

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA



PUCP

**Estudio de la segmentación semántica para la navegación autónoma de un
vehículo que circula en las calles de la provincia de Huamanga
TRABAJO DE INVESTIGACIÓN PARA LA OBTENCIÓN DEL
GRADO DE BACHILLER EN CIENCIAS CON MENCIÓN EN
INGENIERÍA ELECTRÓNICA**

AUTOR

Lui Gustavo Pasapera Huamán

ASESOR:

Donato Andrés Flores Espinoza

Lima, noviembre, 2020

RESUMEN

La detección de objetos y geolocalización son puntos clave en los sistemas de visión por computadora y su aplicación para la conducción autónoma. Le permite a la computadora el análisis de entorno en la medida que detecta objetos que podrían obstaculizar una determinada ruta, trayectorias a través del reconocimiento de carriles y mayor exactitud para la geolocalización a través de la detección de objetos relevantes en un determinado entorno. Una de las herramientas para la visión por computadora es la implementación de las redes de aprendizaje profundo. En los últimos años han tenido mayor acogida por su capacidad en el poco esfuerzo en cuanto a su clasificación manual o métodos clásicos. Si bien es cierto que necesita millones de imágenes para obtener un resultado aceptable, esto se ha ido mejorando a través de la transferencia de aprendizaje ya que reduce los millones de imágenes a miles. Estos miles de imágenes se consideran como los datos locales las cuales se analizan, clasifican y re-entrenan con una arquitectura de red pre-entrada. El presente trabajo de investigación se basa en el estudio del uso de los aprendizajes adquiridos por cada red de aprendizaje profundo tales como Resnet-18, Resnet-50, Mobilenetv2 y Xception con el fin de aprovechar su acceso e implementación en estructuras de segmentación semántica basadas en redes convolucionales tales como Segnet y Deeplab.

ÍNDICE GENERAL

INTRODUCCIÓN.....	1
CAPÍTULO 1: Importancia de la visión artificial para la conducción autónoma.....	2
1.1 El aprendizaje profundo y los conjuntos de datos para el entrenamiento... 2	
1.2 Justificación.....	5
1.3 Objetivos.....	5
1.3.1 Objetivo general.....	5
1.3.2 Objetivos específicos.....	5
CAPÍTULO 2.....	6
2.1 Visión general.....	6
2.2 Visión por computadora.....	6
2.2.1 Funciones para la variación de datos.....	7
2.2.1.1 Rotación.....	7
2.2.1.2 Reflexión.....	7
2.2.1.3 Traslación.....	8
2.2.1.4 Escalamiento.....	8
2.2.2 Aprendizaje profundo.....	9
2.2.3 Transferencia de aprendizaje.....	10
2.2.4 Segmentación semántica.....	11
2.3 Modelo de solución.....	12
CONCLUSIONES.....	14
RECOMENDACIONES Y TRABAJOS FUTUROS.....	15
BIBLIOGRAFÍA.....	16

ÍNDICE FIGURAS

2.1	Métodos para la aplicación del aprendizaje profundo.....	10
2.2	Proceso para la transferencia de aprendizaje.....	11
2.3	Estructura de una arquitectura Segnet y Deeplab [22].....	12
2.4	Modelo de solución.....	13



INTRODUCCIÓN

La visión artificial le permite a la computadora el análisis de entorno en la medida que detecta objetos que podrían obstaculizar una determinada ruta, trayectorias a través del reconocimiento de carriles y, mayor exactitud para la geolocalización a través de la detección de objetos relevantes en un determinado entorno. Uno de los métodos más usados últimamente es la transferencia de aprendizaje que se enfoca en utilizar arquitecturas de red pre-entrenadas de aprendizaje profundo con el fin de disminuir la cantidad de imágenes de una determinada base de datos. Esto es de gran ayuda en cuanto al análisis de una base de datos local ya que, se utiliza el aprendizaje pre adquirido de una determinada red y se la re-entrena con la nueva estructura que este caso se basa en la segmentación semántica por su nivel de eficiencia en la conducción autónoma.

Una de las redes de aprendizaje profundo aplicables a la segmentación semántica es la de Resnet18. Esta red se modificó a una arquitectura de DeepLab y se entrenó con bases de datos tales como Camvid. Arquitecturas de aprendizaje profundo tales como Resnet18, Resnet50, Mobilenetv2 y Xception también son aplicables a esta modificación y entrenamiento de la misma. Por lo que, se podrían implementar estas arquitecturas y evaluarlas con respecto a un conjunto de datos locales.

