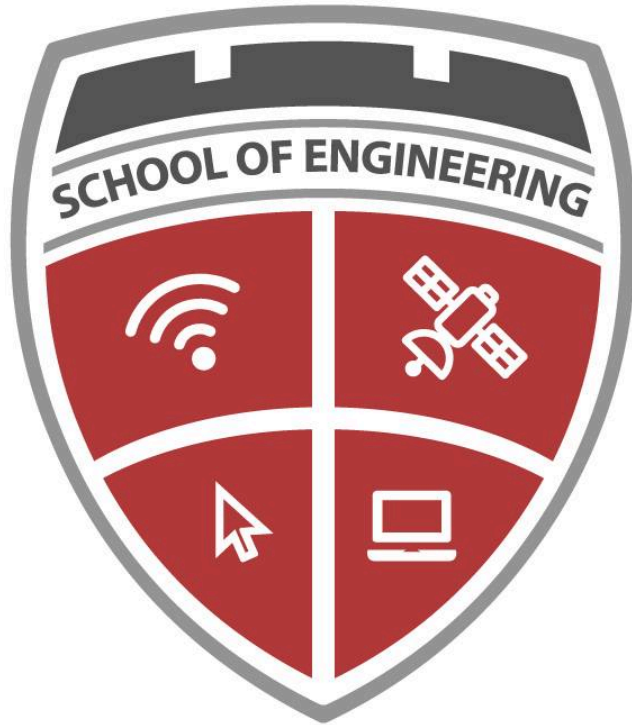


UNIVERSIDAD AUTÓNOMA GABRIEL RENÉ MORENO  
SCHOOL OF ENGINEERING SOE - UNIDAD DE POSGRADO  
MAESTRÍA EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL V1E2



MODELADO GENERATIVO DE ESPACIOS HABITACIONALES CON VAE Y  
GAN USANDO LSUN BEDROOMS  
LABORATORIO 4 - GRUPO 3

**Participantes:** ❖ Yesika Luna  
❖ Elvis Miranda  
❖ Ivan Mamani Condori

**Docente:** Msc. Danny Luis Huanca Sevilla

**Módulo:** GENERATIVE MODELS AND COMPUTER VISION

Febrero, 2026

## ÍNDICE

<b>INTRODUCCIÓN.....</b>	<b>3</b>
<b>1. DESCRIPCIÓN DEL CASO.....</b>	<b>4</b>
<b>2. OBJETIVOS.....</b>	<b>4</b>
<b>3. DATASET.....</b>	<b>5</b>
<b>4. PROBLEMA DE INVESTIGACIÓN.....</b>	<b>6</b>
<b>5. OBJETO DE ESTUDIO.....</b>	<b>7</b>
<b>6. CAMPO DE ACCIÓN.....</b>	<b>7</b>
<b>7. PREGUNTA CENTRAL DE REFLEXIÓN.....</b>	<b>8</b>
<b>8. METODOLOGÍA.....</b>	<b>9</b>
8.1. Preparación del Dataset.....	9
8.2. Modelo VAE (Variational Autoencoder).....	9
8.3. Modelo GAN (Generative Adversarial Network).....	11
Entrenamiento.....	11
Evidencia Visual.....	11
8.4. Comparación Metodológica.....	13
8.5. Consideraciones Experimentales.....	14
<b>9. RESULTADOS Y DISCUSIONES.....</b>	<b>14</b>
<b>10. CONCLUSIONES.....</b>	<b>15</b>
<b>11. ANEXOS.....</b>	<b>16</b>

## INTRODUCCIÓN

En el contexto actual de la inteligencia artificial aplicada a dominios creativos, los modelos generativos profundos han emergido como herramientas capaces de aprender representaciones complejas a partir de grandes volúmenes de datos visuales. En particular, su aplicación en el ámbito del diseño arquitectónico y de interiores abre nuevas posibilidades para automatizar procesos conceptuales que tradicionalmente dependen de la experiencia humana y de la exploración manual de alternativas. La capacidad de estos modelos para identificar patrones espaciales y estructurales en imágenes reales permite no solo analizar configuraciones existentes, sino también proponer nuevas soluciones visuales plausibles dentro de un espacio de diseño aprendido.

