

Counting, Monitoring and Classification of Access Using Artificial Vision

David A. López, Ingeniero Mecatrónico¹, Adrián M. Mora, Ingeniero Mecatrónico², Alicia María Reyes Duke, Magister en Ingeniería³

^{1,2,3} Universidad Tecnológica Centroamericana, UNITEC-Honduras, dalduarte2498@gmail.com, amrivera_1@hotmail.com, alicia.reyes@unitec.edu.hn.

***Abstract—** Through the use of artificial vision and convolutional neural networks supported by data augmentation techniques, it achieves this research as its main objective, it carries out an accurate count of entry and exit of cars, motorcycles / bicycles and people of a commercial premises. The use of these technologies is thanks to the use of artificial vision libraries such as OpenCV in conjunction with the Python programming language provided by the programmer in a friendly and ideal environment for the development of applications of this type. Using YOLOv3 and the weights of the Darknet pre-trained model, faster and more efficient red neural training is achieved. Executing and subsequently analyzing the data collected through the extensive operation of the application. This research aims to be an option to solve a general and recurring problem in terms of monitoring, contemporary classification.*

***Key words:** artificial vision, convolutional neural networks, data augmentation techniques, precision*

Digital Object Identifier (DOI):
<http://dx.doi.org/10.18687/LACCEI2020.1.1.468>
ISBN: 978-958-52071-4-1 ISSN: 2414-6390

Conteo, Monitoreo y Clasificación de Accesos Utilizando Visión Artificial

David A. López, Ingeniero Mecatrónico¹, Adrián M. Mora, Ingeniero Mecatrónico², Alicia María Reyes Duke, Magister en Ingeniería³

^{1,2,3} Universidad Tecnológica Centroamericana, UNITEC-Honduras, dalduarte2498@gmail.com, amrivera_1@hotmail.com, alicia.reyes@unitec.edu.hn.

Resumen— Mediante el uso de visión artificial y redes neuronales convolucionales apoyadas de técnicas de aumento de datos, logra esta investigación como objetivo principal, llevar un conteo preciso del ingreso y salida de automóviles, motocicletas/bicicletas y personas de un local comercial. El uso de estas tecnologías se da gracias a la utilización de librerías de visión artificial tal como OpenCV en conjunto del lenguaje de programación Python que le brinda al programador un entorno amigable e ideal para el desarrollo de aplicaciones de este tipo. Haciendo uso de YOLOv3 y los pesos del modelo pre entrenado de Darknet, se alcanza un entrenamiento de la red neuronal más rápido y eficiente. Ejecutando y posteriormente analizando los datos recopilados a través del extensivo funcionamiento de la aplicación. Esta investigación pretende ser una opción para dar solución a un problema general y recurrente en lo que refiere al monitoreo, clasificación conteo.

Palabras clave— Aumento de datos, precisión, redes neuronales convolucionales, visión artificial.

I. INTRODUCCIÓN

Esta investigación, hará uso de la visión artificial para identificar, contar y finalmente clasificar entre distintos vehículos y personas que entran y salen a un determinado local. Para ello, se explorarán las diversas técnicas existentes en la actualidad de visión artificial para la identificación de objetos, y para realizar el conteo de entradas y salidas de dichos objetos en una zona determinada, con imágenes provenientes de videos de cámaras de seguridad ya instaladas; y así determinar cuál será de mejor uso para la aplicación previamente mencionada. Para el entrenamiento de la red neuronal, se hará uso de las técnicas de aumento de datos, y después se comparará el desempeño del sistema, al ser entrenado con las técnicas de aumento de datos y sin ellas. También se determinará la precisión de detección de objetos al variar el nivel iluminación al cual han sido expuestas las imágenes provenientes de video. Esta implementación de un sistema de control y clasificación de accesos en cualquier comercio, sugiere que sea una alternativa para los propietarios que ya tienen un sistema de monitoreo existente.

II. PRECEDENTES

Existen muchas variedades de negocios que les es de interés llevar un conteo adecuado de la cantidad de clientes que reciben al día. Al tener conocimiento del flujo de personas o vehículos que circulan a través de un local, se pueden obtener datos

estadísticos y cuantitativos que ayudan a tener un mejor control y seguridad, además de brindar datos que facilitan la toma de decisiones en lo que respecta a la administración de los mismos. En la actualidad existen formas muy variadas y diferentes tipos de tecnologías para llevar a cabo la construcción de un sistema de este tipo, cada uno con sus características especiales, que son un punto clave a la hora de querer adquirir uno del mercado.

