

Universidad de Buenos Aires  
Facultad de Ciencias Económicas  
Escuela de Estudios de Posgrado

---

**MAESTRÍA MÉTODOS CUANTITATIVOS PARA LA  
GESTIÓN Y ANÁLISIS DE DATOS EN  
ORGANIZACIONES**

---

**TESIS DE MAESTRÍA**

---

Ciudades inteligentes y detección de infracciones  
viales. Aplicaciones de visión artificial.

---

AUTOR: FEDERICO GARCIA BLANCO  
DIRECTOR: JAVIER IGNACIO GARCIA FRONTI

DICIEMBRE 2020

---

## Contenido

Agradecimientos.....	3
Resumen .....	3
Introducción.....	3
1. Gestión de datos en ciudades inteligentes .....	5
1.1. Motivos y razones para implementaciones de control.....	7
1.2. Política basada en evidencia .....	10
1.3. Variedad de datos exequibles por parte de entes gubernamentales .....	10
1.4. Uso responsable y privacidad a través de la reglamentación.....	13
2. Detección y clasificación de objetos bajo un contexto de ciudades inteligentes. Análisis de diferentes modelos .....	17
2.1. Algoritmos de detección y clasificación de objetos.....	18
2.2. Métrica de error sobre el criterio de selección.....	23
2.3. Aprendizaje de transferencia y sus beneficios .....	26
2.4. Codificación de imágenes. La percepción del humano y de las computadoras....	27
2.5. Descripción de modelos de detección y clasificación de objetos .....	30
2.5.1. Single Shot MultiBox Detector .....	31
2.5.2. You Only Look Once .....	31
2.5.3. Mask_RCNN .....	32
3. Selección del modelo a partir de la detección temprana de infracciones viales.....	32
3.1. Resultados de los modelos a partir del análisis de imágenes.....	35
3.3. Ventajas y desventajas del modelo seleccionado .....	40
3.4. Aplicaciones y usos del modelo de detección de objetos .....	43
Conclusiones.....	44
Referencias Bibliográficas.....	45
Referencias .....	46
Anexos .....	47

### **Agradecimientos**

Quisiera expresar mis agradecimientos a todas aquellas personas que me acompañaron en este proceso de formación tan importante para mi carrera. En principio me gustaría mencionar al director de mi tesis Javier Garcia Fronti que sin su dirección y su ayuda no hubiera podido finalizarla. También agradecerles a actores muy importantes como Pablo Herrera y Facundo Santiago, personas fundamentales no solo para mí sino para todos los alumnos brindando ayuda durante todo el proceso.

A mi novia Antonella que me acompañó y me apoyó durante estos últimos años viendo mis avances y frustraciones. A Pedro Luis Delgado Escobar y Santiago “tato” Manzi que fueron aquellos que me brindaron contención y la posibilidad de tener discusiones muy profundas al respecto de la temática abordada. Finalmente, a mis padres, que tuvieron la gran sabiduría de guiarme en un camino que sin duda es el correcto dentro de la excelencia de la Educación Pública Argentina.

### **Resumen**

El objetivo del presente trabajo es seleccionar un método que sea capaz de identificar infracciones viales de vehículos de tránsito pesado en un contexto de ciudades inteligentes. La forma en que se va a desarrollar la temática consta de la diferenciación de tres algoritmos de visión artificial para poder finalmente optar por el mejor de ellos mencionando las ventajas y desventajas que sostiene cada uno. Además, durante el resto del trabajo se va profundizar sobre aspectos centrales del análisis como lo pueden ser la gestión y el uso de los datos, las bases que transforman a una urbe en una ciudad inteligente y los avances que ha tenido la ciencia con respecto a las nuevas herramientas de la visión artificial.

### **Palabras clave**

Visión artificial; Redes neuronales; Infracciones viales; Detección de objetos

### **Introducción**

El concepto de ciudades inteligentes esta abordado principalmente desde dos aristas fundamentales. Por un lado, se encuentra el uso de datos de forma masiva generados por la sociedad en su conjunto, y por el otro se encuentra la automatización de ciertos procesos que hasta hace algunos años eran imposible de parametrizar. El presente trabajo va a problematizar sobre ambas cuestiones dado que, si bien se va a generar un mecanismo de automatización sobre un proceso como el que comprende la detección de vehículos de tránsito pesado, también se va a discutir que rol debe tener en el Estado en este tipo de funciones.

El debate para dar es muy contemporáneo ya que afecta a la mayoría de la sociedad en la actualidad. Cuando una persona toma un transporte público, navega en internet desde su teléfono móvil, o utiliza alguna red social, está volcando contenido sumamente importante no solo para las corporaciones sino también para el Estado que desea conocer los gustos y preferencias de sus habitantes. Y es ahí donde está el centro de la cuestión, problematizando sobre los motivos por los cuales el Estado está interesado en conocer a la población de forma más certera utilizando las nuevas tecnologías.

Para analizar de qué forma se pueden aplicar algoritmos de identificación de objetos para detectar infracciones viales de forma temprana en ciudades inteligentes, el presente trabajo va a tener tres apartados principales desde los cuales se van a tratar diferentes temáticas que corresponden al problema en cuestión.

En primer lugar, se decide comenzar la discusión dentro del contexto sobre el cual se va a basar el análisis, por lo tanto, se hace mucho énfasis en los conceptos de ciudades inteligentes, privacidad, e interés estatal entre otras cosas. De esta forma, se comienzan a problematizar temas que son de sumo interés para el trabajo como lo puede ser por ejemplo el rol del Estado en conjunto con las reglamentaciones actuales que regulan el acceso y el uso de los datos públicos.

En el segundo apartado, se vuelca información técnica sobre las formas en que se puede tratar la problemática abordada. Para ello, se decide explicar una variedad de conceptos que son de utilidad para entender el contexto del problema. Se profundiza sobre los tipos de imágenes en los cuales se va a estar trabajando y además se analizará que tipo de algoritmos pueden ser utilizados para atacar el problema. Finalmente, en este apartado se presentarán

los modelos a aplicar y se caracterizarán cada uno de ellos con el fin de comprender que diferencia el uno del otro.

El último apartado va a estar dedicado casi exclusivamente a las mediciones de desempeño sobre los modelos generados para poder tomar la decisión sobre cuál es el modelo que mejor se adapta al contexto planteado. Además, se van a enumerar los riesgos por los cuales se considera necesario algún tipo de implementación al respecto, teniendo en cuenta las ventajas y desventajas de las soluciones actuales al problema de detección de tránsito pesado. Finalmente, para mostrar los resultados de formas más intuitiva, se van a graficar las detecciones de objetos por cada uno de los modelos en la misma imagen para verificar que los desempeños de los modelos distan uno del otro.

## **1. Gestión de datos en ciudades inteligentes**

Si bien actualmente la mayoría de los seres humanos están acostumbrados a vivir en sociedad, desde un inicio las relaciones sociales entre los primeros habitantes eran muy diferentes. Viendo la historia de la humanidad en una especie de línea de tiempo, se podría decir que las relaciones sociales tal cual se las conoce hoy en día es un concepto muy moderno. Estos tipos de avances sociales que fue generando la humanidad siempre tuvieron algunos hitos que formaron el comienzo de nuevas eras generando avances significativos para el ser humano y moldeando la forma que logró al relacionarse el uno con el otro. Y continuando con la línea de tiempo, se podrían encontrar innumerables ejemplos donde los seres humanos a través de ciertas tecnologías generaron cambios de paradigmas para lograr avances significativos. El término de la globalización podría ser un gran ejemplo donde la forma en la cual las personas se relacionan ha cambiado notoriamente eliminando obstáculos territoriales o espaciales. Esto sin duda requería de la generación de nuevas formas de comunicación que fueron creadas desde la llegada de un incipiente telegrama para luego convertirse en una comunicación telefónica o una conexión de internet, por ejemplo.

Estos tipos de avances tecnológicos han generado una curva de crecimiento social tan disruptiva que talvez la sociedad se encuentre en una pendiente de crecimiento que se está aplanando. Aunque esta discusión podría tomar años de debate y aun así no se llegaría a una

conclusión acertada, lo cierto es que la forma en que actualmente las personas viven en sociedad está cambiando mucho y parece que va a continuar en ese camino. La forma en la cual las personas se relacionan en la actualidad implica un uso permanente de dispositivos móviles desde las edades más tempranas hasta las personas mayores que se deben adaptar a la situación por la alta demanda de las nuevas tecnologías.

Esta situación actual, genera un aspecto de suma importancia que debe ser analizado. Las barreras de entrada a la comunicación nunca fueron tan laxas como lo son en la actualidad, ya que no se requiere de un dispositivo muy difícil de conseguir como en su momento fueron las radios o los televisores y teléfonos de línea fija, sino que actualmente el avance de la globalización generó que muchas personas pudieran tener acceso a una comunicación más justa y universal.

Se podría problematizar sobre los motivos y las intenciones que han tenido aquellas personas que participaron en este proceso de generar una comunicación más universal dado que talvez ni siquiera ellos pudieron imaginar los impactos de semejante avance. Pero el hecho de relevancia que se quiere analizar en este trabajo está vinculado con las consecuencias que esto generó independientemente de cómo fue su creación.

Actualmente al estar permanentemente conectados a los dispositivos móviles, se generan variedad de datos que hasta hace unos años no eran utilizados por las organizaciones por diversos motivos. Con los avances tecnológicos que ocurrieron en los últimos años, esta situación cambió rotundamente y permitió generar un nuevo mercado, donde la información que generan las personas desde sus dispositivos móviles es de suma importancia.

Si bien al comienzo esto solo les interesó a los grupos económicos para maximizar rentas y mitigar riesgos, en la actualidad las organizaciones estatales también ponen foco en este aspecto porque se percatan de la oportunidad que están desperdiciando. Y eso identifica como sociedad a lo que se tiene hoy en día, países enfocados en la utilización de los datos generados por su población para analizarlos y poder así sacar provecho de ellos. Siendo esto último lo que permite comenzar a mencionar el concepto de ciudades inteligentes.

Con el avance en la recopilación de los datos por parte de los nuevos desarrollos en hardware que han eliminado varias barreras operativas, surgen ciudades que están muy interesadas en monitorear el movimiento de su población. Mas allá de la discusión de fondo donde se analiza si el análisis propuesto mejora o empeora la vida de sus habitantes, el hecho es que

actualmente es un modelo ya establecido y en pleno crecimiento dejando de ser solo para las grandes metrópolis.

Por este motivo, este apartado va a atacar un objetivo muy particular del trabajo que consta de evaluar el impacto de la gestión de datos de forma responsable dado que estas nuevas ciudades se tornan muy poderosas cuando manipulan información tan sensible.

Adicionalmente, estas ciudades modernas se encargan de emplear cada vez más mecanismos que permitan recopilar información sobre su población para luego ser analizada y puesta a disposición de la toma de decisión gubernamentales para, entre otras cosas, reducir costos y optimizar procesos (Zanella, 2014). Ejemplos como cámaras de seguridad, estaciones de recolección de ruidos de ambiente, medidores de calidad de aire y muchos más, son los que actualmente hacen recapacitar sobre la importancia que tiene esta información para la toma de decisiones a niveles estatales. Las ciudades que fomentan e incentivan este tipo de alternativas se las considera ciudades inteligentes dado que aplican mecanismos de análisis complejos a partir del uso de los datos generados por la población.

### **1.1. Motivos y razones para implementaciones de control**

Como se ha mencionado anteriormente, aún no se sabe si las nuevas tecnologías se han creado teniendo en cuenta el poder de los datos, o si las organizaciones se encontraron con cantidades enormes de datos para analizar luego del avance de las tecnologías. Mas allá de eso, lo cierto es que en la actualidad la cantidad de datos recopilado por las organizaciones es creciente y no parece tender a disminuir sino todo lo contrario. Ahora, una cuestión que se va a profundizar en este apartado es los motivos por los cuales estas organizaciones estatales están tan interesadas en recopilar información para analizar. Considerando a las organizaciones estatales como agentes racionales que toman acciones en pos de mejorar su situación actual, se podría pensar que el beneficio que le brinda dichos análisis es tal que permite hacer inversiones en equipamiento y recursos humanos de grandes magnitudes a pesar de sus elevados costos. Dicho esto, se va a problematizar sobre los aspectos en los cuales estas ciudades inteligentes les puede parecer atractivo desarrollar estos análisis.

