

Detección de cambio en superficie costera mediante la segmentación de imágenes aéreas utilizando redes neuronales convolucionales

Coastal surface change detection by aerial image segmentation using convolutional neural networks

J.J. Padilla-Arballo ^{a,*}, S. Martínez-Díaz ^a, M.A. Castro-Liera ^a, J.E. Luna-Taylor ^a

^a Tecnológico Nacional de México, 23080, La Paz, Baja California Sur, México

Resumen

La conservación de ecosistemas en zonas costeras debe contemplarse dentro de un marco sustentable con las actividades antropogénicas. Es relevante cuantificar el impacto que generan agentes externos, por lo que el objetivo de este trabajo es implementar un método para el monitoreo de la degradación en superficie de playa y vegetación litoral adyacente. Fueron recabadas capturas de imágenes aéreas de zonas costeras protegidas, obtenidas periódicamente mediante un vehículo drone. Se integró un dataset que contempla todas las fases estacionales y distingue 5 clases para su monitoreo: mangle, vegetación rastrera, arena, mar y cerro-planicie. Para la segmentación semántica se implementaron y compararon distintas arquitecturas de redes neuronales convolucionales (CNN) empleando aprendizaje transferido. Los resultados han sido robustos en la clasificación, alcanzando una precisión global del 93.9% y entre 89.9-95.8% en las clases individuales. En la métrica Intersección sobre Unión, IoU, el rango fue entre 86.6-92.7%. En la detección de cambio son utilizadas series temporales para el monitoreo de clases. Este método ha sido aplicado al caso de estudio de la playa Ensenada Grande en el Complejo Insular del Espíritu Santo.

Palabras Clave: Redes neuronales convolucionales, Segmentación semántica, Series temporales, Zonas costeras protegidas

Abstract

The conservation of ecosystems in coastal areas must be considered within a sustainable framework with anthropogenic activities. It is relevant to quantify the impact generated by external agents, hence the objective of this work is to implement a method for monitoring the degradation of the beach surface and adjacent coastal vegetation. Captures of aerial images of protected coastal areas have been collected, obtained periodically by means of a drone vehicle. A dataset was integrated that includes all seasonal phases and distinguishes 5 classes for monitoring: mangrove (mangle), creeping vegetation (vegetación rastrera), sand (arena), sea (mar) and hill-plain (cerro-planicie). For semantic segmentation different convolutional neural network (CNN) architectures were implemented and compared using transfer learning. The results have been robust in the classification, reaching an overall accuracy of 93.9% and between 89.9-95.8% in individual classes. In the Intersection over Union, IoU metric, the range was between 86.6-92.7%. In change detection, time series are used for class monitoring. This method has been applied to the case study of Ensenada Grande beach in the Complejo Insular del Espíritu Santo.

Keywords: Convolutional neural networks, Semantic segmentation, Time series, Protected coastal areas

1. Introducción

En la actualidad, el monitoreo de zonas costeras es una tarea imprescindible para el mantenimiento de ecosistemas. La actividad antropogénica, así como los eventos naturales, repercuten en el estado de conservación de estas áreas. Más aún, las zonas que constituyen áreas naturales protegidas

requieren de herramientas que contribuyan a la toma de decisiones, para la planeación y evaluación de los programas de manejo, por parte de las autoridades correspondientes.

Con relación a las actividades de monitoreo, la utilización de vehículos aéreos no tripulados, como los drones, se ha incrementado en los últimos años; esto debido a la practicidad de su funcionamiento, lo asequible de su adquisición y a la alta

*Autor para la correspondencia: M20310005@lapaz.tecnm.mx

Correo electrónico: M20310005@lapaz.tecnm.mx (Jacquelina Jetzabel Padilla-Arballo), saul.md@lapaz.tecnm.mx (Saul Martínez-Díaz), marco.cl@lapaz.tecnm.mx (Marco Antonio Castro-Liera), Jorge.lt@lapaz.tecnm.mx (Jorge Enrique Luna-Taylor)

resolución espacial y temporal obtenida. Adicionalmente, otras ventajas de su utilización, con respecto a la de imágenes satelitales, son: la anulación de problemas de retardos, la eliminación de oclusión por nubosidades, la disponibilidad de capturas en fechas determinadas. La altura alcanzada en los vuelos de un vehículo drone logra una resolución mensurable incluso en centímetros por píxel, lo que permite la revisión minuciosa de las áreas en sobrevuelo. Además, la integración de sensores de posicionamiento global en estos equipos facilita la georreferencia de las capturas.

El ejercicio de clasificar la superficie terrestre se ha convertido en uno de los más demandados en la actualidad, por la amplia variedad de aplicaciones en distintas áreas de estudio. Ecológicamente, posee utilidad como herramienta visual para la detección de degradación en superficies y su cuantificación.

Algunos de los métodos empleados anteriormente para la clasificación, han consistido en técnicas geoespaciales como en Pandey (Pandey et al., 2013). En su estudio, las imágenes provenían de una base de datos satelital y la clasificación supervisada se efectuaba con la aplicación de máxima verosimilitud y la precisión general alcanzaba hasta 91.68%. La detección de cambios era manifestada con las diferencias visualizadas en la serie temporal obtenida.

Técnicas utilizadas en años más recientes, como en Jenčo (Jenčo et al., 2020), involucran el mapeo de propiedades morfométricas de la superficie terrestre, captada en fotogrametría de un vehículo aéreo no tripulado. Lo anterior con el objetivo de evaluar la degradación del suelo y su dinámica temporal.

