

Sistema Automatizado de Inventario en Salones Usando Detección de Objetos con Deep Learning

1st José Alejandro Moreno Mesa

Facultad de Ingeniería

Universidad de Antioquia

Medellín, Colombia

jose.morenom@udea.edu.co

Abstract—This paper presents the design, development, and implementation of an automated classroom inventory system based on deep learning object detection techniques. The proposed implementation identifies common classroom items such as backpacks, chairs, fans, and podiums by leveraging a fine-tuned YOLOv5 object detection model on a custom dataset collected from a classroom at the University of Antioquia in Medellín. The project includes the collection and labeling of images, data preprocessing and augmentation, and the training of the model to achieve accurate detections. Detailed evaluation metrics such as precision, recall, and F1-score validate the system's performance. The results demonstrate high accuracy, showcasing the feasibility of deploying this system for inventory management in educational settings while highlighting different challenges faced during the experiments.

Index Terms—Image, Recognition, Deep, Learning, YOLO, Inventory.

I. INTRODUCTION

El manejo eficiente del inventario en los salones de clase de la Universidad de Antioquia representa un desafío significativo debido a la gran cantidad de objetos presentes y a las limitaciones de los métodos manuales. Actualmente, los inventarios se realizan de forma manual, lo cual implica un gran consumo de tiempo y recursos, además de estar sujetos a errores humanos. Esto resulta en registros inconsistentes y una gestión poco eficiente de los recursos, lo que podría afectar tanto las operaciones administrativas como la experiencia de los estudiantes.

El uso de tecnologías basadas en aprendizaje profundo (deep learning) ofrece una solución prometedora a estos problemas. En este contexto, el aprendizaje profundo ha revolucionado el campo de la visión por computadora, permitiendo la creación de modelos de detección de objetos que son rápidos, precisos y adaptables a diversas aplicaciones del mundo real. Sin embargo, en la Universidad de Antioquia no se cuenta actualmente con un sistema automatizado para gestionar inventarios en salones, lo cual plantea una oportunidad para explorar e implementar estas tecnologías.

Este trabajo propone un Sistema Automatizado de Inventario basado en detección de objetos mediante un modelo

YOLOv5 ajustado a un conjunto de datos personalizado. La elección de YOLOv5 se justifica por su capacidad de realizar detecciones rápidas y precisas, y por su arquitectura optimizada que aprovecha las técnicas más avanzadas de deep learning. La implementación de este sistema no solo ahorra tiempo, sino que también garantiza un registro más preciso de los recursos disponibles en los salones de clase.

II. METODOLOGÍA

A. Recolección de datos

Se tomaron 315 imágenes en los salones de clase de la Universidad de Antioquia bajo diferentes condiciones de iluminación y disposición de los objetos. El conjunto de datos se dividió en dos tipos principales:

- Imágenes individuales: Fotografías específicas de cada tipo de objeto (mochilas, sillas, ventiladores y podiums) desde diferentes ángulos y posiciones).
- Imágenes completas del salón: Fotografías del salón completo, donde varios objetos están presentes simultáneamente, para simular escenarios reales de uso.

Cada imagen fue etiquetada manualmente utilizando la herramienta LabelImg, generando bounding boxes para cada objeto detectado y asignando etiquetas correspondientes. Las etiquetas fueron exportadas en el formato requerido por YOLO.



Fig. 1. Ejemplo imagen individual de una silla

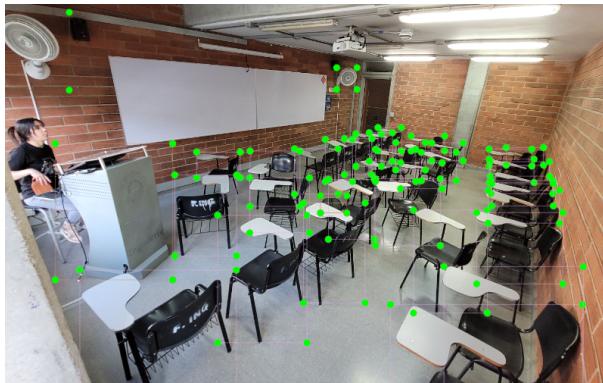


Fig. 2. Ejemplo de imagen completa de salón

B. Preprocesamiento de datos

Una vez se capturaron la totalidad de las imágenes, se realizaron diferentes procesos sobre los datos para transformarlos en información útil para el modelo:

- Limpieza de Datos: Se eliminaron imágenes borrosas o con baja iluminación.
- Data Augmentation: Se aplicaron transformaciones como rotaciones, cambios de brillo, recortes y escalado para aumentar la diversidad del conjunto de datos. Este proceso ajustaba las coordenadas de cada bounding box original de la imagen y resultaba útil para diferentes ejemplo como el de las sillas para personas diestras o zurdas.
- Formato: Las imágenes se adaptaron al tamaño requerido por YOLOv5 (640x640 píxeles), por lo que también se realizaba un escalamiento a las coordenadas de las bounding boxes.
- Reorganización de datos: Se hizo una división final de los datos en dos grupos: un grupo de entrenamiento que representaba el 70% de la información total original (que a su vez se repartía en 70% entrenamiento y 30% validación) y un grupo de test del 30% que permitía evaluar el modelo y obtener las métricas de rendimiento necesarias.



Fig. 3. Ejemplo de data augmentation

C. Entrenamiento del modelo

Se utilizó el modelo YOLOv5s, una versión ligera de YOLO optimizada para hardware con capacidades limitadas. Este modelo fue preentrenado en un conjunto de datos genérico y ajustado mediante fine-tuning con el conjunto de datos personalizado. La configuración empleada fue la siguiente:

- Tasa de aprendizaje inicial: 1x10e-3.
- Número de épocas: 50.
- Tamaño de lote: 16.
- Ejecución: Entrenamiento en GPU de referencia NVIDIA RTX3050.

D. Evaluación

Se utilizaron las siguientes métricas para evaluar el modelo:

- Precisión (Precision): Proporción de predicciones correctas.
- Recall: Proporción de objetos correctamente detectados.
- F1-Score: Promedio armónico de precisión y recall.

III. RESULTADOS

A. Métricas de Desempeño

El modelo ajustado alcanzó los siguientes resultados:

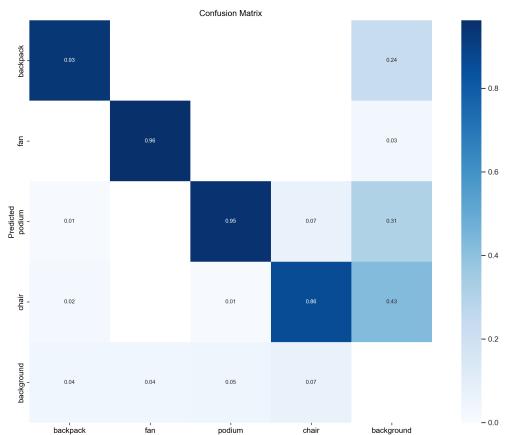


Fig. 4. Matriz de confusión de entrenamiento

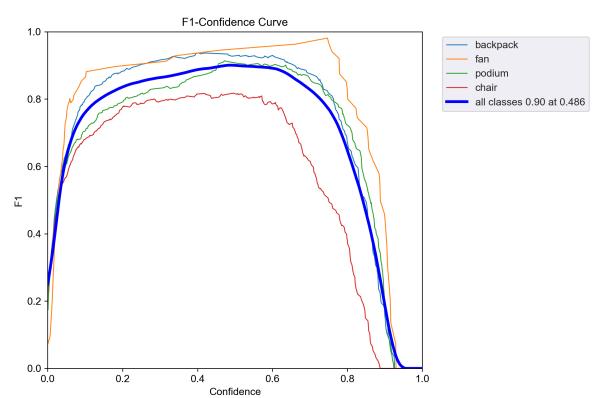


Fig. 5. Gráfica del F1-Score

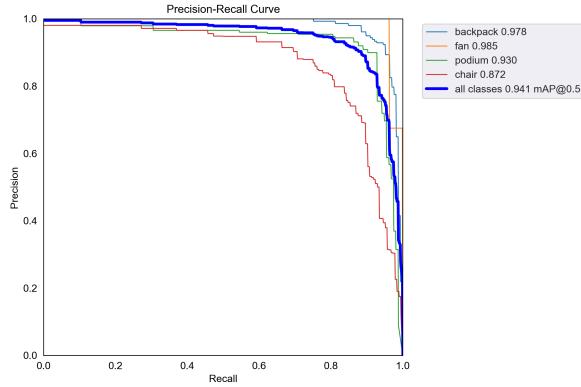


Fig. 6. Precisión vs. Recall

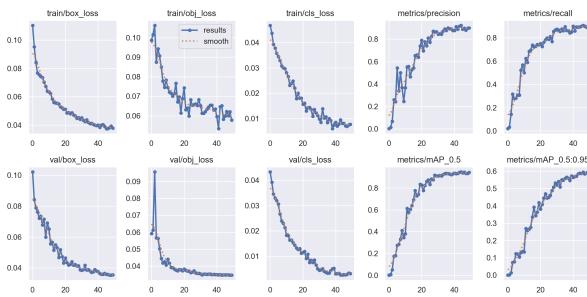


Fig. 7. Pérdidas y métricas individuales

El entrenamiento del modelo YOLOv5 generó una matriz de confusión durante la validación que incluyó las clases de morral, silla, podio y ventilador, junto con una clase adicional para el fondo (background). Esta última representa los casos donde no se detectó ningún objeto de interés, lo cual es esencial para evaluar la capacidad del modelo de discriminar entre objetos y áreas irrelevantes. El mejor modelo guardado logró un F1-score de 0.90, con un umbral de confianza de 0.498, indicando un balance óptimo entre precisión y recall. Además, el desempeño global se reflejó en un mAP en 0.5 de 0.941, lo que demuestra que el modelo realiza detecciones precisas para todas las clases en la mayoría de las imágenes evaluadas.

Los valores de box_loss, object_loss y cls_loss también mostraron un comportamiento estable, indicando un buen ajuste del modelo. El box_loss mide la precisión de las predicciones de los bounding boxes en relación con las etiquetas reales; el object_loss evalúa la confianza asignada a cada predicción, diferenciando objetos relevantes de fondo; y el cls_loss se enfoca en la correcta clasificación de cada objeto detectado. Estos resultados sugieren que el modelo logró un equilibrio entre detectar, localizar y clasificar los objetos correctamente, consolidando su eficacia en el problema planteado.

B. Ejemplos de resultados analizadas

Empleando el mejor modelo, se presentaron algunas resultados como los siguientes:



Fig. 8. Resultado de análisis de salón completo



Fig. 9. Resultado de análisis de elementos unitarios

Con base en el análisis de diferentes casos como los expuestos anteriormente, se evidencia que el modelo pudo generalizar y aprender de manera correcta las características de los diferentes objetos. Además, se evidencia que es fácil para este modelo identificar objetos que están aislados o que no se solapan con otros. No obstante, es evidente que se presenta una confusión persistente al momento de diferenciar sillas de morrales, causando que haya un etiquetado incorrecto dentro de bounding boxes entre ambos tipos de objetos.

Como el objetivo del proyecto es poder tener un inventario final, todas las imágenes analizadas eran resumidas en términos de cantidad de elementos en un archivo de excel que tenía la siguiente estructura:

image	chair	backpack	fan	podium
chair	1	0	0	0
fullA	11	2	1	0
fullB	9	5	1	0
fullC	13	10	1	1
podium	0	0	1	1

Fig. 10. Ejemplo de imagen completa de salón

IV. ANÁLISIS DE RESULTADOS

A. Análisis del modelo implementado

YOLOv5 divide cada imagen en una cuadrícula y predice bounding boxes junto con una probabilidad asociada a cada clase para cada celda. Los resultados se filtran mediante un umbral de confianza, eliminando detecciones redundantes. Este enfoque permite un balance entre precisión y velocidad, ideal para aplicaciones prácticas.

