



# Cálculo de Similitud entre Imágenes Artísticas y Generadas por Stable Diffusion utilizando Redes Convolucionales Siamesas

*Trabajo Terminal No. TT2025–1\_IA–008*

*Alumnos: Castro Elvira Diego <sup>1</sup>, Pineda Rugerio Navil <sup>2</sup>*

*Directores: Ramos Aguilar Ricardo, García Ramírez Jesús*

*email: {diego.castro.elvira <sup>1</sup>, naviladenip <sup>2</sup>}@gmail.com*

## Abstract

El proyecto busca calcular la similitud entre imágenes artísticas originales y aquellas generadas por Stable Diffusion XL Refiner 1.0, utilizando redes convolucionales siamesas con ajuste fino a una ResNet50. Para ello, se crea un conjunto de datos con imágenes de WikiArt y sus versiones generadas por IA. Se seleccionan y extraen descriptores de las imágenes para generar valores de similitud, que son comparados con los resultados de la red siamesa. El rendimiento se evalúa utilizando métricas de regresión como el MSE y MAE, con el objetivo de obtener un valor de la similitud entre imágenes.

**Palabras clave**— Descriptores de imagen, Redes convolucionales siamesas, Similitud de imágenes, Stable Diffusion

## 1 Introducción

El uso de la inteligencia artificial (IA) en la creación artística ha crecido considerablemente, planteando interrogantes sobre la originalidad de las obras y su posible similitud con creaciones humanas, mayormente en el arte digital, donde modelos como *Stable Diffusion* [11] permiten la generación de imágenes realistas y estilizadas. Sin embargo, la falta de herramientas que permitan evaluar la autenticidad de estas obras ha generado la necesidad de desarrollar métodos que ayuden a identificar posibles similitudes en el arte generado por IA.

Este proyecto surge del interés tanto académico como profesional por mejorar los procesos de autenticación de obras artísticas digitales, facilitando la detección de similitudes entre imágenes originales y aquellas generadas por IA. Para ello, se ha optado por emplear un enfoque basado en redes neuronales convolucionales siamesas, capaces de comparar imágenes a partir de descriptores visuales y medir su grado de similitud. La metodología sigue un proceso incremental que incluye la creación de un conjunto de datos, el ajuste fino de un modelo pre-entrenado y la implementación de una red siamesa para el análisis de las imágenes.

A través de este trabajo, se busca aportar una herramienta para la valoración de la originalidad en el arte generado por IA, apoyando a artistas y la comunidad de investigación en la identificación de posibles similitudes.

## 2 Objetivos

### 2.1 Objetivo general

Construir un modelo para calcular la similitud entre imágenes de pinturas artísticas y aquellas generadas por Stable Diffusion XL Refiner 1.0 mediante el uso de descriptores de imagen y una red convolucional siamesa entrenada con fine-tuning a una ResNet50.



## 2.2 Objetivos específicos

- Crear un conjunto de datos con imágenes originales de WikiArt [8] y sus versiones generadas por Stable Diffusion XL Refiner 1.0.
- Seleccionar y extraer los descriptores de las imágenes del dataset de WikiArt y de las imágenes generadas por Stable Diffusion utilizando técnicas de procesamiento de imágenes.
- Diseñar y entrenar una red convolucional siamesa para calcular la similitud entre las imágenes del dataset WikiArt y las imágenes generadas por Stable Difussion.
- Evaluar el rendimiento del modelo mediante el uso de métricas de regresión como Mean Squared Error (MSE)[10] o Mean Absolute Error (MAE)[9] para comparar las similitudes entre imágenes generadas por Stable Diffusion y obras originales.

## 3 Planteamiento del Problema

La IA generativa ha permitido crear obras más complejas estéticamente, comparables a las creadas por artistas humanos, lo que implica nuevas dificultades para evaluar la originalidad y autenticidad de estas. Un aspecto clave al comparar imágenes es la identificación de detalles que, debido a la complejidad de las pinturas artísticas, varían en características perceptibles al ojo humano, como la composición, el color o la iluminación. Sin embargo, comparar estos elementos de manera conjunta se vuelve un desafío considerable.

Aunque ya existen estudios y herramientas de software centrados en la comparación y el análisis del arte digital, persiste una brecha en la investigación, no se han desarrollado modelos capaces de evaluar la similitud entre las obras originales y las imágenes generadas por técnicas de image-to-image, que buscan replicar el estilo de obras creadas por humanos.

Además, la mayoría de los estudios se centran en estilos específicos, sin abarcar una amplia gama de géneros, y no existen conjuntos de datos que incluyan tanto obras originales como generadas por IA, lo que limita el desarrollo de modelos más robustos para el análisis comparativo.

La Tabla 1 presenta algunos trabajos relevantes para contextualizar este panorama:

Tabla 1: Trabajo relacionado.

