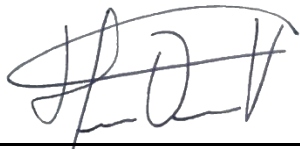


Estimación de imágenes NDVI mediante fusión de imágenes satelitales multiespectrales y SAR para la planeación agrícola

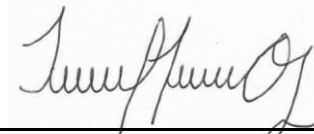
David Ramiro Bastidas Torres

Nota de Aceptación

Certificamos que el presente Trabajo de Grado satisface, en alcances y calidad, todos los requisitos que demanda un Trabajo de Grado de Maestría.



Hernán Darío Vargas Cardona  
Director



Diego Hernán Peluffo Ordoñez  
Co-Director

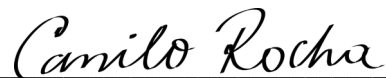


Hernán Darío Benítez Restrepo  
Jurado



Luis Eduardo Tobón Llano  
Jurado

Aprobado en cumplimiento de los requisitos exigidos por la Pontificia Universidad Javeriana Cali, para optar el título de Magister en Ingeniería.



HERNÁN CAMILO ROCHA NIÑO Ph. D.  
Decano Facultad de Ingeniería y Ciencias



JUAN CARLOS MARTÍNEZ ARIAS  
Director Posgrados de Ingeniería y Ciencias

Cali, septiembre 23 de 2021

**Acta de Correcciones al Documento de Trabajo de Grado**

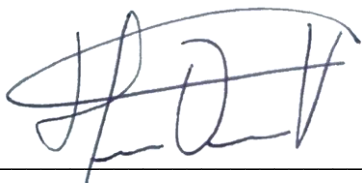
**Santiago de Cali, 25 de septiembre de 2021**

**Autor: David Ramiro Bastidas Torres**

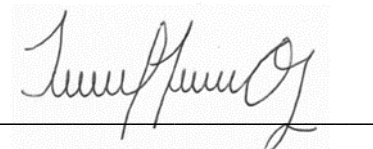
**Título del Trabajo de Grado: “Estimación de imágenes NDVI mediante fusión de imágenes satelitales multiespectrales y SAR para la planeación agrícola”**

**Director: Hernán Darío Vargas Cardona  
Co-Director: Diego Hernán Peluffo Ordóñez**

Como indica el artículo 2.13 de las Directrices para Trabajo de Grado de Maestría, he verificado que el estudiante indicado arriba ha implementado todas las correcciones que los Jurados del Proyecto de Trabajo de Grado definieron que se efectuaran, como consta en el Acta de Evaluación correspondiente.



Firma del Director del Trabajo de Grado



Firma del Co-Director del Trabajo de Grado

## **Datos del estudiante**

**Nombre Completo:** David Ramiro Bastidas Torres

**Dirección:** Cra 16 # 21-06, Pasto, Nariño

**Correo Electrónico:** davba@javerianacali.edu.co

**Número de contacto:** 3145696178

**Profesión:** Ingeniero electrónico

**Empresa:** Ninguna

Pontificia Universidad Javeriana - Cali  
Facultad de Ingeniería y Ciencias  
Maestría en Ingeniería - Énfasis Electrónica  
Proyecto de trabajo de Grado

# Estimación de imágenes NDVI mediante fusión de imágenes satelitales multiespectrales y SAR para la planeación agrícola

David Ramiro Bastidas Torres

Julio, 2021





**ESTIMACIÓN DE IMÁGENES NDVI MEDIANTE FUSIÓN DE IMÁGENES  
SATELITALES MULTIESPECTRALES Y SAR PARA LA PLANEACIÓN AGRÍCOLA**

David Ramiro Bastidas Torres

Proyecto de grado presentado como requisito para optar al título de:  
Magíster en Ingeniería

Director:

Dr. Hernán Darío Vargas Cardona

Co-Director:

Dr. Diego Hernán Peluffo Ordóñez

**Pontificia Universidad Javeriana  
Facultad de Ingeniería y Ciencias  
Cali, Colombia  
2021**

# Resumen

En la actualidad, la mala planificación en la agricultura es causante del 80 % de la deforestación mundial, debido al alto impacto que genera ésta actividad sobre el suelo. En el departamento de Nariño, los problemas ambientales vienen acompañados con problemas en la productividad, originada en una mayoría por la falta de tecnificación. La incorporación de nuevas tecnologías surge como solución para mejorarla planeación de cultivos fundamental en el desarrollo de una agricultura sostenible. Dentro de estas tecnologías las imágenes satelitales multiespectrales (MSI) con el uso del índice de vegetación normalizado (NDVI), han tomado una relevancia significativa debido a que proveen información sobre la vigorosidad de la vegetación, la cual puede ayudar al análisis de diferentes variables en los cultivos. La principal limitación de las imágenes MSI se da en condiciones de alta nubosidad, en donde el satélite es incapaz de captar información por su sensor óptico, esto impide utilizar dicha información de manera continua, clave en la vigilancia de cultivos. En esta investigación se da como solución de este problema utilizar Generative adversarial networks (GANs) mediante su técnica de translación de imagen a imagen. Las arquitecturas Unet, Unet++ y ResNet fueron implementadas como generador de la GAN, con el objetivo de identificar que combinación presenta mejores resultados en las métricas SSIM (Structural Similarity Index), RMSE(root-mean-square error), PSNR (Peak signal-to-noise ratio) y SAM (spectral angle mapper range). Como resultado destacado, las redes Unet++ superan las ResNet y el modelo UNet pues reducen el RMSE de 0.028 a 0.015.

## **Palabras Clave:**

- Fusión
- índice de vegetación normalizado (NDVI)
- Imágenes multiespectrales
- Imágenes satelitales SAR
- Redes neuronales generativas adversarias
- Autoencoders

# Abstract

Currently, poor agricultural planning is responsible for 80 % of the world's deforestation, due to the high impact of this activity on the soil. Nariño department, environmental problems are accompanied by productivity problems, mostly caused by the lack of technology. The incorporation of new technologies has emerged as a solution to improve crop planning, which is fundamental for the development of sustainable agriculture. Among these technologies, multispectral satellite images (MSI) with the use of the normalized vegetation index (NDVI) have become significantly relevant because they provide information on vegetation vigorousness, which can help in the analysis of different crop variables. The main limitation of MSI images occurs in conditions of high cloud cover, where the satellite is unable to capture information by its optical sensor, this prevents the use of such information continuously, which is important in crop monitoring. In this research, the solution to this problem is to use Generative adversarial networks (GANs) through its image-to-image translation technique. The Unet, Unet++ and ResNet architectures were implemented as GAN generators, with the objective of identifying which combination presents the best results in the metrics SSIM (Structural Similarity Index), RMSE (root-mean-square error), PSNR (Peak signal-to-noise ratio) y SAM (spectral angle mapper range). As an outstanding result, the Unet++ networks outperform ResNet and the U-Net model as they reduce the RMSE from 0.028 to 0.015.

## **Keywords:**

- Fusion
- Normalized vegetation index (NDVI)
- Multispectral images
- SAR satellite images
- Adverse generative neural networks
- Autoencoders

# Agradecimientos

Primero, agradezco a Dios por concederme todos los elementos necesarios para culminar mi trabajo de investigación. A mis padres y hermanos por su confianza, y porque siempre estuvieron presentes con sus palabras de apoyo y expresiones de cariño, que alimentaron mi corazón permitiéndome seguir adelante. Doy mis más sinceros agradecimientos a mi director el profesor Hernán Dario Vargas y al codirector el profesor Diego Peluffo por su compañía, de quienes aprendí bastante en términos de mi formación académica y personal. En particular, me gustaría dar las gracias a la Gobernación de Nariño y la organización CEIBA que a través de la beca denominada “BECATE”, me permitieron estudiar en una de las universidades con más reconocimiento de Colombia.

# Índice general

<b>1. Introducción</b>	<b>12</b>
<b>2. Descripción del Problema</b>	<b>13</b>
2.1. Planteamiento del problema	13
2.1.1. Formulación	14
2.1.2. Sistematización	15
<b>3. Objetivos</b>	<b>16</b>
3.1. Objetivo General	16
3.2. Objetivos Específicos	16
<b>4. Justificación</b>	<b>17</b>
4.1. Pertinencia	17
4.2. Viabilidad	18
4.3. Impactos	18
4.4. Delimitaciones y Alcances	19
<b>5. Marco de Referencia</b>	<b>20</b>
5.1. Estado del Arte	20
5.2. Marco Teórico	26
5.2.1. Preprocesamiento de imágenes satelitales	26
5.2.2. Aprendizaje automático	28
5.2.3. Evaluación	36
<b>6. Datos y área de estudio</b>	<b>40</b>
6.1. Dataset	40
6.1.1. Sentinel 1	40
6.1.2. Sentinel 2	41
6.1.3. Area de estudio	42
<b>7. Metodología</b>	<b>44</b>
7.1. Preprocesamiento imágenes satelitales	44
7.2. Entrenamiento de GAN	44
7.2.1. Arquitecturas de discriminador y generador	45
7.3. Evaluación de imágenes NDVI producidas por la GAN	48

<b>8. Resultados y discusión</b>	<b>50</b>
8.1. Unet++ . . . . .	50
8.2. Unet . . . . .	52
8.3. Resnet - 9 . . . . .	54
8.4. Resultados generales . . . . .	56
<b>9. Conclusiones</b>	<b>58</b>
<b>Bibliografía</b>	<b>60</b>
<b>10. Glosario y términos especiales</b>	<b>66</b>
<b>11. Anexo</b>	<b>67</b>
11.1. Gráficas de pérdidas de las arquitecturas conjuntas . . . . .	67
11.2. Diagrama de cajas de las medidas de evaluación para las arquitecturas de generadores . . . . .	68
11.3. Imágenes del grupo de validación con la tabla de sus valores en las medidas de evaluación . . . . .	70
11.3.1. Unet++ . . . . .	70
11.3.2. Unet . . . . .	73
11.3.3. ResNet 9 . . . . .	76
11.3.4. Tabla con medidas de evaluación anteriores imágenes . . . . .	79

# Índice de figuras

2.1. Imagen multi-espectral óptica obtenida del satélite Sentinel -2, el 2020-04-01.	14
2.2. Imagen SAR obtenida del satélite Sentinel -1, el 2020-04-02. . . . .	15
5.1. Arquitectura CGAN, contiene a un discriminador, un generador, la entrada $c$ que alimenta tanto el generador como el discriminador, el conjunto de entrada $z$ la imagen objetivo $x$ . La estructura original no tiene la entrada(c) [46] . . . . .	30
5.2. Arquitectura pix2pix, la figura contiene el generador(G), el discriminador(D), y toma como entrada del generador una imagen satelital SAR(X) y como una entrada del discriminador una NDVI(Y) [46] . . . . .	32
5.3. Arquitectura de una red Unet. Esta contiene una dos redes convolucionales que hacen de codificador y decodificador, en su union hay un cuello de botella, las rayas amarillas muestran los saltos, característica fundamental en estas red. En el cuadro se muestran algunas características de diseño mas profundo relacionados con conceptos de deep learning [48] . . . . .	35
5.4. Estructura principal de una resnet, en esta se muestra una entrada $x$ , esta entra directamente a la primera capa para ser entrenada, pero a su vez se mantiene para hacer una realimentación de la información [48] . . . . .	36
5.5. Arquitectura unet++, en la cual se puede observar una estructura similar que la unet convencional, con la gran diferencia de que se hacen conexiones anidadas hasta llegar a la capa final, la letra L se usa para nombrar esta estructura con el valor de capas anidadas más la final. También tenemos la información de configuración de la red [49] . . . . .	37
6.1. Imagen NDVI, donde se muestra la diferentes colores y sus tonalidades. Entre el color sea mas verde mas saludable es la vegetación, en las zonas rojas o no existe vegetación (ciudad) o esta en malas condiciones . . . . .	42
6.2. Zona del departamento de Nariño que es tomada por la imágenes satelitales, para posteriormente dividirla y hacer todo el entrenamiento de la red. . . . .	43
7.1. Diagrama de flujo de metodología para obtener imágenes NDVI a partir de la fusión de imágenes multispectrales sentinel 1 y sar sentinel 2, en una GAN . . . . .	49
8.1. Gráfica de las pérdidas del discriminador (Azul), generador(Naranja) y el total(Verde) en Unet++. . . . .	51
8.2. Gráficas de ejemplo para Unet++. Como entrada la imagen SAR, en la parte central la referencia (Ground Truth) y en la izquierda la imagen estimada. . . . .	52

8.3. Gráfica de las pérdidas del discriminador (Azul), generador(Naranja) y el total(Verde) en Unet. . . . .	53
8.4. Gráficas de ejemplo para Unet. Como entrada la imagen SAR, en la parte central la referencia (Ground Truth) y en la izquierda la imagen estimada. .	54
8.5. Gráfica de las pérdidas del discriminador (Azul), generador(Naranja) y el total(Verde) en ResNet. . . . .	55
8.7. Gráficas de ejemplo para ResNet-9. Como entrada la imagen SAR, en la parte central la referencia (Ground Truth) y en la izquierda la imagen estimada.	56
8.8. Comparación del SSMI en gráfico de cajas para las tres arquitecturas de generadores. Serie 1= Unet++, Serie = Unet, Serie 3=ResNet 9 . . . . .	57
11.1. Gráficas de pérdida del discriminador y generador para Unet++ . . . . .	67
11.2. Gráficas de pérdida del discriminador y generador para Unet. . . . .	67
11.3. Gráficas de pérdida del discriminador y generador para ResNet-9. . . . .	68
11.4. Comparación del PNSR en gráfico de cajas para las tres arquitecturas de generadores. Serie 1= Unet++, Serie = Unet, Serie 3 = ResNet 9. . . . .	68
11.5. Comparación del SAM en gráfico de cajas para las tres arquitecturas de generadores. Serie 1= Unet++, Serie 2 = Unet, Serie 3 = ResNet 9 . . . . .	69
11.6. Comparación del RMSE en gráfico de cajas para las tres arquitecturas de generadores. Serie 1= Unet++, Serie 2 = Unet, Serie 3 = ResNet 9 . . . . .	69
11.7. Gráficas de ejemplo para Unet++. Como entrada la imagen SAR, en la parte central la referencia (Ground Truth) y en la izquierda la imagen estimada.	72
11.8. Gráficas de ejemplo para Unet. Como entrada la imagen SAR, en la parte central la referencia (Ground Truth) y en la izquierda la imagen estimada. .	75
11.9. Gráficas de ejemplo para ResNet-9. Como entrada la imagen SAR, en la parte central la referencia (Ground Truth) y en la izquierda la imagen estimada.	78

# Índice de tablas

7.1. Arquitectura del discriminador utilizado en esta investigación . . . . .	45
7.2. Arquitectura del generador utilizando la configuración ResNet-9 utilizado en esta investigación . . . . .	45
7.3. Arquitectura del generador utilizando la configuración Unet utilizado en esta investigación . . . . .	46
7.4. Arquitectura del generador utilizando la configuración UNET utilizado en esta investigación . . . . .	48
8.1. Valores de las medidas de evaluación para tres imágenes de ejemplo para Unet++ . . . . .	52
8.2. Valores de las medidas de evaluación para tres imágenes de ejemplo para Unet . . . . .	53
8.3. Valores de las medidas de evaluación para tres imágenes de ejemplo para Resnet . . . . .	56
8.4. Tabla resumen de las medias y su desviación estándar para el grupo de validación . . . . .	57
11.1. Medidas de evaluación de 12 imágenes del grupo de validación para la arquitectura Unet++ . . . . .	79
11.2. Medidas de evaluación de 12 imágenes del grupo de validación para la arquitectura Unet . . . . .	79
11.3. Medidas de evaluación de 12 imágenes del grupo de validación para la arquitectura ResNet-9 . . . . .	80

# Capítulo 1

## Introducción

---

La evolución de la tecnología en los sensores remotos satelitales en la agricultura de precisión, ha contribuido a que millones de hectáreas sean monitoreadas alrededor del mundo [1]. Esto ha permitido que la información generada por los satélites pueda integrarse al funcionamiento rutinario de los servicios de asesoramiento en agricultura, mejorando así la gestión operativa en diferentes tipos de cultivos, con el objetivo de mejorar aspectos como la productividad y cuidado del medio ambiente.

Una de las características de las imágenes satelitales multiespectrales más utilizadas dentro de la agricultura es la generación de índices de vegetación, siendo el más utilizado el índice de vegetación normalizado (NDVI) [2]. Este índice tiene como principal característica que mide la vigorosidad de la vegetación, la cual puede ser usada para diferentes tareas, por ejemplo, conocer el estado de un cultivo. Este índice es obtenido de satélites multiespectrales, como: el LANDSAT 8, Sentinel-2 o SPOT, pero tienen el problema que en zonas de alta nubosidad la información del NDVI queda totalmente cubierta, ya que las frecuencias de sus bandas espectrales no son capaces de penetrar la nubosidad [3].

Una forma efectiva de abordar el problema de nubosidad, es utilizar fusión de imágenes satelitales multiespectrales con imágenes satelitales de radar de apertura sintética (SAR), estas últimas con la característica de no ser afectadas por la nubosidad, pero con la desventaja de no contar con la capacidad de producir el índice NDVI [4]. Entre las técnicas para fusionar imágenes satelitales, en la revisión de literatura se encontró que la utilización de arquitecturas de redes neuronales generativas adversarias (GANs) y autoencoders es prácticamente nulo. Por lo tanto, el utilizar estas arquitecturas para encontrar una imagen NDVI por medio de la fusión de imágenes satelitales es algo novedoso.

En esta investigación se plantea desarrollar una metodología para estimar imágenes NDVI sin presencia de nubosidad, por medio de la fusión de imágenes satelitales multiespectrales e imágenes satélites SAR. Cuyo componente investigativo principal está en determinar una arquitectura GAN o autoencoders que estime una imagen NDVI con resultados de alta precisión en términos de las métricas de comparación de imágenes, utilizando imágenes NDVI de referencia (imágenes reales). Para realizar esta experimentación se obtendrán imágenes de una zona del departamento de Nariño, la cual tiene el problema de alta nubosidad al estar ubicada en el pacífico colombiano, considerada una de las zonas más lluviosas del mundo. Además, de poseer un sector agrícola grande, pero con escasa tecnificación.

## Capítulo 2

# Descripción del Problema

---

### 2.1. Planteamiento del problema

La agricultura de precisión surge como solución en la incorporación de las nuevas tecnologías para hacer prácticas agrícolas sostenibles. A diferencia de la agricultura tradicional, la cual toma la premisa de homogeneidad realizando los procesos de cultivo igual en todo el terreno, la agricultura de precisión parte de un principio de variabilidad, en donde se reconoce la existencia de desigualdades en las propiedades del suelo dentro de un mismo territorio y la demanda de un tratamiento diferente para cada zona. Para detectar estas diferencias la tecnología brinda diferentes herramientas como: sensores, sistemas de posicionamiento global (GPS), equipos electrónicos para medir variables de diferente tipo, sistemas de información geográfica (SIG), y otras tecnologías de información y comunicación (TIC) [5]. El mayor inconveniente de muchos de los sistemas que soportan las prácticas de la agricultura de precisión es el elevado costo que implican ciertos montajes, que tienen, por ejemplo: el uso de sensores, drones de alta tecnología, y en ocasiones equipos que sólo pueden ser manipulados por personal calificado.

Las imágenes satelitales multiespectrales surgen como una buena opción para tomar información de diferentes lugares de forma gratuita, sin tener impedimentos en lugares de difícil acceso por otros medios. Además, que poseen información relevante para la agricultura como los denominados índices de vegetación (NDVI, NDWI, PVI, entre otros) [6]. Estos índices son utilizados para determinar el vigor de vegetación dentro de un píxel, permitiendo evaluar el estado de salud de la vegetación, con base en la medición de la emisión y reflexión que las plantas hacen de la radiación solar. En el NDVI, altos valores en el índice de vegetación del píxel identifica proporciones substanciales de vegetación saludable. Esta característica es clave para determinar la cantidad, calidad y desarrollo de un cultivo, por lo tanto dichos índices pueden ayudar a realizar una mejor planificación agrícola.

A pesar de los grandes beneficios de las imágenes satelitales multiespectrales, se tiene un problema que es inherente al tiempo climatológico en el cual se tome la imagen, ya que si la zona tiene una alta nubosidad, la información del comportamiento de la vegetación queda totalmente cubierta. Según Poveda [7] en Colombia este problema es muy evidente tratándose de una zona de alta nubosidad al estar afectado por el sistema de presión atmosférica denominado “baja anclada de Panamá”, que está justo frente a las costas del Chocó. Básicamente, es una franja de nubosidad que se mueve de sur a norte

en el primer semestre, y de norte a sur en el segundo semestre, convirtiendo al pacifico Colombiano como una de la zonas mas lluviosas del mundo. Por ejemplo, el departamento de Nariño por estar ubicado en esta zona se ve afectado por esta característica particular, lo que impide que se tenga información frecuente a través del uso de imágenes satelitales multiespectrales, impidiendo que se disponga de una herramienta que entregue información adecuada para hacer una buena planificación de cultivos. En la Figura 2.1 se muestra como la alta nubosidad oculta la información de vegetación del territorio sur oriental del departamento de Nariño:

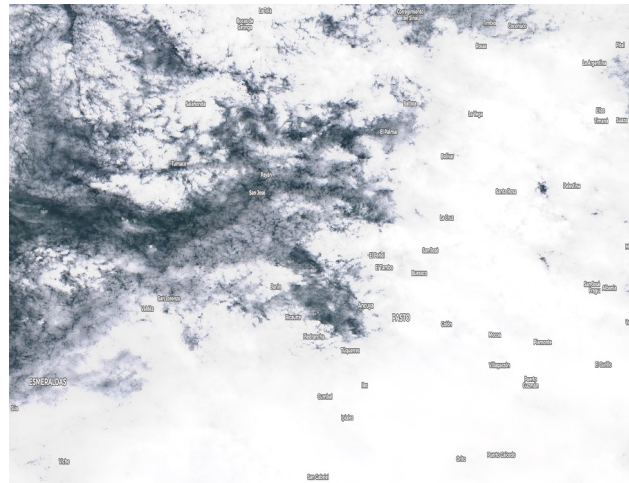


Figura 2.1: Imagen multi-espectral óptica obtenida del satélite Sentinel -2, el 2020-04-01.

Existen otra clase de imágenes satelitales llamadas SAR, estas son tomadas utilizando un sensor de radar ubicada dentro del satélite. Esta tecnología permite que la imagen de la zona sea adquirida sin tener inconvenientes con la nubosidad, debido a que la frecuencias de las ondas del radar son capaces de penetrar las nubes. Pero tienen el problema que no pueden entregar la misma información de los índices de vegetación, llevando a que se investigue como pueden producir información similar. En la Figura 2.2 se muestra que la zona no se ve impactada por la nubosidad:

Esto nos lleva a plantear fórmulas para hacer la unión de las características de los índices de vegetación de las imágenes satélites multiespectrales y la no afectación de la nubosidad por parte de la imágenes satelitales SAR. Encontrar la imagen con estas dos características es un gran aporte a la Teledetección debido al impacto que generaría en procesos como el monitoreo de cultivos.

### 2.1.1. Formulación

¿Como generar imágenes NDVI sin presencia de nubosidad mediante fusión de imágenes satelitales ópticas y SAR, con el fin de generar información para ser utilizadas en la planeación agrícola?

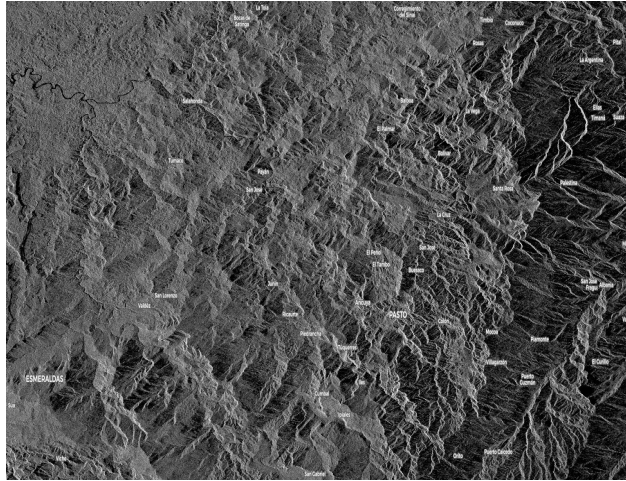


Figura 2.2: Imagen SAR obtenida del satélite Sentinel -1, el 2020-04-02.