Los sistemas de información geográfica (GIS, por sus siglas en inglés) han sido herramientas clave para el desarrollo de trabajos como en Casella (Casella et al., 2016). Funciones predefinidas en GIS, aplicadas a imágenes obtenidas con drone, fueron utilizadas para evaluar cambios topográficos en zona costera. Un estudio con objetivo similar que empleó herramientas de fotogrametría, es definido en Yoo (Yoo & Oh, 2016), con la particularidad de que se adicionan mapeos de elevaciones 3D.

Otra de las formas de realizar la clasificación de superficie terrestre es mediante la segmentación semántica, tarea de la visión por computadora y procesamiento de imágenes, que consiste en la asignación de una etiqueta señalando la clase a la que pertenece cada píxel en una imagen; por lo que cada píxel de un mismo tipo es asignado a un mismo segmento (A.Géron, 2019). Entre los diferentes métodos para efectuar una segmentación semántica, el uso de las redes neuronales convolucionales (CNN, por sus siglas en inglés) va en aumento debido a la precisión y eficiencia demostrada en los últimos años.

Las redes neuronales artificiales son técnicas de aprendizaje máquina que simulan los mecanismos de aprendizaje del cerebro humano (Aggarwal, 2018) y su arquitectura consiste en capa de entrada, capas ocultas y capa de salida, en las cuales se encuentran módulos de procesamiento (neuronas) interconectados entre sí, que darán como resultado una predicción.

Las redes neuronales convolucionales son un modelo de aprendizaje profundo compuestas por una red multicapa con núcleos (*kernels*) que convolucionan para extraer características relevantes de la imagen de entrada.

Se pueden consultar diversos enfoques en la aplicación de CNN tanto para la clasificación de superficies como en

detección de cambios. Siguiendo la línea de clasificación, en Wagner (Wagner et al., 2019) son distinguidos tipos de bosques con la red convolucional U-net mediante su aplicación en imágenes satelitales. También con Younis (Younis & Keedwell, 2019), y en Guo (Guo et al., 2018), es ejemplificado el uso de redes neuronales convolucionales profundas para la clasificación de píxeles en capturas de satélite.

En detección de cambios, con Rubio (Rubio et al., 2019), esta es realizada utilizando vehículos aéreos no tripulados. Su metodología se basa en el alineamiento de imágenes capturadas de forma secuencial y el uso de una red convolucional profunda. Otra alternativa para detectar cambios es la comparación entre pares de imágenes; como fue experimentado en Bayr (Bayr & Puschmann, 2019), en capturas oblicuas de vegetación leñosa. En este caso, la CNN fue entrenada para la detección automática y clasificación de este tipo de vegetación y su posterior cuantificación.

En Valentini (Valentini & Balouin, 2020) es propuesta una CNN cuya entrada son superpíxeles (sets de píxeles) en lugar de píxeles individuales. Son utilizados dispositivos móviles como medio de captura para la detección de sargazo en las costas. El campo de visión es oblicuo en este caso y la segmentación es efectuada diferenciando parches de clases. La precisión se encuentra en un rango del 75 al 96%. La motivación para el uso de superpíxeles en diferentes trabajos es la economización de recursos, aunque con ello exista una cierta reducción de calidad de la imagen inicial. Otra aproximación con superpíxeles como datos de entrada, pero con capturas de origen satelital, se presenta en Gonzalo-Martín (Gonzalo-Martín et al., 2016), obteniendo una precisión cercana al 80%.

Entre los equipos para la obtención de las imágenes, además de los mencionados dispositivos móviles, los vehículos aéreos no tripulados y los satélites, es presentado en Ellenson (Ellenson et al., 2020) un estudio con capturas de cámaras ARGUS para el reconocimiento del estado morfológico de playas con el empleo de redes neuronales convolucionales, logrando un puntaje de 0.68 en la métrica F1.

En este documento es presentada una metodología para el monitoreo de superficie costera y la cuantificación de su degradación con base en el dinamismo de sus clases (mangle, vegetación rastrera, mar, arena y cerro-planicie), inspeccionada en series temporales. Esto es logrado aplicando segmentación semántica para la distinción de clases, por medio de aprendizaje profundo supervisado, con CNN. Considerando las referencias del estado del arte, los resultados obtenidos son consistentes. El dataset (conjunto de imágenes originales con su respectiva imagen etiquetada o *ground truth*) utilizado para el entrenamiento, fue creado y etiquetado manualmente a partir de las capturas efectuadas en planes de vuelo de drone durante un ciclo anual. La utilidad de este dataset puede extenderse a zonas de características similares.



Figura 1: Localización e imagen del área de estudio: Playa Ensenada Grande.

1.1. Caso de estudio

El área de estudio, la playa Ensenada Grande, se localiza dentro del Complejo Insular del Espíritu Santo, ubicado frente a las costas de La Paz, Baja California Sur, México (figura 1). Esta zona es un Área Natural Protegida; forma parte del Área de Protección de Flora y Fauna de las Islas del Golfo de California y se declaró Patrimonio Natural Mundial de la Humanidad de la UNESCO desde el año 2005. La Comisión Nacional de Áreas Naturales Protegidas (CONANP) es la institución a cargo de su administración y resguardo.

Existe amplia variedad de flora en el área, entre las que se destaca el manglar en sus diferentes tipos.

En este complejo también se hallan playas de las denominadas playas de bolsillo, por su tamaño. Durante el año, factores externos, como la actividad antropogénica y la actividad, en ciertos meses, de fenómenos meteorológicos (principalmente huracanes), pueden presentar repercusiones en la región.