En primer lugar, se podría nombrar un aspecto central que es de interés común a todas las ciudades como lo es el control de su población. Este aspecto es crucial y cruza

horizontalmente a todos los motivos que se puedan llegar a mencionar en este apartado porque es el fin más abstracto del análisis. Haciendo una rápida línea de tiempo a través de la historia, al menos desde las épocas de las monarquías donde los reyes eran dueños del poder absolutista que le otorgaban los dioses en diversas civilizaciones, ya existía un control sobre la población y por lo tanto se analizaba este tipo de cuestiones. Y si bien a medida que fue pasando el tiempo, estos controles poblacionales fueron mutando, su aspecto central que implica el tener conocimiento sobre los usos y costumbres de la población no se terminó. Es más, se podría nombrar el caso del filósofo moderno francés Michel Foucault en su obra *Vigilar y Castigar* de 1975 donde introdujo el concepto del panóptico (Foucault, 1975) para controlar los comportamientos de una población. Este tipo de ejemplos permite verificar que las autoridades estatales tienen incentivos claros a la hora de generar un control sobre su población. Es por esto que como uno de los aspectos centrales del motivo por el cual las metrópolis se convierten lentamente en ciudades inteligentes se establece el control de la población.

Como segundo motivo a destacar, se podría indicar el interés de las autoridades estatales en conocer los gustos y preferencias de sus habitantes y por lo tanto lograr la creación de servicios a fines con su población a partir del uso de estas nuevas tecnologías. Este aspecto está ligeramente relacionado con los intereses de cada uno de los mandatarios. Con esto se quiere indicar que no todas las autoridades tienen como objetivo común mejorar los aspectos de cierta parte de la población, tal vez algunos apuntan a medidas más universales y otros apuntan a medidas más sectoriales. Aun así, para ambos casos es de suma importancia comprender los gustos de la población para comenzar a tomar medidas sobre el asunto. Por ejemplo, en ciudades donde el tema de la seguridad es primordial, las autoridades posiblemente prefieran tener cámaras de vigilancia donde pueden identificar siniestros delictivos para actuar con rapidez y lograr mejores tasas de delincuencia. O tal vez en otros casos las autoridades se enfrentan con problemas de mala calidad de aire y por lo tanto deciden actuar en ese sentido y colocar medidores a lo largo y ancho de la ciudad. Si bien esto no quiere decir que ambos aspectos sean mutuamente excluyentes, se desea remarcar que hay áreas dentro de la ciencia de datos que pueden ser más explotadas que otras dentro del contexto de ciudades inteligentes.

Otro aspecto para destacar podría ser el interés fiscal que tiene el Estado sobre sus habitantes. La recopilación de datos que está siendo generada en la actualidad permite que la intromisión estatal pueda ser tal que se involucre en situaciones donde antes era impensado. La facilidad que tiene un Estado en estos días para hacer un cruce de información entre diferentes entidades estatales es muy alta, y por lo tanto esta decisión va a requerir el accionar de un conjunto de organismos coordinados con un mismo fin. En países donde los controles fiscales tienen barreras espaciales y temporales, el uso de los grandes volúmenes de datos para analizar este tipo de situaciones es un principio fundamental para lograr la universalidad en ese aspecto. Muchos podrían debatir si el hecho de que el Estado monitoree los inmuebles de los contribuyentes y los compare con sus declaraciones juradas de bienes personales está bien o directamente no corresponde semejante nivel de intromisión estatal. Aun así, es una práctica que en la actualidad se está haciendo y trae aparejada muchas disputas que van desde los aspectos tecnológicos hasta la problematización ética.

Si bien uno podría mencionar muchos más aspectos donde las autoridades estatales podrían verse interesadas en aplicar mecanismos de uso de grandes volúmenes de datos, tal vez sea preferible mencionar el aspecto ecológico como un bastión creciente en los análisis modernos dado el alto impacto que tienen en las sociedades futuras. Si bien se trata de un aspecto que por diversas razones está más presente en ciertas sociedades que en otras, no cabe duda de que el futuro de consumo y el cuidado del medioambiente tienen que estar equilibrados porque sus consecuencias pueden ser muy costosas desde el punto de vista social y económico. Es por esto por lo que la mayoría de los avances de las ciudades inteligentes en este aspecto se celebra de forma universal porque la mejora que puede generar afecta al resto de la población en mayor o menor medida. En casos de recolección de residuos de manera inteligente, o también los casos de separación automática de residuos entre otros, son celebrados dado que optimiza los recursos y además genera beneficios ecológicos que van a ser valorados a largo plazo. A pesar de esto, existen algunas ciudades que están lejos de aplicar estos avances tecnológicos para mejorar el medioambiente y genera preocupación de todo el mundo porque al fin al cabo, esta generación actual y las posteriores, van a compartir el mismo espacio territorial y las acciones de cada uno afectarán al conjunto de la población mundial.

## **1.2. Política basada en evidencia**

Actualmente, al momento de la redacción del presente trabajo, el mundo está atravesando una pandemia global que afecta a todos los países de manera similar ya que se trata de un virus que no distingue diferencias étnicas ni socio económicas. Esta situación trajo aparejada muchos cambios profundos y repentinos a los cuales las personas debieron adaptarse sin mucha opción de permanecer invariantes al contexto. Ya sea por medidas de cuarentena obligatoria, o distanciamientos sociales preventivos, las personas cambiaron sus hábitos de consumo y su comportamiento dando lugar a nuevas formas de teletrabajo, aumento de consumos a través de internet, mayor conectividad y también aspectos totalmente negativos como depresión juvenil, y aumento en los niveles de soledad de las personas de la tercera edad entre otros. Pero no solo la sociedad fue la que sufrió los cambios, sino que se pudo ver un cambio de paradigma en algunos países con respecto a las decisiones políticas y como están argumentados sus posturas.

El concepto de políticas basadas en evidencia quedó muy expuesto en la presente pandemia dado que se trata de un concepto que se desprende de la medicina basada en evidencia (Pawson, 2006). Conceptos como el aplanamiento de la curva de contagios, letalidad, fatalidad entre otros, fueron muy nombrados en esta época y dado que la visibilidad del problema en los medios de comunicación fue tan grande, hizo que gran parte de la población tenga acceso al debate sobre estos mismos conceptos.

Una discusión para tener por delante podría ser que tan arraigada va a quedar esta forma de política pública que está basada en la evidencia para los años siguientes. Si bien uno podría identificar que las políticas basadas en evidencia parecen haber llegado para quedarse dado su comodidad a la hora de argumentar, no es seguro que así sea dado que la política comprende muchos aspectos muy diversos donde la lógica no siempre tiene un lugar predominante.

## **1.3. Variedad de datos exequibles por parte de entes gubernamentales**

En este apartado se quiere profundizar sobre el tipo de dato al que puede acceder un ente gubernamental y que posible uso le podría dar al mismo. Para eso, es necesario explicar qué

tipos de datos se pueden llegar a utilizar y qué tipos de datos por el momento no se pueden utilizar por falta de tecnologías capaces de recopilarlos.

Si un Estado quisiese comprender los pensamientos de la población, en la actualidad es algo que no se puede lograr porque no existe forma tal de recopilar ese tipo de información. No hay herramientas capaces de comprender los pensamientos que tiene un ser humano a lo largo del día y por lo tanto ese dato, que, si bien se lo puede considerar existente, es imposible de almacenar en la actualidad.

A pesar de esto, existe una amplia variedad de tipos de datos con los que puede llegar a trabajar una administración pública. A continuación, se van a detallar algunos casos donde se identifican tipos de datos relevantes para el análisis en cuestión.

En primer lugar, están los datos de geolocalización. A lo largo de estos años, la tecnología ha avanzado significativamente y ha logrado tener la posibilidad de recopilar datos de ubicación de teléfonos móviles, servidores de internet, y hasta en algunos casos de personas. Si bien se puede problematizar sobre lo bien o mal que esta usar este tipo de información sobre las personas, tener esos datos permitiría a un Estado tener un control mucho más amplio de los movimientos de su población. A su vez, se pueden sumar los datos que genera una ciudad a partir del uso de los transportes públicos utilizando tarjetas magnéticas que a esta altura del trabajo están bastante arraigadas a la sociedad actual.

El hecho de poder monitorear de forma remota el comportamiento espacial de las personas es un hecho que genera bastante atracción por parte de la administración pública ya que podría ayudar a estimar mejor los horarios del transporte público, mejorar el tráfico de vehículos privados y optimizar la distribución de los desplazamientos que las personas realizan diariamente. Aun así, también puede traer aparejados aspectos negativos como el control excesivo y el seguimiento particular de ciertas personas de interés que se encuentran muy expuestas al tener un dispositivo móvil conectado permanentemente que brinda ese tipo de información.

Por otro lado, un aspecto al cual se le suele dar mucha importancia y se puede notar por la cantidad de trabajos que hablan al respecto son los datos financieros que generan las personas. A medida que fue transcurriendo el tiempo, las sociedades se digitalizaron cada vez más y esto generó que la penetración bancaria sea tan grande que se notan año a año números récords de inclusión financiera en conjunto a la expansión de la industria de las

FinTech. En la actualidad algunos autores afirman que el cruzamiento de datos puede permitir desde la creación de un scoring de crédito a partir de las comunicaciones telefónicas que puede generar una persona (Óskarsdóttir, 2020), hasta modelos de recomendación que surgen a partir del consumo diario de cada usuario. Por este motivo, han surgido también empresas que lucran con este tipo de casos recopilando información de las personas y sus movimientos para venderlo posteriormente en el mercado indicando claramente el valor que estos datos componen.

Otro aspecto que no está tan explotado aun, tal vez por cuestiones éticas y legales es el campo de la medicina y los datos que surgen del estado clínico de las personas que se atienden en sanatorios y hospital públicos y privados. Esta información que a priori talvez cause dificultad para comprender su posible uso puede ser de sumo interés para un Estado dado que puede predecir y estimar enfermedades de rápido contagio en una zona particular. Bajo el contexto actual pandémico que todo el mundo está sufriendo, estos datos pasaron a tomar una importancia mayor pero aún hay mucha incertidumbre sobre la forma de abordarlos. Si bien algunos países tienen este aspecto más solucionado, en Argentina aún hay mucho camino por recorrer y muchas discusiones por dar de parte de los gobernantes y las personas idóneas en el tema. Protocolos para anonimizar datos de sanidad, utilización de blockchain como almacenamiento seguro de información tan sensible, historias clínicas digitales, y otros proyectos más, son los que actualmente se discuten entre los legisladores para eventualmente ir aplicándolos de forma segura y regulada respetando sobre todo la privacidad de las personas.

Otro tipo de dato que puede ser de mucho interés para el Estado son aquellos que refieren al medio ambiente y su cuidado. Como ya bien se sabe, el aspecto medioambiental está en auge por los severos daños que recibe el planeta que habitamos a causa de las formas de producción y consumo que rigen hoy en día. El abanico de datos que se pueden recopilar es muy variado, desde imágenes satelitales para identificar el smog de las ciudades, hasta sensores de ruido en el centro de las urbanizaciones para medir la polución sonora que tanto afecta a las personas en su vida diaria. Estos tipos de análisis que se pueden plantear desde el Estado son muy importantes porque los cambios que se pueden generar hoy en día con respecto a este tema son de gran ayuda para la convivencia en el futuro de las personas y las formas de consumir bienes y servicios. Comenzar hoy en día este tipo de proyectos puede

ser muy beneficioso para las generaciones que siguen y dar estos debates últimamente se está considerando de gran importancia.

Si bien se podría seguir ampliando esta temática nombrando más tipos de datos que pueden ser de interés para un Estado, se va a finalizar el debate hablando de las cámaras de vigilancia que tan de moda están últimamente. Este tipo de tecnológicas han ayudado mucho en varios aspectos a los entes que decidieron aplicarlas, pero más aún se lo vincula con la seguridad peatonal y la seguridad vial. Ya se puede notar en algunas ciudades de Argentina el desarrollo de sistema de vigilancia peatonal a través de una red de cámaras de seguridad monitoreada por un grupo de personas encargadas de hacer los seguimientos. Esto permite no solo actuar de forma rápida ante un siniestro, sino también permite generar movimientos preventivos para la disuasión de posibles hechos de inseguridad. Aun así, otro aspecto que tal vez no esta tan explotado aun, sino que más bien parece ser bastante incipiente dado el poco desarrollo en el tema son las cámaras de seguridad para detectar infracciones viales. Y en este aspecto es donde más se va a hacer hincapié dado que el trabajo va a comparar modelos de aprendizaje automatizado capaces de identificar infracciones viales de forma temprana. Si bien hay algunos casos de uso más interesantes que otros, un desarrollo a destacar podría ser la identificación de patentes de autos a la entrada de las urbanizaciones para identificar vehículos que están en infracción por diversos motivos. Esta aplicación es muy novedosa porque permite automatizar un proceso que hasta hace pocos años era totalmente manual y, además de requerir muchos recursos humanos para llevarse a cabo, perjudicaba el libre flujo del tránsito haciendo más comunes los atascos y embotellamientos. A pesar de los aspectos positivos que se le pueda encontrar a este tipo de proyectos, no hay que dejar de lado la problematización sobre la privacidad de los datos que se recopilan y como va a ser su tratamiento. Por más que el desarrollo tecnológico permita la identificación de placas de automóviles, hay que tener mucho cuidado en la forma en que se procesan los datos, y sobre todo el uso que se le va a dar a la información.

