

Generador de ilustraciones para libros utilizando inteligencia artificial

Nayeli Joaquinita Meléndez-Acosta^{1,2}, Edmundo Bonilla-Huerta²,
José Federico Ramírez-Cruz², Yesenia Nohemí González-Meneses²

¹ Universidad del Istmo,
Campus Ixtepec,
México

² Tecnológico Nacional de México,
Campus Apizaco,
México

nayelim@bianni.unistmo.edu.mx, {edmundo.bh,
federico.rc, yesenia.gm}@apizaco.tecnm.mx,

Resumen. En este artículo se presenta un modelo novedoso para la generación de ilustraciones para libros infantiles utilizando modelos de inteligencia artificial, se explica cada etapa del modelo propuesto, pero nos centramos en la revisión del área de investigación de los modelos existentes dentro de la generación automática de ilustraciones a partir de texto o fotografías, específicamente se realizaron pruebas utilizando DALL-E 2, una Red Adversaria Generativa (GAN) y Mini-DALL-E. Los tres modelos realizan la generación automática de ilustraciones, pero se deben considerar aspectos como la calidad de las imágenes y el tiempo de procesamiento. Una GAN requiere de mucho procesamiento, así que si se desea usar una GAN como generador de ilustraciones es necesario en nuestra propuesta pensar en otra variante que sea mucho más rápida. DALL-E 2 y Mini-DALL-E generan una gran variedad de ilustraciones con diferente calidad, estos modelos nos brindan la oportunidad de pensar en la creación de un libro con imágenes dinámicas.

Palabras clave: Generación automática de ilustraciones, red adversaria generativa (GAN), libros, ilustraciones, DALL-E 2, Mini-DALL-E.

Generator of Illustrations for Books Using Artificial Intelligence

Abstract. In this article, a novel model for the generation of illustrations for children's books using artificial intelligence models is presented, each stage of the proposed model is explained, but we focus on reviewing the research area of existing models within the automatic generation of illustrations from text or photographs, specifically tests were performed using DALL-E 2, a Generative Adversarial Network (GAN) and Mini-DALL-E. All models made on automatic generation of illustrations, but aspects such as image quality and processing time must be considered. A GAN requires a lot of processing, so if you want to use a GAN within your model, you need to think of another variant that is much faster. DALL-E 2 and Mini-DALL-E generate a wide variety of illustrations with

different quality, these models give us the opportunity to think about creating a book with dynamic images.

Keywords: Automatic generation of illustrations, generative antagonistic network (GAN), books, illustrations, DALL-E 2, Mini-DALL-E.

1. Introducción

Dibujar ilustraciones de historietas es una tarea compleja y un proceso difícil, esta forma artística es ampliamente utilizada en varios campos, incluyendo la publicidad, el cine y la educación infantil. Los libros ilustrados para niños son una herramienta para el aprendizaje indirecto, ya que mientras miran imágenes o leen las historietas de los cuentos aprender a leer, además brindan un contenido importante que aumenta la conciencia cultural, la conciencia lingüística y la motivación [1].

Actualmente la creación de ilustraciones se basa principalmente en la implementación manual, por lo que no existen algoritmos de aprendizaje automático desarrollados para la creación de ilustraciones, además la creación manual es muy laboriosa, ya que implica una cantidad considerable de habilidades artísticas, requiere de un artista o ilustrador bien entrenado y de mucho tiempo, principalmente si se trata de un libro completo [2, 3, 4].

La Traslación de Imagen a Imagen (Image-to-Image translation I2I) basada en aprendizaje profundo ha logrado grandes resultados [4, 6]. Recientemente, los métodos de transferencia de estilo basados en aprendizaje son un parte importante y problema desafiante de la visión por computadora. I2I es un método que aprende el estilo de la imagen de referencia (imagen de estilo) y aplica el estilo a la imagen de entrada (imagen de contenido) para generar una nueva imagen que fusiona el contenido de la imagen de contenido y el estilo de la imagen de estilo.

Aunque algunos métodos existentes como las GAN han logrado resultados satisfactorios para crear ilustraciones, este modelo no necesariamente genera ilustraciones impecables [5].

Los métodos existentes suelen tener algunos problemas, entre los que se incluyen principalmente tres: (1) las imágenes generadas no tienen texturas bien definidas; (2) las imágenes generadas pierden el contenido de las imágenes originales; y (3) los parámetros de la red requieren de gran capacidad de memoria [2].

Por lo anteriormente descrito, el objetivo principal es describir un modelo para la generación automática de ilustraciones para libros infantiles. Se realiza la revisión del área de investigación de algunos modelos existentes en la generación automática de ilustraciones a partir de texto o imágenes, y también se realizar una comparativa entre modelos que generan ilustraciones de manera automática utilizando inteligencia artificial.

El resto del documento está organizado de la siguiente manera: los trabajos relacionados son descritos en la sección 2. El modelo para la generación automática de ilustraciones para libros infantiles se presenta en la sección 3. El análisis de resultados se aborda en la sección 4. Finalmente, en la sección 5 se incluyen algunas conclusiones y trabajo futuro.