### 2.1.2. Sistematización

- ¿Cómo obtener la mejor calidad de imágenes satelitales, para que puedan entrar al esquema de fusión de imágenes con la mayor exactitud de una zona específica?
- ¿Cómo desarrollar un esquema de fusión de imágenes satelitales que estime una imagen NDVI a partir de una SAR?
- ¿Cómo validar la Imagen NDVI producida por el esquema de fusión de imágenes?

## Capítulo 3

# Objetivos

---

### 3.1. Objetivo General

Desarrollar una metodología para la estimación de imágenes NDVI sin presencia de nubosidad por medio de fusión de imágenes satelitales multiespectrales y SAR orientada a apoyar procesos de planeación agrícola.

### 3.2. Objetivos Específicos

- Implementar una etapa de pre-procesamiento de imágenes multiespectrales y SAR a través de técnicas de corrección digital de imágenes satelitales con el fin de producir datos depurados.
- Proponer un esquema de fusión de imágenes multiespectrales y SAR empleando técnicas de aprendizaje profundo como GAN y Autoencoders para la estimación de imágenes NDVI.
- Evaluar el esquema de fusión de imágenes multiespectrales y SAR a través de métricas de comparación de imágenes con el fin de determinar el esquema de fusión más exacto para propósitos de planeación agrícola.

## Capítulo 4

# Justificación

---

### 4.1. Pertinencia

El tener imágenes NDVI sin afectación de la nubosidad permite que la calidad, cantidad y desarrollo de un cultivo pueda ser vigilado de manera continua [8]. Esto contribuye a realizar una adecuada planificación de cultivos, con el objetivo de mantener una oferta estable de productos sin perjudicar el medio ambiente

Este trabajo surge de la necesidad de impulsar el desarrollo de las nuevas tecnologías para mejorar las prácticas agrícolas. Es importante anotar que en general una mala planificación en el manejo de los cultivos da lugar a problemas de alto impacto ambiental. Según la FAO, OMS y UNICEF [9] la agricultura utiliza el 11 % de la superficie terrestre del mundo para la producción de cultivos, siendo esta responsable de aproximadamente del 80 % de la deforestación a nivel mundial.

En Colombia los problemas derivados del desarrollo de una agricultura no planificada y con escasa tecnificación han impedido alcanzar altos niveles de productividad [10]. Así mismo, han generado un estancamiento económico y social debido al alto impacto proveniente del sector agropecuario al PIB nacional, con alrededor del 25 % enfocado principalmente en el sector de la agricultura. Otro impacto causado por la falta de planeación de los cultivos es el ambiental, consecuencia de la deforestación en departamentos como el Caquetá, Norte de Santander, Magdalena y Nariño [11].

En Nariño el sector agropecuario representa el 14 % del PIB del departamento siendo el segundo sector más relevante después de servicios. Además, el 51 % de la población nariñense se sitúa en el área rural, teniendo alguna relación con la cadena de producción y comercialización de productos agrícolas. Estas razones catalogan el sector como uno de los ejes centrales de desarrollo del Departamento [12]. Sin embargo, en las últimas décadas la falta de tecnificación en la producción no ha permitido el desarrollo pleno del sector, debido a que existe poca investigación que impacte positivamente de manera significativa en el aumento de la productividad.

Observando este panorama se encuentra pertinente realizar esta investigación, la cual pretende entregar una herramienta para ser utilizada por la agricultura para planificar diferentes variables en los cultivos. Debido a que las imágenes NDVI resultantes (sin problemas de nubosidad) pueden ser integradas a diferentes tipos de plataformas, para ser analizadas y posteriormente tomar decisiones sobre el desarrollo de los cultivos, con el objetivo de mejorar la productividad sin impactar de manera significativa en el medio am-

biente.

## 4.2. Viabilidad

Trabajos de investigación previos, han demostrado que es posible fusionar imágenes multiespectrales y SAR, mediante el uso de técnicas de aprendizaje de máquina [13][14][15][16][17]. En estos trabajos se demuestra que existen buenos recursos técnicos para el pre-procesamiento de imágenes satélites y la aplicación de diferentes algoritmos de aprendizaje de máquina. En cuanto a la economía del proyecto, recursos como las imágenes satelitales y la plataforma de programación en Python están disponibles de forma gratuita

## 4.3. Impactos

- **En el campo del conocimiento:** Las redes adversarias generativas (GANs) y auto-encoders son relativamente nuevas, esto provee un amplio campo de investigación ya que se pueden adaptar a diferentes aplicaciones, generando diferentes tipos de arquitecturas. En cuanto a los trabajos que hay sobre su aplicación a la fusión de imágenes satelitales, estos son escasos, debido a que en su gran mayoría utilizan algoritmos de aprendizaje automático clásicos (SVM, Random Forest, CNN, RNN). Por lo tanto el encontrar una arquitectura que entregue buenos resultados para la predicción de imágenes NDVI es algo innovador dentro de la literatura.
- **Económicos:** Las imágenes productos de esta investigación, pueden ser incorporadas a plataformas digitales o sitios web de carácter público o privado, con el fin de ayudar a la planeación agrícola. Con esto se busca optimizar el uso de productos como: herbicidas, fertilizantes, plaguicidas, proporcionando un ahorro en beneficio del agricultor. Otro aspecto a tener en cuenta, es la planeación adecuada de tiempos de los cultivos, la cual puede ser programada con la información de la imagen NDVI y el criterio de un profesional de la agronomía. Con fechas de plantación y cosecha adecuadas, se puede optimizar el uso de la tierra en beneficio de una producción más estable. La consecuencia inmediata será el mantenimiento de la productividad del suelo y así, el incremento de los ingresos de los productores.
- **Sociales:** Es claro que las condiciones sociales de una familia mejoran al recibir más altos ingresos, sustentados en el punto anterior. Por otra parte, se pretende mostrar la importancia de incorporar las nuevas tecnologías para el mejoramiento de la actividad agrícola. También puede llevar a impulsar el desarrollo de otros proyectos que ayuden a mejorar los procesos del sector agropecuario.

- **Ambientales:** Es bien sabido que el cambio climático es uno de los principales problemas que afronta la humanidad en estos momentos. La deforestación es una de las principales causas por las que se ha acelerado dicho fenómeno, por lo tanto, el generar una herramienta para realizar una agricultura sostenible, permite que el fenómeno de la deforestación disminuya al haber una menor degradación del suelo consecuencia de una mejor planificación de los cultivos [18].

## 4.4. Delimitaciones y Alcances

Este proyecto busca desarrollar una metodología y algoritmos con el fin de generar imágenes NDVI de zonas de alta nubosidad. Como punto de partida se realizará el trabajo utilizando imágenes satelitales del departamento de Nariño, sin descartar que se puedan utilizar imágenes de otras zonas que presenten el mismo problema. La investigación estará limitada a generar las imágenes NDVI a partir de imágenes SAR. La verificación de la validez de resultados se hará por medio del uso de un grupo de validación, el cual está compuesto por pares de imágenes SAR y NDVI de una misma zona libre de nubosidad. En este punto es importante destacar que al no tener imágenes NDVI de zonas con alta concentración de nubes, la base de datos será construida con imágenes en diferentes tiempos para que el algoritmo propuesto este alimentado por la mayor cantidad de información, con esto, se pretende que las características de la imagen SAR en una zona con nubosidad puedan estar relacionada con las características aprendidas por el algoritmo. Para realizar el proceso de validación se tomará cada una de las imágenes SAR con el fin de transfórmalas a imágenes NDVI por medio del algoritmo propuesto, para posteriormente comparar la imagen generada con la imagen NDVI original. En este trabajo no se busca definir puntualmente la relación del índice con variables de cultivos, ya que para esto tendría que hacer una investigación de correlación del NDVI con el comportamiento de la variable a investigar para cada cultivo, lo que requiere de un trabajo de campo, que es muy complejo de realizar en este momento.

## Capítulo 5

# Marco de Referencia

---

### 5.1. Estado del Arte

La fusión de imágenes satelitales en la teledetección es un campo de amplia investigación debido a la masificación de su aplicación en diferentes sectores, como por ejemplo: la agricultura, la minería, vigilancia medioambiental, entre otros. En la siguiente revisión de literatura se expondrán diferentes técnicas de fusión de imágenes satelitales, que tienen como principal objetivo predecir el comportamiento vegetal y de cobertura terrestre por medio del uso de diferentes herramientas algorítmicas, con el fin de solucionar el problema de alta nubosidad en algunos lugares.

La técnica de usar imágenes multitemporales es una de las formas de fusión más investigadas en la teledetección, debido al alto contenido de información oculta en la dimensión del tiempo. De acuerdo a la literatura, las aplicaciones de datos multitemporales más ampliamente abordadas son la clasificación de series temporales de imágenes y la detección de cambios en la cobertura terrestre Bovolo y Bruzzone [19] y Ghamisi, Rasti, Yokoya y col. [20]. Swain [21] introdujo un enfoque para la clasificación de datos multitemporales basado en la clasificación en cascada de pares de imágenes. Swain modeló el problema de clasificación desde una perspectiva Bayesiana, que introduce la correlación temporal en el proceso de clasificación de las imágenes adquiridas en diferentes momentos para vincular las probabilidades de clase estimadas en imágenes individuales. Bruzzone, Prieto y Serpico [22] desarrollaron y generalizaron este marco al caso de datos multitemporales y multisensoriales, introduciendo un enfoque para la clasificación compuesta basada en clasificadores de redes neuronales, que pueden fusionarse adecuadamente en un escenario Bayesiano de decisión. De estos trabajos se han producido diferentes investigaciones sobre los índices de vegetación como el NDVI, por ejemplo, Pan, Huang, Zhou y col. [23] utilizaron datos de series temporales de índice de vegetación derivados de imágenes AVHRR, MODIS y SPOT-VEGETATION, los cuales generalmente tienen una baja resolución espacial. En este documento se presentó un método integral para construir un conjunto de datos de series de tiempo NDVI derivado del HJ-1 A / B CCD (satélite Chino), demostrando su aplicación en áreas de tierras de cultivo. Los procedimientos de construcción de datos de series temporales incluyeron pre-procesamiento de imágenes, filtrado de señales e interpolación para imágenes diarias, luego las series temporales NDVI presentaron un ciclo fenológico completo. Es importante recalcar que para analizar dicho ciclo fenológico se tuvieron en cuenta datos obtenidos en campo, lo

que ayuda a realizar una relación acertada del valor del píxel del NDVI con el estado de crecimiento del cultivo.

Profundizando en el enfoque de utilizar series temporales y algoritmos de machine learning, Clerici, Valbuena Calderón y Posada [13] proponen una metodología para fusionar la información de las imágenes de los satélites Sentinel-1 y Sentinel-2, para un mapeo de la cobertura terrestre de una parte de la región del Bajo Magdalena en Colombia. La clasificación de cobertura de la tierra se hizo a través del uso de los índices de vegetación, para posteriormente hacer una comparación del rendimiento de clasificación utilizando diferentes algoritmos de machine learning. El conjunto de datos integrado de radar e infrarrojo visible cercano (NIR) clasificado con un algoritmo de máquina de vectores de soporte (SVM) produce el mapa de cobertura y uso de la tierra, teniendo una precisión de clasificación general del 88,75 %. De igual forma Haas y Ban [24] realizaron un estudio cuyos objetivos principales fueron evaluar el potencial de los datos SAR Sentinel-1A en banda C y Sentinel-2A para la clasificación y el mapeo de espacios urbanos y periféricos urbanos ecológicamente importantes del área metropolitana de Zürich, Suiza. La fusión de datos Sentinel-1 y 2 por medio del algoritmo SVM, facilita tanto la captura de detalles ecológicamente relevantes como al mismo tiempo permite análisis urbanos a gran escala que toman en consideración las regiones circundantes. Por lo tanto el algoritmo SVM demuestra ser una opción interesante en la tarea de fusionar características de imágenes SAR y ópticas, siendo importante destacar que las investigaciones solo utilizaron el kernel lineal y gaussiano, abriendo una posibilidad de investigación con otros tipos de kernels.

Sin lugar a duda el desarrollo del deep learning en los últimos años ha impulsado que la arquitecturas de redes neuronales sean incorporadas en numerosas aplicaciones, incluyendo la predicción de imágenes a partir de imágenes satelitales ópticas y SAR. Scarpa, Gargiulo, Mazza y col. [15] proponen estimar las características ópticas faltantes a través de la fusión de datos y el aprendizaje profundo, a través del uso de diferentes fuentes de información (secuencias ópticas, secuencias SAR, modelo de elevación digital) para explotar las dependencias temporales y de sensores cruzados. Sobre la base de estos datos y una pequeña fracción libre de nubes de la imagen objetivo, una red neuronal convolucional compacta (CNN) fue entrenada para realizar la estimación deseada. Por otra parte Baghdadi, El Hajj, Zribi y col. [14] propusieron la fusión de imágenes RADARSAT-2 y Landsat-7 y 8, por medio de una red neuronal artificial, para estimar la humedad del suelo y el índice de área foliar. El NDVI obtenido de la unión de las bandas en imágenes del satélite Landsat se combinó con diferentes subconjuntos de polarización VH y VV en imágenes SAR para alimentar redes artificiales, y así obtener una imagen SAR con las características del NDVI. En este punto cabe destacar el artículo realizado por He y Yokoya [16], en este se comparan 4 arquitecturas de redes neuronales para la fusión de imágenes Sentinel-1 y Sentinel -2, destacándose la incorporación de la red neuronal generativa antagónica (GAN), que es una de las nuevas arquitecturas con la habilidad para trabajar con datos no supervisados. En este trabajo se destaca que los resultados fueron mejores para la serie temporal CNN, pero según los autores esto se debe en gran medida

a la mala elección del grupo de entrenamiento para la arquitectura GAN. En los últimos años la investigación sobre las redes GAN ha aumentado considerablemente llevando a diseñar diferentes tipos de arquitecturas para diversas aplicaciones, siendo consideradas como uno de los mayores avances dentro del deep learning en los últimos años.

Por otra parte, tal vez uno de los algoritmos de machine learning más utilizados en la fusión de imágenes ópticas y SAR son los algoritmos tipo árbol. Esto se debe a que la identificación de los elementos (tipos de vegetación, bosques, cultivos, etc) que hacen parte de la cobertura terrestre puede representarse por medio de un modelo jerárquico. Dimov, Low, Ibrakhimov y col. [17] compararon el rendimiento de clasificación de series de tiempo con datos SAR polarizados simples, datos ópticos y datos ópticos y SAR fusionados a través de la transformada de Gram-Schmidt. Se aplicaron diferentes algoritmos de aprendizaje automático para la clasificación de cultivos. Específicamente, los árboles de refuerzo de gradiente (GBT) arrojaron buenos resultados. Hirschmugl, Sobe, Deutscher y col. [25] compararon dos métodos diferentes para la combinación de datos ópticos y SAR: en la primera parte hicieron una combinación basada en datos, donde todos los datos preprocesados se incluyeron en el mismo paso de clasificación de un clasificador Random Forest. En segundo lugar, emplearon una combinación basada en resultados utilizando el principio bayesiano. Para la combinación basada en resultados, la probabilidad de un píxel perteneciente a una determinada clase (por ejemplo, bosque o deforestación) se genera a partir de la cadena de procesamiento óptico y SAR independientemente. Irwin, Beaulne, Braun y col. [26] fusionaron datos heterogéneos de misiones satelitales de radar de apertura sintética (SAR), imágenes ópticas basadas en satélites y detección y alcance de luz en el aire (LiDAR) para la detección de aguas superficiales. Cada conjunto de datos se clasificó independientemente para el agua superficial y luego se crearon modelos de clasificación fusionados de los tres conjuntos de datos. Se desarrolló un árbol de decisión multinivel para crear una máscara de agua óptima al minimizar las diferencias entre los modelos que se originan a partir de conjuntos de datos únicos. Los resultados mostraron una incertidumbre en la clasificación del agua de 4 a 9 % utilizando los modelos fusionados finales en comparación con 17 a 23 % de incertidumbre utilizando SAR de polarización única. En general, en los anteriores resultados se destacan las ventajas de fusionar múltiples técnicas heterogéneas de teledetección para detectar aguas superficiales en un paisaje con alta concentración de vegetación. Forget, Shimoni y Gilbert [27] evaluaron la complementariedad de los sensores Landsat 8 y Sentinel-1 para mapear áreas urbanizadas en doce áreas urbanas del África subsahariana, utilizando una clasificación supervisada a nivel de píxeles basada en el clasificador de Random Forest. Los resultados sugieren que la combinación de las características ópticas y SAR mejora constantemente el rendimiento de la clasificación, principalmente al mejorar la diferenciación entre terrenos urbanizados y terrenos descubiertos. Sin embargo, la fusión fue menos beneficiosa en los estudios de casos montañosos. Van Tricht, Gobin, Gilliams y col. [28] utilizan las imágenes de Sentinel-1 y las imágenes ópticas Sentinel-2 para crear un mapa de cultivos para Bélgica. Para garantizar entradas de radar y ópticas homogéneas en todo el país, se

crearon mosaicos de retrodispersión de 12 días para Sentinel-1, y las imágenes del índice de vegetación de diferencia normalizada (NDVI) de Sentinel-2 para producir mosaicos sin nubes durante 10 días. Un clasificador Random Forest optimizado predijo los ocho tipos de cultivos con una precisión máxima del 82 %. La combinación de imágenes de radar y ópticas siempre superó a una clasificación basada en entradas de un solo sensor. De Alban, Connette, Oswald y col. [29] evaluaron la combinación de datos SAR con la banda L y Landsat para estimar el uso de la tierra y cambio de cobertura en el dinámico paisaje tropical de Tanintharyi, en el sur de Myanmar. Se clasificaron datos Landsat y SAR de banda L, utilizando un satélite japonés (JERS-1) y el satélite avanzado de radar de apertura sintética de banda L de banda L-2 (ALOS-2 / PALSAR-2), utilizando Random Forest como clasificador para mapear y cuantificar las transiciones de cambio de uso y cobertura del suelo. Los datos combinados de Landsat y SAR de banda L produjeron las mejores precisiones de clasificación general (92.96 % a 93.83 %), superando a los datos de sensores individuales (91.20 % a 91.93 % solo para Landsat; 56.01 % a 71.43 % solo para SAR). El análisis de datos combinados de Landsat y SAR de banda L proporciona una mejor comprensión asociada a la plantación agrícola. De esta serie de artículos se puede observar, que los árboles de decisión han sido ampliamente explotados para fusionar imágenes satelitales.

La modelación matemática y los índices de correlación entre los índices de vegetación de imágenes ópticas y las bandas de imágenes SAR es otra alternativa para generar una imagen que obtenga las mejores características de cada una. De Bernardis, Vicente-Guijalba, Martínez-Marin y col. [30] proponen la incorporación de fuentes adicionales de información a las series temporales de observaciones agrícolas del índice NDVI para una aplicación de estimación del estado fenológico. La implementación propuesta se basa en el esquema de filtro de partículas (PF) que puede integrar múltiples fuentes de datos. Una serie de imágenes Landsat-5/7 NDVI se complementa con dos fuentes distintas de información: imágenes SAR del satélite TerraSAR-X e información de temperatura del aire de una estación terrestre, obteniendo una mejora en la precisión general de la estimación, especialmente cuando la serie temporal de datos NDVI es incompleta. Reiche [31] presenta un nuevo enfoque basado en probabilidad Bayesiana para combinar imágenes SAR y series de tiempo ópticas para la detección de la deforestación. En este artículo se estiman las probabilidades normalizadas forestales y no forestales para cada observación de series de tiempo individuales utilizando funciones de densidad de probabilidad forestal y no forestal específicas del sensor en el satélite. La actualización Bayesiana calculó la probabilidad de deforestación dadas las observaciones anteriores, la observación actual y las observaciones futuras para confirmar o rechazar un evento de deforestación. El enfoque se aplicó a las series temporales Landsat NDVI y ALOS PALSAR L-band backscatter ratio (HVHH) para detectar la deforestación. Se probó la alta precisión para el caso de NDVI-HVHH fusionados, el resultado general fue de entre 87.4 % y 90.6 %.

Kumar, Rao y Sharma [32] hicieron un estudio explorando la posibilidad del Índice de Vegetación por Radar (RVI) para el monitoreo de la vegetación en los campos de frijol

de algodón y soja como una alternativa al monitoreo con el Índice de Vegetación de Diferencia Normalizada (NDVI en imágenes MODIS). El RVI es la medida de aleatoriedad de la dispersión que se ha propuesto como método para monitorear el nivel de crecimiento de la vegetación, particularmente cuando se dispone de series temporales de datos. La investigación se llevó a cabo en la región nordeste en India, la cual presenta temporadas con alta nubosidad debido a la continua aparición de monzones. Los conjuntos de datos fueron adquiridos del satélite Radarsat-2 en la temporada de monzones de 2011. El RVI derivado de las imágenes SAR posteriormente fue comparado con el producto MODIS NDVI del mismo periodo, con el fin de verificar la validez de los resultados. Como el RVI es una medida de dispersión volumétrica y el NDVI una medida de tonalidad de verde se llevó a cabo una normalización para hacer las medidas comparables en un rango de 0 a 1. Los resultados más destacados muestran que el RVI en la temporada de monzones de 2011 se situó con un valor más bajo de 0,15 en la primera semana de Julio, se incrementó linealmente entre 0.29 y 0.56 en la última semana de agosto. Esta tendencia se atribuye al hecho de crecimiento en el cultivo aumenta la aleatoriedad de la dispersión volumétrica, por lo que RVI aumentó con el crecimiento de la planta. Por otra parte, el NDVI de las imágenes MODIS se incrementó desde el valor medio de 0,35 a 0,62 entre el 13 de julio y el 6 de agosto del 2011. La relación encontrada entre los dos índices permite afirmar que el índice RVI puede utilizarse para el seguimiento del cultivo en todo el ciclo del cultivo. La investigación muestra algunas mejoras en la técnica RVI que NDVI en algunas circunstancias, puesto que el RVI aumenta linealmente a medida que crece el cultivo, a diferencia del NDVI que se satura después de un cierto nivel de crecimiento del cultivo. Por lo tanto, el índice RVI puede ser muy adecuado para el seguimiento de la vegetación en variables muy críticas como la humedad. El resultado del estudio mostró eficacia para el monitoreo de la soja, el algodón y la humedad del suelo.

Wang, Li y Ge [33] relacionaron la retrodispersión ENVISAT ASAR / ALOS PALSAR con NDVI y luego con la biomasa. Se encontró que las series de tiempo de NDVI se correlacionan significativamente con ASAR. La investigación confirmó el potencial de los datos ASAR y PALSAR para monitorear la biomasa y su cambio estacional en el suroeste de Australia templado, y proporcionó una base sólida para futuras investigaciones sobre la estimación cuantitativa de la biomasa. Por otra parte, Mishra y Susaki [34] desarrollaron un marco de integración de datos ópticos y SAR para la detección de cambios y una relación para la detección automática de patrones de cambio automático en áreas urbanas y suburbanas. La relación empírica entre la respuesta de la característica de superficie a las imágenes ópticas y SAR ha delineado con éxito seis clases cambiadas en un área de expansión urbana muy compleja que de otro modo era imposible con imágenes multiespectrales. Por otra parte, Pipia, Muñoz-Marí, Amin y col. [35] proponen el uso de la regresión del Proceso Gaussiano de Salida Múltiple (MOGP), una técnica de aprendizaje automático que usa las relaciones estadísticas entre series de tiempo multisensor, para detectar áreas vegetadas sobre las cuales la sinergia entre imágenes ópticas SAR es rentable. Para este propósito, se utilizaron las series de tiempo del Índice de vegetación de

radar Sentinel-1 (RVI) y el Índice de área de hoja Sentinel-2 (LAI) en un área de estudio en el noroeste de España. Los resultados revelan que el rendimiento de MOGP depende del tipo de cultivo y la etapa de cultivo.

