

Mangrove Semantic Segmentation on Aerial Images

Efrén López-Jiménez , J. Anibal Arias-Aguilar , Oscar D. Ramírez-Cárdenas , J. Carlos Herrera-Lozada , and Nidiyare Hevia-Montiel 

Abstract—In the Yucatán Peninsula, there exists a diverse array of mangroves including *Rhizophora mangle*, *Avicennia germinans*, and *Laguncularia racemosa*. These mangroves play a significant role in restoring natural areas that have been damaged as a result of human activities. Furthermore, they serve as natural habitats for numerous animal and plant species. Studies have highlighted the significance of preserving and restoring these species through traditional methods. Recent advances in remote sensing and deep learning techniques have enabled the automated detection and quantification of mangroves. The application of deep neural network techniques to address computer vision challenges in the field of remote sensing is explored in this study. We focus on the detection and quantification of mangroves in remote image sensing, employing transfer learning and fine-tuning procedures with three distinct deep neural network architectures: SegNet-VGG16, U-Net, and Fully Convolutional Network (R-FCN). We applied some evaluation metrics to assess the performance of each architecture in train and test datasets, such as intersection over union (IoU), Dice coefficient, precision, sensitivity, and accuracy. In the train dataset, SegNet-VGG16 demonstrates superior precision and accuracy, while U-Net excels in IoU, Dice coefficient, and sensitivity. Conversely, R-FCN exhibits the highest sensitivity. In the test dataset, SegNet-VGG16 maintains its high precision and accuracy, while U-Net surpasses in terms of IoU and Dice coefficient, with R-FCN displaying the highest sensitivity.

Link to graphical and video abstracts, and to code: <https://latam.ieceer9.org/index.php/transactions/article/view/8557>

Index Terms—Keywords: deep neural networks, natural areas, remote perception, transfer learning, Seg-Net, U-Net.

I. INTRODUCCIÓN

En los últimos años, los manglares han adquirido una gran relevancia en las zonas costeras de la península de Yucatán debido a que son el hábitat natural de organismos que figuran en listas de protección a nivel nacional e internacional, además de tener un valor comercial reconocido [1], [2], [3]. Algunas de las causas de la degradación de los manglares están relacionadas con la explotación de recursos naturales por parte de los seres humanos, la construcción de infraestructura marítima y la sobrepoblación [4]. Para abordar esta problemática, es fundamental llevar a cabo trabajos de restauración de manglares en diversas áreas afectadas. Para

ello, es importante contar con información precisa sobre la cantidad de manglar degradado y el manglar existente, la cual puede obtenerse a través de monitoreo en el sitio de manera tradicional [5] o, más recientemente, mediante tecnologías e imágenes de percepción remota. Estas tecnologías incluyen Vehículos Aéreos No Tripulados (VANT), imágenes satelitales y técnicas de aprendizaje profundo enfocadas en segmentación semántica [6], [7], [8], [9] [10]. El proceso implica tareas de clasificación y la detección de cada especie de manglar, seguido de la cuantificación de la superficie total de manglar en una ubicación específica.

Las técnicas de segmentación semántica basadas en decodificadores tienen distintos campos de aplicación, por ejemplo, en la detección de edificios y construcciones en marcha con imágenes satelitales mediante aplicaciones en ciencias de la tierra y el acercamiento con otras áreas del conocimiento para explorar y resolver tareas de cuantificación y detección de zonas urbanas analizando además su evolución en el tiempo [11] [12] [13].

En los últimos años, las técnicas de detección y segmentación de objetos en imágenes mediante redes neuronales han mejorado significativamente los resultados en tareas de clasificación y reconocimiento de plantas, tanto de vegetación como de no vegetación. Esto ha permitido el procesamiento de imágenes aéreas y la detección de plantas específicas en las áreas de interés, incluyendo el conteo de hojas de manglares a través de percepción remota con imágenes satelitales y técnicas de visión computacional utilizando el Índice De Vegetación De Diferencia Normalizada (NDVI por sus siglas en inglés) [14] [15]. La investigación sobre algoritmos para la segmentación se centra en evaluar su efectividad, precisión, sensibilidad junto a otras métricas para validar sus resultados. Por ejemplo, en tareas de segmentación de cultivos y hojas de plantas basándose en imágenes, se han desarrollado algoritmos utilizando redes neuronales de segmentación semántica y algoritmos de agrupamiento, logrando una exactitud del 99.19%, lo que sugiere que con estos métodos es posible detectar y segmentar completamente los cultivos y las plantas [16]. A lo largo de los años se han propuesto diversas arquitecturas de redes neuronales profundas para segmentación semántica de imágenes, como SegNet [17], U-Net (basada en una arquitectura codificador-decodificador) [18] y Fully Convolutional Network (FCN) [19]. En el campo de la agricultura, estos modelos aprenden a detectar especies de plantas en imágenes sin necesidad de conocimiento explícito sobre el tipo de planta o las condiciones ambientales. Por ejemplo, U-Net se ha utilizado en tareas de segmentación semántica, mejorando su rendimiento mediante el aumento de datos y la integración de meta-datos provenientes de sistemas de información geográfica

E. López-Jiménez, J. A. Arias-Aguilar, and O. D. Ramírez Cárdenas are with Universidad Tecnológica de la Mixteca, Huajuapán de León, Mexico (e-mails: jmzefren@mixteco.utm.mx, anibal@mixteco.utm.mx and odramirez@mixteco.utm.mx).

J. C. Herrera-Lozada is with Centro de Innovación y Desarrollo Tecnológico en Cómputo, Ciudad de México, México (e-mail: jlozada@ipn.mx).

N. Hevia-Montiel is with Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Mérida, México (e-mail: nidiyare.hevia@iimas.unam.mx).

(GIS por sus siglas en inglés) obtenidos de imágenes satelitales. Asimismo, la segmentación semántica realizada con la red FCN en imágenes de color en los canales rojo, verde y azul (RGB por sus siglas en inglés) permiten extraer información sobre el color, la textura y las características geométricas de los árboles para identificar enfermedades en ellos [20]. Estos avances son especialmente relevantes en aplicaciones de agricultura de precisión, donde los sistemas de segmentación muestran un gran potencial para la detección de cultivos y las características propias que los distinguen entre sí. Para este propósito, se utilizan redes neuronales profundas, algoritmos de segmentación semántica y algoritmos de segmentación de instancias, lo que ha impulsado significativamente la capacidad de detección automática en el campo de la agricultura [21].