El presente estudio se enmarca en la exploración del potencial de arquitecturas generativas profundas, específicamente Variational Autoencoders (VAE) y Generative Adversarial Networks (GAN), para modelar la distribución probabilística de espacios habitacionales a partir del dataset LSUN Bedrooms. Este conjunto de datos, compuesto por cientos de miles de imágenes de dormitorios reales, constituye una fuente rica en variabilidad espacial, estilística y funcional. A través del entrenamiento de estos modelos, se busca aprender representaciones latentes que capturen las regularidades visuales subyacentes, permitiendo tanto la reconstrucción como la generación de nuevas configuraciones arquitectónicas.

Desde una perspectiva metodológica, el trabajo aborda el desafío de traducir estructuras espaciales implícitas en imágenes en representaciones matemáticas manipulables en un espacio latente continuo. Este proceso implica no solo la comprensión de la información visual, sino también la preservación de relaciones semánticas y geométricas que definen la coherencia de un entorno habitacional. En este sentido, la investigación no se limita a la generación de imágenes sintéticas, sino que examina hasta qué punto los modelos generativos pueden internalizar la organización estructural del espacio y utilizarla como base para la creación de nuevas configuraciones visualmente consistentes.

## 1. DESCRIPCIÓN DEL CASO

Una empresa tecnológica dedicada al diseño de interiores y visualización arquitectónica desea desarrollar un sistema inteligente capaz de generar propuestas automáticas de dormitorios realistas a partir de un espacio latente estructurado. Actualmente, el proceso de diseño conceptual depende de catálogos limitados y renderizados manuales, lo que restringe la exploración creativa y la generación de configuraciones novedosas.

La empresa ha decidido explorar modelos generativos profundos utilizando el dataset LSUN Bedrooms, que contiene cientos de miles de imágenes reales de dormitorios en diferentes configuraciones, estilos y distribuciones espaciales. El objetivo es entrenar un Variational Autoencoder (VAE) y una Generative Adversarial Network (GAN) para aprender la distribución de probabilidad de los espacios habitacionales y generar nuevas configuraciones visualmente coherentes.

El reto técnico consiste en:

- ❖ Aprender una representación latente comprimida del espacio visual (VAE).
- ❖ Modelar la distribución real de dormitorios y generar imágenes fotorrealistas (GAN).

Se busca responder si los modelos generativos pueden capturar la estructura semántica y espacial implícita en los entornos habitacionales y si dicha representación permite explorar nuevas configuraciones plausibles.

## 2. OBJETIVOS

Desarrollar e implementar modelos generativos profundos, específicamente Variational Autoencoders (VAE) y Generative Adversarial Networks (GAN), capaces de aprender la distribución visual y estructural de espacios habitacionales a partir del dataset LSUN Bedrooms, con el fin de generar nuevas configuraciones de dormitorios que sean espacialmente coherentes y visualmente plausibles; asimismo, se busca evaluar si estas arquitecturas logran capturar patrones semánticos implícitos en la organización de los entornos interiores y determinar su potencial como herramientas que apoyen

procesos creativos en el diseño arquitectónico mediante la exploración de nuevas propuestas dentro de un espacio latente aprendido.

### 3. DATASET

Para el desarrollo de este estudio se empleó el **dataset LSUN Bedrooms**, un conjunto de datos a gran escala que contiene cientos de miles de imágenes reales de dormitorios en diversas configuraciones espaciales, estilos de diseño y disposiciones de mobiliario. Este dataset resulta especialmente adecuado para tareas de modelado generativo, ya que permite a los modelos aprender patrones visuales complejos asociados a la organización de espacios habitacionales. La fuente oficial del dataset se encuentra disponible en Kaggle a través del siguiente enlace: [https://www.kaggle.com/datasets/jhoward/lsun\\_bedroom](https://www.kaggle.com/datasets/jhoward/lsun_bedroom), desde donde fue descargado e integrado al entorno de entrenamiento para alimentar tanto el modelo VAE como la GAN, permitiendo así la construcción de representaciones latentes y la generación de nuevas propuestas visuales de dormitorios.

