

# GANdela

Alfredo Montero López  
Leonardo Amaro Rodríguez  
Anthuan Montes de Oca  
Yisell Martínez Noa

## Índice

<b>1. Introducción</b>	<b>2</b>
1.1. Motivación . . . . .	2
1.2. Problemática . . . . .	2
1.3. Objetivos Generales . . . . .	2
1.4. Objetivos Específicos . . . . .	2
<b>2. Estado del Arte / Preliminares</b>	<b>3</b>
2.1. Redes Generativas Adversarias (GANs) . . . . .	3
2.2. Pix2Pix GAN . . . . .	4
2.3. Aplicaciones de Pix2Pix GAN y CycleGan en Cartografía . . . . .	4
2.4. Variational AutoEncoders (VAEs) . . . . .	4
2.5. VAE-GAN . . . . .	4
2.6. Multimodal I2I . . . . .	5
2.7. Redes de Convolución Profunda (CNN) - DosGAN . . . . .	5
2.8. Modelos difusos . . . . .	5
2.9. Desafíos y Limitaciones . . . . .	5
<b>3. Propuesta de Solución</b>	<b>6</b>
3.1. ¿Qué es pix2pix-GAN? . . . . .	6
<b>4. Base de Datos</b>	<b>7</b>
<b>5. Resultados de las pruebas</b>	<b>7</b>
5.1. Intentos de configuración . . . . .	8
5.1.1. Root mean square error . . . . .	9
5.2. Modelo final . . . . .	9
<b>6. Discusión de los resultados</b>	<b>9</b>
6.1. Repercusión ética de las soluciones . . . . .	9
6.1.1. Privacidad y Seguridad . . . . .	9
6.1.2. Transparencia y Explicabilidad . . . . .	10

6.1.3. Equidad y Accesibilidad . . . . .	10
6.1.4. Impacto Ambiental . . . . .	10
6.1.5. Uso Responsable de la Tecnología . . . . .	10
<b>7. Conclusiones y trabajo futuro</b>	<b>11</b>
<b>8. Bibliografía</b>	<b>11</b>

# 1. Introducción

## 1.1. Motivación

La generación precisa de imágenes territoriales con representaciones de altura detalladas es crucial para una variedad de aplicaciones, desde la planificación urbanística y la ingeniería civil hasta la gestión de recursos naturales y la respuesta a desastres. Sin embargo, los métodos tradicionales de generación de modelos de elevación del terreno (DEMs) suelen ser limitados en su capacidad para capturar la complejidad del relieve terrestre. En particular, las técnicas actuales carecen de la precisión necesaria para representar pendientes pronunciadas y otros elementos topográficos críticos. La motivación principal de este proyecto radica en la necesidad de superar estas limitaciones mediante la implementación de técnicas avanzadas de inteligencia artificial, específicamente el uso de redes generativas adversarias (GANs). Se necesita poder generar imágenes territoriales realistas con la representación de las alturas de cada punto en el mapa.

## 1.2. Problemática

Los algoritmos matemáticos encargados de realizar operaciones similares, debido a su naturaleza, no pueden generar pendientes muy pronunciadas, lo cual es esencial para una representación fiel del terreno. Esta limitación motiva la búsqueda de métodos alternativos capaces de superar esta barrera. Las redes generativas adversariales (GANs) han demostrado un potencial significativo en la generación de imágenes realistas, lo que las convierte en una opción prometedora para abordar este desafío.

## 1.3. Objetivos Generales

El objetivo principal de este proyecto es la creación de una GAN (Generative Adversarial Network) para la recreación de mapas con elevaciones del terreno a partir de imágenes satelitales.

## 1.4. Objetivos Específicos

1. Desarrollar una arquitectura de GAN adaptada para la generación de modelos de elevación del terreno:

- Diseñar y entrenar una pix2pix-GAN para la traducción de imágenes satelitales en mapas de elevación.
  - Evaluar la efectividad del modelo en la generación de pendientes pronunciadas y otros elementos topográficos complejos.
- 2. Recopilar y preparar un conjunto de datos de alta calidad para el entrenamiento del modelo:**
- Obtener imágenes satelitales y datos de elevación del terreno de fuentes confiables como Google Earth Engine.
  - Preprocesar los datos para asegurar su compatibilidad y adecuación para el entrenamiento de la GAN.
- 3. Validar y evaluar el desempeño del modelo generado:**
- Realizar pruebas extensivas del modelo en diversas áreas geográficas para asegurar su robustez y generalización.
  - Comparar los modelos de elevación generados con datos de referencia y evaluar su precisión.

