



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y
TELECOMUNICACIONES**

Desarrollo de un sistema de monitoreo y conteo de personas
utilizando visión por computadora, para optimizar la gestión de aforo
en la biblioteca de la UPSE.

AUTOR

Panchana Magallanes, Genesis Nicole

Proyecto de Unidad de Integración Curricular

Previo a la obtención del grado académico en
INGENIERO EN TECNOLOGÍAS DE LA INFORMACIÓN

TUTOR

Ing. Orozco Iguasnia, Walter. MsC

Santa Elena, Ecuador

Año 2024



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y
TELECOMUNICACIONES**

TRIBUNAL DE SUSTENTACIÓN

**Ing. José Sánchez Aquino. Mgt.
DIRECTOR DE LA CARRERA**

**Ing. Walter Orozco Iguasnia. Mgt
TUTOR**

**Ing. Carlos Castillo Yagual, Mgt.
DOCENTE ESPECIALISTA**

**Ing. Marjorie Coronel Suárez. Mgti.
DOCENTE GUÍA UIC**



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y
TELECOMUNICACIONES**

CERTIFICACIÓN

Certifico que luego de haber dirigido científica y técnicamente el desarrollo y estructura final del trabajo, este cumple y se ajusta a los estándares académicos, razón por el cual apruebo en todas sus partes el presente trabajo de titulación que fue realizado en su totalidad por Panchana Magallanes Génesis Nicole, como requerimiento para la obtención del título de Ingeniero en Tecnologías de la Información.

La Libertad, a los 12 días del mes de diciembre del año
2024

WALTER ARMANDO Firmado digitalmente por WALTER
OROZCO IGUASNIA ARMANDO OROZCO IGUASNIA
Fecha: 2024.12.12 10:51:42 -05'00'

Ing. Walter Orozco Iguasnia. MsC



UPSE

**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y
TELECOMUNICACIONES**

DECLARACIÓN DE RESPONSABILIDAD

Yo, **Panchana Magallanes Genesis Nicole**

DECLARO QUE:

El trabajo de Titulación, Desarrollo de un sistema de monitoreo y conteo de personas utilizando visión por computadora, para optimizar la gestión de aforo en la biblioteca de la Universidad Estatal Península de Santa Elena, previo a la obtención del título en Ingeniero en Tecnologías de la Información, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

La Libertad, a los 12 días del mes de diciembre del año

2024

GENESIS PANCHANA.

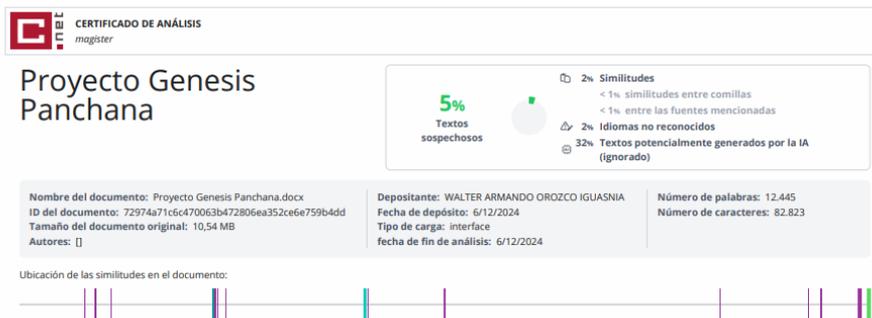
Genesis Nicole Panchana Magallanes



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA**
**FACULTAD DE SISTEMAS Y
TELECOMUNICACIONES**

CERTIFICACIÓN DE ANTIPLAGIO

Certifico que después de revisar el documento final del trabajo de titulación denominado Desarrollo de un sistema de monitoreo y conteo de personas utilizando visión por computadora, para optimizar la gestión de aforo en la biblioteca de la Universidad Estatal Península de Santa Elena, presentado por la estudiante, Panchana Magallanes Génesis Nicole fue enviado al Sistema Antiplagio, presentando un porcentaje de similitud correspondiente al 5%, por lo que se aprueba el trabajo para que continúe con el proceso de titulación.



WALTER ARMANDO OROZCO IGUASNIA Firmado digitalmente por
WALTER ARMANDO OROZCO
IGUASNIA
Fecha: 2024.12.12 10:52:08 -05'00'

**Ing. Walter Orozco Iguasnia. MSc
TUTOR**



UPSE
UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y
TELECOMUNICACIONES

AUTORIZACIÓN

Yo, Panchana Magallanes Génesis Nicole

Autorizo a la Universidad Estatal Península de Santa Elena, para que haga de este trabajo de titulación o parte de él, un documento disponible para su lectura consulta y procesos de investigación, según las normas de la Institución.

Cedo los derechos en línea patrimoniales del trabajo de titulación con fines de difusión pública, dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando mis derechos de autor

Santa Elena, a los 12 días del mes de diciembre del año
2024

GENESIS PANCHANA.

Génesis Panchana Magallanes

AGRADECIMIENTO

Quiero agradecer a Dios por ser mi guía y mi fortaleza a lo largo de este proceso de estudio, quiero extender mi agradecimiento a mi madre María M, mi padre Stalin P, mis abuelitos, mis hermanas, sobrino, cuñado y tíos, por ser ese apoyo incondicional en mi vida y mi mayor inspiración, la confianza que me han brindado fue la motivación para alcanzar esta meta, este logro es tan mío como de ustedes.

Gracias familia por el amor y aliento brindado en estos años de estudio. Este trabajo es el reflejo de su apoyo incansable y su fe en mí.

A mi tutor Ing. Walter Orozco por ser un guía en este proyecto de titulación, sus conocimientos fueron primordiales.

A la Ing. Mónica Jaramillo por ser una excelente maestra, quien impartió conocimientos fundamentales a través de mi formación académica.

Al Ing. Enrique Montenegro quien desde el inicio de mi vida académica observó todo mi proceso.

Génesis Nicole, Panchana Magallanes

DEDICATORIA

Este trabajo está dedicado a Dios, quien ilumina mi camino y me da la fuerza necesaria para seguir este adelante.

A mi madre María M. por creer siempre en mí, por ser la voz de aliento en mi vida, por darme el apoyo moral que siempre he necesitado en los momentos más difíciles, a familia quien siempre ha estado a mi lado, motivándome y brindándome su amor, su apoyo fue fundamental en mi vida académica. A mi mascota Ody quién no puede leer esto, pero también fue parte de mi proceso académico.

A mi prometido Ing. Sabino Ch. tu amor y apoyo fueron esenciales, gracias por estar a mi lado y ser esa fortaleza de amigo, pareja y confidente.

A mis amigos Ing. Génesis P, Christian DP, Yolanda T, Ing. Bruce A, Gregory R, Ing. Robert M, Marcos F, Raul A, Andrés H, Ing. Melanie M, Ing. Lesly M, Ing. Ariel T, Ing Steven T, Ing. Steven R, Ing. Hellen L, Ing. Elizabeth C, Michael M, Christian V y demás compañeros que hicieron de mi vida universitaria un recuerdo muy feliz desde el inicio.

Génesis Nicole, Panchana Magallanes

ÍNDICE GENERAL

TRIBUNAL DE SUSTENTACIÓN	II
CERTIFICACIÓN	III
DECLARACIÓN DE RESPONSABILIDAD	IV
CERTIFICACIÓN DE ANTIPLAGIO.....	V
AUTORIZACIÓN	VI
AGRADECIMIENTO	VII
DEDICATORIA	VIII
ÍNDICE GENERAL.....	IX
ÍNDICE DE TABLAS	XII
ÍNDICE DE FIGURAS.....	XIII
RESUMEN.....	XV
ABSTRACT.....	XVI
INTRODUCCIÓN	1
CAPÍTULO 1. FUNDAMENTACIÓN	3
1.1 Antecedentes	3
1.2 Descripción del Proyecto	5
1.3 Objetivos del Proyecto	9
1.3.1 Objetivo General	9
1.3.2 Objetivos Específicos:.....	9
1.4 Justificación del Proyecto	10
1.5 Alcance del Proyecto.....	12
1.6 Metodología del Proyecto	13
1.6.1 Metodología de Investigación	13

1.6.2 Beneficiarios del Proyecto.....	14
1.6.2.1 Beneficiarios Directos.....	14
1.6.2.2 Beneficiarios Indirectos.....	14
1.6.3 Variables.....	14
1.6.4 Análisis de recolección de datos.....	15
1.7 Metodología de desarrollo.....	18
CAPÍTULO 2. PROPUESTA.....	20
2.1 Marco Contextual.....	20
2.1.1 UPSE (Universidad Estatal Península de Santa Elena).....	20
2.1.2 MISIÓN.....	20
2.1.3 VISIÓN.....	20
2.2 Marco Conceptual.....	20
2.2.1 Inteligencia Artificial.....	20
2.2.2 Visión por computador.....	21
2.2.3 Redes neuronales.....	21
2.2.4 Redes Neuronales Convolucionales (CNN).....	21
2.2.5 Aprendizaje profundo.....	22
2.2.6 Aprendizaje automático.....	22
2.2.7 Aprendizaje Multi-Escala.....	22
2.2.8 Detección de objetos.....	22
2.2.9 Detección y seguimiento de personas.....	23
2.2.10 Re-Identificación de Personas.....	23
2.2.11 Aprendizaje de características en múltiples escalas (OSNet).....	23
2.2.12 Visual Studio Code.....	24
2.2.13 Python.....	24
2.2.14 FastAPI.....	24
2.2.15 Power BI.....	24
2.2.16 Vue.js.....	24

2.2.17 SQL Server	24
2.2.18 CVAT	25
2.2.19 Protocolo RTSP	25
2.3 Marco Teórico	25
2.3.1 Visión por Computadora en Espacios Cerrados	25
2.3.2 Inteligencia Artificial: Visión por Computadora Aplicada en Espacios Educativos	26
2.3.3 Visión por Computadora para Conteo y Reidentificación de Personas ...	26
2.2 Requerimientos	27
2.2.1 Requerimientos Funcionales	27
2.2.2 Requerimientos no Funcionales	31
2.3. Componente de la Propuesta	32
2.3.1. Arquitectura del Sistema	32
2.3.2. Diagramas de casos de uso	33
2.3.3 Modelado de Datos	41
2.3.4. Evaluación y Selección del Modelo de Detección	41
2.4. Diseño de Interfaces	48
2.4.1. Pantalla de Afluencia de Personas	49
2.4.2. Pantalla de Informes de Visitas	49
2.4.3. Pantalla de Informe de Visitas por Fecha	50
2.4.4. Pantalla de Gráfico Anual de Visitas	51
2.4.5. Pantalla de Visitas del Mes: Por Hora de Entrada y Salida	51
2.5. Pruebas	52
2.5.1 Entrenamiento de Modelo	52
2.5.2 Prueba de Detección en Tiempo Real	54
2.6. Resultados	54
2.6.1 Resultados del entrenamiento por época	54
2.6.2 Entrenamiento con 50 Epochs	56

2.6.3 Resultados de detección	59
2.7. Conclusiones	61
2.8. Recomendaciones.....	62
Referencias.....	63
Anexo	69

ÍNDICE DE TABLAS

Tabla 1 Beneficiarios directos del proyecto.....	14
Tabla 2: Requerimientos Funcionales.....	30
Tabla 3: Requerimientos no Funcionales.....	32
Tabla 4: Frontend - Backend.....	33
Tabla 5: Caso de Uso: Detección y Conteo de Personas.	34
Tabla 6: Caso de Uso: Visualización de Datos.	35
Tabla 7: Caso de Uso: Generación de Informes.....	36
Tabla 8: Caso de Uso: Registro de Personas en Tiempo Real.	37
Tabla 9: Caso de Uso: Monitoreo del Aforo.....	38
Tabla 10: Caso de Uso: Registro de Tiempo de Permanencia.	39
Tabla 11: Caso de Uso: Generación de Reportes de Permanencia.	40
Tabla 12: Comparativa de resultados modelos.	42
Tabla 13: Resultados del entrenamiento por época	55

ÍNDICE DE FIGURAS

Figura 1: Metodología en cascada para el entrenamiento del modelo de conteo de personas.....	18
Figura 2: Metodología en cascada para el sistema de conteo de personas en Biblioteca UPSE	19
Figura 3: Diagrama general.....	32
Figura 4: Modelado - Base de datos.....	41
Figura 5: Gráficas de entrenamiento y validación YOLOv8.	42
Figura 6: Matriz de confusión YOLOv8.....	43
Figura 7: Ejemplo de detección de YOLOv8.....	43
Figura 8: Matriz de confusión Faster R-CNN.....	44
Figura 9: Ejemplo de detección de Faster R-CNN.....	44
Figura 10: Matriz de confusión SSD.....	45
Figura 11: Ejemplo de detección de SSD.	45
Figura 12: Ejemplo de detección de RetinaNet.	46
Figura 13: Matriz de confusión RetinaNet.....	47
Figura 14: Pantalla de Afluencia de Personas.....	49
Figura 15: Pantalla de Registro Diario de Visitas.....	50
Figura 16: Pantalla de Informe de Visitas por Fecha.	50
Figura 17: Pantalla Gráfico Anual de Visitas.	51
Figura 18: Pantalla de Visitas del Mes.....	52
Figura 19: Resultados del entrenamiento por época.	53
Figura 20: Prueba de la detección en tiempo real.	54
Figura 21: Gráfico Precision-Recall para 50 Epochs.....	56

Figura 22: Gráfico F1-Confianza para 50 Epochs.	56
Figura 23: Matriz de confusión para 50 Epochs.	57
Figura 24: Correlogramas de etiquetas del modelo.....	57
Figura 25: Resultados de detección predicción.....	58
Figura 26: Resultados de Detección Etiqueta	58
Figura 27: Resultados de detección 1.....	59
Figura 28: Resultados de detección 2.....	59
Figura 29: Biblioteca UPSE.....	69
Figura 30: Registro Manual Biblioteca UPSE	69

RESUMEN

En la actualidad, los datos son fundamentales para la toma de decisiones estratégicas en organizaciones, apoyados en tecnologías avanzadas. La biblioteca de la Universidad Estatal Península de Santa Elena enfrenta problemas con el registro manual de visitantes, lo que genera un control insuficiente del aforo. Este desafío motivó el desarrollo de un sistema web capaz de detectar y contar personas automáticamente mediante redes neuronales convolucionales. La solución incluye funcionalidades como: i) visualización en tiempo real del porcentaje de aforo; ii) reportes de datos sobre la hora de mayor afluencia y, iii) registro de la cantidad diaria de visitantes. El proyecto se centra como una investigación aplicada, enfocada al monitoreo y conteo de personas, para ello, utiliza algoritmos de inteligencia artificial, de manera particular, tecnologías de visión por computador. La evaluación del modelo logró un 80% de precisión en la detección de personas, destacando su efectividad para optimizar la gestión del aforo en la biblioteca.

Palabras claves: Visión por computador, redes convolucionales, dashboard.

ABSTRACT

Nowadays, data is fundamental for strategic decision making in organizations, supported by advanced technologies. The library of the Peninsula St. Helena State University faces problems with the manual registration of visitors, which generates an insufficient control of the capacity. This challenge motivated the development of a web-based system capable of automatically detecting and counting people through convolutional neural networks. The solution includes functionalities such as: i) real-time visualization of the percentage of capacity; ii) data reports on the hour of greatest affluence and, iii) registration of the daily number of visitors. The project is focused as applied research, focused on monitoring and counting people, using artificial intelligence algorithms, in particular, computer vision technologies. The evaluation of the model achieved 80% accuracy in the detection of people, highlighting its effectiveness in optimizing the management of the library's capacity.

Keywords: Computer vision, convolutional networks, dashboard.

INTRODUCCIÓN

La identificación y seguimiento de personas u objetos representan estrategias esenciales dentro de la visión por computador, que permiten detectar y rastrear de manera automática en imágenes o videos [1]. Estas técnicas utilizan algoritmos avanzados, tales como redes neuronales convoluciones para poder detectar a las personas u objetos dentro de una escena [2].

El seguimiento de objetos se encarga de mantener la identificación de esos elementos a lo largo del tiempo, incluso cuando se mueven dentro del área de las cámaras [3]. Al combinar ambas técnicas, se puede obtener un flujo continuo de información sobre la posición y el movimiento de los objetos, lo que es fundamental en aplicaciones como la seguridad, la gestión del tráfico [4] y, en el caso de este proyecto, el monitoreo de personas en entornos como la biblioteca de la UPSE, garantizando un conteo y un análisis en tiempo real de la ocupación del espacio.

El avance de la visión por computador y la inteligencia artificial han hecho posible el desarrollo de sistemas confiables para la detección y conteo de personas en tiempo real. Este proyecto busca diseñar una solución que, mediante el uso de tecnologías, permita monitorear de manera continua cuantas personas se encuentran en biblioteca en cualquier momento. El uso de técnicas de procesamiento de imágenes y algoritmos de aprendizaje profundo espera ofrecer una solución práctica y efectiva, capaz de brindar información en tiempo real sobre el aforo, para la toma de decisiones informadas para mejorar la gestión en biblioteca.

La Universidad Estatal Península de Santa Elena (UPSE) se enfrenta al desafío de gestionar el aforo de su biblioteca, un espacio que no solo es utilizado por estudiantes, sino también por docentes e investigadores, lo que incrementa la complejidad del control de este. Los métodos tradicionales empleados hasta la fecha, como el conteo manual o los sistemas de control basados en entradas y salidas, presentan notables limitaciones en cuanto a precisión y eficiencia. Estas soluciones no permiten un monitoreo en tiempo real ni ofrecen una visión detallada del flujo de personas dentro de las instalaciones, lo que impide que las autoridades tomen decisiones informadas para garantizar la comodidad de los usuarios.

Ante este reto, se presenta el desarrollo de un sistema automatizado de monitoreo y conteo de personas utilizando visión por computadora, como una respuesta innovadora. La visión por computadora, complementada por algoritmos de inteligencia artificial y aprendizaje profundo, ha demostrado avances significativos en su capacidad para detectar, rastrear y contar personas en diferentes entornos. La implementación de estas tecnologías en la biblioteca de la UPSE no solo permitirá monitorear el aforo en tiempo real, sino también obtener datos claves como el tiempo de permanencia de los usuarios en el espacio.

La finalidad de este proyecto es diseñar un sistema de monitoreo y conteo de personas, fundamentado en técnicas de procesamiento de imágenes y algoritmo de redes neuronales, con el fin de optimizar la gestión del aforo en la biblioteca de UPSE. Este sistema realizará detección de personas en el área de enfoque de la cámara.

Se espera que este sistema automatizado no solo beneficie a la biblioteca de la UPSE, sino que también sirva como modelo para la implementación de soluciones similares en otros espacios de la universidad, como auditorios, salas de conferencias o laboratorios. El uso de estas tecnologías avanzadas contribuirá al avance de la transformación digital de la universidad, favoreciendo la creación de espacios más eficientes y sostenibles para toda su comunidad académica.

CAPÍTULO 1. FUNDAMENTACIÓN

1.1 Antecedentes

La Biblioteca de la UPSE es un espacio académico fundamental que proporciona acceso a materiales bibliográficos físicos y digitales para estudiantes, docentes e investigadores, permitiéndoles satisfacer las necesidades de aprendizaje e investigación. Biblioteca a través de sus servicios se consolida como un pilar esencial en el desarrollo académico de la universidad.