Los estudios en el lugar se han referido principalmente al perfilamiento de playas y evaluación de zonas de manglar.

La aplicación de un modelo cuantitativo y automático sobre el área monitoreada proveería información, con mayor precisión, para la toma de decisiones sobre las medidas implementadas en programas de conservación, respecto a las actividades de los factores externos mencionados.

Para este efecto, posterior al entrenamiento de la red convolucional en el método propuesto, en la segmentación semántica son calculadas automáticamente las coberturas en píxeles de cada clase en las imágenes de resolución estandarizada. Los cálculos son realizados en cada una de las zonas monitoreadas. Al ser estas zonas de polígonos predefinidos, el valor de las áreas en metros cuadrados es determinado de forma georeferenciada, de esta manera es viable realizar la cuantificación de superficies, al ser estimadas las correspondencias entre píxeles y metros cuadrados. Estos procesos son realizados de forma iterativa para cada imagen de las regiones monitoreadas. Consecutivamente, las series temporales son generadas considerando la clase de interés en cada zona.

2. Materiales y métodos

El flujo de trabajo es ilustrado en la Figura 2.

2.1. Obtención de datos

La CONANP colabora en este trabajo proporcionando capturas mensuales de imágenes y videos del área de estudio desde febrero de 2021 (para efectos de este documento se integraron datos desde esa fecha hasta el mes de abril de 2022).

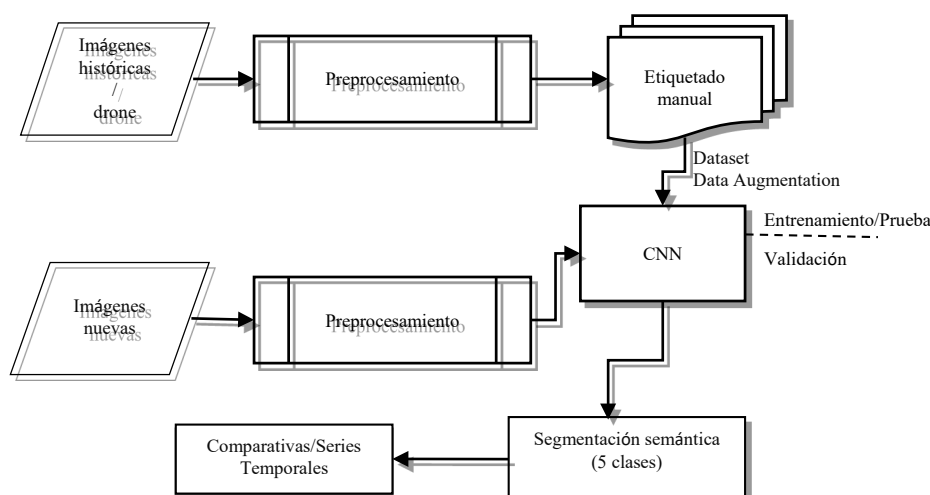


Figura 2: Metodología adoptada en este trabajo.

El equipo utilizado en los planes de vuelo es un drone Inspire (DJI, China), cuadricóptero con cámara (video 4K, fotografías 12MP) y estabilizador basado en Sistema de Posicionamiento Global (GPS, por sus siglas en inglés).

Los planes de vuelo son contemplados a una altura de 150 metros, cubriendo ~1.7 hectáreas. Aunque en este trabajo se indica un período analizado, las capturas siguen actualizándose mes tras mes.

Fueron obtenidas más de 500 fotografías y 9 videos para su procesamiento. Del total de esta colección multimedia, se descartaron aquellas imágenes que presentaron alguna inconsistencia en ángulos de captura o en donde la luz solar intensificara reflejos en superficies acuosas.

También fueron proporcionadas por CONANP, imágenes históricas del caso de estudio, así como del resto de playas del complejo insular, que datan del año 2018, las cuales fueron integradas a las capturas mensuales.

2.2. Preprocesamiento y elaboración de dataset

El conjunto de imágenes resultantes de alta resolución (hasta 4000x3000 pixeles), tras la depuración, fue de 310 fotografías. Su preprocesamiento consistió en:

- Mejora de contraste.
- Rectificación.
- Cambio de formato de imagen: Para homogeneizar el conjunto se estableció el formato PNG.
- Redimensionamiento: Se determinó un tamaño estandarizado de 960x720 pixeles.

Con la herramienta *Image Labeler (Computer Vision Toolbox - MATLAB & Simulink, 2020)* se creó el dataset inicial conformando el ground truth de forma manual, con las etiquetas de clases identificadas por colores. Clases del dataset:

- Mangle
- Vegetación rastrera
- Mar
- Arena
- Cerro-Planicie

2.3. Comparativas previas CNN

Para la segmentación semántica se determinó el uso de la arquitectura CNN DeepLabV3+ (Chen et al., 2018) y se tomó una muestra de 50 elementos del dataset inicial para seleccionar el mejor desempeño de red troncal (ver Tabla 1) entre 5 diferentes CNN: InceptionResNetV2 (Szegedy et al., 2016.), MobileNetV2 (Sandler et al., 2019), Xception (Google, 2014), ResNet18 y ResNet50 (He et al., 2015.).

Tabla 1: Métricas de desempeño CNN

CNN	Precisión	Precisión	Media
	Global	Media	IoU
InceptionResNetV2	0.8678	0.87048	0.73417
MobileNetV2	0.88729	0.90013	0.77613
Xception	0.9065	0.91861	0.80649
ResNet18	0.84096	0.8326	0.68868
ResNet50	0.90022	0.88258	0.78862

En los resultados la CNN Xception tuvo los más altos índices de desempeño (90.65% en precisión global), por lo que se definió como la red troncal en los procesos sucesivos.

