Ciencia Latina Revista Científica Multidisciplinar
Noviembre-Diciembre, 2023, Volumen 7, Número 6
https://doi.org/10.37811/cl_rcm.v7i6.9118

Sistema de Reconocimiento de Objetos como Apoyo a Personas Invidentes

Diana Paulina Martínez Cancino[1]
diana.mar.can@gmail.com
https://orcid.org/0000-0002-1087-9616
Universidad Politécnica de Chiapas
México

RESUMEN

La discapacidad involucra diferentes limitantes con las que convive día a día el individuo que la sufre. La discapacidad visual hace referencia a una pérdida parcial o total de la visión, mediante la cual el individuo interactúa con los elementos de su entorno. Se estima que aproximadamente el 48% de personas que padecen algún tipo de discapacidad presentan limitantes visuales, lo cual les ocasiona dificultades en actividades y espacios cotidianos. Uno de los ámbitos más afectados para estas personas constituye la movilidad, dados los obstáculos variados con los que se pueden encontrar. Para brindar alternativas a esta problemática existen herramientas como los bastones que le permiten al usuario detectar objetos y poder esquivarlos. Sin embargo, estas herramientas presentan limitantes que, en últimos años, han intentado ser cubiertas por técnicas de visión artificial. Este proyecto implementa redes neuronales a través de algoritmos de visión para construir un sistema que pueda ser de utilidad para la población en mención. Los resultados obtenidos nos muestran que hacer uso del algoritmo YOLO en su versión 5 resulta efectivo en un 94.29% para detectar objetos a través de una cámara web y que incluso se pueden realizar detecciones de diferentes objetos al mismo tiempo. Este tipo de algoritmos pueden ser implementados como sistemas móviles para facilitar el tránsito a los usuarios con discapacidad visual. Finalmente, es importante incluir la mayor cantidad de categorías posibles al desarrollar el algoritmo para asegurar que la persona cuente con suficiente información para realizar su desplazamiento.

Palabras clave: algoritmo YOLO; redes neuronales; visión artificial

[1] Autor principal.
Correspondencia: diana.mar.can@gmail.com

Object recognition system as support for blind people

ABSTRACT

Disability involves different limitations with which the individual who suffers it lives on a daily basis. Visual impairment refers to a partial or total loss of vision, through which the individual interacts with the elements of their environment. It is estimated that approximately 48% of people with some type of disability have visual limitations, which causes difficulties in daily activities and spaces. One of the most affected surroundings for these people is mobility, given the various obstacles they may encounter. To provide alternatives to this problem, there are tools such as sticks that allow the user to detect objects and be able to avoid them. However, these tools have limitations that, in recent years, have tried to be covered by artificial vision techniques. This project implements neural networks through vision algorithms to build a system that can be useful for the population in question. This project implements neural networks through vision algorithms to build a system that can be useful for the population in question. The results obtained show us that using the YOLO algorithm in its version 5 is 94.29% effective in detecting objects through a webcam and that it can even detect different objects at the same time. This type of algorithms can be implemented as mobile systems to facilitate transit for users with visual disabilities.

***Keywords*:** YOLO algorithm; artificial vision; neural networks

Artículo recibido 18 noviembre 2023
Aceptado para publicación: 30 diciembre 2023

INTRODUCCIÓN

La Organización Panamericana de la Salud (OPS) (2020) menciona que la discapacidad se trata de diversas limitantes físicas, mentales, intelectuales o sensoriales con las que llega a vivir un individuo y que pueden frenar o incluso disminuir su participación e interacción dentro de la sociedad en la que se desenvuelve. Estas limitaciones o deficiencias son atribuidas a cambios en las estructuras anatómicas que resultan en una disminución significativa o incluso en la pérdida de alguna función (Organización Mundial de la Salud, 2011).

De acuerdo con la Organización Mundial de la Salud (2011) las personas que viven con discapacidad son aquellas que experimentan diversas desigualdades en el ámbito de la salud, cuando se les compara con las personas sin discapacidad; debido a que presentan deficiencias de tipo físico, mental, intelectual o sensorial, las cuales prevalecen a lo largo de su vida. Estas deficiencias pueden llegar a suponer barreras que les limiten o priven totalmente de participar efectivamente en la sociedad en la que se desenvuelven bajo igualdad de condiciones en comparación con el resto de las personas que la integran. Se estima que alrededor del 16% de la población mundial experimenta una discapacidad significativa en la actualidad, lo cual llega a afectar el ejercicio de sus derechos, especialmente en el acceso a condiciones adecuadas de salud (Organización Mundial de la Salud, 2022). En el ámbito nacional, se estima que en México hay 6,179,890 personas con algún tipo de discapacidad, lo que representa 4.9% de la población total del país. De ellas, 53 % son mujeres y 47% son hombres (Instituto Nacional de Estadística y Geografía, 2020).

Dentro de los diferentes tipos de discapacidad podemos encontrar a la discapacidad visual, la cual se relaciona con la pérdida parcial o total del sentido de la vista, mediante el cual el ser humano es capaz de percibir información de su entorno para poder tomas decisiones. Se estima que aproximadamente 314 millones de personas presentan discapacidad visual en menor o mayor grado, alrededor del 48% de la población mundial con discapacidad (Organización Mundial de la Salud, 2020). La OPS menciona que durante el 2014 en América Latina y el Caribe por cada millón de habitantes se registraron 5,000 personas invidentes y 20,000 con discapacidad visual. Otro factor importante a tener en cuenta en esta población es que, a nivel mundial, al menos el 90% vive en condiciones de pobreza y bajos recursos, junto con el factor de edad, ya que se estima que el 82% de los individuos que han perdido la vista

alcanzan más de 50 años (Sareeka et al., 2018).

