

**DETECCIÓN AUTOMÁTICA DEL NIVEL DE ESTRATIFICACIÓN  
SOCIOECONÓMICO URBANO USANDO REDES NEURONALES  
CONVOLUCIONALES SOBRE IMÁGENES SATELITALES CON  
INFORMACIÓN AUMENTADA**

**DANIEL ALCIDES CARVAJAL PATIÑO**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERIAS FISICOMECHANICAS  
ESCUELA DE INGENIERIA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA**

**2018**

**DETECCIÓN AUTOMÁTICA DEL NIVEL DE ESTRATIFICACIÓN  
SOCIOECONÓMICO URBANO USANDO REDES NEURONALES  
CONVOLUCIONALES SOBRE IMÁGENES SATELITALES CON  
INFORMACIÓN AUMENTADA**

**DANIEL ALCIDES CARVAJAL PATIÑO**

**TRABAJO DE GRADO PARA OPTAR POR EL TITULO DE  
INGENIERO DE SISTEMAS**

**Director**

**FABIO MARTINEZ CARRILLO**

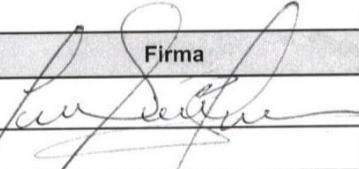
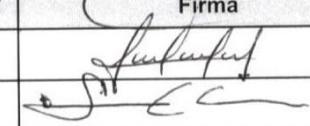
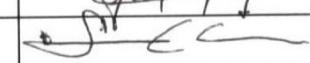
**Ph.D. EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**

**Codirector**

**Ph.D RAUL RAMOS POLLAN**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERIAS FISICOMECHANICAS  
ESCUELA DE INGENIERIA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA**

**2018**

Codigo:	19029	Fecha Presentacion:	15-dic-2017
<b>Título:</b> Detección automática del nivel de estratificación socioeconómico urbano usando redes neuronales convolucionales sobre imágenes satelitales con información aumentada.			
Nota Proyecto:	4.8	Fecha Registro Nota:	23-agosto-2018
Estado:	APROBADO		
Tipo Trabajo:	INVESTIGACION		
<b>Estudiantes</b>			
Código	Nombre	Programa Académico	
2132817	CARVAJAL PATIÑO DANIEL ALCIDES	11-INGENIERIA DE SISTEMAS	
<b>Directores</b>			
Documento	Nombre	Clase	Firma
C-13929892	FABIO MARTINEZ CARRILLO	DIRECTOR	
<b>Calificadores</b>			
Documento	Nombre	Firma	
C-91075781	GABRIEL RODRIGO PEDRAZA FERREIRA		
C-63493826	SONIA CRISTINA GAMBOA SARMIENTO		



## **ENTREGA DE TRABAJOS DE GRADO, TRABAJOS DE INVESTIGACION O TESIS Y AUTORIZACIÓN DE SU USO A FAVOR DE LA UIS**

Yo, Daniel Alcides Carvajal Patiño, mayor de edad, vecino de Bucaramanga, identificado con la Cédula de Ciudadanía No. 1057598147 de Sogamoso - Boyacá, actuando en nombre propio, en mi calidad de autor del trabajo de grado, del trabajo de investigación, o de la tesis denominada(o):

Detección automática del nivel de estratificación socioeconómico urbano usando redes neuronales convolucionales sobre imágenes satelitales con información aumentada, hago entrega del ejemplar respectivo y de sus anexos de ser el caso, en formato digital o electrónico (CD o DVD) y autorizo a LA UNIVERSIDAD INDUSTRIAL DE SANTANDER, para que en los términos establecidos en la Ley 23 de 1982, Ley 44 de 1993, decisión Andina 351 de 1993, Decreto 460 de 1995 y demás normas generales sobre la materia, utilice y use en todas sus formas, los derechos patrimoniales de reproducción, comunicación pública, transformación y distribución (alquiler, préstamo público e importación) que me corresponden como creador de la obra objeto del presente documento. PARÁGRAFO: La presente autorización se hace extensiva no sólo a las facultades y derechos de uso sobre la obra en formato o soporte material, sino también para formato virtual, electrónico, digital, óptico, uso en red, Internet, extranet, intranet, etc., y en general para cualquier formato conocido o por conocer.

EL AUTOR – ESTUDIANTE, manifiesta que la obra objeto de la presente autorización es original y la realizó sin violar o usurpar derechos de autor de terceros, por lo tanto la obra es de su exclusiva autoría y detenta la titularidad sobre la misma. PARÁGRAFO: En caso de presentarse cualquier reclamación o acción por parte de un tercero en cuanto a los derechos de autor sobre la obra en cuestión, EL AUTOR / ESTUDIANTE, asumirá toda la responsabilidad, y saldrá en defensa de los derechos aquí autorizados; para todos los efectos la Universidad actúa como un tercero de buena fe.

Para constancia se firma el presente documento en dos (02) ejemplares del mismo valor y tenor, en Bucaramanga, a los 24 días del mes de Agosto de Dos Mil dieciocho 2018

**EL AUTOR / ESTUDIANTE:**

Daniel Alcides Carvajal Patiño

CONTENIDO	pág.
<b>INTRODUCCION . . . . .</b>	<b>13</b>
<b>1. OBJETIVOS . . . . .</b>	<b>14</b>
1.1. OBJETIVO GENERAL . . . . .	14
1.2. OBJETIVOS ESPECIFICOS . . . . .	14
<b>2. MARCO TEÓRICO . . . . .</b>	<b>15</b>
2.1. ESTRATIFICACIÓN SOCIAL . . . . .	15
2.2. MACHINE LEARNING . . . . .	16
2.2.1. REDES NEURONALES . . . . .	16
2.2.2. REDES NEURONALES CONVOLUCIONALES . . . . .	18
2.2.3. SEGMENTACION SEMANTICA . . . . .	20
<b>3. DESARROLLO DEL PROYECTO . . . . .</b>	<b>22</b>
3.1. FUENTES DE DATOS . . . . .	23
3.2. DATASET . . . . .	24
3.3. REDES NEURONALES . . . . .	29
3.4. DETECCION DEL ESTRATO SOCIAL . . . . .	30
3.4.1. DATOS . . . . .	30
3.4.2. REDES . . . . .	31
3.4.2.1. FC-DenseNet56 . . . . .	31
3.4.2.2. Encoder-Decoder . . . . .	32
3.4.2.3. Encoder-Decoder-Skip . . . . .	33
3.4.2.4. FRRN-A . . . . .	34
3.4.2.5. MobileUNet . . . . .	35
3.4.2.6. MobileUNet-Skip . . . . .	35
3.4.3. METRICA . . . . .	36
<b>4. CONCLUSIONES . . . . .</b>	<b>46</b>

5. RECOMENDACIONES Y TRABAJO FUTURO . . . . .	47
6. LIMITACIONES Y PROBLEMAS . . . . .	48

## LISTA DE FIGURAS

pág.

1. Diagrama de red neuronal . . . . .	17
2. Arquitectura tipica de una CNN . . . . .	18
3. Detección de bordes mediante una convolución . . . . .	19
4. Max pooling with a 2x2 filter and stride = 2 . . . . .	20
5. Problemas de visión por computador . . . . .	21
6. Red Neuronal con arquitectura de convolución y deconvolución . . . . .	21
7. Datos a usar en el proyecto . . . . .	22
8. Modelo a usar en el proyecto . . . . .	23
9. División de Train y Test . . . . .	25
10. Estructura OVERLAECOBO . . . . .	25
11. Imágenes SIMPLE de OVERLAECOBO . . . . .	27
12. Imágenes COMPOUND de OVERLAECOBO . . . . .	27
13. Imágenes LABEL de OVERLAECOBO . . . . .	28
14. Correspondencia geográfica de OVERLAECOBO . . . . .	28
15. Segmentación Semántica . . . . .	29
16. Niveles de ZOOM usados . . . . .	30
17. Arquitectura FC-DenseNet . . . . .	32
18. Arquitectura SegNet . . . . .	33
19. Ejemplo de 'saltos' en una red neuronal . . . . .	33
20. Arquitectura FRRN . . . . .	34
21. Arquitectura UNet . . . . .	35
22. Ecuación IoU . . . . .	36
23. Ejemplo de predicción . . . . .	37
24. Resultados - IoU de cada nivel de ZOOM agrupados por modelo, a la izquierda 'Prueba 1', a la derecha 'Prueba 11' . . . . .	38
25. Resultados - Duración de Entrenamiento y Testeo para cada prueba, a la izquierda 'Prueba 1', a la derecha 'Prueba 11' . . . . .	39
26. Resultados - IoU para diferentes valores de 'Epoch' . . . . .	40

27. Resultados - IoU de cada nivel de ZOOM para la prueba 'Prueba 8' . . . . .	41
28. Resultados - IoU de cada nivel de ZOOM 'Prueba 2' . . . . .	42
29. Predicción para una misma zona urbana con diferente nivel de acercamiento satelital para las pruebas 2 y 8 . . . . .	43
30. Resultados - IoU por clase de la mejor configuración . . . . .	44

## **LISTA DE TABLAS**

	pág.
1. Cantidad de Imágenes en OVERLAECOBO . . . . .	26
2. Proporción de cada clase en Train - OVERLAECOBO . . . . .	26
3. Proporción de cada clase en Test - OVERLAECOBO . . . . .	27
4. Pruebas . . . . .	31
5. Presencia de cada clase en el dataset . . . . .	45

## RESUMEN

**TITULO:** DETECCIÓN AUTOMÁTICA DEL NIVEL DE ESTRATIFICACIÓN SOCIOECONÓMICO URBANO USANDO REDES NEURONALES CONVOLUCIONALES SOBRE IMÁGENES SATELITALES CON INFORMACIÓN AUMENTADA.

**AUTORES:** DANIEL ALCIDES CARVAJAL PATIÑO.

**PALABRAS CLAVE:** DANE, Machine learning, Deep Learning, Red Neuronal Convolucional, .

**DESCRIPCION:** Aplicando segmentación semántica para un dataset de imágenes satelitales de la ciudad de Bogotá se consiguió un nivel de predicción del estrato socioeconómico de las zonas urbanas de la ciudad. Con el dataset, denominado OVER-LAECOBO, el cual fue elaborado durante el desarrollo del proyecto debido a la baja disponibilidad de la información en el formato requerido para la segmentación, se consigue una predicción con un valor de IoU de 0.34. La elaboración del dataset se realizó usando el software SASplanet junto con archivos KML creados en base a la información del nivel socioeconómico disponible en el portal de mapas de la ciudad de Bogotá. Con dicho software se obtuvieron 3 tipos de imágenes satelitales geográficamente correspondidas entre sí y con 5 niveles de acercamiento (*ZOOM*), los 3 tipos de imágenes se denominaron como *SIMPLE*, *COMPOUND* y *LABEL*. Las imágenes *SIMPLE* son imágenes satelitales con 3 canales (RGB), las imágenes *COMPOUND* son imágenes satelitales con 4 canales (RGB + Información extra) donde el cuarto canal es el nivel de riesgo que tienen las mujeres durante horarios nocturnos en las distintas localidades de la ciudad, las imágenes *LABEL* son imágenes de 3 canales (RGB) pero únicamente con 7 colores, donde cada color representa un estrato social. El proceso de segmentación semántica se hizo entrenando las redes neuronales convolucionales ofrecidas por el Framework Semantic-Segmentation-Suite del usuario de github @GeorgeSeif. Se realizaron en total 66 pruebas, las cuales indican que la mejor red para determinar el nivel socioeconómico de una zona urbana es FC-DenseNet56, la cual obtuvo el mejor IoU con un valor de 0.34. Para las pruebas se usaron varias redes neuronales modificando los parámetros de Epoch y batchsize junto con 3 de los zooms disponibles por el dataset.

## ABSTRACT

**TITLE:** AUTOMATIC DETECTION OF THE URBAN SOCIOECONOMIC STRATIFICATION LEVEL USING CONVOLUTIONAL NEURAL NETWORKS ON SATELLITE IMAGES WITH INCREASED INFORMATION.

**AUTHORS:** DANIEL ALCIDES CARVAJAL PATIÑO.

**KEYWORDS:** Container, Cloud Computing, OpenStack, modules and services.

**DESCRIPTION:** A level of prediction was obtained using semantic segmentation for a dataset of satellite images from Bogota city. The dataset was named OVERLAECOBO and was created with information from the maps web page of Bogotá. OVERLACOBO have 3 types of satellite image, SIMPLE, COMPUND and LABEL, The SIMPLE image is RGB satellite image, the COMPOUND image are image with 4 channels, RGB + extra information, in this case the night danger for the women and the LABEL image are color maps where a color is a class for the prediction. The metric IoU (Intersection over Union) is used for determining the quality of predictions. The maximum IoU of the different test was 0.34 and the model used for this value was FC-DenseNet56 with a value of 70 for epoch and 1 for batch-size. The test was elaborated with the Framework Semantic-Segmentation-Suite of the GitHub user @GeorgeSeif.

## INTRODUCCION

La medición del nivel económico de una zona urbana, actualmente, conlleva un trabajo extenso, como lo expresa el DANE, “en el caso de las revisiones generales urbanas, así como en la estratificación rural se apoya en censos de vivienda”<sup>1</sup>. Es decir, se requiere la elaboración de una encuesta de gran tamaño, la cual consume mucho tiempo y personal. Posteriormente, si la encuesta no se realizó usando software de recolección de datos, es necesario realizar su tipeo, lo cual también requiere tiempo. Luego, como lo indica el DANE<sup>2</sup><sup>3</sup>, el cálculo final del estrato se realiza mediante modelos estadísticos y económicos especialmente calibrados para esta tarea.

