los autores, es su incapacidad para competir con modelos del estado de arte. El punto más desfavorable, por otro lado, es que al igual que con la técnica anterior, al limitarse al uso de los parámetros obtenidos durante el preentrenamiento, la operación de feed-forward que se realiza al suministrar de datos a la red no puede modificarse de manera flexible para reducir el tiempo de ejecución,

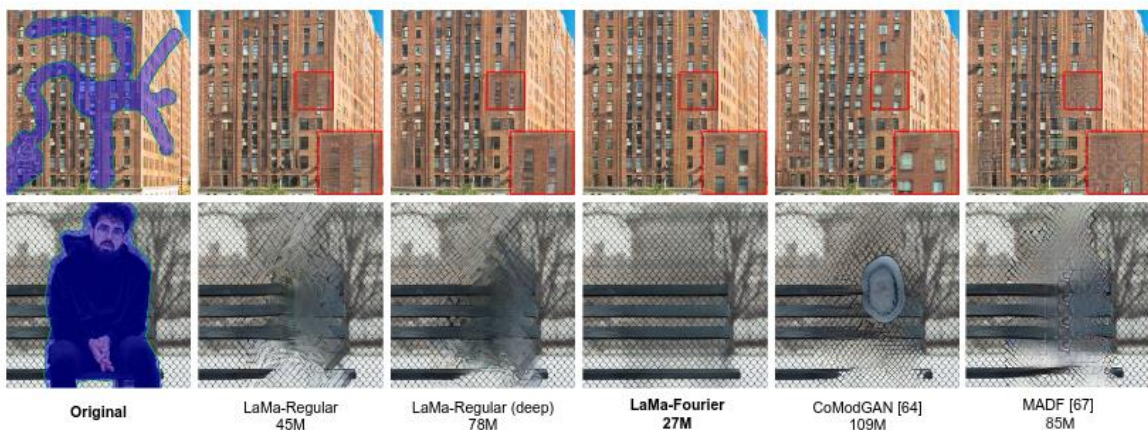
por lo que no pueden realizarse adaptaciones que resulten en una ejecución sin errores del programa. Otro punto desfavorable es que hacer una medición de tiempo directa del programa resulta difícil debido a que no queda claro el punto exacto del programa en el que la predicción comienza a ser llevada a cabo por las numerosas funcionalidades que usa (como checkpoints y procesamiento en paralelo).

3.1.3. LaMa: Restauración de Máscaras Grandes Resistente a la Resolución con Convoluciones de Fourier (Suvorov et al., 2021)

El procedimiento descrito para esta técnica produce máscaras anchas y grandes, que obligan a la red a aprovechar al máximo el alto campo receptivo del modelo y la función de pérdida. Esto produce LaMa (large mask inpainting), un nuevo sistema de restauración de imágenes de una sola etapa. Los principales componentes de LaMa son la arquitectura de alto campo receptivo, la función de pérdida de alto campo receptivo y el algoritmo agresivo de generación de máscaras. El artículo compara meticulosamente LaMa con otros trabajos del estado del arte y analiza la influencia de cada componente propuesto. A continuación, se muestran algunas imágenes restauradas proporcionadas en el artículo científico.

Figura 8

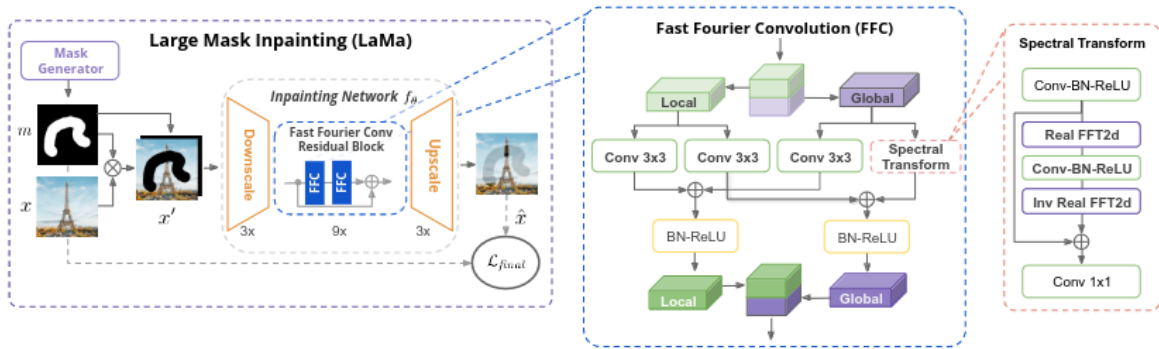
Resultados de LaMa Fourier en comparación con otras arquitecturas.



Nota: Tomado de Suvorov et al., 2021

A través de la evaluación, se descubre que LaMa puede generalizarse a imágenes de alta resolución después de entrenar solo con datos de baja resolución. LaMa puede capturar y generar estructuras periódicas complejas, y es resistente a máscaras grandes. Además, esto se logra con una cantidad significativamente menor de parámetros entrenables y costos de tiempo de inferencia en comparación con trabajos competitivos. El esquema completo de la arquitectura es el siguiente.

Figura 9
Representación de la arquitectura de LaMa Fourier.



Nota: Tomado de Suvorov et al., 2021

Un aspecto positivo de esta técnica de restauración es que supera a muchas otras técnicas consideradas hasta antes como el estado del arte, siendo únicamente comparable (superiormente) por CoModGAN y MADF. Esto puede verse en la siguiente figura 9, proporcionada por los autores de la investigación.

Figura 10
Resultados de la comparación de LaMa Fourier con otras técnicas de restauración digital de imágenes.