Por otro lado, la inspección aérea de alta resolución (a nivel de centímetros) proporciona información detallada sobre la superficie, lo que resulta beneficioso para la agricultura en términos de cuidado y vigilancia de cultivos de interés [22]. Como ejemplo, en el monitoreo del crecimiento de los cultivos se realiza la estimación de áreas con plaga en las hojas de plantas de tomate, escaneando la planta con un sensor con RGB y un canal más de profundidad, aplicando algoritmos de segmentación semántica para el muestreo de puntos uniformes [23].

A. Propuesta

En este trabajo se analiza el desempeño de tres arquitecturas de redes neuronales en la tarea de segmentación semántica. Las arquitecturas a evaluar son: SegNet-VGG16, U-Net y R-FCN. Para mejorar la precisión de la segmentación de manglares en imágenes capturadas por un VANT en la zona de restauración Hunucmá (Sisal) de la Península de Yucatán, México, emplearemos técnicas de aprendizaje por transferencia, ajuste fino y aumento de datos. El enfoque principal de este análisis se centra en la segmentación (en clases manglar y no-manglar) de los manglares, que son una especie de vegetación de particular interés en esta investigación.

II. MÉTODOS

Para llevar a cabo la tarea de segmentación semántica en imágenes aéreas, se propone la implementación del aprendizaje por transferencia en tres arquitecturas de red: SegNet-VGG16, U-Net y R-FCN, [24]. En cada una de estas arquitecturas se utilizó el mismo conjunto de datos, que consiste en imágenes que han sido etiquetadas manualmente, en las cuales está delimitada el área de interés (manglar) así como su complemento (no-manglar), generando mascarar binarias correspondientes a cada clase. Para esto se utilizó la herramienta de polígono del software LabelMe [25]. Una vez definido gráficamente el polígono, se genera un archivo con formato JSON que contiene las etiquetas que definen los vértices del polígono delimitado por el usuario y las clases definidas en él. Para nuestra tarea, los objetos de interés son los manglares contenidos en las imágenes, los cuales en el formato binario corresponden a los valores de píxel blanco (255, 255, 255) con la etiqueta manglar, y negro (0, 0, 0) para la etiqueta no-manglar, asumiendo que el agua, zonas arenosas y plantas

no vivas, se consideran como no-manglar. Las dificultades en el etiquetado lo representan variables como la resolución, la complejidad del ambiente, las sombras de las plantas y el traslape entre los manglares al momento del procesamiento de las imágenes. Los materiales empleados en este proceso incluyen un vehículo aéreo (en este caso, el modelo Mavic Pro de la empresa DJI) a través del cual se capturaron las imágenes mediante planes de vuelo diseñados con la aplicación móvil *Pix4D*. Las características clave del VANT son: una autonomía de vuelo de 27 minutos (a velocidad de viento constante), una distancia máxima de vuelo de 13 km, una cámara a bordo con una resolución de 12.35 Mega Píxeles en formato RGB y un peso de 735 gramos. Por otro lado, para llevar a cabo el proceso de entrenamiento de las redes neuronales, se dispuso de un servidor con sistema operativo Ubuntu 20.04, equipado con dos tarjetas gráficas NVIDIA RTX 2080 Ti y 72 gigabytes de memoria RAM. Para realizar la evaluación de la segmentación semántica es necesario usar métricas empleadas en tareas de clasificación y con ello validar las predicciones de la red. Por lo anterior, el objetivo es puntuar la similitud entre la segmentación predicha (predicción) y la anotada manualmente (etiqueta). Para dicha evaluación se utilizaron las métricas de Intersección sobre la Unión (IoU por sus siglas en inglés), el coeficiente Dice, la precisión, la sensibilidad y la exactitud [26]. En la Tabla I se proporcionan las ecuaciones de cada métrica con los términos clave utilizados y sus siglas en inglés, tales como, verdaderos positivos (TP) representando el número de píxeles de la clase manglar que se han clasificado correctamente como manglar. Los verdaderos negativos (TN) representando el número de píxeles de la clase no-manglar que se clasifican erróneamente como manglar. Los falsos positivos (FP) representando la cantidad de píxeles de la clase manglar que se clasifican erróneamente como no-manglar y los falsos negativos (FN) representando el número de píxeles de la clase no-manglar que se han clasificado correctamente como no-manglar.

A. Estudio de Caso

El área de cobertura se encuentra en el municipio de Hunucmá, Sisal, Yucatán (coordenadas: (21.16138, -90.03674), la cual cuenta con una extensión de 7,395 hectáreas de manglares, como se muestra en la Fig. 1. Además, en dicha superficie se concentra una de las principales áreas de manglares en restauración [27].

B. Descripción del Conjunto de Datos

El conjunto de datos para el entrenamiento, validación y prueba de las redes neuronales está conformado por imágenes obtenidas mediante planes de vuelo del VANT usando el planificador de vuelos *Pix4D*. Los vuelos se realizaron siempre y cuando la velocidad del viento no superara los 10 km/h. El tamaño de las imágenes que se obtienen con el VANT es de 4000×3000 píxeles cada una, que para fines de procesamiento en las tarjetas gráficas, fueron ajustadas de tamaño a 256×192 píxeles. El conjunto de datos para el proceso de entrenamiento y prueba de cada arquitectura de red neuronal fueron un cúmulo de 3400 imágenes. El conjunto de datos fue dividido en tres

TABLA I
MÉTRICAS DE EVALUACIÓN DE LAS ARQUITECTURAS DE RED NEURONAL. EN ESTE TRABAJO, LAS MÉTRICAS MIDEN LAS DIFERENCIAS ESTADÍSTICAS ENTRE LAS CLASES MANGLAR Y NO-MANGLAR OBTENIDAS POR LOS MODELOS DE RED

| Métrica | Ecuación |
|-------------------|---|
| IoU | $\frac{\text{Intersección}}{\text{Unión}}$ |
| Coefficiente Dice | $\frac{2 \times TP}{2 \times TP + FP + FN}$ |
| Precisión | $\frac{TP}{TP + FP}$ |
| Sensibilidad | $\frac{TP}{TP + FN}$ |
| Exactitud | $\frac{TP + TN}{TP + FN + TN + FP}$ |



Fig. 1. La zona de cobertura para el estudio y obtención de datos mediante el uso del VANT se realizó en el municipio de Hunucmá (Sisal), Yucatán, México, en un área de alrededor de 3 km² de extensión territorial (Fuente: Conabio).

partes: un conjunto de entrenamiento con 2720 imágenes que representa el 80 % del total, un conjunto de validación con 340 imágenes que representa el 10 % y un conjunto de prueba con 340 imágenes que representa el 10 % del total del conjunto de datos. Revisando el color, la corteza y la textura registradas en las imágenes, asumimos que en el área fotografiada predomina el *Rhizophora mangle* [28], que es la especie que concentra la mayor cantidad de área en la zona de Hunucmá, Sisal, Yucatán. En resumen, las imágenes aéreas RGB fueron registradas por un VANT, capturadas a diferentes horas del día, evitando la menor sombra posible en las imágenes, las cuales contienen manglares, zonas arenosas, y agua.

