



UNIVERSIDAD NACIONAL DE COLOMBIA

Aprendizaje Profundo para el Mapeo de Maleza Usando Imágenes Multiespectrales Adquiridas por Drones

Victor Alfonso Arias Vanegas

Universidad Nacional de Colombia
Facultad de ingeniería, Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2022

Aprendizaje Profundo para el Mapeo de Maleza Usando Imágenes Multiespectrales Adquiridas por Drones

Victor Alfonso Arias Vanegas

Trabajo final de maestría presentado como requisito parcial para optar al título de:
Magister en Ingeniería - Ingeniería de Sistemas y Computación

Director:

Ph.D. Fabio Augusto González

Co-Director:

Ph.D. Oscar Julián Perdomo Charry

Línea de Investigación:

Sistemas Inteligentes

Grupo de Investigación:

Machine learning, perception and discovery Lab - MindLab

Universidad Nacional de Colombia

Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial

Bogotá, Colombia

2022

Dado a que en ocasiones nos vemos obligados a comportarnos como las bestias, es recomendable que el príncipe se transforme en zorro y en león; por que el león no sabe protegerse de las trampas ni el zorro protegerse de los lobos. Es pues necesario ser zorro para conocer las trampas y león para ahuyentar a los lobos.

Nicolás Maquiavelo

Agradecimientos

A mi co-tutor Ph.D. Oscar Julián Perdomo Charry, quien fue clave en los momentos más críticos de este trabajo. Su experiencia, ideas y consejos me ayudaron a tomar mejores decisiones en distintas fases experimentales, además de servir de ejemplo a seguir como persona. A mi tutor Ph.D. Fabio Augusto González cuya paciencia, consejos y lecciones durante mi etapa en la maestría no solo me sirvieron para llevar a un buen este proyecto, sino que me hicieron un profesional más íntegro y responsable.

Por último, a mi madre y a mi mujer quien con su apoyo y afecto me hicieron la persona que soy hoy en día y me aconsejaron a terminar este reto que había empezado hace unos años, sin ustedes nada de esto hubiera sido posible.

Resumen

Aprendizaje Profundo para el Mapeo de Maleza Usando Imágenes Multiespectrales Adquiridas por Drones

La maleza o malas hierbas se define como una planta que crece de forma silvestre en un lugar indeseable para la actividad agrícola. Esto es debido a que compite por los recursos limitados disponibles en un sector previamente destinado y acondicionado a la producción de alimentos u otras actividades específicas, disminuyendo su rendimiento. Tradicionalmente los granjeros aplican la escarda o eliminación de malas hierbas con herramientas manuales, haciendo de este un proceso lento y costoso debido a la gran cantidad de mano de obra necesaria.

Con el fin de reducir el número de trabajadores en la labor, agentes químicos de acción selectiva son usados directamente sobre el cultivo para matar la planta invasora, sin embargo, en grandes extensiones de terreno es difícil conocer previamente la distribución espacial de la maleza, por lo que la aplicación del agente se hace de manera uniforme en toda la plantación, llevando a un mayor desperdicio del producto y por ende un incremento en los costos.

En este documento se propone una estrategia para la detección automática de la distribución espacial de la maleza en un terreno cultivado usando algoritmos de aprendizaje profundo (DL) en imágenes multiespectrales. Para probar el desempeño de la estrategia se utilizó una base de datos de imágenes recolectada por un vehículo aéreo no tripulado (VANT).

Las bases de datos empleadas proporcionan las imágenes multiespectrales y su respectiva máscara, esta última representa la información semántica de cada uno de los píxeles de la imagen, la información semántica se constituye a partir de tres colores cada uno de ellos pertenecientes a una clase de interés: el rojo representa la maleza, el verde representa el cultivo y el negro representa el fondo o todo aquello que no es vegetal en el mapa.

Adicionalmente, el problema se abordó como un problema de segmentación semántica y la estrategia de solución fue un algoritmo de DL. Al aplicar la solución a las imágenes se evidencia una mejora en las diferentes métricas usadas en la literatura para estas bases de datos tales como el AUC y el F1-score, además se evidencia excelentes resultados en las máscaras predichas para los datos de prueba.

Por último, se analiza el aporte de los diferentes canales multiespectrales y de técnicas clásicas de procesamiento de imágenes a las métricas del modelo, además de la capaci-

dad de este por generar buenas representaciones semánticas del terreno captado por el sensor.

Palabras clave: Mapeo de Maleza, Segmentación Semántica, imágenes Multiespectrales, Aprendizaje Profundo, Vehículo Aéreo No Tripulado , Clasificación Por Píxeles, Aprendizaje Automático En Producción, Redes Neuronales Convolucionales.

Abstract

Deep Learning for Weed Mapping Using Multispectral Imagery Acquired by Drones

A weed is defined as a plant that grows wild in a place undesirable for agricultural crops. This is because it competes for the limited resources available in a sector previously destined and conditioned for food production or other specific activities, decreasing its yield. Traditionally farmers apply weeding or weed removal with hand tools, making this a slow and costly process due to the large amount of labor required.

In order to reduce the number of workers involved, selective action chemical agents are used directly on the crop to kill the invasive plant, however, in large extensions of land it is difficult to know the spatial distribution of the weeds beforehand, so the application of the agent is done uniformly throughout the plantation, leading to a greater waste of the product and therefore an increase in costs.

This thesis presents a strategy for automatic detection of the spatial distribution of weeds in a cultivated field using deep learning (DL) algorithms on multispectral images is proposed. An image database collected by an unmanned aerial vehicle (UAV) was used to test the performance of the strategy. The databases used provide the multispectral images and their respective mask, the latter represents the semantic information of each of the pixels of the image, the semantic information is represented using three colors, each one belonging to a class of interest: red represents the weeds, green represents the crop and black represents the background or everything that is not vegetation on the map.

Additionally, the problem was approached as a semantic segmentation problem and the solution strategy was a DL algorithm. By applying the solution to the images, an improvement in the different metrics used in the literature for these databases such as AUC and F1-score is evidenced, in addition to excellent results in the predicted masks for the test data. Finally, the contribution of the different multispectral channels and classical image preprocessing techniques to the model metrics is analyzed, as well as the model's ability to generate good semantic representations of the terrain captured by the sensor.

Keywords: Weed Mapping, Semantic Segmentation, Multispectral Images, Deep Learning, Unmanned Aerial Vehicle, Pixel-Wise Classification, Machine Learning Engineering for Production, Convolutional Neural Networks.

Esta tesis de maestría se sustentó el 16 de Septiembre de 2022 a las 10:00 am, y fue evaluada por los siguientes jurados:

César Augusto Pedraza Bonilla (Ph.D.)
Universidad Nacional de Colombia

Ángel Alfonso Cruz Roa (Ph.D.)
Universidad de los Llanos

Contenido

Agradecimientos	iv
Resumen	v
Resumen	vii
Lista de Figuras	xi
Lista de Tablas	xii
1 Introducción	1
1.1 Definición del Problema	3
1.2 Objetivos Principales y Específicos	3
1.2.1 Objetivo General	3
1.2.2 Objetivos Específicos	4
1.3 Estructura de la Tesis	4
2 Trabajos Relacionados	5
3 Selección y Preprocesamiento de Bases de Datos de Imágenes multiespectrales	12
3.1 Introducción	12
3.2 Base de Datos Seleccionada	13
3.3 Ingesta y Disponibilización de la Base de Datos	16
4 Etapa Experimental y Resultados	20
4.1 Introducción	20
4.2 Modelos de Línea base: SegNet, UNet y DriUNet	21
4.3 Arquitectura Propuesta	21
4.4 Diseño Experimental	24
4.5 Búsqueda de Hiperparámetros y Ajuste del Umbral	26
4.6 Estrategia de Aumento de Datos	27
4.7 Evaluación de Resultados	28
4.7.1 Análisis Cuantitativo	29
4.7.2 Análisis Pre-Productivo: Mapa Ortomosaico de Prueba RedEdge-MX	30

5 Conclusiones y trabajos futuros **37**

Bibliografía **39**

Lista de Figuras

1-1	Plataformas Robóticas para Mapeo de Maleza	2
2-1	Ejemplo de Imagen Multiespectral	5
3-1	Ejemplo de Entrenamiento	13
3-2	Diagrama de la Tarea de Segmentación Semántica	14
3-3	Estructura de las Carpetas de la Base de Datos	15
3-4	Diagrama del Flujo de los Datos	16
3-5	Aumentación de los Datos por Recorte	17
3-6	Gráfico Comparativo TPU vs. GPU	19
4-1	Diagrama de la Arquitectura de DL	22
4-2	Búsqueda de Hiperparámetros	26
4-3	Ajuste de Umbral de Binarización	27
4-4	Mapa de Maleza Modelo DriUNet	31
4-5	Mapa de Maleza Modelo Propuesto	32
4-6	Mapa de Maleza Modelo SegNet	33
4-7	Mapa de Maleza Modelo UNet	34
4-8	Mapa de Maleza Sección Comparativa 1	34
4-9	Mapa de Maleza Sección Comparativa 2	35
4-10	Mapa de Maleza Sección Comparativa 3	35

Lista de Tablas

2-1	Estado del Arte en Mapeo de Maleza	10
3-1	Resumen Base de Datos WeedMap	16
4-1	Resumen Detallado de la Etapa 1 de la Arquitectura Propuesta	24
4-2	Resultada estrategia de aumentación de datos	28
4-3	Comparación con Resultados Previos	30
4-4	Comparación con Resultados Previos	36

1 Introducción

El uso excesivo de herbicidas puede llevar a una rápida evolución de la resistencia a este producto por parte de la especie vegetal a eliminar. Igualmente, existe un efecto adverso sobre la especie cultivada que se desea proteger, a los hábitats de algunos animales, a la salud de las personas que manipulan el producto e incluso a aquellos que viven cerca del lugar de la aplicación [1]. Anualmente, el costo en el control de una única especie, la *Alopecurus myosuroides* (hierba negra) ascienden a unos 1000 millones de libras esterlinas y las pérdidas del rendimiento anual del trigo es de unas 3.4 millones de toneladas únicamente en Inglaterra, existiendo a nivel mundial unas 253 especies resistentes a los herbicidas [2].

La maleza es un conjunto de plantas que crecen espontáneamente muy cerca de terrenos cultivados. Estas plantas no son deseadas debido a que en todo momento están compitiendo con el cultivo por recursos tales como el agua, la luz, los nutrientes y el espacio, causando efectos adversos en la calidad y en la cantidad de la cosecha [3]. Un agricultor puede tolerar sólo un umbral de pérdidas económicas debido a la maleza, por lo tanto, usa herbicidas para reducir la cantidad de maleza para mantener las pérdidas por debajo de ese umbral al tiempo que se procura evitar daños significativos no deseados a las plantas cultivadas [4].

Sin embargo, la maleza no crece de manera uniforme siendo común una aplicación sobre toda la extensión del cultivo incluso en áreas libres de maleza, empleando como medio de transporte del herbicida un avión o un tractor [5, 6]. El control por precisión de la maleza del inglés *Site Specific Weed Management (SSWM)* es el área de estudio encargada de desarrollar sistemas de gestión de maleza. El SSWM busca disminuir el uso de herbicidas sin comprometer la calidad del cultivo, reduciendo considerablemente la mano de obra y el tiempo requerido para la labor [7, 8].

Para crear una estrategia que logre reducir las pérdidas económicas derivadas de la aplicación indiscriminada del herbicida, es necesario construir un mapa con la localización espacial y la densidad de la maleza en el terreno. Lo anterior se denomina mapeo de maleza o *Weed Mapping*, y puede dar una estimación de donde se necesita una mayor o menor cantidad del producto, o donde realmente no se necesita [9, 10].

La popularización de sistemas robóticos autónomos equipados con sensores capaces de adquirir imágenes que cubren más allá del espectro visual, hacen posible el análisis visual del estado

de un cultivo sin la intervención de un ser humano. Existen dos tipos de plataformas que se emplean normalmente en esta tarea: la primera plataforma es un vehículo terrestre todoterreno de cuatro ruedas que se encuentra equipado con un extenso conjunto de sensores, entre ellos una cámara cuyo lente apunta hacia el suelo. La cámara está rodeada por un elemento protector que obstaculiza la entrada de la luz ambiente en el área de captura, tal como se presenta en la Figura 1-1 [11, 12].



Figure 1-1: Las plataformas terrestres y aéreas son las más populares en la adquisición de imágenes de campos cultivados [11, 12]

La segunda plataforma pertenece al grupo denominado de captura remota. Estos son sistemas de adquisición de imágenes montados sobre satélites, aviones o vehículos aéreos no tripulados (VANTs). Los VANTs son los más populares en esta categoría dado que la tarea a ejecutar requiere de una gran versatilidad, costos bajos y una alta resolución espacial en las imágenes adquiridas [12, 13].

Sin embargo, estas ventajas vienen con limitaciones técnicas asociadas a la adquisición de imágenes a bajas altitudes y la imposibilidad de cubrir el área de estudio en una sola toma. Recientemente, esto ha podido ser solventado uniendo el conjunto de imágenes de todas las secciones del terreno en una única composición homogénea llamada mapa ortomosaico [14]. Una vez obtenido el mapa ortomosaico, la siguiente tarea consiste en desarrollar la metodología para construir el mapa de maleza de la zona.

En adición, las técnicas de aprendizaje profundo (DL) han sido ampliamente utilizadas en el mapeo de maleza con VANTs [15, 16]. Estas técnicas han sentado el estado del arte en esta aplicación, mejorando significativamente las métricas de evaluación y la calidad visual de los mapas resultantes [17].

1.1 Definición del Problema

Aunque las técnicas de DL han mostrado resultados sobresalientes en la construcción de mapas de maleza [18, 19, 20, 12], algunos retos han sido evidenciados. Uno de estos desafíos es el desarrollo y las mejoras para la detección de maleza usando imágenes multispectrales [21]. Estas imágenes brindan mucha más información de elementos vegetales a la que no se puede acceder utilizando el espectro de la luz visible [22]. Sin embargo, en la mayoría de los estudios, el método preferido para adquirir datos fue el uso de una cámara digital montada en un vehículo terrestre para recoger imágenes RGB [23].

En algunos trabajos donde se explora el uso de imágenes multispectrales existen desafíos a la hora de clasificar correctamente los sectores de la imagen con maleza por parte de los modelos y la frecuente confusión entre la maleza y las plantas del cultivo. Esto es causado principalmente por maleza de pequeño tamaño con alta variabilidad en forma, tamaño y apariencia [21]. Por último, trabajar con imágenes multispectrales hace necesario un flujo de datos personalizado que alimente etapas a posteriori dentro del pipeline de entrenamiento sin cuellos de botella que limite el hardware de procesamiento gráfico [24]. Es necesario además el desarrollo de arquitecturas que aprovechen realmente la información extra dada por los canales multispectrales.

Por lo tanto, las siguientes preguntas de investigación son definidas en esta propuesta:

- ¿Cómo entrenar el modelo de manera eficiente haciendo el mejor uso de los recursos computacionales?
- ¿Cómo desarrollar una estrategia usando DL que mejore el rendimiento en el mapeo de maleza con imágenes multispectrales?
- ¿Se pueden aprovechar los canales multispectrales para brindar un mayor rendimiento en modelos basados en DL?
- ¿Cómo se puede evaluar efectivamente los modelos de DL de mapas de maleza que se alimentan de imágenes multispectrales?

1.2 Objetivos Principales y Específicos

1.2.1 Objetivo General

Implementar y evaluar un método de aprendizaje profundo para la generación de mapas de maleza con imágenes multispectrales de diferentes bases de datos públicas.

1.2.2 Objetivos Específicos

- Proponer una estrategia de pre-procesamiento y flujo de datos para imágenes multi-espectrales.
- Proponer una estrategia de DL para mejorar el rendimiento en la creación de mapas de maleza usando imágenes multiespectrales.
- Validar y evaluar sistemáticamente los resultados obtenidos en la generación de mapas de maleza comparado con los algoritmos identificados en la literatura.

