ISSN 1870-4069

Implementación de CNN basada en una arquitectura VGG16 para detección y clasificación de árboles mediante la segmentación semántica en imágenes aéreas

Laritza Pérez-Enríquez, Raquel Díaz-Hernández,
Leopoldo Altamirano-Robles

Instituto Nacional de Astrofísica Óptica y Electrónica,
Posgrado en Ciencias y Tecnología del Espacio,
México

{laritza,raqueld,robles}@inaoep.mx

Resumen. Factores como, el calentamiento global, la contaminación, la sobre población, y la inconciencia humana, son algunas de las causas que han generado impacto de forma negativa en nuestro planeta, provocando consecuencias irreparables como la extinción de recursos naturales no renovables y poner algunas especies en peligro de extinción. Para combatir esta situación se han implementado situaciones legales que protegen áreas y especies en peligro, sin embargo, no es suficiente con realizar la protección a lo que ya se encuentra en riesgo de extinción, también es necesario tomar medidas que ayuden a impedir que más recursos naturales lleguen a estas circunstancias. La segmentación semántica es una técnica que se ha aplicado para la selección de píxeles pertenecientes a una clase, mediante la aplicación de algoritmos de redes neuronales convolucionales (CNN), los cuales son una herramienta que permite realizar mediante entrenamiento con conjuntos de datos el reconocimiento, identificación y selección de objetos. En este trabajo se propone aplicar una red neuronal convolucional basada en una arquitectura VGG16 para realizar segmentación semántica en árboles de palmeras, pinos y naranjos, con la intención de replicar esta técnica en otros tipos de especies de árboles aunado finalmente a la aplicación de la técnica de transferencia de aprendizaje (TL) implementada para mejorar los resultados de segmentación de forma exitosa y eficiente alcanzando una precisión del 94 % a 96.5 %.

Palabras clave: Percepción remota, CNN, VGG16, segmentación semántica, transferencia de aprendizaje.

Implementation of CNN based on a VGG16 Architecture for Detection and Classification of Trees through Semantic Segmentation in Aerial Images

Abstract. Factors such as global warming, pollution, overpopulation, and human unconsciousness are some of the causes that have generated a negative

impact on our planet, causing irreparable consequences such as the extinction of non-renewable natural resources and putting some species in danger. Danger of extinction. To combat this situation, legal situations have been implemented that protect endangered areas and species, however, it is not enough to protect what is already at risk of extinction, it is also necessary to take measures that help prevent more resources natural come to these circumstances. Semantic segmentation is a technique that has been applied to the selection of pixels belonging to a class, through the application of convolutional neural network (CNN) algorithms, which are a tool that allows training with data sets to perform recognition, identification and selection of objects. In this work, it is proposed to apply a convolutional neural network based on a VGG16 architecture to perform semantic segmentation in palm, pine and orange trees, with the intention of replicating this technique in other types of tree species, finally coupled with the application of the transfer learning (TL) technique implemented to improve segmentation results successfully and efficiently reaching an accuracy of 94

Keywords: Remote sensing, CNN, VGG16, semantic segmentation, transfer learning.

1. Introducción

Actualmente, resulta difícil poder realizar inspección de campo de zonas rurales, debido al poco o nulo acceso en algunas de estas áreas. Es mediante recorridos que se puede obtener información y datos para estadísticas de las especies vegetales de la zona y tener un registro que se encuentre disponible para un control eficiente de estás, que en su momento, se encuentren en peligro de extinción o puedan llegar a estarlo.

Hoy en día, el uso de técnicas como la inspección aérea mediante vehículos aéreos no tripulados (VANT) está revolucionando y facilitando la forma de obtener las imágenes de estas zonas. Para realizar la identificación y clasificación de especies se han aplicado, también, técnicas de Inteligencia Artificial (IA). La IA ha evolucionado en los últimos años, y se ha aplicado en distintas áreas tales como: minería espacial, medicina, vehículos autónomos, entre otras.

Con la aplicación de CNNs [1, 2, 3, 5, 6, 7] se ha logrado principalmente realizar análisis de imágenes en cuyas escenas se puede detectar, identificar y clasificar todo tipo de objetos y formas para los cuales la red es entrenada. Estas redes neuronales, son implementadas mediante distintas arquitecturas, las cuales se encuentran formadas por numerosas estructuras y capas convolucionales que es lo que principalmente las hace diferentes una de otra, sin embargo, siguen un mismo objetivo, el detectar el objeto deseado.

No obstante es mediante la técnica de segmentación que se puede lograr obtener y visualizar las clases de objetos de interés [1, 3, 5, 6, 7], ya que la segmentación realizará un agrupamiento en los píxeles característicos que constituyen la forma de la clase indicada. Otro aspecto relevante, es que las arquitecturas pueden ser reentrenables, y además, se puede aplicar técnicas que ayuden a obtener mejor precisión en los resultados de segmentación como se muestra en [2, 3, 5, 8].

