



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas

“Identificación de la cobertura arbórea de Guayaquil urbano usando técnicas de inteligencia artificial”

PROYECTO INTEGRADOR

Previo la obtención del Título de:

INGENIERO EN ESTADÍSTICA INFORMÁTICA

Presentado por:

María Fernanda Jurado Mantilla

GUAYAQUIL - ECUADOR

Año: 2021

DEDICATORIA

A Dios, por ser mi fuente de sabiduría y fortaleza. A mi “rubia” por ser ejemplo de vida como madre, académica y profesional. A mi familia. A JF.

AGRADECIMIENTOS

A Dios, porque gracias a su misericordia y gracia puedo culminar mis estudios.

A mi mamá quien se ha esforzado para que yo sea una buena persona, estudiante y profesional.

A mi hermana por ser mi compañera de carrera, siempre a mi lado para afrontar obstáculos.

Mi familia y novia por su soporte, cariño y ánimos para que no me rinda y finalice esta etapa.

A la Sociedad Ecuatoriana de Estadística quienes me abrieron las puertas para que pueda seguir preparándome y por ello, les debo mucho de este camino. A Néstor M, Roberto E. y Ángel C. porque más que amigos, han sido mis mentores.

A mis profesores de formación, en especial a Alejandro C, Teresa S., Joffre S, Wendy P. porque sin su ayuda, enseñanzas y palabras de aliento, no estaría aquí.

DECLARACIÓN EXPRESA

“Los derechos de titularidad y explotación, me corresponde conforme al reglamento de propiedad intelectual de la institución; Yo, María Fernanda Jurado Mantilla doy mi consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”



María Fernanda Jurado Mantilla

EVALUADORES



Ph.D. Sandra García Bustos
PROFESOR DE LA MATERIA



Ph.D. Sergio Bauz
PROFESOR TUTOR

RESUMEN

Guayaquil es una ciudad que coloquialmente es conocida como un lugar donde carecen los árboles. Según el Instituto Nacional de Estadística y Censos, en 2010, determinó que Guayaquil tenía 1,12 m² de áreas verde por habitante. Posteriormente el Gobierno Autónomo Descentralizado de la ciudad en el 2019 indicó en diversos medios de comunicación que los espacios verdes superaban los 9 m² por habitante sugeridos por la Organización Mundial de la Salud. Como parte de los espacios verdes, está el revestimiento arbóreo que puede ser observado mediante imágenes satelitales. Para poder calcular dicho revestimiento, se tomó ortofotografías de Guayaquil urbano y desde la visión por computadora, se empleó redes neuronales convolucionales para así poder discriminar la cobertura arbórea del resto de elementos presentados en las imágenes, para esta tarea se utilizó Python como herramienta de programación. Como resultado de esta investigación se obtuvo que Guayaquil urbano tiene 85.39 km² de cobertura arbórea en 323.49 km², es decir un 26% en relación con su superficie; pero bajo el principio de Pareto, el 80% del recubrimiento arbóreo se encuentra en el 38% del territorio, es decir que el 62% contiene el 20% de recubrimiento restante.

Palabras claves: Visión por computadora, Redes Neuronales Convolucionales, Organización Mundial de la Salud, cobertura arbórea, zonas verdes.

ABSTRACT

Guayaquil is a city that is commonly known as a place where trees are lacking. According to the National Institute of Statistics and Censuses, in 2010, it determined that Guayaquil had 1.12 m² of green areas per inhabitant. Subsequently, the Decentralized Autonomous Government of the city in 2019 indicated in various media that green spaces exceeded 9 m² per inhabitant suggested by the World Health Organization. As part of the green spaces, there is the tree covering that can be observed through satellite images. In order to calculate said cladding, orthophotographs of urban Guayaquil were taken and from computer vision, convolutional neural networks were used in order to discriminate the tree cover from the rest of the elements presented in the images, all using Python (programming language and software statistical) as a programming tool. As a result of this research, it was obtained that urban Guayaquil has 85.39 km² of tree cover in 323.49 km², that is, 26% in relation to its surface; but under the Pareto principle, 80% of the tree cover is found in 38% of the territory, that is, 62% contains the remaining 20% of cover.

Keywords: Computer vision, Convolutional Neural Networks, World Health Organization, tree cover, green areas.

ÍNDICE GENERAL

RESUMEN.....	vi
ABSTRACT	vii
ÍNDICE GENERAL	viii
ABREVIATURAS.....	x
ÍNDICE DE FIGURAS	xi
ÍNDICE DE TABLAS.....	xii
CAPÍTULO 1.....	1
1. Introducción	1
1.2 Justificación del problema	4
1.3 Objetivos	4
1.1.1 Objetivo General.....	4
1.1.2 Objetivos Específicos.....	5
1.4 Marco Teórico	5
1.1.3 Muestreo Aleatoria Simple.....	5
1.1.4 Visión por Computadora	6
1.1.5 Redes neuronales convolucionales (CNN)	8
1.1.6 Transferencia de aprendizaje.....	12
CAPÍTULO 2.....	14
2. METODOLOGÍA.....	14
2.1 Selección.....	14
2.2 Muestreo Aleatorio Simple	15
2.3 Etiquetado de imágenes.....	19
2.4 Entrenamiento del modelo.....	20
2.5 Validación del modelo	21
2.6 Aplicación del modelo re-entrenado.	22

2.7 Cálculo de la cobertura arbórea	23
CAPÍTULO 3.....	24
3. RESULTADOS Y ANÁLISIS.....	24
3.1 Resultados	24
3.2 Análisis situacional de los resultados	26
CAPÍTULO 4.....	28
4. Conclusiones y recomendaciones	28
4.1 Conclusiones.....	28
4.2 Recomendaciones.....	29
BIBLIOGRAFÍA.....	30

ABREVIATURAS

R-CNN	Región de Redes Neuronales Convolucionales (Siglas en inglés)
TIFF	Formato de imagen etiquetado (Siglas en inglés)
DL	Aprendizaje profundo
AI	Inteligencia Artificial
ML	Aprendizaje de máquina
TL	Transferencia de aprendizaje
MAS	Muestro Aleatorio Simple
SAS Planet	Software de sistema de análisis espacial (Siglas en inglés)
GNU	Licencia General Pública (Siglas en inglés)
COCO	Objetos Comunes en Contexto (Siglas en inglés)
JSON	Notación de objetos JavaScript (Siglas en inglés)
AP	Media de precisión (siglas en inglés)
IoU	Intersección sobre la unión (siglas en inglés)

ÍNDICE DE FIGURAS

Gráfico 1.1 Visión por computadora	7
Gráfico 1.2 Redes Neuronales Convolucionales	9
Gráfico 1.2 Redes neuronales convolucionales rápidas basadas en regiones (Fast y Faster R-CNN).....	10
Gráfico 2.1 Imagen satelital de Guayaquil Urbano, barrio Garzota	15
Gráfico 2.2 Imagen etiquetada de Guayaquil Urbano, barrio Fertisa.....	20
Gráfico 2.3 Fórmula de precisión media.....	21
Gráfico 2.4 Algoritmo de la inferencia del modelo.....	22
Gráfico 3.1 Cobertura arbórea de Guayaquil Urbano en porcentaje.....	26

