

APROXIMACIONES SOBRE LA INTELIGENCIA ARTIFICIAL

APPROACHES TO ARTIFICIAL INTELLIGENCE

ENSAYO

Piñón, Mauricio¹

UVP Universidad del Valle de Puebla

mauricio.vargas@uvp.edu.mx

ORCID: 0000-0002-2386-6910

Recibido el 7 de febrero de 2024. Aceptado el 28 de febrero de 2024. Publicado el 30 de abril de 2024.

Reseña de Autor ¹

Doctor en Investigación e Innovación Educativa por la Universidad de Málaga y actualmente realiza proyectos de aproximación sobre la inteligencia artificial en entornos educativos y productivos.

Resumen

La Inteligencia Artificial ha venido a revolucionar la forma en que se realizan diversas actividades, principalmente en el sector productivo, de sistemas con un uso en el

aprendizaje profundo, análisis de datos, ingeniería de datos, entre otras aplicaciones industriales. El presente documento describe algunos usos prácticos de algunos algoritmos de creación de Inteligencia Artificial y busca compartir la potencialidad de algunos de los más importantes, al difundir de una forma sencilla las aplicaciones para aquellos con poca experiencia sobre estas novedades en el campo tecnológico.

Palabras clave: Inteligencia Artificial, Algoritmos, GANs.

Abstract

Artificial Intelligence has come to revolutionize the way in which various activities are carried out, mainly in the productive sector, of systems industry with a use in deep learning, data analysis, data engineering among other industrial applications.

This document describes some practical uses of some Artificial Intelligence creation algorithms and allows sharing the potential of some of the most important ones, allowing applications to be disseminated in a simple way for those new to these developments in the technological field.

Keywords: Artificial Intelligence, Algorithms, GANs.

Introducción

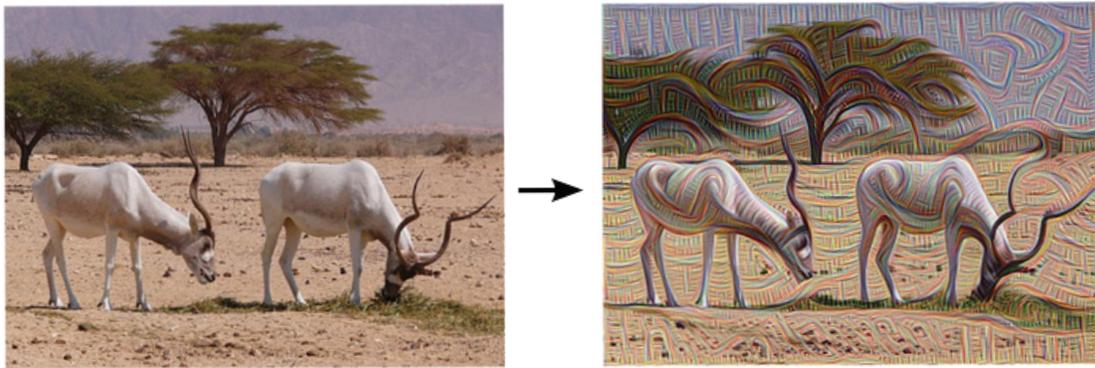
Los deepfake son engaños realizados por Inteligencia Artificial (IA) en donde se utilizan sus potencialidades para engañar a los usuarios con fines maliciosos.

Los algoritmos que se han utilizado en los últimos años se pueden observar a continuación:

Algoritmo DeepDream. El algoritmo de DeepDream fue creado por Google en el año 2013 con la intención de identificar imágenes y es el tipo de modelo que utiliza el Google imágenes en su buscador. La forma de trabajo de este algoritmo radica en que, si se presenta una imagen nueva para realizar el proceso de clasificado, la primera capa lo clasifica y lo modifica para entenderlo. Este proceso no es perfecto por lo que debe de ajustarse él mismo, ajusta la imagen nueva para que entre en las categorías de lo que ha interpretado.

Figura 1

Gacelas oníricas



Nota. Figura que representa a Gacelas con un efecto de pintura. Sin identificación en E. Rodríguez, (2018). El algoritmo de Google que desdibuja la línea entre humanos y computadoras. *Medium*. <https://medium.com/radio-pixel/deep-dream-5c2e7cdaf166>. De libre acceso.

