ISBN: 978-628-96613-0-9. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LEIRD2024.1.1.870

DEVELOPMENT OF A COMPUTER VISION SYSTEM USING YOLOV8 FOR DETECTING AND COUNTING THE NUMBER OF PEOPLE ENTERING AND EXITING.

Ryan Abraham León León[1]; Hans Anderson Olivares Garcia[2] ; Zamyr Edú Tiña Pérez[3], Universidad Privada Del Norte, Perú, ryan.leon@upn.edu.pe[1] , Universidad Privada Del Norte, Perú, N00242937@upn.edu.pe[2] , Universidad Privada Del Norte, Perú, N00244402@upn.edu.pe[3]

Abstract-*This study employed YOLOv8, an advanced neural network, to develop a real-time artificial vision system for detecting and counting people. A total of 7250 images were collected using Roboflow to train the model, enhancing its accuracy through data augmentation techniques. The training process leveraged a Tesla T4 GPU on Google Colab for accelerated processing. The system achieved an average accuracy of 94.6%, with peaks reaching 100% at specific times, albeit encountering some false positives. These findings underscore YOLOv8's effectiveness in enhancing security and crowd management, suggesting future enhancements in model confidence and image quality could further improve performance.*

Keywords-*YOLOv8, Artificial vision system, GPU Tesla T4, Google Colab*

Digital Object Identifier (DOI):
http://dx.doi.org/
ISSN: 2414-6390

DESARROLLO DE UN SISTEMA DE VISIÓN ARTIFICIAL CON YOLOV8 PARA DETECTAR Y CONTABILIZAR INGRESOS Y SALIDAS DE PERSONAS.

Ryan Abraham León León[1]; Hans Anderson Olivares Garcia[2] ; Zamyr Edú Tiña Pérez[3], Universidad Privada Del Norte, Perú, ryan.leon@upn.edu.pe[1] , Universidad Privada Del Norte, Perú, N00242937@upn.edu.pe[2] , Universidad Privada Del Norte, Perú, N00244402@upn.edu.pe[3]

Resumen-*Este estudio empleó YOLOv8, una red neuronal avanzada, para desarrollar un sistema de visión artificial que detecta y cuenta personas en tiempo real. Se recolectaron 7250 imágenes mediante Roboflow para entrenar el modelo, optimizando su precisión con técnicas de aumento de datos y utilizando una GPU Tesla T4 en Google Colab para acelerar el proceso. El sistema logró una precisión promedio del 94.6%, con picos del 100% en momentos específicos, aunque se observaron algunos falsos positivos. Estos resultados destacan la efectividad de YOLOv8 para mejorar la seguridad y la gestión de multitudes, con posibles mejoras futuras en la confianza del modelo y la calidad de las imágenes capturadas.*

Palabras clave-*Yolov8, Sistema de visión artificial, GPU Tesla T4, Google Colab*

INTRODUCCIÓN

La detección y el recuento de personas se considera una de las aplicaciones más importantes de la videovigilancia, incluido el análisis de multitudes, el análisis de comportamiento, la prevención de delitos, el seguimiento y gestión de personas en entornos públicos. Sin embargo, existen muchos desafíos que surgen al detectar personas, incluida la oclusión, la distorsión de la perspectiva, las variaciones en la postura, el tamaño y la orientación del cuerpo humano; estos desafíos afectan los resultados de los modelos de detección y conteo desarrollados. Ante estas problemáticas, el desarrollo de un sistema con visión artificial utilizando redes neuronales, como YOLOv8, para detectar y contabilizar la cantidad de ingresos y salidas de personas, se presenta como una solución crucial para mejorar la eficiencia operativa, optimizar la gestión de recursos y garantizar la seguridad en cualquier establecimiento. Los datos sobre el flujo de personas son fundamentales para optimizar la gestión de multitudes, distribuir equitativamente la afluencia y prevenir congestiones. Además, son esenciales para analizar patrones de comportamiento humano, como las horas pico, las rutas de movilidad y las tendencias de consumo, lo que permite tomar decisiones informadas en diversos escenarios.

En los últimos años, muchas arquitecturas de redes neuronales han mostrado ser un gran potencial. Por ello, se han propuesto varios métodos basados en visión por computadora y aprendizaje profundo cuyo objetivo es proporcionar resultados eficientes y precisos de detección/recuento de personas [1-3]. El desarrollo de algoritmos de detección de objetos basados en redes neuronales convolucionales ha revolucionado la capacidad de los sistemas de visión artificial para identificar y localizar personas en entornos complejos y en tiempo real. Estos algoritmos, han demostrado ser altamente efectivos para detectar personas en diversas condiciones de iluminación, escala y pose, lo que los convierte en una herramienta fundamental [4-6] Además, diversas investigaciones han integrado técnicas de seguimiento de objetos, lo cual no solo ha permitido detectar personas en una sola imagen, sino también rastrear su movimiento a lo largo del tiempo [7]. Esto facilita la contabilización precisa de la cantidad de ingresos y salidas de personas en un área específica, proporcionando información valiosa para la gestión de multitudes y la optimización del flujo de personas en lugares como centros comerciales, aeropuertos, etc. [8,9] Entre estos enfoques, Tsou et al. [10] se centró en abordar el desafío de contar el número de personas utilizando técnicas de detección y clasificación basadas en un sensor de temperatura corporal infrarrojo pasivo (PIR). En este trabajo, se aplicó también una CNN (Red Neuronal Convolucional) logrando una precisión de aproximadamente 92% en la clasificación de situaciones de paso. En otro estudio, se investigó el uso de la red neuronal profunda SSD para detectar y contar personas desde diferentes puntos de vista, incluida la vista superior. El estudio demostró la eficacia del enfoque de aprendizaje profundo [11]. Mohaghegh, desarrolló un sistema de conteo de personas con Mask RCNN, que clasifica entre empleados y clientes y cuenta grupos en intervalos de tiempo específicos. Se reconoce la necesidad de mejorar su rendimiento en situaciones prolongadas o con alta densidad de personas, tema a abordar en futuros estudios [12] . Además, en un antecedente científico previo, se presentó un enfoque técnico centrado en el análisis de vídeo computacional utilizando YOLOv3 para detectar y seguir personas en vídeos. Los resultados mostraron una alta precisión y rendimiento del algoritmo, lo que sugiere su eficacia para estas aplicaciones [13]. Así mismo, en otra investigación, se implementaron sistemas de monitoreo basados en IA que integran algoritmos de detección como YOLOv8. Este modelo propuesto extiende sus capacidades más allá del conteo de personas al abarcar la detección de actividades anormales como armas, incendios, caídas y humo dentro de la multitud [14]. Con lo anteriormente mencionado, se sabe que el avance en el desarrollo de redes neuronales ofrece soluciones prometedoras para la detección y conteo de personas, con impactos significativos en la seguridad y la eficiencia de los sistemas de vigilancia y gestión de multitudes [15].