III. SEGMENTACIÓN SEMÁNTICA CON LA RED SEGNET-VGG16

SegNet-VGG16 [29] es una arquitectura de red diseñada para realizar segmentación semántica que está constituida por un codificador y su decodificador correspondiente. Fue entrenada inicialmente con el conjunto de datos *Camvid*, que tiene once categorías definidas [30]. Esta red consiste de 13 capas convolucionales organizadas en una estructura tipo codificador-decodificador. Cada capa del codificador realiza una convolución de 3×3 , seguida de una normalización por lotes y utilizando como función de activación la Unidad de Rectificación Lineal (ReLU por sus siglas en inglés). Cada capa del codificador repite entonces esta combinación: convolución, normalización por lotes y función de activación ReLU [31]. Al final de cada bloque de codificación se realiza una operación de agrupamiento *max-pooling* de tamaño 2×2 con un paso de recorrido o *stride* de 2. El procedimiento de agrupamiento *max-pooling* consiste en recibir un volumen convolucional como entrada y transformarlo en un mapa de características más compacto tomando en cuenta sólo los píxeles más significativos del volumen de entrada. El codificador realiza una serie de convoluciones y agrupamientos, mientras que el decodificador realiza una serie de deconvoluciones y escalamientos cuya salida es una imagen de K canales, donde K es el número de clases segmentadas, como se muestra en la Fig. 2. La capa de salida utiliza funciones de activación *soft-max* [32].

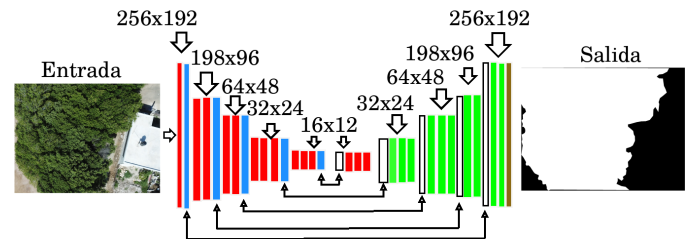


Fig. 2. Arquitectura de la red neuronal SegNet-VGG16.

IV. SEGMENTACIÓN SEMÁNTICA CON LA RED U-NET

U-Net [33] es una arquitectura de red neuronal con etapas codificador-decodificador concatenadas en forma de U, donde el codificador consiste de repetidas implementaciones de dos convoluciones de 3×3 , seguidas con funciones de activación ReLU y una operación de agrupamiento o *max-pooling* de 2×2 con un valor de paso o *stride* de 2 para realizar un muestreo descendente o *down-sampling*. Además, contiene conexiones de salto que concatenan el codificador y el mapa de características del decodificador que ayudan al flujo inverso de gradientes para mejorar el resultado de la salida del decodificador. Cada etapa del codificador duplica en profundidad el número de canales de características. El decodificador implementa un muestreo ascendente o *up-sampling* de 2×2 del mapa de características, seguido de deconvoluciones y una función de activación ReLU. El decodificador obtiene características complejas de la entrada mientras la información de la localización de los píxeles se recupera del codificador. En

cada etapa del decodificador la profundidad de los canales de características se reduce a la mitad, mientras su ancho y largo se duplican. En la etapa final del decodificador, cada vector de características de 64-dimensiones se conecta a una capa de salida usando una convolución de 1×1 , como se muestra en la Fig. 3. A lo largo de su uso han surgido distintas variantes de esta topología, con enfoques distintos para resolver tareas de segmentación semántica.

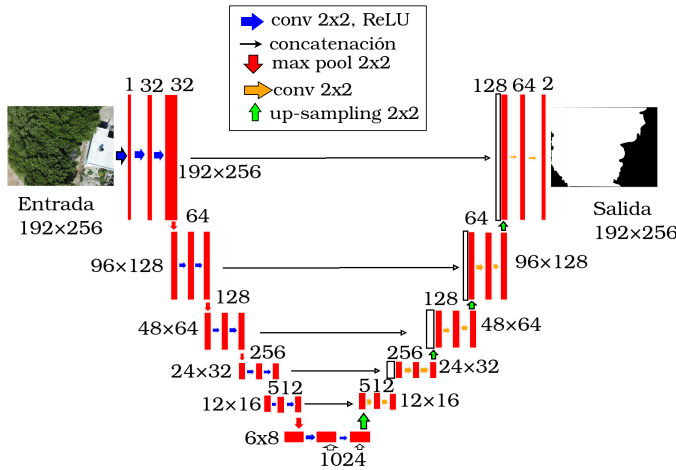


Fig. 3. Arquitectura de la red neuronal U-Net.

V. SEGMENTACIÓN SEMÁNTICA CON LA RED FULLY CONVOLUTIONAL NETWORK

La Fully Convolutional Network (R-FCN) [19] es una arquitectura de red neuronal convolucional profunda enfocada en tareas de segmentación semántica. Está basada en capas pre-entrenadas de la red VGG-16 que son utilizadas como extractores de características (contiene un total de cinco bloques convolucionales terminados con un *max-pooling* de 3×2). Para conseguir que la entrada y salida sean del mismo tamaño se agregan capas para concatenar y deconvolucionar los mapas de características y generar así predicciones a nivel de píxel, como se muestra en la Fig. 4.