Se encontró una discrepancia entre distintas fuentes, mientras que Gobetech (s.f.) describe que la arquitectura es un ciclo de alimentación constante entre las diferentes capas, en donde al inicio se le da una imagen y cada vez que pasa por una capa se

redibuja y que “no está ajustado a ninguna arquitectura específica” y DeepDream, destaca 3 capas, en donde la primera es sensible a bordes y orientación, la segunda interpreta objetos básicos y la tercera detecta elementos más sofisticados.

Ugni.io (2022) menciona que realiza estos procesos:

- Envío de una imagen por la red neuronal preentrenada
- Cálculo de salida de las capas
- Cálculo de gradiente
- Modificación de la imagen para ser comprendida por el algoritmo
- Repetición

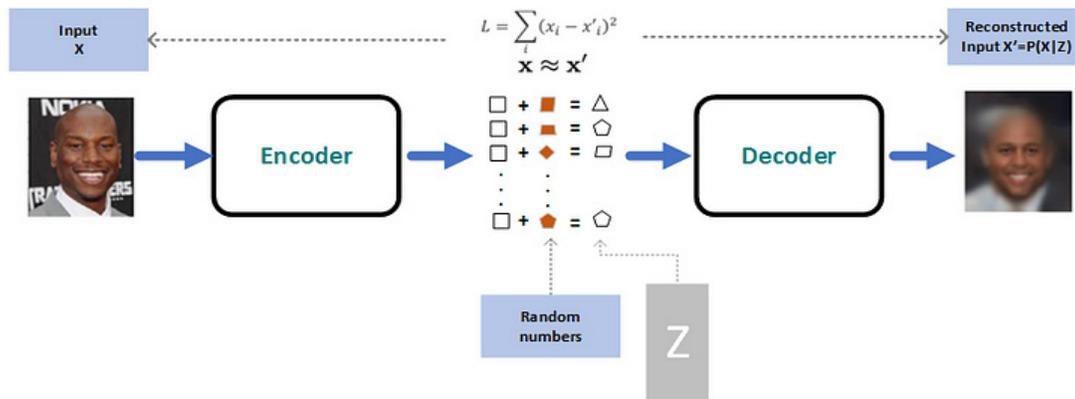
DeepDream fue liberado a los creadores desde el año 2015 y se encuentra disponible no solo para imágenes, sino para la creación de música, texto, reconocimiento de voz, entre otros, (Rodríguez, 2018).

Algoritmo Auto Ecoder Variacional (VAE). Antes de hablar del algoritmo es importante saber que los algoritmos de tipo auto ecoder son redes neuronales se caracteriza por ser parte del grupo de métodos de aprendizaje no supervisado en donde la información que se le proporciona es analizada y se va eliminado todo lo que tiene poca relevancia (Cañadas, 2021).

EL VAE es un algoritmo Auto Ecoder que permite de crear imágenes con base en las imágenes que tiene ya almacenadas o provenientes de una base de datos y a partir de su análisis se pueden crear imágenes nuevas muy parecidas a la realidad pero totalmente nuevas (Ichi.pro, 2021) y esto mediante la estructura que se observa en la figura 2.

Figura 2

Rostro nuevo



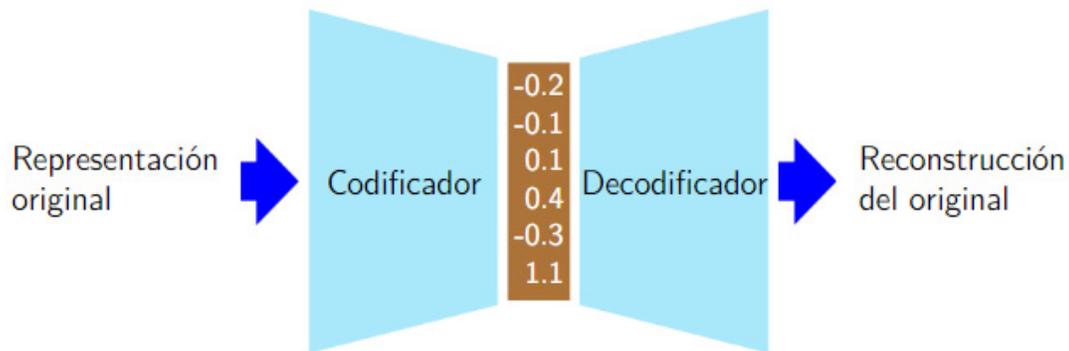
Nota. Figura que representa la forma en que trabaja el algoritmo. De AE a VAE utilizando variables aleatorias en Ichi.pro, (2021). Guía básica de bricolaje con código práctico para construir y entrenar VAE en caras de celebridades con Keras. Ichi.pro. <https://ichi.pro/es/autoencoders-variables-vae-para-principiantes-tutorial-paso-a-paso-116440083060969>. De libre acceso.