2. Estado del arte

En trabajos relacionados para la generación de ilustraciones han aplicado dos enfoques principalmente, el primer enfoque es utilizar herramientas de software, pero algunas requieren un sistema de pago y la generación se realiza de manera manual.

El segundo enfoque es utilizar inteligencia artificial como procesamiento de imágenes o aprendizaje profundo, de este último específicamente las Redes Adversarias Generativas (Generative adversarial networks GAN), que permiten la generación automática a un estilo en particular, en este enfoque podemos encontrar dos categorías para la generación de ilustraciones: a partir de una imagen o a partir del lenguaje natural. Algunos trabajos basados en ambos enfoques y que son relevantes para nuestra investigación, son:

2.1. Generación de ilustraciones utilizando software

Fatimah et al. [1] realizan un estudio que tiene por objetivo conocer el uso de ToonDoo en la enseñanza de cuentos en inglés y descubrir los beneficios de esta herramienta para el desempeño de la enseñanza, además resaltan la importancia de crear comics o historietas ilustradas que benefician al estudiante, ya que el cuento puede brindar un contenido importante que aumenta la conciencia cultural, la conciencia lingüística y la motivación.

ToonDoo es un creador de cómics completamente en línea, que ofrece muchas opciones para la creación de historietas a través de una interfaz fácil de usar. Los resultados mostraron que esta herramienta al generar historias es muy útil para facilitar la imaginación de los estudiantes, promover su capacidad de hablar y producir una mejor experiencia de aprendizaje.

2.2. Generación de ilustraciones a partir de una imagen utilizando Inteligencia Artificial

Uno de los modelos generativos más utilizados y eficientes son las redes adversarias generativas (GANs), formadas por dos redes neuronales que compiten entre sí para crear una salida que se parece mucho a la entrada, así se establece un juego entre las dos redes llamadas: generadora y discriminadora, dichas redes tienen roles adversarios y cada red está representada por una función diferenciable controlada por un conjunto de parámetros.

La red generadora G intenta crear una nueva imagen basada en la imagen de entrada, mientras que la discriminadora D está entrenada para distinguir la diferencia entre imágenes reales y falsas.

El objetivo de una GAN es que la imagen creada por la red generadora se parezca tanto a la imagen real que pueda engañar al discriminador para que piense que la imagen generada es real [6]. A continuación, se muestran algunos trabajos relacionados que hace uso de una GAN:

Chen et al. [2] presentan un enfoque novedoso para transformar fotografías de escenas del mundo real en imágenes de estilo anime, para ello proponen una red

generativa adversaria ligera llamada AnimeGAN que permita lograr una transferencia de estilo rápidamente.

AnimeGAN puede ser entrenada fácilmente con dos conjuntos de datos de entrenamiento no relacionados y sus parámetros requieren una menor capacidad de memoria. Los resultados experimentales muestran que su método puede transformar rápidamente imágenes de fotos del mundo real en imágenes de estilo anime de alta calidad.

Chen et al. [7] proponen CartoonGAN, una red adversaria generativa para la estilización de ilustraciones, su método está basado en la generación de ilustraciones a partir de fotos e imágenes. Hicsonmez et al. [8] exploran las ilustraciones en libros para niños como un nuevo dominio en la traducción de imagen a imagen.

Proponen una nueva red generativa adversaria llamada GANILLA para abordar este problema y mostrar que la red resultante logra un mejor equilibrio entre estilo y contenido. También proponen una nueva métrica para la evaluación cuantitativa de la ilustración generada, donde tanto el contenido como el estilo se toman en cuenta, entonces para evaluar la preservación del contenido ellos proponen un clasificador de contenido *Content-CNN*.

Barzilay et al. [5] proponen una red adversarial generativa de estilo Multi-Ilustrador (MISS GAN), la cual tiene una arquitectura para la traducción no supervisada de imagen a ilustración, que puede generar a partir de una imagen de entrada imágenes de diferente estilo, pero conservando el contenido.

El conjunto de datos de ilustraciones está compuesto por ilustraciones de siete ilustradores diferentes, es decir contiene siete estilos diferentes. Los métodos existentes requieren entrenar a varias redes generadoras para manejar estilos diferentes, lo que limita su uso práctico. MISS GAN usa la imagen de entrada y utiliza la información de otras imágenes empleando sólo un modelo entrenado.

Shu et al. [3] proponen una arquitectura novedosa de red adversaria generativa (GAN) multiestilo, llamada MS-CartoonGAN, que puede transformar fotos en ilustraciones de múltiples estilos. Para el entrenamiento MS-CartoonGAN usa fotos e imágenes de ilustraciones de múltiples estilos.

Su arquitectura de red compartida explota las características comunes de los estilos de las ilustraciones, logrando mejores caricaturas y ser más eficiente que la caricaturización de un sólo estilo.