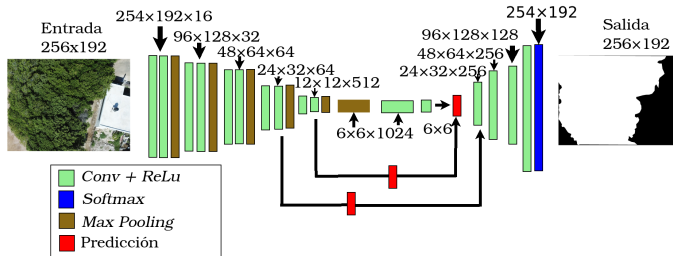


Fig. 4. Arquitectura de la red neuronal R-FCN.

VI. ENTRENAMIENTO Y EVALUACIÓN DE LAS ARQUITECTURAS DE REDES NEURONALES

Las redes neuronales profundas basan su funcionamiento en entrenamientos con grandes cantidades de datos. En nuestra tarea de segmentación hemos implementado distintas técnicas

para mejorar su desempeño, tales como: aprendizaje por transferencia, ajuste fino y aumento de datos. Hay que decir que, para ciertos problemas reales, los datos no son tan abundantes, por lo que el aprendizaje por transferencia y el ajuste fino son una alternativa para transferir el conocimiento de una tarea relacionada a otra similar [34], [13]. Se ha utilizado entonces la técnica de aumento de datos durante el proceso de entrenamiento, la cual consiste en aplicar, de manera aleatoria, transformaciones horizontales, transposición, rotaciones y modificaciones en los niveles de brillo de las imágenes, entre otros cambios. Para la red SegNet, los hiperparámetros con los que se obtuvieron los mejores resultados fueron: un tamaño de lote de 16, una tasa de aprendizaje de 0.003 y el optimizador Adam [35] con la función de costo de entropía cruzada [36], con 120 épocas de iteración. En el segundo caso, la arquitectura U-Net, se usó como función de costo la entropía cruzada y una función de activación sigmoide binaria en la capa de salida. Los demás hiper-parámetros con los que se obtuvieron los mejores resultados fueron: un tamaño de lote de 18, 120 épocas de iteración, el optimizador Adam con una tasa de aprendizaje de 0.03 y un coeficiente de *dropout* de 0.5. En el tercer experimento se utiliza el modelo pre-entrenado R-FCN en la etapa de la extracción de características. Este modelo se re-entrena con nuestro conjunto de datos de manglares como parte del ajuste fino. Los hiper-parámetros con los que se obtuvieron los mejores resultados fueron: un tamaño de lote de 18 y el optimizador Adam con tasa de aprendizaje de 0.0003. Para evaluar los resultados de los modelos de red neuronal se implementaron dos enfoques diferentes, el primero calcula seis métricas de evaluación y con ello obtiene el rendimiento de manera cuantitativa de SegNet-VGG16, U-Net y R-FCN para cada métrica. El segundo enfoque fue evaluar los resultados de manera visual y comparar la segmentación obtenida a la salida de cada red.

VII. RESULTADOS Y DISCUSIÓN

Los resultados cuantitativos de las evaluaciones de las tres arquitecturas se presentan en la Tabla II, donde se muestran las comparativas entre cada red utilizando el conjunto de datos de entrenamiento. U-Net destaca con las métricas más altas en *IoU*, coeficiente Dice y exactitud. Por otro lado, R-FCN registra la métrica de evaluación más alta en sensibilidad, mientras que SegNet-VGG16 lidera en precisión. En cuanto

TABLA II
EVALUACIÓN DEL CONJUNTO DE ENTRENAMIENTO

| Métrica | SegNet-VGG16 | U-Net | R-FCN |
|-----------------|--------------|--------|--------|
| <i>IoU</i> | 97.28 | 97.54* | 91.3 |
| Coficiente Dice | 90.44 | 94.92* | 81.15 |
| Precisión | 99.08* | 94.98 | 75.38 |
| Sensibilidad | 91.13 | 95.84 | 98.37* |
| Exactitud | 95.43 | 97.49* | 91.58 |

*Mejor desempeño

a las métricas de evaluación en el conjunto de prueba, U-Net logra las métricas más altas en *IoU* y coeficiente Dice. Por otro lado, SegNet-VGG16 muestra un rendimiento superior en precisión y exactitud, mientras que R-FCN obtiene la métrica

de evaluación más alta en sensibilidad, como se detalla en la Tabla III.

TABLA III
EVALUACIÓN EN EL CONJUNTO DE PRUEBA

| Métrica | SegNet-VGG16 | U-Net | R-FCN |
|-------------------|--------------|--------|--------|
| IoU | 96.96 | 96.97* | 92.22 |
| Coefficiente Dice | 92.20 | 94.92* | 81.45 |
| Precisión | 98.03* | 96.24 | 75.79 |
| Sensibilidad | 90.29 | 96.81 | 98.25* |
| Exactitud | 97.03* | 92.75 | 91.58 |

*Mejor desempeño

La segmentación resultante de la red SegNet-VGG16 se muestra en la Fig. 5, la cual corresponde a la etapa de prueba, en la que las imágenes procesadas no han sido previamente conocidas por la red. En la primera columna de la figura se presentan las imágenes originales, en la columna del medio se muestran las máscaras de segmentación o etiquetas, y en la tercera columna se representa la segmentación predicha por la red. Al observar las imágenes, se aprecia que las máscaras no logran delimitar con precisión los bordes entre las áreas de manglar y las áreas que no corresponden a la zona de manglares, aunque se aprecia un intento de aproximación. En la Fig. 6, se presentan resultados visuales obtenidos

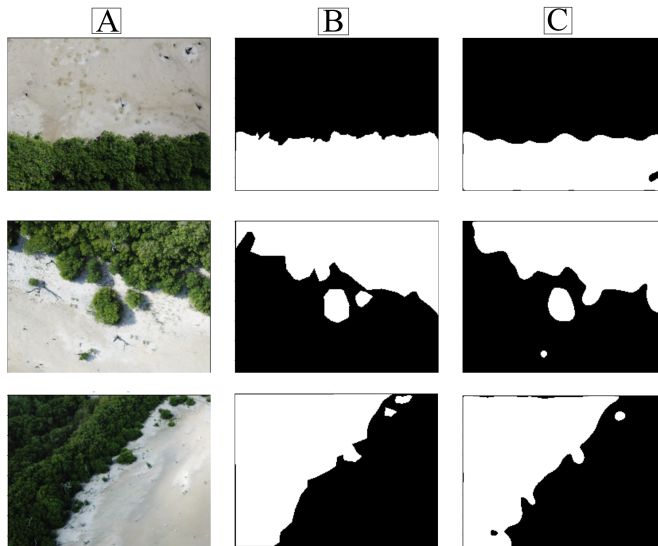


Fig. 5. Resultados en la etapa de prueba con la red SegNet-VGG16, en la columna (A) se muestran ejemplos aleatorios de las imágenes originales, en la columna (B) sus respectivas máscaras de segmentación y en la columna (C) las predicciones hechas por la red.