Detectar y contabilizar la cantidad de ingresos y salidas de personas en diferentes espacios es fundamental para mejorar la gestión y la seguridad en estos mismos. La capacidad de monitorear y analizar el flujo de personas en tiempo real facilita la toma de decisiones estratégicas para optimizar la

distribución de recursos, mejorar la experiencia del cliente y garantizar la seguridad en las diferentes instalaciones. En este contexto, nuestro objetivo principal es desarrollar un sistema de visión artificial utilizando redes neuronales YOLOv8 para detectar y contabilizar de manera precisa la cantidad de ingresos y salidas de personas en un entorno determinado, mejorando así la gestión y el control de flujo de personas. Además, se pretende optimizar la precisión y eficiencia del sistema de visión artificial mediante pruebas exhaustivas y ajustes iterativos, considerando diferentes condiciones de iluminación, oclusión y variaciones en la densidad de personas dentro de cualquier establecimiento. Esto incluye trabajar con la funcion de activación, Leaky ReLU, para mejorar la capacidad del modelo de capturar características no lineales complejas. Así mismo ajustar otros hiperparámetros del modelo, como la tasa de aprendizaje, el tamaño del lote, y la regularización, para mejorar su desempeño y robustez en distintos escenarios.

MATERIALES Y MÉTODOS

El proyecto se inició con la recolección y etiquetado de imágenes para detectar y contabilizar ingresos y salidas de personas, utilizando la plataforma Roboflow [16]. Se capturaron 7250 imágenes de personas en diversas condiciones con una cámara de alta resolución, asegurando la diversidad del conjunto de datos mediante diferentes condiciones de iluminación y ángulos. La cantidad de imágenes empleadas se basó en un estudio anterior que utilizó 7000 imágenes [17]. Dado que las redes neuronales han mejorado significativamente en los últimos años, especialmente en su capacidad de aprendizaje rápido, se decidió entrenar el modelo con imágenes de alta calidad y nitidez para obtener resultados más eficientes y precisos [18]. Durante la selección de imágenes para la detección y contabilización de ingresos y salidas de personas, se eliminaron aquellas que no cumplían con los estándares de calidad, como imágenes borrosas, mal iluminadas, o con obstrucciones que dificultaran la detección precisa, además de aquellas que estaban duplicadas para evitar la redundancia en los datos. En el proceso de etiquetado, se identificaron gráficamente las personas en las imágenes utilizando la herramienta Roboflow (Fig. 1), lo que permitió un etiquetado uniforme y preciso, crucial para entrenar un modelo robusto. Los datos se dividieron en tres conjuntos de manera estratégica: un 70% se utilizó para el entrenamiento del modelo, permitiendo que este aprenda a reconocer patrones y movimientos, un 20% para la validación, ayudando a ajustar los hiperparámetros y prevenir el sobreajuste, y un 10% para el testeo, garantizando que el modelo pueda generalizar bien a datos no vistos. Esta cuidadosa preparación de los datos asegura que el sistema de detección y contabilización funcione de manera confiable en diversas condiciones y escenarios reales.