B. Retos Enfrentados

- Iluminación Variable: Afectó el desempeño del modelo en algunas imágenes.
- Ocultamiento de Objetos: Mochilas parcialmente cubiertas por sillas fueron mal detectadas.
- Confusión de Clases: Algunas sillas fueron identificadas erróneamente como mochilas debido a etiquetas superpuestas durante la recolección de datos.
- Limitación del Conjunto de Datos: Una mayor diversidad en las imágenes podría mejorar el desempeño del modelo en escenarios más complejos.

C. Trabajo Futuro

- Para mejorar la generalización del modelo, se propone aumentar la diversidad del conjunto de datos mediante la inclusión de nuevos tipos de objetos, como video beams, tableros, televisores y diferentes modelos de sillas, así como imágenes tomadas en salones con diversas disposiciones, tamaños y condiciones de iluminación, y capturas que incluyan objetos parcialmente ocultos, lo que permitirá abordar escenarios más desafiantes.
- Se debe implementar un enfoque específico para mejorar la diferenciación entre sillas y morrales, lo cual podría incluir capturas en las que las mochilas no estén colocadas sobre las sillas, así como el uso de técnicas avanzadas de preprocesamiento, como la segmentación semántica, para definir bordes más claros entre objetos superpuestos.
- Considerar el entrenamiento de diferentes instancias de YOLO enfocadas en clases de objetos específicas. Esto permitiría combinar los resultados de varios modelos optimizados para diferentes escenarios o categorías de objetos, aumentando la precisión global.
- Explorar la integración del modelo en dispositivos de borde como NVIDIA Jetson Nano, lo que permitiría desplegar el sistema en tiempo real. Este enfoque garantizaría la practicidad del sistema para uso administrativo, especialmente en auditorías rápidas de salones.
- Implementar técnicas avanzadas de validación cruzada para garantizar que el modelo generalice de manera efectiva. Además, se podría evaluar el desempeño con métricas adicionales como el mAP por clase individual y su evolución a lo largo de diferentes conjuntos de validación.

V. CONCLUSIONES

- Este proyecto demuestra cómo los modelos de deep learning, como YOLOv5, pueden resolver problemas

complejos de manera eficiente en aplicaciones del mundo real. Su capacidad de realizar detecciones rápidas y precisas abre las puertas a sistemas automatizados en sectores como la educación, la seguridad y la logística, transformando la forma en que se gestionan los recursos.

- La principal limitación identificada fue la confusión entre clases similares, como sillas y mochilas, especialmente en imágenes donde ambos objetos estaban superpuestos o mal etiquetados. Estos errores subrayan la importancia de un etiquetamiento preciso durante la preparación de los datos, ya que problemas en esta etapa pueden propagarse a lo largo del entrenamiento y afectar el rendimiento del modelo.
- La calidad y diversidad del conjunto de datos son factores clave para el éxito de cualquier modelo de deep learning. Este proyecto destaca cómo el proceso de etiquetado manual puede introducir sesgos que afectan el desempeño del modelo. Un etiquetado meticuloso y la inclusión de datos más representativos podrían mitigar estos problemas y mejorar la generalización.
- Aunque el modelo funcionó bien en imágenes simples, los resultados muestran cómo una limitación en el conjunto de datos puede tener repercusiones significativas en escenarios complejos. Esto enfatiza la necesidad de abordar problemas específicos de las clases con estrategias especializadas, como el uso de técnicas de segmentación o modelos híbridos.

REFERENCES

- [1] Verma, N. K., Sharma, T., Rajurkar, S. D., Salour, A. (2016). Object identification for inventory management using convolutional neural network. 2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). doi:10.1109/aipr.2016.8010578
- [2] Zhou, X., Gong, W., Fu, W., Du, F. (2017). Application of deep learning in object detection. 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). doi:10.1109/icis.2017.7960069
- [3] Triff, M. (2017). Automatic inventory identification through image recognition.
- [4] Siddheshwar Harkal. (2023). Image Classification with YOLOv8. Medium Website. Available at: <https://sidharkal.medium.com/image-classification-with-yolov8-40a14fe8e4bc>