con la red U-Net. La primera columna está compuesta por las imágenes originales en formato RGB, mientras que la columna del centro muestra las máscaras de segmentación etiquetadas manualmente. La tercera columna está compuesta por imágenes binarias que fueron predichas por la red. Al observar las imágenes de la predicción, se aprecia que los bordes correspondientes a las áreas de manglar están mejor delineados en comparación con SegNet-VGG16 y R-FCN. En la Fig. 7, se presentan resultados del conjunto de prueba utilizando la red R-FCN. La primera columna muestra las

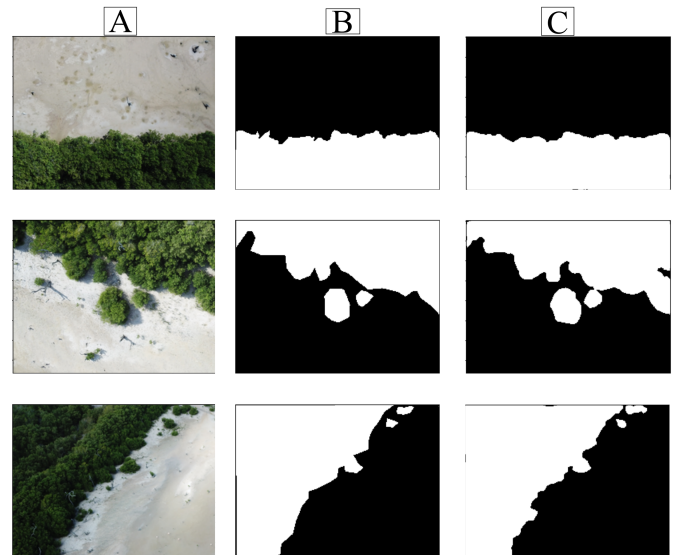


Fig. 6. Resultados del conjunto de datos de prueba para U-Net. En la columna de la izquierda se muestran ejemplos de imágenes RGB, en la columna de en medio se muestran sus respectivas máscaras o etiquetas y la columna de la derecha presenta las predicciones de la red.

imágenes originales, la columna del centro exhibe las máscaras que fueron etiquetadas manualmente y la tercera columna contiene las imágenes con las predicciones realizadas por la red. Al evaluar visualmente las predicciones, se observa que los bordes no se encuentran tan bien definidos en comparación con las máscaras etiquetadas. Desde este punto de vista, podemos concluir que U-Net y SegNet-VGG16 proporcionan aproximaciones más precisas a las máscaras de etiquetado. En

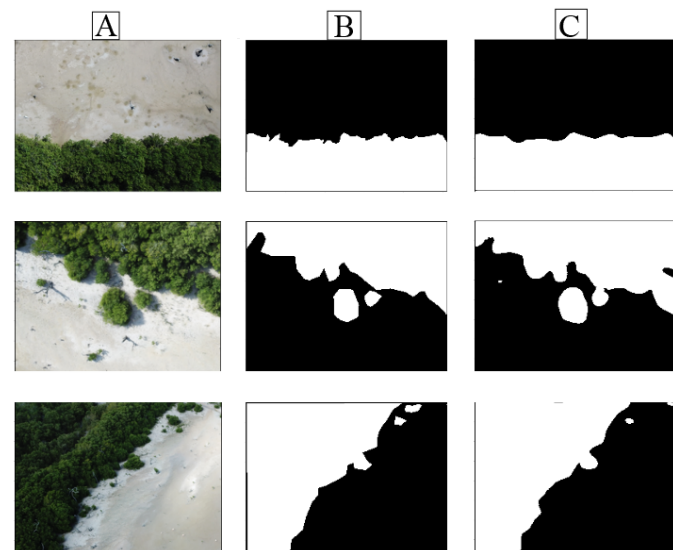


Fig. 7. Resultados del conjunto de prueba para la red R-FCN. La columna de la izquierda presenta ejemplos de imágenes RGB, la columna de en medio muestra sus respectivas máscaras y en la columna de la derecha se presentan las predicciones de la red, donde (255, 255, 255) corresponde a la clase manglar y (0, 0, 0) a no-manglar

la Fig. 8, se presentan las gráficas que muestran los valores

de la función de costo y de la métrica de exactitud durante el proceso de entrenamiento y validación de la arquitectura de SegNet-VGG16. Durante el entrenamiento, la métrica de exactitud alcanzó un 95.43 %, y en la etapa de validación, se logró un 97.03 %. En cuanto a la función de costo, después de 120 épocas de iteración, se registró un error del 2.26 % en el entrenamiento y un 8.08 % en la validación. En la Fig.

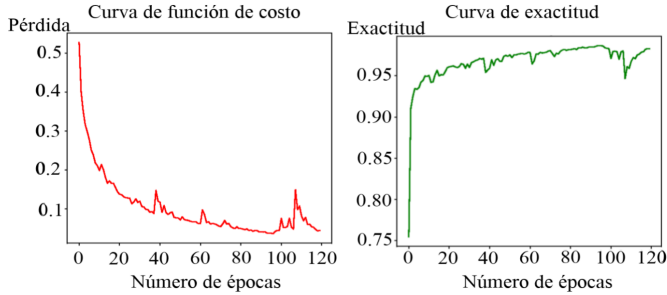


Fig. 8. En esta figura se muestran las gráficas de los valores de la función de costo y de la exactitud en la etapa de entrenamiento de la red SegNet-VGG16.