Sin embargo, enfrenta varias problemáticas en su gestión diaria, especialmente en lo que respecta al registro de visitantes. Actualmente, el registro de usuarios se realiza de manera manual como se puede observar en la **Figura 29: Biblioteca UPSE** y **Figura 30: Registro Manual Biblioteca UPSE**, lo que genera inconsistencias en la recolección de datos y dificulta obtener un control preciso sobre el número de personas que utilizan sus instalaciones. Además, muchos visitantes no se registran en la bitácora, lo que impide obtener estadísticas confiables sobre el uso de la biblioteca.

El proceso de registro manual suele ser lento y propenso a errores humanos. En momentos de alta afluencia, como durante exámenes o eventos especiales, los visitantes deben realizar el registro manual, lo que no solo aumenta la probabilidad de errores, sino que desacelera-disminuye el flujo de entrada [5]. Esto puede generar incomodidad entre los visitantes, quienes frecuentemente prefieren evitar la espera y eludir el registro.

Otro desafío común radica en la falta de datos históricos sobre la afluencia de personas, lo que dificulta la creación de estrategias efectivas para optimizar el servicio. Sin esta información, la administración no puede identificar patrones de uso, lo que limita la planificación para eventos o actividades que exijan un control más riguroso del aforo. Así mismo, esta falta de información afecta negativamente la gestión adecuada del personal y los recursos disponibles.

Estas restricciones también afectan la capacidad de la biblioteca para administrar sus espacios de manera eficiente, ya que la falta de información en tiempo real

complica el control del aforo, especialmente en momentos de alta demanda. Sin datos precisos, resulta imposible identificar las horas de mayor afluencia ni ajustar los recursos necesarios para ofrecer un mejor servicio a los usuarios.

Gracias al desarrollo tecnológico significativo que se ha tenido durante los últimos años, se ha podido constatar que hay trabajos o proyectos de titulación que respaldan el desarrollo de esta tecnología, por tal motivo se sustenta con los siguientes trabajos:

En el trabajo de titulación de la universidad del valle de Guatemala [6] titulada “ALGORITMOS DE VISIÓN POR COMPUTADORA PARA EL RECONOCIMIENTO DE LA POSE DE AGENTE EMPLEANDO PROGRAMACION ORIENTADA A OBJETOS Y MULTIHILOS” el autor hace énfasis que “el aprendizaje automático es utilizada para el procesamiento de imágenes, reconocimiento de objetos, personas, etc. Basándose a esta información se puede realizar acciones o tareas”.

En la Escuela Politécnica Nacional se realizó un proyecto previo a la obtención de ingeniero en sistemas informáticos y de computación [7] Titulado como “VISIÓN ARTIFICIAL APLICADA A LA DETECCIÓN E IDENTIFICACIÓN DE PERSONAS EN TIEMPO REAL”, donde el autor menciona que “el reconocimiento de caras o personas es una herramienta utilizada para el control, la seguridad, video vigilancia” además menciona que: “entidades bancarias como Bando del Pacífico y Banco de Guayaquil usan esta biometría para realizar transacciones en las diferentes plataformas disponibles para sus clientes”.

El proyecto titulado como [8] “DETECCIÓN Y CONTEO DE PERSONAS EN ESPACIOS CERRADOS UTILIZANDO ESTRATEGIAS BASADAS EN VISIÓN ARTIFICIAL” destacan que: “El desarrollo de este tipo sistemas suelen ser más viables que otros métodos de conteo de personas existentes, estos sistemas optimizan el trabajo por no requerir monitoreo humano, además la información que se obtiene se puede procesar de manera más rápida y eficiente”.

Dentro de la Universidad Estatal Península De Santa Elena los estudiantes hacen uso de espacios comunes como es la biblioteca general de la universidad, en este espacio común se ha desarrollado una problemática que es no llevar un control de datos exactos de cuantas personas ingresan a la biblioteca universitaria.

Es evidente que se necesita una solución eficiente para abordar este problema y garantizar que los recursos de la biblioteca se utilicen de manera óptima. La visión por computadora emerge como una tecnología prometedora para lograrlo, es así como su aplicación en la biblioteca general de la Universidad Estatal Península de Santa Elena puede resolver de manera efectiva el problema de datos no exactos acerca de los usuarios que asisten a este espacio, además de no tener datos estimados como el promedio de tiempo que los estudiantes hacen uso de este espacio.

1.2 Descripción del Proyecto

El presente trabajo tiene como finalidad desarrollar un sistema de conteo de personas en biblioteca para la gestión informada del aforo de la biblioteca universitaria. Tomando en cuenta el trabajo [8] titulado como “Detección y conteo de personas en espacios cerrados utilizando estrategias basadas en Visión Artificial” hace mención que la aplicación de sistemas apoyada de la tecnología como visión por computador son más eficientes por el hecho de no depender de un ser humano. A continuación, se detallan los componentes que se utilizaran para lograr el entrenamiento del modelo de este proyecto, estos pasos están basados en el artículo [9] estos pasos fueron adaptados a la metodología CRISP-DM que tiene una secuencia similar a las mencionadas en el artículo:

Componente 1: Adquisición de imágenes

Es crucial asegurarse de que las cámaras estén en ubicaciones estratégicas que permitan cubrir de manera efectiva las áreas donde se llevará a cabo el conteo de personas. La captura de imágenes en tiempo real se realiza mediante cámaras IP que emplean protocolos como RTSP, lo cual asegura una transmisión constante y sin interrupciones para los procesos posteriores. Este componente también incluye

la calibración y configuración de las cámaras, garantizando su óptimo desempeño para capturar imágenes en diversas condiciones de iluminación y desde diferentes ángulos.

Componentes 2: Procesamiento de imágenes

El objetivo de este componente es adaptar las imágenes capturadas para que sean procesadas de manera eficiente por el modelo de visión por computadora. Esto implica tareas como el escalado, la conversión de colores y la aplicación de técnicas de preprocesamiento, tales como la reducción de ruido y el aumento del contraste. Se busca mejorar la calidad visual de las imágenes, optimizando la precisión del sistema para la detección y seguimiento de personas. También se emplean técnicas como la normalización y el ajuste de luminosidad para garantizar la consistencia de las imágenes sin importar las condiciones ambientales.

Componentes 3: Entrenamiento del modelo

La función principal de este componente es identificar a las personas presentes en las imágenes o videos. Para ello, se utiliza un modelo de detección de objetos como YOLO (You Only Look Once), que permite localizar y distinguir personas de otros objetos en la escena. Este componente no solo se encarga de detectar personas, sino también de diferenciarlas. El modelo se entrena con un conjunto de datos previamente etiquetado que contiene imágenes de personas en distintos contextos y situaciones.

Componentes 4: Evaluación del modelo

Este componente asegura que las personas detectadas sean monitoreadas de manera continua, evitando errores como el conteo duplicado. Se utilizan algoritmos avanzados como DeepSort, los cuales asignan una identificación única a cada individuo y rastrean su movimiento dentro del área monitoreada. Este proceso es esencial para garantizar la precisión en el conteo, manteniendo un registro en tiempo real de las personas que ingresan y salen del espacio supervisado, reduciendo errores como pérdidas de seguimiento o doble conteo.

El proceso de los componentes para el sistema de conteo de personas se detalla a continuación:

Componente 1: Entrenamiento del Modelo de Visión por Computadora

En este primer módulo, se lleva a cabo el entrenamiento del modelo utilizando técnicas de inteligencia artificial, como redes neuronales convolucionales. El conjunto de datos empleado contiene imágenes etiquetadas de personas en diferentes escenarios, permitiendo al modelo aprender a identificar individuos y diferenciarlos de otros elementos. Este entrenamiento es esencial para lograr una detección confiable y precisa dentro de la biblioteca. El script del modelo se ajusta adecuadamente para asegurar su correcto funcionamiento con los datos proporcionados.

Componente 2: Implementación del Algoritmo Yolo + DeepSort

En este módulo, se desarrolla el script que integra el modelo YOLO para la detección de objetos y el algoritmo DeepSort para el seguimiento de personas en tiempo real. YOLO se encarga de identificar a las personas en las imágenes capturadas, mientras que DeepSort asigna una identidad única a cada individuo para realizar un seguimiento preciso de su movimiento. Esta combinación permite realizar el conteo de personas de manera eficiente y sin duplicaciones. El script se ajusta para garantizar su correcto funcionamiento con los datos de entrada.

Componente 3: Conexión de la Cámara a través de RTSP

Este módulo establece la comunicación entre el sistema y las cámaras disponibles mediante RTSP. Configura la cámara para enviar el flujo de video en tiempo real al sistema, permitiendo que el modelo procese las imágenes sin interrupciones. Una implementación adecuada del RTSP asegura un flujo continuo y estable de datos que alimentan las etapas de detección y seguimiento, minimizando la latencia en la transmisión.

Componente 4: Desarrollo del Backend

Este módulo se encarga de crear la lógica del backend que conecta el frontend con la base de datos. Se utiliza una arquitectura capaz de manejar múltiples solicitudes

concurrentes mediante el uso de hilos, garantizando un alto rendimiento. Además, se implementan endpoints para interactuar con el sistema, así como WebSockets para enviar actualizaciones en tiempo real sobre el conteo y el estado general del sistema. La base de datos almacena la información sobre el flujo de personas y su permanencia en la biblioteca.

Componente 5: Diseño y Desarrollo del Frontend

Este módulo proporciona la interfaz que permite a los administradores de la biblioteca monitorear el conteo de personas en tiempo real. El diseño se centra en la creación de una experiencia de usuario intuitiva y funcional, que incluya la visualización de reportes generados por el sistema. El frontend asegura que cualquier usuario, incluso sin conocimientos técnicos avanzados, pueda interactuar fácilmente con el sistema.

Componente 6: Registro y Almacenamiento de Datos

Este módulo se encarga de gestionar y almacenar los datos generados por el sistema, incluyendo el conteo de personas, su duración de permanencia y registros históricos. Se utiliza una base de datos SQL Server que organiza y estructura eficientemente la información, permitiendo consultas en tiempo real y la generación de reportes. El sistema asegura que los datos sean almacenados de manera segura y estén disponibles para análisis futuros.

Componente 7: Presentación de Datos

El sistema ofrecerá la capacidad de mostrar los resultados en dashboard basados en los datos almacenados. Estos dashboard serán una herramienta clave para los administradores de la biblioteca, permitiéndoles visualizar el flujo de personas a lo largo del tiempo y tomar decisiones informadas para la gestión del espacio. La presentación de los datos se realizará de forma periódica y se presentarán de manera clara mediante gráficos y tablas accesibles desde la interfaz del sistema.

Herramientas utilizadas para el desarrollo del sistema de conteo de personas:

Visual Studio: Utilizado como el entorno de desarrollo principal para gestionar el código del proyecto, incluyendo el procesamiento de imágenes en Python y el backend con FastAPI.

Python: Lenguaje principal para implementar el modelo de visión por computadora y el entrenamiento de detección. Incluye el uso de bibliotecas especializadas como YOLO y DeepSort.

FastAPI: Framework para desarrollar la API que comunica los diferentes componentes del sistema, incluyendo el procesamiento de imágenes y la actualización en tiempo real del conteo.

Power BI: Herramienta para crear visualizaciones y dashboards basados en los datos almacenados en SQL Server.

Vue.js: Tecnología utilizada para desarrollar un frontend dinámico que permita a los administradores gestionar y visualizar información en tiempo real.

Cvat.io: Herramienta utilizada para la anotación de imágenes en el entrenamiento del modelo, facilitando la creación de datos etiquetados.

SQL Server: Base de datos encargada de almacenar y gestionar la información generada, incluyendo los conteos útiles para generar reportes históricos.

1.3 Objetivos del Proyecto

1.3.1 Objetivo General

Desarrollar un sistema de monitoreo y conteo de personas basado en técnicas de visión por computadora, con el fin de optimizar la gestión del aforo y mejorar a utilización de los espacios disponibles en biblioteca.

1.3.2 Objetivos Específicos:

- Seleccionar tecnologías de visión por computadora para mejorar la gestión del seguimiento de personas en la biblioteca de UPSE.
- Desarrollar un sistema utilizando librerías de Python para la detección y seguimiento de personas, que permita realizar un conteo preciso en tiempo real y la identificación de trayectorias individuales.

- Crear una interfaz con dashboard web, para que los administradores accedan eficientemente a los datos de conteo y seguimiento en la biblioteca.

1.4 Justificación del Proyecto

El avance tecnológico en informática y el acceso a internet han logrado que las personas y/o trabajadores puedan manejar más información además ser más productivos y eficientes en sus actividades laborales o personales. Así como lo menciona [10] las tendencias tecnológicas van a gran velocidad, dando como resultado que las organizaciones quieran adquirir estas tecnologías para obtener ventajas competitivas, además que la toma de decisiones sea más eficiente y ágil.

El avance tecnológico en informática y el acceso a Internet han permitido a las personas y trabajadores manejar una mayor cantidad de información, lo que a su vez ha incrementado su productividad y eficiencia en actividades laborales y personales. Como se menciona en [10], las tendencias tecnológicas evolucionan rápidamente, lo que lleva a las organizaciones a adoptar estas tecnologías para obtener ventajas competitivas y facilitar una toma de decisiones más eficiente y ágil.

Dentro de las evoluciones tecnológicas, se encuentra la inteligencia artificial, que ha avanzado significativamente en los últimos años. [11] Una de sus ramas más destacadas es la visión artificial o visión por computadora. Esta disciplina se puede definir como un campo de la inteligencia artificial que incluye métodos para adquirir, procesar y analizar imágenes del mundo real, con el fin de producir información que pueda ser tratada por máquinas.

Esta técnica está basada en la adquisición de imágenes, [12] generalmente en dos dimensiones, para su posterior procesamiento por el computador, con el fin de extraer y medir determinadas propiedades de la imagen adquirida. Se trata, por lo tanto, de una tecnología que combina las computadoras con las cámaras para adquirir, analizar e interpretar imágenes de una forma equivalente a la inspección visual humana.

[13] La visión por computadora es un campo de la IA que permite a las computadoras y a los sistemas obtener información importante a partir de las imágenes digitales, videos y otras entradas visuales, si la inteligencia artificial permite a las computadoras pensar, este campo permite ver observar y comprender el mundo. La visión por computadora entrena máquinas para realizar funciones que realiza el ser humano, ya que al ser entrenado puede realizar funciones en menos tiempo con cámaras, datos y algoritmos, capacitando a las máquinas para inspeccionar productos, observar actividades, detectando defectos imperceptibles ante las capacidades humanas [14].

La visión artificial tiene como principal finalidad dotar a las máquinas de “ojos” para ver lo que ocurre en el mundo real, y así poder tomar decisiones para automatizar cualquier proceso [11].

Según [15] “Esta disciplina tecnológica ha emergido para dar solución a diversos problemas combinando conceptos de informática y procesamiento de imágenes para desarrollar algoritmos y sistemas capaces de analizar y comprender visualmente el mundo que nos rodea”. Los avances en este campo tienen aplicaciones en diversas industrias, lo que hace que sea una disciplina académica relevante y en constante evolución.

El surgimiento de las nuevas tecnologías ha recobrado una gran relevancia en los últimos años gracias a la evolución tecnológica, la detección y seguimiento de personas en ambientes de interiores es uno de los progresos significativo de la tecnología. Según [8] “El análisis automático del movimiento humano es un área de investigación debido a la cantidad de aplicaciones potenciales como son: vigilancia, control y análisis”.

Actualmente, la visión por computadora en espacios cerrados se está convirtiendo en un elemento clave para la eficiencia, seguridad y automatización en una amplia gama de aplicaciones y sectores. [16] “La visión por computador permite simular o replicar las capacidades humanas, como la forma en que percibimos las cosas y el sentido de lo que ven. Su objetivo es identificar, comprender objetos y personas

mediante el análisis de imágenes y/o videos.”. Estos avances están transformando sectores como la automatización industrial, la seguridad en edificios y la robótica de servicio en interiores, como se menciona [14].

El conteo de personas es una métrica esencial para evaluar el uso de la biblioteca, con estos datos, se pueden identificar cantidad de usuarios que visitan a diario la biblioteca y tomar decisiones informadas que ayuden a mejorar su funcionamiento, además para garantizar que estos recursos se utilicen de manera eficiente, que no solo benefician a la administración, sino también a los usuarios.

El diseño de este sistema basado en visión por computador para el conteo de personas y seguimiento del tiempo de permanencia en la biblioteca no solo aborda a las necesidades mencionadas, sino que también muestra el compromiso de la institución con la adopción de tecnología avanzadas para el beneficio de la comunidad estudiantil y el personal administrativo.

1.5 Alcance del Proyecto

El proyecto se centra en entrenar un modelo para detectar y contar personas en la biblioteca de la universidad, aplicando tecnologías avanzadas de visión por computadora y aprendizaje profundo. Este sistema tiene ayudará a mejorar la gestión de aforo al proporcionar datos en tiempo real, lo que posibilitará la toma de decisiones informadas en la administración del espacio.

El modelo de visión por computadora será entrenado para detectar personas en un plano vertical, garantizando un conteo preciso de los visitantes que ingresan a la biblioteca. La implementación del algoritmo Yolo + DeepSort permitirá un seguimiento eficiente y en tiempo real de las personas en las áreas designadas. La conexión con la cámara se realizará a través del protocolo RTSP, asegurando una transmisión fluida desde la cámara IP compatible.

El desarrollo de la API permitirá gestionar la base de datos, manejar hilos para el procesamiento paralelo y establecer comunicación mediante WebSocket, lo que permitirá actualizar en tiempo real el conteo de personas. Por otro lado, el frontend

proporcionará una interfaz intuitiva que permitirá a los administradores acceder a la información del aforo en tiempo real de forma sencilla. Además, el sistema registrará y almacenará los datos de los visitantes en una base de datos SQL Server, lo que facilitará su análisis posterior.

Finalmente, se generarán informes detallados sobre la afluencia de visitantes, proporcionando estadísticas valiosas para la gestión de la biblioteca. Este proyecto cubre las funcionalidades esenciales para optimizar el control de visitantes, facilitar la toma de decisiones y mejorar la gestión del espacio en la biblioteca.

1.6 Metodología del Proyecto

1.6.1 Metodología de Investigación

El diseño de la investigación se la presenta como investigación aplicada, ya que “se caracteriza por resolver problemas específicos o con necesidades concretas, orientadas a la práctica” [17]. La finalidad principal de este trabajo es desarrollar y validar un sistema de monitoreo y conteo de personas basado en visión por computadora. Este proyecto se considera investigación aplicada, ya que se desarrollará un software con el propósito de abordar un problema práctico específico: mejorar la gestión del aforo en la biblioteca de la UPSE.

El proceso se iniciará con la recolección y análisis de datos preliminares sobre el entorno, que incluirán aspectos como la iluminación, el posicionamiento de las cámaras y la cantidad estimada de usuarios. Con base en esta información, se seleccionarán y ajustarán las herramientas y tecnologías a utilizar, como el modelo de detección YOLO y el algoritmo DeepSort para el seguimiento. A lo largo del proyecto, se monitoreará la capacidad del sistema para contar y gestionar la afluencia de personas, buscando mejorar la precisión del conteo y la gestión del aforo.

1.6.2 Beneficiarios del Proyecto

1.6.2.1 Beneficiarios Directos

Los beneficiarios directos serán la administradores y usuarios operativos de la biblioteca de la UPSE, ya que contarán con una herramienta que optimiza la gestión de aforo. El monitoreo en tiempo real les permitirá tener un control preciso sobre la cantidad de personas en la biblioteca y mejorando la organización. Además, podrán visualizar datos reales que les ayuden a tomar decisiones informadas sobre el uso y disposición del espacio.