A través de extensos experimentos que incluye un estudio de usuario, demuestran la superioridad del método propuesto, superando a los métodos de transferencia de estilo único y de estilo múltiple de última generación. Pictonaut et al. [9] describen un método novedoso para sintetizar automáticamente tomas animadas de imágenes en movimiento.

En lugar de abordar el desafío únicamente como un problema de traducción de imágenes, se propone un enfoque híbrido que combina la estimación de la pose humana en 3D de varias personas y una GAN. Los cuadros de video se procesan con OpenPose y SMPLify-X para obtener los parámetros de la pose 3D (cuerpo, manos y expresión facial) de todos los personajes representados.

El fondo está caricaturizado con una GAN. La evaluación cualitativa muestra que el enfoque es factible, ya que en un pequeño conjunto de datos proporcionan tomas sintetizadas obtenidas de escenas de películas reales.

2.3. Generación de ilustraciones a partir de texto utilizando inteligencia artificial

Proven-Bessel et al. [10] implementan ComicGAN, un nuevo modelo para generar ilustraciones de cómics a partir de texto basado en una GAN que sintetiza cómics según los descriptores de texto.

ComicGAN tiene dos características novedosas: (1) la descripción del texto se crea a partir de etiquetas mediante permutación y aumento, y (2) la codificación de imágenes personalizadas utiliza una Red Neuronal Convolucional.

Para evaluar ComicGAN proponen dos escenarios, la generación de imágenes a partir de descripciones, y la generación de imágenes a partir de un diálogo. Por otro lado, la generación de ilustraciones a partir de descriptores proporciona cómics más claros que incluyen personajes y colores que se especifican en los descriptores.

Xu et al. [11] proponen una Red Adversaria Generativa Atencional (AttnGAN) que permite el refinamiento de varias etapas para la generación de imágenes a partir de texto. Con AttnGAN se pueden sintetizar detalles en diferentes subregiones de la imagen prestando atención a las palabras relevantes en la descripción del texto.

Reed et al. [12] desarrollan una novedosa arquitectura profunda, una red generativa adversaria para generar de manera automática imágenes realistas a partir de texto, traduciendo conceptos a personajes visuales.

En este trabajo están interesados en traducir texto en forma de descripciones escritas por humanos de una sola frase en imágenes. Por ejemplo, “este pequeño pájaro tiene un pequeño pico anaranjado y vientre blanco” o “los pétalos de las flores son rosadas”. El problema de la generación de imágenes a partir de descripciones ganó interés, pero está lejos de resolverse.

Mini Dall-E es un modelo que se puede utilizar para generar imágenes a partir de texto [13][14]. Mini DALL-E es un intento de reproducir el primer modelo para generar imágenes llamado DALL·E, pero Mini Dall-E ofrece una alternativa de código abierto, que permite a cualquier persona la capacidad de generar imágenes [14].

Bansal et al. [15] realizan un estudio para evaluar el sesgo de algunos modelos generativos de texto a imagen al introducir textos neutros ('una foto de un abogado') que carecen de cualquier señal hacia un grupo social. Consideran tres ejes de sesgo: (1) género, (2) color de la piel y (3) cultural.

Sus resultados muestran que modelos como Mini Dall-E cubren diversos grupos sociales mientras preservan la calidad de la imagen. Mini Dall-E es un transformador, una poderosa red neuronal, que se entrena a partir de una gran cantidad de datos del lenguaje natural, aprende información sobre el lenguaje en sí [13].

DALL-E 2 es un generador de imágenes sintéticas originales correspondientes a un texto de entrada creado por OpenAI [16]. Marcus et al. [17] hacen un estudio con catorce preguntas para evaluar su sentido común, razonamiento y capacidad para comprender textos complejos.

Algunas conclusiones a las que llegaron son: la calidad visual de las imágenes es muy buena, DALL-E 2 es, sin duda, extremadamente impresionante en términos de generación de imágenes, si solo se especifican dos o tres objetos, el sistema casi siempre los muestra todos, pero, por otra parte, las relaciones entre entidades son particularmente desafiantes, ya que existen fallas en DALL-E 2 al asociar correctamente propiedades especificadas con objetos como "un cubo rojo encima de un cubo azul".

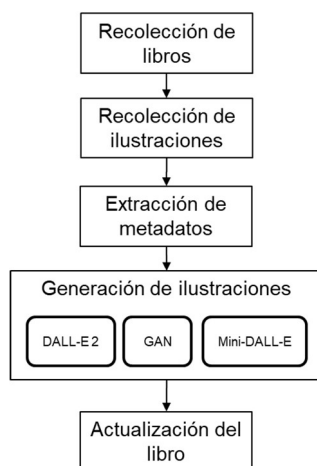


Fig. 1. Diagrama de bloques del modelo propuesto para la generación de ilustraciones para libros infantiles.

DALL-E 2 es un método zero-shot, es decir tienen la capacidad de generar imágenes basadas en descriptores de texto que no se ven durante el entrenamiento, así estos modelos pueden combinar conceptos que aprendió por separado, pero nunca vistos juntos en una imagen generada [16].

