

# Modelos Generativos basados en Mecanismos de Difusión

Presentado por: Alejandro Pequeño Lizcano  
Dirigido por: Guillermo Iglesias Hernández



23 JULIO 2024



POLITÉCNICA

UNIVERSIDAD  
POLÍTÉCNICA  
DE MADRID



Universidad  
Politécnica  
de Madrid  
ETSI SISTEMAS  
INFORMÁTICOS

DEFENSA TFG



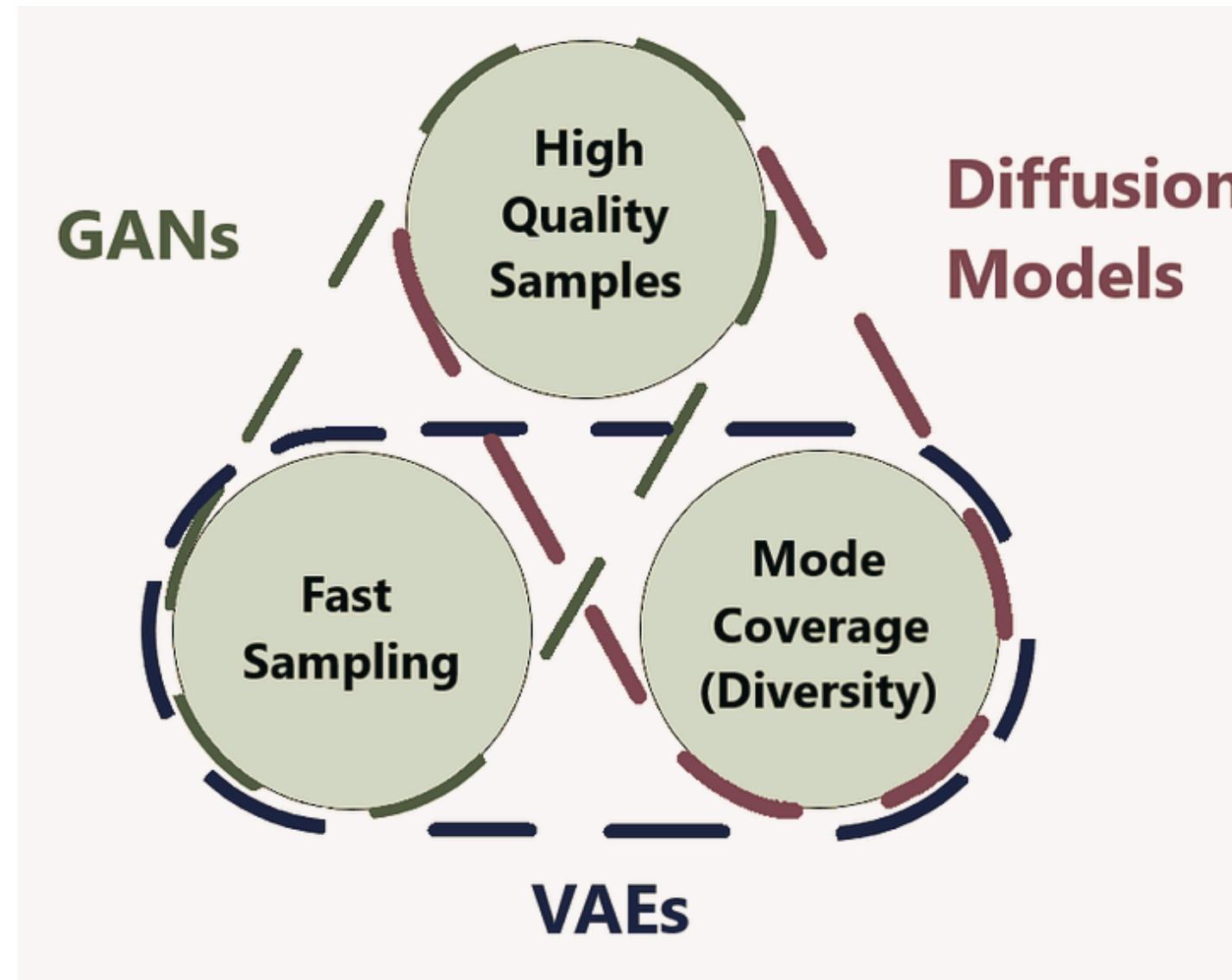
# Abstract

Este proyecto desarrolla un modelo generativo de difusión condicionado (DDPM) para crear imágenes de Pokémon según su tipo (fuego, dragón, planta, etc.), utilizando TensorFlow y Keras para su implementación desde cero.

Se detalla el procesamiento de datos, las metodologías y algoritmos aplicados, y la evolución de los modelos de difusión culminando con la generación de nuevas y únicas imágenes Pokémon. hasta la generación con éxito de imágenes de Pokémon completamente nuevos.

# Introducción

P A R T E   0 1



# ¿Qué es un modelo de difusión?

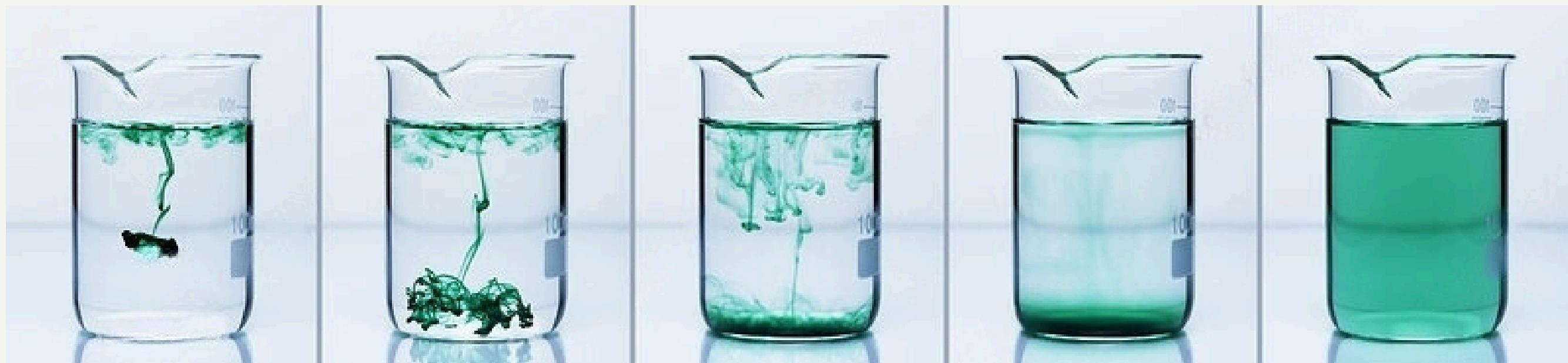
## DENOISING DIFFUSION MODELS

Los modelos de difusión son un tipo de modelo generativo basado en la **termodinámica** cuyo objetivo es aprender la distribución probabilística de un conjunto de datos, y generar nuevas muestras de dicha distribución a través de ruido (**variable latente**).

[i1]

# Termodinámica

Los modelos de difusión tienen su base en la termodinámica, más en concreto en la **entropía**. Pues la difusión no es más que el proceso de mezcla de moléculas en un fluido de manera aleatoria.

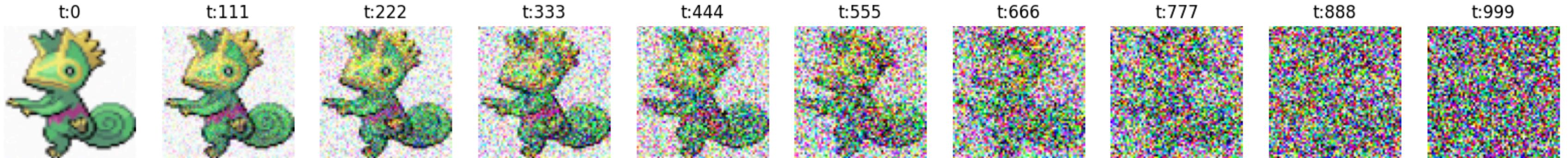


[i2]

# Proceso de Difusión

SE DIVIDE EN DOS PROCESOS:

Forward Diffusion



Inverse Diffusion

# Motivación & Objetivos

P A R T E   0 2

# Motivación



## FORMACIÓN ACADÉMICA

La carrera en Ciencia de Datos e Inteligencia Artificial ha desarrollado una gran pasión por este campo, especialmente en los modelos generativos.



## INNOVACIÓN & DESAFÍO

La dificultad y novedad de los modelos de difusión, siendo el actual estado del arte (SOTA), los hacen especialmente atractivos y desafiantes.



## MODELOS GENERATIVOS

El interés en los modelos generativos surgió en la asignatura de Métodos Generativos, ampliando conocimientos y despertando una verdadera pasión por estos modelos.