ÍNDICE DE TABLAS

Tabla 2.1 Barrio Urdesa, conteo de pixeles por imagen.....	16
Tabla 2.2 Barrio Bastión Popular, conteo de pixeles por imagen.....	17
Tabla 2.3 Barrio Fertisa, conteo de pixeles por imagen.....	18
Tabla 3.1 Guayaquil Urbano, cobertura arbórea/superficie.....	24
Tabla 3.2 Guayaquil Urbano por sector, cobertura arbórea/superficie.....	25
Tabla 3.3 Distribución de la cobertura arbórea de Guayaquil Urbano.....	27

CAPÍTULO 1

1. INTRODUCCIÓN

Gracias al mundo de la ciencia ficción se ha relacionado a la inteligencia artificial (IA) como algo irreal o lejano, pero con el pasar de los años y con el desarrollo de nuevas tecnologías, es algo que está muy presente en la realidad actual. Se conoce que la inteligencia artificial es el conjunto de algoritmos que tienen como fin crear un sistema o máquina en el cual se pretende emular la actividad humana (Universidad Americana de Europa, 2020). Bajo el umbral de la IA, se encuentra el “Deep Learning” o aprendizaje profundo en donde su estructura lógica imita a la organización del sistema nervioso humano, en el cual sus capas de procesamiento se especializan en divisar particularidades en los objetos percibidos (Universidad Austral de Chile, 2020). Dado a los avances en neurociencias, se conoce que hay centros determinados del lenguaje o que existen redes especialistas en descubrir diferentes aspectos de la visión como líneas, bordes, simetrías, áreas, dimensiones e incluso áreas intrínsecamente relacionadas con el reconocimiento facial y la expresión emocional (Byoung Chul Ko, 2018). Esta arquitectura del sistema nervioso es imitada por los modelos computacionales del aprendizaje profundo permitiendo que dentro del sistema existan redes que se especialicen en detectar características determinadas (Revista Xakata, 2016). En el caso particular de la visión por computadora, se emula lo que podría observar el ojo humano, donde con la interpretación de imágenes adquiridas se reconozca objetos y el espacio que ocupa; para lograr esto, el modelo a usar es Redes Neuronales Convolucionales (CNN, siglas en inglés), en el cual a través de la jerarquía de sus capas puede ir reconociendo líneas y curvas en sus capas tempranas y a medida que se va especializando, sus capas profundas reconocen figuras complejas como rostros u objetos (Santander Global Tech, 2020).

Entre las aplicaciones de la visión por computadora se encuentra el reconocimiento facial, tecnología que está presente en los celulares para desbloquear los mismos como también está presente en cámaras de seguridad que son puntos de control en centros comerciales o propiedades privadas. Otra aplicación es en medicina, la mayoría de los diagnósticos se deben a imágenes como radiografías, ecografías o resonancias magnéticas, etc.. Se ha implementado modelos de aprendizaje profundo en el cual se puede determinar con mejores resultados si el paciente presenta o no cierta enfermedad. En otra área donde se ha obtenido beneficio es en el campo, en la agricultura, existen herramientas que comprueban que el grano sea de calidad, que la planta no se rodee de hierbas malas, entre otros. La visión por computadora, gracias al aprendizaje profundo, actualmente es de gran aporte en el desarrollo de la sociedad dado que ha facilitado el trabajo y ha otorgado beneficios.

Así mismo, el diseño de áreas verdes es considerado un factor importante para la sociedad dado que constituye y aportan un sinnúmero de beneficios para la calidad de vida de las personas. Se conoce que sin las plantas ni árboles no existiría el proceso de absorción del dióxido de carbono, purificación y creación de oxígeno; además, controlan la temperatura, evitan la erosión, protegen a la biodiversidad y otros factores favorables para el ambiente. Por los múltiples beneficios obtenidos de las áreas verdes, la Organización Mundial de la Salud (OMS) recomienda un mínimo de 9-15 m²/habitante. Este indicador depende de los planes de manejo de los Municipios. La gestión que realiza el Gobierno Autónomo Descentralizado Municipal de Guayaquil es de implementar y llevar un control de las zonas verdes de la urbe, además de garantizar que sean adecuadas para sus ciudadanos. En Guayaquil, según un informe de su Municipio, se cumple con los 9 m², alcanzando sólo el estándar mínimo establecido por la Organización Mundial de la Salud (El Universo, 2019). Al momento, existen controversias dado que los ciudadanos guayaquileños consideran que su ciudad no cuenta con los suficientes árboles (El Universo 2020).

Por ello, conociendo que humanamente no es posible contar la cobertura arbórea de la urbe, se evidencia el uso de la visión por computadora y reentrenar un modelo de aprendizaje profundo para que pueda calcular el revestimiento arbóreo existente y establecer indicadores en el cual informe si Guayaquil, en su sector urbano, es una ciudad que cumple las condiciones ambientales.

1.1 Descripción del problema

Guayaquil es una ciudad que coloquialmente es conocida como un lugar donde carecen los árboles (Primicias, 2019). Según el Instituto Nacional de Estadísticas y Censos, en 2010, determinó que Guayaquil tenía 1,12 m² de áreas verde por habitante. En un segundo estudio realizado en el 2012 por dicha institución, no se presentó datos de la urbe dado que el municipio no envió la información solicitada, siendo el único cantón del cual no se conocía su Indicador Verde Urbano (IVU) (INEC, 20212). Para el 2019, la dirección encargada de los espacios verdes del cabildo indicó en diversos medios de comunicación que en Guayaquil se superan los 9m² por habitante sugeridos por la OMS (El Universo, Primicias, 2019). Pero la falta de estudios oficiales publicados genera desconocimiento si se acatan las normas internacionales sobre los espacios verdes, por lo que se induce a una falta de conciencia ambiental tanto en los ciudadanos guayaquileños como en instituciones públicas y privadas, quienes cada día necesitan reforzar más sus políticas ambientales, dado que también son responsables de generar desarrollo sostenible que mejore la calidad de vida de la comunidad.

El desarrollo de un instrumento que pueda colaborar con datos muy aproximados sobre la cantidad de cobertura arbórea es un gran paso para conocer la realidad de Guayaquil urbano en esta arista ambiental, además de confirmar o desmentir sobre si la ciudad cuenta con los suficientes árboles en medida de su revestimiento arbóreo.

1.2 Justificación del problema

En la web se puede encontrar varias herramientas que pueden segmentar la vegetación de los distintos elementos que existen en la superficie de la tierra, entre ellas Google Earth Engine que, mediante los satélites Landsat, Sentinel, entre otros, tienen la capacidad de poder clasificar dichos elementos con distintos modelos de aprendizaje de máquina, procesando varias cantidades de pentabytes (Google, 2015). Las limitaciones de esta gran herramienta es que utilizan imágenes de poca resolución por lo que no es posible distinguir árboles en zonas urbanas y sus modelos tampoco están entrenados para que puedan determinar el total del revestimiento arbóreo que pueda existir ni calcular el Índice Verde Urbano (OMS, 2014).

