Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Querétaro

Módulo 2

Momento de Retroalimentación

Implementación de un Modelo de Deep Learning

Transferencia de estilos artísticos con Redes Neuronales (STN)

Autor:

A01368818 Joel Sánchez Olvera

TC3007C

Inteligencia artificial avanzada para la ciencia de datos II

Fecha de Entrega:

25 - Noviembre - 2024

Índice

Abstracto

Este reporte presenta el desarrollo de un modelo de Neural Style Transfer utilizando **PyTorch**, que permite transformar imágenes en una variedad de estilos artísticos, tales como el Cubismo, Impresionismo, Arte Moderno y Barroco. Utilizando una arquitectura **VGG-19** pre-entrenada y un modelo personalizado de transferencia de estilo, se logra extraer características de un conjunto de imágenes que pertenecen a un estilo artístico de referencia y aplicarlas a nuevas imágenes, preservando su estructura original mientras se integran patrones y texturas del estilo deseado.

Introducción

En este proyecto, se presenta la implementación de un modelo de transferencia de estilo neural utilizando **PyTorch** para desarrollar una aplicación que recibe imágenes y las transforma a diferentes estilos artísticos los cuales son: el Barroco, Arte Moderno, Impresionismo y Cubismo.

Objetivo

El objetivo de este proyecto es desarrollar un modelo que aprenda las características de diferentes estilos artísticos, tomando el dataset de WikiArt, podemos extraer del dataset diferentes carpetas con imágenes de obras artísticas correspondientes a cada estilo artístico y darle al modelo una biblioteca sobre la cuala prender y extraer características importantes de los estilos y generar modelos que usen esas características para después aplicarlas a cualquier imagen que se desee.

Transferencia de Estilo

La transferencia de estilo es una técnica de procesamiento de imágenes que combina la estructura de una imagen con el estilo visual de otra. Es decir, esta técnica permite transformar una fotografía o imagen en una obra de arte visualmente similar a una pintura famosa o en el caso del proyecto, un estilo artístico específico, como el impresionismo, el cubismo, o el barroco.

Este proceso se logra utilizando modelos de redes neuronales convolucionales (CNN) que son capaces de descomponer una

imagen en características estructurales y estilísticas. Uno de los modelos más utilizados para este propósito es la **VGG-19**, una red profunda pre-entrenada en grandes conjuntos de datos de imágenes. Está diseñada para reconocer patrones complejos en las imágenes, como texturas, bordes y formas.

En el contexto de la transferencia de estilo, se utiliza para extraer dos tipos de características de las imágenes: las de **contenido** y las de **estilo**.

- Las **características de contenido** se refieren a la estructura general de la imagen, es decir, los objetos y su disposición espacial en la imagen de contenido (por ejemplo, la disposición de montañas y personas en una foto de paisaje).
- Las **características de estilo** se obtienen de una imagen de referencia que representa el estilo deseado (por ejemplo, una pintura en el estilo Barroco). Estas características de estilo capturan texturas, patrones de color, y otras propiedades visuales distintivas que le dan a una obra su "apariencia artística".

¿Cómo Funciona la Transferencia de Estilo?

El proceso de transferencia de estilo se basa en minimizar dos funciones de pérdida: la pérdida de contenido y la pérdida de estilo. La pérdida de contenido asegura que la imagen generada mantenga la estructura de la imagen original, mientras que la pérdida de estilo garantiza que la imagen final tenga los patrones y texturas de la imagen de estilo.

1. **Pérdida de Contenido**: Esta pérdida se calcula al comparar las características extraídas de la imagen de contenido y de la imagen generada en capas específicas de la red VGG-19. A través de esta comparación, el modelo se asegura de que la disposición espacial de los elementos de la imagen generada se mantenga similar a la imagen de contenido.
2. **Pérdida de Estilo**: La pérdida de estilo, por otro lado, se calcula utilizando la "matriz de Gram" para las características de la imagen de estilo y la imagen generada. La matriz de Gram mide la relación entre características en distintas posiciones de la imagen, capturando así texturas y

patrones. Al minimizar esta pérdida, el modelo se asegura de que la imagen generada tenga el estilo visual de la imagen de referencia.

3. **Pérdida de Variación Total**: Esta es una pérdida adicional que se aplica para suavizar la imagen generada, eliminando artefactos y asegurando que las transiciones entre colores sean más naturales.