Las causas que producen la discapacidad visual son múltiples y dentro de las más frecuentes podemos encontrar errores de refracción que no fueron atendidos y corregidos en su debido momento, cataratas, degeneración macular relacionada con el factor de la edad, retinopatía diabética, glaucoma, opacidad de la córnea y tracoma (Hernández et al., 2019). Las personas que padecen algún tipo de discapacidad visual pierden independencia a distintos niveles, lo cual llega a afectar diferentes aspectos de su vida diaria. El lograr incluir dentro de la sociedad de forma efectiva a las personas con discapacidad depende tanto de las barreras actitudinales (Polo y Aparicio, 2018), como de adaptar el entorno a las necesidades y particularidades de cada persona, con la finalidad de abonar al desarrollo personal del individuo.

Las personas que presentan discapacidad visual enfrentan dificultades para desplazarse en espacios públicos ya que se cuenta con una planeación urbana deficiente en lo referente a la obstaculización de objetos, lo cual resulta en un impedimento para el libre tránsito (Laverde, 2013). Para poder llevar a cabo sus actividades referentes a movilidad urbana, las personas con discapacidad han optado por opciones tales como el uso de dispositivos de movilidad, siendo el más importante y utilizado el bastón, así como el apoyo de personas y perros guía. Todas estas opciones, si bien le brindan al usuario opciones para poder transportarse en la ciudad, presentan algunas desventajas a tener en cuenta.

Cuando se toma en cuenta el uso del bastón guía como apoyo a la movilidad urbana se tiene que considerar que es un aditamento que cubre la parte inferior del cuerpo, por lo que es útil para la detección de objetos que se encuentran por debajo de cierta altura. Sin embargo, varios de los objetos y obstáculos con los que se encuentra una persona en su transitar presentan bordes, extensiones y ángulos que pueden llegar a provocar lesiones y que no se pueden detectar con el bastón. Es por ello que resulta de vital importancia contar con herramientas que ayuden también a proteger el tren superior del cuerpo, a partir de poder brindar información del entorno a este nivel.

El objetivo de este proyecto consiste en desarrollar un sistema de apoyo a personas con discapacidad visual que sea capaz de proporcionarles la facilidad del reconocimiento de objetos y obstáculos a través del uso de redes neuronales. Para ello, se planteó un diseño cuasi-experimental con un enfoque cuantitativo utilizando una muestra de 40 datos para las pruebas de la red neuronal entrenada, en donde se busca responder a la pregunta de si el algoritmo YOLO es eficiente al ser utilizado en un sistema de

reconocimiento de objetos en movimiento.

MARCO TEÓRICO

La visión por computadora constituye un componente de la Inteligencia Artificial (IA) el cual está enfocado en el desarrollo de diversas técnicas computacionales encargadas de analizar e interpretar datos visuales los cuales son obtenidos a través de imágenes por fotografía o videos (Prince, 2012). Lo que se busca al aplicar técnicas de visión por computadora es obtener información fidedigna del entorno que rodea al usuario.

Dentro del campo de visión por computadora se pueden distinguir dos diferentes niveles de procesamiento (Aramendiz et al., 2020):

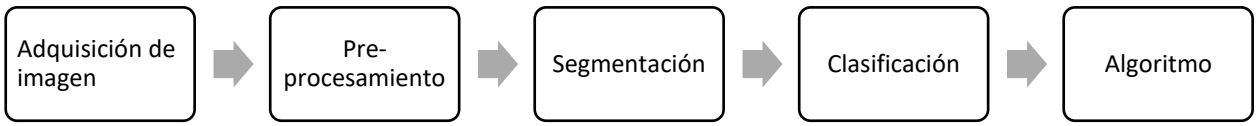
- Visión preliminar, cuyo objetivo consiste en identificar ciertas características de los diversos objetos del entorno en el cual se aplica. Tales características pueden ser posición, tamaño, color, entre algunas otras que son utilizadas para realizar una segmentación progresiva e identificar cambios en los parámetros de la imagen.
- Análisis de la escena, el cual consiste en la abstracción de alto nivel de las características obtenidas por la visión preliminar. En este nivel se pueden analizar formas, reconocer y localizar tanto objetos como patrones que puedan ser utilizados en procesos o taras más complejas dentro de las cuales se debe tomar una decisión.

Por otra parte, el reconocimiento de objetos busca analizar una escena y reconocer los objetos que se encuentran dentro de ella (Trujillo-Romero, 2022). Para ello, se utilizan algoritmos capaces de realizar la toma de deciones sobre objetos que son captadas por una o varias cámares a través de, principalmente, la detección de bordes.

Cuando se trabaja en la detección de bordes, ya sea a escala de grises o escalas a color, se sigue usualmente un proceso básico, como se observa en la Figura 1, en el que primero se debe realizar la captura de la imagen, ya sea a través de imágenes fijas o en movimiento. Una vez obtenida la imagen se deben emplear diferentes técnicas con la finalidad de mejorar la calidad de la imagen, dentro de ellas podemos encontrar la aplicación de filtros, mejora de contraste y reducción de ruido. Posterior al pre-procesamiento la imagen debe ser dividida en diversas regiones con la finalidad de aislar los objetos que resulten de interés y obtener los datos con los cuales se pueden describir rasgos de formas de vectores

numéricos. En la etapa siguiente, las imágenes segmentadas se asignan en categorías a partir de los rasgos obtenidos en etapas previas. Por último, para llevar a cabo el reconocimiento de formas es necesario hacer uso de un algoritmo de detección de bordes a partir de geometrías, dentro de los cuales podemos encontrar diversas opciones (Phadnis et al., 2018).

