



# Title: Semantic segmentation of sugarcane crops with unmanned aerial vehicle (UAV) imagery

**Authors:** LÁRRAGA-ALTAMIRANO, Hugo René, HERNÁNDEZ-LÓPEZ, Dalia Rosario, PIEDAD-RUBIO, Ana María and HERNÁNDEZ-GÓMEZ, Germán David

Editorial label ECORFAN: 607-8695

BECORFAN Control Number: 2022-01

BECORFAN Classification (2022): 131222-0001

Pages: 12

RNA: 03-2010-032610115700-14

**ECORFAN-México, S.C.**

143 – 50 Itzopan Street

La Florida, Ecatepec Municipality

Mexico State, 55120 Zipcode

Phone: +52 1 55 6159 2296

Skype: ecorfan-mexico.s.c.

E-mail: contacto@ecorfan.org

Facebook: ECORFAN-México S. C.

Twitter: @EcorfanC

[www.ecorfan.org](http://www.ecorfan.org)

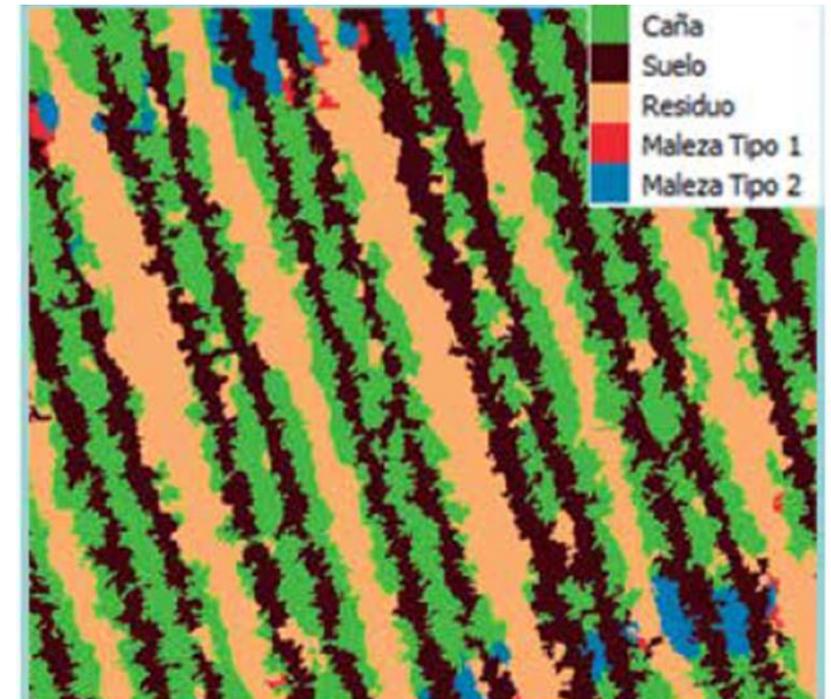
**Holdings**

Mexico	Colombia	Guatemala
Bolivia	Cameroon	Democratic
Spain	El Salvador	Republic
Ecuador	Taiwan	of Congo
Peru	Paraguay	Nicaragua

# Introducción

Una de las aplicaciones de mayor uso actualmente en la percepción remota es la clasificación de cultivos, a través de la cual se identifican aquellas áreas de interés para ser atendidas con mayor precisión (Gu et al., 2019). Las plataformas utilizadas para la obtención de imágenes en agricultura son los satélites y los vehículos aéreos no tripulados conocidos como drones (Anilkumar & Venugopal, 2022).