## 2. Estado del Arte / Preliminares

Una revisión exhaustiva de la literatura sobre el uso de redes generativas adversariales para la traducción de imágenes y la generación de modelos de elevación del terreno es esencial para contextualizar este proyecto. Se analizarán trabajos previos y metodologías relacionadas para identificar las mejores prácticas y los desafíos más comunes.

En los últimos años, el uso de redes generativas adversarias (GANs) ha revolucionado el campo de la inteligencia artificial, particularmente en aplicaciones de generación y traducción de imágenes. Introducidas por Goodfellow et al. en 2014 [1], las GANs han demostrado ser eficaces en la creación de imágenes realistas a partir de datos de entrenamiento, superando las limitaciones de los métodos tradicionales. Se tuvieron en cuenta varias estrategias para realizar el proyecto, las cuales se expondrán, pero la escogida fue Pix2Pix.

### 2.1. Redes Generativas Adversarias (GANs)

Las GANs consisten en dos redes neuronales que compiten entre sí: el generador (G), que crea datos falsos a partir de un ruido aleatorio, y el discriminador (D), que intenta distinguir entre datos reales y generados. Esta configuración adversarial mejora continuamente ambos modelos, resultando en la generación de datos altamente realistas. Las aplicaciones de las GANs abarcan desde la generación de imágenes y videos hasta la síntesis de voz y la creación de obras de arte [2].

## 2.2. Pix2Pix GAN

Una de las variantes más destacadas de las GANs es la *pix2pix* GAN, presentada por Isola et al. en 2017 [3]. Esta arquitectura se utiliza específicamente para la traducción de imágenes de un dominio a otro, manteniendo la estructura y el contexto de la imagen de entrada. A diferencia de las GANs tradicionales, el modelo *pix2pix* emplea una red U-Net como generador y una red PatchGAN como discriminador. La U-Net permite la transferencia de información a diferentes escalas, mientras que la PatchGAN evalúa la autenticidad de pequeños parches de la imagen, en lugar de la imagen completa.

## 2.3. Aplicaciones de Pix2Pix GAN y CycleGan en Cartografía

El uso de *pix2pix* GANs en cartografía y geografía ha abierto nuevas posibilidades para la generación de modelos de elevación del terreno (DEMs) y la simulación de paisajes naturales. Trabajos como los de Isola et al. en 2017 [3] y Zhu et al. [4] han demostrado la capacidad de las GANs para traducir imágenes aéreas en mapas de calles y viceversa. Estos avances han permitido la creación de mapas más precisos y detallados, superando las limitaciones de los métodos tradicionales basados en interpolación y aproximaciones matemáticas. Métodos como CycleGAN son cruciales para situaciones donde no se dispone de pares exactos de imágenes en los dos dominios. La idea es aprender una función de mapeo utilizando ciclos de consistencia para asegurar la calidad de la traducción.

## 2.4. Variational AutoEncoders (VAEs)

Los Variational AutoEncoders [5] utilizan una estructura encoder-decoder para aprender la representación latente de las imágenes en diferentes dominios. Los autoencoders variacionales se pueden usar en I2I debido a su capacidad de representar distribuciones complejas de datos, como las variaciones de terreno en imágenes de satélite. Este enfoque permite aprender una representación compacta de los datos de entrada (imágenes satelitales) que puede ser decodificada en una matriz de alturas correspondiente, proporcionando una aproximación robusta al manejo de la variabilidad y la incertidumbre inherente en este tipo de datos. En comparación con los GANs, los VAEs tienden a generar imágenes menos nítidas y de menor calidad, ya que su objetivo es maximizar la probabilidad de los datos, lo que a menudo lleva a soluciones más suaves o "borrosas".

## 2.5. VAE-GAN

Es un enfoque híbrido que combina la capacidad de generación de VAEs con la capacidad de discriminación de GANs. La estructura VAE-GAN [6] puede generar múltiples salidas para una misma entrada, lo que es útil en tareas multimodales pero la calidad de las imágenes no es la deseada.