Figura 1. Proceso básico a seguir para el reconocimiento de objetos.

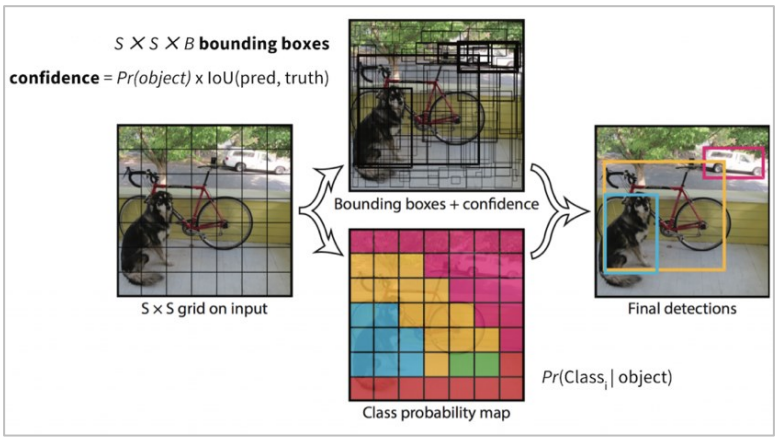


Fuente: Phadnis et al., 2018.

Una red neuronal de tipo convolucional es un algoritmo de aprendizaje profundo (deep learning) el cual está diseñado para manipular imágenes y distinguir objetos a partir de asignar niveles de importancia a elementos específicos que integran a la imagen. Estas redes llevan a cabo el aprendizaje en reconocimiento de diferentes objetos haciendo uso de un conjunto de capas ocultas, lo cual permite que las neuronas de la red distingan y entiendan las propiedades únicas de cada objeto, permitiendo así llevar a cabo la inferencia (Martínez, 2015).

Uno de los principales algoritmos, y que han mostrado ser más efectivos, en la detección de objetos en tiempo real es YOLO (You Only Look Once). Este algoritmo identifica objetos y sus posiciones con la ayuda de recuadros al observar la imagen en una única ocasión, de ahí el nombre. Esta identificación la lleva a cabo a través del uso de redes neuronales convolucionales que predicen la probabilidad de que una clase se presente, a partir de matemática aplicada y del uso de recuadros, en los objetos mostrados en una imagen (Diwan et al., 2022).

Figura 2. Funcionamiento algoritmo YOLO.

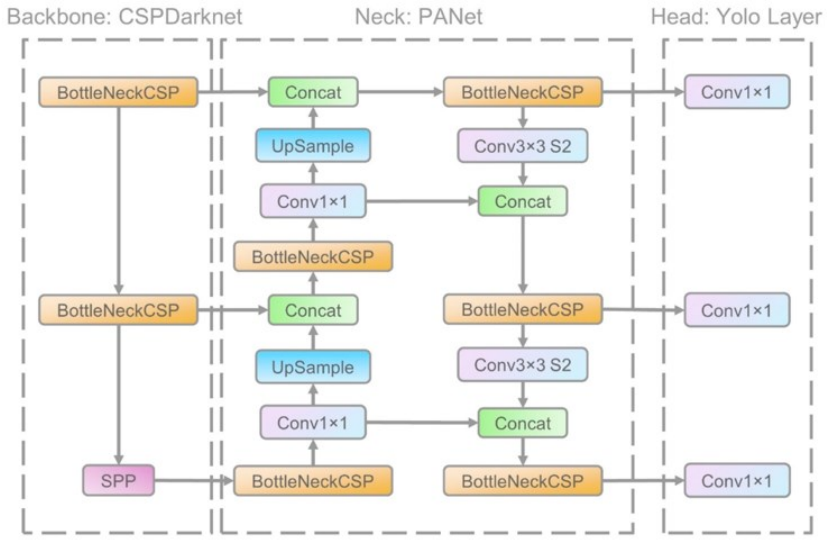


Fuente: Redmon et al., 2016.

Una ventaja de este algoritmo consiste en una reducción del error iteración tras iteración, ya que se van obteniendo más regiones y más segmentos a partir de los cuales la red va profundizando el aprendizaje a partir de las imágenes almacenadas. La imagen es dividida en S x S rejillas, como se observa en la Figura 2, cada rejilla predice B recuadros en conjunto con sus posiciones y dimensiones, probabilidad de un objeto en la rejilla subyacente y las probabilidades de las clases. El centro de un objeto debe permanecer dentro de la rejilla en la que se encuentra (Redmon et al., 2016).

Para extraer las características de la imagen YOLO V5 utiliza CSPDarknet como base, como se observa en la Figura 3. Para agregar las características. Para agregar las características y pasarlas al módulo Head de la predicción, crea una red de pirámides de características utilizando PANet. Intuitivamente, los datos alimentan al módulo Backbone, que consiste en aCSPDarknet, donde se extrae la característica y luego se pasan al módulo Neck, aPANet, donde se produce la fusión de características y, por último, la capa YOLO predice la salida de detección en términos de clase, ubicación, puntuación de confianza y tamaño (Xu et al., 2021).