En este contexto surgen varias interrogantes respecto a la capacidad de actualización de esta metodología: ¿Qué sucede cuando una ciudad tiene una alta tasa de desarrollo urbano?, ¿Cómo mantiene el gobierno actualizada la información de los estratos ante estas circunstancias?, ¿Qué tan efectiva es la metodología actual ante estos casos de alto desarrollo urbano?

Por tanto, el objetivo de este trabajo consiste en seleccionar redes neuronales convolucionales y evaluar su capacidad para determinar automáticamente el estrato socioeconómico usando imágenes satelitales e información adicional (información catastral, presencia y consumo de servicios, etc.), con el fin de presentar una alternativa que haga frente a las inquietudes planteadas.

No es la primera vez que se realiza una predicción del nivel socioeconómico utilizando técnicas de machine learning o Deep learning. Neal Jean en colaboración con varias

---

<sup>1</sup>DANE. Estratificación - Preguntas frecuentes. [en línea]. <[https://www.dane.gov.co/files/geostadistica/Preguntas\\_frecuentes\\_estratificacion.pdf](https://www.dane.gov.co/files/geostadistica/Preguntas_frecuentes_estratificacion.pdf)>[citado en 25 de mayo de 2018]

<sup>2</sup>DANE. Metodología de estratificación. [en línea]. <<http://www.dane.gov.co/index.php/servicios-al-ciudadano/servicios-de-informacion/estratificacion-socioeconomica>>[citado en 25 de mayo de 2018].

<sup>3</sup>DANE. Procedimiento del cálculo. [en línea]. <<http://www.dane.gov.co/files/geoestadistica/estratificacion/procedimientoDeCalculo.pdf>>[citado en 25 de mayo de 2018].

personas e instituciones realizó un modelo<sup>4 5 6</sup> capaz de predecir el nivel de pobreza en cinco países de África, usando imágenes satelitales y datos extra para dicha tarea. En Colombia, más específicamente en Medellín, también se han realizado modelos<sup>7</sup> para determinar niveles socioeconómicos de una zona urbana. En la Universidad EAFIT, usando técnicas de Machine Learning e imágenes satelitales logran medir los índices de pobreza de dicha ciudad.

---

<sup>4</sup>NEAL jean. Combining satellite imagery and machine learning to predict poverty. [en Línea]. <<http://sustain.stanford.edu/predicting-poverty/>>[citado en 25 de mayo de 2018]

<sup>5</sup>NEAL jean. Combining satellite imagery and machine learning to predict poverty. [en línea]. <<https://github.com/nealjean/predicting-poverty>>[citado en 25 de Mayo de 2018]

<sup>6</sup>NEAL Jean, MARSHALL Burke, † MICHAEL Xie, W. Matthew Davis, DAVID B. Lobell, STEFANO Ermon. Combining satellite imagery and machine learning to predict poverty. Science 353 (6301), p. 790-794. 2016

<sup>7</sup>EAFIT. Con imágenes satelitales miden los índices de pobreza en Medellín. [en línea]. <<http://www.eafit.edu.co/investigacion/revistacientifica/edicion-167/Paginas/com-imagenes-satelitales-miden-los-indices-de-pobreza-en-medellin.aspx>>[citado en 25 de mayo de 2018]

## **1. OBJETIVOS**

### **1.1. OBJETIVO GENERAL**

1. Seleccionar y evaluar redes convolucionales para la determinación del nivel socio económico urbano mediante el uso de imágenes satelitales e información adicional.

### **1.2. OBJETIVOS ESPECIFICOS**

1. Identificar fuentes de datos de imágenes satelitales e información adicional.
2. Diseñar y construir datasets integrando los datos obtenidos de las fuentes identificadas.
3. Seleccionar entre distintas arquitecturas de redes neuronales convolucionales existentes en la literatura y repositorios tecnológicos .
4. Entrenar las redes convolucionales probando configuraciones de datasets.
5. Evaluar el desempeño de las redes convolucionales con el uso de los distintos dataset.
6. Elegir la mejor configuración tanto de red convolucional como de conjunto de datos, teniendo en cuenta el desempeño obtenido.

## **2. MARCO TEÓRICO**

### **2.1. ESTRATIFICACIÓN SOCIAL**

“La estratificación social es un fenómeno presente en todas las sociedades. Los miembros se clasifican a sí mismos y a los otros basándose en jerarquías que vienen dadas por diversos factores”<sup>8</sup> y no es algo nuevo, la antigua Mesopotamia contaba con una división social cuyos miembros iban desde el rey y su familia, en el estrato más alto, hasta los esclavos en el más bajo.

Más que una simple división, la estratificación representa la desigualdad existente en una sociedad. Cada uno de los estratos, niveles o grupos sociales indica la capacidad de acceso a recursos, oportunidades, bienes y servicios por parte de las personas pertenecientes a cada nivel. Dicha estratificación es necesaria para la tarea que llevan a cabo los gobiernos en contra de la desigualdad, dado que se suelen crear programas que beneficien a las personas de los niveles más bajos y cobrar mayores impuestos a las personas de los niveles más altos.

En el caso de Colombia, se maneja una estratificación de 6 niveles.

“De éstos 6, los estratos 1, 2 y 3 corresponden a estratos bajos que albergan a los usuarios con menores recursos, los cuales son beneficiarios de subsidios en los servicios públicos domiciliarios; los estratos 5 y 6 corresponden a estratos altos que albergan a los usuarios con mayores recursos económicos, los cuales deben pagar sobrecostos (contribución) sobre el valor de los servicios públicos domiciliarios. El estrato 4 no es beneficiario de subsidios, ni debe pagar sobrecostos, paga exactamente el valor que la empresa defina como costo de prestación del servicio.”<sup>9</sup>

Dicha estratificación no es tarea fácil. Se requiere gran cantidad de variables por cada vivienda, como las características de la zona en la que se ubica, el tamaño, materiales en que fue fabricada, entre otras. Lo que conlleva a una tarea de recolección de datos

---

<sup>8</sup>GONZÁLEZ Vanessa. ¿Qué es la estratificación social? [en línea]. <<https://www.lifeder.com/estratificacion-social/>>[citado en 31 de mayo del 2018].

<sup>9</sup>DANE, Estratificación - Preguntas frecuentes. Op. cit., p. 1.

bastante amplia.

## 2.2. MACHINE LEARNING

“Machine learning es una disciplina científica del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente. Aprender en este contexto quiere decir identificar patrones complejos en millones de datos. La máquina que realmente aprende es un algoritmo que revisa los datos y es capaz de predecir comportamientos futuros. Automáticamente, también en este contexto, implica que estos sistemas se mejoran de forma autónoma con el tiempo, sin intervención humana.”<sup>10</sup>

Se han llevado muchos desarrollos y avances en distintos campos usando técnicas de machine learning. Uno de las más recientes y populares fue la máquina de Google AlphaGo que venció al mejor jugador a nivel mundial de GO, Ke Jie. Wikipedia usa técnicas de machine learning para detectar saboteos en su enciclopedia. Otros usos de machine learning consisten en la detección de objetos, patrones o enfermedades incluso predicción de tráfico urbano y precios de bienes.

Aunque en el fondo, independientemente del campo en el que se trabaje, las técnicas de machine learning son las mismas, existen tareas mucho más complejas que otras. Por ejemplo, hoy en día, es mucho más fácil predecir el valor de una casa que predecir el valor del dólar. Esta dificultad se debe a que cada tarea requiere “ajustar” los datos y los algoritmos para que se puedan realizar predicciones o detecciones con un nivel de tolerancia aceptable.

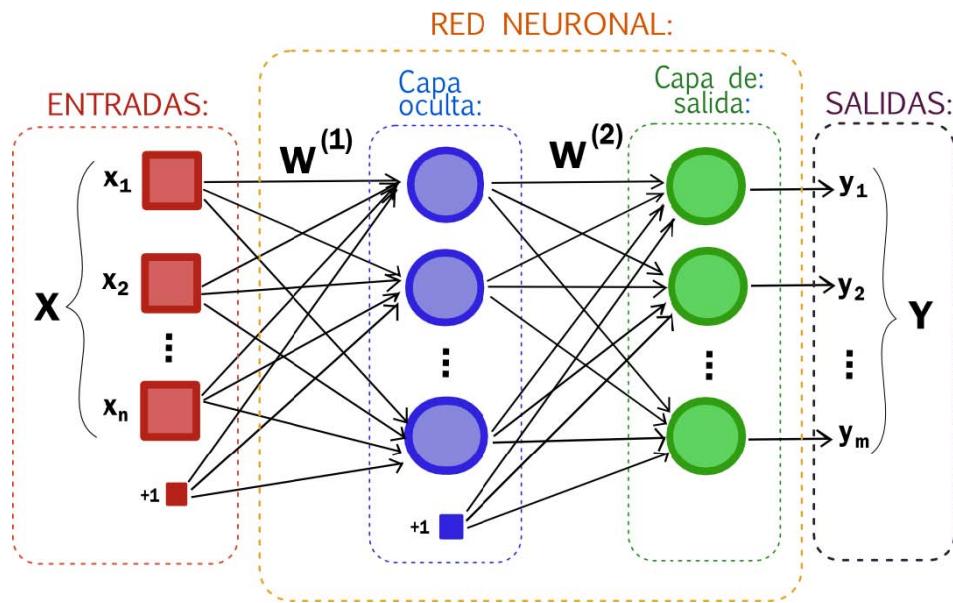
**2.2.1. redes neuronales** es una de las técnicas o algoritmos de machine learning que se pueden emplear en las tareas de predicción o detección. “Una red neuronal es un modelo simplificado que emula el modo en que el cerebro humano procesa la

---

<sup>10</sup>GONZÁLEZ Andrés. ¿Qué es Machine Learning? [en línea]. <<http://cleverdata.io/que-es-machine-learning-big-data/>>[citado en 10 Octubre 2017]

información: Funciona simultaneando un número elevado de unidades de procesamiento interconectadas que parecen versiones abstractas de neuronas”<sup>11</sup>

**Figura 1:** Diagrama de red neuronal



VÍLCHEZ GARCÍA, Víctor Gabriel. *Estimación y clasificación de daños en materiales utilizando modelos AR y redes neuronales para la evaluación no destructiva con ultrasonidos*. [en línea]. <<http://ceres.ugr.es/~alumnos/esclas/>>[citado en 24 Mayo 2018]

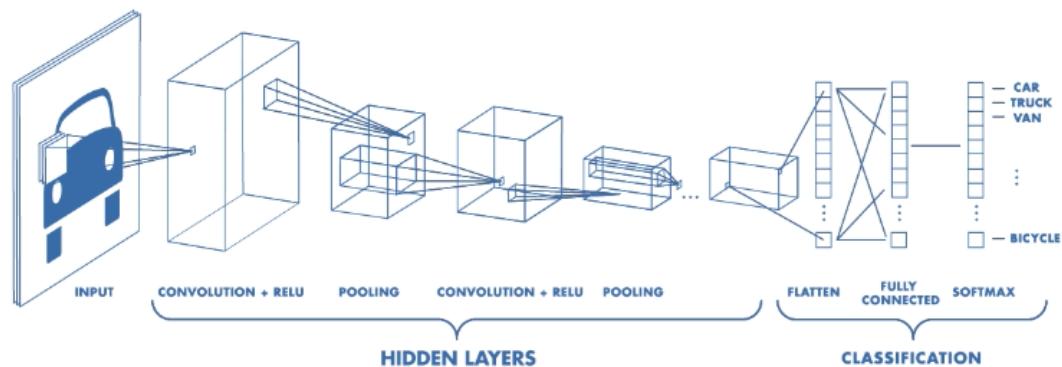
Una red neuronal está constituida por una serie de capas que se activan con determinadas entradas generando determinadas salidas que podrían ser tomadas o no por otras capas, dependiendo de la “profundidad” de la red. Utilizar redes neuronales consiste en hacer que la misma aprenda examinando las entradas, prediciendo las salidas y haciendo ajustes a los distintos parámetros de la misma, este proceso se repite muchas veces hasta que la red sea capaz de predecir o clasificar con un margen de error tolerable.

<sup>11</sup>IBM. El modelo de redes neuronales [en línea]. <[https://www.ibm.com/support/knowledgecenter/es/SS3RA7\\_18.0.0/modeler\\_mainhelp\\_client\\_ddita/components/neuralnet/neuralnet\\_model.html](https://www.ibm.com/support/knowledgecenter/es/SS3RA7_18.0.0/modeler_mainhelp_client_ddita/components/neuralnet/neuralnet_model.html)> [citado en 31 de Mayo de 2018]

La idea del proyecto es obtener la configuración de una red neuronal que sea capaz de predecir el nivel socioeconómico de una zona urbana, para esto se planea alimentar la red con imágenes satelitales e información extra de dichas zonas urbanas. Con esta información, ajustes en los parámetros y bastantes iteraciones la red neuronal aprenderá. Los cambios en los parámetros, profundidad y datos usados en la red neuronal son realizados para encontrar la configuración de red neuronal que mejor desempeño presente para la tarea propuesta. Dado que los datos a utilizar son imágenes, es recomendable utilizar “redes neuronales convolucionales” (CNN).