## 2.6. Multimodal I2I

Un desafío importante en la traducción de imágenes satelitales a matrices de alturas es que un solo conjunto de imágenes satelitales puede tener múltiples posibles soluciones en términos de matrices de alturas, dependiendo de factores como la iluminación o la época del año. Los enfoques multimodales, como MUNIT [7] y DRIT [8], se centran en generar múltiples resultados plausibles para una misma imagen de entrada. Estos modelos aprenden varias representaciones, separando la información de contenido (estructura general de la imagen) de la información de estilo (características específicas como la iluminación), lo cual es crucial para obtener diversas predicciones de alturas desde una misma imagen satelital. Tanto MUNIT como DRIT son más apropiados para tareas donde la variabilidad en el estilo de salida es importante, mientras que Pix2Pix están más enfocados en traducciones donde la precisión y la fidelidad de la salida a la entrada son críticas.

## 2.7. Redes de Convolución Profunda (CNN) - DosGAN

Las redes neuronales convolucionales (CNN) son fundamentales para el análisis de imágenes satelitales, ya que pueden extraer características relevantes como la topografía y las texturas del terreno. Combinadas con GANs o VAEs, estas redes pueden permitir transformar imágenes bidimensionales en matrices de alturas que capturan las características tridimensionales del terreno. Un ejemplo es su uso en DosGAN [9], útil en tareas multimodales.

## 2.8. Modelos difusos

DiffusionSat [10] es un modelo diseñado para la generación de imágenes satelitales. Puede manejar problemas inversos, incluyendo tareas como la super-resolución y la predicción temporal, mediante el condicionamiento de las señales de control de entrada, como una imagen satelital de baja resolución, para producir imágenes de mayor resolución o interpoladas temporalmente. Esto sugiere potencial para la generación de mapas de altura mediante el condicionamiento de datos satelitales. Otro trabajo relevante es DiffI2I [11]. Permite obtener resultados precisos y emplear una red de eliminación de ruido más ligera y con menos iteraciones. En concreto, DiffI2I consta de tres componentes clave: una red compacta de extracción previa I2I (CPEN), un transformador dinámico I2I (DI2Iformer) y una red de eliminación de ruido. DiffI2I o DiffusionSat parecen ser excelentes opciones, debido a su capacidad para generar detalles finos y manejar datos de alta complejidad, aunque son muy lentos y demandantes en términos de recursos, por lo que sigue pareciendo mejor opción Pix2Pix para la tarea actual.

## 2.9. Desafíos y Limitaciones

A pesar de sus éxitos, las GANs y, en particular, las *pix2pix* GANs, enfrentan varios desafíos. La necesidad de grandes cantidades de datos de entrenamiento

de alta calidad, el riesgo de sobreajuste, y la inestabilidad durante el entrenamiento son problemas comunes. Además, la capacidad de las GANs para generar pendientes muy pronunciadas en modelos de elevación sigue siendo limitada. Sin embargo, los avances continuos en técnicas de entrenamiento y arquitectura de redes neuronales prometen mitigar estos desafíos en el futuro [12].

En resumen, la aplicación de *pix2pix* GANs en la generación de imágenes territoriales realistas representa un avance significativo en la cartografía moderna. La investigación actual se centra en mejorar la precisión y eficiencia de estos modelos, con el objetivo de crear representaciones del terreno cada vez más fieles a la realidad.

### 3. Propuesta de Solución

Se implementó una *pix2pix*-GAN que, dada una imagen satelital, genera una matriz de alturas del terreno. Los detalles técnicos de la implementación son los siguientes:

- **Input:** Imagen satelital de 3x256x256 (RGB) en formato GeoTIFF.
- **Output:** Mapa de alturas de 1x256x256.

Dada la imagen satelital y su combinación de colores (RGB), se calcula el tipo de terreno correspondiente. Cada archivo (imagen y altura) se transforma y se ajusta a dimensiones de 256x256, ya que ambos deben tener la misma dimensión debido a la naturaleza del algoritmo. Posteriormente, la matriz resultante del modelo se transforma a 32x32 para cumplir con las restricciones del problema.

#### 3.1. ¿Qué es pix2pix-GAN?

Pix2Pix es un tipo de modelo de red generativa adversaria (GAN) diseñado para tareas de traducción de imágenes, es decir, convertir una imagen de un dominio a otro. Fue introducido en 2017 en el trabajo titulado "Image-to-Image Translation with Conditional Adversarial Networks". A diferencia de las GANs tradicionales, que generan imágenes a partir de un vector de ruido aleatorio, Pix2Pix toma una imagen de entrada y produce una imagen de salida correspondiente, lo que lo hace adecuado para aplicaciones como la conversión de bocetos en imágenes, la colorización de imágenes en blanco y negro, la conversión de fotos de día a noche, entre otros.