2.4. Arquitecturas CNN

DeepLabV3+ es una CNN utilizada para tareas de segmentación semántica, su arquitectura simplificada se muestra en la Figura 3. Está compuesta por un codificador, en donde son realizadas convoluciones de tipo *atrous* (ver Figura 4). En CNN convencionales es frecuente el uso de la convolución en profundidad, en donde el procesamiento es aplicado en cada canal de la imagen. Existen también, en algunas arquitecturas, convoluciones puntuales, en las que es combinada en cada canal la salida de convoluciones en profundidad. La diferencia entre las convoluciones *atrous* y las convoluciones en profundidad radica en que, en su filtro, *kernel*, existen espacios entre los puntos de convolución donde la separación está definida en el parámetro *rate*. Lo anterior permite trabajar en distintas resoluciones, enriqueciendo de esta manera la extracción de características.

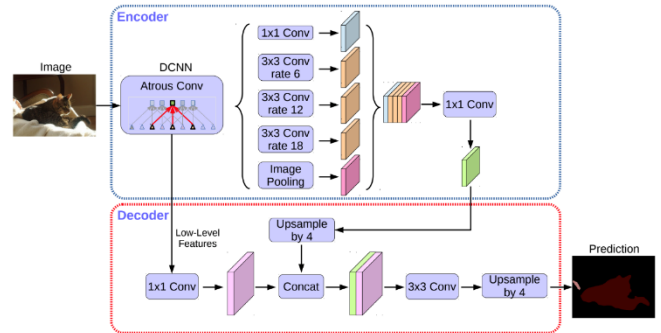


Figura 3: Arquitectura DeepLabV3+ (Chen et al., 2018).

La convolución *atrous* está definida en (1), en donde y es el mapa de características saliente con ubicación i , x es el mapa de características entrante, r es el *rate* de separación aplicado en el filtro convolucional w , con longitud k .

$$y[i] = \sum_k x[i + r \cdot k]w[k] \tag{1}$$

La arquitectura de DeepLabV3+ está basada en su predecesora DeepLabV3, con la distinción de la integración de un módulo de decodificación que obtiene límites más definidos en la segmentación.

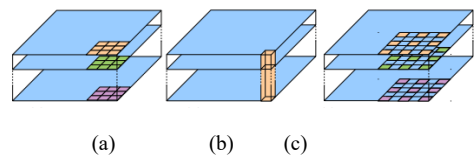


Figura 4: Convoluciones: (a) en profundidad, (b) puntual, (c) *atrous* (Chen et al. 2019)

En cuanto a la red troncal, Xception (Figura 5), su arquitectura radica en la aplicación de convoluciones separables en profundidad, en las cuales, después de una convolución convencional en profundidad, es efectuada una

convolución puntual. Lo anterior resulta en una reducción en la complejidad computacional y en una mayor velocidad de procesamiento.

En la composición final de estas arquitecturas se tienen 205 capas en profundidad y un preentrenamiento con ImageNet, dataset con más de 14 millones de imágenes destinadas para el reconocimiento visual de objetos. Este preentrenamiento es utilizado como aprendizaje transferido en la metodología de este trabajo, lo que reduce en forma considerable el número de épocas y tiempo de entrenamiento.

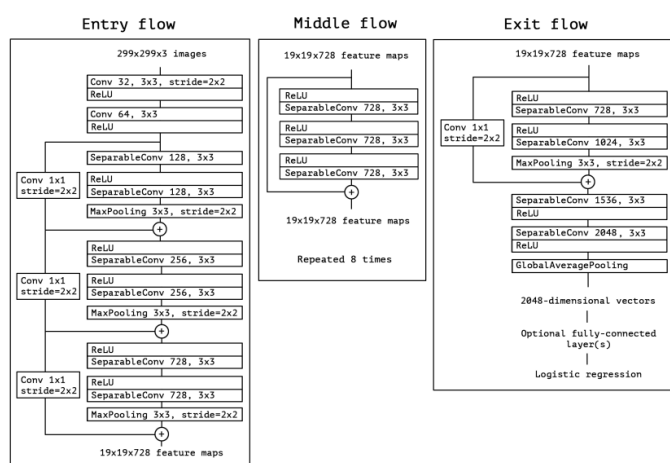


Figura 5: Arquitectura de Xception (Google,2014)

3. Resultados

3.1. Entrenamiento

Por capacidad del equipo de trabajo, del total del dataset originalmente creado fueron seleccionadas 166 imágenes, con sus etiquetas respectivas, las cuales fueron divididas de esta forma: 60% para entrenamiento, 20% para prueba y 20% para validación.

Para evitar un sobreentrenamiento se aplicaron técnicas de aumento de datos (data augmentation), con transformaciones de traslación y rotación en las imágenes del dataset.

El algoritmo optimizador utilizado fue el descenso de gradiente estocástico con momentum (SDGM por sus siglas en inglés); con momentum definido en 0.9; factor de aprendizaje (*learning rate*) 0.01; Regularización L2, 0.0005; función de pérdida, entropía cruzada categórica; *minibatch*, 4; épocas, 30; iteraciones,750. Ver evolución en la Figura 6.

Los procesos fueron codificados en MATLAB (The Mathworks, Inc., 2020). Para la fase de entrenamiento fue utilizado un equipo con procesador Intel Core i7-9750H, RAM 16 GB, así como una Unidad Gráfica de Procesamiento (GPU, por sus siglas en inglés) en tarjeta Nvidia GeForce RTX 2060. El entrenamiento tomó ~104 minutos. Una vez entrenada y guardada la red, el tiempo de predicción en imágenes individuales es de ~0.47 segundos.