Realizar un cálculo manual implicaría tener una gran cantidad de recursos tanto económico, humano y también de tiempo dado que el conteo de árboles y la estimación del área que toma el recubrimiento arbóreo en una ciudad de 344 km² de suelo (Municipio Guayaquil, 2017) implicaría una labor de gran magnitud. Por ello se vuelve determinante elaborar una herramienta tecnológica con ayuda de la inteligencia artificial que pueda reconocer los árboles de las zonas urbanas y que calcule su cobertura. De esta manera, se puede tener un ahorro en los mencionados recursos y, además, reducir el error que podría producir un conteo manual.

1.3 Objetivos

1.1.1 Objetivo General

Identificar el revestimiento arbóreo desde la ortofotografía de Guayaquil urbano utilizando técnicas de inteligencia artificial en segmentación de imágenes.

1.1.2 Objetivos Específicos

- Discriminar las copas de los árboles de los diferentes elementos que se encuentran en la ortofotografía para los datos de entrenamiento del modelo en segmentación de imágenes.
- Construir un modelo de segmentación de imágenes utilizando redes neuronales convolucionales para calcular el área de revestimiento arbóreo.
- Establecer el área de cobertura arbórea de Guayaquil urbano mediante los resultados obtenidos en el modelo de redes neuronales convolucionales para obtener un indicador que establezca si Guayaquil es carente o no de árboles.

1.4 Marco Teórico

1.1.3 Muestreo Aleatoria Simple

El muestreo es un proceso en donde se seleccionan elementos al azar con el fin de obtener una muestra con una cierta probabilidad de ocurrencia (Cochran, 1982). En el muestreo aleatorio simple (MAS) se toma al azar n elementos dentro de una población N , donde cada elemento tiene la misma probabilidad de ser escogido. Es el más sencillo dentro de todas las técnicas de muestreo y es usado como base para los demás métodos.

Para que una muestra sea de tipo MAS se debe cumplir con las siguientes condiciones:

- Los elementos de la población son equiprobables de ser seleccionados para la muestra

$$Prob(U_i) = \frac{1}{N} \quad (1)$$

- Cada muestra posible de tamaño n tiene la misma probabilidad de ser construida. Existen múltiples combinaciones de n casos de posibles muestras entre la población N

$$Prob(U) = \frac{n}{N} \quad (2)$$

1.1.3.1 Tamaño de la muestra

Determinar el tamaño de la muestra para estimar el parámetro de interés es un factor importante para el estudio que se esté realizando. El objetivo es lograr alcanzar la mayor precisión o representatividad posible de los parámetros de la población. Para la determinación del tamaño muestral se define de la siguiente manera:

$$n = \frac{N Z_{\alpha/2}^2 p q}{N e^2 + Z_{\alpha/2}^2 p q} \quad (3)$$

Donde:

n : tamaño de la muestra

N : tamaño de la población

$Z_{\alpha/2}$: variable estandarizada de la distribución normal

e : error

p : probabilidad de éxito

q : probabilidad de fracaso

1.1.4 Visión por Computadora

La visión artificial reside en la deducción automática de las propiedades y estructura del mundo tridimensional por medio de una o varias imágenes bidimensionales de aquel mundo. (Sánchez-Salmerón & Ricolfe-Viala, 2016). El objetivo de la visión computacional es extraer de una imagen las características para su interpretación y descripción por la computadora. Dentro de esta área de conocimiento se entrelazan los conceptos de óptica, electrónica, física de color, algorítmica, geometría, sistemas de computación, etc.

Se define a la imagen como la representación bidimensional del dinámico mundo tridimensional. Como sugiere en el gráfico 1.1, la imagen es el resultado de una

señal otorgada por un sensor, que transforma la información recibida por el espectro electromagnético en códigos numéricos.

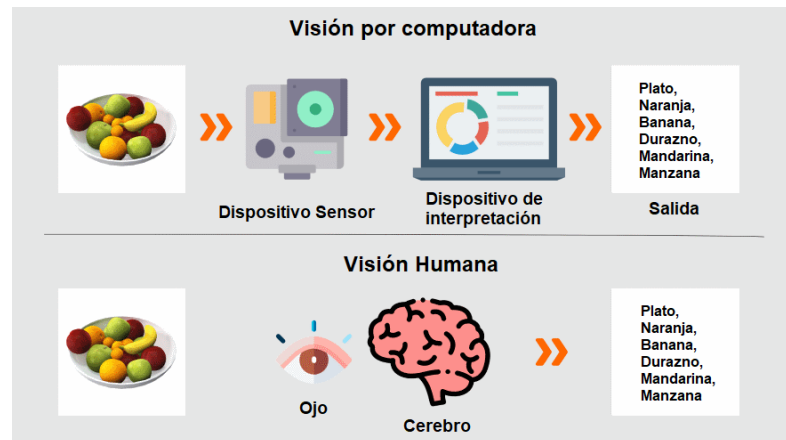


Gráfico 1.1 Visión por computadora

Fuente: <https://www.algorithmxlab.com>

De manera general, una imagen digital es representada por un vector o matriz con una dimensión de $N \times M$, que contiene en cada unidad de la matriz un valor discreto que mide el nivel de información de la correspondiente unidad, representado con un número finito de bits. De esta manera, se puede definir el concepto de imagen como una función de dos dimensiones discreta de la siguiente manera:

$$I(x, y) \mid 0 \leq x \leq N - 1, 0 \leq y \leq M - 1 \quad (4)$$

Donde N y M pueden tomar cualquier valor numérico que pertenezca al conjunto de número naturales y donde los valores de cada elemento son múltiplos de 2:

$$0 \leq I(x, y) \leq p - 1 \mid p = 2^q \quad (5)$$

Los valores que toman cada unidad de la imagen representan desde los niveles más oscuros de luminosidad hasta los niveles más claros. El nivel más oscuro está dado por el valor más bajo del intervalo y es representado por el negro y el nivel más claro es representado por el blanco con un valor más alto.

Es así como una imagen es una función bidimensional que entrega información para uno de sus elementos. En ese sentido, la visión por computadora extrae esa información para su descripción e interpretación por medio de la computadora.

1.1.5 Redes neuronales convolucionales (CNN)

Una red neuronal convolucional (CNN) es un tipo de red neuronal artificial utilizada en el reconocimiento y procesamiento de imágenes que está optimizada para procesar datos de píxeles. Por lo tanto, las redes neuronales convolucionales son los bloques de construcción fundamentales y básicos para la tarea de visión por computadora de la segmentación de imágenes (segmentación de CNN). La arquitectura de red neuronal convolucional consta de tres capas principales:

- **Capa convolucional:** esta capa ayuda a abstraer la imagen de entrada como un mapa de características mediante el uso de filtros y núcleos.
- **Capa de agrupación:** esta capa ayuda a reducir la resolución de los mapas de características al resumir la presencia de características en parches del mapa de características.
- **Capa completamente conectada:** las capas completamente conectadas conectan cada neurona en una capa con cada neurona en otra capa.