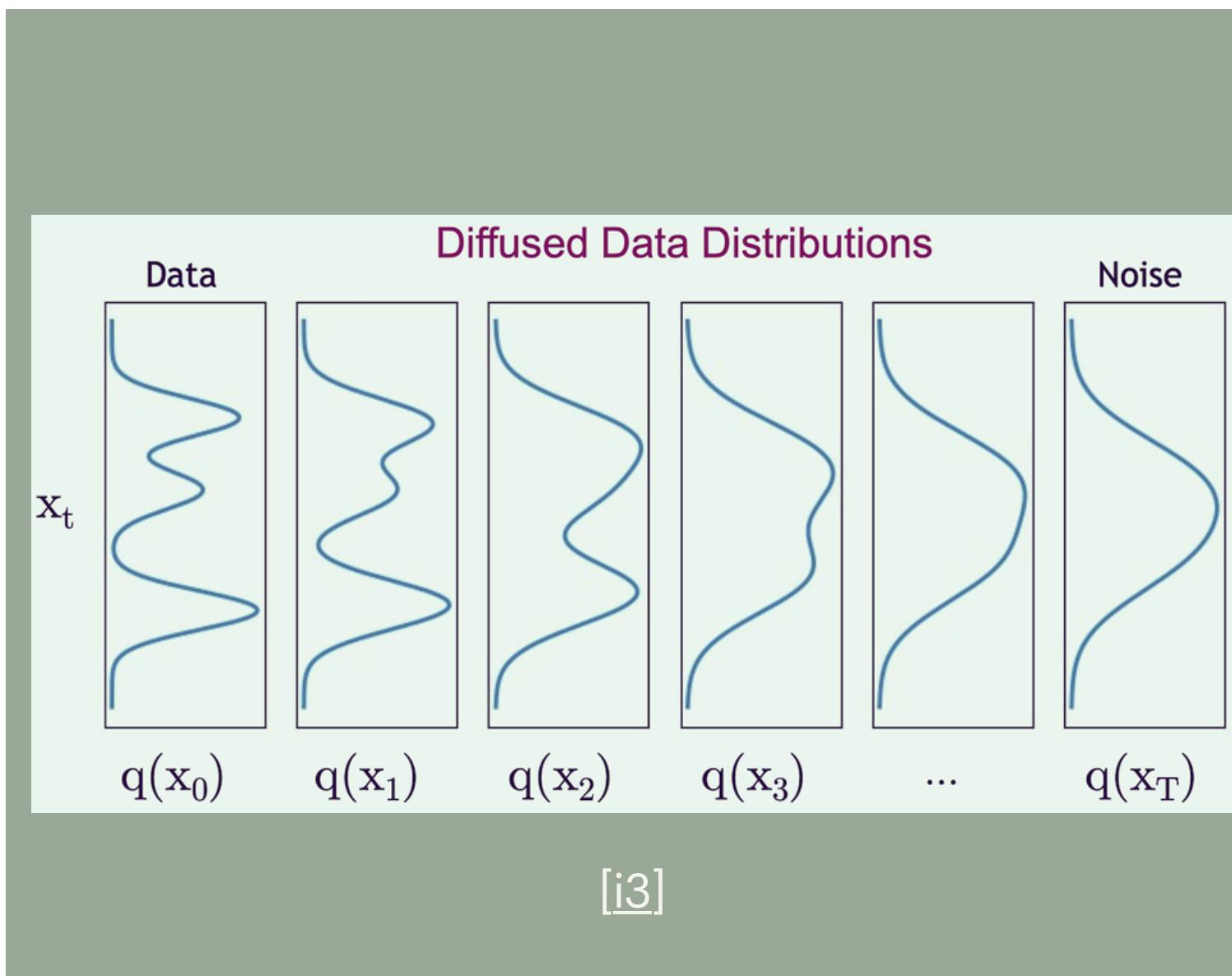


## NOSTALGIA ESPECIAL

Trabajar en este proyecto genera un sentimiento de nostalgia, combinando la pasión de antaño por los Pokémon con la actual desarrollando habilidades académicas y técnicas.

# Objetivo Teórico

Aprender y **entender** los modelos de difusión con una base matemática para poder entender el porqué del funcionamiento de dichos modelos.



## Algorithm 2 Sampling

---

```

1:  $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
2: for  $t = T, \dots, 1$  do
3:    $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  if  $t > 1$ , else  $\mathbf{z} = \mathbf{0}$ 
4:    $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$ 
5: end for
6: return  $\mathbf{x}_0$ 

```

---

[2]

# Objetivo Práctico

Realizar una fiel **implementación** de un modelo de difusión DDPM [2], a través de los algoritmos y prácticas realizadas en los artículos de investigación correspondientes.

# Metodología

P A R T E   0 3

## METODOLOGÍA

# Implementación

## MÉTODOS DE DIFUSIÓN Y ALGORITMOS DDPM

**Deep Unsupervised Learning using Nonequilibrium Thermodynamics**

---

Jascha Sohl-Dickstein  
Stanford University

Eric A. Weiss  
University of California, Berkeley

Niru Maheswaranathan  
Stanford University

Surya Ganguli  
Stanford University

**Abstract**

A central problem in machine learning involves modeling complex data-sets using highly flexible families of probability distributions in which learning, sampling, inference, and evaluation are still analytically or computationally tractable. Here, we develop an approach that simultaneously achieves both flexibility and tractability. The essential idea, inspired by non-equilibrium statistical physics, is to systematically and slowly destroy structure in a data distribution through an iterative forward diffusion process. We then learn a reverse diffusion process that restores structure in data, yielding a highly flexible and tractable generative model of the data. This approach allows us to rapidly learn, sample from, and evaluate probabilities in deep generative models with thousands of layers or time steps, as well as to compute conditional and posterior probabilities under the learned model. We additionally release an open source reference implementation of the algorithm.

**1. Introduction**

Historically, probabilistic models suffer from a tradeoff between two conflicting objectives: *tractability* and *flexibility*. Models that are *tractable* can be analytically evaluated and easily fit to data (e.g. a Gaussian or Laplace). However,

arXiv:1503.03585v8 [cs.LG] 18 Nov 2015

**Denoising Diffusion Probabilistic Models**

---

Jonathan Ho  
UC Berkeley  
[jonathanho@berkeley.edu](mailto:jonathanho@berkeley.edu)

Ajay Jain  
UC Berkeley  
[ajayj@berkeley.edu](mailto:ajayj@berkeley.edu)

Pieter Abbeel  
UC Berkeley  
[pabbeel@cs.berkeley.edu](mailto:pabbeel@cs.berkeley.edu)

**Abstract**

We present high quality image synthesis results using diffusion probabilistic models, a class of latent variable models inspired by considerations from nonequilibrium thermodynamics. Our best results are obtained by training on a weighted variational bound designed according to a novel connection between diffusion probabilistic models and variational score matching with Langevin dynamics. These models naturally admit a progressive lossy decompression scheme that can be interpreted as a generalization of autoregressive decoding. On the unconditional CIFAR10 dataset, we obtain an Inception score of 9.46 and a state-of-the-art FID score of 3.17. On 256x256 LSUN, we obtain sample quality similar to ProgressiveGAN. Our implementation is available at <https://github.com/jonathanho/diffusion>.

**1. Introduction**

Deep generative models of all kinds have recently exhibited high quality samples in a wide variety of data modalities. Generative adversarial networks (GANs), autoregressive models, flows, and variational autoencoders (VAEs) have synthesized striking image and audio samples [14, 27, 3, 58, 38, 25, 10, 32, 44, 57, 26, 33, 45], and there have been remarkable advances in energy-based modeling and score matching that have produced images comparable to those of GANs [11, 55].

Figure 1: Generated samples on CelebA-HQ 256 x 256 (left) and unconditional CIFAR10 (right).

arXiv:2006.11239v2 [cs.LG] 16 Dec 2020

**Improved Denoising Diffusion Probabilistic Models**

---

Alex Nichol \*<sup>1</sup> Prafulla Dhariwal \*<sup>1</sup>

**Abstract**

Denoising diffusion probabilistic models (DDPMs) are a class of generative models which have recently been shown to produce excellent samples. We show that with a few simple modifications, DDPMs can also achieve competitive log-likelihoods while maintaining high sample quality. Additionally, we find that learning variances of the reverse diffusion process allows sampling with an order of magnitude fewer forward passes with a negligible difference in sample quality, which is important for the practical deployment of these models. We additionally use precision and recall to compare how well DDPMs and GANs cover the target distribution. Finally, we show that the sample quality and likelihood of these models scale smoothly with model capacity and training compute, making them easily scalable. We release our code at <https://github.com/openai/improved-diffusion>.

**1. Introduction**

Sohl-Dickstein et al. (2015) introduced diffusion probabilistic models, a class of generative models which match a data distribution by learning to reverse a gradual, multi-step denoising process. More recently, Ho et al. (2020) showed an equivalence between denoising diffusion probabilistic models (DDPM) and score-based generative models (Song & Ermon, 2019; 2020), which learn a gradient of the log-density of the data distribution using denoising score matching (Hyvärinen, 2005). It has recently been shown that this class of models can produce high-quality images (Ho et al., 2020; Song & Ermon, 2020; Jolicoeur-Martineau et al., 2020) and audio (Chen et al., 2020; Kong et al., 2020), but it has yet to be shown that DDPMs can achieve log-likelihoods competitive with other likelihood-based models such as autoregressive models (van den Oord et al., 2016c) and VAEs (Kingma & Welling, 2013). This raises various questions, such as whether DDPMs are capable of capturing all the modes of a distribution. Furthermore, while Ho et al.