Otra técnica es la fusión espacio-temporal, esta es utilizada para mezclar datos de alta resolución pero con tiempos de adquisición largos (p. ej., Landsat, 15 días), con datos de baja resolución pero con tiempo de adquisición cortos (por ejemplo, MODIS, 1 día) [36]. Su implementación se realiza en base a la disponibilidad de al menos un par de imágenes de resolución espacial Alta - Baja (p. ej., par de imágenes MODIS-Landsat adquirido en el mismo día) o un mapa de cobertura terrestre de buena resolución espacial que este temporalmente cerca del día de predicción. Wang y Atkinson [37] propusieron un nuevo algoritmo, llamado Fit-FC, para la fusión espacio-temporal de las imágenes Sentinel-2 y Sentinel-3 para crear Sentinel-2 casi a diario imágenes Fit-FC es un método de tres pasos que consiste en regresión ajuste del modelo (ajuste RM), filtrado espacial (SF) y residual compensación (RC). Se puede implementar el método Fit-FC usando solo un par de imágenes y es particularmente relevante para casos implicando fuertes cambios temporales.

Adrian, Sagan y Maimaitijiang [38] utilizaron las imágenes Sentinel-1 para fusionarlas con los datos de Sentinel-2, con el propósito de mejorar la precisión de clasificación para diferentes tipos de cultivos. Este estudio está dividido en cuatro fases, en la primera se realiza toda la etapa de preprocesamiento de las imágenes satelitales a través de la plataforma Google Earth Engine, en la cual se hicieron actividades como la eliminación de la nubosidad (no muy densa) por medio del método análisis de componentes principales (PCA). También, se destaca el remuestreo a 10m por píxel en las imágenes Sentinel -2 y la conversión de los datos a escala de decibelios en las imágenes Sentinel -1. El objetivo de la segunda fase es eliminar el ruido (ruido de pimienta) por medio del uso de una red neuronal, con esto se busca reducir el moteado provocado por artefactos que interfieren en la calidad de adquisición de la imagen SAR por medio del radar del satélite Sentinel-1. En la sección tres se realiza la fusión de imágenes Sentinel-1 y Sentinel-2 multitemporales por medio de PCA, para lograr este objetivo, se expresan las imágenes como matrices, y se reemplazan las componentes principales de las bandas 9 y 10 de las imágenes Sentinel -2 por las componentes principales de las polarizaciones VH y VV de las imágenes SAR. Por último, se realiza una clasificación de diferentes tipos de cultivos por medio de algoritmos de segmentación como SegNet, Unet, 3D Unet y Random Forest. Como resultado destacado se expuso que la fusión de datos SAR multitemporales y ópticos produce mayores valores de precisión en el grupo de entrenamiento (3D U-Net 0,992, 2D U-Net 0,943, SegNet 0,871) y grupo de prueba (3D U-Net 0,941, 2D U-Net 0. 847, SegNet 0,643) para mapeo de tipo de cultivo en comparación con los datos multitemporales SAR u ópticos independientes.

Teniendo en cuenta las diferentes técnicas utilizadas para la fusión de imágenes satelitales con diferentes tipos de sensores (SAR y óptico), y las oportunidades de investigación en el área de la teledetección, en este trabajo se abordará el problema de nubosidad plan-

teando como solución la estimación de la información por medio del uso de técnicas modernas de Deep Learning. Las arquitecturas tipo GAN o Autoencoders no se encuentran casi presentes en los artículos analizados, esto se debe en gran medida a que el desarrollo y aplicación de estas técnicas se ha concentrado en los últimos 4 años, dedicándose a tareas mas generales como el reconocimiento de caras, generación de imágenes a partir de texto y viceversa [39]. Por lo tanto, incorporar estas arquitecturas a la tarea de fusión de imágenes es algo novedoso, sin olvidar que pueden ser mejoradas al usar dentro de su estructura arquitecturas convencionales como la CNN, RNN, ResNet, entre otras.

## **5.2. Marco Teórico**

En esta sección se presenta la base teórica de conceptos fundamentales para brindar al lector una mejor comprensión del tema de GANs. También se describen las arquitecturas de autoencoders que se propondrán como generador de las GANs en la actividad de traslación imagen a imagen. Por último, se describe la teoría detrás de los métodos de evaluación de las redes de aprendizaje profundo en comparación de imágenes.[40]. [41].

### **5.2.1. Preprocesamiento de imágenes satelitales**

Como los sistemas para la adquisición de imágenes satelitales se encuentran a una gran altura sobre la superficie terrestre, estas pueden contener distorsiones debidas a la vibración de los sistemas de imagen, contaminaciones atmosféricas, la rotación y curvatura terrestre. Por lo tanto, es esencial corregir los posibles defectos de las imágenes de teledetección antes del posprocesamiento y el análisis. Esta etapa de preparación de la imagen se denomina preprocesamiento de la imagen, lo que implica un conjunto de operaciones de procesamiento inicial que pueden utilizarse para corregir o minimizar las distorsiones debidas a los problemas derivados de los sistemas de imágenes y condiciones de observación [4]. En las imágenes satelitales de los sistemas de adquisición Sentinel 1 y 2 se pueden desarrollar operaciones como: corrección atmosférica, corrección radiométrica, corrección geométrica y remuestreo, estas se explicarán con más detalle a continuación.

#### **5.2.1.1. Corrección atmosférica**

En los procesos de adquisición de imágenes satelitales, la radiación de la superficie terrestre sufre una importante interacción con la atmósfera antes de llegar a los sensores de teledetección. Por lo tanto, la radiación medida suele contener al menos dos tipos de energía: una parte de la radiación medida procedente de la reflexión del objetivo de interés y el remanente de la propia atmósfera [42]. Para la mayoría de aplicaciones terrestres, las imágenes obtenidas por teledetección deben recogerse en escenas claras sin otros

contaminantes. Sin embargo, es difícil asegurar que las escenas de observación estén totalmente despejadas, ya que las nubes poco densas, la bruma y los aerosoles de la atmósfera son omnipresentes. Debido a efectos atmosféricos como la atenuación y la dispersión de las nubes y las partículas de aerosol, las radiaciones medidas en estas escenas contienen ruido atmosférico, que a su vez afecta a la precisión de la interpretación de los objetivos terrestres, especialmente en la teledetección del visible y del infrarrojo cercano [43]. Teniendo en cuenta esto, las imágenes pueden limpiarse utilizando técnicas incluidas en la cajas de herramientas de preprocesamiento de cada misión satelital.

#### **5.2.1.2. Corrección radiométrica**

En las imágenes tipo SAR de sentinel 1, debido a las distorsiones radiométricas derivadas de las diferencias en las condiciones de iluminación y las características del sensor del satélite, la radiación observada por el sistema de imágenes puede no coincidir con la energía emitida o energía reflejada de los objetos de interés [42]. Para lograr la coherencia radiométrica entre diferentes imágenes de teledetección, las distorsiones deben ser corregidas antes de proceder a la interpretación y el análisis de las imágenes; los procesos para eliminar estas distorsiones radiométricas se denominan correcciones radiométricas. En general, la corrección radiométrica tiene como objetivo calibrar los datos de teledetección para mejorar la fidelidad de la radiación reflejada o emitida por los objetos, eliminando las distorsiones radiométricas debidas a las irregularidades de los sensores, los efectos topográficos y los ángulos del sol [4].

En las imágenes sentinel -1 la corrección radiométrica del ruido debido a las irregularidades del sensor es la que viene implícita en la toma de la imagen. El ruido radiométrico relacionado con el sensor relacionados con el sensor puede surgir de cambios en la sensibilidad del sensor, como la degradación de los sensores de teledetección sensores después de operar en órbita durante un largo período de tiempo [4]. La señal medida de un sensor degradado contendrá errores radiométricos debido a los cambios en la función de respuesta del sensor. Para hacer frente a estos errores radiométricos deben establecerse nuevas relaciones entre la medición de la irradiancia calibrada y la señal de salida del sensor. Este proceso también se denomina calibración [4].

#### **5.2.1.3. Remuestreo**

Matemáticamente, el remuestreo implica la interpolación y el muestreo para producir nuevas estimaciones de píxeles en diferentes cuadrículas [44]. Se han desarrollado diversos métodos para el remuestreo, y la elección de los núcleos de remuestreo depende en gran medida de la aplicación. En las imágenes sentinel 2 y sentinel 1 se pueden utilizar kernel como el vecino más cercano. El vecino más cercano es un método frecuentemente utilizado para el remuestreo en la teledetección, en este, se estima un nuevo valor para cada píxel corregido (es decir, la nueva cuadrícula) utilizando los datos de los píxeles no

corregidos más cercanos (es decir, las cuadrículas originales). Las ventajas de la técnica del vecino más cercano son su sencillez y la capacidad de conservar los valores originales en la escena inalterada.

#### **5.2.1.4. Filtrado de manchas**

El moteado, que aparece en las imágenes SAR como ruido granular, se debe a la interferencia de ondas reflejadas por muchos dispersores elementales [45]. El filtrado de manchas es un procedimiento para aumentar la calidad de la imagen al reducir las manchas. Cuando dicho procedimiento se realiza en una etapa temprana de procesamiento de los datos SAR, el moteado no se propaga en los procesos en curso (es decir, corrección del terreno o conversión a dB). El filtrado de moteado no es aconsejable cuando existe interés en la identificación de pequeñas estructuras espaciales o textura de imagen, ya que podría eliminar dicha información. Se ha descubierto que el filtro Lee refinado es superior, con respecto a otros filtros de motas de un solo producto, para la interpretación visual debido a su capacidad para preservar los bordes, las características lineales y la información de textura y destino de puntos [45].

#### **5.2.1.5. Corrección de terreno Doppler de rango**

Los datos de SAR generalmente se detectan con un ángulo de visión variable superior a 0 grados, lo que da como resultado imágenes con cierta distorsión relacionada con la geometría lateral. Las correcciones de terreno están destinadas a compensar estas distorsiones para que la representación geométrica de la imagen sea lo más cercana posible al mundo real. La corrección del terreno de rango Doppler es una corrección de las distorsiones geométricas causadas por la topografía, como el escorzo y las sombras, utilizando un modelo de elevación digital para corregir la ubicación de cada píxel. El operador de corrección de terreno Doppler de rango disponible en SNAP implementa el método de ortorrectificación Doppler de rango [45] para geocodificar escenas SAR a partir de imágenes en geometría de radar. La coordenada de destino al Sistema de referencia (CRS) se puede seleccionar y configurar opcionalmente para que coincida con la zona UTM de los gránulos Sentinel-2 superpuestos. El operador permite la selección del método de remuestreo de imágenes y la espaciado de píxeles de destino en el CRS de destino. Este paso de procesamiento permite el ajuste espacial de los productos Sentinel-1 GRD a las cuadrículas de datos Sentinel-2 MSI, con el fin de geolocalizar los datos en una cuadrícula espacial común y promover el uso de constelaciones virtuales satelitales[4].

### **5.2.2. Aprendizaje automático**

El aprendizaje automático sirve para detectar patrones en gran cantidad de datos. Al utilizar los patrones detectados, el algoritmo de aprendizaje automático puede predecir

datos futuros o realizar la toma de decisiones en condiciones de incertidumbre. Los algoritmos de aprendizaje automático a menudo se dividen en dos categorías: supervisados aprendizaje y aprendizaje no supervisado, donde el aprendizaje supervisado es la forma más ampliamente utilizado en la práctica[4]. El aprendizaje automático supervisado es la tarea, donde cada dato de entrenamiento la muestra está asociada con el objetivo o la etiqueta. Por ejemplo, las muestras de entrenamiento pueden estar en forma  $\{(x_i, k_i)\}_{i=1}^N$ , donde  $x_i$  es la muestra de datos y  $k_i$  es su correspondiente clase (dominio o etiqueta). En el proceso de aprendizaje, la tarea es aprender a predecir  $k$  desde  $x$ , donde se puede proporcionar retroalimentación al sistema entrenado en forma de objetivo  $k$ , con una analogía con un maestro, que muestra al sistema capacitado qué proporcionar etiquetas  $k$ . El aprendizaje automático no supervisado es la tarea cuando no hay una etiqueta con las observaciones de entrenamiento, ni información sobre el resultado. La tarea puede ser aprender la distribución de probabilidad  $p(x)$ , agrupamiento, u observación de propiedades interesantes [46].

#### 5.2.2.1. Deep learning

El aprendizaje profundo se refiere a redes neuronales con varias capas ocultas, donde la cantidad de capas ocultas determina la profundidad del modelo . El incremento en el desarrollo de los algoritmos de aprendizaje profundo en los últimos años es debido a su capacidad de manejar conjuntos de datos más grandes y disponibilidad de datos por los diversos sistemas de almacenamiento de datos que se disponen en la actualidad. En el futuro se espera que estas redes neuronales se aplicarán a más tareas en el futuro y se espera que sus desempeños mejorar mediante avances en algoritmos de optimización y diseño de modelos[4].

#### 5.2.2.2. Generative adversarial network (GAN)

El marco propuesto para GANs en 2014 por Goodfellow [47] contiene dos redes neuronales convolucionales entrenadas simultáneamente a través de un proceso de confrontación; generador ( $G$ ) y discriminador ( $D$ ), en la Figura 5.1 se muestra una estructura GAN. El generador está capacitado para generar datos con la misma distribución que los datos de entrenamiento mientras que el discriminador es entrenado como adversario del generador, para poder distinguir entre datos y datos generados por  $G$  [47]. Al entrenar ambas redes simultáneamente, el generador mejora hasta que el discriminador ya no puede diferir entre el real y datos falsos [47]. El nivel del generador está determinado por la experiencia del discriminador. El aprendizaje de GAN se puede expresar de una manera teórica de juegos. Sea  $p_x$  la distribución de datos de entrenamiento y  $p_z$  una distribución de ruido aleatorio. La muestra de datos de entrenamiento  $x$  y el ruido aleatorio  $z$  se pueden representar como un valor, vector o tensor. El objetivo del generador es generar datos  $G(z)$ , por lo que imita a  $x$ , el objetivo del discriminador es clasificar correctamente datos reales

$x$  y datos falsos  $G(z)$ , de modo que la probabilidad  $D(x)$  se aproxima a 1 y la probabilidad  $D(G(z))$  se aproxima a 0. Matemáticamente el anterior planteamiento se puede expresar como [46],

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_x} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))] \quad (5.1)$$

donde  $V$  es la función de valor. Cuando  $D(G(z)) = \frac{1}{2}$ , el discriminador no puede determinar si  $G(z)$  proviene de la distribución de datos de entrenamiento o de distribución de datos; este estado es el óptimo global para la tarea de optimización de la ecuación 5.1.

El modelo de GAN condicional (cGAN) fue propuesto con el objetivo de poder controlar los datos generados por la red (figura 5.1). Acondicionando tanto el generador como el discriminador con alguna información auxiliar, como por ejemplo una clase etiqueta, el modelo GAN condicional puede generar datos con la misma etiqueta. Al realizar el condicionamiento, la información se alimenta en ambas redes como una capa de entrada adicional [47]. El juego minmax resultante entre el generador y el discriminador de la GAN condicional se puede describir como,

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x | y)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z | y)))] \quad (5.2)$$

donde tanto  $x$  como  $z$  están condicionados con la condición  $y$ . El modelo de GAN condicional es una contribución clave para las aplicaciones en traslación de dominio de imágenes utilizando redes adversarias.

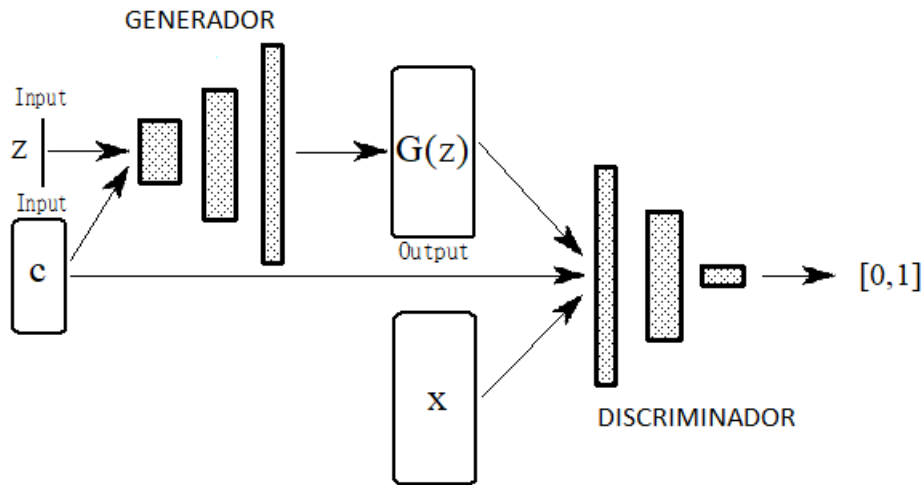


Figura 5.1: Arquitectura CGAN, contiene a un discriminador, un generador, la entrada  $c$  que alimenta tanto el generador como el discriminador, el conjunto de entrada  $z$  la imagen objetivo  $x$ . La estructura original no tiene la entrada( $c$ ) [46]

En los próximos párrafos, se describen un modelo derivados de la GAN original llamado pix2pix, este es utilizado para la traslación del estilo de imagen a imagen en parejas, objeto

fundamental de esta investigación. La traslación de imagen a imagen es la tarea de tomar imágenes de la fuente dominio visual y traducirlos al dominio visual de destino, para que tengan las mismas características (estilo o representación) que el dominio de destino. Los modelos de GAN derivados se pueden dividir en dos grupos principales: GAN basados en la optimización de la arquitectura (por ejemplo, DCGAN) y GAN basados en la optimización de la función objetivo (por ejemplo, WGAN-GP) [46]. Este trabajo, solo está centrado en GANs basadas en la optimización de la arquitectura, debido a que la arquitectura CGAN, estructura base del pix2pix, pertenece a dicha clasificación. Por otra parte, este trabajo no se centra en afrontar problemas como la estabilidad de las GANs característica principal de los modelos GAN pertenecientes a la optimización de la función objetivo como WGAN.

### 5.2.2.3. Imagen a imagen

Las GAN condicionales se han propuesto como una solución de propósito general para la traslación de imagen a imagen [48]. El generador se presenta con una imagen del dominio  $X$  con el propósito de traducir las imágenes en el dominio  $Y$  donde los resultados deberían ser indistinguibles de las reales imágenes del dominio  $Y$ . Para fomentar un menor desenfoque de las imágenes generadas, se agregó un término de pérdida  $L1$ , el error absoluto, a la pérdida adversaria original utilizado en GAN. La pérdida total del modelo propuesto se puede describir como,

$$G^* = \arg \min_G \max_D \mathbb{E}_y[\log D(y)] + \mathbb{E}_{x,z}[\log(1 - D(G(x, z)))] + \mathbb{E}_{x,y,z} [\|y - G(x, z)\|_1] \quad (5.3)$$

Los autores mostraron con éxito que el modelo propuesto, llamado Pix2Pix, podría realizar una buena traducción de imagen a imagen con imágenes generadas de aspecto real en una variedad de conjuntos de datos, incluida la traducción de etiquetas semánticas en fotos, mapas en fotografías aéreas e imágenes en blanco y negro en imágenes en color [48].

### 5.2.2.4. Pix2pix

El modelo pix2pix derivado de CGAN es una arquitectura (Figura 5.2 ) optimizada para la traslación de imagen a imagen de un dominio a otro. Sin embargo, el aprendizaje de imagen a imagen con pix2pix está supervisado y, por lo tanto, pix2pix requiere pares de ejemplos de entrenamiento estrictamente alineados con píxeles[48]. El objetivo de pix2pix es aprender el mapeo del dominio  $X$  al dominio  $Y$  y dados los pares de muestras de entrenamiento  $\{(x_i, y_i)\}_{i=0}^N$ , donde  $x_i \in X$  y  $y_i \in Y$  son muestras de imagen correspondientes. Pix2pix consta de un generador  $G : X \rightarrow Y$  y un discriminador  $D : X \times Y \rightarrow [0, 1]$ . Como pix2pix se deriva de CGAN, necesita información adicional común para el generador y el discriminador, que es la imagen del dominio de entrada  $x$ , y, por lo tanto, el discriminador también recibe la imagen  $x$  [46].

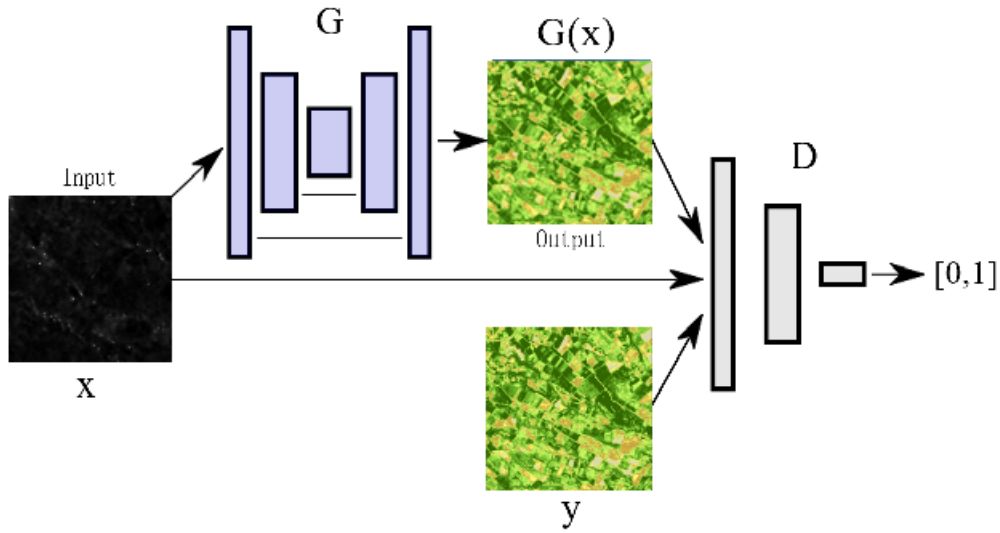


Figura 5.2: Arquitectura pix2pix, la figura contiene el generador(G), el discriminador(D), y toma como entrada del generador una imagen satelital SAR(X) y como una entrada del discriminador una NDVI(Y) [46]

### 5.2.2.5. Pérdida

Considerando el generador definido por Goodfellow [47], es importante notar que  $G$  no toma un vector de ruido. Aún así,  $G$  puede aprender el mapeo de  $X$  a  $Y$ , resultando en  $G$  para producir salidas deterministas. Para evitar esto, se podría agregar un ruido gaussiano  $z$  como entrada adicional con  $x$ . Desafortunadamente, esta estrategia es ineficaz ya que  $G$  aprende a ignorar el ruido  $z$ , por lo que no hay diferencia entre aprender  $G : X \rightarrow Y$  y  $G : X \times Z \rightarrow Y$ , donde  $Z$  es el conjunto de todos los vectores de ruido latente[36]. Para tener ruido en  $G$ , experimento con la deserción en arquitecturas de generador de la misma manera que lo implementaron [46]. Cuando la ecuación 5.2 se adapta para la traducción de imagen a imagen del dominio  $X$  al  $Y$ ,  $G$  y  $D$  juegan el siguiente juego:

$$\begin{aligned} \min_G \max_D V(D, G) = & \mathbb{E}_{x \sim p_X, y \sim p_Y} [\log D(y, x)] \\ & + \mathbb{E}_{x \sim p_X} [\log(1 - D(G(x), x))] \end{aligned} \quad (5.4)$$

Entonces, la pérdida de pix2pix se expresa como,

$$\begin{aligned} \mathcal{L}_{CGAN}(D, G) = & \mathbb{E}_{x \sim p_X, y \sim p_Y} [\log D(y, x)] \\ & + \mathbb{E}_{x \sim p_X} [\log(1 - D(G(x), x))] \end{aligned} \quad (5.5)$$

Cuando  $G$  intenta engañar a  $D$ ,  $G$  puede producir imágenes falsas  $G(x)$  que confunden a  $D$ , sin similitud visual entre  $G(x)$  e  $y$  [48]. Trabajos anteriores resuelven esto utilizando una pérdida con regularización. El enfoque de regularización más común es para mezclar el  $LCGAN(D, G)$ (pérdida del pix2pix) con la distancia  $L2$ . Tenga en cuenta que la pérdida

regularizada diferencia sólo para el  $G$ , el paso de optimización para el  $D$  permanece sin cambios. Sin embargo, la pérdida regularizada con la tarea de minimizar la distancia  $L2$  entre las imágenes falsas y reales dan como resultado imágenes borrosas, porque la distancia  $L2$  se minimiza promediando todos los resultados factibles. Este problema está resuelto con regularización de distancia  $L1$  en lugar de regularización de distancia  $L2$ . La regularizar la pérdida de  $G$  es [46]:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x \sim p_X, y \sim p_Y} [\|y - G(x)\|_1] \quad (5.6)$$

Tenga en cuenta que no es necesario utilizar el discriminador con regularización. La tarea de optimización resultante se puede expresar como:

$$\min_G \max_D \mathcal{L}_{CGAN}(D, G) + \lambda \mathcal{L}_{L1}(G) \quad (5.7)$$

donde  $\lambda \in R$  es el hiperparámetro del multiplicador de regularización

#### 5.2.2.6. Convergencia

Encontrar el equilibrio entre  $G$  y  $D$  es un problema complejo. Es probable que los métodos numéricos de descenso de gradiente para GAN fallen. Específicamente, de la tarea de optimización de la ecuación 5.1, asumiendo que  $D$  no es una función lineal o afín, podemos ver que la función de valor no es convexa, porque  $D(x)$  y  $-D(G(z))$  son funciones opuestas que no pueden ser convexas simultáneamente. Por lo tanto, encontrar el mínimo de la función general no convexa con parámetros continuos de alta dimensión es un problema difícil.