3. Modelo propuesto

La ilustración de libros es una forma de arte que se utiliza principalmente para crear imágenes y dibujos para libros. Sin embargo, el uso de ilustraciones no es sólo para embellecer los libros; son fundamentales para mejorar el contenido escrito del libro. En esta sección se muestra el diagrama general del modelo propuesto para la generación de ilustraciones.

La primera etapa es la recolección de libros en formato pdf que permiten realizar la recolección de imágenes y la generación de metadatos, los cuales son necesarios para la siguiente etapa la generación de ilustraciones usando modelos de inteligencia artificial, finalmente se debe realizar la actualización de las imágenes en el libro. El diagrama de bloques del modelo propuesto se presenta en la Fig. 1.

3.1. Recolección de libros

En esta etapa se realiza la recolección de libros ilustrados que serán utilizados para las pruebas y la recolección de ilustraciones, este tipo de libros contienen texto que se acompaña de ilustraciones que reflejan imágenes de la historia que se está narrando.

Algunos de ellos son Alicia en el País de la Maravillas, el Principito, la Vuelta al Mundo y el Mago de Oz, por ahora nos centramos en libros ilustrados de cuentos clásicos como los mencionados, la recolección se realiza de manera manual, buscando en internet los libros en formato pdf para su descarga.

Tabla 1. Metadatos de las imágenes del libro “El Principito”.

#	pág	ancho	alto	formato	objeto	etiqueta
1	7	257	158	jpg	26	Boa eating a mouse cartoon
2	7	299	98	jpg	27	Boa with an elephant in the stomach cartoon
3	11	338	365	jpg	38	Little Prince cartoon
4	12	129	114	jpg	41	White lamb jumping cartoon
5	12	135	127	jpg	42	White lamb cartoon
6	12	135	116	jpg	43	White lamb cartoon
7	12	146	62	jpg	44	White box with three holes cartoon
8	15	262	357	jpg	51	A boy walked on the planet earth cartoon
⋮	⋮	⋮	⋮	⋮	⋮	⋮
40	82	278	272	jpg	222	A boy sitting on the sand cartoon
41	84	476	660	jpg	228	A boy gently falling into the sand cartoon

3.2. Recolección de ilustraciones

Los libros en pdf recolectados en la etapa anterior son usados para realizar la recolección de imágenes, para ello primero usamos un manipulador de archivos pdf llamado *PDFtf*, esta herramienta descomprime el archivo y hace más fácil su lectura y manipulación.






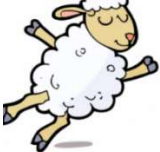
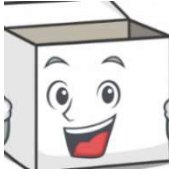





La extracción de ilustraciones se realiza de manera semiautomática haciendo uso de la herramienta *pdfimages* (parte de *poppler-utils*), la cual permite extraer los archivos de todas las imágenes en sus formatos originales, creando la base de datos de ilustraciones.

3.3. Extracción de metadatos

En esta etapa se realiza la construcción de una tabla con los metadatos de las imágenes, esta construcción se realiza de manera semiautomática utilizando la herramienta *pdfimages*, la cual también nos proporciona una lista de todas las imágenes del pdf con sus metadatos.

Una vez que se extraen los metadatos del pdf se crea una tabla, esta tabla está compuesta por siete metadatos: (1) número de imagen, (2) página de localización de la imagen, (3) las dos dimensiones (ancho, alto) de la imagen, (4) formato de la imagen, (5) número de objeto (localización de la imagen en el pdf) y (5) la etiqueta de la descripción de la imagen.

La Tabla 1 muestra la tabla construida con los metadatos que describen el contenido de imágenes del libro “El Principito”.

Etiqueta	Little Prince cartoon	White lamb jumping cartoon	White box with three holes cartoon	A boy walked on the planet earth cartoon
Imagen Generada 1				
Imagen Generada 2				
Imagen Generada 3				

3.4. Generación de ilustraciones

La generación automática de imágenes es una tarea compleja y en la que muchos investigadores aún están trabajando. En esta etapa se realiza la revisión de tres modelos para la generación de ilustraciones, los cuales utilizan inteligencia artificial: DALL-E, una GAN y Mini-DALL-E.

Generación de imágenes utilizando DALL-E 2. DALL-E 2 puede crear imágenes realistas de una descripción en lenguaje natural, primero se realiza una consulta a los metadatos creados en la etapa anterior para obtener los parámetros necesarios para poder ejecutar DALL-E 2.

Los parámetros utilizados para generar la imagen son: las dos dimensiones (ancho, alto) de la imagen y la etiqueta de la descripción de la imagen, una vez creadas las nuevas imágenes se salvan en el banco de ilustraciones.