1.3 Estructura de la Tesis

Esta tesis está estructurada de la siguiente manera: El primer capítulo presenta la introducción de la tesis, la identificación del problema, los objetivos principales y específicos, así como las aportaciones del trabajo. El segundo capítulo presenta los antecedentes y los trabajos relacionados. El tercer capítulo presenta la estrategia de disponibilidad y preprocesamiento de datos utilizada para servir al modelo las imágenes multiespectrales junto con sus máscaras al tiempo que se reducen los tiempos no productivos del acelerador gráfico. El cuarto capítulo explica la estrategia propuesta junto con la evaluación de la misma comparándola con resultados obtenidos en trabajos previos que usan esta base de datos y modelos de segmentación comúnmente empleados para esta tarea. Por último, el quinto capítulo presenta las conclusiones de la tesis y los trabajos futuros.

2 Trabajos Relacionados

El uso de VANTs es de gran utilidad en el campo del mapeo de maleza dada su versatilidad. Estos vehículos tienen la posibilidad de volar a bajas altitudes en comparación con otros vehículos aéreos, mantenerse estáticos en un punto en el cielo en el caso de los cuadricópteros y capturar imágenes aéreas de muy alta resolución que abarcan un área de terreno considerablemente mayor a un vehículo terrestre.

Adicionalmente, estos vehículos pueden equiparse con cámaras multispectrales que ofrecen un mayor abanico de información que el proporcionado por una cámara digital RGB (ver Figura 2-1). Estas bandas adicionales escapan del espectro de la luz visible que puede percibir el ojo humano, tal como las bandas cercanas al infrarrojo (NIR), pero permite obtener información vegetal que puede ayudar a diferenciar la planta del entorno que la rodea. Estas bandas espectrales son la base de un gran número de índices de vegetación, además de las características de color o textura extraídas a partir de las bandas RGB.

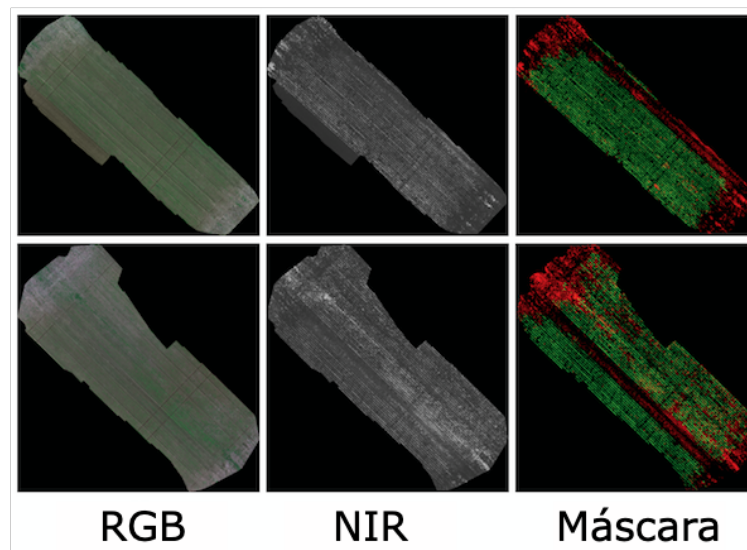


Figure 2-1: Las imágenes multispectrales tienen canales espectrales más allá de la luz visible, brindando información adicional de los elementos vegetales presentes.

Fuente: [12]

Los modelos clásicos de aprendizaje de máquina (ML) han sido usados con gran éxito durante

mucho tiempo en la tarea de clasificación de maleza usando imágenes multispectrales [25]. La estrategia para implementar estos modelos sigue básicamente un conjunto de cinco pasos de manera secuencial: adquisición de las imágenes, preprocesamiento, extracción de características, implementación del modelo, y, por último, una manera de evaluar el desempeño del modelo. Uno de estos pasos, la extracción de características requiere de la intervención de un experto que defina a partir de su conocimiento del área la utilidad de una u otras. Estas características subjetivas extraídas a partir de las imágenes deben ser capaces de discriminar la maleza invasora de la planta cultivada. Esto plantea grandes retos a estos clasificadores clásicos debido a la complejidad y variabilidad de la información, un caso común es el hecho de que la planta cultivada y la maleza pueden ser muy similares [23].

Por otro lado, las estrategias de DL extraen automáticamente las características discriminatorias, sin la intervención de algún experto. Una familia de arquitecturas comúnmente empleadas son las redes neuronales convolucionales (CNN), que procesan datos en forma de arreglos multidimensionales tales como las imágenes. Las CNN han ganado gran popularidad estos últimos años en el campo de la visión por computadora, y sus distintas configuraciones han ganado bastante aceptación por el éxito en distintos desafíos académicos en áreas como el reconocimiento de imágenes, detección de objetos y segmentación semántica [26, 27].

La arquitectura tradicional de una CNN es de la forma imagen/etiqueta, es decir, el modelo recibe una imagen o un conjunto de imágenes y produce una distribución de probabilidad sobre un cierto número de clases o etiquetas pertenecientes a un grupo de objetos de interés que se presumen se podrían encontrar dentro de la imagen [28]. A continuación, se expone una recopilación de trabajos previos siguiendo la taxonomía general de técnicas de detección de maleza basada en DL hecha por Zhang et al. [27] en el año 2021, en esta búsqueda se priorizaron artículos que apliquen clasificación a nivel de píxeles comúnmente conocida como segmentación semántica [29] sobre imágenes multispectrales adquiridas desde VANTs.

El campo de mapeo de maleza ha sido un campo bastante activo en estos últimos años [30, 31], existen numerosos trabajos escritos en el campo del mapeo de maleza usando imágenes multispectrales [32, 33, 31], VANTs [34, 35] y segmentación semántica con DL [36, 37] de manera separada. En el 2017, Sa et al. [19] con su implementación de la *WeedNet* desarrollaron un framework que utiliza de manera conjunta estos tres pilares. Los autores realizaron la adquisición de una base de datos etiquetada manualmente con las clases cultivo (remolacha azucarera) y maleza (de una especie no especificada), a esta base de datos se le varió el nivel de maleza en ciertas áreas del terreno con ayuda de pesticidas y se aplicaron estrategias de preprocesamiento para que cada ejemplo de entrenamiento y de prueba tuvieran los canales NIR, Rojo y NDVI. Esta base de datos alimentó una red SegNet [38], la cual esta basada en las capas para la extracción de características de la VGG16 y fue entrenada con diferentes combinaciones de canales obteniendo todos resultados mayor a 0.8 en F1-score, siendo el

modelo con la combinación los canales NIR+Rojo, el mejor modelo para la clasificación de la maleza.

Es importante en este punto mencionar trabajos que abordan un problema crítico: el tiempo y la mano de obra especializada en el etiquetado manual de este tipo de tareas. Di Cicco et al. [39] también en 2017 generó una base de datos sintética que reproduce distintas especies de malezas y cultivos, etapas de crecimiento, condiciones atmosféricas y tipos de suelos. El resultado son imágenes fotorrealistas de campos agrícolas de las que se obtienen máscaras segmentadas similares a una base de datos real. Sin embargo, únicamente se pueden reproducir canales del espectro visible afectando el desempeño del modelo.

En 2018, San et al. [12] publicó una versión extendida de su trabajo anterior que incluía una metodología para la creación de bases de datos para el mapeo de maleza en alta resolución y con una amplia cobertura de terreno cultivado. Esta metodología fue llevada a cabo nuevamente sobre cultivos de remolacha azucarera (*Beta vulgaris*) usando VANTs con cámaras multiespectrales y a baja altitud, las imágenes suministradas por los drones fueron agrupadas en un único mapa del terreno (mapa orto-mosaico) y expertos etiquetaron el mapa a nivel de píxeles. Finalmente, las imágenes alimentaron una red SegNet en distintas configuraciones de canales, siendo la concatenación de nueve canales (incluyendo canales redundantes) la que alcanzó el área bajo la curva (AUC) más alta [Fondo=0.839, Cultivo=0.863, maleza=0.782]. Los mapas ortomosaicos agregan gran valor al mapeo de maleza, permitiendo a algoritmos de alto nivel planear rutas óptimas para la aplicación de pesticidas.

El trabajo de Huang et al. [28] de 2018 exploró el mapeo de maleza en un campo de arroz en China y demostró que una red FCN [29] con una arquitectura que extrae mapas de características en secciones intermedias de la red reduce el efecto de la pérdida de información espacial mejorando la salida del segmentador. Una continuación de este trabajo por parte del mismo autor y en el mismo año es publicado, en esta ocasión con mejoras tales como el procesamiento multi-escala, nuevas capas convolucionales y modelos gráficos a la salida de la red. Como resultado se obtiene un incremento considerablemente de las métricas [40].

En el 2019, Skovsen et al. [41] publicó un desafío que incluye la segmentación semántica de maleza junto con distintas especies de tréboles, la base de datos consta de imágenes con un tamaño de píxel en el terreno (GSD) de 4-8 píxeles por milímetro (mm). Rist et al. [42] en ese mismo año investigó los efectos de la baja resolución en el mapeo de maleza, para ello se utilizó una base de datos satelital relativamente pequeña. Esta base de datos en su momento amplió la cobertura en comparación con otras disponibles, abarcando un terreno amplio ($30km^2$) pero con una resolución baja con respecto a los VANTs ($500mm/pixel$ de GSD). Cada una de las imágenes de entrenamiento contaba con 4 canales (RGB+NIR) y máscaras binarias etiquetadas manualmente (píxeles blancos corresponden a la maleza), dis-

tintas técnicas de pre-procesamiento se usaron para aumentación de datos (desplazamientos espaciales, rotaciones, acercamientos y alejamientos) y una variante de la U-Net [43] especialmente diseñada para datos adquiridos por sensores montados de manera remota [43] se implementa como modelo segmentador, los mejores resultados son de un 0.535 de intersección sobre la unión (IoU).

En 2019, Adhikari et al. [44] expuso cómo en un campo de arroz las estrategias de control biológico para un tipo de maleza, el mijo común, fallaban. Esto se debe a que su aspecto es similar al de las plantas de arroz, para solucionar el problema se buscó una estrategia de DL que consistía en detectar las líneas de arroz para que un vehículo robótico se desplazara entre las líneas, luego se detectaba la maleza y se extraía automáticamente. Además, el autor en este trabajo introdujo el concepto de gráficos semánticos en el área, una manera de disminuir la carga del etiquetado manual expresando un concepto objetivo en forma de figura para facilitar el aprendizaje. Para este caso en concreto, se trazaron las líneas de arroz reales y se marcaron como maleza todas aquellas plantas de aspecto similar fuera de las trazas de arroz. Un modelo personalizado llamado ES-Net fue propuesto como modelo segmentador, obteniendo una media de la intersección sobre la unión (mIoU) de 0.5178 superando a modelos en la misma tarea como la U-Net [45], DeepLabv3 [46] y FCN-8 [29].

Siguiendo con los trabajos claves en 2019, Brilhador et al. [47] exploró el impacto de técnicas de aumentación de datos en el mapeo de maleza: transformaciones sobre la imagen de entrada y extracción de parches en imágenes de alta resolución, el uso de estas estrategias incrementó un 19.96% el promedio del coeficiente de similitud de Dice (Avg. DSC) con respecto al modelo que no las usa.

El trabajo de Fawakherji et al. [48] en 2020 abordó el mapeo de maleza en dos pasos usando dos CNN en secuencia. El primer paso utilizó una U-Net modificada cuyo objetivo fue separar los píxeles entre dos clases: el suelo y lo vegetal. Después, cada planta se clasificó como cultivo o maleza utilizando una VGG-16 [49]. De forma similar, Abdalla et al. [20] en 2019 implementó la SegNet para separar los elementos vegetales del suelo y otros elementos no deseados, esta máscara resultante hizo las veces de entrada a un clasificador clásico (una SVM en este caso), disminuyendo el número de parámetros entrenables y por ende una menor cantidad de las costosas etiquetas manuales es necesaria.

Asad et al. [50] en 2020 reportó un análisis comparativo en la tarea de mapeo de maleza en un campo de canola con el extractor de la ResNet-50 [51], en este caso ambas se usaron como red base tanto en la SegNet como en la U-Net. Como resultado se obtuvo un 0.6648 de IoU en la clase maleza cuando la SegNet usa de base la ResNet-50 frente a los 0.5930 que se obtienen con la VGG-16.

De forma independiente, Ma et al. [52] en 2019 llegó a conclusiones similares esta vez en un campo de arroz, donde una red SegNet obtiene un mIoU de 0.618 contra un 0.538 y 0.530 de IoU de la FCN y la U-Net respectivamente. Además del riesgo de perder una parte de sus cosechas, los agricultores deben tratar con maleza tóxica, tal es el caso de la especie *Colchicum autumnale* una planta que crece en prados y pastos extensos y que pone en peligro la salud del ganado.

Petrich et al. [53] en su trabajo del 2020 aplicó una estrategia de segmentación semántica con una U-Net modificada para identificar los píxeles en imágenes de drones que pertenecían a la indeseada especie, el modelo se probó sobre una sección de la pradera nunca antes vista y se obtuvo un recall de 0.886 y una precisión de 0.692 en promedio. Por otro lado, Ramírez et al. [54] en ese mismo 2020 usó tres modelos de segmentación semántica (SegNet, UNet y DeepLab v3) para la detección de maleza con imágenes multispectrales. Los resultados de 0.89 y 0.81 en términos de AUC y F1-score promedio del modelo DeepLab v3 mostró su gran potencial de uso en la tarea.

Osorio et al. [21] en el año 2020 presentó una estrategia de detección de maleza novedosa, donde un sensor multispectral se usa para extraer el índice NDVI. Este índice se usó como un filtro que busca eliminar todo lo que no es vegetal en la imagen. A esta máscara de objetos vegetales se le aplicó un extractor de características llamado histograma de gradientes orientados (HOG) [55] que sirve como entrada para un modelo de máquina de soporte vectorial (SVM) que discrimina la clase planta cultivada. La salida del modelo eliminó todas las plantas del índice NDVI dejando únicamente la clase maleza. Aunque los demás algoritmos usados son de DL estos siguieron prácticamente la misma metodología, solo que en estos casos el índice NDVI se usó únicamente para filtrar la clase planta cultivada generada tanto por el modelo YOLO [56] como por el modelo Mask-RCNN [57]. Los resultados fueron de 0.88, 0.96, y 0.96 de F1-score para los modelos HOG+SVM, YOLO y Mask-RCNN respectivamente.

En adición, Zou et al. [58] en 2021 usó mapas ortomosaicos recopilados a partir de imágenes tomadas por drones con una resolución espacial de $5mm/pixel$ en GSD en un campo de cultivo de caléndulas. El mapa ortomosaico fue dividido en parches de 2560×256 píxeles para alimentar una red neuronal convolucional de tipo U-Net. En el artículo se expone una metodología que consiste en extraer primero todo elemento vegetal de la imagen (tanto plantas como maleza) llamado índice ExG. La UNet se encarga de segmentar únicamente las plantas y la diferencia entre el índice ExG y el resultado de la UNet crea el mapa de maleza, los resultados de segmentación de maleza fueron de 0.9340 en IoU.

Otro importante aspecto que afecta la implementación de estos algoritmos en el problema de mapeo de maleza es el crecimiento, es decir, la capacidad del modelo de segmentar correctamente el mismo cultivo en distintas etapas de su desarrollo. Este enfoque espacio-temporal

fue abordado por Lottes et al. [11] en 2020, donde se usa un robot sobre cultivos sembrados en hileras y cuya distancia inter-planta es la misma. Aprovechando esta característica geométrica se entrena un modelo el cual fue alimentado con una secuencia de imágenes tomadas por el robot en movimiento sobre la misma hilera en un rango de tiempo uniforme. Esto implicó que el segmentador pudo además discriminar maleza visualmente similar a la planta cultivada usando la dispersión de estas con respecto a las plantas capturadas en fotogramas anteriores, los resultados de esta estrategia mejoraron hasta en un 18% en mAP.

Bosilj et al. [59] en 2020 aplicó una estrategia de *transfer learning* usando modelos ya entrenados con diferentes bases de datos que contienen distintos tipos de cultivos y de maleza. El resultado obtenido dejó patente de que es posible entrenar un modelo hasta un 80% más rápido, teniendo máscaras parcialmente anotadas y aun así obtener resultados comparables a anotaciones más laboriosas. Transcurriendo todavía el 2020, Huang et al. [18] hizo una comparación entre el método clásico de análisis de imágenes basado en objetos (OBIA) y las nuevas técnicas de DL dejando de manifiesto la superioridad de esta última usando imágenes tomadas por VANTs. Trabajos del mismo año llegan a conclusiones similares [60, 61].

