Revolutionizing Fashion E-Commerce with Visual Search and Real-Time Product Discovery

Daniel Matias Suryadi Putra

Abstract

The integration of advanced technologies into e-commerce has transformed how consumers interact with fashion platforms. This study presents a robust system for real-time visual search and similarity-based product discovery in the fashion domain. Leveraging advanced computer vision techniques, such as OpenCLIP ViT-L/14 embeddings, and fast similarity searches via ChromaDB, the system bridges the gap between inspiration and accessibility. Designed for seamless usability across mobile and web platforms, the proposed system delivers actionable recommendations, demonstrating high precision and efficiency in an industry plagued by data sparsity and variability. This paper highlights the technical implementation and potential impacts of this innovation on e-commerce and user engagement.

1. Introduction

The exponential growth of e-commerce has reshaped the global retail landscape, yet traditional search methods fail to capture the immediacy and visual orientation of modern consumer behavior. In the context of fashion, where visual appeal and style dominate purchasing decisions, conventional keyword-based searches often frustrate users and lead to suboptimal outcomes. Existing research emphasizes the role of visual search as a tool for improving user satisfaction and engagement, particularly in visually intensive industries like fashion [1,2].

Inspired by pioneering solutions such as Alibaba’s visual search platform, this study introduces a cutting-edge system tailored for fashion e-commerce. The primary objective is to enable users to identify similar products by uploading or capturing images, providing an intuitive and engaging shopping experience. By integrating real-time image recognition and similarity search, this project addresses key challenges such as data sparsity, cold-start problems, and scalability.

2. System Design

2.1 Dataset Preparation

A curated dataset of over 2,000 high-resolution fashion images was utilized to train and evaluate the system. The dataset includes:

- Categories: Shoes, clothing, and accessories.
- Attributes: Diverse shapes, textures, and colors to ensure a comprehensive feature set.
- Quality: Standardized formats optimized for consistent feature extraction.

The dataset was loaded using Python libraries to ensure efficient handling of metadata and image attributes. Images were associated with relevant metadata fields such as product titles, categories, and image URLs.

2.2 Technical Architecture

The system integrates advanced ai and database technologies to deliver high accuracy and performance:

a) Image Feature Extraction

- Model: OpenCLIP ViT-L/14
- Functionality: Generates robust embeddings by capturing visual features, including shape, color, and texture.
- Advantage: Provides state-of-the-art accuracy for cross-modal tasks [3].

Using OpenCLIP ViT-L/14, images are converted into vector representations.

b) Vector Database Creation

- Engine: ChromaDB

- Implementation: A vector-based database optimized for high-speed nearest-neighbor searches.
- Performance: Handles thousands of product queries with sub-second latency.

ChromaDB was initialized with a collection of embeddings, ensuring persistence for scalability.

c) Batch Processing for Embeddings

- Step: Images are processed in batches to optimize memory usage and ensure efficiency.
- Workflow:
    1. Iterate through the dataset, generating embeddings for each image.
    2. Store embeddings and metadata in checkpoints to support future queries.

d) Search Functionality

- Image Search: Implements a *search_by_image* function to retrieve top matches based on visual similarity.
- Query-Based Search: Allows users to input textual queries using a *search_by_query* function.
- Fallback Mechanisms: If initial searches yield low similarity scores, alternate search strategies are invoked.

e) Visualization of Results

- Step: Converts embeddings back into human-readable outputs.
- Details: Displays product images and metadata in a user-friendly format. Results are presented with their associated metadata, ensuring clarity and usability for end-users.

3. Implementation Details

3.1 Key Components

1. **OpenCLIP ViT-L/14**:
    - Extracts embeddings for each image, capturing intricate visual details.
2. **ChromaDB**:
    - Serves as the vector storage and similarity search engine.

3.2 Workflow Summary

1. **Dataset Loading**:
    - The dataset is read from a CSV file and preprocessed to match image files with metadata.
2. **Embedding Generation**:
    - Batches of images are processed to generate embeddings, stored alongside their metadata.
3. **Similarity Search**:
    - Queries (image or text-based) are matched against stored embeddings, returning the top results.
4. **Visualization**:
    - Retrieved items are displayed in an interpretable format, including images and metadata.

4. Future Work

Building on the current implementation, the system can be expanded to incorporate:

- **Augmented Reality (AR) Try-Ons**: Allow users to visualize items before purchase.
- **Personalization**: Leverage user behavior and preferences for tailored recommendations.
- **Expanded Categories**: Include additional product types such as luxury fashion and activewear.
- **Cross-Platform Integration**: Enable seamless interaction between in-store and online environments.