#### 5.2.2.7. Colapso de modo

El generador puede caer en colapso de modo (la red cae en la configuración de parámetros cuando devuelve la misma salida para una entrada arbitraria), donde el gradiente de la función de pérdida del generador se acerca a cero que da como resultado una retro propagación vacía. En el entrenamiento pix2pix, dado que la red se entrena con una fuente emparejada e imágenes de dominio de destino, es menos probable que el generador pix2pix entre en modo colapso [46].

#### 5.2.2.8. Arquitectura

Se requiere una estructura para el generador, el cual renderice imágenes con la misma resolución que la imagen de entrada y también que preserve la estructura de la imagen de entrada. Una red generadora que cumple estos requisitos es una red codificador-decodificador, donde la información de entrada se muestrea a través de una serie de capas

hasta que llega a la capa de cuello de botella desde la que se muestra en forma inversa a la disminución de resolución. Los autoencoders cumplen con las condiciones antes mencionadas, pero tienen el problema que información relevante podría perderse en la capa de cuello de botella. En consecuencia para preservar la estructura de la imagen de entrada, se ha planteado utilizar estructuras avanzadas como la arquitectura Unet, la cual utiliza conexiones de entre el submuestreo y su respectiva capa de upsampling para solucionar el problema del cuello de botella, adicionalmente se plantean utilizar arquitecturas como la ResNet y la Unet++ [46] .

En el discriminador está diseñado de tal manera que hace una predicción en muchos parches de un determinado tamaño para tomar la decisión de discriminación (PatchGAN). Se supone que esto permite predicciones más precisas. Consiste en una CNN de 5 capas de 64, 128, 256, 512 y 512 filtros con kernels 4x4 y 2x2 strides. La normalización por lotes y las activaciones ReLU con fugas aparecen en cada una de estas capas. Luego, una capa de aplanamiento con activación sigmoidea. Los parches son 70x70 debido al tamaño de la entrada, las imágenes de 256x256 píxeles y la combinación del tamaño del kernel y el paso aplicado[48].

#### 5.2.2.9. Autoencoder

Es una red neuronal que está capacitada para intentar copiar su entrada a la salida. Internamente, tiene una capa oculta  $h$  que describe un código utilizado para representar la entrada. La red está compuesta de dos partes: una función del codificador  $h = f(x)$  y un decodificador que produce una reconstrucción  $r = g(h)$ . Si un autoencoder tiene éxito simplemente aprendiendo a establecer  $g(f(x)) = x$  en todas partes, entonces no es especialmente útil. En cambio, los codificadores automáticos están diseñados para no poder aprender a copiar perfectamente. Por lo general, están restringidos de manera que les permite copiar solo aproximadamente, y copiar solo la entrada que se parece a los datos de entrenamiento. Debido a que el modelo se ve obligado a priorizar qué aspectos de la entrada deben copiarse, a menudo aprende propiedades útiles de los datos [48].

#### 5.2.2.10. Unet

El modelo de generador para Pix2Pix GAN se implementa como un autoencoder especial llamado U-Net. El modelo U-Net es un modelo de codificador-decodificador (Figura 5.3) que se utiliza para la traslación de imágenes a imágenes y segmentación, en el que se utilizan conexiones de salto para conectar capas en el codificador con las capas correspondientes en el decodificador que tienen mapas de características del mismo tamaño. La parte del codificador del modelo se compone de capas convolucionales que utilizan strides de  $2 \times 2$  para reducir la resolución de la imagen de la fuente de entrada hasta una capa de cuello de botella. La parte del decodificador del modelo lee la salida del cuello de

botella y usa capas convolucionales de transposición para aumentar la muestra al tamaño de imagen de salida requerido [48].

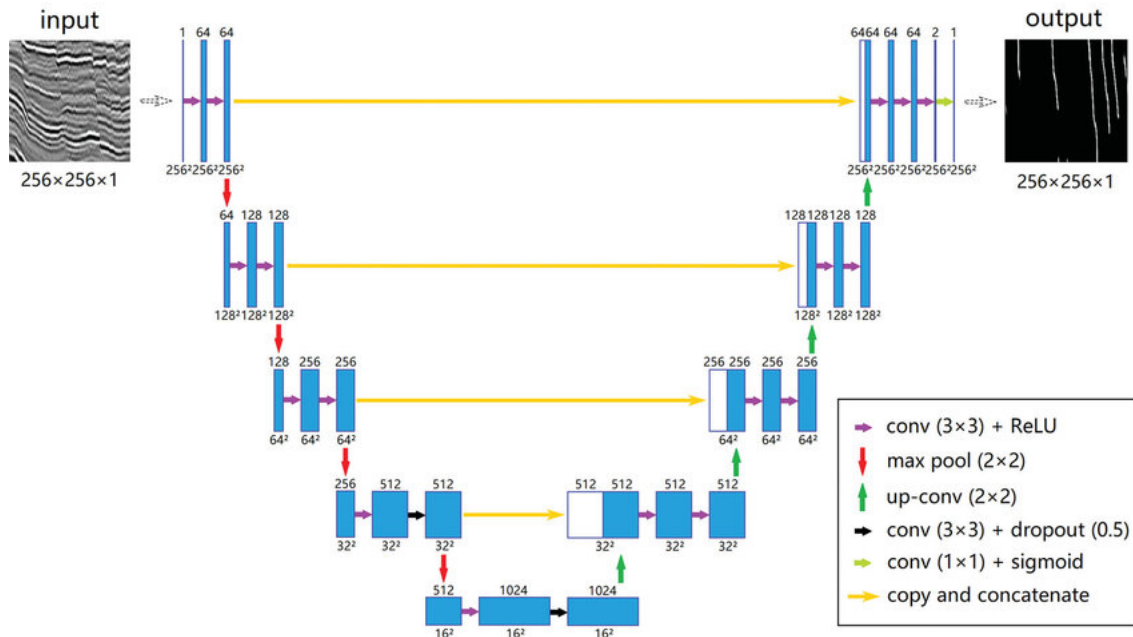


Figura 5.3: Arquitectura de una red U-Net. Esta contiene una dos redes convolucionales que hacen de codificador y decodificador, en su union hay un cuello de botella, las rayas amarillas muestran los saltos, característica fundamental en estas red. En el cuadro se muestran algunas características de diseño mas profundo relacionados con conceptos de deep learning [48]

### 5.2.2.11. Redes basadas en residuos para imágenes(ResNet)

Las redes residuales profundas han conquistado el mundo del aprendizaje profundo. La solidez de ResNets ha sido probada por varias tareas de reconocimiento visual y por tareas no visuales que involucran el habla y el lenguaje. La profundidad de la red es de crucial importancia en las arquitecturas de redes neuronales, pero las redes más profundas son más difíciles de entrenar. El marco de aprendizaje residual facilita la formación de estas redes y les permite ser mucho más profundas, lo que lleva a una mejora rendimiento en tareas visuales y no visuales. Estas redes residuales son mucho más profundas que sus contrapartes "simples", pero requieren un número similar de parámetros (pesos). En la Figura 5.4 se muestra la estructura en medio de las capas y que permite las conexiones de salto o atajos para brincar sobre algunas capas. Esta estructura se puede utilizar en una red de autoencoder y así, utilizarse para realizar un generador de una GAN para la tarea de translación de imagen a imagen[48].

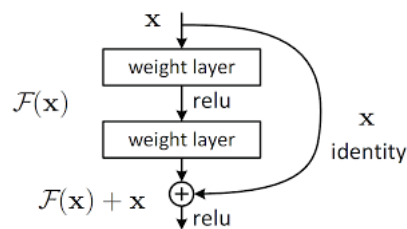


Figura 5.4: Estructura principal de una resnet, en esta se muestra una entrada  $x$ , esta entra directamente a la primera capa para ser entrenada, pero a su vez se mantiene para hacer una realimentación de la información [48]

### 5.2.2.12. Unet++

UNet ++ consta de un codificador y decodificador que están conectados a través de una serie de bloques convolucionales densos anidados (Figura 5.5). La idea principal detrás de UNet ++ es tender un puente entre los mapas de características del codificador y el decodificador antes de la fusión. En el resumen gráfico, el negro indica la U-Net original, verde y azul muestran bloques densos de convolución en las vías de salto, y el rojo indica una supervisión profunda. Los componentes rojo, verde y azul distinguen a UNet ++ de U-Net[49]. Esta estructura según sus autores presenta una mejor actuación para la segmentación en imágenes médicas que su antecesora Unet. En esta investigación adaptaremos esta estructura como generador, para observar si las buenas prestaciones de la arquitectura sirven para mejorar los valores de evaluación.

## 5.2.3. Evaluación

La evaluación del GAN es difícil porque el objetivo del generador cambia en cada iteración y la calidad de la imagen es subjetiva de definir. Los generadores no utilizan una función objetivo, sino un discriminador, en lugar de ser entrenados directamente, lo que dificulta la comparación entre ellos [4]. El examen visual humano de las muestras generadas es la forma más sencilla, pero lenta y costosa de evaluar los GANs ya que la definición de la calidad de la imagen generada es subjetiva y no está definida matemáticamente. Sin embargo, ninguna de las métricas propuestas está de acuerdo en ser el único punto de referencia común de los GAN, utilizado para la comparación de los modelos generadores, capturando todas sus fortalezas y debilidades[36].

### 5.2.3.1. Error cuadrático medio

El error cuadrático medio (RMSE o rmse), también conocido como desviación media cuadrática, es una medida dependiente de la escala que suele utilizarse para estimar la diferencia entre los valores predichos por un modelo y los valores realmente observados[4]. Además, el RMSE se ha utilizado ampliamente para evaluar el rendimiento de las técnicas

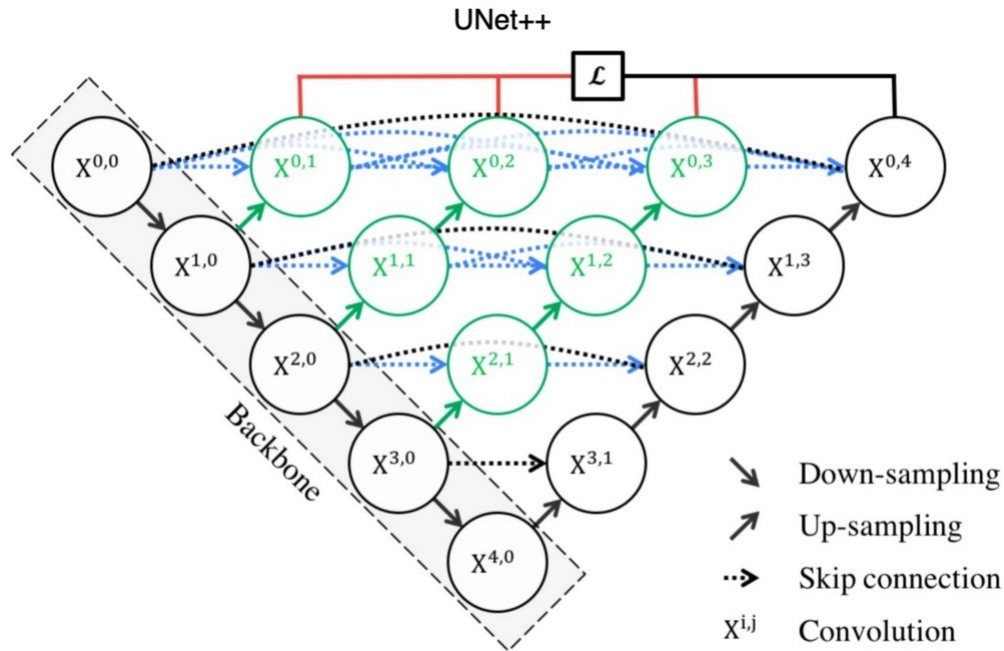


Figura 5.5: Arquitectura unet++, en la cual se puede observar una estructura similar que la unet convencional, con la gran diferencia de que se hacen conexiones anidadas hasta llegar a la capa final, la letra L se usa para nombrar esta estructura con el valor de capas anidadas más la final. También tenemos la información de configuración de la red [49]

de fusión de imágenes midiendo la desviación de los valores de los píxeles de la imagen fusionada en contraste con la imagen de referencia. Dada la imagen de referencia  $X$  y la imagen fusionada  $X'$  con un tamaño de  $N \times M$ , el RMSE entre ellas puede calcularse como

$$rmse = \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^M (x_{i,j} - x'_{i,j})^2}{N \times M}} \quad (5.8)$$

Debido a la naturaleza del cuadrado medio, el valor del RMSE es no negativo, con un valor perfecto de cero. El rango utilizado por esta métrica es de 0 a infinito y no tiene unidades. Cuanto mayor sea el valor del RMSE, más graves serán las distorsiones espectrales en la imagen fusionada de la imagen fusionada con respecto a la de referencia. En la mayoría de los casos, el RMSE se utiliza junto con  $r$  para describir mejor las similitudes entre las imágenes de referencia y las fusionadas

### 5.2.3.2. Índice de similitud estructural

El índice de similitud estructural fue desarrollado para evaluar la similitud de imágenes. Como criterio de percepción para evaluar similitud de imagen, SSIM refleja adecuadamen-

te la distorsión percibida por un observador humano[4]. Cuando se aplica para evaluar la imagen fusionada, permite la comparación de los patrones locales de intensidades de píxeles entre la referencia y el fusionado imágenes. Matemáticamente, SSIM se define como,

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (5.9)$$

donde,  $C_1$  es una constante que se incluye para evitar la inestabilidad cuando  $\mu_x^2 + \mu_y^2$  está cerca de cero y  $C_2$  es una constante que se incluye para evitar la inestabilidad cuando  $\sigma_x^2 + \sigma_y^2$  está cerca de cero.  $\mu_x$  y  $\mu_y$  denota las intensidades medias de la referencia y las imágenes fusionadas, mientras que  $\sigma_x$  y  $\sigma_y$  denotan desviaciones estándar. El rango utilizado por esta métrica es de 0 a 1 y no tiene unidades. Para el índice SSIM, los valores más cercanos a 1 indican una buena conservación de detalles espaciales, con una puntuación de 1 cuando ambas imágenes idénticas[4].

### 5.2.3.3. Spectral Angle Mapper

El spectral angle mapper (SAM) está motivado por observación que cambia en las condiciones de iluminación debido a factores como sombras, pendiente la variación, la posición del sol y la nube de luz, dan como resultado cambios en solo la magnitud de un píxel vector, en lugar de la dirección [4]. Por lo tanto, SAM se ha utilizado ampliamente para evaluar la fidelidad espectral de las imágenes fusionadas calculando los ángulos espectrales entre la imagen de referencia y la imagen fusionada para medir las distorsiones espectrales. Matemáticamente, SAM se puede expresar como,

$$SAM \triangleq \arccos \left( \frac{\langle x_{i,j} \cdot x'_{i,j} \rangle}{\|x_{i,j}\| \|x'_{i,j}\|} \right) \quad (5.10)$$

donde  $x_{i,j}$  y  $x'_{i,j}$  denotan los valores espectrales en el píxel  $(i, j)$  en la referencia y el imágenes fusionadas. Comúnmente, SAM se puede medir en grados o radianes y es generalmente promediado sobre toda la imagen para producir una medición global de la distorsión espectral. El rango utilizado por esta métrica es de 0 a 90 en unidades de grados y de 0 a 1 en unidades de radianes. Cuanto menor sea el valor de SAM, menor será la distorsión espectral.

### 5.2.3.4. Relación señal-ruido máxima

La relación señal-ruido máxima (PSNR) es un término de ingeniería con respecto a la relación entre la potencia máxima posible de una señal y la potencia de ruido corruptor que afecta la fidelidad de su representación. En la fusión de imágenes, PSNR se aplica comúnmente para evaluar la distorsión radiométrica de la imagen fusionada en comparación con la referencia uno, asumiendo que la distorsión solo es causada por ruido aditivo

independiente de la señal [4]. Para los fusionados imagen  $X'$  y la imagen de referencia  $X$  de tamaño  $N \times M$ , PSNR se puede calcular como,

$$PSNR = 10 \log_{10} \left( \frac{(\text{máx}(X'))^2}{\sum_{i=1}^M \sum_{j=1}^N (x(i, j) - x'(i, j))^2} \right) \quad (5.11)$$

donde  $x(i, j)$  y  $x'(i, j)$  son valores de píxeles de la imagen de referencia  $X$  y la imagen fusionada  $X'$ , respectivamente.  $\text{máx}(X')$  denota el valor máximo de píxeles de la imagen fusionada (por ejemplo, 255 para Imágenes de 8 bits). El rango utilizado por esta métrica es de 0 a infinito en unidades de decibel (db). En general, un valor PSNR más grande indica un mejor rendimiento de la fusión. Por ejemplo, los valores típicos de PSNR en una compresión con pérdidas de imagen y video oscilan entre 30 y 50 dB, para una profundidad de 8 bits, donde un mayor valor de PSNR es resultado de una mejor compresión. Para 16 bits de profundidad, el valor del PSNR oscila entre 60 y 80 dB [50].

## Capítulo 6

# Datos y área de estudio

---

El capítulo anterior introdujo terminología básica de las preprocesamiento de imágenes satelitales, GANs y evaluación de imágenes. También presentó el estado del arte según literatura. Este capítulo proporcionará información sobre los conjuntos de datos utilizados para responder las preguntas de investigación esbozadas en esta tesis, y también describe la áreas de interés elegida.

### 6.1. Dataset

Los productos de datos de Sentinel-1(s1) y Sentinel-2(s2) están disponibles gratuitamente en la página web de la organización Copernicus Open Access Hub. Los usuarios registrados pueden elegir el lugar, la fecha y otros parámetros requeridos para descargar los datos. Teniendo en cuenta, que para hacer la traslación de imagen a imagen se necesitan imágenes SAR y ópticas del mismo día y ante la imposibilidad de obtenerlas, ya que por las distintas órbitas que manejan es complejo que los satélites pasen el mismo día y en el mismo lugar, se tomaron imágenes que tienen fechas de no más de cinco días de distancia. Por lo tanto, se construyeron los siguientes grupos de pares de imágenes: s1 2019-07-15 con s2 2019-07-11, s1 2020-01-09 con s2 2020-01-07, s1 2020-02-02 con s2 2020-02-06 y s1 2020-03-21 con s2 2020-03-22. Con estos grupos se asegura información del sitio en diferentes épocas del año. El tamaño de las imágenes individualmente es 256 \* 256 píxeles, y los pares de imágenes tienen 256\*512 píxeles. Los grupos para el entrenamiento, validación y prueba fueron escogido a través de un algoritmo de selección aleatoria, el cual tomó el 30 % de imágenes del total como grupo de validación y el 70 % como grupo de entrenamiento, adicionalmente se eligieron 32 aleatoriamente para ser imágenes de prueba, en consecuencia los grupos de entrenamiento, validación y prueba quedaron compuestos por 680 , 250 y 32 imágenes respectivamente. La base de datos se encuentra en el siguiente enlace: (<https://github.com/David121189/GANs-for-agriculture>). Ahora se expondrán las características principales de estas imágenes satelitales.

#### 6.1.1. Sentinel 1

Es una constelación de satélites, cuyos mas representativos son: Sentinel-1A lanzado el 3 de abril de 2014 y Sentinel-1B lanzado el 25 de abril de 2016 . Ambos satélites vuelan en la misma órbita circular síncrona con el sol, a una altura de 693 km, con un ciclo de

repetición de 12 días para un satélite y de 6 días para el par (ESA, 2012, 2013). El instrumento SAR funciona en la banda C, que se encuentra dentro del rango de frecuencia de 4 a 8 GHz (longitud de onda de 7,5 a 3,75 cm), en la porción de microondas del espectro electromagnético [51].

Las imágenes entregadas por los sentinel -1 se utilizan en una amplia gama de aplicaciones debido a su cómodo acceso a datos y la gratuidad de los mismos. Los datos se pueden recopilar en cuatro modos: modo Stripmap (SM), modo de onda (WV), modo de barrido extra ancho (EW) y modo de franja ancha interferométrica (IW) [51].

En este trabajo solo se utiliza el modo IW. El IW tiene una franja de hasta 250 km con resolución espacial de 5 x 20 m. Este modo usa Terrain Observación con escaneos progresivos SAR (TOPSAR) para crear una mayor calidad datos. El modo de onda está disponible en polarización única (VV o HH) y modos en polarización simple y dual (VV + VH o HH + HV).

## **6.1.2. Sentinel 2**

Sentinel-2, similar a Sentinel-1, es una misión satelital desarrollada por la ESA Iniciativa de Copérnico. Se trata de dos satélites (Sentinel-2A y Sentinel-2B) en lados opuestos de la órbita sincrónica con el sol a una altitud de 786 km. Sentinel- 2A se lanzó el 23 de junio de 2015 y Sentinel-2B el 7 de marzo de 2017. Cada uno de ellos llevan un instrumento multiespectral (MSI) que se recolecta en 13 bandas espectrales a 10 m (4 bandas), 20 m (6 bandas) y 60 m de resolución espacial (3 bandas) (ESA, 2015b). Sentinel-2 cubre áreas terrestres y costeras de 56° Sur a 84° Norte con una franja orbital ancho igual a 290 km y un tiempo de revisión de 10 días, o 5 para el par Sentinel-2. Los datos se pueden utilizar para aplicaciones tales como monitoreo de agua, bosques y vegetación. observación, así como gestión de desastres naturales e infraestructura, frontera, y vigilancia marítima [51].

### **6.1.2.1. NDVI**

A través de la combinación de las bandas de Sentinel 2 se puede obtener el NDVI (Normalizado Diferencial de Vegetación), uno de los índices más utilizados para determinar la vigorosidad de las plantas. Los valores del NDVI están en función de la energía absorbida o reflejada por las plantas en diversas partes del espectro electromagnético. Para determinar dicha cantidad de energía se relacionan una banda del espectro visible, especialmente la roja y el infrarrojo cercano (NIR) mediante la siguiente fórmula expuesta en 6.1. El valor resultante caracteriza el comportamiento radiométrico de la vegetación, relacionado con la actividad fotosintética y la estructura foliar de las plantas, en consecuencia, se permitiría obtener una buena representación para determinar la calidad de

salud de la planta [52].

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \quad (6.1)$$

En las imágenes Sentinel -2 de la respuesta espectral que tiene la vegetación sana, muestra un claro contraste entre la banda 4(banda roja), y la banda 8 (NIR). Mientras que en la banda 4 los pigmentos de la hoja absorben la mayor parte de la energía que reciben, en la banda 8, las paredes de las células de las hojas, que se encuentran llenas de agua, reflejan la mayor cantidad de energía, esto generalmente se presenta gráficamente en una tonalidad de verde más intenso (Figura 6.1). Por otra parte, cuando la vegetación sufre algún tipo de estrés o daño, la cantidad de agua disminuye en las paredes celulares por lo que la reflectividad disminuye en la banda 8 y aumenta paralelamente en la banda 4 al tener menor absorción clorofílica. Habitualmente está representado por una coloración más roja, resultado de la diferencia en la respuesta espectral que permite separar con relativa facilidad la vegetación sana de otras cubiertas.

