

# Deep Learning

Marco Monroy Ayala - CU  
Diego Villa Lizárraga - CU  
Rodrigo Suárez Segovia - 191351  
Maestría en Ciencia de Datos  
Instituto Tecnológico Autónomo de México

May 2020

## Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Problemática a resolver</b>	<b>3</b>
<b>3. Propuesta de solución</b>	<b>3</b>
<b>4. Experimentos y resultados</b>	<b>3</b>
<b>5. Conclusiones</b>	<b>3</b>

## Resumen

Generar imágenes de fotografías reales a partir de texto es un problema importante y tiene enormes aplicaciones como la edición de fotos o el diseño asistido por computadora. A pesar de que las muestras generadas provenientes de enfoques de texto pueden reflejar aproximadamente el significado de descripciones dadas, en muchos casos, no contienen los detalles finos y granulares que son parte de la imagen real. En este trabajo, proponemos el uso de Generative Adversarial Networks (GAN's) para generar imágenes fotorrealistas de 64 x 64 condicionadas a descripciones de texto. Los resultados ...

# 1. Introducción

La capacidad de asociar imágenes con oraciones de lenguaje natural que describen lo que se representa en ellas es un sello distintivo de la comprensión de imágenes y un requisito previo para aplicaciones como la búsqueda de imágenes basada en oraciones. Dicha capacidad, que pretende describir automáticamente las entidades, eventos y escenas representadas en una imagen es posiblemente la prueba más ambiciosa de comprensión de imágenes. Cualquier avance en esta tarea tiene implicaciones prácticas significativas, ya que hay miles de millones de imágenes en la web y colecciones de fotos personales. Nuestra capacidad para acceder de manera eficiente a la gran cantidad de información que contienen se ve obstaculizada por las limitaciones de los motores de búsqueda, (hardware y software) de imágenes estándar.

Considerando lo anterior, el principal reto o problemática en este proyecto consistió en generar imágenes basadas en oraciones o subtítulos como una tarea de clasificación. Para llevar a cabo esta tarea se presenta como punto de partida una nueva colección de referencia para la búsqueda y descripción de imágenes basadas en oraciones, que consta de 8,000 imágenes combinadas con cinco subtítulos diferentes que proporcionan descripciones claras de las entidades y eventos más destacados.

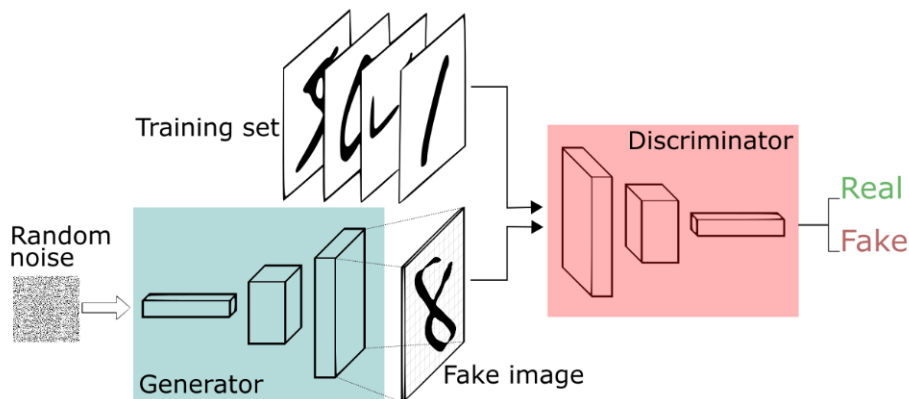


Figura 1: GAN's

Actualmente existen algunas soluciones con resultados que indican claramente la importancia de la capacitación en subtítulos múltiples por imagen, y de capturar características sintácticas (basadas en el orden de las palabras) y semánticas de estos subtítulos<sup>1</sup>. También realizaron una comparación en profundidad de las métricas de evaluación humana y automática para esta tarea, y se proponen

<sup>1</sup><https://www.jair.org/index.php/jair/article/view/10833/25854>

estrategias para recopilando juicios humanos a bajo costo y a gran escala, lo que permitió aumentar la colección con juicios de relevancia adicionales sobre qué subtítulos describen qué imagen.

Basado en estas investigaciones previas, nuestro trabajo propone como solución lo siguiente: (espacio para solución).

- 2. Problemática a resolver**
- 3. Propuesta de solución**
- 4. Experimentos y resultados**
- 5. Conclusiones**

“I always thought something was fundamentally wrong with the universe”  
[1]

## **Referencias**

- [1] D. Adams. *The Hitchhiker’s Guide to the Galaxy*. San Val, 1995.