#### **1.4. Uso responsable y privacidad a través de la reglamentación**

El concepto de lo público y lo privado fue cambiando a lo largo del tiempo. Si bien hay autores que proponen un concepto de privacidad intrínseco del ser humano y muy relacionado al uso del bien en cuestión (Locke, 1689), en la actualidad estas visiones fueron cambiando principalmente porque el mundo fue cambiando y la manera de consumir que tiene el ser humano es totalmente diferente a lo que sucedía hace cientos de años atrás. La propiedad privada estaba muy vinculada a los bienes que el ser humano era capaz de consumir y que a su vez tenían un cierto valor económico como lo puede ser un lote o una manzana. Con el avance de la tecnología, se empezaron a dar discusiones más profundas sobre la propiedad privada de lo intangible. ¿Es una marca propiedad de alguna persona en particular? ¿El concepto de panóptico que establece Foucault en sus escritos le pertenece? Todas estas cuestiones en su época fueron disruptivas y por lo tanto en algún momento los gobernantes de los Estados modernos tuvieron que reglamentar estas situaciones para darle un marco teórico a estas problemáticas.

A lo largo de los años, la temática de la privacidad fue avanzando cada vez más identificando países pioneros en el tema que contribuyeron de gran manera a que otros países tomen medidas similares. De a poco los conceptos intangibles como son los datos en la gran mayoría de los casos fueron tomando relevancia y lentamente fueron surgiendo las normativas. Como suele pasar en la mayoría de los temas disruptivos, los países hegemónicos fueron aquellos que comenzaron el debate reglamentando el uso y la privacidad de los datos montando así las bases para la problematización sobre el tema. Seguramente, en el momento que se comenzó a debatir al respecto, no se tenía una visión tan acabada de cómo iba a resultar la explotación de los datos a nivel mundial, es por eso que estas discusiones son más bien dinámicas y variantes a lo largo de los años. No se puede pretender que las mismas leyes de protección de datos personales sigan vigentes a medida que se profundiza el uso de las redes sociales, el uso de los teléfonos móviles y el acceso a internet de forma universal generando conectividad en gran parte del planeta.

Dicho esto, el caso de Argentina en América Latina es destacable dado que fue uno de los primeros en sancionar una ley que reglamente la privacidad de los datos a través de la Ley N° 25.326 llamada “Protección de los Datos Personales”. Esta ley entre otras cosas permitió enmarcar las cuestiones legales bajo un mismo criterio protegiendo de esa forma a los consumidores y generando incentivos formando reglas claras para el uso de datos privados.

Otro apartado para destacar es el habeas data que se menciona en el artículo 43 de la Constitución Nacional Argentina donde se menciona que “Toda persona podrá interponer esta acción para tomar conocimiento de los datos a ella referidos y de su finalidad, que consten en registros o bancos de datos públicos, o los privados destinados a proveer informes, y en caso de falsedad o discriminación, para exigir la supresión, rectificación, confidencialidad o actualización de aquéllos. No podrá afectarse el secreto de las fuentes de información periodística.” (Argentina C. d., 1994).

Teniendo en cuenta ambos aspectos, todo el territorio argentino está contemplado bajo esta ley de protección de datos personales y este artículo de la constitución nacional a pesar de que también algunas provincias o ciudades autónomas ya se manifestaron independientemente con una diversidad de normas que complementan ciertos aspectos de las dos reglamentaciones mencionadas previamente. Todas en su conjunto fueron haciendo que el panorama tan dinámico que se mencionó pueda ser esclarecido atacando necesidades de la actualidad para lograr un mejor funcionamiento de la sociedad en lo que respecta a esta temática. Además, dentro de la misma ley nacional de datos, se incorporó la creación de la Dirección Nacional de Protección de Datos Personales (PDP por sus siglas en español) que continua vigente hoy en día dependiendo del Ministerio de Justicia y Derechos Humanos en el país.

Como ya se ha mencionado previamente, existen diferentes tipos de datos y por consecuente, también existen diferentes formas de tratar cada uno de ellos legalmente. Si bien se podría ejemplificar cada uno de los casos donde esta ley nacional hace énfasis, se pretende tratar el caso particular de las imágenes de video captadas a través de cámaras de video filmación. Para este caso en particular donde las personas se pueden ver expuestas a filmaciones por parte de terceros, la PDP se ha involucrado mientras el auge de videos de vigilancia se hacía notar sobre todo en grandes ciudades, y en conjunto con el Gobierno Nacional lanzaron la Disposición N°10/2015 donde se regula el uso de este tipo de datos. En pocas palabras, toda recolección de imágenes digitales de las personas a través de cámaras de seguridad debe contemplar un consentimiento previo del titular del dato teniendo 3 excepciones:

- Cuando las imágenes provienen de un evento privado.
- Cuando las imágenes las realice el Estado bajo alguna de sus funciones.

- Cuando las imágenes son captadas en un perímetro privado sin invadir el espacio público o de terceros salvo que ello sea inevitable; en ese caso se deberá restringir la captación de imágenes al mínimo.

Como se puede notar, la Disposición deja mucho a interpretación y es ahí donde comienzan los problemas dado que en algunos casos se puede encontrar una empresa privada que brinda servicios públicos como lo pueden ser colectivos o metros filmando a los pasajeros con el fin de evitar hechos de inseguridad entre otras cosas. Ese tipo de cuestiones se tienen que analizar aparte porque no son del todo comprendidas por la Disposición mencionada dado que nadie hace firmar a los pasajeros la voluntad de estar expuestos a la recolección de imágenes digitales.

Mas allá de estos inconvenientes, en el mundo se está planteando la calidad de la segunda excepción que menciona la Disposición generada en 2015 en Argentina. El rol que toma el Estado puede ser muy amplio y hasta en algunos casos puede confrontar con otro tipo de derechos que entran en conflicto con este tipo de soluciones legales. Particularmente en China, donde para el año 2018 ya contaban con más de 170 millones de cámaras de videovigilancia (Gershgorin, 2021), estas discusiones se están tornando muy importantes y se extienden a otros territorios donde el control del Estado está cumpliendo un rol similar. La gran discusión que ronda esta temática está dada por si corresponde que el Estado sepa permanentemente que está haciendo la población y peor aún, si puede llegar a utilizar dicha información para amedrentar a ciertos sectores reactivos de la población. En adición a este aspecto, también cabe destacar que los datos que reúnen este tipo de gobiernos no solo son útiles para monitorear físicamente a la población, sino que también entrecruzan datos de otros rubros pudiendo así generar métricas de score crediticio (Creemers, 2018), multas por incumplimientos de normas locales entre otras cosas. Cuando se suman todas estas cuestiones, el planteamiento del problema se hace aún más profundo y requiere de un accionar prudente y consensuado con la población para que tanto los generadores de datos en el aspecto más crudo que son las personas, y los gobiernos que hacen uso de ellos puedan coexistir y tengan incentivos a generar más y mejores servicios para simplificar la vida de la población generando un equilibrio en el sistema entre protección de los datos y la privacidad individual. Este tema planteado en tan pocas líneas es un debate que tiene que ser dado en un trabajo completo al respecto ya que las aristas por donde se puede tratar el tema son

múltiples y hay una necesidad de considerar muchos más factores que los recientemente mencionados.

## **2. Detección y clasificación de objetos bajo un contexto de ciudades inteligentes. Análisis de diferentes modelos**

Si bien el proceso de detección y clasificación de objetos está en auge en la actualidad donde el Internet de las Cosas está penetrando fuertemente en la sociedad, es un concepto que hace ya muchos años se viene utilizando pero que a su vez fue mutando en este último tiempo.

Hace algunos años no era muy novedoso que exista algún equipo de trabajo vinculado sobre todo al área de seguridad en entidades públicas o privadas que se dedicaban a analizar videos e identificar situaciones de riesgo. Para ejemplificar esta situación se puede nombrar el caso de los aeropuertos que desde hace muchos años utilizan sistemas de seguridad que monitorean los movimientos de los pasajeros en tránsito identificando equipajes, personas y elementos de interés. Estos casos podrían ser tal vez uno de los primeros usos prácticos que se podían encontrar en la temática abordada.

Aun así, con el paso del tiempo y la irrupción de la tecnología, estos procesos se comenzaron a automatizar cada vez más hasta la actualidad donde algunas regiones del mundo están siendo monitoreadas permanentemente.

En este apartado se va a hacer énfasis en el desarrollo teórico de algunos algoritmos para cumplir con un objetivo particular de identificar un algoritmo útil para el procesamiento de imágenes dentro del contexto de una ciudad inteligente.

La automatización de la detección y clasificación de objetos en la actualidad comenzó a verse beneficiada por el uso de las nuevas tecnologías como el mayor poder de procesamiento de las computadoras y su rapidez a la hora de analizar la información. Esto hizo que este proceso vuelva a estar en boga, y por lo tanto los desarrollos tecnológicos alrededor de esta temática fueron creciendo. Casos como el monitoreo de pasajeros en el transporte público, identificación de riesgos de trabajo en una planta de manufacturas, o

también la captación de placas de automóviles al ingreso de una ciudad, son una parte de la amplia variedad de ejemplos que se pueden ver en la actualidad al respecto.

A pesar de que para el consumidor final de este tipo de herramientas tal vez parezca un desarrollo fácil y sencillo de implementar, hay una variedad de situaciones complejas a solucionar antes de decidir aplicar este tipo de procesos a un ambiente en particular. Los algoritmos que estos sistemas comprenden suelen ser muy complejos y requieren de mucha información para ser llevados a cabo.

### **2.1. Algoritmos de detección y clasificación de objetos**

Según Nosson S. Yanofsky, se puede definir informalmente un algoritmo (Yanofsky, 2010) como cualquier proceso computacional bien definido que toma un valor cierto o un conjunto de valores como campo de entrada, y produce un cierto valor o serie de valores como su conjunto de salida. Si bien esta definición parece ser un poco abstracta y no apunta específicamente al tema que se está abordando, cuando se la relaciona con términos matemáticos y de cálculo, comienza a tener más sentido su aplicación.

Los algoritmos llevados al ambiente de la computación se pueden identificar como un proceso generalizado donde la computadora recibe un conjunto de datos de entrada para lograr un conjunto de salida a través de ciertos mecanismos internos. Esto es importante destacarlo porque en los próximos apartados de este trabajo se va a hacer mucho hincapié en el funcionamiento de ciertos algoritmos y las ventajas y desventajas que cada uno trae aparejado.

El hecho que dos algoritmos reciban el mismo conjunto de entrada y tal vez tengan el mismo conjunto de salida, no nos habilita a pensar que ambos procesos funcionan de la misma forma. Es decir, si bien los algoritmos están fuertemente influenciados por los conjuntos de entrada, los procesos internos que tienen cada uno de ellos los pueden diferenciar por completo. Esto hace que la gran diferenciación de los algoritmos por analizar en este trabajo se centre en identificar los procesos internos y caracterizar su comportamiento ante el conjunto de entrada. Esto hará que se pueda llegar a un buen nivel de comprensión de cada uno de los algoritmos y poder así caracterizarlos de forma más precisa.

A lo largo de toda la historia, siempre se quiso comprender como funciona el cerebro humano dado que implica una complejidad tal que a veces es muy difícil de comprender su funcionamiento. Desde la acumulación de recuerdos, movimientos corporales o hasta los sentimientos que el humano capta a través de los sentidos, todos ellos fueron parte de una amplia gama de situaciones que desde un inicio fueron procesos muy difíciles de comprender. A medida que fue pasando el tiempo y en conjunto con los avances tecnológicos, el ser humano estuvo cada vez más cerca de comprender la totalidad de las acciones que puede realizar con el cerebro (Stuss, 2000). Dicho esto, se puede indicar que el cerebro humano funciona a través de conexiones de neuronas que transmiten impulsos para lograr una acción que puede ir desde el movimiento de una extremidad a la sensación de frío cuando se expone a un clima hostil.