\*Equal contribution. <sup>1</sup>OpenAI, San Francisco, USA. Correspondence to: <[alex@openai.com](mailto:alex@openai.com)>, <[prafulla@openai.com](mailto:prafulla@openai.com)>.

(2020) showed extremely good results on the CIFAR-10 (Krizhevsky, 2009) and LSUN (Yu et al., 2015) datasets, it is unclear how well DDPMs scale to datasets with higher diversity such as ImageNet. Finally, while Chen et al. (2020b) found that DDPMs can efficiently generate audio using a small number of sampling steps, it has yet to be shown that the same is true for images.

In this paper, we show that DDPMs can achieve log-likelihoods competitive with other likelihood-based models, even on high-diversity datasets like ImageNet. To more tightly optimise the variational lower-bound (VLB), we learn the reverse process variances using a simple reparameterization and a hybrid learning objective that combines the VLB with the simplified objective from Ho et al. (2020).

We find surprisingly that, with our hybrid objective, our models obtain better log-likelihoods than those obtained by optimizing the log-likelihood directly, and discover that the latter objective has much more gradient noise during training. We show that a simple importance sampling technique reduces this noise and allows us to achieve better log-likelihoods than with the hybrid objective.

After incorporating learned variances into our model, we surprisingly discovered that we could sample in fewer steps from our models with very little change in sample quality. While DDPM (Ho et al., 2020) requires hundreds of forward passes to produce good samples, we can achieve good samples with as few as 50 forward passes, thus speeding up sampling for use in practical applications. In parallel to our work, Song et al. (2020a) develops a different approach to fast sampling, and we compare against their approach, DDIM, in our experiments.

While likelihood is a good metric to compare against other likelihood-based models, we also wanted to compare the distribution coverage of these models with GANs. We use the improved precision and recall metrics (Kynkäinniemi et al., 2019) and discover that diffusion models achieve much higher recall for similar FID, suggesting that they do indeed cover a much larger portion of the target distribution. Finally, since we expect machine learning models to consume more computational resources in the future, we evaluate the performance of these models as we increase model size and training compute. Similar to (Henighan et al.,

Proceedings of the 32<sup>nd</sup> International Conference on Machine Learning, Lille, France, 2015. JMLR: W&CP volume 37. Copyright 2015 by the author(s).

Deep unsupervised learning using nonequilibrium thermodynamics. [1]

Denoising diffusion probabilistic models (DDPM). [2]

Improved Denoising Diffusion Probabilistic Models. [3]

# Dataset

## DE LOS DATOS AL MODELO



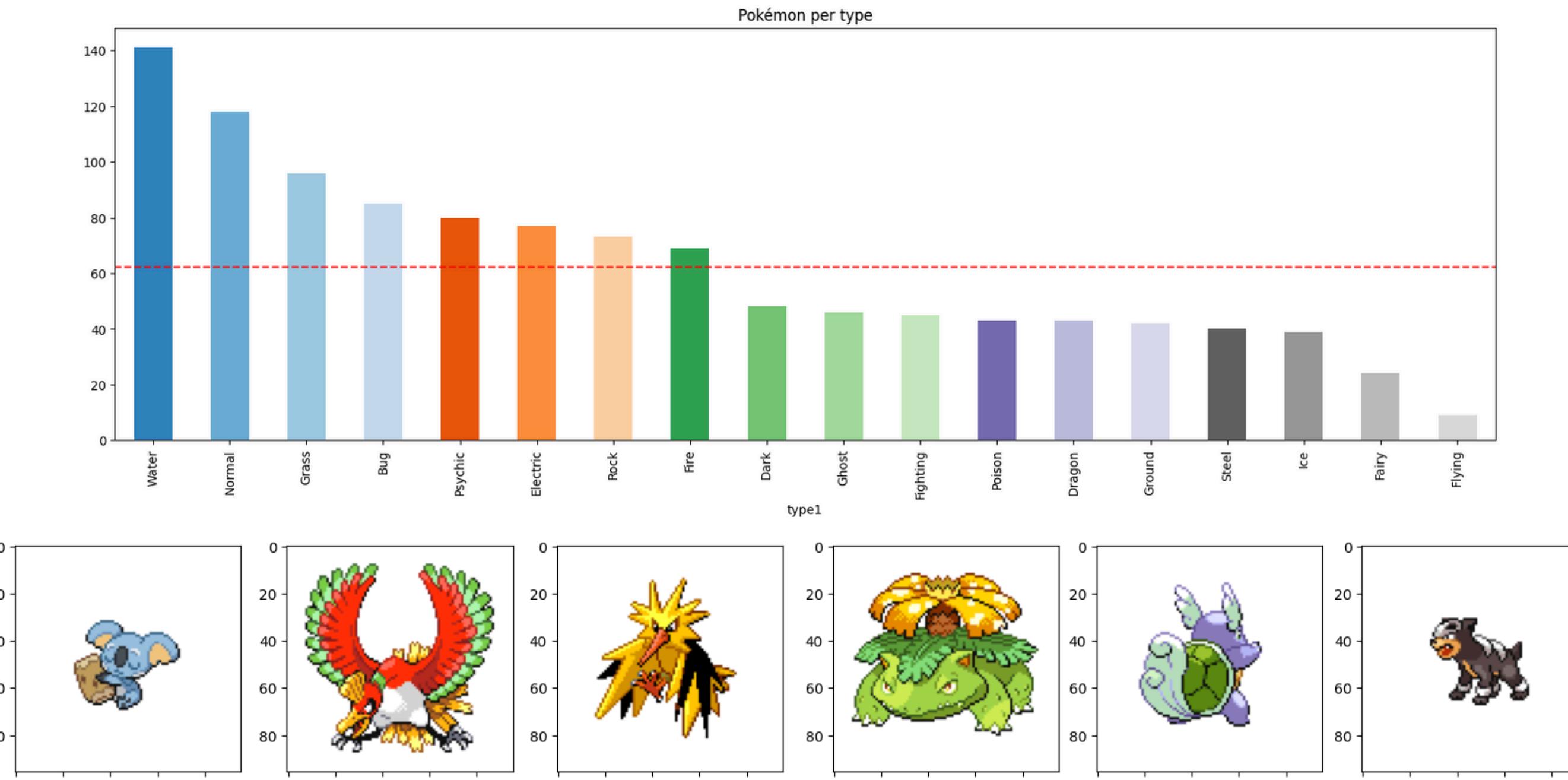
Se analiza el conjunto de datos Pokémon extraído de Kaggle [4]. Todo ello con el objetivo de orquestar la preparación de los datos necesaria para el posterior uso del modelo DDPM.

Atendiendo a las conclusiones del análisis previo, se escogen los datos más representativos y relevantes alineados con los objetivos del proyecto.

Se implementan las soluciones propuestas en el análisis del conjunto de datos para que el modelo sea capaz de procesarlo de manera correcta y eficiente.

# 01. Análisis del dataset

Se estudia la **cantidad** (10.437), **características** del dataset y **distribución** de tipos Pokémon a lo largo este, con el objetivo de obtener información para su posterior procesado y poder sacar conclusiones de valor tras la obtención de resultados.



## 02. Selección de imágenes

Tras el análisis, se establece el criterio de selección de imágenes Pokémon a normales (no shiny) y de frente, con el objetivo de obtener la mayor representatividad de las características de conjunto de datos. Reduciendo el conjunto de **10.437** a **4.086** imágenes Pokémon.