Generación de ilustraciones utilizando una GAN. Una GAN es un modelo de aprendizaje profundo que pretende construir una réplica $x = g(y)$ para generar las muestras deseadas de x a partir de la variable y , su enfoque es encontrar el equilibrio entre un generador y un discriminador.

Para probar la generación de ilustraciones usando una GAN, tomamos como referencia la GAN de [8], la cual ya contiene 10 modelos pre-entrenado de 10 artistas diferentes.

Finalmente, **generación de imágenes utilizando Mini-DALL-E**. Este proceso se realiza de manera semejante a como lo hicimos utilizando DALL-E 2, realizamos la consulta a los metadatos para obtener los parámetros necesarios (las dimensiones ancho y alto; y la etiqueta de descripción de la imagen).

3.5. Actualización del libro

Finalmente, en esta última etapa las ilustraciones son actualizadas por las nuevas ilustraciones generadas utilizando los modelos generadores de ilustraciones revisados en la etapa anterior.

Para realizar la actualización de la ilustración es necesario consultar la tabla de metadatos para identificar por cada imagen: el número de objeto (localización de la imagen dentro del pdf) y las dimensiones (ancho, alto) de la imagen.

La identificación del número de objeto en el pdf que representa a la imagen es muy importante para localizar la posición de la imagen dentro del archivo pdf. Así se genera una nueva versión del libro que contiene imágenes generadas ya sea por DALL-E 2, la GAN o Mini-DALL-E. Resumiendo, después de revisar cada una de las etapas de la generación de ilustraciones a continuación se presenta el algoritmo de todo el proceso:

Algoritmo 1. Generación de ilustraciones

#	Algoritmo
1	Recolección de libros en internet (archivos en PDF)
2	Descomprimir el archivo PDF usando la herramienta <i>PDFf</i>
3	Extraer las ilustraciones originales del libro usando la herramienta <i>pdfimages</i>
4	Extraer los metadatos de las ilustraciones contenidas en el libro usando la herramienta <i>pdfimages</i>
5	Crear la tabla de metadatos usando los datos del paso anterior
6	Generar las imágenes usando un modelo de inteligencia artificial
7	Abrir el archivo PDF descomprimido a actualizar
8	Identificar la imagen a actualizar a través del número de objeto guardado en la tabla de metadatos
9	Abrir el archivo PDF y actualizar la ilustración

4. Análisis de resultados

En esta sección se muestra nuestros experimentos preliminares sobre la generación automática de ilustraciones, la implementación se realiza en Python y los modelos pre entrenados de la GAN usan PyTorch.

Para realizar los experimentos se ha utilizado el libro “El principito”. Las imágenes generadas son 256 x 256 píxeles, así que se debe realizar un redimensionamiento al tamaño de la imagen original. A continuación, se muestran algunos resultados de la generación de ilustraciones utilizando cada uno de los modelos generadores que son revisados en este trabajo.

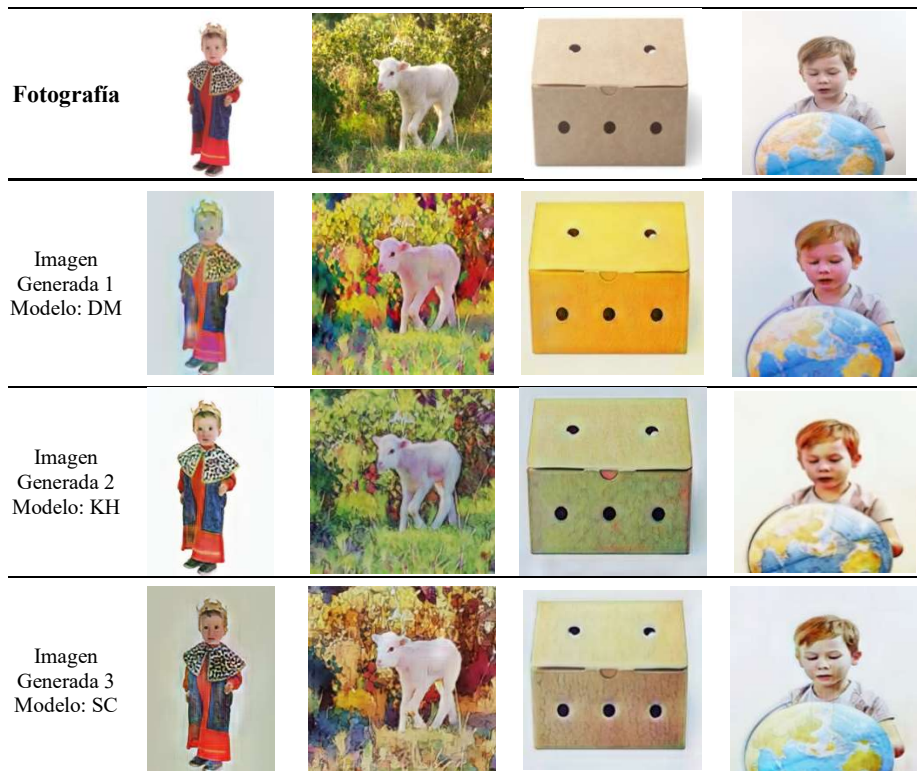


Fig. 3. Ejemplos de imágenes generadas por la GAN utilizada en [8] usando tres modelos pre-entrenados: DM, KH y SC.