Figura 3. Red de arquitectura de YOLOV5.

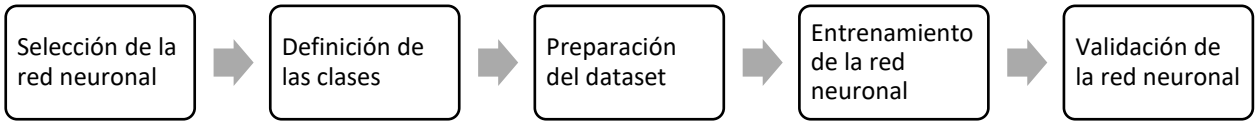


Fuente: Xu et al., 2021.

METODOLOGÍA

El desarrollo de este trabajo se dividió en las cinco etapas que se describen en la Figura 4.

Figura 4. Etapas del proyecto.



Fuente: Elaboración propia.

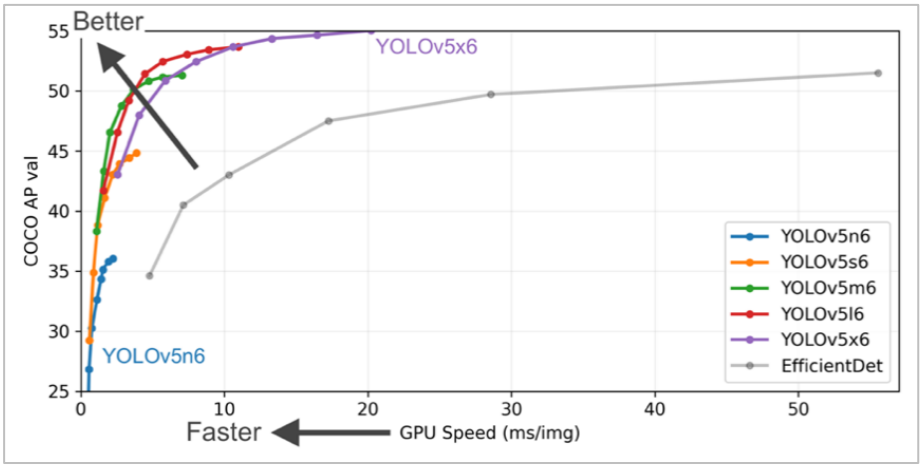
Selección de la red neuronal

Se utilizará un modelo preentrenado de YOLO en su versión 5, pudiendo elegir alguna de las siguientes opciones:

- Nano: 640 pixeles, 28.4 mAP y velocidad de 6.3 ms.
- Small: 640 pixeles, 37.2 mAP y velocidad de 6.4 ms.
- Medium: 640 pixeles, 45.2 mAP y velocidad de 8.2 ms
- Large: 640 pixeles, 48.8 mAP y velocidad de 10.1 ms
- XLarge: 640 pixeles, 50.7 mAP y velocidad de 12.1 ms

De las opciones que ofrece YOLOV5, se puede observar en la Figura 5 que la que ofrece un mejor rendimiento es la versión XLarge (YOLOv5x6), la más grande de ellas, por lo cual para este proyecto se hace uso de los pesos otorgados por esta versión.

Figura 5. Rendimiento de las versiones YOLOV5.



Fuente: Jocher, 2020.

Definición de las clases

Al situar la problemática de la movilidad de las personas con discapacidad visual, se tomaron en cuenta algunas de las clases que encontramos comúnmente al transitar por vialidades y que pueden resultar de mayor dificultad al momento de identificar dichos objetos.

Las categorías fueron entonces definidas de la siguiente forma: categoría 0 persona, categoría 1 carro, categoría 2 animal, categoría 3 árbol, categoría 4 semáforo, categoría 5 casa y categoría 6 poste.

Es posible incluir más categorías para brindarle mayor información al usuario y falicitarle la movilidad, teniendo en cuenta otros objetos, obtscáulos o barreras que se pudieran encontrar en las vialidades.

Preparación del dataset

Para la creación de la red neuronal se recopilaron 500 imágenes de 640 pixeles para cada una de las siete categorías, las cuales fueron obtenidas de diversos bancos de imágenes de internet. De estas imágenes el 80% fueron utilizadas para el entrenamiento de la red neuronal y el 20% para la validación. Para cada una de las clases se tomaron 400 imágenes para entrenamiento y 100 imágenes para validación, teniendo un total de 2800 imágenes para entrenamiento y 700 imágenes para validación.

Para poder llevar a cabo el entrenamiento de la red neuronal es necesario realizar el etiquetado de las imágenes. Para ello se utilizó la herramienta online gratuita Make Sense AI, en la cual, para cada una de las imágenes, se va señalando a partir de recuadros los objetos de interés. Como resultado de este etiquetado se obtiene un archivo de texto con los parámetros de interés para la red neuronal: la clase del objeto de interés señalado y la posición en la cual se encuentra dicho objeto dentro de la imagen.

Una vez etiquetadas las imágenes, es necesario organizar la información en una estructura de carpetas especificada, ya que este es el orden que tomará el modelo para llevar a cabo el entrenamiento de la red. Se cuenta con dos carpetas principales, una para las imágenes y otra para las etiquetas, dentro de las cuales se colocan dos carpetas adicionales, una para los datos de entrenamiento y otra para los datos de validación, las cuales serán utilizadas durante la etapa de entrenamiento.