Beneficiarios Directos	Cantidad
Personal de Jefatura	1 persona
Personal Operativo	4 personas
Total	5 personas

Tabla 1 Beneficiarios directos del proyecto

1.6.2.2 Beneficiarios Indirectos

Los beneficiarios indirectos de este sistema de conteo de personas incluyen a profesores, estudiantes y visitantes de la biblioteca, quienes experimentarán un entorno más organizado y una gestión más eficiente del espacio. Además, las instituciones educativas y académicas podrán aprovechar los datos generados por el sistema para desarrollar estrategias orientadas a la optimización de recursos en espacios de aprendizaje y recreación.

1.6.3 Variables

- **Variable Dependiente:** Gestión del aforo, medida a través de la solución informática diseñada.
- **Variable Independiente:** Número de personas que ingresan a la biblioteca.

1.6.4 Análisis de recolección de datos

La recolección de datos para este proyecto se llevó a cabo mediante entrevistas directas con actores clave, como la encargada de la biblioteca, lo que proporcionó información sobre el flujo de personas, horarios de mayor afluencia y los desafíos en la gestión de aforo. Esta información fue crucial para definir los requisitos del sistema de conteo.

A demás, se realizó una observación directa del proceso de acceso a la biblioteca, lo que permitió comprender mejor el comportamiento de los usuarios al entrar y salir. Esta fase de recolección de datos fue valiosa para identificar posibles desafíos en la implementación del sistema, como la variabilidad en las condiciones de iluminación o el comportamiento impredecible de los usuarios al ingresar.

1.6.4.1 Análisis de Entrevista

Entrevistado: Nivel Ejecutivo.

Objetivo: Obtener información sobre los métodos actuales de registro de visitantes, los problemas enfrentados, y las necesidades para mejorar la gestión del aforo mediante un sistema automatizado de monitoreo y conteo de personas.

1. ¿Cómo se realiza actualmente el registro de visitantes en la biblioteca?

Respuesta:

- El registro de visitantes en la biblioteca se realiza de manera manual.
- Los usuarios ingresan sus datos en una hoja de registro.
- La información se almacena en formato físico.
- Este método dificulta el acceso rápido a la información.

Análisis: Revela una deficiencia en la velocidad y precisión del registro de visitantes. El uso de un sistema manual limita la capacidad de la biblioteca para gestionar y analizar los datos en tiempo real.

2. ¿Existen momentos en los que la afluencia de visitantes es más alta?

- Exámenes

Análisis: Se observa que existen momentos con mayor concentración de usuarios.

3. ¿Qué obstáculos encuentran al registrar manualmente a los usuarios durante los picos de afluencia?

- Relentiza el registro manual
- Filas dentro de la biblioteca

Análisis: La respuesta resalta la necesidad de un sistema automatizado eficiente que reduzca los tiempos de espera y aumente la precisión en la recopilación de datos.

4. ¿Qué tipo de información adicional sobre los visitantes les gustaría poder recolectar con mayor exactitud?

- Horas de ingreso de usuarios
- Horas de salida de usuarios
- Tiempos de permanencia
- Numero o porcentajes de visitantes diarios

Análisis: Esto indica que el sistema debe incluir funcionalidades avanzadas de seguimiento y generación de informes, basadas en datos precisos y en tiempo real.

5. ¿Cómo afecta la falta de datos pasados sobre el uso de la biblioteca a la toma de decisiones?

- Limitación de planificación de eventos
- Mejorar servicios

Análisis: La necesidad de un sistema que registre datos históricos y facilite el análisis para una toma de decisiones informada.

6. ¿Qué mejoras consideraría necesarias para optimizar el control y gestión de visitantes en la biblioteca?

- Sistema digital que permita recopilar datos en tiempo real
- Reportes automáticos

Análisis: Se sugiere que el sistema debe incluir herramientas de automatización y generación de gráficos estadísticos para satisfacer las necesidades de los administradores.

7. ¿Qué tan importante es para ustedes contar con información en tiempo real sobre el número de personas en la biblioteca?

- Asegurar el cumplimiento de normativas de aforo
- Asegurar la tranquilidad y orden para la comodidad del estudio de los usuarios

Análisis: Esto enfatiza la necesidad de desarrollar un sistema que pueda actualizar y mostrar información en tiempo real.

8. ¿Han recibido quejas o comentarios de los usuarios sobre el proceso de registro?

- Si

Análisis: Refuerza la urgencia de implementar un sistema eficiente que elimine las molestias asociadas al registro manual.

1.6.4.2 Conclusión del Análisis de la Entrevista

A partir del análisis de la entrevista, se identificaron las siguientes necesidades y aspectos clave relacionados con la gestión del aforo en la biblioteca:

1. **Horas pico:** Identificación de los momentos de mayor afluencia de usuarios, lo que afecta la gestión del espacio.
2. **Capacidad de puestos disponibles:** Evaluar la cantidad de asientos y espacios disponibles para garantizar el cumplimiento de las normativas de aforo.
3. **Tecnología disponible:** Examinar las herramientas y sistemas actuales utilizados para registrar y monitorear a los usuarios.

Los principales problemas identificados incluyen la ineficiencia del registro manual y la falta de datos históricos. La implementación de un sistema digital resolverá estas cuestiones, mejorando la precisión de los datos, reduciendo los tiempos de espera y permitiendo que los administradores tomen decisiones informadas sobre la gestión de recursos y espacios.

1.7 Metodología de desarrollo

El desarrollo del sistema de conteo de visitantes en la biblioteca universitaria utilizará la metodología en cascada. Donde se podrá ver la planificación y ejecución secuencial en diferentes etapas del proyecto, importante se debe completar cada fase para poder pasar a la siguiente. Así ayudando a tener una buena estructura dando facilidad a la gestión del proyecto logrando que los requisitos se cumplan desde el inicio.

Posteriormente, se explicará a detalle estas etapas de la metodología:

- **Investigación:** Se elabora la recopilación de información enlazada con las tecnologías de visión por computadora y el uso de conteo de personas.
- **Selección:** Con base en la investigación previa, se seleccionan las herramientas y tecnologías adecuadas para el entrenamiento del modelo.
- **Implementación:** Se entrena el modelo de paso a paso, iniciando con la operatividad de la selección de imágenes, etiquetado y procesamiento de reconocimiento de personas y seguimiento, Incluyendo mejoras en cada iteración.
- **Prueba y validación:** Se procede con las pruebas sistemáticas para confirmar que el sistema cumpla con las condiciones establecidas y funcione en diferentes escenarios.

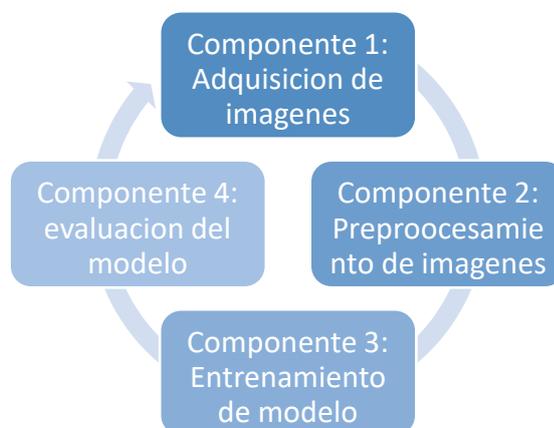


Figura 1: Metodología CRISP-DM entrenamiento del modelo de conteo de personas.

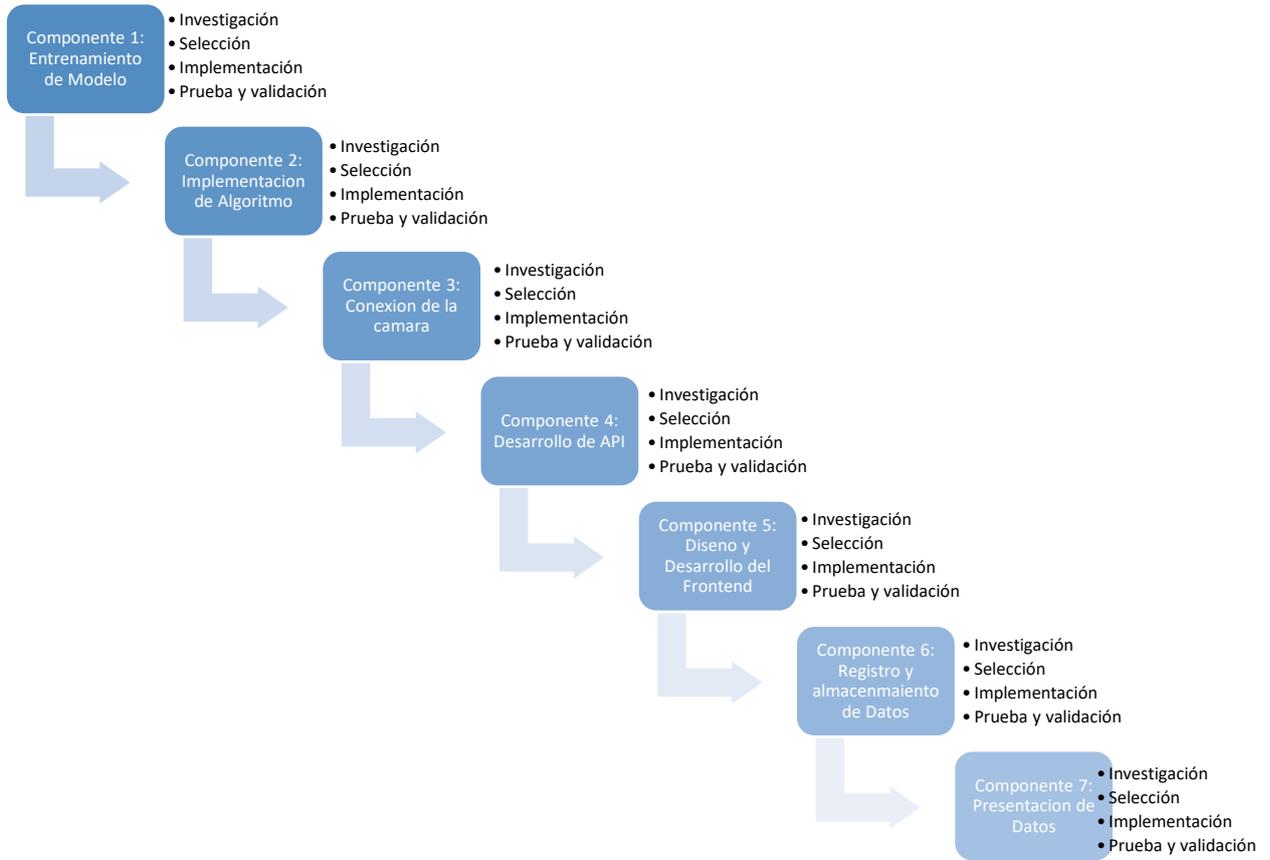


Figura 2: Metodología en cascada para el sistema de conteo de personas en Biblioteca UPSE

CAPÍTULO 2. PROPUESTA

2.1 Marco Contextual

2.1.1 UPSE (Universidad Estatal Península de Santa Elena)

Es una institución educativa ubicada en la Provincia De Santa Elena, Ecuador. La universidad fue fundada el 22 de Julio de 1998 y su principal objetivo es proporcionar educación superior de calidad. La universidad se ha empeñado y esforzado en ofrecer una amplia gama de programas académicos en diversas áreas de estudio incluida la ingeniería de tecnologías de la información que pertenece a la facultad de sistemas y telecomunicaciones donde brindan grandes aprendizajes académicos del área tecnológica [18].

2.1.2 MISIÓN

Formar profesionales que aportan al desarrollo sostenible, contribuye a la solución de los problemas de la comunidad y promueve la cultura [19].

2.1.3 VISIÓN

Ser reconocida por su calidad académica, impacto de sus investigaciones y su aporte al desarrollo de la sociedad [19].

2.2 Marco Conceptual

2.2.1 Inteligencia Artificial

La inteligencia artificial (IA) es un campo de la informática que se dedica a desarrollar sistemas capaces de realizar tareas que, en principio, requieren inteligencia humana. Estas tareas pueden incluir el razonamiento, la toma de decisiones, el reconocimiento de patrones, la comprensión de lenguaje natural, la percepción visual y auditiva, entre otras [20].

La IA permite a las máquinas aprender datos, adecuando a nuevas mejoras o en caso de diferentes situaciones. También tiene relación con el machine learning, la computer vision, el procesamiento de lenguaje natural (NLP), con distintos procesamientos que asimilan habilidades humanas en el ámbito tecnológico [21].

2.2.2 Visión por computador

Busca reemplazar las habilidades humanas para así que las maquinas sepan que a través de la información guardada tenga la capacidad de comprender las visualidades humanas. Permitiendo que las maquinas con las capturas o imagen las procesen y genere datos significativos [22].

La visión por computador busca replicar las funciones de la percepción visual humana, logrando resultados que se asemejan a la interpretación humana. Su utilidad se ha evidenciado en aplicaciones como sistemas de vigilancia, reconocimiento facial y reconocimiento de huellas dactilares, los cuales han sido integrados exitosamente en diversas industrias [23].

Según un artículo del Journal de ESPE [24], “La visión por computador se desempeña en múltiples campos, ya que permite a las máquinas interpretar y comprender el entorno visual de manera similar a los humanos. Esta tecnología sirve para practicas desde procesamientos de multitudes de personas hasta la vigilancia de seguridad. La visión por computadora logra procesar y analizar imágenes y videos y ser un soporte a la toma de decisiones en tiempo real”.

2.2.3 Redes neuronales

Las redes neuronales son un modelo computacional inspirado en la estructura del cerebro humano, diseñado para procesar información de manera similar a como lo harían las neuronas biológicas. Se componen de nodos (neuronas artificiales) organizados en capas (capa de entrada, capas ocultas y capa de salida). Cada nodo en una capa está conectado con los de la siguiente capa, y las conexiones tienen pesos que se ajustan durante el entrenamiento para optimizar la salida del modelo en tareas específicas como clasificación, regresión o reconocimiento de patrones [25].

2.2.4 Redes Neuronales Convolucionales (CNN)

Las CNN están diseñadas para manejar los datos de las imágenes. Haciendo uso de convoluciones que conecta a cada neurona con todas las demás, facilitando la identificación de características locales en las imágenes. Estas se encuentran por

capas pooling, que disminuyen la dimensión y da la facilidad a la captura de patrones espaciales que se profundiza en la red [26].

2.2.5 Aprendizaje profundo

Deep learning se utiliza para formar relaciones complejas dentro de los datos. A distinción de perspectiva tradicionales de aprendizaje automático, por lo que se define manualmente, el aprendizaje profundo interpreta los datos automáticamente. Así realizando el trabajo como el procesamiento de img, reconocimiento y lenguaje natural [21].

2.2.6 Aprendizaje automático

El aprendizaje automático tiene la función de prepararse sin que lo programen para ejecutar una tarea. Este algoritmo compagina modelos en base a la información que entra para poderla clasificar o en tal caso predecir. Así como hay diferentes tipos, tiene agregado el aprendizaje por refuerzo, supervisado o no [27].

2.2.7 Aprendizaje Multi-Escala

Utiliza una particular técnica relacionada con las redes neuronales como es la captura de las características de un nivel local y global. Esta metodología mejora la capacidad de reconocimiento permitiendo que los modelos analicen desde los patrones más pequeños hasta los más amplios dando como resultado la Re-identificación por medio de detalles específicos o por detalles generales de la apariencia [28].

2.2.8 Detección de objetos

Detección de objetos constituye una de las funciones esenciales en el ámbito de la visión por computadora. Este proceso se ejecuta mediante diversos métodos, los cuales típicamente analizan las características de todos los objetos presentes en la imagen. Se trata de un procedimiento empleado para reconocer la existencia de determinados elementos en una imagen [29].

2.2.9 Detección y seguimiento de personas

La detección y seguimiento de personas con el uso de visión por computador implica la aplicación de algoritmos de aprendizaje profundos y cámaras de vigilancia para la identificación y seguimiento de personas en tiempo real dentro de establecimientos cerrados. Este enfoque tecnológico se emplea extensamente en diversos sectores como el comercio minorista transporte hotelería etcétera. Los sistemas de conteo de personas, como los sistemas 2D y 3D, se utilizan cámaras y redes neuronales profundas para precisar la cantidad de personas en una ubicación específica estos sistemas son útiles para supervisar filas y puntos de acceso en aeropuertos áreas de ventas minoristas y áreas de Atención al Cliente ofreciendo información valiosa a las organizaciones [30] [31] [32].

2.2.10 Re-Identificación de Personas

La re-identificación de personas es una tarea en la que se busca reconocer a una misma persona en diferentes imágenes o videos capturados por múltiples cámaras, en diversos momentos y desde distintos ángulos. Esto implica la capacidad de correlacionar imágenes de una persona con otras vistas previas, permitiendo su rastreo en entornos complejos, como aeropuertos, estaciones de tren o espacios públicos. Uno de los principales retos es la variación en las posturas, ropa, iluminación y el cambio en las perspectivas de las cámaras. La re-identificación se apoya fuertemente en modelos de aprendizaje profundo y redes neuronales para analizar las características únicas de los individuos y permitir su identificación precisa [33].

2.2.11 Aprendizaje de características en múltiples escalas (OSNet)

El aprendizaje de características multiescalar se usa comúnmente para la re-identificación de personas, coloca bloques convolucionales que da la facilidad de la captura de detalles, por ejemplo, en la ropa, y características específicas o la estructura corporal. Tiene como ventaja la adaptación a diferentes condiciones de iluminación, perspectivas y posturas, mejora la precisión al momento de reconocer en diferentes entornos [34].

2.2.12 Visual Studio Code

Se desempeña como un editor de códigos para el desarrollo de aquellos o también de aplicaciones web y móviles, incluso trabaja en la nube en diferentes lenguajes de programación como Python, C++, y otros[35].

2.2.13 Python

Es un lenguaje de programación muy conocido por su particularidad de ser legible y tener una sintaxis clara. Python es ampliamente utilizado en una variedad de aplicaciones, incluyendo desarrollo web, análisis de datos, inteligencia artificial y más [36].

2.2.14 FastAPI

Es un framework moderno y de alto rendimiento para la creación de aplicaciones web en Python. Fue diseñado para facilitar el desarrollo rápido y eficiente de APIs, utilizando las características más avanzadas del lenguaje [37].

2.2.15 Power BI

Es una herramienta de análisis de datos y visualización también desarrollada por Microsoft. Está diseñada para permitir a los usuarios conectar, modelar y visualizar datos de diversas fuentes.

2.2.16 Vue.js

Es un framework progresivo de JavaScript utilizado para construir interfaces de usuario interactivas y de una sola página (SPA). Vue.js es conocido por su enfoque modular y su curva de aprendizaje gradual, lo que lo hace muy popular entre los desarrolladores web [38].

2.2.17 SQL Server

Es una base de datos ampliamente utilizado para almacenar, organizar y administrar datos. Se maneja más dentro del ambiente empresariales para las BD, así almacenando los datos y su posterior análisis [39].

2.2.18 CVAT

Es útil en relación sobre las anotaciones de las img que da la facilidad tanto de etiquetado de información por decir de los datos para utilizar en un proyecto que va en base a la visión por computadora. Accede a distintas anotaciones por decir los puntos, delimitadores, líneas hasta incluso los polígonos, proporcionando los etiquetados o anotaciones automáticas para que el proceso se acelere e mejore la producción de preparar los conjuntos que se entrenaran en el modelos [40].