Method	# Params $\times 10^6$	Places (512 \times 512)						CelebA-HQ (256 \times 256)			
		Narrow masks		Wide masks		Segm. masks		Narrow masks		Wide masks	
		FID \downarrow	LPIPS \downarrow	FID \downarrow	LPIPS \downarrow	FID \downarrow	LPIPS \downarrow	FID \downarrow	LPIPS \downarrow	FID \downarrow	LPIPS \downarrow
LaMa-Fourier (ours)	27	0.63	0.090	2.21	0.135	5.35	0.058	7.26	0.085	6.96	0.098
CoModGAN [64]	109 \blacktriangle	0.82 \blacktriangle 30%	0.111 \blacktriangle 23%	1.82 \blacktriangledown 18%	0.147 \blacktriangle 9%	6.40 \blacktriangle 20%	0.066 \blacktriangle 14%	16.8 \blacktriangle 131%	0.079 \blacktriangledown 7%	24.4 \blacktriangle 250%	0.102 \blacktriangle 4%
MADF [67]	85	0.57 \blacktriangledown 10%	0.085 \blacktriangledown 5%	3.76 \blacktriangle 70%	0.139 \blacktriangle 3%	6.51 \blacktriangle 22%	0.061 \blacktriangle 5%	—	—	—	—
AOT GAN [60]	15 \blacktriangledown	0.79 \blacktriangle 25%	0.091 \blacktriangle 1%	5.94 \blacktriangle 169%	0.149 \blacktriangle 11%	7.34 \blacktriangle 37%	0.063 \blacktriangle 10%	6.67 \blacktriangledown 8%	0.081 \blacktriangledown 4%	10.3 \blacktriangle 48%	0.118 \blacktriangle 20%
GCPR [17]	30 \blacktriangle	2.93 \blacktriangle 363%	0.143 \blacktriangle 59%	6.54 \blacktriangle 196%	0.161 \blacktriangle 19%	9.20 \blacktriangle 72%	0.073 \blacktriangle 27%	—	—	—	—
HiFill [54]	3 \blacktriangledown	9.24 \blacktriangle 1361%	0.218 \blacktriangle 142%	12.8 \blacktriangle 479%	0.180 \blacktriangle 34%	12.7 \blacktriangle 137%	0.085 \blacktriangle 49%	—	—	—	—
RegionWise [30]	47 \blacktriangle	0.90 \blacktriangle 42%	0.102 \blacktriangle 14%	4.75 \blacktriangle 115%	0.149 \blacktriangle 11%	7.58 \blacktriangle 42%	0.066 \blacktriangle 14%	11.1 \blacktriangle 53%	0.124 \blacktriangle 46%	8.54 \blacktriangle 23%	0.121 \blacktriangle 23%
DeepFill v2 [57]	4 \blacktriangledown	1.06 \blacktriangle 68%	0.104 \blacktriangle 16%	5.20 \blacktriangle 135%	0.155 \blacktriangle 15%	9.17 \blacktriangle 71%	0.068 \blacktriangle 18%	12.5 \blacktriangle 73%	0.130 \blacktriangle 53%	11.2 \blacktriangle 61%	0.126 \blacktriangle 28%
EdgeConnect [32]	22 \blacktriangledown	1.33 \blacktriangle 110%	0.111 \blacktriangle 23%	8.37 \blacktriangle 279%	0.160 \blacktriangle 19%	9.44 \blacktriangle 76%	0.073 \blacktriangle 27%	9.61 \blacktriangle 32%	0.099 \blacktriangle 17%	9.02 \blacktriangle 30%	0.120 \blacktriangle 22%
RegionNorm [58]	12 \blacktriangledown	2.13 \blacktriangle 236%	0.120 \blacktriangle 33%	15.7 \blacktriangle 613%	0.176 \blacktriangle 31%	13.7 \blacktriangle 156%	0.082 \blacktriangle 42%	—	—	—	—

Nota: Tomado de Suvorov et al., 2021

Esto posibilita descartar cualquiera de las otras técnicas con las que LaMa fue comparada y usarla en su lugar. Una excepción a esto es la técnica de EdgeConnect.

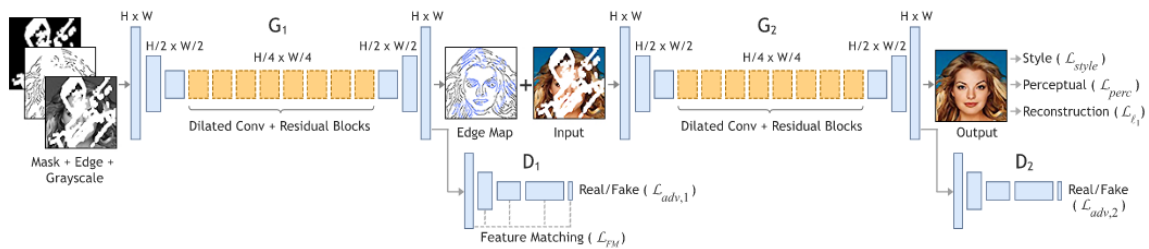
El segundo aspecto positivo de esta técnica es que es posible realizar adaptaciones menores durante la etapa de Fast Fourier Convolution, reduciendo parámetros que hacen posible la ganancia significativa de tiempo que se busca. Finalmente, el tercer aspecto positivo es que las máscaras pueden ser dadas como entrada al programa junto con las

fotografías dañadas, lo que hace posible su comparación con otras técnicas bajo la misma métrica.

3.1.4. EdgeConnect: Relleno Generativo de Imágenes con Aprendizaje Adversarial de Bordes (Nazeri et al., 2019)

Con esta técnica, los autores presumen que la recuperación de bordes es una tarea más fácil que la completación de imágenes. El modelo propuesto esencialmente desacopla la recuperación de información de alta y baja frecuencia de la región rellenada. La siguiente figura 10 muestra un esquema de la técnica propuesta.

Figura 11
Diagrama de la arquitectura de EdgeConnect.



Nota: Tomado de Nazeri et al., 2019

Una de las ventajas de utilizar esta técnica es la capacidad de alterar la arquitectura del modelo utilizado para la generación de bordes (Edge Generator) simplemente reduciendo el número de capas, sin causar errores en tiempo de ejecución. Otra ventaja es que las máscaras pueden ser proporcionadas al programa sin necesidad de adaptaciones adicionales, ya que su uso es compatible con los requisitos de nuestro proyecto. Sin embargo, una desventaja es que esta técnica, a pesar de estar a la altura con otras del estado del arte, a diferencia de LaMa, EdgeConnect no las supera en general.

3.2. Selección de técnicas en base a las características de interés

La siguiente es una comparación de las características de interés del proyecto realizadas sobre todas las técnicas consideradas para su adaptación. Se consideraron los siguientes criterios.

- A. **Ejecutabilidad.** La ejecución del código fuente se llevó a cabo con el hardware disponible y siguiendo las instrucciones proporcionadas o con ayuda externa.
- B. **Modificabilidad.** Hubo adaptaciones significativas a algunos de los parámetros del programa que no dieron lugar a errores en tiempo de ejecución.
- C. **Uniformidad.** Es posible uniformizar el uso de la máscara para los propósitos del proyecto (por ejemplo, la máscara puede proporcionarse como entrada al programa sin muchas adaptaciones).
- D. **Comparabilidad.** Es posible hacer comparaciones de tiempo sin muchas adaptaciones al programa.

La elección de estos criterios para seleccionar y evaluar técnicas de restauración digital resultó fundamental para garantizar su aplicación práctica y la rigurosidad de los resultados obtenidos. En primer lugar, la ejecutabilidad asegura que los modelos puedan implementarse en entornos computacionales realistas, permitiendo su reproducibilidad y uso práctico (Schmidhuber, 2015). Este criterio es esencial cuando los recursos de hardware son limitados y se requiere evaluar modelos en situaciones concretas. Por otra parte, la modificabilidad posibilita ajustar y optimizar las redes neuronales en función de las necesidades específicas de un proyecto sin comprometer su funcionamiento (Haykin, 1994). Esta adaptabilidad es crucial en trabajos donde se requiere personalizar los parámetros de los modelos para alcanzar un equilibrio entre rendimiento y eficiencia. Además, la uniformidad asegura la consistencia de los resultados al aplicar las técnicas seleccionadas bajo las mismas condiciones experimentales, lo que facilita la comparación objetiva entre diferentes arquitecturas (Krizhevsky et al., 2012). Finalmente, la comparabilidad permite identificar las fortalezas y debilidades de las técnicas evaluadas bajo métricas estandarizadas, favoreciendo la toma de decisiones informadas al momento de elegir la mejor solución para una tarea específica (Zhao et al., 2019). La combinación de estos cuatro criterios no solo cubre los aspectos esenciales para la evaluación efectiva de modelos de redes neuronales, sino que también optimiza el proceso experimental al evitar sesgos y asegurar la transparencia en los resultados.