Tecnologías utilizadas

PyTorch: Una biblioteca de aprendizaje profundo que permite diseñar, entrenar y desplegar modelos complejos de manera flexible y eficiente. También se ha utilizado **torchvision**, un conjunto de herramientas para la manipulación de imágenes que facilita el procesamiento y la aplicación de transformaciones. Estas tecnologías son ideales para proyectos de visión por computadora debido a su rendimiento y capacidad de optimización.

VGG-19: Una red pre-entrenada en tareas de clasificación de imágenes, ha sido elegida como extractora de características. La importancia de VGG-19 radica en su capacidad para capturar características visuales a diferentes niveles de abstracción, lo cual es esencial en la transferencia de estilo, ya que los niveles iniciales capturan texturas, mientras que los niveles más profundos capturan estructuras más complejas.

Matriz de Gram: Una herramienta esencial en la transferencia de estilo neuronal, es utilizada para capturar y representar las características de estilo de una imagen. Esta matriz se calcula a partir de las activaciones de una capa de red neuronal convolucional y describe la correlación entre diferentes filtros de esa capa.

Al calcular el producto interno entre las activaciones, la **matriz de Gram** permite capturar patrones y texturas característicos de la imagen de estilo, independientemente de su ubicación espacial. En la transferencia de estilo, la pérdida de estilo se mide comparando las matrices de Gram de la imagen de estilo y la imagen generada, lo que ayuda a que el modelo aprenda a replicar el aspecto visual de la imagen de referencia en la imagen final. Gracias ésto, el modelo puede mantener la consistencia en texturas y colores,

logrando una representación visual del estilo artístico.

Red Neuronal Convolucional (CNN): Una arquitectura de red neuronal diseñada específicamente para procesar datos estructurados en forma de imágenes, permitiendo la detección y reconocimiento de patrones visuales complejos.

Las CNNs constan de capas de convolución que aplican filtros a pequeñas secciones de la imagen, extrayendo características locales como bordes, texturas y formas. A medida que las capas se profundizan, la red aprende a reconocer patrones más abstractos, como la estructura de objetos completos. Las CNNs también incluyen capas de pooling que reducen la dimensionalidad, manteniendo las características más relevantes y mejorando la eficiencia de procesamiento, así como capas de activación, que introducen no linealidades, permitiendo que la red capture relaciones complejas en los datos. En combinación con capas completamente conectadas al final, las CNNs logran analizar y clasificar imágenes con alta precisión.

Recursos computacionales

- **Equipo**: Macbook Pro
- **Procesador**: Apple M3 Pro
- **Memoria**: 18 Gb

Procesamiento y análisis de los datos

Descripción del Dataset

Para realizar la transferencia de estilo, se utilizó el dataset WikiArt que contiene imágenes organizadas en carpetas, donde cada carpeta representa un estilo artístico diferente (**Cubismo**, **Impresionismo**, **Arte Moderno** y **Barroco**, entre otros.)Como los ejemplos mostrados a continuación:

Figura 1.1(Ejemplo Estilo Barroco)

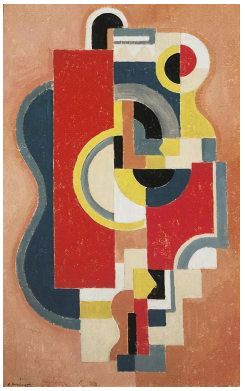


Figura 1.2(Ejemplo Estilo Cubismo)

Para el desarrollo del proyecto se tomaron 4 de estos estilos debido a las limitaciones de poder de cómputo que se tienen, Sin embargo, puede ser escalado mucho más a otras clases, lo cual se pretende implementar y mejorar en futuras iteraciones del proyecto.

De los estilos artísticos tomados pudimos obtener para cada estilo:

- **Barroco** : 4,241 imágenes
- **Arte Moderno** : 4,335 imágenes
- **Cubismo** : 2,236 imágenes
- **Impresionismo** : 13,061 imágenes

Las imágenes que se encontraron son representativas del estilo correspondiente, son imágenes como pinturas, esculturas y otras obras de arte. Éstas sirven como referencia visual para que el modelo aprenda patrones y texturas características de cada estilo. para el procesamiento y entrenamiento de los modelos, solo se utilizó una cantidad definida de 500 imágenes para no sobrecargar el sistema actual sobre el cual entrena el modelo.