Algunos trabajos como [2, 6] implementan la arquitectura VGG16 para detección y selección mediante la aplicación de la técnica de transferencia de aprendizaje, con la cual logran utilizar parámetros adquiridos y aplicarlos para identificar otros objetos. Recientemente métodos basados en una red totalmente convolucional (FCN) realizaron un gran progreso en la segmentación semántica [1, 6, 7], para la clasificación de imágenes, utilizan características que no se encuentran en métodos tradicionales, sin embargo, la aplicación de estas técnicas para segmentación semántica a nivel de píxel trae como consecuencia un alto costo computacional.

Métodos como la segmentación de instancias [5] y segmentación débilmente supervisada [3] exploran la unión entre la supervisión a nivel de píxel y la supervisión mediante cuadros delimitadores con la finalidad de disminuir el uso de recursos computacionales, con esto se ha logrado obtener una mejora en la precisión de la detección y clasificación, sin embargo, aún quedan algunos rasgos sin identificar.

Por lo anterior, en este trabajo se propone la implementación de una red neuronal convolucional aplicando una arquitectura VGG16 para realizar la segmentación semántica en imágenes aéreas que incluyen zonas de vegetación para detectar árboles de naranjo y finalmente aplicar transferencia de aprendizaje para mejorar los resultados de precisión obtenidos.

2. Trabajos relacionados

El reconocimiento de objetos mediante la segmentación semántica, es un término general para describir una colección de tareas relacionadas con la visión por computadora, que involucran el reconocimiento de imágenes y análisis de datos cómo se menciona en [6, 12]. Esta acción ha sido la base del estudio de nuevos métodos para realizar estas tareas acercándose cada vez más a la automatización.

La segmentación semántica es realizada con un algoritmo de aprendizaje profundo como en [8, 10, 11] que es implementado usando redes neuronales convolucionales para realizar la tarea de identificación y clasificación en imágenes, como menciona Long J. at al. (2015) donde el enfoque principal de su trabajo está relacionado a la aplicación de redes totalmente convolucionales para extender la clasificación hasta la segmentación y mejorar la técnica realizando combinaciones de arquitecturas como AlexNet[14], VGG net [15], GoogLeNet [16], para obtener segmentación semántica con capas profundas gruesas y capas poco profundas finas para producir segmentaciones precisas y detalladas.

Ellos reportan mejora en los resultados y simplificó y acelero el aprendizaje de inferencia, sin embargo sus resultados reportados se encuentran alrededor del 62.2 % IoU promedio (métrica de intersección sobre la unión). Por otro lado, se ha abordado la segmentación semántica implementando una red SegNet [13], la cual se basa en una arquitectura de CNN [8, 11], para realizar la comprensión de escenas mediante la segmentación semántica por medio de una red de codificación.

Esta red es idéntica topológicamente a la arquitectura VGG16, solo se omiten las capas completamente conectadas para que esta sea más pequeña y más fácil de entrenar. He, K. et al. (2017) [5], introduce técnicas de Mask R-CNN, que realiza instancias de segmentación en las imágenes con algoritmos que pueden ser reentrenables.

Con esta técnica se mejoró de forma significativa los resultados de precisión de la segmentación aplicando Faster R-CNN [9], pues realizan máscaras de segmentación en cada Región de Interés (RoI) para mejorar la predicción. Por otro lado, Guo, R., et al. (2020) [3], realiza segmentación semántica con múltiples etiquetas usando algoritmos que utilizan cuadros delimitadores y técnicas que mejoran la detección de objetos tales como: Grab Cut, GradCAMD y GrabCutC.

El entrenamiento es realizado con un modelo R-CNN y utilizan una red troncal Resnet50. Con esta técnica demostraron mejoraras en la segmentación y rendimiento aplicando la técnica al conjunto de datos iSAID. Lobo Torres, at al (2020) [17], evalúo cinco redes totalmente convolucionales: SegNet, U-Net, FC-DenseNet y dos variantes de DeepLabv3+ para realizar segmentación semántica de una especie de árbol; además, verificó los beneficios del fully connected conditional random (CFR) como un paso de procesamiento posterior para mejorar los mapas de segmentación.

En su análisis experimental mostró resultados promedio de precisión de entre 88.9 % a 96.7 %. Además indico que CFR mejoró el rendimiento pero a un alto costo computacional. Osco, L., et al. (2021) [1], resalta la importancia de la cartografía en áreas agrícolas; utilizó vehículos aéreos no tripulados (VANT) con la integración de cámaras multiespectrales que facilitan el mapeo de plantas en paisajes agrícolas.