Tabla 1*Características de interés de las técnicas consideradas para su modificación.*

Nro	Técnica	Ejecutabilidad	Modificabilidad	Uniformidad	Comparabilidad
1	Context Encoders	NO	-	-	-
2	MSNP	NO	-	-	-
3	GLIC	NO	-	-	-
4	Shift-Net	NO	-	-	-
5	GICA	SÍ	NO	NO	-
6	FFGC	NO	-	-	-
7	BrushNet	SÍ	NO	SÍ	SÍ
8	MAE	SÍ	NO	SÍ	NO
9	LaMa	SÍ	SÍ	SÍ	SÍ
10	EdgeConnect	SÍ	SÍ	SÍ	SÍ

Debido a que los cuatro puntos son necesarios para los propósitos de esta investigación, únicamente se consideraron las técnicas de LaMa y EdgeConnect para la siguiente fase del proyecto.

3.3. Propuesta de adaptación

3.3.1. Adaptaciones propuestas para LaMa

Las siguientes son las tres adaptaciones propuestas para LaMa. Los detalles de la implementación de la técnica original relevante a estas adaptaciones pueden ser obtenido en <https://github.com/advimman/lama/blob/main/saicinpainting/training/modules/ffc.py>.

1. Una de las contribuciones de arquitectura de LaMa es la convolución rápida de Fourier (Chi et al., 2020). Este es un operador propuesto recientemente que permite utilizar el contexto global en las capas iniciales. FFC se basa en una transformada rápida de Fourier (Brigham & Morrow, 1967) por canales y tiene un campo receptivo que cubre toda la imagen. FFC divide los canales en dos ramas paralelas. La rama local usa convoluciones convencionales de Fourier de dimensión (3, 2, 1) (en imágenes 2d). La rama global usa la transformada rápida de Fourier en la misma dimensión para tener en cuenta el contexto global. La primera modificación propuesta es reducir la dimensión de la convolución de Fourier a una dimensión de

(2, 2, 1) sobre imágenes 2d. Claramente la reducción de una dimensión de tres a dos puede únicamente tener un impacto positivo en el tiempo. Tener dos valores unidimensionales es equivalente a un valor bidimensional, por lo que el segundo parámetro no puede ser reducido aún más. Por otro lado, tener una dimensión creciente es equivalente a alguna secuencia de dimensiones no creciente, por lo que el primer valor no puede ser reducido más allá de dos.

2. Como se explicó en el caso anterior, el operador de FFC propuesto por Lama, es un operador de dos fases, en las cuales se usa la transformada de Fourier en dimensión (3, 2, 1). La segunda modificación propuesta es reducir la dimensión de la transformada rápida de Fourier en la segunda fase a una dimensión de (2, 2, 1). El motivo de la reducción únicamente el primer parámetro y la no reducción de los demás parámetros es igual al caso anterior.
3. La última modificación tiene que ver con la dimensión de las unidades de convolución. Se usa unos bloques de 3x3 para estas unidades, por lo que se propone usar bloques de dimensión 2x2. Tener un bloque de 1x1 no tiene efecto alguno en la convolución, por lo que 2x2 es el mínimo valor al que se puede reducir este parámetro.

Debido a que, en los tres casos, las redes ya fueron entrenadas, es necesario hacer una redimensión a los pesos cada vez que son usados por la red. Esto se puede conseguir fácilmente con el comando `resize` de `numpy`, que ignora los valores fuera de la dimensión.

3.3.2. Adaptaciones propuestas para EdgeConnect

Las siguientes son las adaptaciones propuestas para EdgeConnect.

1. Como se ha mencionado previamente, EdgeConnect emplea dos fases para la restauración de imágenes. En la primera fase, reconstruye los contornos de la imagen, y en la segunda fase, estos contornos se utilizan junto con la imagen original para generar la imagen final. Cada fase emplea una red profunda que utiliza ocho bloques residuales. La primera modificación propuesta se centra en la primera fase, denominada EdgeGenerator, la cual utiliza ocho bloques de dimensiones $H/4 \times W/4$, donde H y W representan la altura y el ancho de la imagen, respectivamente. Se sugiere reducir el número de bloques a seis para esta etapa. El algoritmo utilizado genera errores en tiempo de ejecución cuando se intenta reducir el número de bloques a menos de seis, por lo que este será el límite en este caso.

2. Similarmente, para la segunda fase, ImageCompletion Network, se usan ocho bloques residuales de la misma dimensión que antes, por lo que la propuesta para este caso es reducir el número de bloques a 6. Al igual que en el caso anterior, el algoritmo produce errores en tiempo de ejecución con menos de seis bloques.

En ambos casos, los cambios hechos a la red fueron inmediatos y también fue posible la redimensión de los pesos de la red.

Todas las adaptaciones están basadas en el estudio de cada arquitectura y conceptos básicos que nos brinda Sipser, M. (2006) y Arora, S., & Barak, B. (2007)

3.3.3. Proceso de ejecución y experimentación

Es importante destacar que, aunque existen otros factores que pueden ser medidos en las predicciones generadas por la Inteligencia Artificial, en esta experimentación se optó por utilizar únicamente la calidad y el tiempo, dado que estos son los parámetros principales de la investigación. Esta elección se realizó para enfocar el estudio en aspectos clave, permitiendo un análisis más profundo de cómo ambos factores afectan los resultados obtenidos. Aunque otras variables podrían haber sido consideradas, el objetivo de la investigación es explorar la relación directa entre la calidad y el tiempo, debido a su relevancia central en el contexto de los procesos evaluados.

El procedimiento comenzó con la selección y obtención de las redes neuronales previamente entrenadas. Estas redes, que ya habían sido optimizadas para realizar tareas específicas de predicción, sirvieron como base para el experimento, Lama demostró un desempeño superior a las técnicas consideradas como el estado del arte en su momento, tales como U-Net, DeepFill v2 y RegionWise, lo que sugiere que su implementación puede reemplazar cualquier análisis realizado con estas metodologías. De manera similar, EdgeConnect superó a técnicas como CA, GLCIC y Canny, que también representaban el estándar más avanzado al momento de su introducción. Este patrón de superioridad se observa igualmente en las diez técnicas seleccionadas para su posible optimización. Una vez que se seleccionaron Lama y Edge Connect en base al cumplimiento de 4 criterios, se procedió a implementar las modificaciones propuestas.

Después de aplicar las modificaciones tanto en Lama como en Edge Connect, la red modificada quedó alterada, lo que afectó su rendimiento. Es importante señalar que, a pesar de los cambios realizados, los pesos que aún se mantenían en la red modificada no eran necesariamente los más óptimos, lo que ocasionó una degradación en la capacidad de la red

para realizar predicciones con la misma precisión que antes. Sin embargo, aunque su rendimiento se vio comprometido, la red modificada siguió siendo capaz de realizar predicciones, pero con una calidad inferior en comparación con la red original. El objetivo principal de la experimentación fue realizar una comparación entre las predicciones generadas por la red original y las predicciones de la red modificada. Esta comparación tuvo como propósito identificar y cuantificar la pérdida de calidad en las predicciones, lo cual fue crucial para evaluar si las modificaciones valían la pena o no. Además de la calidad de las predicciones, se prestó especial atención al tiempo de ejecución de ambas redes. El tiempo de predicción fue otro factor importante a considerar, ya que la modificación de la red pudo haber llevado a mejoras en la eficiencia, incluso si eso implicaba una leve pérdida en la precisión de las predicciones.

