

Cálculo De Similitud Entre Imágenes Artísticas Y Generadas Por Stable Diffusion Utilizando Redes Convolucionales Siamesas

Navil Pineda Rugerio, Diego Castro Elvira
Ingeniería en Inteligencia Artificial

Dr. Ricardo Ramos Aguilar

Dr. Jesús García Ramírez



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"



ÍNDICE

1. Introducción
2. Planteamiento del problema
3. Objetivos
4. Hipótesis
5. Marco Teórico
6. Estado de Arte
7. Metodología
8. Cronograma
9. Referencias

1. Introducción

- Crecimiento de los modelos generativos de imágenes en el arte
- Surge la necesidad de evaluar la originalidad de obras generadas por IA.
- Desarrollar un método para comparar imágenes originales de arte y creadas por un modelo generativo, centrándose en medir su grado de similitud.

2. Planteamiento del Problema



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



Definición del Problema

- La IA genera obras artísticas comparables a las humanas, dificultando evaluar su autenticidad y originalidad.
- Comparar en conjunto los elementos que componen una obra es complejo, debido a las variaciones perceptibles en cada pieza.

Justificación

- Falta de modelos especializados para la generación y evaluación de similitud en pinturas de diferentes géneros artísticos.
- Se propone un enfoque que combina la generación *image-to-image* y la evaluación de su similitud usando redes convolucionales siamesas.
- Se desarrollará un conjunto de datos que incluya obras originales y versiones generadas por IA.

3. Objetivos



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

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 ajuste fino a una *ResNet50*.

Objetivos específicos

- Crear un conjunto de datos.
- Seleccionar y extraer los descriptores de las imágenes.
- Crear y entrenar una red convolucional siamesa para calcular la similitud entre las imágenes.

4. Hipótesis



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

Hipótesis

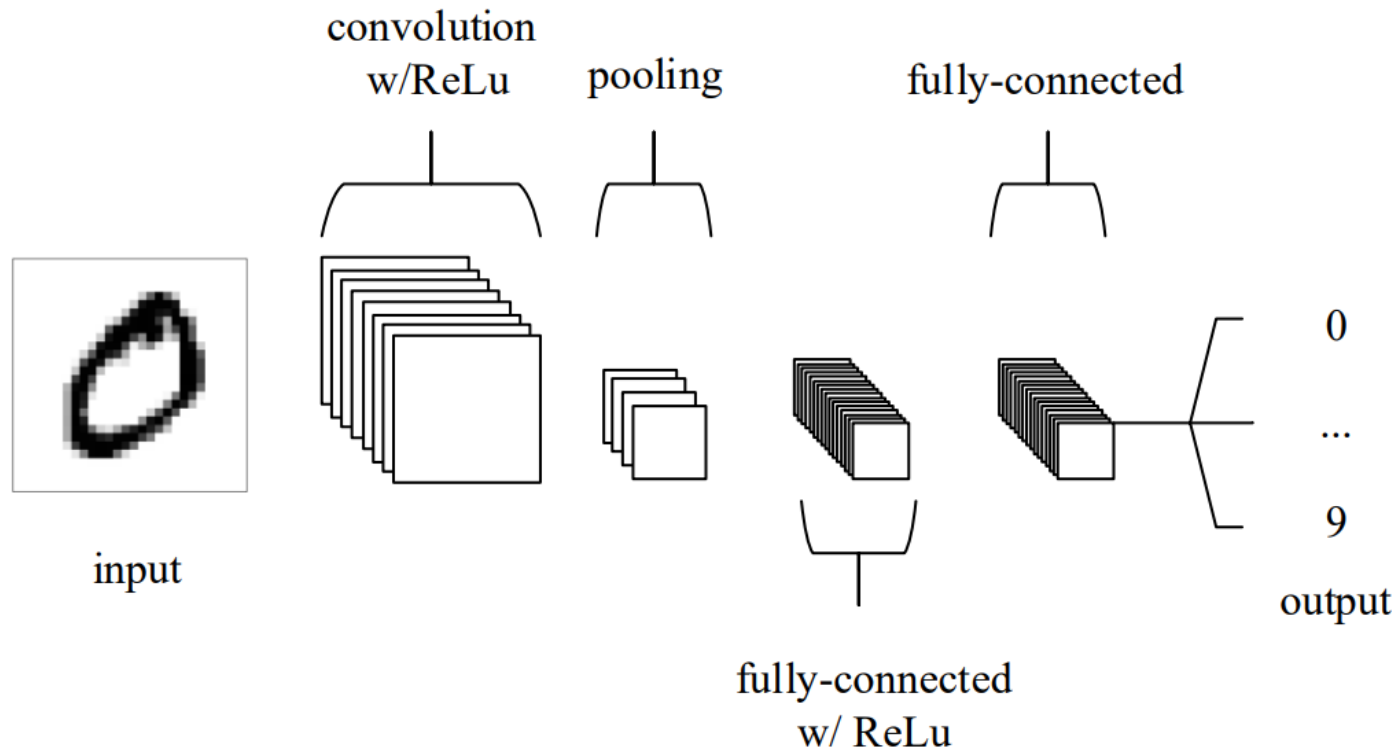
Las redes neuronales siamesas permiten medir cuantitativamente el nivel de similitud compositiva y estilística entre pinturas artísticas originales y sus contrapartes generadas por *Stable Diffusion XL Refiner 1.0*, mediante la identificación y comparación de sus patrones visuales o semánticos.