A continuación, En la Tabla **2-1** se resume a grandes rasgos los últimos 5 años en el estado del arte del mapeo de maleza usando estrategias de DL.

Table 2-1: Resumen del estado del arte para la tarea de mapeo de maleza usando DL

Año	Referencia	Modelo	Especie de Maleza	Especie de Cultivo
2017	Di Cicco et al. [39]	SegNet	Capsela bursa-pastoris, galium aparine	Remolacha azucarera
	Sa et al. [19]	FCN	No especificada	Trigo de invierno
2018	San et al. [12]	SegNet	Galinsoga spec., Amaranthus retroflexus, Atriplex spec., Polygonum spec., Gramineae	Remolacha azucarera
	Huang et al. [28]	FCN	No especificado	Arroz
	Huang et al. [40]	FCN+SVM	No especificado	Arroz
2019	Skovsen et al. [41]	FCN-8s	Hierba	Treboles
	Rist et al. [42]	U-Net	Hierba Gamba	No especificada
	Adhikari et al. [42]	ESNet, U-Net, FCN-8s, DeepLab v3, Faster R-CNN, EDNet	Mijo común	Arrozal
	Brilhador et al. [47]	U-Net	No especificada	Zanahoria
	Fawakherji et al. [48]	SegNet, U-Net, Bon-Net, FCN8s	No especificada	Girasol, Zanahoria, Remolacha azucarera
	Abdalla et al. [20] Ma et al. [52]	FCN SegNet, FCN, U-Net	No especificada Sagittaria trifolia	Semillas oleaginosas Arroz
	Asad et al. [50]	U-Net, SegNet	No especificada	Canola
	Petrich et al. [53]	U-Net	Colchicum autumnale	No especificado

Table 2-1 continued from previous page

	Ramirez et al. [54]	SegNet, U-Net, DeepLab v3	Galinsoga spec., Amaranthus retroflexus, Atriplex spec., Polygonum spec., Gramineae	Remolacha azucarera
	Osorio et al. [54]	HOG-SVM, CNN-YOLOv3, Mask R-CNN	No especificado	Lechuga
	Lottes et al. [11]	FCN modificada	Maleza dicotiledóneas, Maleza monocotiledóneas	Remolacha azucarera
	Bosilj et al. [59]	SegNet	No especificada	Remolacha, zanahorias, cebollas
	Huang et al. [18]	FCN	Leptochloa chinensis, Cyperus iria, Digitaria sanguinalis (L), Barnyard Grass	Arroz
	Umamaheswari et al. [60]	SegNet-512, SegNet-256	No especificado	Zanahorias
	Wang et al. [61]	FCN	No especificado	Semillas oleaginosas, remolacha azucarera
2021	Zou et al. [58]	U-Net modificada	Setaria, Asclepias, juncia	Caléndula

En conclusion:

- Existen muy pocos autores usando imágenes multiespectrales y mapas ortomosaicos como entrada a los modelos, debido a la dificultad de encontrar datos públicos.
- La mayoría de modelos usados son los mas comunes en las distintas aplicaciones de segmentación. La VGG-16 como extractor de características es las mas utilizada y la UNET junto con variantes es la red de segmentación mas popular.
- Las estrategias en dos pasos segmentador + clasificador es un enfoque que ha ganado auge en los últimos años.
- Existen muy pocos trabajos que aborden un enfoque temporal, es decir, que logren probar la efectividad de la segmentación a medida que el cultivo cambia en el tiempo.
- Existen estrategias que logran capturar la geometría del cultivo para ayudar a la red a tomar mejores decisiones, pero son pocas.

3 Selección y Preprocesamiento de Bases de Datos de Imágenes multispectrales

3.1 Introducción

Al momento de redactar este documento de tesis se encontraban disponibles tan solo 21 bases de datos públicas que abordan el área de control de maleza [62, 63, 64]. Seis (6) de estas bases de datos contienen anotaciones a nivel de píxeles que incluyen tanto la clase correspondiente a la maleza, como la clase correspondiente a las plantas del cultivo. Estas anotaciones sirven como requisito primario en el entrenamiento de modelos DL que abordan problemas de segmentación semántica multiclase.

De estas seis (6) bases de datos, solo tres (3) corresponden a imágenes aéreas, solo cinco (5) bases de datos contienen imágenes multispectrales y una (1) es generada artificialmente con la ayuda de un motor gráfico. Para nuestra investigación se seleccionaron únicamente las bases de datos que contienen imágenes multispectrales y hayan sido adquiridas desde una plataforma aérea (un cuadricóptero en este caso). Otro requisito fundamental es que la base de datos haya sido usada en diferentes investigaciones para que posibilite la comparación de los resultados reportados en la literatura, con respecto a los obtenidos en la presente investigación.

Finalmente, el mapa de maleza generado deberá potencialmente servir como insumo a un algoritmo generador de trayectorias para, por ejemplo, disponerlo en un futuro en un robot dispensador de herbicida. Por lo tanto, una base de datos de mapas ortomosaicos también es requerida para este propósito. En las secciones siguientes se exponen las siguientes contribuciones:

- Selección de la base de datos de trabajo que cumple con los requisitos planteados en la anterior introducción.
- Metodología para la ingestión y puesta a disposición de una base de datos multispectral que reduzca los tiempos no productivos del hardware acelerador (TPU).

3.2 Base de Datos Seleccionada

Existen una buena variedad de bases de datos para tareas de clasificación en el área de agricultura. Sin embargo, para el mapeo de maleza las bases de datos públicas son limitadas. Aquellas disponibles por lo general son adquiridas únicamente con sensores de luz visible, poseen una baja resolución, el terreno cubierto es pequeño, la cantidad de imágenes es insuficiente o no están compuestas en forma de mapa ortomosaico.

Este último punto es importante dado que al tener una visión del estado del terreno en su conjunto es posible darle un uso efectivo a los resultados en una instancia superior donde sea necesario guiar con precisión hacia su objetivo el herbicida, independientemente de la plataforma usada para almacenarlo y distribuirlo. La Figura 3-1 muestra un ejemplo de un mapa ortomosaico junto con la respectiva máscara con el etiquetado manual.

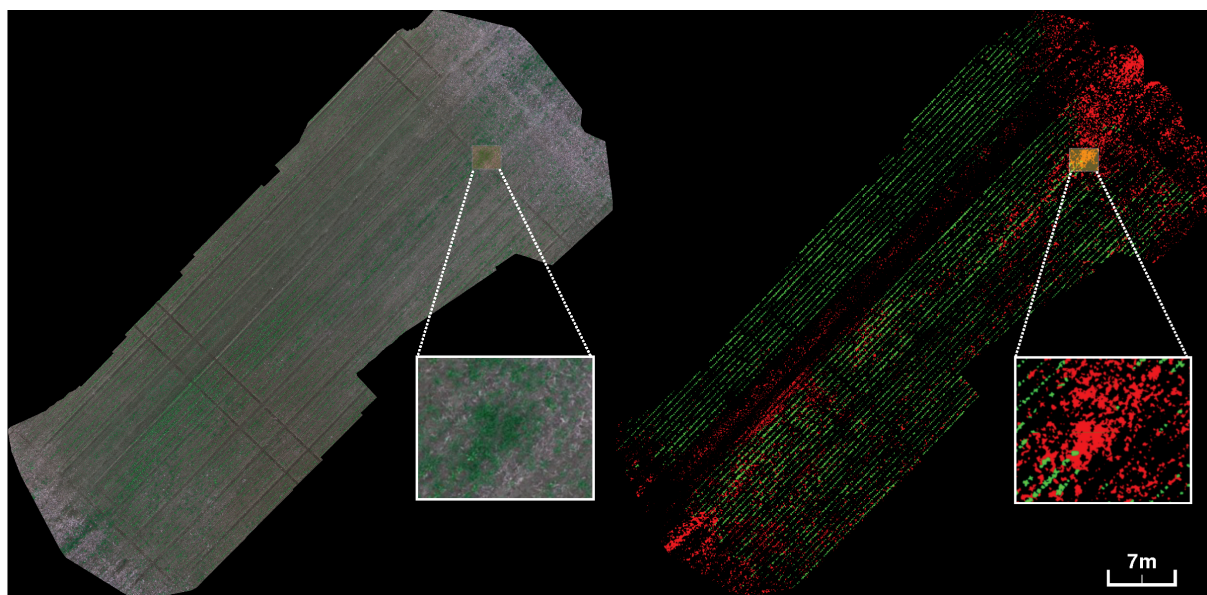


Figure 3-1: Ejemplo de entrenamiento en forma de mapa ortomosaico de la base de datos. A la izquierda, el mapa ortomosaico únicamente con los canales RGB. A la derecha, la máscara resultante del etiquetado por parte del experto, el color rojo indica presencia de maleza en el pixel, el color verde cultivo y el color negro es fondo. **Fuente:** [12]

Antes de pasar a la selección de la base de datos, es importante conocer cómo se relacionan ambas imágenes de la Figura 3-1 y como esta relación es la base de la estrategia planteada. El *weed mapping* puede ser considerado una tarea de clasificación a nivel de píxeles, es decir, el objetivo es asignar una etiqueta categórica dentro de las etiquetas existentes a cada pixel en el mapa ortomosaico [65]. La tarea es comúnmente llamada segmentación semántica y

puede ser formalizada como sigue: Sea una imagen $\mathbf{X} \in \mathbb{R}^{N \times B}$ una matriz compuesta por N vectores de características de longitud B $X_{n=i,j}^b$. Donde N es el número de píxeles y B el número de bandas independientes que componen la imagen multiespectral.

Por lo tanto, el algoritmo de segmentación semántica es una función cuya entrada es la imagen multiespectral \mathbf{X} , y la salida es la máscara predicha por el modelo \mathbf{X}' (ver Figura 3-2). Donde $\mathbf{X}' \in \mathbb{R}^{N \times 3}$ es una matriz compuesta por N vectores etiqueta de longitud 3 $X_{n=i,j}'^{(R,G,B)} = \{X_n^R, X_n^G, X_n^B\}$, donde cada pixel/vector $X_{n=i,j}'^{(R,G,B)}$ puede tener una combinación de valores de los canales rojo, verde y azul. Esta combinación representa la probabilidad de que un pixel de entrada $X_{n=i,j}^b$ pertenezca a la clase $l \in L = \{maleza, cultivo, fondo\}$ [66, 67].

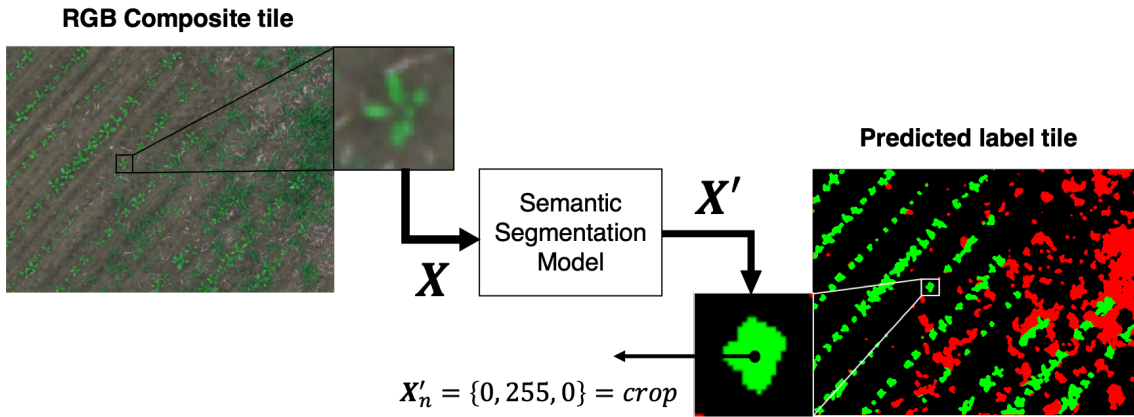


Figure 3-2: La tarea de segmentación semántica consiste en generar etiquetas interpretables por el ser humano para cada píxel de una imagen determinada. **Fuente:** [12]

La base de datos de *WeedMap* [12] brinda un conjunto de mapas ortomosaicos que cubren un terreno de $16.500m^2$. La base de datos fue adquirida en dos campañas:

- Eschikon (Suiza): Usando la plataforma aérea Mavic Pro y la cámara multiespectral Sequoia Parrot[®] de cuatro (4) canales espectrales (R, G, RE, NIR).
- Rheinbach (Alemania): Usando la plataforma Inspire 2 montando la cámara MicaSense[®] RedEdge-MX de cinco canales (R, G, B, RE, NIR), ambas plataformas aéreas pertenecen al fabricante DJI[®].

Las tomas se realizaron manteniendo el vuelo a 10 metros sobre un cultivo de remolacha azucarera. Sin embargo, en una inspección visual se notó que un mapa ortomosaico de la primera campaña tiene errores en la corrección radiométrica, imposibilitando su uso. Es por ello que en esta investigación se usó los datos de la segunda campaña, donde el sensor capta hasta cinco bandas espectrales y es capaz de entregar índices de vegetación sin ningún

preprocesamiento adicional. Cada uno de estos mapas es etiquetado a nivel de píxeles manualmente por expertos en el área, donde cada pixel de la máscara puede pertenecer a una de tres clases: maleza, cultivo, fondo.

Las máscaras se configuraron de la siguiente manera: se asignaron a los píxeles en donde se encontró maleza, el color rojo, los píxeles donde se encontró plantas del cultivo, el color verde y finalmente negro para el resto. La estructura de la base de datos entregada por el autor consta de cinco carpetas, una por cada mapa, manteniendo separadas las máscaras de las imágenes multispectrales.

Con el fin de simplificar la gestión de los datos de entrenamiento en los experimentos con modelos de DL, Sa et al. [12] dividió los mapas ortomosaicos en secciones de 480×360 píxeles de ancho por alto. Cada una estas secciones o *tiles* contienen ocho (8) imágenes en formato *.PNG*: seis (6) imágenes de un (1) canal (R, G, B, NIR, RE, NDVI) y dos (2) imágenes de tres (3) canales (RGB, CIR).

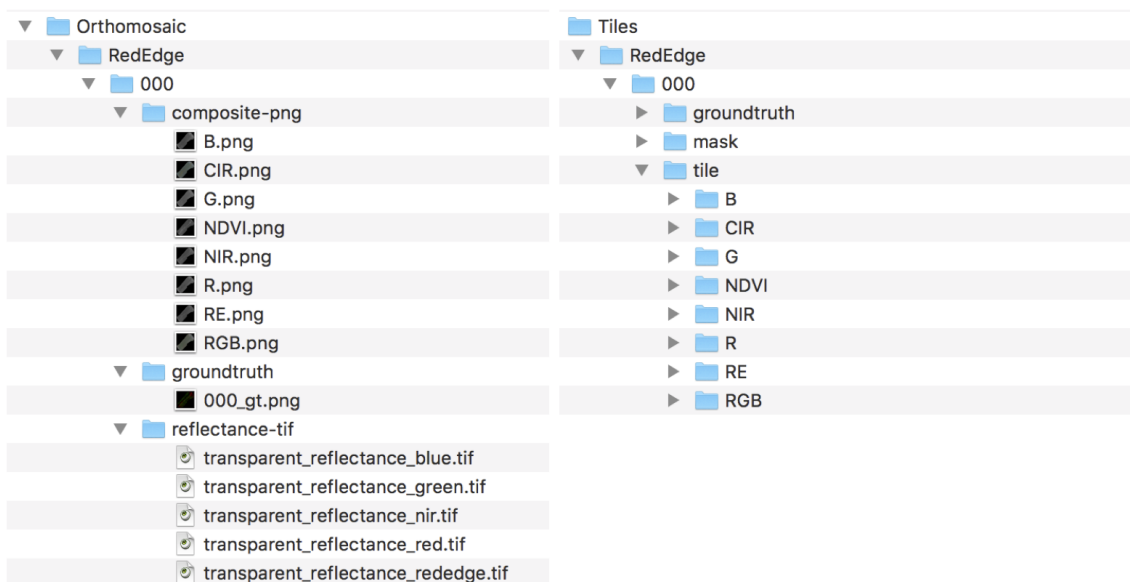


Figure 3-3: Estructura de las carpetas en la base de datos **WeedMap**. los canales R, G, B y NIR se definen como canales redundantes, esto es debido a que las imágenes RGB y CIR son combinaciones lineales de estos. **Fuente:** [12]

La Figura 3-3 muestra la estructura de las carpetas de la base de datos, Un total de 403 secciones efectivas para entrenamiento y 94 para prueba son extraídas de los mapas. Secciones efectivas se refiere a imágenes que contienen en sus máscaras al menos un pixel con un valor diferente al negro $[0, 0, 0]$ (perteneciente a la clase *fondo*). La Tabla 3-1 resume la información referente a la base de datos.