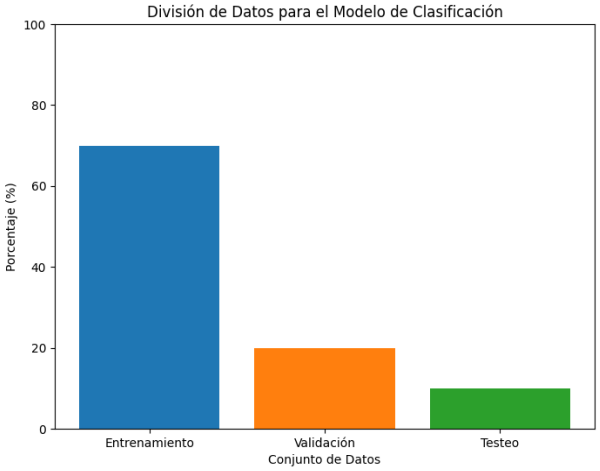

Fig 1. Cantidad de imágenes divididas para el entrenamiento

Para la detección y contabilización de ingresos y salidas de personas, el modelo se entrenó utilizando criterios visuales y morfológicos como la postura, el tamaño y la ropa de las personas. Se aplicaron técnicas de aumentación de datos, incluyendo rotación, cambio de brillo y contraste, para incrementar la variabilidad y robustez del conjunto de datos. Estas técnicas permitieron al modelo aprender a distinguir entre diferentes individuos y condiciones de iluminación. Durante el proceso de etiquetado, se identificaron gráficamente las personas en las imágenes, clasificándolas según las etiquetas "entrando" y "saliendo". El software detecta y marca las personas en las imágenes, asignando etiquetas específicas basadas en su dirección de movimiento y otras características visuales, lo que permite al modelo distinguir entre los diferentes estados de las personas de manera precisa, tal como lo señalan trabajos recientes [19]. Esta clasificación visual, fundamentada en criterios como la postura y el movimiento, sienta las bases para el entrenamiento efectivo de la red neuronal, asegurando que el sistema pueda contar y detectar con precisión las entradas y salidas de personas en diversas condiciones.

Fig. 2 Etiquetado de personas en centro commercial

YOLO (You Only Look Once) es una serie de modelos de redes neuronales convolucionales (CNN) diseñados para la detección de objetos en imágenes y videos en tiempo real [20]. En este proyecto, se ha implementado la versión YOLOv8, la iteración más reciente de esta tecnología, destacada por su alta precisión y eficiencia. YOLOv8 se estructura en una arquitectura de red neuronal convolucional (Fig. 3) que permite la detección de objetos con una sola pasada a través de la red, en lugar de requerir múltiples etapas de procesamiento como otros métodos. La selección de esta versión se basa en investigaciones recientes que evidencian la superioridad de YOLOv8 en términos de precisión y velocidad en comparación con versiones anteriores y otros modelos de detección de objetos [21]. Las características de esta arquitectura incluyen:

Convoluciones: Estas capas extraen características importantes de las imágenes mediante el uso de filtros que detectan bordes, texturas y patrones.

Capas de Anclaje: YOLOv8 utiliza cajas de anclaje predefinidas para predecir la ubicación y el tamaño de los objetos en la imagen. Las coordenadas predichas para una caja de anclaje

(bx, by, bw, bh) se ajustan de la siguiente manera.

$$bx = \sigma(tx) + cx \quad (1)$$

$$by = \sigma(ty) + cy \quad (2)$$

$$bw = pw\, e^{(tw)} \quad (3)$$

$$bh = ph\, e^{(th)} \quad (4)$$

donde (cx, cy) es la ubicación de la celda de la cuadrícula, (pw, ph) son las dimensiones de la caja de anclaje predeterminada, y (tx, ty, tw, th) son las predicciones de la red.

Predicción Simultánea: En contraste con otros métodos que inicialmente crean propuestas de regiones y posteriormente las clasifican, YOLOv8 predice simultáneamente las clases y posiciones de los objetos, lo que incrementa tanto la rapidez como la eficacia.

Función de Pérdida: La función de pérdida de YOLOv8 combina la precisión de la clasificación y la precisión de la localización en un único valor escalar, optimizando así ambos aspectos durante el entrenamiento. La pérdida total L se define como:

$$L = Lloc + Lconf + Lcls \quad (5)$$

$Lloc$ es la pérdida de localización, que mide el error en las predicciones de las coordenadas de las cajas. $Lconf$ es la pérdida de confianza, que mide el error en las predicciones de la confianza de que una caja contiene un objeto y $Lcls$ es la pérdida de clasificación, que mide el error en las predicciones de las clases de los objetos.

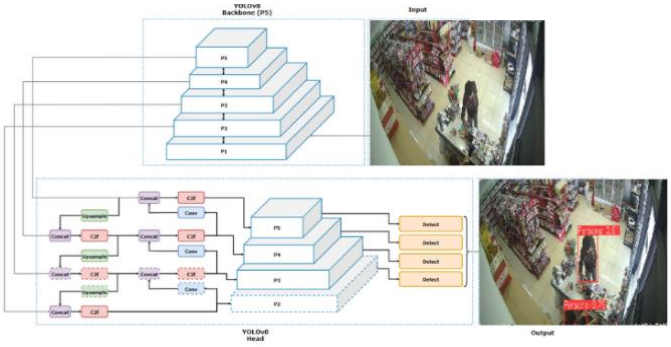

Fig. 3 Arquitectura YOLOv8. Se *presenta una representación esquemática de cómo el sistema YOLOv8 procesa una imagen de entrada para detectar personas, detallando el flujo de datos desde la extracción de características en el Backbone hasta la detección final en el Head, con ejemplos visuales de los resultados obtenidos.*

La convolución es una operación matemática que combina dos conjuntos de datos. En el contexto de las CNNs, se refiere a la aplicación de un filtro (o núcleo) sobre una imagen para extraer características esenciales. Este filtro se desplaza por la imagen de entrada y, en cada posición, calcula un valor de salida resultante de una suma ponderada entre los valores de la imagen y los del filtro (Fig. 4). Durante el entrenamiento de YOLOv8, las capas convolucionales se emplean para extraer características visuales cruciales de las imágenes de personas, como bordes, texturas y formas que facilitan la detección y contabilización de entradas y salidas.