Entrenamiento de la red neuronal

Para llevar a cabo el entrenamiento de la red neuronal es necesario personalizar los datos de entrada del modelo. Esta etapa fue realizada a través de Google Colab, herramienta a través de la cual se puede programar y ejecutar Python desde el navegador. Google Colab es Entorno de Desarrollo Integrado

(IDE) de libre acceso utilizado como apoyo a la investigación y al desarrollo de proyectos de inteligencia artificial.

Proporciona un entorno similar a Jupyter Notebook y tiene la libertad de utilizar la Unidad de Procesamiento Gráfico (GPU) y la Unidad de Procesamiento Tensorial (TPU). Google Colab tiene preinstaladas librerías utilizadas en Deep Learning como son OpenCV, PyTorch, TensorFlow y Keras. Este proyecto hace uso principalmente de OpenCV y PyTorch (Thuan, 2021).

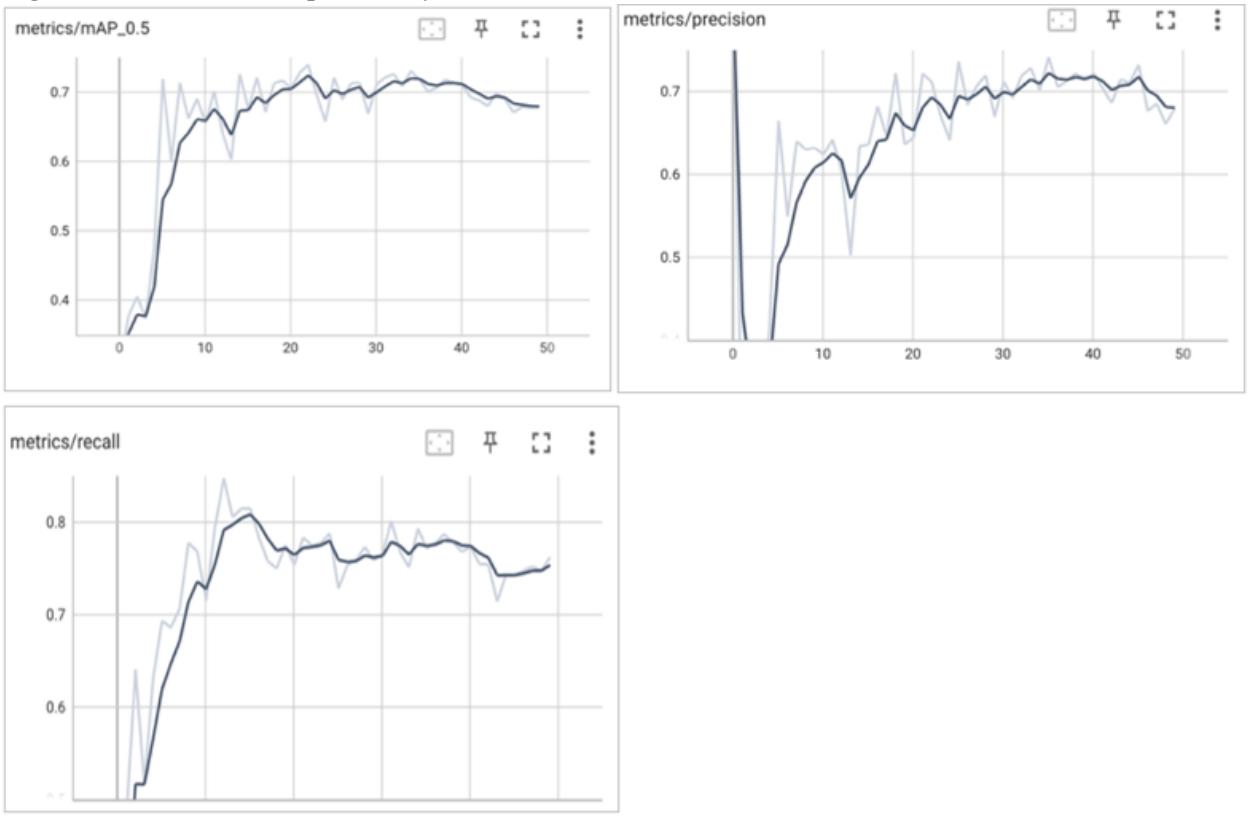
En este proyecto se implementó el algoritmo YOLO en su versión 5, por lo que se utilizó el modelo pre-entrenado obtenido de Ultralytics, sitio web enfocado en la creación de modelos de inteligencia artificial, personalizando el archivo COCO128.yaml, en el cual se deben definir las rutas en las cuales se encuentran la carpeta raíz, así como las carpetas de validación y de entrenamiento, de las cuales se obtendrán las capas de entrada.

Además, deben definirse las clases con las que deberá entrenarse el modelo, definiendo en este apartado las 7 clases establecidas en la etapa anterior. Finalmente, renombramos a este archivo como custom.yaml, para identificar las rutas personalizadas.

Para llevar a cabo el entrenamiento del modelo desde Google Colab es necesario definir algunos parámetros:

- img – tamaño de la imagen de entrada (640 pixels de acuerdo a YOLO V5).
- Batch – tamaño del batch (16).
- epochs – número de épocas (50).
- data – archivo YAML con los parámetros del data set (custom.yaml).
- weights – archivo de pesos relacionado con la versión de yolo a utilizar, en nuestro caso la versión XLarge (yolov5x).

Figura 6. Métricas mAP, precisión y recall obtenidas del modelo.



