

# Redes Generadoras

Visión por Computador, curso 2024-2025

---

Silvia Martín Suazo, [silvia.martin@u-tad.com](mailto:silvia.martin@u-tad.com)

6 de diciembre de 2024

U-tad | Centro Universitario de Tecnología y Arte Digital

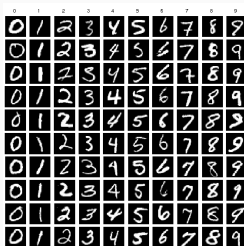


# Introducción

Además de las tareas de clasificación y segmentación anteriormente vistas, existe la tarea basada en la **generación de nuevas imágenes**. Estas imágenes son sintéticas y se **asemejan a las características** del dataset de entrenamiento.

Se utilizan en distintos ámbitos, ya sean:

- **Generación de imágenes**

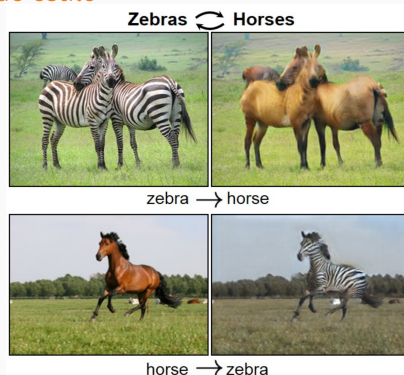


# Introducción

Además de las tareas de clasificación y segmentación anteriormente vistas, existe la tarea basada en la **generación de nuevas imágenes**. Estas imágenes son sintéticas y se **asemejan a las características** del dataset de entrenamiento.

Se utilizan en distintos ámbitos, ya sean:

- **Transferencia de estilo**



# Introducción

Además de las tareas de clasificación y segmentación anteriormente vistas, existe la tarea basada en la **generación de nuevas imágenes**. Estas imágenes son sintéticas y se **asemejan a las características** del dataset de entrenamiento.

Se utilizan en distintos ámbitos, ya sean:

- **Aumentar la resolución**



# Introducción

Además de las tareas de clasificación y segmentación anteriormente vistas, existe la tarea basada en la **generación de nuevas imágenes**. Estas imágenes son sintéticas y se **asemejan a las características** del dataset de entrenamiento.

Se utilizan en distintos ámbitos, ya sean:

- **Texto a imagen**

the flower has petals that  
are bright pinkish purple  
with white stigma



this white and yellow flower  
have thin white petals and a  
round yellow stamen

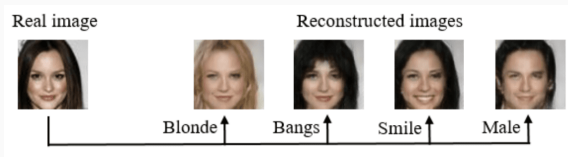


# Introducción

Además de las tareas de clasificación y segmentación anteriormente vistas, existe la tarea basada en la **generación de nuevas imágenes**. Estas imágenes son sintéticas y se **asemejan a las características** del dataset de entrenamiento.

Se utilizan en distintos ámbitos, ya sean:

- **Edición de imágenes**

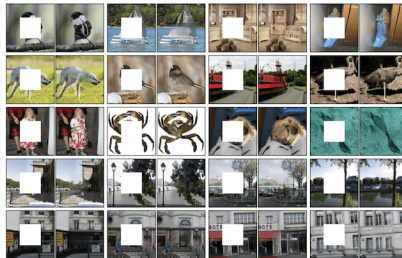


# Introducción

Además de las tareas de clasificación y segmentación anteriormente vistas, existe la tarea basada en la **generación de nuevas imágenes**. Estas imágenes son sintéticas y se **asemejan a las características** del dataset de entrenamiento.

Se utilizan en distintos ámbitos, ya sean:

- **Inpainting**



# Introducción

Además de las tareas de clasificación y segmentación anteriormente vistas, existe la tarea basada en la **generación de nuevas imágenes**. Estas imágenes son sintéticas y se **asemejan a las características** del dataset de entrenamiento.