La arquitectura de Pix2Pix consta de dos componentes principales:

- **Generador (G):** Toma una imagen de entrada y genera una imagen que se espera se asemeje a una imagen realista del dominio objetivo. En Pix2Pix, el generador suele ser una red de tipo "U-Net", una arquitectura de red neuronal que permite la propagación de contextos a través de las diferentes escalas de la imagen.

- **Discriminador (D):** Su función es distinguir entre las imágenes reales del dominio objetivo y las imágenes generadas por el generador. El discriminador en Pix2Pix es una CNN (red neuronal convolucional) que clasifica segmentos de la imagen (en lugar de la imagen completa) como real o falsa, lo cual se conoce como PatchGAN.

El entrenamiento de Pix2Pix se realiza de manera adversaria: el generador intenta producir imágenes que el discriminador no pueda distinguir de las reales, mientras que el discriminador se mejora para distinguir entre imágenes reales y generadas. Este proceso se guía mediante una función de pérdida que combina la pérdida adversaria (para engañar al discriminador) y la pérdida L1 (para penalizar las diferencias pixel a pixel entre la imagen generada y la imagen real objetivo), ayudando así a producir resultados más fieles al dominio objetivo.

El resultado es un sistema capaz de transformar imágenes de un dominio a otro manteniendo la estructura y el contexto de la imagen de entrada, lo que lo hace muy efectivo para una amplia gama de aplicaciones de traducción de imágenes.

## 4. Base de Datos

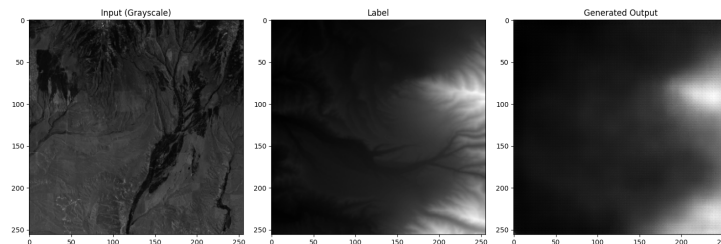
La base de datos se obtuvo de Google Earth Engine:

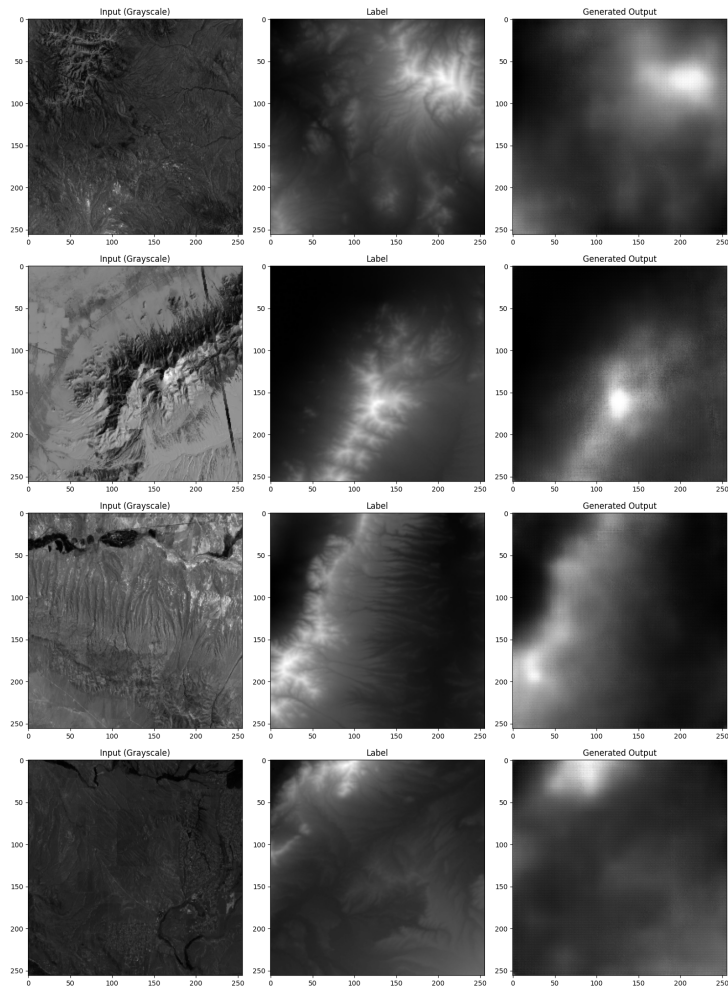
- **Imágenes Satelitales:** LANDSAT/LC09/C02/T1.
- **Alturas:** CGIAR/SRTM90\_V4.