Fuente: Elaboración propia obtenida a través de Google Colab.

El modelo fue entrenado en un tiempo de 0.278 horas y se obtuvo una métrica mAP (precisión media) por arriba del umbral 0.5, como se muestra en la Figura 6, comprobando así el rendimiento del modelo respecto a su precisión (Shah, 2022).

Validación de la red neuronal

Una vez que la red neuronal ha sido entrenada, se toma el archivo que contiene los mejores pesos para proceder a probar el modelo. Para realizar la inferencia es necesario hacer uso de las librerías OpenCV, Torch y Numpy.

Se realizaron pruebas preliminares con imágenes obtenidas de la web y también se realizaron pruebas con imágenes obtenidas por cámara web, ya que este es el objetivo principal del proyecto, que el algoritmo, a través de una cámara web, pueda identificar los objetos con los que el usuario se encuentra en su transitar. Para ello, se hicieron 40 pruebas con objetos reales a través de la cámara web. De estas pruebas 35 se realizaron con un solo objeto de las clases que forman parte de la red neuronal y 5 se realizaron en combinaciones de objetos.

RESULTADOS Y DISCUSIÓN

Tabla 1. Pruebas realizadas a la red neuronal con un objeto, mostrando el nivel de confianza obtenido y si se trata de un acierto (1) o desacierto (0).

No. Prueba	Respuesta esperada	Respuesta obtenida	Nivel de confianza	Acierto
1	Persona	Persona	0.95	1
2	Carro	Carro	0.93	1
3	Animal	No detectó	0.86	0
4	Árbol	Árbol	0.87	1
5	Semáforo	Semáforo	0.91	1
6	Casa	Casa	0.93	1
7	Poste	Poste	0.85	1
8	Persona	Persona	0.81	1
9	Carro	Carro	0.91	1
10	Animal	Animal	0.81	1
11	Árbol	Árbol	0.88	1
12	Semáforo	Semáforo	0.94	1
13	Casa	Casa	0.91	1
14	Poste	Poste	0.75	1
15	Persona	Persona	0.89	1
16	Carro	Carro	0.92	1
17	Animal	Animal	0.91	1
18	Árbol	Árbol	0.93	1
19	Semáforo	Semáforo	0.9	1
20	Casa	Casa	0.87	1
21	Poste	Poste	0.92	1
22	Persona	Persona	0.95	1
23	Carro	Carro	0.91	1
24	Animal	Animal	0.93	1
25	Árbol	Árbol	0.92	1
26	Semáforo	Semáforo	0.89	1
27	Casa	Casa	0.93	1
28	Poste	No detectó	0.95	0
29	Persona	Persona	0.98	1
30	Carro	Carro	0.97	1
31	Animal	Animal	0.92	1
32	Árbol	Árbol	0.91	1
33	Semáforo	Semáforo	0.95	1
34	Casa	Casa	0.92	1
35	Poste	Poste	0.87	1

Fuente: Elaboración propia

Como resultado del entrenamiento se pudo observar que la efectividad de la red neuronal es de 95.76%, cuando se realizaron pruebas con imágenes cargadas de la web, lo cual indica que el algoritmo es bastante efectivo al momento de predecir el objeto que se le presenta. Estas pruebas se realizaron con 21 imágenes obtenidas de sitios web, de las cuales se probaron 3 por clase.

De las 40 pruebas realizadas, a continuación se presentan los resultados obtenidos en las pruebas realizadas con un solo objeto Tabla 1 y los resultados obtenidos en las pruebas realizadas con cinco objetos (Tabla 2).

Como se puede observar de los resultados obtenidos, el algoritmo es capaz de determinar correctamente el objeto de la imagen en 33 de las 35 pruebas realizadas, indicando así una efectividad de 94.29% cuando se coloca un único objeto de las categorías incluidas en la red neuronal. En el caso de las pruebas que se realizaron con la detección de dos objetos de las clases incluidas en la red neuronal, se detectó un fallo en las 5 pruebas realizadas.

Cabe destacar que para determinar si el algoritmo detecta o no el objeto, se tomó la respuesta obtenida a los 2 segundos de mostrarse el objeto en la cámara web, por lo que es probable que el objeto no sea reconocido en ese tiempo debido al movimiento e inestabilidad de la cámara web.