Se utilizan en distintos ámbitos, ya sean:

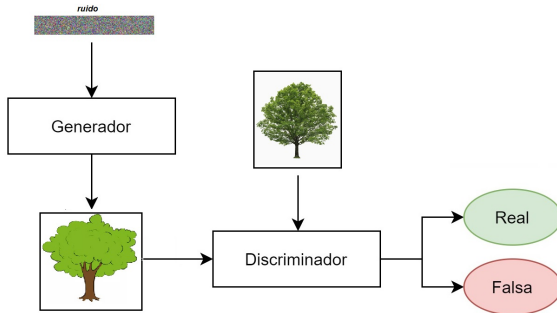
- **Predicción de video**





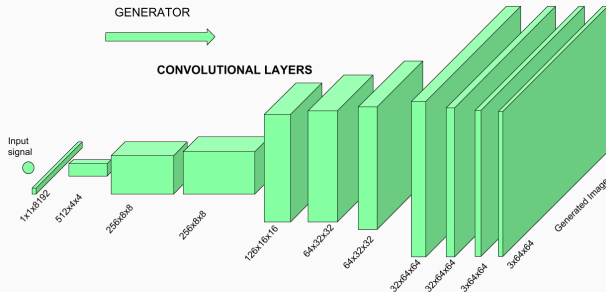
# Redes generativas adversariales

En 2014, Ian Goodfellow y sus compañeros proponen la red pionera en tareas generativas, **Generative Adversarial Networks (GAN)**[1]. Estas redes aprenden a modelar la distribución de entrada mediante dos subredes que compiten entre ellas.



# Redes generativas adversariales: Generador

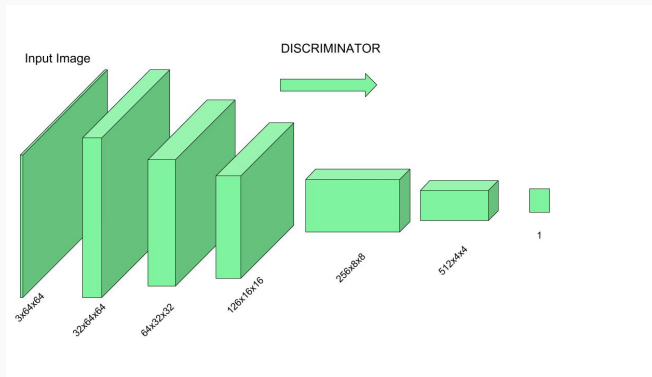
El generador es una subred cuyo principal objetivo es la **generación de datos**, normalmente imágenes falsas que sean realistas en el contexto del dataset.



Su arquitectura general es de **upsampling** para generar una imagen a partir de un vector de ruido (normalmente producido por una distribución gaussiana) que influye en las características de la imagen generada. Cada vector de ruido puede generar unos datos distintos.

# Redes generativas adversariales: Discriminador

Por otro lado, el discriminador se encarga de **distinguir si los datos son reales o falsos**. Para ello se le alimentan tanto los datos generados como los reales.



Su arquitectura es de un **clasificador binario basado en CNN**. Los datos reales reciben una etiqueta de 1, y los falsos de 0.

# Redes generativas adversariales: competición

El objetivo de estas redes es competir entre ellas, de tal forma que el **generador sea capaz de engañar al discriminador**, y el discriminador sea capaz de identificar las imágenes falsas. Por lo tanto, lo que se busca es **crear datos tan realistas que el discriminador no sea capaz de distinguirlas de las reales**. Cuando se llega a este punto se considera que la red ha finalizado su entrenamiento.

Uno de los principales es encontrar el equilibrio entre rendimiento de ambas subredes. Un discriminador que converge muy rápido haría que el generador no entrene y un sí el discriminador no está bien entrenado la red generadora no producirá datos de calidad.