La extracción de datos se realizó mediante la API de Google Earth Engine, con los datos almacenados en la cuenta de PixelCampione y descargados en la carpeta del *dataset* para su posterior uso. La región se dividió en partes debido a las restricciones de la API, requiriendo múltiples llamados a la API.

## 5. Resultados de las pruebas

El modelo fue probado con diferentes funciones de pérdidas tanto para el Discriminador como para el Generador. Para la evaluación visual del modelo, se tienen las imágenes siguientes:





### 5.1. Intentos de configuración

El problema principal de estos modelos es que durante las primeras iteraciones, el Discriminador era capaz de superar fácilmente al Generador, provocando que estuviera muy seguro de si la entrada era falsa o no. Dándole valores como 0.999 a las entradas reales y 0.001 a las falsas, restringiendo el aprendizaje del Generador.

Mejores parámetros:

Epochs = 700  
 Step Ahead = 3  
 Generator Lr =  $2e-4$



Discriminator Lr =  $2e-4$

Batch Size = 16

Beta1 = 0.5

### 5.1.1. Root mean square error

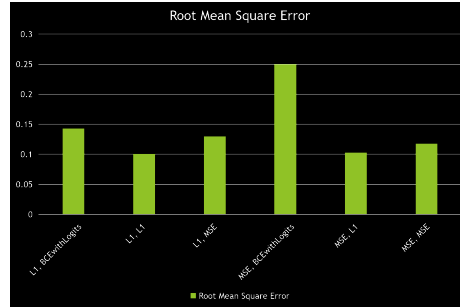


Figura 1: Root mean square error

## 5.2. Modelo final

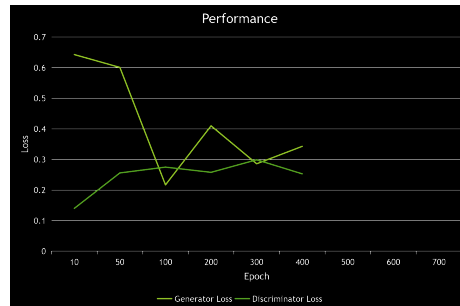


Figura 2: Performance

## 6. Discusión de los resultados

### 6.1. Repercusión ética de las soluciones

#### 6.1.1. Privacidad y Seguridad

El uso de imágenes satelitales y datos geoespaciales puede plantear preocupaciones de privacidad, especialmente si se utilizan imágenes que incluyen áreas

habitadas o sensibles. Es esencial asegurarse de que los datos utilizados provienen de fuentes autorizadas y que se cumplen todas las normativas de privacidad y protección de datos vigentes. Además, los resultados generados por las GANs no deben comprometer la seguridad de individuos o comunidades, evitando la divulgación de información que podría ser utilizada con fines malintencionados.

#### **6.1.2. Transparencia y Explicabilidad**

Las técnicas basadas en IA, particularmente las GANs, son conocidas por su naturaleza de “caja negra”, donde los procesos internos pueden ser opacos y difíciles de interpretar. Esta falta de transparencia puede generar desconfianza entre los usuarios y limitar la aceptación de la tecnología. Es fundamental trabajar en la explicabilidad de los modelos, proporcionando información clara sobre cómo se generan los mapas de elevación y qué datos se utilizan. Los usuarios deben ser conscientes de las limitaciones y posibles errores del sistema para tomar decisiones informadas.

#### **6.1.3. Equidad y Accesibilidad**

La implementación de estas tecnologías debe asegurar que sus beneficios se distribuyen de manera equitativa y que no perpetúan o agravan desigualdades existentes. Por ejemplo, la disponibilidad y calidad de las imágenes satelitales pueden variar entre regiones, lo que podría resultar en una representación desigual del terreno y afectar negativamente a las comunidades menos favorecidas. Es crucial desarrollar políticas que garanticen un acceso equitativo a los recursos y tecnologías, así como mecanismos para corregir posibles sesgos en los datos.