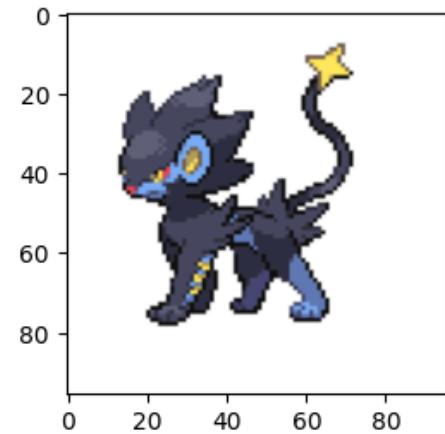
DATASET

## 03. Preprocesado

PASO 01



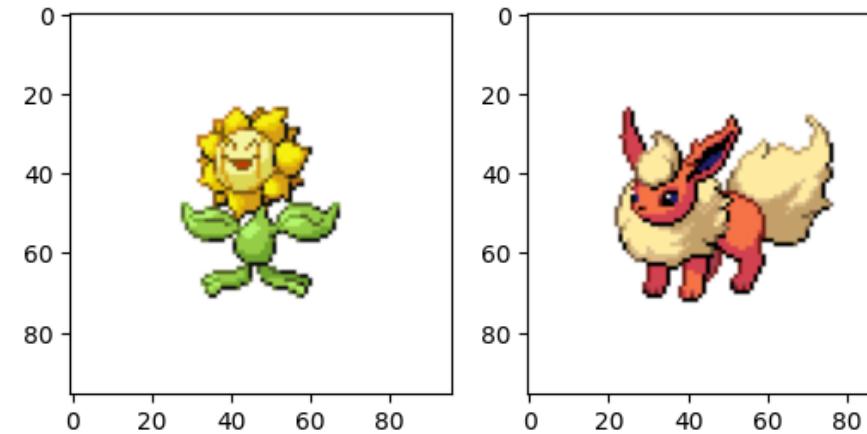
BINARIZACIÓN DE  
ETIQUETAS



PASO 02



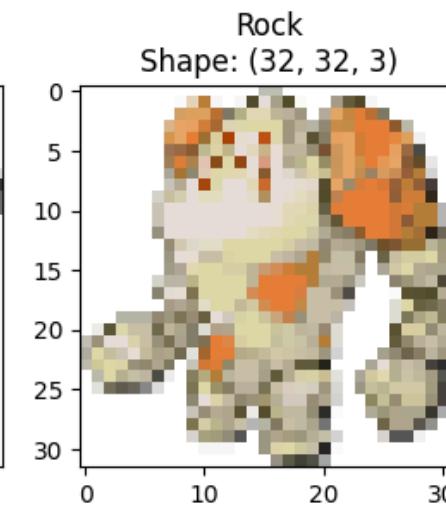
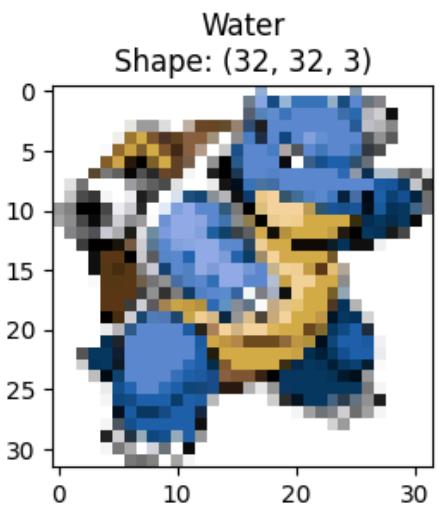
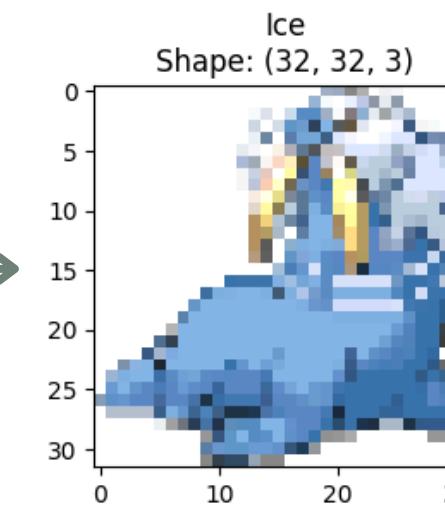
RECORTE DE FONDO



PASO 03



CONVERSIÓN Y ESCALADO  
DE IMÁGENES



PASO 04

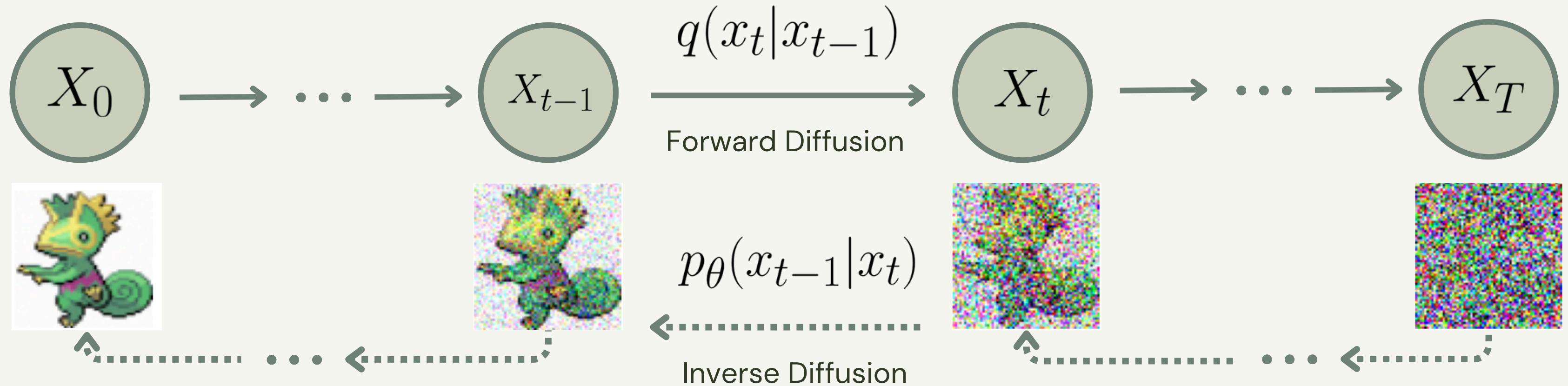


NORMALIZACIÓN DE  
PÍXELES

M E T O D O L O G Í A

# Modelo DDPM

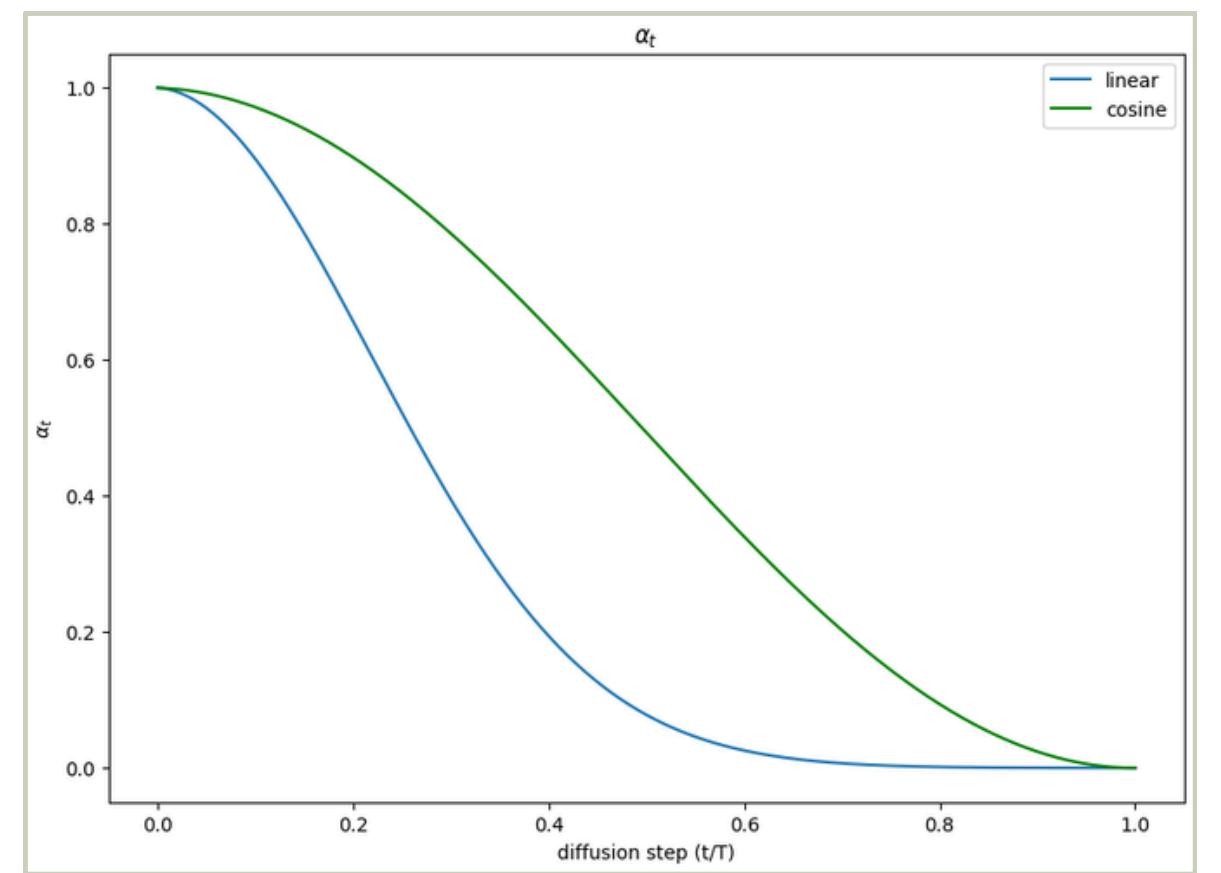
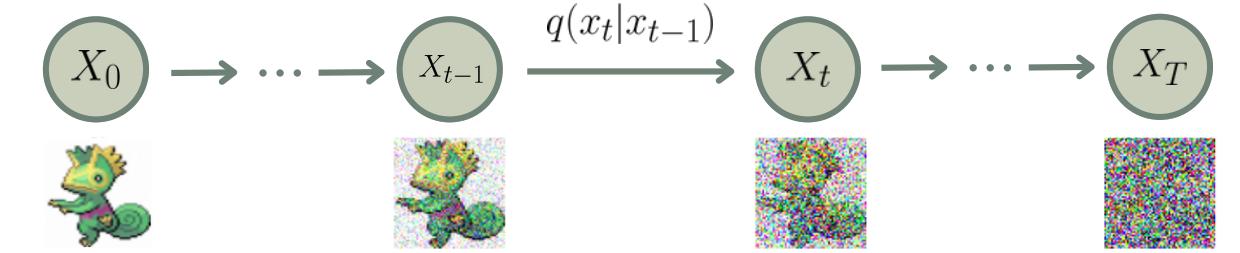
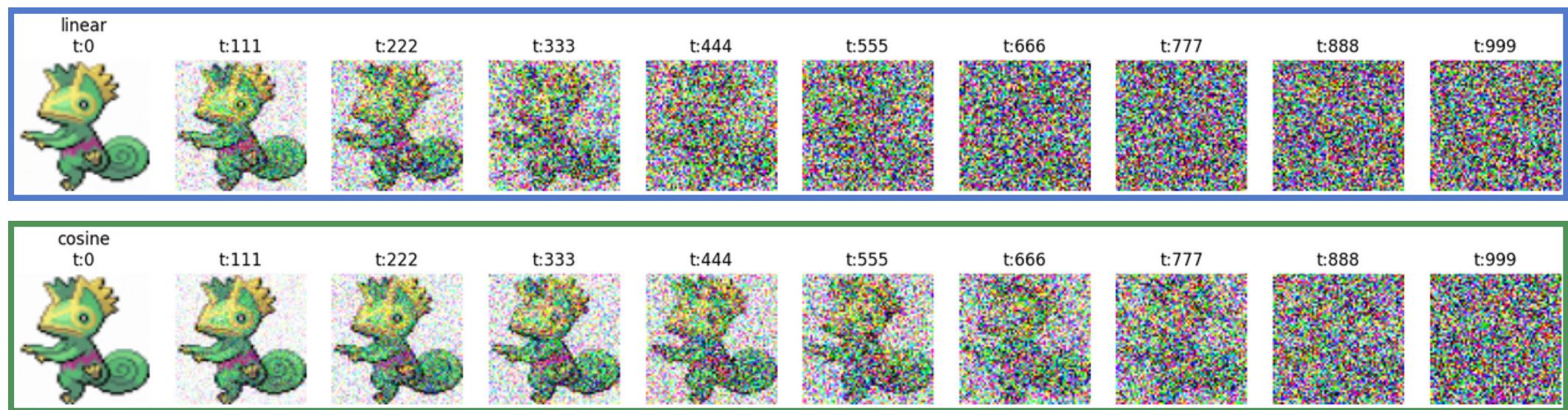
IMPLEMENTACIÓN DEL PROCESO DIFUSIVO