**2.2.2. redes neuronales convolucionales** son un tipo de red neuronal que se adapta mejor al uso de imágenes dado que “las CNN eliminan la necesidad de una extracción de características manual, por lo que no es necesario identificar las características utilizadas para clasificar las imágenes. La CNN funciona mediante la extracción de características directamente de las imágenes. Las características relevantes no se entrena previamente; se aprenden mientras la red se entrena con una colección de imágenes”<sup>12</sup>

**Figura 2:** Arquitectura típica de una CNN



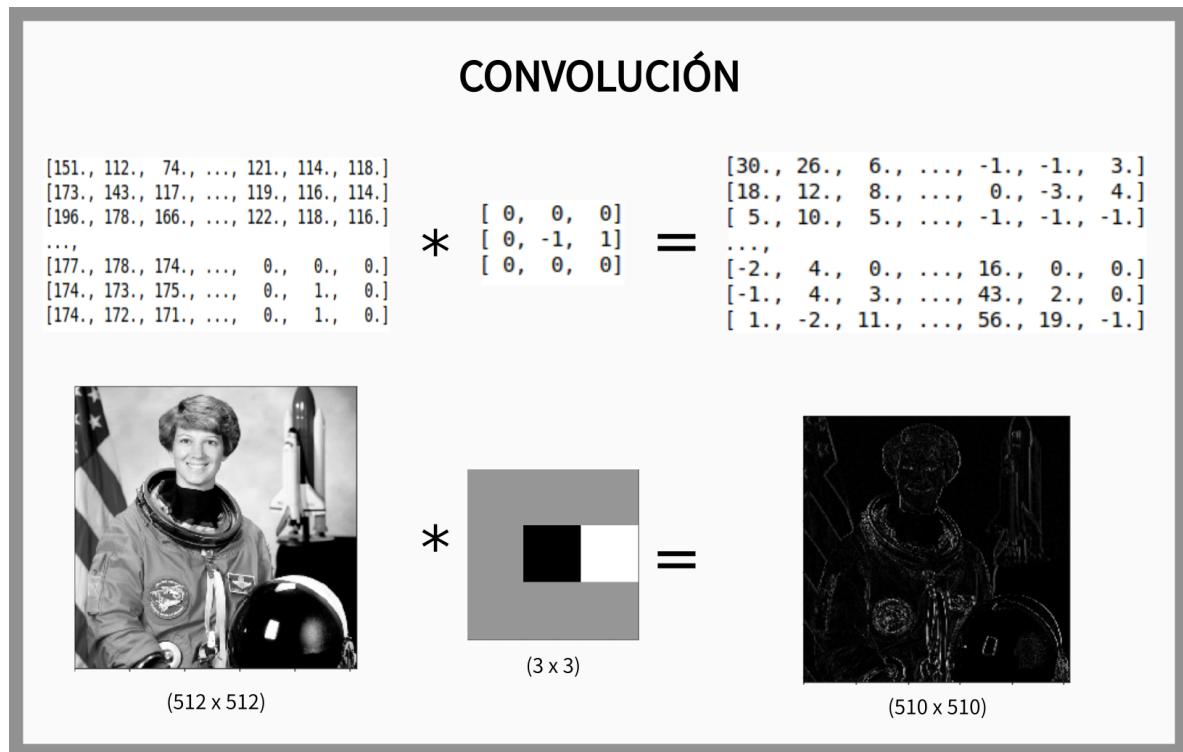
MATHWORKS. Architecture of a CNN. [en línea] <https://la.mathworks.com/solutions/deep-learning/convolutional-neural-network.html> [Citado en 30 de junio de 2018]

<sup>12</sup>MATHWORKS. Aprendizaje profundo [en línea]. <<https://es.mathworks.com/discovery/deep-learning.html>>[citado en 31 de Mayo de 2018]

Las CNN pueden contar con un sin número de capas (Neuronas), sin embargo, existen 3 tipos de neuronas fundamentales. Las **Neuronas convolucionales**, **Neuronas de reducción de muestreo** y **Neuronas de clasificación**.

Las **neuronas convolucionales** son funciones que se encargan de extraer características de cada imagen que pasa por la red, estas características se toman realizando operaciones de convolución, de ahí el nombre de la red, a cada imagen. Un ejemplo de convolución es la detección de bordes.

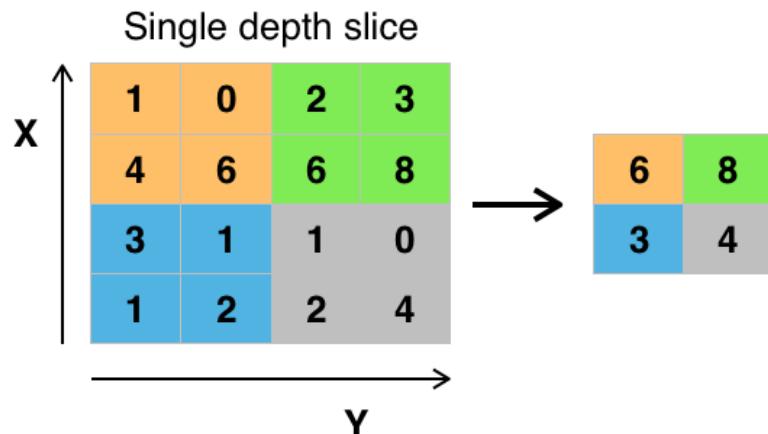
**Figura 3:** Detección de bordes mediante una convolución



La salida de cada neurona convolucional puede conectarse a otra neurona convolucional o a una neurona de reducción de muestreo o *Pooling layer*. Las **neuronas de reducción de muestreo** se encargan de reducir o 'resumir' las características obtenidas en las neuronas convolucionales optimizando así el aprendizaje y futura clasificación. Esta reducción de características es realizada con funciones como *max-pooling*, la cual

selecciona el valor máximo de un conjunto de datos.

**Figura 4:** Max pooling with a 2x2 filter and stride = 2



Aphex34. Max pooling with a 2x2 filter and stride = 2 [en línea] [https://upload.wikimedia.org/wikipedia/commons/e/e9/Max\\_pooling.png](https://upload.wikimedia.org/wikipedia/commons/e/e9/Max_pooling.png) [citado en 30 de julio de 2018]

la salida de una neurona de reducción de muestreo puede conectarse con una neurona convolucional, otra neurona de reducción de muestreo o con una neurona de clasificación. Las **neuronas de clasificación** se encargan de asignar una etiqueta o clase a cada imagen que entra en la red. Esta asignación se consigue gracias al aprendizaje de las características obtenidas en las capas anteriores durante el proceso de entrenamiento de la red. La salida de una neurona de clasificación puede ser un valor que representa una clase, un vector de valores que indican el porcentaje de pertenencia de cada clase o como es el caso de la segmentación semántica una matriz de valores.

**2.2.3. segmentacion semantic** es un problema de visión por computador el cual consiste en asignar a cada pixel de una imagen una etiqueta o clase, obteniendo una nueva imagen de 2 dimensiones, donde se puede observar la delimitación de cada clase en la imagen.

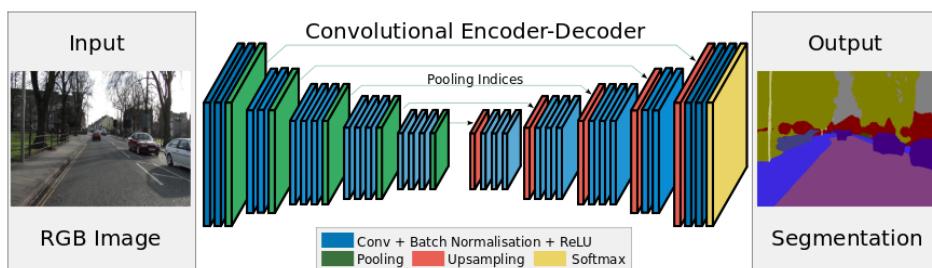
**Figura 5:** Problemas de visión por computador



VENTURA Royo Carles Ventura. La segmentación semántica y sus benchmarks [en línea] <http://informatica.blogs.uoc.edu/2016/05/26/la-segmentacion-semantica-y-sus-benchmarks/> [citado en 30 de julio de 2018]

Las tareas de segmentación semántica usando machine learning consisten en realizar la predicción de dicha imagen segmentada, como se mencionó anteriormente para esta tarea se usan redes neuronales convolucionales, sin embargo, estas redes convolucionales para segmentación semántica tienen una estructura algo diferente. Estas redes cuentan con una sección de 'convolución' (*Encode*) y 'deconvolución' (*Decode*), esto para obtener una imagen con las dimensiones de la imagen original y con el problemas de segmentación resuelto. Estas redes con sección de convolución y deconvolución toman el nombre de autoencoders.

**Figura 6:** Red Neuronal con arquitectura de convolución y deconvolución



KENDALL Alex, BADRINARAYANAN Vijay, CIPOLLA Roberto. SegNet [en línea] <http://mi.eng.cam.ac.uk/projects/segnet/> [citado en 30 de julio de 2018]

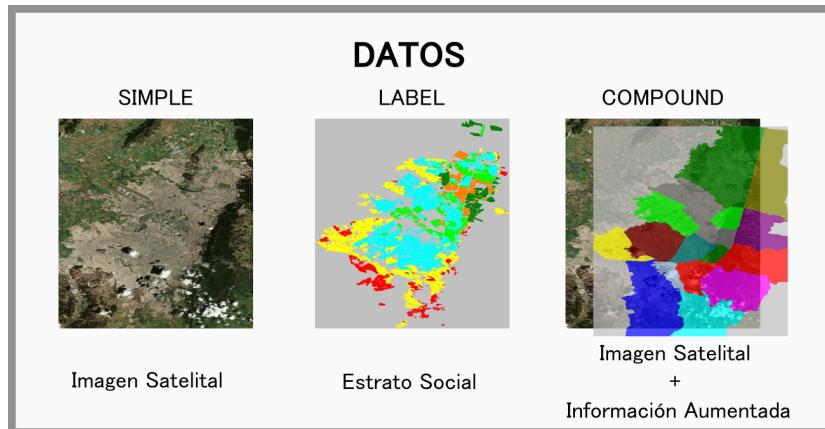
### 3. DESARROLLO DEL PROYECTO

Está claro que la recolección de datos mediante encuestas a vivienda es una metodología que conlleva bastante tiempo y que el cálculo del estrato debe tener en cuenta un sin número de variables. Como alternativa se propone usar imágenes satelitales dado que se puede obtener mucha información socioeconómica con el análisis de las mismas. Se pueden identificar patrones, detectar construcciones específicas, clasificar materiales de construcción en los tejados y más, todos estos análisis son variables usadas en el cálculo del estrato.

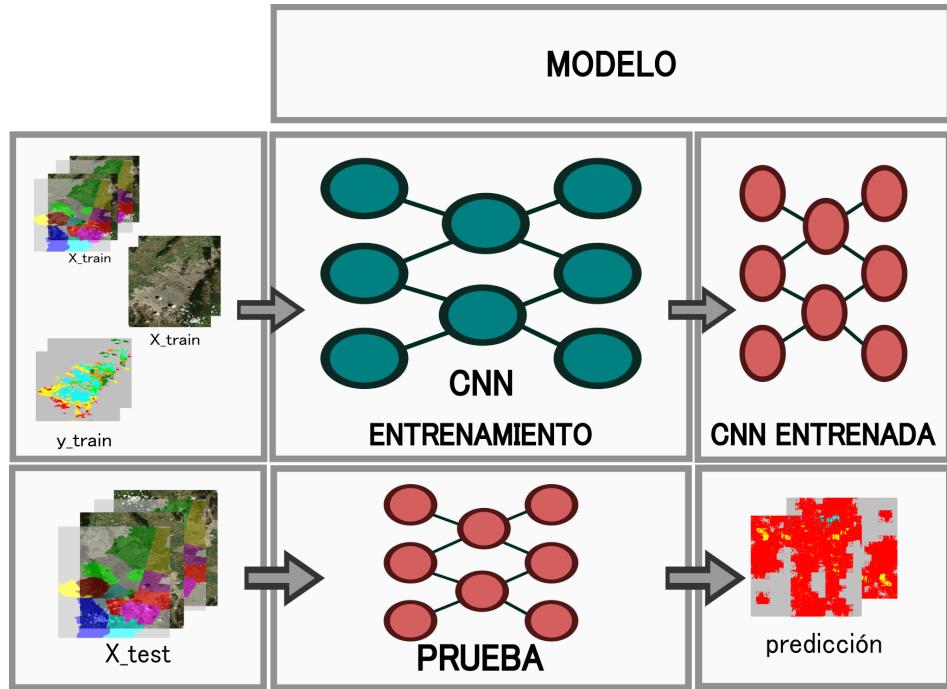
Como se mencionó existe un par de trabajos sobre el nivel socioeconómico usando técnicas de machine learning. El trabajo de Neal Jean usa redes neuronales convoluciones y Transfer Learning para lograr una predicción de la pobreza en 5 países de África. El trabajo de la EAFIT usa técnicas de machine learning para detectar características que puedan dar a conocer el nivel socioeconómico de una zona. Hay que tener en cuenta que en ambos trabajos la principal fuente de datos son las imágenes satelitales.

A grandes rasgos lo que se planteó realizar con las imágenes satelitales, la información aumentada, la información de los estratos y las redes neuronales, se muestra en las siguientes figuras.