# Redes generativas adversariales: competición

Durante el entrenamiento del discriminador se actualizan sus parámetros como clasificador binario. Por otro lado, el optimizador calcula los parámetros del generador basándose en la predicción presentada por el discriminador.

La función de pérdida utilizada es la entropía cruzada:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

Donde  $D(x)$  es el resultado del discriminador y  $G(x)$  del generador. El discriminador procura maximizar la función, mientras que el generador intenta minimizarla.

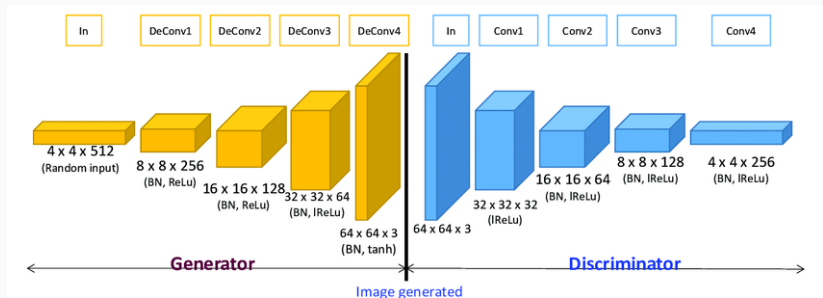
Los gradientes se pasan de discriminador a generador alternativamente. En primer lugar se actualizan los pesos del discriminador cada paso de una iteración, y antes de finalizarla se actualizan los pesos del generador.

Algunos de los problemas de estas redes son:

- Problemas para entrenar correctamente ambas redes.
- Necesidad de un gran dataset para generalizar correctamente.
- Existen versiones que aumentan el rendimiento.

# Red generativa antagónica profunda

Una de las variaciones de las GAN son las *Deep convolutional GANs* (DCGAN)[2]. Estas redes propuestas en 2016, se diferencian principalmente por la *eliminación de las capas de pooling* y la utilización de *capas convolucionales profundas*.



# Red generativa antagónica profunda

Otras modificaciones destacables son:

- Utilización de *strides* en las convoluciones del discriminador.
- Utilización de *strides fraccionadas* para el generador.
- Se utiliza *batch normalization*.
- Se eliminan las FFCC (DCNN).
- Se utilizan funciones de activación ReLu en todas las capas del generador menos en la *salida que se utiliza tanh*.
- Las capas del *discriminador* utilizan la función de activación *LeakyReLu*.



# Redes generativas antagónicas condicionales

Mientras que en las GAN originales **no hay control sobre los datos generados** y el resultado es **dependiente del ruido**, en las cGAN sí.

Las **Conditional Generative Adversarial Network (cGAN)**[3] se proponen en 2014, y se basan en la utilización de **un vector one-hot** donde cada posición está asociado con una característica de la imagen generada.

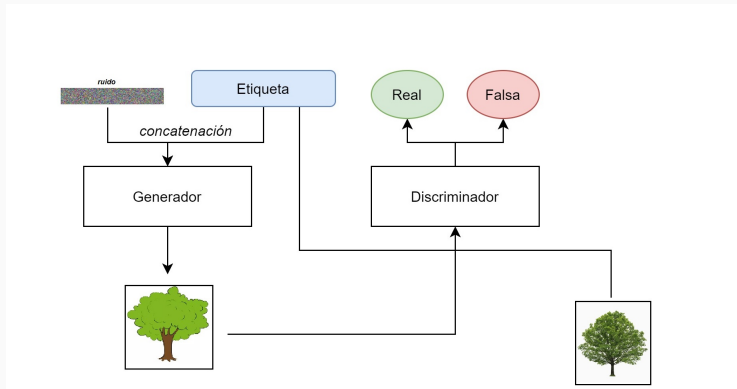
Por ejemplo, si generamos imágenes de camisetas algunas de las características pueden ser el color o el largo de la manga.