Software	Descripción
DeepMatch in Text-to-Image Generative Models [6]	Detector de copyright de imágenes generadas con tres modelos text-to-image, Stable Difussion v1.4, Stable Difussion v2 y OpenJourney. Para ello realiza una comparación utilizando la técnica DeepMatch.
ResNet50 para clasificación de obras artísticas [2]	Clasificación de obras artísticas del dataset WikiArt según su género, y la creación de un modelo basado en una ResNet50 para realizar una detección de autores de las obras de este dataset.
Preventing AI Art Plagiarism [3]	Modelo de machine learning para detectar la autenticidad de imágenes de obras artísticas generadas por modelos de IA inspiradas en el estilo de Leonardo Da Vinci.

## 4 Justificación

Debido a la falta de modelos especializados que aborden la generación y evaluación de similitud en pinturas de diversos géneros artísticos, este proyecto propone un nuevo enfoque al combinar la generación de imágenes similares mediante técnicas de image-to-image con la evaluación de su similitud, utilizando redes convolucionales siamesas, y plantea desarrollar una compilación de un conjunto de datos que incluya obras originales y sus versiones generadas por IA en diversos estilos, ya que a diferencia de estudios previos, este proyecto propone una gama de géneros artísticos más diversa.

Con ello, se busca abordar el desafío de la autenticación y originalidad en el arte digital, beneficiando a artistas, galerías, plataformas en línea y la comunidad de investigación. Su viabilidad está asegurada por el uso de recursos computacionales accesibles y un cronograma adecuado para su desarrollo y prueba.

## 5 Productos o Resultados esperados

La Figura 1 muestra el digrama de componentes del proyecto.

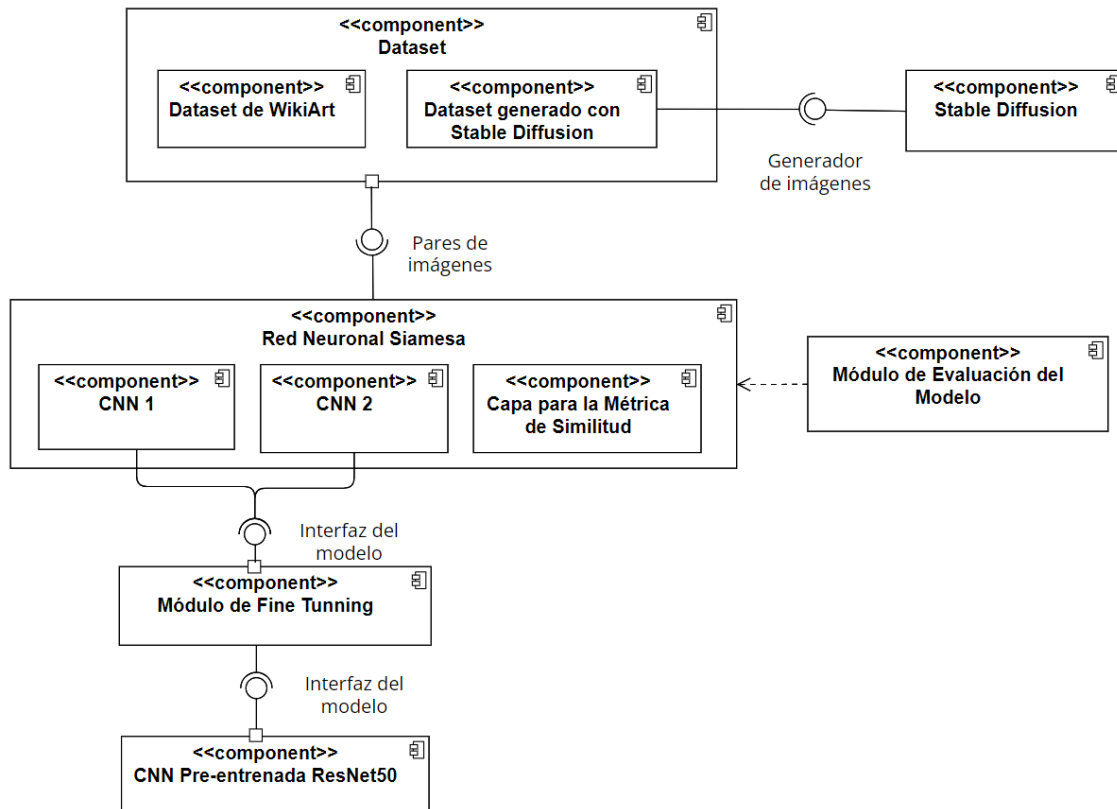


Figura 1: Diagrama de Componentes del proyecto.



A continuación, se describen los productos esperados.

1. Dataset que incluye imágenes originales de WikiArt junto con sus versiones generadas por Stable Diffusion XL Refiner 1.0.
2. Modelo de Red Convolutiva Siamesa entrenada para el cálculo de similitud.
3. Pruebas de validación del modelo.
4. Reporte técnico del sistema desarrollado.