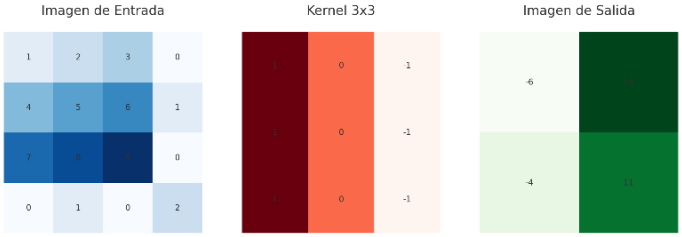

Fig. 4 Un kernel (o filtro) de convolución

las funciones de activación juegan un papel fundamental en la capacidad de la red para aprender y realizar predicciones precisas. La función Sigmoid transforma las salidas de las neuronas a un rango entre 0 y 1, lo cual es esencial para la tarea de clasificación binaria involucrada en la detección de objetos. La función Tanh (Tangente hiperbólica) mapea las salidas a un rango entre -1 y 1, lo que puede mejorar la convergencia del modelo al centrar los datos en torno a cero. La función ReLU (Rectified Linear Unit), ampliamente utilizada en YOLOv8, ayuda a mitigar el problema del desvanecimiento de gradiente al establecer los valores negativos a 0 y dejar los valores positivos sin cambios, permitiendo una propagación de gradientes más eficiente y una mejor activación de las neuronas. La función Leaky ReLU, una variación de la ReLU, permite un pequeño gradiente cuando la entrada es negativa, evitando que las neuronas queden completamente inactivas, lo cual es crucial

para mantener la capacidad de la red de detectar personas en diferentes condiciones y posiciones (Fig. 5).

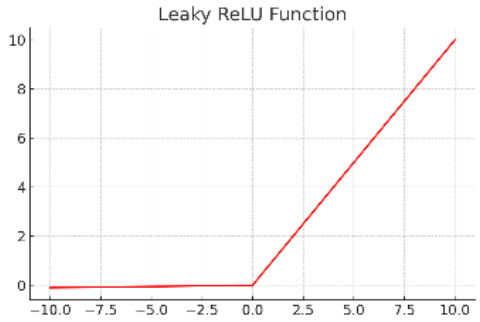

Fig. 5 Función activación

Los filtros convolucionales son matrices pequeñas aplicadas sobre las imágenes. Cada filtro se entrena para identificar características específicas. Investigaciones anteriores han demostrado su eficacia en la detección y clasificación de diversas características, destacando su utilidad en el análisis de imágenes. En el contexto de la detección de personas, algunos filtros pueden aprender a reconocer la silueta de una persona, mientras que otros pueden identificar texturas y detalles relevantes para diferenciar entre individuos y contar con precisión las entradas y salidas.

Para el entrenamiento del modelo YOLOv8 (Fig. 6), se aprovechó la potencia computacional de Google Colab, que ofrece recursos especializados para el entrenamiento de redes neuronales, incluyendo GPUs. En particular, se utilizó una GPU Tesla T4 con 15 GB de memoria, compatible con CUDA, lo que permitió acelerar significativamente el proceso de entrenamiento. La capacidad de procesamiento de esta GPU aseguró una rápida convergencia del modelo y tiempos de entrenamiento eficientes.

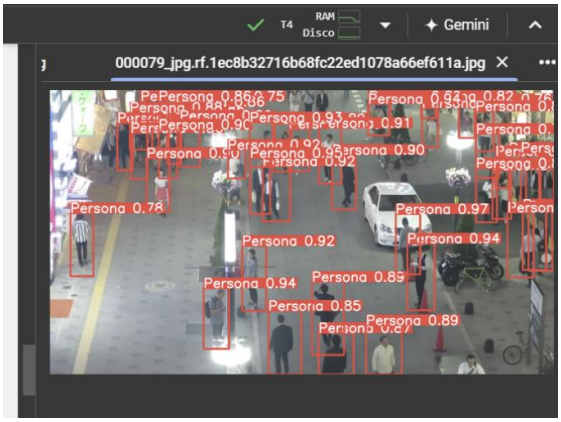

Fig. 6 Entrenamiento de la red neuronal

Una vez configurado el entorno en Google Colab, se importaron las bibliotecas esenciales, como torch y yolov8, para el desarrollo y entrenamiento del modelo [22]. PyTorch, conocido como torch, es una biblioteca de código abierto ampliamente utilizada para el aprendizaje profundo, destacada por su facilidad de uso y flexibilidad, lo que la hace ideal para la investigación y el desarrollo de modelos de redes neuronales. Su eficacia y versatilidad están respaldadas por numerosas evaluaciones [23]. Además, se integró Google Drive para facilitar el almacenamiento y el acceso a los conjuntos de datos y modelos (Fig 7).

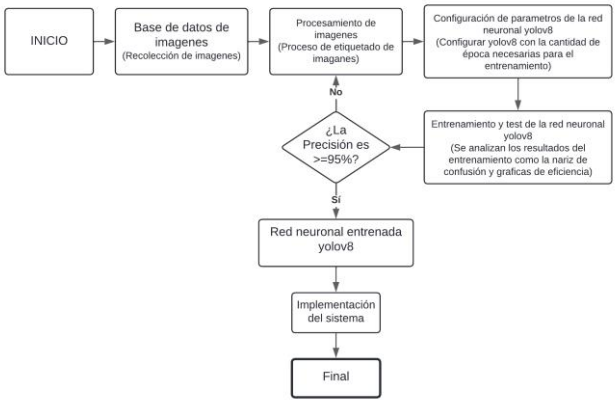

Fig. 7 Actividades realizadas

El proceso de entrenamiento se llevó a cabo en Google Colab, donde se descargaron los datos etiquetados desde Roboflow y se cargaron en el entorno de trabajo. Posteriormente, se configuró el modelo YOLOv8 ajustando parámetros como el tamaño del lote, la tasa de aprendizaje y el número de épocas, fijando este último en un total de 150. Esta configuración se fundamentó en indagaciones anteriores que demostraron la eficacia de este número de épocas para optimizar el rendimiento del modelo. Durante el entrenamiento, se monitorearon métricas como la pérdida de entrenamiento y la precisión en el conjunto de validación, permitiendo ajustes en los hiperparámetros según fuera necesario para mejorar el rendimiento del modelo (Fig.8).