Table 3-1: Resumen de la base de datos **WeedMap** usada. **Fuente:** [12]

Mapa	Resolución (Ancho/Alto)	Area cubierta(<i>ha</i>)	Secciones efectivas (480 × 360)	Partición
000	5995 × 5854	0.312	107	Entrenamiento
001	4867 × 5574	0.1108	90	Entrenamiento
002	6403 × 6405	0.2096	145	Entrenamiento
003	5470 × 5995	0.1303	94	Prueba
004	4319 × 4506	0.1307	61	Entrenamiento
Total	-	-	497	-

3.3 Ingesta y Disponibilización de la Base de Datos

En esta sección se explica el procedimiento para poner a disposición y optimizar el flujo de datos hacia el modelo a entrenar. La Figura 3-4 es un diagrama que resume el enfoque utilizado. El flujo inicia con un archivo de configuración que almacena los requerimientos del usuario, en él se encuentran tres parámetros principales: el primer parámetro es el protocolo, concepto introducido previamente en un trabajo realizado por Ramírez et al. [54] en 2020, donde además de la separación original de los mapas en entrenamiento y prueba (ver Tabla 3-1) se definió un nuevo "protocolo experimental" de división que reducía el número de mapas de entrenamiento de cuatro (4) a tres (3) y se evaluaba el impacto de este cambio en las métricas. El protocolo por defecto es la distribución original hecha por Sa et al. [12].

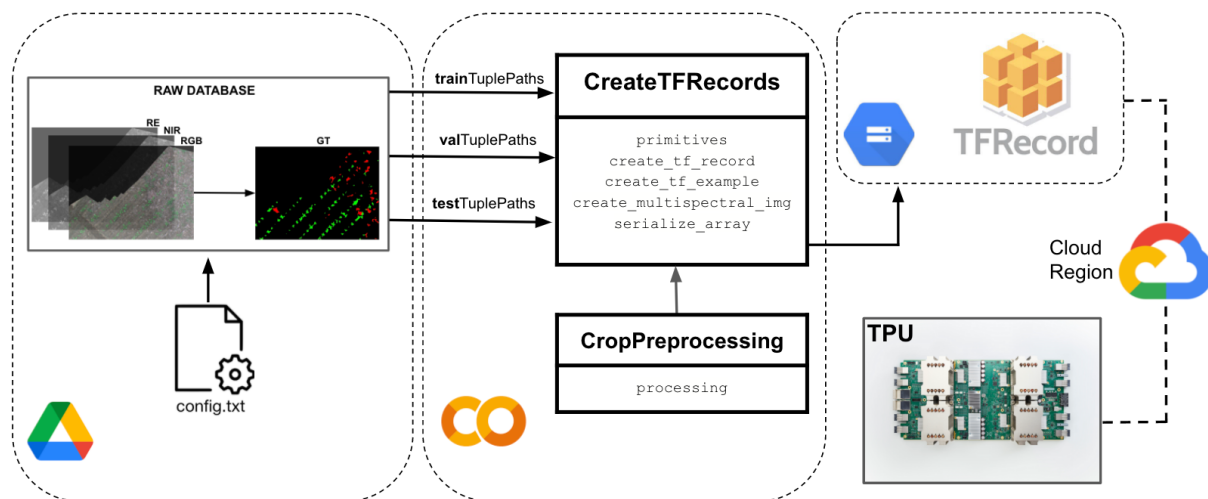


Figure 3-4: Procedimiento de puesta a disposición de datos para reducir los tiempo no productivos de la TPU.

El segundo parámetro es el patrón, es decir, la combinación de canales que conforman la imagen multiespectral que alimentara el modelo. Este conjunto es seleccionado de las ocho (8) imágenes independientes disponibles en la base de datos, incluyendo aquellas con canales redundantes (ver Figura 3-3). En el archivo de configuración se encuentran disponibles los patrones utilizados en trabajos previos, así como patrones adicionales creados para la presente investigación. El último parámetro principal es el número de canales, este parámetro contiene el número total de canales para un patrón dado. Otros parámetros adicionales como el tamaño de la partición de validación, el directorio de las secciones originales, y la carpeta donde se almacenarán los TFRecords también se encuentran en el archivo de configuración.

Una vez se definen estos parámetros y se ejecuta la herramienta, se crean listas con las rutas a las imágenes requeridas en un formato de pares: *imagen multiespectral/máscara multiclase*, siendo cada uno de estos pares un ejemplo de entrenamiento. El tamaño de las listas creadas depende del número de ejemplos asignados a cada una de las tres (3) particiones (entrenamiento, validación y prueba). Los pares se cargan en memoria volátil en subgrupos de cinco (5), esto se debe a que cada archivo en el formato TFRecord contendrá (5) ejemplos de entrenamiento.

Antes de crear los archivos TFRecord, la herramienta utiliza una funcionalidad adicional (CropPreprocessing). Esta funcionalidad extrae de las secciones de 480×360 píxeles, cinco subsecciones de 224×224 que aumentan el número de datos de entrenamiento, al tiempo que mantienen la resolución original (ver Figura 3-5). Esto se aplica tanto a la imagen multiespectral como a la máscara.

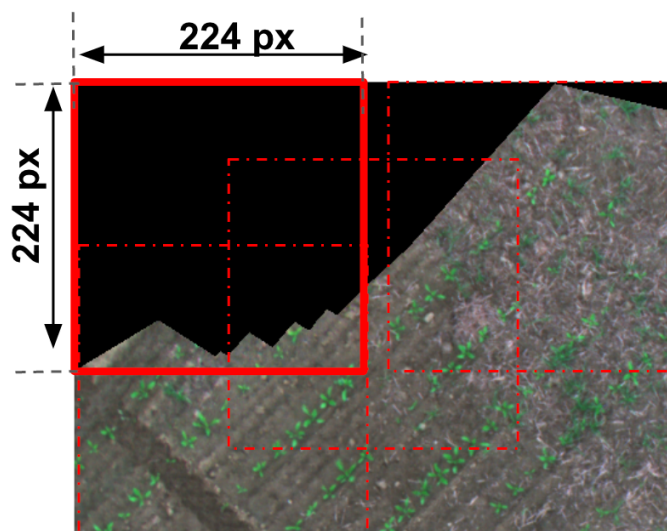


Figure 3-5: Estrategia que extrae subsecciones de 224×224 píxeles de las cuatro (4) esquinas y el centro de la sección original.

Como se mencionó anteriormente, en esta investigación se utiliza el formato TFRecord para consumir datos en los procesos de entrenamiento, validación y prueba. El formato TFRecord es el formato de almacenamiento binario propio de Tensorflow, entre cuyas ventajas está el ser auto contenidos, un menor espacio de almacenamiento en comparación con los datos originales y operaciones de lectura/escritura en paralelo (útiles al utilizar dispositivos de procesamiento como GPUs y TPUs). Es importante aclarar en este punto que el hecho de dividir los archivos en grupos de cinco (5) ejemplos de entrenamiento es por conveniencia, y en nada afecta a la manera de proveer los datos al modelo en el proceso de entrenamiento.

Los archivos TFrecords son almacenados en la herramienta de almacenamiento de la nube de Google (GCS), la razón para hacer esto es asegurar que los datos de entrenamiento se encuentren en la misma región que el hardware acelerador que dispone Google Colab, en este caso una unidad de procesamiento tensorial (TPU) en su versión 2. Dada la rapidez de esta unidad el reto suele ser alimentarlas con datos lo suficientemente rápido como para mantenerlas ocupadas.

En este punto de la investigación nos planteamos las siguientes preguntas: ¿Por qué utilizar una TPU y no una GPU en Colab?, ¿Qué ventajas tienen las TPUs?, ¿Qué limitaciones impone el uso de TPUs en el formato de los datos de entrenamiento?. En principio cualquier dispositivo de cómputo puede ejecutar operaciones matriciales propias de modelos de DL. Sin embargo, las unidades de procesamiento computacional (CPU) realizan estas operaciones secuencialmente en cada uno de sus núcleos. Las CPU modernas de consumo masivo pueden alcanzar hasta 18 núcleos de procesamiento, esto contrasta con los más de 10496 núcleos de procesamiento que posee la RTX-3090TI, la GPU de consumo masivo más potente del mercado. Las GPUs, por tanto, son unidades que ejecutan operaciones matriciales de manera muy eficiente dado el número de núcleos en paralelo.

Las TPUs, por otro lado, son circuitos integrados de aplicación específica (ASIC) ¹ cuya arquitectura física encapsula la esencia del cálculo de redes neuronales, es decir, secuencias de multiplicadores y acumuladores de matrices que van descendiendo en tamaño a medida que se acercan a la salida del modelo. La segunda generación de TPUs (disponibles en Colab) alcanza los 180 TeraFlops en operaciones de punto flotante y 64 GB de memoria de gran ancho de banda (HBM) en una sola placa. En Jangamreddy [68] se usaron benchmarks con modelos de DL comúnmente utilizados y se demostró la reducción de los tiempos de entrenamiento con respecto a la NVIDIA V100 (ver Figura **3-6**).

Una vez vistas todas las ventajas de las TPUs es importante conocer algunas restricciones que imponen su uso, una de estas restricciones es la localización, el usuario en todo momento

¹<https://cloud.google.com/blog/products/ai-machine-learning/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu>

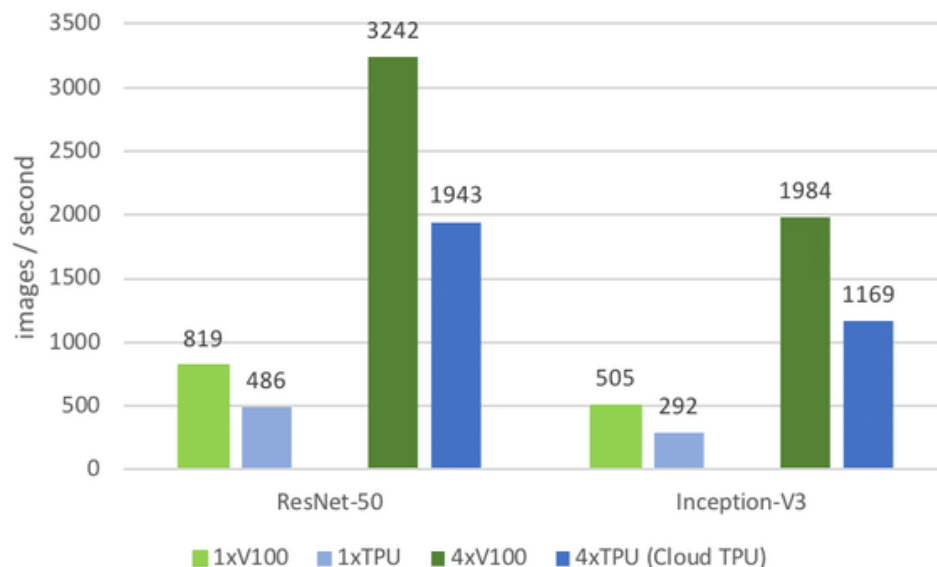


Figure 3-6: Gráfico que muestra la reducción en los tiempos de entrenamiento en dos modelos de DL usando diferentes configuraciones de TPUs y GPUs [68].

debe asegurar que la base de datos de entrenamiento se encuentre en la misma región que la TPU. También que la estructura de la base de datos permita acceder a la información en paralelo y no acceda a la memoria de la TPU de manera dispersa (requisito que aseguramos al tener la base de datos en el formato TFRecord). Finalmente, el usuario no debe utilizar operaciones personalizadas en el bucle de entrenamiento, una vez nuestro pipeline cumpla todas estas reglas es posible sacar el máximo provecho de este tipo de hardware.

4 Etapa Experimental y Resultados

4.1 Introducción

Hoy en día, se estima que en el mundo las cosechas pierden entre un 20% y un 80% de su productividad debido a enfermedades, pestes y maleza [69]. El agente de control estándar para la maleza son los herbicidas químicos, llegando a ser dos tercios del total de químicos aplicados en los campos agrícolas de los Estados Unidos [70]. Sin embargo, el uso indiscriminado de estos químicos está provocando una rápida evolución de la maleza resistente a los herbicidas, la contaminación del medio ambiente y la afección en la salud humana [71].

Por otro lado, las imágenes multiespectrales de grandes superficies cultivadas adquiridas por medio de VANTs y el desarrollo de algoritmos de inteligencia artificial para extraer automáticamente información de interés, es esencial en el tratamiento focalizado de la maleza en los cultivos. Esto hará que el herbicida sea usado únicamente donde sea necesario debido al enfoque de localización espacial de estos desarrollos, reduciendo significativamente los impactos mencionados anteriormente derivados de su uso. Las secciones que se verán a continuación generan las siguientes contribuciones:

- Una estrategia de entrenamiento distribuido para la búsqueda de hiper-parámetros.
- Una estrategia para la búsqueda de umbral óptimo para modelos de segmentación semántica.
- Una estrategia de aumentación de datos del conjunto de entrenamiento para la evaluación del impacto en las métricas de desempeño.
- Un análisis comparativo de nuestro modelo propuesto con respecto a modelos reportados en el estado del arte.

4.2 Modelos de Línea base: SegNet, UNet y DriUNet

Tal y como se demostro en el capitulo 2 la SegNet y la UNet con el backbone de la VGG-16 son los modelos mas utilizados en el problema de mapeo de maleza. Adicionalmente, se agrega el trabajo de Maninis et al. [72] como linea base, ya que la hipotesis es que la propuesta no empeora los resultados de la DriUNet usada de manera solitaria.

La DriUNet es una arquitectura de DL para la tarea de segmentación binaria desarrollada para el campo de la imagenología biomédica, específicamente en imágenes de fondo de ojo. La particularidad de esta red radica en poder segmentar al mismo tiempo el disco óptico y los vasos sanguíneos, sobrepasando el desempeño de un grupo de expertos al etiquetar la misma base de datos. La red está inspirada en la VGG-16 excepto que se mantiene únicamente los bloques convolucionales encargadas de extraer las características, estos cinco (5) bloques usan una función de activación ReLU y una capa maxpooling entre bloques. De cada uno de estos bloques se extrae una capa convolucional de 16 canales, a estas capas se les deben ajustar las dimensiones en caso de no coincidir con las dimensiones de la imagen de entrada (224×224). Estas capas convolucionales son la base de las *capas especializadas*, estas capas consisten en concatenar un grupo de cuatro (4) de las cinco (5) capas de 16 canales. Se crean entonces dos (2) capas especializadas, la primera tiene como objetivo segmentar los vasos sanguíneos y está constituida por las cuatro (4) capas provenientes de los primeros bloques de la VGG-16, es decir, de la información más fina a la más compactada, siendo la salida de esta capa especializada la capa más cercana a la entrada de la red. la segunda capa especializada tiene como fin segmentar el disco óptico, por lo tanto, su composición es inversa a la capa especializada anterior, concatenando las cuatro (4) capas más cercanas a la salida. Finalmente, para cada una de estas capas especializadas de 64 canales, se adiciona una capa de salida de un (1) canal y una función de activación sigmoideal, obteniendo la salida binaria.

4.3 Arquitectura Propuesta

La Figura 4-1 muestra la arquitectura general de la red de mapeo de maleza diseñada e implementada en este trabajo. La red está basada en la DriUNet [72] pero se ha modificado para generar una salida multiclase.