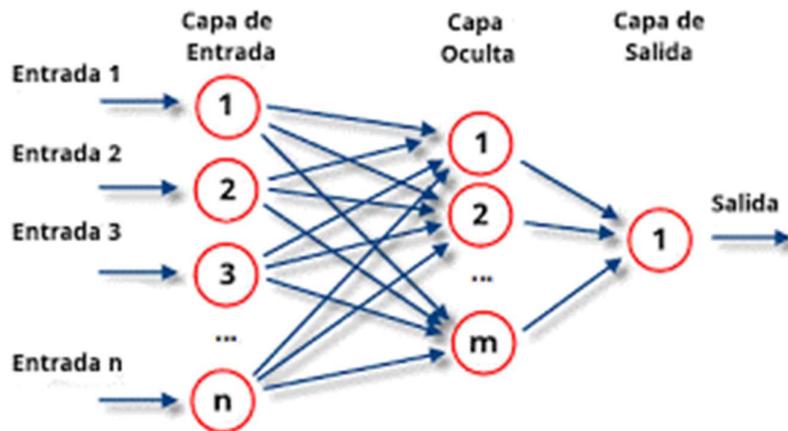
Haciendo una analogía con el termino de algoritmo que se mencionó anteriormente, no parecería estar muy alejado de lo que realizan estos procesos bajo un cierto contexto. Considerar una estimulación de bajas temperaturas como un conjunto de entrada y realizar un proceso interno que indique que el usuario tiene frío, podría asemejarse mucho al funcionamiento de un algoritmo básico donde todos los componentes están dados. Es por eso que desde hace ya muchos años algunos autores han introducido el concepto de redes neuronales en el campo de la estadística y de la programación (Pitts, 1943).

Si bien las primeras menciones al respecto fueron algo incipientes, cabe destacar que el panorama cambió radicalmente con la incorporación de las nuevas tecnologías. Hace ya muchos años se viene hablando de algoritmos muy complejos que simulan la interconexión neuronal, pero eran de poca aplicabilidad por la falta de poder de cómputo que hasta ese momento solo las personas eran capaces de tener. Con el avance de la tecnología esto fue cambiando y hoy en día, existen computadoras tan potentes que pueden replicar ciertos procesos que realiza el ser humano con su cerebro utilizando sus neuronas.

Para diagramar el funcionamiento de una red neuronal es fundamental comprender que existe una capa de entrada y una capa de salida con un numero cierto de neuronas cada una de ellas. En el medio de ambas capas están las capas ocultas que pueden tener un numero variado de neuronas y la existencia de estas capas son las que brindan mayor complejidad al proceso. El diagrama queda representado a continuación:

**Figura 1.**

Diagrama de red neuronal



*Nota:* El grafico representa el diagrama de una red neuronal artificial. Tomado de *Desarrollo de la inteligencia artificial*, por Paola Itzel, 2020, Timetoast.

Se puede notar que cada una de las neuronas están conectadas con todas las combinaciones posibles de neuronas que existan en las capas anteriores y posteriores simulando así la interconexión neuronal del ser humano. Esto está indicando que va a existir algún tipo de relación entre las capas a través de las neuronas dado que la capa de entrada va afectando hacia su derecha al resto de las neuronas para llegar finalmente a la capa de salida con la ansiada respuesta.

Para entender mejor su funcionamiento hay que detallar que tipo de relación tienen las neuronas entre sí. Para ello, es necesario comprender que existen pesos vinculados a cada una de las neuronas de la capa de entrada. Esos pesos van a ser utilizados para que mediante una función de activación se logre un resultado que termine siendo dirigido hacia la capa siguiente, y así sucesivamente.

Dicho esto, cabe destacar que la importancia de la función de activación es fundamental dado que va a ser la metodología con la cual se aplican los pesos para generar los resultados. Si se pensase sobre una función de activación del estilo lineal podría decirse que las redes neuronales funcionan como una regresión lineal donde los pesos son los coeficientes de cada una de las variables. Si bien en parte esto es cierto, la ventaja de usar redes neuronales más complejas es que pueden saltar la linealidad en sus modelos a través del cambio de la función

de activación ya que la concatenación de muchas regresiones lineales da como resultado una regresión lineal. Es por este motivo que se utilizan funciones de activación alternativas como la sigmoide, RELU, o escalonada entre otras permitiendo así la generación de regresiones no lineales.

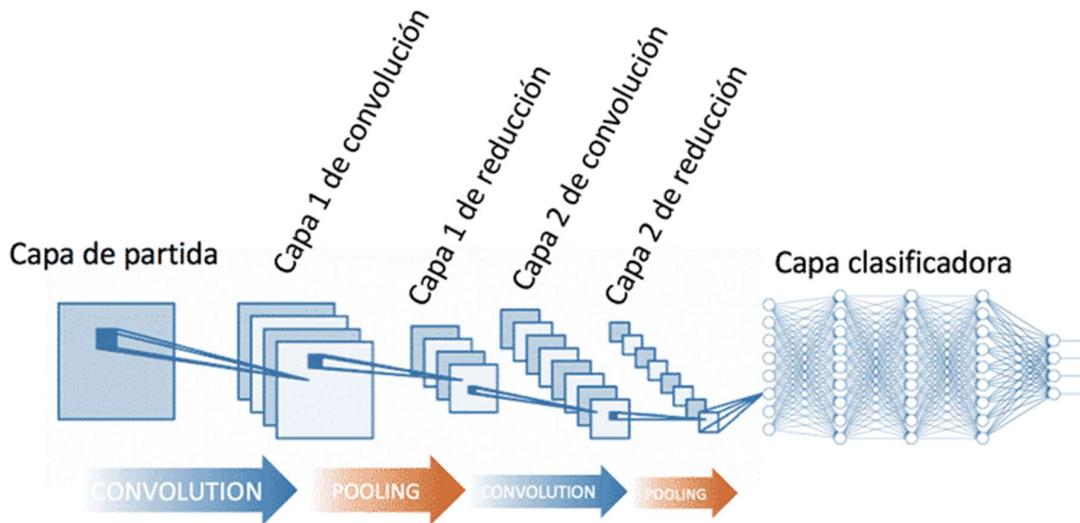
Un último aspecto para destacar es la forma de aprendizaje que tienen las redes neuronales artificiales. Para eso se utiliza el sistema de propagación hacia atrás, que consta de notificar el error alcanzado en la capa de salida y hacer modificaciones en los pesos de forma retrospectiva. Es decir, si en un momento se tiene una observación donde cada uno de los valores de sus atributos ingresan en la capa de entrada y luego de iniciar con unos pesos al azar para cada uno de los valores se le aplica la función de activación seleccionada obteniendo así el resultado en la capa de salida, el sistema arrojará un error con respecto al valor real de esa observación. En ese caso, el sistema de propagación hacia atrás va a ir modificando los pesos del último al primero para lograr una mejora en ese nivel de error hasta achicarlo lo máximo posible. De esa forma es como se va a desarrollar la tarea, y consecuentemente el sistema va a haber aprendido después de tanto equivocarse y luego generará pesos que van a dejar de ser al azar para contener mucha información sobre cómo lograr una capa de salida adecuada.

Por otro lado, las redes neuronales convolucionales son un tipo de red neuronal que ganó mucho protagonismo desde el nuevo auge de la visión artificial dado que aplican justamente para este tipo de problemas donde se desean analizar imágenes digitales. A diferencia de las redes neuronales convencionales, las redes neuronales convolucionales aplican una nueva capa de convolución generando así un filtro que afecta según como están ordenados los valores. No le va a ser indiferente encontrar valores en una cierta posición dado que según como estén posicionados, el filtro va a actuar de una forma u otra. En el caso específico de las imágenes, esto permitirá identificar contrastes, cambios de colores, texturas entre otras cosas.

Si bien las redes neuronales convolucionales están dentro del mercado hace ya unos años, la adopción de estas ha generado nuevos desarrollos que profundizan más aun estos casos y permiten realizar actividades de mayor complejidad siguiendo este mismo análisis. En el presente trabajo varios de los métodos que van a ser presentados para la identificación temprana de infracciones viales recurren a estos algoritmos para llevar a cabo las tareas.

**Figura 2.**

Diagrama de red neuronal convolucional



*Nota:* El grafico representa el diagrama de una red neuronal artificial convolucional identificando las capas de convolución y reducción. Tomado de *Detección Automática de Órganos en adquisiciones de Tomografía Computarizada con Métodos de Machine Learning*, por Rafael Lopez Gonzalez, 2017, Universitat Politècnica de València.

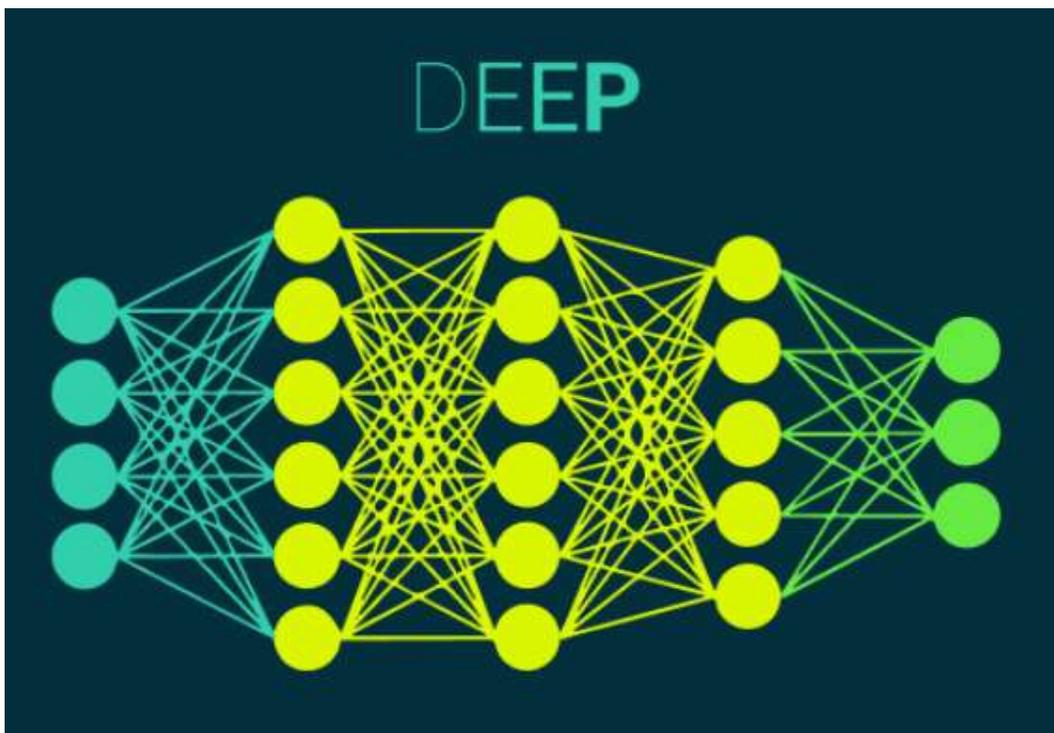
Por último, existe otro tipo de red neuronal que se caracteriza por su composición de capas ocultas y las interconexiones que tienen sus neuronas. Cuando una red neuronal comprende muchas capas ocultas, se la denomina red neuronal profunda porque analiza muy profundamente todas las combinaciones posibles entre cada una de las neuronas.

Como ya se analizó anteriormente, el hecho de no utilizar capas ocultas hace que solamente se tenga una forma de segmentar la información porque solo se genera una relación entre la capa de entrada y la capa de salida. En cambio, si se utilizan muchas capas ocultas, esto generará una amplia gama de posibilidades para segmentar las observaciones según el objetivo que se quiera tomar. Aun así, estos algoritmos son tan complejos que tienen dos aspectos muy negativos, uno es el alto poder de cómputo que se requiere para procesar toda la información bajo este diagrama, y el otro es la falta de capacidad para explicar cómo el sistema está tomando las decisiones. Estos algoritmos si bien pueden ser muy buenos desde

el punto de vista empírico, carecen de la validez para explicar los resultados dado que las capas ocultas se llevan mucho de esa argumentación. Es por eso que se suelen aplicar en procesos muy particulares donde hay una falencia para obtener los resultados dado la complejidad de la observación.

**Figura 3.**

Diagrama de red neuronal profunda



*Nota:* El grafico representa el diagrama de una red neuronal artificial profunda. Tomado de *Introducción al Deep Learning I: Funciones de Activación*, por Mario Campos Soberanis, 2020, Medium.

## 2.2. Métrica de error sobre el criterio de selección

En el universo de la ciencia de datos, existe una amplia variedad de mecanismos que se pueden utilizar para realizar una medición de los errores cometidos en las inferencias

estadísticas. Si bien esta temática es muy amplia y se podría hacer un trabajo puramente al respecto, la idea de este apartado es mencionar cuáles son las métricas de error que se van a utilizar en este trabajo brindando más información sobre qué significa cada una y problematizando sobre el uso apropiado de la métrica.

