.UBAfiuba
FACULTAD DE INGENIERÍA

Automatización del Etiquetado de Productos mediante Transformadores de Visión

Visión por Computadora III

Carrera de Especialización en Inteligencia Artificial

Autores:

Abril Noguera
Pedro Lucas Barrera
Lautaro Gabriel Medina

Ciudad de Buenos Aires, diciembre de 2025

Resumen

El proyecto desarrollado consiste en la construcción de un sistema capaz de identificar y asignar atributos descriptivos a productos a partir de sus imágenes mediante un modelo de aprendizaje profundo basado en transformadores de visión. El sistema aborda una necesidad concreta del comercio digital, donde el etiquetado manual de artículos representa una tarea costosa, lenta y con alta probabilidad de errores. La solución propuesta automatiza este proceso, mejora la calidad del catálogo y permite acelerar la incorporación de nuevos productos en plataformas de venta en línea.

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Introducción general

El presente capítulo introduce el contexto general del proyecto, expone el problema que motivó su desarrollo, presenta el caso de negocio y describe los datos utilizados. Asimismo, se detallan los objetivos del trabajo y su alcance, estableciendo el marco conceptual necesario para comprender las decisiones técnicas y metodológicas que se desarrollan en los capítulos posteriores.

1.1. Contexto y motivación

El crecimiento del comercio electrónico y la disponibilidad masiva de catálogos digitales han incrementado la necesidad de organizar y clasificar grandes volúmenes de imágenes de productos. En este escenario, la automatización del etiquetado visual se ha convertido en una herramienta fundamental para mejorar la eficiencia operativa y garantizar la consistencia en la gestión de catálogos.

El proyecto desarrollado surgió a partir de esta necesidad: se implementó un sistema de etiquetado automático de productos basado en modelos de visión por computadora. Dicho sistema permitió reducir la dependencia del etiquetado manual, disminuir errores humanos y acelerar el procesamiento de catálogos. Esta iniciativa se orientó a evaluar la utilidad de arquitecturas modernas como los *Vision Transformers* en problemas de clasificación multi-etiqueta.

1.2. Caso de negocio

El crecimiento acelerado del comercio electrónico ha generado catálogos con miles de productos que requieren actualización constante. En este contexto, la correcta clasificación y etiquetado de imágenes es un proceso crítico: determina cómo los productos se muestran, cómo se encuentran mediante búsquedas internas y cómo son recomendados por los motores de recomendación. Sin embargo, este proceso suele realizarse de manera manual, lo que introduce varias limitaciones operativas.

En primer lugar, el etiquetado manual implica costos elevados en horas-hombre, particularmente en empresas que gestionan catálogos dinámicos donde ingresan cientos o miles de productos nuevos por semana. En segundo lugar, este proceso presenta altos niveles de inconsistencia debido a la subjetividad de los operadores: productos similares pueden recibir etiquetas distintas o incompletas, lo que perjudica la calidad del catálogo. Además, el tiempo requerido para clasificar grandes volúmenes genera cuellos de botella que ralentizan el lanzamiento de nuevos productos.

Las consecuencias de un etiquetado deficiente se reflejan en múltiples áreas del negocio. Un producto mal clasificado puede no aparecer en las búsquedas relevantes, reducir su tasa de conversión o ser excluido de sistemas automáticos de recomendación, impactando directamente en ventas. Asimismo, un catálogo inconsistente genera fricción en la navegación del usuario, lo que disminuye la satisfacción y deteriora la percepción de calidad del sitio.

En este contexto, resulta necesario contar con un sistema automático capaz de identificar atributos visuales de manera rápida, coherente y escalable. El presente proyecto se enmarca en esta problemática, evaluando la capacidad de modelos basados en *Vision Transformers* para realizar un etiquetado automático y confiable de productos de moda a partir de sus imágenes.

1.3. Propuesta de valor

El sistema desarrollado busca reemplazar o complementar el etiquetado manual mediante un modelo de visión por computadora capaz de asignar automáticamente atributos clave como categoría, tipo, color y género del producto. Esta automatización agrega valor en distintos niveles:

- **Reducción de costos operativos:** Disminuye la necesidad de intervención humana, especialmente en etapas iniciales de carga masiva de catálogo.
- **Consistencia y estandarización:** El modelo aplica criterios homogéneos sin variación entre operadores, reduciendo errores y ambigüedades.
- **Mayor velocidad de procesamiento:** La clasificación automática permite acelerar el tiempo desde la recepción del producto hasta su publicación en el catálogo.
- **Mejor experiencia de usuario:** Catálogos coherentes mejoran las búsquedas, la navegación y la relevancia de las recomendaciones.
- **Escalabilidad:** El sistema puede procesar miles de imágenes sin aumentar el costo marginal, algo imposible con procesos manuales.

En conjunto, la propuesta de valor consiste en un pipeline automatizado capaz de fortalecer la calidad del catálogo digital y optimizar procesos internos, alineándose con prácticas modernas de comercio electrónico basadas en datos y automatización inteligente.

1.4. Objetivos y alcance

El proyecto tuvo como objetivo general desarrollar un sistema capaz de etiquetar automáticamente imágenes de productos de moda utilizando modelos basados en *Vision Transformers*. Este sistema buscó reproducir y estandarizar el proceso de etiquetado que habitualmente se realiza de manera manual, evaluando su capacidad para predecir atributos clave del catálogo tales como categoría, tipo, color y género del producto.

En términos más específicos, el trabajo se propuso:

- Construir un pipeline reproducible de procesamiento de datos que incluyera la descarga, validación y limpieza del dataset original.