La combinación de las capas de una “CNN” permite que la red neuronal diseñada aprenda a identificar y reconocer el objeto de interés en una imagen. Las redes neuronales convolucionales simples están diseñadas para la clasificación de imágenes y la detección de objetos tal como hace referencia el gráfico 1.2.

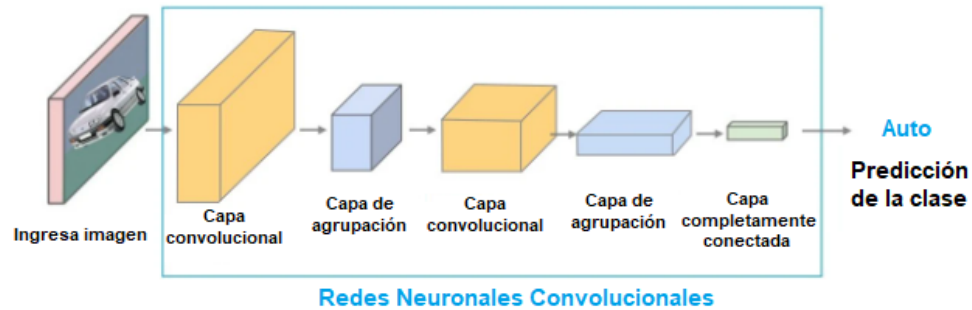


Gráfico 1.2 Redes Neuronales Convolucionales

Fuente: <https://www.researchgate.net>

1.1.5.1 Redes neuronales basadas en regiones (R-CNN)

“R-CNN” es un tipo de modelo de aprendizaje automático que se utiliza para tareas de visión por computadora, específicamente para la detección de objetos. Este modelo utiliza cuadros delimitadores en las regiones del objeto, que luego evalúa las redes convolucionales de forma independiente en todas las Regiones de interés (ROI) para clasificar múltiples regiones de imagen en la clase propuesta. La arquitectura “RCNN” fue diseñada para resolver tareas de detección de imágenes.

1.1.5.2 Redes neuronales convolucionales rápidas basadas en regiones (Fast y Faster R-CNN)

“Fast R-CNN” es una versión mejorada de las arquitecturas R-CNN con dos etapas:

- **Red de regiones propuestas (RPN).** RPN es una red neuronal que propone múltiples objetos que están disponibles dentro de una imagen en particular.
- **Rápido R-CNN.** Esto extrae características utilizando “RoIPool” (agrupación de regiones de interés) de cada cuadro candidato y realiza una clasificación y regresión del cuadro delimitador. RoIPool

es una operación para extraer un pequeño mapa de características de cada Rol en detección.

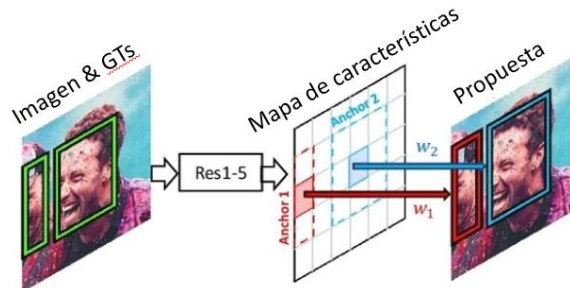


Gráfico 1.2 Redes neuronales convolucionales rápidas basadas en regiones (Fast y Faster R-CNN)

Fuente: <https://viso.ai>

“Faster R-CNN” aprende el mecanismo de atención con una red de regiones propuestas y una arquitectura “Fast R-CNN”. La razón por la que “Fast R-CNN” es más rápido que “R-CNN” es porque no tiene que enviar 2000 propuestas de región a la red neuronal convolucional cada vez. En cambio, la operación de convolución se realiza solo una vez por imagen y se genera un mapa de características a partir de ella. Además, Faster R-CNN es una forma optimizada de R-CNN porque está diseñado para mejorar la velocidad de cálculo (ejecutar R-CNN mucho más rápido). La principal diferencia entre Fast y Faster RCNN es que Fast R-CNN usa la búsqueda selectiva para generar regiones de interés, mientras que Faster R-CNN usa una “Red de región propuesta” (RPN).

1.1.5.3 Redes neuronales con máscara basadas en regiones.

“Mask R-CNN” es una red neuronal convolucional (CNN) y de vanguardia en términos de segmentación de imágenes y segmentación de instancias. “Mask R-CNN” se desarrolló sobre “Faster R-CNN”, una red neuronal convolucional basada en regiones. Esta variante de una red neuronal profunda detecta objetos en una imagen y genera una máscara de segmentación de alta calidad para cada instancia. La tarea de visión artificial en segmentación de imágenes es el proceso de dividir una imagen digital en múltiples segmentos (conjuntos de píxeles, también conocidos como objetos de imagen). Esta segmentación se utiliza para localizar objetos y límites (líneas, curvas, etc.). Hay 2 tipos principales de segmentación de imágenes que se incluyen en “Mask R-CNN”:

- Segmentación semántica
- Segmentación de instancias

“Mask R-CNN” se creó utilizando “Faster R-CNN”. Mientras que “Faster R-CNN” tiene 2 salidas para cada objeto candidato, una etiqueta de clase y un desplazamiento de cuadro delimitador, “Mask R-CNN” es la adición de una tercera rama que genera la máscara del objeto. La salida de máscara adicional es distinta de las salidas de clase y caja, lo que requiere la extracción de un diseño espacial mucho más fino de un objeto. En otras palabras, “Mask R-CNN” es una extensión de “Faster R-CNN” y funciona agregando una rama para predecir una máscara de objeto (región de interés) en paralelo con la rama existente para el reconocimiento del cuadro delimitador.

El elemento clave de “Mask R-CNN” es la alineación píxel a píxel, que es la principal pieza que falta en “Fast / Faster R-CNN”. “Mask R-CNN” adopta el mismo procedimiento de dos etapas con una primera etapa idéntica (que es RPN). En la segunda etapa, en paralelo a la predicción de la clase y el desplazamiento de la caja, “Mask R-CNN” también genera una máscara binaria para cada región de interés. Esto contrasta con los sistemas más

recientes, donde la clasificación depende de las predicciones de la máscara.

“Mask R-CNN” es fácil de implementar y entrenar dado el marco “Faster R-CNN”, que facilita una amplia gama de diseños de arquitectura flexible. Además, la rama de máscara solo agrega una pequeña sobrecarga computacional, lo que permite un sistema y una experimentación rápida.

1.1.5.3.1 Segmentación semántica

La segmentación semántica clasifica cada píxel en un conjunto fijo de categorías sin diferenciar las instancias de objetos. En otras palabras, la segmentación semántica se ocupa de la identificación / clasificación de objetos similares como una sola clase desde el nivel de píxeles.

