

Uso de Redes Neuronales Convolucionales para Interpretación de Imágenes Satelitales

Primera Presentación - Proyecto Final de Carrera

Alumno: Giovanni Dueck

Tutor: Alberto Ramírez - **Cotutor:** Félix Carvallo
Ing. Informática

Fecha [1]

1. Introducción

Imágenes satelitales o teledetección se refiere a imágenes capturadas por un sensor montado en un satélite artificial, para extraer información. Estas imágenes contienen información multispectral, es decir, además de la luz visible se toman imágenes de bandas invisibles como por ejemplo la luz infrarroja. [2]

Para la captura de estas imágenes se emplean varios métodos, que se dividen en dos categorías: sensores pasivos recolectan radiación electromagnética reflejada del sol, mientras que sensores activos emiten su propia radiación y captan la reflexión de la tierra. Sensores pasivos requieren de una cantidad importante de energía para operar, pero tienen la ventaja de operar a cualquier hora del día y la capacidad de crear imágenes en bandas que el sol no emite. [2]

Diferentes suelos, vegetación, o humedad reflejan diferentes bandas de radiación. Estas imágenes son usadas en varias aplicaciones, desde Sistemas de Información Geográfica y mapas a meteorología y monitoreamiento de la salud de vegetación forestal. Un índice bastante común es el Índice de Vegetación de Diferencia Normalizada, o NDVI por sus siglas en inglés, el cual es usado para estimar la cantidad, calidad y desarrollo de la vegetación con base a la mediación. Estos datos ya se usan en sistemas de advertencia temprana de sequías y la predicción del rendimiento de la agricultura en los Estados Unidos a partir de los datos de la NASA. [3]

La importancia de la producción agropecuaria y agroganadera en el Paraguay también invita a considerar estas tecnologías para el monitoreamiento de la salud de la vegetación y el uso adecuado de la tierra. Actualmente, ya se están empleando tecnologías de teledetección y el NDVI en el sector agrícola en aplicaciones como la detección de malezas y predicción del orden ideal de cosecha de campos cultivados. [4]

Con aproximadamente la mitad del territorio paraguayo hacia el norte del Río Paraguay en la región semi-árida del Chaco, tecnologías que alivien las sequías y precipitación baja son muy valiosas, tanto para la agricultura y ganadería en las estancias chaqueñas como para centros poblacionales más aislados como por ejemplo las misiones indígenas. Estos pueblos más aislados generalmente son caracterizados por pobreza, lo que se manifiesta en la falta de alimento y agua potable.

Un paleocauce es un cauce por el cual antiguamente fluía agua, como por ejemplo un antiguo lecho de un río. Los paleocauces han sido propuestos como reservorios o conductos para el flujo subterráneo de agua dulce. Se consideran de interés principalmente los paleocauces arenosos, y estos pueden ser aprovechados para acceder al agua en áreas en las que la distribución habitual del agua no existe o está dificultada de alguna forma. [5] Con la abundancia de paleocauces en el Chaco central (ocupan un 15 % de la región), esta propuesta es una bastante prometedora que ya ha sido considerada en investigaciones anteriores. [6]

La detección de estos paleocauces se haría a partir de imágenes satelitales en una serie temporal por medio de redes neuronales. Las redes convolucionales son una categoría de redes neuronales especializadas para el procesamiento de imágenes. El principio básico de su funcionamiento consiste en la convolución de grupos píxeles cercanos, una operación que permite tener en cuenta no solo el valor de cada píxel individual, sino el contexto de los mismos. [7]

El resto del escrito está compuesto por las siguientes secciones: (2) una breve descripción del proyecto, (3) los objetivos general y específicos, (4) las bases conceptuales del proyecto, (5) el estado del arte relacionado al proyecto, (6) la metodología de la solución, (7) la importancia del presente proyecto, y finalmente (8) se presenta el estado actual del proyecto.

2. Descripción del proyecto

La motivación principal del proyecto es la necesidad de mitigar las sequías prolongadas del Chaco Paraguayo, que impacta de forma más severa a zonas remotas o rurales. El objetivo del trabajo es aplicar técnicas de clasificación e interpretación de imágenes satelitales por medio de redes neuronales al problema de la identificación de usos de suelo, particularmente para identificar paleocauces.

3. Objetivos

3.1. Objetivo General

Creación de modelos a partir de redes neuronales convolucionales para la clasificación y caracterización de imágenes satelitales.

3.2. Objetivos Específicos

1. Análisis de series temporales a partir de imágenes satelitales correspondientes a la región occidental del Paraguay
2. Identificación y clasificación de componentes de uso de suelo
3. Determinación de áreas de ocurrencia de paleocauces

4. Marco Conceptual

En esta sección se exploran brevemente los conceptos más fundamentales para el proyecto. En particular, los temas a desarrollarse son una revisión del estado de la teledetección y los recursos disponibles, las bases de las redes neuronales convolucionales y su uso en el análisis de imágenes satelitales, las técnicas modernas en uso en la investigación y en aplicaciones en el mundo real.

4.1. Teledetección

Teledetección se refiere a la captación o detección remota de alguna señal o imagen. En este contexto nos referimos específicamente a imágenes captadas por medio de un sensor montado en un satélite artificial o algún vehículo aéreo como un avión o un dron, para extraer información. Estas imágenes contienen información multiespectro, es decir, además de la luz visible se toman imágenes de bandas invisibles como por ejemplo la luz infrarroja. [2] A lo largo de este proyecto, el término “teledetección” se refiere a la captación de imágenes por medio de satélites.