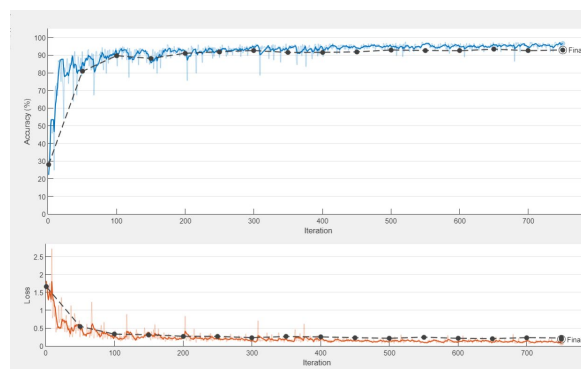


Figura 6: Fase de entrenamiento

3.2. Métricas

Para evaluar la tarea de segmentación se contemplaron valores de verdaderos positivos (VP), falsos positivos (FP), falsos negativos (FN) y verdaderos negativos (VN) en las ecuaciones.

Precisión global: Razón de pixeles correctamente clasificados del total.

$$Precisión = \frac{VP}{VP+FP} \tag{2}$$

Precisión media: Razón de pixeles correctamente clasificados por clase, promediado todas las clases.

IoU: Intersección sobre unión (3), que especifica el traslape entre los pixeles de la predicción y los del *ground truth* (4). IoU, o índice Jaccard, es una de las métricas más utilizadas en tareas de segmentación semántica.

$$IoU = \frac{VP}{VP+FN+FP} \tag{3}$$

$$IoU = \frac{\text{Área de traslape}}{\text{Área de unión}} \tag{4}$$

IoU media. Promedio de la intersección sobre unión de todas las clases.

IoU ponderada: Promedio IoU de todas las clases, ponderada por el número de pixeles en la clase.

Puntaje BF Medio: Promedio del puntaje (6) *Boundary F1* (BF) de todas las imágenes. Requiere cálculo previo de exhaustividad (5).

$$Exhaustividad = \frac{VP}{VP+FN} \tag{5}$$

$$Puntaje = 2 \cdot \frac{\text{precisión} \cdot \text{exhaustividad}}{\text{exhaustividad} + \text{precisión}} \tag{6}$$

Las métricas correspondientes se evaluaron por clases individuales y de forma global.

En la Tabla 2 se indican los resultados por clases.

Cerro-planicie es la clase que alcanza los valores máximos en IoU y precisión. Lo anterior tiene explicación en la frecuencia de clases en las imágenes del dataset. Debido a la dificultad de contemplar un conjunto de capturas con clases completamente equilibradas en el ambiente, en el dataset, en su versión más balanceada posible, predominó ligeramente la

clase Cerro-Planicie, lo que resultó favorecedor en sus predicciones.

La clase Arena obtuvo los valores menores, aunque su presencia en el dataset también fue frecuente. Los valores menores se manifiestan en esta clase en particular, porque, por su naturaleza, la definición de sus límites en la imagen es más intrincada que en el resto de las clases; además, en algunas capturas, se presentaron oclusiones por presencia de personas y objetos llevados por éstas a la zona.

La clase Mar tuvo buenos resultados en su predicción general, por las mínimas variaciones en el ciclo anual.

Mangle y Vegetación Rastrera lograron ser efectivamente predichas; algunas de las inconsistencias se visualizaron en estas clases principalmente por las fotografías con ligero desenfoque o sombras.

Tabla 2: Métricas por clase

Clases	Precisión	IoU	Puntaje BF Medio
Cerro-Planicie	0.95797	0.92728	0.73427
Mar	0.95605	0.92719	0.81250
Mangle	0.93622	0.79804	0.67817
VRastrera	0.95850	0.84949	0.70865
Arena	0.89861	0.86669	0.65889

Las métricas globales se muestran en la Tabla 3, en donde los resultados obtenidos por clase se promedian o ponderan entre el número de clases. La precisión global alcanzada fue del 93.93%.

Tabla 3: Métricas globales en entrenamiento de CNN

Precisión Global	Precisión Media	IoU Media	IoU Ponderada	Puntaje BF medio
0.9393	0.94147	0.87374	0.88732	0.70573

3.3. Pruebas de segmentación

En la Figura 7 se incluyen algunos ejemplos de los resultados obtenidos con las diferentes clases.

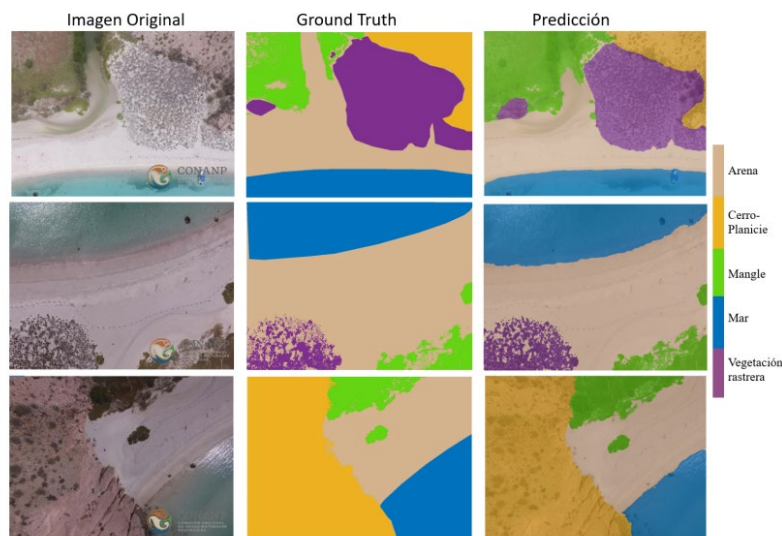


Figura 7. *Ground Truth* vs predicciones resultantes

3.4. Series temporales

Una vez logrados resultados satisfactorios en las pruebas, el sitio de estudio fue dividido en 6 zonas (Figura 8), para el monitoreo de clases por su relevancia en cada una de ellas. Por cada zona se segmentó una imagen de cada mes transcurrido, para cuantificar los cambios en metros cuadrados en el período.