2.2.19 Protocolo RTSP

Controla a momento de transmitir los datos multimedia, como audio y video, a través de Internet. Su función principal es gestionar y controlar las sesiones de transmisión, permitiendo a los usuarios interactuar con el contenido en tiempo real [41].

2.3 Marco Teórico

2.3.1 Visión por Computadora en Espacios Cerrados

Se enfoca en la interpretación y análisis de imágenes y videos por parte de las máquinas. Esta tecnología se ocupa específicamente para mecanizar tareas de control de aforo, también de vigilancia o incluso de detección de diferentes comportamientos tanto como irregulares.

Según Szeliski [42], dice que la visión por computadoras que aprovecha por decir las bibliotecas, agencias o centros comerciales agarran la capacidad de la función estable e mejora de la precisión tanto de la detección como el algoritmo.

En estos ambientes cerrados, la estabilidad en la iluminación y la estructura favorece a la utilización del uso de modelos como las redes neuronales convolucionales (CNN), que pueden ser entrenadas para reconocer patrones específicos y realizar el seguimiento de objetos o personas [1].

Zhu [43] demostró que los sistemas de visión por computadora aplicados en interiores tienen un rendimiento significativamente mayor en la reidentificación de

personas debido a las condiciones más constantes, lo que permite que los algoritmos de detección y seguimiento operen de manera eficiente.

2.3.2 Inteligencia Artificial: Visión por Computadora Aplicada en Espacios Educativos

El uso de la visión por computadora en entornos educativos ha ido en aumento debido a su capacidad para mejorar la seguridad, la gestión del espacio y la eficiencia operativa. Los sistemas de IA en campus educativos permiten el monitoreo en tiempo real del aforo y la identificación de patrones de uso, lo que contribuye a una mejor asignación de recursos. La visión por computadora aplicada en estos entornos ayuda a mejorar la distribución del espacio y facilita la optimización de actividades como la asignación de aulas, el control de acceso a laboratorios y la gestión de auditorios durante eventos masivos [44].

2.3.3 Visión por Computadora para Conteo y Reidentificación de Personas

La visión por computadora para el conteo y la reidentificación de personas es un campo en constante evolución, impulsado por la creciente necesidad de gestionar la cantidad de personas en espacios públicos, comerciales y educativos. Los sistemas de conteo de personas emplean algoritmos avanzados que permiten identificar, rastrear, clasificar y contar individuos en tiempo real mediante el análisis de imágenes y videos capturados por cámaras de vigilancia [45].

Estos sistemas resultan particularmente útiles en entornos donde el control del aforo es esencial, como bibliotecas, auditorios y eventos masivos. Utilizando técnicas de detección de objetos, como las redes neuronales convolucionales (CNN), y algoritmos de seguimiento como DeepSort, es posible lograr un conteo preciso y eficiente de personas, reduciendo los errores comunes en los métodos manuales [46].

Además, la reidentificación de personas es crucial cuando es necesario seguir el movimiento de los individuos a través de diferentes cámaras y a lo largo de distintos momentos. Esta capacidad permite, por ejemplo, analizar el comportamiento de los

usuarios en un espacio determinado, identificar patrones de uso y optimizar la disposición del espacio [47].

La reidentificación se basa en características únicas de cada individuo, como la vestimenta y la estructura corporal, y se apoya en modelos de aprendizaje profundo que mejoran su precisión a medida que se entrenan con más datos [48].

2.2 Requerimientos

2.2.1 Requerimientos Funcionales

N.º R.F	REQUERIMIENTOS	DETALLES
RF 1	Detección de Personas	El sistema debe ser capaz de detectar y clasificar personas en tiempo real utilizando algoritmos de visión por computadora, específicamente redes neuronales convolucionales (CNN).
RF 2	Conteo de Personas	El sistema debe contar el número de personas presentes en la biblioteca en tiempo real y actualizar esta información continuamente.
RF 3	Seguimiento de Personas	Ocupa los algoritmos de Yolo y DeepSort para en si la persecución de los individuos vistos, logrando una identificación o semejanza única para la persona.
RF 4	Conexión de Cámara	Ve la necesidad del protocolo RTSP para poder logra la conexión tanto del sistema

		y la cámara, así visualizando la captura del video en tiempo real.
RF 5	Interfaz de Usuario	Debe permitir la visualización esencial que es el conteo de los individuos de la captura en tiempo real de la cámara, para poder acceder a las ventanas con la información de los detalles de la posterior gestión de forma fácil e intuitiva.
RF 6	Almacenamiento de Datos	El sistema debe registrar y almacenar datos sobre el conteo de personas, la duración de su permanencia y datos históricos en una base de datos SQL Server.
RF 7	Presentación de datos	Proporcionar la capacidad de mostrar datos de manera automática sobre la afluencia de personas.
RF 8	Comunicación en Tiempo Real	Implementar WebSockets en la API para enviar actualizaciones instantáneas sobre el conteo de personas y el estado del sistema.
RF 9	Visualización de Datos en Tiempo Real	Muestra los individuos detectados que están presentes en la biblioteca dentro de la interfaz.
RF 10	Pruebas de Sistema	Ejecutar pruebas necesarias para poder verificar la funcionalidad correcta en diferentes escenas.

RF 11	Implementación de Detección De Objetos En Tiempo Real	Debe acceder a la observación de la detección de personas con el modelo YOLO.
RF 12	Desarrollo de API	La realización de esta API debe facilitar al usuario con sus funcionamientos y tanto que debe estar bien la relación entre el frontend y el backend del sistema.
RF 13	Acceso a Reportes	Autorizar al usuario la visualización de los informes históricos de lo que sucede dentro de la biblioteca.
RF 14	Monitoreo de Aforo	Concede los resultados en tiempo real del número de personas dentro para poder tener el control tanto de la cantidad de personas que debe estar dentro.
RF 15	Gestión de Base de Datos	Administra los datos de la base para el almacenamiento de datos del monitorio de personas.
RF 16	Optimización del Proceso de Captura de Imágenes	Incorporar técnicas de preprocesamiento de imágenes, como reducción de ruido y ajuste de brillo, para mejorar la calidad de las imágenes y asegurar una detección precisa en diversas condiciones de iluminación.
RF 17	Almacenamiento de Datos Históricos	Registrar y almacenar los datos sobre el conteo de personas y su tiempo de permanencia en la biblioteca, permitiendo un fácil acceso a la información histórica para análisis posteriores.

RF 18	Visualización de Uso de la biblioteca	Se muestran gráficos y estadísticas sobre el uso facilitando la identificación de patrones de afluencia.
RF 19	Entrenamiento del Modelo	El sistema debe permitir la utilización de un conjunto de datos etiquetados de imágenes de personas.
RF 20	Acceso a Datos en Tiempo Real	Permitir a los administradores acceder a datos en tiempo real sobre el conteo de personas y su permanencia.
RF 21	Capacitación de Usuarios	Proporcionar recursos y documentación para capacitar a los usuarios en el uso del sistema.
RF 22	Optimización de Rendimiento	El sistema debe ser capaz de procesar los videos y realizar el conteo de personas sin retrasos significativos, asegurando un rendimiento eficiente incluso con una gran cantidad de usuarios.
RF 23	Interfaz de Usuario	La interfaz debe ser flexible y adaptable para garantizar una experiencia de usuario uniforme y coherente.
RF24	Actualización en Tiempo Real del Conteo	El sistema debe actualizar el número de personas presentes en la biblioteca en tiempo real, reflejando de inmediato cualquier cambio en el aforo.
RF 25	Historial de Acceso y Permanencia	El sistema debe registrar un historial detallado de acceso y permanencia de las personas en la biblioteca, permitiendo a los administradores consultar datos históricos para análisis de uso.

Tabla 2: Requerimientos Funcionales.

2.2.2 Requerimientos no Funcionales

N.º R.NF	REQUERIMIENTOS	DETALLES
RNF 1	Rendimiento	El sistema debe ser capaz de procesar el flujo de video y realizar la detección y conteo de personas con una latencia mínima, asegurando una respuesta en tiempo real.
RNF 2	Escalabilidad	Este sistema debe dejar acceder a la expansión en el caso de la utilización de otros espacios dentro de la universidad sin que se haga tanta modificación o rediseñar en tal caso el sistema.
RNF 3	Optimización de Recursos Computacionales	Tiene la obligación de ser capaz de la minimización tanto del CPU y la memoria, hasta cuando ocurra una alta carga dentro del entorno para que se puede ver que el rendimiento es aceptable.
RNF 4	Usabilidad	Conceder al usuario que interactúa con el programa sin que se le exija conocimientos avanzados.
RNF 5	Mantenibilidad	El código del sistema debe estar bien documentado y estructurado para facilitar futuras actualizaciones y mantenimiento.

RNF 6	Compatibilidad	Tiene la misión de emplearse a ser compatible tanto de los diferentes tipos de navegadores web para que el usuario se sienta seguridad a utilizar diferentes plataformas.
RNF 7	Fiabilidad	Su funcionalidad debe ser estable hasta incluso continua así poder garantizar el buen servicio de disponibilidad.

Tabla 3: Requerimientos no Funcionales.

2.3. Componente de la Propuesta

2.3.1. Arquitectura del Sistema

La arquitectura del sistema va a permitir en tiempo real la visualización de datos tanto como el procesamiento, captura y almacenamiento. Además, se va a describir los componentes fundamentales de la arquitectura que vendrían ser los siguiente:

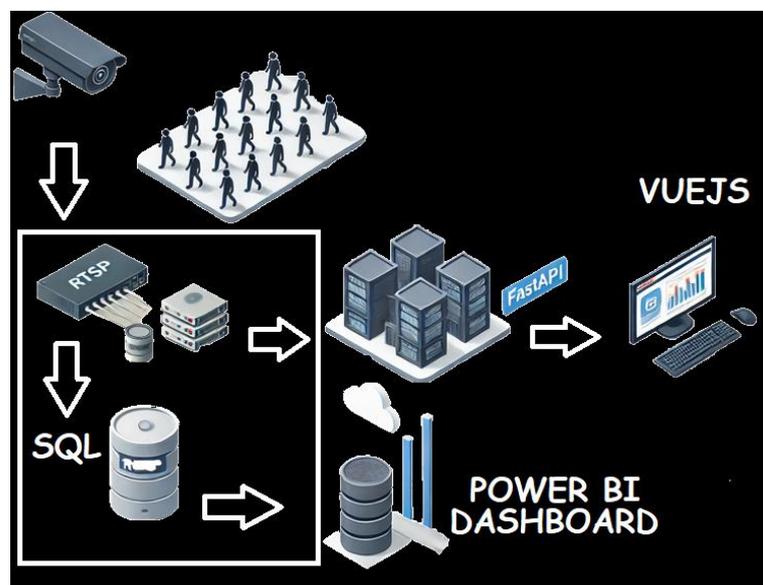


Figura 3: Diagrama general

Desde un punto de vista funcional, la estructura del sistema de seguimiento y recuento de personas se fundamenta en un modelo Cliente-Servidor que facilita una interacción eficaz entre los usuarios y el sistema. Esta estructura debe conseguir los datos en tiempo real, así mejorando tanto como la gestión dentro de la biblioteca.

	FUNCIONES
CLIENTE- FRONTEND	Aquí se va a representar como se refleja el conteo de los individuos para que proceda a dar información tanto como actualizada de la multitud dentro de la biblioteca y la procreación de informes de la afluencia de aquello para la toma de decisiones.
SERVIDOR - BACKEND	Se lleva a cado el FastAPI que admite y procesa la petición de usuario. La utilización de visión por computadora que es YOLO y DeepSot, localizando las personas con la captura de la cámara de vigilancia. Produce respuesta actualizadas para el cliente e informes. Acumulando el reporte de datos y la duración dentro de la biblioteca en el sistema así mejorando la eficiencia.

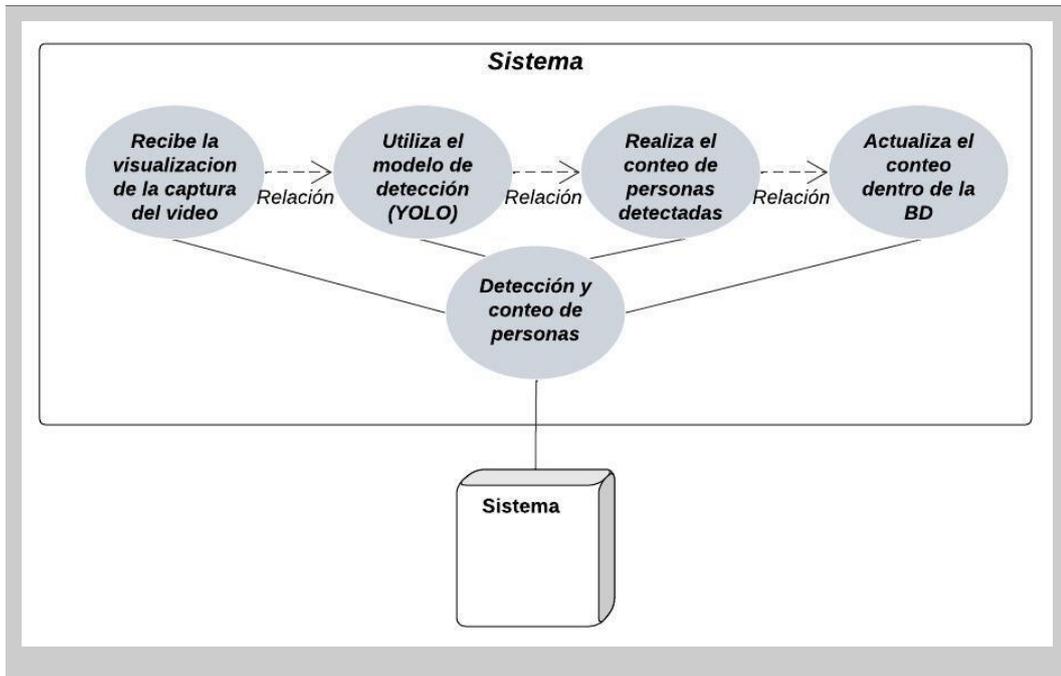
Tabla 4: Frontend - Backend

2.3.2. Diagramas de casos de uso

Los diagramas de casos de uso son herramientas esenciales para representar las interacciones entre los actores y el sistema. A continuación, se describen los casos de uso para el sistema de monitoreo y conteo de personas:

Caso de Uso: Detección y Conteo de Personas

Actor	Sistema de Procesamiento
Función	El sistema recibe la captura del video de las cámaras para luego ser procesadas las imágenes y detectar de manera correcta el número de personas dentro de la biblioteca.



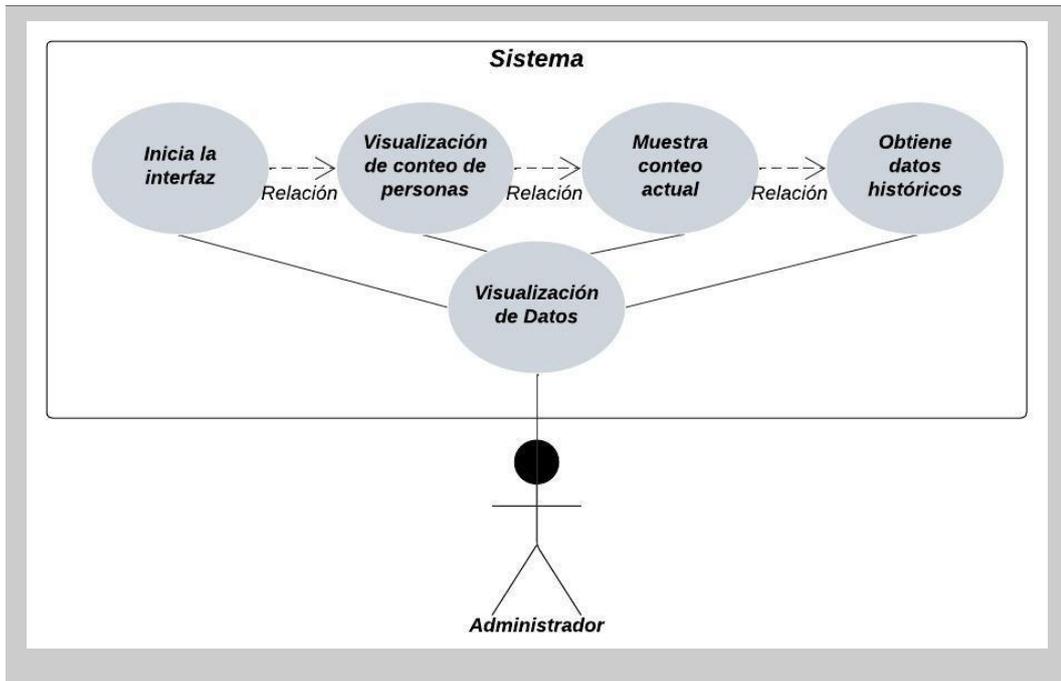
Pasos realizados

1. El sistema recibe el flujo de video.
2. Se aplica el modelo de detección (YOLO).
3. Se cuenta el número de personas detectadas.
4. Se actualiza el conteo en la base de datos.

Tabla 5: Caso de Uso: Detección y Conteo de Personas.

Caso de Uso: Visualización de Datos

Actor	Administrador
Función	Permite al administrador a que acceda a la interfaz del usuario para poder visualizar el conteo de personas en tiempo real.



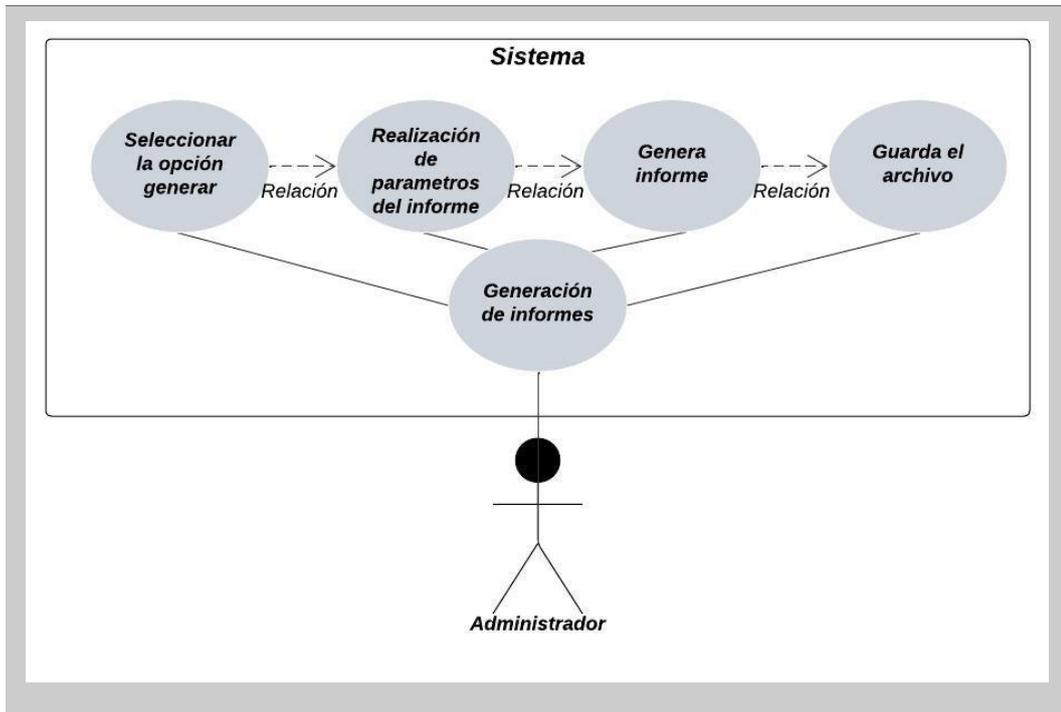
Pasos realizados

1. El administrador inicia sesión en la interfaz.
2. Se solicita la visualización del conteo de personas.
3. El sistema muestra el conteo actual.
4. Obtiene datos históricos.