Para la captura de estas imágenes se emplean varios métodos, que se dividen en dos categorías: sensores pasivos recolectan radiación electromagnética reflejada del sol, mientras que sensores activos emiten su propia radiación y captan la reflexión de la tierra. Sensores pasivos requieren de una cantidad importante de energía para operar, pero tienen la ventaja de operar a cualquier hora del día y la capacidad de crear imágenes en bandas que el sol no emite. [2]

Los primeros programas de observación de la tierra por medio de satélites surgieron en los años 70 y 80. El primero fue el programa Landsat de los Estados Unidos en 1972, y le siguieron programas similares en India, Francia y la Unión Europea. [8]

4.1.1. Aplicaciones

Imágenes satelitales proveen información muy útil para todo tipo de estadísticas en áreas relacionadas con el territorio, como por ejemplo la agricultura, silvicultura y el estudio de uso del suelo. El estudio de la agricultura a gran escala por medio de la teledetección se realizó por primera vez entre 1974 y 1977 por medio de datos de Landsat 1, a cargo de la NASA, la Oficina Nacional de

Administración Oceánica y Atmosférica (NOAA) y el Departamento de Agricultura de los Estados Unidos (USDA). [9]

Dado que las imágenes producidas generalmente cubren toda o casi toda el área de estudio, y que suelen ser multiespectrales, lo que provee datos que fotografías ordinarias no contienen, cualquier aplicación que involucre estudiar un área vasta puede beneficiarse de ellas. Dependiendo de la resolución, aplicaciones que involucren detalles más finos también las pueden aprovechar, como por ejemplo su uso en aplicaciones de mapas digitales.

4.1.2. Características de los datos

La calidad de imágenes recolectadas por teledetección se mide de cuatro formas, estas son su resolución espacial, espectral, radiométrica y temporal.

Resolución espacial: el tamaño de un píxel en una imagen rasterizada. Típicamente corresponde a un área cuadrada de entre 1 y 1000 m^2 .

Resolución espectral: la longitud de onda de las diferentes bandas de frecuencia capturadas, normalmente relacionada a la cantidad de bandas de frecuencia. El sensor Hyperion en “Earth Observing-1”, por ejemplo, observa 220 bandas entre 0,4 y 2,5 μm , con una resolución espectral de 0,10–1,11 μm por banda. [10]

Resolución radiométrica: la cantidad de niveles de intensidad de radiación detectable por el sensor. Típicamente entre 8 y 14 bits de información, correspondiente a 256 a 16384 niveles en cada banda. La cantidad de ruido en el sensor también afecta la resolución radiométrica.

Resolución temporal: la cantidad de sobrevuelos del avión o satélite, importante solamente cuando se realizan series de tiempo, promedios o mosaicos, como por ejemplo en el monitoreamiento de la agricultura.

4.1.3. Disponibilidad de recursos

Existen varios repositorios de datos de teledetección disponibles para usos comerciales como académicos. Los programas de observación terrestre de la NASA y de la ESA, Landsat y Copernicus respectivamente, disponibilizan recursos por medio de portales en la internet. Para los datos de Landsat, uno de los recursos más accesibles es Google Earth Engine, que permite el procesamiento de imágenes en línea, de forma gratuita para usos no comerciales. [11] El programa Copernicus por otro lado provee un navegador de imágenes, una forma de descargar datos con algunos filtros, y todo esto de forma gratuita tanto para fines académicos como comerciales. [12] También ofrecen un espacio de trabajo en línea, similar en propósito a Google Earth Engine. [13]

4.2. Redes Neuronales Convolucionales

Una Red Neuronal Convolutiva (o CNN por sus siglas en inglés) es un tipo de red neuronal artificial en la cual las neuronas procesan datos de entrada por medio de filtros de convolución. Esto implica el procesamiento de un grupo de datos cercanos, lo que permite interpretar el contexto de un dato, en contraste con redes neuronales típicas. Esta propiedad hace que las CNN sean el método preferido para el procesamiento de imágenes por medio de redes neuronales. [14] [15]

Este método de procesamiento permite procesar una gran cantidad de entradas con una cantidad reducida de neuronas, comparado con una red neuronal típica con la misma capacidad.

Por ejemplo, considerando una red neuronal con una capa de entrada y una capa siguiente con la misma cantidad de neuronas N en ambas, una red neuronal densa, es decir donde cada neurona de una capa está conectada a cada neurona de la siguiente capa, contiene $N \times N$ conexiones. En contraste, una red neuronal convolutiva equivalente estaría compuesta de tan solo N neuronas, mientras que al mismo tiempo captura un grupo de píxeles en cada neurona en lugar de uno solo.

Para el procesamiento se utilizan los filtros de convolución, matrices de dimensiones reducidas comparadas con la imagen, cuyas celdas contienen coeficientes. Este filtro se superpone sobre una sección de la imagen, y los valores de los píxeles se multiplican con los de la celda superpuesta del filtro, y la suma de los productos es el resultado de la convolución del grupo de píxeles. Con

una representación adecuada de los datos de cada píxel, estos filtros, también llamados *kernels*, pueden usarse en la detección de bordes en cualquier orientación, reducción de ruido, aumentación de intensidad de píxeles de cierto color o brillo, entre otros. [15]

4.2.1. Aplicaciones

CNN ya se han utilizado en todo tipo de aplicaciones relacionadas con imágenes y videos, entre ellas clasificación de imágenes, segmentación de imágenes, detección de objetos e inclusive en análisis de imágenes de inundación para predecir la gravedad de uno de estos tipos de desastre natural. [16]

También se ha usado extensivamente en aplicaciones relacionadas a la teledetección, con una gran colección de técnicas, conjuntos de datos, material de aprendizaje y software de libre acceso en artículos web, videos y repositorios de código. [17] [18] Queda claro que los modelos convolucionales son muy eficaces en el procesamiento de imágenes, y la cantidad de material estudiado relacionado a las imágenes satelitales facilitaría enormemente la aplicación en el tema de este proyecto.

4.2.2. Técnicas y arquitecturas

Las primeras CNN surgieron en los años 90, con LeNet siendo la primera implementación que ganó atención. Esta red se desarrolló para el reconocimiento de dígitos escritos a mano, y consistía de capas convolucionales, de pooling y capas completamente conectadas.

Recién en 2012 con AlexNet se logró el siguiente salto, con una competencia de reconocimiento visual. AlexNet se diseñó con conjuntos de imágenes de gran escala en mente, compuesta de capas similares a LeNet, con algunas optimizaciones en las funciones de activación y en medidas contra overfitting.