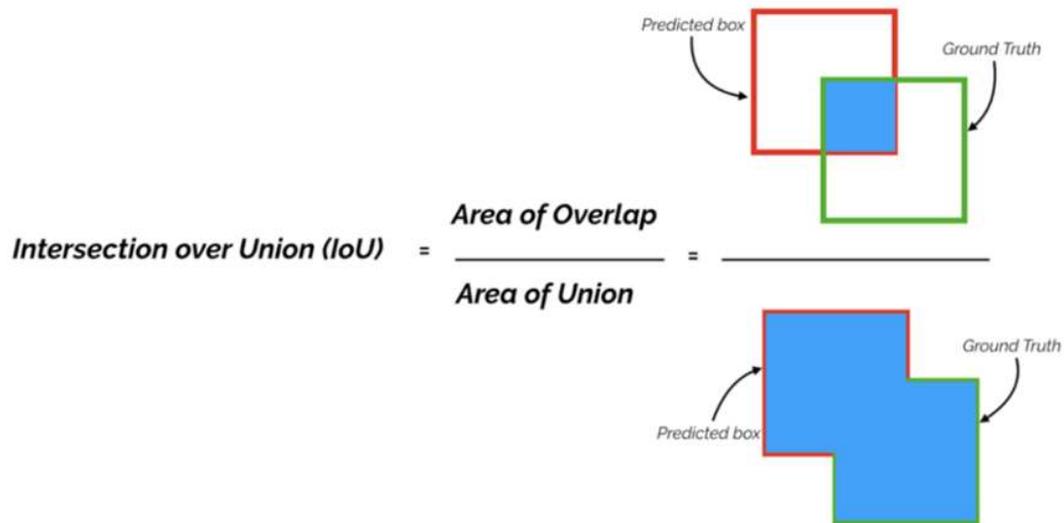
En un problema de visión artificial hay dos aristas fundamentales sobre cómo se debe analizar la inferencia. Por un lado, está la detección del objeto que consiste en identificar la locación de este dentro de la imagen, pero por otro lado está la clasificación de este. En la mayoría de los casos son problemas que se tratan al mismo tiempo porque están muy relacionados entre sí.

Inicialmente, a la hora de verificar la locación del objeto, pueden existir una variedad de errores que van desde la no detección del objeto hasta la mala detección de este. Son dos casos que, si bien parecen ser similares, distan mucho uno del otro. En el caso de la falta de detección, el algoritmo está fallando en ubicar al objeto dentro del contexto de la imagen por lo tanto para el algoritmo no existe dicho objeto cuando analiza la situación porque no lo pudo ubicar en el espacio. Por el otro lado, cuando se localiza erróneamente un objeto, se refiere a los casos donde, si bien el algoritmo interpretó que hay un objeto en la imagen analizada, no pudo detallar con precisión que área comprende dicho objeto dentro del contexto de la imagen. A este tipo de error se lo suele llamar área de intersección de unión o IoU (Intersection of Union) por sus siglas en inglés.

Como se puede notar se tratan de errores que no pueden ser analizados en conjunto ya que sus dominios son muy diferentes porque la no detección de un objeto es simplemente una variable binaria donde se puede o no encontrar al objeto, pero cuando se trata del IoU el dominio de error cambia totalmente brindando una variable continua que identifica el error. Mas específicamente, el IoU se genera a través de la división de dos áreas resultantes de la inferencia. En el numerador se encuentra la intersección entre el área de la inferencia y el área del objeto real, y en el denominador se encuentra el área comprendida por ambas localizaciones en conjunto. Esto hace que la formula se vea de la siguiente forma:

#### **Figura 4.**

Diagrama de métrica de intersección sobre la unión



*Nota:* El grafico representa el diagrama de la métrica de intersección sobre la unión para medir la precisión de una identificación de objeto. Tomado de *mAP (mean Average Precision) might confuse you!*, por Shivy Yohanandan, 2020, TowardsDataScience.

A medida que el IoU comienza a estar cercano al valor unitario, significa que la localización del objeto es excelente y ubica el objeto exactamente dónde está representado en la imagen. En cambio, si ese valor se aleja del valor unitario, significa que está fallando en mayor o menor medida al ubicar al objeto dentro de la imagen.

A pesar de esto, aun ya habiendo localizado el objeto a través de una tarea que requiere mucha capacidad y poder de cómputo, resta la segunda tarea a realizar que consta de identificar al objeto localizado. Esto significa identificar al objeto con algún rotulo que lo clasifique y diferenciarlo del resto de los objetos. Usualmente este proceso se realiza utilizando una cantidad de etiquetas fijas dado que el universo de identificación si no es acotado puede extenderse a millones de opciones posibles y no es la idea de este trabajo. Para este caso en particular se desea identificar solamente los vehículos que están vinculados al tránsito y a la identificación temprana de infracciones viales.

Cuando la tarea de clasificación finaliza, el algoritmo puede estar muy seguro de que dicho objeto se trata de un auto, por ejemplo, pero puede haber casos donde el algoritmo no tiene muchas herramientas para asegurar de forma univoca la clasificación del objeto. Es por eso que se introduce la variable confianza en este campo dado que, si bien se desea saber dónde

está localizado el objeto con un cierto margen de error, también se desea saber qué tipo de objeto es con otro margen de error adecuado. En este caso la confianza va a indicar que tan confiable es la identificación del objeto por parte del algoritmo.

De esta forma, no va a ser indiferente para medir el desempeño del modelo el grado de confianza ya que lo deseable es que ese valor sea lo más cercano a 100% que sería el valor máximo donde la clasificación del objeto es 100% confiable y no hay margen de error. A medida que la métrica se vaya alejando de ese valor, está indicando que no es tan confiable identificar a ese objeto bajo esa clasificación.

Otra métrica de suma importancia será la precisión (Jonathan, 2018) del modelo dado que va a ser una de las claves para poder tomar la decisión de seleccionar uno de ellos por sobre los otros. En este caso la precisión está identificada como la métrica que indica el índice de verdaderos positivos sobre la suma de todos los casos positivos que existen en la imagen. Este indicador es posteriormente va a ser utilizado en conjunto con la métrica de recall.

En el campo de la estadística, el recall es una métrica que indica que tan bien fueron identificados los casos positivos dentro de la imagen teniendo en cuenta los niveles de confianza encontrados. Este índice está compuesto por el cociente entre los verdaderos positivos sobre la suma de los verdaderos positivos y los falsos negativos.

Conjuntamente entre la precisión y el recall se va a poder identificar al promedio ponderado de la media de precisión (mAP por sus siglas en inglés) que se identifica como una métrica relevante a la hora de analizar la comparación de modelos de identificación de objetos. Esta misma se define como el promedio ponderado del área bajo la curva de la relación entre precisión y recall para cada una de las observaciones. Cuando esta métrica esté más cerca del valor unitario, va a identificar un proceso de identificación de objetos muy bueno ya que no solo va a identificar con mucha precisión a los objetos, sino también va a fallar muy poco a la hora de confundirse casos positivos con casos negativos.

### **2.3. Aprendizaje de transferencia y sus beneficios**

El concepto de aprendizaje de transferencia es muy nuevo y hace pocos años que se está utilizando. Se trata de un atajo a la hora de utilizar una red neuronal secuencial para que los pesos iniciales no sean asignados de forma azarosa, sino que ya comprendan algún tipo de

aprendizaje previo. Estos tipos de mecanismos son de mucha utilidad porque no solo permite usar el aprendizaje de otro proceso, sino también que ahorra mucho tiempo de aprendizaje con respecto a la situación donde uno lo quiere hacer por su propia cuenta. Son estrategias que comúnmente se desean aplicar en caso de ser posible porque tienen muy pocas contraindicaciones.

Teniendo en cuenta la utilidad de esta estrategia, se dispara una bifurcación acerca de su aplicación. Están aquellos procesos que utilizan los pesos de una red neuronal ya entrenada para aplicar ese conocimiento a nuevos grupos de datos, y también están aquellos procesos donde se utilizan dichos pesos para no comenzar a aprender de forma azarosa sobre el nuevo conjunto de datos. Vale la pena destacar esta diferencia, porque puede cambiar rotundamente el desempeño de un modelo si se opta por una opción u otra. Probablemente el hecho de entrenar sobre un nuevo conjunto de datos utilizando los pesos ya predeterminados pueda dar lugar a mejores resultados dado que el aprendizaje todavía continúa existiendo, aun así, esto no siempre concluye de esa forma.

Por otro lado, cuando se inicia un proceso de esta magnitud, hay que asegurarse que los pesos de la red neuronal previamente entrenada y su aprendizaje tenga relación con el nuevo conjunto de datos que se desea analizar. Si por ejemplo en este caso se desea identificar vehículos de gran tamaño, sería de gran utilidad que los pesos del aprendizaje por transferencia hayan sido entrenados viendo variedad de vehículos de gran tamaño, porque si su aprendizaje fue realizado tomando como observaciones pelotas de tenis, va a ser de poca utilidad los pesos resultantes del modelo ya entrenado.

Si bien esto último es cierto y requiere de gran atención previo al inicio del nuevo proceso, también cabe destacar que en lo que respecta a imágenes, talvez los pesos de la red neuronal ya entrenada sepan diferenciar texturas y contrastes a pesar de estar mirando observaciones muy distantes entre cada grupo. Es por eso que tampoco se desestima completamente utilizar pesos surgidos del proceso de mirar otro tipo de observaciones.

## **2.4. Codificación de imágenes. La percepción del humano y de las computadoras**

Si una persona se pone a analizar profundamente como está compuesta una imagen digital llegaría a la conclusión de que es un archivo digital de cierto formato compuesto por ceros y unos. Si bien esta conclusión parece muy lejana a lo que realmente se ve cuando se tiene en frente una fotografía, por ejemplo, el proceso de segmentación de la imagen en miles o millones de puntos de colores es lo que sucede realmente, pero el ojo humano algunas veces no es capaz de identificarlos, sino que la imagen simula perfectamente un entorno de continuidad donde es muy difícil diferenciar un punto de otro.

Sabiendo ya que todas las imágenes por más grandes o chicas que sean están compuestas por puntos o recuadros de colores, se podría decir que la nitidez de la imagen va a estar fuertemente relacionada al tamaño de cada uno de esos recuadros porque no va a ser lo mismo utilizar 100 recuadros de colores para ver una figura de 10cm de alto por 10cm de ancho que utilizar 1.000 recuadros para ver lo mismo. A medida que más recuadros tenga, más nitidez tendrá la imagen sobre el mismo plano. Esos recuadros se los denomina píxeles, y por lo tanto para analizar la nitidez de la imagen es una característica que suele ser tenida en cuenta en conjunto con el tamaño final.

El peso digital de ese archivo va a estar dado en gran medida por la cantidad de píxeles que contenga la imagen porque cada píxel tiene su propio valor numérico que es traducido en un cierto color. Y a su vez, ese color que está siendo representado en el píxel, suele ser la combinación de los tres colores primarios que dan como resultado el color original, es decir, la combinación de cantidades de rojo verde y azul que dan como resultado el color que se desea representar. De esa forma, se tiene que una imagen va a estar compuesta por la cantidad de píxeles en el ancho y en el largo de la imagen, pero a su vez por la composición de cada uno de los píxeles identificando que color contiene cada uno.

Una vez que se desgrana la imagen como se mencionó anteriormente, ya se está en condiciones de analizarla pixel por pixel para poder asemejarlo a una tabla de valores de 3 dimensiones, el ancho, el largo, y la profundidad según el color que comprenda. En este entorno es que las computadoras pueden ser de utilidad dado que les es muy simple poder identificar este tipo de información organizada de esa forma y por lo tanto también les va a ser sencillo aplicar ciertos procesos sobre esa información ya desgranada.

Si bien este tipo de codificación está vigente hace muchos años, la capacidad de las computadoras para procesar semejante volumen de información no es tan antigua dado que,

hasta hace poco, mucho de los procesos que se ven hoy en día serían inviables. Además, a pesar de que las computadoras hubieran tenido ese poder de procesamiento tan deseado, también hubiese hecho falta la capacidad de almacenamiento de grandes volúmenes de datos para aprender de ellos. Y es ahí donde nace el proyecto de image net en el año 2009 abriendo una gran base de datos de libre acceso para fomentar el desarrollo de la visión artificial que era incipiente en esa época (Jia Deng, 2009). A partir de ese hito, la actividad creció ampliamente generando cada vez más y mejores modelos que trabajen con el análisis de imágenes.

A lo largo de la historia siempre se quiso entender el funcionamiento del cerebro humano. Comprender como funciona la visión del ser humano a partir de los impulsos de su cerebro es algo totalmente avanzado, pero a su vez se trata de una tarea que se está estudiando hace ya mucho. En el año 1962, el neurofisiólogo canadiense David Hunter Hubel y el neurobiólogo sueco Torsen Wiesel, ambos ganadores del premio nobel de medicina en 1981, estaban experimentando en animales la forma de captar los impulsos neuronales que afectaban a la visión. Es por eso que hicieron un experimento con un felino al cual le medían la frecuencia de los impulsos nerviosos de ciertas zonas del cortex y a la vez le mostraban imágenes para ver como ambas situaciones se desenvolvían. Bajo este experimento pudieron identificar que ciertas zonas del cerebro se activaban cuando el animal veía líneas horizontales, pero no así con líneas verticales. De forma análoga notaron que otra parte del cerebro se activaba cuando veía solamente líneas verticales. A partir de esto establecieron una hipótesis que tenía mucho sentido diciendo que el cerebro interpreta la imagen por partes y no toda en un solo proceso.