Tabla 6: Caso de Uso: Visualización de Datos.

Caso de Uso: Generación de Informes

Actor	Administrador
Función	Dar acceso al administrador para generar informes sobre la cantidad de personas en la biblioteca.



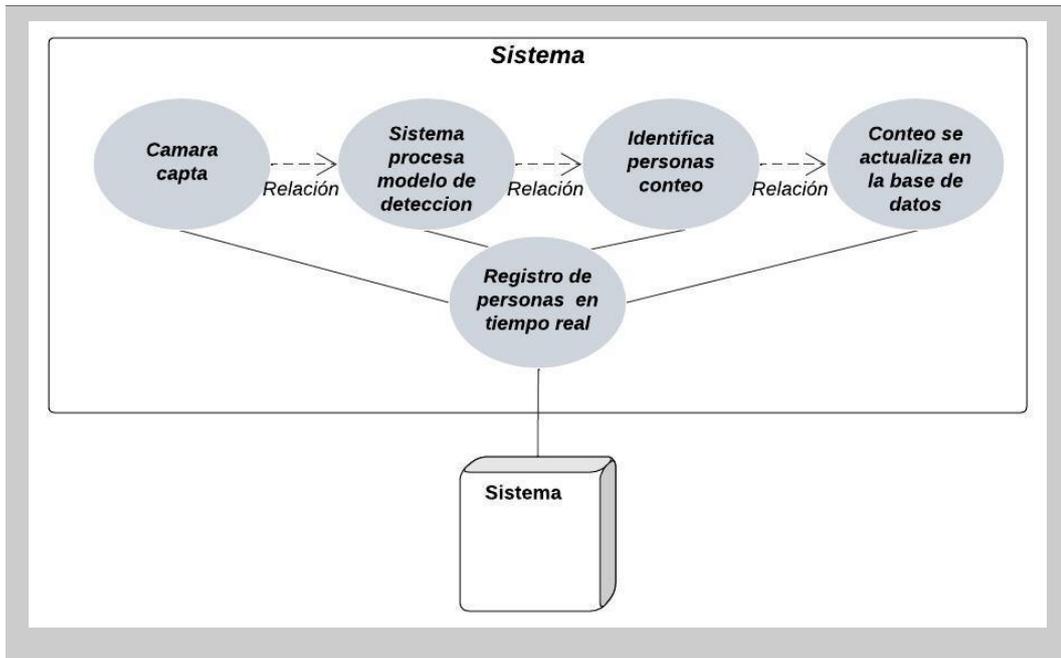
Pasos realizados

1. El administrador selecciona la opción de generar informes.
2. En el archivo especifica los parámetros del informe (fecha de entrada, hora, tiempo total y hora de salida).
3. El sistema genera el informe.
4. Guarda el archivo.

Tabla 7: Caso de Uso: Generación de Informes.

Caso de Uso: Registro de Personas en Tiempo Real

Actor	Sistema de visión por computadora
Función	El sistema detecta y registra automáticamente a cada persona que ingresa a la biblioteca utilizando cámaras de vigilancia y algoritmos de visión por computadora.



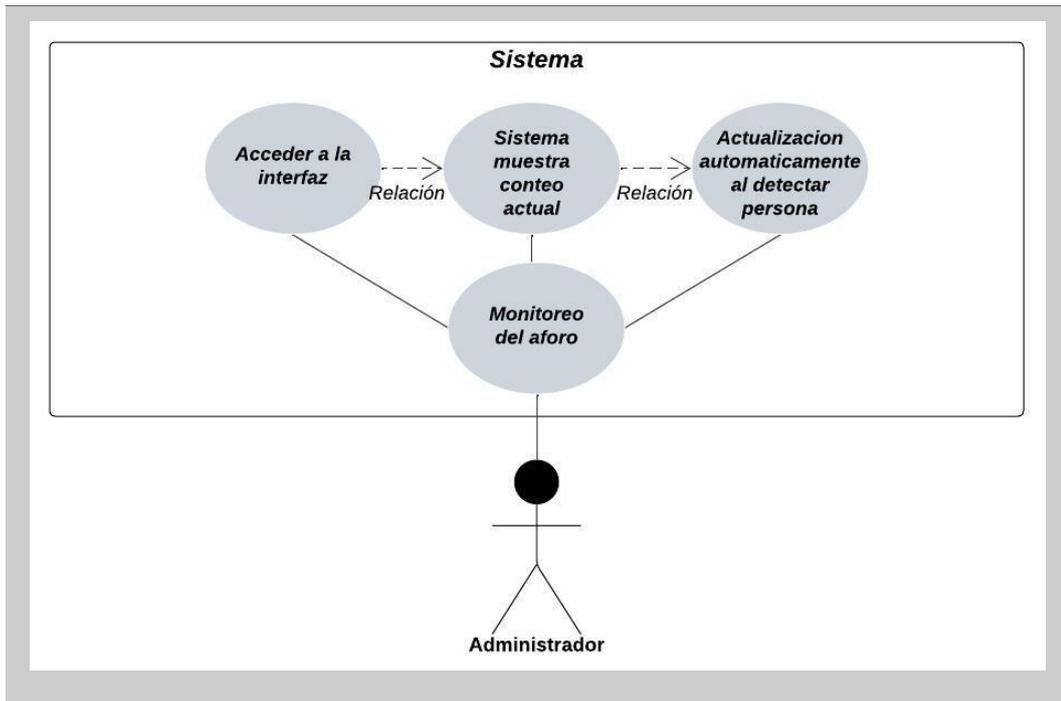
Pasos realizados

1. La cámara capta el video en tiempo real.
2. El sistema procesa el video utilizando el modelo de detección (YOLO).
3. Se identifica y cuenta a cada persona que entra.
4. El conteo se actualiza en la base de datos después de 1 minuto de ser detectada en pantalla en tiempo real.

Tabla 8: Caso de Uso: Registro de Personas en Tiempo Real.

Caso de Uso: Monitoreo del Aforo

Actor	Administrador de la biblioteca
Función	El administrador puede visualizar en tiempo real el número de personas que se encuentran en la biblioteca.



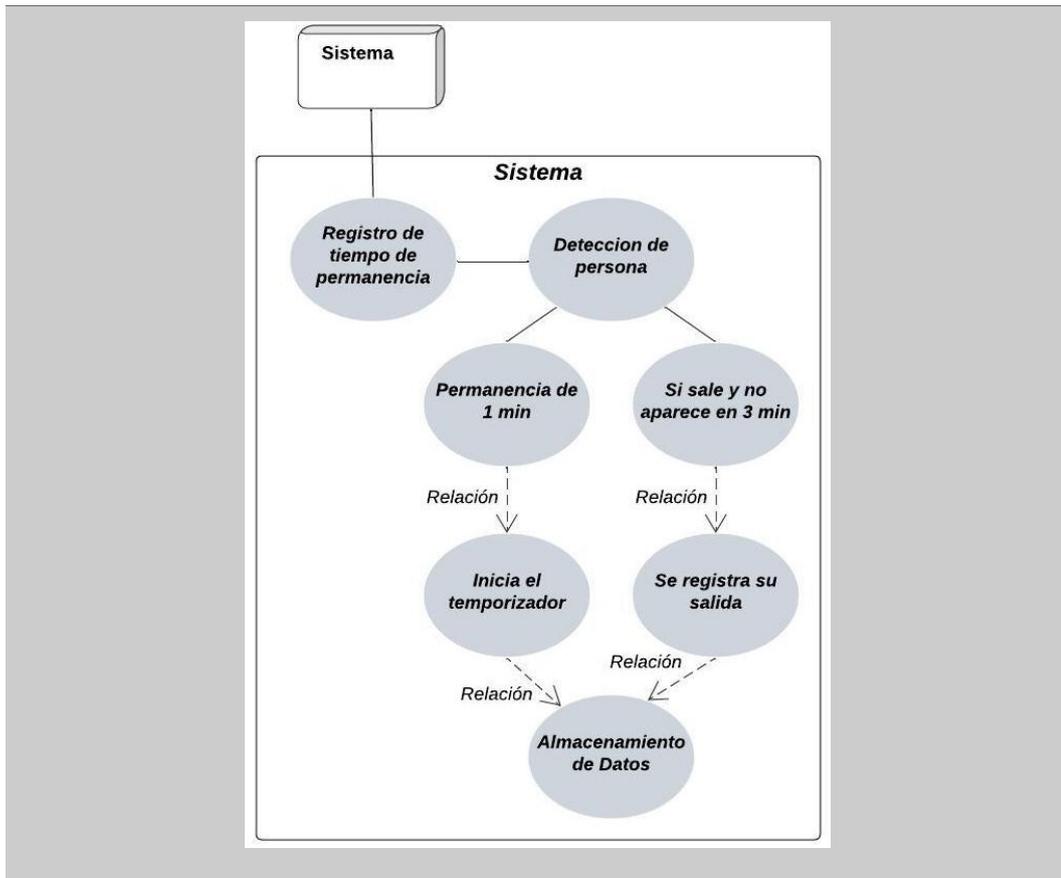
Pasos realizados

1. El administrador accede a la interfaz del sistema.
2. El sistema muestra el conteo actual de personas en la biblioteca.
3. Se actualiza automáticamente cada vez que se detecta un nuevo ingreso o salida.

Tabla 9: Caso de Uso: Monitoreo del Aforo.

Caso de Uso: Registro de Tiempo de Permanencia

Actor	Sistema de Monitoreo
Función	Registrar el tiempo de permanencia de cada persona en la biblioteca.



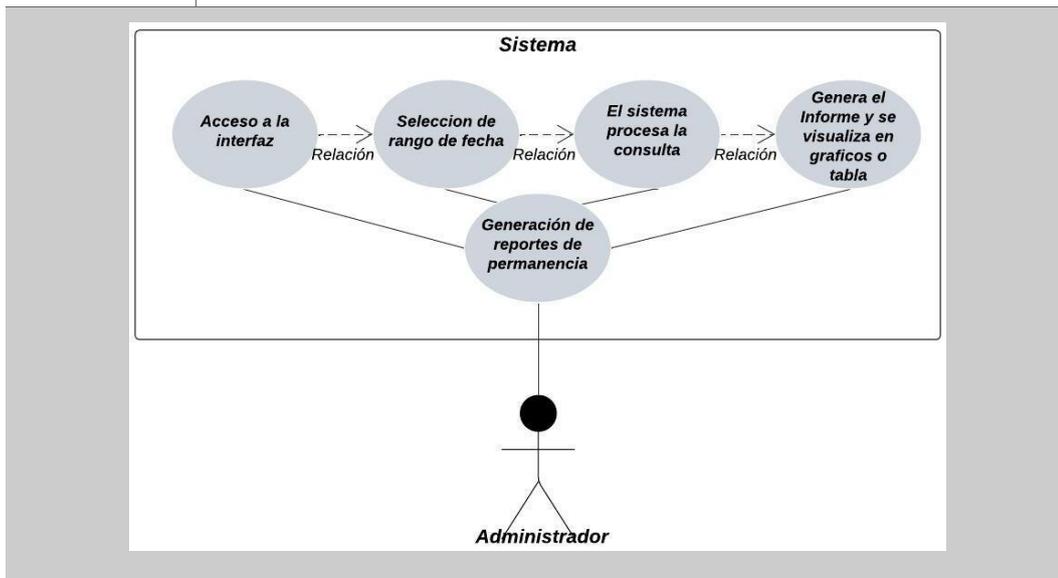
Pasos realizados

1. El sistema detecta a una persona y guarda su ID en la base de datos.
2. Si la persona permanece en la pantalla durante más de 1 minuto, se registra su tiempo de entrada.
3. Si la persona desaparece de la pantalla, se inicia un temporizador.
4. Si la persona no vuelve a aparecer en un plazo de 3 minutos, se registra su tiempo de salida y se calcula el tiempo total de permanencia.
5. Los datos se almacenan en la base de datos para futuras consultas y generación de informes.

Tabla 10: Caso de Uso: Registro de Tiempo de Permanencia.

Caso de Uso: Generación de Reportes de Permanencia

Actor	Administrador
Función	Generar reportes sobre el tiempo de permanencia de los usuarios en la biblioteca.



Pasos realizados

1. El administrador accede a la interfaz de generación de reportes.
2. Selecciona el rango de fechas para el cual desea obtener el reporte.
3. El sistema consulta la base de datos y extrae los datos de tiempo de permanencia de los usuarios.
4. Se genera un informe que incluye:
 - ID de la persona
 - Hora de entrada
 - Hora de salida
 - Tiempo total de permanencia
5. El informe se presenta en un formato visual (gráficos, tablas)

Tabla 11: Caso de Uso: Generación de Reportes de Permanencia.

2.3.3 Modelado de Datos

El análisis del modelo de datos se distribuye en módulos para poder diferenciar las partes del sistema que realmente funcionan. El registro de visitas recoge los datos de cada persona, mientras que el encabezado de visitas agrupa los datos del día.

Relación:

La tabla **registro visitas** se relaciona con la tabla **encabezado visitas** a través de una relación uno a muchos, donde un encabezado de visita puede tener múltiples registros de visitas.

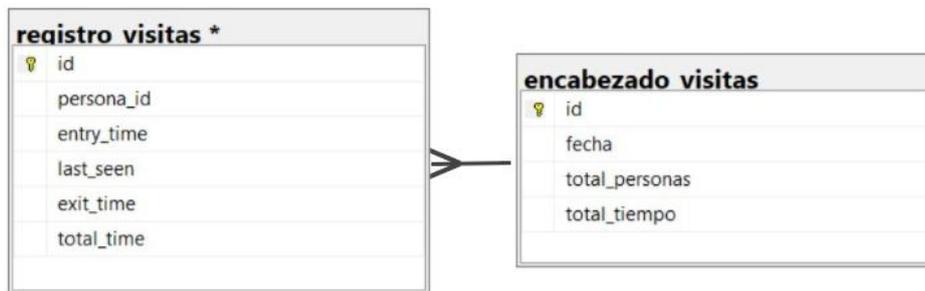


Figura 4: Modelado - Base de datos.

Para el presente estudio del proyecto, la estructura del modelo de datos que sirve como raíz de almacenamiento de las consultas para el aplicativo está desarrollado en base de datos relacionales. La información almacenada en estas tablas se puede consultar e interactuar con ella a través de un sistema de gestión de bases de datos.

2.3.4. Evaluación y Selección del Modelo de Detección

2.3.4.1 Comparativa de Resultados

Para comparativa de los modelos de detección se aplicó un conjunto de datos que contiene imágenes, con el tipo de etiqueta que es este caso son las personas. Los 5 modelos se entrenaron con el mismo conjunto de datos para controlar las variables. La relación entre el conjunto de entrenamiento y el conjunto de validación.

A continuación, se presenta una tabla con los resultados de cada modelo y se indicara el proceso de cada uno de ellos:

Modelo	Precisión	Exhaustividad	F1-Score	mAP	FPS
YOLOv8	0.94	0.88	0.84	0.90	30
Faster R-CNN	0.85	0.88	0.90	0.87	5
SSD	0.77	0.83	0.85	0.81	22
RetinaNet	0.82	0.87	0.88	0.84	12

Tabla 12: Comparativa de resultados modelos.

1) Proceso de entrenamiento YOLOv8

Acerca de YOLOv8, se proporcionan principalmente cuatro tipos de pesos de preentrenamiento (YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large) y YOLOv8x (extra large)). Debido a que la mayoría de las imágenes en el conjunto de datos para entrenar el modelo contenían objetos pequeños, elegimos específicamente YOLOv8m porque podría ayudar al modelo a reconocer el objeto pequeño con mayor precisión.

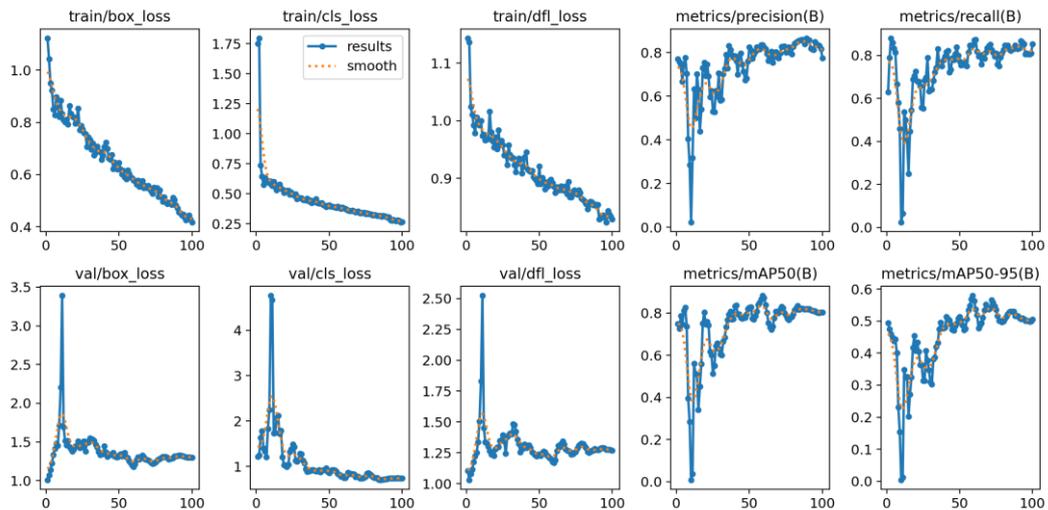


Figura 5: Gráficas de entrenamiento y validación YOLOv8.

En la figura podemos visualizar una serie de gráficas de entrenamiento y validación de un modelo de detección de objetos. Estas gráficas presentan diferentes métricas y pérdidas a lo largo de las épocas. Se guardo el modelo cada 5 épocas. El modelo se entrenó 100 épocas, incluidas 50 épocas de congelación. La pérdida se redujo a alrededor de 0,068 y casi dejó de caer. No es seguro que la pérdida haya alcanzado el mínimo global.

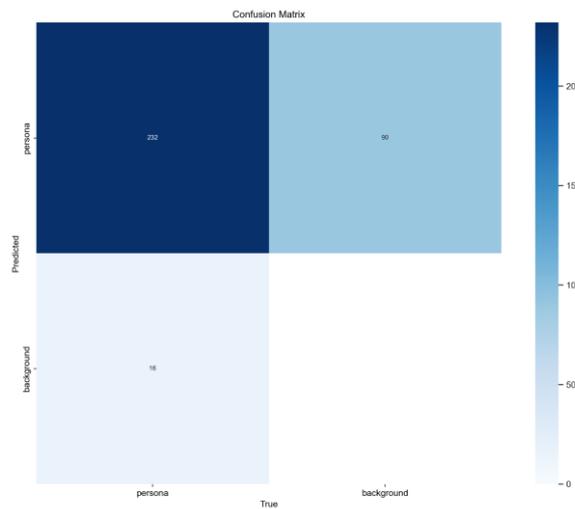


Figura 6: Matriz de confusión YOLOv8.