Nuevos desarrollos en los años siguientes se enfocaron en la optimización y la solución de problemas específicos. VGGNet, originando en Oxford, se popularizó por su simplicidad, con kernels pequeños de 3x3 y capas convolucionales en secuencias. Google introdujo GoogLeNet demostrando la efectividad del paralelismo con sus módulos *inception*, que además mejoraron la capacidad de generalización usando kernels de diferentes tamaños al mismo tiempo. Otra arquitectura, Redes Residuales o ResNets, abordaron el desafío de entrenar redes muy profundas por medio de conexiones que saltan una o varias capas, facilitando el entrenamiento de redes de hasta cientos de capas. Otra red diseñada por Google es MobileNet, una arquitectura diseñada para ejecutarse en ambientes de recursos limitados como dispositivos móviles que busca equilibrar el rendimiento con la eficiencia.

Otra arquitectura importante es la U-Net, diseñada para la segmentación de imágenes biomédicas, que funciona mediante reducciones sucesivas de resolución de la imagen de entrada, seguidas por aumentos sucesivos que se combinan con las reducciones respectivas ya procesadas. Esta técnica permite entrenar modelos con mejor segmentación y menos datos de entrenamiento, y con tarjetas gráficas modernas (de 2015 en adelante), procesar una imagen de 512×512 toma menos de un segundo. [19] También se han utilizado U-Nets en aplicaciones de reducción de ruido en imágenes en modelos de difusión, lo cual se sigue utilizando en tecnologías de generación de imágenes como *DALL-E*, *Midjourney* y *Stable Diffusion*. [20] También se ha usado la arquitectura U-Net en segmentación de imágenes satelitales para identificar rasgos de imágenes, como recursos de agua, bosques o agricultura con una intersección de entre 81 y 96 % con marcaciones manuales. [21]

4.3. Análisis de imágenes satelitales

Las formas más comunes de análisis de imágenes son la clasificación, la segmentación, la detección de cambios y las series de tiempo. Existen muchas técnicas usadas en aplicaciones más específicas, como la predicción del rendimiento de una plantación o la salud de la vegetación, la reducción de ruido o redes generativas, que no se aplican tan directamente para el objetivo de este trabajo.

4.3.1. Clasificación

La clasificación es una tarea fundamental en el análisis de datos de teledetección, en el cual el objetivo es etiquetar cada imagen, como por ejemplo “área urbana”, “bosque”, “agricultura”, etc. El proceso de asignar etiquetas a imágenes se conoce como clasificación a nivel de imagen. [18]

Sin embargo, en algunos casos una imagen puede contener más de un tipo de uso de suelo, como por ejemplo un bosque con un río que lo divide, o una ciudad con áreas comerciales y residenciales. En estos casos, clasificación a nivel de imagen se vuelve más compleja e implica asignar múltiples etiquetas a cada imagen. Esto se puede lograr por medio de una combinación de extracción de características y algoritmos de *Machine Learning* para identificar los diferentes tipos de uso de suelo. [18]

Es importante no confundir la clasificación a nivel de imagen con la clasificación a nivel de píxel, también conocida como segmentación semántica. Mientras que clasificación a nivel de imagen asigna una etiqueta a una imagen entera, la segmentación semántica asigna una etiqueta a cada píxel de la imagen, lo que resulta en una representación detallada y precisa del uso de suelo en una imagen. [22]

4.3.2. Segmentación

La segmentación consiste en dividir una imagen en segmentos o regiones semánticamente significativas. El proceso de segmentación de imágenes asigna una etiqueta de clase a cada píxel de una imagen, transformándola de una grilla 2D de píxeles a una grilla 2D de etiquetas. Una aplicación común es la segmentación de calles o edificios, donde el objetivo es separar las calles y los edificios de otras características de la imagen. [18]

Para realizar esta tarea, modelos de una clase única son frecuentemente entrenados para detectar y diferenciar entre calles y el ambiente, o edificios y el ambiente. Estos modelos se diseñan para reconocer características específicas como el color, la textura y la forma que son típicas de una calle o un edificio para que puedan etiquetar los píxeles que forman parte de estas estructuras en una imagen. [22]

Otras aplicaciones comunes se encuentran en la agricultura o clasificación de uso de suelo en una imagen. En este caso, se utilizan modelos multiclase que son capaces de diferenciar entre varias clases en una imagen, como por ejemplo bosques, áreas urbanas y tierra agrícola. Estos modelos son capaces de reconocer relaciones más complejas entre tipos de uso de suelo, y permiten un entendimiento más integral del contenido de la imagen. [22]

4.3.3. Detección de cambios

Detección de cambios es un componente vital del análisis de teledetección, permitiendo el monitoreo de cambios de un paisaje a lo largo del tiempo. Esta técnica se puede aplicar para identificar una amplia gama de cambios, entre otros el cambio de uso de suelo, desarrollo urbano, erosión costal y deforestación. [18]

Detección de cambios puede ser realizada entre dos imágenes tomadas en diferentes momentos, o analizando una serie de imágenes tomadas a lo largo de un periodo de tiempo. [18]

Una consideración importante es que la detección de cambios puede verse afectada por la presencia de nubes y sombras. Estos factores dinámicos pueden alterar la apariencia de un paisaje y causar falsos positivos en los resultados. Por ende, es importante considerar estos factores y emplear técnicas que puedan mitigar estos efectos. [18]

4.3.4. Serie de tiempo

La serie de tiempo consiste en una serie de datos ordenados por el tiempo. A menudo se trata de muestras tomadas en intervalos regulares, pero no necesariamente debe ser así. El análisis de series de tiempo se persigue con el fin de extraer estadísticas, patrones o características generales de los datos.

El análisis de series de tiempo en teledetección tiene numerosas aplicaciones, incluyendo mejorar la exactitud de modelos de clasificación y el pronóstico de patrones y eventos futuros, especialmente en la agricultura, por ejemplo en la predicción de la producción de una plantación. [18]

5. Estado del Arte

En este capítulo se explora el estado del arte del uso de imágenes satelitales en diversas áreas y las técnicas de análisis relevantes para este proyecto. Esta investigación tiene el fin de entender la forma en que se aplican en sus diversos campos de aplicación y cuáles técnicas son las más eficaces en el campo a estudiarse.

