Avance 6. Conclusiones clave

Imagen que contiene Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto., Picture, Picture, Imagen  
   
   
   
 

Análisis de Anaqueles BimboNet

Equipo 14

Integrantes del equipo:

Giancarlo Franco Carrillo A01638108

Luis Fernando Rivera Albarrán A01209133

Brandon Vladimir Montoya López A01209115

## **Resumen**

El presente trabajo detalla el diseño, desarrollo y evaluación de un sistema de Visión por Computadora para la segmentación de instancias del producto “Mantecadas” en entornos de retail. El objetivo de negocio principal es la verificación de la presencia o ausencia del producto en anaquel, una tarea crítica para la gestión de la cadena de suministro y la estrategia de ventas. Ante el desafío de la recolección masiva de datos etiquetados, se implementó una metodología de dos fases para la generación del dataset, combinando un modelo fundacional (Segment Anything Model - SAM) con reconocimiento óptico de caracteres (OCR) para una pre-selección automática. Los hallazgos de ciclos de entrenamiento iterativos con el modelo YOLOv8, revelaron que la capacidad de discriminación del modelo era el principal factor limitante. Los resultados demostraron que la mejora del rendimiento estaba intrínsecamente ligada a la calidad y especificidad del dataset, más que al simple escalado de la arquitectura del modelo. Se concluye con una discusión sobre la viabilidad de la solución, una hoja de ruta para su mejora continua, y un análisis estratégico para su despliegue en una infraestructura de nube.

**1. Introducción**

**1.1. El Problema de Negocio: La Incertidumbre en el “Último Metro”**

Para una empresa de consumo masivo como Grupo Bimbo, el “último metro” de la cadena de suministro —el anaquel en el punto de venta (PDV)— es el momento de la verdad. La falta de visibilidad a esta escala, multiplicada por miles de tiendas de conveniencia, supermercados y pequeños comercios, genera una incertidumbre costosa: ¿Está el producto en el anaquel? ¿Se ha agotado? ¿Está exhibido correctamente? Tradicionalmente, responder a estas preguntas ha dependido de auditorías manuales, un proceso lento, caro y con un margen de error humano considerable. La capacidad de automatizar esta verificación no solo optimiza la gestión de inventarios, sino que también proporciona datos valiosos para la estrategia de ventas y marketing.

### **1.2. Objetivo del Proyecto: Hacia una Verdad Fundamental Automatizada**

El objetivo central de este proyecto es responder a una pregunta de negocio aparentemente simple —“¿Hay Mantecadas en esta foto, sí o no?”— de la forma más robusta y fiable posible. En lugar de abordar esto como un problema de clasificación binaria, que podría ser frágil y opaco, se optó por una solución más avanzada: la segmentación de instancias.

Este enfoque no solo confirma la presencia del producto, sino que lo delinea píxel a píxel, ofreciendo una prueba visual irrefutable de su existencia. Adicionalmente, sienta las bases para capacidades futuras como el conteo automático de unidades, la identificación de variantes específicas del producto (original, nuez, marmoleada) y el análisis de la disposición en el anaquel. El objetivo, por tanto, es desarrollar un sistema de IA capaz de identificar, contar y clasificar por tipo el producto “Mantecadas” en imágenes capturadas en los heterogéneos y caóticos entornos del retail.