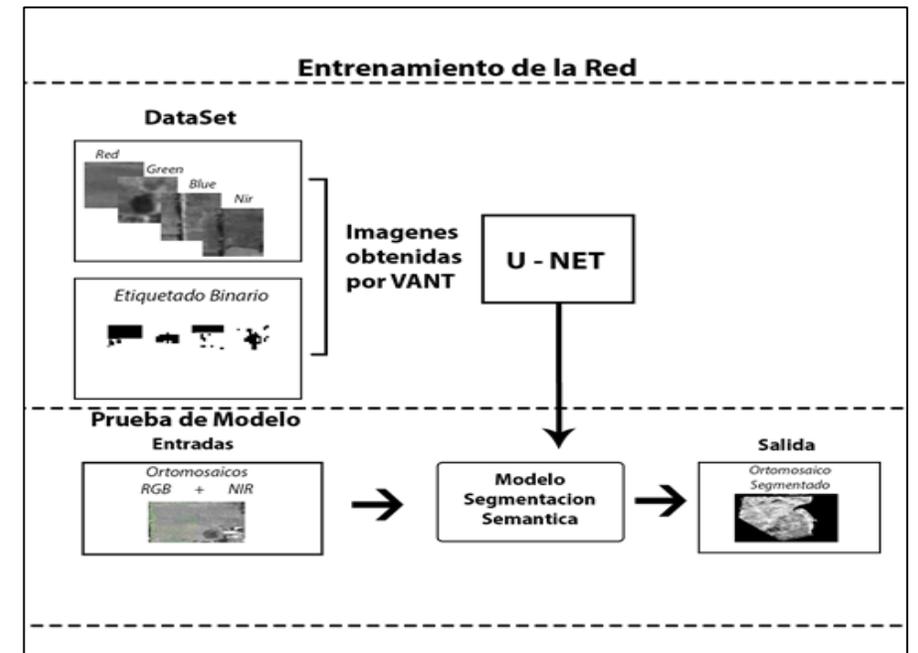
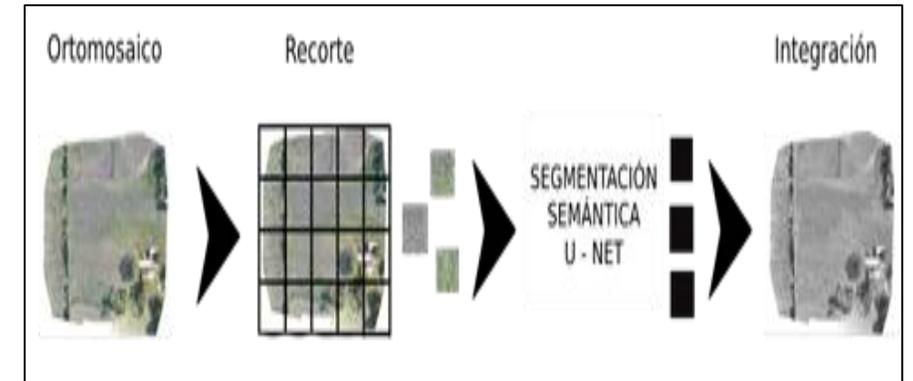
Uno de los modelos de redes CNN (redes neuronales convolucionales) con capacidad para identificar varias clases de objetos, por ejemplo, la vegetación, de la imagen de un cultivo, es el diseño U-Net. Este modelo utiliza una rama de codificación y otra de decodificación integradas con operaciones de convolución y pooling, para procesos de segmentación semántica (Das & Bais, 2021).



# Introducción

El proceso de segmentación semántica consiste en etiquetar cada píxel de la imagen, de manera que sea posible distinguir a qué clase pertenece. De las ventajas de este enfoque es la detección de características en las imágenes mediante información contextual que distinguen a los objetos (Ramirez et al., 2020).

La propuesta aquí expuesta consiste en el diseño de una CNN de tipo U-Net para segmentación semántica de cultivos de caña de azúcar. A partir de un ortomosaico generado por fotografías provenientes de vehículos aéreos no tripulados (VANT), sea posible reconocer solo los píxeles que pertenecen al cultivo del resto de la ortofoto. Este enfoque, entrena al clasificador para etiquetar de manera binaria los píxeles de la imagen, siendo el valor 1 aquellos que identifican la caña y el 0 para todo lo demás.