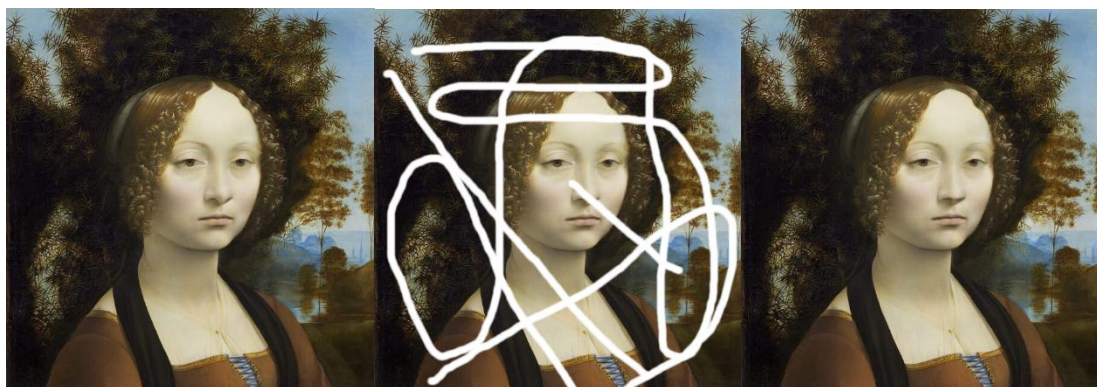
Una vez obtenidos los resultados en cuanto a calidad de predicción y tiempos de ejecución, se realizó un análisis exhaustivo para calcular la ganancia en términos de tiempo al comparar la red modificada con la red original. Este análisis buscó determinar si la mejora en la velocidad de predicción compensaba la pérdida en la calidad de las mismas. Fue posible que en algunos casos una ligera disminución en la precisión fuera aceptable si se traducía en un aumento significativo en la rapidez de las predicciones, lo cual pudo haber sido beneficioso en aplicaciones donde la eficiencia era más crítica que la exactitud absoluta. En resumen, el proceso buscó un equilibrio entre la calidad de las predicciones y el tiempo de ejecución, con el objetivo de optimizar la red para obtener un rendimiento adecuado dentro de los límites establecidos por las condiciones experimentales y los requisitos específicos de la aplicación.

3.4. Resultados experimentales

El dataset utilizado fue “Best Artworks of All Time”, disponible gratuitamente en <https://www.kaggle.com/datasets/ikarus777/best-artworks-of-all-time>. Este dataset es una colección de 8663 imágenes re-escaladas de pinturas de los 50 artistas más influyentes de todos los tiempos. La máscara que se utilizó para este dataset es uniforme en todos los experimentos.

Figura 12

Restauración de ejemplo realizado con la máscara aplicada.



Durante el proceso de experimentación, se realizaron pruebas con las dos técnicas de forma independiente. Para cada técnica, se consideraron todas las posibles combinaciones de adaptaciones, a partir de las cuales se obtuvieron los resultados aplicando la técnica a todas las imágenes del conjunto de datos.

Para llevar a cabo este proceso, se utiliza una técnica en dos ocasiones sobre la imagen de una pintura: una vez sin modificar y otra con las adaptaciones aplicadas. Como resultado, se obtienen dos predicciones de la imagen: una basada en la técnica sin adaptaciones y otra en la técnica modificada. Además, se registran los tiempos de ejecución correspondientes.

3.4.1. Evaluación el error de la restauración - calidad

Para evaluar el error de la restauración, se compara la imagen original de la pintura con la imagen restaurada mediante la técnica. Dado que la técnica solo modifica los píxeles de la máscara, se calcula la distancia Euclidiana entre los valores de color de cada píxel de la máscara de acuerdo a (Alman, 1993). Es decir, si el primer píxel tiene el color (r_1, g_1, b_1) , donde cada valor va desde 0 hasta 255, y el segundo píxel tiene el color (r_2, g_2, b_2) , entonces la distancia entre los dos colores es

$$\text{Distancia} = \sqrt{(r_2 - r_1)^2 + (g_2 - g_1)^2 + (b_2 - b_1)^2}.$$

Se calcula una lista de distancias correspondientes a cada píxel de la máscara entre la imagen restaurada y la imagen original. El error de la restauración se representa mediante el promedio aritmético de estos valores. El tiempo se mide en segundos. De esta manera, se obtienen los errores de la restauración realizada por la técnica, tanto con y sin adaptaciones, para una sola pintura. Dado el error y el tiempo consumido por la técnica, se calcula la

diferencia en porcentaje de tiempo con la técnica sin modificar. Luego, se repite el mismo proceso con todas las pinturas del conjunto de datos, generando una tabla de comparación con 8663 entradas

La primera técnica tiene tres adaptaciones propuestas: 1, 2 y 3. Por lo que todas las 7 posibles adaptaciones son {1}, {2}, {3}, {1, 2}, {1, 3}, {2, 3}, {1, 2, 3}. Esto genera 7 tablas de comparación. La segunda técnica tiene dos adaptaciones propuestas: 1 y 2. Por lo que todas las 4 posibles adaptaciones son {1}, {2}, {1, 2}. Esto genera 3 tablas de comparación. En total, se obtienen 10 tablas de comparación, con 8663 entradas cada una.

Tabla 2

Ejemplo del formato de la información obtenida de los experimentos hechos en LaMa con la modificación 1.

Pintura	Técnica original	Técnica modificada		Diferencia de tiempo (D.T.) (%)
	Tiempo	Error	Tiempo	
Pintura 1	24	230	20	20
Pintura 2	30.69	233	30	2.3
...
Pintura 8663	27	210	26	1.03

Nota: Este es el ejemplo de una de las diez tablas que serían en total con cada modificación, y cada una de ellas de 8663 filas (cantidad de imágenes del dataset)

Es relevante destacar que cada tabla de las 10 puede resumirse en una estructura de dos columnas y 8663 filas, ya que solo los valores de error y las diferencias de tiempo (D.T. en la tabla mostrada) son relevantes para la comparación. A partir de estos 8663 valores, se calculan las siguientes estadísticas resumidas: media, desviación estándar, mediana, mínimo y máximo. Cabe resaltar que los valores se comparan con el valor máximo de la pérdida de calidad, 442. La información resumida para las 10 tablas es la siguiente.

Tabla 3

Comparaciones de error y tiempo junto con estadísticas de promedio, desviación estándar, mediana, mínimo y máximo.

Técnica	Adaptaciones	Param.	Promedio	Desviación	Mediana	Min	Max
LaMa	1	Error	233.046	94.352	232.174	68.342	394.303
		D. T.	5.41	5.689	5.396	-4.344	15.32
	2	Error	180.995	93.962	181.228	19.331	345.282
		D. T.	0.669	4.003	0.676	-6.319	7.649
	3	Error	232.737	84.485	231.686	88.35	380.282
		D. T.	8.001	7.337	8.079	-4.909	20.69
	1, 2	Error	252.587	86.304	253.835	100.5	402.157
		D. T.	14.817	10.299	14.949	-3.186	32.678
	1, 3	Error	303.147	92.523	303.538	142.425	441.0
		D. T.	13.175	8.881	13.102	-2.188	28.686
	2, 3	Error	284.148	93.317	285.445	121.485	441.0
		D. T.	6.628	6.71	6.714	-5.28	18.098
	1, 2, 3	Error	328.801	70.646	330.054	205.298	441.0
		D. T.	20.515	8.76	20.529	5.285	35.797
Edge Connect	1	Error	349.268	57.489	348.898	250.177	441.0
		D. T.	10.644	4.203	10.708	3.287	17.8
	2	Error	352.517	59.377	352.339	248.919	441.0
		D. T.	11.736	4.31	11.67	4.284	19.199
	1, 2	Error	377.929	58.126	387.398	266.235	441.0
		D. T.	25.66	10.008	25.707	8.284	42.808

Nota: D.T. (Diferencia de tiempo)

3.5. Análisis y discusión de resultados

Como se puede observar en la tabla resultante de los experimentos, a medida que se incrementa el número de adaptaciones, la diferencia de tiempo se hace mayor. Para un análisis de los resultados obtenidos, es necesario contrastar las pérdidas de calidad con las diferencias de tiempo. Por este motivo mostramos los valores de la tabla anterior con un parámetro adicional de diferencia de tiempo sobre error, que representa la ganancia de tiempo comparada con la pérdida de calidad: $D.T. / Error * 100 \%$

Cabe resaltar que cuanto mayor este número, mejores resultados se habrán obtenido economizando tiempo sin pérdidas significativas de la calidad.