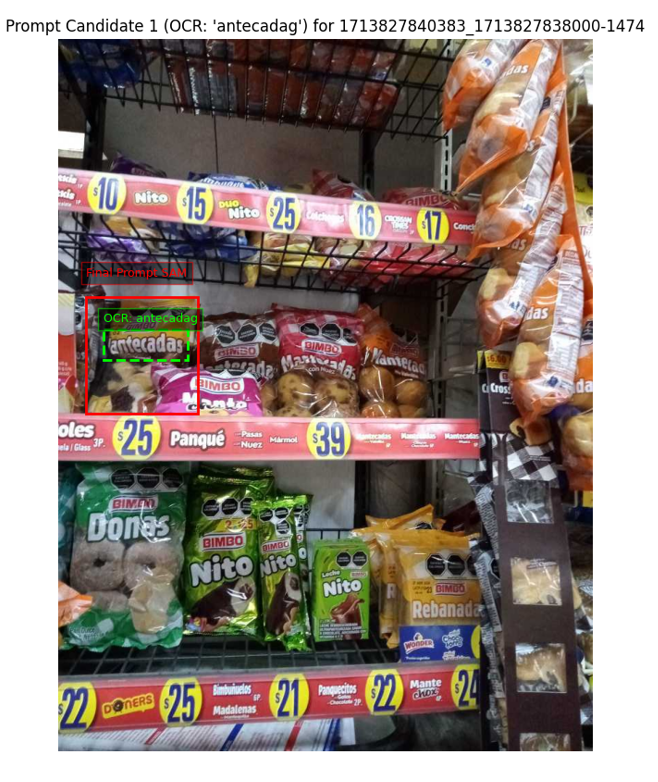
**2. Metodología y Desarrollo: El Viaje de los Datos al Modelo**  
El proyecto se concibió como un viaje iterativo, donde cada paso informaba al siguiente, siguiendo una filosofía de IA centrada en los datos (Data-Centric AI).

### **2.1. El Desafío del “Lienzo en Blanco”: Generación de Datos a Escala**

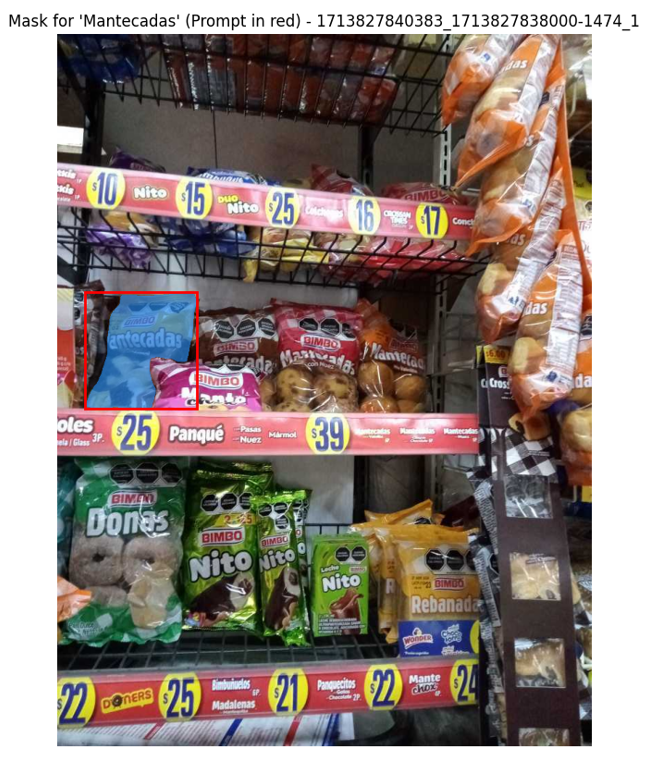
Todo proyecto de IA comienza con datos, y raramente se dispone de un dataset perfectamente etiquetado. Para superar el desafío de la creación de datos a gran escala, se diseñó un pipeline de dos etapas que transformaba un vasto conjunto de imágenes no etiquetadas en un dataset preliminar.

1. El Ojo que Lee (OCR): En la primera etapa, se utilizó la librería EasyOCR como un “explorador” de bajo nivel. Su misión era escanear miles de imágenes y simplemente encontrar la palabra “Mantecada”. Este método, aunque ruidoso, funcionó como un filtro masivo y de bajo costo para identificar regiones de interés.

El Pincel Inteligente (SAM): Las coordenadas de las palabras detectadas por OCR se convirtieron en “prompts” (pistas) para un modelo de IA mucho más potente: Segment Anything Model (SAM) de Meta. SAM, un modelo fundacional de segmentación, utilizó estas pistas para “pintar” una máscara sobre los objetos asociados a la palabra. El resultado fue la generación de miles de máscaras de segmentación preliminares, nuestro primer gran avance.