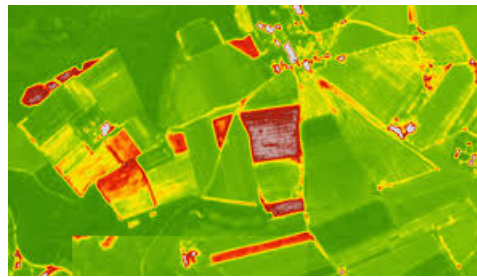


Figura 6.1: Imagen NDVI, donde se muestra la diferentes colores y sus tonalidades. Entre el color sea mas verde mas saludable es la vegetación, en las zonas rojas o no existe vegetación (ciudad) o esta en malas condiciones

### 6.1.3. Area de estudio

El Departamento de Nariño se encuentra ubicado al sur occidente de Colombia. Tiene una superficie de 33.268 km<sup>2</sup> lo que representa el 2,9 % del territorio nacional. Su posición astronómica está entre los 0° 21´ y 2° 40´ de latitud norte (desembocadura del río Iscuandé); y entre los 76° 50´ (cerro Petacas y Ánimas) y 79° 02´ (cabo Manglares) de longitud oeste, del meridiano de Greenwich. El departamento esta ordenado administrativamente en 13 subregiones, siendo San Juan de Pasto, su capital. En total el departamento tiene 64 municipios que se ubican en las subregiones administrativas: Sanquianga, Pacifico Sur, Telembí, Pie de Monte Costero, Exprovincia de Obando, Sabana, Abades, Occidente, Coordillera, Centro, Juanambú, Rio Mayo y Guambuyaco [53].



Figura 6.2: Zona del departamento de Nariño que es tomada por la imágenes satelitales, para posteriormente dividirla y hacer todo el entrenamiento de la red.

Para el objeto de este estudio se tienen en cuenta las subregiones de Obando, Abades, Centro, Juanambú, Río mayo y Guambuyaco (Figura 6.2). Estas regiones unidas tienen una extensión aproximada de 13000 km<sup>2</sup> y alrededor de 1 millón de habitantes. Las actividades económicas más importantes de esta subregión están basadas en el sector agropecuario destacándose el cultivo de papa, maíz, cebolla, yuca, trigo, cebada, frijol, alverja; igualmente es significativa la explotación de ganado bovino, porcino, caballar y especies menores. También cabe resaltar la actividad comercial y artesanal[53].

# Capítulo 7

## Metodología

---

En esta sección se expondrá la metodología para abordar el problema de eliminar la nubosidad por medio de la fusión de imágenes satelitales SAR y ópticas multiespectrales, con el objetivo de generar una imagen NDVI que dentro del área de la teledetección es utilizada para el monitoreo de cultivos. Como se mencionó anteriormente, los algoritmos propuestos para lograr esta fusión son los algoritmos de Deep Learning: GANs y autoencoders. Sobre estos algoritmos girará la parte central de investigación, ya que obtener una arquitectura con buenos resultados en la estimación de la imagen NDVI, puede generar un buen impacto tanto en el sector agrícola y académico. Para esto se tendrán en cuenta las siguientes actividades, con fin de cumplir los objetivos planteados.

### 7.1. Preprocesamiento imágenes satelitales

El preprocesamiento de datos SAR de la imagen sentinel- 1 se realizará utilizando los métodos de procesamiento estándar de la caja de herramientas (SNAP Sentinel Application Platform), en donde se hará la calibración, eliminación de ruido térmico, así como remuestreo a una resolución de  $10 \times 10m^2$  y parametrización multilook para imágenes de polarización VV y VH. El programa de procesamiento de imágenes satelitales ESA SEN2COR se utilizará para la corrección atmosférica y del terreno de la imagen Sentinel-2 para obtener una imagen con menor cantidad de ruido. Se debe realizar un nuevo muestreo basado en interpolación bilineal a una resolución espacial de  $10 \times 10m^2$ , en las bandas 5, 6, 7 y 8a para lograr la misma resolución de las bandas 2, 3, 4 y 8 y las imágenes de radar Sentinel-1A remuestreadas.

### 7.2. Entrenamiento de GAN

En la búsqueda de alcanzar este objetivo, se propone utilizar dentro de las redes neuronales GAN y Autoencoders, diferentes redes neuronales clásicas como CNN para determinar cual es la arquitectura más adecuada para la fusión de las imágenes multi-espectrales y SAR. En estos esquemas también se tuvo en cuenta algunas arquitecturas que han sido utilizadas en otras áreas, las cuales han entregado buenos resultados como: pix2pix, para observar si con modificaciones entregan buenos resultados para determinar la imagen NDVI más exacta.

### 7.2.1. Arquitecturas de discriminador y generador

Las arquitecturas que se utilizaron como generador son la Unet, ResNet y Unet++. Estas estructuras se han adaptado para la actividad de transferencia de estilo de imágenes. La estructura del discriminador es una configuración estándar planteada por Ian Goodfellow [47]. En consecuencia, en este estudio se plantean tres diferentes configuraciones de GAN, manteniendo el mismo esquema en el discriminador y cambiando la configuración del generador. Teniendo en cuenta que las imágenes son de tamaño  $256 \times 256$ , la capa inicial de cada estructura esta diseñada con el tamaño  $256 \times 256 \times 3$ . A continuación se muestran con mas detalles las arquitecturas utilizadas en este trabajo.

Capa	Descripción de cada capa en el discriminador
1	Convolución-(Filters-64, Kernel size-4, Strides-2), Leaky-ReLU
2	Convolución-(Filters-128, Kernel size-4, Strides-2), Batch-Normalization, LeakyReLU
3	Convolución-(Filters-256, Kernel size-4, Strides-2), Batch-Normalization, LeakyReLU
4	Convolución-(Filters-512, Kernel size-4, Strides-2), Batch-Normalization, LeakyReLU
5	Convolución-(Filters-1, Kernel size-4, Strides-1), sigmoid

Tabla 7.1: Arquitectura del discriminador utilizado en esta investigación

Capa	Descripción de cada capa en el generador tipo ResNet-9
1	Convolución-(Filters-64, Kernel size-7, Strides-2), ReLU
2	Convolución-(Filters-128, Kernel size-4, Strides-2), Batch-Normalization, LeakyReLU
3	Convolución-(Filters-256, Kernel size-4, Strides-2), Batch-Normalization, LeakyReLU
4-12	Convolución-(Filters-256, Kernel size-3, Strides-2), Batch-Normalization, ReLU
13	Deconvolución-(Filters-128, Kernel size-4, Strides-2), BatchNormalization, ReLU
14	Deconvolución-(Filters-64, Kernel size-4, Strides-2), Batch-Normalization, ReLU
15	Deconvolución-(Filters-3, Kernel size-7, Strides-1), Tanh

Tabla 7.2: Arquitectura del generador utilizando la configuración ResNet-9 utilizado en esta investigación

Capa	Descripción de cada capa en el generador tipo Unet
1	Convolución-(Filters-64, Kernel size-4, Strides-2), Leaky-ReLU
2	Convolución-(Filters-128, Kernel size-4, Strides-2), Batch-Normalization, LeakyReLU
3	Convolución-(Filters-256, Kernel size-4, Strides-2), Batch-Normalization, LeakyReLU
4-7	Convolución-(Filters-512, Kernel size-4, Strides-2), Batch-Normalization, LeakyReLU
8	Deconvolución-(Filters-512, Kernel size-4, Strides-2), BatchNormalization, Dropout(Rate-0.5), ReLU
9-10	Deconvolución-(Filters-1026, Kernel size-4, Strides-2), BatchNormalization, Dropout(Rate-0.5), ReLU
11	Deconvolución-(Filters-1026, Kernel size-4, Strides-2), BatchNormalization, ReLU
12	Deconvolución-(Filters-512, Kernel size-4, Strides-2), BatchNormalization, ReLU
13	Deconvolución-(Filters-256, Kernel size-4, Strides-2), BatchNormalization, ReLU)
13	Deconvolución-(Filters-1, Kernel size-4, Strides-1), Tanh

Tabla 7.3: Arquitectura del generador utilizando la configuración Unet utilizado en esta investigación

Debido a la complejidad en las conexiones internas de la arquitectura Unet++, las capas no pueden expresarse de igual forma que los anteriores modelos, por lo tanto, se tuvo en cuenta la notación presente en la Figura 5.5, puesto que facilita la descripción de los parámetros en cada sección. La configuración Unet++ se describe a continuación.

Capa	Descripción de cada capa en el generador tipo Unet++
x00	Convolución-(Filters-64, Kernel size-2, Strides-2), Batch-Normalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x10	Convolución-(Filters-128, Kernel size-2, Strides-2), Batch-Normalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x01	Deconvolución-(Filters-64, Kernel size-2, Strides-2)
x01	Convolución-(Filters-concatenación[x00, x01], Kernel size-2, Strides-2), Batch-Normalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2

Continúa en la página siguiente.

Capa	Descripción de cada capa en el generador tipo Unet++
x20	Convolución-(Filters-256, Kernel size-2, Strides-2), Batch-Normalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x11	Deconvolución-(Filters-64, Kernel size-2, Strides-2)
x11	Convolución-(Filters-concatenación[x10, x11] , Kernel size-2, Strides-2), BatchNormalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x02	Deconvolución-(Filters-64, Kernel size-2, Strides-2)
x02	Convolución-(Filters-concatenación[x00, x01, x02] , Kernel size-2, Strides-2), BatchNormalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x30	Convolución-(Filters-512, Kernel size-2, Strides-2), Batch-Normalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x21	Deconvolución-(Filters-64, Kernel size-2, Strides-2)
x21	Convolución-(Filters-concatenación[x20, x21] , Kernel size-2, Strides-2), BatchNormalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x12	Deconvolución-(Filters-64, Kernel size-2, Strides-2)
x12	Convolución-(Filters-concatenación[x10, x11, x12] , Kernel size-2, Strides-2), BatchNormalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x03	Deconvolución-(Filters-64, Kernel size-2, Strides-2)
x03	Convolución-(Filters-concatenación[x00, x01, x02, x03] , Kernel size-2, Strides-2), BatchNormalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x40	Convolución-(Filters-1024, Kernel size-2, Strides-2), Batch-Normalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x31	Deconvolución-(Filters-512, Kernel size-2, Strides-2)
x31	Convolución-(Filters-concatenación[x31, x30] , Kernel size-2, Strides-2), BatchNormalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x22	Deconvolución-(Filters-256, Kernel size-2, Strides-2)
x22	Convolución-(Filters-concatenación[x22, x20, x21] , Kernel size-2, Strides-2), BatchNormalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x13	Deconvolución-(Filters-128, Kernel size-2, Strides-2)

Continúa en la página siguiente.

Capa	Descripción de cada capa en el generador tipo Unet++
x13	Convolución-(Filters-concatenación[x13, x10, x11, x12] , Kernel size-2, Strides-2), BatchNormalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
x04	Deconvolución-(Filters-64, Kernel size-2, Strides-2)
x04	Convolución-(Filters-concatenación[x04, x00, x01, x02, x03] , Kernel size-2, Strides-2), BatchNormalization, LeakyReLU-0.01, maxpooling size - 2, dropout-0.2
Output	Deconvolución-(Filters-4, Kernel size-4), Tanh

Tabla 7.4: Arquitectura del generador utilizando la configuración UNET utilizado en esta investigación

### 7.3. Evaluación de imágenes NDVI producidas por la GAN

Se utilizó una imagen obtenida en otras fases temporales cercanas (imagen de referencia) con imagen producida en el anterior objetivo, para hacer una comparación de imágenes. Con esto se evaluó el rendimiento de los esquemas de fusión mediante métricas estadísticas, para indicar similitudes espectrales o espaciales entre la salida fusionada y la imagen de referencia. Las medidas estadísticas basadas en imágenes de referencia se han utilizado ampliamente en la literatura debido a su eficacia y mayor precisión para demostrar la calidad de la salida fusionada. Para este trabajo se plantean las siguientes métricas estadísticas: Structural similarity index (SSIM), Spectral Angle Mapper (SAM) y Peak Signal-to-Noise Ratio (PSNR).

A continuación se muestra las tres partes de la metodología propuesta en un diagrama de flujo, donde se resume todos los pasos realizados para obtener los resultados de esta investigación.

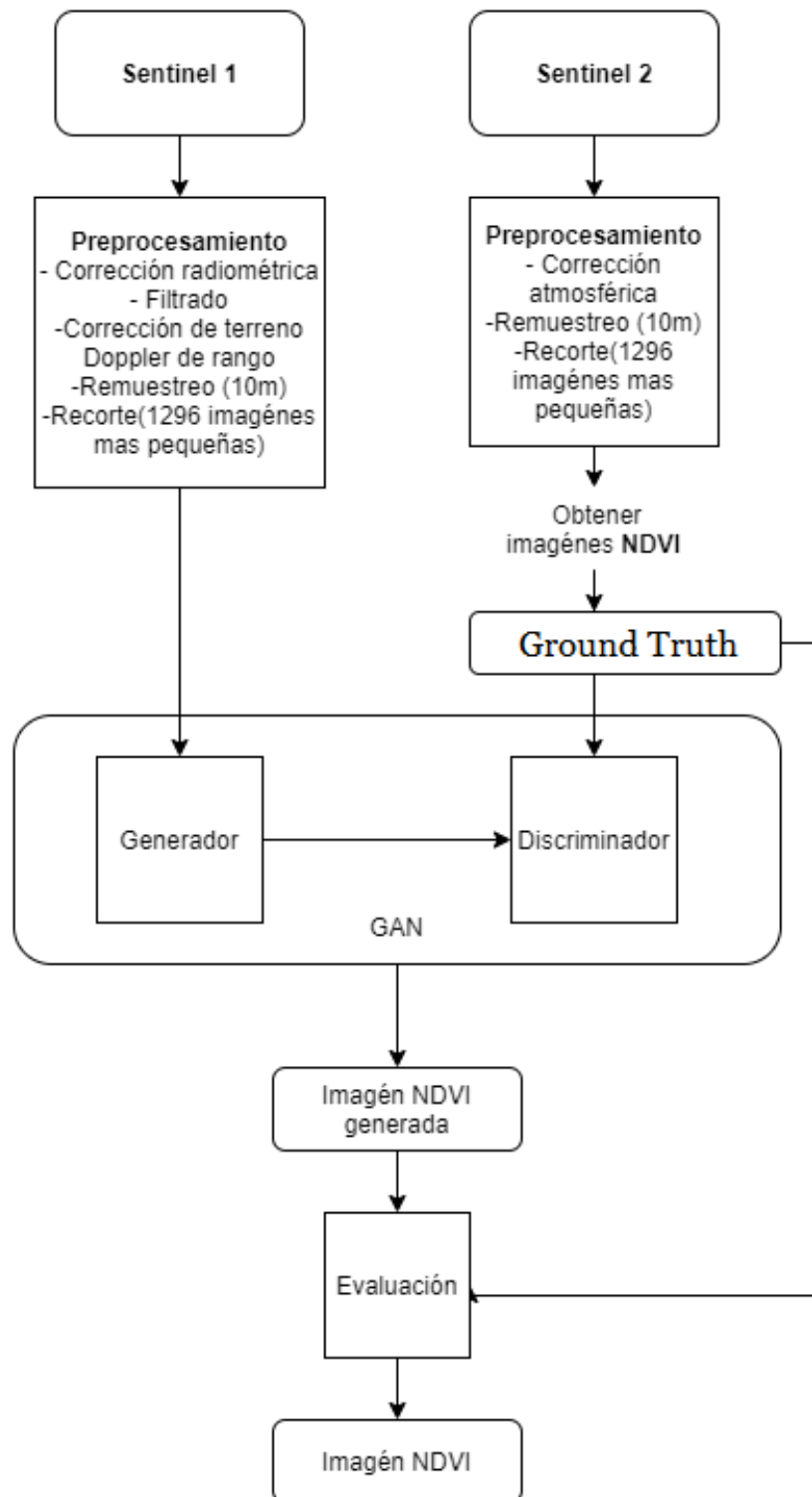


Figura 7.1: Diagrama de flujo de metodología para obtener imágenes NDVI a partir de la fusión de imágenes multiespectrales sentinel 1 y sar sentinel 2, en una GAN

## Capítulo 8

# Resultados y discusión

---

En este capítulo se realizará la exposición de los resultados de la implementación de la metodología propuesta. En los anteriores capítulos se mostraron diferentes definiciones que serán fundamentales para que el lector pueda tener clara la relación de los conceptos y los resultados presentados. La organización del capítulo presenta resultados individuales para cada arquitectura de generador, y finalmente se presenta el enfoque general. Cabe destacar que las pruebas se hicieron con 500 épocas en todos los experimentos, este valor fue determinado después de realizar entrenamientos con múltiplos de 100. Por cada múltiplo, se fueron evaluando las métricas, llegando a la conclusión que después de la época 500 no habían mejores significativas.

### 8.1. Unet++

En el desarrollo de esta arquitectura y debido a que se tenían imágenes de  $256 \times 256$  pixeles, se decidió utilizar la misma estructura expuesta por [47], para construir la estructura Unet. Teniendo esta estructura, se procedió a realizar todas las estructuras anidadas presentes en la Unet++. Todos los parámetros fueron obtenidos a través del entrenamiento partiendo de la base que la mayoría de artículos encontrados en este tema toma los mismos que el artículo [47]. En nuestro caso y al realizar diversos entrenamientos se comprobó que, en el proceso de traslación de imagen a imagen, los parámetros no tienen mucho margen para el cambio.

La primera gráfica para mostrar y analizar es la Figura 8.1 que es una presentación de las pérdidas (loss) del discriminador y generador. En estas se puede ver que el loss del generador a medida que pasan las épocas va decreciendo, pero trata de converger en el valor 3 a partir de la época 400, mientras tanto el discriminador a partir de la época 200 ya empieza a converger alrededor de 1, estos valores son resultado del comportamiento de la ecuación de pérdida del pix2pix (Ecuación 5.5), la convergencia presente asegura estabilidad en el proceso de entrenamiento independientemente del valor al cual converge. Este comportamiento es adecuado porque no se presenta colapso en el entrenamiento (converger a 0), un problema muy común en esta clase de entrenamientos. Además, de que en las gráficas, aunque se presentan ondulaciones o pequeños saltos sobre la línea de entrenamiento, estas no presentan grandes cambios lo que convierte a este entrenamiento como estable dentro de las GANs.

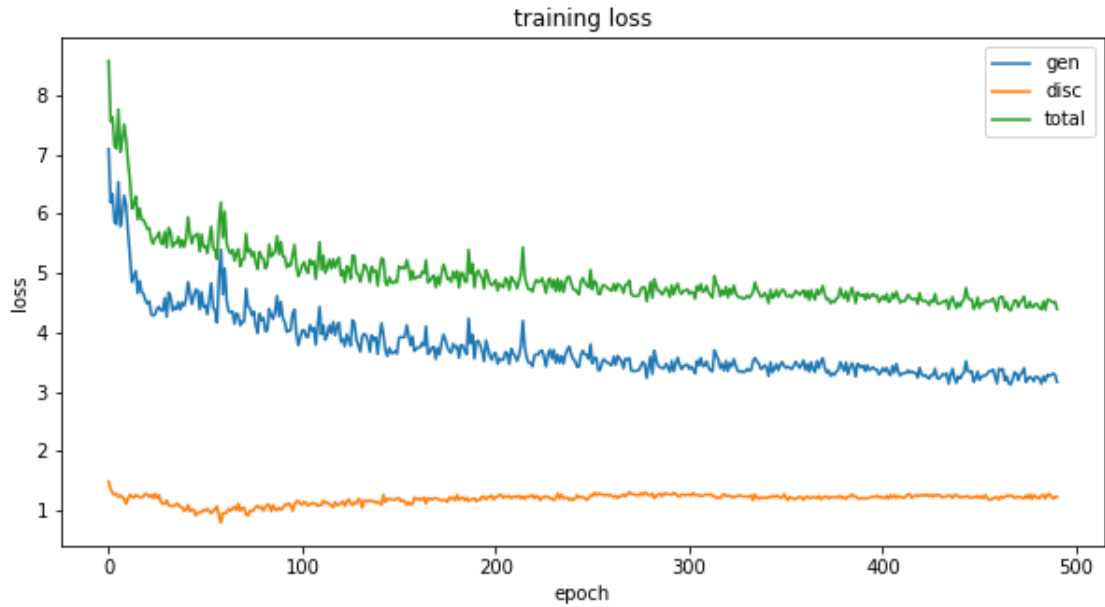
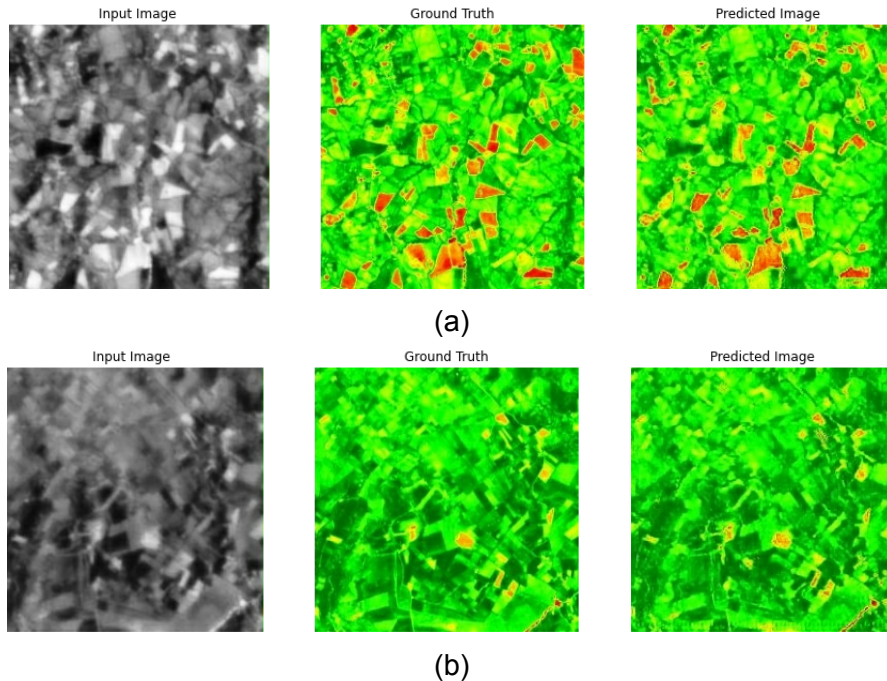


Figura 8.1: Gráfica de las pérdidas del discriminador (Azul), generador(Naranja) y el total(Verde) en Unet++.

Por otra parte en la Figura 8.2 y Tabla 8.1, se presentan algunas imágenes en su última época de entrenamiento con sus respectivos valores en las medidas de evaluación.



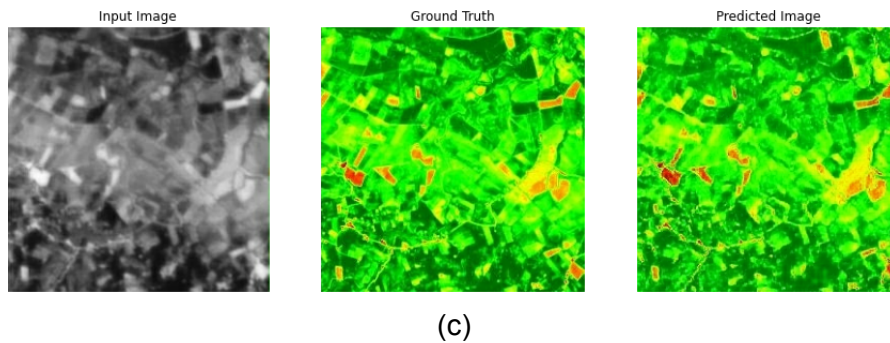


Figura 8.2: Gráficas de ejemplo para Unet++. Como entrada la imagen SAR, en la parte central la referencia (Ground Truth) y en la izquierda la imagen estimada.

Imagen	SSIM	RMSE	SAM [Grados]	PSNR [db]
a	0.662	0.023	63.716	67.683
b	0.670	0.011	61.597	71.892
c	0.676	0.019	65.307	68.652

Tabla 8.1: Valores de las medidas de evaluación para tres imágenes de ejemplo para Unet++

Los valores SSIM están alrededor del 0.66, algo que es relativamente alto en un trabajo de estas características, ya que se trata de llevar una imagen de un dominio a otro. Para el RMSE los valores rondan el 0.015, el valor del error es bajo, esto significa que los valores de los píxeles se acercan bastante al valor original. Entre estas dos medidas vemos que la SSIM es mucho más estricto. Los valores de la métrica SAM(alrededor de 63 grados) son altos debido a que, aunque tienen características similares en cuanto a la coloración (valores de los pixeles), se esta haciendo la comparación de dos imágenes con orígenes espectrales diferentes. Para el PSNR (valores entre el 67db y 72db) los valores se consideran buenos, si se tiene en cuenta que en una actividades como la compresión de imágenes y vídeo resultados entre 60 db y 80 db son considerados óptimos.