Tabla 4
Comparación porcentual de la diferencia de tiempo con respecto al error.

Técnica	Adaptaciones	Parámetros	Promedio	D.T. / Error * 100 %
LaMa	1	Error	233.046	2.321 %
		D. T.	5.41	
	2	Error	180.995	0.370 %
		D. T.	0.669	
	3	Error	232.737	3.438 %
		D. T.	8.001	
	1, 2	Error	252.587	5.866 %
		D. T.	14.817	
	1, 3	Error	303.147	4.346 %
		D. T.	13.175	
	2, 3	Error	284.148	2.332 %
		D. T.	6.628	
1, 2, 3	Error	328.801	6.239 %	
	D. T.	20.515		
Edge Connect	1	Error	349.268	3.048 %
		D. T.	10.644	
	2	Error	352.517	3.329 %
		D. T.	11.736	
	1, 2	Error	377.929	6.790 %
		D. T.	25.66	

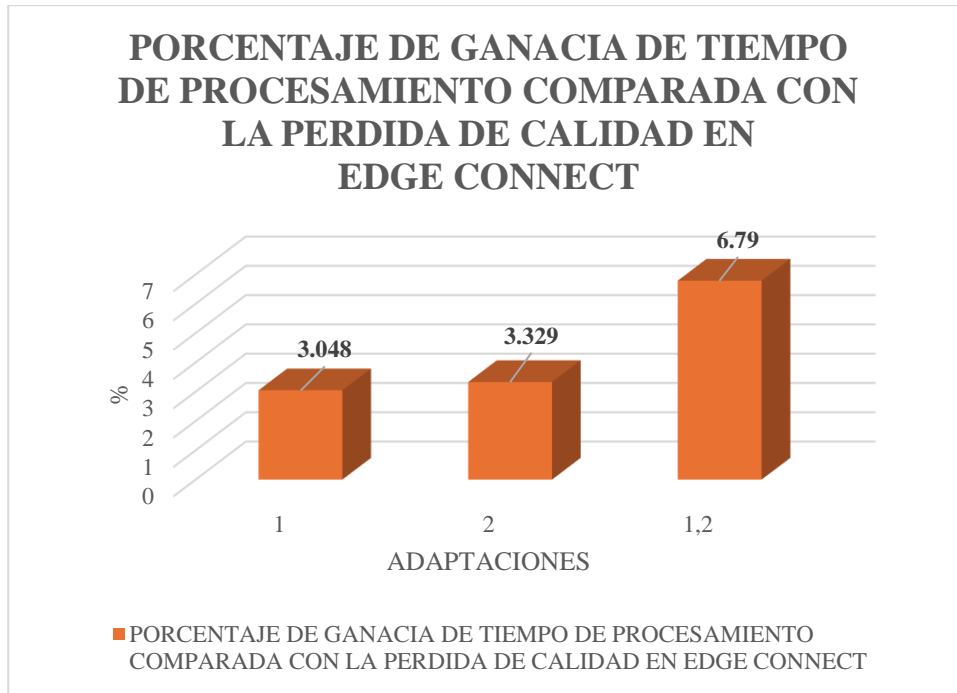
Nota: D.T. (Diferencia de tiempo)

3.5.1. Interpretación de resultados con Edge Connect

Como se aprecia en la tabla, al utilizar EdgeConnect junto con las adaptaciones 1 y 2 simultáneamente, se maximiza el ahorro de tiempo relativo a la pérdida de calidad, lo cual se ve reflejado en la figura 12.

Figura 13

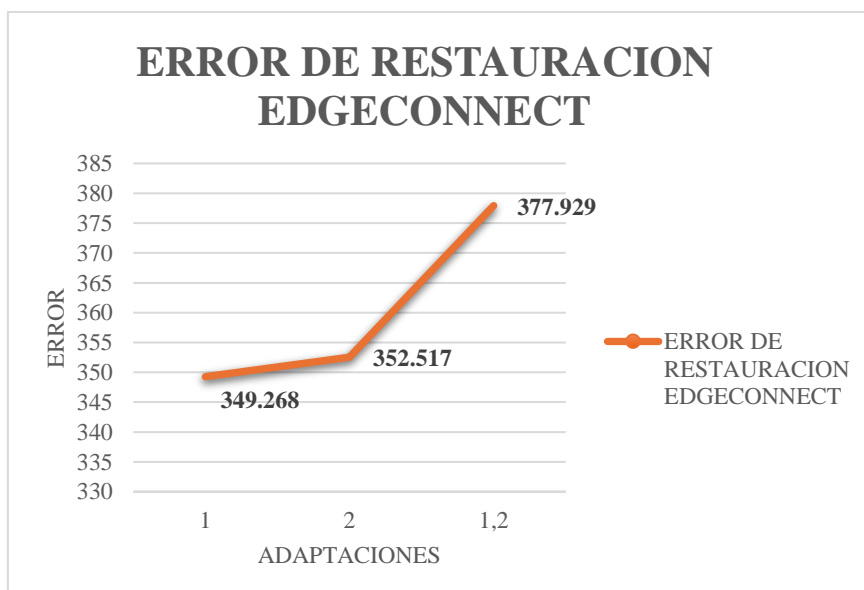
Gráfico de porcentaje de ganancia de tiempo de procesamiento comparado con la pérdida de calidad - Edge Connect



Sin embargo, es importante notar que, con todas las adaptaciones realizadas a Edge Connect, la calidad de la restauración se ve severamente afectada. Es decir, el error promedio en la restauración en todo el conjunto de datos oscila entre 350 y 380, como se muestra en la figura 13. Este fenómeno se controla gracias a la considerable ganancia de tiempo obtenida.

Figura 14

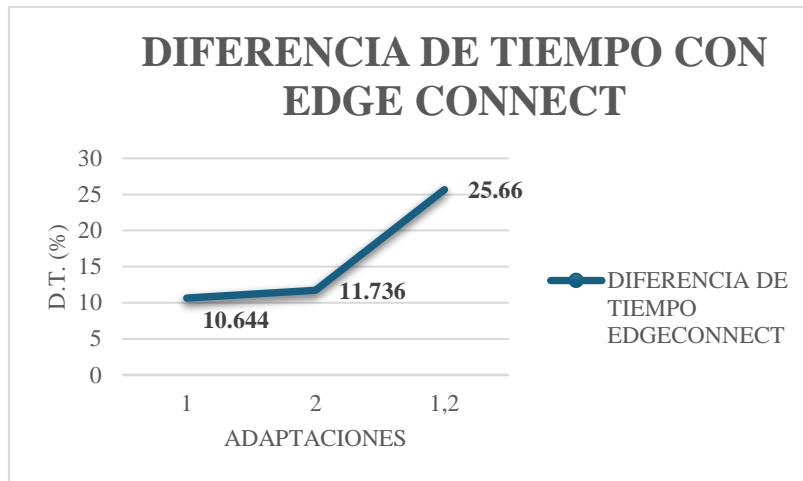
Gráfico de error de restauración con Edge Connect



Por ejemplo, al utilizar la primera modificación, se logra un ahorro promedio del 10% en tiempo. Este resultado es similar al usar la segunda modificación. Al emplear ambas adaptaciones simultáneamente, se obtiene un ahorro promedio del 25% en tiempo como se observa en la figura 14. Estos resultados compensan cualquier pérdida de calidad derivada de las adaptaciones.