#### **6.1.4. Impacto Ambiental**

El entrenamiento de modelos de IA, especialmente aquellos que requieren grandes cantidades de datos y poder computacional, puede tener un impacto ambiental significativo debido al consumo de energía. Es importante considerar la sostenibilidad y buscar formas de minimizar el impacto ambiental, como la optimización de los algoritmos y el uso de fuentes de energía renovable. Fomentar prácticas de desarrollo sostenibles es esencial para reducir la huella ecológica de estas tecnologías.

#### **6.1.5. Uso Responsable de la Tecnología**

Finalmente, es crucial promover un uso responsable de las GANs en la generación de DEMs. Los resultados de estos modelos deben ser utilizados con fines éticos y beneficiosos para la sociedad, como la planificación urbanística, la gestión de recursos naturales y la respuesta a desastres. Se debe evitar el uso de esta tecnología para actividades que puedan causar daño o explotación indebida de los recursos naturales o de las comunidades.

## 7. Conclusiones y trabajo futuro

Para mejorar y expandir este proyecto, se proponen las siguientes ideas:

1. **Entrenamiento de un modelo para clasificación de terrenos:** Se puede entrenar otra instancia del modelo donde la salida sea el tipo de terreno correspondiente. Esto se puede realizar de manera que a cada terreno le corresponda un número. Incluso se puede modificar el número asignado para que terrenos similares tengan representaciones numéricas cercanas. Luego se tendrían dos GAN, una para la generación de alturas y otra para la clasificación de terrenos.
2. **Generación de mapas de alturas sin imágenes satelitales:** Utilizando imágenes satelitales y su clasificación según el tipo de terreno, se podría generar un mapa de alturas sin necesidad de ingresar una imagen satelital. Esto implicaría la creación de vectores representativos de cada mapa en la base de datos, considerando la presencia de diferentes características del terreno. La vectorización de la consulta permitiría obtener mapas relacionados, facilitando la redacción e interpretación de consultas mediante técnicas de procesamiento de lenguaje natural o incluso usando ChatGPT.

## 8. Bibliografía

- **Pix2Pix:** <https://arxiv.org/abs/1611.07004>
- **GAN definition:** <https://arxiv.org/abs/1406.2661>
- **Google Earth catalog:** <https://developers.google.com/earth-engine/datasets/catalog>
- **Earth Engine API:** <https://developers.google.com/earth-engine/guides>
- **Dataset in personal Drive:** [https://drive.google.com/drive/folders/1RnP7gD8rTcWRNzTr9LC65Z3RDC\\_HgxgX](https://drive.google.com/drive/folders/1RnP7gD8rTcWRNzTr9LC65Z3RDC_HgxgX)

## Referencias

- [1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. *Generative Adversarial Nets*. In Advances in Neural Information Processing Systems, 2014.
- [2] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. *Generative Adversarial Networks: An Overview*. IEEE Signal Processing Magazine, 2018.

- [3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. *Image-to-Image Translation with Conditional Adversarial Networks*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [4] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*. In Proceedings of the IEEE International Conference on Computer Vision, 2017.
- [5] M.-Y. Liu, T. Breuel, and J. Kautz, “Unsupervised image-to-image translation networks,” in Advances in neural information processing systems, 2017, pp. 700–708.
- [6] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, “Auto-encoding beyond pixels using a learned similarity metric,” in International conference on machine learning. PMLR, 2016, pp. 1558–1566.
- [7] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, “Multimodal unsupervised image-to-image translation,” in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 172–189.
- [8] H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M. Singh, and M.-H. Yang, “Diverse image-to-image translation via disentangled representations,” in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 35–51.
- [9] J. Lin, Z. Chen, Y. Xia, S. Liu, T. Qin, and J. Luo, “Exploring explicit domain supervision for latent space disentanglement in unpaired image-to-image translation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2019.
- [10] Samar Khanna, Patrick Liu, Linqi Zhou, Chenlin Meng, Robin Rombach, Marshall Burke, David Lobell, Stefano Ermon, “DiffusionSat: A Generative Foundation Model for Satellite Imagery,” in ICLR 2024
- [11] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, Radu Timotfe, Luc Van Gool, “DiffI2I: Efficient Diffusion Model for Image-to-Image Translation,” in 2023
- [12] Karol Kurach, Mario Lucic, Xiaohua Zhai, Marcin Michalski, and Sylvain Gelly. *A Large-Scale Study on Regularization and Normalization in GANs*. In Proceedings of the 36th International Conference on Machine Learning, 2019.