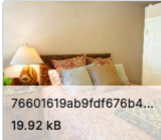
**LSUN bedroom scene 20% sample**

79<> CodeDownload

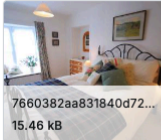
Data CardCode (8)Discussion (0)Suggestions (0)

About this directory

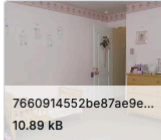
This directory does not have a description yet.



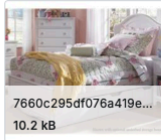
76601619ab9fd676b4...  
19.92 kB



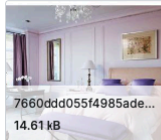
7660382aa831840d72...  
15.46 kB



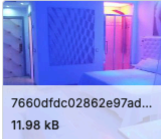
7660914552be87ae9e...  
10.89 kB



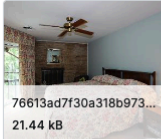
7660c295df076a419e...  
10.2 kB



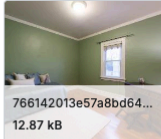
7660dd055f4985ade...  
14.61 kB



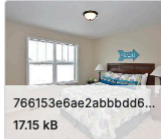
7660dfdc02862e97ad...  
11.98 kB



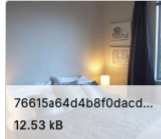
76613ad7f30a318b973...  
21.44 kB



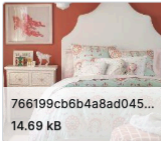
766142013e57a8bd64...  
12.87 kB




766153eae2abbbdd6...  
17.15 kB



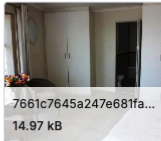
76615a64d4b8f0dacc...  
12.53 kB



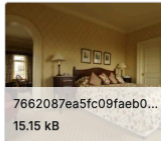
766199cb6b4a8ad045...  
14.69 kB



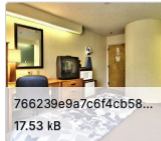
76619fcc894b7999aa...  
12.78 kB



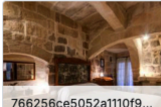
7661c7645a247e681fa...  
14.97 kB




7662087ea5fc09faeb0...  
15.15 kB



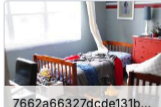
766239e9a7c6f4cb58...  
17.53 kB



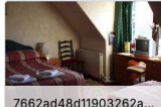
766256ce5052a1110f9...  
14.69 kB




76626784afe754d647...  
14.69 kB



7662a66327dcd131b...  
14.69 kB



7662ad48d11903262a...  
14.69 kB



7662dd61741f6b1d801...  
14.69 kB

data0lsunbedroom01234567012345667660161976603827660914

Summary303k files

#### 4. PROBLEMA DE INVESTIGACIÓN

El modelado de configuraciones espaciales de dormitorios reales mediante arquitecturas generativas profundas plantea el desafío de representar de manera probabilística la compleja distribución visual que caracteriza a los entornos habitacionales. A diferencia de tareas tradicionales de clasificación o detección, en este contexto no se busca identificar patrones explícitos, sino aprender la estructura subyacente que define la organización espacial, la disposición de objetos y la coherencia estética presente en los dormitorios reales. Esto implica capturar relaciones semánticas implícitas entre elementos como camas, iluminación, mobiliario y distribución del espacio, sin recurrir a reglas predefinidas o modelados geométricos manuales.