- Implementar un modelo de clasificación multi-etiqueta basado en *Vision Transformers*, adaptado a las características del dataset.
- Entrenar y validar el modelo mediante particiones estratificadas, asegurando una evaluación equilibrada de las distintas clases.
- Analizar el desempeño del modelo utilizando métricas estandarizadas como *accuracy*, F1 y *mean Average Precision* (mAP).
- Examinar los errores y sesgos presentes en las predicciones, identificando limitaciones del enfoque.
- Desarrollar una herramienta de inferencia que permitiera aplicar el modelo a nuevas imágenes en un entorno práctico.

El alcance del proyecto se limitó al análisis y desarrollo de un prototipo funcional entrenado sobre un dataset público. No se abordaron aspectos propios de un sistema productivo, tales como el entrenamiento continuo, la integración con plataformas de comercio electrónico, la optimización de tiempos de inferencia para altos volúmenes ni la incorporación de retroalimentación humana en el ciclo de etiquetado. Tampoco se realizaron experimentos con arquitecturas alternativas ni con técnicas de aumento extensivo de datos debido a restricciones de tiempo y recursos computacionales.

A pesar de estas limitaciones, el desarrollo realizado permite demostrar la viabilidad de automatizar el etiquetado de productos mediante modelos de visión por computadora modernos, sentando las bases para futuras mejoras orientadas a escenarios reales de operación.

1.5. Datos utilizados

Para el desarrollo del sistema se utilizó el dataset público *Fashion Product Images (Small)*, disponible en Kaggle[1]. Este conjunto incluye más de 44.000 imágenes de productos de moda junto con un archivo tabular `styles.csv` que contiene información estructurada del catálogo.

El dataset original presenta atributos como categoría general del producto, subcategoría, tipo, género, temporada y color dominante. Cada imagen está identificada mediante un `productId`, lo que permite vincularla con sus metadatos tabulares. Antes de poder utilizarse en el pipeline de entrenamiento, se aplicaron procesos de limpieza, normalización y verificación de integridad debido a la presencia de valores faltantes, clases poco representadas y rutas inconsistentes.

A partir de este conjunto inicial, se construyó un dataset final que incluyó: (i) imágenes validadas y preprocesadas, (ii) atributos seleccionados para la predicción y (iii) una partición estratificada en *train*, *validation* y *test*. Esta estructura permitió evaluar de manera rigurosa el desempeño del modelo entrenado bajo condiciones reales.

[1] https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-small

Apéndice A

Planificación del Proyecto

A continuación se presenta la tabla detallada de planificación del proyecto, que incluye todas las tareas necesarias para completar el trabajo desde la inicialización del repositorio hasta la presentación final.

TABLA A.1. Planificación del Proyecto

Título	Descripción	Responsable	Estado
Inicializar cookiecutter	Ejecutar `cookiecutter-data-science` para generar la estructura base del proyecto.	Abril	Completado
Configurar entorno Conda	Crear entorno `product_tagger` e instalar dependencias.	Abril	Completado
Crear repositorio GitHub	Crear repositorio, subir estructura inicial, conectar con VSCode.	Abril	Completado
Definir proyecto final	Acordar que el proyecto será un sistema Product Tagger basado en Vision Transformers.	Equipo	Completado
Definir Business Case	Documentar problema de negocio, costos y beneficios.	Abril	En Progreso
Obtención de Datos	Implementar descarga automática del dataset desde `KaggleHub`.	Abril	En progreso
EDA	Explorar `styles.csv`, distribución de atributos y ejemplos de imágenes.	XX	Pendiente
Definir atributos objetivo	Seleccionar atributos a predecir.	XX	Pendiente
Limpieza del CSV	Corregir faltantes y normalizar strings.	XX	Pendiente
Unir imágenes con CSV	Crear columna `image_path` y validar archivos.	XX	Pendiente
Generar dataset final	Guardar CSV procesado en `data/processed`.	XX	Pendiente
Train/Val/Test Split	Crear particiones estratificadas.	XX	Pendiente
Implementar PyTorch Dataset	Construir clase Dataset para imágenes y etiquetas.	XX	Pendiente
Implementar augmentations	Agregar normalización, resize, flips y transforms.	XX	Pendiente
Cargar Vision Transformer	Cargar ViT/DeiT preentrenado y adaptar classifier.	XX	Pendiente
Pipeline de entrenamiento	Entrenar ViT con optimizer, scheduler y early stopping.	XX	Pendiente
Registrar métricas	Guardar loss, accuracy, F1 y mAP por época.	XX	Pendiente
Evaluar modelo	Evaluar rendimiento sobre test set.	XX	Pendiente
Visualizar resultados	Graficar curvas, métricas y matriz de confusión.	XX	Pendiente
Ejemplos de predicción	Mostrar aciertos y errores.	XX	Pendiente
Guardar modelo entrenado	Exportar modelo final a `models/`.	XX	Pendiente
Implementar CLI	Script Typer de inferencia para imágenes nuevas.	XX	Pendiente
Escribir README	Instrucciones de instalación, entrenamiento e inferencia.	XX	Pendiente
Redactar Objetivo	Escribir objetivo del proyecto en el informe.	XX	Pendiente
Redactar Arquitectura	Diagramar y explicar el pipeline.	XX	Pendiente
Redactar Implementación técnica	Detallar módulos y diseño del sistema.	XX	Pendiente
Redactar Evaluación	Explicar métricas y análisis de resultados.	XX	Pendiente
Redactar Resultados	Incluir visualizaciones y análisis final.	XX	Pendiente
Redactar Conclusiones	Limitaciones y líneas futuras.	XX	Pendiente
Preparar presentación	Diapositivas y narrativa final.	XX	Pendiente
Practicar presentación	Ensayo del equipo.	XX	Pendiente
Entrega final código	Limpiar repo y hacer tagging final.	XX	Pendiente
Entrega final PDF	Compilar y entregar memoria final.	XX	Pendiente