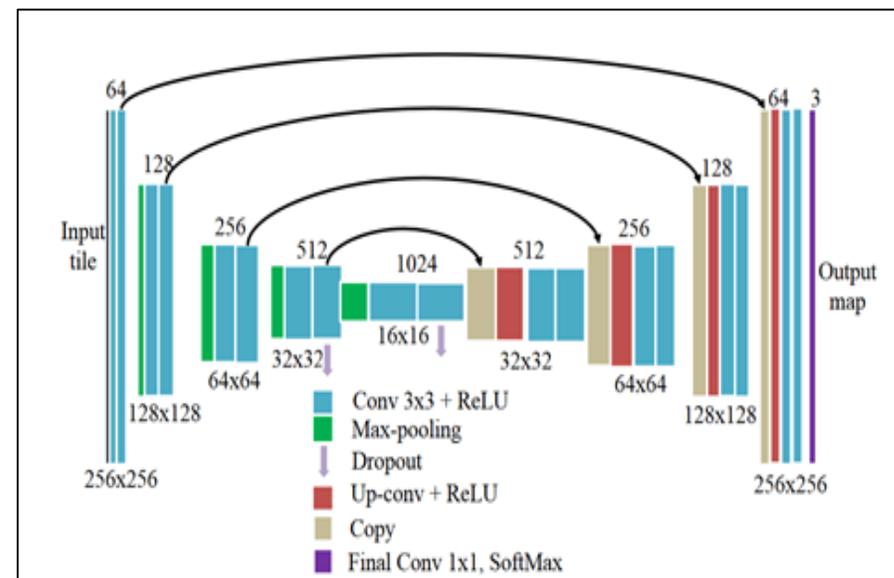


# Metodología

## Diseño de la red

Dado que se desea analizar pixel por pixel de una imagen para ser clasificado en la clase 'cultivo' o en la clase 'no cultivo', se propone usar una red neuronal convolucional por su capacidad de extraer con mayor detalle características espectrales. La CNN específica para esta tarea de segmentación semántica es el modelo U-net.

Para la fase de entrenamiento de la red las entradas están compuestas por imágenes de 4 canales: Red, Green, Blue y NIR. Las salidas son imágenes binarias las cuales representan con el valor 1 a la clase 'cultivo' y 0 a la clase 'no cultivo' (Su et al., 2021).



# Metodología

## Prueba del modelo

Posterior a realizar las pruebas de entrenamiento y generar el modelo con mayor nivel de exactitud en la segmentación semántica del cultivo de caña de azúcar, este debe de probarse en un ambiente real. Para tal fin se construye una aplicación de software mediante Python para procesar un nuevo ortomosaico no utilizado en el proceso de entrenamiento de la red neuronal. Las etapas de la aplicación son las siguientes:

- Las imágenes de entrada provienen de ortomosaicos en las bandas Red, Green, Blue y NIR.
- Recorte de ortomosaicos en subimágenes de 256x256 px, ya que estas son las dimensiones esperadas por el modelo.
- Cada imagen es cargada a la memoria del sistema para ser procesada por el modelo. El resultado es una matriz de valores de probabilidad, se consideran aquellos mayores del 65% como píxeles de la clase cultivo para crear la imagen binaria.
- Se ensamblan las imágenes binarias nuevamente respetando la posición original para así, integrar un nuevo ortomosaico binario que muestra sólo los píxeles del cultivo de caña de azúcar.

# Resultados

## Imágenes para entrenamiento (DataSet)

El conjunto de imágenes para entrenamiento (DataSet) se construyó mediante fotogrametría aérea utilizando un drone modelo DJI Matrice 600 Pro con una autonomía de vuelo de 23 minutos con baterías cargadas al 100%.

El cultivo de caña de azúcar estudiado pertenece al rancho "La Aurora" ubicado en el municipio de Ciudad Valles, S.L.P, México con una superficie de 23 hectáreas. El sensor multiespectral es una cámara Sequoia de 5 lentes: rojo (Red), verde (Green), rojo cercano (NIR), rojo límite (RedEdge) y RGB, la cual fue instalada en el drone, adicionalmente del sensor de luz y una batería independiente.

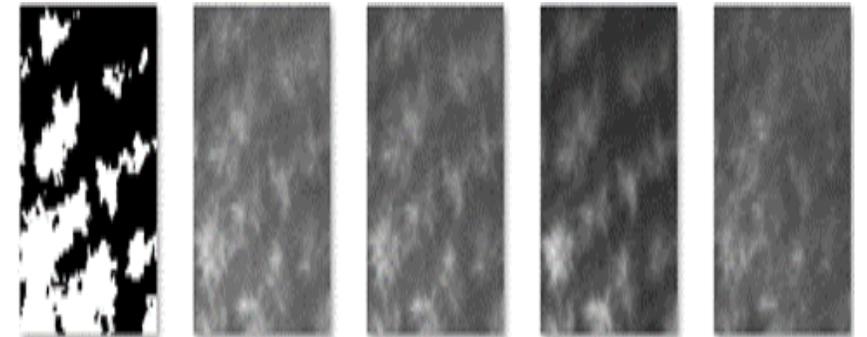


# Resultados

## Imágenes para entrenamiento (DataSet)

Para el procesamiento de la galería de imágenes adquiridas por el VANT se efectuaron varias etapas:

- La integración de los ortomosaicos de las bandas RGB, Red, Green, NIR y Red Edge con el software Pix4D Mapper.
- Mejoramiento de la resolución de las ortofotos multiespectrales.
- Se diseñó un modelo de regresión lineal multivariable para generar datos sintéticos para la banda azul a partir del ortomosaico RGB.
- El método de segmentación utilizado para generar la salida esperada del modelo, fue la umbralización por Otsu.
- Se generaron subimágenes de 256x256 píxeles de cada una de las ortofotos. El DataSet final se integró por 7650 imágenes de cada banda Red, Green, Blue, NIR y las imágenes binarias.