Desde una perspectiva probabilística, el problema consiste en aprender una función generativa que permita aproximar la distribución de probabilidad de las imágenes reales del dataset, de modo que nuevas muestras puedan ser generadas a partir de un espacio latente estructurado. Modelos como el Variational Autoencoder (VAE) permiten construir representaciones comprimidas que preservan las características esenciales del espacio visual, mientras que las Generative Adversarial Networks (GAN) facilitan la generación de imágenes con alto nivel de realismo. No obstante, el reto radica en equilibrar la capacidad de generalización del modelo con la preservación de coherencia estructural, evitando la generación de configuraciones visualmente inconsistentes o semánticamente inválidas.

En este sentido, la investigación se centra en determinar cómo estas arquitecturas generativas profundas pueden no solo reproducir la apariencia superficial de los dormitorios, sino también capturar la lógica espacial que rige su organización. La preservación del realismo visual debe ir acompañada de una coherencia estructural que refleje disposiciones plausibles del espacio, permitiendo que las configuraciones generadas mantengan sentido funcional y estético. Por tanto, el problema no se limita a la generación de imágenes sintéticamente convincentes, sino que aborda la capacidad

de los modelos para internalizar y modelar la estructura probabilística del entorno habitacional real.

## **5. OBJETO DE ESTUDIO**

El objeto de estudio de la presente investigación está constituido por la representación latente de configuraciones espaciales de dormitorios reales obtenidas a partir de modelos generativos profundos. En particular, se analiza cómo las arquitecturas de aprendizaje no supervisado, como los Variational Autoencoders (VAE) y las Generative Adversarial Networks (GAN), son capaces de aprender y codificar las características estructurales, semánticas y visuales que definen los espacios habitacionales, utilizando como base imágenes reales provenientes del dataset LSUN Bedrooms.

En este contexto, el interés se centra en comprender la forma en que estos modelos internalizan la distribución probabilística de los elementos presentes en los dormitorios, como disposición del mobiliario, iluminación, profundidad espacial y estilo visual, dentro de un espacio latente que permita la generación de nuevas configuraciones plausibles. Así, el objeto de estudio no es únicamente la imagen generada como resultado final, sino el proceso de aprendizaje de la estructura subyacente que posibilita la síntesis de entornos coherentes y visualmente consistentes.

## **6. CAMPO DE ACCIÓN**

El campo de acción de esta investigación se sitúa en la intersección entre el aprendizaje profundo generativo y la visualización arquitectónica asistida por inteligencia artificial. Específicamente, aborda la aplicación de modelos probabilísticos como los Variational Autoencoders (VAE) y las Generative Adversarial Networks (GAN) en el análisis y generación de configuraciones espaciales de entornos habitacionales, utilizando grandes volúmenes de datos visuales como base para el aprendizaje de patrones estructurales.

Desde una perspectiva tecnológica, el estudio se desarrolla dentro del ámbito del modelado generativo de imágenes, donde el objetivo no es únicamente reproducir

datos existentes, sino aprender representaciones abstractas que permitan explorar nuevas soluciones de diseño. Esto implica trabajar sobre espacios latentes que capturen relaciones semánticas entre elementos visuales, posibilitando la síntesis de dormitorios que mantengan coherencia espacial, distribución funcional y plausibilidad estética.

Asimismo, el campo de acción se extiende hacia el uso práctico de estas arquitecturas en procesos de diseño conceptual automatizado, donde la generación de propuestas visuales puede apoyar la toma de decisiones en arquitectura y diseño de interiores. En este sentido, la investigación contribuye al desarrollo de herramientas inteligentes capaces de ampliar la creatividad proyectual mediante la exploración de configuraciones espaciales no presentes explícitamente en los datos originales, pero derivadas de su estructura probabilística.