En [1] aplican métodos de aprendizaje profundo para realizar segmentación semántica a imágenes espectrales de árboles cítricos, utilizando cinco arquitecturas diferentes para realizar la segmentación: FCN, U-Net, SegNet, dynamic dilated convolution network (DDCN) y Deep-LabV3 +. Con estos métodos de aprendizaje profundo, obtuvieron precisiones que van desde el 94.88 % al 95.46 %.

De esta forma se puede ver que se ha implementado CNN con las distintas arquitecturas conocidas y mencionadas en este trabajo con aplicaciones a distintas áreas, pero sobre un mismo objetivo, realizar segmentación para la detección y clasificación de distintos objetos de diferentes clases, una ventaja de la red VGG16 respecto a otras es que su arquitectura es fácil de comprender y de implementar, logró excelentes resultados en la competencia ImageNet (ILSVRC-2014) entre el 96 % y 97 %, ademas podemos encontrar esta red disponible en Keras, también fue entrenada para resolver problemas de clasificación de 1000 clases en el conjunto de datos de ImageNet el cual se conforma por mas de 1.4 millones de imágenes, por esto es utilizada para realizar entrenamiento en áreas a fines a los objetos etiquetados o para continuar con la segmentación de otras clases, se reduce el tiempo de entrenamiento de la red gracias al uso de los pesos ImageNet. La finalidad principal es mejorar los tiempos de procesamiento de la información, obtener una mejor precisión en el resultado y reducir el uso de recursos computacionales durante el procesamiento.

3. Método propuesto

Existen asociaciones que se dedican al cuidado y preservación de las especies vegetales en peligro de extinción y es mediante la recolección de datos que pueden obtener información del lugar y la especie para poder realizar registros y obtener estadísticas de los cambios pertinentes de la evolución de la diversidad de la flora en el área.

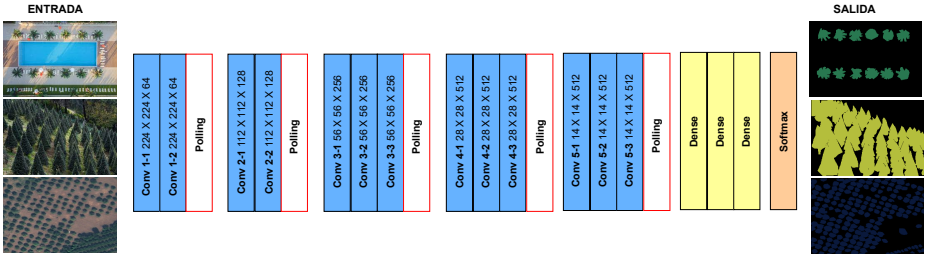

Fig. 1. Diagrama de arquitectura VGG16.

Sin embargo, esta actividad resulta no ser tan fácil, cómoda y segura de realizar en la forma tradicional, es por ello que gracias al uso de la tecnología hoy en día se puede explorar la zona de manera segura sin exponer la integridad del personal a cargo de esta labor, ya que con el uso de drones podemos obtener imágenes de estas áreas y analizar e identificar mediante algoritmos de inteligencia artificial los elementos que se encuentran presente en las imágenes.

Es por esto que principalmente se propone la implementación de una CNN basada en un algoritmo VGG16 pre-entrenado con el conjunto de datos ImageNet para la segmentación semántica de imágenes con vegetación, cuya forma general de funcionamiento se puede ver en la Fig. 1 que se muestra a continuación.

Como se observa en la Fig. 1, a la entrada de la arquitectura se tiene la imagen re-dimensionada a $224 \times 224 \times 3$ que contiene la escena donde se pueden apreciar zonas con árboles de palmeras, pinos y naranjos. Se realiza con cada una de las imágenes el recorrido en la arquitectura por cada una de las capas donde se aplican filtros convolucionales que permitirán la extracción de características particulares de cada una de las clases.

Según se avanza en la arquitectura el número de convoluciones será mayor, esto para poder extraer características más particulares que puedan lograr hacer la distinción entre un objeto y otro. Así pues, se lleva a cabo el proceso hasta finalizar con la última imagen que conforma el conjunto de datos de entrenamiento para enseñar al algoritmo a identificar cada clase.

Este procedimiento es repetido con el conjunto de datos de validación para poder saber y medir la precisión que obtuvo en el entrenamiento, así se puede determinar si se debe volver a entrenar y/o ajustar los parámetros de número de pasos, épocas u otro dato.