**Figura 7:** Datos a usar en el proyecto



**Figura 8:** Modelo a usar en el proyecto



El proyecto se desarrolló encontrando fuentes de datos, generando dataset con imágenes satelitales, información aumentada y distintos niveles de zoom, seleccionando redes neuronales y realizando pruebas para determinar la mejor manera de predecir el estrato social. Hay que mencionar que el desarrollo del proyecto está documentado en el siguiente repositorio de GitHub [https://github.com/DaielChom/proyecto\\_uis](https://github.com/DaielChom/proyecto_uis). El desarrollo se llevó de la siguiente manera.

### 3.1. FUENTES DE DATOS

Era necesario tener imágenes satelitales, imágenes con los estratos sociales e imágenes con información extra que correspondan geográficamente y que tengan distintos niveles de zoom, como se muestran en la Figura 7. Se usan distintos niveles de zoom para determinar si esto influye en la predicción del estrato social. Existen muchas plataformas para obtener imágenes satelitales tanto de alta como baja resolución, varias de estas

plataformas tienen la opción de insertar información en dichos mapas en forma de marcador, polígono o línea. Esta información por lo general son archivos en formato kml o shp. Gracias a la política de datos abiertos<sup>13</sup> es posible encontrar archivos kml o shp en distintas páginas web del gobierno. sin embargo no todos los departamentos o ciudades cuentan con la misma cantidad de datos disponibles y menos con datos del estrato social. El único portal web donde se encontró una buena cantidad de información, incluyendo la del estrato social fue en el portal de mapas de Bogotá, disponible en <http://mapas.bogota.gov.co>.

No fue posible descargar los archivos kml mostrados en la plataforma, por ende, y aprovechando la licencia Creative Commons que presentan los datos, se crearon nuevos archivos kml usando los mapas de la plataforma como guía. Los archivos kml realizados están disponibles en <https://drive.google.com/open?id=15VnvN6ZRTbsqqd9k13ukNWbj3oqADsy0>. se crearon dos kml, uno con información de los estratos sociales y otro con los índices de condiciones de seguridad nocturna por localidad que tienen las mujeres, mas específicamente los datos de la categoría *riesgo alto* de dicho mapa.

### 3.2. DATASET

Como en la mayoría de trabajos de machine learning se debe tener un dataset, un conjunto de datos organizados y con una estructura dividida en train (datos para entrenamiento) y test (datos para pruebas) con los cuales se entrena o prueba un algoritmo de aprendizaje. El dataset que se construyó lleva por nombre OVERLAECOBO y cuenta con 3 tipos imágenes (Figura 7) geográficamente correspondidas y de distintos niveles de zoom o acercamiento satelital. Las imágenes del dataset fueron obtenidas usando el software Ruso SASplanet<sup>14</sup>, los archivos kml diseñados y *Bing maps* como fuente de imágenes satelitales, fuente ofrecida en SASplanet. Se tomo el mapa de Bogotá y se trazó una línea que dividiera la ciudad en dos partes, de tal manera que en ambas hubiera información de todos los estratos, una parte se escogió para train y otra la para

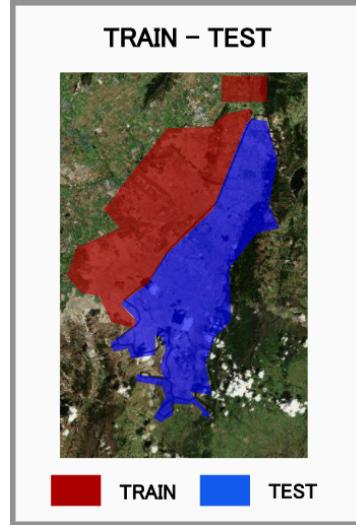
---

<sup>13</sup>GOBIERNO DIGITAL. Datos Abiertos [en linea] <<http://estrategia.gobiernoenlinea.gov.co/623/w3-article-9407.html>>[citado en 30 de Mayo de 2018]

<sup>14</sup>Disponible en <http://www.sasgis.org/>

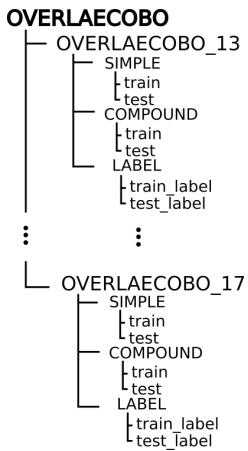
test, como se muestra en la siguiente imagen.

**Figura 9:** División de Train y Test



Se extrajeron imágenes de 7 Zooms diferentes, del 13 al 20, cada nivel de zoom cuenta con imágenes satelitales (*SIMPLE*), satelitales + información extra (Riesgo nocturno para las mujeres) (*COMPOUND*) y de estrato social, las cuales son usadas como *Ground Truth* (*LABEL*). Cada nivel está dividido en train y test. OVERLAECOBO cuenta con la siguiente estructura.

**Figura 10:** Estructura OVERLAECOBO



**Tabla 1:** Cantidad de Imágenes en OVERLAECOBO

Cantidad de Imágenes en OVERLAECOBO			
ZOOM	Configuración	Train	Test
13	SIMPLE	7	2
13	COMOUND	7	2
13	LABEL	7	2
14	SIMPLE	24	24
14	COMOUND	24	24
14	LABEL	24	24
15	SIMPLE	98	90
15	COMPOUND	98	90
15	LABEL	98	90
16	SIMPLE	347	321
16	COMPOUND	347	321
16	LABEL	347	321
17	SIMPLE	1338	1203
17	COMPOUND	1338	1203
17	LABEL	1338	1203

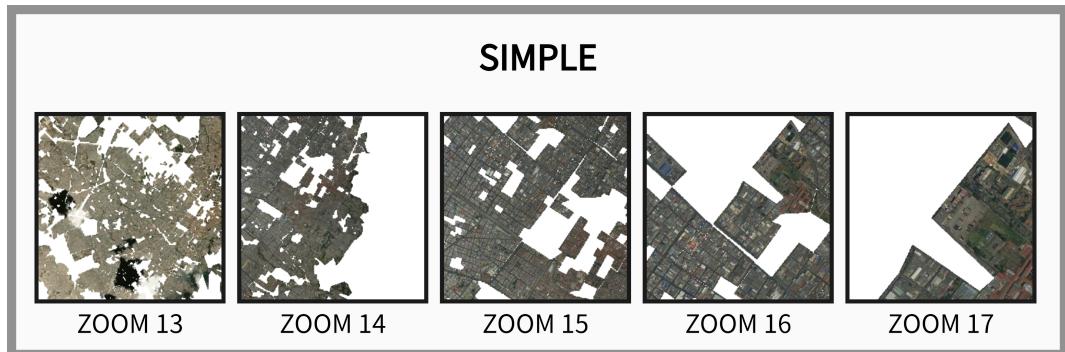
**Tabla 2:** Proporción de cada clase en Train - OVERLAECOBO

Proporción de cada clase en Train - OVERLAECOBO							
Zoom	No Definido	Estrato 1	Estrato 2	Estrato 3	Estrato 4	Estrato 5	Estrato 6
13	41.495	7.423	21.502	25.908	2.605	1.067	0.001
14	54.759	5.254	10.547	15.508	6.889	3.953	3.091
15	52.886	5.690	11.388	16.564	6.038	3.621	3.815
16	50.778	6.418	11.745	17.078	6.330	3.311	4.340
17	47.853	6.898	12.553	18.316	6.388	3.322	4.669

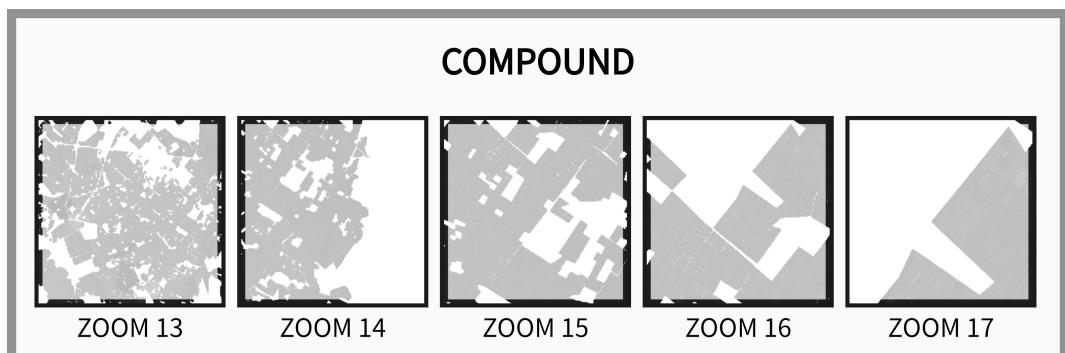
**Tabla 3:** Proporción de cada clase en Test - OVERLAECOBO

Proporción de cada clase en Test - OVERLAECOBO							
Zoom	No Definido	Estrato 1	Estrato 2	Estrato 3	Estrato 4	Estrato 5	Estrato 6
13	66.858	2.897	12.080	13.635	3.168	0.955	0.407
14	58.255	3.074	14.557	16.085	3.692	2.585	1.752
15	56.453	2.623	14.526	16.844	4.660	2.877	2.018
16	52.951	2.924	15.380	18.017	5.250	3.092	2.386
17	51.200	2.928	15.571	19.340	5.689	2.851	2.421

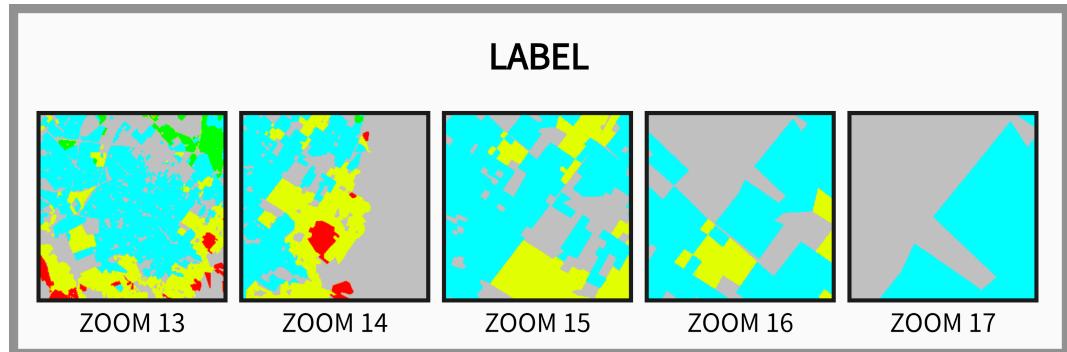
**Figura 11:** Imágenes SIMPLE de OVERLAECOBO



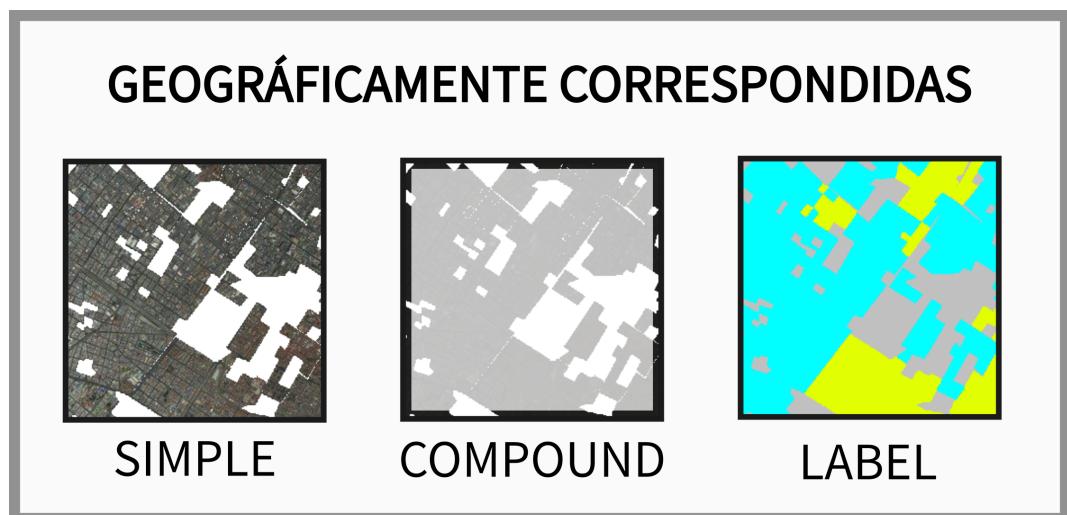
**Figura 12:** Imágenes COMPOUND de OVERLAECOBO



**Figura 13:** Imágenes LABEL de OVERLAECOBO



**Figura 14:** Correspondencia geográfica de OVERLAECOBO

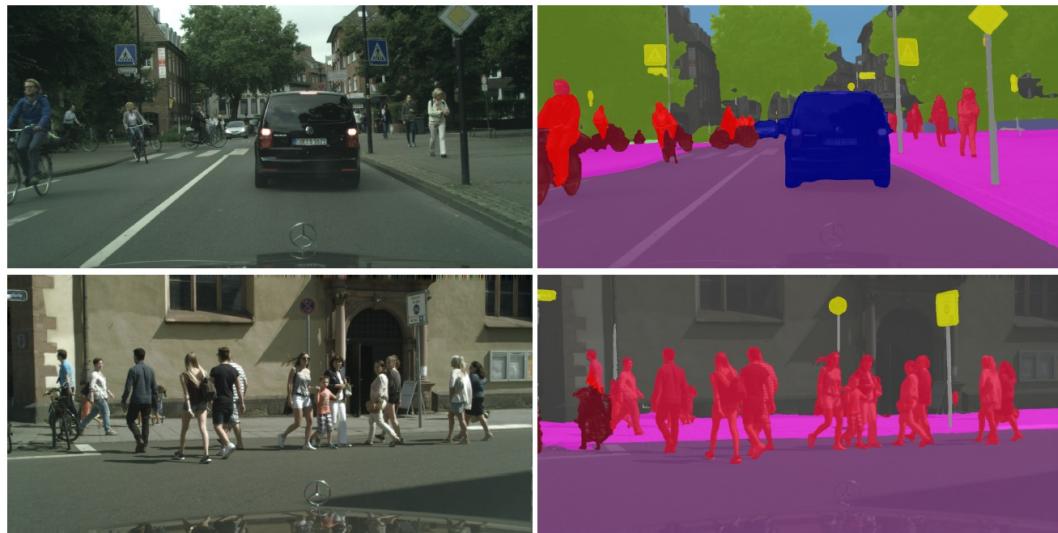


Una vez el dataset estuvo elaborado se prosiguió con la búsqueda de redes neuronales y posteriormente pruebas para determinar la mejor configuración para la detección del estrato social.