## 7. PREGUNTA CENTRAL DE REFLEXIÓN

***¿Un modelo generativo comprende la imagen o solo aprende a imitar su distribución?***

La interrogante constituye un eje conceptual fundamental dentro del estudio del aprendizaje profundo. En el contexto de arquitecturas como VAE y GAN, estos modelos no poseen una comprensión semántica en el sentido humano del término; es decir, no interpretan el significado funcional de un dormitorio ni reconocen explícitamente conceptos como comodidad, estilo o habitabilidad. En cambio, aprenden representaciones estadísticas que capturan regularidades presentes en los datos, tales como formas, texturas, disposiciones espaciales y relaciones visuales entre objetos.

No obstante, esta imitación no es meramente superficial. A través del aprendizaje de espacios latentes estructurados, los modelos generativos pueden internalizar patrones complejos que reflejan la organización implícita de los entornos habitacionales. Esto les permite generar nuevas configuraciones que, aunque no existen en el dataset original, mantienen coherencia visual y plausibilidad espacial. En este sentido, el modelo no “comprende” la imagen como lo haría un diseñador humano, pero sí



construye una representación probabilística que le permite recrear y extrapolar la lógica visual del dominio estudiado.

Por lo tanto, puede afirmarse que los modelos generativos operan en un nivel intermedio entre la mera copia y la comprensión semántica. Su capacidad radica en modelar distribuciones de datos de manera que emergen estructuras significativas sin necesidad de conocimiento explícito del mundo real. Esta reflexión permite entender que la inteligencia de estos sistemas no reside en la interpretación conceptual, sino en su habilidad para aprender y reproducir la estructura estadística del espacio visual, generando así nuevas instancias coherentes dentro del mismo dominio.

## 8. METODOLOGÍA

### 8.1. Preparación del Dataset

Se utilizó el conjunto de imágenes correspondiente a dormitorios (LSUN Bedrooms), el cual fue previamente redimensionado y normalizado en el rango  $[0,1]$ . Las imágenes fueron ajustadas a una resolución uniforme para garantizar consistencia en el entrenamiento de ambas arquitecturas generativas.

Se implementó un pipeline de carga optimizado utilizando batches, permitiendo el procesamiento eficiente en GPU dentro del entorno Google Colab.

### 8.2. Modelo VAE (Variational Autoencoder)

#### Arquitectura

El VAE implementado estuvo compuesto por:

- ❖ **Encoder:** red convolucional que proyecta la imagen al espacio latente aprendiendo la media ( $\mu$ ) y la desviación estándar ( $\sigma$ ).
- ❖ **Sampling Layer:** aplicación del truco de reparametrización.
- ❖ **Decoder:** red deconvolucional que reconstruye la imagen desde el espacio latente.

La función de pérdida combinó:

- ❖ Error de reconstrucción (MSE o BCE)
- ❖ Divergencia KL (regularización del espacio latente)