Con el resultado de la matriz se puede representar el rendimiento de aquella clasificación del modelo en este punto que es la detección. Un balance tanto de predicciones para contribuir un mejor rendimiento.



Figura 7: Ejemplo de detección de YOLOv8

2) Proceso de entrenamiento Faster R-CNN

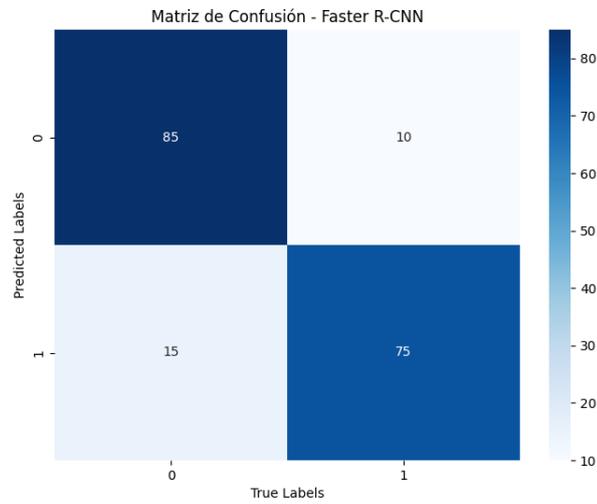


Figura 8: Matriz de confusión Faster R-CNN.

En cuanto a la configuración de parámetros del modelo antes del entrenamiento, configuramos la red de extracción de características de la red troncal como resnet50. Guardamos el modelo cada 5 épocas. El modelo se entrenó durante 100 épocas, incluidas 50 épocas de congelamiento. La pérdida se redujo a alrededor de 0,715 y casi dejó de caer. No es seguro que la pérdida haya alcanzado el mínimo global.



Figura 9: Ejemplo de detección de Faster R-CNN.

3) Proceso de Entrenamiento SSD

En este caso se prioriza el trabajo en la eficiencia que procesa el tiempo de respuesta, así mismo desempeñando en tiempo real.

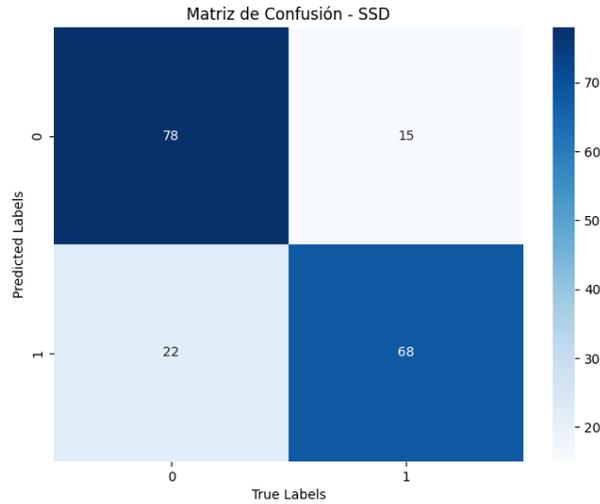


Figura 10: Matriz de confusión SSD.

Tiene sus desventajas que minimiza la precisión a comparación de otros. Entrenado con 100 épocas, donde maximiza la clasificación de personas con tamaño medio a grande. Tomando que YOLOv8, se adapta a objeto pequeños.

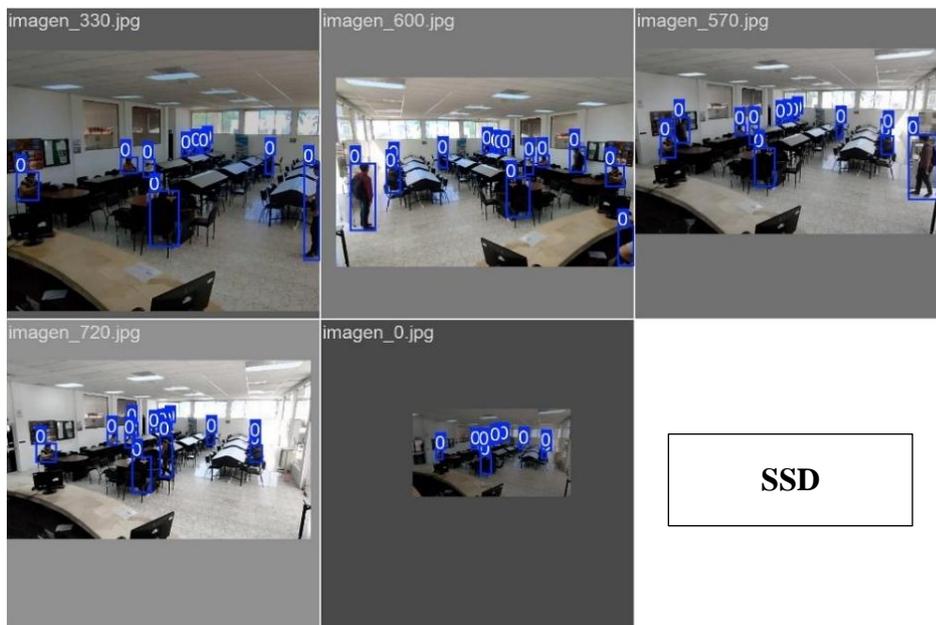


Figura 11: Ejemplo de detección de SSD.

La pérdida del modelo convergió a un valor de 0,124. En cuanto a la velocidad de procesamiento, el modelo alcanzó 22 FPS, lo que indica un rendimiento razonable para aplicaciones en tiempo real, aunque con una precisión inferior en comparación con otros modelos.

4) Proceso de Entrenamiento RetinaNet

El RetinaNet se distingue por utilizar la técnica de focal loss, que está diseñada para abordar el problema del desequilibrio de clases en tareas de detección. Este modelo fue especialmente efectivo para conjuntos de datos con categorías que contienen pocas muestras.

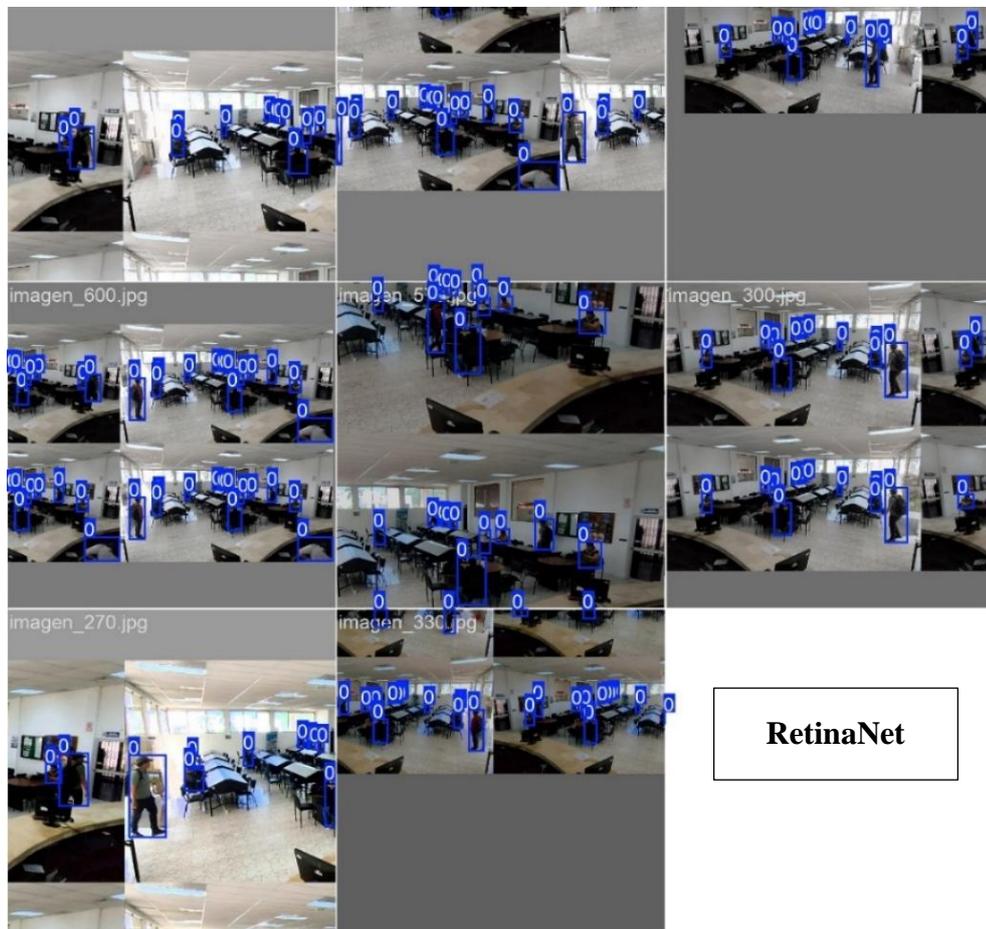


Figura 12: Ejemplo de detección de RetinaNet.

El modelo RetinaNet fue entrenado durante 100 épocas y alcanzó una pérdida final de aproximadamente 0,098. Aunque su velocidad de procesamiento fue de 12 FPS,

que es menor en comparación con modelos como SSD y YOLOv8, mantiene un buen equilibrio entre precisión y exhaustividad.

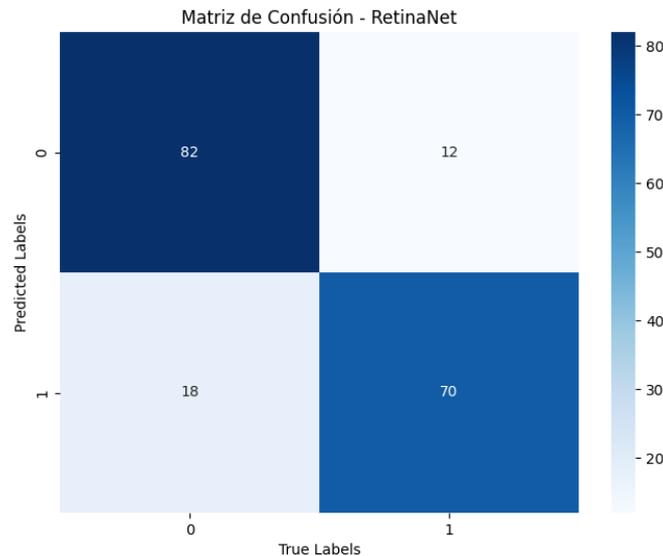


Figura 13: Matriz de confusión RetinaNet.

Para evaluar el rendimiento de cada modelo, se emplearon varias métricas, como precisión, exhaustividad y F1-Score. Como se detalla en la **Tabla 12: Comparativa de resultados modelos.**, cada uno presenta características distintas en términos de velocidad y precisión. YOLOv8, por ejemplo, se destaca en entornos de alta velocidad, con una precisión notable al detectar objetos pequeños, mientras que Faster R-CNN proporciona un análisis más detallado, aunque no es tan adecuado para tareas en tiempo real debido a su menor FPS.

Análisis y Selección del Modelo

Basándonos en los resultados obtenidos, podemos observar que:

1. YOLO (YOLOv8) demuestra un rendimiento equilibrado con una alta precisión, exhaustividad y F1-Score, comparable a los modelos más precisos como Faster R-CNN.
2. En términos de velocidad de procesamiento (FPS), YOLO supera significativamente a los otros modelos, alcanzando 30 FPS, lo cual es crucial para nuestro sistema de conteo en tiempo real.

3. Faster R-CNN muestra la precisión más alta, pero su velocidad de procesamiento es extremadamente baja (5 FPS), lo que lo hace inadecuado para aplicaciones en tiempo real.
4. SSD ofrece una velocidad de procesamiento de aproximadamente 22 FPS, pero sus métricas de precisión son las más bajas entre los modelos comparados.
5. RetinaNet presenta un buen equilibrio entre precisión y velocidad, pero aún queda por detrás de YOLO en ambos aspectos.

Considerando estos factores, se escogió YOLOv8, como modelo que se utilizara para el conteo de personas en la biblioteca de la UPSE. Las principales razones de esta selección son:

1. Alto rendimiento: Por la precisión y exhaustividad de manera óptima, alcanzando un F1-Score de 0.90, lo que asegura detecciones confiables.
2. Velocidad superior: Con una capacidad de 30 FPS, YOLO puede procesar video en tiempo real, una característica indispensable para el monitoreo continuo del sistema.
3. Eficiencia computacional: En comparación con modelos como Faster R-CNN, YOLO requiere menos recursos, lo que lo convierte en una opción ideal para sistemas con limitaciones de hardware.
4. Versatilidad: YOLO ha demostrado un desempeño sólido bajo diversas condiciones de iluminación y con diferentes ángulos de cámara, un aspecto clave para su implementación en el entorno dinámico de una biblioteca.

2.4. Diseño de Interfaces

El diseño de las interfaces se desarrolló siguiendo principios de usabilidad y simplicidad, con el objetivo de garantizar una experiencia de usuario fluida y accesible. A continuación, se detallan las pantallas principales del sistema, enfocadas en la detección y el registro de visitas a la biblioteca en tiempo real:

2.4.1. Pantalla de Afluencia de Personas

La primera pantalla muestra una interfaz sencilla. En la parte de abajo, el sistema presenta un menú de navegación que permite al usuario acceder fácilmente a las secciones de "Aforo", "Por fechas", "Diario", "Anual", "Días con más visitas". En el centro de la pantalla se incluye la fecha actual y el número total de personas detectadas, así como el porcentaje de ocupación de la biblioteca y el detalla del día.



Figura 14: Pantalla de Afluencia de Personas.

2.4.2. Pantalla de Informes de Visitas

Esta pantalla ofrece un informe detallado de las visitas registradas en la biblioteca. La interfaz incluye un filtro por fecha que permite seleccionar un rango de días para observar las visitas correspondientes.

Los datos mostrados comprenden la hora de entrada y salida, así como el tiempo total de permanencia en la biblioteca, lo que facilita el análisis del flujo de usuarios. La tabla de registros es generada automáticamente, permitiendo visualizar múltiples entradas por fecha.

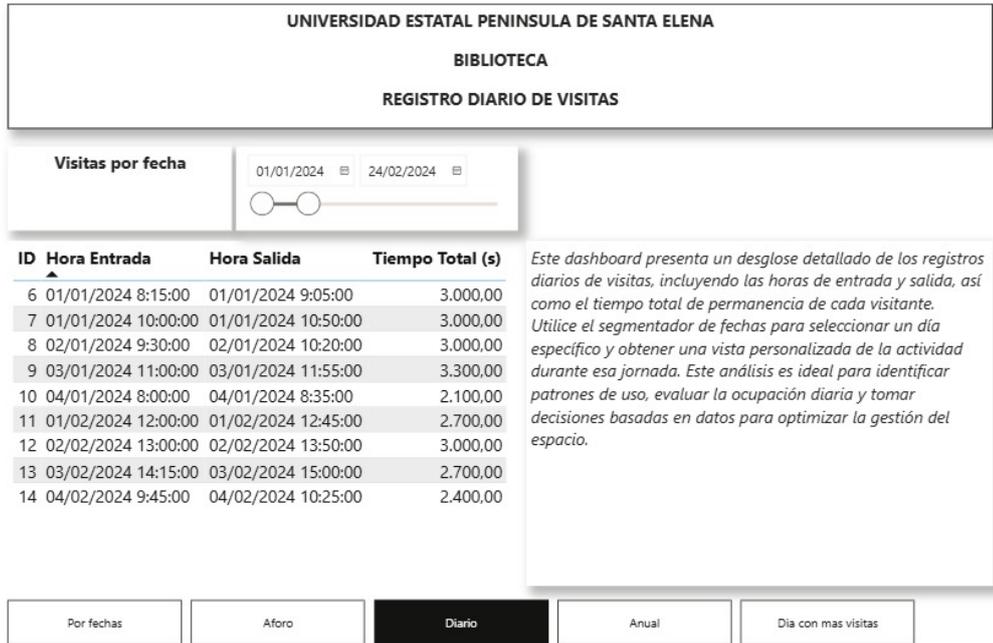


Figura 15: Pantalla de Registro Diario de Visitas.

2.4.3. Pantalla de Informe de Visitas por Fecha

La visualización de la pantalla de informes de vista cuenta con un diagrama con el rango de fecha que puede seleccionar el usuario.



Figura 16: Pantalla de Informe de Visitas por Fecha.

2.4.4. Pantalla de Gráfico Anual de Visitas

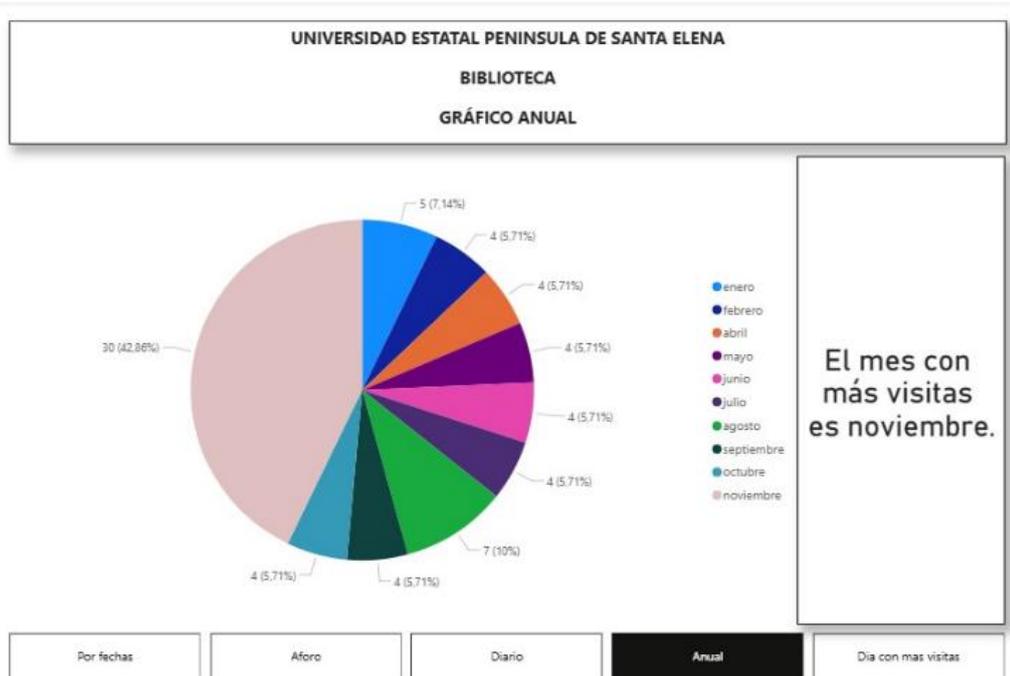


Figura 17: Pantalla Gráfico Anual de Visitas.

Lo que proyecta esta pantalla es un gráfico en forma de pastel con sus porcentajes de cada mes para facilitar que mes utilizan más los estudiantes la biblioteca.

2.4.5. Pantalla de Visitas del Mes: Por Hora de Entrada y Salida

Cuenta con lo siguiente según la hora de entrada y salida. También tiene elementos claves:

- **Tabla de datos:** Presenta detalles de cada visita, incluyendo ID, hora de entrada, hora de salida y el tiempo total de permanencia en segundos.
- **Resumen numérico:** Muestra el total de personas que ingresaron en el rango de tiempo seleccionado.
- **Gráfico:** Un gráfico circular ilustra la distribución de la duración de las visitas en diferentes intervalos de tiempo, facilitando el análisis visual.