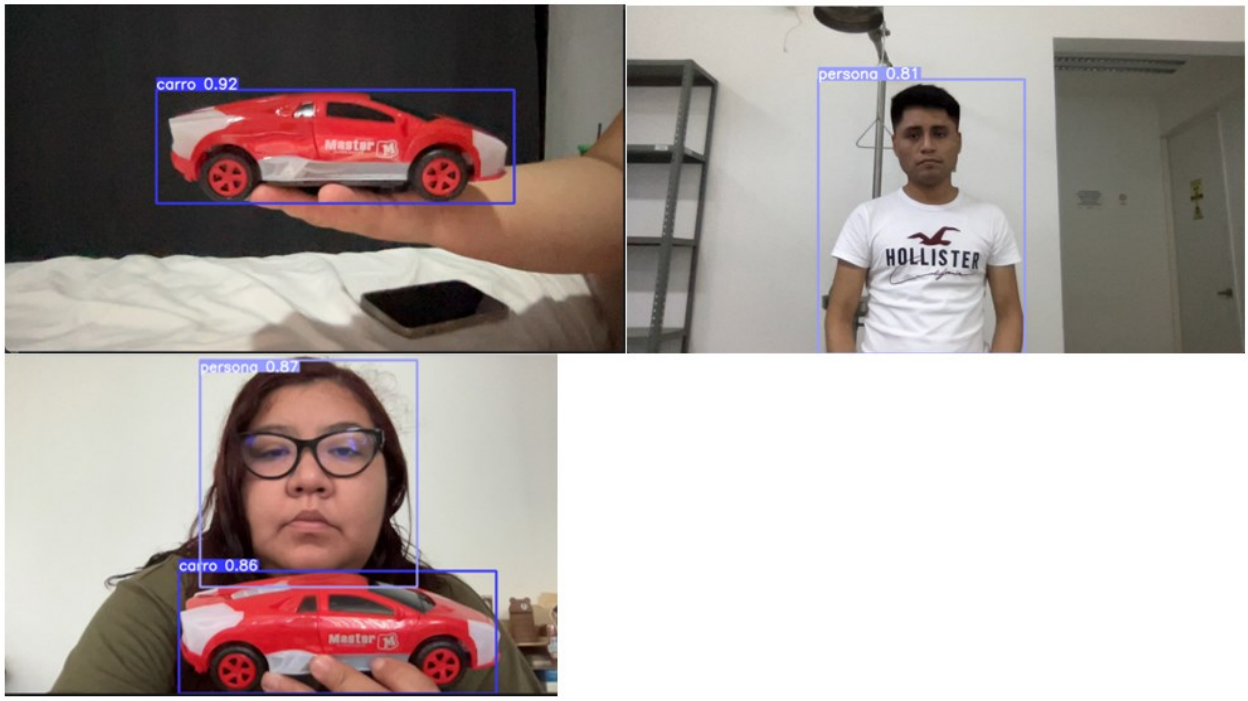
Tabla 2. . Pruebas realizadas a la red neuronal con cinco objetos, mostrando el nivel de confianza obtenido para cada uno de los objetos y si se trata de un acierto (1) o desacierto (0).

	Respuesta esperada		**Respuesta obtenida**		**Nivel de confianza**		**Acierto**	
No. Prueba	**Objeto 1**	**Objeto 2**	**Objeto 1**	**Objeto 2**	**Objeto 1**	**Objeto 2**	**Objeto 1**	**Objeto 2**
1	Persona	Carro	Persona	Carro	0.87	0.86	1	1
2	Carro	Semáforo	Carro	Semáforo	0.88	0.86	1	1
3	Semáforo	Poste	Semáforo	Poste	0.89	0.91	1	1
4	Persona	Animal	Persona	No se detectó	0.92	--	1	0
5	Casa	Árbol	Casa	Arbol	0.89	0.91	1	1

Fuente: Elaboración propia.

En la Figura 7 se pueden observar algunas pruebas realizadas a través de la cámara web con imágenes obtenidas de un entorno real en donde se puede observar el característico recuadro que se genera con el algoritmo YOLO, así como la clase a la que corresponde el objeto detectado y el nivel de confianza de la detección. Al tratarse de imágenes obtenidas en tiempo real se puede observar que el movimiento y la inestabilidad de la cámara juega un papel importante al momento de realizar la detección ya que en ocasiones no logra detectar satisfactoriamente los objetos captados.

Figura 7. Pruebas de detección de objetos de acuerdo a las clases con imágenes obtenidas a través de la cámara web.



Fuente: Elaboración propia.

Sin embargo, se observa también que YOLO es un algoritmo bastante rápido al momento de detectar objetos, por lo que resulta de utilidad su implementación. Para poder integrar un sistema portable, es conveniente que la red neuronal pueda montarse en una pequeña y potente computadora que permita ejecutar aplicaciones de inteligencia artificial en paralelo. Uno de estos sistemas corresponde a la NVIDIA Jetson Nano la cual, en conjunto con una cámara web y auriculares, puede detectar objetos en tiempo real y desplegarlo mediante audio para construir un sistema portable y de fácil uso.

CONCLUSIONES

Los avances tecnológicos brindan herramientas que facilitan el manejo o resolución de diversas problemáticas. El lograr que las personas con algún tipo de discapacidad logren llevar a cabo sus actividades cotidianas de la manera más sencilla posible y que sea el entorno el que se adapte a sus necesidades, es el objetivo que se persigue para lograr una vida libre de discriminación que les permita realmente ser incluidos como parte de la sociedad. Es por ello que se contempla que este sistema sirva de ayuda funcional para las personas que viven con discapacidad visual, tanto en el tema de movilidad como en el reconocimiento de objetos en otros entornos.

La implementación de redes neuronales constituye una herramienta de la inteligencia artificial que ha

dado pie a lo que hoy día conocemos como visión artificial y deep learning. Dentro de los algoritmos más utilizados para el uso de redes neuronales en aplicaciones de reconocimiento de objetos, YOLO constituye uno de los más estudiados debido a que presenta una alta efectividad y no requiere de tanto tiempo para llevar a cabo el procesamiento, esto debido a su principio de funcionamiento, por lo cual resulta perfecta para aplicaciones portables. Además, permite que el entrenamiento de la red neuronal sea más rápido que otro tipo de algoritmos debido a la estructura de la neurona y a las capas de entrada que posee.