Luego de haber transcurrido algunos años, esta hipótesis fue tomando cada vez más fuerza dado que también se verificó que el ser humano actúa de forma muy similar y también procesa de la misma forma las texturas, los contrastes y las líneas verticales y horizontales (Gagnon, 2001). Dicho se puede decir que el ser humano tiene su cerebro compuesto por partes que están interconectadas pero que cada una cumple una cierta función para luego poder interpretar una fotografía, por ejemplo. Lo mismo sucede con las otras funciones que tiene el cerebro como la personalidad, los gustos y preferencias, dado que con el tiempo se comprobó que ciertas partes del cerebro se dedican a tareas muy específicas y es por eso que

la computación al iniciar un camino hacia la visión artificial decidió tomar la misma ruta (A., 2012).

Habiendo visto cómo funciona un cerebro humano a la hora de interpretar una imagen, al crear el concepto de visión artificial se decidió simular la forma en que el humano capta las imágenes a través del cerebro. Es decir, si la visión ocular depende de zonas del cerebro que identifiquen características de una imagen, la visión artificial va a generar artificialmente zonas que se especialicen en identificar ciertas características de la imagen. Hasta el momento este es el paradigma que reina en el ambiente.

Para llevar a cabo este proceso, es de suma importancia la adopción de redes neuronales y sobre todo aquellas que aplican al menos un proceso de convolución dentro del mecanismo. La capa de convolución va a generar zonas del modelo que sean específicas para identificar ciertos aspectos de la imagen tal cual lo hace el ser humano dado que la convolución actúa como un filtro para respetar la estructura de la imagen. De hecho, si se quisiera ver cada capa únicamente, se podría apreciar que una de las capas solo identificó bordes, otra capa solo identificó contrastes y así sucesivamente. De esta forma se logra identificar a un objeto en su totalidad pudiendo diferenciar muchas características de este a través de cálculos matemáticos generados por la red neuronal.

Sin la codificación de imágenes que reina en la actualidad y estos últimos avances tecnológicos, sería muy difícil lograr que una máquina artificialmente pueda identificar objetos a través de un lente. Es más, si se piensa en el pasado tal vez haya sido muy improbable contar con un contexto como el actual donde este campo en particular crece día a día. Aun así, como en todo evento tecnológico, no se está exento a nuevas herramientas y cambios de paradigma que el futuro pueda traer consigo.

## **2.5. Descripción de modelos de detección y clasificación de objetos**

A lo largo de los años el ser humano ha tratado con muchos modelos aplicados en diversas áreas que pueden ser muy diferentes entre sí. Es que un modelo, refiere al conjunto de normas aplicables a un proceso dado, para lograr un resultado final que permanecerá invariante siempre y cuando el contexto siga siendo el mismo.

En lo que respecta a la ciencia de datos, el modelaje de ciertas situaciones es una de las bases fundamentales del avance científico porque muchas veces se quiere replicar una situación en un contexto y para eso hacen falta modelos que simulen una realidad ficticia. Es por eso que a continuación, se van a presentar 3 modelos diferentes donde todos ellos apuntan al mismo objetivo, pero las reglas y los procesos van a ser algo diferentes entre sí.

### **2.5.1. Single Shot MultiBox Detector**

El modelo Single Shot MultiBox Detector (SSD) fue creado en el año 2015 (Liu, 2016). El objetivo del trabajo fue desarrollar un método capaz de identificar objetos en una misma imagen digital utilizando una sola red neuronal profunda. Además de esto, el método aplicado puede trabajar con múltiples tamaños de imagen lo cual hace que sea muy versátil al momento de la aplicación de este.

Como toda red neuronal, este proceso fue entrenado para que el algoritmo aprenda a actuar en ciertas situaciones y replicar lo aprendido en nuevas imágenes, y para eso fueron utilizadas las bases de datos de imágenes de PASCAL VOC, MS COCO, y ILSVRC. El resultado que generó este modelo fue muy bueno dado que no solo mejoró en algunos puntos la precisión con respecto a modelos similares, sino que también pudo actuar de forma más rápida permitiendo mejorar en terminos de tiempo y en terminos de resolución de imagen.

Al momento de publicar estos avances, esta nueva tecnología tuvo mucho éxito porque fue comparada con el estado actual del arte de su momento de las redes neuronales convolucionales basadas en regiones. Este tipo de redes neuronales eran las herramientas más avanzadas en su momento y el modelo de Single Shot MultiBox Detector tuvo rendimientos muy similares por lo cual fue muy tenido en cuenta por la industria al momento de desarrollar nueva tecnología.

### **2.5.2. You Only Look Once**

Si bien tiene permanentes cambios y las nuevas versiones van mejorando algunos aspectos de las versiones anteriores, YOLO (You Only Look Once) siempre es tendencia como uno de los algoritmos de detección de objetos más utilizados en la industria (Joseph Redmon,

2018). Naciendo a mediados de 2015 en su versión inicial, se propuso una forma de abordar la temática holísticamente generando así una sola red neuronal encargada de identificar la existencia de objetos y clasificarlos en el mismo proceso. Esto generó una optimización de los recursos que captó la atención de los desarrolladores. Además, crearon una versión rápida de YOLO que permitía procesar imágenes a una velocidad de hasta 155 fotogramas por segundo con una precisión sumamente buena.

Todo esto hizo que se desarrollen nuevas versiones de YOLO donde fueron mejorando ciertos aspectos de rapidez y precisión, llevando así a la versión tres la cual va a ser utilizada en este trabajo para identificar objetos de forma rápida y poco costosa.

### **2.5.3. Mask\_RCNN**

Un algoritmo que fue posterior a los anteriormente presentados es el Mask RCNN, que como particularidad presenta una red neuronal recurrente basada en regiones que adopta una máscara previa para el cálculo de la localización de objetos (Kaiming He, 2017). Esto genera que esta red neuronal sea muy fácil de aplicar y de adaptar a otras situaciones como lo pueden ser la detección de poses realizadas por un ser humano, por ejemplo.

Una vez que se desarrolló esta tecnología, estuvo durante un tiempo prolongado considerada como una de las mejores técnicas para la detección de objetos a través de imágenes digitales dado que comprende una buena precisión en un tiempo acotado. Aun así, los mecanismos continúan avanzando y en poco tiempo estas tecnologías probablemente queden obsoletas con nuevos descubrimientos.

Este método se va a aplicar en el trabajo y se espera que sea el más preciso de todos por todo el respaldo de la comunidad científica que tiene y todas las menciones que se ha ganado en estos últimos tiempos.

## **3. Selección del modelo a partir de la detección temprana de infracciones viales**

En este apartado se selecciona el mejor modelo de los anteriormente presentados para resolver una temática en particular que se trata de la detección de infracciones viales y el análisis de posibles alternativas a la aplicación de proyectos de visión artificial en un contexto de ciudades inteligentes. Para esto, es necesario contemplar muchos aspectos dentro de los cuales se encuentran el desempeño del algoritmo y la aplicación del mismo teniendo en cuenta el trade-off que se genera al momento de tomar la decisión de qué tanto se pondera la precisión sobre la facilidad de aplicación o viceversa. Este debate es fundamental dado que en el contexto de ciudades inteligentes es un tema para tener en cuenta porque en muchos de los casos los recursos para desarrollar este tipo de proyectos son limitados y hay necesidad de administrarlos de la mejor forma posible.

Si bien los costos de las nuevas tecnologías como las cámaras de seguridad son relativamente bajos y están decreciendo, las ciudades son cada vez más grandes y requieren de más sitios para monitorear, lo que hace que esta discusión tome sentido. Va a ser importante la decisión de ceder un poco de precisión y complejidad del modelo para obtener más capacidad a la hora de embeber el sistema en dispositivos de baja capacidad computacional. Esto puede llevar a usar modelos reducidos que en algunos casos son llamados modelos mobile o tiny porque justamente son fácilmente aplicables a dispositivos móviles de baja capacidad de procesamiento.

Como se mencionó anteriormente, la selección del modelo está dada con el propósito de identificar infracciones viales de forma temprana para poder actuar rápidamente en el contexto de ciudades inteligentes. Particularmente las infracciones viales que se van a determinar son la identificación de tránsito pesado en aquellos lugares de mucha frecuencia vehicular donde no corresponde que un camión de gran tamaño circule según ordenanzas locales.

El manejo del tránsito pesado es una política de Estado de mucha importancia dado que afecta tanto a la circulación vehicular normal de una ciudad como al estado del pavimento y la seguridad vial. Si bien el tema es muy amplio, se puede identificar fácilmente los riesgos a los cuales se expone la ciudad cuando permite que vehículos de gran porte ingresen y circulen libremente por las calles de un municipio.

Por un lado, se puede mencionar las dimensiones de las calles céntricas de una ciudad comparado con las dimensiones que tiene un camión con acoplado, por ejemplo, donde

resulta incompatible una circulación vial del camión por el centro de la ciudad por cuestiones físicas que pueden impedir el viraje o ciertas maniobras por parte del conductor. Pero más allá de eso, en el supuesto caso que el tránsito pesado pueda realizar maniobras en dichas calles esto también puede traer aparejado una situación donde se generen mas embotellamientos y caos vehicular por la presencia de semejantes automotores.

Teniendo en cuenta los siniestros viales, también hay un riesgo a contemplar que se trata de los posibles accidentes que se pueden dar en una ciudad. Los habitantes del municipio que suelen estar expuestos a siniestros viales con autos particulares o colectivos locales de pasajeros, estarían expuestos a un riesgo más grande dado que la intensidad de un posible siniestro cambia rotundamente ya que se incorpora un actor de importancia como lo son los vehículos de gran porte. No es lo mismo tener un accidente con un auto particular con dimensiones pequeñas que tener un accidente con un camión que suele superar en tamaño y peso a la media de los vehículos que circulan en la ciudad.

Adicionalmente, en la actualidad sería difícil encontrar una ciudad que prescindiera del movimiento de los camiones dado que aportan gran valor al sector de la industria y al sector del comercio. No sería fácil hacer llegar ciertos productos a las góndolas de los supermercados a los cuales asisten los habitantes sin tener en cuenta el aporte de los camiones en la cadena de suministro. De la misma forma no sería fácil desarrollar un polo industrial que no permita el paso de los camiones para retirar los productos fabricados y permitir el ingreso de insumos necesario para la producción. Por estos motivos, los municipios suelen tener políticas muy claras a la hora de tratar con este tipo de cuestiones, brindando caminos seguros y aptos para la circulación de tránsito pesado.

Esto último, remonta al último riesgo a identificar que se trata del estado del pavimento urbano que está diseñado exclusivamente para que los automotores de pequeño porte puedan circular libremente. Cuando los pesos sobre el pavimento no son adecuados, se generan daños en el material que perjudican severamente a los vecinos del municipio ya que transitan esas calles diariamente y tienen que sufrir el mal estado de estas.

Si bien estos problemas mencionados anteriormente son atacados por parte de las autoridades estatales, en algunas ciudades es difícil encontrar reglas que funcionen bien para evitar este tipo de situaciones. En muchos casos las soluciones que se proponen son poco

efectivas, o difíciles de llevar a cabo, o bien son costosas desde el punto de vista de recursos económicos y humanos.

En algunos municipios el control se realiza a través de agentes de tránsito en zonas muy específicas donde son más propensas las infracciones de este tipo. Tal vez esta forma de solucionarlo sea la más eficaz porque evita el paso del tránsito pesado a las calles que no son aptas para este tipo de pesos. Pero a su vez no es del todo eficiente porque para llevar a cabo este proceso se requiere de una gran cantidad de recursos y la fluidez del tránsito también se verá afectada.

Otra solución que se puede establecer es la creación de un camino sinuoso previo a la zona de conflicto. Si bien esta solución parece ser adecuada, apunta a controlar los vehículos de gran porte, pero no diferencia el problema del pesaje del vehículo. Con eso se quiere decir que este sistema le va a poner las mismas trabas a un colectivo de pasajeros con un peso totalmente adecuado para el tránsito urbano que a un camión con acoplado que transporte cargas pesadas.

Por último, otra solución viable puede ser la puesta en marcha de un arco vial donde se establezca una altura máxima permitida de circulación prohibiendo así el tránsito de vehículos de mayor altura. Esta posible solución también tiene las desventajas de la anterior en el sentido que no diferencia por peso ni por tipo de vehículo prohibiendo el paso también a colectivos de pasajeros.

