1. **INTRODUCCIÓN**
   1. **En qué consiste el problema**

El objetivo principal del presente proyecto consiste en el desarrollo y entrenamiento de un modelo, basado en un enfoque de Aprendizaje Profundo, que genere imágenes consistentes con descripciones de texto.

* 1. **Qué tipo de soluciones existen actualmente. Qué nivel de calidad se obtiene, cómo lo obtienen, cuáles son sus limitaciones.**

Las Redes Neuronales Generativas Adversariales (GANN por sus siglas en inglés) en conjunto con los Autoencoders Variacionales (VAE por sus siglas en inglés) son modelos generativos que han adquirido relevancia y su análisis ha sido profundo dada la versatilidad de sus aplicaciones. En los últimos años, gran parte de la literatura tanto teórica como empírica se ha enfocado en clarificar el entendimiento sobre su gran capacidad para resolver tareas complejas. El primer trabajo de interés para este proyecto es el desarrollado por Goodfellow et al (2014), que exhibe una de las primeras representaciones formales de las redes neuronales tipo GAN. En términos generales, los autores consideran un nuevo enfoque para la estimación de modelos generativos basado en el entrenamiento simultaneo de 2 redes neuronales: (i) una red generativa (G) que captura la distribución de los datos, y (2) un modelo discriminativo (D) que funge como clasificador entre los datos provenientes del muestreo de entrenamiento y aquellos generados de manera sintética. La arquitectura de ambas redes neuronales está caracterizada por un Multilayer Perceptron (MLP), y el objetivo radica en maximizar la probabilidad de asignar la etiqueta correcta a cada uno de los datos de entrada, considerando la siguiente función objetivo vista en clase:



Los resultados del trabajo concluyen que, durante el proceso de entrenamiento, el discriminador alcanza un óptimo dado G, y considerando que PG se actualiza constantemente para optimizar implícitamente la función objetivo, entonces la probabilidad PG converge a la probabilidad de los datos (PDatos).

Posteriormente, Zhang et al (2016) desarrollan una extensión a la especificación anterior y proponen un nuevo modelo denominado Stacked Generative Adversarial Networks (StackGAN), para generar imágenes de alta calidad condicionadas a descripciones de texto. El experimento descompone el proceso generativo en 2 etapas, constituidas ambas por redes neuronales de tipo GAN. La primera etapa codifica la entrada de texto en una representación de embedding que emplea un encoder pre entrenado. Éste se combina a su vez con un muestreo aleatorio proveniente de una distribución Gaussiana, que se cruza con un generador para esbozar la forma y los colores primitivos de la imagen, produciéndose así representaciones de baja resolución. La GAN en la etapa 2 suaviza las distorsiones existentes e incluye los detalles contenidos en el texto que pudieran haberse omitido. Para esto, el modelo utiliza las representaciones resultantes de la etapa anterior junto con el embedding de texto, para corregir dichos defectos y generar imágenes de mejor resolución. Si bien el modelo obtiene representaciones fotográficas de alta calidad, los autores señalan como posible restricción el limitado número de pares de texto-imagen que pueden ser utilizadas durante el proceso de entrenamiento, lo que podría traducirse como una escasez de vocabulario y dificultar la elaboración de ciertas representaciones.

* 1. **Propuesta**

**Referencias.**

Goodfellow, I. J., Mirza, M., Courville, A., Bengio, Y. (2013). “Multi-prediction deep Boltzmann machines”, In NIPS’2013.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Courville, A., Begio, Y. (2014): “Generative Adversarial Nets”, In NIPS’2014.

Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., Metaxas, Y. (2016): “StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks”.