Extracción de Características de Estilo: Matriz de Gram Promedio

El objetivo principal del procesamiento de los datos de estilo es capturar las correlaciones entre las características visuales de las imágenes de cada estilo mediante una **matriz de Gram promedio**. La matriz de Gram se utiliza para calcular las relaciones entre diferentes filtros de características en las capas de una red neuronal convolucional (CNN). Estas relaciones capturan las texturas, colores y patrones visuales de cada imagen, que son esenciales para replicar el estilo.

Para calcular la matriz de Gram promedio de cada estilo, usamos una función llamada `calculate_average_gram_matrix`. Esta función toma el conjunto de imágenes de cada estilo, las procesa a través de un modelo **VGG-19** pre-entrenado para extraer las características, y luego calcula la matriz de Gram para cada imagen individual. Posteriormente, estas matrices de Gram se combinan para obtener una matriz de Gram promedio que representa el estilo en su conjunto.

Esto permite que el modelo generalice las características de cada estilo y las aplique a nuevas imágenes, independientemente de la estructura específica de las imágenes de estilo utilizadas en el entrenamiento.

Modelo

Descripción del modelo y arquitectura

Para lograr la transferencia de estilo artístico en imágenes, se ha implementado una red neuronal personalizada denominada **StyleTransferNetwork**. Red que puede ser consultada en el archivo StyleTransferNetwork.py

Esta red está diseñada para capturar las características visuales distintivas de estilos artísticos específicos y aplicarlas a nuevas imágenes de contenido, permitiendo que adopten los patrones y texturas del estilo deseado sin perder la estructura de la imagen original.

Esta red está compuesta por varios módulos principales:

1. **Capas de Convolución y Normalización:** La primera etapa de la red utiliza varias capas convolucionales, seguidas de una normalización de instancia y activaciones ReLU, para extraer las características básicas de la imagen de entrada. Estas capas convolucionales permiten al modelo aprender texturas y patrones específicos del estilo artístico.
2. **Bloques Residuales:** Los bloques residuales preservan la coherencia visual y estructural de la imagen. Cada bloque residual contiene capas adicionales de convolución y normalización, permitiendo al modelo aprender características complejas sin degradar la información visual de la imagen original. Al reutilizar y preservar características intermedias, los bloques residuales garantizan que la estructura de la imagen de contenido no se vea alterada de manera significativa, a la vez que permiten que los patrones de estilo se integren de forma más natural en la imagen.
3. **Capas de Deconvolución (Upsampling):** Después de capturar y manipular las características de la imagen mediante las capas de convolución y los bloques residuales, se emplean capas de deconvolución para restaurar la resolución original de la imagen.

Función de Pérdida y Optimización

El éxito del modelo depende en gran medida de la función de pérdida, que ha sido diseñada para equilibrar adecuadamente la estructura de la imagen y los elementos estilísticos. La función de pérdida se compone de tres componentes principales, cada uno de los cuales cumple una función específica en la transferencia de estilo:

1. **Pérdida de Contenido:** Este componente asegura que la estructura de la imagen original se mantenga en la imagen estilizada. Para ello, se extraen características de

una capa específica de la red VGG-19 preentrenada, que se utiliza como referencia de contenido. Al minimizar la diferencia entre las características de contenido de la imagen de entrada y las de la imagen generada, el modelo logra preservar la disposición espacial de los elementos de la imagen.

2. **Pérdida de Estilo:** La pérdida de estilo es responsable de capturar las texturas y patrones del estilo deseado en la imagen de salida. Esto se logra mediante la **matriz de Gram**, el modelo aprende a replicar los elementos visuales característicos del estilo (colores, patrones de pinceladas, texturas) en la imagen de salida, logrando una apariencia artística coherente.
3. **Pérdida de Variación Total (TV):** La pérdida de variación total actúa como una regularización, suavizando la imagen de salida y eliminando posibles artefactos de ruido generados durante el proceso de estilización. El uso de la pérdida de variación total contribuye a que la imagen generada tenga una apariencia profesional y estética.

Para optimizar esta función de pérdida, se utiliza el optimizador **Adam**. Además, se emplea un programador de tasa de aprendizaje (**StepLR**), que reduce progresivamente la tasa de aprendizaje a medida que avanza el entrenamiento, facilitando una convergencia más estable y evitando oscilaciones en la pérdida.