Una vez que se obtiene la matriz que contiene los resultados del entrenamiento, se guardan los pesos para usarlos durante la aplicación de la transferencia de aprendizaje. A continuación se enumeran los pasos que se realizaron previo para realizar este trabajo:

1) Adquisición de datos: En esta etapa se conformó el conjunto de datos, mediante la recolección de imágenes de distintas fuentes para el caso de datos de palmeras y pinos; para el conjunto de datos de árboles de naranjo se pudieron obtener las imágenes de una fotografía aérea capturada de una plantación de naranjos en la zona del estado de Veracruz,

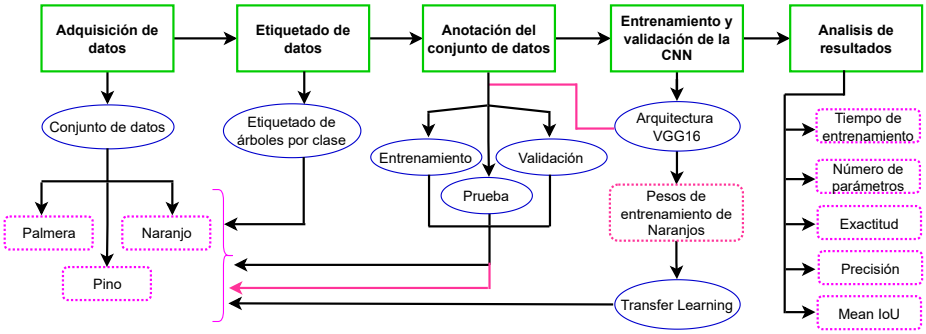

Fig. 2. Esquema del flujo de trabajo.

2) Etiquetado de datos: Esta acción se realiza de forma manual con la ayuda del software: ImageAnnotator, este permite realizar el etiquetado de las clases (palmera, pino y naranjo) mediante la selección de forma delimitada con puntos alrededor del borde del árbol, generando polígonos y finalmente se le asigna el nombre que le corresponde a la clase de árbol,
3) Anotación del conjunto de datos: A partir del etiquetado de las imágenes se realizaron máscaras de segmentación utilizando los puntos de intersección en las imágenes etiquetadas. Con estas imágenes y sus máscaras de segmentación de cada clase, se forman los conjuntos de datos de entrenamiento y validación,
4) Entrenamiento y validación de la CNN: Se ejecuta la CNN, a partir de la arquitectura VGG16 seleccionada y se establecen y/o ajustan parámetros específicos para la ejecución del entrenamiento de la red con los distintos conjuntos de datos que hemos formado para este trabajo. Una vez realizado el entrenamiento se emplea la validación para determinar el correcto funcionamiento del algoritmo,
5) Análisis de resultados: Finalmente se analizan los resultados obtenidos mediante las métricas que nos indican la precisión del algoritmo dependientes de valores como el tiempo de entrenamiento, número de parámetros, etc.

En la Fig. 2 se detalla el procedimiento de forma esquematizada.

3.1. Adquisición de datos

Para formar los conjuntos de datos, se realizó una recopilación de imágenes aéreas en las cuales se visualizará árboles de palmeras, pinos y en el caso de naranjos, las imágenes fueran obtenidas de una captura aérea de una zona de cultivos de naranjos en el estado de Veracruz. Estas imágenes se encuentran en formato RGB y de distintas dimensiones. En el caso de las palmeras y pinos las imágenes tenían un tamaño inicial de 480×480, y las de los naranjos 255×255. En la Fig. 3 se muestra algunas imágenes que conforman la base de datos de cada clase.

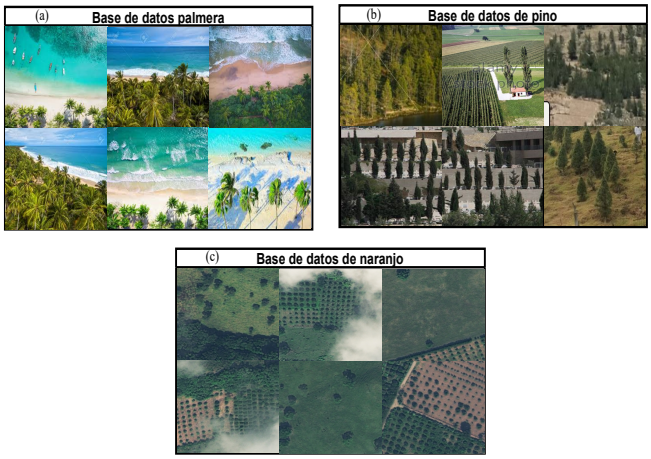

Fig. 3. Imágenes aéreas: (a) palmeras, (b) pinos, (c) naranjos.