La arquitectura se compone de dos etapas: La **Etapa 1** consiste de un AutoEncoder (AE), cuyo objetivo es generar automáticamente una representación de tres (3) canales extraída de una capa intermedia de esta red (*Upsamp3 + Conv8*), la entrada del AE es de tamaño 224×224 píxeles. En este punto es importante aclarar la importancia de esta etapa, del estado del arte, se puede concluir que de los trabajos que aportan a la mejora de los mapas de maleza hay dos enfoques principales. Por un lado, están los trabajos que agregan infor-

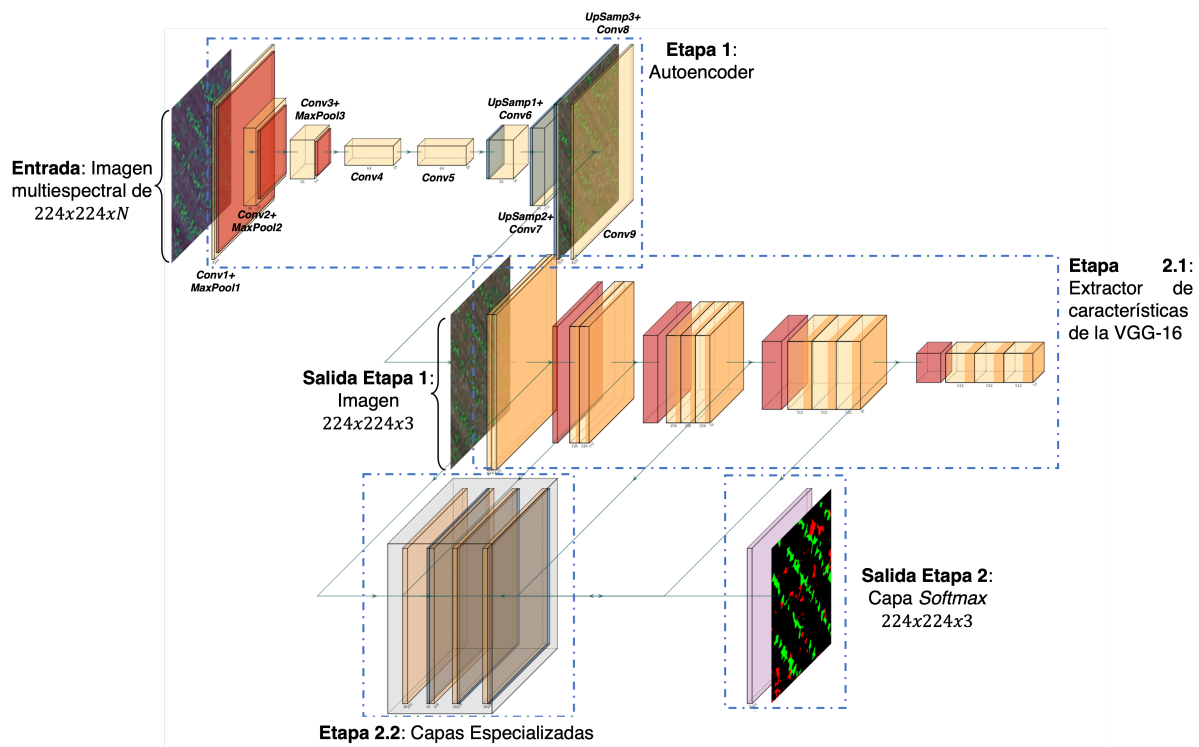


Figure 4-1: Diagrama de arquitectura propuesta con dos CNNs entrenadas en simultáneo para generación de mapa de maleza.

mación adicional al modelo para ayudarlo a tomar una mejor decisión, por ejemplo, indicar donde se encuentran las líneas del cultivo, o almacenar información temporal de imágenes tomadas en los tiempos $t - 1, t - 2, t - 3, \dots, t - N$ a partir de la toma actual. El otro enfoque consiste en dividir el problema en partes, es decir, primero segmentar lo vegetal y a partir de las secciones vegetales usar un clasificador que discrimine el cultivo de la maleza. En el caso de este trabajo, el autoencoder como aporte externo, busca generar una representación de tres canales (3) de la imagen multiespectral independientemente del número de canales de entrada, esta representación debe maximizar la tarea del segmentador al brindarle únicamente la información de utilidad de los canales adicionales al RGB. Para ello es necesario que el entrenamiento de la arquitectura en dos etapas sea de tipo *end-to-end*. La salida de la **Etapa 1** es una representación de la imagen multiespectral con tamaño $224 \times 224 \times 3$ que sirve como entrada para la CNN de la **Etapa 2**.

La estructura del AE está compuesta de nueve (9) capas convolucionales, el primer bloque recibe imágenes de $224 \times 224 \times N$ seguidas por tres sub-bloques que contienen capas convolucionales con 3, 16 y 32 filtros respectivamente, para cada sub-bloque hay un tamaño de kernel de 3×3 y una capa maxpooling a la salida, esta capa de maxpooling cuenta con una ventana de 2×2 . El segundo bloque consiste de dos (2) capas convolucionales idénticas con 64 filtros y un kernel de 3×3 . El tercer bloque está compuesto por tres sub-bloques idénticos al conjunto del primer bloque, pero en lugar de una capa maxpooling a la salida hay una capa de *upsampling* a la entrada de cada sub-bloque y el orden en el número de filtros está invertido (32, 16 y 3), siendo el último de estos sub-bloques la entrada a la siguiente etapa (con unas dimensiones espaciales de $224 \times 224 \times 3$). Por último una capa convolucional con dimensiones idénticas a la imagen de entrada se encuentra a la salida del AE, lo anterior se presenta en detalle en la Tabla 4-1.

La **Etapa 2** es una DriuNet [72] fue modificada para usar únicamente la capa especializada de *compactada a fina*, y una salida multiclase. Esta DriuNet modificada recibe a la entrada una imagen de $224 \times 224 \times 3$ empleando como red base la VGG-16 [73] para realizar la extracción de características. De cada uno de los bloques del extractor se obtiene una capa convolucional con dieciséis (16) filtros de características y un *upsampling* de ser necesario en caso de no tener las mismas dimensiones en alto/ancho de la imagen de entrada, como se presenta en la figura 4-1. Como se mencionó anteriormente, Estas capas son concatenadas en un solo bloque llamado *capas especializadas*, y la salida de este bloque es recibida por una última capa convolucional, que a diferencia de todas las capas en la DriuNet tiene una función de activación *softmax*, el número de filtros de esta capa es igual al número de clases de interés para el problema en cuestión, tres (3) en nuestro caso $\{maleza, cultivo, fondo\}$. En total, las dos etapas contienen aproximadamente un total de 14.9 millones de parámetros, los cuales son valores cercanos, valores muy similares a la SegNet y DenseNet169, pero mucho

Table 4-1: Resumen detallado de la **Etapa 1** de la arquitectura propuesta

Etapa	Capa	Tamaño	Canales	Tamaño de Kernel	Tamaño de Ventana
AE	Entrada	224x224	n	-	2x2
	Conv_1	224x224	3	3x3	-
	MaxPool_1	112x112	-	-	2x2
	Conv_2	112x112	16	3x3	-
	MaxPool_2	56x56	-	-	2x2
	Conv_3	56x56	32	3x3	-
	MaxPool_3	28x28	-	-	2x2
	Conv_4	28x28	64	3x3	-
	Conv_5	28x28	64	3x3	-
	UpSamp_1	56x56	-	-	2x2
	Conv_6	56x56	32	3x3	-
	UpSamp_2	112x112	-	-	2x2
	Conv_7	112x112	16	3x3	-
	UpSamp_3	224x224	-	-	2x2
	Conv_8	224x224	3	3x3	-
Conv_9	224x224	n	3x3	-	

más bajos que redes como xception, ResNet, Inception¹.

4.4 Diseño Experimental

El modelo está dividido en dos etapas entrenadas de manera simultánea, como se describió en la sección 4.3. La primera etapa es un autoencoder, el objetivo de esta etapa es obtener una representación de tres (3) canales a partir de una imagen multiespectral, haciendo las veces de reductor de dimensionalidad. Lo anterior se realiza con el fin de cargar los pesos pre-entrenados de la VGG-16 para el desafío ImageNet, al tiempo que disminuye la complejidad de los datos de entrada a la segunda etapa, la red DriuNet modificada con el segmentador multiclase.

El diseño experimental se divide en tres (3) partes: la primera parte hace uso de la arquitectura propuesta con una imagen de entrada de cinco (5) canales espectrales (R , G , B , NIR , RE). Durante todos los experimentos se usa esta misma entrada, que fue seleccionada a criterio del autor debido a que se busca analizar imágenes multiespectrales. Esta primera parte consiste en una búsqueda de hiperparámetros, siguiendo los siguientes criterios: la tasa de aprendizaje se fija usando una secuencia de valores entre $1e^{-7}$ y $1e^{-3}$ en una escala

¹<https://keras.io/api/applications/>

logarítmica, y 20 muestras tomadas aleatoriamente usando la técnica de *random search* [74]. El tamaño del lote o en inglés *batch size* se fija variando su valor entre $2e^3$ y $2e^5$, el total de épocas por entrenamiento se fijó en 200 épocas para los experimentos exploratorios debido al tiempo computacional de los experimentos, la métrica de entrenamiento es el coeficiente de Dice, y se definió un criterio de 10 épocas para la detención temprana, en caso de no obtener mejoras en las métricas de desempeño durante el entrenamiento. El objetivo de la búsqueda de hiperparámetros es disminuir la pérdida en la partición de validación, la pérdida es constituida por el *error medio cuadrático* para el autoencoder y la *entropía cruzada categórica* para el caso de la DriuNet modificada. Una vez obtenido el modelo con la pérdida más baja, se realiza una búsqueda del umbral aplicado a la salida de la capa convolucional con *softmax*.

La segunda parte del entrenamiento busca medir el impacto de la aumentación en los datos de entrada. Para ello se utiliza la estrategia desarrollada en Brillhador et al. [47], realizando un número de experimentos igual al total de estrategias de aumentación utilizadas y analizando qué combinación resultante da mejores resultados. Para nuestro caso en específico, se utilizan las siguientes técnicas de aumentación de datos: giro de 180° en horizontal, giro de 180° en vertical, rotación (de 0° a 90°), desplazamiento a lo ancho (rango de 1/10 del ancho), desplazamiento a lo alto (rango de 1/10 del alto), Zoom (-0.2 a 0.2).

En la tercera parte se entrenan otros modelos del estado del arte con los mismos hiperparámetros como (*SegNet*, *DriuNet*), donde se evalúan en la partición de prueba usando métricas tales como el *área bajo la curva* (AUC), *F1-score* y el coeficiente de Dice. Los resultados son comparados con nuestro modelo propuesto usando diferente número de canales en las imágenes de entrada. Posteriormente, se realiza un análisis cualitativo comparando respuestas de diferentes modelos a secciones de imágenes en donde sea un desafío realizar la segmentación.

Todos los experimentos aquí propuestos fueron realizados usando una TPU v2 con ocho (8) núcleos, 64GB de memoria de alto ancho de banda (*HBM*) y 180 teraflops suministrada en la versión Pro de la herramienta **Colab** ². El número de núcleos es importante debido a que se debe usar como multiplicador al valor del *batch size* tanto en la búsqueda de hiperparámetros como en un entrenamiento regular. El tiempo en completar la búsqueda de hiperparámetros inicial fue de 2.5 horas aproximadamente, utilizando una estrategia de entrenamiento distribuido implementada en Tensorflow ³ y la herramienta *keras-tuner* ⁴. En total se realizaron sesenta (60) experimentos de 200 épocas cada uno hechos con la arquitectura propuesta en la sección 4.3. Los experimentos realizados comprenden un conjunto diferente de *batch size / taza_de_aprendizaje* definidos por el setup de la sección 4.4. La

²<https://colab.research.google.com/signup>

³https://www.tensorflow.org/guide/distributed_training

⁴https://keras.io/keras_tuner/

obtención del mejor modelo se hace usando la pérdida en validación, a menor pérdida en validación mejor es el conjunto de hiperparámetros. Los resultados de estos experimentos se pueden observar en la Figura 4-2, donde se observan las pérdidas en validación para cada uno de los conjuntos posibles. El mejor modelo se obtiene con una tasa de aprendizaje de $1.13 \cdot 10^{-3}$ y un *batch size* de 4.

4.5 Búsqueda de Hiperparámetros y Ajuste del Umbral

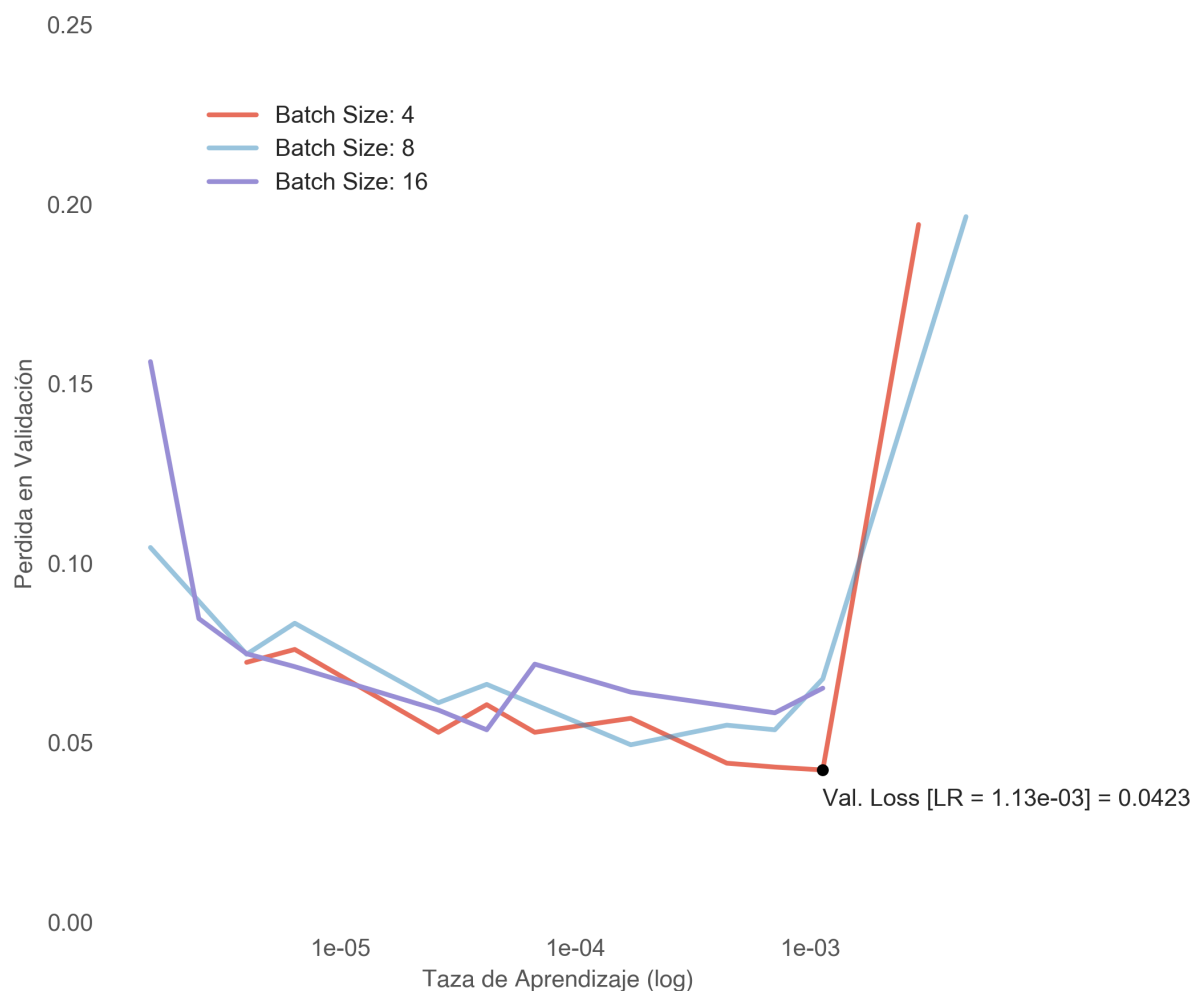


Figure 4-2: La búsqueda de hiperparámetros tienen como objetivo alcanzar la menor pérdida en validación (entropía cruzada categórica) dado un único conjunto *batch size* / *taza_de_aprendizaje*.