5. Marco Teórico



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

Redes Neuronales Convolucionales (CNN)



- Las CNN aprenden automáticamente características.
- Esta capacidad para extraer patrones visuales hace que sean utilizadas para tareas de clasificación y detección de objetos.

Fine-tunning

- Ajuste de modelos preentrenados para adaptarlos a un nuevo conjunto de datos específico.
- Las primeras capas suelen capturar características generales y las capas finales se ajustan específicamente al dominio de las imágenes objetivo.

Red Residual (*ResNet*)

- Modelo de CNN introducida por Kaiming He et al. en 2015.
- Una *ResNet50* se compone de 50 capas organizadas en bloques residuales.
- Entrenada con el conjunto de datos *ImageNet*, con más de un millón de imágenes.

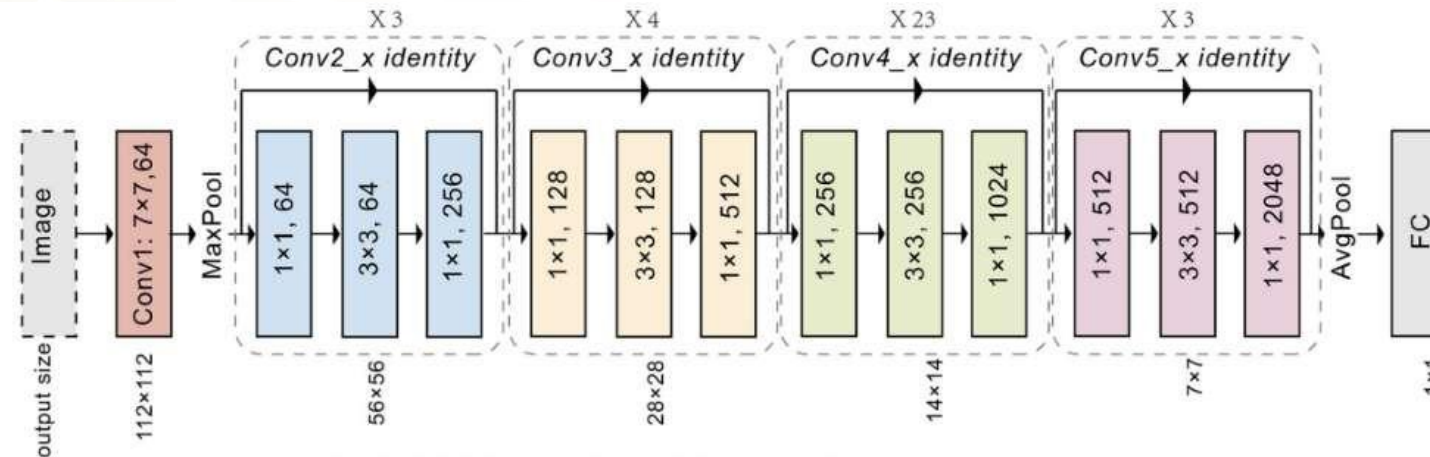


Figura 2. Understanding resnet: A milestone in deep learning and image recognition.
General Architecture Design. Allan, K.