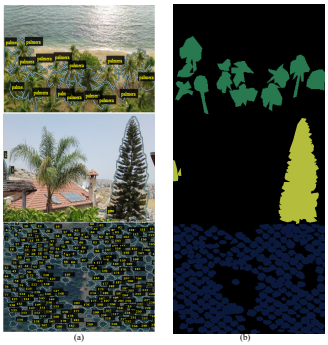

Fig. 4. (a) Etiquetado de las clases, (b) Máscaras de las clases.

3.2. Etiquetado de datos

El etiquetado de los árboles de las imágenes consiste en ir asignando a cada imagen la clase a la que pertenecen para cada tipo: palmera, pino y naranjo. Se utilizó el sotfware Image Annotator [18] para realizar esta tarea.

Una vez finalizado el etiquetado de cada imagen se genera un archivo descargable que contiene los puntos que indican y seccionan el área que abarca cada árbol en la escena. Este archivo es procesado en un programa realizado en lenguaje Python para obtener las máscaras de segmentación de las clases. En la Fig. 4 se muestra ejemplo del etiquetado por clase, y la máscara obtenida.

3.3. Anotación del conjunto de datos

Para la generación de los conjuntos de datos una vez que se ha realizado el etiquetado de las 420 imágenes y generado las máscaras, se forma el conjunto de datos de entrenamiento, este contiene 340 imágenes de distintas zonas y el objeto principal

Tabla 1. Datos de entrenamiento y validación de red con imágenes de pino.

Prueba	Batch	Epoch	Precisión de pixel	IoU	Tiempo
1	6	50	94 %	86 %	1560 seg
2	6	80	96 %	90 %	1980 seg
3	10	120	96.5 %	92 %	2975 seg

Tabla 2. Datos de entrenamiento y validación de red con imágenes de naranjo.

Prueba	Batch	Epoch	Precisión de pixel	IoU	Tiempo
1	6	50	87 %	64 %	2280 seg
2	6	80	89 %	70 %	2640 seg
3	10	120	91 %	73 %	3040 seg

son árboles de palmeras, pinos y naranjos lo que corresponde al 80 % del conjunto de datos, estas se utilizan para realizar el entrenamiento de la red. El conjunto de datos para validación, está conformado por 80 imágenes del conjunto de datos que corresponde a el 20 % del conjunto, estas son utilizadas para evaluar la precisión de la segmentación de las distintas clases de árboles trabajados.

Para el conjunto de datos de prueba se utilizaron 15 imágenes RGB que contienen en la escena las 3 clases de árboles, las imágenes de prueba son empleadas finalmente para probar el funcionamiento de la red mediante la segmentación de las clases palmera, pino y naranjo y así verificar de forma visual la precisión en el trabajo, estos conjuntos de datos podrán estar disponibles para uso público(una vez que el comité correspondiente acredite su publicación, la cual ya esta en proceso) en la plataforma GitHub.

3.4. Entrenamiento y validación

Una vez conformado el conjuntos de datos, se inicia la ejecución del programa de CNN para poder realizar el entrenamiento, validación y finalmente la prueba, con la cual podremos ver si se cumple el objetivo que es obtener la segmentación semántica de la imagen con clase de árbol que corresponde, esto toma un tiempo que va de 30 a 50 min.

La ejecución del programa se realiza en Python con la paquetería de TensorFlow [1] utilizando la arquitectura VGG16 pre-entrenada con datos de ImageNet. Para este trabajo se implementa finalmente la técnica de transferencia de aprendizaje [2], con la cual se pretende usar bajos recursos computacionales, disminuir el tiempo de procesamiento durante el entrenamiento y adaptar pesos del modelo para aplicarlos en el aprendizaje de las clases trabajadas.

3.5. Análisis de resultados

Para realizar la evaluación del algoritmo, se utilizaron las siguientes métricas:

– Tiempo de entrenamiento.- El tiempo de entrenamiento es el tiempo que le tomó al programa realizar el entrenamiento del algoritmo, este varía dependiendo del número de imágenes a procesar y va desde los 30 min a 1:40 min,

Tabla 3. Datos de entrenamiento y validación de red con imágenes de palmera.

Prueba	Batch	Epoch	Precisión de pixel	mean IoU	Tiempo
1	6	50	91 %	76 %	1680 seg
2	6	80	93 %	82 %	2040 seg
3	10	120	95 %	86 %	2400 seg