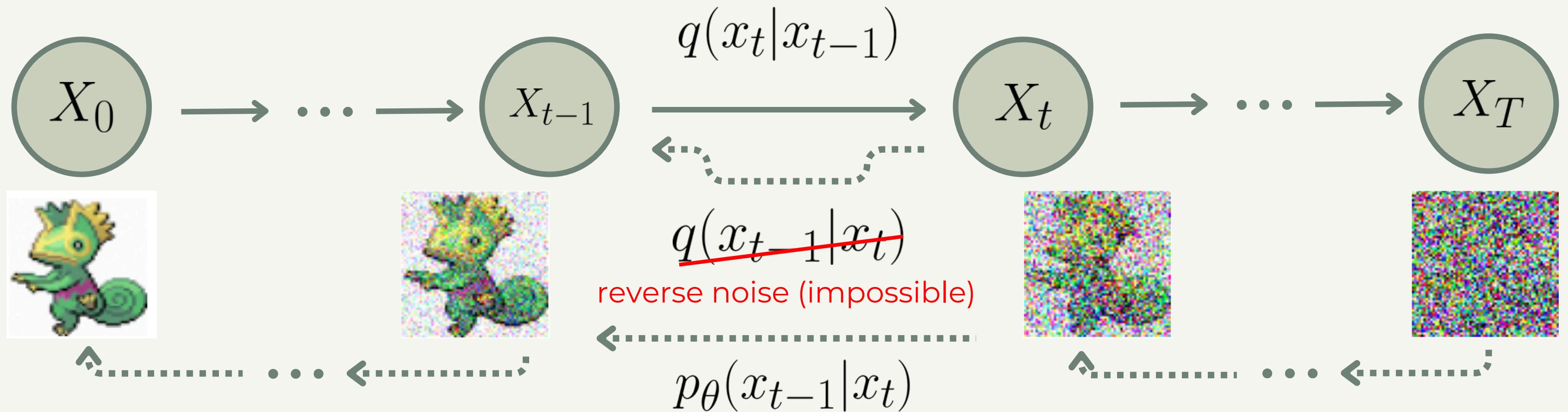
# Forward Diffusion

El proceso de adición de ruido va regulado por un **scheduler** (variance scheduler) que describe la manera en la que se añade ruido a la imagen de entrada dependiendo del instante de tiempo en el que se encuentre.

Para el proyecto se decidió por el uso de un **cosine scheduler** por sus claras ventajas visuales frente al linear a la hora de añadir ruido a la imagen.

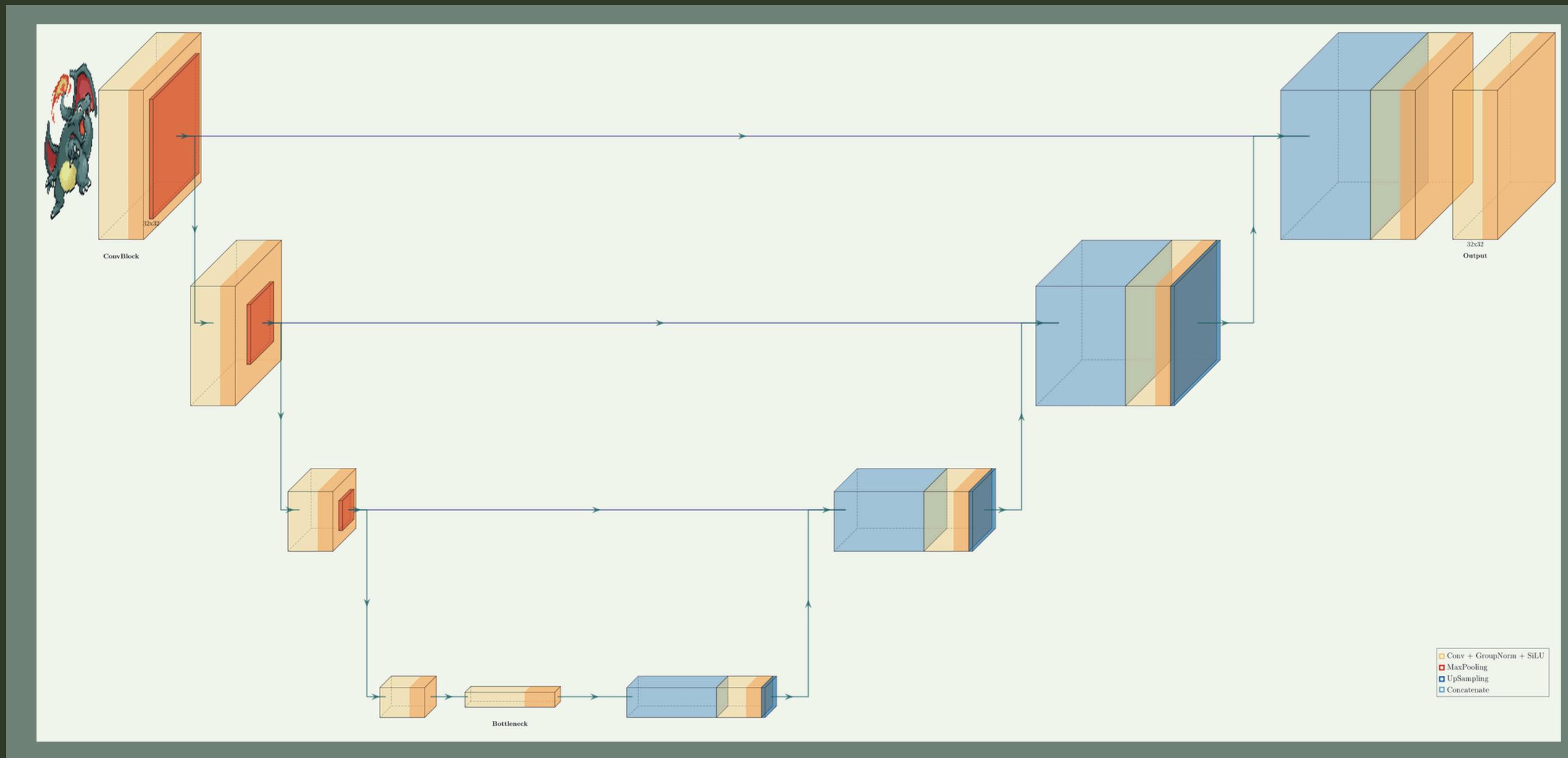


# Inverse Diffusion



# Inverse Diffusion

EL PROCESO CONTRARIO DEBE DE SER ESTIMADO POR UNA RED NEURONAL, LA MÁS USADA PARA LA SÍNTESIS DE IMÁGENES ES LA U-NET.