Redes Siamesas

- Se empleará una red convolucional siamesa basada en una *ResNet50* para calcular el valor de similitud.
- Las redes siamesas son arquitecturas diseñadas para aprender una función de similitud entre pares de entradas.

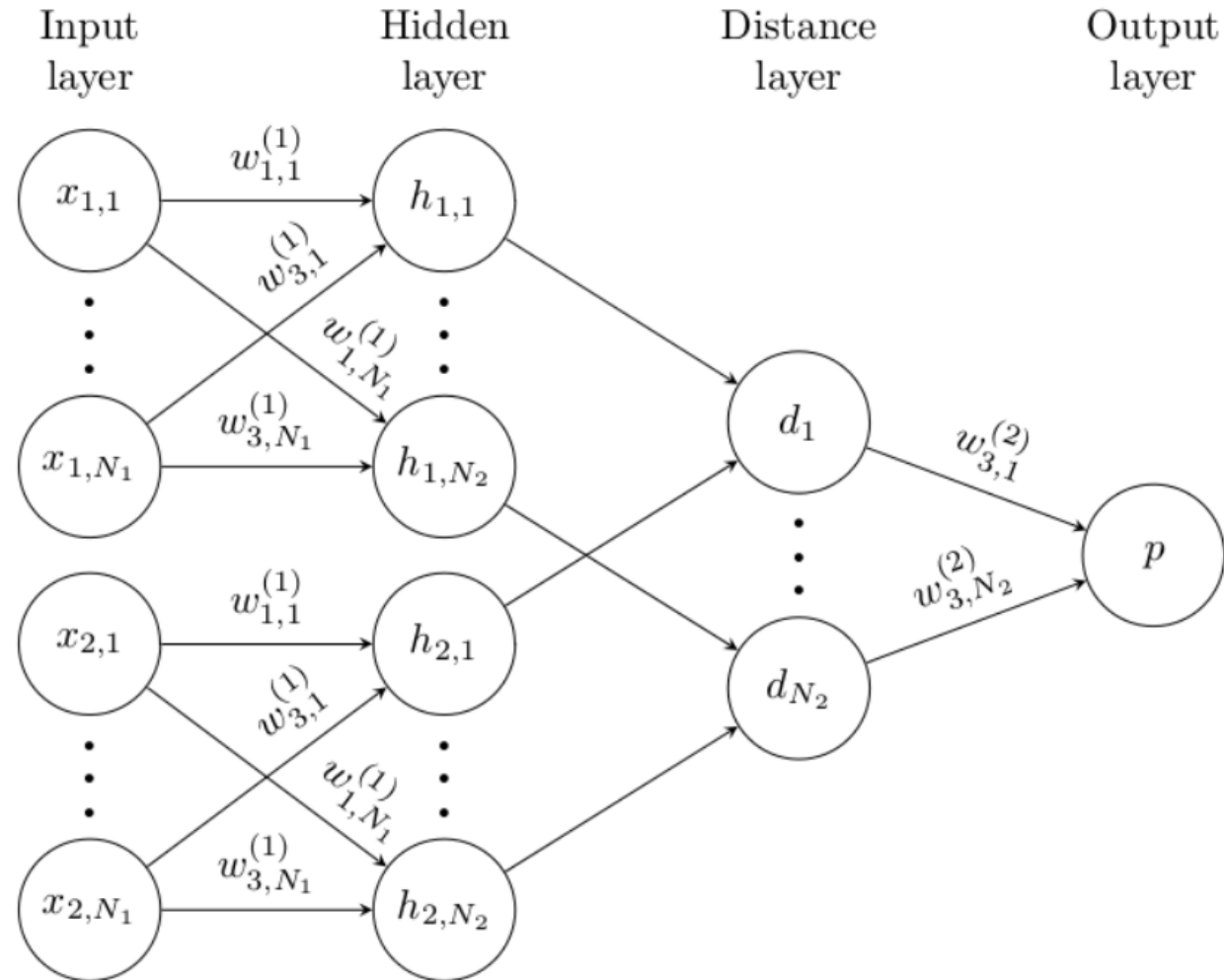


Figura 3. Siamese Neural Networks for One-shot Image Recognition. Red Siamesa simple de 2 capas ocultas. Gregory Koch, et.al.