5.1. Estrategias de búsqueda

Para la revisión de literatura se utilizaron términos referentes a [Redes Neuronales], [Teledetección], y [Clasificación y Detección]. Se tuvieron en cuenta principalmente obras en el idioma inglés, aunque se incluyen obras en español también. Para la búsqueda se usaron los siguientes términos:

Términos	Sinónimos
Neural Network	Convolutional Neural Network Deep Learning
Remote Sensing	Satellite Imagery
Classification	Detection
Lack of data	Small data

Las cadenas de búsqueda se construyen a partir de los términos y sus sinónimos. Las cadenas con mejores resultados fueron [*Neural Network AND Remote Sensing AND Classification*], [*Convolutional Neural Network AND Remote Sensing AND Classification*], [*Convolutional Neural Network AND Satellite Imagery AND Classification*], [*Convolutional Neural Network AND Satellite Imagery AND Detection*], y [*Deep Learning AND Remote Sensing AND Small Data*]

El motor de búsqueda utilizado fue Google Scholar, poniendo mayor enfoque en resultados provenientes de bases de datos reconocidas y establecidas como IEEE Xplore, ScienceDirect y ArXiv.

También se incluyeron trabajos relevantes para la universidad y la realidad local del país, proveídos por los tutores.

Criterios de selección Se incluyen artículos, papers, conferencias, y otros trabajos formales debidamente documentados. Se establecen los siguientes criterios para juzgar si un trabajo es incluido o excluido de esta investigación:

Inclusión 1 Trabajos que se enfoquen en la clasificación de imágenes satelitales por medio de redes neuronales en el rango de publicación de 2014 a 2024.

Inclusión 2 Trabajos que coincidan en su contenido con los términos de búsqueda.

Inclusión 3 Trabajos cuyo contenido sea relevante para la investigación.

Exclusión 1 Trabajos que no contengan las palabras claves o son irrelevantes para el campo de investigación.

Exclusión 2 Trabajos que se centran en un término de búsqueda pero no incluyen alguno de los demás.

Exclusión 3 Trabajos con una cantidad mayoritaria de información irrelevante para el tema estudiado.

Procedimientos de selección El proceso de selección de trabajos se basa en responder las preguntas de investigación presentadas en la siguiente sección, con el fin de responderlas con información válida y actual.

Se limita el número de artículos incluidos a 20, y en caso de que se supere la cantidad encontrada se filtran por medio de los siguientes criterios:

P.S.1. Los trabajos deben responder la mayor cantidad de preguntas de investigación.

P.S.2. Los trabajos deben contar con la mayor cantidad de incidencia de términos definidos anteriormente.

P.S.3. Artículos que incluyan las palabras clasificar, interpretar, imágenes satelitales, redes neuronales, redes neuronales convolucionales en su resumen, conclusión.

Extracción y síntesis de datos Para la planilla de extracción de datos de cada estudio, se guardaron título, autores, año de publicación, resumen, palabras claves, fuente y conclusiones relacionadas a las preguntas de investigación. En el cuadro de la sección de resultados se listan las informaciones relevantes para responder las preguntas de investigación de este proyecto. Para determinar la inclusión de cada artículo se realizó un análisis de los objetivos y resultados de cada trabajo, teniendo en cuenta los criterios de selección. Para realizar la síntesis de los datos se realizó la estrategia descriptiva, que detalla y ordena las conclusiones principales de los autores de los artículos para una mejor comprensión de las ideas principales.

5.2. Preguntas de investigación

El objetivo principal de este estudio es determinar cual es el estado del arte en técnicas utilizadas para clasificar y caracterizar o interpretar imágenes satelitales por medio de redes neuronales convolucionales. Con este fin en mente, se plantean las siguientes preguntas de investigación:

P.1. ¿Qué proyectos se están llevando adelante para clasificar y caracterizar imágenes satelitales usando redes neuronales convolucionales?

P.2. ¿Cuáles son las ventajas y/o desventajas de la clasificación y caracterización de imágenes satelitales usando redes neuronales convolucionales en comparación con las alternativas?

P.3. ¿Qué soluciones existen para abordar la falta de datos de entrenamiento para las redes neuronales convolucionales?

5.3. Resultados

Las respuestas a las preguntas de investigación se encuentran resumidas en los cuadros 1, 2, y 3.

Discusión de los resultados

El campo de análisis de datos de teledetección ha visto muchos proyectos en los últimos años, y las redes convolucionales han sido una pieza crucial para crear modelos mas poderosos y eficientes que las redes neuronales convencionales. La exploración de alternativas a redes neuronales, como RF y SVN, indica que la CNN es la tecnología más prometedora para la clasificación de una gran cantidad de datos.

La posibilidad de usar datos sin etiquetamiento en el entrenamiento indica que la falta de datos etiquetados no es razón para dejar de considerar el entrenamiento de modelos independientes. Sin embargo, gracias a la cantidad de cómputo necesaria para la tarea, el exploramiento que se realice en este trabajo necesariamente deberá ser limitado. Aun así, gracias a estos factores, el resultado debería reflejar con bastante precisión los métodos necesarios para crear modelos con objetivos poco comunes en terreno nuevo para la teledetección a nivel nacional.

También de importancia es la naturaleza altamente dinámica de los datos. Cada año la cantidad de datos de alta resolución se multiplica y se vuelve más accesible, y para mantenerse al tanto de estos cambios, son necesarios el mantenimiento y la actualización periódica de estos modelos. Por otro lado, una de las técnicas más utilizadas en los trabajos mencionados es la transferencia de modelos preentrenados por medio de “fine-tuning”, lo que hace el mantenimiento un obstáculo más fácil de superar.