*Figura 1: OCR en funcionamiento*



*Figura 2: Mascara final usada para entrenamiento*

### **2.3. La Arquitectura del Entrenamiento: De Píxeles a Polígonos**

Con un dataset limpio y, el siguiente paso fue prepararlo para el “lenguaje” que entiende YOLOv8. Este proceso técnico se automatizó con dos scripts:

1. **organize\_segmentation\_data.py:** Puso orden en los archivos, asegurando que cada imagen tuviera una única máscara correspondiente (fusionando instancias si era necesario) y que los nombres de los archivos coincidieran perfectamente.
2. **convert\_masks\_to\_yolo.py:** Actuó como un “traductor”, convirtiendo las máscaras de imagen (.png) a archivos de texto (.txt). Cada archivo de texto describe los contornos de las mantecadas como una serie de coordenadas de polígonos normalizadas, precedidas por el índice de su nueva clase (0 para original, 1 para nuez, 2 para marmoleada).

Finalmente, el script **split\_dataset.py** dividió todo este conjunto de datos en subconjuntos de entrenamiento (80%) y validación (20%), dejando todo listo para el entrenamiento.

### **2.3.1 Evaluación de Clasificadores Convencionales sobre Datos Sintéticos**

Como parte de la exploración técnica del proyecto, se evaluó la viabilidad de utilizar clasificadores supervisados entrenados sobre un conjunto de datos sintético generado completamente en BlenderProc. Este enfoque buscaba identificar productos directamente a partir de recortes individuales, segmentados por las máscaras de BlenderProc, sin necesidad de aplicar detección en imágenes completas.

Se generaron aproximadamente 200 imágenes sintéticas a partir de escenas que contenían objetos de interés renderizados desde distintas perspectivas y bajo iluminación uniforme. A partir de las máscaras asociadas, se extrajeron recortes automáticos de cada objeto, los cuales fueron etiquetados en función del nombre del archivo y utilizados como entrada para entrenar tres clasificadores: ResNet18, EfficientNet\_b0 y MobileNet\_v3.

Los modelos fueron entrenados utilizando un esquema multiclase, donde cada tipo de objeto se trató como una clase independiente. Sin embargo, los resultados obtenidos fueron insatisfactorios. En todos los casos, los clasificadores tendieron a predecir con alta confianza una única clase dominante (por ejemplo, obj\_255), ignorando la variabilidad presente en los datos. Este comportamiento persistió incluso al evaluar sobre imágenes del propio conjunto de entrenamiento, lo cual evidencia un severo sobreajuste y falta de capacidad discriminativa.

Entre las causas identificadas se encuentran:

* Calidad limitada en la representación visual de los objetos sintéticos.
* Recortes ambiguos o poco representativos derivados de máscaras incompletas o ruidosas.
* Distribución desequilibrada de clases en el dataset generado.
* Ausencia de condiciones realistas como ruido, deformaciones, iluminación variada o fondo natural.

Estos hallazgos permitieron descartar el enfoque de clasificación directa sobre datos sintéticos como una alternativa viable para el despliegue del sistema. En su lugar, se reforzó la necesidad de adoptar métodos de detección de objetos sobre imágenes completas, como los desarrollados posteriormente mediante arquitecturas YOLO, los cuales no dependen de una segmentación previa y están diseñados para localizar múltiples instancias simultáneamente.