9, se presentan las curvas correspondientes a la red U-Net durante la etapa de entrenamiento. Después de 120 épocas se alcanzó una Exactitud del 97.67 %, mientras que en el conjunto de validación, se obtuvo un 96.81 %. El valor de la función de costo en la etapa de entrenamiento, fue del 2.43 %, y en la etapa de validación se mantuvo en un 2.43 %. La Fig. 10

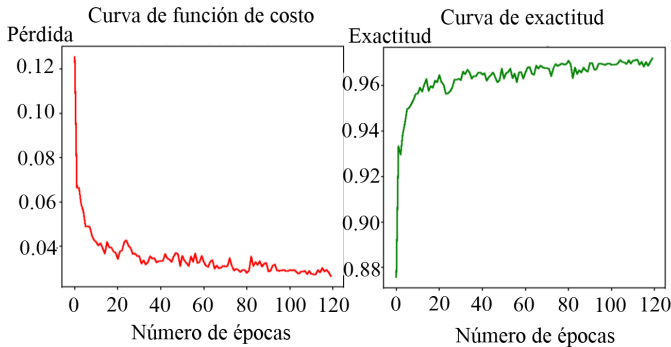


Fig. 9. Gráficas de los valores de la función de costo y de la Exactitud en la etapa de entrenamiento de la red U-Net.

muestra los valores de la exactitud y de la Función de costo en las etapas de entrenamiento y validación de la arquitectura R-FCN. Durante la etapa de entrenamiento, la exactitud alcanzó un máximo del 97.49 %, mientras que en la etapa de validación llegó hasta el 96.81 % después de 120 iteraciones. En cuanto al valor de la función de costo, se registró un 2.33 % en la etapa de entrenamiento y un 3.04 % en la etapa de validación.

Los resultados obtenidos a partir de cada enfoque de aprendizaje profundo pueden explicarse en gran medida por los datos utilizados, ya que la tarea de segmentación se limitó a dos clases: manglar y no-manglar. El desafío futuro radica en la obtención de un mapa completo y geo-referenciado de toda el área de cobertura, que incluya información sobre el porcentaje total de cobertura de manglar. Además, se busca determinar

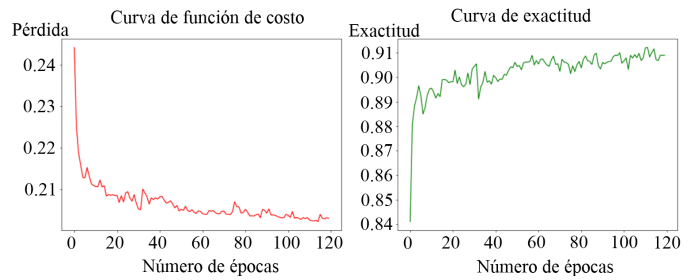


Fig. 10. Gráficas de los valores de la función de costo y de la exactitud en la etapa de entrenamiento de la red R-FCN.

la cantidad de manglar en función de su morfología, lo que proporcionaría un análisis más detallado de la vegetación.

Una vez que se haya generado este mapa (uniendo inteligentemente pequeñas imágenes capturadas con al menos 75 % de porcentaje de sobre-posición), se podrá llevar a cabo una comparación con las salidas de programas comerciales de procesamiento de imágenes aéreas, como *Pix4D*, con el fin de evaluar la eficacia de la propuesta en relación con soluciones establecidas en el mercado. Sin embargo, es importante señalar que las predicciones de las arquitecturas de redes neuronales actualmente generan imágenes binarias, lo que limita su capacidad para proporcionar información detallada sobre la vegetación.

Un ejemplo de enfoque similar se encuentra en el trabajo de [37], que emplearon técnicas de aprendizaje por transferencia en la segmentación de cultivos de semilla de aceite. En su evaluación destacaron tres enfoques diferentes de aprendizaje por transferencia, incluyendo el entrenamiento previo de VGG16, una versión modificada de SegNet y un clasificador de aprendizaje profundo con pesos entrenados previamente. Realizaron experimentos con y sin aumento de datos, y observaron un mejor rendimiento en los casos que incluyeron aumento de datos. En contraste, el trabajo de [38] se centra en la percepción remota utilizando datos satelitales combinados con técnicas de sistemas de información geográfica, para detectar manglares y monitorizar sus cambios a lo largo del tiempo. Los resultados de estas propuestas permiten observar las transformaciones temporales del área de estudio, lo que a su vez facilita la planificación, gestión y regulación de las zonas costeras.

VIII. CONCLUSIONES

El artículo presenta resultados del desempeño, con métricas de evaluación cuantitativas y una evaluación visual, de la salida de tres arquitecturas de redes neuronales entrenadas previamente para llevar a cabo la segmentación de manglares en imágenes aéreas de la región de la Península de Yucatán. Mediante el uso de técnicas de aprendizaje por transferencia, ajuste fino y aumento de datos, se han mejorado significativamente las métricas de evaluación en comparación con los modelos de red originales. Al comparar los tres enfoques de aprendizaje por transferencia en el conjunto de prueba, se observó que SegNet-VGG16 obtuvo el mejor desempeño en precisión (98.03 %) y exactitud (97.03 %), mientras que U-Net lideró en IoU (96.97 %), coeficiente Dice (92.20 %) y R-FCN

en sensibilidad (98.25 %). Los experimentos visuales realizados para cada red mostraron que la arquitectura U-Net destacó en la delimitación precisa de los bordes entre los manglares y lo que no corresponde a la clase manglar en la salida de la red neuronal. Estas técnicas han destacado la importancia de la segmentación semántica como una alternativa para monitorear la salud y la distribución de los manglares. Sin embargo, aún quedan retos por superar, como la mejora de la precisión en la detección de bordes y la adaptación a diferentes condiciones de iluminación y cobertura vegetal.

El futuro de esta investigación se enfoca en la reconstrucción completa de mapas, considerando otras condiciones, como la presencia de diferentes tipos de vegetación y de suelo. Además, se explorará la generación de orto-mosaicos mediante el uso de imágenes satelitales, complementando la información obtenida a partir de las imágenes capturadas por un Vehículo Aéreo No Tripulado (VANT).

AGRADECIMIENTOS

Los autores agradecen al CONAHCYT por la beca mixta otorgada para realizar una estancia en la UNAM campus Mérida Yucatán. Se agradece también el apoyo de la Dra. Claudia Teutli (UNAM) y el Dr. Jorge Herrera (CINVESTAV) por el apoyo en el acceso a los sitios de manglares en restauración.