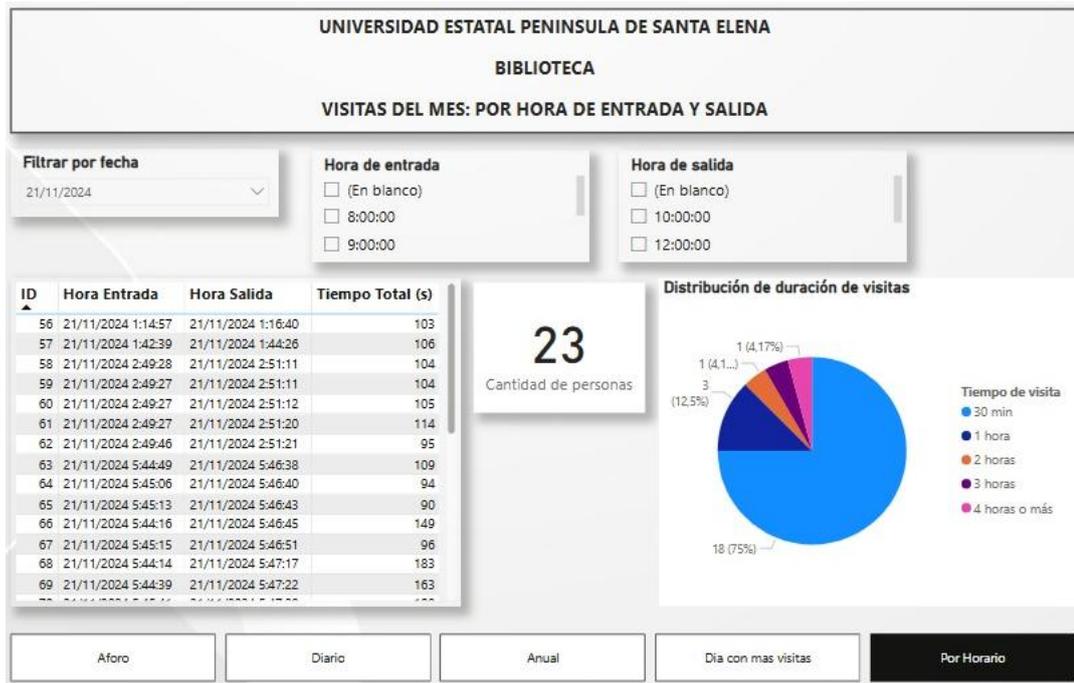


Figura 18: Pantalla de Visitas del Mes

Funcionalidades Clave:

- **Selección de rango de fechas:** Permite al usuario definir un rango personalizado para consultar las estadísticas de visitas.
- **Interactividad con gráficos:** Los gráficos brindan una herramienta visual intuitiva para un análisis rápido, facilitando la toma de decisiones informadas sobre la gestión de la biblioteca.
- **Enlaces a informes:** Proporciona acceso directo a reportes generados con Power BI, mejorando la integración con herramientas avanzadas de análisis de datos.

2.5. Pruebas

2.5.1 Entrenamiento de Modelo

El modelo fue entrenado utilizando una unidad de procesamiento gráfico (GPU) y la versión 8.0 de YOLOv8m, que es una arquitectura avanzada desarrollada por Ultralytics. El proceso de entrenamiento abarcó un total de 100 épocas, donde cada

época representa una iteración completa sobre el conjunto de datos de entrenamiento. En cada iteración, los datos se dividieron en pequeños lotes o "batches", los cuales fueron procesados por el modelo para calcular el error en las predicciones. Posteriormente, se ajustaron los parámetros del modelo con el objetivo de minimizar dicho error.

Es importante destacar que el número de épocas debe ser cuidadosamente configurado. Un mayor número de iteraciones puede mejorar el rendimiento del modelo; sin embargo, existe el riesgo de sobreajuste, una situación en la que el modelo se adapta excesivamente a los datos de entrenamiento, perdiendo su capacidad de generalizar en nuevos conjuntos de datos.

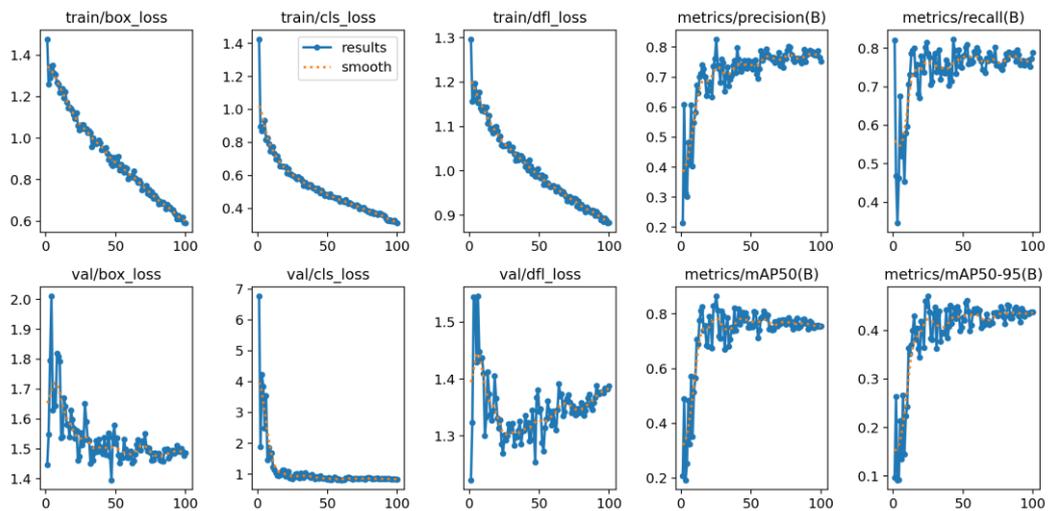


Figura 19: Resultados del entrenamiento por época.

Durante este entrenamiento, se realizó un seguimiento continuo de los parámetros de precisión, recall y mAP (Mean Average Precision) en cada época, con el objetivo de evaluar el rendimiento del modelo. A continuación, se presentan los resultados obtenidos a lo largo de 100 épocas, resaltando la evolución de métricas clave, como la precisión y la pérdida (loss), tanto en el conjunto de datos de entrenamiento como en el de validación.

Como se puede visualizar el proceso entrenamiento en **Anexo 2: Captura de resultados de entrenamiento.**

2.5.2 Prueba de Detección en Tiempo Real

Como parte del proceso de validación del modelo, se llevó a cabo su implementación en un entorno de detección en tiempo real. En este contexto, el sistema se encargó de identificar y contar la cantidad de personas presentes dentro de un intervalo temporal específico. La imagen presentada a continuación ilustra los resultados obtenidos durante una prueba realizada el 29 de noviembre de 2024, en la cual se contabilizaron un total de 47 personas.



Figura 20: Prueba de la detección en tiempo real.

Esta pantalla también permite observar el porcentaje de ocupación con respecto al aforo máximo definido (80 personas), alcanzando un 0% de ocupación en el momento de la captura, demostrando la capacidad del sistema para adaptarse a escenarios con alta variabilidad en el flujo de personas.

2.6. Resultados

2.6.1 Resultados del entrenamiento por época

A continuación, se muestra un resumen de los valores obtenidos durante el proceso:

Época	Pérdida (box)	Pérdida (cls)	Pérdida (dff)	Precisió n	Recall	mAP50	mAP50-95
1	1.504	1.521	1.303	0.578	0.637	0.529	0.25
2	1.406	1.091	1.245	0.455	0.54	0.449	0.182
3	1.428	1.051	1.249	0.0228	0.387	0.0154	0.00584
4	1.482	1.043	1.264	0.0271	0.00895	0.00134	0.000463
5	1.474	1.006	1.244	0.295	0.297	0.203	0.0862
6	1.368	0.9394	1.212	0.279	0.269	0.179	0.0928
7	1.365	0.8831	1.214	0.427	0.439	0.368	0.169
...
44	0.8499	0.4465	0.9776	0.732	0.762	0.744	0.418
45	0.8139	0.4319	0.9841	0.7334	0.7813	0.763	0.425
46	0.8208	0.4322	0.9812	0.745	0.763	0.754	0.423
47	0.8059	0.4192	0.972	0.731	0.758	0.744	0.421
48	0.8007	0.4172	0.9796	0.76	0.752	0.754	0.427
49	0.7897	0.4158	0.9711	0.744	0.746	0.756	0.432
50	0.7674	0.4037	0.9612	0.748	0.746	0.745	0.426

Tabla 13: Resultados del entrenamiento por época

2.6.2 Entrenamiento con 50 Epochs

Se analizaron los resultados obtenidos como se muestra en la **Figura 21: Gráfico Precision-Recall para 50 Epochs.**, que permite la evaluación del modelo tanto de la capacidad y principalmente la precisión de localizar los objetos. Evidenciando el buen desempeño que se aproxima a 1 tanto de los dos ejes. Adaptado bien la precisión y sensibilidad.

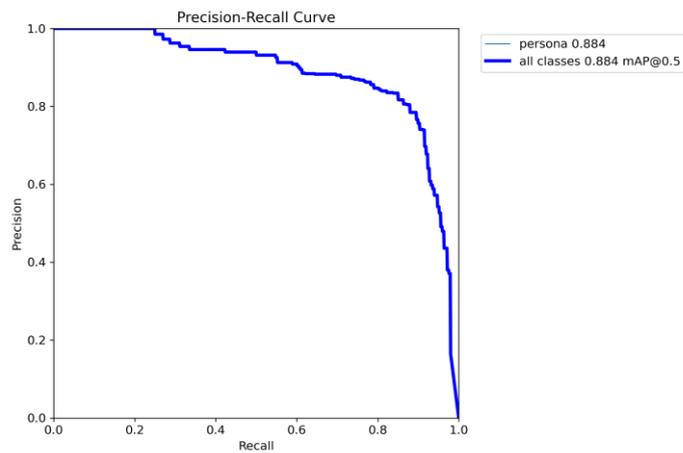


Figura 21: Gráfico Precision-Recall para 50 Epochs.

Por otro lado, en la **Figura 22: Gráfico F1-Confianza para 50 Epochs.**, se presenta el F1-score con los niveles de confianza. Aunque exactamente no llega al ideal absoluto, se contemplan valores suficientes para respaldar la predicción.

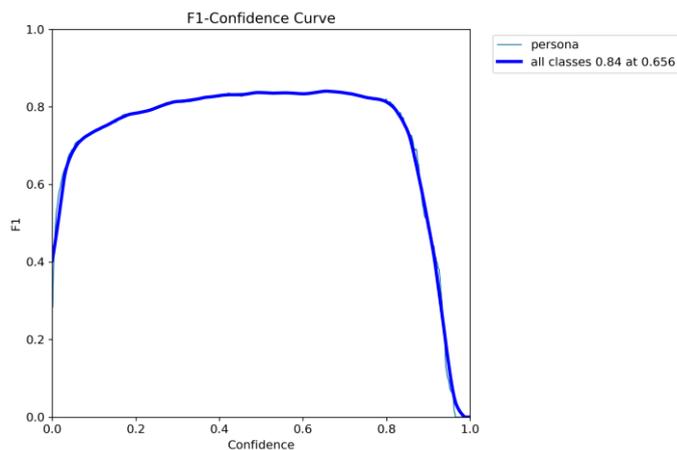


Figura 22: Gráfico F1-Confianza para 50 Epochs.

Incluso en la **Figura 23: Matriz de confusión para 50 Epochs.** se observa la probabilidad de predicción de clases reales. Que se asocia a altos valores como 0.8, dando en efecto una buena observación en el gráfico.

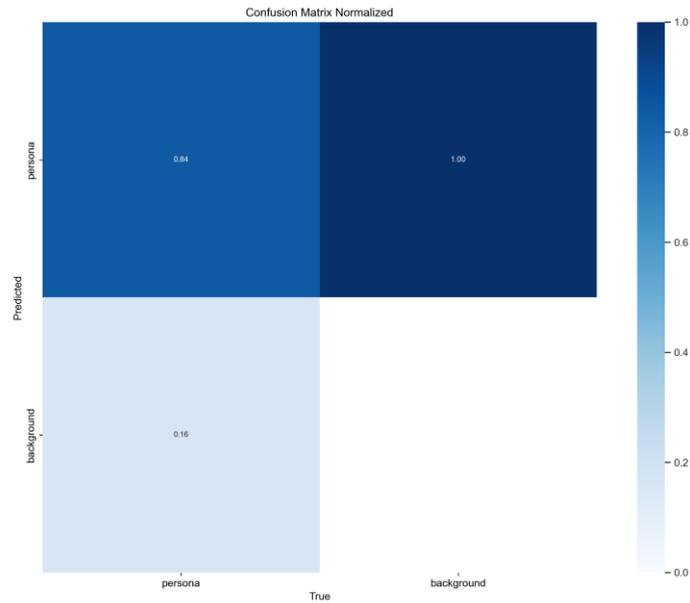


Figura 23: Matriz de confusión para 50 Epochs.

Finalmente, los gráficos de correlación que proporciona la relación entre dichas variables, mostrando patrones relevantes del desempeño del modelo.

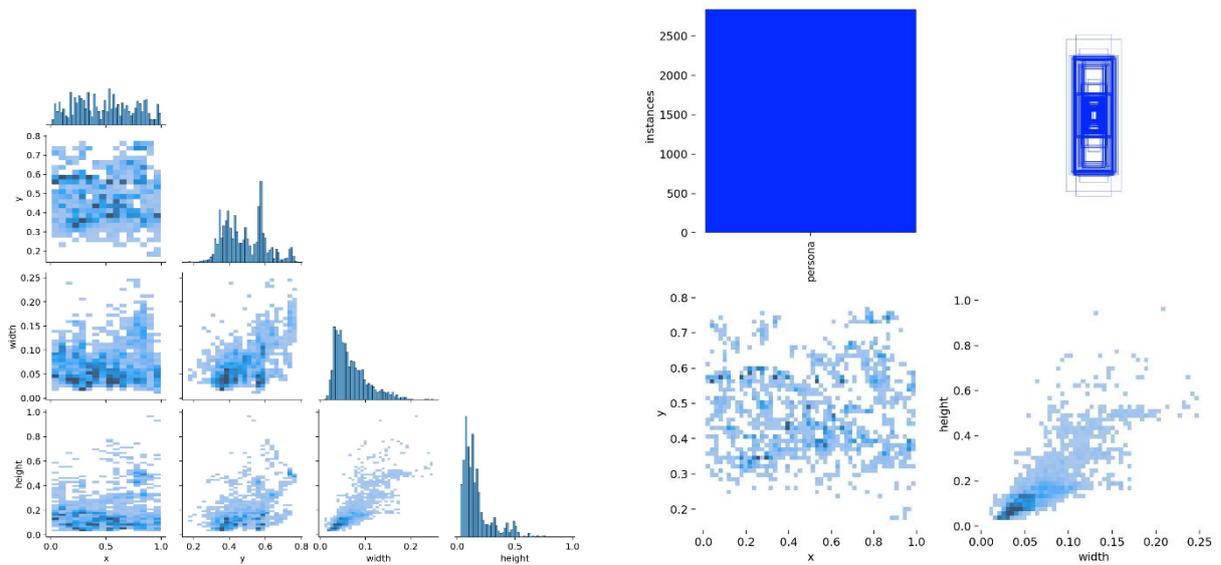


Figura 24: Correlogramas de etiquetas del modelo.

En la **Figura 25: Resultados de detección predicción.** se presenta las detecciones del modelo entrenado tras 50 epochs. Proyecta los objetos detectados, delimitados por cuadros con sus etiquetas hasta el nivel de confianza. Demuestra el rendimiento de los escenarios complejos, que se caracteriza por múltiples detecciones simultáneas en cada img.



Figura 25: Resultados de detección predicción.

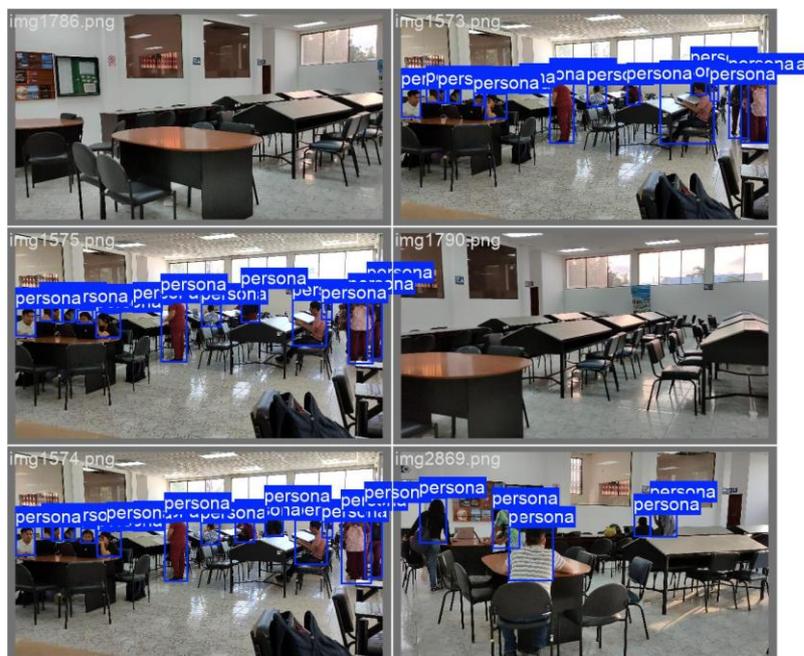


Figura 26: Resultados de Detección Etiqueta

2.6.3 Resultados de detección

Se presentan ejemplos visuales de las detecciones realizadas tras el entrenamiento del modelo. En las imágenes, se observan las etiquetas de las clases detectadas, los niveles de confianza, y las áreas delimitadas (bounding boxes) de los objetos identificados.



Figura 27: Resultados de detección 1.

Como se puede visualizar en la imagen muestra que mantiene la fiabilidad en la identificación de personas, mostrando las probabilidades asociadas a cada clase detecta.



Figura 28: Resultados de detección 2.

En la figura 28 representa el ejemplo de detección en un entorno controlado con múltiples personas presentes. La imagen refleja la precisión del modelo al identificar a cada individuo con un nivel de confianza asociado que varía entre 0.70 y 0.78.

2.7. Conclusiones

- La implementación de modelos de visión por computador como Yolov8 permite garantizar una alta precisión y velocidad en la detección de personas, adaptándose al monitoreo en tiempo real, esta tecnología permite operar de manera eficiente en ambientes dinámicos y con variabilidad de condiciones como es el caso de biblioteca.
- El empleo de librerías como PyTorch y DeepSort permite realizar la detección y el seguimiento individualizado de las personas. Estas tecnologías operan de manera conjunta para garantizar que cada individuo identificado sea monitoreado de forma precisa a lo largo de su trayectoria, minimizando posibles errores. Además, su integración asegura la escalabilidad del sistema, permitiendo manejar mayores volúmenes de datos de manera eficiente durante periodos de incremento en el número de usuarios.
- El sistema propuesto no solo hace enfoque al conteo y detección de personas, sino también al seguimiento de los usuarios basado en el tiempo de permanencia dentro de biblioteca, obteniendo la cantidad de tiempo que han usado las instalaciones para la evaluación del uso de la biblioteca.

2.8. Recomendaciones

- Para garantizar un monitoreo más preciso, se debería usar al menos tres cámaras con más de 30FPS y con alta resolución preferiblemente resolución 2k, esto nos asegurara captar más detalles en tiempo real.
- Para evitar pérdidas de FPS, interrupciones de señales y retraso en la transmisión del video en tiempo real, es fundamental hacer uso de conexiones cableadas. Esto nos asegura que las cámaras puedan operar de manera óptima y eficaz en periodos de alta demanda en la red, además mejora la estabilidad del sistema.
- Es fundamental ubicar las cámaras en lugares o puntos estratégicos, como entradas, salidas y zonas de mayor actividad dentro de biblioteca. De esta manera obtendremos una cobertura total del espacio y datos más precisos del flujo de personas, potenciando la efectividad del sistema.

