

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 [10]

1. INTRODUCCIÓN Y MOTIVACIÓN

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

1.1. Objetivos

1.1.1. Objetivo General

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

1.1.2. 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

2. 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.

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.

2.1. Teledetección

Imágenes satelitales o teledetección se refiere a imágenes capturadas por un sensor montado en un satélite artificial, 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. [14]

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. [14]

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. [3]

2.1.1. 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). [4]

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.

2.1.2. 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 pixel 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. [16]

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.

2.1.3. 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. [9] 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. [8] También ofrecen un espacio de trabajo en línea, similar en propósito a Google Earth Engine. [2]

2.2. 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. [11] [5]

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. [5]

2.2.1. 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. [15]

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. [1] [7] 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.

2.2.2. 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. [17] 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*. [12] 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. [13]

2.3. 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.

2.3.1. 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. [7]

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 involucra 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. [7]

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. [6]

2.3.2. 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. [7]

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. [6]

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. [6]

3. PROPUESTA DE ANÁLISIS DE DATOS

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

4. PASOS A SEGUIR

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Referencias

- [1] 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.
- [2] *About the Copernicus Data Space Ecosystem*. Disponible en <https://dataspace.copernicus.eu/about> (accedido 2024-08-03).
- [3] 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).
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Robin Cole y Mikolaj Czerkawski. Disponible en <https://github.com/satellite-image-deep-learning/techniques> (accedido 2024-08-10).
- [8] *Copyright and licences*. Disponible en <https://www.copernicus.eu/en/access-data/copyright-and-licences> (accedido 2024-08-03).
- [9] *Data Access*. Disponible en <https://landsat.gsfc.nasa.gov/data/data-access/> (accedido 2024-08-03).
- [10] Crash Dummy. *On creating a filler bibliography entry*. Disponible en <https://example.com/> (accedido 2024-07-28).
- [11] Aurélien Géron. *Hands-on Machine Learning with Scikit-Learn, Keras, and Tensor-Flow*. Sebastopol, CA: O'Reilly Media, 2019.
- [12] Jonathan Ho, Ajay Jain y Pieter Abbeel. «Denoising Diffusion Probabilistic Models». En: *arXiv preprint arxiv:2006.11239* (2020).
- [13] 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>.
- [14] Global Forest Link. *How does satellite imaging work?* Disponible en <https://globalforestlink.com/how-does-satellite-imaging-work/> (accedido 2024-06-24).
- [15] 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>.
- [16] 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).

- [17] 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>.