1.1.5.3.2 Segmentación de imágenes

La segmentación de instancias, o reconocimiento de instancias, se ocupa de la detección correcta de todos los objetos en una imagen y al mismo tiempo segmenta con precisión cada instancia. Es, por tanto, la combinación de detección de objetos, localización de objetos y clasificación de objetos. En otras palabras, este tipo de segmentación va más allá para dar una clara distinción entre cada objeto clasificado como instancias similares.

1.1.6 Transferencia de aprendizaje

El aprendizaje por transferencia es la reutilización de un modelo previamente entrenado sobre un nuevo problema. Es muy popular en el aprendizaje profundo porque puede entrenar redes neuronales profundas con relativamente pocos datos. En lugar de comenzar el proceso de aprendizaje desde cero, comienza con patrones aprendidos al resolver una tarea relacionada. Con el aprendizaje por

transferencia, se intenta aprovechar lo aprendido en una tarea para mejorar la generalización en otra. Transferimos los pesos que una red ha aprendido en la "tarea A" a una nueva "tarea B".

En la visión por computadora, por ejemplo, las redes neuronales generalmente intentan detectar bordes en las capas anteriores, formas en la capa intermedia y algunas características específicas de la tarea en las capas posteriores. En el aprendizaje por transferencia, se utilizan las capas inicial y media y solo reentrenamos las últimas capas. Ayuda a aprovechar los datos etiquetados de la tarea en la que se entrenó inicialmente. Además, se trata de transferir la mayor cantidad de conocimiento posible de la tarea anterior en la que se entrenó el modelo a la nueva tarea en cuestión. Este conocimiento puede presentarse de diversas formas según el problema y los datos.

1.1.7 Media de precisión

La precisión media es una medida que combina la recuperación y la precisión para obtener resultados de recuperación clasificados. Para una necesidad de información, la precisión promedio es la media de las puntuaciones de precisión después de que se recupera cada documento relevante (Zhang E, 2019).

$$Precisión\ media = \frac{\sum_r P@r}{R} \quad (6)$$

CAPÍTULO 2

2. METODOLOGÍA

En el presente capítulo, se describirá los procedimientos empleados para el estudio de la cobertura arbórea de la ciudad de Guayaquil urbano, tomando la ortofotografía a través del software SAS Planet, con licencia GNU.

Para la calcular la dimensión que tiene el recubrimiento arbóreo en la ciudad de Guayaquil urbano, se utiliza aprendizaje profundo, el cual permite que la visión por computadora pueda observar y calcular de forma eficiente las grandes cantidades de datos que comprenden dentro de la ortofotografía de la urbe.

2.1 Selección

Para obtener el set de datos se usó la ortofotografía de Guayaquil, a través de SAS Planet en calidad de 0.30 metros píxel dado que, con esta calidad de fotografía, el modelo de segmentación podrá reconocer árboles de tamaños pequeños. Delimitando la ciudad de Guayaquil en un polígono, se obtuvo un total de 54,845 imágenes en tamaño de 256 x 256 pixeles.

En el gráfico 2.1.1 se comparte una imagen aleatoria de las obtenidas correspondiente al barrio la Garzota. Dicha imagen es una referencia para demostrar el tipo de imagen a usar en el presente proyecto para la detección de las copas de los árboles.

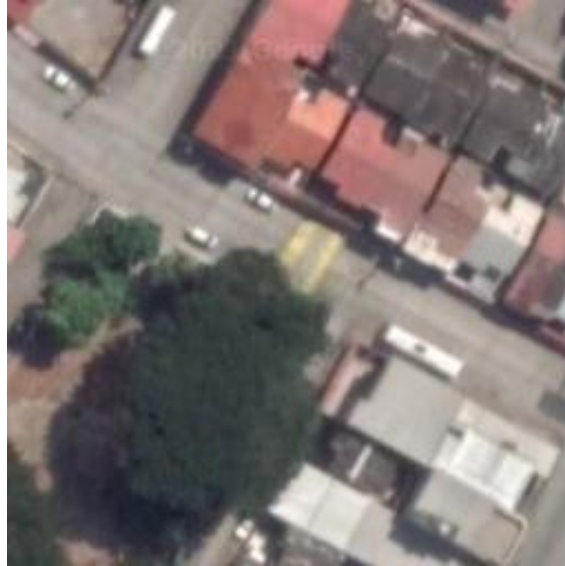


Gráfico 2.1 Imagen satelital de Guayaquil Urbano, barrio Garzota, coordenadas X: 145802, Y: 265265

Fuente: SAS Planet

2.2 Muestreo Aleatorio Simple

Considerando que el entrenamiento del modelo de redes neuronales para segmentación de imágenes se realiza con el 30% de los datos, se trabajaría con un total de imágenes etiquetadas, pero humanamente este trabajo no se podría realizar en el corto plazo. Por esta razón se aplicará MAS con las parroquias que tienen más de 100 imágenes para su entrenamiento ya que, con esta metodología de muestreo estadístico, se tomará una muestra que represente a la población.

A fin de obtener el tamaño de muestra requerido, se trabaja en una prueba piloto para determinar la probabilidad de éxito en el MAS, tomando 45 imágenes aleatorias de tres barrios distintos de la ciudad. En el norte, se toma 15 imágenes del barrio Bastión Popular, del centro, se toma 15 imágenes del barrio Urdesa y del sur, se toma 15 imágenes del barrio Fertisa. Seleccionadas las ortofotografías, se realiza el conteo de píxeles de cada una en el revestimiento arbóreo mediante la librería "Pillow" de Python y se asigna el color a contar, en este caso, el verde.

Dada la tabla 2.2.1, en las 15 imágenes seleccionadas aleatoriamente del barrio Urdesa se promediaron los píxeles que cumplen la condición de ser color verde en relación con el total de píxeles de la muestra, obteniendo como resultado que 19.31% de los píxeles cumplen con dicha condición.

	IMAGEN	PIXELES_SEG	PARTICIPACION
0	x_145777_y_265297.jpg	16216	0.247437
1	x_145776_y_265295.jpg	7610	0.116119
2	x_145775_y_265304.jpg	8857	0.135147
3	x_145775_y_265301.jpg	7894	0.120453
4	x_145775_y_265302.jpg	9676	0.147644
5	x_145776_y_265293.jpg	5675	0.086594
6	x_145775_y_265303.jpg	7656	0.116821
7	x_145775_y_265300.jpg	13482	0.205719
8	x_145777_y_265293.jpg	6764	0.103213
9	x_145777_y_265294.jpg	24081	0.367447
10	x_145776_y_265291.jpg	27943	0.426376
11	x_145776_y_265292.jpg	16788	0.256165
12	x_145777_y_265295.jpg	13382	0.204193
13	x_145776_y_265294.jpg	3834	0.058502
14	x_145777_y_265296.jpg	20034	0.305695

Tabla 2.1 Barrio Urdesa, conteo de píxeles por imagen

Fuente: Información propia

Considerando los resultados de la tabla 2.2.2, se calcula el promedio de la cantidad de pixeles de color verde que aparecen en las imágenes satelitales del barrio Bastión Popular tomadas aleatoriamente, como resultado se obtuvo que 15.84% de los pixeles cumplen con la condición en relación con el total de pixeles de la muestra.