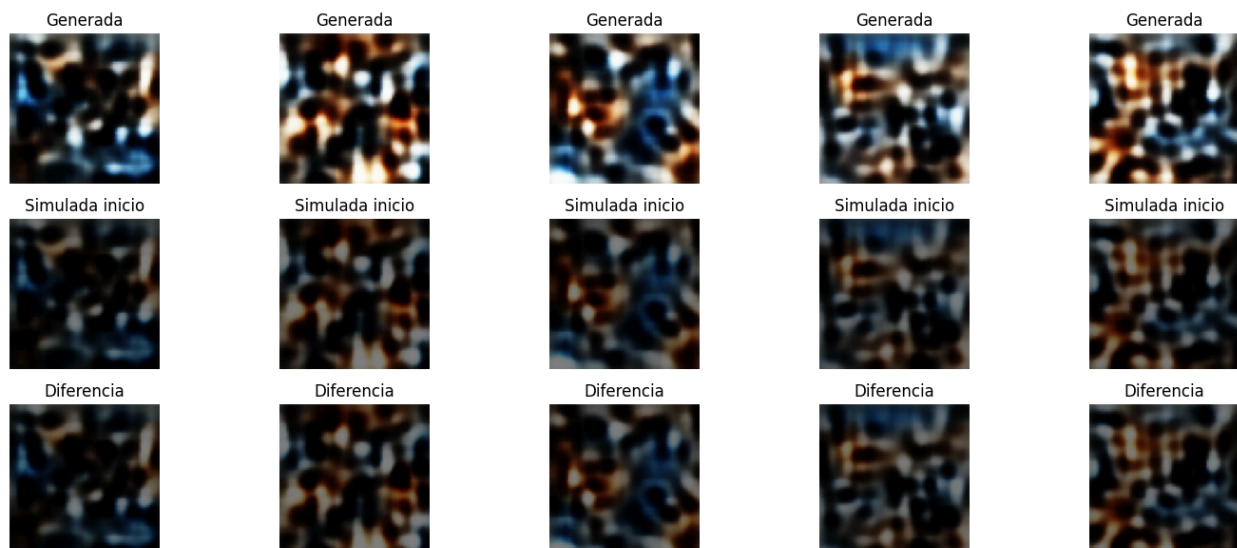
## Entrenamiento y Limitaciones

Inicialmente se programó el entrenamiento para 40 épocas; sin embargo, debido a las restricciones de uso de GPU en Google Colab, la ejecución fue interrumpida al alcanzar el límite de recursos disponibles. Esto impidió que el modelo completara el proceso de convergencia previsto.

Se observó que:

- ❖ Con 10 épocas → las imágenes generadas permanecen difusas.
- ❖ Aun con mayor entrenamiento parcial → no se alcanza el nivel de nitidez observado en la GAN.

## Evidencia Visual



Las capturas muestran:

- ❖ Imágenes con textura homogénea y baja definición.

- ❖ Bordes poco marcados.
- ❖ Estructuras globales apenas distinguibles.

Esto se explica porque el VAE tiende a producir resultados más suaves debido a la regularización impuesta por la divergencia KL, lo que favorece estabilidad probabilística pero sacrifica nitidez visual.

### 8.3. Modelo GAN (Generative Adversarial Network)

#### Arquitectura

La GAN implementada estuvo compuesta por:

- ❖ **Generador:** red neuronal profunda basada en capas densas y deconvolucionales (Conv2DTranspose), cuya función es transformar un vector latente aleatorio en una imagen sintética.
- ❖ **Discriminador:** red convolucional encargada de distinguir entre imágenes reales del dataset e imágenes generadas por el generador.

Ambos modelos fueron entrenados de forma adversarial mediante la función de pérdida binaria cruzada.

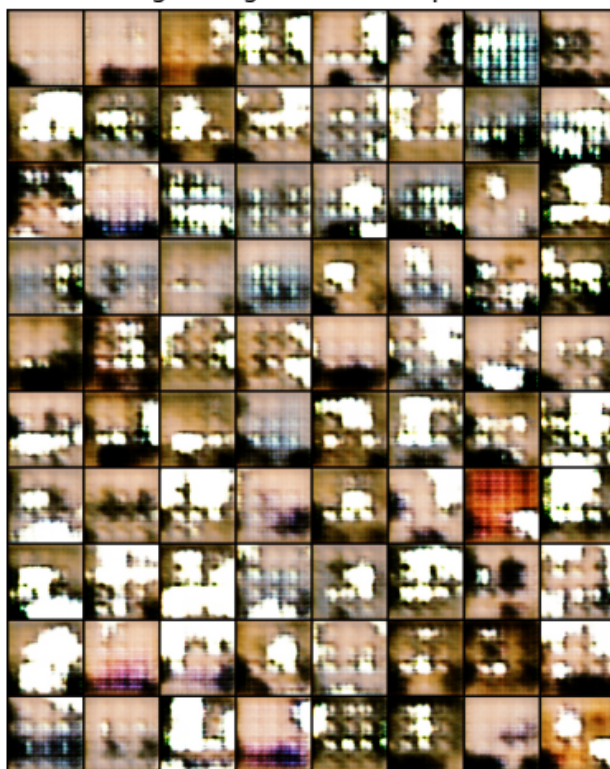
#### Entrenamiento

El entrenamiento se realizó durante múltiples épocas, observándose una evolución progresiva en la calidad visual de las imágenes generadas.