Otra concepción del VAE, proveniente de Meggi (s.f.) describe que su función principal como modelo generativo es crear nuevos elementos a partir de la información con la que cuenta mediante dos procedimientos la codificación e interpretación en donde la información de entrada va a ser analizada y posteriormente se reinterpretará para crear una representación totalmente nueva sin requerir supervisión.

Su arquitectura se resume en la siguiente figura:

Figura 2

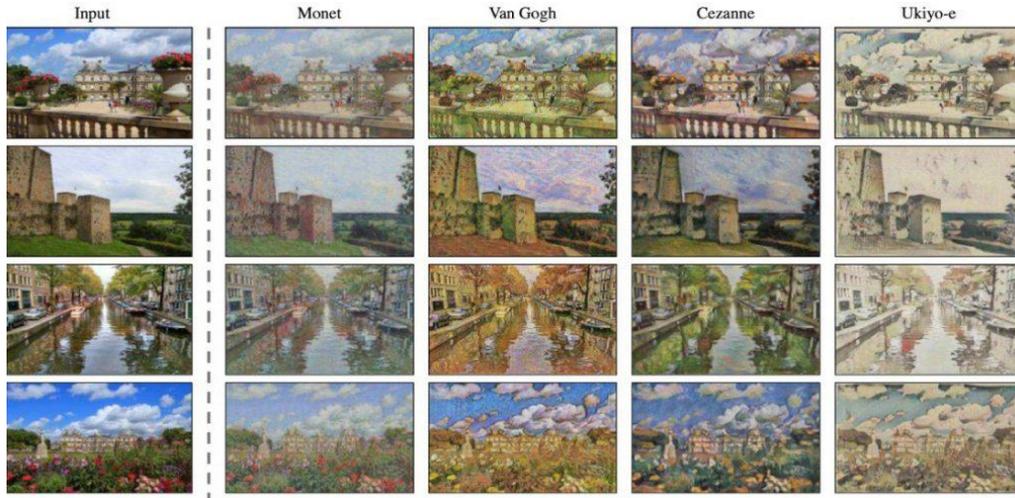
Arquitectura VAE



Nota. Representación de VAE. Esquema general de un auto-codificador en J. de la Torre, (2023). *Autocodificadores Variacionales (VAE) Fundamentos teóricos y aplicaciones*. https://www.researchgate.net/publication/368664523_Autocodificadores_Variacionales_VAE_Fundamentos_Te%27oricos_y_Aplicaciones. De libre acceso.

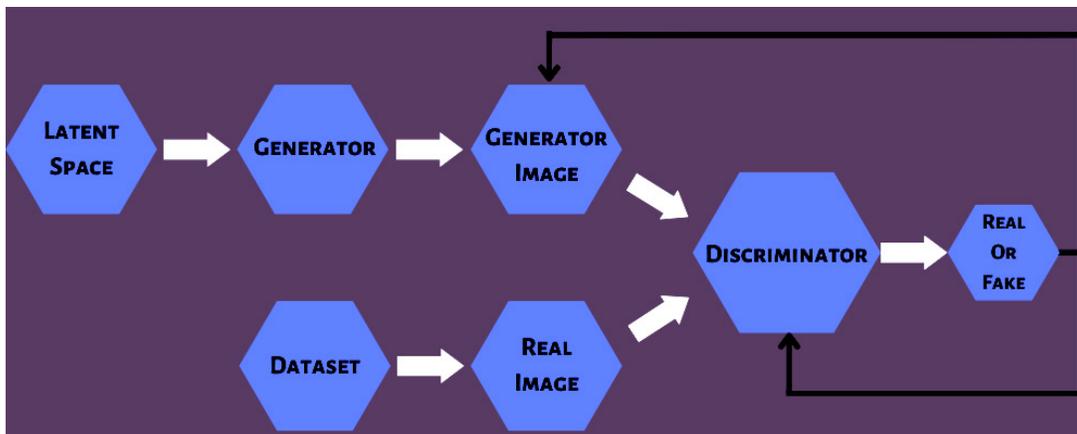
Algoritmo de Redes generativas adversarias (Gans). Este tipo de algoritmo tiene una diversidad de usos que va desde la creación de imágenes con un tono sepia que recuerda a fotos antiguas, imágenes sumamente realistas, rostros humanos, aumento de la voz, música, predicción de video, objetos 3D, hasta aplicaciones médicas (Pathak, s.f.; Artiga, 2023). Fue creado en el año 2014 por Ian Goodfellow y su estructura básica permite crear nuevos elementos, una vez que ha sido entrenado.