Figura 15

Gráfico de diferencia de tiempo (D.T.) con Edge Connect

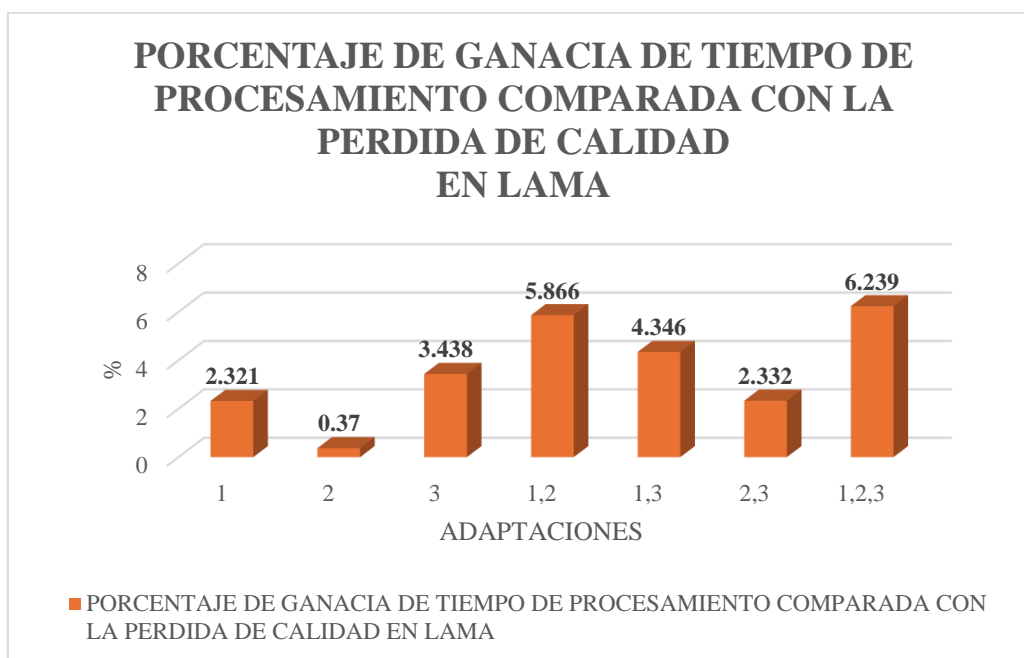


3.5.2. Interpretación de resultados con LaMa

Por otro lado, la calidad de la restauración realizada por LaMa es más consistente.

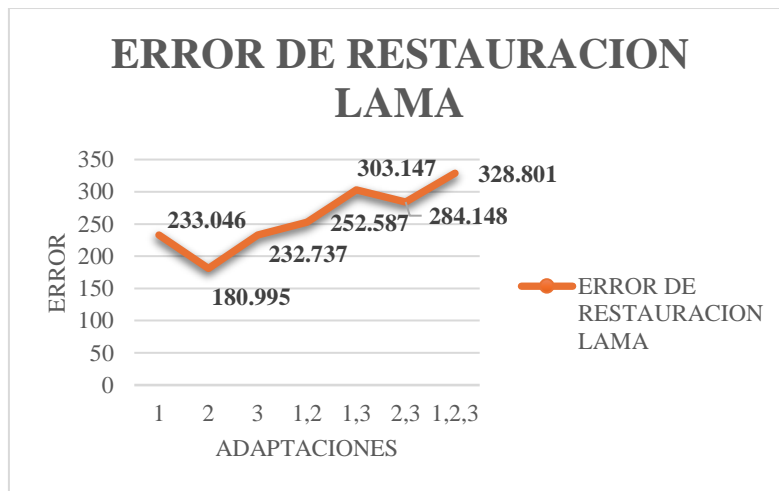
Figura 16

Gráfico de porcentaje de ganancia de tiempo de procesamiento comparado con la pérdida de calidad - LaMa



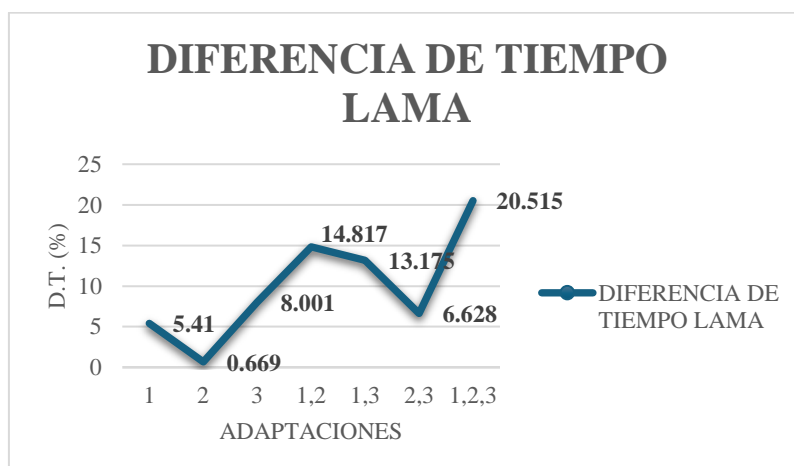
Incluso en el caso más extremo, el error de restauración solo alcanza los 330, en contraste con EdgeConnect, cuyo mejor caso llega a 350. El mejor resultado de LaMa en términos de calidad de restauración se obtiene con la segunda modificación propuesta, que muestra un error de restauración de 180, no obstante, en este caso, la ganancia en tiempo es casi nula, por lo que no hay ganancias ni pérdidas en términos porcentuales. como se muestra en la figura 16.

Figura 17
Gráfico de error de restauración con LaMa



El caso más favorable para LaMa en términos de ahorro de tiempo se da al usar las tres adaptaciones simultáneamente, lo que resulta en una ganancia del 20% en tiempo como se muestra en la figura 17. Aunque la pérdida de calidad en este caso es significativa, la ganancia de tiempo la supera porcentualmente, casi igualando los resultados obtenidos por EdgeConnect y con una calidad de restauración muy superior.

