Descubriendo la distribución espacial de comercio en ciudades con economías informales.

Estudiante: Juan Camilo Saldarriaga[1]
Director Tesis: Juan Carlos Duque[1]
Codirector Tesis: Jairo Alejandro Gómez[1]

Resumen
Un censo económico tiene como objetivo registrar la actividad económica en una ciudad recopilando datos de encuestas georreferenciadas. Aunque sus beneficios son importantes, los costos son muy elevados y, por esto, es muy raro que la información del censo económico esté actualizada y completa. En este trabajo proponemos una nueva metodología para detectar y georreferenciar la actividad comercial visible en una ciudad o región de manera eficiente, generando reportes exhaustivos automatizados de la actividad comercial visible en una zona de interés. Esta metodología intenta estimar la distribución espacial que permite tener un censo económico pero únicamente para el comercio visible. Contrastamos los resultados de nuestra metodología con información oficial de Cámara de Comercio para estimar la distribución espacial del comercio visible informal o no registrado públicamente en el municipio de Envigado.

Keywords
Estimación de comercio, distribución de comercio, economía espacial, informalidad empresarial, visión por computador, aprendizaje profundo, sistemas de información geográfica, retos urbanos.

[1] RiSE-group, Departmento de Ciencias Matemáticas, Universidad EAFIT, Medellín, Colombia

Corresponding author:
Autor correspondiente: (JCS) RiSE Group"(http://www.rise-group.org/). Universidad EAFIT, Carrera 49 No. 7 Sur-50, 050022 Medellin, Colombia.
Email: jcsaldarrm@eafit.edu.co

Prepared using sagej.cls *[Version: 2020/03/11 v3.00]*

1. Introducción

Un censo económico tiene como objetivo capturar completamente la actividad económica en una ciudad mediante una extensa recopilación de datos de encuestas georreferenciadas. Su valor como herramienta de política se ha resaltado con anterioridad (DANE 2019a), mostrando que permite a los responsables de políticas identificar vecindarios prometedores, áreas afectadas por nuevas políticas, problemas de uso de la tierra o valor de la tierra, etc. Los censos económicos proporcionan una mejor comprensión de la dinámica comercial al tomar una imagen completa de su presencia en el territorio. Así mismo, permiten tener una medida sobre los subregistros públicos, la evasión de impuestos y la informalidad comercial en entornos urbanos. Los censos económicos que se vienen realizando, tienen un retraso entre el momento en el que se censa y en el que se publican los resultados que puede ser de hasta 1 año (Bureau 2019), también se realizan con poca frecuencia, aproximadamente cada 4 años (DANE 2019b).

El mejor entendimiento de la dinámica comercial que proporciona un censo es particularmente importante en contextos de alta informalidad como el de América Latina. Según cifras oficiales, América Latina es una de las regiones del mundo con mayor informalidad laboral, cerca del 80 % (Perry et al. 2010) de los trabajadores son contratados en condiciones informales. También tiene cifras altas de informalidad comercial pues cerca del 75 % de las empresas no cuentan con un registro mercantil (Fernández 2018). Así mismo, en esta región se evidencia una tendencia en la que se crean muchas empresas pero sobreviven poco tiempo, por ejemplo para Colombia cerca de la mitad de las microempresas no logra sobrevivir más allá de 5 años (CCMA 2018), lo que a su vez puede estimular la operación informal. Adicionalmente, cerca del 92,8 % del total de las empresas en Colombia son pequeñas y familiares (CCMA 2018), las cuales por su naturaleza son mucho más difíciles de monitorear que las más grandes y estables. Debido a estas características, las empresas en América Latina merecen una atención especial y acompañamiento, y se hace muy relevante contar con registros actualizados frecuentemente de las empresas comerciales que ofrecen sus servicios en las ciudades.

Aunque los beneficios de realizar un censo económico completo son muy significativos, los costos son muy elevados y, por esto es muy raro que la información disponible esté actualizada y completa. Los censos económicos tradicionales son costosos, pues necesitan de un gran número de encuestadores para capturar toda la información necesaria (DANE 2019b), por ejemplo, en 2021 se realizará un censo económico para Colombia el cuál se estima que costará unos 280.000 millones de pesos (Portafolio 2019) y el cual no se realizaba desde 1991. Por otra parte, la informalidad en América Latina agrava el problema de desinformación, debido a que existe un porcentaje importante de establecimientos comerciales que nunca se han registrado en cifras públicas. Además, la tendencia a la creación y cierre rápido de las empresas desactualiza la información de los censos más de lo deseado, propiciando una pérdida acelerada de vigencia en la información sobre la actividad comercial. Los costos elevados para adquirir y actualizar la información hacen que la mayoría de las ciudades latinoamericanas opten por no realizar censos económicos debido a la falta de recursos financieros. Por lo tanto, existen grandes dificultades para seguir con frecuencia el comercio en América Latina, en especial el de la pequeña y mediana empresa.

En esta tesis de maestría, proponemos una nueva metodología que utiliza herramientas de la inteligencia artificial, en particular del área de visión por computador y de aprendizaje de máquina, para detectar y georreferenciar la actividad comercial en una ciudad o región de forma eficiente y a un costo menor que el tradicional. Esta metodología consiste de un algoritmo que detecta la presencia de comercio en una escena urbana con vista desde la calle y de un modelo matemático que usa las detecciones comerciales para producir un mapa de la distribución espacial para una ciudad o región de estudio. Nosotros usamos las escenas con vista desde la calle para detectar las empresas comerciales vistas por los peatones que cruzan las calles de las ciudades, las cuales son un conjunto importante del total de empresas comerciales. Esta metodología automatizada, permite realizar exploraciones de la actividad comercial sin ser un censo económico, facilitando la recolección de información valiosa a un menor costo y con un menor tiempo de retraso entre la adquisición de los datos y la publicación de resultados, en comparación a las formas más tradicionales de censar el comercio. Como se verá en la Sección 4.4 las imágenes son actualizadas anualmente y tienen un costo aproximado de siete dólares americanos por cada mil imágenes. Aplicaremos la metodología propuesta en el municipio de Envigado, Colombia, comparando las cifras oficiales del municipio con las predicciones de nuestro exploración de la actividad comercial automatizada para estimar la distribución espacial de las diferencias, aportando nueva evidencia sobre la distribución espacial del comercio informal.

La posibilidad de realizar exploraciones comerciales automatizadas de bajo costo, al poder hacerse varias mediciones en cortos periodos de tiempo, abre la oportunidad de estudiar cómo es la dinámica empresarial en una ciudad? qué tanto se cierran locales comerciales? cada cuánto se abren nuevos comercios? entre otros. Es decir, al hacer mediciones más frecuentes de la distribución espacial del comercio, se puede estudiar la demografía empresarial en una ciudad. Adicionalmente, esta posibilidad puede permitir a los gobiernos locales identificar oportunamente las empresas evasoras e informales, por su menor retraso entre el censo y la publicación de resultados, lo que puede permitir mayores recaudos de impuestos, mejores controles a la informalidad, entre otros. El valor de esta innovación es particularmente relevante en el contexto de América Latina, donde gran proporción de la actividad comercial es realizada informalmente, y por lo tanto nunca han existido registros oficiales para ellas. Así mismo, las exploraciones de la actividad comercial automatizadas se pueden aplicar con una mayor frecuencia que los tradicionales, permitiendo en América Latina mejorar el seguimiento a la actividad comercial debido a las altas tasas de creación y muerte de las empresas. Su bajo costo permite también a las ciudades más pequeñas y de menores recursos incorporar esta información tan importante para sus economías.

Esta tesis busca brindar información para contribuir con los Objetivos de Desarrollo Sostenible (ODS ó SDG) planteados por la Organización de Naciones Unidas (ONU) para el 2030. En especial, el ODS 8 que busca asegurar el trabajo decente, pleno empleo, empleo digno y de calidad (Martínez Agut 2015), deja en una situación compleja a las ciudades de América Latina debido a los altos niveles de informalidad con los que cuentan. El reto es grande debido a que las ciudades latinoamericanas están altamente construidas y urbanizadas, haciendo necesario revisar con otro enfoque lo ya existente, a fin de proponer soluciones innovadoras a los problemas tradicionales. Esta tesis busca brindar herramientas a los gobiernos para que tengan información actualizada que les permita tomar mejores decisiones y puedan así cumplir con los ODS.

La estructura de esta tesis se describe a continuación. Inicialmente, se introduce el concepto de censo económico automatizado y su importancia en contextos altamente informales como el de América Latina. Después, se presenta una revisión de literatura en detección automática de comercio en imágenes tomadas desde la calle. Posteriormente, se explica el marco teórico y se introducen algunos conceptos que se abordan en la metodología. Luego se describen los datos usados, la metodología propuesta y el lugar de estudio dónde se hicieron las pruebas. Finalmente, se muestran los principales resultados y se resaltan algunas conclusiones.

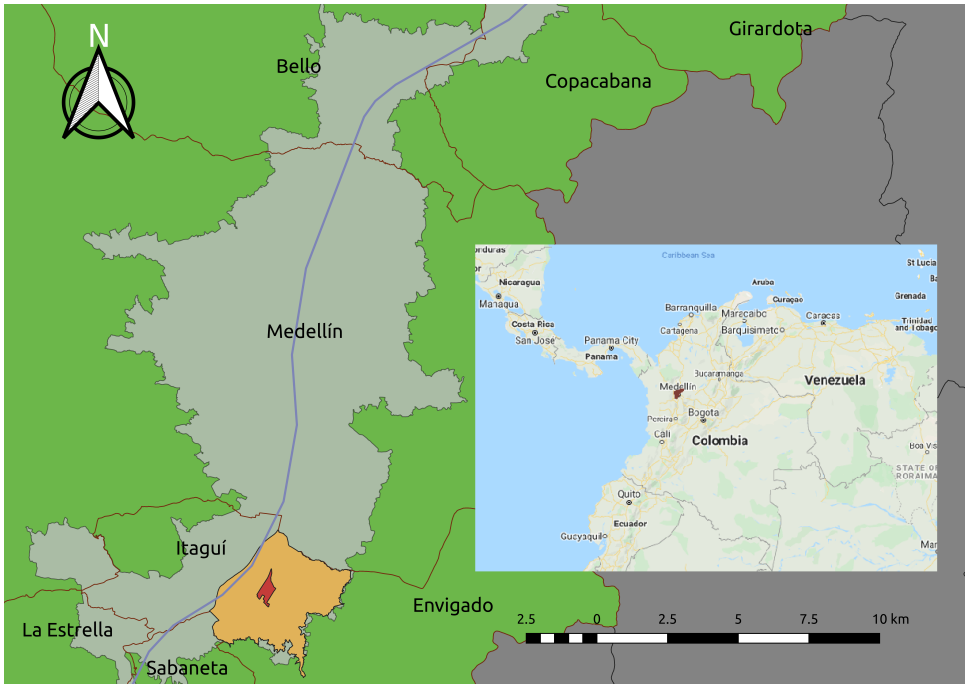

Figura 1. Mapa(s) que representa a Colombia, la ubicación del Valle de Aburra (verde), la ubicación de Envigado dentro del Valle (amarillo), y la ubicación del Centro de Envigado (rojo).

2. Revisión de literatura

La necesidad de conocer dónde está ubicada la actividad comercial es una pregunta recurrente en la literatura (Lee & Pace 2005). En algunos estudios previos, se reportan estrategias extensivas como los censos comerciales, y otras muestrales como las encuestas. La actividad comercial se puede estimar cruzando bases de datos oficiales, como por ejemplo las de Cámara de Comercio, Seguridad Social, Industria y Comercio, etc. (Fernández 2018). Sin embargo, la estimación de la información de comercios a partir de las bases de datos con información oficial, no siempre genera la confianza suficiente en los gobiernos para utilizarla como insumo para la política pública. Adicionalmente, estas aproximaciones son

muy efectivas en ciudades con altos niveles de formalidad empresarial, en los cuales los registros oficiales contienen todo el universo de empresas, pero no funcionan tan bien en contextos de informalidad. La ONU resalta la importancia de los censos económicos como fuente de datos que permite diseñar y ejecutar políticas publicas (UN 2010), pues cubre tanto a las empresas formales, como informales.

Otra forma poco tradicional de construir la información de distribución espacial del comercio, es a partir de la clasificación del uso del suelo, que es un insumo importante de política (Grippa et al. 2018). Si se conoce en que usos está distribuido el suelo urbano de las ciudades, se puede identificar qué suelo se está usando comercialmente, estos usos del suelo son mapas de la ciudad que muestran por colores las diferentes categorías. Sin embargo, estos mapas no siempre están disponibles, y los que se tienen muchas veces no se construyen de la mejor manera (Hu et al. 2016). En la literatura existen una serie de estudios que utilizan algoritmos de inteligencia de máquina junto con datos de sensores remotos y diferentes fuentes de datos libres, para entrenar clasificadores automáticos que determinan el uso del suelo, principalmente en países desarrollados. Es de rescatar que los mapas resultado de las predicciones proveen información más detallada del patrón espacial del uso del suelo que los que usan los gobiernos locales, pero se centran en unidades espaciales muy grandes como lo son bloques de calle o edificios.

Algunos de estos algoritmos de inteligencia artificial también se están usando para identificar y ubicar el comercio. Las imágenes con vista de calle o imágenes de calle, son una fuente importante de datos que están disponibles para todo el mundo. Algunos proveedores de este tipo de imágenes incluye: Google Street View, Mapillary, NavVis, y Microsoft Live Earth (Anguelov et al. 2010). En muchas ocasiones estas imágenes capturan información de algunos lugares de los que no se tiene registro en ninguna otra parte. Sus usos van encaminados a la extracción de información del mundo real tal como: estimación de variables socio-económicas (Ilic et al. 2019), niveles de ingreso (Acharya et al. 2017), mediciones de zonas verdes (Lu 2018), evaluaciones sociales de los niveles de accesibilidad de personas con movilidad reducida (Hara et al. 2013), percepciones de seguridad (Porzi et al. 2015), estudios de criminalidad (Kang & Kang 2017), riesgo de accidentes peatonales (Mooney et al. 2016), riesgo sísmico, eficiencia energética, y cambios en el tiempo de las fachadas (Tang & Long 2018), entre otras.

Entre los usos que se le dan a las imágenes de calle, también están los estudios relacionados con identificar fachadas comerciales en escenas urbanas, en los cuales nos centraremos en esta tesis. Su taxonomía se puede resumir en tres tipos, según la metodología con la que se aborda la problemática: Reconocimiento Óptico de Caracteres (OCR), Clasificación, y Detección. Zamir et al. (2011) identifican el texto de un conjunto de imágenes georrefenciadas usando OCR, lo usan como un letrero comercial y lo buscan en una base de datos de páginas amarillas (Yellow Pages)(thryv 2020), los autores aplicaron esta metodología en algunas calles de San Francisco y Pittsburgh. Por su parte Iovan et al. (2012), entrenan un clasificador de escenas urbanas a partir de imágenes de calle, entre los que se incluye una categoría de comercio, y evalúan en un distrito de París con imágenes tomadas por ellos mismos. En otro estudio, Movshovitz-Attias et al. (2015), utilizan las imágenes de Google StreetView (GSV) para entrenar una clasificación ontológica con la cual logran clasificar los comercios en 208 categorías únicas, proponen una metodología para propagar etiquetas de categorías comerciales, y exponen las principales dificultades asociadas a la clasificación de fachadas comerciales.

A continuación se mencionan algunos estudios que utilizan algoritmos de detección sobre las imágenes con vista desde la calle, de forma similar a como se hace en esta tesis. Laupheimer et al. (2018) usan un detector de objetos basado en un Faster R-CNN (Ren et al. 2017) con las etiquetas de la base de datos ImageNet (Deng et al. 2009) para clasificar los edificios en una base de datos de imágenes de GSV, posteriormente los edificios detectados los pasan por un clasificador que entrenan, y que lo ubica en una de cinco categorías entre la que se encuentra la comercial. Un trabajo similar al de esta tesis, es el de Yu et al. (2015), en el cual hacen un descubrimiento de negocios a gran escala usando imágenes de GSV, etiquetan una base de datos de cerca de dos millones de panorámicas en 12 países. Este estudio hace parte de una serie de estudios priorizados por Google para detectar el comercio en sus imágenes, con el fin de actualizar automáticamente su servicio de Google Maps. Liao et al. (2018) proponen una metodología para emparejar automáticamente información de imágenes de calle con contenido web, detectando los letreros comerciales de imágenes al interior de un centro comercial y buscando estos letreros en una base de datos de referencia, el estudio se hace en un centro comercial japonés.

El aporte de esta tesis es el diseño de una metodología que, usando tecnologías de Inteligencia de Máquina, permita hacer barridos exhaustivos automatizados de una zona de interés, contando y estimando la distribución espacial de la actividad comercial. Esta contribución es particularmente importante en el contexto de las ciudades latinoamericanas, en las que la distribución de la actividad comercial se desconoce en muchos casos. A diferencia de otros estudios, nuestra metodología no se aplica sobre imágenes panorámicas sino sobre imágenes que corresponden a fachadas laterales. Adicionalmente buscamos crear nueva información en contextos de alta informalidad, a un menor costo, que permita crear estadísticas económicas para los gobiernos.

3. Marco teórico

En esta sección mostramos la fundamentación matemática sobre la que se soporta esta tesis, la cual es producto del proceso de formación recibida en la Maestría en Matemáticas Aplicadas con énfasis en Estudios Espaciales. En la Figura 2 se evidencian los grandes bloques de este proceso, los cuales son explicados en esta Sección, los sub-bloques se abordarán en la Sección 5. En la Tabla 1 se presenta un resumen con la terminología y principales variables utilizadas durante el desarrollo de esta tesis. El marco teórico inicia con una descripción de las técnicas de procesamiento de imágenes empleadas en esta investigación, las cuales son usadas para generar los datos de referencia para esta tesis. Posteriormente, se hace una breve explicación de los principios fundamentales que utiliza un modelo de detección de objetos usando Aprendizaje Profundo y las formas de caracterizar su desempeño. También se hace una breve introducción sobre la fundamentación económica de informalidad empresarial. Finalmente, se explica el componente espacial de la investigación, su descripción y los modelos de análisis espacial empleados.

Tabla 1. Terminología empleada en esta tesis.