CAPÍTULO I

Importancia de la visión artificial para la conducción autónoma

En los últimos años, empresas de las industrias automotrices han mostrado sus desarrollos en el campo de los vehículos autónomos, proporcionando prototipos y productos para diversas aplicaciones tales como taxis autónomos y vehículos de reparto. Por ejemplo, empresas como Ocado, Walmart, Ford, Amazon vienen desarrollando nuevos sistemas de entrega autónoma que facilitan el servicio de reparto. Estas facilidades son las de realizar la entrega con vehículos silenciosos, automatizados, controlados y libres de emisiones que dañan el medio ambiente [1], [2]. Una componente importante para la conducción autónoma es la visión artificial o visión por computadora, pues esta le permite al vehículo emular la visión propia de las personas. Para una computadora, esta forma de trabajo se da a través del procesamiento de las imágenes digitales. Es decir, interpreta los datos captados por los sensores tales como cámaras y sensores para que así se puedan definir qué objetos se están presentes en un determinado entorno. Este proceso se realiza a través de la extracción de características y clasificación de las imágenes.

1.1 El aprendizaje profundo y los conjuntos de datos para el entrenamiento

Hoy en día, el aprendizaje autónomo permite que las aplicaciones de software mejoren, en cuanto a su precisión y predicción de resultados [6, p. 2]. Se puede decir que la programación en este aspecto, ha ido cambiando, ya que; por ejemplo, con este nuevo sistema ya no se usan miles de sentencias condicionales; sino que, es el aprendizaje autónomo el que se encarga de realizar este tipo de tareas.

Dentro de este sistema de aprendizaje autónomo, dos técnicas que han ido variando con respecto al uso e implementación. Estas técnicas son el aprendizaje profundo y el uso de clasificadores manuales. Su uso y aplicación depende de las necesidades y aplicaciones donde se desean implementar y estudiar.

El uso de clasificadores manuales implica el entrenamiento a través de clasificación de datos estructurados mediante etiquetas y patrones característicos. Así mismo, los clasificadores manuales aprenden a clasificar cuando se le indica el resultado frente a una entrada y una respuesta. Por ejemplo, para saber que se está procesando la imagen de un perro, se le relaciona con datos de perros (ejemplos similares de respuesta) previamente registrada en un conjunto de datos. Seguidamente, se le hace saber manualmente al sistema que es un perro a través de características individuales y así, se refuerza más el algoritmo de reconocimiento de perros y su predicción futura. De esta forma, su aplicación abarca el reconocimiento de personas, objetos, correos no deseados, etc. a través de un pre-entrenamiento y uso de clasificadores manuales. Los métodos clásicos para este aprendizaje son SVM, Random Forest, KNN y HOG.

El aprendizaje profundo, no utiliza, necesariamente, un conjunto de datos estructurados para su proceso de clasificación; es decir, no existen datos a priori caracterizados manualmente. Este método proporciona un conocimiento a partir de varias entradas, las filtra por cada capa de la red neuronal de manera que las va jerarquizando a medida de la similitud entre ellas. Su nivel de efectividad, en comparación a métodos clásicos, tiene mayor impacto por el gran número de capas. Los métodos de mayor uso son k-means, codificadores automáticos, agrupación jerárquica, etc. [6, p. 2].

En los últimos años, el aprendizaje profundo viene teniendo mayor acogida por su nivel de efectividad. Su integración a la visión por computadora tiene un alto potencial de poder generar soluciones sólidas y de mayor accesibilidad para la nueva tendencia de conducción autónoma [10],[11],[12], [14], [15]. La principal ventaja es la alta capacidad de realizar procesos de clasificación sin la necesidad de extraer manualmente las características de cada dato a procesar; es decir, no necesitan tener necesariamente una respuesta por entrada y, además, pueden ser más eficientes cuando de entornos cambiantes se trata. Esto es porque el aprendizaje profundo se entrena con un alto conjunto de datos que se pueden representar como

millones de imágenes, lo que significa que se entrenan con mayor número de entornos y características. De esta forma, se adquiere un mayor conocimiento a nivel computacional y entrenamiento por parte de una arquitectura de aprendizaje profundo. Sin embargo, la tarea disminuida por parte del aprendizaje y profundo y lo que respecta a su no extracción manual de características se podría ver en desventaja al ser necesario que se tengan millones de imágenes para el respecto entrenamiento de una red neuronal. Este punto ha sido muy estudiado por lo que ya se han implementado grandes conjuntos de datos disponibles en internet tales como Cityspaces dataset e Imagenet. Estos grandes conjuntos de datos se utilizan muy a menudo para la evaluación de las arquitecturas de redes neuronales profundas.