# Resultados

## Diseño de la red

El diseño de la red correspondió a una arquitectura tipo U-Net tradicional. Se integró en dos partes:

- Descendente utilizando operaciones de convolución mediante un kernel de 3x3, stride 1 y padding same y RELU como función de activación, además, de realizar maxpooling para extraer las características semánticas de las partes de la imagen, en esta fase se reduce la resolución.
- Ascendente, tiene como finalidad aumentar la resolución de la imagen producto del proceso anterior, para lo cual se realizaron operaciones de convolución transpuesta con concatenaciones de capas procesadas de la fase descendente del mismo nivel.

La red finaliza con una capa de salida con una operación de convolución 1x1 y una función de activación sigmoid ya que lo que se pretende es obtener una imagen binaria como salida. La implantación del modelo utilizando keras tensorflow dio como resultado los parámetros de la figura .

```
Total params: 31,032,321  
Trainable params: 31,032,321  
Non-trainable params: 0
```

# Resultados

## Diseño de la red

El método de retropropagación que permite actualizar los pesos de la red fue Adam, como algoritmo de optimización stochastic gradient descent, la función de coste o pérdida binary\_crossentropy y las métricas utilizadas para monitorear el proceso de entrenamiento son los índices de exactitud y error. Se realizaron dos modelos con diferentes configuraciones en los parámetros de número de epochs y la tasa de aprendizaje (learning rate). El segundo modelo presentó el valor de exactitud mayor con un 97% de los píxeles clasificados correctamente como cultivo de caña de azúcar.

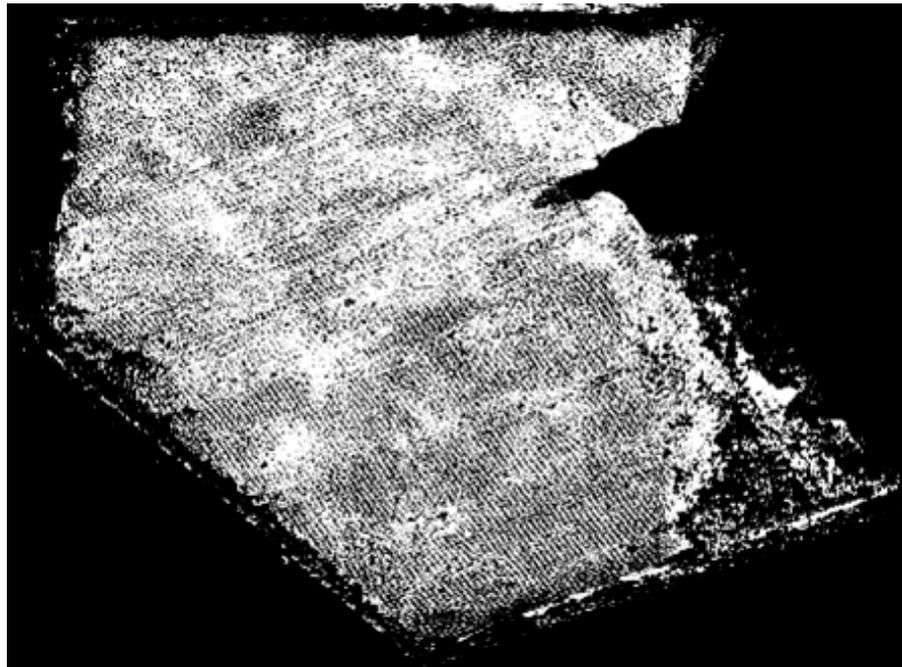
Para cuantificar la eficacia de cada modelo diseñado se utilizó una matriz de confusión y se calculó el factor de exactitud.