Proyecto	Objetivos	Métodos y Observaciones
Clasificación y Segmentación de Ortofotografía Satelital Usando Redes Neuronales Convolucionales. [23]	Explorar el uso de CNN para la clasificación por píxel completa, rápida y exacta de una ciudad pequeña.	Un CNN es apto para el análisis de imágenes multispectrales corregidas ortográficamente, junto con un modelo de superficie digital de una pequeña ciudad.
Identificación y Mapeo de Paleocauces Utilizando Imágenes Satelitales de Alta Resolución en la Llanura Costera de la Bahía Samborombón, Argentina [24]	Reconocimiento y análisis de paleocauces por medio de sensores remotos de los paleocauces de los ríos Samborombón y Salado, en la zona de su desembocadura, para la reconstrucción paleoambiental.	Se utilizan imágenes satelitales multispectrales de alta resolución en combinación con algunos filtros de dirección convolucionales. Sin embargo, no se utilizan redes neuronales.
Redes Neuronales Completamente Convolucionales para Clasificación de Imágenes Satelitales [25]	Demostrar un modelo CNN que utiliza solamente capas convolucionales, sin capas tradicionales, para la clasificación de imágenes satelitales.	El modelo resultante tiene mejor rendimiento que modelos CNN con capas tradicionales y menor tiempo de entrenamiento.
Detección Automática de Objetos en Imágenes Aéreas basada en CNN [26]	Aprovechamiento de la cantidad de imágenes de teledetección por medio de CNN para la detección automática de objetos en diversas aplicaciones.	CNNs son útiles para la detección de objetos, mejor que métodos basados en “features” y varios otros modelos CNN.
SatCNN: clasificación de datos de imágenes satelitales usando CNN ágiles [27]	Diseño de una arquitectura CNN especializada para el análisis de imágenes satelitales en vez de adaptar un modelo existente creado para la clasificación de escenas naturales.	Pruebas en varias muestras de datos muestran una efectividad de más de 99.5 %.
Un CNN basado en pedazos (patch-based) para la clasificación de datos de teledetección [28]	Creación de un modelo para la clasificación de imágenes de resolución media, donde estructuras finas no existen para la clasificación por píxel.	Mejora de hasta 25 % por encima de CNN basados en clasificación por píxel.
Clasificación de Imágenes Satelitales con Deep Learning [29]	Creación de modelos CNN para la automatización del análisis de grandes cantidades de imágenes satelitales de alta resolución.	El modelo creado gana el segundo puesto en la competición de Functional Map of the World (fMoW) TopCoder, 15 de 63 clases clasificadas con exactitud de 95 %.
CNN para la Clasificación de Humedales Complejos Usando Imágenes de Teledetección Óptica [30]	Entrenamiento de un modelo preentrenado para la clasificación de imágenes de alta resolución.	Rendimiento mejor que RF, inclusive con menos características de entrada, con una exactitud de 94 %.
DeepSat V2: CNN de Características Aumentadas para la Clasificación de Imágenes Satelitales [31]	Creación de conjuntos de datos de imágenes satelitales etiquetadas y un modelo CNN con varias técnicas nuevas para su análisis.	El modelo resultante logra clasificar correctamente con más de 99 % de exactitud en los nuevos conjuntos de datos.
Un CNN para la Detección de Anomalías Térmicas Volcánicas en Imágenes Satelitales [32]	Entrenamiento de un modelo CNN para analizar imágenes de actividad volcánica en infrarrojo.	Un CNN preentrenado adaptado a nuevos datos seleccionados a mano logra buena clasificación.

Cuadro 1: Algunos proyectos de clasificación y segmentación de imágenes de teledetección y CNN

6. Propuesta de Solución

El proyecto consiste principalmente de dos fases: la recopilación y procesamiento de datos, y la clasificación de imágenes. El procesamiento de datos puede abarcar un amplio espectro de posibilidades, incluyendo preprocesamiento para eliminar defectos o disminuir la cantidad de datos,

Tipo	Característica	Referencias
Ventaja	CNN patch-based (basado en pedazos) mejor que NN convencional o CNN basados en píxeles, SVN o RF	[28]
Ventaja	Clasificación de datos multifuente por medio de CNN mejor que SVN y ELM	[33]
Ventaja	Clasificación de uso de suelo por CNN mucho mejor que RF, especialmente para terrenos difíciles	[30]
Desventaja	CNN basado en píxeles comparable o peor que NN convencional, SVN o RF	[28]
Desventaja	Entrenamiento de modelos basados en redes neuronales es más computacionalmente costoso que SVN o RF	[28, 33, 30]

Cuadro 2: Ventajas y desventajas de CNN en comparación con otras técnicas

Técnica	Descripción, (+) Ventajas, (−) Desventajas	Referencias
Transferencia (Transfer, Fine-tuning)	Uso de modelo preentrenado con un conjunto de datos relevante y ajustado con un conjunto de datos nuevo. (+) Mejor rendimiento, menos datos de entrenamiento, mejor generalizabilidad. (−) Riesgo de reducción de rendimiento con transferencia a dominio diferente, tamaño de modelo grande.	[34, 35, 36] [37, 27, 32]
Auto supervisado (Self-supervised)	Creación de un modelo con etiquetas creadas por el modelo, seguido de entrenamiento supervisado con etiquetas proveídas. (+) Uso de datos no etiquetados, reconocimiento de patrones sin necesidad de etiquetación, mejor generalizabilidad. (−) Computacionalmente caro, posibilidad de que el modelo deje de entrenarse con algunas técnicas.	[34]
Semi supervisado (Semi-supervised)	Mezcla de entrenamiento supervisado y no supervisado con conjuntos de datos etiquetados y no etiquetados. (+) Uso de datos etiquetados y no etiquetados, mejor generalizabilidad. (−) Computacionalmente caro, riesgo de “overfitting”, sensible a calidad de datos.	[34]
Débilmente supervisado (Weakly-supervised)	Desarrollo de un modelo con datos etiquetados parcialmente, de manera imprecisa o con ruido. (+) Costo de etiquetamiento reducido, permite un modelo inexacto para permitir escalabilidad. (−) Computacionalmente caro, menos exacto que entrenamiento (completamente) supervisado.	[34]
Conjunto (Ensemble)	Combinación de muchos modelos individuales que aprendieron patrones de forma diferente para la predicción. (+) Mejor generalizabilidad, robustez contra perturbación de datos e incertidumbre. (−) Computacionalmente caro, peor interpretabilidad que un modelo simple.	[34, 23, 29]
Validación cruzada (Cross-validation)	Entrenar y validar un modelo varias veces usando diferentes particiones de datos para el entrenamiento y la validación. (+) Modelo menos sesgado, evita reportaje sobre-optimista de rendimiento, mejor generalizabilidad. (−) Computacionalmente caro	[34]