Variable	Descripción
W, H	Número de píxeles de ancho y alto en la imagen original.
W_d, H_d	Número de píxeles de alto en la imagen cubo de destino.
C	Número de píxeles en una cara del cubo.
X, Y, Z	Las coordenadas o ejes de una esfera 3-D.
i, j, k	Variables auxiliares que representan índices.
m, n	Variables auxiliares que representan número de datos.
θ	Inclinación, ángulo latitudinal o lineas horizontales de una esfera.
ϕ	Azimuth, ángulo polar o longitudinal, o lineas verticales de una esfera.
R	El radio de la tierra.
$\hat{x}_1, \hat{x}_2$	El valor mínimo y máximo de X del recuadro predicho.
$\hat{y}_1, \hat{y}_2$	El valor mínimo y máximo de Y del recuadro predicho.
$K(m, n)$	Función de Kernel o filtro de tamaño m$\times n$
I	Imagen o colección de entrada
$\mathbb{Y}$	Es el valor real o de referencia del modelo.
$\hat{\mathbb{Y}}$	Es la predicción del modelo.
$J(\mathbb{Y}, \hat{\mathbb{Y}})$	Función de costo que relaciona a $\mathbb{Y}$ con $\hat{\mathbb{Y}}$.
w_i	Parámetro i de la red neuronal.
A_n	Activación para la categoría n.
A	Vector de activaciones de todas las categorías.
r_i	i-ésimo valor de *recall* o sensibilidad ordenado.
ρ	Métrica precisión.
$\hat{\mathbb{B}}_i, \mathbb{B}_i$	Recuadro de interés predicho y verdadero.
τ	Parámetro de balanceo.
L_{cls}, L_{reg}	Funciones de costo de clasificación y de ubicación.
N_{cls}, N_{reg}	Factores de normalización de las clasificaciones y las ubicaciones.
p_{comm}	Participación del comercio predicho sobre el comercio real.
err_{comm}	Errores en el conteo de comercios.
c_i, $\hat{c}_i$	Número real y de predicciones de establecimientos comerciales en la imagen i-ésima.
P	Es el número total de puntos muestreados.
L	El número de puntos muestreados de las aristas.
N	El número de puntos muestreados de los nodos.
α	Tasa de aprendizaje.
∇J	Gradiente de J.
$\hat{\lambda}_k(\mathrm{x})$	Estimación de densidad de Kernel espacial para x.
d	Es la distancia medida en metros.
s_i	Sobrante de una arista i.
h	Amplitud de banda en unidades de distancia.
p_i, p_i^*	El número de puntos en una arista i con sus decimales y su versión entera.
l_i	Longitud de una arista i medida en metros.

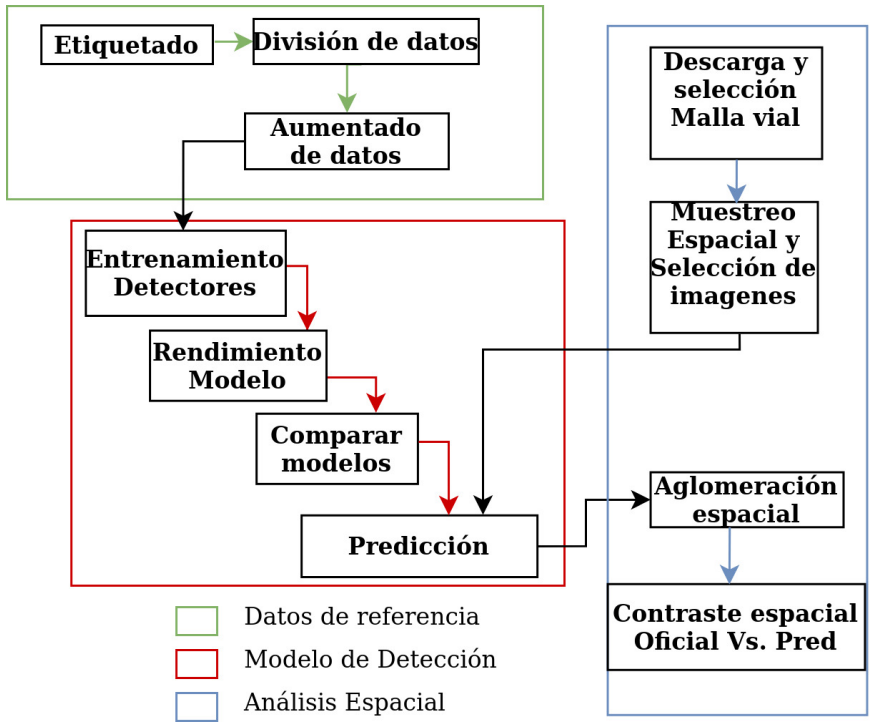

Figura 2. Diagrama de flujo de la metodología.

3.1. Procesamiento de imágenes

El procesamiento digital de imágenes (Gonzalez 2002) es un proceso que puede ser entendido con álgebra matricial, en el cual cada canal de una imagen es representado en una matriz de intensidades, y cada elemento de la matriz representa un píxel. A su vez, la visión por computador busca emular a través de algoritmos la capacidad humana que nos permite interpretar información visual compleja. La intensidad de cada píxel en cada canal se codifica usualmente con 8 bits, lo que se traduce en 256 niveles, y por ende, la codificación en RGB requiere 24 bits por píxel. Este formato de imágenes son las que usan las cámaras fotográficas convencionales. Sin embargo, existen otras imágenes con más de tres canales como es el caso de las imágenes satélites, que capturan bandas adicionales del espectro electromagnético. También hay imágenes de un sólo canal que se conocen como imágenes en escala de grises que pueden representar información de cualquier tipo. Para esta tesis entendemos ancho de la imagen (W) y altura de la imagen (H), como los tamaños de las imágenes RGB usadas en número de píxeles, o de manera equivalente, las dimensiones de las matrices con las que se representa ($H \times W$).

Las imágenes capturadas con perspectiva panorámica de 360° son llamadas imágenes panorámicas y reflejan la información de una esfera tridimensional, definida en los ejes X, Y, Z como la de la Figura 3.

Estas imágenes suelen ser tomadas por cámaras de 360° que capturan la escena con dos lentes de ojo de pescado, lo que les permite fotografiar en todos los ángulos. El formato equirectangular es la representación de una imagen panorámica, en una imagen 2-D o plana, en la que se ajusta la curvatura de la esfera a las márgenes de la imagen, generando deformaciones en los extremos. Las deformidades del formato equirectangular capturan la profundidad de la escena fotografiada, de modo que con la adecuada transformación, se puede representar en visores 3-D de realidad aumentada. Como muestran Yu et al. (2015), las imágenes 360° de Google Street View vienen en formato equirectangular, la cual es una proyección en un espacio panorama.

3.1.1. De esfera a cubo

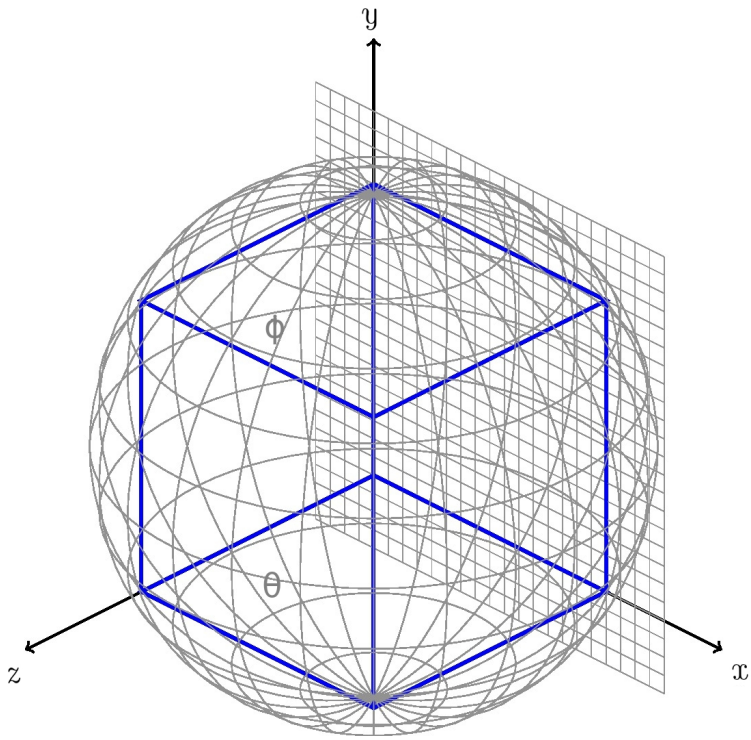

Figura 3. La representación de un cubo dentro de una esfera.

El primer paso para procesar las imágenes panorámicas consiste en realizar una proyección geométrica de una esfera a un cubo, como se ilustra en la Figura 3, al usar coordenadas esféricas, que permiten llevar la imagen de la esfera a las seis caras de un cubo que solo estén definidas en dos ejes, X y Y. Para la proyección se utiliza el sistema de ecuaciones, descrito en las ecuaciones 1, 2 y 3, con las cuales se logra reubicar cada píxel en una nueva imagen (Hartley & Zisserman 2004). El proceso inicia en la Ecuación 1 extrayendo para cada píxel las coordenadas X, Y, Z de la esfera a partir de la imagen equirectangular.

Seguidamente en la Ecuación 2 se calculan las proyecciones en coordenadas esféricas de cada píxel, y finalmente se calculan las ubicaciones de cada píxel dentro del cubo proyectado: $u_{i,j}$, $v_{i,j}$. Este sistema de ecuaciones produce dos matrices de ubicaciones de píxeles, una para X y una para Y. La matriz de ubicaciones de píxeles de X, contiene las ubicaciones en X que tendrá cada píxel original, y se construye con $u_{i,j}$ para todos los valores de i, j. Similarmente, la matriz de ubicaciones de Y, se construye con $v_{i,j}$ para todos los valores de i, j. Estas dos matrices se utilizan como mapas de píxeles para la proyección de la imagen al cubo, la cual tendrá un tamaño en número de píxeles de $3C \times W$, donde C es el número de píxeles en una cara del cubo,

La proyección usa los siguientes parámetros:
W, H: número de píxeles de ancho y alto en la imagen equirectangular original.
W_d, H_d: número de píxeles de ancho y alto en la imagen cubo de destino.
C: número de píxeles en una cara del cubo, $C = W/4$
$W_d = W$
$H_d = 3c$
i: índice de píxeles en X, iniciando desde la esquina inferior izquierda, $i = \{0, 1, ..W - 1\}$.
j: índice de píxeles en Y, $j = \begin{cases} \{0, 1, 2, ..., 3C - 1\} & \text{si } 2C < i < 3C \\ \{C, C + 1..., 2C - 1\} & \text{en otro caso.} \end{cases}$
$X_{i,j}, Y_{i,j}, Z_j$: las coordenadas X, Y, Z de la esfera.
θ: inclinación, ángulo latitudinal o lineas horizontales de la esfera, $0 \leq \theta \leq \pi$.
ϕ: azimuth, ángulo polar o longitudinal, o lineas verticales de la esfera, $0 \leq \phi \leq 2\pi$.

$$
\begin{aligned}
X_{i,j} &= \begin{cases} -1 & \text{si } i < C \\ \frac{2i}{C} - 3 & \text{si } C < i < 2C \\ \frac{2j}{C} - 1 & \text{si } 2C < i < 3C \text{ y } j < C \\ 1 & \text{si } 2C < i < 3C \text{ y } C < j < 2C \\ 5 - \frac{2j}{C} & \text{si } 2C < i < 3C \text{ y } j \geq 2C \\ 7 - \frac{2j}{C} & \text{si } i \geq 3C \end{cases} \\
Y_{i,j} &= \begin{cases} 1 - \frac{2i}{C} & \text{si } i < C \\ -1 & \text{si } C < i < 2C \\ \frac{2i}{C} - 5 & \text{si } 2C < i < 3C \\ 1 & \text{si } i \geq 3C \end{cases} \\
Z_j &= \begin{cases} 3 - \frac{2j}{C} & \text{si } C < j < 2C \\ 1 & \text{si } j < C \text{ ó } j \geq 2C \end{cases}
\end{aligned}
\tag{1}
$$

$$\begin{aligned}\phi_{ij} &= \tan^{-1}\left(\frac{Z_j}{\sqrt{X_{ij}^2 + Y_{ij}^2}}\right) \\ \theta_{ij} &= \tan^{-1}\left(\frac{Y_{ij}}{X_{ij}}\right)\end{aligned} \tag{2}$$

$$\begin{aligned}\mathrm{u}_{i,j} &= \left\lfloor \frac{2c(\theta_{ij} + \pi))}{\pi} \right\rfloor \\ \mathrm{v}_{i,j} &= \left\lfloor \frac{2c(\pi/2 - \phi_{ij})}{\pi} \right\rfloor\end{aligned} \tag{3}$$

3.1.2. Transformación de perspectiva

Una transformación de perspectiva permite que cualquier cuadrilátero se transforme en otro cuadrilátero (ambos convexos). Para hacer la transformación de perspectiva es necesario calcular una máscara, usando cuatro puntos de control de la imagen original y los cuatro puntos de destino. El algoritmo de procesamiento de imágenes lleva los píxeles de los puntos de referencia a los cuatro puntos de destino en el mismo orden que se le dieron, y transforma todos los demás píxeles de forma acorde. Para los píxeles en la nueva imagen que no tienen datos, se hace una interpolación con los valores de los píxeles cercanos, haciendo que visualmente la imagen adquiera la deformación de perspectiva.

3.1.3. Detección de Objetos

Una de las tareas más importantes en procesamiento de imágenes y visión por computador es la detección de objetos (Forsyth & Ponce 2012). En términos generales, un detector de objetos busca ubicar recuadros que encierren un objeto de interés en una imagen. Esta tarea es diferente a la de clasificación, en la cual se busca asignar una clase a la imagen completa. La detección de objetos combina dos tareas de visión por computador: la ubicación de objetos y la clasificación. Usando algoritmos de aprendizaje supervisado, se le enseña al detector de objetos los recuadros que debe aprender a predecir, mediante imágenes de entrenamiento, con una base de datos llamada verdad absoluta (*Ground Truth*). La Ecuación 4 y la Figura 4 representan la formulación de un recuadro de interés (o *bounding box*).

$$\begin{aligned}\hat{x}_1, \hat{x}_2 \in 0 \leq \hat{x}_1 < \hat{x}_2 \leq W \\ \hat{y}_1, \hat{y}_2 \in 0 \leq \hat{y}_1 < \hat{y}_2 \leq H\end{aligned} \tag{4}$$

$\hat{x}_1$: el valor mínimo de X del recuadro predicho.
$\hat{x}_2$: el valor máximo de X del recuadro predicho.
$\hat{y}_1$: el valor mínimo de Y del recuadro predicho.
$\hat{y}_2$: el valor máximo de Y del recuadro predicho.

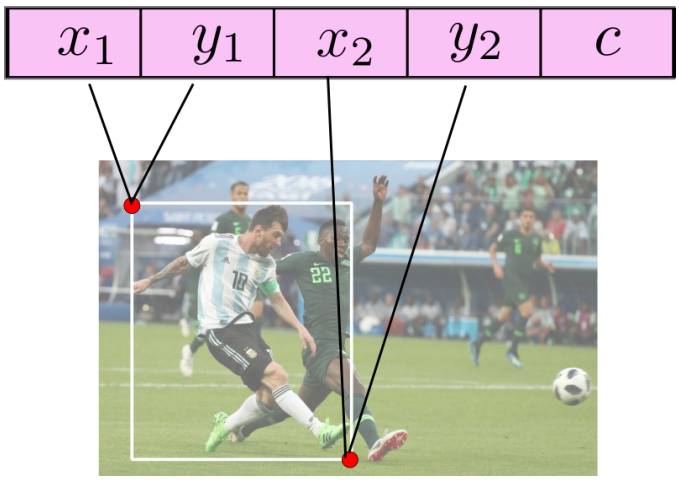

Figura 4. Ejemplo de los recuadros de interes, c es la categoria a la que pertenece el recuadro. Tomado de: https://blog.paperspace.com/data-augmentation-for-bounding-boxes/, fecha de acceso: 4 de Marzo de 2020.

En la actualidad, los mejores detectores de objetos se basan en aprendizaje profundo y utilizan redes neuronales artificiales. Este tipo de detectores se optimizan y predicen en un menor tiempo, logrando el estado del arte en precisiones en las detecciones (Sermanet et al. 2014). Estos detectores operan en dos fases, primero proponen regiones de interés sobre una imagen, y posteriormente las clasifican.

3.2. Redes neuronales convolucionales

Una red neuronal puede ser entendida como una función matemática que mapea un conjunto de valores de entrada a valores de salida. Esta función es un compuesto de funciones más simples llamadas neuronas, donde cada neurona es definida por un peso y un sesgo. El término red viene del hecho de que es un compuesto de funciones, y el neuronal de que el concepto se basa en la Neurociencia (Goodfellow et al. 2016). Una de las arquitecturas más representativas en redes neuronales es el perceptrón multicapa (*Multilayer Perceptron* -MLP) (Marsland 2014), el cual está compuesto por unas neuronas de entrada, unas capas ocultas o intermedias y unas neuronas de salida. La información fluye al interior de la red desde las neuronas de entrada hacia las de salida, usando funciones de activación, haciendo que cada neurona aprenda cosas diferentes, y por ende se activen de manera diferente ante los estímulos de las neuronas anteriores. Cada neurona tiene pesos que hacen las veces de conexiones entre las neuronas,

sus valores reflejan que tanto reacciona una neurona ante los estímulos de las neuronas anteriores. Las neuronas de entrada de la red son una para cada variable explicativa. La capa del final suele estar configurada con una arquitectura completamente conectada (*fully connected*) con una neurona de salida para cada clase que se deseada aprender. Entre más neuronas, mayor número de parámetros se deben estimar para la red.