En cuanto al tema de interés de la presente tesis, el cual es la visión por computadora, podemos decir que el uso de los métodos tradicionales puede implicar mucho atascamiento en cuanto a su implementación, pues la variación de la forma de los carros, la iluminación, afectan en gran medida su modo de implementación en cuanto a su entrenamiento e interpretación [13]. Esto implica que muchas veces no es muy óptimo implementar un sistema que utilice técnicas tradicionales (Random Forest, KNN, etc.) cuando se trata de entornos con características que puedan variar.

Por otro lado, existe una gran diferencia entre usar datos tales como Cityspaces Dataset e Imagenet y, los datos locales. No siempre pueden funcionar arquitecturas de aprendizaje profundo, entrenadas con datos globales, cuando se quiere estudiar una determinada área local. Esto es por la variación de los objetos a estudiar ya que no son iguales en todas las ciudades e incluso pueden variar en color y forma. Frente a este problema, es necesario añadir un conjunto de datos locales para que así, las arquitecturas de redes neuronales se puedan adaptar y tener un mejor desempeño en un entorno local. Por lo tanto, hay una necesidad de desarrollar y crear un conjunto de datos locales concernientes a los objetos dentro de la provincia de Huamanga y seguidamente evaluar su precisión con arquitecturas neuronal de aprendizaje profundo.

1.2 Justificación

El estudio de un determinado entorno local es primordial para la visión por computadora ya que determina si una determinada arquitectura neuronal puede funcionar de forma correcta. La ampliación de un conjunto de datos globales a través de los locales, implica una mayor precisión y desempeño por parte de una arquitectura neuronal. Es necesario el estudio en cada tipo de entorno, pues los métodos y entrenamiento aplicados a experimentos ya hechos fueron diseñados para un sistema para poder determinar la eficiencia y precisión de esta técnica [16]. En este sentido, se desea implementar y desarrollar un conjunto de datos locales en la provincia de Huamanga y evaluarlos con arquitecturas neuronales de redes profundas que permitan el estudio e implementación de futuros proyectos de sistemas de conducción autónoma de vehículos que circulan en las calles de Huamanga.

1.3 Objetivos

1.3.1 Objetivo general

Realizar un estudio sobre el reconocimiento de objetos en las calles de la provincia de Huamanga a través de redes pre-entrenadas

1.3.2 Objetivos específicos

- Realizar un estudio de redes de aprendizaje profundo para la visión por computadora
- Investigar sobre el hardware y software para la adquisición y procesamiento de imágenes
- Realizar estudios sobre métodos de entrenamiento y acoplamiento de arquitecturas de segmentación semántica

CAPÍTULO 2

Marco teórico

2.1 Visión general

La detección de objetos y geolocalización son puntos clave en los sistemas de visión por computadora y su aplicación para la conducción autónoma. Le permite a la computadora el análisis de entorno en la medida que detecta objetos que podrían obstaculizar una determinada ruta, trayectorias a través del reconocimiento de carriles y, una mayor exactitud para la geolocalización a través de la detección de objetos relevantes en un determinado entorno tales como edificios, veredas y pistas. En cuanto al conjunto de datos, es de suma relevancia que estos varíen en forma, tamaño, texturas, brillo, etc. Esto se puede realizar tanto manualmente como a nivel de computadora. Para el primer caso, se adquieren los datos en diferentes entornos y diferentes ángulos; por ejemplo, se podrían tomar datos de una determinada avenida en diferentes horarios y así obtener imágenes con diferentes brillos y formas. Para el segundo caso, se podrían enfocar los datos en un proceso de cambio de brillo, texturas y formas con el uso de la computadora. Todos estos cambios y variaciones hacen que el entrenamiento de cada red basada en el aprendizaje profundo, sea más eficiente. Para la clasificación y detección de imágenes se utilizará la segmentación semántica que permite analizar la imagen de forma más exacta. Su arquitectura de red, está basada en el uso de redes pre-entrenadas de redes de aprendizaje profundo. Esto implica la aplicación de transferencia de aprendizaje que permite ahorrar costos en cuanto a la adquisición de datos.

2.2 Visión por computadora

El término visión por computadora, dentro del campo de la inteligencia artificial, se basa en un conjunto de técnicas y uso de modelos orientados al procesamiento, análisis e interpretación

de imágenes digitales para un determinado entorno o medio que se desea estudiar [17]. Sus aplicaciones se pueden orientar a las imágenes médicas, agricultura y vehículos autónomos.