	IMAGEN	PIXELES_SEG	PARTICIPACION
0	x_145764_y_265203.jpg	30889	0.471329
1	x_145764_y_265201.jpg	11109	0.169510
2	x_145765_y_265194.jpg	9064	0.138306
3	x_145765_y_265195.jpg	4760	0.072632
4	x_145764_y_265202.jpg	38796	0.591980
5	x_145766_y_265201.jpg	5798	0.088470
6	x_145767_y_265195.jpg	3615	0.055161
7	x_145770_y_265199.jpg	1543	0.023544
8	x_145769_y_265199.jpg	5819	0.088791
9	x_145769_y_265200.jpg	3931	0.059982
10	x_145768_y_265200.jpg	10194	0.155548
11	x_145767_y_265196.jpg	8156	0.124451
12	x_145769_y_265198.jpg	4236	0.064636
13	x_145767_y_265197.jpg	11612	0.177185
14	x_145766_y_265200.jpg	6253	0.095413

Tabla 2.2 Barrio Bastión Popular, conteo de pixeles por imagen

Fuente: Información propia

De acuerdo con los resultados mostrados en la tabla 2.2.3, se calcula el promedio de la participación de los pixeles por imagen, obteniendo como resultado que 13.34% corresponden a pixeles de color verde del total de las imágenes satelitales tomadas aleatoriamente.

	IMAGEN	PIXELES_SEG	PARTICIPACION
0	x_145783_y_265442.jpg	7293	0.111282
1	x_145784_y_265413.jpg	6314	0.096344
2	x_145785_y_265417.jpg	3475	0.053024
3	x_145783_y_265443.jpg	21779	0.332321
4	x_145784_y_265412.jpg	10320	0.157471
5	x_145785_y_265416.jpg	9299	0.141891
6	x_145783_y_265429.jpg	5445	0.083084
7	x_145783_y_265431.jpg	10678	0.162933
8	x_145783_y_265430.jpg	3322	0.050690
9	x_145783_y_265441.jpg	2227	0.033981
10	x_145785_y_265415.jpg	8896	0.135742
11	x_145783_y_265432.jpg	19626	0.299469
12	x_145785_y_265414.jpg	18135	0.276718
13	x_145783_y_265428.jpg	1648	0.025146
14	x_145785_y_265413.jpg	2716	0.041443

Tabla 2.3 Barrio Fertisa, conteo de pixeles por imagen

Fuente: Información propia

Como resultado de los 65,536 píxeles que representa cada imagen de Guayaquil tomadas para este muestreo, en promedio 10,590 píxeles correspondían al revestimiento arbóreo, es decir el 16%.

Conociendo la probabilidad de éxito, se establece los diferentes parámetros para el entrenamiento del modelo de redes neuronales:

N: 54,845

$Z_{\alpha/2}$: Considerando el 95% de confianza, el valor a tomar es de 1.96.

p: 0.16 de éxito

q: 0.84 de fracaso

e: 0.03 error

Finalmente, la cantidad de imágenes que se usa para el entrenamiento del modelo es de 567.

2.3 Etiquetado de imágenes

Para que el modelo a emplear pueda aprender a qué elemento debe segmentar dentro de la imagen, se realiza el etiquetado de imágenes en el cual se hace uso de la herramienta "Labelme". Herramienta de fuente abierta, en la cual se cargaron las 567 imágenes correspondiente al entrenamiento del modelo de acuerdo con el muestreo, y con ellas se delimitaron los revestimientos arbóreos que aparecían en las imágenes mediante polígonos irregulares, asignando su clase como "árbol" y su identificador como "1". Una vez rotulada la ortofotografía, se procede a descargar un json con la información de la imagen y así para cada una de ellas. Así como se destaca en la imagen a continuación.



Gráfico 2.2 Imagen etiquetada de Guayaquil Urbano, barrio Fertisa, coordenadas X:145783 Y:265442

Fuente: Label me – edición propia

2.4 Entrenamiento del modelo

Para el inicio de este proceso, implica tener dos sets de datos: uno para entrenamiento y otro para la validación, siendo el 70% de las imágenes etiquetadas (567) con sus respectivos json para el primer set y el 30% para el segundo set (256). Además de que sean cargados los dos datasets, se toma el modelo “Máscara – Región de Redes Neuronales Convolucionales” usando “ResNet-101 (He et al., 2016) para entrenar el set de datos de Objetos Comunes en Contexto, “COCO”, ya que este modelo ha sido entrenado con más de 330,000 imágenes que contienen más de 200,000 etiquetas para segmentación de objetos, reconocimiento de objetos de acuerdo con el contexto, segmentación de imágenes en superpixel, entre otras particularidades. (Lin et al., 2014). Conociendo los beneficios de este modelo, se usará transferencia de aprendizaje para reentrenar dicho modelo ya que se pretende enseñar al modelo a que reconozca las copas de los árboles desde una toma vertical de los satélites.

Para el reentrenamiento se usa la librería “PixelLib” del software “Python” en el cual se definió el “batch size”, o el número de datos que tiene una iteración en cada ciclo, se estableció en 4 y el “epoch” o iteración de cada ciclo en 30. Una vez que el reentrenamiento ha concluido, se valora la precisión del reentrenamiento usando el indicador “mAP”

2.5 Validación del modelo

Una vez entrenado el modelo de Máscara – Región de Redes Neuronales Convolucionales, se valida el mismo usando el método “AP” o media de precisión en el cual se determina la calidad del modelo. Para calcular dicho método, se establece el grado de coincidencia entre la predicción de la imagen y su etiqueta. Este grado de coincidencia o también llamado intersección sobre la unión, se define de la siguiente manera:

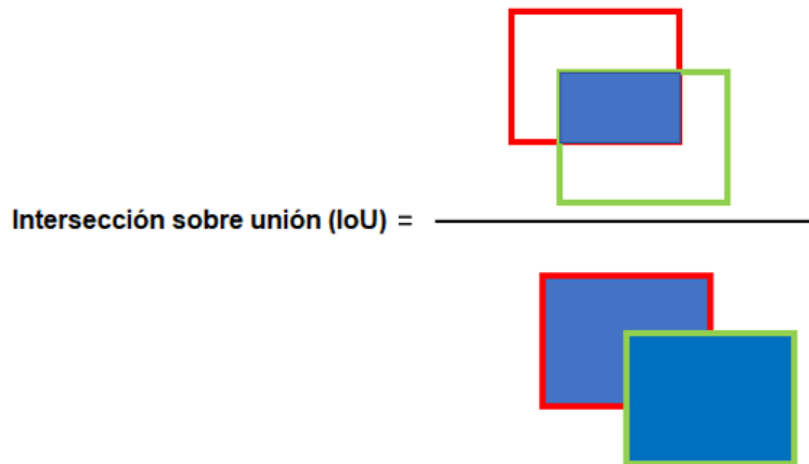


Gráfico 2.3 Fórmula de precisión media

Fuente: <https://medium.com>

Para la IoU se establece el umbral 0.5, es decir una precisión del 50%. Si la predicción del modelo supera esta participación de la etiqueta asignada se

considera como verdadero positivo, caso contrario, es decir, menos del umbral, se considera falso positivo. Por lo tanto, tenemos:

- Verdadero positivo: predicción que supera el 50% de precisión.
- Falso positivo: predicción que está por debajo del 50% de precisión.
- Falso negativo: elemento no detectado.
- Verdadero negativo: Todos los cuadros posibles que determinan lo que no se quiere identificar. Este indicador no es usado en el método de AP.