Referencias

- [1] D. Forsyth y J. Ponce, *Computer Vision: A Modern Approach*, Pearson, 2012.
- [2] I. Goodfellow, Y. Bengio y A. Courville, *Deep Learning*, MIT Press., 2016.
- W. M. Smeulders, A. D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan y
- [3] M. Shah, «Visual Tracking: An Experimental Survey,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1442-1468, 2014.
- X. Liu, Y. Zhu y Y. Zhan, «Deep Learning-Based Object Detection and
- [4] Tracking: A Comprehensive Review,» *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, n° 7, pp. 2376-2394, 2020.
- [5] «Procesos Manuales Y Errores,» FasterCapital, 2024. [En línea]. Available: como se puede observar en la figura.
- J. P. G. Jordán, «Repositorio UVG,» 2021. [En línea]. Available:
- [6] <https://repositorio.uvg.edu.gt/static/flowpaper/template.html?path=/bitstream/handle/123456789/3933/TESIS%20JOSE%20PABLO%20GUERRA.pdf?sequence=1&isAllowed=y>.
- D. A. A. ESCOBAR, «ESCUELA POLITECNICA NACIONAL,» Agosto
- [7] 2018. [En línea]. Available: <https://bibdigital.epn.edu.ec/bitstream/15000/20098/1/CD-9539.pdf>.
- A. F. Luna Camacho y N. E. Rodríguez Menjura, «PONTIFICIA UNIVERSIDAD JAVERIANA,» Noviembre 2017. [En línea]. Available:
- [8] <https://repository.javeriana.edu.co/bitstream/handle/10554/38771/Andr%c3%a9s%20Luna%2c%20%20Nicol%c3%a1s%20Rodr%c3%adguez.pdf?sequence=1&isAllowed=y>.

- [9] «Pasos de un proyecto de visión artificial,» Ultralytics YOLO Docs, 29 mayo 2024. [En línea]. Available: <https://docs.ultralytics.com/es/guides/steps-of-a-cv-project/#what-are-the-best-practices-for-monitoring-and-maintaining-a-deployed-computer-vision-model>.
- [10] E. M. Moreno, *El impacto de las tendencias tecnológicas y organizacionales en el trabajo administrativo*, Azcapotzalco: Azcapotzalco, 2003.
- [11] CONTAVAL, «CONTAVAL,» 18 FEBRERO 2016. [En línea]. Available: <https://www.contaval.es/que-es-la-vision-artificial-y-para-que-sirve/>.
- [12] L. A. Cárdenas Díaz y R. M. Tirado Ballesta, «Repositorio de Universidad Tecnológica de Bolívar,» 2004. [En línea]. Available: <https://repositorio.utb.edu.co/bitstream/handle/20.500.12585/3166/0028381.pdf?sequence=1&isAllowed=y>. [Último acceso: Septiembre 2023].
- [13] Microsoft, «Microsoft,» [En línea]. Available: <https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-is-computer-vision#clasificaci%C3%B3n-de-objetos>.
- [14] IBM, «IBM,» [En línea]. Available: <https://www.ibm.com/topics/computer-vision#:~:text=Computer%20vision%20is%20a%20field,recommendations%20based%20on%20that%20information..>
- [15] R. Delgado, «INISEM-BUSINESS SCHOOL,» 11 MAYO 2023. [En línea]. Available: <https://www.inesem.es/revistadigital/gestion-integrada/vision-artificial/>.
- [16] T. B. Moeslund, A. Hilton y V. Krüger, «ELSEVIER,» 22 Feb 2006. [En línea]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S1077314206001263>.

- R. Hernández Sampieri, C. Fernández Collado y M. d. P. Baptista Lucio,
 [17] Metodología de la Investigación, México: McGRAW-HILL /
 INTERAMERICANA EDITORES, 2014.
- UPSE, «UPSE,» [En línea]. Available:
 [18] https://www.upse.edu.ec/index.php?option=com_content&view=article&id=10&Itemid=166.
- UPSE, «UPSE-MISION VISION,» [En línea]. Available:
 [19] https://www.upse.edu.ec/index.php?option=com_content&view=article&id=12&Itemid=190.
- S. J. Russell y P. Norving, Artificial Intelligence A Modern Approach,
 [20] Pearson, 2020.
- [21] I. Goodfellow, Y. Bengio y A. Courville, *Deep Learning*.
- O. Najmath , A.-A. Abdulla , . A.-M. Somaya, E. Omar y M. Amr, «Journal of
 [22] King Saud University - Computer and Information Sciences,» *ELSEVIER*, p.
 35, 2023.
- P. P. A. Thakkar, «The upsurge of deep learning for computer vision
 [23] applications,» *IAES*, vol. 10, pp. 538-548, 2020.
- ACADEMIA DE GUERRA DEL EJERCITO, *El terrorismo y la seguridad
 [24] del Estado y de la sociedad*, 2024.
- S. Haykin, NEURAL NETWORKS A COMPREHESIVE FOUNDATION,
 [25] PEARSON EDUCATION .
- Y. Lecun, L. Bottou, Y. Bengio y P. Haffner, «Gradient-based learning applied
 [26] to document recognition,» *IEEE XPLORE*, vol. 86, nº 11, pp. 2278-2324.

- [27] T. M. Mitchell, *Machine Learning*, Book News, Inc, 1997.
- [28] X. Tong , L. Shuang, W. Bochao, L. Liang y W. Xiaogang , *Joint Detection and Identification Feature Learning for Person Search*, IEEE SPLORE.
- [29] A.-Z. Zeyad , V. Branislav y H. Imad , «Image Processing Based Ambient Context-Aware People Detection and Counting,» *International Journal of Machine Learning and Computing*,, vol. 8, nº 11, 2018.
- [30] viso.ai, «VISO.AI,» [En línea]. Available: <https://viso.ai/application/people-counting/>.
- [31] V. Meel, 13 Febrero 2021. [En línea]. Available: <https://viso.ai/applications/people-counting-system/>.
- [32] A. TECHVISION, «AIVID TECHVISION,» 28 Septiembre 2021. [En línea]. Available: <https://www.aividtechvision.com/people-counting-using-computer-vision/>.
- [33] L. Zheng, L. Shen, L. Tian , S. Wang, J. Wang y T. Qi , *Scalable Person Re-identification: A Benchmark*, 2015.
- [34] K. Zhou, Y. Yang, A. Cavallaro y T. Xiang, *Omni-Scale Feature Learning for Person Re-Identification*.
- [35] Microsoft-VS, «Microsoft-VS,» [En línea]. Available: <https://visualstudio.microsoft.com/es/>.
- [36] S. Kelly, «Springer Link,» [En línea]. Available: https://link.springer.com/chapter/10.1007/978-1-4842-2517-2_2.
- [37] FastAPI, «FastAPI,» [En línea]. Available: <https://fastapi.tiangolo.com/es/>.

- [38] VUEJS, «VUEJS,» [En línea]. Available:
<https://cli.vuejs.org/guide/installation.html>.
- [39] SQL-Microsoft, «Microsoft,» [En línea]. Available:
<https://www.microsoft.com/es-es/sql-server/sql-server-downloads>.
- [40] Nanobaly, «Cómo realizar anotaciones en imágenes con CVAT,» innovatiana,
04 Marzo 2024. [En línea]. Available: <https://es.innovatiana.com/post/cvat-a-guide-by-innovatiana>.
- [41] A. Lenihan, «¿Qué es el protocolo de transmisión en tiempo real (RTSP)?,»
dacast, 04 Octubre 2022. [En línea]. Available:
<https://www.dacast.com/es/blog-es/what-is-real-time-streaming-protocol/>.
- [42] R. S. A. a. Applications, Computer Vision, Springer, 2011.
- [43] B. Yang, S. Yang, X. Zhu, M. Qi, H. Li, Z. Lv, X. Cheng y F. Wang,
«Computer Vision Technology for Monitoring of Indoor and Outdoor
Environments and HVAC Equipment: A Review,» vol. 23, 2023.
- [44] T. Zhao, «AI in Educational Technology,» 2023.
- [45] G. Mesika, «Monitoreo inteligente,» Maprix, 11 Junio 2024. [En línea].
Available: <https://www.maprix.com/es/blog/intelligent-monitoring-how-automated-video-surveillance-is-revolutionizing-threat-detection>.
- [46] D. Ameijeiras Sanchez, H. González Díez y Y. Hernández Heredia, «Revisión
de algoritmos de detección y seguimiento de objetos con redes profundas para
videovigilancia inteligente,» *Revista Cubana de Ciencias Informáticas*, vol.
14, nº 3, pp. 165-195, 2020.

- D. A. Moctezuma-Ochoa, «Re-identificación de personas a través de sus
[47] características soft-biométricas en un entorno multi-cámara de video-
vigilancia,» *Ingeniería, investigación y tecnología*, vol. 17, nº 2, 2022.
- J. Rodríguez y C. Sampedro, «Métodos de aprendizaje profundo para la
[48] segmentación semántica de personas,» Industriales ERSII - UPM, 2020.
- D. Forsyth y J. Ponce, *Computer Vision: A Modern Approach*, Pearson
[49] Education., 2012.

Anexo

Anexo 1: Fotografías de registro manual



Figura 29: Biblioteca UPSE



Figura 30: Registro Manual Biblioteca UPSE

Anexo 2: Captura de resultados de entrenamiento.

```
Starting training for 50 epochs...
```

Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size
1/50	7.08G	1.504	1.521	1.303	125	640: 100% [██████████] 28/28 [00:11<00:00, 2.34it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 5.24it/
	all	110	782	0.578	0.637	0.529 0.25
2/50	7.16G	1.406	1.091	1.245	124	640: 100% [██████████] 28/28 [00:11<00:00, 2.45it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 5.15it/
	all	110	782	0.455	0.54	0.449 0.182
3/50	7.19G	1.428	1.051	1.249	68	640: 100% [██████████] 28/28 [00:11<00:00, 2.47it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 5.45it/
	all	110	782	0.0228	0.387	0.0154 0.00584
4/50	7.12G	1.482	1.043	1.264	73	640: 100% [██████████] 28/28 [00:11<00:00, 2.48it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 5.56it/
	all	110	782	0.0271	0.00895	0.00134 0.000463
5/50	7.08G	1.474	1.006	1.244	63	640: 100% [██████████] 28/28 [00:11<00:00, 2.50it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 5.43it/
	all	110	782	0.295	0.297	0.203 0.0862
6/50	7.11G	1.368	0.9394	1.212	78	640: 100% [██████████] 28/28 [00:11<00:00, 2.49it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 5.19it/
	all	110	782	0.279	0.269	0.179 0.0928
7/50	7.12G	1.365	0.8831	1.214	82	640: 100% [██████████] 28/28 [00:11<00:00, 2.47it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 4.07it/
	all	110	782	0.427	0.439	0.368 0.169
8/50	7.09G	1.312	0.8679	1.191	89	640: 100% [██████████] 28/28 [00:11<00:00, 2.37it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 4.10it/
	all	110	782	0.639	0.569	0.588 0.307
9/50	7.09G	1.274	0.8115	1.153	76	640: 100% [██████████] 28/28 [00:11<00:00, 2.35it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:01<00:00, 3.85it/
	all	110	782	0.556	0.613	0.562 0.286
10/50	7.13G	1.265	0.7878	1.158	109	640: 100% [██████████] 28/28 [00:11<00:00, 2.37it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 4.28it/
	all	110	782	0.631	0.604	0.611 0.314
11/50	7.09G	1.271	0.78	1.172	96	640: 100% [██████████] 28/28 [00:11<00:00, 2.34it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 4.08it/
	all	110	782	0.658	0.697	0.675 0.327
12/50	7.09G	1.253	0.7729	1.157	101	640: 100% [██████████] 28/28 [00:11<00:00, 2.35it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 4.06it/
	all	110	782	0.675	0.695	0.686 0.348
13/50	7.1G	1.278	0.7737	1.175	96	640: 100% [██████████] 28/28 [00:11<00:00, 2.34it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 4.01it/
	all	110	782	0.653	0.611	0.618 0.319
14/50	7.09G	1.208	0.7243	1.126	78	640: 100% [██████████] 28/28 [00:11<00:00, 2.35it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:01<00:00, 4.00it/
	all	110	782	0.643	0.702	0.689 0.372
15/50	7.12G	1.162	0.6827	1.116	119	640: 100% [██████████] 28/28 [00:11<00:00, 2.35it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 4.10it/
	all	110	782	0.629	0.673	0.647 0.349
16/50	7.09G	1.146	0.6718	1.117	76	640: 100% [██████████] 28/28 [00:11<00:00, 2.38it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% [██████████] 4/4 [00:00<00:00, 4.11it/
	all	110	782	0.677	0.72	0.723 0.378

Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
17/50	7.12G	1.146	0.6789	1.105	82	640: 100% 28/28 [00:11<00:00, 2.38it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.12it/
	all	110	782	0.705	0.638	0.71 0.379
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
18/50	7.13G	1.152	0.673	1.124	59	640: 100% 28/28 [00:11<00:00, 2.37it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.00it/
	all	110	782	0.662	0.719	0.723 0.396
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
19/50	7.12G	1.151	0.6626	1.117	117	640: 100% 28/28 [00:11<00:00, 2.34it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.09it/
	all	110	782	0.722	0.684	0.723 0.381
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
20/50	7.1G	1.118	0.659	1.097	82	640: 100% 28/28 [00:11<00:00, 2.36it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.06it/
	all	110	782	0.733	0.702	0.733 0.401
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
21/50	7.09G	1.104	0.6385	1.098	76	640: 100% 28/28 [00:11<00:00, 2.35it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.12it/
	all	110	782	0.729	0.724	0.753 0.413
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
22/50	7.08G	1.088	0.6239	1.087	68	640: 100% 28/28 [00:11<00:00, 2.35it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:01<00:00, 3.78it/
	all	110	782	0.75	0.711	0.771 0.434
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
23/50	7.11G	1.052	0.6126	1.069	52	640: 100% 28/28 [00:11<00:00, 2.34it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.17it/
	all	110	782	0.72	0.703	0.738 0.396
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
24/50	7.03G	1.064	0.6175	1.075	113	640: 100% 28/28 [00:11<00:00, 2.37it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.03it/
	all	110	782	0.763	0.732	0.77 0.425

Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
25/50	7.12G	1.073	0.6064	1.078	65	640: 100% 28/28 [00:11<00:00, 2.38it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.09it/
	all	110	782	0.701	0.776	0.761 0.418
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
26/50	7.11G	1.065	0.5881	1.07	121	640: 100% 28/28 [00:11<00:00, 2.37it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:01<00:00, 4.00it/
	all	110	782	0.723	0.783	0.783 0.437
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
27/50	7.11G	1.052	0.5844	1.068	94	640: 100% 28/28 [00:11<00:00, 2.37it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:01<00:00, 3.94it/
	all	110	782	0.705	0.746	0.761 0.429
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
28/50	7.04G	1.02	0.5803	1.05	60	640: 100% 28/28 [00:11<00:00, 2.38it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.17it/
	all	110	782	0.731	0.762	0.789 0.443
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
29/50	7.1G	1.01	0.5716	1.047	73	640: 100% 28/28 [00:11<00:00, 2.35it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.12it/
	all	110	782	0.739	0.726	0.748 0.411
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
30/50	7.12G	0.9917	0.5553	1.042	92	640: 100% 28/28 [00:11<00:00, 2.37it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.04it/
	all	110	782	0.694	0.744	0.73 0.403
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
31/50	7.09G	0.9752	0.5375	1.043	112	640: 100% 28/28 [00:11<00:00, 2.37it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.27it/
	all	110	782	0.725	0.735	0.75 0.426
Epoch	GPU_mem	box_loss	cls_loss	dfll_loss	Instances	Size
32/50	7.1G	0.9733	0.5401	1.038	107	640: 100% 28/28 [00:11<00:00, 2.37it/s]
	Class	Images	Instances	Box(P)	R	mAP50 mAP50-95): 100% 4/4 [00:00<00:00, 4.20it/
	all	110	782	0.75	0.775	0.78 0.432

Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
33/50	7.12G	0.9636	0.5324	1.031	105	640: 100%	28/28 [00:11<00:00, 2.37it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.705	0.749		0.764	0.423
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
34/50	7.11G	0.9473	0.5246	1.027	106	640: 100%	28/28 [00:11<00:00, 2.35it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.737	0.752		0.754	0.415
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
35/50	7.08G	0.9366	0.5194	1.025	112	640: 100%	28/28 [00:11<00:00, 2.37it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.738	0.729		0.744	0.408
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
36/50	7.11G	0.9234	0.5228	1.025	77	640: 100%	28/28 [00:11<00:00, 2.37it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.75	0.711		0.744	0.424
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
37/50	7.07G	0.9203	0.5112	1.018	86	640: 100%	28/28 [00:11<00:00, 2.37it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.749	0.711		0.734	0.408
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
38/50	7.07G	0.8904	0.4998	1.004	60	640: 100%	28/28 [00:11<00:00, 2.38it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.736	0.784		0.767	0.413
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
39/50	7.1G	0.8781	0.4798	0.9856	72	640: 100%	28/28 [00:11<00:00, 2.41it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.725	0.742		0.737	0.404
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
40/50	7.1G	0.8858	0.4843	1.005	75	640: 100%	28/28 [00:11<00:00, 2.48it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.747	0.73		0.736	0.403
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
41/50	7.01G	0.893	0.4888	1.008	47	640: 100%	28/28 [00:11<00:00, 2.39it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.751	0.75		0.76	0.423
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
42/50	7.01G	0.837	0.4565	0.9878	44	640: 100%	28/28 [00:11<00:00, 2.49it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.745	0.748		0.748	0.424
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
43/50	7.02G	0.8424	0.4459	0.9896	40	640: 100%	28/28 [00:11<00:00, 2.50it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.697	0.767		0.739	0.414
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
44/50	7.04G	0.8499	0.4465	0.9976	44	640: 100%	28/28 [00:11<00:00, 2.51it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.732	0.762		0.744	0.418
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
45/50	7.01G	0.814	0.4319	0.9841	52	640: 100%	28/28 [00:11<00:00, 2.50it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.745	0.763		0.754	0.423
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
46/50	7.01G	0.8208	0.4322	0.9812	44	640: 100%	28/28 [00:11<00:00, 2.49it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.745	0.763		0.754	0.423
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
47/50	7.01G	0.8059	0.4192	0.972	45	640: 100%	28/28 [00:11<00:00, 2.51it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.731	0.758		0.744	0.421
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
48/50	7.01G	0.8007	0.4172	0.9796	40	640: 100%	28/28 [00:11<00:00, 2.51it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.76	0.752		0.754	0.427
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
49/50	6.99G	0.7897	0.4158	0.9711	58	640: 100%	28/28 [00:11<00:00, 2.42it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.744	0.745		0.756	0.432
Epoch	GPU_mem	box_loss	cls_loss	df1_loss	Instances	Size	
50/50	7G	0.7674	0.4037	0.9612	43	640: 100%	28/28 [00:11<00:00, 2.38it/s]
Class	Images	Instances	Box(P	R		mAP50	mAP50-95): 100%
all	110	782	0.748	0.746		0.745	0.426