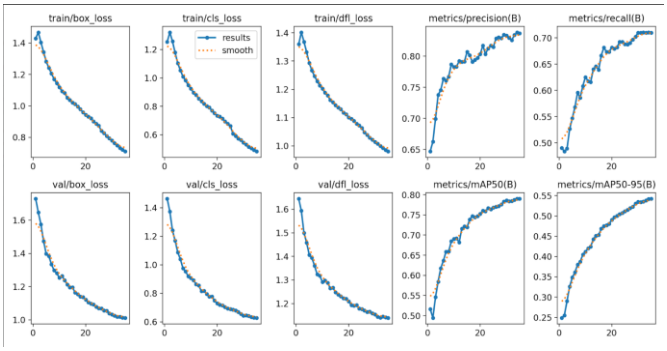

Fig. 8 Resultados del entrenamiento

Al culminar el proceso de entrenamiento, se evaluó el modelo utilizando un conjunto de datos de prueba separado, asegurando que no hubiera sobreajuste y validando su capacidad para generalizar a datos no vistos. Este enfoque

garantizó la robustez y precisión del modelo en la detección de la madurez de los arándanos.

Para implementar el algoritmo de detección y contabilización de entradas y salidas de personas, se utilizó Python en el entorno de desarrollo PyCharm. Se configuró PyCharm con un entorno virtual y se instalaron las dependencias necesarias, incluyendo opencv-python y pytorch [24,25]. Se cargó el modelo entrenado de YOLOv8 y se desarrolló un script en Python para procesar nuevas imágenes de personas, capturadas desde una cámara o cargadas desde el sistema de archivos (Fig. 9). Utilizando el modelo YOLOv8, se detectaron y contabilizaron las personas en las imágenes, mostrando los resultados mediante cuadros delimitadores y etiquetas sobre las imágenes originales.

Fig.9 Prueba de algoritmo en centro commercial

RESULTADOS

Los resultados que se muestran en la Figura 10, indican que 8064 instancias de "Persona" fueron clasificadas correctamente. Sin embargo, se identificaron algunos errores: 2592 instancias de "Persona" fueron incorrectamente etiquetadas como "Fondo", y 1898 instancias de "Fondo" fueron clasificadas erróneamente como "Persona". Estos resultados sugieren que, aunque el modelo tiene una alta precisión en la clase "Persona", todavía hay margen de mejora. Para reducir estos errores, se podría aumentar el nivel de confianza del modelo y mejorar la calidad de las cámaras utilizadas para proporcionar imágenes más nítidas y claras, facilitando al modelo la identificación precisa de objetos en escenarios complejos. Estas medidas podrían conducir a una mayor fiabilidad y exactitud en la detección de personas, esenciales para aplicaciones que requieren una monitorización precisa y confiable. En la Fig. 10 La matriz de confusión muestra el desempeño del algoritmo en la detección y conteo de personas, dividiendo las instancias en dos clases: "Persona" y "Fondo".

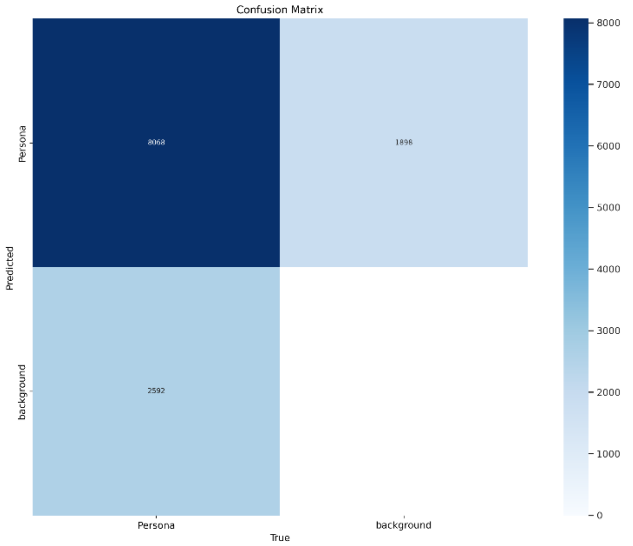

Fig. 10 La matriz de confusion

La implementación de YOLOv8 para la detección y conteo de personas ha demostrado ser altamente efectiva. La precisión promedio del 94.6% mencionada en el artículo fue medida utilizando un conjunto de datos separado destinado específicamente para la validación, el cual representó el 20% del total de imágenes capturadas. Estos datos no fueron usados durante el entrenamiento, lo que permitió asegurar que el modelo no estuviera sobreajustado. Este enfoque es común en los modelos de aprendizaje profundo para evaluar la capacidad de generalización en datos no vistos.

La precisión se calculó usando la métrica de "precisión" (accuracy), la cual toma en cuenta el número total de predicciones correctas (verdaderos positivos y verdaderos negativos) sobre el total de predicciones realizadas (verdaderos y falsos positivos y negativos). Las condiciones bajo las cuales se alcanzó esta precisión incluyen variaciones de iluminación, ángulos de cámara y oclusiones parciales, lo que refleja un entorno de operación en tiempo real. Cabe destacar que la precisión máxima del 100% se logró en momentos específicos del día, como a las 8:00 y a las 13:00 horas, cuando las condiciones de iluminación eran óptimas.