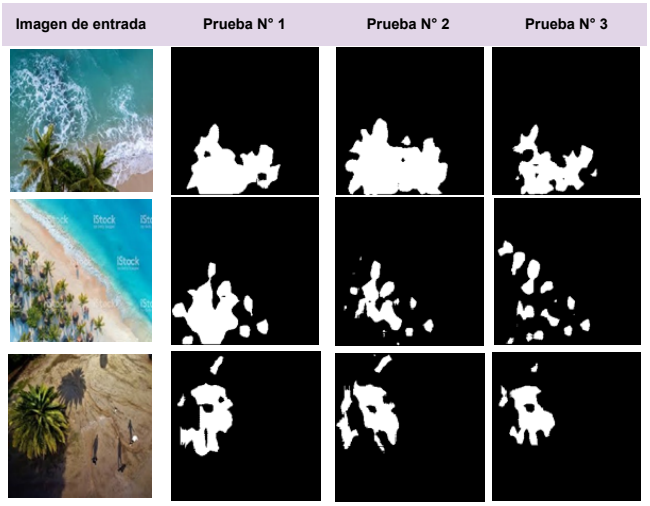

Fig. 5. Resultado de segmentación semántica de árboles de palmera.

– Número de pasos (batch size).- Define el número de muestras que se propagara a través de la red,
– Número de épocas (epoch).- Este determina el número de veces que el conjunto completo de imágenes será procesado por la red,
– Exactitud (accuracy).- La exactitud se refiere a lo cerca que esta el resultado de una medición del valor verdadero. En términos estadísticos esta se relaciona con el sesgo de una estimación,
– Precisión.- Se representa por la proporción de verdaderos positivos dividido entre todos los resultados positivos (tanto verdaderos positivos como falsos positivos), de esta manera se puede ver de forma porcentual que tan precisa fue la predicción respecto al conjunto de datos,
– IoU.- Es una métrica de evaluación utilizada comunmente para la segmentación semántica de imágenes. Esta métrica se define como:

$$IoU = \frac{VP}{VP + FP + FN}, \tag{1}$$

donde VP son los verdaderos positivo, FP los falsos positivo y FN representados por los falsos negativo.

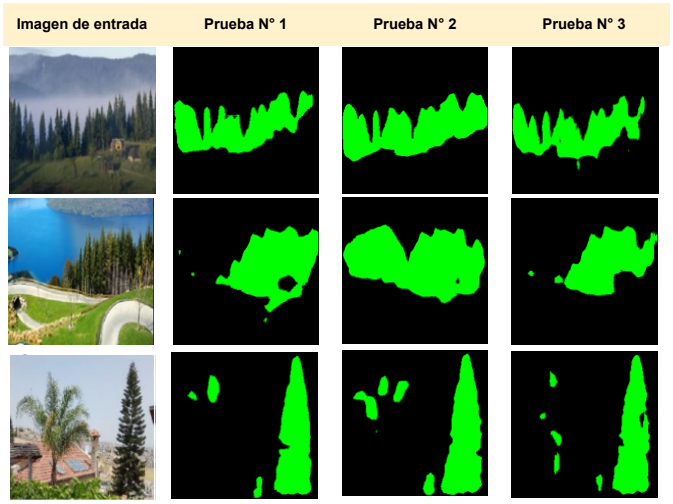

Fig. 6. Resultado de segmentación semántica de árboles de pino.

4. Pruebas y resultados

Se realizaron las pruebas con los 3 conjuntos de datos de árboles de palmeras, pinos y naranjos. Ahora se explica y muestra los resultados obtenidos en cada prueba.

4.1. Resultados con conjunto de datos de palmeras usando pesos ImageNet

Para el entrenamiento con el conjunto de datos de palmeras, se considero en el programa 3 pruebas estableciendo los parámetros que se muestran en la Tabla 3. Además se puede ver el resultado obtenido reflejado en porcentajes, teniendo entonces 91 %, 93 % y 95 % de precisión de píxel en las pruebas 1, 2 y 3.

En la Fig. 5 se pueden visualizar los resultados obtenidos en las imágenes del conjunto de datos de palmeras. El tiempo de ejecución para obtener la segmentación semántica en cada caso estuvo alrededor de los 15 a 20 min.

4.2. Resultados con conjunto de datos de pinos usando pesos ImageNet

Para el entrenamiento con el conjunto de datos de pinos también fueron consideradas 3 pruebas, estableciendo los parámetros que se muestran en la Tabla 1. Se refleja de forma porcentual la mejora de la precisión del resultados teniendo 94 %, 96 % y 96.5 % de precisión de píxel en las pruebas 1, 2 y 3 respectivamente. A continuación, en la Fig. 6 se muestran parte de los resultados obtenidos para cada una de las pruebas en las imágenes del conjunto de datos de pino.

4.3. Resultados con conjunto de datos de naranjos usando pesos ImageNet

Para el entrenamiento con el conjunto de datos de naranjo también se consideraron las 3 pruebas con los parámetros establecidos en los conjuntos de palmera y pino,

Tabla 4. Datos aplicando transferencia de aprendizaje con base de datos de naranjo.