2.2.1 Funciones para la variación de datos

En cuanto a la variación a nivel de computadora del conjunto de datos, se muestra a continuación, funciones que permiten obtener imágenes rotadas, reflejadas, trasladadas y escaladas.

2.2.1.1 Rotación. - Es el proceso en el cual se permite cambiar el sentido de una imagen a través de un ángulo especificado. Dentro del preprocesamiento de imágenes digitales, se utiliza para obtener una mejor vista de la imagen. Se muestra a continuación su forma matemática.

$$X_n = \cos(\theta) * (X_a - X_o) - \sin(\theta) * (Y_a - Y_o) + X_o \quad (1)$$

$$Y_n = \sin(\theta) * (X_a - X_o) + \cos(\theta) * (Y_a - Y_o) + Y_o \quad (2)$$

donde (X_o, Y_o) son las coordenadas del centro de rotación, θ es el ángulo de rotaciones en sentido horario que tienen los ángulos positivos, (X_a, Y_a) representan la posición inicial de un elemento de una imagen y (X_n, Y_n) , la posición nueva para un elemento de una imagen [18].

2.2.1.2 Reflexión. - Es el proceso que permite hacer el efecto espejo para una determinada imagen. Se muestra a continuación su forma matemática.

- Reflexión sobre el eje vertical de abscisa X_o .

$$X_n = -X_a + (2 * X_o) \quad (3)$$

$$Y_n = Y_a \quad (4)$$

- Reflexión sobre el eje vertical de abscisa Y_o .

$$X_n = X_a \quad (5)$$

$$Y_n = -Y_a + (2 * Y_o) \quad (6)$$

- Reflexión sobre el eje orientado en cualquier dirección arbitraria θ que pasa por (X_o, Y_o) .

$$X_n = X_a + 2 * \Delta * (-\sin(\theta)) \quad (7)$$

$$\Delta = (X_a - X_o) * \sin(\theta) - (Y_a - Y_o) * \cos(\theta) \quad (8)$$

donde (X_o, Y_o) son las coordenadas de las abscisas por las cuales se hará la reflexión, θ es el ángulo de rotaciones en sentido horario que tienen los ángulos positivos, (X_a, Y_a) representan la posición inicial de un elemento de una imagen y (X_n, Y_n) , la posición nueva para un elemento de una imagen [18].

2.2.1.3 Traslación. - Es el proceso en el cual se le asigna a cada elemento de la imagen, una nueva posición representada en la imagen de salida los pixeles de los datos. Se muestra a continuación su forma matemática.

$$X_n = X_a + X_x \quad (9)$$

$$Y_n = Y_a + Y_y \quad (10)$$

donde (X_x, Y_y) representan el desplazamiento para cada elemento de la imagen, (X_a, Y_a) , la posición inicial de un elemento de una imagen y (X_n, Y_n) , la posición nueva para un elemento de una imagen [18].

2.2.1.4 Escalamiento. - Se utiliza para ampliar o reducir el tamaño o parte de una determinada imagen. Puede ampliarse mediante dos métodos: la replicación de pixeles o la interpolación. El primero reemplaza cada pixel de la

imagen original por un grupo de píxeles del mismo valor, mientras que el segundo, lo hace a través de un grupo expandido de píxeles.

2.2.2 Aprendizaje profundo

El término aprendizaje profundo, es un tipo de aprendizaje de máquina que aprende a realizar tareas de clasificación directamente a partir de datos digitales. Las redes convolucionales, son muy eficientes cuando se desea el reconocimiento de objetos. El aprendizaje profundo es una herramienta muy útil para la visión por computadora ya que permite la detección de objetos de manera muy eficiente. De manera general, cada arquitectura de red posee 3 partes principales: La entrada, detección de características y la clasificación. La detección de características posee capas tales como las convolucionales, unidad lineal rectificadora y de agrupación. Su estructura y orden jerárquico depende de cada arquitectura.

- Convulsión. – ubica las imágenes en la entrada y las relaciona a través de filtros convolucionales. Estos filtros activan características de interés de cada imagen.
- Unidad lineal rectificadora. – Asigna el valor de cero (0) a los valores negativos.
- Agrupación. –Simplifica la salida con un muestreo no lineal en la forma de eliminar la cantidad de parámetros que una red necesita para realizar el proceso de aprendizaje.

La clasificación utiliza las siguientes capas.

- Capa totalmente conectada. – genera un número 'N' de conexiones que representas el número de clases que la red podrá predecir.