### **3.1. Resultados de los modelos a partir del análisis de imágenes**

Este apartado va a identificar los resultados obtenidos luego del proceso de implementación de los 3 modelos explicados anteriormente. Esta tarea no será sencilla, dado que la forma de seleccionar no es única, sino que hay que tener en cuenta principalmente dos tipos de variables. Por un lado, está la confianza del modelo que indica que tan bueno es para la identificación de objetos expresado en términos de probabilidad, determinando como ideal aquel modelo que identifica un vehículo con 100% de certeza y no se equivoca en su clasificación. Por otro lado, también hay que ponderar el hecho de la practicidad del modelo porque puede haber modelos muy precisos a la hora de clasificar, pero muy difíciles a la

hora de implementar dado que requieren recursos técnicos y de equipamiento que pueden llevar a la no preferencia de dicho modelo.

Habiendo decidido utilizar el aprendizaje resultante del modelo y no realizar un aprendizaje propio, se reducen mucho las necesidades de equipamiento ya que no se genera tal aprendizaje, sino que se utiliza un aprendizaje realizado en otro contexto. Aun así, si bien el objetivo del trabajo apunta a resolver otra cuestión, cabe destacar que, para una implementación completa de un servicio de identificación temprana de infracciones viales, hay una parte del proceso que no fue analizada donde existen necesidades técnicas al momento de capturar la imagen. Ese proceso puede traer aparejado complicaciones en el aspecto técnico, pero principalmente en el aspecto de equipamiento.

Teniendo en cuenta la precisión de los modelos, los resultados generados por los códigos ejecutados en el lenguaje de programación Python utilizando las librerías adecuadas, fueron muy tajantes, ya que el modelo Mask RCNN fue ampliamente mejor que el modelo de YOLO y el modelo de SSD.

Para analizar esa diferencia de forma gráfica, se va a ejemplificar el problema con una misma imagen y los resultados generados por cada modelo. A continuación, se puede ver claramente la diferencia que se menciona.

La siguiente imagen refiere a la imagen original (Martín, 2019). No está clasificada y está en su versión inicial. Claramente se identifican dos camiones en la parte inferior de la imagen, autos en la parte izquierda de la misma, y difícilmente por la lejanía se pueden identificar dos camiones más y algunos autos también. Los modelos posteriores deberían identificar tanto a los autos como a los camiones a pesar de que el objetivo del trabajo sea solamente identificar a los vehículos de gran porte.

### **Figura 5.**

Imagen objetivo



*Nota:* La imagen es la utilizada para visualizar fácilmente el impacto de cada uno de los modelos de detección de objetos. Tomado de *Hemos preguntado a los camioneros qué piensan del camión eléctrico de Tesla, y esto es lo que nos han dicho*, por Jesus Martin, 2019, MotorPasion.

En la siguiente imagen clasificada por el modelo de YOLO versión 3 se identifica con una precisión bastante alta uno solo de los dos camiones principales y fáciles de captar por el ojo humano. Cabe destacar que, si bien no es el objetivo del trabajo, el modelo también tiene un buen desempeño identificando autos, y hasta también se equivoca identificando una persona en el medio de la autopista. Esas equivocaciones se van a notar también en el resto de los modelos lo cual supondría que ese objeto es difícil de identificar ya que es muy confundible con una persona a pesar de estar en el medio de la autopista.

Dicho esto, se considera que el modelo YOLO es el segundo mejor de los tres presentados.

### **Figura 6.**

Imagen objetivo procesada por YOLO



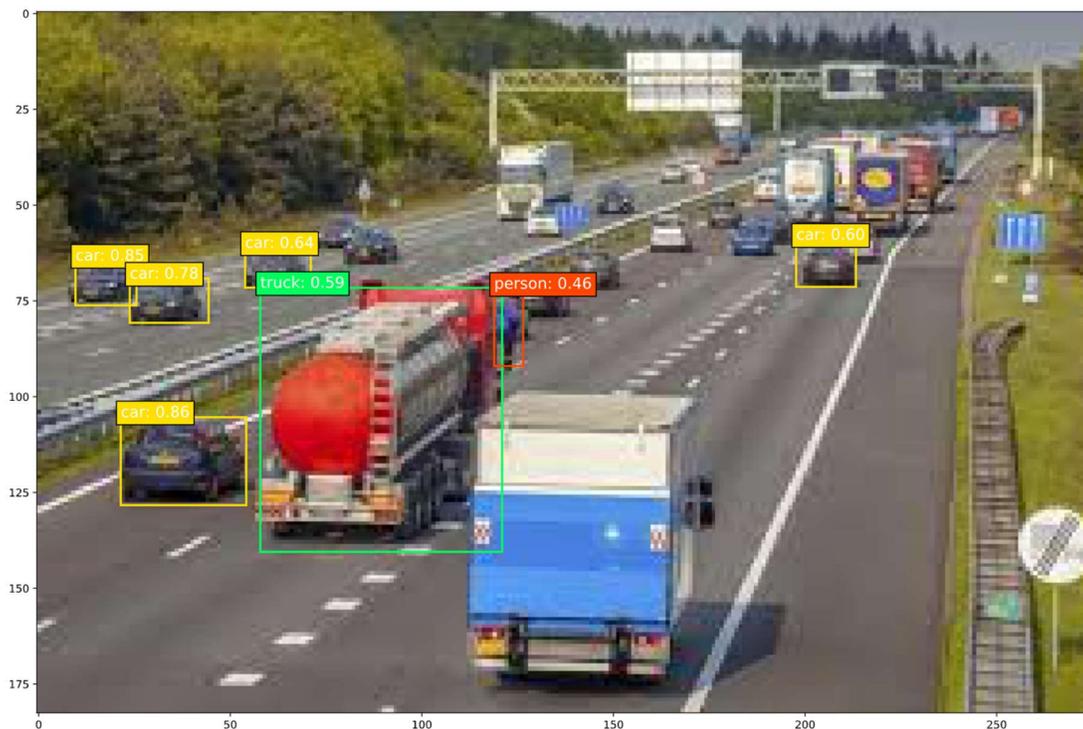
Nota: La imagen es la indicada para visualizar el comportamiento del modelo YOLO utilizando como base la imagen objetivo. Tomado de *Hemos preguntado a los camioneros qué piensan del camión eléctrico de Tesla*, y esto es lo que nos han dicho, por Jesus Martin, 2019, MotorPasion.

En la siguiente imagen clasificada por el modelo SSD300, se identifica un camión en la parte inferior de la pantalla, pero no se identifica al camión con partes azules que está al lado. Evidentemente se trata de un objeto difícil de identificar porque tanto el modelo anterior como éste no fue capaz de identificar dicho camión.

Cabe destacar que los recuadros generados por cada modelo tienen por detrás un umbral de confianza que genera un límite para mostrar la identificación del objeto. Esto quiere decir que aquellos objetos que son identificados por el modelo, pero cuya confianza no supera el límite, no se van a ver reflejados en la imagen a pesar de que el modelo los esté identificando. Este modelo llamado SSD se encuentra como última opción entre los tres modelos elegidos.

### Figura 7.

Imagen objetivo procesada por SSD



*Nota:* La imagen es la indicada para visualizar el comportamiento del modelo SSD utilizando como base la imagen objetivo. Tomado de *Hemos preguntado a los camioneros qué piensan del camión eléctrico de Tesla*, y esto es lo que nos han dicho, por Jesus Martin, 2019, MotorPasion.

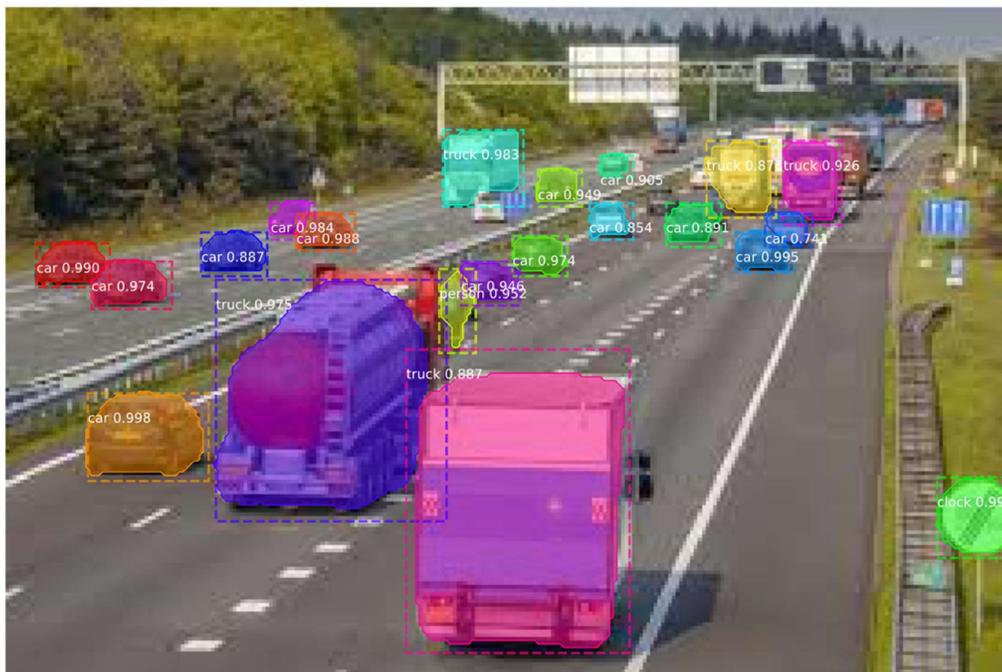
En la siguiente imagen se puede notar la identificación del modelo Mask RCNN. A diferencia del resto de los modelos, tiene la particularidad que rellena el contorno del objeto lo cual hace que el modelo sea un poco más complejo porque no solo identifica y clasifica al objeto, sino que también posee esta funcionalidad. Este modelo pudo identificar no solo a los dos camiones que están en la parte inferior de la imagen, sino también a aquellos vehículos de gran tamaño que están algo alejados en la imagen y que son de difícil identificación. Además, cuenta con una confianza elevada dado que por ejemplo identifica al mismo camión que el resto de los modelos con una confianza de 97.5% lo cual lo convierte en el mejor modelo de los tres.

Se puede notar que no solamente identifica camiones de muy buena forma, sino que también sucede de la misma manera con los autos que son identificados con una confianza muy elevada. Y como detalle se puede mencionar el caso del cartel de normas de tránsito en la parte inferior derecha de la imagen que es identificado como un reloj. Dicho esto, se puede

notar que el modelo no tiene mucho contexto sobre lo que está analizando porque no debería identificar personas o relojes principalmente en las autopistas dado que tiene poco sentido encontrar un reloj grande al lado de una ruta.

**Figura 8.**

Imagen objetivo procesada por Mask RCNN



*Nota:* La imagen es la indicada para visualizar el comportamiento del modelo YOLO utilizando como base la imagen objetivo. Tomado de *Hemos preguntado a los camioneros qué piensan del camión eléctrico de Tesla*, y esto es lo que nos han dicho, por Jesus Martin, 2019, MotorPasion.

### 3.3 Ventajas y desventajas del modelo seleccionado

Con el afán de seleccionar un modelo apto para solucionar la problemática de la identificación de tránsito pesado en un contexto de ciudad inteligente se probaron diversos modelos que, si bien todos funcionan muy bien detectando el problema, no funcionan de la misma forma y tienen detalles que convierte a uno en la mejor opción dada la situación.

En este caso el modelo que mejor se adaptó al problema fue Mask RCNN dado que tiene un buen trade off entre precisión y facilidad de aplicación. Aun así, cabe destacar que el resto de los modelos también tuvieron un buen desempeño, pero no lograron superar al modelo seleccionado principalmente por la alta precisión de este.

Si bien el modelo de Mask RCNN ha sido seleccionado entre la terna propuesta como el mejor a la hora de identificar tránsito pesado en el contexto de ciudades inteligentes, cabe destacar que a pesar de eso existen diversas cuestiones a tener en cuenta que complejiza el uso del mismo y por lo tanto generan problemas a tener en cuenta antes de aplicarlo.

En primer lugar, una desventaja que tal vez tengan en común los modelos mencionados es la dificultad a la hora de la aplicación práctica para generar un caso de uso utilizando el algoritmo. Si una persona ajena al mundo de las matemáticas y la programación computacional quisiese abordar esta temática haciendo uso de dicho modelo, seguramente encontraría una variedad de barreras teóricas y prácticas que dificultaría mucho llevar a cabo el proceso. La complejidad del modelo hace que el operador del sistema deba conocer al menos abstractamente cómo funciona una red neuronal y como se pueden aplicar procesos de aprendizajes de transferencia, por ejemplo.

Por otro lado, el desarrollo del algoritmo no tiene una comunidad muy activa donde poder evacuar dudas sobre el funcionamiento de este mientras que tal vez otros modelos si la tienen y eso genera que la aplicación del modelo no pueda ser llevada a cabo por una persona que talvez carezca de recursos técnicos.