Prueba	Batch	Epoch	Precisión de pixel	IoU	Tiempo
1	6	50	93 %	81 %	1896 seg
2	6	80	95 %	85 %	2338 seg
3	10	120	96 %	87 %	2720 seg

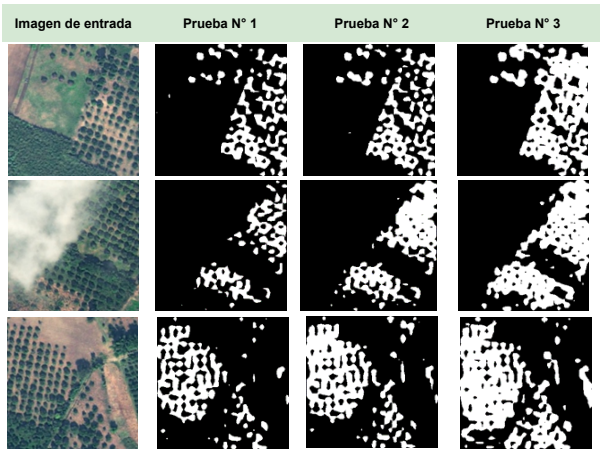

Fig. 7. Resultado de segmentación semántica de árboles de naranjo.

ver Tabla 2, donde los resultados de precisión de píxel obtenidos fueron 87 %, 89 % y 91 % respectivamente. En la Fig. 7 son mostrados algunos resultados de segmentación semántica obtenidos para cada una de las pruebas con los parámetros establecidos en la Tabla 3. Una vez analizado los casos para los 3 conjuntos de datos de árboles, se toma la decisión de realizar la implementación de la transferencia de aprendizaje para el caso del conjunto de datos de árboles de naranjo.

4.4. Resultados aplicando transferencia de aprendizaje a conjunto de datos de naranjos

Se decidió aplicar la técnica de transferencia de aprendizaje con la finalidad de mejorar la precisión de la detección y segmentación de la clase e implementar una técnica que resulte significativa para trabajos futuros y que puedan ayudar a reducir los tiempos de ejecución costos y recursos computacionales. En la Tabla 4, se muestran los parámetros establecidos para realizar el entrenamiento utilizando los pesos que se obtuvieron del entrenamiento con el conjunto de datos de naranjos pero ahora implementando la transferencia del aprendizaje con árboles de naranjos.

Se observa como se mejora los resultados de precisión en cada prueba comparada con el obtenido en la prueba usando pesos de ImageNet. Ver Tabla 2. Ahora en la Fig. 8 se muestran los resultados que se obtuvieron aplicando la transferencia de aprendizaje en la arquitectura VGG16 para obtener la detección de los árboles de naranjo mediante la segmentación semántica.

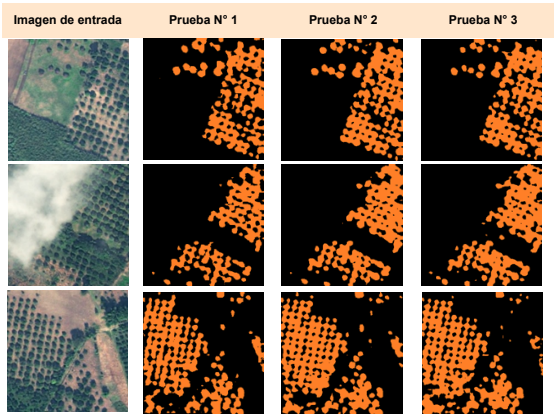

Fig. 8. Resultado de segmentación semántica en árboles de naranjo aplicando transferencia de aprendizaje.

En la imagen de entrada 1 se puede ver que en la prueba 1 se detectan algunos píxeles de árboles que están pegados a otros y no se pueden visualizar como uno solo, sin embargo para esa misma imagen en la prueba 2 aún se siguen visualizando árboles segmentados no separados pero en menor cantidad, para el caso de la prueba 3 de la misma imagen se visualiza una mejor precisión y se observa cómo se ha definido un poco mejor entre un árbol y otro, esto rectifica los datos obtenidos en la Tabla 4 y por lo tanto también si comparamos la Tabla 4 con la Tabla 3 se demuestra que al aplicar la técnica de transferencia de aprendizaje se logra mejorar de forma significativa los resultados.

5. Conclusiones y trabajo a futuro

Se puede observar que la aplicación de las CNN mediante la arquitectura VGG16 es un procedimiento que logró el objetivo principal que era detectar y segmentar las clases de las especies de árboles utilizadas en este trabajo, obteniendo resultados con una precisión alrededor del 87 % hasta 96.5 %, lo cual demuestra que fue efectivo y eficiente el uso de los datos y la aplicación de la técnica de transferencia de aprendizaje.