- Capa de función exponencial normalizada. – Proporciona una salida para la clasificación.

2.2.3 Transferencia de aprendizaje. – este término implica el uso de un conocimiento ya existente para generar uno nuevo. Dentro del campo de la inteligencia artificial, es posible reutilizar una red para reentrenarla solo si se van a utilizar clases similares dentro de una determinada categoría; por ejemplo, si se tiene una red pre-entrenada para la detección de flores, y esta no ha sido entrenada con algunas variedades de flores, es posible reutilizar esta red para que se puedan detectar las nuevas variedades de flores.

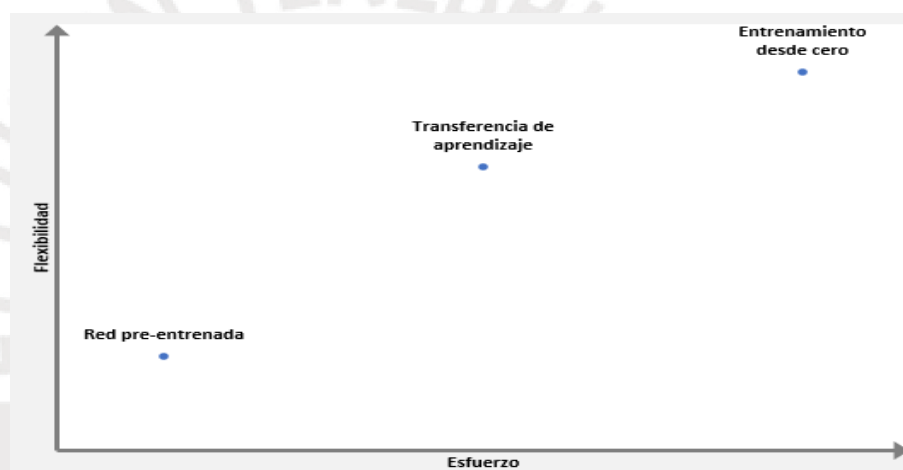


Figura 2.1. Métodos para la aplicación del aprendizaje profundo. Fuente: Propia

En la figura 2.1, se justifica su uso como gran potencial para la implementación de proyectos. La implementación de solo una red pre-entrenada es de poco esfuerzo; sin embargo, posee poca flexibilidad ya que se depende directamente del proceso de entrenamiento de una red, sin la posibilidad de cambio o reajuste. El entrenamiento desde cero implica una mayor flexibilidad. Su uso e implementación es favorable siempre y cuando no existan redes pre-entrenadas que puedan clasificar imágenes de interés por clase. Como punto adicional, se puede decir que es posible el uso de una arquitectura de red para un

procesamiento desde cero; sin embargo, se perdería el aprendizaje, medido en el peso por cada clase, de una determinada arquitectura de red. El proceso de transferencia de aprendizaje se detalla a continuación.

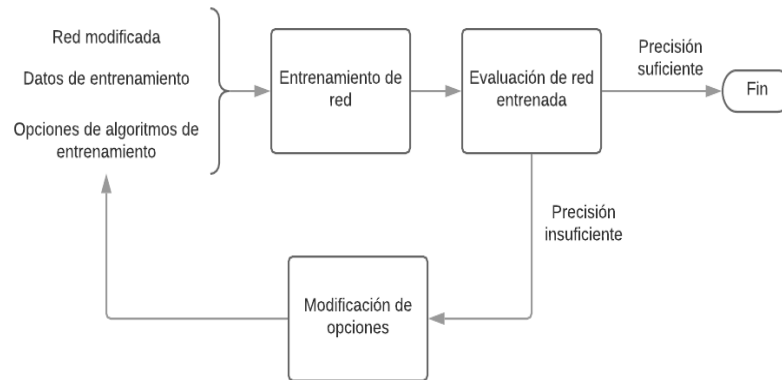


Figura 2.2. Proceso para la transferencia de aprendizaje. Fuente: Propia

En la primera parte, se puede apreciar la Red modificada, los datos de entrenamiento y opciones de algoritmos de entrenamiento. Estos representan una arquitectura ya modificada. Seguidamente se pasa a reentrenar la red y se procede a evaluar la precisión hasta que se logre un resultado aceptable.