# Arquitectura U-Net

## DESARROLLO Y EVOLUCIÓN

### Vanilla U-Net



#### LA ARQUITECTURA MÁS SENCILLA

- Bloques de procesamiento básicos
- Etapas convolucionales ligeras
- Función de activación **ReLU**
- Normalización de capa (**Layer Normalization**)

### Sinusoidal embedding & Attention U-Net



#### MÁS COMPLEJIDAD Y FUNCIONALIDAD

- Mayor procesamiento
- Función de activación **SiLu**
- Normalización en grupo (**Group Normalization**)
- Embedding sinusoidal y Módulos de atención

### U-Net final



#### ESTABILIZANDO EL ENTRENAMIENTO

- Ajuste de hiperparámetros (reducción del número de neuronas y learning rate)
- Mayor normalización de grupo
- **Dropout**
- Modelo **EMA**

# Algoritmos DDPM

IMPLEMENTACIÓN Y DESARROLLO DE LOS ALGORITMOS DDPM

## ALGORITMO 01

Se hace el forward diffusion a las imágenes de entrada en cada  $t$ . Se predice el ruido generado mediante el modelo, y se ajusta dicho modelo minimizando la diferencia entre el ruido predicho y el real (MSE).

---

### Algorithm 1 Training

---

```

1: repeat
2:    $\mathbf{x}_0 \sim q(\mathbf{x}_0)$ 
3:    $t \sim \text{Uniform}(\{1, \dots, T\})$ 
4:    $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
5:   Take gradient descent step on
       $\nabla_{\theta} \|\epsilon - \epsilon_{\theta}(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t)\|^2$ 
6: until converged

```

---

## ALGORITMO 02

Inicia con ruido puro gaussiano  $y$ , en cada paso inverso, utiliza el modelo para estimar y restar el ruido del paso anterior, refinando gradualmente esta estimación para reconstruir la imagen original desde el ruido.

---

### Algorithm 2 Sampling

---

```

1:  $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
2: for  $t = T, \dots, 1$  do
3:    $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  if  $t > 1$ , else  $\mathbf{z} = \mathbf{0}$ 
4:    $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\bar{\alpha}_t}} \left( \mathbf{x}_t - \frac{1 - \bar{\alpha}_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_{\theta}(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$ 
5: end for
6: return  $\mathbf{x}_0$ 

```

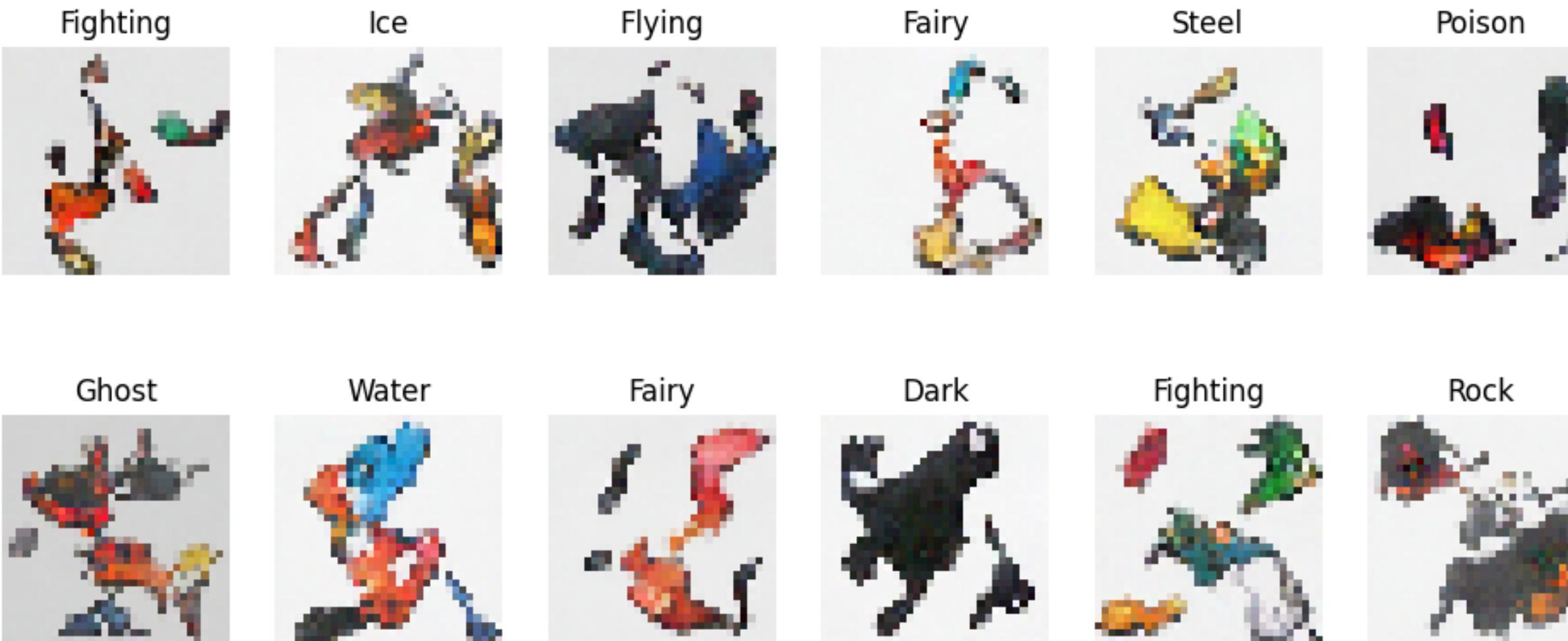
---

# Resultados

P A R T E   0 4

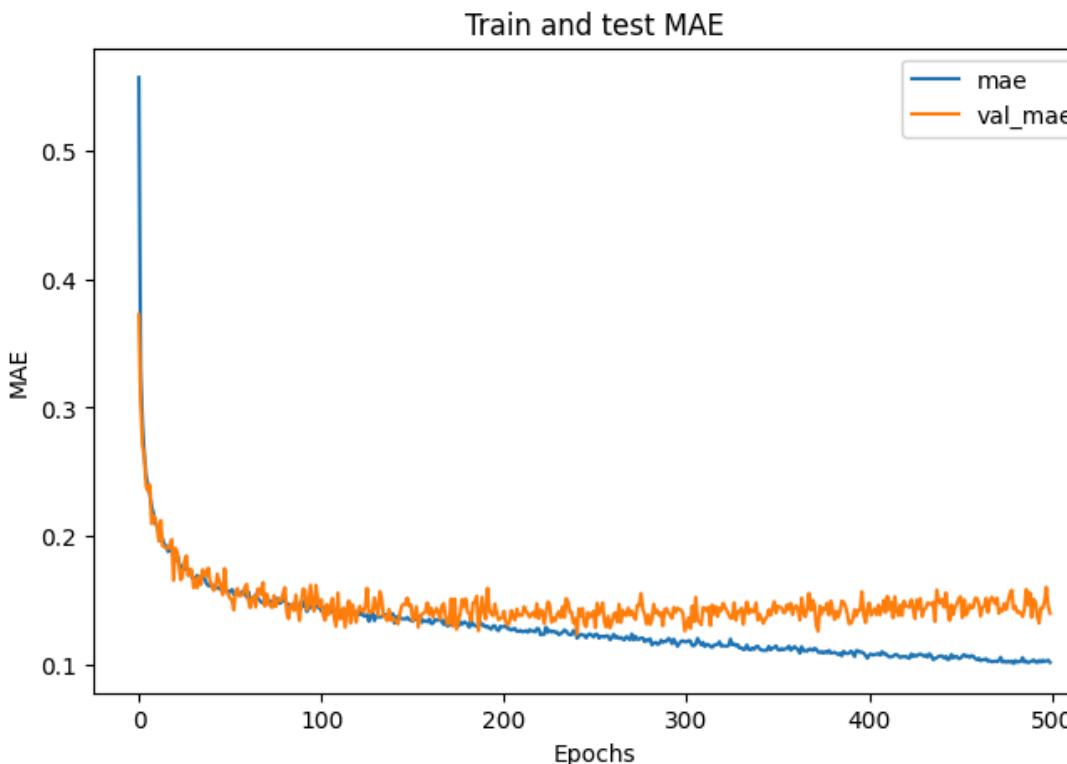
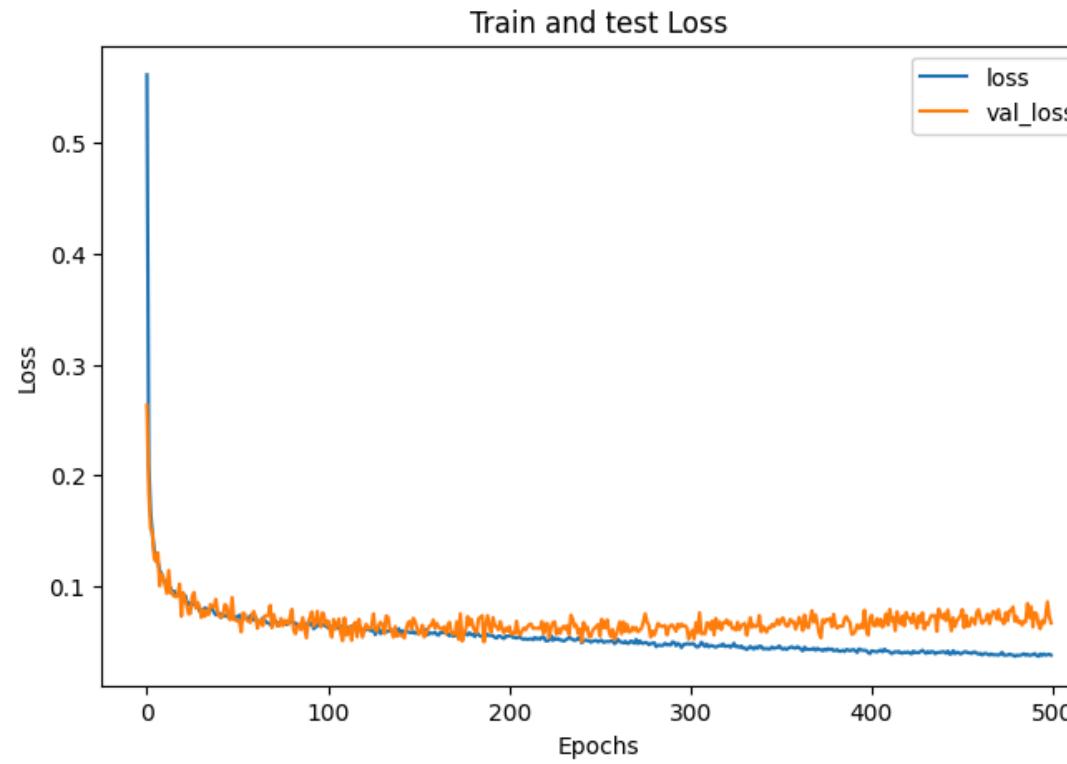
# Resultados de arquitecturas previas