Catucumbamba & Quinga [1] logran identificar los diferentes tipos de contadores existentes en el mercado actual: visuales, mecánicos, electrónicos y los híbridos resultantes de entre los 3. Siendo el contador visual el cual es que es actualmente utilizado en el local de este estudio.

A. Sistemas de Conteo

Los sistemas de conteo electrónicos convencionales se basan fundamentalmente en el uso de sensores. Los sistemas de conteo de vehículos no intrusivos se basan en sensores como los laser, infrarrojos, ultrasonido, radares, o cámaras [2]. La elección de dichos sensores, mostrados en la tabla 1 ha propiciado un debate en cuanto al uso de sensores pasivos o activos en el sistema de detección de carretera. Los sensores pasivos, tales como las videocámaras, tienen la ventaja de poder adquirir los datos de manera no intrusiva. Estos ofrecen una perspectiva más amplia de la escena, teniendo la posibilidad de usar estos sistemas para otras funciones, tales como: detección de las líneas laterales del carril, detección de señales de tráfico y reconocimiento de objetos [3].

TABLA 1. SENSORES UTILIZADOS PARA EL SISTEMA DE CONTEO.

Método	Conteo	Velocidad	Clasificación	Múltiples carriles de detección
Circuitos Inductores	X	X	X	
Magnetómetros	X	X		
Radar de microondas	X	X	X	X
Infrarrojo activo	X	X	X	X
Infrarrojo pasivo	X	X		
Ultrasonido	X	X		
Acústicos	X	X		X
Procesamiento de Imágenes	X	X	X	X

Digital Object Identifier (DOI):
<http://dx.doi.org/10.18687/LACCEI2020.1.1.46>:
ISBN: 978-958-52071-4-1 ISSN: 2414-6390

B. Visión Artificial

El enfoque actual del reconocimiento de objetos hace uso esencial de los métodos de *Machine Learning*, siendo los modelos de *Deep Learning* los que han logrado los resultados más remarcables en la visión artificial[4].

Jian et al. [5] en China, Peña [6] en Perú, Chamorro [7] y Sisalima [8] en Ecuador, son ejemplos de la implementación de sistemas de visión artificial para la detección, y conteo de personas y vehículos; demuestran que la viabilidad de estos sistemas en la actualidad yace en el menor costo de implementación y mantenimiento al ser comparado con métodos más tradicionales como el uso de barreras vehiculares o sensores.

Goodfellow, Bengio, & Courville [9] afirman que las redes de avance son la base de las aplicaciones comerciales, como las redes convolucionales utilizadas para el reconocimiento de objetos en las fotos. Las redes neuronales convolucionales son un tipo especial de red de avance para el procesamiento de datos con topología de cuadrícula, como las imágenes.

Rosebrock [10] menciona que existen tres principales modelos basados en *Deep Learning* para el reconocimiento de objetos: Red Neuronal Convolutiva de Región (R-CNN), y sus demás variantes (Fast R-CNN y Faster R-CNN); Detector de Disparo Único (SSD); YOLO.

Redmon & Farhadi [11] en 2018 presentan YOLOv3, la versión más reciente y mejorada de YOLO, la cual resulta ser más rápida de ejecutar que sus contrapartes antes mencionadas. La red predice 4 coordenadas para cada cuadro delimitador, t_x , t_y , t_w , t_h . Después se lleva a cabo una predicción para la puntuación de los objetos para cada una de las cajas delimitadoras usando regresión logística.

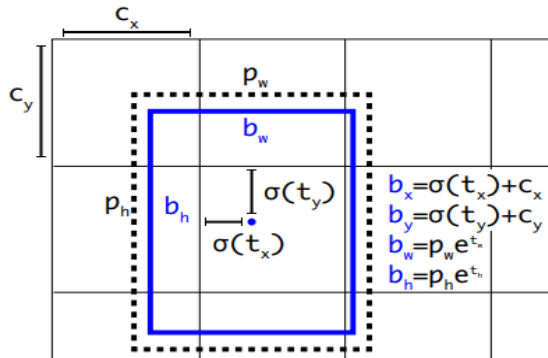


Fig. 1 Cálculo de cajas delimitadoras en YOLOv3

C. Técnicas de Aumento de Datos

Abarcan una variedad de técnicas usadas para generar un mayor número de muestras para el entrenamiento a partir de los datos originales. La meta es incrementar la generalización en la clasificación del modelo, ya que, en la práctica, la red estará viendo versiones modificadas de manera leve en la información que esta tenga que procesar [12]. Las técnicas de aumento de datos son una manera de reducir el sobreajuste en los modelos,

una práctica común para el aumento de datos es realizar un ajuste en la geometría o color de las imágenes, como reflejarlas sobre sus ejes, recortándolas, reflejándolas, y cambiar su paleta de colores[13]

En el sentido de esta investigación, el uso de las técnicas de aumento de datos será con el motivo de disminuir el posible error generado para la identificación de personas o vehículos en horario nocturno, haciendo uso de imágenes modificadas de un data set de imágenes existentes.