## 8.2. Unet

Se decidió utilizó la misma estructura expuesta por [47]. Todos los parámetros se basan en el artículo de Goodflew de 2014 [47]. Para la gráfica de las pérdidas de discriminador y generador (figura 8.3) se observa que la pérdida de generador es creciente, aunque tiende a estabilizarse en 6, el discriminador tiende a estabilizarse en un valor muy pequeño. Este comportamiento nos indica que el generador se está imponiendo sobre el discriminador, esto puede provocar que el generador pueda producir imágenes basura y que el discriminador no sea capaz de distinguirlo. Esto en gran parte se puede ver en las ondulaciones

y los fuertes saltos en la línea del entrenamiento.

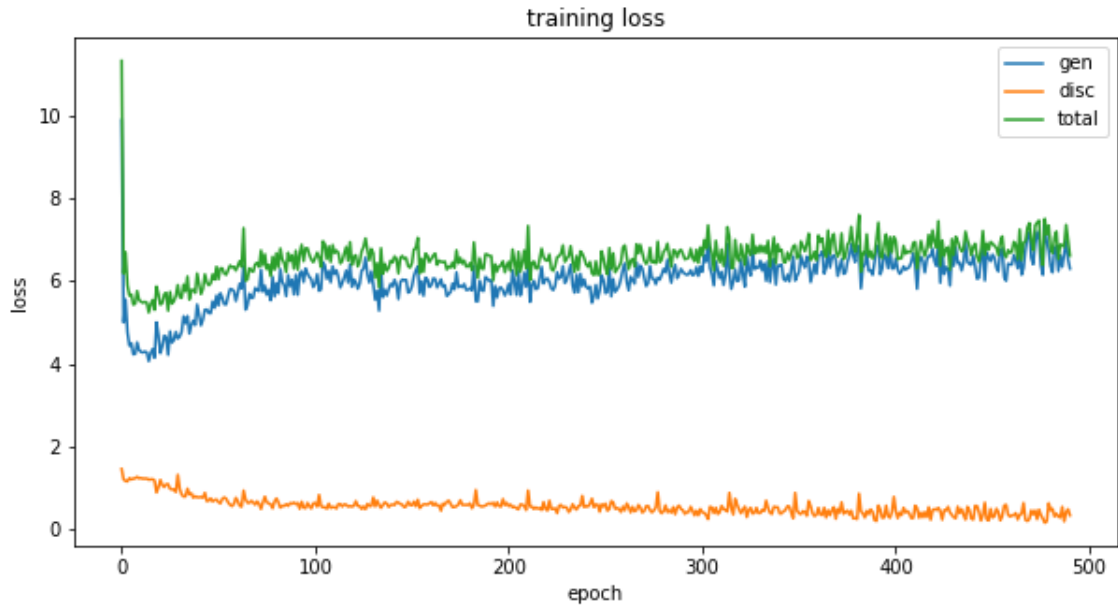


Figura 8.3: Gráfica de las pérdidas del discriminador (Azul), generador(Naranja) y el total(Verde) en Unet.

Imagen	SSIM	RMSE	SAM [Grados]	PSNR [db]
a	0.582	0.030	66.110	66.128
b	0.602	0.017	61.632	69.748
c	0.603	0.027	67.138	66.768

Tabla 8.2: Valores de las medidas de evaluación para tres imágenes de ejemplo para Unet

En la Tabla 8.2 se puede ver que los valores en general son más pobres que los de la red Unet ++. En términos del SSIM el valor medio se encuentra alrededor 0.59, valor que no se puede considerar bajo, pero que se ve superado por el anterior, para el RMSE los valores están alrededor del 0.025 que es bajo. Con respecto a la SAM y PSNR los valores son buenos. Pero como medida principal se tendrá el SSIM ya que nuestro objetivo es tratar que la imagen producida tenga la mayor cantidad de información parecida a la referencia.

Por otra parte, se presentan las imágenes correspondiente a los valores presentados en la anterior tabla.

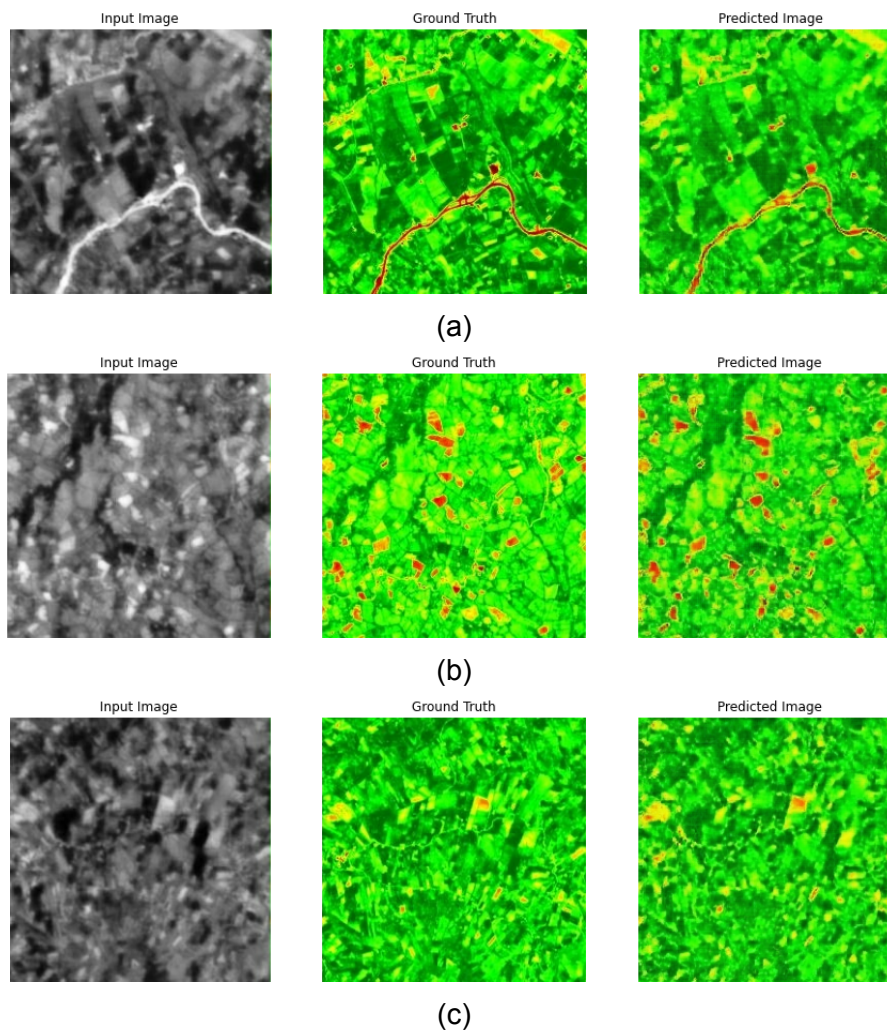


Figura 8.4: Gráficas de ejemplo para Unet. Como entrada la imagen SAR, en la parte central la referencia (Ground Truth) y en la izquierda la imagen estimada.

### 8.3. Resnet - 9

Esta arquitectura utiliza un codificador y decodificador igual que la Unet, pero con la gran diferencia que no tiene saltos y en la parte que comunica las dos estructuras (cuello de botella), existen 9 estructuras de realimentación, expuesta en un anterior capítulo de este trabajo. En términos generales presentó unos mejores resultados que la Unet en cuanto a todas las medidas de evaluación, pero cuando se observa la gráfica de pérdidas, tanto el discriminador como el generador poseen los mismos inconvenientes que la Unet. En la Figura 8.5 se observa un comportamiento muy similar a la pérdida de la Unet, indicando que la estabilidad en el entrenamiento es aproximadamente igual.

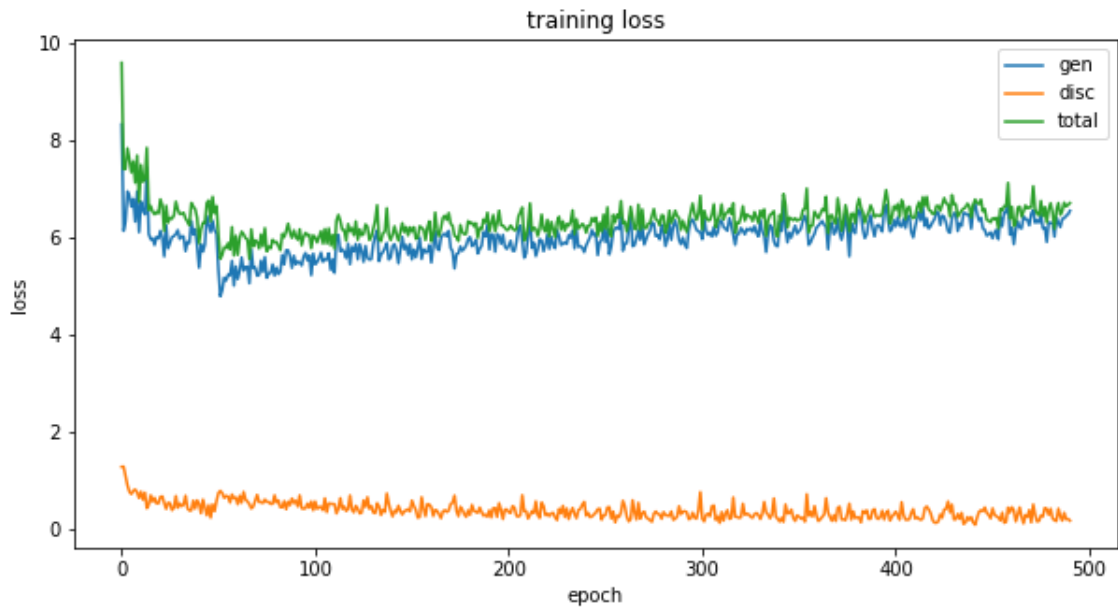
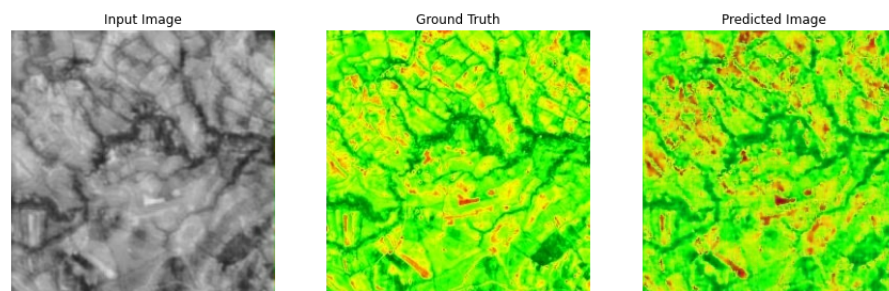
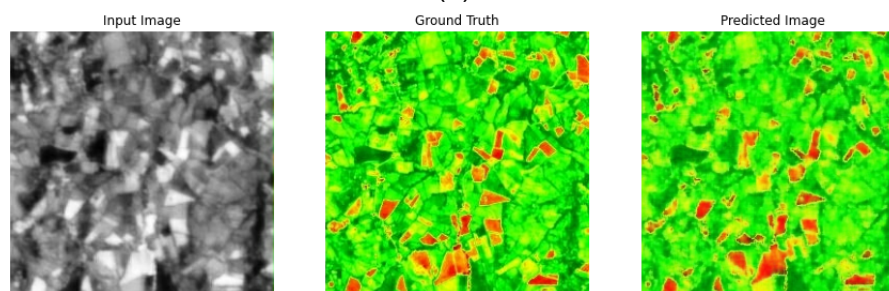


Figura 8.5: Gráfica de las pérdidas del discriminador (Azul), generador(Naranja) y el total(Verde) en ResNet.

Por otra parte en la Figura 8.7 y Tabla 8.3, se presentan algunas imágenes en su última época de entrenamiento con sus respectivos valores en las medidas de evaluación.



(a)



(b)

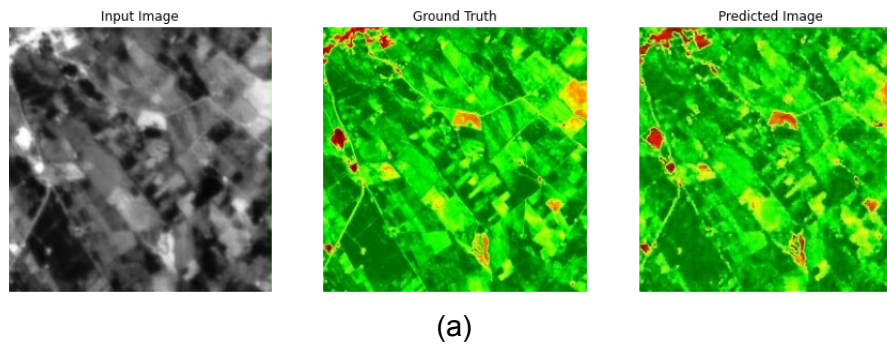


Figura 8.7: Gráficas de ejemplo para ResNet-9. Como entrada la imagen SAR, en la parte central la referencia (Ground Truth) y en la izquierda la imagen estimada.

Imagen	SSIM	RMSE	SAM [Grados]	PSNR [db]
a	0.657	0.023	63.646	67.538
b	0.627	0.014	61.547	70.059
c	0.655	0.022	65.174	67.701

Tabla 8.3: Valores de las medidas de evaluación para tres imágenes de ejemplo para Resnet

## 8.4. Resultados generales

Como comentarios generales cabe aclarar que el análisis visual en esta clase de imágenes es complejo. Esto se debe a que no existen contornos muy específicos en este dataset, por lo tanto, es difícil detectar donde termina una parcela de la otra. En algunas gráficas el ojo humano es capaz de detectar las diferencias entre coloración, pero este criterio de calificación no es el más confiable porque no existen fronteras bien definidas. En consecuencia, este trabajo solo se centra en la parte cuantitativa (valores entregados por las métricas de evaluación), en el anexo se pueden ver algunas de las imágenes del grupo de validación y la respectiva tabla con las métricas de evaluación.

Por otra parte, la principal diferencia en las medidas de evaluación fue la SSIM ya que en la arquitectura Unet++ se obtuvo  $0.643 \pm 0.045$ , 0.4 unidades mejor que la ResNet-9 y 0.9 mejor que la Unet convencional. La comparación se expone en el siguiente gráfico de caja:

Otra mejoría importante es en el RMSE, ya que se mejoró a 0.015, muy importante si se tiene en cuenta que se utilizan el mismo grupo de validación y con este valor nuestras imágenes resultantes serán más parecidas a las de referencia. Con respecto a las otras medidas no hay un cambio relevante para ser tomada en cuenta en la elección de una estructura. Por ejemplo, si se tomaran como valores de decisión para SAM o PSNR, prácticamente sería indiferente tomar uno u otro modelo.



Figura 8.8: Comparación del SSIM en gráfico de cajas para las tres arquitecturas de generadores. Serie 1= Unet++, Serie = Unet, Serie 3=ResNet 9

Imagen	SSIM	RMSE	SAM [Grados]	PSNR [db]
Unet++	$0.643 \pm 0.045$	$0.015 \pm 0.005$	$64.684 \pm 8.116$	$69.978 \pm 1.761$
Unet	$0.562 \pm 0.060$	$0.028 \pm 0.019$	$64.765 \pm 7.092$	$67.663 \pm 2.067$
ResNet	$0.606 \pm 0.039$	$0.019 \pm 0.008$	$63.324 \pm 7.260$	$68.755 \pm 1.604$

Tabla 8.4: Tabla resumen de las medias y su desviación estándar para el grupo de validación

## Capítulo 9

# Conclusiones

---

- Las imágenes satelitales y su evolución para la adquisición de información de la superficie terrestre han permitido su utilización en muchos sectores de la actividad humana. La etapa de preprocesamiento de imágenes satelitales contribuye a entregar la información requerida por el usuario con la menor cantidad de impurezas. Los procesos que permiten realizar el mejoramiento son muy restringidos a las recomendaciones y herramientas brindadas por los propietarios de los satélites. Por consiguiente, es necesario hacer una investigación de la clase de datos que se necesitan en cualquier proyecto y si, con las etapas de preprocesamiento se puede alcanzar a obtener dichos requerimientos de información. Por ejemplo, en este trabajo después de la etapa de preprocesamiento las características entregadas por las imágenes Sentinel son adecuadas para la planeación agrícola, ya que se puede alcanzar una resolución de  $10\text{m} \times 10\text{m}$  en las bandas 4 y 8 presentes en la fórmula NDVI fundamental para conocer la vigorosidad de las plantas.
- Los resultados sugieren que las redes adversas condicionales son un enfoque prometedor para muchas tareas de traslación de imagen a imagen, especialmente aquellas que involucran salidas gráficas. Estas redes aprenden una pérdida adaptada a la tarea y los datos en cuestión, lo que los hace aplicables en una amplia variedad de entornos. En nuestro estudio, primero generamos resultados para el modelo utilizando la arquitectura U-Net++ para el generador. Luego usamos la arquitectura U-Net y por último, la ResNet para el generador. Se observó que el uso del modelo de generador basado en U-Net++ mejoró el rendimiento y redujo el error cuadrático medio del modelo de generador basado en U-Net de 0,028 a 0,015, que es una mejora significativa. Por lo tanto, las redes U-Net++ superan las ResNet y el modelo U-Net.
- El desempeño insatisfactorio en algunas métricas de escenas complejas, se debe principalmente a la distribución de datos del conjunto de entrenamiento utilizado, ya que no se encontraron los suficientes ejemplos con diferentes geometrías en las fronteras de las parcelas. En Generative adversarial networks (GANs) la diversidad en el grupo de entrenamiento mejora los resultados, el problema de la zona estudiada, es que los bordes de las parcelas no eran muy nítidas. En consecuencia, se recomienda realizar un base de datos en zonas la actividad agrícola este más organizada y en donde las zonas agrícolas estén bien limitadas.

- En este proyecto se propuso una metodología capaz de producir imágenes NDVI sintéticas a partir de imágenes SAR. Las imágenes generadas pueden ser empleadas en investigaciones encaminadas a identificar clases de cultivos por medio de algoritmos de segmentación y clasificación. Con la información obtenida, se pueden utilizar métricas como la exactitud o precisión para comparar los resultados entre las imágenes NDVI originales y sintéticas, con el fin de verificar la calidad de la información entregada por las imágenes generadas en la metodología propuesta. Por otra parte, uno de los aspectos más críticos en entrenamientos para GAN es la base de datos, por lo tanto, es necesario que el algoritmo basado en GAN pueda ser entrenado con imágenes de otras regiones, para observar el comportamiento de las métricas ante el cambio de la base de datos.

# Bibliografía

- [1] J. Hatfield, A. A. Gitelson, J. S. Schepers y C. Walthall, “Application of spectral remote sensing for agronomic decisions,” *Agronomy Journal*, vol. 100, n.º Supplement\_3, S-117, 2008 (vid. pág. 12).
- [2] P. Muñoz Aguayo, “Apuntes de Teledetección: Índices de vegetación,” *Centro de información de recursos Naturales*, pág. 15, 2013 (vid. pág. 12).
- [3] C. Lanaras, J. Bioucas-Dias, E. Baltsavias y K. Schindler, “Super-Resolution of Multispectral Multiresolution Images from a Single Sensor,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2017-July, n.º i, págs. 1505-1513, 2017, issn: 21607516 (vid. pág. 12).
- [4] N.-B. Chang y K. Bai, *Multisensor data fusion and machine learning for environmental remote sensing*. CRC Press, 2018 (vid. págs. 12, 26-29, 36, 38, 39).
- [5] E. C. Mantovani y C. Magdalena, “Manual de agricultura de precisión.,” *Embrapa Milho e Sorgo-Livro científico (ALICE)*, 2014 (vid. pág. 13).
- [6] P. Muñoz Aguayo, “Apuntes de teledetección: índices de vegetación,” 2013 (vid. pág. 13).
- [7] G. Poveda, “La hidroclimatología de Colombia: una síntesis desde la escala inter-decadal hasta la escala diurna,” *Rev. Acad. Colomb. Cienc*, vol. 28, n.º 107, págs. 201-222, 2004 (vid. pág. 13).
- [8] H. Bolívar, “Metodologías e indicadores de evaluación de sistemas agrícolas hacia el desarrollo sostenible,” *CICAG: Revista del Centro de Investigación de Ciencias Administrativas y Gerenciales*, vol. 8, n.º 1, págs. 1-18, 2011 (vid. pág. 17).
- [9] F. FAO, P. OMS, UNICEF y col., “El estado de la seguridad alimentaria y la nutrición en el mundo 2017,” *Fomentando la resiliencia en aras de la paz y la seguridad alimentaria*. Roma: FAO, 2017 (vid. pág. 17).
- [10] E. Martínez Barrada, S. Morales Chica y col., “Análisis del sector agroindustrial en Colombia,” 2016 (vid. pág. 17).

- [11] S. L. C. Díaz, C. Z. Prada, J. B. López, G. D. R. OTALORA, J. E. R. Cala, R. C. O. Vengoechea, L. M. Ibatá y col., “Dividendos ambientales de la paz Retos y oportunidades para construir una paz sostenible,” Departamento Nacional de Planeación, inf. téc., 2016 (vid. pág. 17).
- [12] G. de Nariño, “Plan Participativo de Desarrollo Departamental–Nariño Corazón del Mundo. 2016-2019,” *Plan de Desarrollo Departamental de Nariño [Internet]*, pág. 255, 2016 (vid. pág. 17).
- [13] N. Clerici, C. A. Valbuena Calderón y J. M. Posada, “Fusion of sentinel-1a and sentinel-2A data for land cover mapping: A case study in the lower Magdalena region, Colombia,” *Journal of Maps*, vol. 13, n.º 2, págs. 718-726, 2017, issn: 17445647 (vid. págs. 18, 21).
- [14] N. N. Baghdadi, M. El Hajj, M. Zribi e I. Fayad, “Coupling SAR C-band and optical data for soil moisture and leaf area index retrieval over irrigated grasslands,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, n.º 3, págs. 1229-1243, 2015 (vid. págs. 18, 21).
- [15] G. Scarpa, M. Gargiulo, A. Mazza y R. Gaetano, “A CNN-based fusion method for feature extraction from sentinel data,” *Remote Sensing*, vol. 10, n.º 2, págs. 1-20, 2018, issn: 20724292 (vid. págs. 18, 21).
- [16] W. He y N. Yokoya, “Multi-temporal sentinel-1 and -2 data fusion for optical Image Simulation,” *ISPRS International Journal of Geo-Information*, vol. 7, n.º 10, págs. 1-11, 2018, issn: 22209964. arXiv: 1807.09954 (vid. págs. 18, 21).
- [17] D. Dimov, F. Low, M. Ibrakhimov, G. Stulina y C. Conrad, “SAR and optical time series for crop classification,” *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2017-July, n.º August, págs. 811-814, 2017 (vid. págs. 18, 22).
- [18] D Dunchev, *Evaluation of innovation technologies in precision agriculture*, 2020 (vid. pág. 19).
- [19] F. Bovolo y L. Bruzzone, “The time variable in data fusion: A change detection perspective,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, n.º 3, págs. 8-26, 2015 (vid. pág. 20).

- [20] P. Ghamisi, B. Rasti, N. Yokoya, Q. Wang, B. Hofle, L. Bruzzone, F. Bovolo, M. Chi, K. Anders, R. Gloaguen y col., "Multisource and Multitemporal Data Fusion in Remote Sensing," *arXiv preprint arXiv:1812.08287*, 2018 (vid. pág. 20).
- [21] P. H. Swain, "Bayesian classification in a time-varying environment," 1978 (vid. pág. 20).
- [22] L. Bruzzone, D. F. Prieto y S. B. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Transactions on Geoscience and remote Sensing*, vol. 37, n.º 3, págs. 1350-1359, 1999 (vid. pág. 20).
- [23] Z. Pan, J. Huang, Q. Zhou, L. Wang, Y. Cheng, H. Zhang, G. A. Blackburn, J. Yan y J. Liu, "Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 34, n.º 1, págs. 188-197, 2015, issn: 1872826X (vid. pág. 20).
- [24] J. Haas e Y. Ban, "Sentinel-1A SAR and sentinel-2A MSI data fusion for urban ecosystem service mapping," *Remote Sensing Applications: Society and Environment*, vol. 8, n.º July, págs. 41-53, 2017, issn: 23529385 (vid. pág. 21).
- [25] M. Hirschmugl, C. Sobe, J. Deutscher y M. Schardt, "Combined use of optical and synthetic aperture radar data for REDD+ applications in Malawi," *Land*, vol. 7, n.º 4, págs. 1-17, 2018, issn: 2073445X (vid. pág. 22).
- [26] K. Irwin, D. Beaulne, A. Braun y G. Fotopoulos, "Fusion of SAR, optical imagery and airborne LiDAR for surface water detection," *Remote Sensing*, vol. 9, n.º 9, págs. 1-18, 2017, issn: 20724292 (vid. pág. 22).
- [27] Y. Forget, M. Shimoni y M. Gilbert, "Complementarity Between Sentinel-1 and Landsat 8 Imagery for Built-Up Mapping in Sub-Saharan Africa," n.º October, págs. 1-13, 2018 (vid. pág. 22).
- [28] K. Van Tricht, A. Gobin, S. Gilliams e I. Piccard, "Synergistic use of radar Sentinel-1 and optical Sentinel-2 imagery for crop mapping: A case study for Belgium," *Remote Sensing*, vol. 10, n.º 10, pág. 1642, 2018 (vid. pág. 22).
- [29] J. D. T. De Alban, G. M. Connette, P. Oswald y E. L. Webb, "Combined Landsat and L-band SAR data improves land cover classification and change detection in dynamic tropical landscapes," *Remote Sensing*, vol. 10, n.º 2, 2018, issn: 20724292 (vid. pág. 23).