La correlación-cruzada (*cross-correlation*) es una operación lineal en la cual se utiliza un conjunto de pesos para multiplicar los elementos de una colección de entrada, y luego se suman sus productos (Gonzalez 2002). La Ecuación 5 muestra la formula de una correlación-cruzada para el caso 2-D. En el caso de colecciones 2-D o imágenes, el conjunto de pesos tiene la forma de una matriz de pesos $K(m,n)$, que tiene el nombre de filtro o Kernel, con el cual se hace un producto punto con una parte (del mismo tamaño del filtro) de la matriz de entrada o la imagen, lo que produce un valor escalar, la Figura 5 ilustra esta operación en el caso 2-D. La convolución matemática es muy similar a la función de correlación-cruzada, como se muestra en la Ecuación 6, con la diferencia de que el filtro se gira 180° antes de multiplicar elemento a elemento. En Aprendizaje Profundo (*Deep Learning*) y bibliotecas de Aprendizaje de Máquina suelen referirse a convolución cuando lo que se hace es correlación-cruzada (Goodfellow et al. 2016). En procesamiento de imágenes, esta operación también es llamada filtrado espacial. En esta tesis, se usará el termino de convolución para referirnos al proceso de correlacionar un Kernel.

$$(I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n) \tag{5}$$

$$(I * K)(i,j) = \sum_m \sum_n I(i-m, j-n) K(m,n) \tag{6}$$

K: Kernel o filtro convolucionado.
I: Imagen o colección de entrada.
m, n: dimensiones del Kernel o filtro.
i, j: cada posible ubicación en la Imagen o colección de entrada.

Una capa convolucional consiste en aplicar un proceso de filtrado espacial, o convolución, a todas las posibles partes de la entrada, produciendo un mapa de activación (*feature map*), el cual tiene una menor dimensión que la entrada y refleja las partes de la entrada que se activaron con los filtros. Se produce un mapa de activación por cada filtro en cada capa convolucional. Una red neuronal convolucional (*Convolutional Neural Networks* - CNN) es una en la que al menos una de sus capas intermedias es una capa convolucional, y en las capas de niveles superiores se pueden aplicar capas convolucionales sobre los mapas de activación de las capas anteriores. Los parámetros que se modifican en el proceso de optimización de la red son los pesos que contiene cada filtro, en cada capa, ajustando los parámetros de modo que mejor le sirvan para diferenciar entre categorías de interés. En el caso de que la entrada de una CNN sean imágenes, los filtros de primer nivel aprenden características visuales básicas, como por

ejemplo lineas, bordes, degradaciones de color, entre otras. En las capas de más hacia el final, la CNN aprende características visuales más avanzadas y elaboradas, como por ejemplo ojos, llantas, bloques de edificios, etc.

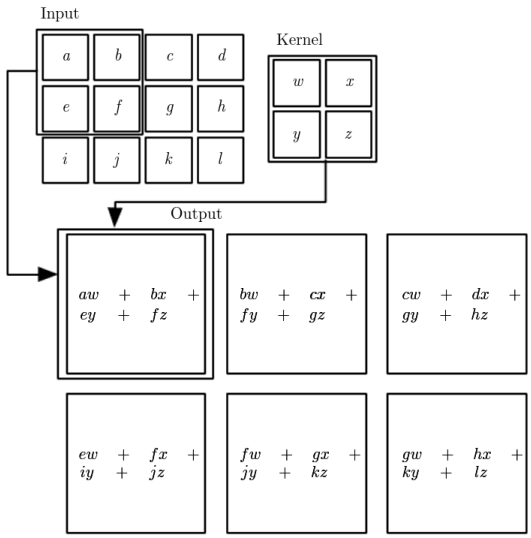

Figura 5. Ejemplo del proceso de convolución 2-D. Tomado de Goodfellow et al. (2016), p330.

3.2.1. Optimización de un modelo de CNN

$$\min_{w_0,...,w_n} J(\mathbb{Y}, \hat{\mathbb{Y}}(w_0,...,w_n)) = J(w_0,...,w_n) \tag{7}$$

$\mathbb{Y}$: es el valor real.
$\hat{\mathbb{Y}}$: es la predicción del modelo.
J: función de costo que relaciona a $\mathbb{Y}$ con $\hat{\mathbb{Y}}$.
w_i: parámetro i de la red neuronal.

Cuando se optimiza una red neuronal, el objetivo es minimizar una función de costo que está asociada al error de predicción del modelo, que se obtiene al comparar los valores reales con los estimados, como se indica en la Ecuación 7. Esta función de costo depende de los parámetros de la red neuronal. En tareas de clasificación, una de las funciones de costo más usadas en redes neuronales y teoría de la información, es la de entropía cruzada categórica (*categorical cross-entropy*) (Marsland 2014). La Ecuación 8 refleja esta entropía cruzada, la cual mide que tan lejos está el valor real (que puede tomar valores 0 o 1) de la predicción de la clasificación.

$$J(\mathbb{Y}, \hat{\mathbb{Y}}) = -\sum_{j=0}^{M}\sum_{i=0}^{N}(\mathbb{Y}_{ij} * \log_2(\hat{\mathbb{Y}}_{ij})) \tag{8}$$

i, N: índice y conjunto de observaciones, i = {1,...,N }.
j, M: índice y conjunto de categorías, i = {1,...,M }.
$\mathbb{Y}_{ij}$: Es el valor real.
$\hat{\mathbb{Y}}_{ij}$: Es la predicción del modelo.

También, en el caso de clasificación, se suelen usar funciones softmax, como la de la Ecuación 9 al final de la red neuronal en una etapa de post-procesamiento. Esta función normaliza los valores de las activaciones de las neuronas de salida (uno para cada categoría de interés) de modo que calcula la predicción como un vector de confianza de la pertenencia a cada categoría, en el cual la posición con el valor más alto corresponde a la categoría predicha.

$$softmax(A_n) = \frac{e^{A_n}}{\|e^A\|} \tag{9}$$

A_n: activación para la categoría n.
A: vector de activaciones de todas las categorías.

El algoritmo de gradiente descendente busca minimizar la función de costo, ajustando los parámetros del modelo de forma iterativa, de forma tal que en cada paso se desplace la solución en dirección contraria a la de máxima crecimiento de la función de costo, es decir en la dirección contraria al gradiente. El gradiente es un vector de derivadas parciales, definido por la Ecuación 10. En este punto se introduce un concepto muy importante denominado tasa de aprendizaje (*learning rate*), el cual está representado por la constante α que multiplica a la derivada parcial, toma valores entre 0 y 1, y se puede interpretar como la velocidad a la cual se mueve la solución por la función de costo. También se le conoce como el tamaño del paso. La actualización de los parámetros se expresa mediante la Ecuación 11.

$$\nabla J = \begin{bmatrix} \frac{\partial J}{\partial w_0} \\ \frac{\partial J}{\partial w_1} \\ \dots \\ \frac{\partial J}{\partial w_n} \end{bmatrix} \tag{10}$$

$$w_i = w_i - \alpha\frac{\partial J(w_0, ..., w_n)}{\partial w_i} \tag{11}$$

$\frac{\partial J}{\partial w_i}$: derivada parcial del costo con respecto al parámetro i.
α: tasa de aprendizaje.

La red converge cuando se han estimado los parámetros que minimizan la función de costo. En este punto el modelo está listo para hacer inferencia y se pueden usar datos de entrada que la red nunca ha visto. Una red neuronal puede tener miles o millones de parámetros dependiendo de su complejidad. La configuración de la red neuronal y la estimación de los parámetros, así como el proceso de inferencia, se puede hacer en herramientas de software como TensorFlow, PyTorch, Caffe, entre otras. En el caso de detección de objetos, luego de optimizar los parámetros la red está lista para predecir en nuevas imágenes, estimando los nuevos recuadros que encierran los objetos de interés, y un valor asociado a cada recuadro que refleja la confianza con la que el detector predice la categoría a la que pertenece el objeto.

3.2.2. Métricas para evaluar el desempeño de un detector de objetos

Existen varias métricas para evaluar el desempeño de un detector de objetos (Kelleher et al. 2015). Estas se calculan al comparar los recuadros de la base de datos de referencia, con los recuadros predichos por el modelo usando por ejemplo la intersección de las áreas de los recuadros sobre la unión de las áreas de los mismos (*Intersection Over Union* - IoU). Para su cálculo se usa la fórmula expresada en la Ecuación 12. Entre las métricas más usadas en la literatura de visión por computador se encuentran la Precisión Promedio (*Average Precision* - AP), Sensibilidad (también conocida como tasa de verdaderos positivos o *Recall* en Inglés), precisión, medida F1 (*F1 score*), entre otras. Estas métricas parten de las tipologías de errores y aciertos posibles: Verdadero Positivo (*TP*), Falso Positivo (*FP*), Falso Negativo (*FN*), Verdadero Negativo (*TN*). En el contexto de detección de objetos se aproximan con los mejores emparejamientos entre los objetos conocidos en la imagen y las predicciones. En detección de objetos el TP son todos aquellos recuadros que en verdad están en la imagen y que el detector los ubicó con una IoU mayor a un umbral definido que suele ser 0.5 (Hui 2018a). FP son aquellos recuadros que el detector ubicó con una IoU entre 0 y el umbral definido. FN son aquellos recuadros que en verdad están en la imagen y que el detector no ubicó. TN no se tiene en cuenta en detección de objetos.

$$IoU(B_0, B_1) = \frac{\text{Area of } B_0 \cap \text{Area of } B_1}{\text{Area of } B_0 \cup \text{Area of } B_1} \tag{12}$$

B_0: Recuadro que encierra la verdad absoluta.
B_1: Recuadro que encierra la predicción del modelo.

La Precision, la Sensibilidad, y la medida F1 son métricas importantes para caracterizar los errores de predicción. La precisión se puede interpretar como la tasa de aciertos con respecto al total de predicciones hechas, reflejado en la Ecuación 13. El recall o sensibilidad se puede interpretar como la tasa de aciertos con respecto al total de objetos que en verdad debía encontrar, expresado en la Ecuación 14. La medida F1 o *F1 score* busca ponderar un poco las medidas de precisión y de recall, como ejemplificada en la Ecuación 15. Una forma gráfica de analizar los resultados es construir la gráfica que relaciona precisión vs. sensibilidad (Davis & Goadrich 2006), en la que se encuentra el nivel de precisión correspondiente a cada nivel de sensibilidad, como la que se muestra en la Figura 6, en la cual es deseable que la gráfica sea lo más cercana al valor de $precision = 1$.

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F1 = 2\frac{(Precision * Recall)}{Precision + Recall} \tag{15}$$

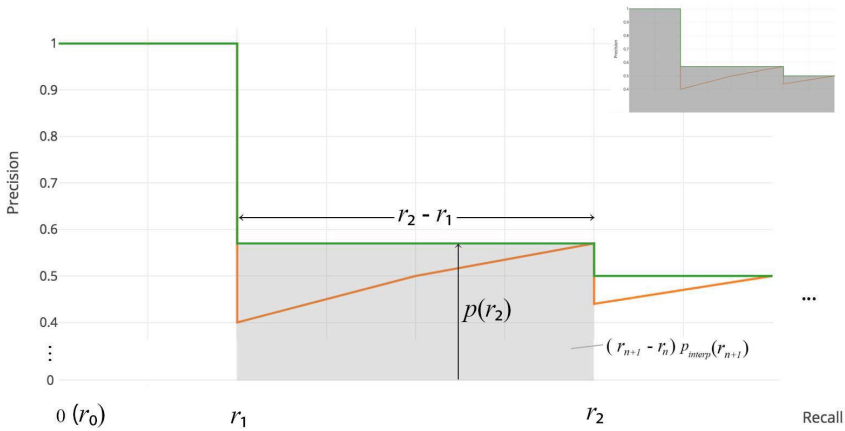

Figura 6. Ejemplo de curva Precisión vs. Sensibilidad, Área bajo la curva (AUC). Tomado de https://medium.com/@jonathan_hui/map-mean-average-precision-for-object-detection-45c121a31173, fecha de acceso: 9 de Marzo de 2020.

En detección de objetos, existe métrica importante usada para comparar la calidad en las detecciones, la cual es la Precisión Promedio (*Average Precision* - AP). El AP se calcula para cada clase, se ordena por el valor de score, se cálcula el recall y la precision para cada detección y el acumulado. Luego se busca por el nivel máximo de precisión a cada valor de recall, como se muestra en la Ecuación 16. Esta medida se suele interpolar con 11 niveles de recall: $r \in \{0, 0.1, 0.2, ..., 1\}$ y se encuentra el máximo valor de precisión para esos niveles de recall, es decir, cual es la precisión más alta alcanzable a cada nivel de recall. Este medida suele ser interpolarse con el objetivo de evitar los grandes quiebres que puede tener esta gráfica. Finalmente, se hace el valor esperado de estos niveles de recall para ajustar el valor de precisión promedio para toda la base de datos, como se representa en el área bajo la curva de la Figura 6.

$$AP = \sum (r_{n+1} - r_n)\rho_{interp}(r_{n+1})$$
$$\rho_{interp}(r_{n+1}) = \max_{\widetilde{r} \geq r_{n+1}} \rho(\widetilde{r}) \tag{16}$$

r_i: i-ésimo valor de recall ordenado.
$\widetilde{r}$: valor de recall interpolado.
ρ: métrica precisión.

La media de la precisión promedio (*mean Average Precisión* - mAP) es la media del AP de cada clase con la que se entrenó el detector de objetos. En muchas competencias entienden el mAP como el AP (Everingham et al. 2010), pero en otras competencias hacen pequeñas variaciones en su cálculo dependiendo del propósito de la competencia. El AP@0.5 es el AP usando como umbral para el IoU el valor de 0.5 para considerar la detección como valida.

3.2.3. Arquitecturas del estado del arte

Existen varias arquitecturas de aprendizaje profundo para la detección de objetos, dentro del estado del arte (Hui 2018b), se destacan: Faster Regions-Convolutional Neuronal Networks (Faster R-CNN), Single Shot Detector (SSD) y You Only Look Once (Yolo) v3. Estas arquitecturas mostraron desempeños sobresalientes en competencias como el Imagenet Large Scale Visual Recognition Challenge 2015 (ILSVRC2015) (ICCV 2015) y Microsoft Common Objects in Context (MS COCO)-2015 (Microsoft 2015), que se realizan sobre bases de datos públicas como COCO (Lin et al. 2014), que tiene recuadros de muchos tipos diferentes de objetos.

El Faster R-CNN es una arquitectura que propone y refina regiones de interés (ROI) durante el proceso de entrenamiento. En esta arquitectura, la red neuronal propone recuadros de interés y luego clasifica cada recuadro propuesto (Ren et al. 2017). Es una de las arquitecturas que tiene los menores tiempo de entrenamiento y predicción en el estado del arte. Sus capas de entrada son las de una red neuronal convolucional, las cuales generan el resultado de la clase a la que pertenece la imagen que se la pasa. Para realizar la detección de objetos, se deben incorporar capas de neuronas que le permitan a la red hacer propuestas acerca de las posibles ubicaciones y tamaños de los objetos en las imágenes. En las capas finales se incluyen, además de las funciones de activación, capas que permiten inferir mediante una regresión, la ubicación y el tamaño de la detección representado en los cuatro puntos del recuadro asociado a cada predicción. Por lo tanto, el detector retorna el recuadro estimado y un valor de activación para la clase a la que se estima pertenece.

La función de costo que busca minimizar el Faster R-CNN está dada por la Ecuación 17 (Ren et al. 2017). Esta función de costo es una combinación lineal de dos funciones de costo mediante un parámetro τ, el primero es el costo de clasificación L_{cls} y el segundo es de ubicación L_{reg}. El costo de clasificación

usualmente se calcula mediante una función de entropía cruzada categórica, como se muestra en la Ecuación 8, en la cual se comparan las categorías predichas y las de las verdades absolutas. El costo en ubicación suele calcularse usando la IoU como en la Ecuación 12. El parámetro τ sirve para darle más peso al proceso de clasificación o al proceso de localización, si es grande hace que los costos asociados a los errores de ubicación sean más altos, y cuando es bajo le da mayor importancia a los errores de clasificación.

$$L\left(\hat{\mathbb{Y}}_i, \mathbb{B}_i\right) = \frac{1}{N_{cls}} \sum_i L_{cls}(\hat{\mathbb{Y}}_i, \mathbb{Y}_i) + \tau \frac{1}{N_{reg}} \sum_i \mathbb{Y}_i L_{reg}(\hat{\mathbb{B}}_i, \mathbb{B}_i) \tag{17}$$

$\hat{\mathbb{Y}}_i, \mathbb{Y}_i$: categoría de interés predicha y verdadera.
$\hat{\mathbb{B}}_i, \mathbb{B}_i$: recuadro de interés predicho y verdadero.
τ: parámetro de balanceo.
L_{cls}, L_{reg}: funciones de costo de clasificación y de ubicación.
N_{cls}, N_{reg}: factores de normalización de las clasificaciones y las ubicaciones.

3.2.4. Generalización del modelo

Para evaluar el desempeño y la generalización del modelo, tradicionalmente se dividen los datos de verdad absoluta en tres grupos de datos: entrenamiento/selección del modelo(validación) /prueba (*o train/model selection/test*) con porcentajes como por ejemplo de 60 %, 20 %, 20 % o 50 %, 25 %, 25 %, respectivamente (Marsland 2014). Los datos de entrenamiento son los que se procesan durante todo el proceso de aprendizaje de la máquina. En cada iteración el algoritmo evalúa y predice sobre las imágenes de entrenamiento, compara con la referencia para esa imagen, y ajusta los pesos de la red en función del error que se obtenga. Los datos de prueba se guardan y no se usan hasta el final de la investigación/experimentos. La base de datos de selección del modelo, también conocida como división de validación sirve para seleccionar los mejores hiperparámetros del modelo y monitorear que no exista sobre entrenamiento. La base de datos de prueba se usa una vez se haya finalizado el proceso de entrenamiento, para evaluar su desempeño en imágenes que el modelo nunca ha visto.

Esta base de datos nos permite saber si el modelo está generalizando bien, o si el modelo está Sobreajustando (*Overfiting*) (Marsland 2014). El sobreajuste, es un problema que pueden presentar los modelos de aprendizaje de máquina, cuando los modelos entrenados únicamente se desempeñan bien sobre el conjunto de entrenamiento. En el caso de las redes neuronales, este fenómeno se atribuye al gran número de parámetros que tienen. En un modelo que generalice bien, los resultados en entrenamiento, selección del modelo, y prueba deberían ser muy parecidos. Con el objetivo de generalizar y de que el modelo no se aprenda de memoria los ejemplos de entrenamiento, las funciones de costo incluyen un factor de regularización que castiga los valores grandes de los parámetros.