Stable Diffusion XL

- En 2022, surge *Stable Diffusion*, un modelo de difusión *text-to-image*.
- Surge una versión mejorada: *Stable Diffusion XL (SDXL)*.
- Código abierto.

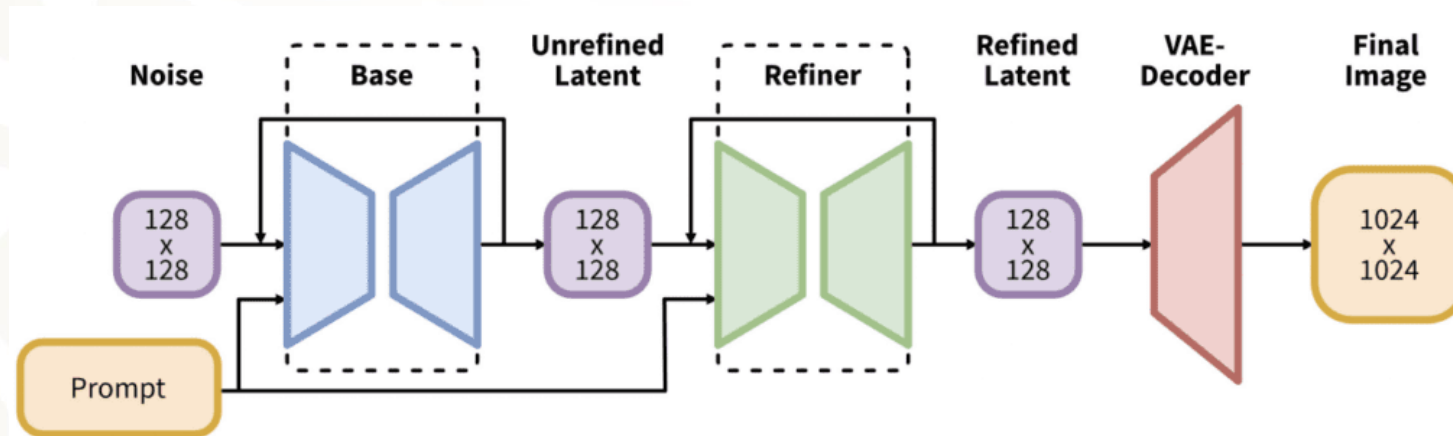


Figura 4. *Sdxl: Improving latent diffusion models for high-resolution image synthesis*. Arquitectura del modelo base y del modelo refinador. Podell, D, et.al.

6. Estado de Arte



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

Trabajos relacionados

Referencia	Año	Descripción
<i>Measuring the Success of Diffusion Models at Imitating Human Artists</i>	2023	Se evalúa la imitación de <i>Stable Diffusion</i> de estilos de 70 artistas, logrando 81 % de precisión y 90 % de similitud con obras originales, analizando implicaciones legales.
<i>StyleAuditor</i>	2024	Se utilizan redes como VGG para auditar la imitación de <i>Stable Diffusion</i> de estilos de 30 artistas, logrando más del 90 % de precisión y baja tasa de falsos positivos.
<i>On Copyright Risks of Text-to-Image Diffusion Models</i>	2024	Se evalúan infracciones de derechos de autor en imágenes generadas por <i>Stable Diffusion</i> , encontrando un 70 % de similitud con contenido protegido, pese a <i>prompts</i> de baja sensibilidad semántica.
<i>ArtSavant</i>	2024	Se analiza si imágenes generadas por IA replican estilos de 372 artistas usando modelos como <i>DeepMatch</i> (89.3 % precisión) y <i>TagMatch</i> , detectando posibles infracciones de derechos de autor con <i>Stable Diffusion</i> .