- [30] C. De Bernardis, F. Vicente-Guijalba, T. Martinez-Marin y J. M. Lopez-Sanchez, "Contribution to Real-Time Estimation of Crop Phenological States in a Dynamical Framework Based on NDVI Time Series: Data Fusion with SAR and Temperature," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, n.º 8, págs. 3512-3523, 2016, issn: 21511535 (vid. [pág. 23](#)).
- [31] J. Reiche, "Combining SAR and optical satellite image time series for tropical forest monitoring," 2015 (vid. [pág. 23](#)).
- [32] S. D. Kumar, S. S. Rao y J. Sharma, "Radar Vegetation Index as an alternative to NDVI for monitoring of soyabean and cotton," en *Proceedings of the XXXIII INCA International Congress (Indian Cartographer)*, Jodhpur, India, 2013, págs. 19-21 (vid. [pág. 23](#)).
- [33] X. Wang, X. Li y L. Ge, "Relating envisat ASAR and ALOS PALSAR backscattering coefficient to spot NDVI for monitoring seasonal change of pasture biomass in Western Australia," *International Geoscience and Remote Sensing Symposium (IGARSS)*, págs. 3744-3747, 2012 (vid. [pág. 24](#)).
- [34] B Mishra y J Susaki, "Optical and SAR data integration for automatic change pattern detection," 2014 (vid. [pág. 24](#)).
- [35] L. Pipia, J. Muñoz-Marí, E. Amin, S. Belda, G. Camps-Valls y J. Verrelst, "Fusing optical and SAR time series for LAI gap filling with multioutput Gaussian processes," *Remote Sensing of Environment*, vol. 235, [pág. 111452](#), 2019 (vid. [pág. 24](#)).
- [36] F. Gao, T. Hilker, X. Zhu, M. Anderson, J. Masek, P. Wang e Y. Yang, "Fusing Landsat and MODIS data for vegetation monitoring," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, n.º 3, págs. 47-60, 2015 (vid. [págs. 25, 32, 36](#)).
- [37] Q. Wang y P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sensing of Environment*, vol. 204, págs. 31-42, 2018 (vid. [pág. 25](#)).
- [38] J. Adrian, V. Sagan y M. Maimaitijiang, "Sentinel SAR-optical fusion for crop type mapping using deep learning and Google Earth Engine," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 175, págs. 215-235, 2021 (vid. [pág. 25](#)).
- [39] Z. Wang, Q. She y T. E. Ward, "Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy," 2019. arXiv: [1906.01529](#) (vid. [pág. 26](#)).

- [40] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort y col., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote sensing of Environment*, vol. 120, págs. 25-36, 2012 (vid. pág. 26).
- [41] R. Torres, P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown y col., "GMES Sentinel-1 mission," *Remote Sensing of Environment*, vol. 120, págs. 9-24, 2012 (vid. pág. 26).
- [42] D. G. Hadjimitsis, G Papadavid, A Agapiou, K Themistocleous, M. Hadjimitsis, A Retalis, S Michaelides, N Chrysoulakis, L Toullos y C. Clayton, "Atmospheric correction for satellite remotely sensed data intended for agricultural applications: impact on vegetation indices," *Natural Hazards and Earth System Sciences*, vol. 10, n.º 1, págs. 89-95, 2010 (vid. págs. 26, 27).
- [43] Y. J. Kaufman y D. Tanre, "Strategy for direct and indirect methods for correcting the aerosol effect on remote sensing: from AVHRR to EOS-MODIS," *Remote sensing of Environment*, vol. 55, n.º 1, págs. 65-79, 1996 (vid. pág. 27).
- [44] S. S. Baboo y M. R. Devi, "An analysis of different resampling methods in Coimbatore, District," *Global Journal of Computer Science and Technology*, 2010 (vid. pág. 27).
- [45] F. Filipponi, "Sentinel-1 GRD preprocessing workflow," en *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 18, 2019, pág. 11 (vid. pág. 28).
- [46] K. Usui, "Data augmentation using image-to-image translation for detecting forest strip roads based on deep learning," *International Journal of Forest Engineering*, vol. 32, n.º 1, págs. 57-66, 2021 (vid. págs. 29-34).
- [47] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville e Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014 (vid. págs. 29, 30, 32, 45, 50, 52).
- [48] R. Atienza, *Advanced deep learning with Keras : apply deep learning techniques, autoencoders, GANs, variational autoencoders, deep reinforcement learning, policy gradients, and more*. Packt Publishing, 2018 (vid. págs. 31, 32, 34-36).

- [49] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh y J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," en *Deep learning in medical image analysis and multimodal learning for clinical decision support*, Springer, 2018, págs. 3-11 (vid. págs. 36, 37).
- [50] R. A. Olivera Solís e Y. López Pérez, "Codificación de video en HEVC/H. 265 utilizando FFMPEG," *Ingeniería Electrónica, Automática y Comunicaciones*, vol. 40, n.º 2, págs. 22-33, 2019 (vid. pág. 39).
- [51] K. Cwik, "Flood mapping with the sentinel-1 time-series data in arid areas," 2017 (vid. pág. 41).
- [52] G. M. Gandhi, S Parthiban, N. Thummalu y A Christy, "Ndvi: Vegetation change detection using remote sensing and gis—A case study of Vellore District," *Procedia computer science*, vol. 57, págs. 1199-1210, 2015 (vid. pág. 42).
- [53] A Castillo y A Jurado, "Caracterización social y económica del departamento de Nariño," *Análisis de información secundaria Available: <https://www.google.com/url>*, 2014 (vid. págs. 42, 43).

## Capítulo 10

# Glosario y términos especiales

---

- GANs - Generative adversarial network
- NDVI - Normalized difference vegetation index
- SAR - Synthetic-aperture radar
- MSI - Multispectral Instrument
- SSIM - Structural similarity index
- MSE - Mean-Squared Error
- PSNR - Peak Signal-to-Noise Ratio
- SAM - Spectral Angle Mapper

# Capítulo 11

## Anexo

---

### 11.1. Gráficas de pérdidas de las arquitecturas conjuntas

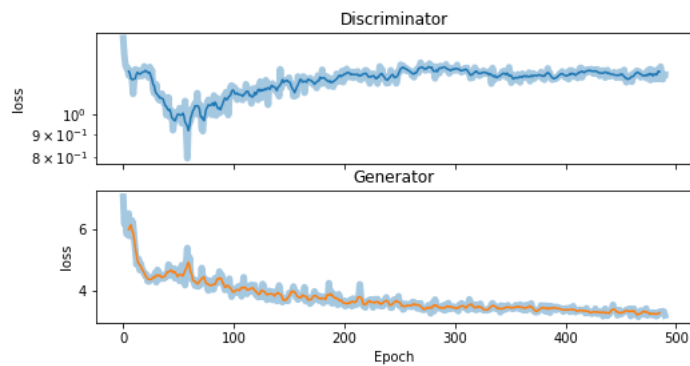


Figura 11.1: Gráficas de pérdida del discriminador y generador para Unet++

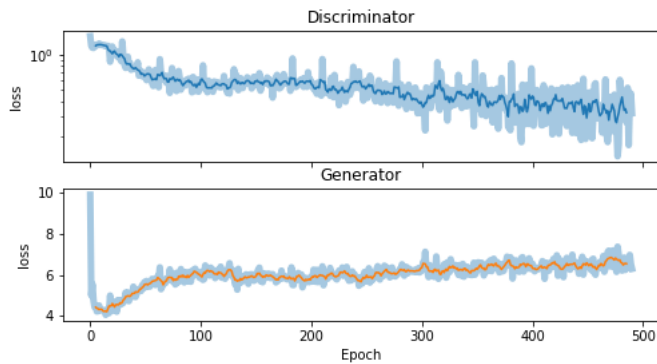


Figura 11.2: Gráficas de pérdida del discriminador y generador para Unet.

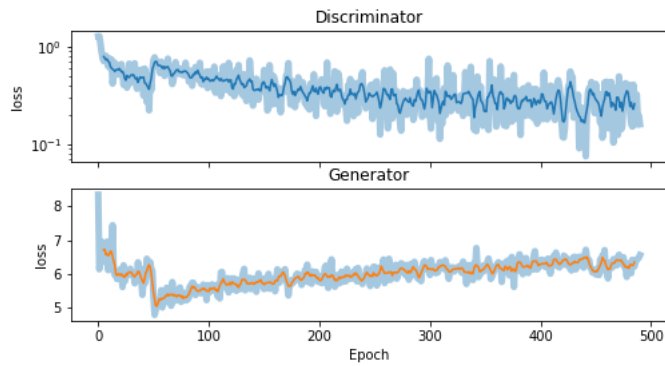


Figura 11.3: Gráficas de pérdida del discriminador y generador para ResNet-9.

## 11.2. Diagrama de cajas de las medidas de evaluación para las arquitecturas de generadores

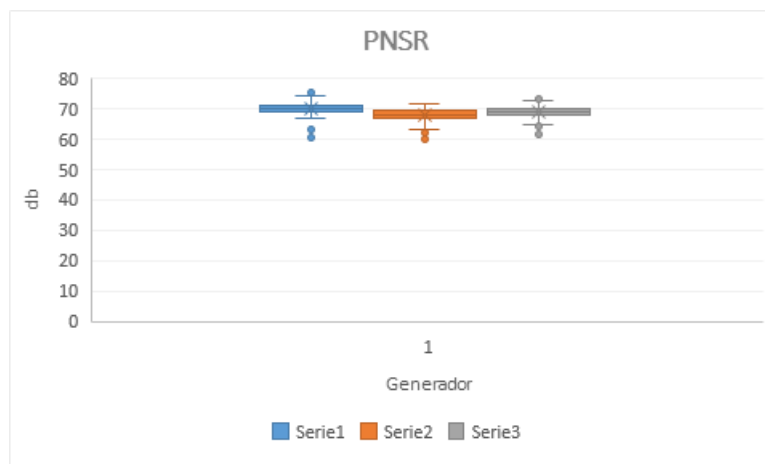


Figura 11.4: Comparación del PNSR en gráfico de cajas para las tres arquitecturas de generadores. Serie 1= Unet++, Serie = Unet, Serie 3 = ResNet 9.

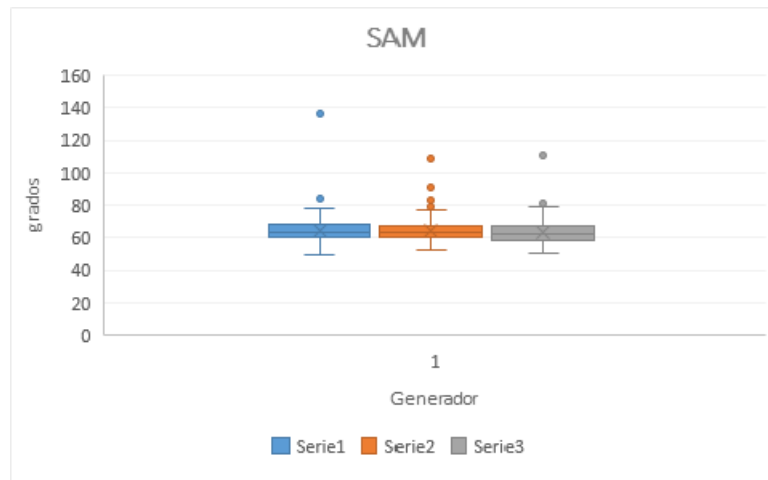


Figura 11.5: Comparación del SAM en gráfico de cajas para las tres arquitecturas de generadores. Serie 1= Unet++, Serie 2 = Unet, Serie 3 = ResNet 9

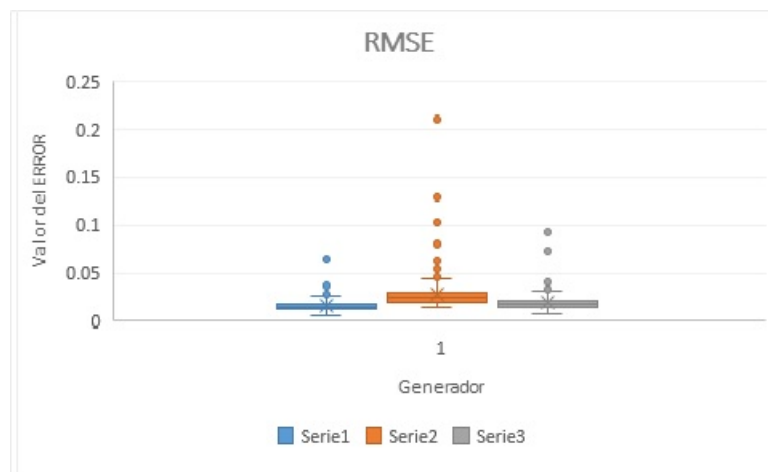
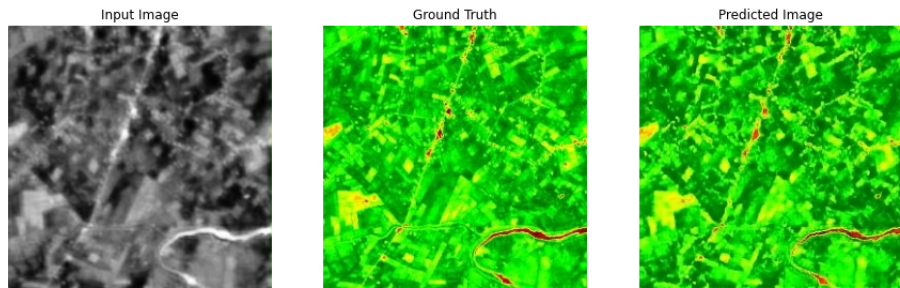


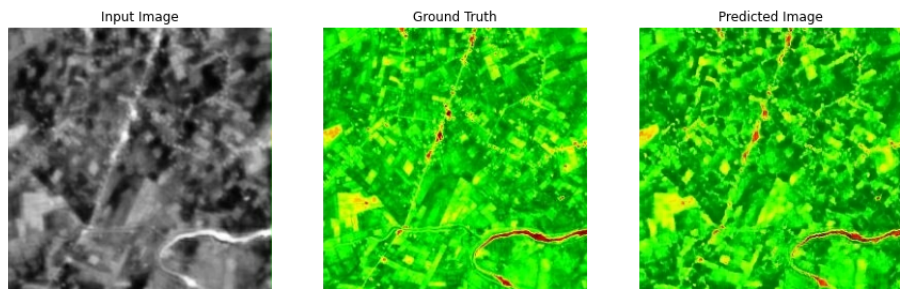
Figura 11.6: Comparación del RMSE en gráfico de cajas para las tres arquitecturas de generadores. Serie 1= Unet++, Serie 2 = Unet, Serie 3 = ResNet 9

### 11.3. Imágenes del grupo de validación con la tabla de sus valores en las medidas de evaluación

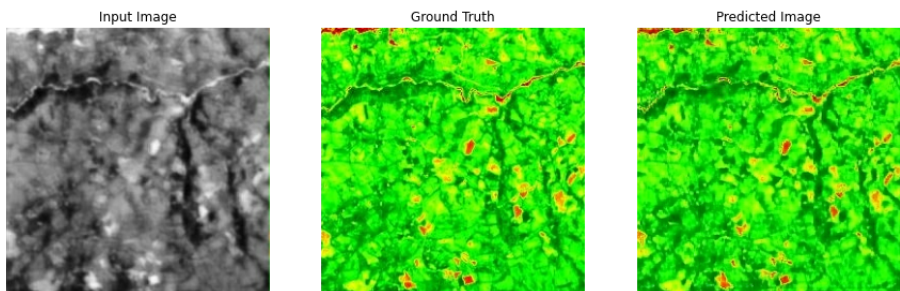
#### 11.3.1. Unet++



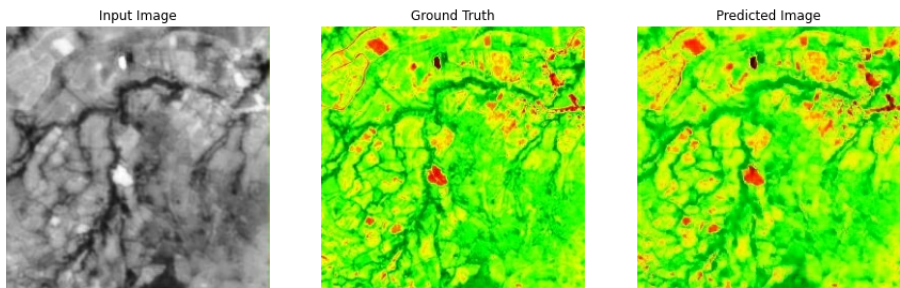
(a)



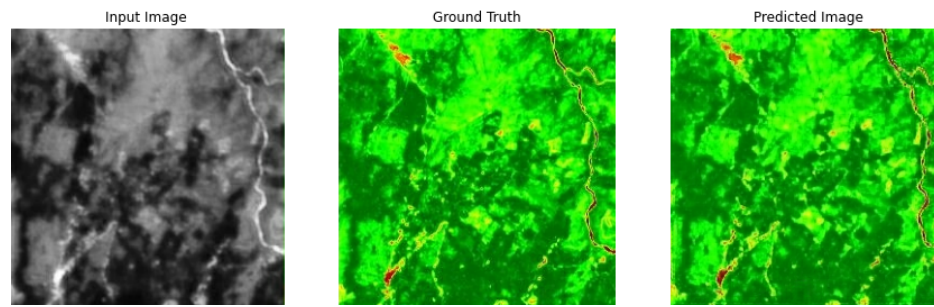
(b)



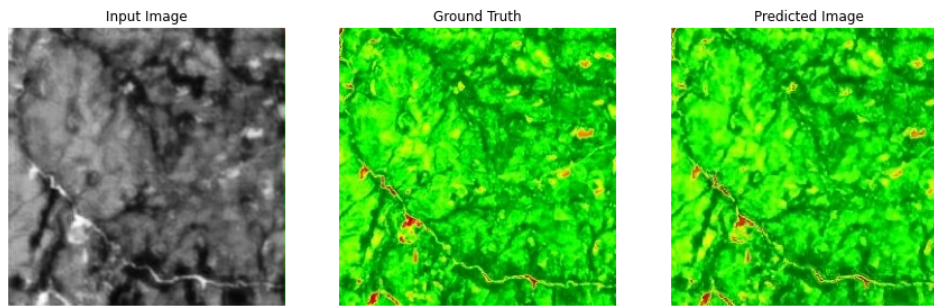
(c)



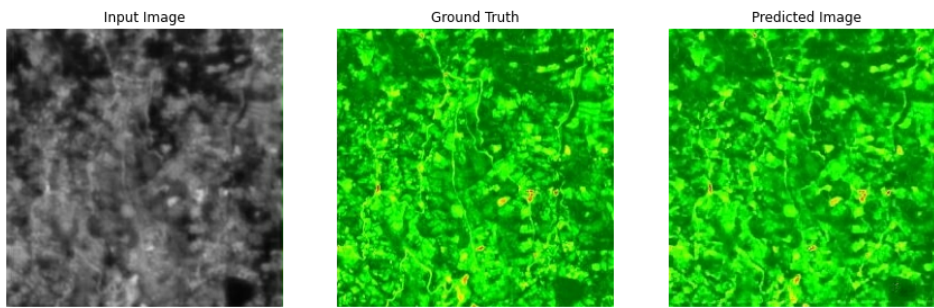
(d)



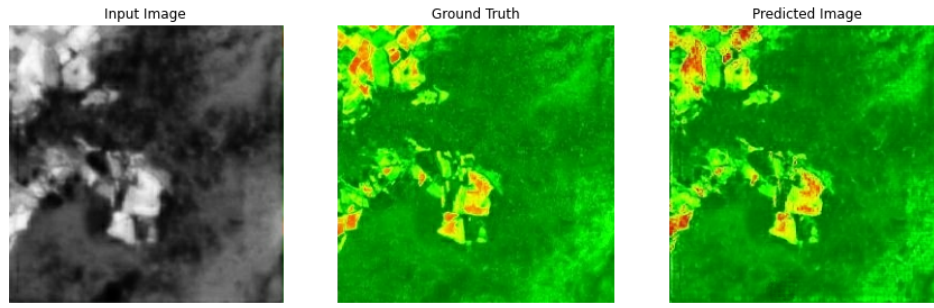
(e)



(f)



(g)



(h)

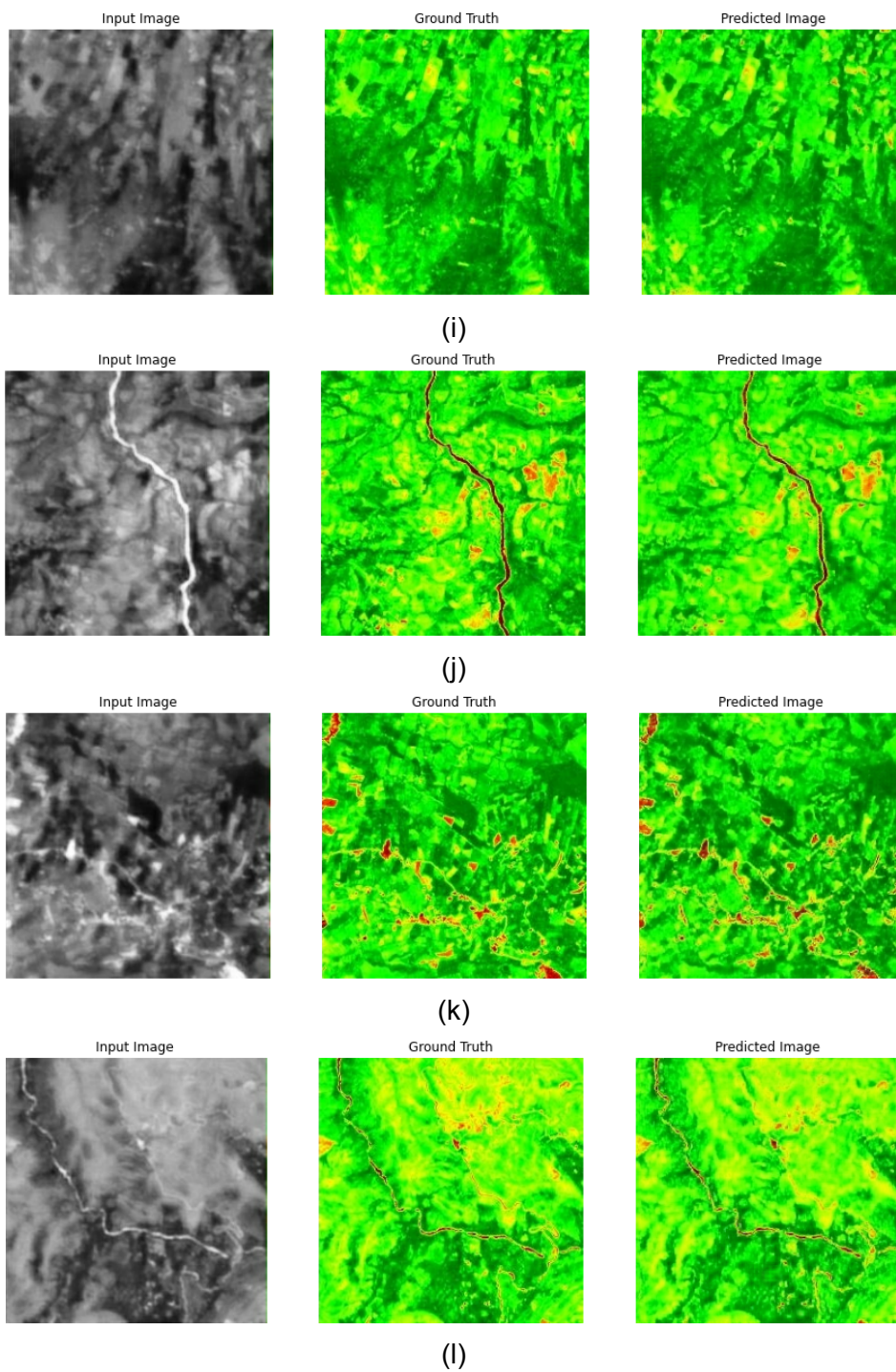
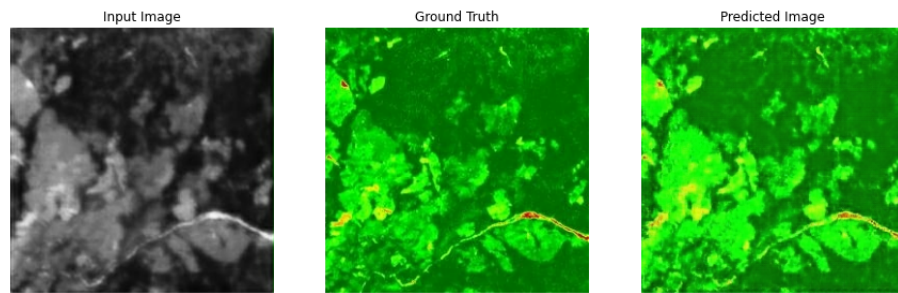


Figura 11.7: Gráficas de ejemplo para Unet++. Como entrada la imagen SAR, en la parte central la referencia (Ground Truth) y en la izquierda la imagen estimada.