Buscando que los modelos de predicción sean más robustos y generalicen mejor, se pueden complementar los datos originales de entrenamiento con datos sintéticos que se obtienen a partir de

los primeros aplicando algunas transformaciones. Este proceso se denomina aumentado de datos (*data Augmentation*) (Shorten & Khoshgoftaar 2019). El aumentado de datos se hace con el objetivo que las redes se entrenen en una base de datos abundante y con suficiente variabilidad que les permita generalizar bien y desempeñar mejor con escenas que nunca han sido vistas en el proceso de entrenamiento. Entre las principales técnicas de aumentado de datos se encuentran rotación y translación, eliminar regiones de forma aleatoria a las imágenes (*crop*), invertir horizontalmente las imágenes (*flip*), eliminar píxeles de forma aleatoria (*dropout*), difuminado (*blur*), entre otros. En el caso de la detección de objetos, no basta con transformar sólo la imagen, sino que es necesario también transformar los recuadros de referencia, ya que si por ejemplo la imagen original se rota, los recuadros asociados a esa imagen se deberían rotar también en la misma proporción.

3.3. Informalidad empresarial

Existen varias definiciones de informalidad, y esta tesis se enfoca en la informalidad empresarial que es diferente por ejemplo a la informalidad laboral, o la economía informal. Para esta tesis se entiende informalidad como toda actividad económica sin registros públicos que contribuye a las cifras oficiales de Producto Interno Bruto (PIB). Esta es una de las definiciones que más se usa en la literatura (Schneider 2016). Así pues, se puede aproximar la informalidad empresarial como toda aquella empresa que no está registrada en cifras oficiales. De la informalidad empresarial es difícil tener cifras confiables y frecuentes, por dos razones principalmente. La primera es que la naturaleza informal de la actividad económica implica que dichos establecimientos comerciales nunca se han registrado en bases de datos oficiales como las de Cámara de Comercio. La segunda, los comercios informales son de naturaleza dinámica e inestable, lo cual los hace difícil de monitorear por medio de métodos convencionales como encuestas o censos.

La forma en que los gobiernos buscan medir y cuantificar la informalidad se basa en el vacío o diferencia (*gap*) que existe entre las cifras de producción nacionales y las cifras de producción reportadas por las empresas formales. Otra forma se basa en encuestas de hogares y empresariales (Fernández 2018), que suelen ser costosas de implementar, no recogen información espacial, no son representativas, pueden tener cambios metodológicos en como capturan la información, no son muy frecuentes, su procesamiento y publicación suele tardar mucho mientras se tabulan los resultados. De modo que no se tiene información del tipo de: cuántas son las empresas que operan informalmente, dónde están ubicadas, si están aglomeradas espacialmente, si comparten alguna característica, si duran mucho o poco en el mercado, etc.

3.4. Análisis espacial

Los Sistemas de Información Geográfica (SIG) nos permiten estudiar los fenómenos espaciales, gracias a sistemas de coordenadas de referencia estándares para todo el mundo, de modo que cada par de coordenadas X, Y son únicas para cada punto en nuestro espacio de trabajo que es la tierra. Usualmente

se utiliza la Latitud y Longitud de un punto, para referirnos a su ubicación en la superficie de la tierra (Olaya 2009). Estas variables se miden en grados, minutos y segundos. La Latitud se mide de forma relativa a la línea del Ecuador, aumenta hacia el polo Norte y disminuye hacia el polo Sur. La Longitud se mide relativa al meridiano de Greenwich en Inglaterra, y aumenta hacia el este y disminuye hacia al oeste. Este es un sistema geodésico que busca dividir la superficie de la tierra usando los grados de una esfera. Uno de los sistemas geodésicos que más se utiliza es el Sistema Geodésico Mundial 84 (WGS84).

3.4.1. Proyección a coordenadas planas

Sin embargo, al usar un sistema geodésico, la curvatura de la tierra dificulta los cálculos de distancias y superficies, pues requiere tener en cuenta la altura de cada punto y hacer uso de la trigonometría esférica (Fotheringham et al. 2010). Es por esto que existen otros sistemas de coordenadas que utilizan proyecciones planas de la tierra, conocidos como Sistemas de Coordenadas Cartesianos, los cuales permiten calcular distancias sin tener variaciones debidas a las diferentes alturas. El sistema Universal Transversal de Mercator (UTM) es uno de estos, y hace una proyección de coordenadas Mercator (proyección cilíndrica conformal), ver Ecuación 18.

$$\begin{aligned} X &= R\phi \\ Y &= R \ln \left[\tan \left(\frac{\pi}{4} + \frac{\theta}{2}\right)\right] \end{aligned} \tag{18}$$

X, Y: coordenadas proyectadas.
R: el radio de la tierra.
ϕ: longitud de la ubicación.
θ: latitud de la ubicación.

Este sistemas de coordenadas de referencia también suele ser relativo a la línea del Ecuador y al meridiano de Greenwich. Este se mide sobre el nivel del mar y sus valores hacen referencia a metros, de modo que las diferencias se pueden interpretar como distancias en metros. Por ejemplo, se puede usar la distancia euclidiana, definida por la Ecuación 19, para calcular la distancia en metros entre los puntos $p1$ con coordenadas (X_1, Y_1) y el punto $p2$ con coordenadas (X_2, Y_2).

$$d_{1,2} = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \tag{19}$$

3.4.2. Geografía cuantitativa

Los fenómenos espaciales se han tratado de modelar de muchas maneras, siempre buscando simplificar su complejidad dimensional (Fotheringham et al. 2010). Un ejemplo de simplificar el espacio es crear

cuadriculas regulares, para todo el espacio de estudio, y calcular los atributos para cada cuadricula, pasando a un problema discreto. Otra forma emplea la teoría de grafos que busca modelar el espacio y simplificarlo (Duque et al. 2007), al hacer que polígonos sean nodos y las distancias sean los enlaces. Los enlaces son los vínculos que conectan los vértices unos con otros. Hay diferentes tipos de grafos dependiendo de cómo se definan sus posibles conexiones. La teoría de grafos permite simplificar el espacio de forma tal que cada entidad espacial (país, ciudad, individuo) es un nodo en la red, y las conexiones, carreteras, vías, o distancias son los vínculos entre ellos.

La literatura académica acerca de Patrones de Puntos, busca a partir de los datos estimar los procesos subyacentes que generaron las distribuciones espaciales de los puntos observados. Esta área de estudio de la Geografía cuantitativa se trabaja en Ecología, Biología, y Epidemiología, entre otros. En cada caso se parte del estudio de nubes de puntos que se representan con los pares de coordenadas X, Y de cada punto.

3.4.3. Densidad de kernel

La densidad de kernel es un proceso en el cual se busca realizar una estimación suavizada de la densidad de probabilidad usando funciones matemáticas o kernels centradas en cada punto y luego promediar estas funciones (Fotheringham et al. 2010). La aplicación espacial de la densidad de Kernel es usar una función de densidad de probabilidad 2-dimensional. Similar a un histograma, a la función de densidad se le debe especificar la amplitud de la banda a analizar (*bandwidth*) y los puntos de interés para los cuales se quiere calcular la densidad promedio. La Ecuación 20 muestra la formulación matemática de la densidad de Kernel espacial. La amplitud de banda en el caso del kernel espacial hace las veces del radio a partir del cual se toman los elementos para calcular la densidad, también controla la amplitud del Kernel. Los resultados de la densidad de Kernel espacial, $\hat{\lambda}_k(\mathrm{x})$, son las estimaciones de densidades de intensidad de kernel por m^2 calculadas para los puntos de interés.

$$\hat{\lambda}_k(\mathrm{x}) = \sum_{i=1}^{n} \frac{1}{nh^2} K\left(\frac{\mathrm{x} - x_i}{h}\right) \tag{20}$$

i, n: índice y conjunto de puntos en la base de datos.
h: amplitud de banda en unidades de distancia.
K: función de Kernel.
x: vector de coordenadas X,Y.
x_i:i-ésimo par de coordenadas.

Por otro lado, existen algoritmos de agrupamiento (*clustering*) espacial que no buscan inferir la densidad de la distribución, sino que buscan agrupamientos en los datos basados en umbrales de distancia y de puntos por grupo. Un ejemplo de estos es el DBSCAN (Daszykowski & Walczak 2009), en el cual el agrupamiento se hace para altas densidades, se asignan los puntos a ciertos grupos a partir de un umbral de distancia y de un número de puntos mínimos para definir cada grupo.

4. Datos

4.1. Área de estudio - Envigado

El municipio de Envigado está ubicado al sur del Valle de Aburrá, como se ilustra en la Figura 1. Este municipio se escogió debido a que se contaba con el apoyo de la Alcaldía y de la Cámara de Comercio Aburrá Sur para la puesta en marcha del piloto de esta tesis en su municipio, y nos suministraron los datos oficiales y las autorizaciones por escrito para hacer nuestro análisis. Adicionalmente, las características del comercio en Envigado son ideales para nuestra aplicación, pues sus habitantes son muy tradicionales a la hora de comprar, recorriendo las diferentes tiendas que están ubicadas por todas sus calles. Con esto en mente, la metodología que proponemos se aplicó en la Zona Centro del Municipio de Envigado, Colombia, la cual cuenta con una densidad de comercio significativa y buena disponibilidad de imágenes de GSV para el año 2017 que es para el año para el cual tenemos la base de datos de registros oficiales. Sobre esta zona de interés se realizó el análisis de informalidad empresarial.

4.2. Comercios en el Valle de Aburrá

Para el entrenamiento de nuestros detectores de objetos usamos imágenes con comercio de todo el Área Metropolitana del Valle de Aburrá del departamento de Antioquia, Colombia. Al utilizar todo el Valle de Aburrá en nuestra base de entrenamiento, incluimos todos los rangos de ingresos posibles de la población y de comercio. El Valle de Aburrá es un grupo de diez municipios, como se muestra en la Figura 7, que quedan en el centro del departamento de Antioquia, en un valle dibujado por el Río Medellín, y que juntos conforman el Área Metropolitana, compartiendo políticas ambientales, de movilidad y de seguridad. El centro del Valle del Aburrá es Medellín, la capital de los antioqueños, es el centro administrativo, laboral, educativo, y el que impulsa el desarrollo de los demás municipios dentro del Valle. El Área Metropolitana del Valle de Aburrá es idóneo para nuestro ejercicio pues es un área con alta actividad comercial, ya que tiene uno de los índices de complejidad económica sectorial más altos de Colombia, la segunda ciudad en 2017 (CID & Bancoldex 2017), lo que garantiza una gran variedad de actividades. Además, tiene una alta heterogeneidad urbana, en el sentido social, económico y geográfico, lo que se manifiesta en una variedad de estructuras urbanas, fachadas y tipos de construcción, indispensable para entrenar un buen detector de objetos. También, la región está bien cubierta por las imágenes de Google Street View.

Los comercios que servirán de referencia pertenecerán al Valle de Aburrá, por la alta variedad de comercios, pues van desde los comercios de bajo ingreso, como el de barrio que vende abarrotes de primera necesidad, hasta los comercios más lujosos, como el de los concesionarios que ofrecen coches importados. Usar sólo Medellín o sólo Envigado sesgaría la muestra, pues el comercio en su interior es del tipo de ciudad desarrollada y no contiene casi comercios de bajo ingreso, que si tienen municipios aledaños como Caldas y Girardota. Los comercios del Valle de Aburrá que se usaron paran entrenar los detectores se describen en la Tabla 2. Se usaron 1500 imágenes panorámicas inicialmente. A estas 1500 panorámicas se les hizo la proyección de fachadas laterales, produciendo 3000 imágenes, de las

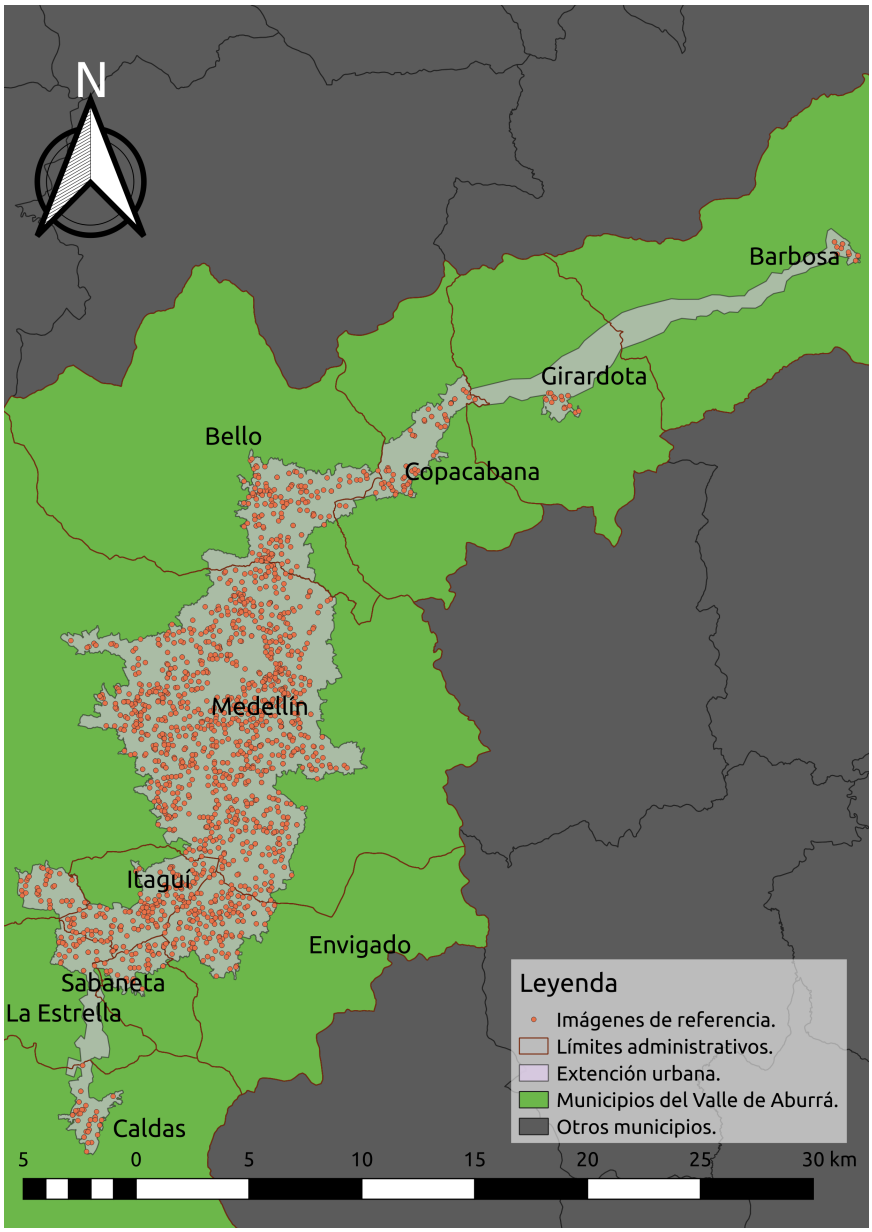

Figura 7. Mapa con los lugares donde se hizo el muestreo espacial y el proceso de etiquetado (comercio vs. no-comercio) para esta tesis.

cuales se usaron 2236 que tenían información relevante. En total se usan 3914 recuadros de fachadas no comerciales y 2249 recuadros de fachadas comerciales para entrenar los modelos de detección de objetos.

Tabla 2. Resumen descriptivo de los datos usados para entrenar los detectores de comercio.

Estadística	Valor
Número de panorámicas etiquetadas	1500
Número de imágenes de fachada lateral etiquetadas	2236
Número máximo de comercios etiquetados por imagen	16
Comercio promedio etiquetado por imagen	1.01
Número total de etiquetas de comercio	2249
Número total de etiquetas de no comercio	3914

4.3. Open Street Maps

En este trabajo se utilizó la base de datos libre de OpenStreetMaps (OSM) (OpenStreetMap contributors 2017) que contiene cerca del 80 % del total de la malla vial mundial (BarringtonLeigh & Millard-Ball 2017). Usando un polígono de extensión urbana, por ejemplo del centro de Envigado, se extrae la malla vial de OSM para este polígono, devuelve las aristas y los nodos dentro del polígono. Posteriormente, se hace un proceso de filtrado de vías en las que se retienen únicamente las vías sobre la que se ubica el comercio, es decir aquellas a las que se puede acceder públicamente. Esta malla vial depurada y lista para usar, se usa como insumo en toda la metodología, iniciando con un proceso de muestreo espacial que permita hacer un barrido en todo el territorio bajo estudio.

4.4. Google Street View

Uno de los principales proveedores de imágenes con vista desde la calle, es Street View de Google (Anguelov et al. 2010) que permite acceder a las imágenes asociadas a un territorio. Google ofrece imágenes 360° utilizando un formato estándar, con los mismos tamaños, bajo las mismas condiciones y con cámaras de 360° de características muy similares. Adicionalmente, las imágenes de Google Street View están altamente disponibles para muchos territorios, Google y los usuarios actualizan cada año las imágenes en los diferentes territorios. Debe señalarse que Google no es el único proveedor de imágenes con vista desde la calle, existen otros como Mapillary, NavVis, Microsoft Live Earth (Laupheimer et al. 2018). Las imágenes panorámicas vienen en formato equirectangular de 13312×6656 píxeles. En esta tesis se descargaron las imágenes usando la interfaz de programación de aplicaciones (API) para Python de los desarrolladores de Google para Street View. Esta descarga tiene un costo aproximado de siete dólares americanos por cada mil imágenes (Google 2019).

4.5. Cámara de Comercio

Para la elaboración de esta tesis se contó con la base de datos de registros oficiales de la Cámara de Comercio de Aburrá Sur para Envigado y de una carta autorización de la misma entidad para su uso en esta investigación. Esta base de datos de comercios registrados oficialmente, está debidamente anonimizada, y con las direcciones geo-codificadas, usando el software geocoder del grupo de investigación RiSE (RiSE-group 2015). En esta base de datos se encuentran las empresas que contaban con su registro mercantil para el año 2017 en un formato tabular junto con sus coordenadas geográficas, nombre del establecimiento, y actividad económica en la que se desempeñaba. La Figura 8 muestra una estimación de la distribución espacial de los comercios registrados en Cámara de Comercio para Envigado en 2017, y el polígono perteneciente al centro de Envigado, dónde se evidencia la alta concentración del comercio en el centro. Estos datos corresponden a los registros oficiales de comercio que usamos como referencia para comparar los resultados de nuestra metodología, y posteriormente, intentar identificar patrones del comercio sin registrar oficialmente.