Figura 3
GANs



Nota. Usos de GANs. Aplicaciones de estilos de arte con una entrada y salida diferente en R. Artiga (2023). Que son las redes adversarias generativas. Una guía completa. <https://techaffinity.com/blog/redes-adversarias-generativas/>. De libre acceso.

Figura 4
Arquitectura



Nota. Arquitectura de GANs obtenido de A. Pathak (2022). Redes adversarias generativas (GAN): una introducción. <https://geekflare.com/es/generative-adversarial-networks/>. De uso libre.

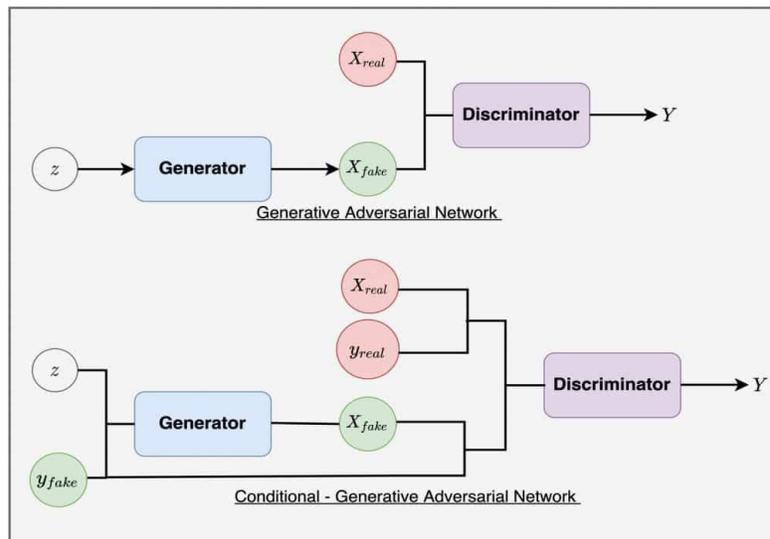
La arquitectura crea una figura en base al ruido latente y trata de engañar al discriminador, haciéndolo creer que es real y el discriminador trata de identificar si la figura es real o ficticia y aquí donde el generador es más sofisticado al crear mejores engaños y el discriminador se vuelve más ágil en identificar los engaños.

Algoritmo de redes generativas adversarias condicionales (cGans).

Esta es una red neuronal utilizada en el aprendizaje profundo en el cual se insertan etiquetas específicas o condicionadores dentro de una red de Gans, lo que hace que se actúe con la información generada y empiezan actuar el generador y el discriminador, uno para engañar y el otro para identificar el engaño con la adición de las condicionales para ambos elementos lo que hace que sea más sofisticada la generación.

Figura 5

Arquitectura CGANS



Nota. Estructura de un CGANS. En A. Sharma (2021). Conditional GAN (CGAN) in PyTorch and TensorFlow. *LearnOpenCV*. <https://learnopencv.com/conditional-gan-cgan-in-pytorch-and-tensorflow/>. De uso libre.

Sus usos son muy variables ya que permite rellenar espacios en una imagen como en un árbol, colocando las hojas, colocando mayor cantidad de detalles en un boceto o crear imágenes realistas a partir de texto que es el condicionante, mejorando las fotografías viejas aumentando la resolución entre otros (TaskUs, 2023).

Algoritmo de Redes generativas de transferencia (StyleGan). Es un algoritmo de creación de información altamente sofisticado que utiliza las bases del GANS y que utiliza el enfoque de crecimiento progresivo y el de Normalización de Instancias Adaptativas, lo que le permite generar imágenes de alta resolución SADA IA (2024).

Lo anterior lo hace por un entrenamiento por capas, comenzando con una baja resolución y pasando por varias más para crear detalles más exactos y con mayor resolución y poco a poco con el entrenamiento se van agregando cada vez más detalles o capas. Lo anterior mediante un mapeo más específico que se enfocan a distintas características lo que hace que no se vea influenciado por el resto de los elementos, es decir por el cabello, anteojos, etc.

Su uso es diverso, desde el punto de vista de la mercadotecnia, para crear opciones de diversos isotipos (Maksym, 2023) para empresas, para crear diversas opciones de imágenes ya que esta tecnología permite crear una diversidad de opciones muy cercanas a la realidad, diferentes y con gran detalle.