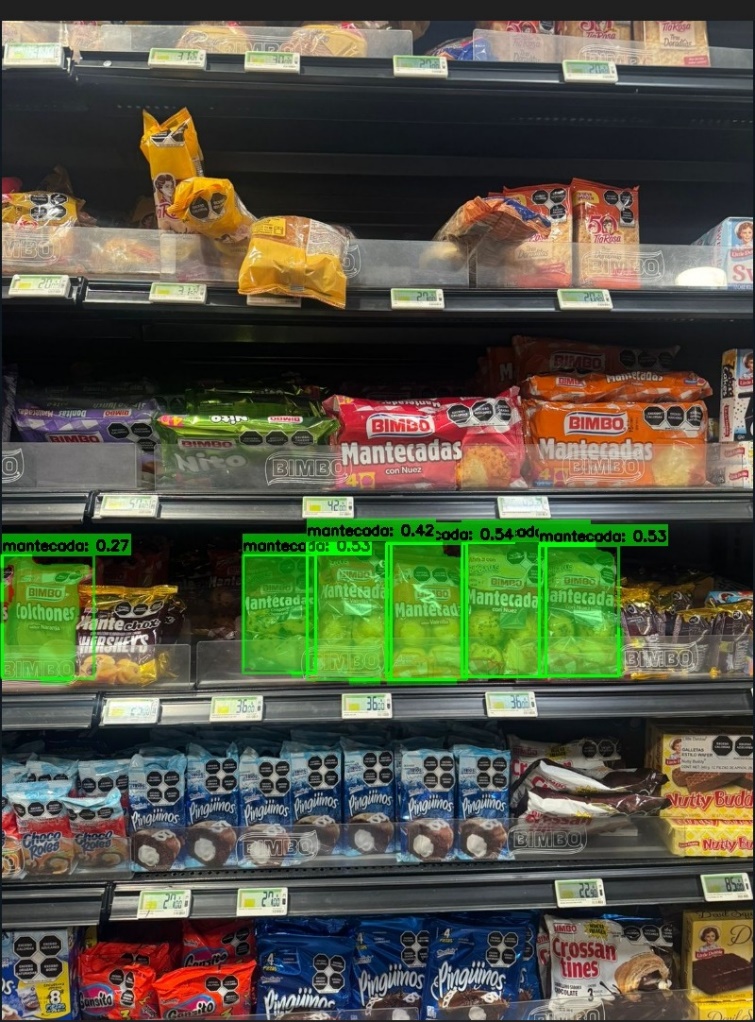
### **2.4. Ciclos de Entrenamiento Iterativo**

El entrenamiento se planteó como una serie de experimentos científicos en una estación de trabajo local equipada con una GPU NVIDIA RTX 4080 SUPER.

* Experimento 1: La Línea Base (yolov8n-seg)
  + Se entrenó el modelo más pequeño de YOLOv8 con el objetivo de establecer un punto de partida y validar el pipeline.
  + El resultado, un Mask mAP50-95 de 0.146, fue modesto pero crucial. Confirmó que el modelo podía aprender, pero el análisis de sus errores (confundir “Panqués” con “Mantecadas”) reveló su incapacidad para captar detalles finos.
* Experimento 2: La Prueba de Fuerza (yolov8m-seg + Aumentación)
  + La hipótesis era simple: ¿Necesita el modelo un “cerebro” más grande? Se escaló al modelo mediano y se activó la aumentación de datos para mejorar la robustez.
  + El resultado fue la revelación más importante del proyecto: el rendimiento no mejoró (mAP50-95 de 0.147) y el modelo se estancó en la epoch 44.

Esta conclusión fue fundamental: demostró que el cuello de botella no era la arquitectura del modelo, sino la falta de especificidad en los datos. El modelo no podía aprender a diferenciar productos porque, hasta ese momento, no le habíamos dado las herramientas para hacerlo. Este hallazgo justificó plenamente la decisión de invertir tiempo en el re-etiquetado multi-clase.

### **3. Análisis de Resultados Futuros**

  
*Figura 1: Ejemplo del modelo en funcionamiento en una imagen tomada por nosotros en un supermercado*

  
*Figura 2: Ejemplo de predicciones del modelo final sobre un lote de validación, mostrando la clasificación.*



*Figura 2: Modelo en funcionamiento en webcam.*



*Figura 2: Prueba simple del modelo en escritorio con otro producto.*

### **4. Discusión y Recomendaciones Estratégicas**

#### **4.1. Rendimiento del Modelo para Producción**

Incluso para el objetivo de negocio de detectar “presencia o ausencia”, un modelo de bajo rendimiento es inutilizable. Un modelo de clasificación simple podría ser engañado fácilmente por productos similares. La gran ventaja de nuestro enfoque de segmentación es que la “presencia” se valida con una “prueba” visual: la máscara delineada.

Un mAP50-95 superior a 0.60 (60%) podría considerarse para un piloto. Un valor superior a 0.75 (75%) podría considerarse robusto para producción. El modelo actual no cumple estos requisitos, pero hemos establecido la ruta para lograrlo.