Para el indicador AP, se necesita conocer la exactitud y la sensibilidad:

$$Exactitud = \frac{VP}{VP+FP} \quad (7)$$

$$Sensibilidad = \frac{VP}{VP+FN} \quad (8)$$

Una vez calculados los indicadores previamente mencionados, se desea calcular el área bajo la curva que generan ambos. Dado que esto es un cálculo complejo, se usa la interpolación de 11 puntos que intenta resumir la curva de la precisión por la sensibilidad, promediando la precisión entre dichos 11 niveles de recuperación igualmente espaciados: $[0, 0.1, 0.2, \dots, 1]$

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \dots, 1\}} p_{interp}(r) \quad (9)$$

Finalmente, aplicando la fórmula previa, se determina el AP y se conoce la calidad del modelo re-entrenado. Para el presente proyecto, el modelo tuvo una precisión del 0.7 identificando el revestimiento arbóreo de las imágenes satelitales.

2.6 Aplicación del modelo re-entrenado.

Luego de conocer el AP, se realiza la inferencia en la segmentación del revestimiento arbóreo aplicando el modelo obtenido en las ortofotografías de Guayaquil urbano que previamente se han seleccionado, pero para un resultado integral en el que no solo consten las imágenes con el reconocimiento de la

cobertura arbórea, se ha creado una matriz de datos en el cual consta el nombre la imagen segmentada, la cantidad de elementos detectados y la cantidad total de pixeles de los elementos detectados.

1	Crear data frame
2	Ingreso de imagen 1 hasta total de imágenes
3	Inferencia: detección de cobertura arbórea en imagen
4	Conteo de elementos detectados
5	Suma de pixeles en los elementos detectados
6	Salida Data frame con "Nombre de imagen", "Número de elementos detectados", "Total pixeles en elementos detectados"

Gráfico 2.4 Algoritmo de la inferencia del modelo

Fuente: Elaboración propia

2.7 Cálculo de la cobertura arbórea

Como se mencionó en el punto 1.1, las ortofotografías son de calidad 0.3 metros por píxel, por lo que la superficie en un solo píxel sería de 0.09 m². Conociendo esto, desde la matriz de datos se toma la cantidad de pixeles totales de la segmentación de árboles que identifico el modelo entrenado, con ello se realiza el cálculo:

$$\text{Revestimiento arbóreo} = 0.09m^2 * \text{total pixeles de los elementos detectados}$$

CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS

En el capítulo actual se muestran los resultados del total de la cobertura arbórea de Guayaquil urbano mediante el modelo re-entrenado de segmentación de imágenes usando redes neuronales convolucionales.

3.1 Resultados

Dada la tabla 3.1 se muestra como resultado que, en 54,845 imágenes satelitales se obtuvo 948,774,590 píxeles de revestimiento arbóreo, esto quiere decir que en 323.49 km² de superficie territorial de Guayaquil urbano, 85.39 km² corresponden a la cobertura arbórea.

SECTOR	N° DE IMÁGENES	KM2 SUPERFICIE	N° PÍXELES COBERTURA ARBÓREA	KM2 COBERTURA ARBOREA	COBERTURA/SUPERFICIE
GUAYAQUIL URBANO	54,845	323.49	948,774,590	85.39	26.4%

Tabla 3.1 Guayaquil Urbano, cobertura arbórea/superficie

Fuente: Información propia

Para un análisis más exhaustivo y conocer como se distribuye el revestimiento arbóreo, se dividió en 4 sectores la ciudad: Norte, Sur, Centro y Oeste.

En la tabla 3.1.1, se evidencia que el Centro de la ciudad es un sector crítico dado que en 8.87 km² de superficie, tiene 1.12 km² de cobertura arbórea, es decir, tiene

un 12.6% de revestimiento arbóreo con relación a su extensión territorial. El sector sur, también tiene poca participación considerando que en los 62.90 km², su cobertura arbórea es de 8.62 km², en valor porcentual del 13.7%. El sector del norte tiene una cobertura arbórea media del 24.7% dado que su superficie es de 170.74 km² y su revestimiento arbóreo es de 42.13 km². Finalmente, el sector con mayor recubrimiento arbóreo es el oeste ya que en sus 80.98 km² de territorio tiene 33.53 km² extensión arbórea, es decir el 41.4% con relación a su superficie.

SECTOR	KM2 SUPERFICIE	% KM2 SUPERFICIE	KM2 COBERTURA ARBOREA	% KM2 COBERTURA ARBOREA	COBERTURA/SUPERFICIE
CENTRO	8.87	2.7%	1.12	1.3%	12.6%
NORTE	170.74	52.8%	42.13	49.3%	24.7%
OESTE	80.98	25.0%	33.53	39.3%	41.4%
SUR	62.90	19.4%	8.62	10.1%	13.7%
TOTAL GENERAL	323.49	100.0%	85.39	100.0%	26.4%

Tabla 3.2 Sectores de Guayaquil Urbano, cobertura arbórea/superficie

Fuente: Información propia

De forma gráfica, se proyecta lo indicado en la tabla 3.1.2 en el gráfico 3.1 demostrando los límites de los sectores indicados y los km² de territorio y de revestimiento arbóreo contiene dichos sectores.

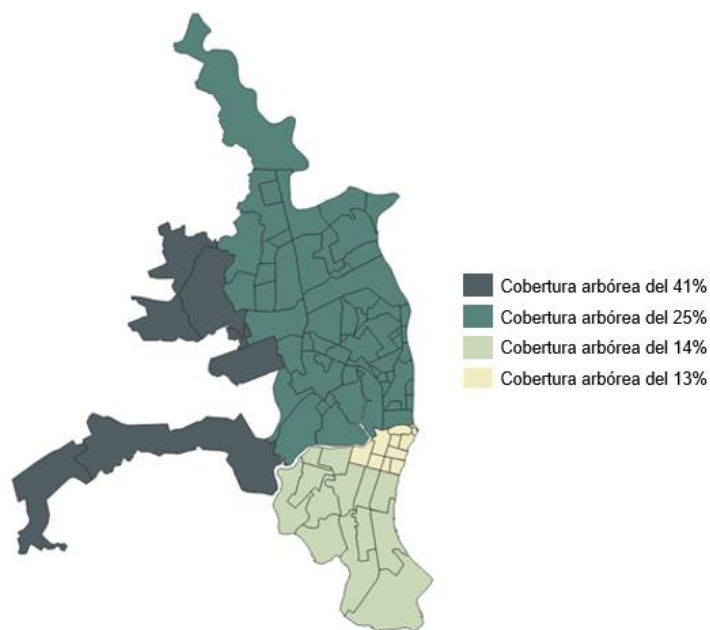


Gráfico 3.1 Cobertura arbórea de Guayaquil Urbano en porcentaje

Fuente: Información propia

3.2 Análisis situacional de los resultados

Para una mejor comprensión de los resultados, se organizó los datos de mayor a menor en recubrimiento arbóreo. El fin de este orden, es aplicar el principio de Pareto. Dado al gran volumen de datos, se realiza una tabla resumen con la información obtenida del mencionado principio.