Para mostrar la generación automática de ilustraciones se han seleccionado las imágenes 3, 5, 6, y 7 de la tabla de metadatos (Tabla 1).

La Fig. 2 muestra los resultados de generar tres ilustraciones por cada imagen utilizando DALL-E-2, que hace uso de la etiqueta de descripción de cada imagen, cabe mencionar que las etiquetas están en inglés debido a que la generación de imágenes es mucho mejor, los resultados muestran que la calidad visual de las imágenes generadas es muy buena, ya que los objetos que componen a la ilustración están bien definidos [18].

La generación de ilustraciones usando algunos de los modelos en [8] que utilizan una GAN se muestran en la Fig. 3. Estos modelos no necesitan imágenes relacionadas, sino solo dos conjuntos de imágenes diferentes, uno para el dominio origen y otro para el dominio destino.

Así esta GAN utiliza fotografías como dominio de origen e ilustraciones como dominio de destino. Para probar estos modelos se han utilizado fotografías que representan a las imágenes del libro ilustrado, debido a que las imágenes del libro ya son ilustraciones, se han seleccionado los tres modelos: DM (David MCKee), KH (Kevin Henkes) y SC (Stephen Cartwright) que a nuestro criterio muestran los mejores resultados y podemos ver que algunas imágenes generadas tienen algunos pequeños defectos o detalles en la generación de bordes [5, 7], incluso a veces los resultados







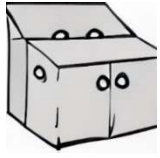





Etiqueta	Little Prince cartoon	White lamb jumping cartoon	White box with three holes cartoon	A boy walked on the planet earth
Imagen Generada 1				
Imagen Generada 2				
Imagen Generada 3				

Fig. 4. Ejemplo de imágenes generadas por Mini-DALL-E.

aparentan sólo cambiar la tonalidad de los colores y la textura, y hay ocasiones en las que se observa que el estilo ha cambiado con éxito.

Continuando con el último modelo generador de ilustraciones, la Fig. 4 muestra las ilustraciones generadas por Mini-DALL-E, podemos ver que las ilustraciones generadas son de buena calidad [15], pero a veces los bordes de los elementos que componen a la ilustración no están bien definidos, principalmente en las manos, los ojos, la textura del cabello y el rostro, pero en particular los rostros no son simétricos y las expresiones no son naturales. Concluyendo que las caras y las personas no se generadas correctamente [14].

Los tres modelos revisados en este trabajo realizan la generación automática de ilustraciones, pero hay ventajas y desventajas que se deben considerar, de esta forma DALL-E-2 genera ilustraciones con elementos bien definidos, pero la etiqueta debe ser lo más detallada posible.

Por otro lado, la GAN genera la ilustración; en ocasiones parece que no realiza ningún cambio, solo modifica colores y texturas; los resultados de las ilustraciones se obtienen en un tiempo aproximado de 1 a 2 minutos, dependiendo del número de imágenes. Finalmente, los elementos de las ilustraciones generadas por Mini-DALL-E no están bien definidos y pueden presentar ilustraciones con elementos borrosos. También es importante mencionar que la generación de imágenes es mucho mejor en el caso de DALL-E -2 y Mini-DALL-E con las etiquetas en inglés [14].

Puede ser difícil decidir qué modelo podría ser el más adecuado. La Tabla 2 proporciona una comparación de los modelos mencionados anteriormente en términos del método utilizado, el tamaño del base de datos, las capacidades del modelo y su

Tabla 2. Resumen de los modelos generativos

Modelo	Método	Tamaño de la base de datos	Código abierto	Capacidades
DALL-E 2	Zero-Shot	650M	Parcialmente	Texto a Imagen
GAN	No supervisado	Varias	Parcialmente	Imagen a Imagen
Mini DALL-E	Transformador	3M, 12M y 15M	SI	Texto a Imagen