REFERENCIAS

- [1] A. Zaldívar-Jiménez, J. A. Herrera-Silveira, R. Pérez-Ceballos, and C. Teutli-Hernández, "Evaluación del uso de los humedales de manglar como biofiltro de efluentes de camarónicas en Yucatán, México," *Revista De Biología Marina Y Oceanografía*, vol. 47, pp. 395–405, 12 2012. doi: 10.4067/s0718-19572012000300003.
- [2] M. T. R. Zúñiga, J. A. Velázquez, C. Galindo-Leal, S. Cerdeira-Estrada, M. I. C. López, J. D. Gallegos, R. J. Rosenberg, J. D. M. Mendoza, R. Ressler, C. T. Souza, A. U. Martínez, L. Valderrama-Landeros, B. V. Balderas, A. D. V. Lule, and S. V. Salazar, *Manglares de México : Extensión, distribución y monitoreo*. 1 2013. doi: 10.5962/bhl.title.111178.
- [3] J. L. Portillo and E. Ezcurra, "Los manglares de México: una revisión," *Madera Y Bosques*, vol. 8, pp. 27–51, 9 2016. doi: 10.21829/myb.2002.801290.
- [4] S. Cinco-Castro and J. A. Herrera-Silveira, "Vulnerability of mangrove ecosystems to climate change effects: The case of the Yucatan Peninsula," *Ocean & Coastal Management*, vol. 192, p. 105196, 7 2020. doi: 10.1016/j.ocecoaman.2020.105196.
- [5] M. A. Zaldívar-Jiménez, J. A. Herrera-Silveira, C. Teutli-Hernández, F. A. Comín, J. L. Andrade, C. C. Molina, and R. Pérez-Ceballos, "Conceptual framework for mangrove restoration in the Yucatan Peninsula," *Ecological restoration, North America*, vol. 28, pp. 333–342, 8 2010. doi: 10.3368/er.28.3.333.
- [6] M. Mohan, G. Richardson, G. Gopan, M. M. Aghai, S. Bajaj, G. P. Galmuwa, M. Vastaranta, P. S. P. Arachchige, L. Amorós, A. P. D. Córte, S. De-Miguel, R. V. Leite, M. K. Ganyago, E. N. Broadbent, W. Doaemo, M. A. B. Shorab, and A. Cardil, "UAV-Supported Forest regeneration: Current trends, challenges and implications," *Remote Sensing*, vol. 13, p. 2596, 7 2021. doi: 10.3390/rs13132596.
- [7] A. S. Ayub, A. Anggoro, A. H. Lukman, A. Ariasari, A. N. N. Suci, N. T. Agustini, F. Nugroho, A. M. Muslih, C. C. Hanami, and R. Zuhendri, "Mapping The Potential of Mangrove Planting in The Rehabilitation of Coastal Ecosystems Using Drone Technology," *Journal of Sylva Indonesiana*, vol. 6, pp. 164–177, 8 2023. doi: 10.32734/jsi.v6i02.10515.
- [8] D. Yin and L. Wang, "Individual mangrove tree measurement using UAV-based LIDAR data: Possibilities and challenges," *Remote Sensing of Environment*, vol. 223, pp. 34–49, 3 2019. doi: 10.1016/j.rse.2018.12.034.
- [9] A. C. C. Viodor, C. J. G. Aliac, and L. T. Santos-Feliscuzo, "Mangrove species identification using deep neural network," in *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, pp. 1–6, 2022. doi: 10.1109/ICITISEE57756.2022.10057793.
- [10] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *Isprs Journal of Photogrammetry and Remote Sensing*, vol. 152, pp. 166–177, 6 2019. doi: 10.1016/j.isprsjprs.2019.04.015.
- [11] Y. Yu, C. Wang, Q. Fu, R. Kou, F. Huang, B. Yang, T. Yang, and M. Gao, "Techniques and Challenges of Image Segmentation: A review," *Electronics*, vol. 12, p. 1199, 3 2023. doi: 10.3390/electronics12051199.
- [12] U. Sehar and M. L. Naseem, "How deep learning is empowering semantic segmentation," *Multimedia Tools and Applications*, vol. 81, pp. 30519–30544, 4 2022. doi: 10.1007/s11042-022-12821-3.
- [13] Y. Guo, H. Shi, A. Kumar, K. Grauman, T. Rosing, and R. Feris, "SpotTune: Transfer Learning Through Adaptive Fine-Tuning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6 2019. doi: 10.1109/cvpr.2019.00494.
- [14] S. Malek, Y. Bazi, N. Alajlan, H. Alhichri, and F. Melgani, "Efficient framework for palm tree detection in UAV images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 4692–4703, 12 2014. doi: 10.1109/jstars.2014.2331425.
- [15] N. A. Binh, L. T. Hauser, P. V. Hoa, G. T. P. Thao, N. N. An, H. S. Nhut, T. A. Phuong, and J. Verrelst, "Quantifying mangrove leaf area index from Sentinel-2 imagery using hybrid models and active learning," *International Journal of Remote Sensing*, vol. 43, pp. 5636–5657, 3 2022. doi: 10.1080/01431161.2021.2024912.
- [16] S. G. Sodjinou, V. Mohammadi, A. T. S. Mahama, and P. Gouton, "A deep semantic segmentation-based algorithm to segment crops and weeds in agronomic color images," *Information Processing in Agriculture*, vol. 9, pp. 355–364, 9 2022. doi: 10.1016/j.inpa.2021.08.003.
- [17] S. Kolhar and J. Jagtap, "Convolutional neural network based encoder-decoder architectures for semantic segmentation of plants," *Ecological Informatics*, vol. 64, p. 101373, 9 2021. doi: 10.1016/j.ecoinf.2021.101373.
- [18] W. Li, C. He, J. Fang, J. Zheng, H. Fu, and L. Yu, "Semantic Segmentation-Based building footprint extraction using very High-Resolution satellite images and Multi-Source GIS data," *Remote Sensing*, vol. 11, p. 403, 2 2019. doi: 10.3390/rs11040403.
- [19] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 640–651, 4 2017. doi: 10.1109/tpami.2016.2572683.
- [20] R. Saleem, J. H. Shah, M. Sharif, and G. J. Ansari, "Mango leaf disease identification using fully resolution convolutional network," *Computers, materials & continua*, vol. 69, pp. 3581–3601, 1 2021. doi: 10.32604/cmc.2021.017700.
- [21] J. Fuentes-Pacheco, J. Torres-Olivares, E. Román-Rangel, S. Cervantes, P. Juárez-López, J. H. Valadez, and J. M. Rendón-Mancha, "Fig Plant Segmentation from Aerial Images Using a Deep Convolutional Encoder-Decoder Network," *Remote Sensing*, vol. 11, p. 1157, 5 2019. doi: 10.3390/rs11101157.
- [22] T. Anand, S. Sinha, M. Mandal, V. Chamola, and F. R. Yu, "AGRISEG-NET: Deep Aerial Semantic Segmentation Framework for IoT-Assisted Precision Agriculture," *IEEE Sensors Journal*, vol. 21, pp. 17581–17590, 8 2021. doi: 10.1109/jsen.2021.3071290.
- [23] T. Masuda, "Leaf Area Estimation by Semantic Segmentation of Point Cloud of Tomato Plants," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10 2021. doi: 10.1109/iccvw54120.2021.00159.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6 2016. doi: 10.1109/cvpr.2016.90.
- [25] B. Russell, A. Torralba, K. Murphy, and W. T. Freeman, "LabelMe: A Database and Web-Based Tool for Image Annotation," *International Journal of Computer Vision*, vol. 77, pp. 157–173, 10 2007. doi: https://doi.org/10.1007/s11263-007-0090-8.
- [26] D. Müller, I. Soto-Rey, and F. Krämer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, p. 6 2022. doi: https://doi.org/10.1186/s13104-022-06096-y.
- [27] J. A. Herrera-Silveira, "Overview and characterization of the hydrology and primary producer communities of selected coastal lagoons of Yucatán, México," *Aquatic Ecosystem Health & Management*, vol. 1, pp. 353–372, 12 1998. doi: 10.1016/s1463-4988(98)00014-1.