El conjunto de validación también fue clave para ajustar los hiperparámetros del modelo, como la tasa de aprendizaje, el tamaño del lote, y la función de activación (Leaky ReLU), lo que ayudó a prevenir el sobreajuste y optimizar el rendimiento del modelo. Estos ajustes contribuyeron significativamente a alcanzar la precisión promedio del 94.6%. Además, en ciertas horas del día se identificaron algunos falsos positivos, especialmente a las 12:00 y 14:00 horas, que se atribuyen a factores como las variaciones en la iluminación y la presencia de objetos que pudieron confundirse con personas. Para mitigar estos errores, se sugiere en futuras investigaciones aumentar la calidad de las

cámaras y ajustar el nivel de confianza del modelo durante la inferencia. (Tabla 2). Estos resultados sugieren que YOLOv8 es capaz de realizar un seguimiento preciso del flujo de personas en entornos concurridos con una mínima tasa de error. Los falsos positivos podrían deberse a factores ambientales o a la presencia de objetos que el modelo pudo haber confundido con personas. No se reportaron falsos negativos, lo cual es un indicador positivo de la sensibilidad del modelo.

Tabla 1. Desempeño de la red neuronal YOLOv8 para detectar y contabilizar la cantidad de ingresos y salidas de personas

Fecha	*Hora*	*Total Personas Detectadas*	*Personas Reales*	*Precisión (%)*	*Falsos Positivos*	*Falsos Negativos*
17/06/2024	08:00	15	15	100	0	0
17/06/2024	09:00	22	20	90.9	2	0
17/06/2024	10:00	23	20	87	3	0
17/06/2024	11:00	30	28	93.3	2	0
17/06/2024	12:00	45	40	88.9	5	0
17/06/2024	13:00	50	50	100	0	0
17/06/2024	14:00	60	55	91.7	5	0
17/06/2024	15:00	40	38	95	2	0

DISCUSIÓN

Los resultados obtenidos del estudio utilizando YOLOv8 para la detección y conteo de personas muestran un rendimiento notablemente sólido, con una precisión promedio del 94.6% a lo largo de un día completo de operación. Este rendimiento es respaldado por la observación de picos de precisión del 100% a las 08:00 y 13:00 horas, indicando una capacidad de detección perfecta en esos momentos específicos del día.

En varios escenarios de prueba, como entradas a tiendas, área de cajeros automáticos, vehículos en movimiento, calles y pasadizos, el sistema funcionó consistentemente a 94.6% de efectividad, superando a los modelos existentes como Yolov5, SSD, R-CNN, Tini Yolo. Los cuales son inferiores en la detección de objetos por ser versiones menores en la detección de todo ámbito, dependiendo en primeras instancias los entrenamientos que realizan para poder tener una mayor tasa de efectividad.

En contraste con otros modelos y enfoques utilizados en estudios anteriores, la implementación de YOLOv8 ha demostrado ser efectiva en términos de precisión global. Investigaciones anteriores [26], han reportado niveles de precisión que variaban entre el 85% y el 92% en situaciones similares de detección de personas [27]. La capacidad de YOLOv8 en nuestro estudio para alcanzar consistentemente un 94.6% de precisión promedio posiciona este modelo como una opción robusta para aplicaciones que requieren alta exactitud en la clasificación de objetos, particularmente personas, en entornos dinámicos y desafiantes. A pesar de estos logros, es crucial mencionar los casos de falsos positivos observados durante las pruebas, siendo más prominentes a las 12:00 y 14:00 horas, con 5 detecciones incorrectas en cada período.

El sistema está diseñado para ser adaptable, lo que permite su uso en una variedad de entornos. Ya sea que esté trabajando en entornos con bastante o poca luz, el sistema puede ajustarse automáticamente conforme se realicen más entrenamientos de forma directa con casos en vivo y no por fotos o videos en donde se lograría ayudarlo a trabajar de manera eficiente. Además, se adapta fácilmente desde entornos pequeños y controlados hasta áreas más grandes y de alto tráfico, lo que garantiza una precisión y un rendimiento constantes independientemente del tamaño de la operación. Estos errores pueden atribuirse a factores como variaciones en la iluminación, ángulos de cámara o la presencia de objetos que puedan ser confundidos con personas. Para reducir la probabilidad de falsas positivos, especialmente en condiciones de iluminación inestable o que puedan causar interferencias visuales, el sistema utiliza redes neuronales convolucionales, así como filtros que detecten bordes, texturas y patrones, incorporando un comportamiento automático de aprendizaje lo cual permite lograr la adaptabilidad y mejorar la precisión. Además, se utiliza un componente de aprendizaje automático que permite que el sistema aprenda y se adapte en función de hallazgos anteriores, aumentando gradualmente la precisión y reduciendo los falsos positivos con el tiempo. La mitigación de estos falsos positivos podría mejorar aún más la fiabilidad del sistema en situaciones de uso continuo y en tiempo real. Para futuras investigaciones y desarrollos, se recomienda explorar métodos para aumentar el nivel de confianza del modelo durante la inferencia, así como mejorar la calidad de las cámaras utilizadas para capturar imágenes más nítidas y consistentes. Estos pasos podrían no solo reducir los errores identificados, sino también elevar aún más la precisión y la confiabilidad de YOLOv8 en escenarios prácticos.