Figura 6

Usos de StyleGan



Nota. Usos de StyleGan en Sada IA (2024) *What is StyleGAN? Sada IA.* <https://sadaai.com/what-is-stylegan/>. De uso libre.

Existen diferentes algoritmos que permiten el desarrollo de fakes tanto en imágenes, como diagramas, sonidos y otros elementos que ciertamente se observan claramente en el mundo de las redes sociales y que pasan no solamente como un elemento de entretenimiento o de creatividad, ya que pueden ayudar a crear *catfish*, suplantación de identidad, préstamos fantasmas u otros por lo que es conveniente estar atentos.

A continuación, se muestra una tabla donde se pueden observar diferentes métodos utilizados para la detección de Deepfake.

Tabla 1

Comparación de algoritmos para detectar Deepfake

Algoritmo	Fortalezas	Debilidades
Redes neuronales convolucionales	Se ha observado su eficacia para la detección de deepfake en imágenes y video con un rendimiento superior permitiendo diferentes formatos de input. Su arquitectura en 3 capas la hace muy eficiente con imágenes.	Limitado a la tecnología disponible.
Aprendizaje por transferencia	Se basa en la obtención de la experiencia de un dominio madre para trasladarlo al dominio hijo reduciendo la cantidad de datos que se producen o manejan. Muy útil para hacer pruebas, innovar o mejorar.	No se pudieron detectar parámetros que permitieran identificar posibles debilidades, inconvenientes o malos usos.
Técnicas de forense digital	Los algoritmos detectan las falsificaciones mediante un análisis de imágenes activas (análisis previo con marcaje imperceptible) y pasivas (imágenes desconocidas). Su amplio uso no solamente para las Tecnologías de la Información sino en el mundo de la criminología y criminalística.	Restricciones por la fuente del material original, su calidad y la tecnología disponible. En algunos países que no se tiene bien desarrollada la técnica, no funciona como pruebas válidas.
Modelos de aprendizaje profundo	Herramientas como Tensorflow y Pytorch permite identificar los deepfakes mediante un entrenamiento en el que se le alimenta con elementos reales y generados por IA.	El entrenamiento es largo, lento y debe de ser alimentado para permitir la identificación adecuada, aunque siempre es susceptible a error.
Modelos de aprendizaje automático	Sensity, se especializa en identificar Deepfakes mediante modelos avanzados de aprendizaje automático.	La herramienta tiene un costo mensual y se observa para usos individuales y no corporativos.

Análisis de parpadeo y de movimiento irregular	Deepware es una herramienta que permite identificar el deepfake en videos observando la manipulación del rostro humano. Es de código abierto por lo que muchas personas se pueden beneficiar.	El modelo no es preciso en su totalidad y no puede hacer análisis de elementos de voz.
--	---	--

Nota. Tabla elaborada a partir de la información de Salako (2023); Gutiérrez (2023); Chavez et al. (2022) y Martín (2022).

Ahora, si se toma en cuenta las diferentes aplicaciones disponibles para la creación de DeepFake se revisaron 19 aplicaciones diferentes para crear video, avatares, arte y demás elementos con AI.

Para encontrar estas aplicaciones, del mar de opciones disponibles, se analizó la información de *influencers* en Instagram, especializados en recomendaciones de generación de Inteligencia artificial, obteniendo una diversidad de recursos diferentes.

Se encontró que varios de ellos, aunque mencionan gratuitos, no funcionan y otros ciertamente funcionan muy bien, pero son de paga, pero fue necesario navegar visualizar y encontrar las potencialidades de cada uno y fue grato identificar los modelos de creación de IA que se trabajaron en el anterior apartado.