Dado que el sistema presenta resultados favorables en la primera etapa presentada en este trabajo, es recomendable realizar la migración hacia un sistema embebido para que el usuario pueda transportarlo cómodamente. Para ello, se utilizará el módulo NVIDIA Jetson Nano, kit de desarrollo utilizado en aplicaciones de inteligencia artificial, en conjunto con una cámara web y auriculares. Se opta por este sistema debido a que se trata de una computadora portátil que cuenta con GPU con arquitectura NVIDIA Maxwell™ de 128 núcleos, procesador ARM® Cortex®-A57 MPCore de cuatro núcleos, memoria LPDDR4 de 4 GB y 64 bits, red Gigabit Ethernet, M.2 Key E y tiene un tamaño de 100 × 79 × 30.21 mm, características que le hacen una excelente opción para esta aplicación. Es por todo ello que resulta conveniente continuar trabajando en el desarrollo tecnológico de herramientas que faciliten la interacción del usuario con su entorno.

REFERENCIAS BIBLIOGRAFICAS

Aramendiz, C., Escorcia, D., Romero, J., Torres, K., Triana, C., & Moreno, S. (2020). Sistema basado en reconocimiento de objetos para el apoyo a personas con discapacidad visual. En Investigación y Desarrollo en TIC (Vol. 11, Números 2, pp. 75-82).

Diwan, T., Anirudh, G., & Tembhurne, J. V. (2022). Object detection using YOLO: challenges, architectural successors, datasets and applications. En Multimedia Tools and Applications (Vol. 82, Números 6, pp. 9243-9275). https://doi.org/10.1007/s11042-022-13644-y

Hernández, H., Hernández, J. R., Ramos, M., & Fundadora, Y. (2019). Calidad de vida y visual en pacientes operados de catarata por facoemulsificación bilateral simultánea con implante de lente intraocular. En Revista Cubana de Oftalmología (Vol. 32, Números 2, p. e311).

Instituto Nacional de Estadística y Geografía. (2020). Censo de Población y Vivienda 2020. INEGI.

Recuperado mayo de 2023 https://www.inegi.org.mx/app/areasgeograficas/#collapse-Indicadores

Jocher, G. (2020). Ultralytics YOLO V5. GitHub. Recuperado septiembre de 2022, de https://github.com/ultralytics/yolov5?ref=blog.roboflow.com

Martínez, J. (2015). Reconocimiento de imágenes mediante redes neuronales convolucionales (tesis de pregrado). Universidad Politécnica de Madrid, Madrid, España.

Laverde, O. D. (2013). Personas con discapacidad visual y su accesibilidad al entorno urbano. En Revista TECKNE (Vol. 11, Número 1, pp. 48-53).

Organización Mundial de la Salud. (2011). World report on disability and rehabilitation. WHO. Recuperado mayo de 2023, de https://www.who.int/teams/noncommunicable-diseases/sensory-functions-disability-and-rehabilitation/world-report-on-disability

Organización Mundial de la Salud. (2020). Informe mundial sobre la visión [World report on vision]. WHO. Recuperado mayo de 2023, de https://apps.who.int/iris/bitstream/handle/10665/331423/9789240000346-spa.pdf

Organización Mundial de la Salud. (2022). Global report on health equity for persons with disabilities. WHO. Recuperado mayo de 2023, de https://www.who.int/publications/i/item/9789240063600

Organización Panamericana de la Salud. (2020). Plan de Acción para la Prevención de la Ceguera y de las Deficiencias Visuales. PAHO. Recuperado mayo de 2023, de https://www.paho.org/es/documentos/cd58inf2-plan-accion-para-prevencion-ceguera-deficiencias-visuales-evitables-informe

Phadnis, R., Mishra, J., & Bendale, S. (2018). Objects Talk - Object Detection and Pattern Tracking Using TensorFlow. En 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT) (pp. 1216-1219). https://doi.org/10.1109/icicct.2018.8473331

Polo, M. T., & Aparicio, M. (2018). Primeros pasos hacia la inclusión: Actitudes hacia la discapacidad de docentes en educación infantil. En Revista de Investigación Educativa(Vol. 36, Números 2, pp. 365-379). https://doi.org/10.6018/rie.36.2.279281

Prince, S. (2012). Computer vision: Models, learning, and inference (1.ª ed.). Cambridge University

Press.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. En 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)(pp. 779-788). https://doi.org/10.1109/cvpr.2016.91

Sareeka, A. G., Kirthika, K., Gowthame, M. R., & Sucharitha, V. (2018). PseudoEye — Mobility assistance for visually impaired using image recognition. En 2018 2nd International Conference on Inventive Systems and Control (ICISC) (pp. 174-178). https://doi.org/10.1109/icisc.2018.8399059

Shah, D. (2022). Mean Average Precision (mAP) Explained: Everything You Need to Know. V7 Labs. Recuperado junio de 2023, de https://www.v7labs.com/blog/mean-average-precision

Thuan, D. (2021). Evolution of YOLO algorithm and YOLOV5: the state-of-the-art object detection algorithm (tesis de pregrado). Oulu University of Applied Sciences, Oulu, Finlandia.

Trujillo-Romero, F. (2022). Reconocimiento de objetos usando conjuntos pequeños de entrenamiento. En Reaxión Revista de Divulgación Científica (Vol. 9, Números 3). Recuperado mayo de 2023, de http://reaxion.utleon.edu.mx/Art_Reconocimiento_de_objetos_usando_conjuntos_pequenos_de_entrenamiento.html#

Xu, R., Lin, H., Lu, K., Cao, L., & Liu, Y. (2021). A Forest Fire Detection System Based on Ensemble Learning. En Forests (Vol. 12, Números 2, p. 217). https://doi.org/10.3390/f12020217