### 3.3. REDES NEURONALES

Hay que mencionar que la metodología que se usó para la detección del estrato social se conoce en machine learning como Segmentación Semántica (Figura 15), la cual consiste en segmentar una imagen con el fin de delimitar un objeto o diferenciar entre elementos. En Segmentación semántica, en el proceso de entrenamiento se tienen dos imágenes, la imagen que se quiere segmentar ( $X_{train}$ ) y la imagen segmentada ( $y_{train}$ ) y con estas se realiza el aprendizaje. En el proceso de pruebas con imágenes que se requieren segmentar ( $X_{test}$ ) se crea una nueva imagen segmentada, luego en el proceso de medición del error con las imágenes segmentadas creadas se comparan con las imágenes segmentadas de dataset ( $y_{test}$ ). En otras palabras, cada pixel RGB de una imagen tiene su propia etiqueta. La segmentación semántica es una técnica muy utilizada en el campo de la conducción autónoma, ya que representa una buena metodología para descubrir los distintos objetos que percibe la cámara del vehículo y así tomar decisiones.

**Figura 15:** Segmentación Semantica



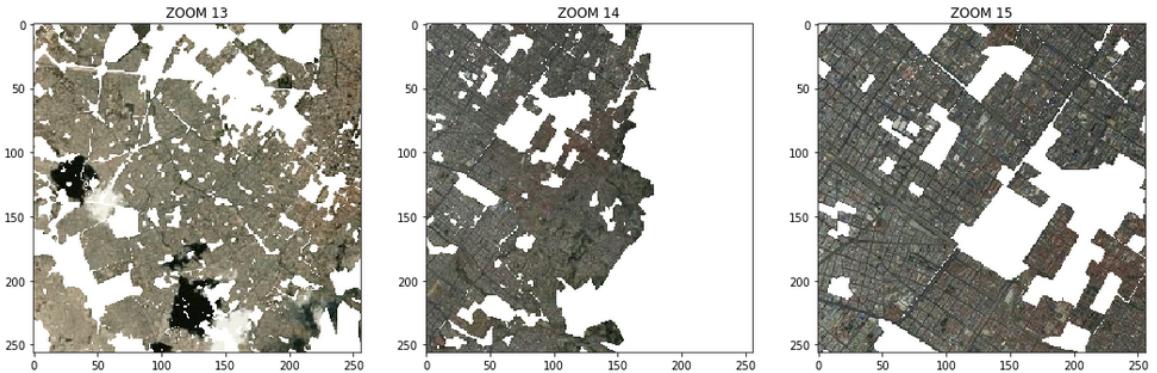
ABHIJIT Kundu, VIBHAV Vineet, and VLANDLEN Koltun Feature Space Optimization for Semantic Video Segmentation [en linea] <<http://vladlen.info/publications/feature-space-optimization-for-semantic-video-segmentation/>> [citado en 01 de junio de 2018]

Existe muchas redes neuronales convolucionales con las que se puede aplicar Segmentación semántica, entre ellas SegNet, uNet, DenseNet, AlexNet, cada una de ellas con diferentes arquitecturas que les dan ventajas o desventajas a la hora de solucionar un problema de segmentación semántica. En GitHub hay varios repositorios con implementaciones de estas redes; como es de esperarse estos repositorios, aunque implementen la misma red neuronal, tienen varias diferencias; implementados en distintos lenguajes, usan diferentes librerías e incluso entranan con diferentes dataset, sin embargo, cumplen la tarea de Segmentación semántica que proponen. Tras una búsqueda y selección entre varios repositorios para seleccionar la mejor red neuronal o la mejor implementación se encontró un Framework que permite usar distintas redes neuronales para la misma tarea. Dicho framework está disponible en el siguiente enlace <https://github.com/GeorgeSeif/Semantic-Segmentation-Suite> y es el que se usó para las pruebas.

### 3.4. DETECCION DEL ESTRATO SOCIAL

**3.4.1. datos.** Para los entrenamientos de las distintas redes usadas Se utilizaron 3 niveles de ZOOM para las pruebas y cada nivel se entrenó con imágenes *SIMPLE* y *COMPOUND* (satelital, satelital + información aumentada).

**Figura 16:** Niveles de ZOOM usados



**3.4.2. redes.** Para cada configuración del dataset se realizaron 11 pruebas variando la red neuronal y los parámetros *epoch* y *batch size*. En total se realizaron 66 pruebas.

**Tabla 4:** Pruebas realizadas por cada configuración de dataset

Pruebas			
Nº	RED 1	EPOCH	BATCH SIZE
1	FC-DenseNet56	50	1
2	FC-DenseNet56	70	1
3	Encoder-Decoder	70	1
4	Encoder-Decoder	70	2
5	Encoder-Decoder-Skip	70	1
6	Encoder-Decoder-Skip	70	2
7	FRRN-A	70	1
8	MobileUNet	70	1
9	MobileUNet	70	2
10	MobileUNet-Skip	70	1
11	MobileUNet-Skip	70	2

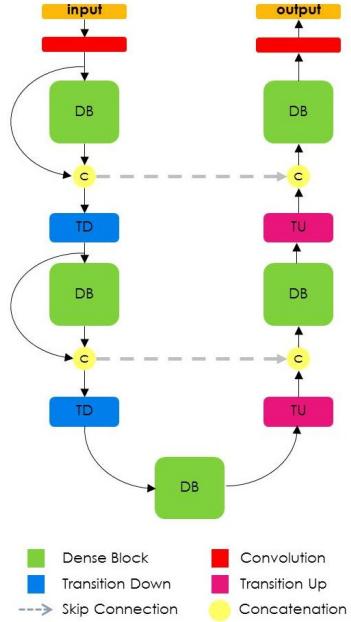
**3.4.2.1. fc-densenet56** <sup>15</sup> Es una red neuronal que aprovecha las ventajas de la arquitectura DenseNet<sup>16</sup> y las combina con las propiedades de las redes Fully Convolutional<sup>17</sup>, como se puede ver en la imagen utiliza bloques densos y saltos de encoder a decoder, elementos clásicos de la arquitectura DenseNet.

<sup>15</sup>JEGOU Simon, DROZDZAL Michal, VAZQUEZ David, ROMERO Adriana, BENGIO Yoshua. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation [en línea] <<https://arxiv.org/abs/1611.09326>>[citado en 31 de julio de 2018]

<sup>16</sup>HUANG Gao, LUI Zhuang, VAN DER Maaten Laurens, Q. Weinberger Kilian. Densely Connected Convolutional Networks [en línea] <<https://arxiv.org/abs/1608.06993>>[citado en 31 de julio de 2018]

<sup>17</sup>SHELHAMER Evan, LONG Jonathan, DARRELL Trevor. Fully Convolutional Networks for Semantic Segmentation [en línea] <<https://arxiv.org/abs/1605.06211>>[citado en 31 de julio de 2018]

**Figura 17:** Arquitectura FC-DenseNet



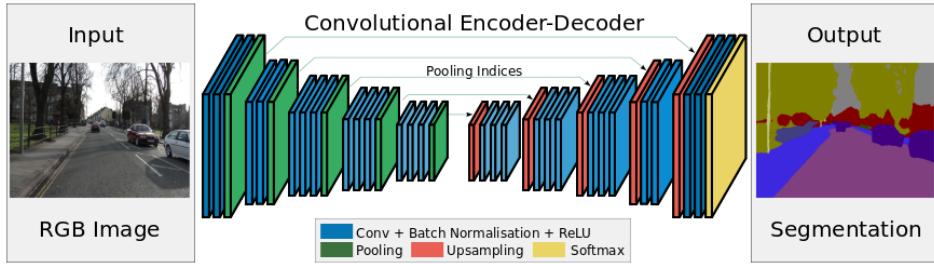
JEGOU Simon, DROZDZAL Michal, VAZQUEZ David, ROMERO Adriana, BENGIO Yoshua. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation [en línea] <<https://arxiv.org/abs/1611.09326>>[citado en 31 de julio de 2018]

**3.4.2.2. encoder-decoder** Red neuronal diseñada en base a la arquitectura SegNet<sup>18</sup> la cual cuenta con una sección de convolución seguida por una de deconvolución, como se muestra en la siguiente figura

---

<sup>18</sup>KENDALL Alex, BADRINARAYANAN Vijay, CIPOLLA Roberto. SegNet [en línea] <http://mi.eng.cam.ac.uk/projects/segnets/> [citado en 30 de julio de 2018].

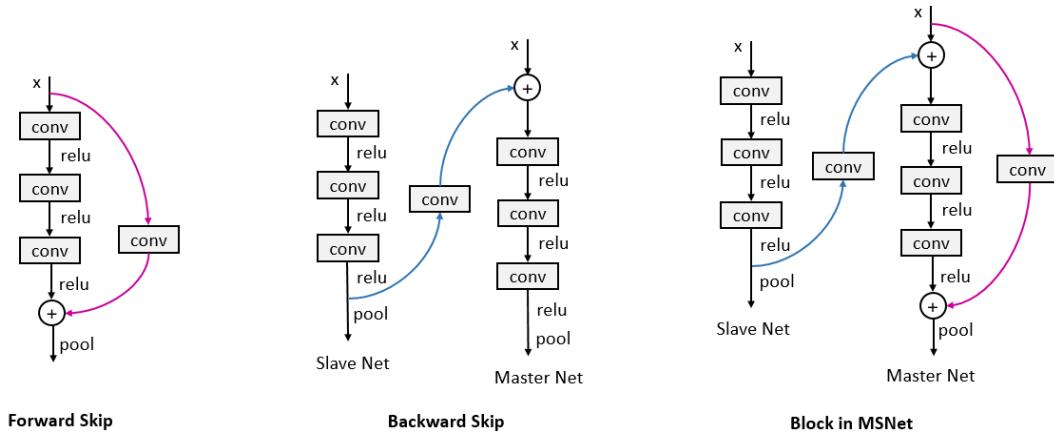
**Figura 18:** Arquitectura SegNet



KENDALL Alex, BADRINARAYANAN Vijay, CIPOLLA Roberto. SegNet [en línea] <http://mi.eng.cam.ac.uk/projects/segnets/> [citado en 30 de julio de 2018]

**3.4.2.3. encoder-decoder-skip** De igual manera está basada en SegNet con una sección de convolución seguida por una de deconvolución, la diferencia es que esta cuenta con la capacidad de 'saltarse' alguna capa entre la convolución y la deconvolución.

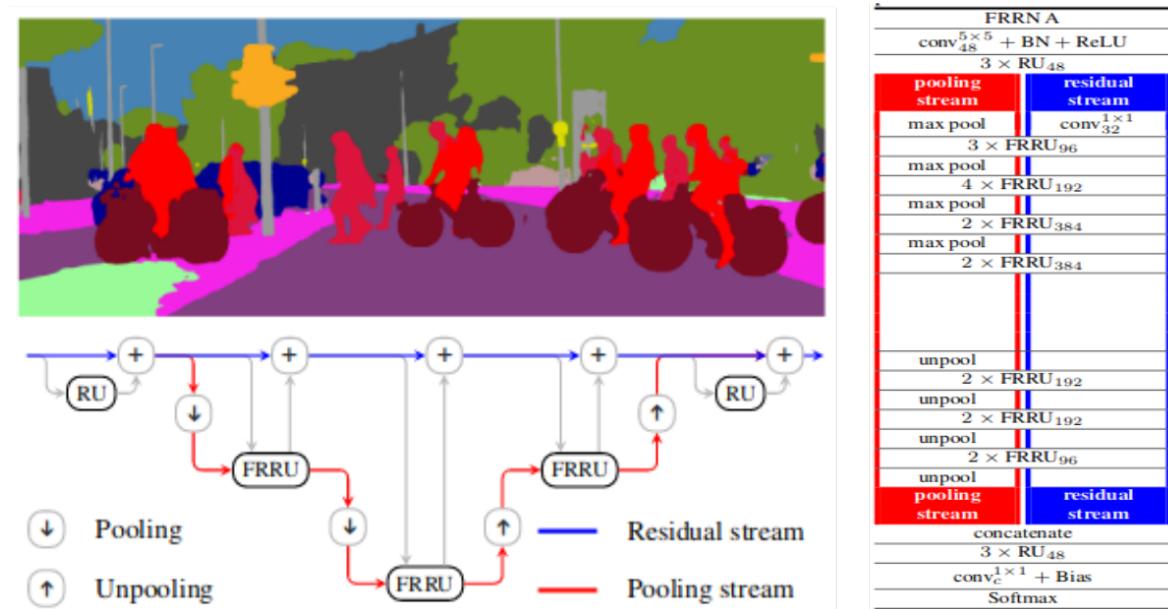
**Figura 19:** Ejemplo de 'saltos' en una red neuronal



H.Abdulnabi Abrar, WINKLER Stefan, WANG Gang. Beyond Forward Shortcuts: Fully Convolutional Master-Slave Networks (MSNets) with Backward Skip Connections for Semantic Segmentation [en línea] <<https://arxiv.org/abs/1707.05537>>[citado en 31 de julio de 2018]

**3.4.2.4. frrn-a**<sup>19</sup> Es una red neuronal con una arquitectura de doble flujo que utiliza las propiedades de ResNet, los cuales son bloques de capas diseñados para ayudar en el aprendizaje y diseño de las redes. En la siguiente imagen se puede observar la arquitectura de FRRN.