Proceso de Entrenamiento

Durante el entrenamiento, la red recibe pares de imágenes de contenido y de estilo, extrayendo características estilísticas mediante el modelo VGG-19. A través de un proceso iterativo, la red ajusta sus parámetros para minimizar la función de pérdida, equilibrando los componentes de contenido y estilo. El modelo aprende a aplicar texturas, colores y patrones característicos del estilo artístico en la imagen de contenido, generando una representación visual que refleja tanto la esencia del estilo como la estructura original de la imagen de entrada.

La **StyleTransferNetwork** es capaz de generalizar el estilo aprendido a cualquier imagen de contenido, lo que le permite adaptarse a múltiples aplicaciones artísticas y visuales, tales como la personalización de imágenes y la creación de obras de arte digitales.

Una vez entrenada, esta red puede transformar imágenes en tiempo real, aplicando el estilo artístico de manera rápida y eficiente, haciendo que la transferencia de estilo sea accesible para una amplia variedad de usuarios y aplicaciones.

Éste proceso debe repetirse para cada uno de los estilos artísticos deseados, lo que se hace con el dataset es leer completa la carpeta de las imágenes, tomar el nombre de las carpetas para tobtener el nombre de las clases a entrenar, y posteriormente el modelo entrena sobre el dataset de cada estilo artístico, éste proceso puede ser consultado en : Art_Generator_STN.ipynb o en Art_Generator_STN.py, de ésta manera, se genera un modelo para cada estilo artístico y se guarda en el mismo directorio del proyecto como un archivo .pth para ser utilizado luego en el script Art_Generator_Tester.py, que toma los modelos generados para cada uno de los modelos artísticos y al darle una ruta de una image, usa los modelos para aplicarle las características de cada estilo y después nos muestra la imagen original comparada con las imágenes estilizadas.

Resultados

Al finalizar el proceso de entrenamiento para cada estilo, evaluamos el desempeño de los modelos generados mediante la observación de la pérdida de error en función de las épocas. Este análisis permite identificar la eficiencia con la que el modelo captura y replica los patrones de cada estilo artístico.

Análisis de la pérdida de error

Para monitorear el progreso del modelo, se registraron los valores de pérdida cada 100 épocas, generando gráficos de la pérdida acumulada sobre las épocas para cada uno de los estilos artísticos.

A continuación, se presentan los resultados para el estilo **Modern Art:**

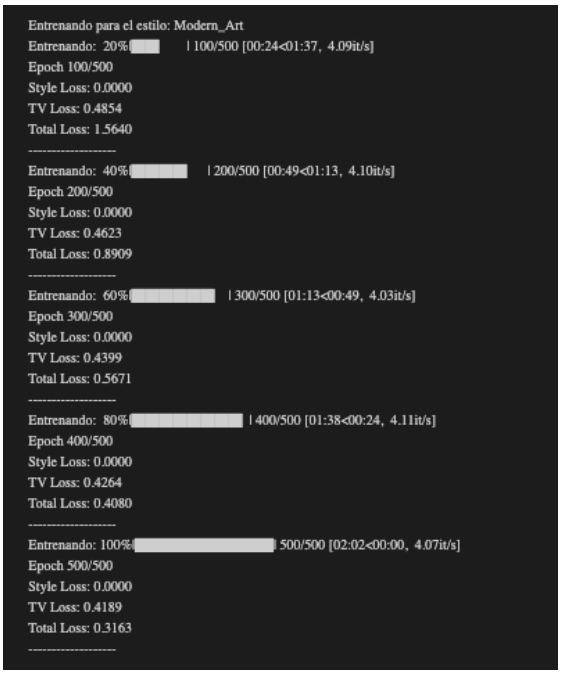


Figura 2.1. Resultado de Entrenamiento

Podemos ver que en la figura2.1 el error va disminuyendo cada 100 épocas de manera constante.

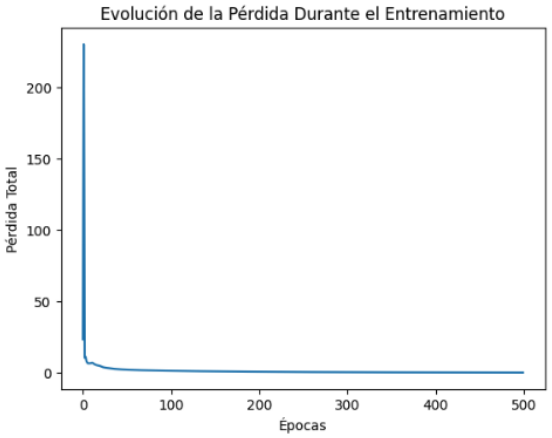


Figura 2.2. Gráfica de Pérdida (TV) sobre épocas.