Cuadro 3: Técnicas para abarcar el problema de pocos datos

y la mejora de imágenes, que incluye la manipulación de los datos para aumentar el contraste, resaltar bordes, o combinar datos de varias bandas. A continuación se exploran algunos pasos a seguir.

6.1. Recopilación de Imágenes

La recopilación de datos principalmente consistirá de fuentes primarias de imágenes satelitales. La fuente de estos datos será el portal de datos abiertos de Copernicus, el programa de teledetección y observación terrestre de la Agencia Espacial Europea (ESA), la cual provee acceso libre y completo a cualquier persona u organización a los datos recolectados.[38] Estas imágenes pueden ser descargadas con varios formatos, bandas espectrales, o segmentación básica como capas adicionales,

con resoluciones de entre 10 y 60 metros por píxel.

Adicionalmente, varias fuentes secundarias se utilizarán para la verificación de los resultados producidos por el modelo, como por ejemplo mapas y diagramas geológicos.

6.2. Preprocesamiento de Imágenes

El objetivo del preprocesamiento es la corrección de errores y la exclusión de datos irrelevantes o demasiado degradados como para ser de uso. Se desea mejorar la utilidad de las imágenes sin alterar su esencia ni amplificar características individuales. Algunas de las técnicas a utilizarse son:

Eliminación de ruido: ruido en una imagen satelital puede ser causado por una variedad de razones como interferencia electrónica entre componentes o mal funcionamiento de los sensores. Este ruido puede degradar una imagen e incluso ocultar información, y la eliminación de ruido tiene por objetivo disminuir el efecto del ruido sobre los datos recolectados. Las técnicas específicas dependen del tipo de ruido, es decir si es periódico (patrones de ruido que se repiten) o aleatorio, o una combinación de ambos. [39]

Corrección radiométrica: implica la corrección de datos erróneos causados por distorsiones específicas provenientes de efectos atmosféricos o errores de calibración que pueden causar píxeles perdidos o bandedado en la imagen. [39]

Georreferenciación: las imágenes satelitales son combinadas con una referencia en un sistema de coordenadas estándar, cada píxel es asociado con una coordenada. Debido a que las imágenes capturadas por el programa Copernicus ya contienen este dato, este paso se realiza al descargar las imágenes. [39]

Subconjuntos de bandas y creación de mosaicos: el análisis de un subconjunto de capas y la omisión de datos de poco interés ayudan a disminuir el volumen de datos para analizar, y la creación de mosaicos a partir de múltiples imágenes puede servir para cubrir áreas más amplias, o analizar estructuras que no fueron capturadas en su totalidad en una imagen. [39]

6.2.1. Mejora de Imágenes

Estos pasos se toman para aumentar la utilidad de las imágenes, resaltando ciertas características con el objetivo de ampliar la cantidad de información que se puede interpretar a partir de las imágenes. Algunas formas de mejora de imágenes incluyen las siguientes:

Manipulación de contraste: consiste en resaltar diferencias entre diferentes partes de una imagen. Existen varias técnicas de manipulación de contraste, como la umbralización de niveles de gris, la segmentación, o el estiramiento de contraste. La primera permite definir un nivel un nivel mínimo o máximo de brillo para cada píxel, aislando ciertas áreas. Esto permite, por ejemplo, resaltar cuerpos de agua. La segunda permite definir ciertos niveles de los píxeles de la imagen para realizar una segmentación. La tercera consiste en ampliar el rango de valores en una imagen, que puede hacer algunos detalles más visibles. [39]

Manipulación de características espaciales: incluye filtros espaciales, filtros convolucionales y el realce de bordes. Filtros espaciales pueden incluir un amplio rango de filtros, y se usan para aumentar o eliminar ciertas características de una imagen, como la aumentación de nitidez o filtro de paso bajo. Filtros convolucionales también cumplen diversas funciones, y consisten de un cálculo por cada píxel en función a sus vecinos, lo que permite realzar cierta información local, como la detección de bordes, y permite mantener baja la cantidad de datos, combinando datos de muchos píxeles en uno. El realce de bordes consiste en realzar el contraste entre bordes de la imagen para aumentar su nitidez percibida. [39, 40]

Manipulación de múltiples imágenes: la creación de imágenes compuestas por la combinación de datos de varias bandas puede ser útil para realzar detalles que de otra manera no son visibles. Por ejemplo, la combinación de bandas de color rojo y bandas de infrarrojo cercano permite evaluar la salud de vegetación. Varios índices como NDVI, SAVI o NDSI presentan datos normalizados al rango de 0 a 1 que evalúan relaciones de este estilo entre varias bandas de una imagen. [39]

6.2.2. Clasificación

Este es el objetivo principal del proyecto, crear un modelo capaz de reemplazar la interpretación visual por parte de una persona entrenada por técnicas cuantitativas para automatizar la identificación de características en una escena. Para imágenes satelitales, esto implica el análisis de múltiples bandas de la imagen, que son en efecto dimensiones adicionales. El objetivo final es asignar etiquetas a cada píxel de una imagen de acuerdo a una lista de posibilidades. [39]

La implementación de este modelo se hará en Python, que gracias a la multitud de librerías de Machine Learning y manipulación de imágenes facilita el prototipado y desarrollo de modelos de clasificación.