## 6 Metodología

La metodología seleccionada tendrá la siguiente estructura:

1. Investigación preliminar. Se realizará una revisión de la literatura sobre técnicas de redes convolucionales siamesas y métodos de comparación de imágenes, enfocándose en proyectos previos relacionados con la similitud entre imágenes generadas por inteligencia artificial y obras de arte originales.
2. Creación del conjunto de datos. Se generará un conjunto de datos que consiste en la generación de imágenes por Stable Diffusion XL Refiner 1.0, utilizando prompts basados en las descripciones de las imágenes originales obtenidas de WikiArt.
3. Selección de descriptores de imágenes. Se estudiarán diferentes descriptores de características visuales (e.g., colores dominantes, textura, bordes, etc.) y se elegirán aquellos más apropiados para cuantificar la similitud entre las imágenes originales y las generadas. Esto podría incluir descriptores sencillos o más complejos, dependiendo de los resultados de la investigación.
4. Generación de valores de similitud preliminares. A partir de los descriptores seleccionados, se calcularán valores de similitud aplicando factores ponderados a cada descriptor. Los resultados se utilizarán para obtener un valor numérico que represente la similitud entre cada par de imágenes.
5. Fine-tuning del modelo ResNet50. Se utilizará un modelo pre-entrenado (ResNet50) y se realizará un ajuste fino con las imágenes de WikiArt para especializar el modelo en la clasificación de géneros artísticos. Se seguirán los siguientes pasos:
  - División del conjunto de datos en entrenamiento, validación y prueba.
  - Ajuste de hiperparámetros como la tasa de aprendizaje, tamaño de lote y número de épocas.
  - Guardado de los pesos del modelo entrenado.
6. Entrenamiento de la red convolutiva siamesa. Se implementará una red siamesa utilizando las imágenes originales y generadas por Stable Diffusion. El modelo se entrenará con los pesos pre-entrenados de ResNet50 y se configurará para producir un valor de regresión que mida la similitud entre las imágenes.



7. Evaluación del modelo. Se evaluará el rendimiento de la red siamesa utilizando métricas de regresión como: MSE y MAE. Además, los resultados obtenidos serán comparados con los valores de similitud preliminares obtenidos a partir de los descriptores.
8. Conclusiones y mejoras futuras. Con base en los resultados obtenidos, se identificarán posibles mejoras al modelo, como la inclusión de más descriptores, la mejora en el ajuste fino del modelo o la implementación de nuevas técnicas de similitud de imágenes.

## 7 Cronograma



Figura 2: Cronograma del proyecto para Navil y Diego



## Referencias

- [1] Duran: Cómo clasificar obras de arte por estilo en 7 líneas de código. Medium (2021)
- [2] Duran: ¿podrías adivinar el pintor de un cuadro que no conoces? las redes neuronales sí. Medium (2021b)
- [3] Ehan: Preventing ai art plagiarism. Medium (2023)
- [4] Knuth, D.E.: The Art of Computer Programming, vol. 2. Addison-Wesley (1997)
- [5] Lamport, L.: Latex: A document preparation system. Addison Wesley **2** (1994)
- [6] Moayeri: Rethinking artistic copyright infringements in the era of text-to-image generative models. University of Maryland, Computer Science Department (2024)
- [7] Somepalli: Understanding and mitigating copying in diffusion models. NeurIPS (2023)
- [8] WikiArt: Wikiart. <https://www.wikiart.org/>, [Online; accessed 13-September-2024]
- [9] Wikipedia contributors: Mean absolute error — Wikipedia, the free encyclopedia. [https://en.wikipedia.org/w/index.php?title=Mean\\_absolute\\_error&oldid=1216949643](https://en.wikipedia.org/w/index.php?title=Mean_absolute_error&oldid=1216949643) (2024), [Online; accessed 15-September-2024]
- [10] Wikipedia contributors: Mean squared error — Wikipedia, the free encyclopedia. [https://en.wikipedia.org/w/index.php?title=Mean\\_squared\\_error&oldid=1228454019](https://en.wikipedia.org/w/index.php?title=Mean_squared_error&oldid=1228454019) (2024), [Online; accessed 15-September-2024]
- [11] Wikipedia contributors: Stable diffusion — Wikipedia, the free encyclopedia. [https://en.wikipedia.org/w/index.php?title=Stable\\_Diffusion&oldid=1245124206](https://en.wikipedia.org/w/index.php?title=Stable_Diffusion&oldid=1245124206) (2024), [Online; accessed 13-September-2024]



## 8 Alumnos y directores

*Castro Elvira Diego.*- Alumno de la carrera de Ing. en Inteligencia Artificial  
Boleta: 2022710168, Tel. 2411081478, email [diego.castro.elvira@gmail.com](mailto:diego.castro.elvira@gmail.com)

Firma: \_\_\_\_\_

*Pineda Rugerio Navil.*- Alumna de la carrera de Ing. en Inteligencia Artificial  
Boleta: 2022710240, Tel. 2461509006 , email [naviladenip@gmail.com](mailto:naviladenip@gmail.com)

Firma: \_\_\_\_\_

*Ricardo Ramos Aguilar.*- email [rramosa@ipn.mx](mailto:rramosa@ipn.mx)

Firma: \_\_\_\_\_

*García Ramírez Jesús.*- email [jegarciara@ipn.mx](mailto:jegarciara@ipn.mx)

Firma: \_\_\_\_\_