- 2.2.4 Segmentación semántica. - La segmentación semántica resulta muy útil cuando se trata de analizar a profundidad conjuntos de datos digitales ya que, se analiza la imagen a nivel de píxeles. Dentro de estas arquitecturas de red, se encuentran algunas como Segnet y Deeplab. Estas arquitecturas se pueden reutilizar para el entrenamiento a través de redes convolucionales. Es decir, aprovechan el entrenamiento previo por parte de las redes de aprendizaje profundo y ajustan su propia arquitectura al entrenamiento en el campo de la segmentación semántica.

Las arquitecturas Segnet y Deeplab se asemejan mucho, pues la arquitectura Deeplab es una modificación en cuanto al número de capas de la arquitectura Segnet. Su estructura se muestra a continuación.

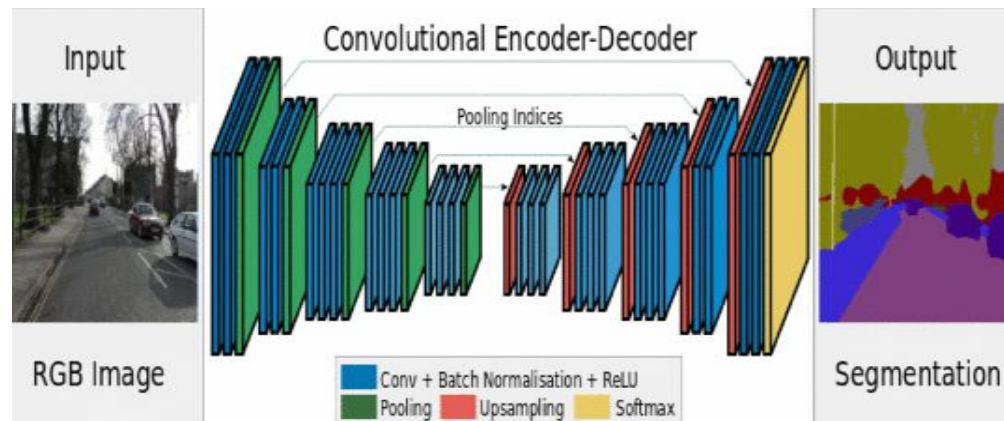


Figura 2.3. Estructura de una arquitectura Segnet y Deeplab. Fuente: [22]

El codificador extrae características de una imagen digital a través de una secuencia de filtros cada vez más estrechos y profundos. En el codificador utiliza arquitecturas de aprendizaje profundo pre-entrenadas con el fin de implementar la transferencia de aprendizaje y aprovechar los pesos de cada red. La capa de muestreo (Upsampling) realiza el proceso inverso a la de agrupación que se explicó en el punto 2.2.2. Esto hace que el codificador realice un proceso inverso al codificador hasta llegar a una resolución de píxeles igual al de la imagen de entrada.

2.3 Modelo de solución

La propuesta del modelo de solución se basa en utilizar los aprendizajes adquiridos por cada red de aprendizaje profundo tales como Resnet-18, Resnet-50, Mobilenetv2 y Xception. Esto con el fin de aprovechar su acceso e implementación en estructuras basadas en redes convolucionales tales como Segnet y Deeplab. El entorno que se desea estudiar es la provincia de Huamanga por lo que queda como pendiente

desarrollar un conjunto de datos de al menos 1200 imágenes. De la misma forma se utilizarán funciones para el aumento de datos tales como los rotación, traslación y reflexión explicados en el punto 2.2.1. Por otro lado, se tendrán en cuenta las opciones de entrenamiento tales como el minibatchsize y el número de épocas y sus variaciones conforme vaya avanzando el entrenamiento de la red. Finalmente, en caso de ser necesario, se incrementarán el número de imágenes para el entrenamiento adecuado de cada red de aprendizaje profundo en su implementación a la segmentación semántica.