Otra desventaja que se exterioriza al momento de trabajar con el código en Python es la dificultad para extraer caracterizaciones de importancia que pueden ser útiles para el análisis como por ejemplo la importancia de los pixeles al momento de identificar un objeto. Cuando se generan las cajas de reconocimiento, no se indica cuáles fueron los motivos por los cuales el modelo tomó dicha decisión y eso tal vez sea algo para tener en cuenta al momento de medir la precisión del modelo.

Si bien no es del todo correcto mencionar la siguiente situación como una desventaja propia del modelo, debería tenerse en cuenta que el modelo desarrollado capta muy bien el tamaño de los vehículos de gran porte en un contexto, pero es incapaz de analizar el peso de los vehículos que en estos casos sería una característica importante para tener en cuenta. Dicho esto, el modelo no va a captar la diferencia entre aquellos camiones que están transportando

mercaderías de gran peso y aquellos camiones que, si bien tienen acoplado o parecen trasladar productos, están vacíos o transportan manufacturas de gran volumen, pero poco pesaje.

Por último, se desea destacar como una desventaja las necesidades en términos de equipamiento para llevar a cabo un proceso de estas características dado que no son procesos que se puedan ejecutar en computadoras simples, sino que requieren de un mínimo de componentes para poder llevarse a cabo. Si bien esta desventaja con los nuevos avances en hardware va quedando algo obsoleta, cabe destacar que, para aplicaciones en mayor escala, los requerimientos de equipamiento son un hecho muy importante porque se tiene que afrontar con un presupuesto abultado sobre todo a la hora de la captación de las imágenes.

Por otro lado, se mencionarán algunas ventajas propias del funcionamiento del modelo, y otras que son ventajas resultantes de medir el desempeño de este.

Si bien la tecnología de aplicar cajas de reconocimiento en las imágenes digitales ya está siendo utilizada, tal vez una ventaja muy destacada que tiene este modelo sobre los otros es la aplicación de máscaras generativas que resultan de cajas de reconocimiento más abstractas para poder luego ser aplicadas en la imagen de forma tal de generar un resaltado que comprende el contenido del objeto. Es decir, si bien se utilizan las cajas de reconocimiento, finalmente el objeto es resaltado a través de su contorno y generando una especie de sombreado sobre su contenido.

Otra ventaja para destacar propia del modelo es la utilización de una gran base de datos para generar el aprendizaje de reconocimiento de objetos. Para dicho proceso se utilizó la base de datos de MS COCO donde existen más de 330 mil imágenes clasificadas entre 80 categorías generando así una gran base para el aprendizaje automático de los procesos de detección de objetos.

Para culminar, se desea destacar la mayor ventaja que se le observa al modelo que refiere a la precisión de este teniendo en cuenta las dificultades de su implementación. Habiendo comparado anteriormente la precisión de los modelos, en este caso el modelo de Mask RCNN es ampliamente superior al resto de los modelos implementados dado que tiene más confianza a la hora de clasificar un objeto y la dificultad a la hora de realizar la implementación es moderadamente baja.

### **3.4 Aplicaciones y usos del modelo de detección de objetos**

En este apartado se quiere mencionar a grandes rasgos algunas posibles mejoras que puede tener este proceso de reconocimiento de infracciones viales en un contexto de ciudades inteligentes. Algunas mejoras van a estar vinculadas al modelo en sí, pero otras van a estar dirigidas sobre la idea de ampliar un poco la problemática del trabajo y poder así abarcar nuevas ideas.

Inicialmente, se desea mencionar que se utilizaron los pesos ya establecidos en los modelos para la identificación de objetos sin un entrenamiento posterior que resulte mejor para la temática. Como ya se mencionó anteriormente, en este proceso solo se utilizan los aprendizajes obtenidos de un modelo de entrenamiento sobre una base de datos específica con etiquetas ya preestablecidas. Lo que se está proponiendo en esta primera mejora es utilizar un mecanismo de aprendizaje propio a través de imágenes de aquellos objetos que se desea analizar para lograr una precisión aun mayor y evitar aquellos errores que todavía puede cometer este modelo. Cabe destacar que para iniciar este proceso hace falta no solo una base de datos muy específica con elementos de interés dentro, sino también una mayor capacidad de cómputo que pueda ejecutar una red neuronal compleja como lo puede ser Mask RCNN. Por otro lado, sería de mucha utilidad analizar cómo funcionan los modelos tiny o mobile en estos casos ya que requieren menor poder de cómputo y están desarrollados para ser embebidos en sistemas de baja capacidad. Esto apuntaría sobre todo a resolver problemas de equipamiento a la hora de ejecutar un proyecto completo de identificación de tránsito pesado incluyendo la toma de las imágenes.

Uno de los problemas que se identificó al momento de aplicar el algoritmo ganador, fue que el modelo es capaz de identificar objetos que no son propios de una autopista, por lo tanto, esto indica que el modelo no tiene contexto ya que identifica a una persona y a un reloj en la imagen. Desde un set de entrenamiento especializado, se podría llegar a añadir dicho contexto para evitar que identifique objetos que no deberían estar dentro de la imagen.

Como última mejora que tal vez está relacionada con el caso anterior, se podría agregar también al modelo análisis de video en vez de estar ejecutando con imágenes solamente. Si bien un video está compuesto de muchas imágenes, el hecho de aplicar el aprendizaje sobre un video evitaría tal vez esas situaciones donde el modelo confunde el espejo retrovisor de

una automóvil con una persona en el medio de la autopista dado que no sería un análisis estático, sino más bien se trataría de un análisis dinámico de la situación.

Con el afán de mencionar donde puede ser útil este tipo de implementaciones, no hay que olvidar aquellas soluciones que se presentan hoy en día con respecto a la temática que ya fueron mencionadas previamente. Estas nuevas herramientas pueden ser principalmente instaladas en aquellos sectores de los municipios donde actualmente se realizan periódicamente controles viales para evitar el tránsito pesado dentro de una ciudad. Esto ayudaría ampliamente a una gestión estatal a reasignar mejor sus partidas presupuestarias en términos de dinero y de recursos asignados para cada proyecto dado que se puede complementar muy bien con los controles que se establecen en la actualidad.

## Conclusiones

El presente trabajo de investigación ha estudiado con profundidad aspectos relativos a la visión artificial y como esta nueva tecnología pueda aplicarse dentro de un contexto de ciudades inteligentes. El problema abordado esta relacionado con el objetivo principal del trabajo que es la búsqueda de alternativas para aplicar métodos de detección temprana de infracciones viales a través del uso de algoritmos de detección de objetos. Como objetivos específicos, se ha planteado la evaluación del impacto que tiene la gestión de datos dentro del sector público, la identificación de un algoritmo útil para la detección de infracciones viales, y por último el análisis de alternativas de aplicación de este tipo de proyectos dentro del contexto de ciudades inteligentes.

Respecto del objetivo general, en la tesis se ha analizado y aplicado los tres modelos propuestos para resolver la problemática inicial. Una vez verificados los resultados que se encuentran en el anexo, se decidió que el modelo que más se adecúa al contexto detallado es el algoritmo Mask RCNN desarrollado por Kaiming He y su equipo. Si bien el resto de los modelos analizados no tuvieron resultados erráticos, la mejora en niveles de precisión y de complejidad que implica el uso del modelo seleccionado fue determinante para poder tomar la decisión entre la terna planteada.

Por un lado, el modelo de Mask RCNN ha identificado con una confianza del 97.5% y del 88.7% a los dos vehículos de mayor envergadura que aparecen en la parte inferior de la fotografía (Martín, 2019). Además, el resto de los vehículos de tránsito pesado que fueron identificados en la parte superior de la imagen, siendo estos más difíciles de identificar, tuvieron una confianza de 92.6% 87% y 98.3% respectivamente. Esto hace que el modelo ganador haya superado al segundo modelo que mejor desempeño tuvo por un amplio margen, dado que no solo identificó a más vehículos de tránsito pesado sino también tuvo una mejora de 11.28% sobre el promedio de confianza en la identificación de vehículos de gran porte. No obstante, también arrojó un promedio ponderado de la media de precisión (mAP) de 71% cuando el resto de los modelos para la misma métrica arrojaron resultados de 41% y 26%. Para finalizar, vale la pena mencionar que la tasa de falsos negativos es considerablemente menor en el modelo Mark RCNN dado que alcanza un valor de 29% cuando el resto de los modelos tienen valor final de 59% y 74% bajo el mismo análisis.

Cabe destacar que, si bien los resultados obtenidos fueron contundentes y positivos, existen diversas formas de ampliar la problemática y utilizar herramientas que puedan mejorar aún más la aplicación del propio modelo. Es decir, si bien se puede poner a prueba otros modelos diferentes a los planteados en este trabajo, también se puede agregar como trabajo futuro la posibilidad de añadir un aprendizaje propio utilizando una base de datos adecuada para realizar las pruebas. Esto no solo generaría mejores métricas de precisión sino también la posibilidad de especificar aún más el objeto de estudio.

Por último, a lo largo de todo el trabajo se destaca como fue creciendo la industria de la ciencia de datos y el uso de estos. Dado esto se considera que algunos aspectos tratados dentro de este trabajo como lo son el libre acceso a los datos, el rol que toma el Estado para el tratamiento de estos, o también el rol que pueden tomar los privados sin incurrir en faltas éticas o legales, podrían profundizarse aún más dejando como trabajos futuros muchas inquietudes que si bien fueron planteadas y analizadas, dan lugar a problematizaciones más complejas y necesarias dentro del rubro.

## Referencias Bibliográficas

### Referencias

- A., G.-M. (2012). Phineas Gage y el enigma del córtex prefrontal. En G.-M. A., *Neurología* (págs. 370-375). España.
- Argentina, C. d. (15 de Diciembre de 1994). Artículo 43. *Artículo 43*. Ciudad Autónoma de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina: Constitución de la Nación Argentina.
- Argentina, C. N. (4 de Octubre de 2000). Ley 25.326. *Protección de los Datos Personales*. Ciudad Autónoma de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina: Boletín Oficial.
- Argentina, M. d. (24 de Febrero de 2015). Disposición 10/2015. *Disposición 10/2015*. Ciudad Autónoma de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina: Dirección Nacional de Protección de Datos Personales.
- Creemers, R. (2018). China's Social Credit System: An Evolving Practice of Control. *Leiden Institute for Area Studies*.
- Foucault, M. (1975). *Vigilar y Castigar*.
- Gagnon, L. L.-C. (2001). Procedure to detect anatomical structures in optical fundus images. *Medical Imaging 2001: Image Processing*.
- Gershgor, D. (02 de Marzo de 2021). *Medium*. Obtenido de Medium: <https://onezero.medium.com/chinas-sharp-eyes-program-aims-to-surveil-100-of-public-space-ddc22d63e015>
- Jia Deng, W. D.-J.-F. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*.
- Jonathan, H. (06 de Marzo de 2018). *Medium*. Obtenido de Medium: <https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173>
- Joseph Redmon, A. F. (2018). YOLOv3: An Incremental Improvement. *University of Washington*.
- Kaiming He, G. G. (2017). Mask R-CNN.
- Liu, W. (2016). SSD: Single Shot MultiBox Detector. En E. C. Vision, *Computer Vision – ECCV 2016* (págs. 21-37). Springer, Cham.
- Locke, J. (1689). *Segundo Tratado Sobre el Gobierno Civil*.
- Martín, J. (27 de Junio de 2019). *motorpasion*. Obtenido de <https://www.motorpasion.com/industria/hemos-preguntado-a-los-camioneros-que-piensan-del-camion-de-tesla-y-esto-es-lo-que-nos-han-dicho>
- Óskarsdóttir, M. B. (2020). Credit scoring for good: Enhancing financial inclusion with smartphone-based microlending. *Thirty Ninth International Conference on Information Systems*.
- Pawson, R. (2006). Evidence-Based Policy: A Realist Perspective. En R. Pawson, *Evidence-Based Policy: A Realist Perspective* (pág. 38). New Delhi: Sage.
- Pitts, W. S. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*.
- Stuss, D. T. (2000). Executive functions and the central lobe: a conceptual view. *Psychological Research*, 289-298.
- Yanofsky, N. S. (2010). Towards a Definition of an Algorithm. *Journal of Logic and Computation*, 253-286.

Zanella, A. B. (2014). Internet of Things for Smart Cities. *IEEE Internet of Things Journal*, 22-32.

## Anexos

- [https://github.com/fedegarciablanco/Mask\\_RCNN](https://github.com/fedegarciablanco/Mask_RCNN)
- [https://github.com/fedegarciablanco/ssd\\_keras](https://github.com/fedegarciablanco/ssd_keras)
- <https://github.com/fedegarciablanco/darknet>
- <https://docs.google.com/spreadsheets/d/1BnUZ1ZPmGvJxXxjQKfdBMBgT8xNW-qAaKqEe5PmkfWQ/edit#gid=0>