Luego, el siguiente paso consiste en reentrenar el modelo propuesta con 500 épocas utilizando los mejores hiperparámetros encontrados anteriormente. Una vez reentrenados los modelos,

se busca demostrar la capacidad predictiva en los datos de prueba, que en este caso son las secciones de 224×224 extraídas del mapa ortomosaico de prueba. Las métricas utilizadas son el Área bajo la curva (AUC) y *F1-Score*. Estas métricas son calculadas para cada una de las clases de salida, es decir, los canales de la máscara multiclase resultante son imágenes que tienen escala de grises codificadas de tal manera que valores cercanos al blanco [255, 255, 255] corresponden a una clase en particular de las tres, y lo restante es definido como *background* o áreas de no interés.

El umbral con el que se binariza estos canales en escala de grises normalmente se fija en un valor intermedio de 127 para un formato de enteros sin signos de 8 bits. El ajuste, en este caso, se hace modificando el umbral de binarización de cada clase a lo largo de los 255 valores que puede tomar un pixel. Los resultados obtenidos se pueden observar en la Figura 4-3 donde el umbral seleccionado es de **133**. Es decir, el máximo F1-score que obtiene la clase *maleza* (0.89 en este caso), siendo esta clase la que generalmente obtiene los valores de métricas más bajas.

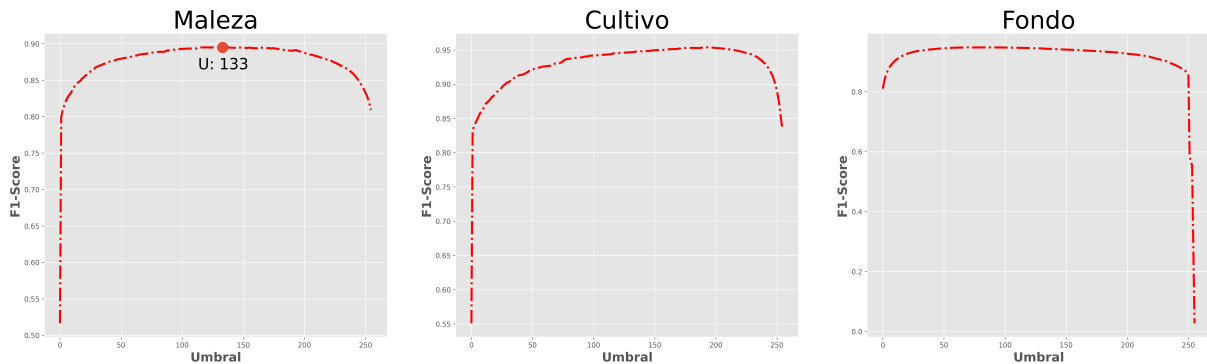


Figure 4-3: El umbral se obtiene a partir de la clase maleza, ya que generalmente es la que obtiene las métricas más bajas.

4.6 Estrategia de Aumento de Datos

Como se mencionó anteriormente, en la sección 4.4 se aplicarán una serie de estrategias de aumento de datos (AD). Lo anterior, pretende mejorar las métricas del modelo con los mejores hiperparámetros (tasa de aprendizaje de $1.13 e^{-3}$ y un *batch size* de 4) sin ningún tipo de estrategia de aumentación (SA), con cinco (5) canales en las imágenes multispectrales de entrada y usando las mismas secciones de prueba. Los experimentos están basados en el trabajo de Brilhador et al. [47], por lo tanto, las estrategias aplicadas estarán limitadas únicamente a algunas transformaciones espaciales y a un rango específico de valores. Las transformaciones se enumeran a continuación:

- Giro de 180° en horizontal [GH].

- Giro de 180° en vertical [GV].
- Rotación (de 0° a 90°) [ROT].
- Desplazamiento a lo ancho (rango de 1/10 del ancho) [DAN].
- Desplazamiento a lo alto (rango de 1/10 del alto) [DAL].
- Zoom (-0.2 a 0.2) [ZOOM].

La tabla 4-2 resumen el impacto positivo en un conjunto de métricas al aplicar cada una de estas estrategias de aumentación: *F1-Score* promedio (F1S-P), *F1-Score* maleza (F1S-M), *F1-Score* cultivo (F1S-C), *F1-Score* fondo (F1S-F), AUC promedio (AUC-P), AUC maleza (AUC-M), AUC cultivo (AUC-C), AUC fondo (AUC-F). Algunos de los mejores resultados son obtenidos en el experimento que combina el giro de 180° en el eje horizontal y en el eje vertical, otros experimentos como el Zoom también obtienen un grupo de mejores resultados. Sin embargo, el unir ambas estrategias no mejora los resultados de cada una de ellas aplicadas de manera individual. Uniendo todas las estrategias se reducen incluso las métricas con datos sin ningún tipo de aumentación aplicada, es decir, se introduce información redundante o ruido.

Table 4-2: Estrategias de aumentación de datos para mejora de métricas

Estrategia de AD	F1-Score				AUC			
	Promedio	Maleza	Cultivo	Fondo	Promedio	Maleza	Cultivo	Fondo
SA	0.9306	0.8954	0.9495	0.9469	0.9476	0.9033	0.9734	0.9660
GH/V	0.9344	0.8952	0.9563	0.9517	0.9500	0.9054	0.9773	0.9671
DAN/L	0.9328	0.8957	0.9509	0.9518	0.9494	0.9091	0.9720	0.9670
ROT	0.9278	0.8936	0.9447	0.9450	0.9517	0.9060	0.9779	0.9713
ZOOM	0.9261	0.8961	0.9318	0.9506	0.9549	0.9127	0.9802	0.9718
TODOS	0.8145	0.8104	0.8395	0.7935	0.8184	0.8134	0.8445	0.7974

4.7 Evaluación de Resultados

Una vez ya definido el mejor modelo propuesto, es decir, después de realizar sistemáticamente la búsqueda de hiper-parámetros, el ajuste del umbral y la estrategia de aumentación de datos, se procede a evaluar el impacto de los experimentos hechos en la predicción de mapas de maleza y su aporte al estado del arte actual.

4.7.1 Análisis Cuantitativo

El modelo propuesto es sometido a un análisis comparativo usando trabajos previos, tal y como se puede observar en la Tabla 4-3. La red SegNet usada en Sa et al. [12] obtiene sus mejores métricas (en términos de AUC) al combinar nueve (9) canales, incluyendo canales redundantes: B, CIR, G, NDVI, NIR, R y RE. Ramírez et al [54] usa tres modelos: SegNet, UNet y DeepLab-v3, obteniendo sus mejores métricas de este último modelo y una combinación de siete (7) canales espectrales en la imagen de entrada (no especifica la combinación).

Además, se entrenaron cada una de las arquitecturas de redes convolucionales de línea base descrita en la sub-sección 4.2, es decir, la SegNet, la UNet y la DriUNet. Todas ellas usando la misma configuración de datos de prueba, hiperparámetros, función de pérdida, aumentación de datos y umbral de la estrategia propuesta, cambiando únicamente la arquitectura del modelo. Todos los modelos descritos se evaluaron con un número diferente de canales en la imagen de entrada (12, 9, 5, 4 y 1), tal y como se especifica a continuación:

- 12 : B, CIR, G, NDVI, NIR, R, RE, RGB.
- 9 : B, G, NDVI, NIR, R, RE, RGB.
- 5 : NIR, RE, RGB.
- 4 : NIR, RGB.
- 1 : NDVI.

El *F1-Score* y el *AUC* fue evaluado para cada una de las tres clases que componen el mapa de salida, además del promedio de todas las clases en ambas métricas. La hipótesis inicial era que a medida que los modelos eran entrenados con más canales, la propuesta sacaría mayor provecho del autoencoder, dado que el modelo segmentador siempre trataría con una versión comprimida de tres 3 canales de las imágenes multiespectrales y por ende facilitaría el aprendizaje.

Sin embargo, en los experimentos que involucran 12, 9, 5 y 4 canales, se obtienen mejores métricas con los modelos de línea base en comparación con el modelo propuesto. Caso totalmente contrario ocurre con los modelos entrenados con un 1 canal, el modelo propuesto saca mejor provecho del índice NDVI en comparación con la DriUNet, especialmente en la clase *maleza*.

Table 4-3: Evaluación de resultados del modelo propuesto en métricas como el *F1-Score* y el AUC, y comparación con trabajos previos que han utilizado la base de datos.

Arquitectura	Canales	F1-Score				AUC			
		Promedio	Maleza	Cultivo	Fondo	Promedio	Maleza	Cultivo	Fondo
SegNet [12]	9	-	-	-	-	0.8280	0.7820	0.8630	0.8390
DeepLabV3 [54]	7	0.8233	0.8800	0.7800	0.8100	0.8800	0.8300	0.8900	0.9200
DriUNet	12	0.9290	0.9007	0.9390	0.9471	0.9582	0.9128	0.9863	0.9755
SegNet		0.9333	0.8991	0.9492	0.9518	0.9428	0.9073	0.9643	0.9567
UNet		0.9195	0.8976	0.9206	0.9403	0.9560	0.9051	0.9864	0.9765
Propuesta		0.9275	0.8939	0.9386	0.9502	0.9519	0.9096	0.9778	0.9682
DriUNet	9	0.9356	0.9038	0.9503	0.9527	0.9568	0.9127	0.9852	0.9726
SegNet		0.9280	0.8979	0.9407	0.9455	0.9486	0.9129	0.9691	0.9637
UNet		0.9115	0.8674	0.9190	0.9481	0.9471	0.8773	0.9862	0.9779
Propuesta		0.9302	0.8981	0.9442	0.9483	0.9568	0.9190	0.9796	0.9719
DriUNet	5	0.9370	0.9007	0.9560	0.9541	0.9533	0.9044	0.9835	0.9720
SegNet		0.9307	0.8955	0.9481	0.9485	0.9484	0.9026	0.9761	0.9664
UNet		0.9278	0.8939	0.9390	0.9506	0.9595	0.9198	0.9814	0.9774
Propuesta		0.9261	0.8961	0.9318	0.9506	0.9549	0.9127	0.9802	0.9718
DriUNet	4	0.9345	0.8980	0.9493	0.9562	0.9548	0.9080	0.9840	0.9723
SegNet		0.9346	0.8996	0.9534	0.9507	0.9479	0.9067	0.9840	0.9644
UNet		0.9289	0.8867	0.9477	0.9524	0.9563	0.9170	0.9783	0.9737
Propuesta		0.9266	0.8921	0.9398	0.9477	0.9497	0.9056	0.9764	0.9671
DriUNet	1	0.9195	0.8759	0.9385	0.9442	0.9300	0.8750	0.9681	0.9470
SegNet		0.9047	0.8607	0.9303	0.9229	0.9034	0.8564	0.9385	0.9154
UNet		0.9240	0.8832	0.9390	0.9500	0.9380	0.8843	0.9726	0.9571
Propuesta		0.9316	0.8966	0.9461	0.9521	0.9495	0.9108	0.9712	0.9664

4.7.2 Análisis Pre-Productivo: Mapa Ortomosaico de Prueba RedEdge-MX

En experimentos se llegó a la conclusión de que no es posible usar los mapas ortomosaicos del sensor Sequoia para un análisis pre-productivo debido a que la variabilidad en estos datos no es capturada por los modelos hasta ahora entrenados con el sensor RedEdge-MX. Por lo tanto, se debe recurrir una vez más a los de datos de prueba del sensor RedEdge-MX y seleccionar cada uno de los modelos entrenados con el número de canales que maximizo las métricas en la clase maleza, es decir, que tanto para el modelo propuesto, como para la DriUNet y la SegNet se usa el modelo entrenado con nueve (9) canales espectrales, en el caso de la UNet se usa el modelo entrenado con cinco (5) canales espectrales.

Para llevar a cabo el análisis se usa directamente el mapa ortomosaico de prueba completo, es decir, por fuera de la infraestructura desarrollada para entrenar con TFRecords y una TPU. Para ello se utiliza como entrada a los modelos secciones de 224×224 cuyas coordenadas están referenciadas dentro de las dimensiones del mapa ortomosaico, permitiendo reconstruir

el mapa de maleza a la salida. En la Figura 4-4 se pueden observar los resultados obtenidos para la DriUNet, de izquierda a derecha la primera imagen corresponde al mapa etiqueto por el experto, la segunda imagen corresponde al mapa de salida de la DriUNet. En la Figura 4-5 se puede ver el mapa de maleza del modelo propuesto. Finalmente la Figura 4-6 y la Figura 4-7 corresponden a los resultados obtenidos por la SegNet y la UNet respectivamente.

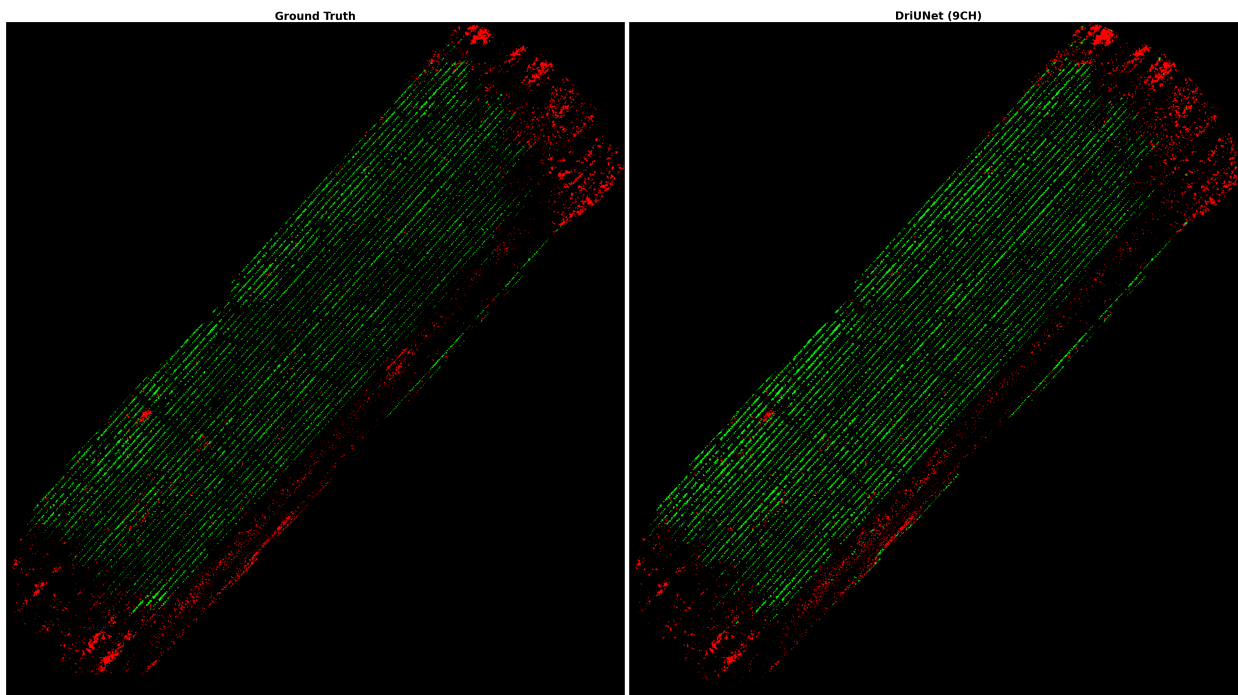


Figure 4-4: Comparación entre la referencia (Groundtruth) y el mapa de maleza generado por la DriUNet multiclase. A la izquierda la referencia etiquetada por el experto, a la derecha el mapa generado por la DriUNet.

Finalmente se realizan recortes sobre el mapa ortomosaico y las mascararas resultantes de dichos recortes, esto con el fin de hacer una comparativa entre la DriUNet y la propuesta, los resultados se pueden ver en la Figura 4-8, 4-9, 4-4. En todos los casos se observan resultados similares, siendo los errores mas comunes los falsos positivos en la clase *cultivo*.