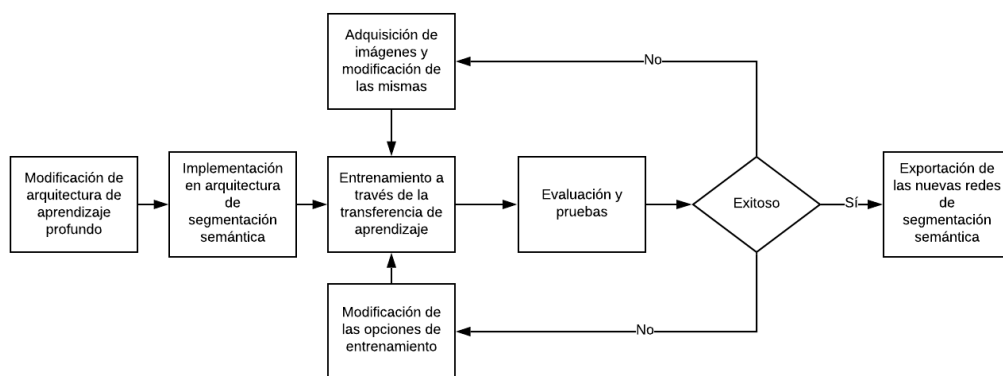
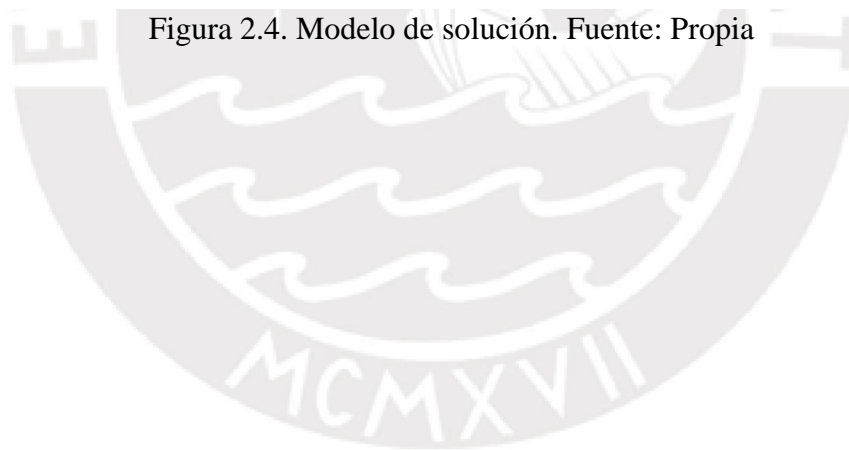


Figura 2.4. Modelo de solución. Fuente: Propia



CONCLUSIONES

- Se puede concluir que, si se quiere analizar la visión por computadora en un entorno local en cuanto a sus clases, es más factible la transferencia de aprendizaje ya que utiliza un número muy reducido de imágenes con respecto al del aprendizaje desde cero y presenta una gran efectividad en cuanto a precisión.
- El número de imágenes necesarias para realizar este proceso debe ser proporcional a las variaciones del entorno y, éstas deben variar en cuando a luminosidad, giro y forma. Esto ayuda a que la red re-entrada con redes de aprendizaje profundo pre-entrenadas tenga una mayor precisión cuando el entorno local varíe con las características mencionadas anteriormente.
- Se tienen que tener dos bases de datos. La primera debe contener imágenes tomadas por diferentes cámaras con el fin de aumentar la variabilidad de las mismas. Y finalmente, la segunda debe contener información sobre la segmentación manual referida a cada imagen colocada dentro de la primera base de datos.
- La evaluación debe enfocarse en cuanto a la precisión total y parcial de todas y cada clase respectivamente. Las métricas de clasificación recomendadas son la exactitud y el IoU.

RECOMENDACIONES Y TRABAJOS FUTUROS

- Es recomendable que, al utilizar arquitecturas pre-entrenadas, estas tengan dentro de sus clases, las clases que se desean clasificar en cuanto a la transferencia de aprendizaje.
- Se recomienda conservar el dimensionamiento de las imágenes adquiridas por los diferentes tipos de cámaras. En caso de reducirlas o incrementarlas asignarles un valor proporcional y, en el peor de los casos aplicar un relleno con ceros de tal forma de no perder el dimensionamiento proporcional de cada imagen.
- El presente trabajo de investigación forma una de cinco principales partes de la conducción autónoma. Su estudio puede acoplarse de manera flexible a la implementación de un sistema total de conducción autónoma.