VANILLA U-NET



# Resultados de arquitecturas previas

## SINUSOIDAL EMBEDDING & ATENTION U-NET



EPOCHS: 500

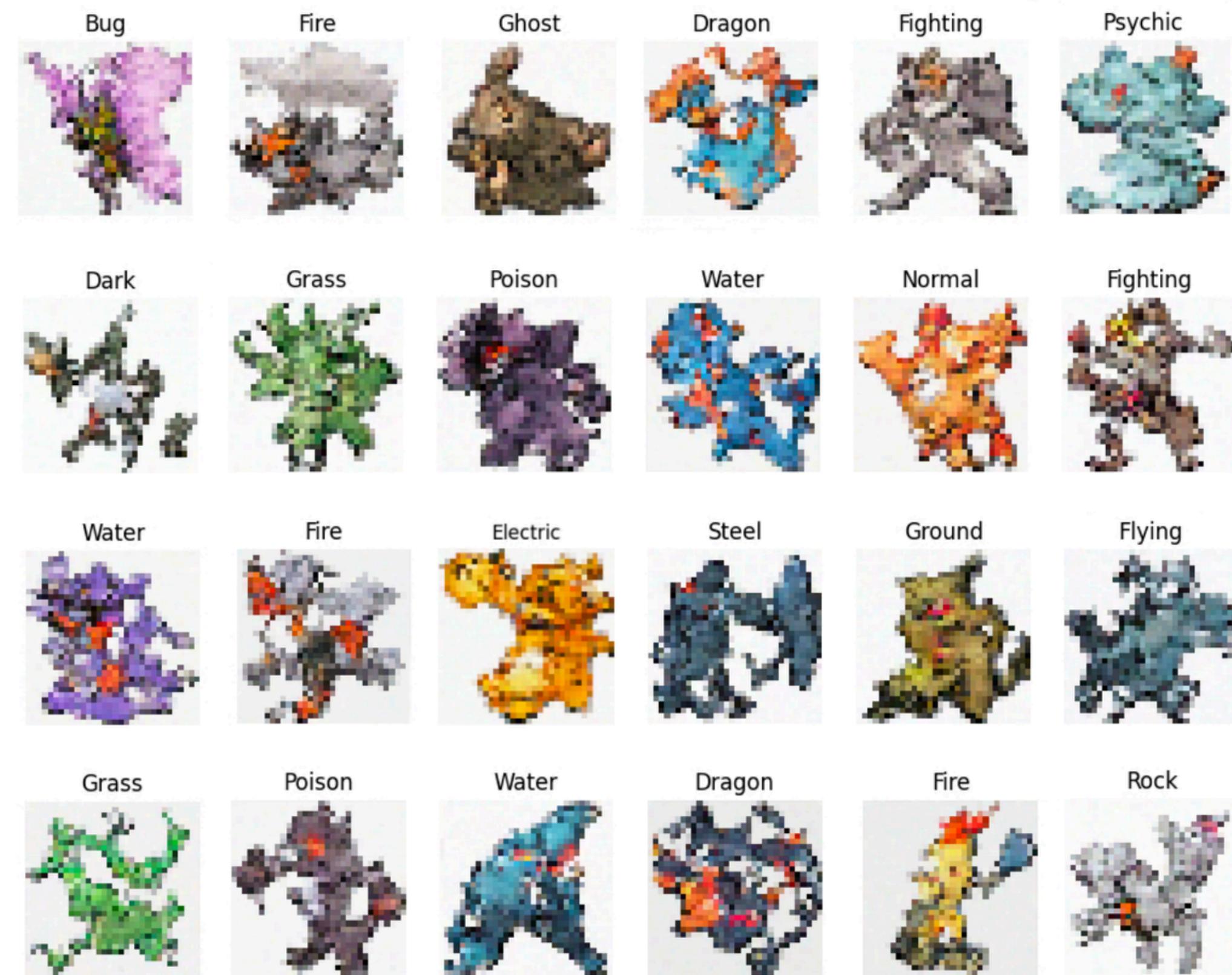
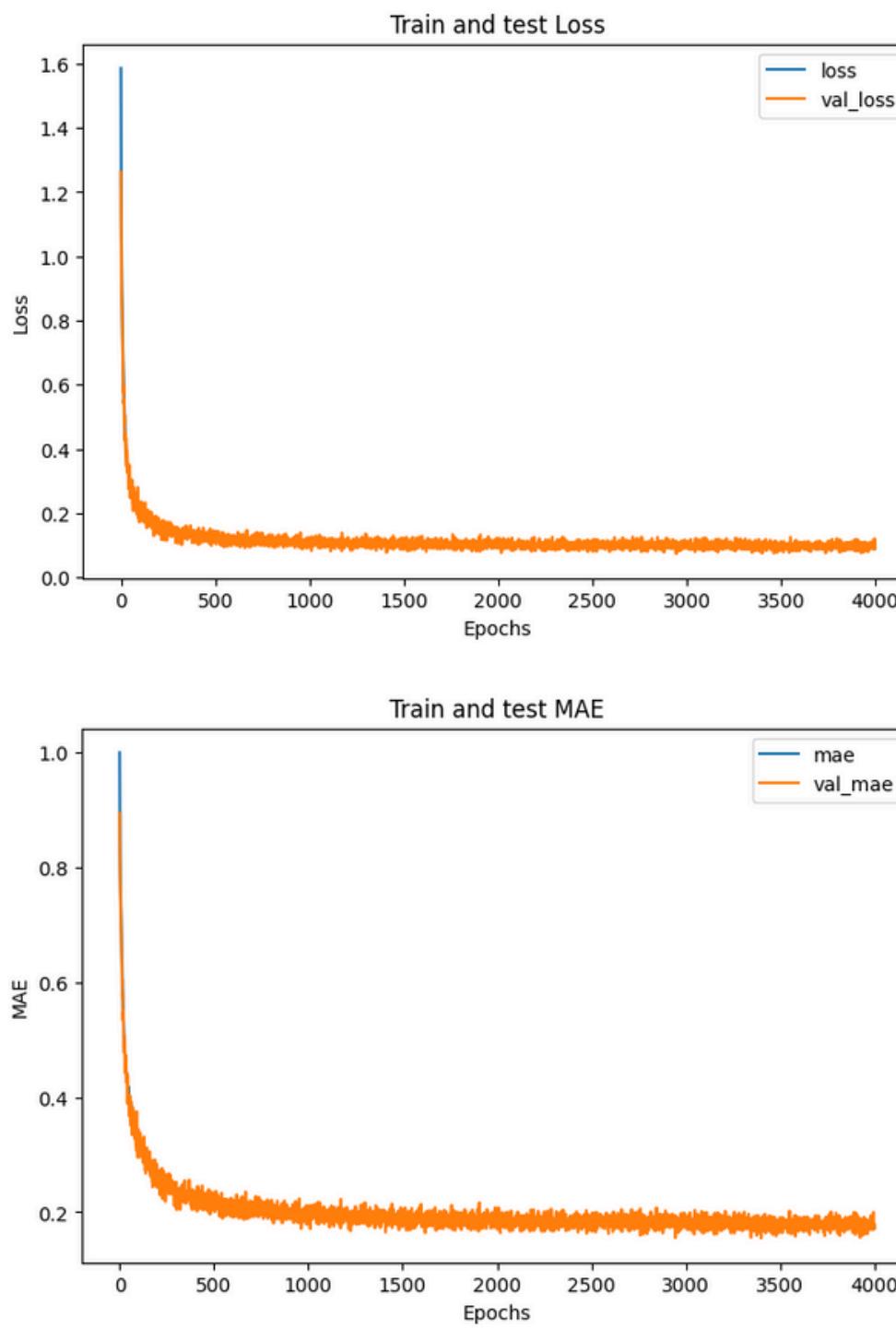


EPOCHS: 1500

RESULTADOS

# Resultados finales

A continuación se muestran los resultados producidos con la **U-Net final**.