Podemos observar en la figura la gráfica de pérdida y como en ella vemos la pérdida para la clase Modern Art disminuyendo. Esto demuestra que el modelo trabaja de una manera correcta y va disminuyendo la pérdida con cada iteración sobre la información que tiene.

Podemos ver de éstas gráficas que estamos disminuyendo la pérdida de manera eficiente y controlada.

Al usar el Script Art_Generator_Tester.py con la test_img.jpg podemos analizar los resultados arrojados por elmodelo después de entrenar para cada uno de los estilos artísticos mencionados

Podemos ver que el modelo aprendió correctamente de cada

uno de los estilos y puede aplicar sus características y texturas a la imagen para crear una obra de arte única con elementos de estilos artísticos específicos.

Textura y patrón: En el caso del estilo Barroco, el modelo replicó con precisión los patrones de pinceladas y texturas propias de este estilo, aplicando efectos visuales que evocan los acabados complejos y la riqueza visual de las obras barrocas. En contraste, en el estilo **Modern Art**, se observó una incorporación de patrones más abstractos y una paleta de colores vivos, característicos de la libertad expresiva de este estilo.

Aplicación de color: En el estilo **Impresionismo**, se observó una aplicación precisa de colores pastel y un tratamiento de luz suave, aspectos que son distintivos de este movimiento artístico.

Conclusión

El proyecto logró implementar un modelo de transferencia de estilo que puede replicar con éxito los estilos artísticos de Barroco, Arte Moderno, Cubismo e Impresionismo. Utilizando una red neuronal convolucional de transferencia de estilo, el modelo fue capaz de aprender las texturas, patrones y paletas de color específicos de cada estilo y aplicarlos a nuevas imágenes.

La reducción consistente en la pérdida durante el entrenamiento confirma que el modelo se optimizó correctamente, logrando una una buena integración de las características de los estilos en la imagen generada.

Sin embargo, el proyecto puede ser mejorado para que a nivel visual las imágenes sean un poco más detalladas y más definidas después de ser pasadas por los modelos, mejora que puede ser implementada en iteraciones futuras del proyecto.

Mejoras, Fine-Tuning y Resultados

El objetivo principal de las mejoras implementadas en el modelo fue lograr imágenes más claras y definidas tras la aplicación del estilo artístico, abordando las limitaciones observadas en los resultados iniciales. En la versión original del modelo, los resultados presentaban características estilísticas difusas y patrones que no capturaban plenamente las particularidades de cada estilo artístico. Por esta razón, se realizaron ajustes significativos en los parámetros y configuraciones del modelo, priorizando la mejora visual.

Estas mejoras permiten que los patrones estilísticos característicos de cada estilo artístico (como las pinceladas suaves y los colores pastel en el Impresionismo, o las formas geométricas y abstractas en el Cubismo) se repliquen de manera más fiel en las imágenes generadas.

Al garantizar que las texturas, colores y patrones estilísticos sean más precisos, el modelo logra transformar las imágenes de manera más efectiva, respetando la estructura original de la imagen y fusionándola con los elementos del estilo deseado.

Además de mejorar la calidad visual, estas modificaciones facilitan un análisis más detallado de los resultados generados, lo que aumenta la utilidad del modelo en aplicaciones artísticas, educativas y analíticas.

Estas mejoras destacan la importancia del ajuste fino (fine-tuning) en el desarrollo de modelos de deep learning.

1. Cambios Realizados

Hiperparámetros Ajustados:

1. **Pesos del estilo (`style_weight`)**:
   - Modelo original: 1×1051 \times 10^51×105.
   - Modelo mejorado: 1×10121 \times 10^{12}1×1012.
2. **Pesos de variación total (`tv_weight`)**:
   - Modelo original: 1×10−61 \times 10^{-6}1×10−6.
   - Modelo mejorado: 1×10−21 \times 10^{-2}1×10−2.
3. **Épocas de entrenamiento (`epochs`)**:
   - Modelo original: 500.
   - Modelo mejorado: 600 y 1200.
4. **Capas Seleccionadas**:
   - En el modelo original, se utilizaron las capas `conv1_1`, `conv2_1`, `conv3_1`, `conv4_1`, y

`conv5_1` de la red VGG-19, asignando pesos iniciales estáticos.