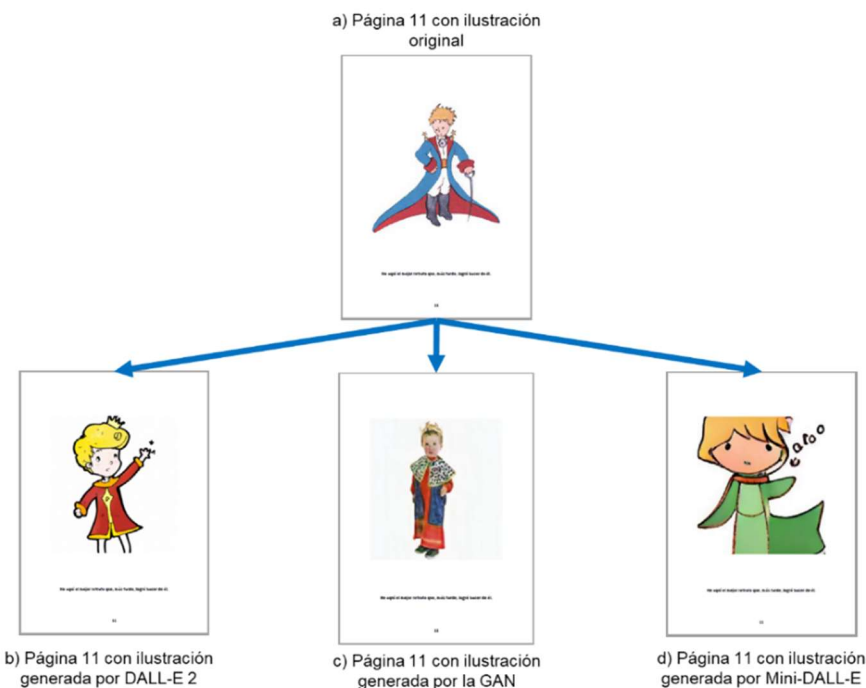


Fig. 5. Actualización de una ilustración en la página 11 del libro “El Principito”.

accesibilidad, Además, se puede acceder a estas herramientas a través de Python, pero DALL-E 2 sólo permite generar un número limitado de imágenes.

Finalmente, después de la generación de las ilustraciones se realiza la actualización de ilustraciones en el libro, generando una nueva versión del libro. La actualización de ilustraciones se realiza a través de la identificación del objeto del número de objeto de la imagen a actualizar, luego se reemplaza la imagen con la nueva imagen generada.

La Fig. 5 muestra la actualización de una ilustración por una ilustración generada por DALL-E -2 (ver Fig. 5a), la GAN (ver Fig. 5b) y Mini-DALL-E (ver Fig. 5d) de la página 11 del libro “El Principito”. La Fig. 6 muestra la actualización de las ilustraciones de la página 12 del libro “El Principito”, se han considerado una ilustración generada por DALL-E -2, dos generadas por la GAN y una generada por Mini-DALL-E.

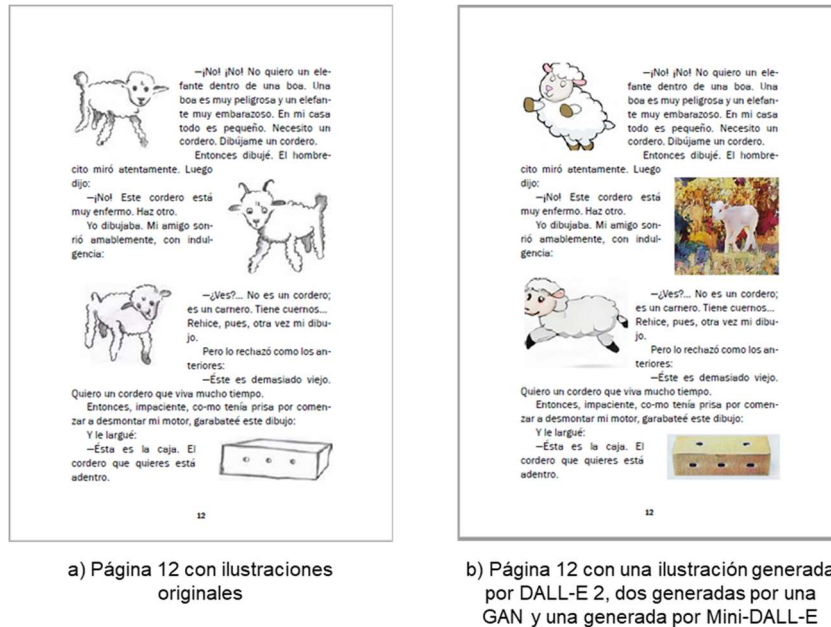


Fig. 6. Actualización de las ilustraciones en la página 12 del libro “El Principito”.

Es importante mencionar que es muy conveniente tener un banco de imágenes para generar un libro dinámico en el que de manera aleatoria cada vez que un niño lea un cuento pueda observar una gran variedad de ilustraciones diferentes.

5. Conclusiones y trabajo a futuro

La generación automática de ilustraciones es una tarea muy difícil, que se puede realizar a través del lenguaje natural o utilizando una imagen de referencia, por lo que hemos revisado algunos modelos existentes dentro de la generación automática de ilustraciones, para poder seleccionar un modelo que mejor se adapte como generador de ilustraciones para libros.

Los resultados experimentales nos indican que utilizar una GAN requiere de mucho procesamiento, ya que tarda demasiado en el proceso de entrenamiento, de esta forma si se pretende utilizar una GAN es necesario pensar en otra variante, realizando modificaciones a la arquitectura para que sea mucho más rápida o explorar la generación automática de ilustraciones utilizando sólo procesamiento digital de imágenes.