- Zona 1: Monitoreo de mangle en 2749.8m², Figura 9.
- Zona 2: Monitoreo de mangle en 2786.6m², Figura 10.
- Zona 3: Monitoreo de mangle en 3247m², Figura 11.
- Zona 4: Monitoreo de vegetación rastrera en 3336.36m², Figura 12.
- Zona 5: Monitoreo de vegetación rastrera en 1738.45m², Figura 13.
- Zona 6: Monitoreo de superficie de playa en 6659.1m², Figura 14.

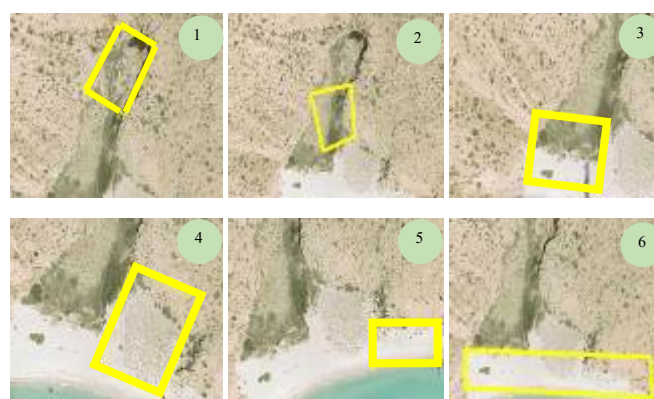


Figura 8. División del área de estudio en 6 zonas.

Al segmentar de forma semántica las imágenes por medio de la red entrenada, en la codificación es realizado el cálculo de coberturas en metros cuadrados por clase. La cobertura considera el número total de píxeles de la clase monitoreada, multiplicada por el área en metros cuadrados de la zona de interés correspondiente, dividida entre el número total de píxeles de las dimensiones estandarizadas de las imágenes de entrada a la red. Una vez calculadas las áreas por clase supervisada en su zona correspondiente, son almacenados estos valores para la generación de las series temporales por

regiones de interés, con la frecuencia mensual y su estimación de cobertura en metros cuadrados. El porcentaje de cambio es referido mes a mes o en promedio contemplando el ciclo anual. Es importante señalar que el análisis estadístico requiere la prolongación de los ciclos para recabar más información.

En cuanto al monitoreo por clases, el mangle, contemplado en un solo tipo, es inspeccionado a través del ciclo anual, en 3 de las 6 zonas. El incremento o decremento de su indicador tiene correspondencia con las fases estacionales y los fenómenos meteorológicos presentados.

La clase vegetación rastrera corresponde a aquella que crece en terreno de duna, por lo que una disminución en esta clase implica una reducción en estos depósitos.

En la superficie de playa se identifica la ampliación, superficial, o pérdida de área de la región arenosa con relación a las clases de cobertura vegetal. Si el tamaño de la cobertura del área de mangle o de vegetación rastrera se ve disminuida, la cobertura de arena será visualmente mayor y viceversa. En esta zona en particular es donde se manifiesta de forma más frecuente la actividad antropogénica.