III. METODOLOGÍA

A. Técnicas e Instrumentos Aplicados

Las técnicas e instrumentos aplicados para el desarrollo de este proyecto son los siguientes:

- OpenCV 4.1.1: es un software de fuente abierta para visión artificial y también una librería de machine learning. OpenCV se construyó para proveer una infraestructura básica para aplicaciones de visión artificial y de esta manera acelerar el uso de la perspectiva artificial en los productos comerciales.
- Python 3.7
- Jupyter Notebook
- Microsoft LifeCam HD-3000
- YOLOv3
- Google Colab: es un servicio de nube gratis, basado en el entorno de desarrollo Jupyter para *machine-learning*, educación e investigación. Google Colab le facilita al usuario un entorno previamente configurado para aplicaciones de aprendizaje profundo y aparte cuenta con acceso a un GPU robusto de 12 GB de GDDR5.
- Técnicas de Aumento de Datos
- Concepto de Zona Monitoreada: Se utiliza la imagen vista en el metraje de la cámara web. Para lograr el monitoreo, se delimitan dos líneas paralelas, una de entrada y la otra de salida. Estas líneas delimitadoras consisten en líneas virtuales posicionadas sobre las imágenes en OpenCV. Existen dos líneas delimitantes:
 - a. Línea de Entrada: referencia la línea que define el límite para que un objeto entre a la zona monitoreada.
 - b. Línea de Salida: referencia la línea que define el límite para que un objeto salga de la zona monitoreada.

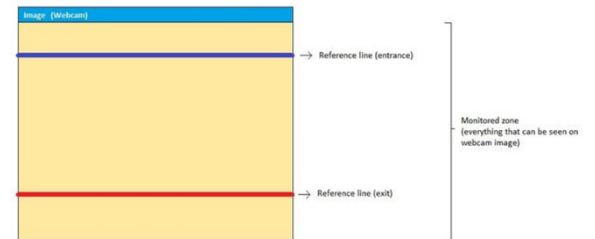


Fig. 2 Zona monitoreada con líneas de entrada y salida

B. Población y Muestra

Los conjuntos de información que se utilizaron para el entrenamiento de la red neuronal fueron obtenidos a través de MS COCO (*Common Objects in Context*), un conjunto de datos utilizado para la segmentación y reconocimiento de objetos en aplicaciones de visión artificial. Se cuenta con mil imágenes de automóviles en diferentes ángulos, mil imágenes de motocicletas/bicicletas y mil imágenes de peatones. Recopiladas con aproximadamente 72 horas de video extraído.

C. Metodología de Estudio

Tomando como base el proceso de la investigación de desarrollo, ilustrado en la figura 3, se siguen cinco pasos:

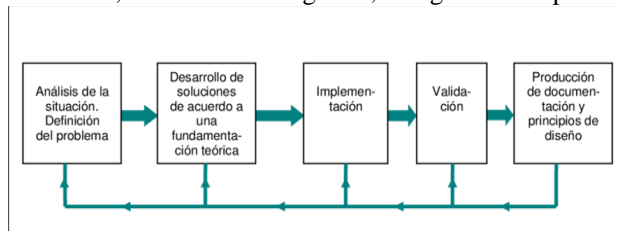


Fig. 3 Proceso de la investigación de desarrollo.