6.2.3. Entrenamiento y Validación del Modelo

Para el entrenamiento del modelo, el conjunto de datos recolectados será particionado aproximadamente en subconjuntos de 70 %, 20 % y 10 % para su uso en entrenamiento, validación y prueba respectivamente. Esto se hace para que la verificación y prueba del modelo no sean influenciadas por imágenes sobre las cuales el modelo ya fue entrenado. Esta configuración será dinámica, es decir, podrá ser ajustada dependiendo de cuál configuración demuestra mejores resultados.

6.2.4. Modelo y Propuesta de Solución

Para la solución proponemos el uso de Redes Neuronales Convolucionales debido a que su eficacia en la interpretación de imágenes ya ha sido demostrada extensamente, tanto para la clasificación binaria como para la segmentación con múltiples etiquetas. Se utilizarán técnicas del estado del arte para explorar la clasificación y caracterización de imágenes satelitales del territorio Paraguay, yendo más allá de la segmentación simple ya disponible en servicios de imágenes satelitales.

Se utilizará un conjunto de imágenes preprocesadas y mejoradas por las técnicas mencionadas en las secciones anteriores, etiquetados correctamente para permitir un enfoque sobre el aprendizaje supervisado para la generación de un modelo clasificador. El resultado será un modelo capaz de clasificar imágenes de diversos lugares, no solo el área de interés de este proyecto, lo que permitirá entrenar modelos para su uso en nuevos proyectos de forma más fácil.

7. Relevancia del Presente Proyecto

Este proyecto es de importancia para ... ?

8. Estado Actual del Proyecto

El cuadro 4 contiene una lista de las actividades que denotan el avance de la propuesta y el estado de cada una de ellas.

	Actividad	Estado
1	Revisión de trabajos previos y análisis del estado del arte.	Finalizado
2	Recopilación del conjunto datos para entrenamiento y de prueba.	En curso
3	Preprocesamiento de datos: limpieza y preparación de los datos.	No iniciado
4	Selección del algoritmo de aprendizaje automático más adecuado.	En curso
5	Entrenamiento del modelo y ajuste de parametros.	No iniciado
6	Evaluación del rendimiento del modelo entrenado en el conjunto de prueba.	No iniciado
7	Evaluación de predicción en datos nuevos.	No iniciado
8	Despliegue en producción.	No iniciado
9	Documentación del proceso de implementación.	En Curso

Cuadro 4: Estado de las actividades planificadas para este trabajo

Referencias

- [1] Crash Dummy. *On creating a filler bibliography entry*. Disponible en <https://example.com/> (accedido 2024-07-28).
- [2] Global Forest Link. *How does satellite imaging work?* Disponible en <https://globalforestlink.com/how-does-satellite-imaging-work/> (accedido 2024-06-24).
- [3] Earth Data. *Vegetation*. Disponible en <https://www.earthdata.nasa.gov/learn/find-data/near-real-time/hazards-and-disasters/vegetation> (accedido 2024-06-24).
- [4] Catherine Stepanova. *Ricardo Rodríguez: "Gracias a la aplicación de OneSoil ahorramos USD 100 por hectárea"*. Disponible en <https://blog.onesoil.ai/es/interview-with-ricardo-rodriguez> (accedido 2024-06-24).
- [5] S.M. White et al. "Small Muddy Paleochannels and Implications for Submarine Groundwater Discharge near Charleston, South Carolina, USA. *Geosciences*". En: 13.232 (2023).
- [6] Antero J. N. Cabrera et al. "Sistemas de Captación y Almacenamiento de Agua en el Chaco Central". En: (2021), pág. 7. URL: https://www.conacyt.gov.py/sites/default/files/upload_editores/u454/MANUAL-SISTEMAS-CAPTACION-ALMACENAMIENTO-AGUA-CHACO.pdf.
- [7] Vladimir Iglovikov, Sergey Mushinskiy y Vladimir Osin. "Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition". En: (jun. de 2017). URL: <https://arxiv.org/abs/1706.06169>.
- [8] European Space Agency. *50 years of Earth Observation*. Disponible en https://web.archive.org/web/20120130034957/http://www.esa.int/esaMI/Space_Year_2007/SEMP4FEVL2F_0.html (accedido 2024-07-30).
- [9] J. Donald Allen. "A Look at the Remote Sensing Applications Program of the National Agricultural Statistics Service". En: *Journal of Official Statistics* 6.4 (1990), págs. 393-409.
- [10] Holli Riebeek. *Earth Observing-1: Ten Years of Innovation*. Disponible en <https://web.archive.org/web/20220418062231/https://earthobservatory.nasa.gov/features/E01Tenth> (accedido 2024-07-30).
- [11] Landat Science. *Data Access*. Disponible en <https://landsat.gsfc.nasa.gov/data/data-access/> (accedido 2024-08-03).
- [12] Copernicus. *Copyright and licences*. Disponible en <https://www.copernicus.eu/en/access-data/copyright-and-licences> (accedido 2024-08-03).
- [13] Copernicus. *About the Copernicus Data Space Ecosystem*. Disponible en <https://dataspace.copernicus.eu/about> (accedido 2024-08-03).
- [14] Aurélien Géron. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. Sebastopol, CA: O'Reilly Media, 2019.
- [15] Dan Ciresan et al. "Flexible, High Performance Convolutional Neural Networks for Image Classification". En: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence* 2 (2011), págs. 1237-1242.
- [16] R.J. Pally y S. Samadi. "Application of image processing and convolutional neural networks for flood image classification and semantic segmentation". En: *Environmental Modelling & Software* 148 (2022). DOI: <https://doi.org/10.1016/j.envsoft.2021.105285>. URL: <https://www.sciencedirect.com/science/article/pii/S1364815221003273>.
- [17] Abdishakur. *Land use/Land cover classification with Deep Learning*. Disponible en <https://towardsdatascience.com/land-use-land-cover-classification-with-deep-learning-9a5041095ddb> (accedido 2024-08-10). Ago. de 2018.