BIBLIOGRAFÍA

- [1] C. Hopping, "Ocado trials self-driving deliveries in London," *IT Pro*, 2017. [En línea], Disponible en: <https://www.itpro.co.uk/strategy/28963/ocado-trials-self-driving-deliveries-in-london> [Accedido: 02-oct-2019]
- [2] Anonymous "SoftBank fund invests big in self-driving deliveries," *AFP International Text Wire in English*, 2019. [En línea], Disponible en: <https://phys.org/news/2019-02-softbank-fund-invests-big-self-driving.html> [Accedido: 02-oct-2019]
- [3] Anonymous "Ford to test self-driving deliveries with postmates," *Manufacturing.Net*, 2018. [En línea], Disponible en: <https://www.manufacturing.net/news/2018/01/ford-test-self-driving-deliveries-postmates> [Accedido: 02-oct-2019]
- [4] A. García, "Evaluación del uso de la tecnología LIDAR en la estimación del contenido de carbono forestal," Universidad de Alcalá [En línea]. Disponible en: http://www.redccadgiz.org/documentos/doc_1350723744.pdf [Accedido: 03-oct-2019]
- [5] A. González *et al*, "Técnicas y algoritmos básicos de visión artificial," Universidad de la Rioja Press, 2006 [En línea]. Disponible en: <https://publicaciones.unirioja.es/catalogo/online/VisionArtificial.pdf> [Accedido: 03-oct-2019]
- [6] M. Perc, M. Ozer and H. Hojnik, "Social and juristic challenges of artificial intelligence," *Palgrave communications*, vol. 61, pp. 1-3, 2019 [En línea]. Disponible en: <https://www-nature-com.ezproxybib.pucp.edu.pe/articles/s41599-019-0278-x> [Accedido: 03-oct-2019]
- [7] Amazon Web Service, 2019 [En línea]. Disponible en: https://aws.amazon.com/es/?nc2=h_lg [Accedido: 03-oct-2019]
- [8] ImageNet [En línea]. Disponible en: <http://www.image-net.org/about-overview> [Accedido: 03-oct-2019]
- [9] N. Akhtar and A. Mian, "Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey," in *IEEE Access*, vol. 6, pp. 14410-14430, 2018. doi: 10.1109/ACCESS.2018.2807385
- [10] P. Viswanath, S. Nagori, M. Mody, M. Mathew and P. Swami, "End to End Learning based Self-Driving using JacintoNet," 2018 IEEE 8th International Conference on Consumer Electronics - Berlin (ICCE-Berlin), Berlin, 2018, pp. 1-4. doi: 10.1109/ICCE-Berlin.2018.8576190
- [11] T. Okuyama, T. Gonsalves and J. Upadhyay, "Autonomous Driving System based on Deep Q Learning," 2018 International Conference on Intelligent Autonomous Systems (ICoIAS), Singapore, 2018, pp. 201-205. doi: 10.1109/ICoIAS.2018.8494053
- [12] S. Chen, "Multimedia for Autonomous Driving," in *IEEE MultiMedia*, vol. 26, no. 3, pp. 5-8, 1 July-Sept. 2019. doi: 10.1109/MMUL.2019.2935397
- [13] G. Prabhakar, B. Kailath, S. Natarajan and R. Kumar, "Obstacle detection and classification using deep learning for tracking in high-speed autonomous driving," 2017 IEEE Region 10 Symposium (TENSYMP), Cochin, 2017, pp. 1-6. doi: 10.1109/TENCONSpring.2017.8069972
- [14] D. Dong, X. Li and X. Sun, "A Vision-Based Method for Improving the Safety of Self-Driving," 2018 12th International Conference on Reliability, Maintainability, and Safety (ICRMS), Shanghai, China, 2018, pp. 167-171. doi: 10.1109/ICRMS.2018.00040
- [15] R. Kulkarni, S. Dhavalikar and S. Bangar, "Traffic Light Detection and Recognition for Self-Driving Cars Using Deep Learning," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-4. doi: 10.1109/ICCUBEA.2018.8697819

[16] Z. Unnisa, Z. Akhtar, H. Riaz and T. Zulfiqar, "Obstacle detection for self-driving car in Pakistan's perspective," *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Sukkur, 2018, pp. 1-8. doi: 0.1109/ICOMET.2018.8346334

[17] Mathworks, "Computer Vision", 2018. [En línea]. Disponible en: <https://es.mathworks.com/products/computer-vision.html>. [Accedido: 20-nov-2019]

[18] R. Gonzales y R. Woods, "Digital Image Processing," en *an introduction to the Mathematical Tools Used in Digital Image Processing*, vol 3, New York: Pearson,2008, pp. 93-96

[19] R. Gonzales y R. Woods, "Digital Image Processing," en *Intensity transformations and spatial filtering*, vol 3, New York: Pearson,2008, pp. 105-160

[20] A. Canziani, E. Culurciello y A. Paszke, "An análisis of Deep neural network models for practical applications", Indiana, 2017, pp. 1-2. [En línea]. Disponible en <https://arxiv.org/pdf/1605.07678.pdf>. [Accedido: 20-nov-2019]

[21] Mathworks, "Semantic Segmentation", 2018. [En línea]. Disponible en: https://es.mathworks.com/solutions/image-video-processing/semantic-segmentation.html?s_tid=srchtitle. [Accedido: 20-nov-2019]

[22] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, 1 Dec. 2017. doi: 10.1109/TPAMI.2016.264461