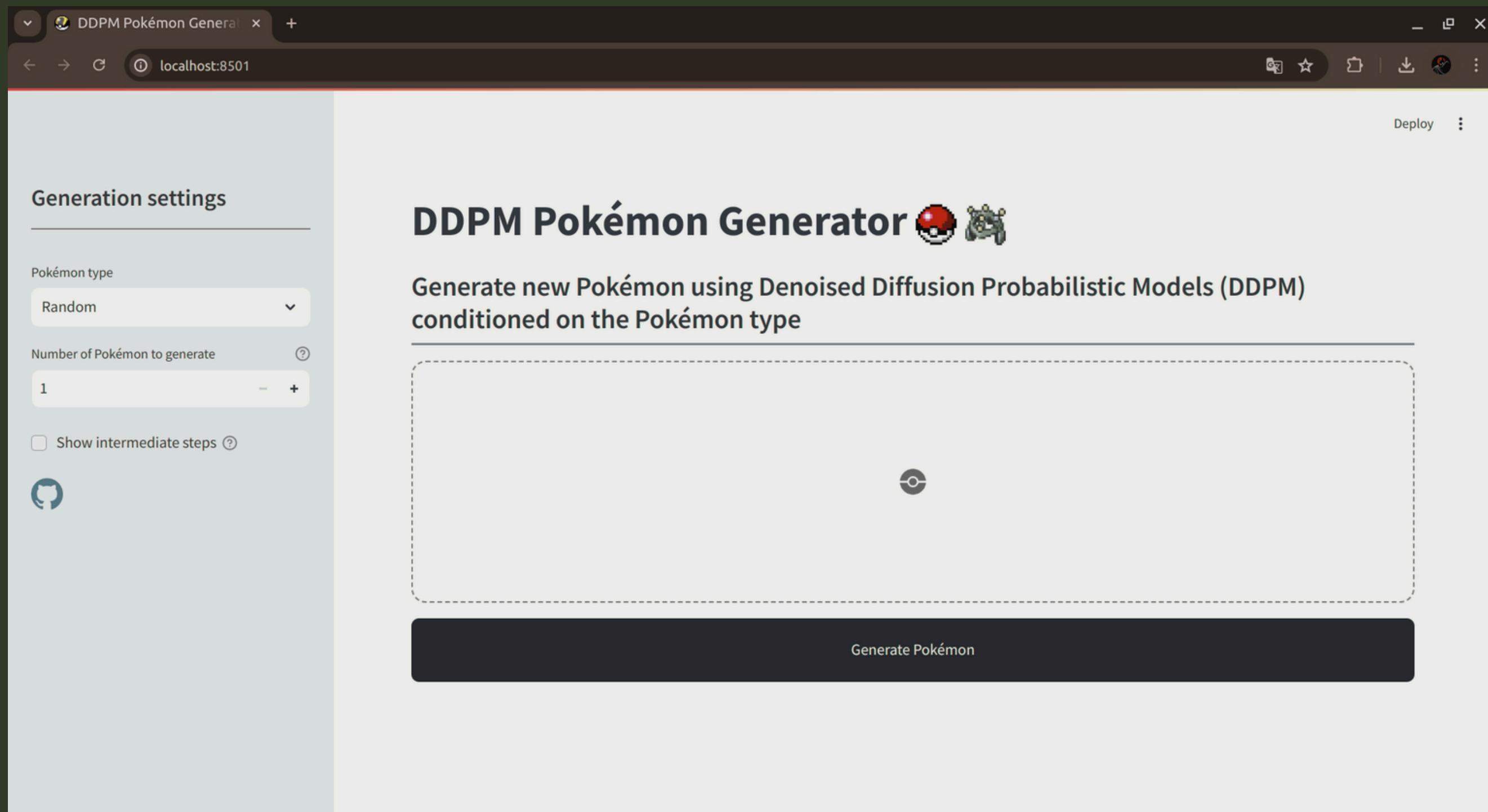


EPOCHS: 4 0 0 0

R E S U L T A D O S

# Interfaz gráfica

Se desarrolla una interfaz de usuario con el objetivo de **facilitar** la generación de Pokémon haciéndolo de una manera más intuitiva y visual



RESULTADOS

# Líneas de investigación futuras

P A R T E   0 5

# Supresión del Overfitting



## EL CAMINO DE LA GENERALIZACIÓN

Continuar ajustando hiperparámetros y buscar más en profundidad técnicas más eficientes para el entrenamiento de modelos de deep learning y en concreto de difusión.



## ANÁLISIS PROFUNDO Y OBJETIVO

Con el objetivo de evaluar la calidad de los resultados generados por el modelo de difusión implementado, se habría deseado implementar varias métricas de evaluación específicas para modelos generativos como: **KID, FID** y la métrica de **Wasserstein**.



## AUMENTO DEL CONJUNTO DE DATOS

Implementación de técnicas avanzadas de data augmentation para aumentar la cantidad y la diversidad de los datos de entrenamiento sin reducir la potencia de la red.

# Métricas de Evaluación

# Data Augmentation

# Aumento de la condicionalidad del modelo

## Modelo DDIM

- **AMPLIACIÓN DE LA FUNCIONALIDAD**  
Incorporación de factores de condicionamiento adicionales para mejorar el potencial creativo del modelo. Por ejemplo, un segundo tipo de Pokémon, características de color, o incluso atributos específicos como el tamaño.
- **REDUCCIÓN DEL TIEMPO DE INFERENCIA**  
Una de las principales limitaciones del modelo DDPM es su tiempo de inferencia relativamente largo, lo que puede ser un obstáculo para aplicaciones en tiempo real. Para abordar este problema, una posible línea de investigación sería la implementación de un modelo DDIM.  
  
Esta mejora en la velocidad de inferencia podría hacer que el modelo sea más adecuado para aplicaciones en tiempo real, como videojuegos o aplicaciones móviles.

# Conclusiones

P A R T E   0 6

# Conclusiones

DEL PROYECTO TRAS SU  
INVESTIGACIÓN Y DESARROLLO



## DDPM EN POKÉMON

Se ha explorado a nivel matemático y práctico la aplicación de modelos de difusión DDPM al ámbito único del análisis y modelización de datos de Pokémon utilizando técnicas como embeddings sinusoidales, módulos de atención entre otras.



## RESULTADOS Y DESAFÍOS

Los modelos de difusión demostraron ser efectivos en generar sprites Pokémon, aunque enfrentaron problemas de overfitting debido a un conjunto de datos limitado y variado. Estos problemas se minimizaron lo máximo posible a través del ajuste de hiperparámetros.



## CONOCIMIENTO Y HABILIDADES ADQUIRIDAS

Se adquirió un profundo entendimiento teórico y práctico de los modelos generativos, en concreto los modelos de difusión DDPM y durante el proyecto se mejoraron habilidades en programación, escritura académica y resolución de problemas.



¡Gracias por  
vuestra atención!

¿PREGUNTAS?

## REFERENCIAS

- [1] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli.  
**Deep unsupervised learning using nonequilibrium thermodynamics.**  
In International conference on machine learning, pages  
2256–2265. PMLR, 2015.
- [2] Jonathan Ho, Ajay Jain, and Pieter Abbeel.  
**Denoising diffusion probabilistic models.**  
Advances in neural information processing systems,  
33:6840–6851, 2020.
- [3] Alex Nichol, Prafulla Dhariwal.  
**Denoising diffusion probabilistic models.**  
arXiv:2102.09672v1 [cs.LG] 18 Feb 2021.
- [4] Pokémon sprite images (<https://www.kaggle.com/datasets/yehongjiang/pokemon-sprites-images>)

## REFERENCIAS (IMÁGENES)

[i1] Ainur Gainetdinov.

**Diffusion Models vs. GANs vs. VAEs: Comparison of Deep Generative Models.**

12 May 2023.

[i2] Diffusion (<https://www.sciencephoto.com/media/860120/view/diffusion>)

[i3] Nikhil Verma.

**Diffusing the mathematical equations of Diffusion Modelling.**

18 Nov 2022.