1. Análisis de la situación. Mediante entrevistas a figuras gerenciales dentro del comercio del país, se llegó a la conclusión que llevar un control eficiente de la cantidad de clientes que ingresan a cierto local en el transcurso del día no es una tarea fácil.
2. Desarrollo de soluciones de acuerdo con una fundamentación teórica. Posteriormente a realizar una investigación sobre las mejores alternativas para el conteo de ingresos y salidas a un local, se llegó a la conclusión de que la mejor y más económica opción es desarrollar una aplicación de visión artificial que se encargue del constante monitoreo de las cámaras de seguridad ya instaladas en dicho local.
3. Implementación. Durante la implementación del sistema, se buscaron las mejores alternativas en cuanto a los materiales utilizados para dicha aplicación. Se llegó a la conclusión que, posicionando la cámara de manera vertical, direccionada hacia el suelo, era la manera más eficiente de documentar la cantidad de clientes que ingresaban al local, utilizando una técnica de monitoreo de zona con líneas paralelas.
4. Validación. Haciendo un esfuerzo por validar la aplicación, se adquirió una cámara web con una resolución de 720p con el fin de poder comparar los resultados recopilados por el sistema desarrollado al analizar la información de manera objetiva pudiendo comparar con la información recopilada en las grabaciones de las cámaras de seguridad del local.

5. Producción de documentación y principios de diseño. Finalmente, a través de lo largo de la investigación se produjo la documentación necesaria que abarca desde el marco teórico, hasta los resultados y análisis. Los principios del diseño de la aplicación y el montaje de la misma son la base fundamental de esta investigación.

D. Metodología de Validación

Para realizar la validación del funcionamiento de la aplicación, se recopilan los datos tabulados por la aplicación, posteriormente se analizan a la par de la información que se obtiene de manera tradicional. Al comparar dicha información se busca obtener como máximo un 5% de margen de error, tomando en cuenta el hecho de que un 3% de las personas que ingresan al local de estudio, no se quedan [15].

En modelos de identificación de objetos, es común utilizar lo que es llamado el Promedio Medio de Precisión, o mAP por sus siglas en inglés (*mean Average Precision*). Para poder calcular el mAP, primeramente, se debe comprender el termino de Intersección sobre Unión, IoU por sus siglas en inglés (*Intersection over Union*) para luego cuantificarlo. IoU es una proporción entre la intersección y la unión de la caja delimitadora que predice el algoritmo y la caja de verdad que identifica correctamente al objeto en la imagen, como se muestra en la figura 4.



Fig. 4 Intersección sobre Unión.

Para poder cuantificar estos valores, se utiliza la siguiente formula:

$$IoU = \frac{\text{área de superposición}}{\text{área de unión}} \quad (1)$$

Al utilizar esta fórmula es posible entrenar al algoritmo para identificar si una detección es positiva o negativa. El límite más comúnmente utilizado es de 0.5. Si el IoU > 0.5, entonces este es considerado un Verdadero Positivo, de lo contrario se le denomina un falso positivo.

Para poder calcular la exhaustividad, también se deben tomar en cuenta los negativos. Ya que cada parte de la imagen donde el algoritmo no detecta un objeto es considerado un negativo. Finalmente se calcula el IoU utilizando la verdadera clasificación de la imagen en cada una de las detecciones positivas que el modelo reporta. Utilizando este valor obtenido y el límite definido, se calcula el número de detecciones

correctas para cada clase en la imagen, este valor se utiliza para calcular la precisión y la exhaustividad del sistema.

IV. RESULTADOS Y ANÁLISIS

La elaboración del software utilizado para las pruebas realizadas por esta investigación se llevó a cabo a lo largo de tres semanas. El software elaborado se divide en dos etapas. La primera etapa consiste en la detección y clasificación del objeto que se encuentra en movimiento en la imagen. Mientras que la segunda etapa cumple el propósito de determinar si el objeto en movimiento está ingresando o egresando del local comercial.

A. Detección y Clasificación

La red neuronal que se entrenó para la aplicación descrita fue hecha utilizando el modelo de YOLOv3 en conjunto con los pesos de *Darknet* y un conjunto de información de aproximadamente 3,000 imágenes obtenidas de MS COCO, el conjunto total se logra ampliar un 50% al hacer uso de técnicas de aumento de información, se logra alcanzar una mayor cantidad de datos para el entrenamiento al hacer pequeñas variaciones geométricas y de color con el fin de incrementar la precisión de la red neuronal. El tiempo de entrenamiento de la red neuronal fue de 11 horas utilizando Google Colab que cuenta con un GPU de 12gb de memoria DDR5.

Para las pruebas iniciales, se tomaron distintas fotografías dentro del campus de Unitec SPS en las que hubiera distintos objetos a clasificar. Como se observa en la figura 5, el algoritmo logra diferenciar con bastante precisión distintos objetos, en algunos casos incluso si los objetos están siendo parcialmente bloqueados u obstaculizados.