CONCLUSIÓN

Se logró desarrollar un sistema de visión artificial basado en redes neuronales utilizando YOLOv8 para detectar y contar la cantidad de personas que ingresan y salen de un espacio. Además se optimizó la precisión y eficiencia del sistema de visión artificial mediante pruebas exhaustivas y ajustes iterativos. Para ello, se trabajó con la funcion de activación, Leaky ReLU, mejorando la capacidad del modelo para capturar características no lineales complejas. Así mismo, se ajustaron otros hiperparámetros del modelo, como la tasa de aprendizaje, el tamaño del lote y la regularización, mejorando su desempeño y robustez en distintos escenarios. Los resultados del estudio para la detección y conteo de personas

revelan un rendimiento notable, con una precisión promedio del 94.6% a lo largo de un día completo de operación. La implementación de YOLOv8 permitió una combinación eficiente de precisión y velocidad, lo que resulta fundamental para aplicaciones en tiempo real. A través de la detección de objetos y la segmentación de imágenes, se logró identificar y contabilizar las entradas y salidas de personas. Este sistema ofrece ventajas significativas en comparación con enfoques tradicionales, ya que su arquitectura simplificada y su capacidad para procesar imágenes en tiempo real lo hacen ideal para aplicaciones prácticas.

El sistema está diseñado para que pueda seguir mejorando, lo que lo hace aplicable a una amplia gama de entornos, desde pequeños eventos en interiores hasta grandes espacios al aire libre o centros de transporte concurridos. Para tener en cuenta las variaciones en el tamaño de la multitud, la iluminación y otros factores ambientales, el sistema utiliza métricas con la precisión y la perdida de entrenamientos durante el proceso del modelado, lo que le permite mantener la precisión y el rendimiento independientemente de la escala. Esto lo hace particularmente útil para la implementación a gran escala en conferencias, centros comerciales, aeropuertos, estacionamientos, etc. Además, la utilización de redes neuronales profundas como YOLOv8 garantiza una mayor precisión en la detección de personas, incluso en condiciones de iluminación variable o fondos complejos.

Para abordar el problema de los falsos positivos en situaciones de alta densidad y condiciones adversas, como iluminación variable o interferencias visuales, proponemos varias mejoras técnicas. En primer lugar, se planea ajustar el umbral de confianza del modelo YOLOv8. Al incrementar este umbral, se descartan detecciones con baja confianza, reduciendo la probabilidad de falsos positivos generados por objetos irrelevantes o sombras. Este ajuste, combinado con el uso de imágenes de mayor calidad y resolución, permitirá al modelo identificar mejor los contornos y características clave de las personas, diferenciándolos más efectivamente de otros objetos en entornos con iluminación difícil o bajo contraste.

Además, se ampliará el conjunto de datos de entrenamiento, incluyendo imágenes capturadas en escenarios con alta densidad de personas y variaciones de iluminación. Estas imágenes permitirán al modelo aprender a distinguir entre personas en condiciones más complejas y reducir las confusiones con objetos circundantes. Complementariamente, se evaluará la integración de técnicas de seguimiento de objetos (tracking) junto con YOLOv8, lo que permitirá un análisis más coherente de los movimientos en secuencias de video. De este modo, las detecciones incorrectas que no sigan un patrón lógico de movimiento podrían descartarse, mitigando aún más los falsos positivos en tiempo real.

REFERENCIAS BIBLIOGRÁFICAS

[1] M. Andriluka, S. Roth and B. Schiele, "Pictorial structures revisited: People detection and articulated pose estimation,". IEEE Conference on Computer Vision and Pattern Recognition. 2019: 1014-1021. https://doi.org/10.1016/j.knosys.2017.06.001

[2] Bouazizi M, Ye C, & Ohtsuki T. Low-resolution infrared array sensor for counting and localizing people indoors: When low end technology meets cutting edge deep learning techniques. *Information*. 2022;13(3): 132. https://doi.org/10.3390/info13030132

[3] Chato P, Chipantasi D, Velasco N, Rea S, Hallo V and Constante P. "Image processing and artificial neural network for counting people inside public transport".IEEE Third Ecuador Technical Chapters Meeting (ETCM). 2018: 1-5, doi: 10.1109/ETCM.2018.8580287.

[4] Hu, Yaocong, et al. Dense crowd counting from still images with convolutional neural networks. Journal of Visual Communication and Image Representation, 2016, vol. 38, p. 530-539. https://doi.org/10.1016/j.jvcir.2016.03.021

[5] Zou, Zhikang, et al. Attend to count: Crowd counting with adaptive capacity multi-scale CNNs. Neurocomputing, 2019, vol. 367, p. 75-83. https://doi.org/10.1016/j.neucom.2019.08.009

[6] Filipic, Joaquín, et al. People counting using visible and infrared images. Neurocomputing, 2021, vol. 450, p. 25-32. https://doi.org/10.1016/j.neucom.2021.03.089

[7] Liu, Guojin, et al. Passenger flow estimation based on convolutional neural network in public transportation system. Knowledge-Based Systems, 2018, vol. 123, p. 102-115. https://doi.org/10.1016/j.knosys.2017.02.016

[8] Hsu, Ya-Wen; WANG, Ting-Yen; PERNG, Jau-Woei. Passenger flow counting in buses based on deep learning using surveillance video. Optik, 2020, vol. 202, p. 163675. https://doi.org/10.1016/j.ijleo.2019.163675