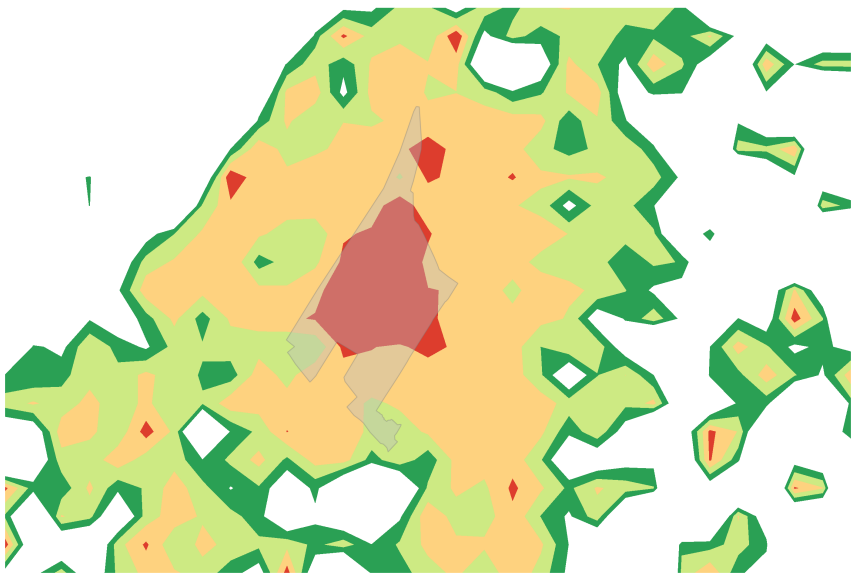

Figura 8. Distribución espacial continua de los registros de Cámara de Comercio Aburrá Sur para Envigado y el polígono del barrio Centro de este municipio.

5. Metodología

A continuación se describe la metodología matemática que se utilizó para explorar automáticamente la actividad comercial en una ciudad o región. La metodología tiene tres grandes etapas que se ilustran en la Figura 2, más una cuarta etapa de aplicación. La primera etapa corresponde al proceso de adquisición,

pre-procesamiento, y etiquetado manual de las imágenes. La segunda etapa agrupa el proceso de entrenamiento de los detectores de comercio, análisis de desempeño y comparación del mejor detector de comercio visible urbano. La tercera etapa de análisis espacial aborda la forma como se entiende el espacio en esta tesis, la forma en la que se realizan muestreos exhaustivos automatizados, y cómo se estima la distribución espacial del comercio visible. Finalmente, se explica como se aplicó la metodología de esta tesis en el centro del municipio de Envigado, Colombia.

5.1. Datos de referencia

5.1.1. Espacio de fachada lateral

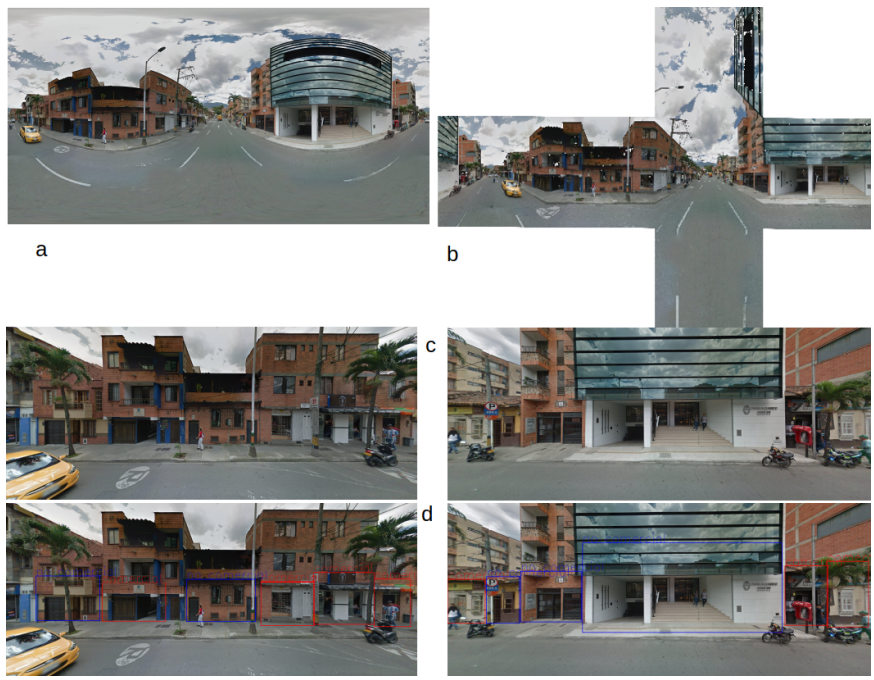

Figura 9. Procedimiento para extraer imágenes de fachadas laterales. a) Imagen 360°, b) Cubo generado intermedio sin deformaciones curvas, c) Imágenes de fachada lateral proyectadas del cubo, d) Imágenes de fachada lateral etiquetadas manualmente.

En esta tesis se trabaja sobre imágenes de fachada lateral que contienen información que se captura desde la calle como si hubiera sido tomada por una cámara convencional. La Figura 9 ilustra la proyección de un espacio 360° al de fachada lateral. La ventaja de utilizar imágenes convencionales y no imágenes panorámicas es que éstas se pueden adquirir con cámaras de bajo costo como cámaras fotográficas, cámaras de celular, cámaras embebidas en plataformas móviles (sean estos vehículos o drones) e incluso

las mismas cámaras 360. Las imágenes de fachada lateral contienen el comercio al interior de las ciudades, que es el objeto de estudio de esta tesis.

Para obtener las imágenes de fachada, el primer proceso consiste en aplicar una transformación geométrica de un espacio esférico a un espacio cúbico, como la que se explica en la Sección 3.1.1. Del cubo proyectado se utilizan las caras que contienen las fachadas, es decir las caras izquierda y derecha, paralelas a la malla vial. Las caras superior e inferior no se usan, pues son las que apuntan al cielo y al piso. De la imagen frontal y trasera, sólo nos interesan las partes que contienen información de fachadas, pero estas incorporan efectos del punto de desvanecimiento (Zhou et al. 2017), producto de los puntos de fuga de la escena fotografiada.

Para corregir el efecto del punto de desvanecimiento aplicamos unas transformaciones de perspectiva (ver Sección 3.1.2), para producir imágenes planas. La máscara que se usa para esta transformación se calcula con los puntos de un trapecio. La imagen de una cara del cubo tiene dimensiones $c \times c$, donde $c = W/4$. La Figura 10 representa la corrección aplicada a cada imagen frontal y trasera, así como los puntos de referencia X, Y que definen la mascara que se usa para la transformación de perspectiva. Así, tenemos las nuevas imágenes planas, que junto con las caras izquierda y derecha, permiten ensamblar las imágenes de fachadas laterales. La Figura 9, ilustra este proceso.

5.1.2. Etiquetado manual

Para el entrenamiento de un modelo de detección de objetos supervisado que utilice CNN es necesario tener una base de datos de ejemplos etiquetados de referencia o verdad absoluta (*ground truth*), como se explica en la Sección 3.2, que le permita al modelo estimar los pesos de la red neuronal necesarios para identificar las fachadas que son de comercio de las que no lo son. Es por esto que, se muestrearon aleatoriamente 10.000 coordenadas geográficas que pertenecían al área urbana del Valle de Aburrá, Colombia, para solicitar las imágenes de Google Street View correspondientes, con lo cual se descargaron las imagenes representadas en los puntos de la Figura 7. Se conservaron las imágenes que contenían información relevante, descartando aquellas que tenían obstáculos en el medio o que tenían baja calidad. Luego se proyectó cada imagen panorámica a dos imágenes de fachada lateral, usando las caras izquierda y derecha, junto con las partes que contenían fachadas de las caras frontal y trasera sin efecto de punto de desvanecimiento. A estas imágenes de fachada lateral se les asoció un identificador único en una base de datos que también contenía la coordenada y la fecha de adquisición, y se renombraron las imágenes usando su identificador único.

Posteriormente se construyó manualmente la base de datos de etiquetas de locales comerciales, dibujando cada uno de los límites rectangulares de las fachadas en las imágenes, indicando en cada caso si la fachada es o no comercial. Se etiquetaron imágenes hasta obtener 6700 fachadas no comerciales, y 3300 fachadas comerciales, generando así la base de datos de referencia o de verdad absoluta. Estas cotas heurísticas buscan obtener un modelo robusto (Devroye et al. 1996), y para cada ejemplo de lo que queremos detectar (comercios), tendremos dos ejemplos de lo que no queremos detectar (no comercios).

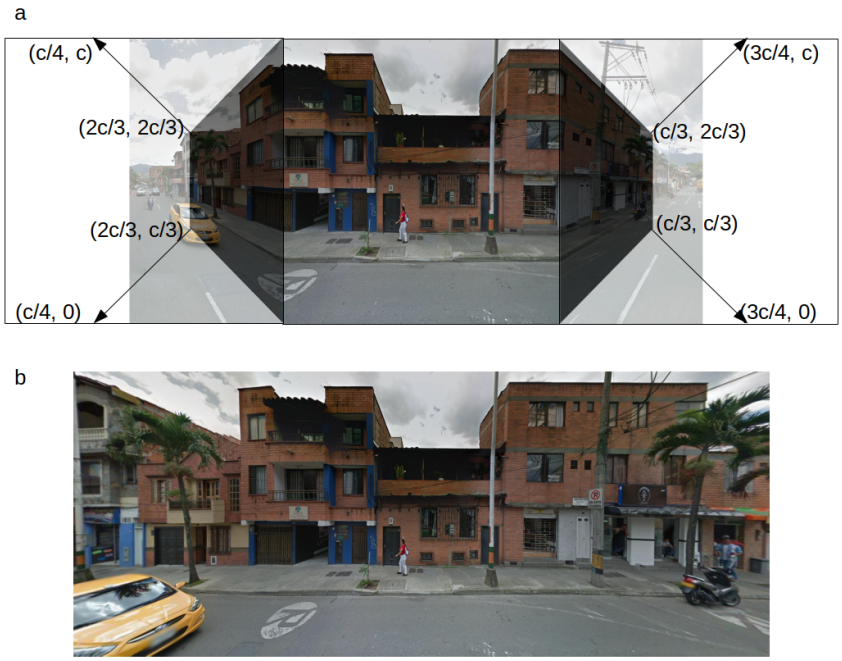

Figura 10. Procedimiento para corregir el efecto de punto de desvanecimiento. a) Imagen con efecto de punto de desvanecimiento y mascaras calculadas (trapecios negros) para corregirlo, b) Imagen sin efecto de punto de desvanecimiento.

5.1.3. División de la base de datos

Para evaluar que el modelo generalice bien lo haremos dividiendo la base de datos etiquetada en particiones independientes, de modo que no se usen los datos de prueba en el entrenamiento del detector, de la forma en que se explica en la Sección 3.2.4. Se dividirá la base de datos de referencia de los comercios en el Valle de Aburrá en tres bases de datos dejando 60 % para Entrenamiento (*Train*), 20 % para Selección del Modelo (*Model Selection*) (también conocido como Validación - *Validation*) y 20 % para Prueba (*Test*), la selección de las imágenes en cada conjunto se hace de forma aleatoria. La división se hace a partir de las fachadas laterales y no de los comercios individuales buscando que todas las etiquetas de una imagen entren juntas a la misma división.

5.1.4. Aumentado de datos

La base de datos de entrenamiento, que corresponde al 60 % de la base de datos de referencia, es la que se utilizará para hacer los entrenamientos de los detectores. Para esto es necesario un proceso de generación de datos sintéticos que le permitan al detector generalizar mejor ante diferentes escenas, tal

como se comenta en la Sección 3.2.4. El proceso de generación de imágenes sintéticas se conoce como aumentado de datos y está ampliamente documentado en la literatura (Shorten & Khoshgoftaar 2019). Los algoritmos que contemplamos hacer son rotación y translación, eliminar regiones de forma aleatoria a las imágenes, invertir horizontalmente las imágenes, eliminar píxeles de forma aleatoria, difuminado, entre otros. La Figura 11 ilustra los resultados de aplicar algunos de los algoritmos de aumentado que se usaron en esta tesis.

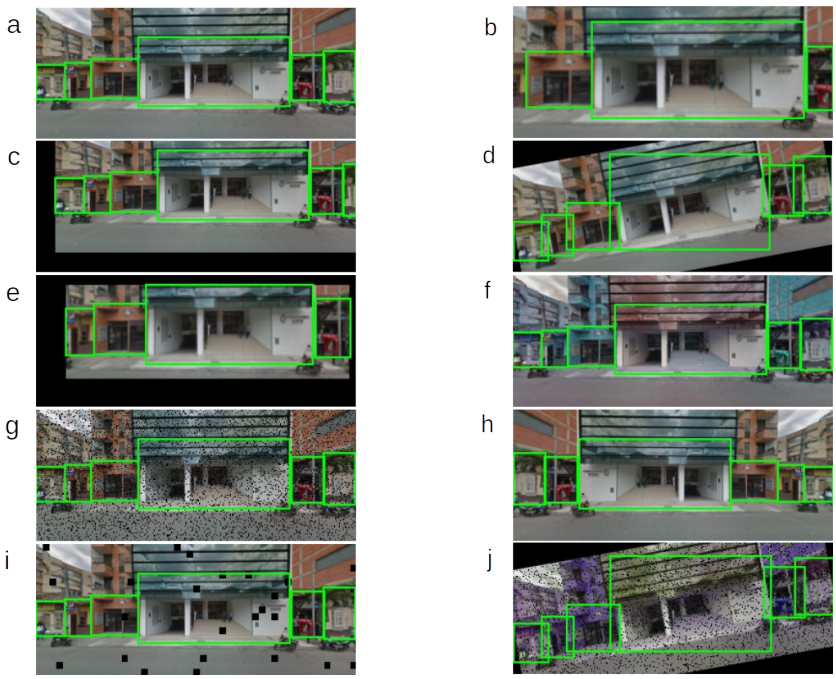

Figura 11. Algunos ejemplos del proceso de aumentado de datos. a) Original, b) Acercarse, c) Translación, d) Rotación, e) Eliminar regiones de forma aleatoria, f) Proyección en espacios de color, g) Eliminar píxeles aleatoriamente, h) Invertir horizontalmente, i) Eliminar recuadros aleatoriamente, j) Ejemplo de combinaciones.

5.2. Detector de comercio

El siguiente paso en la metodología, es proceder con el entrenamiento de modelos de detección de objetos en imagen con vista de calle usando la base de datos aumentada sintéticamente. Nuestro objetivo es encontrar el mayor número de coincidencias entre las predicciones de comercio y las regiones etiquetadas manualmente como tal en las imágenes con vista de calle de estudio. Un detector de objetos que seleccionamos se basa en redes neuronales convolucionales y está en capacidad de encontrar cada fachada en la imagen con vista de calle y posteriormente clasificarla como comercial o no comercial. Este proceso se explicó con más detalle en la Sección 3.2.

Para esta etapa contemplamos el entrenamiento de varias arquitecturas del estado del arte, como lo son Faster R-CNN, Single Shot Detector (SSD), y Yolo v3. Estos tres detectores de objetos se entrenaron usando los parámetros por defecto, y se eligió el que mejor desempeño mostró, luego de hacer el conteo de comercio. La arquitectura Faster R-CNN cuenta con las mejores características de desempeño/tiempo entre los demás detectores del estado del arte, como se muestra en la Sección 3.2.3. Aun así, entrenamos también el SSD y el Yolo v3 para comparar su desempeño contando comercios en una imagen. En todos los casos, el ajuste de hiperparámetros se hizo mediante búsquedas en grilla (*grid search*).

Luego de escoger el mejor modelo entre las tres alternativas usando el conjunto de datos de selección del modelo (validación), se evalúa su comportamiento en el conjunto de datos de prueba, y se extraen las métricas de desempeño. Finalmente, las métricas de desempeño obtenidas durante entrenamiento, selección del modelo y pruebas se comparan entre sí para evaluar si el modelo seleccionado generaliza bien.

5.2.1. Métricas de conteo de comercios

Debido al objetivo de esta tesis, es más relevante que el detector de objetos indique el número correcto de establecimientos comerciales, más que su ubicación exacta en la imagen. Con esta premisa se propusieron dos métricas para evaluar el comportamiento del detector. La primera métrica se representa en la Ecuación 21 y busca calcular el valor esperado de la participación del comercio predicho sobre el comercio real, en imágenes en donde en realidad si hay comercio. Adicionalmente calculamos otra métrica, expresada en la Ecuación 22 que refleja los errores en el conteo de comercios, calculados como el valor esperado de los comercios predichos en dónde en verdad no hay comercios. Para un detector de comercio ideal se esperaría que $p_{comer} = 1$ y $err_{comer} = 0$.

$$p_{comer}(\hat{c}_i, c_i) = E(\frac{\hat{c}_i}{c_i}) = \frac{1}{n_r} \sum_{i}^{n_r} \frac{\hat{c}_i}{c_i}, \quad c_i > 0 \tag{21}$$

$$err_{comer}(\hat{c}_i, c_i) = E(\hat{c}_i | c_i = 0) = \frac{1}{n_{nr}} \sum_{i}^{n_{nr}} \hat{c}_i, \quad c_i = 0 \tag{22}$$

c_i: número de establecimientos comerciales en la imagen i-esima.
$\hat{c}_i$: número de predicciones de establecimientos comerciales en la imagen i-ésima.
n_r: número de imágenes en la base de datos que en realidad tienen comercio.
n_{nr}: número de imágenes en la base de datos que en realidad no tienen comercio.

5.3. Análisis espacial

Para la dimensión espacial del problema proponemos una metodología matemática que a partir de la extensión geográfica de la zona de interés, construye cada uno de los puntos de referencia (y sus coordenadas), necesarios para hacer un barrido exhaustivo automatizado de la zona de interés. Esta estrategia construye los puntos de referencia a partir de los datos de malla vial disponibles en OSM, consultados con la extensión geográfica de la zona de interés, proceso que puede replicarse fácilmente en otros lugares. Esta metodología matemática, utiliza elementos de la teoría de grafos para modelar la malla vial disponible, crear nuevas capas de datos espaciales a partir de su análisis y producir una base de datos de puntos para la zona de interés, que llamaremos puntos geográficos de referencia. Esta etapa se hizo de forma paralela al proceso de entrenamiento de los modelos de detección de objetos.