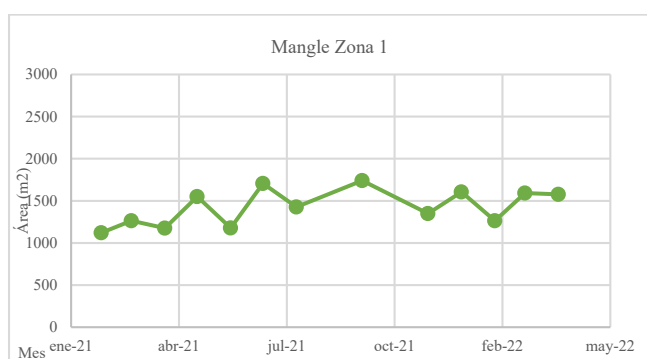


Figura 9. Monitoreo zona 1, Feb 2021-Abr 2022

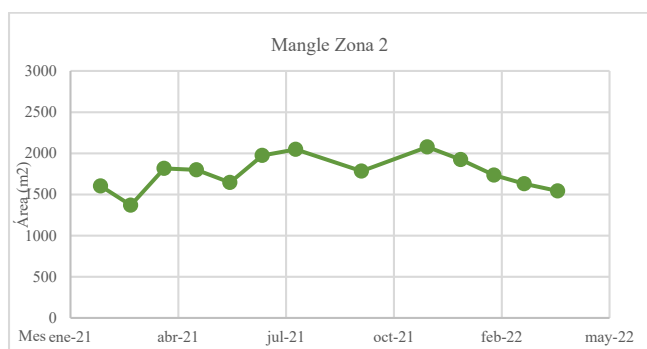


Figura 10. Monitoreo zona 2, Feb 2021-Abr 2022

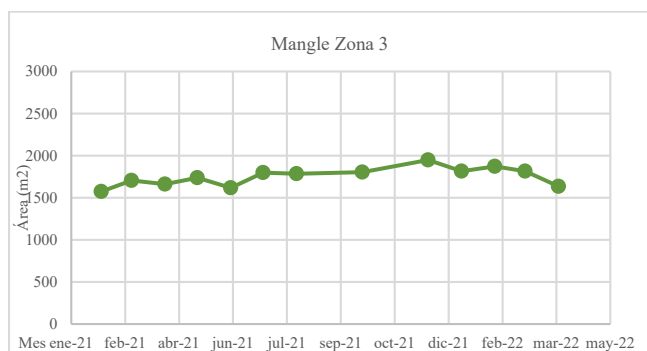


Figura 11. Monitoreo zona 3, Feb 2021-Abr 2022

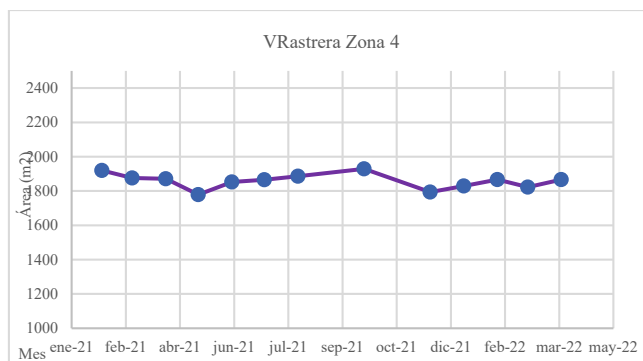


Figura 12. Monitoreo zona 4, Feb 2021-Abr 2022

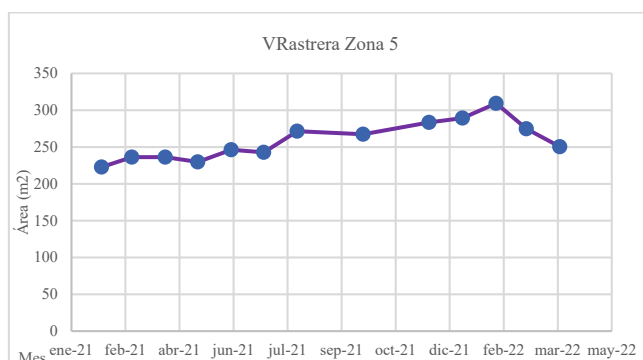


Figura 13. Monitoreo zona 4, Feb 2021-Abr 2022

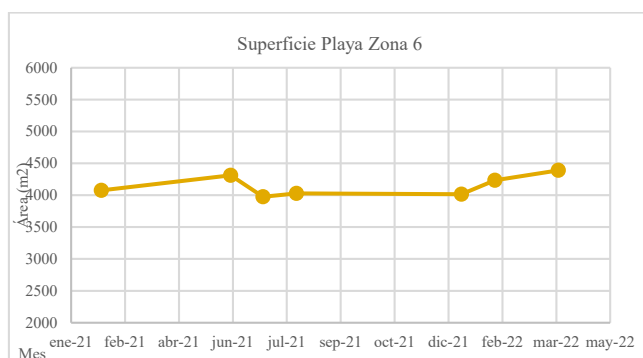


Figura 14. Monitoreo zona 6, Feb 2021-Abr 2022

4. Conclusiones

El uso de las redes neuronales convolucionales aplicadas en capturas de dispositivos aéreos no tripulados constituye una alternativa asequible y de alta precisión para el análisis de las imágenes y la generación de series temporales largas.

Efectos de eventos naturales como huracanes pueden ser cuantificados tras su paso en las clases monitoreadas: En el caso de estudio se presentó un fenómeno atmosférico en la segunda mitad del ciclo anual observado.

Los resultados obtenidos fueron satisfactorios, por lo que el método implementado supone una contribución a las herramientas de monitoreo de las áreas costeras locales, administradas por las autoridades correspondientes.

El dron de este trabajo es el modelo Inspire I, que utiliza una cámara Zennuse X3, con rango de 360°, 3 ejes estabilizadores y sensor GPS para ubicaciones georreferenciadas. La resolución en video es de 4k y en

fotografías de 12MP. Fue utilizado en planes de vuelo de 150m de altura, y las imágenes de captura se ajustaron a una resolución de 960x720 píxeles en su preprocesamiento. No obstante, es posible aplicar la misma metodología considerando otras alturas de capturas y resoluciones de imágenes de entrada distintas, mientras éstas se estandaricen y sean actualizados los valores correspondientes en los cálculos de coberturas.

Como consideraciones futuras se proyectaría la diversificación de escenarios para aumentar la robustez en predicciones.

Agradecimientos

Agradecemos a la Comisión Nacional de Áreas Naturales Protegidas (CONANP), con especial atención a la Biól. Irma González López, directora de Islas del Golfo de California en B.C.S., y de la Zona Marina del Archipiélago de Espíritu Santo, por el apoyo brindado a este proyecto. Nuestra gratitud al Lic. Raol David Preciado Carrillo, por su importante contribución al proporcionar el material digital histórico, así como la recabación mensual multimedia del área de estudio, en el transcurso de la realización de este trabajo.

Se agradece al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el soporte académico proporcionado.