#### **4.2. Oportunidades de Mejora**

El camino hacia un modelo de alto rendimiento está claramente definido:

1. **Finalizar el Re-etiquetado Multi-Clase:** Inicialmente, se planteó una simple validación. Sin embargo, los primeros experimentos con el modelo revelaron una debilidad fundamental: la confusión con productos de apariencia similar. Esto condujo a una decisión estratégica clave: **pivotar de un problema de una sola clase (“mantecada”) a un problema multi-clase.**
   1. **Aprobar y Clasificar:** Asignar a cada máscara de alta calidad una etiqueta específica: **“original”**, **“nuez”** o **“marmoleada”**.
   2. **Descartar:** Eliminar máscaras de mala calidad, como las de productos parcialmente visibles, empaques rotos o errores evidentes de SAM.
2. **Entrenar y Analizar el Modelo Multi-Clase:** Lanzar el entrenamiento con los datos refinados y analizar los nuevos resultados y confusiones.
3. **Incorporar “Negativos Difíciles”:** Basado en los nuevos errores, recolectar e incluir imágenes de productos de la competencia o de otras líneas que aún causen confusión.

**Aumentar el Volumen:** Una vez que el modelo demuestre una alta capacidad de diferenciación, escalar el volumen de imágenes por clase solidificará su rendimiento en una gama más amplia de escenarios.

#### **4.3. Recomendaciones Clave para Implementación**

1. **API de Inferencia:** Encapsular el modelo best.pt en una API RESTful (usando FastAPI) que reciba una imagen y devuelva un JSON con una lista de objetos detectados, cada uno con su clase, confianza y polígono.
2. **Aplicación Móvil Cliente:** Desarrollar una aplicación simple para los dueños de las tiendas que les permita tomar una foto, enviarla a la API y recibir una confirmación visual o un simple “OK / Faltante”.
3. **Infraestructura en la Nube:** Desplegar la API en un servicio de contenedores o una función serverless para garantizar la escalabilidad.

#### **4.4. Tareas Accionables para los Stakeholders**

Dado que el objetivo es verificar la **presencia o ausencia** de Mantecadas, el modelo de segmentación ofrece un valor añadido que debe ser aprovechado:

1. **Definir las Reglas de Negocio:** ¿Cómo se traduce una detección en una acción? Por ejemplo: “Si el modelo no detecta NINGUNA ‘Mantecada Original’ en la imagen, enviar una alerta al cliente.
2. **Aprovechar la Data Granular:** El modelo no solo dirá “hay Mantecadas”, sino que podrá decir “hay 5 ‘originales’ y 2 de ‘nuez’, pero ninguna ‘marmoleada’”. Los stakeholders deben diseñar las preguntas de negocio para explotar esta granularidad. ¿Queremos monitorear el lanzamiento de un nuevo sabor? Este modelo puede hacerlo.
3. **Planificar el Futuro:** Este pipeline (OCR-SAM -> Curación -> Entrenamiento YOLO) es un activo reutilizable. Los stakeholders deben identificar el **siguiente producto** de alto valor que deseen rastrear para aplicar la misma metodología.

**Diseñar el Piloto:** Colaborar en el diseño de un programa piloto con un grupo selecto de dueños de tiendas para probar la aplicación, recolectar retroalimentación y medir el impacto real en la reducción de agotados.

#### **4.5. Análisis Comparativo de Proveedores Cloud para Despliegue**

|  |  |  |  |
| --- | --- | --- | --- |
| **Característica** | **AWS (Amazon Web Services)** | **Azure (Microsoft)** | **GCP (Google Cloud Platform)** |
| **Servicios de IA/ML** | Amazon SageMaker: Plataforma muy completa y madura. | Azure Machine Learning: Excelente integración con ecosistema Microsoft. | **Vertex AI:** Plataforma unificada y muy amigable para el desarrollador. |
| **Facilidad de Uso** | Curva de aprendizaje media. | Curva de aprendizaje media. | **Alta.** Considerada la más intuitiva para empezar y escalar. |
| **Escalabilidad** | Muy Alta. Líder del mercado. | Muy Alta. | Muy Alta. |
| **Ventaja Clave** | Ecosistema más grande. | Integración nativa con herramientas Microsoft. | **Excelencia en IA/ML y datos.** |

##### **Justificación de la Elección Recomendada**

Para este proyecto, se reitera la recomendación de Google Cloud Platform (GCP). Su plataforma Vertex AI simplifica enormemente el despliegue de modelos personalizados como el desarrollado, y su enfoque en herramientas de IA amigables facilitará la creación de la API y el ciclo de re-entrenamiento continuo, lo cual es fundamental para la vida de este proyecto.