Adicionalmente, se incorpora una capa de postprocesamiento de los puntos devueltos por GSV, pues GSV no retorna la imagen en la coordenada exacta que se solicita, como se explica en la Sección 5.3.2. También, en la Sección 5.3.3, se propone una agregación espacial del comercio predicho, correspondiente a la estimación de la distribución espacial continua del comercio visible. Finalmente, en la Sección 5.3.4, se propone una metodología para comparar dos distribuciones espaciales continuas.

5.3.1. Puntos geográficos de referencia

Tomando prestados elementos de la teoría de grafos, similar a Duque et al. (2011), modelamos la red vial cómo un grafo interconectado, en donde los nodos representan las esquinas o cruces de vías y las aristas representan las calles o carreras. El problema a solucionar es minimizar el número de puntos muestreados al interior de la malla vial sujeto a que con el total de imágenes de Street View muestreadas se cubran todas las fachadas de la zona de interés, sujeto también a la disponibilidad de datos y a que no se repitan fachadas. Para hacer nuestra exploración de calle extensiva (o exhaustiva) de la zona de interés, proponemos un muestreo de puntos que use la información de los nodos (N) y la información de las aristas (L) cubriendo toda la malla vial de la zona de estudio, de la forma:

$$P = L + N \tag{23}$$

P: número total de puntos muestreados.
L: número de puntos muestreados de las aristas.
N: número de puntos muestreados de los nodos.

Proponemos un método de división de las aristas que permita tener puntos cada distancia d para maximizar el uso de la información disponible en la imagen. Sean $A_1, A_2, ..., A_n$, aristas en una red vial M, con longitudes medidas en metros $l_1, l_2, ..., l_n$ respectivamente, y $N_1, N_2, ..., N_m$ los nodos de esta red. Proponemos usar el número de puntos por arista i dado por:

$$p_i = \frac{l_i}{d} \tag{24}$$

$$p_i^* = floor\left(\frac{l_i}{d}\right) = \left\lfloor \frac{l_i}{d} \right\rfloor \tag{25}$$

d: es la distancia euclidiana medida en metros.
p_i: el número de puntos en una arista i con sus decimales.
p_i^*: es su versión entera.

Usaremos una función de aproximación al entero inferior (*floor*) ya que permite separar los valores enteros de sus decimales sobrantes. Usando p_i^* cada arista es divida según su longitud, de modo que se conserve una distancia constante d entre cada punto. Usaremos como medida de distancia, la distancia euclidiana entre dos puntos i y j medida en metros, de modo que se debe trabajar en un sistema de coordenadas cartesianas como el que se explica en la Sección 3.4.1. Este d dependerá de la distancia que abarque la imagen, en la que se pueda diferenciar una fachada de otra, por ejemplo para GSV es aproximadamente 20 metros. Definimos el sobrante de una arista i como s_i como en la Ecuación 26.

$$s_i = p_i - p_i^* = \frac{l_i}{d} - floor\left(\frac{l_i}{d}\right) = \frac{l_i}{d} - \left\lfloor \frac{l_i}{d} \right\rfloor, \quad 0 \leq s_i < 1 \tag{26}$$

i, n: índice y conjunto de aristas, $i = \{1, 2, ..., n\}$.
j, m: índice y conjunto de nodos, $j = \{1, 2, ..., m\}$.

Los sobrantes los repartiremos la mitad al inicio y la mitad al final de la arista, de modo que los puntos inicien a la misma distancia del nodo. Usando todos los nodos y los puntos provenientes de aristas, obtenemos que el total de puntos para la zona de estudio, reemplazando en la Ecuación 23, estará dado por la Ecuación 27. La ubicación de los puntos producidos a partir de las aristas, estarán dados por las Ecuaciones 28 y 29.

$$P = \sum_{i=1}^{n} \left\lfloor \frac{l_i}{d} \right\rfloor + \sum_{j=1}^{m} N_j \tag{27}$$

$$X_{ki} = X_{0i} + \frac{k - (\frac{1-s_i}{2})}{p_i}(X_{1i} - X_{0i}) \tag{28}$$

$$Y_{ki} = Y_{0i} + \frac{k - (\frac{1-s_i}{2})}{p_i}(Y_{1i} - Y_{0i}) \tag{29}$$

k: el punto k-ésimo generado a partir de la arista i, $k = \{1, 2, ...p_i^*\}$.
i, n: índice y conjunto de aristas, $i = \{1, 2, ..., n\}$.
p_i: el número de puntos en una arista i con sus decimales.
X_{0i}, Y_{0i}: coordenadas del punto donde inicia la arista i.
X_{1i}, Y_{1i}: coordenadas del punto donde finaliza la arista i.

Para los puntos de los nodos no se necesitan ecuaciones de X y Y, ya que se conocen de antemano. Un ejemplo de estos puntos se muestra en la Figura 12. Estas ecuaciones 28 y 29, junto con los puntos de los nodos, permiten calcular cada una de las coordenadas necesarias para dividir la zona de interés, en puntos que estén a una distancia cercana a d entre ellos. Así se construye la base de datos de puntos geográficos de referencia que corresponde a la unidad de análisis básica de esta tesis, del barrido exhaustivo automatizado, y para la cual se construye toda la información de interés.

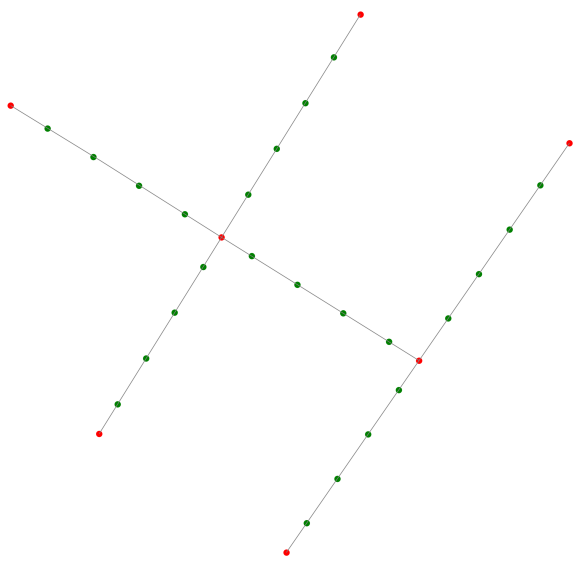

Figura 12. Ejemplo de la metodología de puntos de referencia. Los puntos verdes son puntos de arista y los puntos rojos son los nodos.

5.3.2. Barrido espacial exhaustivo

Los puntos de referencia son los lugares para los que se debe disponer imágenes con vista de calle apuntando a las fachadas de ambos costados de la carretera. Estas imágenes podrán ser o una imagen panorámica de 360°, o dos imágenes planas, una para cada fachada lateral. Obteniendo estas imágenes para cada punto, podemos hacer un barrido espacial automatizado buscando detectar los objetos de interés, y luego guardar los resultados como atributos para los puntos geográficos de referencia. Las imágenes recolectadas para todos estos puntos, equivalen a tener una vista de calle de toda la zona de interés, pues hay una imagen con vista a las fachadas para cada posible punto de calle en la zona geográfica de interés. Los atributos calculados, si se ubican espacialmente en un mapa, dan una primera aproximación a una distribución espacial de los objetos de interés.

Debido a que las imágenes disponibles de Google no se encuentran ubicadas necesariamente sobre los puntos geográficos de interés, es necesario realizar una etapa de procesamiento matemático para ajustar este desfase espacial. Debe indicarse que GSV retorna la información del punto más cercano disponible al que se solicita. Como es posible que algunos de los metadatos de las imágenes que devuelva GSV estén a una menor distancia que la distancia de muestreo d, las imágenes resultantes no se pueden usar directamente para el conteo de objetos de interés en ellas, ya que se podría procesar y contar más de una vez los objetos de interés en estas imágenes.

Es por esto que se incluyó un post-procesamiento espacial que toma los metadatos resultantes de las consultas a GSV, e identifica cuáles de estos están a una distancia menor a d, selecciona el de la mejor ubicación y borra los demás que puedan tener duplicados. Este procedimiento lo formulamos usando el algoritmo de agrupamiento espacial DBSCAN (Daszykowski & Walczak 2009), que se explica en la Sección 3.4.3. Con este algoritmo, creamos grupos que estén a una distancia de d metros, para cada grupo seleccionamos el punto que está mas lejos de los puntos alrededor del grupo, y borramos los demás puntos del grupo. Esto genera una nube de puntos en la cual los puntos se encuentran mínimo a una distancia euclidiana de d metros. Con esta nube de puntos podemos estar tranquilos de que no van a haber objetos o imágenes con vista de calle repetidas.

5.3.3. Descubriendo el comercio

El objetivo de esta tesis, más que ubicar e identificar cada comercio, es descubrir dónde se encuentra localizada la actividad comercial al interior de una zona urbana de interés, que nos dé cuenta de su patrón espacial. Para descubrir la distribución espacial de la actividad comercial, se usó la metodología matemática propuesta en las Secciones 5.3.1 y 5.3.2. Con la cual se construyeron los puntos de interés para toda la extensión geográfica, que además están distanciados lo suficiente entre ellos de modo que no se repitan los establecimientos comerciales visibles. Adicionalmente se usó el detector de comercios que mejor se desempeñó en las métricas de conteo de comercios (propuestas en la Sección 5.2.1) para hacer inferencia y detectar los comercios en cada una de las imágenes. Esto para contar la cantidad de establecimientos comerciales en cada imagen con vista de calle. Con esto obtendremos una distribución espacial discreta de los comercios visibles al interior de la zona de interés.

Usamos una función de densidad de kernel espacial para agregar y suavizar los resultados como se explica en la Sección 3.4.3. Esta función al suavizar los datos, permite obtener una estimación de la distribución espacial continua de los comercios visibles. La estimación la hacemos con los valores de la distribución espacial discreta del comercio visible, con una amplitud de banda de 20 metros y con unos puntos de interés muestreados con una grilla de 5x5 metros para la zona geográfica de estudio. Diferentes valores de la amplitud de banda, capturarán patrones a diferentes escalas geográficas. Al usar una amplitud de banda de 20 metros y una grilla de 5x5 metros se obtiene una estimación que captura un patrón espacial local, permitiendo resaltar los resultados a nivel de una calle o de una manzana. Todo esto usando un sistema de coordenadas UTM, para poder usar los valores en metros y no tener que preocuparnos por las curvaturas de la tierra, como se mostró en la Sección 3.4.1. También aglomeramos los datos para estudiar los patrones espaciales y evitar infringir posibles políticas de uso de datos personales. Se habla de descubrir el comercio porque en ambientes altamente informales como el estudiado, es muy difícil que los gobiernos tengan un mapa con su distribución espacial. Con esta metodología brindamos la posibilidad de tener una estimación de dichos mapas.

5.3.4. Diferencia espacial

Una ventaja de tener una estimación continua de la distribución espacial del comercio, es que se puede comparar fácilmente con otra distribución espacial si se manejan las mismas escalas. Un ejercicio interesante de hacer con los datos producidos por nuestra metodología, es el de compararlos con otra distribución espacial de comercios, por ejemplo, los comercios registrados. Como vimos en la literatura de informalidad empresarial, la diferencia entre los comercios reales y los comercios registrado nos permite tener una medición aproximada del nivel de informalidad empresarial. Por ello, al substraer de la distribución espacial de los comercios reales, la distribución espacial de los comercios registrados oficialmente, podemos tener una estimación aproximada de la distribución espacial de los comercios informales al interior de la zona de interés. La Ecuación 30 refleja la diferencia entre las distribuciones.

$$h(x,y) = f(x,y) - g(x,y), \quad f(x,y), g(x,y) > 0 \tag{30}$$

$h(x,y), f(x,y), g(x,y)$: son funciones de distribución espacial.

Como se expuso en la Sección 3.4.3, los resultados de la densidad kernel espacial de dos distribuciones de datos espaciales no se pueden comparar directamente, pues vienen en valores de densidades relativas al número de datos de cada distribución. Por esto ajustamos las estimaciones, multiplicando por constantes positivas toda la distribución espacial, lo que no altera su variabilidad, solo su escala. En la Ecuación 20, $\hat{\lambda}_k(\mathrm{x})$ es la estimación de las densidades de intensidad de kernel por m^2 calculadas para cada punto de interés, entonces $n\hat{\lambda}_k(\mathrm{x})$ en la Ecuación 31 (de multiplicar n a ambos lados del igual en la Ecuación 20) es la estimación de las intensidades de kernel promedio por m^2 (Fotheringham et al. 2010). Esta estimación ya no está en densidades relativas al número de datos de la distribución, pero sigue siendo relativa al área que abarca el kernel, es decir son intensidades promedio por m^2 (al tener el h^2 dividiendo). Entonces

al multiplicar h^2 a ambos lados del igual en la misma Ecuación 31, obtenemos $nh^2\hat{\lambda}_k(\mathrm{x})$, el cual es la estimación de las intensidades de kernel para los puntos de interés. Al usar esta estimación de intensidades podemos substraer dos distribuciones espaciales con la seguridad de que están en las mismas unidades, intensidad de kernel.

$$\begin{aligned} \hat{\lambda}_k(\mathrm{x}) &= \sum_{i=1}^{n} \frac{1}{nh^2} K\left(\frac{\mathrm{x}-x_i}{h}\right) \Rightarrow \\ n\hat{\lambda}_k(\mathrm{x}) &= \sum_{i=1}^{n} \frac{1}{h^2} K\left(\frac{\mathrm{x}-x_i}{h}\right) \Rightarrow \\ nh^2\hat{\lambda}_k(\mathrm{x}) &= \sum_{i=1}^{n} K\left(\frac{\mathrm{x}-x_i}{h}\right) \end{aligned} \tag{31}$$

Proponemos analizar el contraste entre el comercio predicho y el comercio registrado con una diferencia de las distribuciones espaciales en intensidades de kernel. Al utilizar el $nh^2\hat{\lambda}_k(\mathrm{x})$ de la Ecuación 31 como estimador de las distribuciones espaciales continuas de $f(x,y)$ y $g(x,y)$ de la Ecuación 30, podemos estimar las diferencias de intensidades de las distribuciones del comercio predicho o real y el de registros oficiales con la resta en la Ecuación 32. Espacialmente, la diferencia entre las dos distribuciones refleja nuestra aproximación a la distribución del comercio informal en el territorio, siguiendo las definiciones de la Sección 3.3. Cuando es cero, existe un acuerdo entre los registros oficiales y las predicciones. Las diferencias positivas pueden evidenciar la presencia de comercio que no existe en cifras oficiales, es decir comercio informal. Mientras que las diferencias negativas se pueden interpretar como comercios que están en cifras oficiales pero que no fueron detectados, es decir, que nuestra metodología no logró detectar por diferente razones.

$$\hat{h}(x,y) = \hat{f}(x,y) - \hat{g}(x,y) = nh^2\hat{\lambda}_R(\mathrm{x}) - mh^2\hat{\lambda}_O(\mathrm{y}) \tag{32}$$

$\hat{\lambda}_R(\mathrm{x})$: densidad de kernel estimada por m^2 para el comercio real.
$\hat{\lambda}_O(\mathrm{y})$: densidad de kernel estimada por m^2 para el comercio oficial.
n: número de comercios reales en la zona de estudio.
m: número de comercios oficiales en la zona de estudio.

5.4. Caso de estudio - análisis de informalidad para el centro de Envigado

El último proceso en esta tesis, consistió en aplicar nuestra metodología para hacer un análisis espacial de la informalidad empresarial en el municipio de Envigado. El objetivo era descubrir el comercio para una zona para la cual también se tenía la distribución espacial del comercio oficial registrado en datos

del gobierno. Como se explicó en la Sección 3.3, una definición usada ampliamente de informalidad empresarial es la de toda empresa que produce en el territorio pero que no esta registrada en las bases de datos del gobierno. Por lo tanto, si tenemos el conjunto de todos los comercios visibles en un territorio, y los comercios oficiales visibles registrados en cifras públicas, su diferencia espacial nos da una estimación del comercio informal visible, como se propone en la Sección 5.3.4.

La metodología se aplicó en Envigado porque tenemos el apoyo de la administración municipal materializado en la base de datos de Cámara de Comercio, y unas cartas de aprobación de uso de los datos de Envigado para nuestro piloto. Cómo se describe en la Sección 4.5, la base de datos de comercios oficiales de las que disponemos es la del registro mercantil para Envigado de 2017. Es por esto que necesitábamos imágenes del mismo año para hacer sobre ellas el descubrimiento de la actividad comercial visible. Después de consultar en GSV, encontramos que los datos disponibles de imágenes de GSV para Envigado en 2017 eran escasos para todo el municipio a excepción del centro de Envigado, es por esto que este análisis de informalidad empresarial se enfocó en la zona centro de Envigado.

El comercio visto desde la calle excluye aquel que se encuentra en lugares como centros comerciales, pasajes comerciales, o en general locales al interior de edificios. Es por esto que es necesario hacer ciertos filtros a los comercios en la base de datos oficial, de modo que sólo queden los que son visibles desde la calle. Los filtros son hechos sobre los datos de direcciones del registro mercantil, eliminando de la base aquellos comercios que tengan más de dos puntos para la misma coordenada X,Y. Esto porque el geocodificador que se utilizó (ver Sección 4.5), asigna la misma coordenada a las direcciones que terminen en Interior (IN) X, Apartamento (APTO) X, Oficina (OF) X, o Local (LC) X, a la dirección cómo si no tu tuviera estas terminaciones. Dejamos dos porque es el máximo de posibles comercios que puedan compartir la misma dirección y ser vistos desde la calle.