Por otro lado, la revisión de estos modelos nos hace considera la generación automática de ilustraciones utilizando lenguaje natural, ya que DALL-E 2 o Mini-DALL-E pueden generar una gran cantidad y variedad de ilustraciones, así estos modelos nos brindan la oportunidad de pensar en la creación de un libro con imágenes dinámicas, es decir, el usuario podría elegir que ilustración ver de un conjunto en un libro determinado utilizando realidad aumentada.

Es importante mencionar que es pertinente investigar la motivación de lectura de los niños y realizar muestreos en escuelas con el apoyo de pedagogos y psicólogos para ver las reacciones de los niños al leer y ver un libro con ilustraciones dinámicas. Incluso también se está considerando permitir que el niño haga sus propios dibujos y usando estas herramientas generar la ilustración.

Además, también como trabajo futuro se pretenden realizar pruebas utilizando otro tipo de libros a ilustrar, por ejemplo, considerar en generar ilustraciones a color para libros con ilustraciones en blanco y negro. Por último, se piensa en la implementación de una función discriminadora que permita seleccionar la mejor imagen generada por alguno de los modelos generadores automáticos de ilustraciones revisados en este trabajo.

Referencias

1. Fatimah, A. S., Santiana, S., Saputra, Y.: Digital comic: An innovation of using toondoo as media technology for teaching english short story. *English Review: Journal of English Education*, vol. 7, no. 2, pp. 101–108 (2019) doi: 10.25134/erjee.v7i2.1526
2. Chen, J., Liu, G., Chen, X.: AnimeGAN: A novel lightweight GAN for photo animation. *Communications in Computer and Information Science*, vol. 1205, pp. 242–256 (2020) doi: 10.1007/978-981-15-5577-0_18
3. Shu, Y., Yi, R., Xia, M., Ye, Z., Zhao, W., Chen, Y., Lai, Y., Liu, Y.: GAN-based multi-style photo cartoonization. *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 10, pp. 3376–3390 (2022) doi: 10.1109/tvcg.2021.3067201
4. Yi, Z., Zhang, H., Tan, P., Gong, M.: DualGAN: Unsupervised dual learning for image-to-image translation. In: *IEEE International Conference on Computer Vision*, pp. 2868–2876 (2017) doi: 10.1109/iccv.2017.310
5. Barzilay, N., Shalev, T. B., Giryes, R.: MISS GAN: A multi-illustrator style generative adversarial network for image to illustration translation. *Pattern Recognition Letters*, vol. 151, pp. 140–147 (2021) doi: 10.1016/j.patrec.2021.08.006
6. Pang, Y., Lin, J., Qin, T., Chen, Z.: Image-to-image translation: Methods and applications. *IEEE Transactions on Multimedia*, vol. 24, pp. 3859–3881 (2022) doi: 10.1109/tmm.2021.3109419
7. Chen, Y., Lai, Y., Liu, Y.: CartoonGAN: Generative adversarial networks for photo cartoonization. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9465–9474 (2018) doi: 10.1109/cvpr.2018.00986
8. Hicsonmez, S., Samet, N., Akbas, E., Duygulu, P.: GANILLA: Generative adversarial networks for image to illustration translation. *Image and Vision Computing*, vol. 95 (2020) doi: 10.1016/j.imavis.2020.103886
9. Tous, R.: Pictonaut: Movie cartoonization using 3D human pose estimation and GANs. *Multimedia Tools and Applications*, vol. 82, no. 14, pp. 21101–21115 (2023) doi: 10.1007/s11042-023-14556-1
10. Proven-Bessel, B., Zhao, Z., Chen, L.: ComicGAN: Text-to-comic generative adversarial network (2021) doi: 10.48550/ARXIV.2109.09120
11. Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., He, X.: AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1316–1324 (2018) doi: 10.1109/cvpr.2018.00143

12. Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H.: Generative adversarial text to image synthesis. In: Proceedings of The 33rd International Conference on Machine Learning, vol. 48, pp. 1060–1069 (2016) doi: 10.48550/arXiv.1605.05396
13. Zhang, M., Zhang, Z. L. X.: Transformers, Dall-E mini, and next level text to video. pp. 1–8 (2023) <https://www.researchgate.net/publication/369299281>
14. Dayma, B., Pcuena, Saifullah, K., Ghosh, R., Abraham, T., Abid, A., Patil, S., Khác, P. L.: Borisdayma/dalle-mini: Initial release (2021) doi: 10.5281/ZENODO.5146400
15. Bansal, H., Yin, D., Monajatipoor, M., Chang, K.: How well can text-to-image generative models understand ethical natural language interventions? In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pp. 1358–1370 (2022) doi: 10.48550/arXiv.2210.15230
16. Maerten, A., Soydaner, D.: From paintbrush to pixel: A review of deep neural networks in AI-generated art (2023) doi: 10.48550/arXiv.2302.10913
17. Marcus, G., Davis, E., Aaronson, S.: A very preliminary analysis of DALL-E 2 (2022) doi: 10.48550/arXiv.2204.13807