Con esto se puede decir que se puede implementar la metodología propuesta entrenando con otros conjuntos de datos de áreas de vegetación principalmente para tratar con especies en peligro de extinción, y poder abordar y proponer soluciones relacionadas a este tema importante, mediante el monitoreo constante de zonas protegidas donde se encuentren estas especies.

Se demostró que la transferencia de aprendizaje ayuda a mejorar los resultados de detección de objetos, segmentando además de forma semántica las clases específicas. Es importante mencionar que también, esta técnica ayudo a reducir los tiempos de ejecución y a evitar el uso excesivo de recursos computacionales. Debido a que esta técnica utiliza el conocimiento obtenido mediante el entrenamiento para realizar la segmentación de otras clases, solo se deberá aplicar a un trabajo relacionado al área para no iniciar todo el procedimiento desde cero en el entrenamiento.

Como trabajo a futuro se propone realizar mediante la aplicación de la técnica de transferencia de aprendizaje, la detección y segmentación de distintas clases de árboles en imágenes aéreas de zonas con vegetación, además de implementar este método con arquitecturas como: DeepLabV3, SegNet, Faster R-CNN, aplicado a imágenes RGB y Multiespectrales. También se propone la aplicación de la técnica de Fine tuning.

Referencias

1. Osco, L. P., Nogueira, K., Marques-Ramos, A. P., Faita-Pinheiro, M. M., Furuya, D. E. G., Gonçalves, W. N., dos-Santos, J. A.: Semantic segmentation of citrus-orchard using deep neural networks and multispectral UAV-based imagery. Precision Agriculture, vol. 22, no. 4, pp. 1171–1188 (2021) doi: 10.1007/s11119-020-09777-5
2. Tammina, S. Transfer learning using VGG-16 with deep convolutional neural network for classifying images. International Journal of Scientific and Research Publications (IJSRP), vol. 9, no. 10, pp. 143–150 (2019) doi: 10.29322/IJSRP.9.10.2019.p9420
3. Guo, R., Sun, X., Chen, K., Zhou, X., Yan, Z., Diao, W., Yan, M. Jmlnet: Joint multi-label learning network for weakly supervised semantic segmentation in aerial images. Remote Sensing, MDPI, vol. 12, no. 19, pp. 3169 (2020) doi: 10.3390/rs12193169
4. Campillo, L. M. G., Torres, R. A. C., López, H. M. D: Percepción remota: Elementos básicos. Kuxulkab', vol. 21, no. 40 (2015) doi: 10.19136/kuxulkab.a21n40.1001
5. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Proceedings of the IEEE international conference on computer vision, pp. 2961–2969 (2017)
6. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440 (2015)
7. Martins, J. A. C., Nogueira, K., Osco, L. P., Gomes, F. D. G., Furuya, D. E. G., Gonçalves, W. N., Junior, J. M., et al.: Semantic segmentation of tree-canopy in urban environment with pixel-wise deep learning. Remote Sensing, vol. 13, no. 16, pp. 3054 (2021) doi: 10.3390/rs13163054
8. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. In: International conference on machine learning, PMLR, vol. 32, no. 1, pp. 647–655 (2014)
9. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, vol. 28 (2015)
10. Lateef, F., Ruichek, Y.: Survey on semantic segmentation using deep learning techniques. Neurocomputing, vol. 338, pp. 321–348 (2019) doi: 10.1016/j.neucom.2019.02.003
11. Audebert, N., Saux, B. L., Lefèvre, S.: Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. Asian conferen-

ce on computer vision, Springer, Cham, vol. 10111, pp. 180–196 (2016) doi: 10.1007/978-3-319-54181-5_12
12. Butler, M. J. A., Mouchot, M. C., Barale, V., LeBlanc, C.: Aplicación de la tecnología de percepción remota a las pesquerías marinas: manual introductorio. Organización de las Naciones Unidas para la Agricultura y la Alimentación, no. 639.2028 BUTa (1990)
13. Badrinarayanan, V., Kendall, A., Cipolla, R.: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 12, pp. 2481–2495 (2017)
14. Krizhevsky, A., Sutskever, I., Hinton, G. E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. Communications of the ACM, vol. 60, no. 6, pp. 84–90 (2017) doi: 10.1145/3065386
15. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014) doi: 10.48550/arXiv.1409.1556
16. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9 (2015)
17. Lobo-Torres, D., Queiroz-Feitosa, R., Nigri-Happ, P., Cué-laRosa, L. E., Marcato-Junior, J., Martins, J., Liesenberg, V.: Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery. Sensors, vol. 20, no. 2, pp. 563 (2020) doi: 10.3390/s20020563
18. Sydorenko, I.: Image annotation tools for machine learning. Label Your Data (2021)