

# **Generación de Imágenes Realistas de Aves Utilizando Modelos Generativos de Difusión**



**Javier Betancurth Cano**

Profesor:

**Raúl Ramos Pollán**

Línea de investigación:

**Informática - Deep Learning**

Grupo de investigación:

**Intelligent Information Systems Lab (In2Lab)**

Universidad de Antioquia (UDEA)  
Facultad de Ingeniería, Departamento de Sistemas  
Fundamentos de Deep Learning  
Medellín, Colombia  
Octubre 2023

## Contexto de Aplicación

El proyecto se enmarca en el campo del aprendizaje profundo y tiene aplicaciones interdisciplinarias en áreas como la ornitología, la conservación de la biodiversidad, el ecoturismo y la monitorización ambiental. Se enfoca en la generación de imágenes realistas de especies de aves utilizando modelos generativos de difusión (Stable Diffusion), con el objetivo de proporcionar datos de entrenamiento de alta calidad para modelos de clasificación de aves. Esto puede ser especialmente útil en la identificación automática de aves en imágenes capturadas en la naturaleza o en la monitorización de poblaciones de aves en entornos naturales (p. ej. Cornell Lab of Ornithology Macaulay Library).

## Objetivo de Machine Learning

El objetivo principal de este proyecto es desarrollar un modelo generativo de difusión basado en Stable Diffusion que sea capaz de generar imágenes realistas de especies de aves a partir del dataset "BIRDS 525 SPECIES - IMAGE CLASSIFICATION" de Kaggle. Además, se busca explorar la utilidad de estas imágenes generadas en tareas posteriores de clasificación de aves.

## Dataset

- Tipo de Datos: El dataset utilizado es una colección de imágenes de aves de 525 especies diferentes. Todas las imágenes son en formato JPG, en color (RGB), y tienen dimensiones de 224x224 píxeles.
- Tamaño del Dataset: El dataset consta de 84,635 imágenes para entrenamiento, 2,625 imágenes para pruebas y 2,625 imágenes para validación. Cada conjunto de datos se divide en 525 subdirectorios, uno por cada especie de ave (El tamaño en disco es de 2GB).
- Distribución de Clases: El conjunto de entrenamiento no está equilibrado en términos de número de archivos por especie, cada especie tiene al menos 130 imágenes de entrenamiento. Además, aproximadamente el 80% de las imágenes son de machos y el 20% de hembras.

## Métricas de Desempeño

- Métricas de Machine Learning: Se evaluará la calidad de las imágenes generadas utilizando métricas específicas para la generación de imágenes, como la puntuación de similitud estructural (SSIM) y la distancia de Frechet Inception (FID). Estas métricas se utilizarán para medir la similitud entre las imágenes generadas por el modelo y las reales del dataset.
- Métricas de Negocio: En términos de aplicaciones prácticas, el éxito se medirá mediante la capacidad del modelo para generar imágenes realistas y útiles para la generación de imágenes sintéticas que puedan ser utilizadas en tareas de reconocimiento y clasificación de especies de aves (p. ej. Merlin Bird ID app de Cornell Lab of Ornithology).

## Resultados Previos

Este proyecto se basa en investigaciones previas en el campo de la generación de imágenes utilizando modelos generativos de difusión, como el modelo basado en Stable Diffusion. Además, se toma como referencia el dataset "BIRDS 525 SPECIES - IMAGE CLASSIFICATION" de Kaggle, que proporciona una amplia variedad de imágenes de aves para este proyecto.

Resultados previos en la generación de imágenes utilizando técnicas similares han demostrado la capacidad de los modelos generativos para crear imágenes realistas. Sin embargo, este proyecto busca mejorar aún más la calidad y utilidad de las imágenes generadas y explorar su aplicación en la clasificación de aves (Generación de imágenes sintéticas).

## Referencias

### Modelado generativo:

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015, June). Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning* (pp. 2256-2265). PMLR.

Song, Y., & Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33, 6840-6851.

Nichol, A. Q., & Dhariwal, P. (2021, July). Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning* (pp. 8162-8171). PMLR.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10684-10695).

### Other techniques:

Van Den Oord, A., & Vinyals, O. (2017). Neural discrete representation learning. *Advances in neural information processing systems*, 30.

Razavi, A., Van den Oord, A., & Vinyals, O. (2019). Generating diverse high-fidelity images with vq-vae-2. *Advances in neural information processing systems*, 32.

Gregor, K., Papamakarios, G., Besse, F., Buesing, L., & Weber, T. (2018). Temporal difference variational auto-encoder. arXiv preprint arXiv:1806.03107.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748-8763). PMLR.

Ronneberger, O., Fischer, P., & Brox, T. (2022). Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015 Conference Proceedings*.

Sauer, A., Karras, T., Laine, S., Geiger, A., & Aila, T. (2023). Stylegan-t: Unlocking the power of gans for fast large-scale text-to-image synthesis. arXiv preprint arXiv:2301.09515.

Machine learning and Deep learning frameworks:

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Kaggle Dataset:

Gerry. (s.f.). BIRDS 525 SPECIES- IMAGE CLASSIFICATION. Kaggle. URL: <https://www.kaggle.com/datasets/gpiosenska/100-bird-species/data>