- [28] A. M. Gill and P. B. Tomlinson, "Studies on the growth of red mangrove (*Rhizophora mangle* L.) 3. Phenology of the shoot," *Biotropica*, vol. 3, p. 109, 12 1971. doi: 10.2307/2989815.
- [29] V. Badrinarayanan and R. Cipolla, "SEgNet: a deep convolutional Encoder-Decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 2481–2495, 12 2017. doi: 10.1109/tpami.2016.2644615.
- [30] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, *Segmentation and recognition using structure from motion point clouds*. 1 2008. doi: 10.1007/978-3-540-88682-2_5.
- [31] F. Millstein, *Convolutional Neural Networks In Python: Beginner's Guide To Convolutional Neural Networks In Python*. North Charleston, SC, USA: CreateSpace Independent Publishing Platform, 2018.
- [32] X. Qi, T. Wang, and J. Liu, "Comparison of Support Vector machine and Softmax classifiers in Computer Vision," *Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, 12 2017. doi:https://doi.org/10.1109/icmccce.2017.49.
- [33] O. Ronneberger, P. Fischer, and T. Brox, "U-NET: Convolutional Networks for Biomedical Image Segmentation," *arXiv (Cornell University)*, 5 2015. doi: 10.48550/arxiv.1505.04597.
- [34] L. Shao, F. Zhu, and X. Li, "Transfer Learning for Visual Categorization: A Survey," *IEEE transactions on neural networks and learning systems*, vol. 26, pp. 1019–1034, 5 2015. doi:https://doi.org/10.1109/tnnls.2014.2330900.
- [35] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv (Cornell University)*, 12 2014. doi:https://arxiv.org/abs/1412.6980.
- [36] U. R. DrA, "Binary cross entropy with deep learning technique for Image classification," *International journal of advanced trends in computer science and engineering*, vol. 9, pp. 5393–5397, 8 2020. doi:https://doi.org/10.30534/ijatcse/2020/175942020.
- [37] A. Abdalla, H. Cen, L. Wan, R. B. Rashid, H. Weng, W. Zhou, and Y. 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*, vol. 167, p. 105091, 12 2019. doi: 10.1016/j.compag.2019.105091.
- [38] V. T. Thi, A. Xuan, H. Nguyen, F. Dahdouh-Guebas, and N. Koedam, "Application of remote sensing and GIS for detection of long-term mangrove shoreline changes in Mui Ca Mau, Vietnam," *Biogeosciences*, vol. 11, pp. 3781–3795, 7 2014. doi: 10.5194/bg-11-3781-2014.



Nidiyare Hevia Montiel Researcher at the Academic Unit of the Institute for Research in Applied Mathematics and Systems (IIMAS) of the state of Yucatan, belonging to the Universidad Autónoma de México (UNAM). She obtained her PhD at the University of Orsay - Paris XI in France. Her research areas are Image Processing, Computer Vision and Deep Learning with biomedical, biological and environmental applications.



Juan Carlos Herrera Lozada He is a Communications and Electronics Engineer, graduated from the Escuela Superior de Ingeniería Mecánica y Eléctrica of the Instituto Politécnico Nacional in Mexico City in 1996. He obtained a Master's degree in Computer Engineering with a specialization in Digital Systems in 2002 and a PhD degree in Computer Science in 2011, both from the Computer Research Center of the Instituto Politécnico Nacional in Mexico City. He is currently a member of the National System of Researchers. His general areas of interest are intelligent computing and embedded systems.



Efrén López Jiménez received the B.S. degree in Electronics engineering at Technology Institute Puebla and his master's degree in computing technology from Instituto Politecnico Nacional, Mexico, in 2015. He is currently pursuing the Doctorate of Robotics at Universidad Tecnológica de la Mixteca, México. His research interest are artificial intelligence applications and autonomous navigation.



Jose Anibal Arias-Aguilar received his PhD in Computer Science of Image and Language at Paul Sabatier University (Toulouse, France). He is currently professor-researcher at Universidad Tecnológica de la Mixteca and teaches courses in the master's degrees in Robotics, Artificial Intelligence and Interactive Media, as well as in the doctorates in Artificial Intelligence, Electronics and Robotics.



Oscar D. Ramírez-Cárdenas received his M.Sc. degree in electronics from the Universidad Tecnológica de la Mixteca, Oaxaca, México, in 2015. In 2020, he earned a PhD degree in Robotics from the same institution. He is currently professor-researcher at Universidad Tecnológica de la Mixteca. His research interests encompass theoretical and practical aspects of feedback regulation in linear and nonlinear dynamic systems, with a particular focus on backstepping control techniques. His expertise extends to applications in mobile robotics and multi-agent systems.