7. Metodología



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

Investigación preliminar

Revisión de la literatura:

- Técnicas de redes convolucionales siamesas
- Métodos de comparación de imágenes
- Descriptores de imágenes

Creación del conjunto de datos



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"

Obtención del dataset de *Wikiart*

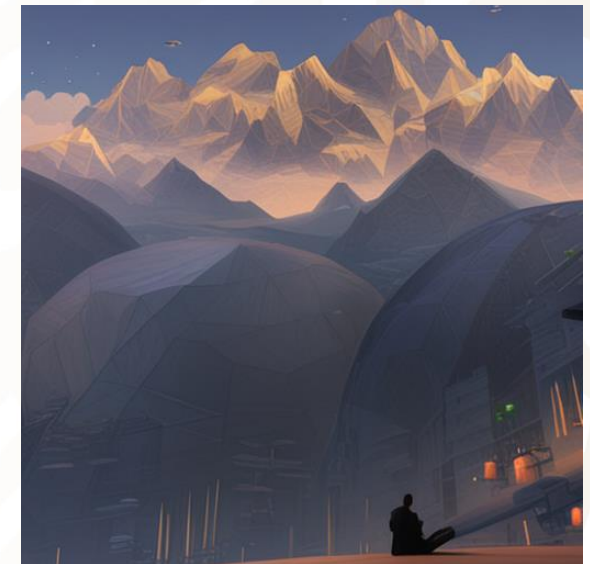


Pintura de Nicholas-roerich.

Generación de imágenes con *Stable Diffusion*



"Add mythical creatures such as dragons and fairies to the scene."
Stable Diffusion



"Reimagine in a futuristic setting, with advanced technology and neon architecture." *Stable Diffusion*.

Selección de descriptores de imágenes



Descriptores de color

Descriptores de textura

Descriptores de forma

Descriptores basados en bordes

Descriptores semánticos

[0.123, 0.422 , ...]

Generación de valores de similitud



→ [0.65 , 0.12 , ...]



→ [0.23 , 0.92 , ...]

Cálculo de valor de similitud

Fine-tunning del modelo ResNet50.

Especializar un modelo en la clasificación de géneros artísticos, usando *ResNet50* y las imágenes de *WikiArt*.

Subtareas

- División del conjunto de datos
- Ajuste de hiperparámetros
- Guardado de los pesos

Entrenamiento de la red convolucional siamesa.

Implementación de la red con red siamesa utilizando las imágenes originales y generadas por *Stable Diffusion*

Subtareas

- Entrenamiento con los pesos preentrenados de *ResNet50*
- Configuración para producir valor que mida la similitud entre las imágenes