		Modelo: modelo_unet1.h5	Modelo: unet_cana2.h5
Parámetros de la CNN	Arquitectura	U-Net	U-Net
	Profundidad	64-128-256-512-1024-512-256-128-64	64-128-256-512-1024-512-256-128-64
	Tamaño	256x256: Normalizada	256x256: Normalizada
	Batch-size	16	16
	Epochs	100	130
	Learning Rate	.0001	Calendarizado
	Dropout	No	No
	Entrenamiento	2100	2100
	Validación	10%	10%
	Prueba	900	900
Parametros	31,031,745	31,031,745	
Desempeño de la red	Loss	0.0360	0.0089
	Accuracy	0.9877	0.9971
	Val_Loss	0.1703	0.2628
	Val_Loss	0.9633	0.9681
Matriz de confusión	VP	32283754	32482856
	VN	24760950	24893877
	FP	943026	743924
	FN	994670	861743
	Accuracy	0.9671478949652778	0.9727771843804254

# Resultados

## Prueba del modelo

La prueba final del modelo de segmentación semántica generado consistió en procesar todas las imágenes del DataSet RGB + NIR. Una vez segmentadas se integran a una sola matriz para formar la ortoimagen del cultivo de caña de azúcar en formato binario.



# Conclusiones

La red neuronal convolucional de arquitectura U-Net se implementó exitosamente como clasificador binario, asignando un valor de 1 aquellos pixeles que son parte del cultivo de caña de azúcar y el valor 0 al resto de ellos.

La configuración de los parámetros de la red tiene un papel importante en el control del sobreentrenamiento, siendo el número de epochs y el learning rate dos de los que más influyeron en el diseño del modelo.

A partir de esta aplicación se obtiene una máscara binaria que puede ser utilizada para extraer solo los valores del cultivo de por ejemplo un mapa de índices o bien evaluar con mayor exactitud el índice de producción.

# Referencias

- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1), 53. <https://doi.org/10.1186/s40537-021-00444-8>.
- Anand, T., Sinha, S., Mandal, M., Chamola, V., & Yu, F. R. (2021). AgriSegNet: Deep Aerial Semantic Segmentation Framework for IoT-Assisted Precision Agriculture. *IEEE Sensors Journal*, 21(16), 17581-17590. <https://doi.org/10.1109/JSEN.2021.3071290>.
- Anilkumar, P., & Venugopal, P. (2022). Research Contribution and Comprehensive Review towards the Semantic Segmentation of Aerial Images Using Deep Learning Techniques. *Security and Communication Networks*, 2022, 1-31. <https://doi.org/10.1155/2022/6010912>.
- Baeta, R., Nogueira, K., Menotti, D., & dos Santos, J. A. (2017). Learning Deep Features on Multiple Scales for Coffee Crop Recognition. 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 262-268. <https://doi.org/10.1109/SIBGRAPI.2017.41>.
- Das, M., & Bais, A. (2021). DeepVeg: Deep Learning Model for Segmentation of Weed, Canola, and Canola Flea Beetle Damage. *IEEE Access*, 9, 119367-119380. <https://doi.org/10.1109/ACCESS.2021.3108003>.

# Referencias

- Diez, D. S. G. (s. f.). EVALUACIÓN DEL USO DE REDES NEURONALES CONVOLUCIONALES PARA CLASIFICACIÓN DE CULTIVOS MEDIANTE IMAGEN MULTIESPECTRAL. 70.
- Ding, L., & Bruzzone, L. (2019). A Deep Architecture Based on a Two-Stage Learning for Semantic Segmentation of Large-Size Remote Sensing Images. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing .



**ECORFAN®**

© ECORFAN-Mexico, S.C.

No part of this document covered by the Federal Copyright Law may be reproduced, transmitted or used in any form or medium, whether graphic, electronic or mechanical, including but not limited to the following: Citations in articles and comments Bibliographical, compilation of radio or electronic journalistic data. For the effects of articles 13, 162,163 fraction I, 164 fraction I, 168, 169,209 fraction III and other relative of the Federal Law of Copyright. Violations: Be forced to prosecute under Mexican copyright law. The use of general descriptive names, registered names, trademarks, in this publication do not imply, uniformly in the absence of a specific statement, that such names are exempt from the relevant protector in laws and regulations of Mexico and therefore free for General use of the international scientific community. BECORFAN is part of the media of ECORFAN-Mexico, S.C., E: 94-443.F: 008- ([www.ecorfan.org/booklets](http://www.ecorfan.org/booklets))