**Figura 20:** Arquitectura FRRN

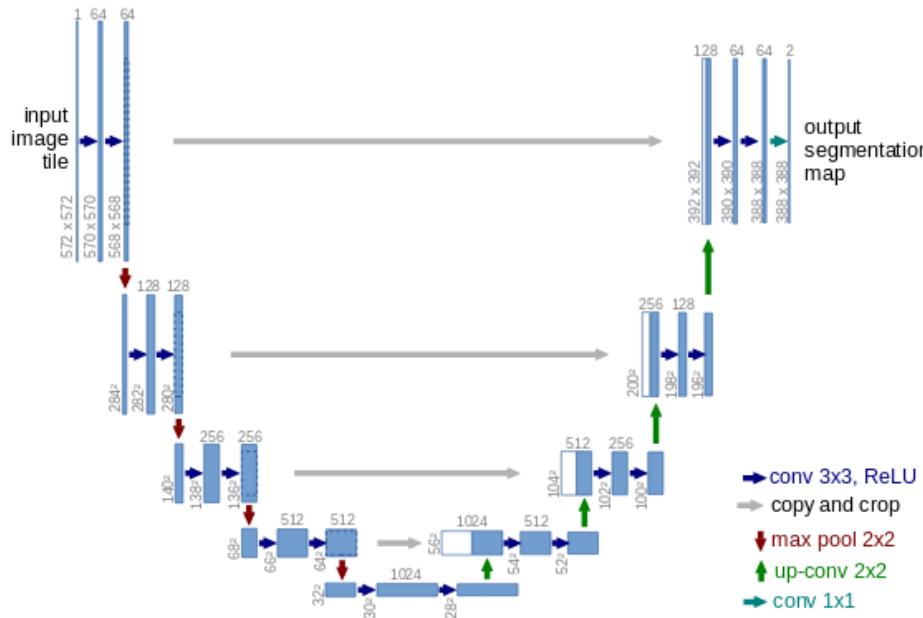


POHLEN Tobias, HERMANS Alexander, MATHIAS Markus, LEIBE Bastian. Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes [en línea] <<https://arxiv.org/abs/1611.08323>>[citado en 31 de julio de 2018]

<sup>19</sup>POHLEN Tobias, HERMANS Alexander, MATHIAS Markus, LEIBE Bastian. Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes [en línea] <<https://arxiv.org/abs/1611.08323>>[citado en 31 de julio de 2018]

**3.4.2.5. mobileunet** Red diseñada en base a las redes Unet<sup>20</sup> y MobileNet<sup>21</sup>, esta combinación de arquitecturas logra una alta velocidad y buenos resultados con valores de parámetros bajos. A continuación, se puede observar la arquitectura de UNet.

**Figura 21:** Arquitectura UNet



RONNEBERGER Olaf, FISCHER Philipp, BROX Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation [en línea] <<https://arxiv.org/abs/1505.04597>>[citado en 31 de julio de 2018]

**3.4.2.6. mobileunet-skip** Usa la misma arquitectura de MobileUnet pero con la posibilidad de 'saltos', tal cual se describió anteriormente.

<sup>20</sup> RONNEBERGER Olaf, FISCHER Philipp, BROX Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation [en línea] <<https://arxiv.org/abs/1505.04597>>[citado en 31 de julio de 2018]

<sup>21</sup> G. Howard Andrew, ZHU Menglong, CHEN Bo, KALENICHENKO Dmitry, WANG Weijun, WE-YAND Tobias, ANDREETTO Marco, ADAM Hartwig. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [en línea] <<https://arxiv.org/abs/1704.04861>>[citado en 31 de julio de 2018]

**3.4.3. metrica** Todas las pruebas fueron realizadas en la misma máquina, la cual cuenta con un procesador Intel Core i7, una Tarjeta gráfica NVIDIA GEFORCE 940MX y 8GBs de RAM, las pruebas fueron ejecutadas sobre el Sistema Operativo Xubuntu 18-04. Una vez realizado el entrenamiento de cada red neuronal se midió la capacidad de cada una para predecir el estrato social (generar aproximaciones a las imágenes *LABEL*) utilizando el índice de jaccard, más conocido como *IoU - Intersection over Union*, una métrica<sup>22</sup> muy utilizada en la detección de objetos y la segmentación. El índice de jaccard consiste en encontrar el grado de similitud entre dos conjuntos, en este caso entre dos imágenes, calculando el cociente entre la intersección y unión de los dos conjuntos. En el caso de segmentación semántica se usan las áreas cubierta por cada clase en las imágenes predichas y *test* para calcular dicho cociente.

**Figura 22:** Ecuación IoU

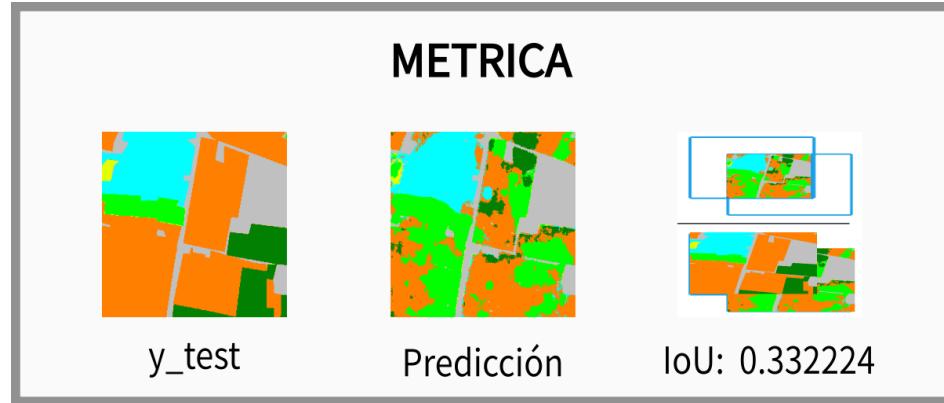
$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

ROSEBROCK Adrian. Intersection over Union (IoU) for object detection [en linea] <<https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>>[citado en 17 de julio de 2018]

---

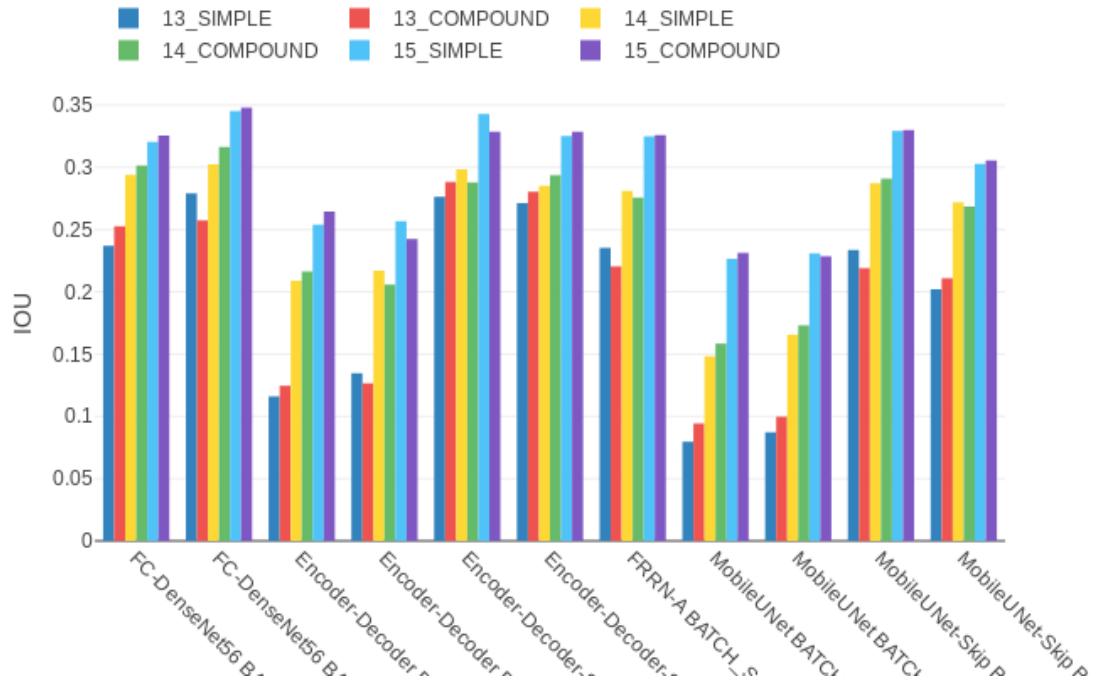
<sup>22</sup>ROSEBROCK Adrian. Intersection over Union (IoU) for object detection [en linea] <<https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>>[citado en 17 de Julio de 2018]

**Figura 23:** Ejemplo de predicción



Cada red entrenada generó una cantidad de imágenes equivalente a las de *test* del dataset con el que se entrenó. Con estas imágenes generadas y las imágenes del dataset se calculó el índice de jaccard para cada una de las 66 pruebas. Como se mencionó dichos procesos se pueden observar en el repositorio del proyecto. [https://github.com/DaielChom/proyecto\\_uis/](https://github.com/DaielChom/proyecto_uis/). Las pruebas entregaron los siguientes resultados.

**Figura 24:** Resultados - IoU de cada nivel de ZOOM agrupados por modelo, a la izquierda 'Prueba 1', a la derecha 'Prueba 11'

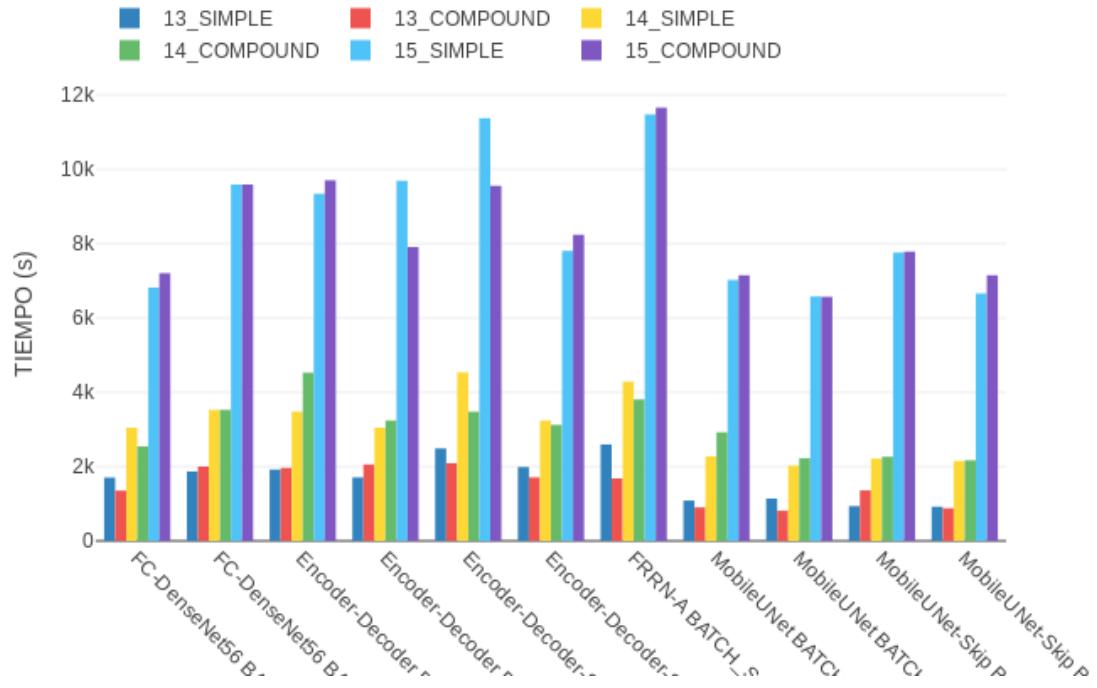


Analizando la Figura 24, se observa como el aumento del nivel de ZOOM provoca un incremento en la variable IoU, lo que significa un aumento en la precisión de cada predicción. Aunque los valores de IoU son bajos (puesto que se esperan valores cercanos a 1) este resultado no es extraño dado a la cantidad de imágenes de cada configuración de dataset y a los valores bajos de *Epoch* y *Batch Size*, los cuales son debido a los recursos computacionales con los que se cuenta. A excepción de una prueba, se puede observar el aumento de IoU en el uso de capas de información extra, como es el caso de la configuración *COMPOUND*. También se puede decir que el modelo que mejor se acopla a la tarea de predicción del nivel socio económico de una zona urbana es FC-DensetNet56<sup>23</sup>, mientras que MobileUNet<sup>24</sup> no es recomendable.