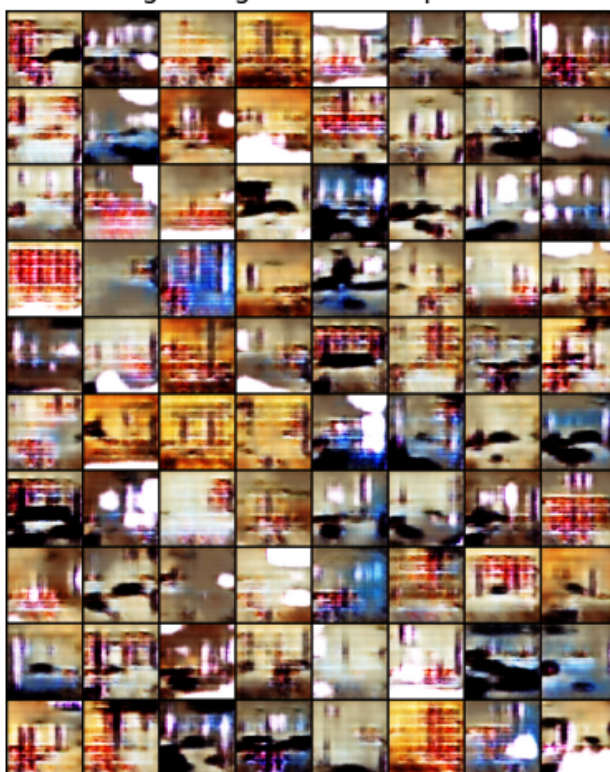
- ❖ **Epoch inicial:** las imágenes presentaban ruido estructurado sin coherencia espacial.
- ❖ **Epoch final:** se observaron patrones más definidos, mejor distribución de iluminación y mayor coherencia estructural.

#### Evidencia Visual

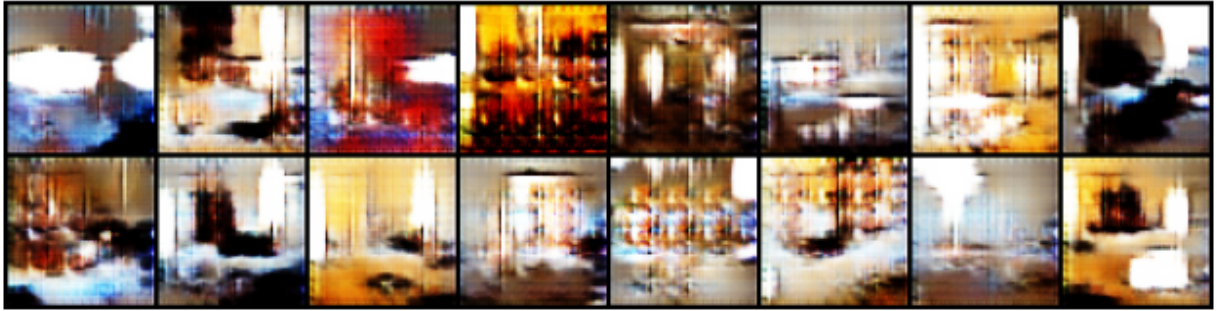
Imágenes generadas - Epoch 6



Imágenes generadas - Epoch 46



### Dormitorios generados



Las capturas comparativas muestran:

- ❖ Primera iteración → imágenes altamente borrosas y sin estructura definida.
- ❖ Última iteración → mayor organización espacial y aparición de formas similares a camas, paredes y ventanas.

Esto confirma que el generador logró aproximar la distribución real de datos mediante la dinámica competitiva con el discriminador.