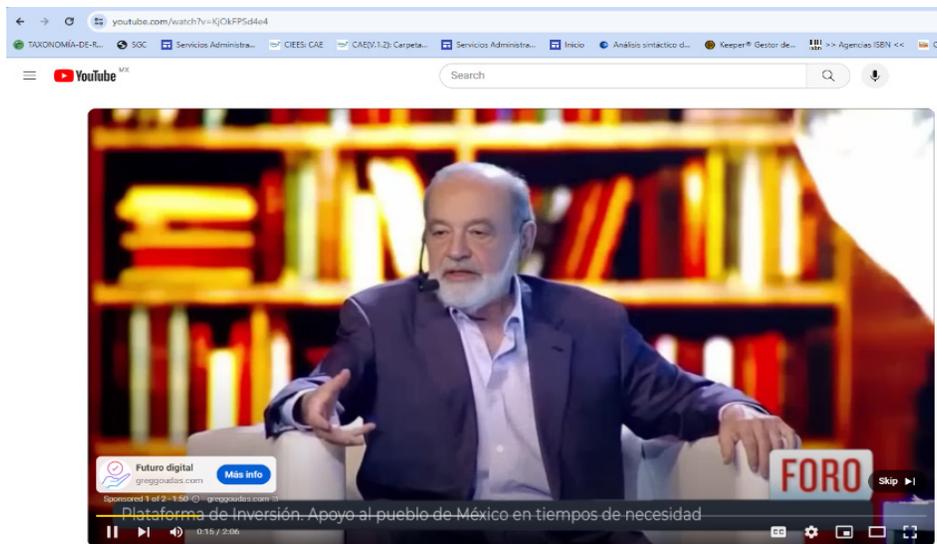
Las aplicaciones que se encontraron susceptibles para crear DeepFake fueron Leonardo.AI para crear arte, Pika que crea un video muy cortito pero muy bello y finalmente Studio D-ID fue excelente porque se carga con personas reales y se coloca el texto o se carga la voz, permitiendo que el avatar se mueva con mucha naturalidad. Realmente el tiempo que se pudo obtener fue corto, pero se observan las potencialidades de la herramienta, por ejemplo, colocar alguna figura pública y colocar un discurso diferente o mal intencionado.

Lo anterior lleva a reflexionar sobre la facilidad, potencialidad e implicaciones del uso de estas herramientas porque vaya, sin acceder a las aplicaciones pro, se puede hacer mucho y ahora con paga, seguramente los resultados pueden ser muy

impresionantes pero catastróficos si se tienen malas intenciones, como la publicidad que ha estado apareciendo en Youtube sobre inversión en Petróleo en México y se observa al presidente de México Andrés Manuel López Obrador y al empresario Carlos Slim alentando a la inversión y claramente se observa el DeepFake.

Figura 7

Carlos DeepFake



Nota. Siendo un anuncio es casi imposible realizar la cita que redirige al anuncio.

Se realizó una imagen con Leonardo.ai, posteriormente se colocó en Studio D-ID y se creó un video con movimiento, aunque realmente fue de un personaje ficticio, el resultado fue muy impresionante.

Figura 8

Blancanieves IA



Figura 9

Blancanieves en video



Funcionamiento de Open AI Glide

Openai, y su función de Chat GPT, se utiliza comúnmente para solicitar la creación de textos, corrección de estos, apoyar en la creación de estrategias y diversas tareas de creación. Su variante Glide, fue lanzada en 2022 tuvo una evolución que le permitió transformar texto en imágenes, crear páginas web e incluso aplicaciones, por medio de elementos visuales (The Black Box Lab, s.f.).

Para los noveles de la programación, una tarea como crear una página web, resultaba técnicamente imposible por lo que era necesario contratar a un profesional de la rama. Hace algunos años se lanzaron sitios como jimdo, los cuales, con un formato visual, con cajas y arrastrando elementos, se podía crear sitios web, sin embargo, hemos avanzado al punto que Glide puede hacer la misma tarea, pero con una hoja de cálculo abriendo las posibilidades a la personalización.

Glide utiliza lenguaje natural para la creación de imágenes realistas sin la utilización de Prompts, lo que permite añadir elementos, texturas, formas en donde su logaritmo es basado en probabilidad, creando imágenes diversas. Su algoritmo se base en un modelo de difusión hablado (ADM) que genera imágenes pequeñas de $64 * 64$, que continua con un mensaje transformador y que continua con un transformador, que amplía la imagen inicial a $256 * 256$ (Arrendajo, 2022).

Si se sintetiza la información, se encuentra que la forma en que el algoritmo de Glide trabaja es a base de lenguaje natural en donde igual que los modelos GAN que se generan imágenes a partir del ruido, en este caso el ruido es agregado paulatinamente en forma Gaussiana (densidad de probabilidad que sigue una distribución estadística en forma de campana) y luego es revertido, de manera que a una imagen se le agrega ruido hasta que solo queda ruido y luego se revierte (AI Coffe Break with Letitia, 23 de marzo de 2022) eliminando ruido por medio de variables específicas (los condicionantes que le colocamos) y esto se hace por CLIP (Difusión guiada) para crear una puntuación de similitud entre el texto y la imagen.