<sup>23</sup>JEGOU Simon, DROZDZAL Michal, VAZQUEZ David, ROMERO Adriana, BENGIO Yoshua. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation [en Línea]<<https://arxiv.org/abs/1611.09326>>[citado en 23 de Julio de 2018]

<sup>24</sup>HOWARD Andrew G., ZHU Menglong, CHEN Bo, KALENICHENKO Dmitry, WANG Weijun,

**Figura 25:** Resultados - Duración de Entrenamiento y Testeo para cada prueba, a la izquierda 'Prueba 1', a la derecha 'Prueba 11'

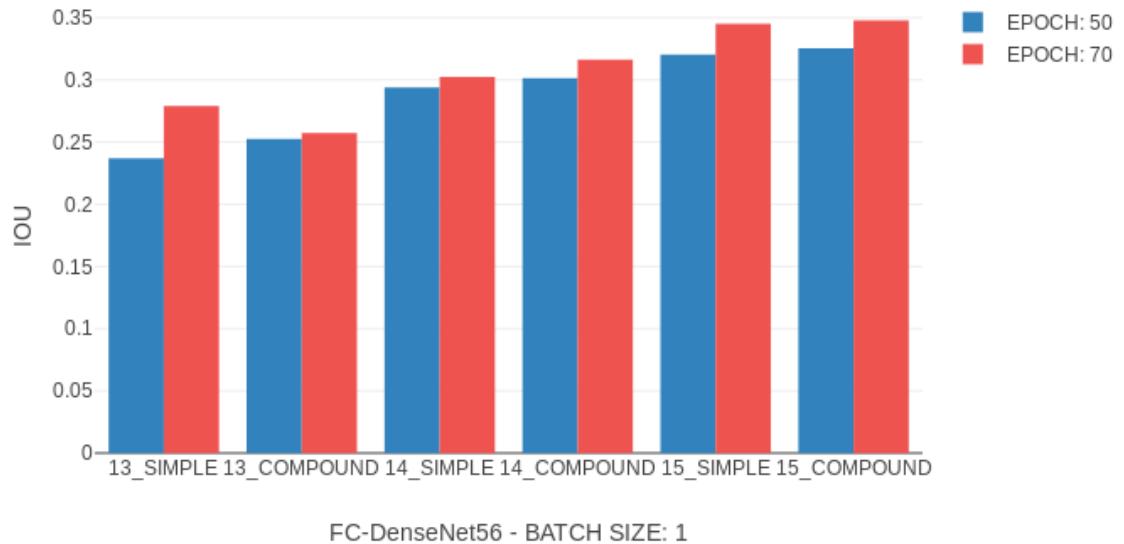


Como era de esperarse y como lo demuestra la gráfica la cantidad de imágenes usadas en el entrenamiento afecta la duración del proceso de entrenamiento y testeo, siendo las configuraciones con más imágenes las que más tiempo gastan. Hay que aclarar que la duración de cada prueba se ve afectada por el poder computacional con el que se cuenta.

---

WEYAND Tobias, ANDRETTTO Marco, ADAM Hartwig. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [en línea]<<https://arxiv.org/abs/1704.04861>>[citado en 23 de Julio de 2018]

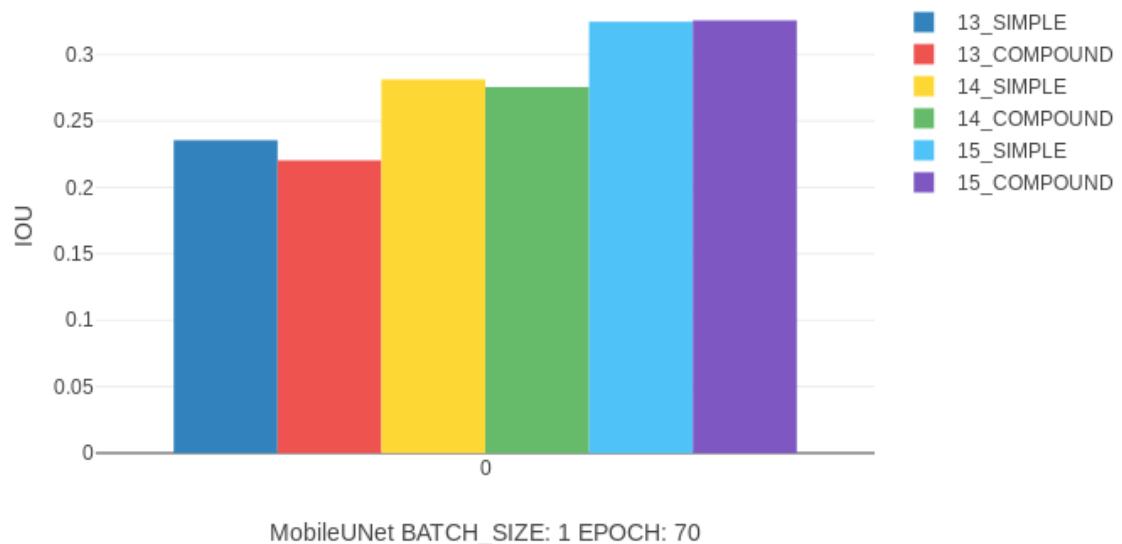
**Figura 26:** Resultados - IoU para diferentes valores de 'Epoch'



Aunque se hicieron pocas pruebas con distintos valores de *Epoch*, se puede observar una mejora en el IoU con el incremento de dicha variable, resultado que era de esperarse dado que se recomiendan<sup>25</sup> valores altos para este parámetro. Un número mayor de 70 podría generar mejores resultados, aunque también podría sobreentrenar el modelo.

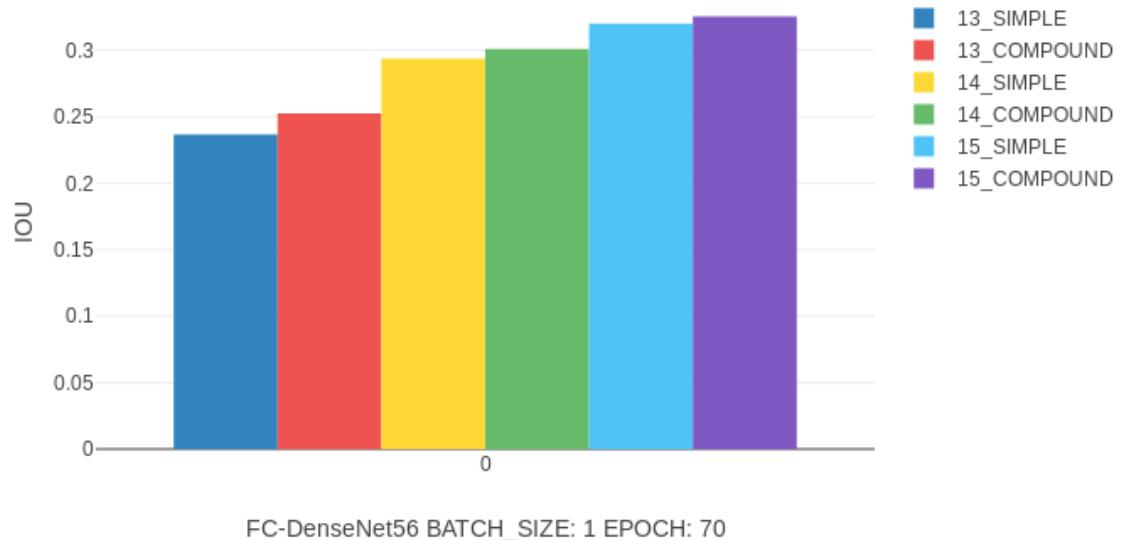
<sup>25</sup>SHARMA Sagar. Epoch vs Batch Size vs Iterations [en linea]<<https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9>>[citado en 23 Julio de 2018]

**Figura 27:** Resultados - IoU de cada nivel de ZOOM para la prueba 'Prueba 8'



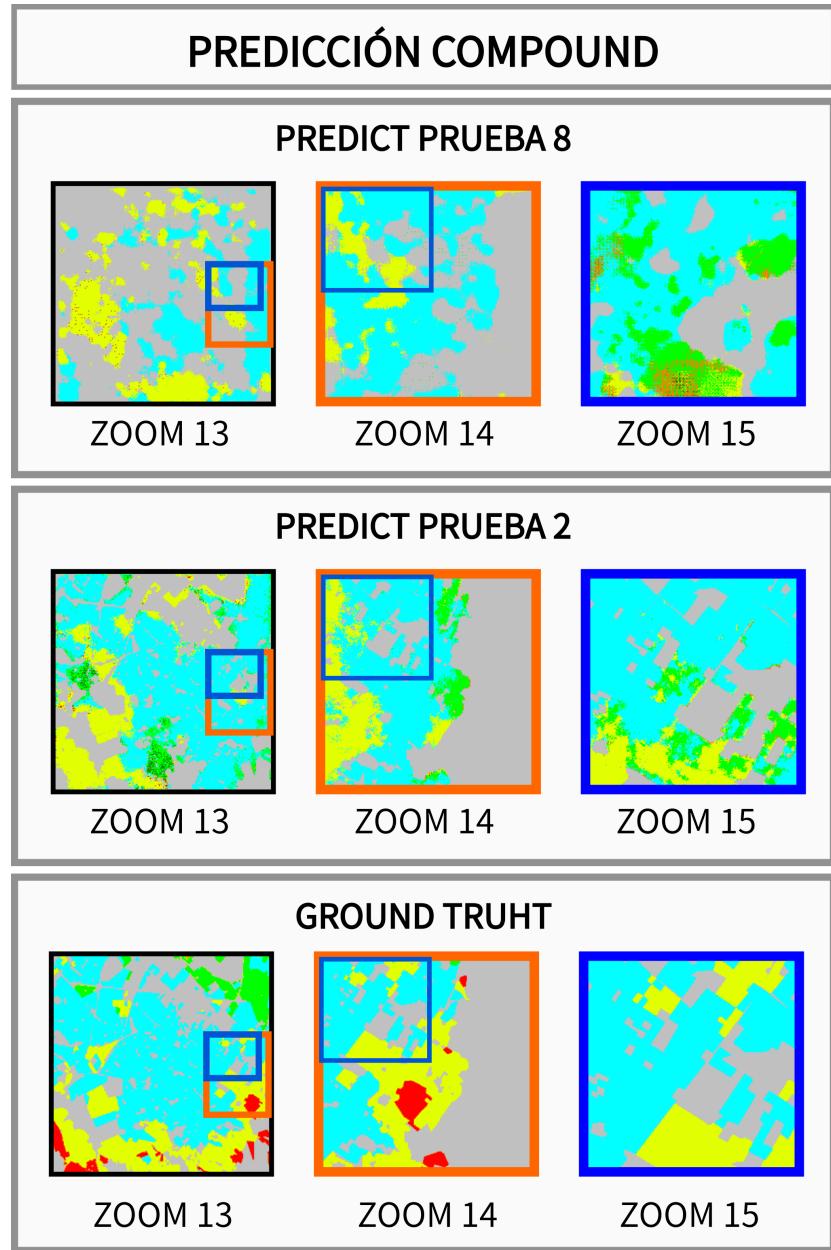
Como se mencionó anteriormente la peor configuración para la tarea propuesta corresponde al modelo MobileUNet, el cual registra los valores más bajos de IoU para las distintas configuraciones de dataset, como se puede observar en la Figura 24. Sin embargo, se puede notar el incremento de IoU con el uso de configuraciones con mayor cantidad de imágenes.

**Figura 28:** Resultados - IoU de cada nivel de ZOOM 'Prueba 2'



Contrario a MobileUnet, FC-DenseNet56 presenta los mejores valores de IoU en las distintas configuraciones. FC-DenseNet56 fue la mejor red para la tarea propuesta. De igual manera se puede observar el incremento de la métrica IoU con el aumento de imágenes usadas en entrenamiento y testeо. De la Figura 26 se conoce la mejoría de IoU con el incremento del parámetro *Epoch* usando esta misma red neuronal.

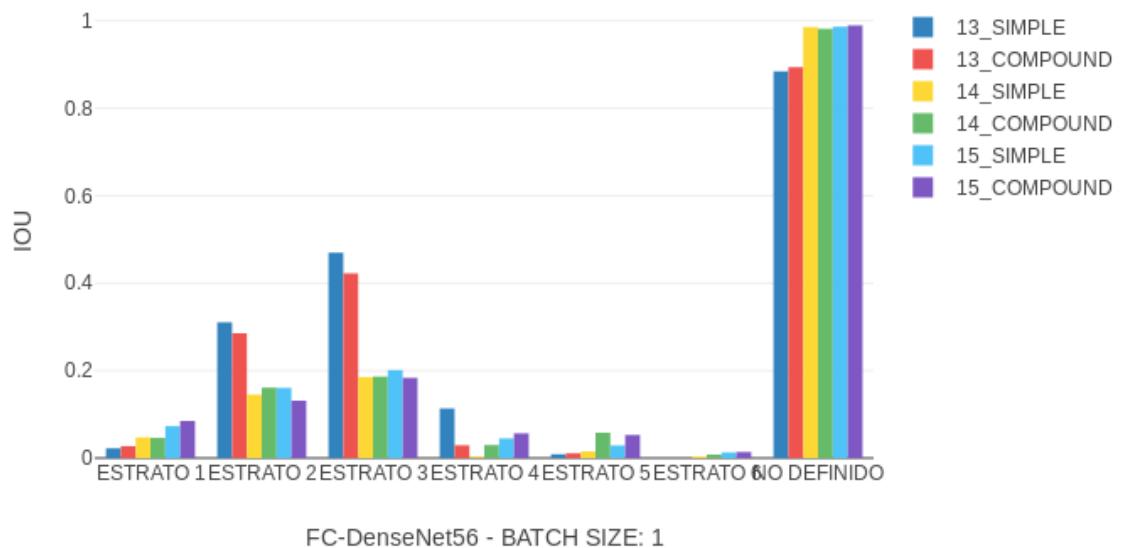
**Figura 29:** Predicción para una misma zona urbana con diferente nivel de acercamiento satelital para las pruebas 2 y 8



La Figura 29 se puede observar la predicción de FC-DenseNet (Prueba 2) y de MobileUNet (Prueba 8) mostrando sus diferencias a la hora de predecir el estrato social. En las predicciones de la prueba 8 se observa una mala predicción de la clase *NO DEFIT*.