6. Resultados

El primer objetivo específico de esta tesis era etiquetar manualmente una base de datos de comercios y no comercios que pueda servir para entrenar un detector de objetos que ubique con recuadros la presencia del comercio en una imagen con vista de calle, como se indicó en la Sección 5.1.2. En la Tabla 3 se muestra un resumen estadístico del total de imágenes etiquetadas. De los 10.000 puntos que se muestrearon aleatoriamente, se descargaron efectivamente 4357 imágenes panorámicas, de las cuales 1941 se usaron para dibujar las etiquetas en ellas. A estas 1941 se les hizo la proyección de fachadas laterales, produciendo 3882, imágenes de las cuales se usaron 3726 que tenían información relevante. En total se dibujaron 6553 recuadros de fachadas no comerciales y 3761 recuadros de fachadas comerciales. Las diferencias con la Tabla 2 corresponden a los datos de Selección del Modelo (también conocido como validación) y de Prueba.

El segundo objetivo era entrenar algoritmos supervisados de redes neuronales, como los que se explican en la Sección 3.2, para la tarea de detectar el comercio en una imagen de calle, como fue expuesto en la Sección 5.2. La Tabla 4 muestra una comparación entre tres de las principales arquitecturas

Tabla 3. Resumen descriptivo de los datos usados para entrenar los detectores de comercio.

Estadística	Valor
Número de puntos muestreados	10000
Número de panorámicas descargadas	4357
Número de panorámicas etiquetadas	1941
Número de imágenes de fachada lateral etiquetadas	3726
Número máximo de comercios etiquetados por imagen	16
Comercio promedio etiquetado por imagen	1.01
Número total de etiquetas de comercio	3761
Número total de etiquetas de no comercio	6553

del estado del arte para detección de objetos, usando nuestras métricas en conteo de comercios, propuestas en la Sección 5.2.1, ya que reflejan mejor el desempeño para nuestra tarea. La métrica p_{comer} puede tomar valores continuos de cero a uno, mientras que err_{comer} puede tomar valores continuos de cero a n. Para un detector de comercio ideal se esperaría que $p_{comer} = 1$ y $err_{comer} = 0$. Usando esta lógica, se evidencia que el Faster R-CNN se desempeña mejor en el p_{comer} (celdas verdes), y aunque no se desempeño tan bien en el err_{comer} como el SSD, el SSD se desempeña muy mal en el p_{comer} (celdas rojas). Por lo tanto, elegimos el Faster R-CNN como arquitectura para nuestros detectores de comercio visible.

Un ejemplo de como se puede interpretar el rendimiento del detector de comercio se ilustra en la Figura 13. En esta figura, los cuadros azules representan que el detector debía encontrar esa fachada y no lo hizo (sin importar el tipo). Los cuadros rojos indican que ubico bien la fachada pero la clasifico mal. Los cuadros verdes representan que que ubicó bien la fachada y que la clasificó en la categoría correcta. Usando nuestras métricas en conteo de comercios, expuestas en la Sección 5.2.1, estas predicciones tendrían un $p_{comer} = 0.5$, es decir dos comercios predichos correctamente de cuatro que existen en la imagen.

También en la Tabla 5 se muestran como son los resultados de evaluar las métricas de comercio propuestas sobre el Faster R-CNN con un parámetro de regularización $l_2 = 0.001$, el cual fue el detector de comercio con mejor rendimiento, prediciendo sobre los grupos de datos de entrenamiento, selección del modelo (también conocido como validación) y prueba. Adicionalmente, en la Tabla 6 se reportan algunas de las métricas más convencionales en detección de objetos para este detector, las indicadas en la Sección 3.2.2.

Tabla 4. Resumen con las métricas de rendimiento en el conteo de comercios para las arquitecturas propuestas en la Sección 5.2.1. Se esperaría que un detector de comercio ideal tenga un $p_{comer} = 1$ y un $err_{comer} = 0$. Usando una escala de colores para resaltar los resultados, donde los verdes son los mejores, amarillos no tan malos, naranja son malos, y rojos son muy malos resultados.

Métrica	Estimador	Faster R-CNN	SSD	Yolov3
$\mathbf{p}_{comer}$	**Media**	0.8220	0.5486	1.4218
	Desv. estdr.	0.7488	0.6353	1.3349
$\mathbf{err}_{comer}$	**Media**	0.1656	0.1046	1.4118
	Desv. estdr.	0.4798	0.3589	1.0687

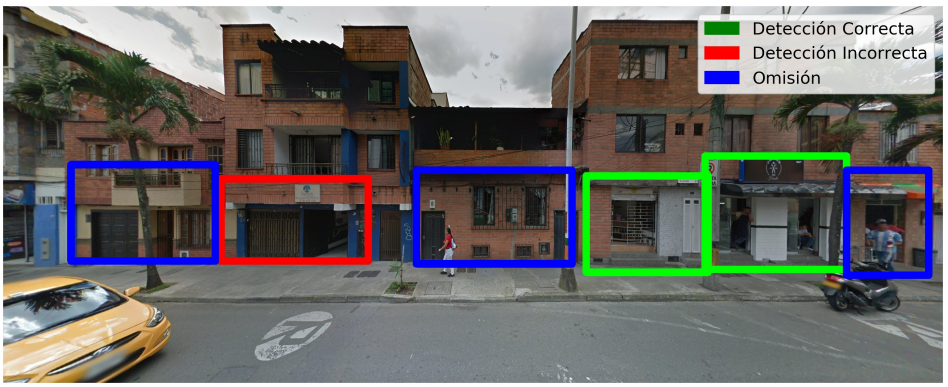

Figura 13. Imagen con vista de calle de ejemplo para caracterizar el detector de comercio.

Tabla 5. Resumen de métricas de rendimiento en el conteo de comercios para el Faster R-CNN con un parámetro de regularización $l_2 = 0.001$, evaluadas en las divisiones de la verdad absoluta para entrenamiento, selección del modelo (o validación) y pruebas.

Métrica	Estimador	Entrena-miento	Selección del Modelo	Prueba
$\mathbf{p}_{comer}$	**Media**	0.9859	1.0119	0.9556
	Desv. estdr.	0.3708	0.8310	0.8349
$\mathbf{err}_{comer}$	**Media**	0.0345	0.2222	0.2000
	Desv. estdr.	0.2191	0.5436	0.5082

El siguiente objetivo es hacer un modelado matemático para producir un mapa de la distribución espacial del comercio en una ciudad o región, el cuál está descrito en la metodología, en la Sección 5.3.1. El modelado consta de unas operaciones espaciales sobre los datos de entrada para producir, inicialmente los puntos geográficos de referencia. La Figura 14 muestra los puntos geográficos de referencia para el centro de Envigado. Estos puntos de referencia se usan para conseguir las imágenes de GSV,

Tabla 6. Resumen métricas de rendimiento en detección de objetos con respecto a la categoría Comercio, evaluadas en las divisiones de la verdad absoluta para entrenamiento, selección del modelo (o validación) y pruebas.

Métrica	Entrenamiento	Selección del Modelo	Prueba
AP@0.5	0.8978	0.4907	0.5039
Precision	0.9965	0.9706	0.9816
Recall	0.8871	0.6094	0.5941
Medida F1	0.9386	0.7487	0.7402

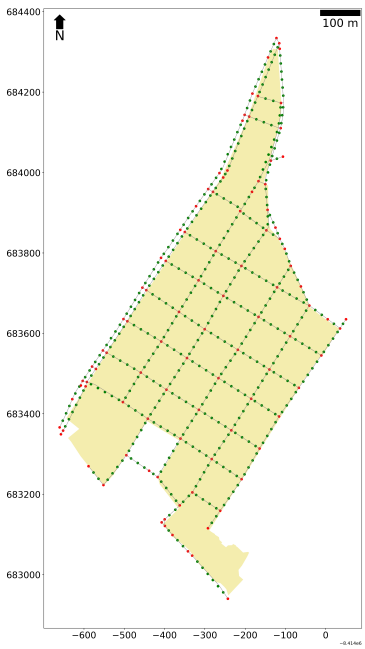

Figura 14. Ejemplo de nube de puntos muestreadas para Envigado.

correspondientes a cada punto, y sobre ellas se hace inferencia sobre la presencia del comercio, como se explica en las Secciones 5.3.2 y 5.3.3. Luego se cuentan los resultados del comercio y se llevan a una función de suavizado con Kernel espacial, con la que se estima para una distribución continua para el espacio estudiado, como se explica en la Sección 3.4.3.

Una de las aplicaciones de la metodología propuesta, es para el caso del Área Metropolitana del Valle de Aburrá con el objetivo de tener un mapa con la distribución espacial del comercio en una ciudad latinoamericana. Inicialmente se usó la extensión urbana para consultar las mallas viales y generar los puntos geográficos de referencia para el Valle de Aburrá. Posteriormente se hizo el barrido exhaustivo

para poder tener imágenes para todo el Valle de Aburrá, del año más reciente que estuviera disponible para cada punto, y a una distancia entre puntos de 20 metros. La distribución espacial del comercio visible en el Valle de Aburrá se muestra en la Figura 15. En esta figura se evidencia para cuales puntos no hay imágenes de GSV disponibles, en especial en las zonas de los extremos Nororiental y Norocciental de Medellín.

Otra aplicación de la metodología propuesta, como un caso de estudio fue solo para el centro del Envigado y con imágenes de sólo el año 2017, como se explica en la Sección 5.4. Esto porque el último objetivo específico era comparar las distribuciones espaciales del comercio real (entendido como el comercio predicho por nuestra metodología) y el comercio oficial (entendido como el comercio registrado en Cámara de Comercio), para producir un mapa de diferencias, el cual es nuestra estimación de comercio informal para Envigado. Para la representación de los mapas se grafican los contornos correspondientes a los diferentes niveles de intensidad. Con el objetivo de que los mapas de calor del comercio real, el comercio oficial y la diferencia sean gráficamente comparables, usamos los mismos niveles y la misma escala de colores para graficar los contornos de unos y otros. En la Figura 16 se muestra la comparación gráfica de, a la izquierda del comercio real visible (predicho por nuestra metodología), y a la derecha los comercios oficiales visibles (registros visibles de la Cámara de Comercio), ambas utilizando intensidad de Kernel con una amplitud de banda de 20 metros.

La Figura 17 muestra nuestra estimación del comercio visible informal para el centro de Envigado en el año 2017, como se propuso en la Sección 5.4. Los niveles que se reflejan en la leyenda de colores son valores de intensidad, por lo que, por ejemplo, donde es más rojo podríamos decir que hay aproximadamente 12 comercios. Esta distribución espacial continua fue estimada de la forma propuesta en la metodología, en la Sección 5.3.4. Los valores que se usaron para colorear esta figura son los mismos que los de la Figura 16, con el fin comparar los colores y poder hacer un mejor análisis.

7. Análisis de resultados

El entrenamiento y las comparaciones a nivel de rendimiento del detector, se hacen con los datos del proceso de etiquetado visual y manual que se hizo de imágenes de comercios del Valle de Aburrá, como el indicado en la Sección 5.1.2. La diferencia entre los resultados para las divisiones de las bases de datos de entrenamiento y pruebas muestran que tan bien logró generalizar el modelo, al aplicarse a datos que nunca había visto, y de los cuales teníamos las verdades absolutas. Cómo se observa en la Tabla 5, las diferencias entre el rendimiento del detector en la base de datos de entrenamiento y el de la de pruebas, es de cerca de un 3 %, lo cual no es muy elevado y nos da a pensar que el modelo está generalizando bien.

El modelo no se desempeñó tan bien en las métricas de detección de objetos, que fueron explicadas en la Sección 3.2.2, como se ilustra en la Tabla 6. Aunque un AP@0.5 de 0.5039 para test es razonable con los resultados del estado del arte, una Sensibilidad de 0.5941 y una Precision de 0.9816, nos dice que los recuadros que encontró lo hizo con una alta similaridad geométrica, pero hay algunos recuadros que no

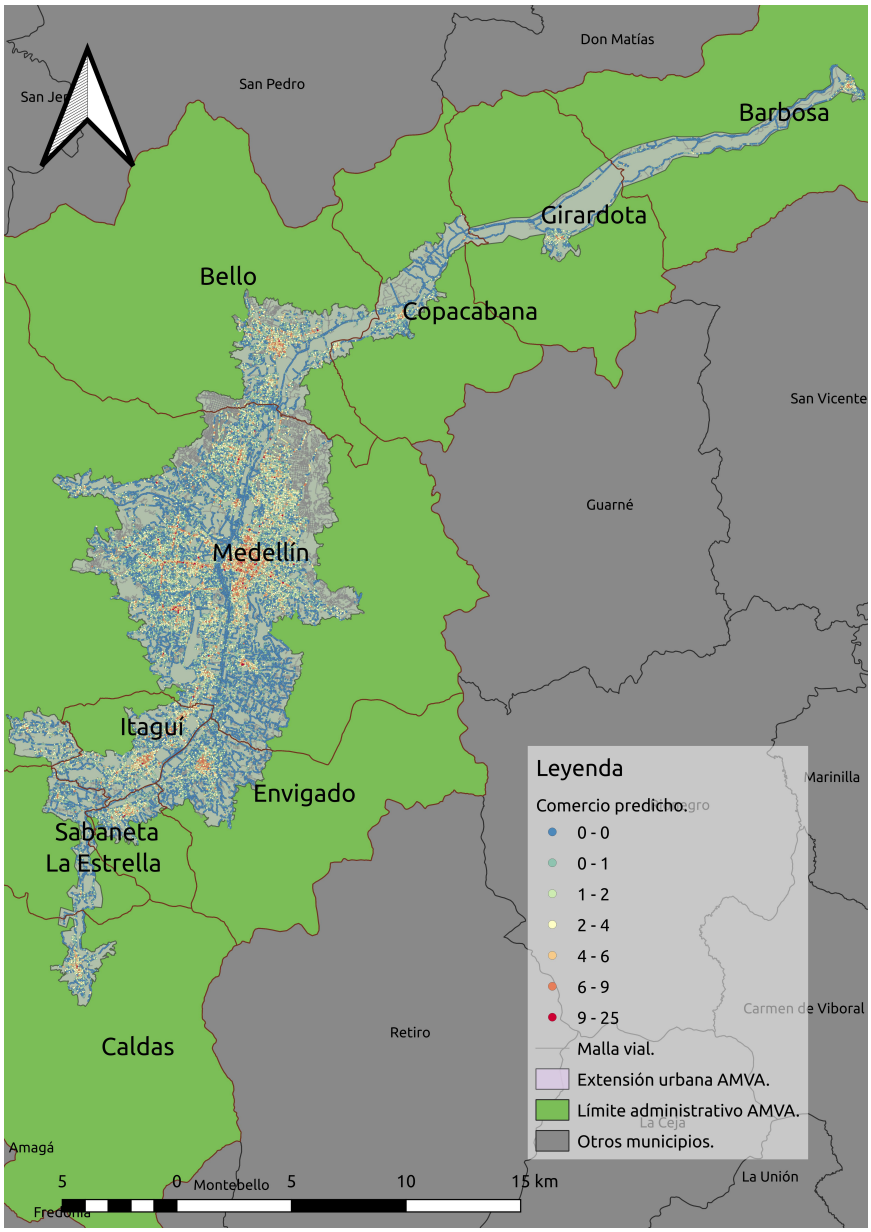

Figura 15. Predicciones discretas del comercio visible para toda el Área Metropolitana del Valle de Aburrá.

alcanzo a detectar. Este resultado nos da la noción de que aunque se este desempeñando bien en conteos,

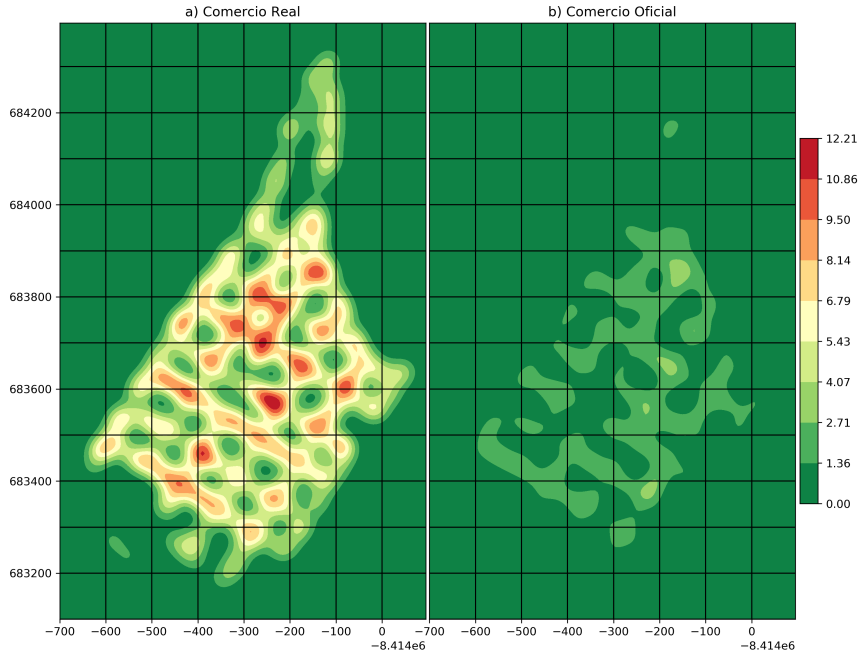

Figura 16. Mapas de calor de las estimaciones de intensidades del comercio real (imagen izquierda) y comercio oficial (imagen derecha) de Cámara de Comercio, ambos en la misma escala de colores.

es posible que las ubicaciones propuestas se pueden mejorar. Una medida F1 de 0.7402 para los datos de prueba nos da a pensar que este modelo podría ser útil para detectar el comercio visible en otras ciudades colombianas con características de estructuras de fachadas similares.

Visualmente se evidencia cómo la distribución espacial del comercio visible al interior del Valle de Aburrá se concentra alrededor de los parques principales de cada municipio y de cada comuna de Medellín. También se evidencian patrones interesantes, en los cuales se ve como las altas concentraciones de comercios dibujan las principales vías arterias de la ciudad (como por ejemplo la Avenida 33, la Avenida 80, la calle San Juan, la calle 30, etc.). La única excepción es la autopista Regional del Valle de Aburrá, la cual hace parte de las carreteras de la Red Vial Nacional (INVIAS 2020), también llamadas vías nacionales. Aunque requeriría un análisis más profundo, los datos sugieren que los comercios se asientan más en los lugares que son altamente transitados por peatones y automóviles, exceptuando vías nacionales. A una conclusión similar llegaron Omer & Goldblatt (2016) usando datos de comercio tomados de encuestas.