Figura 18
Gráfico de diferencia de tiempo (D.T.) con LaMa



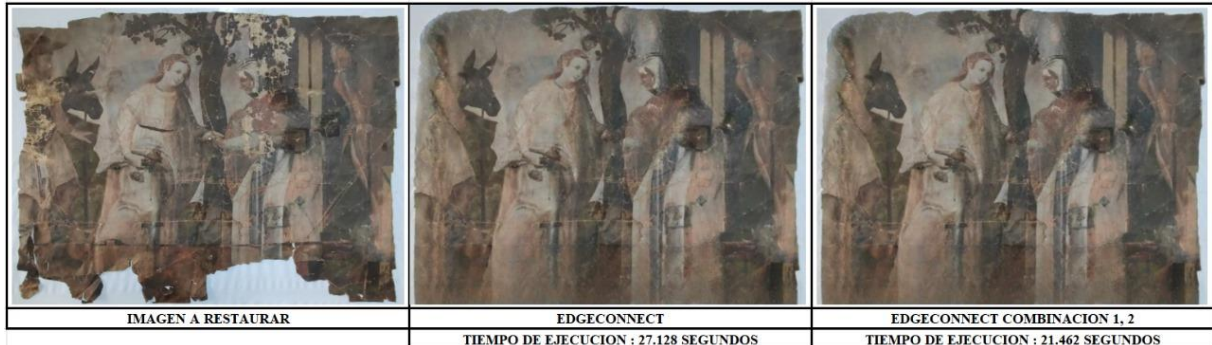
Finalmente al hacer un análisis comparativo de la calidad entre las técnicas, se observó una mejora de más del 6% en el tiempo de procesamiento en relación con la pérdida de calidad en ambos modelos de inteligencia artificial, alcanzando así el objetivo general del proyecto que se ve reflejado en la Tabla 4. Y observando los resultados obtenidos en la Tabla 3, tenemos que, para la primera técnica evaluada, LaMa, se recomienda implementar las tres adaptaciones simultáneamente, ya que esto resulta en una ganancia del 20% en el tiempo de procesamiento y la pérdida de calidad es mínima en comparación con otras adaptaciones. Las tres adaptaciones combinadas optimizan significativamente el rendimiento sin comprometer la integridad visual de las imágenes restauradas, lo que las convierte en la opción preferida para esta técnica. En el caso de la segunda técnica, EdgeConnect, se sugiere aplicar ambas adaptaciones propuestas de manera conjunta. Esta combinación genera una mejora del 25% en el tiempo de procesamiento, con una pérdida de calidad relativamente baja en comparación con otras combinaciones posibles. Las adaptaciones optimizan el uso de recursos y mejoran la eficiencia del modelo, asegurando que el proceso de restauración sea más rápido sin degradar significativamente la calidad de la imagen final. Además, se recomienda utilizar LaMa en lugar de EdgeConnect, ya que, aunque ambas técnicas presentan mejoras similares en el tiempo de procesamiento, LaMa ofrece una superioridad en la calidad de restauración. Al comparar los resultados de ambas técnicas, LaMa no solo proporciona una reducción del tiempo de procesamiento comparable a la de EdgeConnect, sino que también mantiene una mejor calidad en la restauración de imágenes. Esto hace que LaMa sea la opción preferida para aplicaciones donde la calidad visual es crítica, además de la eficiencia temporal. En resumen, la implementación conjunta de las adaptaciones propuestas para ambas técnicas no solo cumple con el objetivo de mejorar el tiempo de procesamiento, sino que también minimiza la pérdida de calidad, con LaMa destacándose como la técnica superior debido a su mejor desempeño en la restauración de imágenes.

3.5.3. Resultados obtenidos en pintura de caballete real

- Con Edge Connect

Figura 19

Resultado visual de restauración con Edge Connect sin y con adaptaciones



- Con LaMa

Figura 20

Resultado visual de restauración con LaMa sin y con adaptaciones



Conclusiones

Como conclusiones se tienen las siguientes.

- Se propusieron tres adaptaciones a la técnica LaMa y dos adaptaciones a la técnica EdgeConnect, las cuales se combinaron para las comparaciones de tiempo y calidad. Las adaptaciones propuestas para LaMa son cambios a las dimensiones de las convoluciones de Fourier en la arquitectura. Las adaptaciones realizadas a EdgeConnect se basan en alterar el número de bloques residuales, en ambos casos, de ocho a seis bloques.
- Se realizaron comparaciones detalladas de tiempo y calidad entre las técnicas actuales de restauración digital de imágenes. A partir de estos análisis, se extrajeron estadísticas clave, como el valor promedio, que indica cómo es típicamente una restauración, y la desviación estándar, la cual refleja la consistencia y confiabilidad de los resultados. Estos hallazgos, que muestran que es difícil obtener resultados significativamente diferentes del promedio, se pueden observar en la Tabla 3.
- Se realizó un análisis comparativo entre las técnicas LaMa y EdgeConnect, mejorando el tiempo de procesamiento y minimizando la pérdida de calidad. LaMa, con tres adaptaciones simultáneas, mejora un 20% el tiempo de procesamiento sin comprometer mucho la calidad, mientras que EdgeConnect, con dos adaptaciones, mejora un 25% pero con una ligera pérdida de calidad. Aunque ambas técnicas muestran mejoras similares en tiempo, LaMa es preferida debido a su superior calidad en la restauración de imágenes, siendo la opción más eficiente para mantener una buena calidad visual.

Recomendaciones

- Este trabajo establece las bases para futuras investigaciones y desarrollos en la mejora de herramientas de restauración digital más rápidas y accesibles, por tanto, se recomienda desarrollar herramientas que no solo sean más eficientes, sino también más accesibles para restauradores con pocos recursos, facilitando la adopción de estas tecnologías en aplicaciones reales. Esto podría incluir la creación de versiones ligeras de estas herramientas, que se puedan utilizar en equipos de bajo rendimiento, permitiendo su implementación en museos y proyectos de restauración de menor escala.
- Se recomienda analizar la dependencia mutua entre los bloques modificados de las redes. De ser el caso de existir una fuerte dependencia entre dos adaptaciones, es posible que únicamente baste realizar la modificación que tenga mayor efecto en el procesamiento a nivel de tiempo en lugar de dos adaptaciones dependientes para obtener resultados comparables (o hasta iguales). Al igual que con la recomendación anterior, esta observación no está justificada y tener evidencia concreta podría ser un problema de interés.
- ¿Es la proporcionalidad del procesamiento de tiempo consumido al predecir una etiqueta en una red neuronal tanto horizontal como verticalmente similar? A nuestro conocimiento, no existen estudios sobre la proporcionalidad del procesamiento de tiempo según la variación del número de capas con relación al número de neuronas por capa. Se recomienda estudiar este problema separadamente a la investigación presente.
- Se recomienda ejecutar el proceso en la tarjeta gráfica en lugar de asignar la carga de trabajo al procesador central, con el fin de mejorar el rendimiento y la eficiencia del sistema, así como limitar el uso de todos los núcleos de la tarjeta gráfica, dado que su empleo completo puede dificultar la comunicación con el procesador central.
- Se recomienda evitar el uso de pinturas de arte abstracto y contemporáneo.