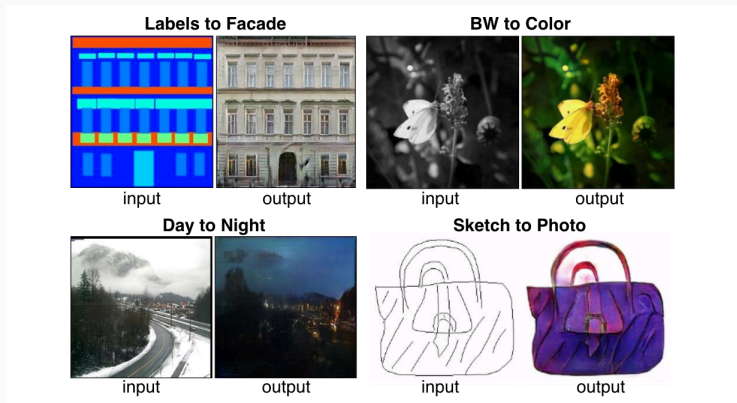
# Redes generativas antagónicas condicionales

Para utilizar este vector de características, **se concatena con el vector latente**. Además, la función de pérdida utilizada incorpora la información condicional.

El resto de la arquitectura sigue la estructura GAN vista anteriormente.



Es un caso de **cGAN** para "traducción" de imágenes.



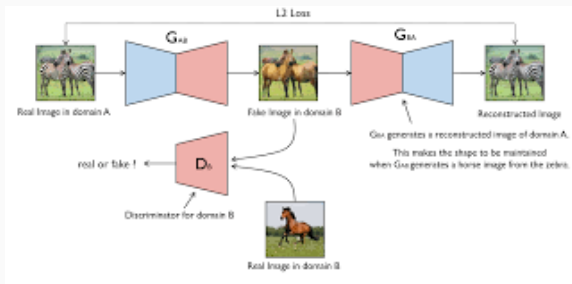
La CycleGAN[4] es un tipo de cGAN presentado en 2017 para la "traducción" de estilo de una imagen. Por ejemplo, podríamos modificar el estilo artístico de una pintura a otro completamente distinto.

Utiliza aprendizaje no supervisado al no necesitar de ejemplos con correspondencia en ambos dominios, como si ocurre en otras arquitecturas.

Para realizar esta "traducción", utiliza generadores y discriminadores dobles, uno por dominio.

# CycleGAN

La imagen es traducida del primer dominio al segundo, y de vuelta al dominio original, tratando de que sea lo más cercana a la imagen original posible. Por lo tanto, se utiliza una función de pérdida específica denominada **pérdida de consistencia cíclica**.



*\* El entrenamiento puede ser asimétrico, centrándose más en un dominio.*

Las **StyleGAN**[5] son una de las redes GAN más utilizadas y novedosas. Fueron propuestas por primera vez en 2018 por Nvidia. Tanto ésta como su mejora StyleGAN2 hacen posible la generación de imágenes de mayor calidad y una modificación mayor del estilo de las imágenes generadas.

Sus principales novedades fueron la introducción de una capa de **mappeo de estilo**, una **generación progresiva de imágenes de baja resolución a alta** y la utilización de **normalización por instancia**.

**Ejemplo StyleGAN**

- [1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. **Generative adversarial networks**, 2014.
- [2] Alec Radford, Luke Metz, and Soumith Chintala. **Unsupervised representation learning with deep convolutional generative adversarial networks**, 2016.
- [3] Mehdi Mirza and Simon Osindero. **Conditional generative adversarial nets**, 2014.
- [4] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. **Unpaired image-to-image translation using cycle-consistent adversarial networks**, 2020.

- [5] Tero Karras, Samuli Laine, and Timo Aila.  
A style-based generator architecture for generative adversarial networks, 2019.