Glide con Dall-e y VQGAN+CLIP

Dall-e fue lanzado en el año 2021, con una gran popularidad, debido a su potencia para crear imágenes a partir de lenguaje natural obteniendo resultados únicos y creativos. Tiene algunas debilidades como la falta de interpretación del lenguaje, entendimiento de las especificaciones, requiere de una gran cantidad de información para hacer el entrenamiento, entre otros.

VQGAN+Clip hace la fusión de un vector variacional con auto encoder, en donde se incorpora el CLIP para hacer una alineación de la figura con el texto. Tiene algunas limitaciones como la necesidad de grandes cantidades de información, sesgo en los resultados debido a la interpretación semántica y ajuste de parámetros por la arquitectura con la que fue desarrollado.

CLIP, entendido como Preentrenamiento de imágenes y lenguaje contrastivo es utiliza el lenguaje natural del usuario para identificar, detectar y proyectar las características más representativas de la imagen analizada pero ciertamente con algunos problemas con las palabras dibujadas, identificación de los géneros sexuales y de un marcaje claro del color por la cantidad de información involucrada y al mismo tiempo por los parámetros y entendimiento del lenguaje natural (7 Minutos, s.f.).

De acuerdo con AI Coffe Break with Letitia (23 de marzo de 2022) Glide puede generar más fotos realistas a comparación de Dall-e, ya que los resultaos obtenidos por este último, son más borrosos y desordenados, lo que habla de una mayor exactitud del modelo Glide debido a la inclusión del CLIP y fue lanzado a inicio de 2022.

Los resultados de Glide se traduce en mayor calidad, pero se debe de sacrificar en tiempo de espera en la generación ya que Glide debe de tener más procesos de análisis, más ruido, pero que vale la pena en los resultados. La versión que se encuentra disponible no tiene tanto entrenamiento como las versiones full por lo

que actualmente Glide tiene bastantes restricciones que llevarán a ralentizar la tecnología hasta que se lance el modelo completo y abierto para todas las personas y más para los investigadores.

Por lo anterior se puede observar que la evolución comenzó con la creación de imágenes creativas con Dall-e pero tiene discrepancias con los resultados en base a lo solicitado, tiempo después con VQGAN+CLIP, incorpora a este último que permite aumentar la exactitud de los resultados obtenidos, por lo que se termina con Glide, que crea las imágenes como Dall- e, incorporando las potencialidades de CLIP de VQGAN+CLIP.

El surgimiento de VQGAN+CLIP?

Después del auge de los algoritmos para el desarrollo de imágenes a través de texto que surgió por la implementación de GAN con CLIP se creó en el 2021 VQGAN, un modelo que permite desarrollar imágenes de diferentes tamaños con unas características muy interesantes como es calidad en la nitidez de la imagen obtenida, coherencia de acuerdo con los términos establecidos en lenguaje natural y lo generado. Surgió con la necesidad de ampliar el alcance de los GLAN y la fusión con CLIP generando mejores resultados. La inclusión de CLIP, permite insertar discriminadores en el proceso de creación de imagen, lo que beneficia a una creación mucho más cercana a los términos establecidos por el usuario y evita la creación de nubes, ruido o elementos borrosos de baja calidad.

Dall-e en el mundo

Dall-e es un modelo generacional autorregresivo parecido a GPT, excelente para la creación diversa de resultados visuales por medio de palabras, que tiene ciertas restricciones con los detalles y con la creación de realismo, por lo que se observa su potencia para generar elementos nuevos, creativos, novedosos, pero ciertamente, más cercanos a la caricatura. Su potencialidad para crear elementos totalmente nuevos fomenta a la controversia de la originalidad de los autores, muy debatida en la actualidad, donde si los autores realmente obtienen los créditos por crear algo por medio de palabras o es realmente el algoritmo quien lo hace. Al mismo tiempo, se observa tiene la capacidad de fomentar, establecer y consolidar estereotipos por su marcada tendencia a crear resultados con estereotipos occidentales, roles de género y de ser explícito con violencia y otros temas sensibles.

La evolución de los GANs como Glide

GANs genera fotorealismo por medio de ruido aleatorio y su proceso es mucho relativamente rápido y si lo comparamos con los Glide que tiene un tiempo de respuesta mucho más elevado por el número de procesos que requiere hacer para obtener una respuesta, entonces se podría observar en un futuro una arquitectura de GANs con un tiempo de respuesta parecido elevado, pero con los resultados y la calidad de Glide. En donde GAN requiere es en detalles, Glide pierde en velocidad, por lo que si se mejoran estos detalles podremos tener un proceso mucho más rápido, pero al mismo tiempo con calidad insuperable.