Bibliografía

- Alman, D. H. (1993). CIE technical committee 1-29, industrial color-difference evaluation progress report. *Color Res. Appl.*, 18, 137–139.
- Arora, S., & Barak, B. (2007). *Computational complexity: A modern approach* (draft). Princeton University. Retrieved from <https://theory.cs.princeton.edu/complexity/book.pdf>
- Ballester, C., Bertalmío, M., Caselles, V., Sapiro, G., & Verdera, J. (2001). Filling-in by joint interpolation of vector fields and gray levels. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, 10(8), 1200–1211. <https://api.semanticscholar.org/CorpusID:16484841>
- Bertalmío, M., Sapiro, G., Caselles, V., & Ballester, C. (2000). Image inpainting. *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, 417–424. <https://doi.org/10.1145/344779.344972>
- Bertalmío, M., Vese, L., Sapiro, G., & Osher, S. (2003). Simultaneous structure and texture image inpainting. *IEEE transactions on image processing*, 12(8), 882–889.
- Brandi, C. (1995). *Teoría de la restauración*. Editorial Alianza.
- Brandi, C. (2005). *Teoría de la restauración*. Editorial Gustavo Gili.
- Brigham, E. O., & Morrow, R. E. (1967). The fast Fourier transform. *IEEE spectrum*, 4(12), 63–70.
- Burkov, A. (2020). *Machine learning engineering* (Vol. 1). True Positive Incorporated Montreal, QC, Canada.
- Cao Chenjie and Dong, Q. and F. Y. (2022). Learning Prior Feature and Attention Enhanced Image Inpainting. En G. and C. M. and F. G. M. and H. T. Avidan Shai and Brostow (Ed.), *Computer Vision – ECCV 2022* (pp. 306–322). Springer Nature Switzerland.
- Chi, L., Jiang, B., & Mu, Y. (2020). Fast Fourier Convolution. *Neural Information Processing Systems*. <https://api.semanticscholar.org/CorpusID:227276693>
- Coremans, P. (1996). Intervenciones sobre pintura: criterios y técnicas en la conservación de pinturas de caballete. *Revista de Conservación*, 34(3), 45-59.
- COREMANS, P. (2018). Criterios de intervención en pintura de caballete.
- Criminisi, A., Perez, P., & Toyama, K. (2003). Object removal by exemplar-based inpainting. *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, 2, II–II.
- Darabi, S., Shechtman, E., Barnes, C., Goldman, D. B., & Sen, P. (2012). Image melding: combining inconsistent images using patch-based synthesis. *ACM Trans. Graph.*, 31(4). <https://doi.org/10.1145/2185520.2185578>
- Edwards, B. (1997). Drawing on the Right Side of the Brain. En *CHI'97 Extended Abstracts on Human Factors in Computing Systems* (pp. 188–189).
- Elharrouss, O., Almaadeed, N., Al - Maadeed, S., y Bouridane , A. (2020) . Reproducción de imágenes: una revisión. *Neural Processing Letters*, 51 (2), 2007–2028 . <https://doi.org/10.1007/s11063-020-10275-6>

- Esedoglu, S., & Shen, J. (2002). Digital inpainting based on the Mumford–Shah–Euler image model. *European Journal of Applied Mathematics*, *13*, 353–370. <https://api.semanticscholar.org/CorpusID:7017711>
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. “O’Reilly Media, Inc.”
- Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. Prentice Hall.
- Hays, J., & Efros, A. A. (2007). Scene completion using millions of photographs. *ACM Transactions on graphics (TOG)*, *26*(3), 4–es.
- Huang, J.-B., Kang, S. B., Ahuja, N., & Kopf, J. (2014). Image completion using planar structure guidance. *ACM Trans. Graph.*, *33*(4). <https://doi.org/10.1145/2601097.2601205>
- ICOM. (2020). Código de ética para los museos. Recuperado de <https://icom.museum/es/recursos/>
- Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2017). Globally and locally consistent image completion. *ACM Trans. Graph.*, *36*(4). <https://doi.org/10.1145/3072959.3073659>
- Instituto Andaluz del Patrimonio Histórico. (s.f.). La restauración de pintura: Santa Isabel de Hungría curando a los tiñosos. Instituto Andaluz del Patrimonio Histórico.
- Ju, X., Liu, X., Wang, X., Bian, Y., Shan, Y., & Xu, Q. (2024). *BrushNet: A Plug-and-Play Image Inpainting Model with Decomposed Dual-Branch Diffusion*.
- Kaddour, J., Lynch, A., Liu, Q., Kusner, M. J., & Silva, R. (2022). Causal machine learning: A survey and open problems. *arXiv preprint arXiv:2206.15475*.
- Kelleher, J. D., Mac Namee, B., & D’arcy, A. (2020). *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT press.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*, 1097–1105.
- Liu, D., Sun, X., Li, S., & Zhang, Y.-Q. (2007). Image Compression With Edge-Based Inpainting. *Circuits and Systems for Video Technology, IEEE Transactions on*, *17*, 1273–1287. <https://doi.org/10.1109/TCSVT.2007.903663>
- Liu, H., Jiang, B., Song, Y., Huang, W., & Yang, C. (2020). Rethinking image inpainting via a mutual encoder-decoder with feature equalizations. *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II* *16*, 725–741.
- Nazeri, K., Ng, E., Joseph, T., Qureshi, F. Z., & Ebrahimi, M. (2019). EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning. *CoRR*, *abs/1901.00212*. <http://arxiv.org/abs/1901.00212>
- Ortiz, AS (2012). *Restauración de obras de arte: pintura de caballete (Vol. 11)*. Ediciones AKAL.
- Panda, S. K., Mishra, V., Balamurali, R., & Elngar, A. A. (2021). *Artificial Intelligence and Machine Learning in business management: Concepts, challenges, and case studies*. CRC Press.

- Pathak, D., Krähenbühl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context Encoders: Feature Learning by Inpainting. *CoRR*, *abs/1604.07379*. <http://arxiv.org/abs/1604.07379>
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Pearson.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Sipser, M. (2006). *Introduction to the theory of computation*. Retrieved from <https://theory.cs.princeton.edu/complexity/book.pdf>
- Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., & Lempitsky, V. (2021). Resolution-robust Large Mask Inpainting with Fourier Convolutions. *CoRR*, *abs/2109.07161*. <https://arxiv.org/abs/2109.07161>
- Theobald, O. (2017). *Machine learning for absolute beginners: a plain English introduction* (Vol. 157). Scatterplot press United States.
- Wang, W., Zhang, J. y Tang, Z. (2021). Técnicas eficientes de pintura de imágenes basadas en aprendizaje profundo para la conservación del patrimonio cultural. *Journal of Cultural Heritage*, 54, 251–264. <https://doi.org/10.1016/j.culher.2021.01.002>
- Yan, Z., Li, X., Li, M., Zuo, W., & Shan, S. (2018). Shift-Net: Image Inpainting via Deep Feature Rearrangement. *CoRR*, *abs/1801.09392*. <http://arxiv.org/abs/1801.09392>
- Yang, C., Lu, X., Lin, Z., Shechtman, E., Wang, O., & Li, H. (2016). High-Resolution Image Inpainting using Multi-Scale Neural Patch Synthesis. *CoRR*, *abs/1611.09969*. <http://arxiv.org/abs/1611.09969>
- Yi, Z., Tang, Q., Azizi, S., Jang, D., & Xu, Z. (2020). Contextual residual aggregation for ultra high-resolution image inpainting. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7508–7517.
- Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. S. (2018a). Free-Form Image Inpainting with Gated Convolution. *CoRR*, *abs/1806.03589*. <http://arxiv.org/abs/1806.03589>
- Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. S. (2018b). Generative Image Inpainting with Contextual Attention. *CoRR*, *abs/1801.07892*. <http://arxiv.org/abs/1801.07892>
- Zeng, X., Zhao, H. y Li, P. (2023). Marcos de aprendizaje profundo rápidos para la reproducción de imágenes de alta resolución. *IEEE Transactions on Image Processing*, 32, 1321–1335. <https://doi.org/10.1109/TIP.2023.3244108>
- Zeng, Y., Lin, Z., Yang, J., Zhang, J., Shechtman, E., & Lu, H. (2020). High-resolution image inpainting with iterative confidence feedback and guided upsampling. *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX 16*, 1–17.
- Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2019). A comparative review of recent neural network-based scene parsing methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(6), 1578–1606.