De acuerdo con la tabla 3.1.2, se evidencia que el 80% de la cobertura arbórea se encuentra en el 38% de territorio urbano, en valores totales sería que en 122.93 km² de extensión territorial, se encuentra el 68.31 km² de revestimiento arbóreo. Y a su vez, esto indica que en el 62% del territorio contiene el 20% de la cobertura arbórea restante, en valores totales sería que 200.56 km² de la superficie territorial, 17.08 km² es del revestimiento arbóreo.

DISTRIBUCIÓN	KM2 COBERTURA ARBOREA	KM2 SUPERFICIE	% KM2 SUPERFICIE	N° IMÁGENES
80% DE COBERTURA ARBÓREA	68.31	122.93	38%	20,841
20% DE COBERTURA ARBÓREA	17.08	200.56	62%	34,004
TOTAL GENERAL	85.39	323.49	100%	54,845

Tabla 3.3 Distribución de la cobertura arbórea de Guayaquil Urbano

Fuente: Información propia

CAPÍTULO 4

4. CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones

El principal objetivo de este proyecto es identificar el revestimiento arbóreo desde la ortofotografía de Guayaquil urbano utilizando técnicas de inteligencia artificial en segmentación de imágenes; para lo cual se realizó un muestreo que con su resultado se tomó la cantidad de ortofotografías necesarias para el entrenamiento del modelo, una vez entrenado el modelo se evaluó la precisión. Conociendo dicha precisión, se ejecutó el modelo en las imágenes satelitales; con el total de los píxeles obtenidos de las imágenes segmentadas, se calcula la cobertura arbórea; y con ello, se llegaron a las siguientes conclusiones:

- Se implementó la técnica de muestreo aleatorio simple, conociendo que 597 imágenes satelitales contienen suficiente información de la cobertura arbórea de Guayaquil urbano para entrenar el modelo de segmentación de imágenes.
- Se realizó un modelo de identificación de la cobertura arbórea mediante transferencia de aprendizaje, usando redes neuronales convolucionales del cual su precisión fue del 70%.
- A partir del resultado de la ejecución del modelo en las ortofotografías de Guayaquil urbano, se conoció que 948,774,590 píxeles correspondían al revestimiento arbóreo, dando como resultado 85.39 Km². Con ello se divide en 4 sectores, de las cuales el centro y sur de la ciudad tienen una baja extensión de cobertura arbórea, específicamente un 14% para el sur y 13% para el centro en relación con su superficie territorial. Para el norte de la ciudad, se conoce que la participación de revestimiento arbóreo es del 25% y finalmente, el sector con mayor extensión arbórea es el oeste de la urbe con 41%.

- Partiendo del Principio de Pareto, se conoce que el 38% de la extensión territorial contiene el 80% de la cobertura arbórea, en otras palabras, el 20% de la cobertura restante se encuentra en el 62% de la superficie de Guayaquil urbano.

4.2 Recomendaciones

De acuerdo con los resultados obtenidos y a las conclusiones previamente descritas, se tienen las siguientes recomendaciones:

- Dado que la precisión del modelo es del 0.70 se recomienda optimar este indicador para mejores resultados y reducir errores.
- Para trabajos futuros se recomienda trabajar sectorizando por barrios y zonas censales dado que con ello se podría identificar zonas con mayor concentración arbórea, así como identificar zonas críticas.
- Como información adicional al proyecto, se recomienda crear indicadores relacionados entre la cobertura arbórea y la cantidad de habitantes en la ciudad.

BIBLIOGRAFÍA

Universidad UNADE. (2020, noviembre 17). Qué es la inteligencia artificial y sus ventajas. <https://unade.edu.mx/que-es-la-inteligencia-artificial/>

Universidad Austral de Chile. (2021, octubre 26) Uach.cl. El futuro computacional está en el Deep Learning, de <https://www.uach.cl/direccion-de-tecnologias-de-informacion/canalnoticia/18960>

Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. Sensors (Basel, Switzerland), <https://doi.org/10.3390/s18020401>

Arrabales, R. (2016, marzo 29). Deep Learning: qué es y por qué va a ser una tecnología clave en el futuro de la inteligencia artificial. Xataka.com; Xataka. <https://www.xataka.com/robotica-e-ia/deep-learning-que-es-y-por-que-va-a-ser-una-tecnologia-clave-en-el-futuro-de-la-inteligencia-artificial>

Anónimo. (2011, julio 9). Déficit de áreas verdes persiste, pese a que se suman proyectos. El Universo. <https://www.eluniverso.com/2011/07/09/1/1445/deficit-areas-verdes-persiste-pese-suman-proyectos.html/>

González, D. (2020, enero 16). Entre más árboles, menos calor. La importancia de reforestar y cómo hacerlo en Guayaquil. El Universo. <https://www.eluniverso.com/noticias/2020/01/16/nota/7693365/siembra-arbol-guayaquil/>

Guayaquil aún lucha por aumentar sus espacios verdes. (2020, diciembre 20). Primicias.ec. <https://www.primicias.ec/noticias/sociedad/guayaquil-aumentar-espacios-verdes/>

Sánchez-Salmerón & Ricolfe-Viala, (2016). Conceptos y métodos de la visión por computadora. Grupo de visión del Comité Español de automática. ISBN: 978-84-608-8933-5

Morales, A. (2020, enero 30). Google Earth Engine y cómo acceder a sus algoritmos desde QGIS. Mappinggis.com. <https://mappinggis.com/2020/01/google-earth-engine-y-como-acceder-a-sus-algoritmos-desde-qgis/>

Wen Q, Jiang K, Wang W, Liu Q, Guo Q, Li L, Wang P. (2019, febrero 19) Automatic Building Extraction from Google Earth Images under Complex Backgrounds Based on Deep Instance Segmentation Network. Sensors. <https://doi.org/10.3390/s19020333>

Xu, Y., Sun, Z., Hoegner, L., Stilla, U., & Yao, W. (2018). Instance segmentation of trees in urban areas from MLS point clouds using supervoxel contexts and graph-based optimization. 2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS), 1–5.