Adicionalmente, se hizo otra tarea usando nuestros resultados con el detector de comercio, y es un caso de estudio preeliminar de informalidad empresarial, detallado en la Sección 5.4, aprovechando el hecho que para el municipio de Envigado, teníamos los datos oficiales de Cámara de Comercio.

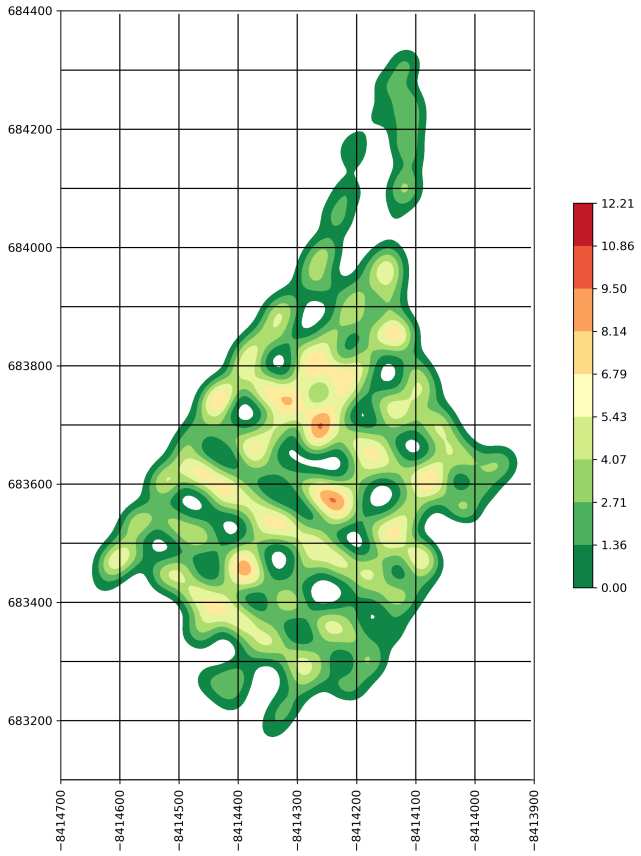

Figura 17. Diferencia de intensidades estimadas entre el comercio real y el comercio oficial. Estimación de la distribución espacial del comercio informal visible en el centro de Envigado.

Debe señalarse que como nuestras predicciones de comercio contienen un superconjunto de los datos de comercio oficial de Cámara de Comercio, su diferencia nos da una idea de la distribución del comercio informal. Para este caso de estudio de informalidad comercial en el centro de Envigado, podemos ver como la distribución espacial de los comercios visibles informales es similar a la distribución espacial del comercio real visible. Esto debido esencialmente a que los comercios registrados en la cámara de comercio están distribuidos de forma similar en el espacio que las predicciones de comercio, solo que en menor intensidad. Por lo que no tenemos evidencia para afirmar que la informalidad siga algún patrón de concentración diferente al de los comercios en general.

8. Conclusiones

Esta tesis muestra que es posible generar una herramienta que permita hacer una buena estimación de la distribución espacial del comercio en ciudades altamente informales, a partir de imágenes con vista de calle. Sin embargo, los resultados de las métricas de detección de objetos permiten inferir que aunque se este desempeñando bien en conteos, es posible que las ubicaciones propuestas se pueden mejorar. Los datos sugieren que los comercios se asientan más en vías que son altamente transitadas por peatones y automóviles, exceptuando las vías nacionales. La distribución espacial del comercio visible informal en el centro de Envigado sigue un patrón espacial similar al del comercio visible, solo que en menor cuantía. De modo que no hay evidencia de que la informalidad comercial se concentre en lugares específicos.

La escalabilidad en este ejercicio es un tema a tener en cuenta. Para hacer este ejercicio en grandes extensiones de territorio (como el caso de áreas metropolitanas), el manejo de la gran cantidad de imágenes agrega un problema logístico. Se espera que con el avance de la tecnología debería ser cada vez menor el problema de la logística en el manejo de la información. En esta tesis se solucionó procesando imágenes por municipio y luego agregando los resultados.

Esta tesis abre nuevas lineas de investigación, al crear esta información de comercio que es tan difícil de tener disponible en países en desarrollo. La estimación de la distribución espacial de la actividad comercial abre la oportunidad de estudiar la demografía empresarial en una ciudad si se realizan de forma periódica. Los datos que se logran generar con esta metodología, permiten la evaluación de impacto de políticas públicas en el comercio, estudiar patrones de aglomeración del comercio, modelarlo con otras variables espaciales para identificar determinantes de la ubicación del comercio, entre otros. Una importante aplicación sería poder refinar las predicciones de comercio, usando algoritmos de OCR para leer el texto en los letreros comerciales, con la idea de inferir su actividad comercial.

Adicionalmente, esta metodología como herramienta de política puede permitir a los gobiernos locales enfrentar con mejor información los retos que trae el Objetivo de Desarrollo Sostenible 8 que busca asegurar el trabajo decente, pleno empleo, empleo digno y de calidad. Por ejemplo esta metodología puede permitir identificar posibles conflictos entre dónde esta ubicado el comercio y dónde debería estar según los planes maestros de cada ciudad. También, si se compara la distribución de los comercios identificados con las cifras de oficiales de tributación, se puede tener una medición del nivel de evasión de impuestos en una ciudad. Esta metodología es fácil de replicar incluso dónde no hay imágenes de GSV, por el hecho de que solo se necesitan imágenes de las fachadas en la zona de estudio, para lo que se podrían fotografiar con una cámara tradicional o con cámaras embebidas en plataformas móviles (sean estos vehículos o drones).

Agradecimientos

Esta tesis de maestría se completó con el apoyo del programa PEAK Urban, respaldado por el Fondo de Investigación Global Challenge de UKRI, Grant Ref: ES / P011055 / 1.

Referencias

Acharya, A., Fang, H., & Raghvendra, S. (2017). Neighborhood Watch: Using CNNs to Predict Income Brackets from Google Street View Images, . URL: http://cs231n.stanford.edu/reports/2017/pdfs/556.pdf.

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., & Weaver, J. (2010). Google Street View: Capturing the World at Street Level. *Computer*, *43*, 32–38. URL: http://ieeexplore.ieee.org/document/5481932/. doi:10.1109/MC.2010.170.

BarringtonLeigh, C., & Millard-Ball, A. (2017). The world's user-generated road map is more than 80 % complete. *PLoS ONE*, *12*, 1–20. URL: http://dx.doi.org/10.1371/journal.pone.0180698. doi:10.1371/journal.pone.0180698.

Bureau, U. S. C. (2019). Using Economic Census Bureau Statistics. URL: https://www.census.gov/programs-surveys/economic-census/information/using-census-bureau-statistics.html.

CCMA (2018). Informe de Gestión año 2017 Cámara de Comercio de Medellin para Antioquia. *CamCom*, (pp. 1–56). URL: http://www.camaramedellin.com.co/site/Portals/0/Documentos/2018/Informedegesti{\unhbox\voidb@x\bgroup\let\unhbox\voidb@x\setbox\@tempboxa\hbox{o\global\mathchardef\accent@spacefactor\spacefactor}\accent19o\egroup\spacefactor\accent@spacefactor\futurelet\@let@token\penalty\@M\hskip\z@skip}ngeneralCCMA2017fin.pdf.

CID, & Bancoldex (2017). Centre for International Development at Harvard University. El Atlas Colombiano de Complejidad Económica. URL: http://datlascolombia.com/#/location/34.

DANE (2019a). Encuesta anual manufacturera (EAM). URL: https://www.dane.gov.co/index.php/estadisticas-por-tema/industria/encuesta-anual-manufacturera-enam.

DANE (2019b). Metodología Censo económico de Colombia 2021, . (pp. 1–27). URL: https://www.dane.gov.co/index.php/estadisticas-por-tema/comercio-interno/censo-economico-de-colombia-2021-documento-metodologico.

Daszykowski, M., & Walczak, B. (2009). Density-Based Clustering Methods. *Comprehensive Chemometrics*, *2*, 635–654. doi:10.1016/B978-044452701-1.00067-3.

Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and roc curves. In *Proceedings of the 23rd International Conference on Machine Learning* ICML '06 (p. 233–240). New York, NY, USA: Association for Computing Machinery. URL: https://doi.org/10.1145/1143844.1143874. doi:10.1145/1143844.1143874.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.

Devroye, L., Györfi, L., & Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition* volume 31 of *Stochastic Modelling and Applied Probability*. Springer.

Duque, J. C., Church, R. L., & Middleton, R. S. (2011). The p -Regions Problem. *Geographical Analysis*, *43*, 104–126.

Duque, J. C., Ramos, R., & Suriñach, J. (2007). Supervised regionalization methods: A survey. *International Regional Science Review*, *30*, 195–220. doi:10.1177/0160017607301605.

Everingham, M., Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *Int. J. Comput. Vision*, *88*, 303–338. URL: https://doi.org/10.1007/s11263-009-0275-4. doi:10.1007/s11263-009-0275-4.

Fernández, C. (2018). Informalidad empresarial en Colombia. *Revista Desarrollo y Sociedad*, (pp. 211–243). doi:10.13043/dys.63.5.

Forsyth, D. A., & Ponce, J. (2012). *Computer Vision - A Modern Approach, Second Edition.*. Pitman.

Fotheringham, A. S., Brundson, C., & Chalrton, M. (2010). *Quantitative Geography : Perspectives on Spatial Data Analysis*. URL: http://www.sage-ereference.com/view/hdbk{_}qualgeography/n18.xml. doi:10.1111/j.1467-9787.2009.00642.x.

Gonzalez, R. C. (2002). Digital_Image_Processing_2ndEd.pdf. URL: www.prenhall.com/gonzalezwoods.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

Google (2019). Street View Static API Usage and Billing. URL: https://developers.google.com/maps/documentation/streetview/usage-and-billing?hl=es_419.

Grippa, T., Georganos, S., Zarougui, S., Bognounou, P., Diboulo, E., Forget, Y., Lennert, M., Vanhuysse, S., Mboga, N., & Wolff, E. (2018). Mapping Urban Land Use at Street Block Level Using OpenStreetMap, Remote Sensing Data, and Spatial Metrics. *ISPRS International Journal of Geo-Information*, *7*, 246. doi:10.3390/ijgi7070246.

Hara, K., Le, V., & Froehlich, J. (2013). Combining crowdsourcing and google street view to identify street-level accessibility problems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, (p. 631). URL: http://dl.acm.org/citation.cfm?doid=2470654.2470744. doi:10.1145/2470654.2470744.

Hartley, R., & Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. (2nd ed.). Cambridge University Press. doi:10.1017/CBO9780511811685.

Hu, T., Yang, J., Li, X., & Gong, P. (2016). Mapping urban land use by using landsat images and open social data. *Remote Sensing*, *8*. doi:10.3390/rs8020151.

Hui, J. (2018a). mAP (mean Average Precision) for Object Detection. URL: https://medium.com/@jonathan_hui/map-mean-average-precision-for-object-detection-45c121a31173.

Hui, J. (2018b). Object detection: speed and accuracy comparison (Faster R-CNN, R-FCN, SSD, FPN, RetinaNet and YOLOv3). URL: https://medium.com/@jonathan_hui/object-detection-speed-and-accuracy-comparison-faster-r-cnn-r-fcn-ssd-and-yolo-5425

ICCV (2015). ImageNet and MS COCO Visual Recognition Challenges Joint Workshop. URL: http://image-net.org/challenges/ilsvrc+mscoco2015.

Ilic, L., Sawada, M., & Zarzelli, A. (2019). Deep mapping gentrification in a large Canadian city using deep learning and Google Street View. *PloS one*, *14*, e0212814. doi:10.1371/journal.pone.0212814.

INVIAS (2020). Red Vial Nacional - Instituto Nacional de Vias INVIAS. URL: https://www.invias.gov.co/index.php/red-vial-nacional.

Iovan, C., Picard, D., Thome, N., & Cord, M. (2012). Classification of urban scenes from georeferenced images in urban street-view context. *Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012*, *2*, 339–344. doi:10.1109/ICMLA.2012.171.

Kang, H.-W., & Kang, H.-B. (2017). Prediction of crime occurrence from multi-modal data using deep learning. *Plos One*, *12*, e0176244. URL: http://dx.plos.org/10.1371/journal.pone.0176244. doi:10.

1371/journal.pone.0176244.

Kelleher, J. D., MacNamee, B., & D'Arcy, A. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. Cambridge, MA: MIT Press.

Laupheimer, D., Tutzauer, P., Haala, N., & Spicker, M. (2018). NEURAL NETWORKS for the CLASSIFICATION of BUILDING USE from STREET-VIEW IMAGERY. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (pp. 177–184). volume 4. doi:10.5194/isprs-annals-IV-2-177-2018.

Lee, M. L., & Pace, R. K. (2005). Spatial distribution of retail sales. *Journal of Real Estate Finance and Economics*, *31*, 53–69. doi:10.1007/s11146-005-0993-5.

Liao, C., Wang, W., Sakurada, K., & Kawaguchi, N. (2018). Image-Matching Based Identification of Store Signage Using Web-Crawled Information. *IEEE Access*, *6*, 45590–45605. doi:10.1109/ACCESS.2018.2865490.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8693 LNCS*, 740–755. doi:10.1007/978-3-319-10602-1_48. arXiv:1405.0312.

Lu, Y. (2018). Using Google Street View to investigate the association between street greenery and physical activity. *Landscape and Urban Planning*, . doi:10.1016/j.landurbplan.2018.08.029.

Marsland, S. (2014). *Machine Learning An Algorithmic Perspective Second Edition*. (1st ed.). Taylor & Francis Group.

Martínez Agut, M. d. P. (2015). OBJETIVOS DE DESARROLLO SOSTENIBLE (ODS , 2015-2030) Y AGENDA DE DESARROLLO POST 2015 A PARTIR DE LOS OBJETIVOS DE DESARROLLO DEL MILENIO (2000-2015) M[a] del Pilar Martínez Agut Universitat de València. *Quadernsaminacio.net*, (p. 16).

Microsoft (2015). COCO 2015 Object Detection Task. URL: http://cocodataset.org/#detection-2015.

Mooney, S. J., DiMaggio, C. J., Lovasi, G. S., Neckerman, K. M., Bader, M. D., Teitler, J. O., Sheehan, D. M., Jack, D. W., & Rundle, A. G. (2016). Use of google street view to assess environmental contributions to pedestrian injury. *American Journal of Public Health*, *106*, 462–469. doi:10.2105/AJPH.2015.302978.

Movshovitz-Attias, Y., Yu, Q., Stumpe, M. C., Shet, V., Arnoud, S., & Yatziv, L. (2015). Ontological supervision for fine grained classification of Street View storefronts. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June*, 1693–1702. doi:10.1109/CVPR.2015.7298778.

Olaya, V. (2009). Sistemas de información geográfica. *Cuadernos internacionales de tecnología para el desarrollo humano, ISSN 1885-8104, Nº. 8, 2009 (Ejemplar dedicado a: Tecnologías de la información geográfica)*, .

Omer, I., & Goldblatt, R. (2016). Spatial patterns of retail activity and street network structure in new and traditional Israeli cities. *Urban Geography*, *37*, 629–649. URL: http://dx.doi.org/10.1080/02723638.2015.1101258. doi:10.1080/02723638.2015.1101258.

OpenStreetMap contributors (2017). Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org.

Perry, G. E., Maloney, W. F., Arias, O. S., Fajnzylber, P., & Saavedra-chanduvi, A. D. M. J. (2010). *Informality: Exit and Exclusion* volume 57. URL: https://www.openknowledge.worldbank.org/bitstream/handle/10986/6730/400080Informal101OFFICIAL0USE0ONLY1.pdf?sequence=1. doi:10.1596/978-0-8213-7092-6.

Portafolio (2019). Así será el censo económico que realizará el Dane en el 2021. URL: https://www.portafolio.co/economia/asi-sera-el-censo-economico-que-realizara-el-dane-en-el-2021-529656.

Porzi, L., Rota Bulò, S., Lepri, B., & Ricci, E. (2015). Predicting and Understanding Urban Perception with Convolutional Neural Networks. *Proceedings of the 23rd ACM international conference on Multimedia - MM '15*, (pp. 139–148). URL: http://dl.acm.org/citation.cfm?doid=2733373.2806273. doi:10.1145/2733373.2806273.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*, 1137–1149. doi:10.1109/TPAMI.2016.2577031. arXiv:1506.01497.

RiSE-group (2015). AI-Geocoder Research in Spatial Economics - RiSE group - EAFIT. URL: http://dev.geo.ai-rise.com/.

Schneider, F. (2016). Estimating the size of the shadow economy: Methods, problems and open questions. *Turkish Economic Review*, *3*, 256–280. doi:10.1453/ter.v3i2.832.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2014). Overfeat: Integrated recognition, localization and detection using convolutional networks. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, . arXiv:1312.6229.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*. URL: https://doi.org/10.1186/s40537-019-0197-0. doi:10.1186/s40537-019-0197-0.

Tang, J., & Long, Y. (2018). Measuring visual quality of street space and its temporal variation: Methodology and its application in the Hutong area in Beijing. *Landscape and Urban Planning*, (pp. 0–1). URL: https://doi.org/10.1016/j.landurbplan.2018.09.015. doi:10.1016/j.landurbplan.2018.09.015.

thryv (2020). The Real Yellow Pages. URL: https://www.yellowpages.com/.

UN, U. N. (2010). Economic census: challenges and good practices. *As of United Nations*, (pp. 1–177).

Yu, Q., Szegedy, C., Stumpe, M. C., Yatziv, L., Shet, V., Ibarz, J., & Arnoud, S. (2015). Large Scale Business Discovery from Street Level Imagery, . URL: http://arxiv.org/abs/1512.05430. arXiv:1512.05430.

Zamir, A. R., Darino, A., & Shah, M. (2011). Street view challenge: Identification of commercial entities in street view imagery. *Proceedings - 10th International Conference on Machine Learning and Applications, ICMLA 2011*, *2*, 380–383. doi:10.1109/ICMLA.2011.181.

Zhou, Z., Farhat, F., & Wang, J. Z. (2017). Detecting dominant vanishing points in natural scenes with application to composition-sensitive image retrieval. *IEEE Transactions on Multimedia*, *19*, 2651–2665. URL: http://dx.doi.org/10.1109/TMM.2017.2703954. doi:10.1109/tmm.2017.2703954.