Evaluación

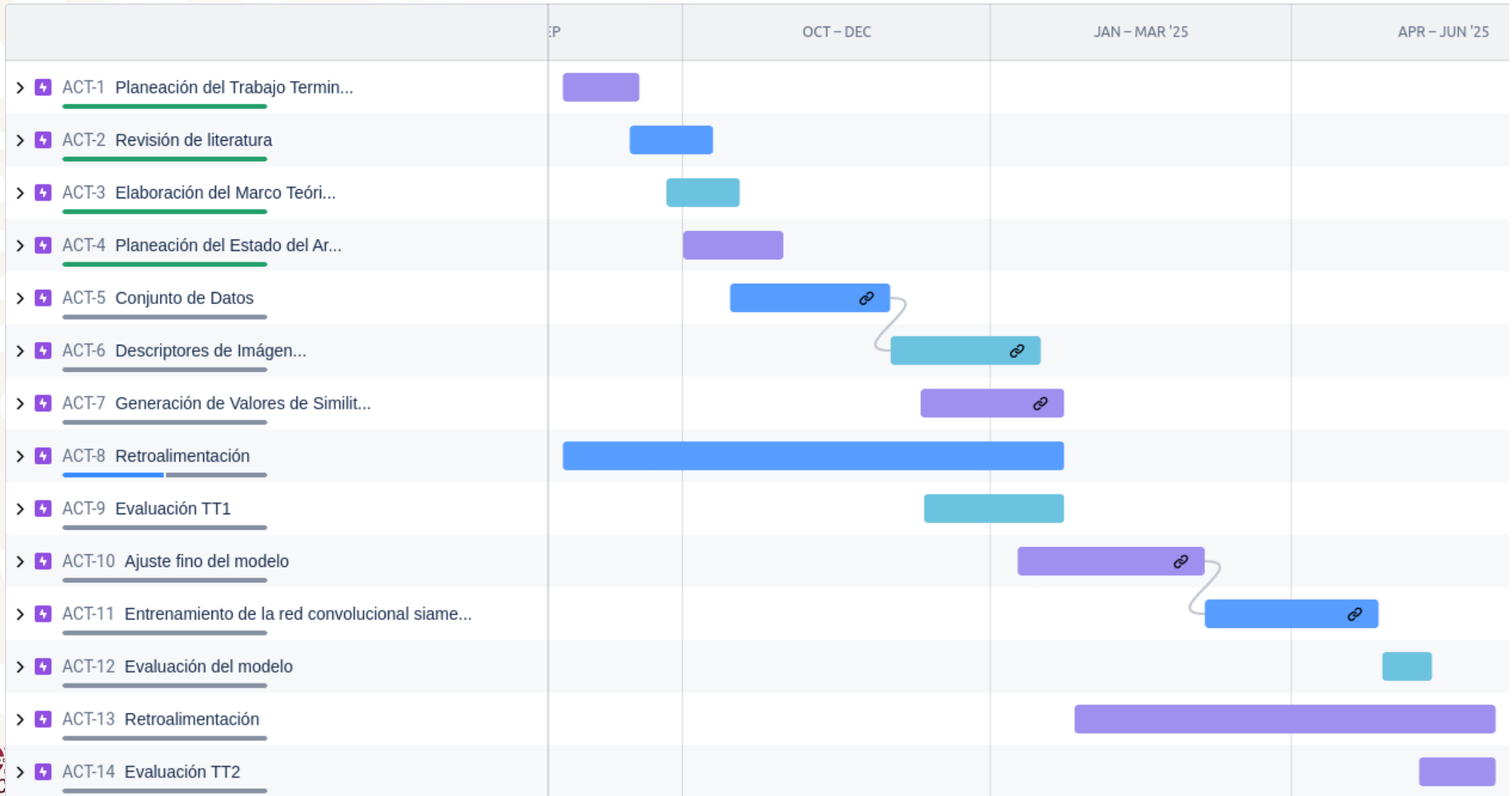
- Evaluación del modelo usando funciones de evaluación como: MSE y MAE
- Comparación con los valores de similitud preliminares generados con los descriptores



8. Cronograma



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



9. Referencias

O'Shea, K. and Nash, R. (2015). An Introduction to Convolutional Neural Networks. Article arXiv:1511.08458v2. <https://doi.org/10.48550/arXiv.1511.08458>

Alshalali, T., & Josyula, D. (2018). Fine-tuning of pre-trained deep learning models with extreme learning machine. 2018 International Conference on Computational Science and Computational Intelligence(CSCI), 469–473. <https://doi.org/10.1109/CSCI46756.2018.00096>

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. <https://doi.org/10.1109/CVPR.2016.90>

Koch, G. R. (2015). Siamese neural networks for one-shot image recognition. <https://api.semanticscholar.org/CorpusID:13874643>

Sherly A P, P. A. R. (2024). Siamese augmented network (saugnet) for jpeg steganalysis. <https://doi.org/10.36227/techrxiv.171994747.74344541/v1>

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., & Rombach, R. (2023). Sdxl: Improving latent diffusion models for high-resolution image synthesis. arXiv preprint arXiv:2307.01952.

Zhang, Y., Teoh, T. T., Lim, W. H., Wang, H., & Kawaguchi, K. (2024). On Copyright Risks of Text-to-Image Diffusion Models. National University of Singapore.

Du, L., Zhu, Z., Chen, M., Ji, S., Cheng, P., Chen, J., & Zhang, Z. (2024). WIP: Auditing Artist Style Pirate in Text-to-image Generation Models. CISPA Helmholtz Center for Information Security, Saarbrücken.

Casper, S., Guo, Z., Mogulothu, S., Marinov, Z., Deshpande, C., Yew, R. J., Dai, Z., & Hadfield-Menell, D. (2023). Measuring the Success of Diffusion Models at Imitating Human Artists. International Conference on Machine Learning

Moayeri, M., Basu, S., Balasubramanian, S., Kattakinda, P., Chengini, A., Brauneis, R., & Feizi, S. (2024). Rethinking Artistic Copyright Infringements in the Era of Text-to-Image Generative Models. University of Maryland, Computer Science Department.



Gracias



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"