#### 8.4. Comparación Metodológica

Aspecto	GAN	VAE
Tipo de aprendizaje	Adversarial	Probabilístico
Calidad visual	Mayor nitidez	Mayor suavizado
Convergencia visual	Más rápida	Más lenta
Estabilidad	Puede ser inestable	Más estable matemáticamente
Costo computacional	Alto	Muy alto (por KL + reconstrucción)

Las diferencias observadas en las capturas de imagen se justifican por la naturaleza de cada modelo:

- ❖ La GAN optimiza directamente la calidad visual mediante competencia.
- ❖ El VAE optimiza una aproximación de distribución latente regularizada.

## 8.5. Consideraciones Experimentales

Las limitaciones de hardware influyeron particularmente en el desempeño del VAE. El entorno Google Colab impuso restricciones de tiempo y uso de GPU que impidieron completar un entrenamiento extendido (>80 épocas), requerido típicamente para que el VAE alcance una representación latente estructurada en datasets complejos.

Por tanto, los resultados del VAE deben interpretarse como parciales dentro del contexto experimental disponible.

## 9. RESULTADOS Y DISCUSIONES

Los resultados obtenidos evidencian diferencias significativas en la calidad visual generada por ambas arquitecturas. En el caso de la GAN, se observó una evolución progresiva desde imágenes con ruido estructurado en las primeras épocas hacia representaciones con mayor coherencia espacial, mejor definición de bordes y patrones visuales reconocibles en las etapas finales. Las capturas comparativas muestran que el generador logró aproximar de manera efectiva la distribución del dataset, produciendo imágenes con mayor nitidez y detalles estructurales más claros.

En contraste, el VAE generó imágenes con características más difusas y menor definición, incluso tras el entrenamiento parcial alcanzado. Aunque se aprecia una ligera organización global en las estructuras, los resultados presentan suavizado excesivo y pérdida de detalle fino. Este comportamiento es consistente con la naturaleza probabilística del VAE, cuya función de pérdida incluye la divergencia KL, promoviendo un espacio latente regularizado que tiende a producir salidas más estables pero menos nítidas visualmente. Adicionalmente, el entrenamiento no pudo completarse en su totalidad debido a las restricciones de uso de GPU en el entorno de ejecución, lo que limitó su convergencia.

En términos comparativos, la GAN demostró mayor capacidad para generar imágenes visualmente realistas en un menor número de épocas, mientras que el VAE requiere un entrenamiento más prolongado para estructurar adecuadamente su espacio latente y



mejorar la calidad de reconstrucción. Por tanto, las diferencias observadas no solo responden a la arquitectura subyacente de cada modelo, sino también a las condiciones experimentales y recursos computacionales disponibles durante la implementación.

## 10. CONCLUSIONES

El desarrollo e implementación de los modelos generativos permitió evidenciar diferencias sustanciales en su comportamiento y calidad de generación. La GAN demostró una mayor capacidad para producir imágenes visualmente más realistas y estructuralmente coherentes en un menor número de épocas, gracias a su mecanismo de entrenamiento adversarial. La competencia entre generador y discriminador favoreció una mejora progresiva en la nitidez, definición de bordes y organización espacial de las imágenes.

Por su parte, el VAE mostró un desempeño más estable desde el punto de vista probabilístico, pero con resultados visualmente más difusos. La regularización impuesta por la divergencia KL, aunque beneficiosa para estructurar el espacio latente y garantizar continuidad en la representación, conlleva una pérdida de detalle fino en la generación. Además, las limitaciones computacionales del entorno de ejecución impidieron completar un entrenamiento extendido, lo que restringe la evaluación completa de su potencial en este experimento.

En conclusión, mientras la GAN resulta más adecuada cuando el objetivo principal es la calidad visual y el realismo perceptual, el VAE ofrece ventajas en términos de estabilidad matemática y estructuración del espacio latente. La elección entre ambos modelos debe considerar no solo la calidad visual deseada, sino también los recursos computacionales disponibles y el propósito específico de la aplicación generativa.

## 11. ANEXOS

Se adjunta notebooks en Python con el script aplicado al laboratorio en el siguiente repositorio:

<https://github.com/mc-ivan/zero-shot/tree/main/lab4>