Referencias

- 7 Minutos. (s.f.). ¿Los modelos CLIP ‘repiten como loros’ el texto de las imágenes? Este artículo explora el sesgo de detección de texto en los sistemas visión-lenguaje. *7 Minutos*. <https://7minutos.es/2023/12/30/los-modelos-clip-repiten-como-loros-el-texto-de-las-imagenes-este-articulo-explora-el-sesgo-de-deteccion-de-texto-en-los-sistemas-vision-lenguaje/>
- AI Coffe Break with Letitia. (23 de marzo de 2022). Diffusion models explained. How does OpenAi’s Glade work? [Video]. *Youtube*. <https://www.youtube.com/watch?v=344w5h24-h8>
- Arrendajo. (2022). Open Ai Glide (difusión): generación de imágenes con solo una entrada de texto. *HashDork*. <https://hashdork.com/es/deslizamiento-abierto-de-difusi%C3%B3n-ai/>
- Artiga, R. (2023). Que son las redes adversarias generativas. Una guía completa. *Techaffinity*. <https://techaffinity.com/blog/redes-adversarias-generativas/>
- Cañadas, R. (2021). Qué son los autoencoders. *AbDatum*. <https://abdatum.com/tecnologia/autoencoders>
- Chavez, S.M., Lozano, F. y García, J.J. (2022). Detección de deepfakes [Archivo PDF]. <https://repositorio.uniandes.edu.co/server/api/core/bitstreams/77954568-c15c-4493-b4f7-1e0b515ea597/content>
- De la Torre, J. (2023). *Autocodificadores Variacionales (VAE) Fundamentos teóricos y aplicaciones* [Archivo PDF]. https://www.researchgate.net/publication/368664523_Autocodificadores_Variacionales_VAE_Fundamentos_Te%27oricos_y_Aplicaciones
- DeepDream. (23 de octubre de 2019). En *Wikipedia*. https://es.wikipedia.org/wiki/Deep_Dream
- Gobetech. (s.f.). ¿Cuál es el diagrama de red de DeepDream de Google? *Gobetech*. https://tech.gobetech.com/33721/cual-es-el-diagrama-de-red-de-deepdream-de-google.html?expand_article=1

- Gutiérrez, M. (2023). Artificial Intelligence Techniques for detecting deepfakes and their impact on the veracity of online content. *International Journal of Systems Engineering*, 4(2), 6-9. <https://systemjournal.org/index.php/ijse/article/view/39/26>
- Ichi.pro. (2021). Guía básica de bricolaje con código práctico para construir y entrenar VAE en caras de celebridades con Keras. *Ichi.pro*. <https://ichi.pro/es/autoencoders-variables-vae-para-principiantes-tutorial-paso-a-paso-116440083060969>
- Maksym, T. (2023). Generación de imágenes GAN con Stylegan2. *Mobidev*. <https://mobidev.biz/blog/gan-image-generation-with-stylegan2>
- Martín, A. (2022). *Análisis forense de imágenes y videos manipulados (Deepfake)* [Tesis de Máster, Universidad Nacional de Educación a Distancia]. Repositorio institucional de la Universidad Nacional de Educación a Distancia. <http://e-spacio.uned.es/fez/view/bibliuned:master-ETSInformatica-CBS-Amartin>
- Meggi. (s.f.). Autoencoder variacional. *TechEdu*. <https://techlib.net/techedu/autoencoder-variacional-vae/>
- Pathak, A. (s.f.). Redes adversativas generativas (GAN): una introducción. *Geekflare*. <https://geekflare.com/es/generative-adversarial-networks/>
- Rodríguez, E. (2018). El algoritmo de Google que desdibuja la línea entre humanos y computadoras. *Medium*. <https://medium.com/radio-pixel/deep-dream-5c2e7cdaf166>
- Sada IA. (2024). *What is StyleGAN? Sada IA*. <https://sadaai.com/what-is-stylegan/>
- Salako, E. (2023). Herramientas de inteligencia artificial para detectar deepfakes. *IJ Net*. <https://ijn.net.org/es/story/herramientas-de-inteligencia-artificial-para-detectar-deepfakes#:~:text=TensorFlow%20y%20PyTorch%20son%20odos,de%20detectar%20indicios%20de%20manipulaci%C3%B3n>.
- Sharma, A. (2021). Conditional GAN (CGAN) in PyTorch and TensorFlow. *LearnOpenCV*. <https://learnopencv.com/conditional-gan-cgan-in-pytorch-and-tensorflow/>