Dependiendo de la calidad de la imagen varía el tiempo de procesamiento del algoritmo de clasificación, para las imágenes presentadas anteriormente, el tiempo de procesamiento de cada una fue de aproximadamente 7 segundos. Esto resulta ser preocupante puesto que, si se tiene un retraso tan grande con una imagen de baja calidad, es de esperarse que al utilizar un video cuyo tamaño de archivo es considerablemente más grande este tiempo sea incluso más alto.



Fig. 5. Pruebas de Algoritmo de Clasificación.

B. Precisión de la Red Neuronal

Para realizar el cálculo de la precisión de la red neuronal se utiliza la ecuación 2. Como se menciona anteriormente, se utilizan los Verdaderos Positivos (VP) y Falsos Positivos (FP) provenientes del análisis de intersección sobre unión.

$$\text{Precisión} = \frac{VP}{VP + FP} \quad (2)$$

La precisión evaluada en esta investigación se divide en dos categorías con tres incisos cada una.

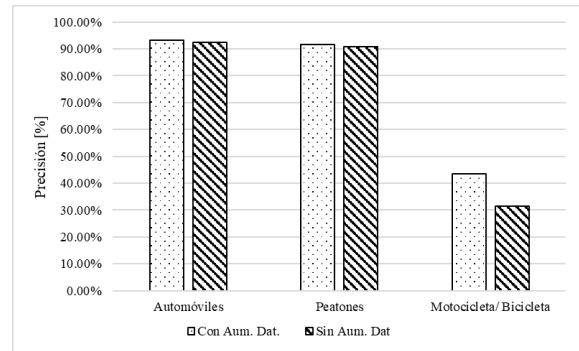


Fig. 6. Comparación de precisión.

La figura 6, deja en evidencia la efectividad de las técnicas de aumento de datos a la hora de ampliar para posteriormente entrenar una red neuronal identificadora de objetos. Los cambios de color principalmente dan una mayor variedad de alteraciones lo que conlleva a un porcentaje más alto de precisión.

C. Exhaustividad de la Red Neuronal

Para realizar el cálculo de la exhaustividad de la red neuronal se utiliza la ecuación 3. Para su cálculo se toman en cuenta los Verdaderos Positivos (VP) y los Falsos Negativos (FN)

$$\text{Exhaustividad} = \frac{VP}{VP + FN} \quad (3)$$

Al igual que la precisión, la exhaustividad de la red neuronal se divide en las mismas dos categorías con tres incisos:

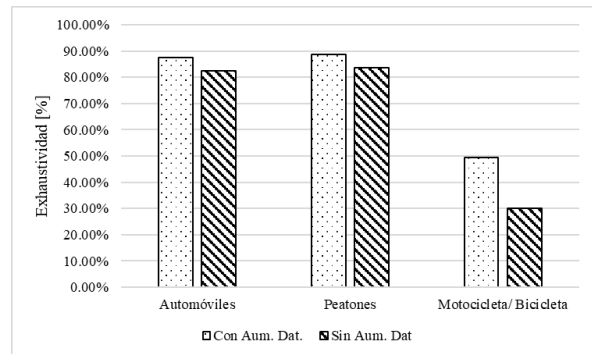


Fig. 7. Comparación de exhaustividad.

Como se puede observar en la figura 7, al igual que el porcentaje de precisión, los resultados siguen siendo mejores al utilizar aumento de datos. Dichos porcentajes representan la cantidad de las muestras que se clasificaron correctamente.

D. Monitoreo de Accesos

El monitoreo de los accesos de la aplicación en uso fue creado a través de Python en el entorno de programación de Jupyter. Para las pruebas preliminares del contador de accesos, se definió el área donde se harían las primeras pruebas del funcionamiento de la zona monitoreada.

Al marcar las líneas delimitadoras en la aplicación de clasificación de objetos, se obtiene un contador de entradas y salidas que a la vez cumple la función de identificar y asignarle una variable a cada uno de los objetos clasificados, como se puede observar en las figuras 8 y 9.

La tabulación individual de cada uno de estos objetos es llevada a cabo por el software al ir haciendo una suma o resta y presentándola en la imagen a medida van cambiando los valores.

E. Resultados

Los resultados expuestos en esta sección son producto de extensivas pruebas realizadas utilizando grabaciones de las cámaras de video del local.



Fig. 8. Pruebas de Campo.



Fig. 9. Pruebas de Campo.

Se alcanza el objetivo principal, se crea una aplicación que es capaz de contar entradas y salidas que además también clasifica los objetos que están ingresando al local comercial.