[9] Massa, L., Barbosa, A., Oliveira, K. et al. LRCN-RetailNet: A recurrent neural network architecture for accurate people counting. Multimed Tools Appl. 2021; 80: 5517–5537. https://doi.org/10.1007/s11042-020-09971-7

[10] Tsou PR, Wu CE, Chen YR, Ho YT, Chang JK, Tsai HP. Counting people by using convolutional neural network and A PIR array. En: 2020 21st IEEE International Conference on Mobile Data Management (MDM). IEEE; 2020: 342–7. doi: 10.1109/MDM48529.2020.00075

[11] M. Ahmad, I. Ahmed, K. Ullah y M. Ahmad, "Un enfoque de red neuronal profunda para la detección y el recuento de personas en vista superior", décima conferencia anual de informática ubicua, electrónica y comunicación móvil (UEMCON) del IEEE de 2019 . 2019: 1082-1088, doi: 10.1109/UEMCON47517.2019.8993109.

[12] M. Mohaghegh. "A Four-Component People Identification and Counting System Using Deep Neural Network," *2018 5th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*. 2018: 10-17, doi: 10.1109/APWConCSE.2018.00011.

[13] Kajabad, Ebrahim Najafi; Ivanov, Sergey V. People detection and finding attractive areas by the use of movement detection analysis and deep learning approach. Procedia Computer Science, 2019, vol. 156, p. 327-337. https://doi.org/10.1016/j.procs.2019.08.209

[14] Sudharson, D., et al. Proactive Headcount and Suspicious Activity Detection using YOLOv8. Procedia Computer Science, 2023, vol. 230, p. 61-69. https://doi.org/10.1016/j.procs.2023.12.061

[15] Kanatov M, & Atymtayeva L. Deep convolutional neural network based person detection and people counting system. Advanced Engineering Technology and Application. 2018; 7(3), 5-9. https://doi.org/10.1016/j.knosys.2017.02.016

[16] Pavithra, M., et al. Implementation of Enhanced Security System using Roboflow. En *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2024. p. 1-5. doi: 10.1109/ICRITO61523.2024.10522313.

[17] A. A. Protik, A. H. Rafi and S. Siddique, "Real-time Personal Protective Equipment (PPE) Detection Using YOLOv4 and TensorFlow," *2021 IEEE Region 10 Symposium (TENSYMP)*, Jeju, Korea, Republic of, 2021, pp. 1-6, doi: 10.1109/TENSYMP52854.2021.9550808.

[18] Tripathi, G., Singh, K. & Vishwakarma, D.K. Convolutional neural networks for crowd behaviour analysis: a survey. *Vis Comput.* 2019, vol. 35, pág. 753-776. https://doi.org/10.1007/s00371-018-1499-5

[19] Ab razak, Muhammad Naqib Syahmi Bin, et al. Prayer Hall Vacancy Detection. En *2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED)*. IEEE, 2023. p. 1-6. doi: 10.1109/ICCED60214.2023.10425587.

[20] M. M. Ali, M. S. Qaseem, M. Zeeshan, A. A. Quraishi and A. U. Rahman, "Real-Time Crowd-Counting and Management in Sacred places using Computer Vision & ESP32 cameras," *2023 9th International Conference on Signal Processing and Communication (ICSC)*, NOIDA, India, 2023, pp. 434-439, doi: 10.1109/ICSC60394.2023.10441236.

[21] A. Elaoua, M. Nadour, L. Cherroun and A. Elasri, "Real-Time People Counting System using YOLOv8 Object Detection," *2023 2nd International Conference on Electronics, Energy and Measurement (IC2EM)*, Medea, Algeria, 2023, pp. 1-5, doi: 10.1109/IC2EM59347.2023.10419684.

[22] Fotia, L., Percannella, G., Saggese, A., Vento, M. (2023). Highly Crowd Detection and Counting Based on Curriculum Learning. In: Tsapatsoulis, N., *et al.* Computer Analysis of Images and Patterns. CAIP 2023. Lecture Notes in Computer Science, vol 14185. Springer, Cham. https://doi.org/10.1007/978-3-031-44240-7_2

[23] Alhawsawi, Abdullah N.; Khan, Sultan Daud; UR Rehman, Faizan. Crowd Counting in Diverse Environments Using a Deep Routing Mechanism Informed by Crowd Density Levels. *Information*, 2024, vol. 15, no 5, p. 275. https://doi.org/10.3390/info15050275

[24] Nigam, Nitika; Dutta, Tanima. Crowd crush detection in large mass gatherings via federated learning across multicamera environment. En *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 2022. p. 297-298. https://doi.org/10.1145/3563357.3567753

[25] Amil, F.G., Rana, Z., Zhao, Y. (2023). Seamless Passenger Experience for the Airport Environment: Research at DARTeC. In: Karakoc, T.H., Atipan, S., Dalkiran, A., Ercan, A.H., Kongsamutr, N., Sripawadkul, V. (eds) Research Developments in Sustainable Aviation. ISSASARES 2021. Sustainable Aviation. Springer, Cham. https://doi.org/10.1007/978-3-031-37943-7_37

[26] M. Alruwaili et al., "Deep Learning-Based YOLO Models for the Detection of People With Disabilities," in IEEE Access, vol. 12, pp. 2543-2566, 2024, doi: 10.1109/ACCESS.2023.3347169.

[27] Vasantha, B. Kiranmai, M. A. Hussain, S. S. Hashmi, L. Nelson and S. Hariharan, "Face and Object Detection Algorithms for People Counting Applications," 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2023, pp. 1188-1193, doi: 10.1109/ICACRS58579.2023.10405114.