*NIDO* a pesar de ser esta clase la más común en el dataset y la que en las imágenes de entrenamiento es simplemente el color blanco. Las predicciones de la prueba 2 tienen mejor resultado, logra identificar muy bien la clase *NO DEFINIDO*, sin embargo tiene problemas en la diferenciación de los demás estratos. Hay que mencionar que tanto para la prueba 2 y la prueba 8 la calidad de la predicción mejora cuando se trata de niveles de zoom mayores.

**Figura 30:** Resultados - IoU por clase de la mejor configuración



Con la Figura 30 se puede realizar un mayor análisis de los resultados generados por la red FC-DenseNet56. dicha grafica muestra el IoU de cada una de las clases presente en el dataset, en otras palabras, se muestra la capacidad de FC-DenseNet56 para predecir cada uno de los distintos estratos. Este valor de IoU se ve reflejado tanto por la cantidad de imágenes usadas como por la presencia de los distintos estratos en dichas imágenes. Como se puede ver en la siguiente tabla las zonas sin estrato o clase 'NO DEFINIDO' cuenta con una presencia casi del 50 por ciento del dataset, lo que se refleja en su IoU,

siendo este el mayor entre todas las clases.

**Tabla 5:** Presencia de cada clase en el dataset

Porcentaje por clase	
CLASE	PORCENTAJE
Estrato 1	0.061223
Estrato 2	0.144790
Estrato 3	0.193266
Estrato 4	0.051771
Estrato 5	0.028800
Estrato 6	0.023021
No Definido	0.497130

En cambio, el 'ESTRATO 6' que tiene menor presencia en el dataset cuenta con el menor valor de IoU. Hay que aclarar que no se usaron técnicas para modificar los pesos de cada clase, ya que esto suele ser una alternativa para mejorar la calidad de las predicciones ante la ausencia o presencia de las clases.

## 4. CONCLUSIONES

- La mejor manera de predecir el nivel socio económico de una zona urbana usando imágenes satelitales es utilizar niveles de acercamiento satelital altos, capas de información extra en las imágenes satelitales y la red neuronal FC-DenseNet56 con valores de *Epoch* y *Batch-size* mayores o iguales a 70 y 1 respectivamente.
- Con los archivos KML que se crearon y usando el software ruso SASplanet se crea un dataset nombrado OVERLAECOBO, el cual consta de imágenes satelitales de 5 diferentes niveles de acercamiento satelital. El dataset cuenta con dos tipos de imágenes satelitales (*SIMPLE* y *COMPOUND*) e imágenes *Ground Truth* (*LABEL*) todas geográficamente correspondidas. las imágenes *COMPOUND* son imágenes satelitales con un canal extra el cual consiste en información sobre el riesgo que corren las mujeres en horarios nocturnos de las distintas localidades de Bogotá.
- Como método de predicción para el nivel socio económico de zonas urbanas se utiliza segmentación semántica, redes neuronales convoluciones y el Framework Semantic-Segmentation-Suite disponible en <https://github.com/GeorgeSeif/Semantic-Segmentation-Suite> el cual facilita el trabajo con los datos que se poseen.
- El uso de niveles de acercamiento satelital más amplios genera una mayor calidad en la predicción, esto es debido a que los *ZOOM* más grandes cuentan con un número mayor de imágenes y a la vez una mayor cantidad de información de la cual el modelo puede aprender.
- Como lo enuncia la teoría el uso de valores altos en la variable *Epoch* genera mejores resultados y el porcentaje de presencia de las clases en el dataset también afecta la predicción.

## 5. RECOMENDACIONES Y TRABAJO FUTURO

- Utilizar maquinas con mayor poder computacional.
- Realizar entrenamiento con FC-DenseNet56 variando los valores de *Epoch* y *Batch-size*, usando los *ZOOM* 16 o 17.
- Dar más peso a las clases que tienen menor presencia en el dataset.
- Cambiar de Framework dado a las limitaciones del mismo respecto a las capas de información extra.
- Entrenar las redes neuronales con imágenes satelitales con más de una capa de información extra.

## **6. LIMITACIONES Y PROBLEMAS**

- Los datos en el formato necesario no están disponibles en las páginas del gobierno, ni en alguna otra página en internet, debido a esto se crearon varios archivos KML de la ciudad de Bogotá, los cuales quedan disponibles para el público. Esta falta de datos generó retrasos en la elaboración del proyecto.
- El Framework únicamente permite el uso de imágenes con 4 canales.
- El poder computacional con el que se contaba

## BIBLIOGRAFIA

DANE. Estratificación - Preguntas frecuentes. [en línea]. <[https://www.dane.gov.co/files/geoestadistica/Preguntas\\_frecuentes\\_estratificacion.pdf](https://www.dane.gov.co/files/geoestadistica/Preguntas_frecuentes_estratificacion.pdf)>[citado en 25 de Mayo de 2018]

DANE. Metodología de estratificación. [en línea]. <<http://www.dane.gov.co/index.php/servicios-al-ciudadano/servicios-de-informacion/estratificacion-socioeconomica>>[citado en 25 de Mayo de 2018]

DANE. Procedimiento del cálculo. [en línea]. <<http://www.dane.gov.co/files/geoestadistica/estratificacion/procedimientoDeCalculo.pdf>>. [citado en 25 de Mayo de 2018].

NEAL jean. Combining satellite imagery and machine learning to predict poverty. [en Linea]. <<http://sustain.stanford.edu/predicting-poverty/>>[citado en 25 de Mayo de 2018].

NEAL jean. Combining satellite imagery and machine learning to predict poverty. [en linea]. <<https://github.com/nealjean/predicting-poverty>>[citado en 25 de Mayo de 2018]

NEAL Jean, MARSHALL Burke, † MICHAEL Xie, W. Matthew Davis, DAVID B. Lobell, STEFANO Ermon. Combining satellite imagery and machine learning to predict poverty. Science 353 (6301), p. 790-794. 2016

EAFIT. Con imágenes satelitales miden los índices de pobreza en Medellín. [en linea]. <<http://www.eafit.edu.co/investigacion/revistacientifica/edicion-167/Paginas/con-imagenes-satelitales-miden-los-indices-de-pobreza-en-medellin.aspx>>[citado en 25 de Mayo de 2018]

GONZÁLEZ Vanessa. ¿Qué es la estratificación social? [en línea]. <<https://www.lifeder.com/estratificacion-social/>>[citado en 31 de Mayo del 2018].

GONZÁLEZ Andrés. ¿Qué es Machine Learning? [en línea]. <<http://cleverdata.io/que-es-machine-learning-big-data/>>[citado en 10 de Octubre de 2018]

IBM. El modelo de redes neuronales [en línea]. <[https://www.ibm.com/support/knowledgecenter/es/SS3RA7\\_18.0.0/modeler\\_mainhelp\\_client\\_ddita/components/neuralnet/neuralnet\\_model.html](https://www.ibm.com/support/knowledgecenter/es/SS3RA7_18.0.0/modeler_mainhelp_client_ddita/components/neuralnet/neuralnet_model.html)>[citado en 31 de Mayo de 2018]

VÍLCHEZ GARCÍA, Víctor Gabriel. Estimación y clasificación de daños en materiales utilizando modelos AR y redes neuronales para la evaluaciónno destructiva con ultraso-nidos. [en linea]. <<http://ceres.ugr.es/~alumnos/esclas/>>[citado en 24 de Mayo de 2018]

MATHWORKS. Aprendizaje profundo [en línea]. <<https://es.mathworks.com/discovery/deep-learning.html>>[citado en 31 de Mayo de 2018]

MATHWORKS. Architecture of a CNN. [en linea] <https://la.mathworks.com/solutions/deep-learning/convolutional-neural-network.html> [Citado en 30 de junio de 2018]

Aphex34. Max pooling with a 2x2 filter and stride = 2 [en linea] [https://upload.wikimedia.org/wikipedia/commons/e/e9/Max\\_pooling.png](https://upload.wikimedia.org/wikipedia/commons/e/e9/Max_pooling.png) [citado en 30 de julio de 2018]

VENTURA Royo Carles Ventura. La segmentación semántica y sus benchmarks [en linea] <http://informatica.blogs.uoc.edu/2016/05/26/la-segmentacion-semantica-y-sus-benchmarks/> [citado en 30 de julio de 2018]

KENDALL Alex, BADRINARAYANAN Vijay, CIPOLLA Roberto. SegNet [en linea] <http://mi.eng.cam.ac.uk/projects/segnets/> [citado en 30 de julio de 2018]

GOBIERNO DIGITAL. Datos Abiertos [en linea] <<http://estrategia.gobiernoenlinea.gov.co/623/w3-article-9407.html>>[citado en 30 de Mayo de 2018]

ABHIJIT Kundu, VIBHAV Vineet, and VLANDLEN Koltun Feature Space Optimization for Semantic Video Segmentation [en linea] <<http://vladlen.info/publications/feature-space-optimization-for-semantic-video-segmentation/>>[citado en 01 de junio de 2018]

JEGOU Simon, DROZDZAL Michal, VAZQUEZ David, ROMERO Adriana, BENGIO Yoshua. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation [en linea] <<https://arxiv.org/abs/1611.09326>>[citado en 31 de julio de 2018]

HUANG Gao, LUI Zhuang, VAN DER Maaten Laurens, Q. Weinberger Kilian. Densely Connected Convolutional Networks [en linea] <<https://arxiv.org/abs/1608.06993>>[citado en 31 de julio de 2018]

SHELHAMER Evan, LONG Jonathan, DARRELL Trevor. Fully Convolutional Networks for Semantic Segmentation [en linea] <<https://arxiv.org/abs/1605.06211>>[citado en 31 de julio de 2018]

JEGOU Simon, DROZDZAL Michal, VAZQUEZ David, ROMERO Adriana, BENGIO Yoshua. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation [en linea] <<https://arxiv.org/abs/1611.09326>>[citado en 31 de julio de 2018]

KENDALL Alex, BADRINARAYANAN Vijay, CIPOLLA Roberto. SegNet [en linea] <http://mi.eng.cam.ac.uk/projects/segn/> [citado en 30 de julio de 2018]

H. Abdulnabi Abrar, WINKLER Stefan, WANG Gang. Beyond Forward Shortcuts: Fully Convolutional Master-Slave Networks (MSNets) with Backward Skip Connections for Semantic Segmentation [en linea] <<https://arxiv.org/abs/1707.05537>>[citado en 31 de julio de 2018]

POHLEN Tobias, HERMANS Alexander, MATHIAS Markus, LEIBE Bastian. Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes [en linea]

<<https://arxiv.org/abs/1611.08323>>[citado en 31 de julio de 2018]

POHLEN Tobias, HERMANS Alexander, MATHIAS Markus, LEIBE Bastian. Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes [en linea] <<https://arxiv.org/abs/1611.08323>>[citado en 31 de julio de 2018]

RONNEBERGER Olaf, FISCHER Philipp, BROX Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation [en linea] <<https://arxiv.org/abs/1505.04597>>[citado en 31 de julio de 2018]

G. Howard Andrew, ZHU Menglong, CHEN Bo, KALENICHENKO Dmitry, WANG Weijun, WEYAND Tobias, ANDRETTI Marco, ADAM Hartwig. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [en linea] <<https://arxiv.org/abs/1704.04861>>[citado en 31 de julio de 2018]

RONNEBERGER Olaf, FISCHER Philipp, BROX Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation [en linea] <<https://arxiv.org/abs/1505.04597>>[citado en 31 de julio de 2018]

ROSEBROCK Adrian. Intersection over Union (IoU) for object detection [en linea] <<https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>>[citado en 17 de Julio de 2018]

JEGOU Simon, DROZDZAL Michal, VAZQUEZ David, ROMERO Adriana, BENGIO Yoshua. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation [en Linea]<<https://arxiv.org/abs/1611.09326>>[citado en 23 de Julio de 2018]

HOWARD Andrew G., ZHU Menglong, CHEN Bo, KALENICHENKO Dmitry, WANG Weijun, WEYAND Tobias, ANDRETTI Marco, ADAM Hartwig. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [en linea]<<https://arxiv.org/abs/1704.04861>>[citado en 23 de Julio de 2018]

SHARMA Sagar. Epoch vs Batch Size vs Iterations [en linea] <<https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9>>[citado en 23 Julio de 2018]