F. Limitaciones y Dificultades

Se presentan dos principales limitaciones al momento de la clasificación de motocicletas/bicicletas. La primera y mayor dificultad yace en el hecho que la red neuronal le da prioridad a la persona que detecta que va conduciendo la motocicleta.

La segunda limitación se presenta en la tabulación de los datos cuando es una persona la que está en movimiento. Debido a que los humanos tienen un leve tambaleo al momento de desplazarse de un lado a otro, esto causa que el centroide de la imagen esté en constante cambio. Esta particularidad puede ocasionar que se registren falsos positivos en el momento que el peatón cruce una de las líneas delimitadoras ya que su centroide posiblemente cruce la línea múltiples veces.

V. CONCLUSIONES Y RECOMENDACIONES

- La aplicación cumple su función en un porcentaje alto sin embargo queda a discreción del usuario la utilización del mismo software puesto que la tabulación de los datos no tiene la precisión esperada.
- La iluminación artificial por las noches incrementa las posibilidades de que existan falsos positivos especialmente al momento de hacer la clasificación, la red neuronal se ve ligeramente afectada por la cantidad de luz y el posicionamiento del alumbrado público con respecto a las cámaras de vigilancia del local en particular.
- Las técnicas de aumento de información son una herramienta invaluable al momento de entrenar una red neuronal puesto que amplían considerablemente el conjunto de datos utilizado y obtener una mayor precisión.
- Se recomienda utilizar un conjunto de datos mayor al momento del entrenamiento de la red neuronal convolucional con el fin de intentar simular las condiciones óptimas y de esta manera alcanzar una mayor precisión sin sacrificar tanto la exhaustividad del sistema.
- Se recomienda reducir la cantidad de sombras a las que están expuestas las cámaras debido a que las mismas distorsionan y dificultan la identificación de objetos.

APLICABILIDAD

Debido a que el software de reconocimiento alcanza porcentajes de precisión elevados en el caso de automóviles y peatones, sin embargo, se le dificulta al momento de clasificar motocicletas/bicicletas, se concluye que el software descrito en la investigación cumple únicamente parte de su función. En el caso del local comercial utilizado para este estudio, se podría decir que, cumple su función más básica, el conteo de entradas y salidas, pero da resultados mixtos a la hora de clasificar ciertos vehículos lo cual afecta directamente en la factibilidad para ese negocio en particular.

VI. REFERENCIAS

- [1] B. Catucuamba y A. Quinga, “Diseño y construcción de un prototipo contador de personas para sitios cerrados, utilizando sensores PIR(Pirloelectric Infrared Radial) y sensores de ultrasonido.”, Quito, Ecuador, 2010.
- [2] S. Ishida, S. Liu, S. Tagashira, y A. Fukuda, “Design of simple vehicle counter using sidewalk microphones”, Kyushi University, Japón, 2016.
- [3] G. Ballesteros, “Verificación de vehículos mediante técnicas de visión artificial.”, 2014.
- [4] A. Krizhevsky, I. Sutskever, y G. Hinton, “ImageNetClassification with Deep Convolutional Neural Networks”, University of Toronto, 2012.
- [5] B. L. Jian, Z. Zhang, y Y. Xu, “A people counting method based on head detection and tracking”, Hong Kong, China, 2014.
- [6] J. Peña, “Sistema de Detección y Conteo de vehículos Utilizando Visión Artificial”, Piura, Perú, 2017.
- [7] J. Chamorro, “Sistema para conteo de personas basado en visión por computador”, Ibarra, Ecuador, 2017.
- [8] F. Sisalima, “Sistema para detección y conteo vehicular aplicando técnicas de Visión Artificial”, Loja-Ecuador, 2018.
- [9] I. Goodfellow, Y. Bengio, y A. Courville, *Deep Learning*. MIT-Press, 2016.
- [10] A. Rosebrock, “YOLO object detection with OpenCV”, 2018. .
- [11] J. Redmon y A. Farhadi, “YOLOv3: An Incremental Improvement”, University of Washington, 2018.
- [12] A. Rosebrock, *Deep Learning for Computer Vision with Python*, 1a ed. 2017.
- [13] J. Wang y L. Perez, “The Effectiveness of Data Augmentation in Image Classification using Deep Learning”, Stanford University, 2017.
- [14] Bertoleti, “<https://www.hackster.io/phfbertoleti/counting-objects-in-movement-using-raspberry-pi-opencv-015ba5>”, 2017. .
- [15] V. Padilla, “Entrevista a propietario de Brisas S.A”, 2019.