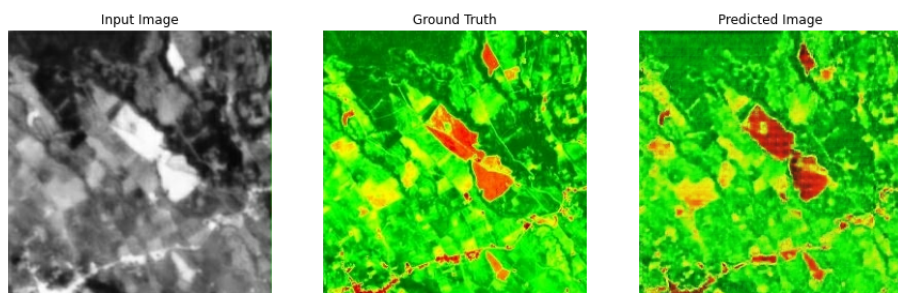
## 11.3.2. Unet



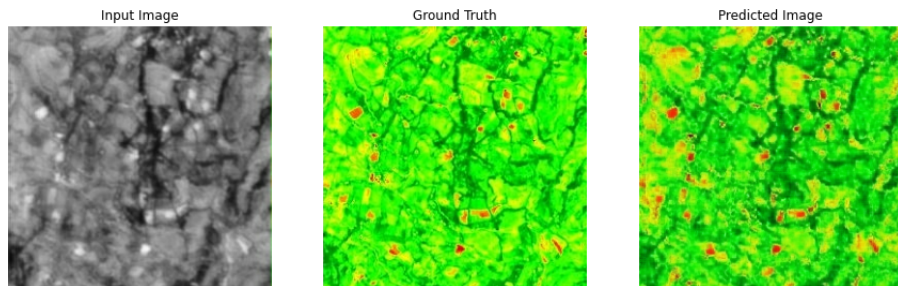
(a)



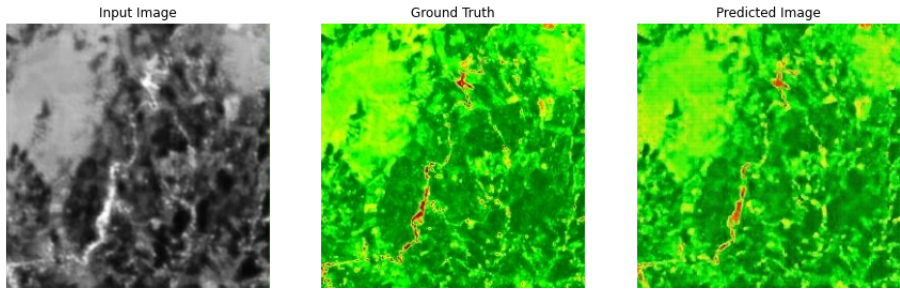
(b)



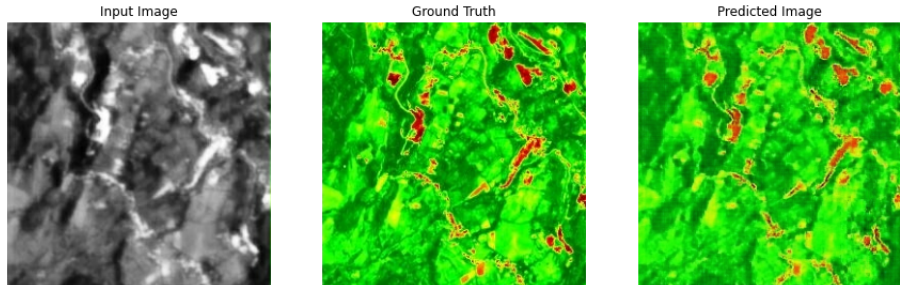
(c)



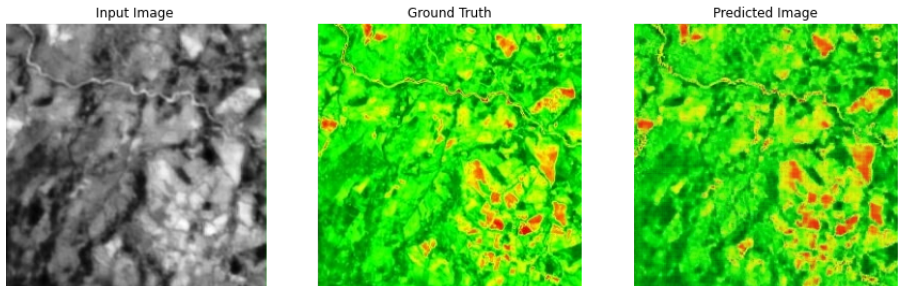
(d)



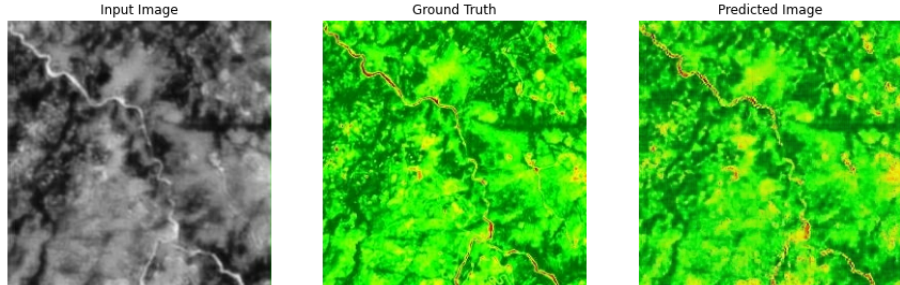
(e)



(f)



(g)



(h)

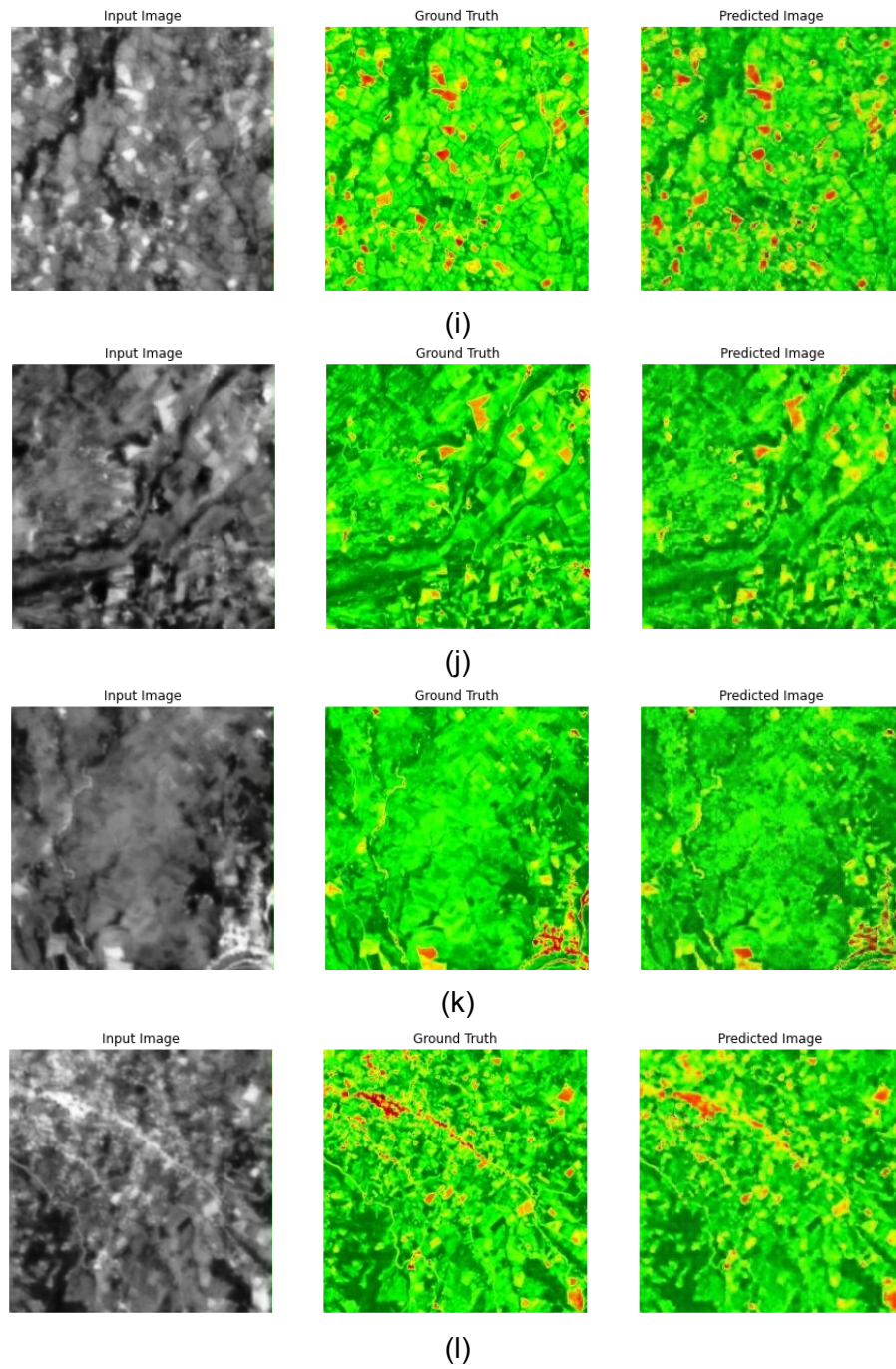
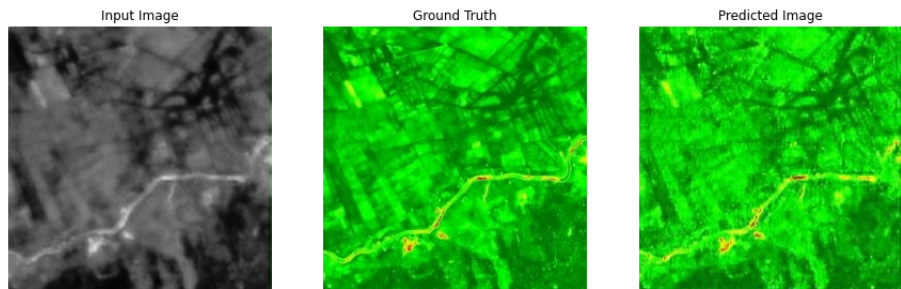
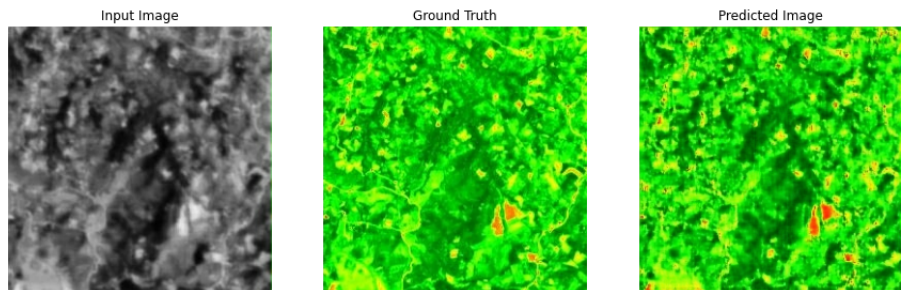


Figura 11.8: Gráficas de ejemplo para Unet. Como entrada la imagen SAR, en la parte central la referencia (Ground Truth) y en la izquierda la imagen estimada.

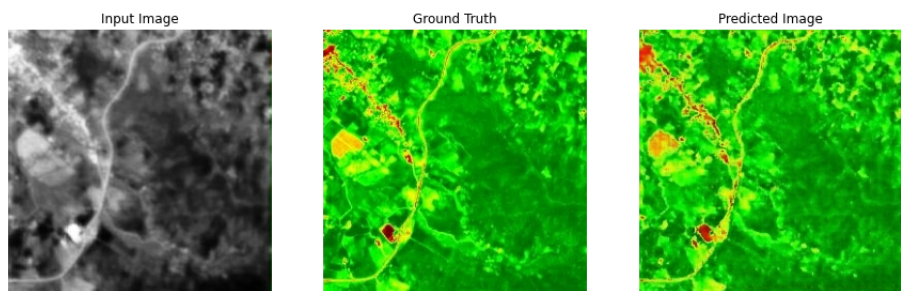
### 11.3.3. ResNet 9



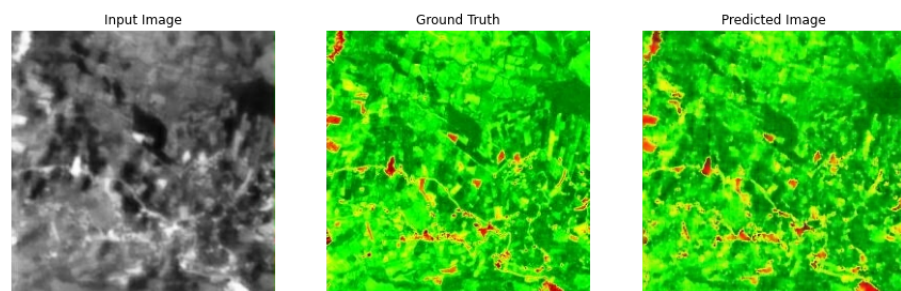
(a)



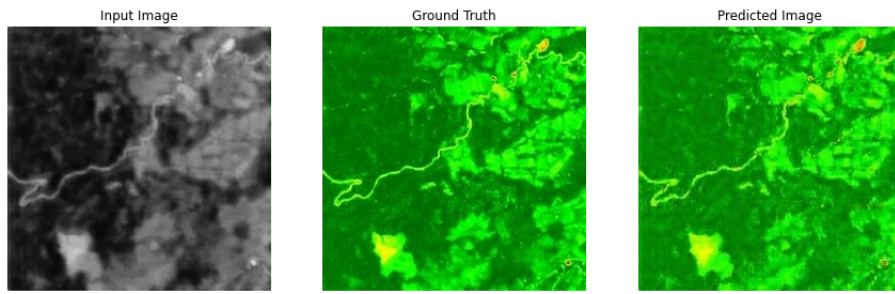
(b)



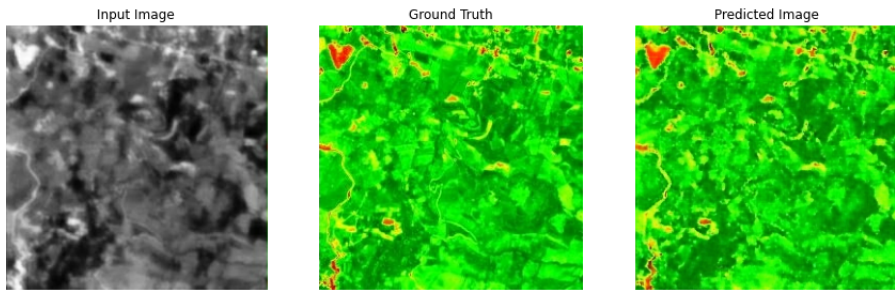
(c)



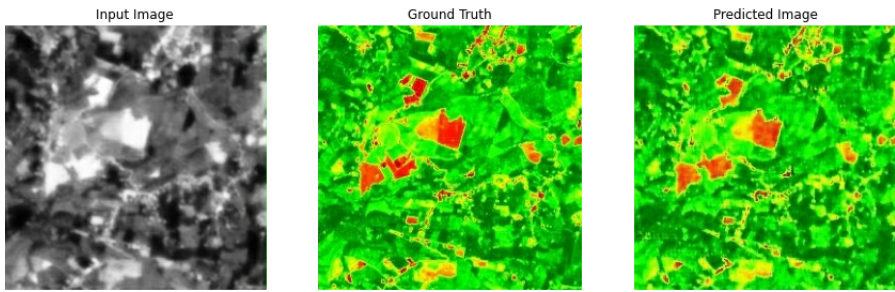
(d)



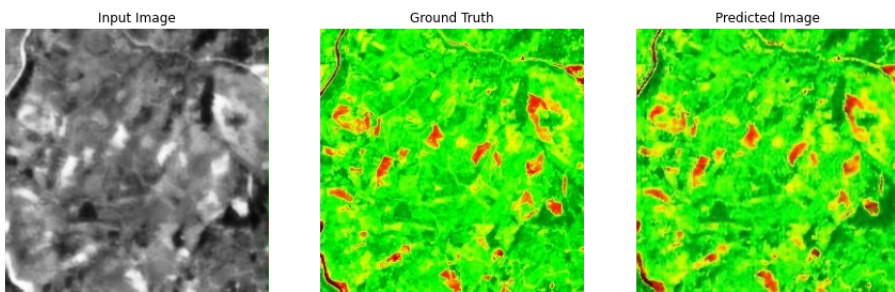
(e)



(f)



(g)



(h)

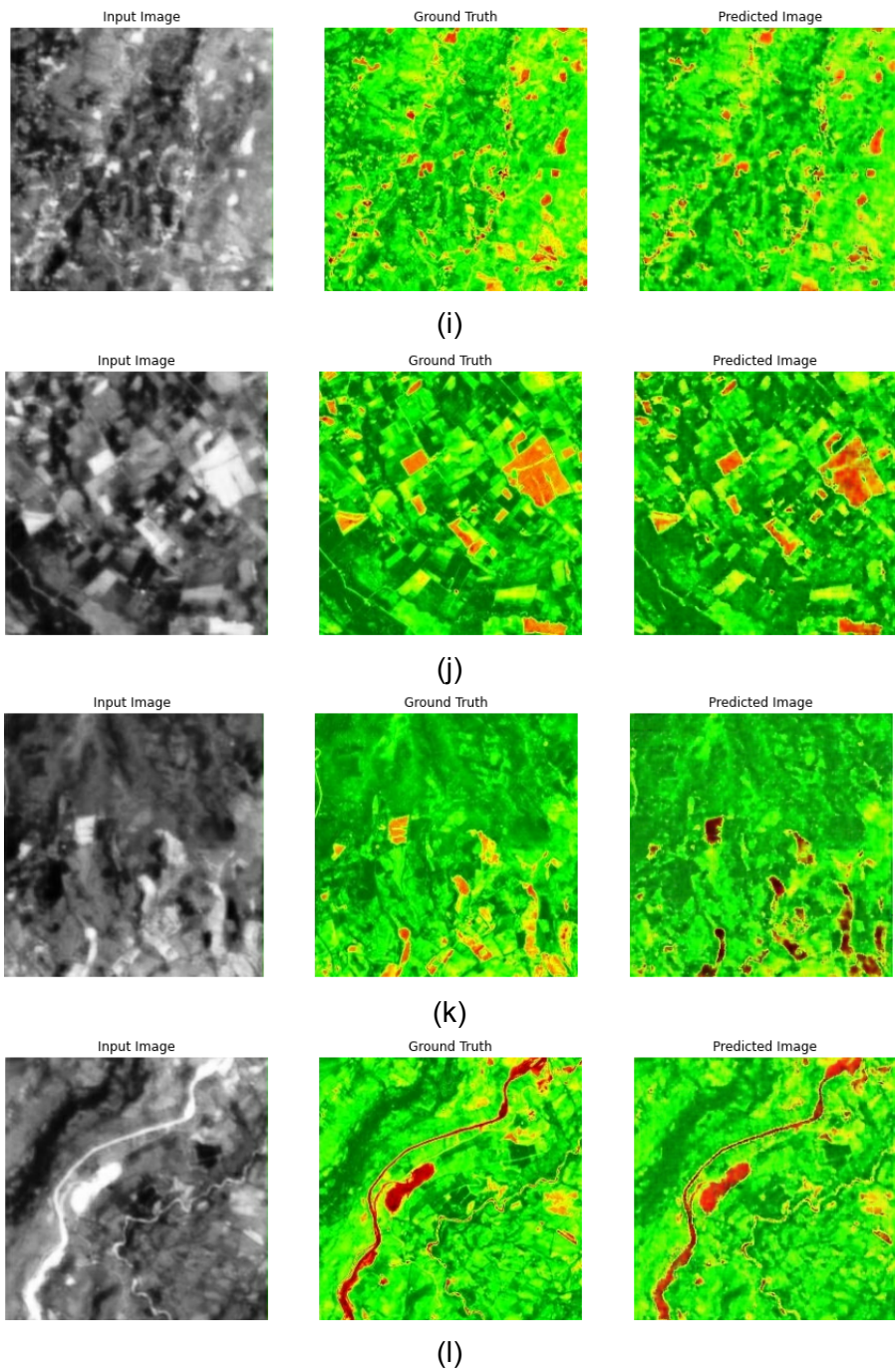


Figura 11.9: Gráficas de ejemplo para ResNet-9. Como entrada la imagen SAR, en la parte central la referencia (Ground Truth) y en la izquierda la imagen estimada.

### 11.3.4. Tabla con medidas de evaluación anteriores imágenes

Gráfica	SSIM	RMSE	SAM	PSNR
a	0.664	0.013	58.794	68.214
b	0.645	0.016	58.058	65.224
c	0.633	0.018	65.369	70.785
d	0.653	0.024	69.962	71.585
e	0.605	0.013	71.110	69.406
f	0.660	0.017	57.345	71.136
g	0.634	0.020	64.350	60.551
h	0.663	0.018	65.690	71.994
i	0.664	0.008	61.039	70.432
j	0.662	0.015	63.952	71.491
k	0.606	0.018	68.529	68.183
l	0.626	0.012	72.963	70.170

Tabla 11.1: Medidas de evaluación de 12 imágenes del grupo de validación para la arquitectura Unet++

Gráfica	SSIM	RMSE	SAM	PSNR
a	0.517	0.025	62.316	67.654
b	0.600	0.037	84.599	65.344
c	0.550	0.021	65.399	64.075
d	0.601	0.019	57.619	70.115
e	0.482	0.030	59.702	60.213
f	0.544	0.023	65.310	69.089
g	0.598	0.103	58.541	68.038
h	0.593	0.025	71.589	65.376
i	0.572	0.055	64.246	67.384
j	0.571	0.016	56.780	67.020
k	0.580	0.020	57.864	66.322
l	0.507	0.027	71.927	69.733

Tabla 11.2: Medidas de evaluación de 12 imágenes del grupo de validación para la arquitectura Unet

Gráfica	SSIM	RMSE	SAM	PSNR
a	0.613	0.013	57.557	66.932
b	0.596	0.018	72.264	69.374
c	0.574	0.018	56.999	69.698
d	0.607	0.021	60.162	65.465
e	0.628	0.021	61.043	70.412
f	0.623	0.014	78.187	70.097
g	0.616	0.020	55.897	68.777
h	0.614	0.016	53.809	68.125
i	0.602	0.019	60.173	68.951
j	0.640	0.026	54.938	69.343
k	0.535	0.014	64.662	68.071
l	0.500	0.017	56.916	67.530

Tabla 11.3: Medidas de evaluación de 12 imágenes del grupo de validación para la arquitectura ResNet-9