- [18] Robin Cole y Mikolaj Czerkawski. Disonible en <https://github.com/satellite-image-deep-learning/techniques> (accedido 2024-08-10).
- [19] Olaf Ronneberger, Philipp Fischer y Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV]. URL: <https://arxiv.org/abs/1505.04597>.
- [20] Jonathan Ho, Ajay Jain y Pieter Abbeel. “Denoising Diffusion Probabilistic Models”. En: *arXiv preprint arxiv:2006.11239* (2020).
- [21] Vladimir Khryashchev et al. “Convolutional Neural Network for Satellite Imagery”. En: *FRUCT Conference 22* (2018). URL: <https://fruct.org/publications/volume-22/acm22/files/Khr.pdf>.
- [22] Robin Cole. *A brief introduction to satellite image segmentation with neural networks*. Disponible en <https://medium.com/@robmarkcole/a-brief-introduction-to-satellite-image-segmentation-with-neural-networks-33ea732d5bce> (accedido 2024-09-01). 2023.
- [23] Martin Långkvist et al. “Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks”. En: (2016). DOI: <https://doi.org/10.3390/rs8040329>. URL: <https://www.mdpi.com/2072-4292/8/4/329>.
- [24] Mariel Samanta Luengo et al. “Identificación y Mapeo de Paleocauces utilizando Imágenes Satelitales de Alta Resolución en la Llanura Costera de la Bahía Samborombón, Este de la Provincia de Buenos Aires, Argentina.” En: (2016). URL: https://www.researchgate.net/publication/325827963_IDENTIFICACION_Y_MAPEO_DE_PALEOCAUCES_UTILIZANDO_IMAGENES_SATELITALES_DE_ALTA_RESOLUCION_EN_LA_LLANURA_COSTERA_DE_LA_BAHIA_SAMBOROMBON_ESTES_DE_LA_PROVINCIA_DE_BUENOS_AIRES_ARGENTINA.
- [25] Emmanuel Maggiori et al. “Fully convolutional neural networks for remote sensing image classification”. En: (2016). DOI: <https://doi.org/10.1109/IGARSS.2016.7730322>. URL: <https://ieeexplore.ieee.org/abstract/document/7730322>.
- [26] Igor Ševo y Aleksej Avramović. “Convolutional Neural Network Based Automatic Object Detection on Aerial Images”. En: (2016). DOI: <https://doi.org/10.1109/LGRS.2016.2542358>. URL: <https://ieeexplore.ieee.org/abstract/document/7447728>.
- [27] Yanfei Zhong et al. “SatCNN: satellite image dataset classification using agile convolutional neural networks”. En: (2016). DOI: <https://doi.org/10.1080/2150704X.2016.1235299>. URL: <https://www.tandfonline.com/doi/abs/10.1080/2150704X.2016.1235299>.
- [28] Atharva Sharma et al. “A patch-based convolutional neural network for remote sensing image classification”. En: (2017). DOI: <https://doi.org/10.1016/j.neunet.2017.07.017>. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0893608017301806>.
- [29] Mark Pritt y Gary Chern. “Satellite Image Classification with Deep Learning”. En: (2017). DOI: <https://doi.org/10.1109/AIPR.2017.8457969>. URL: <https://ieeexplore.ieee.org/abstract/document/8457969>.
- [30] Mohammad Rezaee et al. “Deep Convolutional Neural Network for Complex Wetland Classification Using Optical Remote Sensing Imagery”. En: (2018). DOI: <https://doi.org/10.1109/JSTARS.2018.2846178>. URL: <https://ieeexplore.ieee.org/abstract/document/8401505>.
- [31] Qun Liu et al. “DeepSat V2: feature augmented convolutional neural nets for satellite image classification”. En: (2019). DOI: <https://doi.org/10.1080/2150704X.2019.1693071>. URL: <https://www.tandfonline.com/doi/abs/10.1080/2150704X.2019.1693071>.
- [32] Eleonora Amato et al. “A Deep Convolutional Neural Network for Detecting Volcanic Thermal Anomalies from Satellite Images”. En: (2023). DOI: <https://doi.org/10.3390/rs15153718>. URL: <https://www.mdpi.com/2072-4292/15/15/3718>.

- [33] Xiaodong Xu et al. “Multisource Remote Sensing Data Classification Based on Convolutional Neural Network”. En: (2017). DOI: <https://doi.org/10.1109/TGRS.2017.2756851>. URL: <https://ieeexplore.ieee.org/abstract/document/8068943>.
- [34] Anastasiia Safonova et al. “Ten deep learning techniques to address small data problems with remote sensing”. En: (2023). DOI: <https://doi.org/10.1016/j.jag.2023.103569>. URL: <https://www.sciencedirect.com/science/article/pii/S156984322300393X>.
- [35] Emmanuel Maggiori et al. “Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification”. En: (2016). DOI: <https://doi.org/10.1109/TGRS.2016.2612821>. URL: <https://ieeexplore.ieee.org/abstract/document/7592858>.
- [36] Marco Castelluccio et al. “Land Use Classification in Remote Sensing Images by Convolutional Neural Networks”. En: (2015). DOI: <https://doi.org/10.48550/arXiv.1508.00092>. URL: <https://arxiv.org/abs/1508.00092>.
- [37] Keiller Nogueira, Otávio A.B. Penatti y Jefersson A. dos Santos. “Towards better exploiting convolutional neural networks for remote sensing scene classification”. En: (2017). DOI: <https://doi.org/10.1016/j.patcog.2016.07.001>. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0031320316301509>.
- [38] Copernicus. *Access to data*. Disponible en <https://www.copernicus.eu/en/access-data> (accedido 2025-02-21).
- [39] T. Lillesand, R.W. Kiefer y J. Chipman. *Remote Sensing and Image Interpretation*. Wiley, 2015. ISBN: 9781118343289. URL: <https://books.google.com.py/books?id=AFHDCAAAQBAJ>.
- [40] U.S. Geological Survey. *The Power of Spatial Filters*. Disponible en <https://astrogeology.usgs.gov/docs/concepts/image-processing/the-power-of-spatial-filters/> (accedido 2025-02-17).