Referencias

- A.Géron. (2019). Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems. *O'Reilly Media*, pp. 450-478. <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>
- Aggarwal, C. C. (2018). Neural Networks and Deep Learning. *Neural Networks and Deep Learning*. <https://doi.org/10.1007/978-3-319-94463-0>
- Bayr, U., & Puschmann, O. (2019). Automatic detection of woody vegetation in repeat landscape photographs using a convolutional neural network. *Ecological Informatics*, 50, 220–233. <https://doi.org/10.1016/j.ecoinf.2019.01.012>
- Casella, E., Rovere, A., Pedroncini, A., Stark, C. P., Casella, M., Ferrari, M., & Firpo, M. (2016). Drones as tools for monitoring beach topography changes in the Ligurian Sea (NW Mediterranean). *Geo-Marine Letters*, 36(2), 151–163. <https://doi.org/10.1007/S00367-016-0435-9>
- Chen, L.-C., Zhu, Y., Papandreu, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. https://github.com/tensorflow/models/tree/master/Computer_Vision_Toolbox_-_MATLAB_&_Simulink. (2020.). Retrieved May 9, 2022, from <https://la.mathworks.com/products/computer-vision.html>
- Ellenson, A. N., Simmons, J. A., Wilson, G. W., Hesser, T. J., & Splinter, K. D. (2020). Beach State Recognition Using Argus Imagery and Convolutional Neural Networks. *Remote Sensing* 2020, Vol. 12, Page 3953, 12(23), 3953. <https://doi.org/10.3390/RS12233953>
- Gonzalo-Martin, C., Garcia-Pedrero, A., Lillo-Saavedra, M., & Menasalvas, E. (2016). Deep learning for superpixel-based classification of remote sensing images. <https://doi.org/10.3990/2.401>
- Google, C. (2014). Xception: Deep Learning with Depthwise Separable Convolutions. *Computer Vision and Pattern Recognition*, 1251–1258.
- Guo, R., Liu, J., Li, N., Liu, S., Chen, F., Cheng, B., Duan, J., Li, X., & Ma, C. (2018). Pixel-Wise Classification Method for High Resolution Remote Sensing Imagery Using Deep Neural Networks. *ISPRS International Journal of Geo-Information* 2018, Vol. 7, Page 110, 7(3), 110. <https://doi.org/10.3390/IJGI7030110>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. <http://image-net.org/challenges/LSVRC/2015/>
- Jenčo, M., Fulajtár, E., Bobál'ová, H., Matečný, I., Saksá, M., Kožuch, M., Gallay, M., Kaňuk, J., Piš, V., & Oršulová, V. (2020). Mapping soil degradation on arable land with aerial photography and erosion models, case study from Danube Lowland, Slovakia. *Remote Sensing*, 12(24), 1–17. <https://doi.org/10.3390/RS12244047>
- Pandey, P. C., Rani, M., Srivastava, P. K., Sharma, L. K., & Singh Nathawat, M. (2013). Land degradation severity assessment with sand encroachment in an ecologically fragile arid environment: a geospatial perspective. <https://doi.org/10.5339/connect.2013.43>
- Rubio, V. G., Antonio, J., Ferrán, R., Manuel Menéndez García, J., Sánchez Almodóvar, N., María, J., Mayordomo, L., & Álvarez, F. (2019). Automatic Change Detection System over Unmanned Aerial Vehicle Video Sequences Based on Convolutional Neural Networks. <https://doi.org/10.3390/s19204484>
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks.
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.
- The Mathworks, Inc. (2020). MATLAB (No. R2020b). la.mathworks.com/
- Valentini, N., & Balouin, Y. (2020). Assessment of a Smartphone-Based Camera System for Coastal Image Segmentation and Sargassum monitoring. *Journal of Marine Science and Engineering* 2020, Vol. 8, Page 23, 8(1), 23. <https://doi.org/10.3390/JMSE8010023>
- Wagner, F. H., Sanchez, A., Tarabalka, Y., Lotte, R. G., Ferreira, M. P., Aidar, M. P. M., Gloor, E., Phillips, O. L., & Aragão, L. E. O. C. (2019). Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. *Remote Sensing in Ecology and Conservation*, 5(4), 360–375. <https://doi.org/10.1002/rse2.111>
- Yoo, C. I., & Oh, T. S. (2016). Beach volume change using UAV photogrammetry Songjung beach, Korea. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 41, 1201–1205. <https://doi.org/10.5194/ISPRSARCHIVES-XLI-B8-1201-2016>
- Younis, M. C., & Keedwell, E. (2019). Semantic segmentation on small datasets of satellite images using convolutional neural networks. *Journal of Applied Remote Sensing*,

Apéndice A.

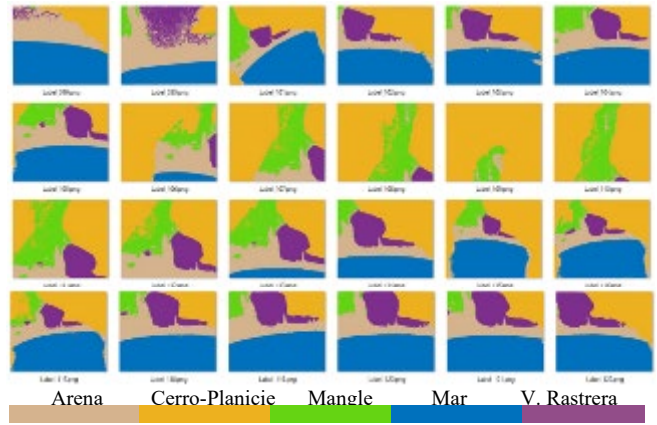
Figura A1. Algunos elementos del conjunto de datos de entrenamiento:

(a) Imágenes preprocesadas.

(b) Imágenes etiquetadas con identificación de colores por clase.



(a)



(b)