5. **Pesos Estilísticos por Capa**:
    - Modelo original:
        - `conv1_1`: 1.0, `conv2_1`: 0.8, `conv3_1`: 0.4, `conv4_1`: 0.2, `conv5_1`: 0.2.
    - Modelo mejorado:
        - `conv1_1`: 0.2, `conv2_1`: 0.2, `conv3_1`: 0.4, `conv4_1`: 0.3, `conv5_1`: 0.2.

2. Impacto en los Resultados

Los cambios realizados resultaron en una mejora significativa en la calidad visual y definición de las imágenes generadas. A continuación, se comparan los resultados antes y después de las mejoras para cada estilo artístico.

Imágenes Generadas:

1. **Impresionismo**:

    Antes: Colores deslavados y patrones incompletos.

    Después: Colores más vibrantes y pinceladas mejor definidas.

2. **Arte Moderno**:

    Antes: Detalles difusos y texturas inconsistentes.

    Después: Incorporación precisa de colores vivos y patrones abstractos más nítidos.

3. **Barroco**:

    Antes: Pérdida de detalle en las texturas, con fondos oscuros poco definidos.

    Después: Mejor representación de sombras y riqueza visual en las texturas.

4. **Cubismo**:

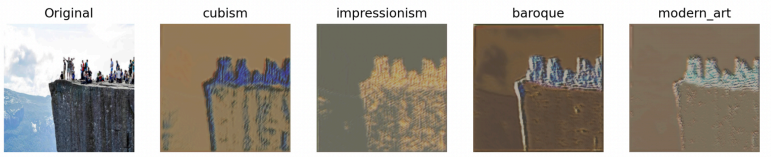
    Antes: Líneas y formas geométricas dispersas.

    Después: Representación más clara de patrones cubistas y colores distintivos.

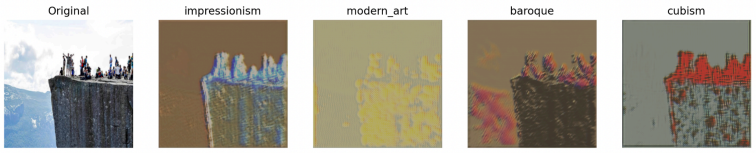
3. Comparativa Visual

Se incluye una comparación visual entre el modelo original y el modelo mejorado.

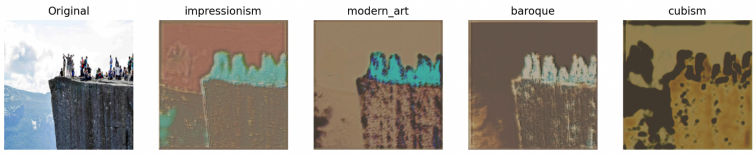
Resultados Modelo Original



Resultados Modelo Mejorado 600 epochs.



Resultado Modelo Mejorado 1200 epochs.



1. **Modelo Original (Primera Fila)**:
   En los resultados del modelo original, se observa que las imágenes estilizadas presentan un claro **dominio de superposición de colores**, lo cual se logra con relativa facilidad. Sin embargo, el modelo no logra capturar las texturas ni los estilos de líneas característicos de los estilos artísticos seleccionados.
2. **Modelo Mejorado con 600 Épocas** :
   El modelo mejorado, entrenado durante 600 épocas, muestra una **mejora significativa en los detalles texturales y las estructuras estilísticas** en comparación con el modelo original.
3. **Modelo Mejorado con 1200 Épocas (Tercera Fila)**:
   Tras extender el entrenamiento a 1200 épocas, los resultados muestran una **representación mucho más refinada de los estilos artísticos**, con un enfoque más claro en las texturas y los estilos de líneas. Este modelo logra acercarse más a los elementos distintivos de cada corriente artística.

4. Conclusiones

Las mejoras implementadas, como el ajuste de los pesos estilísticos, el incremento en las épocas de entrenamiento, y los cambios en las capas seleccionadas, resultaron en una mejora significativa en la fidelidad estilística de las imágenes generadas. Esto demuestra la importancia del fine-tuning en modelos de transferencia de estilo neural, especialmente cuando se busca capturar detalles visuales específicos de estilos artísticos.