Es un hecho que las métricas obtenidas en la Tabla 4-3 logran captar la diferencia entre el mapa de maleza obtenido por los modelos de linea base y el mapa de maleza generado por la propuesta desarrollada en este trabajo. Finalmente, es necesario obtener las métricas directamente sobre los mapas de maleza que se observan en la Figura 4-4. Para esto se utilizan un conjunto de métricas adicionales como el *índice de Jaccard* y el *coeficiente de Dice* tanto para los modelos de linea base como para el modelo propuesto, los resultados se muestran en la Tabla 4-4.

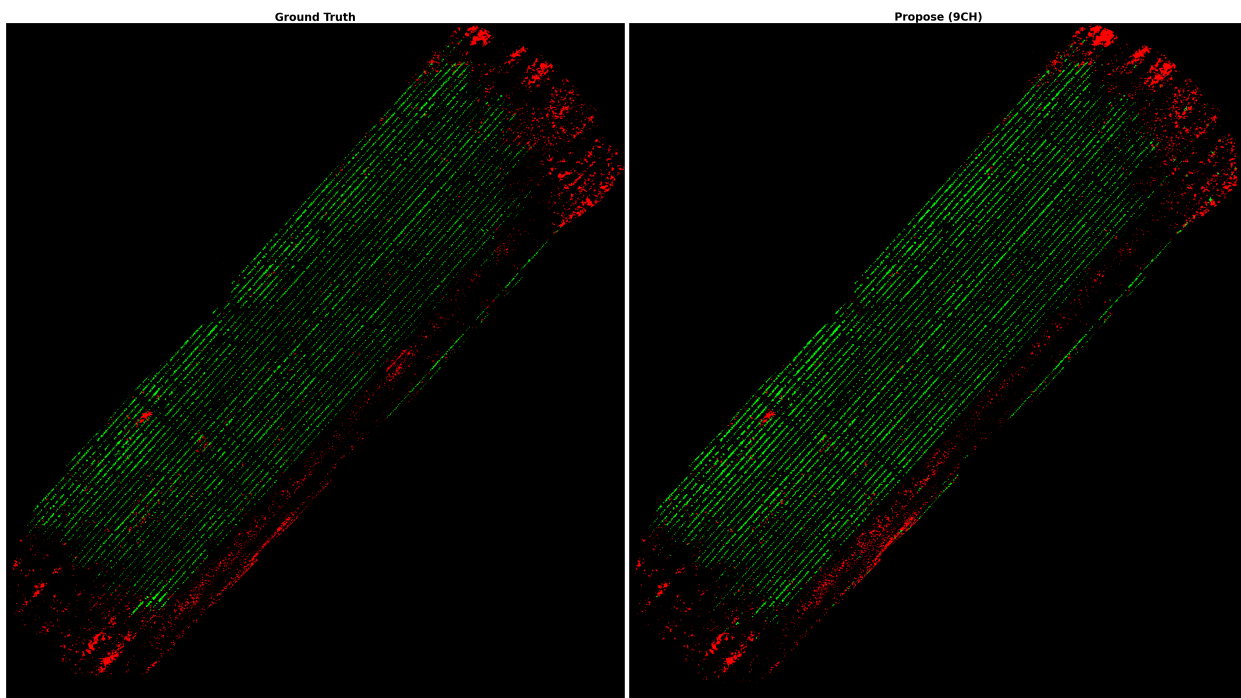


Figure 4-5: Comparacion entre el groundtruth y el mapa de maleza generado por la propuesta planteada en este proyecto. A la izquierda la referencia etiquetada por el experto, a la derecha el mapa generado por la propuesta.

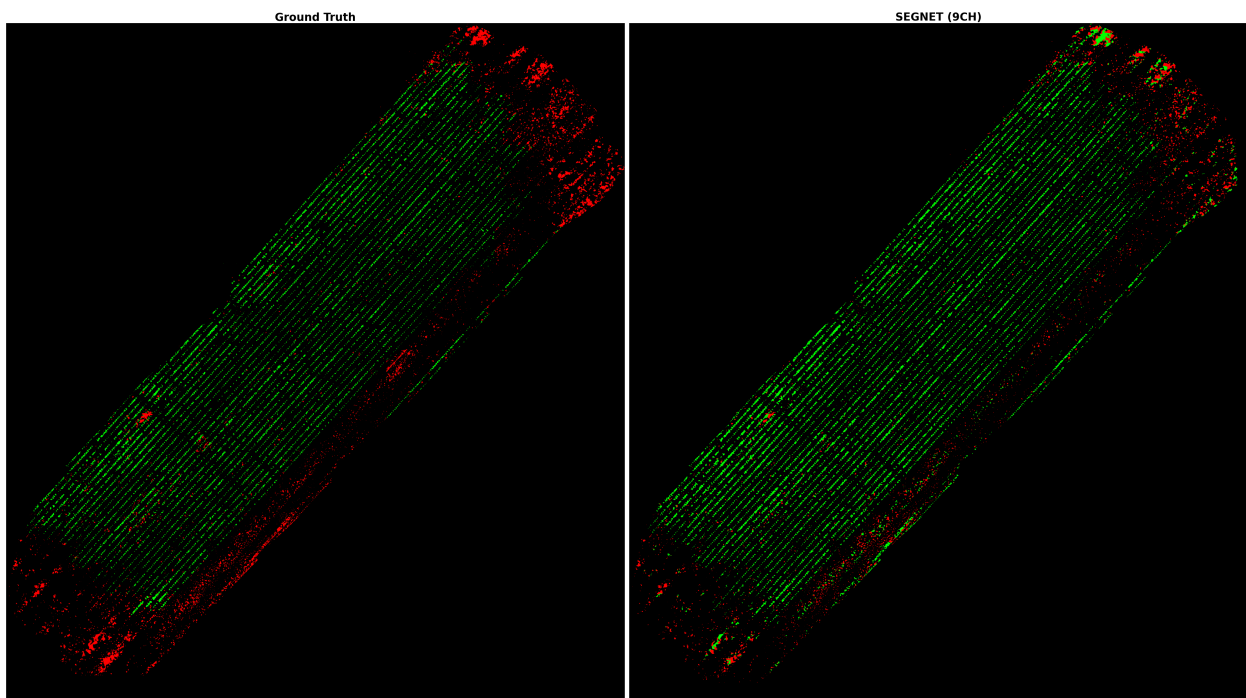


Figure 4-6: Comparacion entre el groundtruth y el mapa de maleza generado por la SegNet. A la izquierda la referencia etiquetada por el experto, a la derecha el mapa generado por la SegNet.

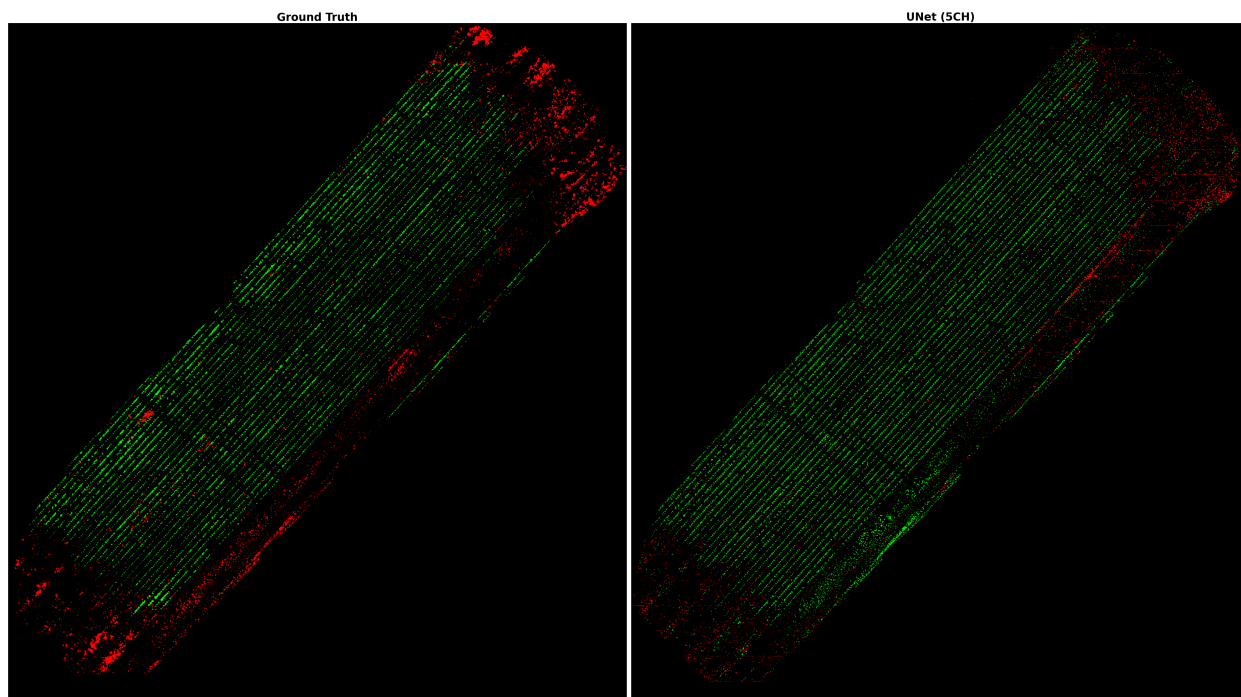


Figure 4-7: Comparación entre el groundtruth y el mapa de maleza generado por la UNet. A la izquierda la referencia etiquetada por el experto, a la derecha el mapa generado por la UNet.

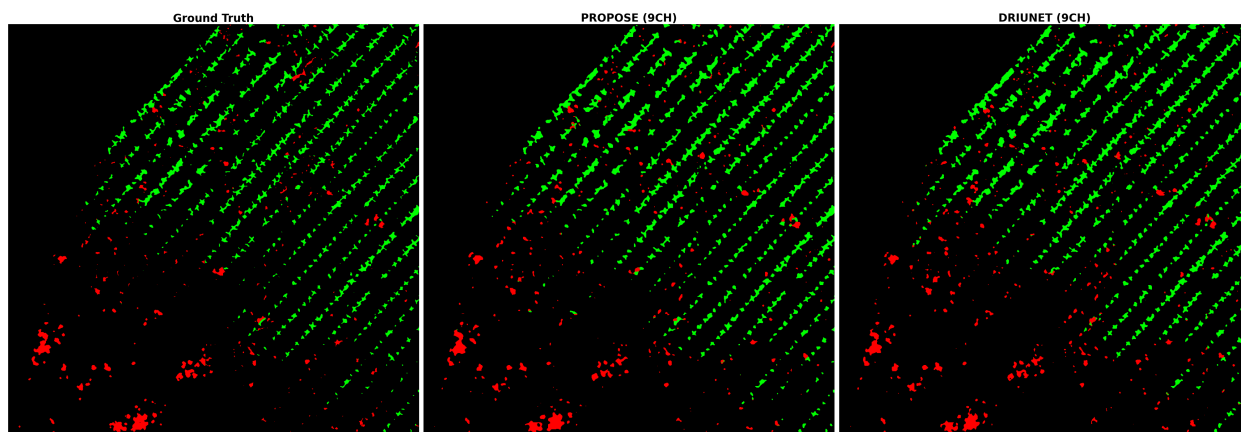


Figure 4-8: Comparación de una sección del mapa de maleza para los modelos que mejor resultados de segmentación obtuvieron Parte 1.

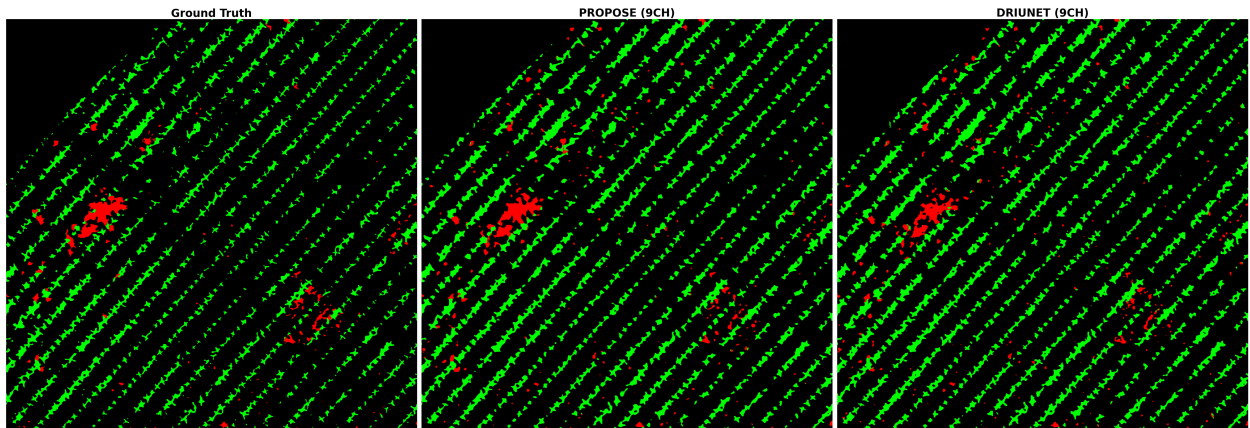


Figure 4-9: Comparación de una sección del mapa de maleza para los modelos que mejor resultados de segmentación obtuvieron Parte 2.

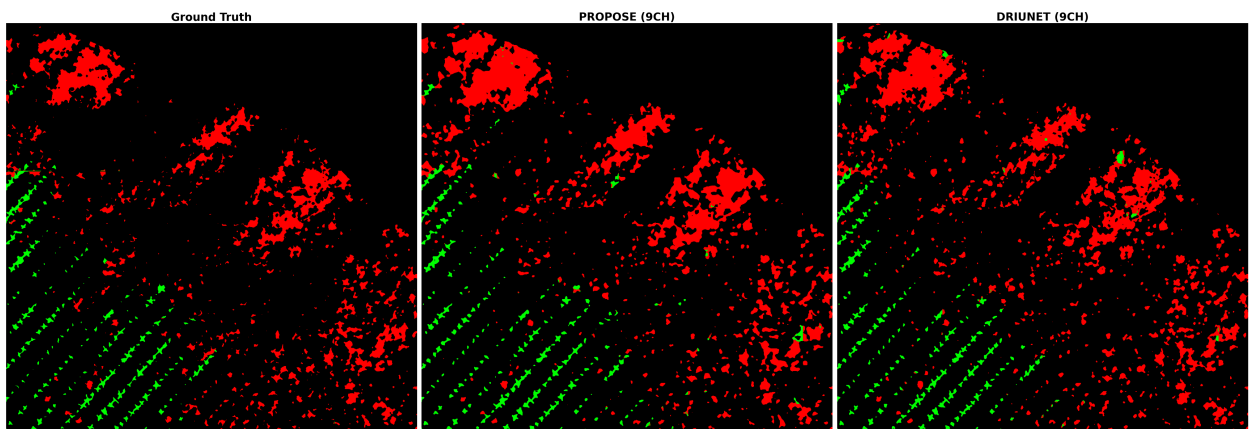


Figure 4-10: Comparación de una sección del mapa de maleza para los modelos que mejor resultados de segmentación obtuvieron Parte 3.

Table 4-4: Evaluación de resultados del modelo propuesto en métricas alternativas, y comparación con las redes de línea base.

Modelo	Coeficiente de Dice				Índice de Jaccard			
	Promedio	Maleza	Cultivo	Fondo	Promedio	Maleza	Cultivo	Fondo
DriUNet [9CH]	0.8846	0.7951	0.8575	0.9914	0.8414	0.7278	0.8131	0.9833
UNet [5CH]	0.8258	0.6665	0.8216	0.9892	0.7941	0.6271	0.7759	0.9793
SegNet [9CH]	0.8258	0.7463	0.8215	0.9909	0.7941	0.6889	0.7745	0.9824
Propuesta [9CH]	0.8788	0.7877	0.8461	0.9901	0.8321	0.7167	0.7986	0.9809

5 Conclusiones y trabajos futuros

Esta investigación estudió y exploró varios enfoques que proporcionaron una metodología para entrenar de manera eficiente modelos de aprendizaje profundo usando imágenes multiespectrales. Por un lado, el modelo desarrollado aprovecha la información proveniente de canales multiespectrales para generar automáticamente un mapeo de maleza. Por otro lado, generamos una estrategia para la evaluación cuantitativa y cualitativa con diferentes métricas de desempeño.

La estimación objetiva de un mapa de maleza requiere criterios enfocados en métricas de desempeño para encontrar maleza, cultivo e información de fondo o no relevante para nuestra investigación. Sin embargo, se encontró que es posible evaluar la generación del mapa de maleza a través de métricas de desempeño cuantitativas como f1-score y área bajo la curva, coeficiente de Dice e índice de Jaccard. Adicionalmente, se realizó la evaluación cualitativa en la base de datos que contaba con una etapa temprana de maleza, es decir, menos del 1% de información de maleza.

El modelo propuesto propone dividir el tratamiento de la información multiespectral y la segmentación semántica en dos etapas. La primera etapa consiste en comprimir la imagen multiespectral de N canales en una versión equivalente de tres canales usando un autoencoder. El modelo segmentador, por lo tanto, entrena con una imagen de tres canales, para lo cual fue diseñado, obteniendo mejores resultados al disminuir la cantidad de información que tiene que compactar. Otro punto importante es que tanto el autoencoder como el segmentador se entrenan simultáneamente (end-to-end), dándole una mayor prioridad a la pérdida del segmentador.

Una evaluación exhaustiva de diversas arquitecturas reportadas en el estado del arte como SegNet, DeepLab V3 y DriUNet con diferentes combinaciones experimentales de canales fue realizada en este trabajo. Lo anterior, nos permitió obtener un resultado más robusto, el cual a nuestro conocimiento supera lo reportado en los diferentes trabajos del estado del arte, como se reportó en la tabla 4-3.

Durante el experimento se pudo constatar que el modelo es dependiente de los datos, por lo tanto para poder segmentar maleza en otros campos de cultivo, con otro tipo de maleza, a otra altura, con diferentes sensores y incluso en una hora diferente del día, es necesario

reentrenar con nuevos datos el modelo.

Finalmente, este trabajo es la primera etapa de un proyecto de control por precisión de la maleza validado con bases de datos reportadas en la literatura. Para trabajos futuros, se sugiere la exploración con otras bases de datos más extensas, otras modalidades o imágenes hiperespectrales. Otra posible línea de trabajo es integrar este trabajo en un flujo de trabajo en un dron o un robot e incorporarlo con otros modelos de diagnóstico de otro tipo de enfermedades en plantas para apoyo en el mapeo automático de maleza.

Bibliography

- [1] Stephen O Duke. Perspectives on transgenic, herbicide-resistant crops in the united states almost 20 years after introduction. *Pest management science*, 71(5):652–657, 2015.
- [2] Alexa Varah, Kwadjo Ahodo, Shaun R Coutts, Helen L Hicks, David Comont, Laura Crook, Richard Hull, Paul Neve, Dylan Z Childs, Robert P Freckleton, et al. The costs of human-induced evolution in an agricultural system. *Nature sustainability*, 3(1):63–71, 2020.
- [3] Alessandro dos Santos Ferreira, Daniel Matte Freitas, Gercina Gonçalves da Silva, Hemerson Pistori, and Marcelo Theophilo Folhes. Weed detection in soybean crops using convnets. *Computers and Electronics in Agriculture*, 143:314–324, 2017.
- [4] Shirley A Briggs. *Basic guide to pesticides: their characteristics and hazards*. CRC Press, 2018.
- [5] Isabelle Schuster, Henning Nordmeyer, and Thomas Rath. Comparison of vision-based and manual weed mapping in sugar beet. *Biosystems engineering*, 98(1):17–25, 2007.
- [6] David Pimentel, Herbert Acquay, Michael Biltonen, P Rice, M Silva, J Nelson, V Lipner, S Giordano, A Horowitz, and M D’amore. Environmental and economic costs of pesticide use. *BioScience*, 42(10):750–760, 1992.
- [7] K Neil Harker and John T O’Donovan. Recent weed control, weed management, and integrated weed management. *Weed Technology*, 27(1):1–11, 2013.
- [8] Mulham Fawakherji, Ali Youssef, Domenico D Bloisi, Alberto Pretto, and Daniele Nardi. Crop and weed classification using pixel-wise segmentation on ground and aerial images. *Int. J. Robot. Comput*, 2(1):39–57, 2020.
- [9] David R Shaw. Remote sensing and site-specific weed management. *Frontiers in Ecology and the Environment*, 3(10):526–532, 2005.
- [10] Colin Birch, Ian Cooper, Gurjeet Gill, Stephen Adkins, and Madan Gupta. Weed management in rainfed agricultural systems. In *Rainfed Farming Systems*, pages 215–232. Springer, 2011.

-
- [11] Philipp Lottes, Jens Behley, Nived Chebrolu, Andres Milioto, and Cyrill Stachniss. Robust joint stem detection and crop-weed classification using image sequences for plant-specific treatment in precision farming. *Journal of Field Robotics*, 37(1):20–34, 2020.
- [12] Inkyu Sa, Marija Popović, Raghav Khanna, Zetao Chen, Philipp Lottes, Frank Liebisch, Juan Nieto, Cyrill Stachniss, Achim Walter, and Roland Siegwart. Weedmap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sensing*, 10(9):1423, 2018.
- [13] Jorge Torres-Sánchez, José Manuel Pena, Ana Isabel de Castro, and Francisca López-Granados. Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from uav. *Computers and Electronics in Agriculture*, 103:104–113, 2014.
- [14] Chang-chun Li, Guang-sheng Zhang, Tian-jie Lei, and A-du GONG. Quick image-processing method of uav without control points data in earthquake disaster area. *Transactions of Nonferrous Metals Society of China*, 21:s523–s528, 2011.
- [15] Andreas Kamilaris and Francesc X Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147:70–90, 2018.
- [16] Konstantinos G Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. Machine learning in agriculture: A review. *Sensors*, 18(8):2674, 2018.
- [17] Dimosthenis C Tsouros, Stamatia Bibi, and Panagiotis G Sarigiannidis. A review on uav-based applications for precision agriculture. *Information*, 10(11):349, 2019.
- [18] Huasheng Huang, Yubin Lan, Aqing Yang, Yali Zhang, Sheng Wen, and Jizhong Deng. Deep learning versus object-based image analysis (obia) in weed mapping of uav imagery. *International Journal of Remote Sensing*, 41(9):3446–3479, 2020.
- [19] Inkyu Sa, Zetao Chen, Marija Popović, Raghav Khanna, Frank Liebisch, Juan Nieto, and Roland Siegwart. weednet: Dense semantic weed classification using multispectral images and mav for smart farming. *IEEE Robotics and Automation Letters*, 3(1):588–595, 2017.
- [20] Alwaseela Abdalla, Haiyan Cen, Liang Wan, Reem Rashid, Haiyong Weng, Weijun Zhou, and Yong He. Fine-tuning convolutional neural network with transfer learning for semantic segmentation of ground-level oilseed rape images in a field with high weed pressure. *Computers and Electronics in Agriculture*, 167:105091, 2019.
- [21] Kavir Osorio, Andrés Puerto, Cesar Pedraza, David Jamaica, and Leonardo Rodríguez. A deep learning approach for weed detection in lettuce crops using multispectral images. *AgriEngineering*, 2(3):471–488, 2020.

-
- [22] Sigurbjörn Jónsson. Rgb and multispectral uav image classification of agricultural fields using a machine learning algorithm. *Student thesis series INES*, 2019.
- [23] ASM Mahmudul Hasan, Ferdous Sohel, Dean Diepeveen, Hamid Laga, and Michael GK Jones. A survey of deep learning techniques for weed detection from images. *Computers and Electronics in Agriculture*, 184:106067, 2021.
- [24] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* ” O’Reilly Media, Inc.”, 2019.
- [25] Calvin Hung, Zhe Xu, and Salah Sukkarieh. Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a uav. *Remote Sensing*, 6(12):12037–12054, 2014.
- [26] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [27] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021.
- [28] Huasheng Huang, Jizhong Deng, Yubin Lan, Aqing Yang, Xiaoling Deng, and Lei Zhang. A fully convolutional network for weed mapping of unmanned aerial vehicle (uav) imagery. *PloS one*, 13(4):e0196302, 2018.
- [29] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [30] Sebastian Haug and Jörn Ostermann. A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks. In Lourdes Agapito, Michael M. Bronstein, and Carsten Rother, editors, *Computer Vision - ECCV 2014 Workshops*, pages 105–116, Cham, 2015. Springer International Publishing.
- [31] Sebastian Haug, Andreas Michaels, Peter Biber, and Jörn Ostermann. Plant classification system for crop/weed discrimination without segmentation. In *IEEE winter conference on applications of computer vision*, pages 1142–1149. IEEE, 2014.
- [32] José Manuel Peña, Jorge Torres-Sánchez, Ana Isabel de Castro, Maggi Kelly, and Francisca López-Granados. Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (uav) images. *PLOS ONE*, 8(10):null, 10 2013.
- [33] Maria Pérez-Ortiz, JM Peña, Pedro Antonio Gutiérrez, Jorge Torres-Sánchez, César Hervás-Martínez, and Francisca López-Granados. A semi-supervised system for weed

- mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. *Applied Soft Computing*, 37:533–544, 2015.
- [34] Thomas K Alexandridis, Afroditi Alexandra Tamouridou, Xanthoula Eirini Pantazi, Anastasia L Lagopodi, Javid Kashefi, Georgios Ovakoglou, Vassilios Polychronos, and Dimitrios Moshou. Novelty detection classifiers in weed mapping: *Silybum marianum* detection on uav multispectral images. *Sensors*, 17(9):2007, 2017.
- [35] Philipp Lottes, Raghav Khanna, Johannes Pfeifer, Roland Siegwart, and Cyrill Stachniss. Uav-based crop and weed classification for smart farming. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3024–3031. IEEE, 2017.
- [36] Anders Krogh Mortensen, Mads Dyrmann, Henrik Karstoft, R Nyholm Jørgensen, René Gislum, et al. Semantic segmentation of mixed crops using deep convolutional neural network. In *CIGR-AgEng Conference, 26-29 June 2016, Aarhus, Denmark. Abstracts and Full papers*, pages 1–6. Organising Committee, CIGR 2016, 2016.
- [37] M. Dyrmann, R. N. Jørgensen, and H. S. Midtiby. Roboweedsupport - detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. *Advances in Animal Biosciences*, 8(2):842–847, 2017.
- [38] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.
- [39] Maurilio Di Cicco, Ciro Potena, Giorgio Grisetti, and Alberto Pretto. Automatic model based dataset generation for fast and accurate crop and weeds detection. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5188–5195. IEEE, 2017.
- [40] Huasheng Huang, Yubin Lan, Jizhong Deng, Aqing Yang, Xiaoling Deng, Lei Zhang, and Sheng Wen. A semantic labeling approach for accurate weed mapping of high resolution uav imagery. *Sensors*, 18(7):2113, 2018.
- [41] Soren Skovsen, Mads Dyrmann, Anders K Mortensen, Morten S Laursen, René Gislum, Jorgen Eriksen, Sadaf Farkhani, Henrik Karstoft, and Rasmus N Jorgensen. The grassclover image dataset for semantic and hierarchical species understanding in agriculture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.
- [42] Yannik Rist, Iurii Shendryk, Foivos Diakogiannis, and Shaun Levick. Weed mapping using very high resolution satellite imagery and fully convolutional neural network. In

- IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 9784–9787. IEEE, 2019.
- [43] Foivos I Diakogiannis, François Waldner, Peter Caccetta, and Chen Wu. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:94–114, 2020.
- [44] Shyam Prasad Adhikari, Heechan Yang, and Hyongsuk Kim. Learning semantic graphics using convolutional encoder–decoder network for autonomous weeding in paddy. *Frontiers in plant science*, 10:1404, 2019.
- [45] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [46] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.
- [47] Anderson Brillhador, Matheus Gutoski, Leandro Takeshi Hattori, Andrei de Souza Inácio, André Eugênio Lazzaretti, and Heitor Silvério Lopes. Classification of weeds and crops at the pixel-level using convolutional neural networks and data augmentation. In *2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pages 1–6. IEEE, 2019.
- [48] Mulham Fawakherji, Ali Youssef, Domenico Bloisi, Alberto Pretto, and Daniele Nardi. Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation. In *2019 Third IEEE International Conference on Robotic Computing (IRC)*, pages 146–152. IEEE, 2019.
- [49] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [50] Muhammad Hamza Asad and Abdul Bais. Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network. *Information Processing in Agriculture*, 7(4):535–545, 2020.
- [51] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [52] Xu Ma, Xiangwu Deng, Long Qi, Yu Jiang, Hongwei Li, Yuwei Wang, and Xupo Xing. Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields. *PloS one*, 14(4):e0215676, 2019.

-
- [53] Lukas Petrich, Georg Lohrmann, Matthias Neumann, Fabio Martin, Andreas Frey, Albert Stoll, and Volker Schmidt. Detection of colchicum autumnale in drone images, using a machine-learning approach. *Precision Agriculture*, 21(6):1291–1303, 2020.
- [54] W Ramirez, P Achanccaray, LF Mendoza, and MAC Pacheco. Deep convolutional neural networks for weed detection in agricultural crops using optical aerial images. In *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)*, pages 133–137. IEEE, 2020.
- [55] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005.
- [56] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.
- [57] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [58] Kunlin Zou, Xin Chen, Fan Zhang, Hang Zhou, and Chunlong Zhang. A field weed density evaluation method based on uav imaging and modified u-net. *Remote Sensing*, 13(2):310, 2021.
- [59] Petra Bosilj, Erchan Aptoula, Tom Duckett, and Grzegorz Cielniak. Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture. *Journal of Field Robotics*, 37(1):7–19, 2020.
- [60] S Umamaheswari and Ashvini V Jain. Encoder–decoder architecture for crop-weed classification using pixel-wise labelling. In *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*, pages 1–6. IEEE, 2020.
- [61] Aichen Wang, Yifei Xu, Xinhua Wei, and Bingbo Cui. Semantic segmentation of crop and weed using an encoder-decoder network and image enhancement method under uncontrolled outdoor illumination. *IEEE Access*, 8:81724–81734, 2020.
- [62] Yuzhen Lu and Sierra Young. A survey of public datasets for computer vision tasks in precision agriculture. *Computers and Electronics in Agriculture*, 178:105760, 2020.
- [63] Zhangnan Wu, Yajun Chen, Bo Zhao, Xiaobing Kang, and Yuanyuan Ding. Review of weed detection methods based on computer vision. *Sensors*, 21(11):3647, 2021.

-
- [64] Merima Kulin, Tarik Kazaz, Eli De Poorter, and Ingrid Moerman. A survey on machine learning-based performance improvement of wireless networks: Phy, mac and network layer. *Electronics*, 10(3):318, 2021.
- [65] Panqu Wang, Pengfei Chen, Ye Yuan, Ding Liu, Zehua Huang, Xiaodi Hou, and Garrison Cottrell. Understanding convolution for semantic segmentation. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 1451–1460. IEEE, 2018.
- [66] Peng Liu, Hui Zhang, and Kie B Eom. Active deep learning for classification of hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(2):712–724, 2016.
- [67] Nguyen Thanh Toan and Nguyen Thanh Tam. Early bushfire detection with 3d cnn from streams of satellite images.
- [68] Nikhil Jangamreddy. A survey on specialised hardware for machine learning. 2019.
- [69] E-C Oerke. Crop losses to pests. *The Journal of Agricultural Science*, 144(1):31–43, 2006.
- [70] Craig D Osteen and Jorge Fernandez-Cornejo. Herbicide use trends: a backgrounder. *Choices*, 31(4):1–7, 2016.
- [71] John Peterson Myers, Michael N Antoniou, Bruce Blumberg, Lynn Carroll, Theo Colborn, Lorne G Everett, Michael Hansen, Philip J Landrigan, Bruce P Lanphear, Robin Mesnage, et al. Concerns over use of glyphosate-based herbicides and risks associated with exposures: a consensus statement. *Environmental Health*, 15(1):1–13, 2016.
- [72] Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Pablo Arbeláez, and Luc Van Gool. Deep retinal image understanding. In *International conference on medical image computing and computer-assisted intervention*, pages 140–148. Springer, 2016.
- [73] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [74] Liam Li and Ameet Talwalkar. Random search and reproducibility for neural architecture search. In *Uncertainty in artificial intelligence*, pages 367–377. PMLR, 2020.