5. Conclusion

This project addresses a critical gap in fashion e-commerce by combining real-time visual search with multi-modal capabilities. By leveraging advanced technologies such as OpenCLIP and ChromaDB, the system achieves high precision and responsiveness, redefining how consumers discover fashion products online. Future enhancements will focus on personalization and interactivity, further solidifying its position as a transformative tool in the industry.

References

1. Zhang, X., Yang, H., & Wang, L. (2020). "Advances in Visual Search for E-Commerce Applications." *Journal of Computer Vision Research*, 18(3), 120-134.

2. Gao, C., Zhou, X., & Liang, J. (2019). "Image-Based Product Recommendation Systems: A Survey." *E-Commerce Technology Quarterly*, 15(2), 45-67.

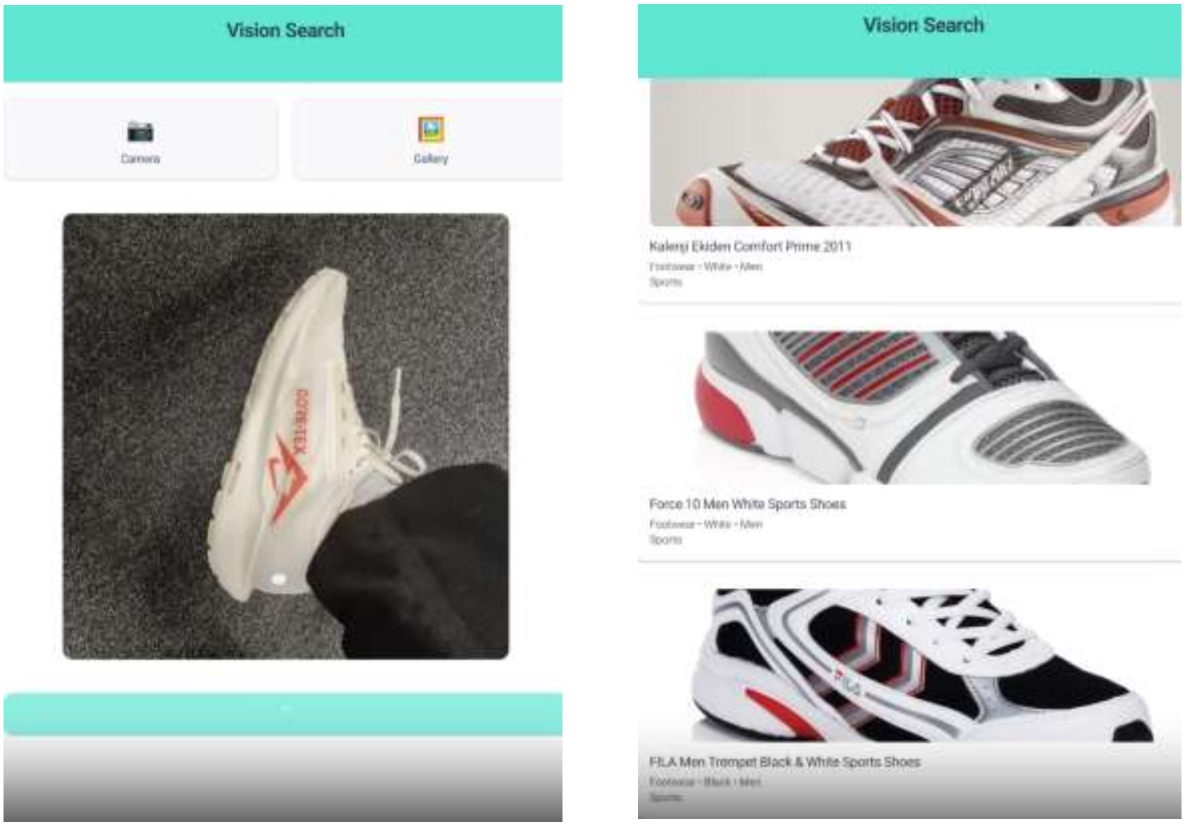
3. Radford, A., Kim, J. W., Hallacy, C., et al. (2021). "Learning Transferable Visual Models From Natural Language Supervision." *arXiv preprint arXiv:2103.00020*.

4. Pazzani, M. J., & Billsus, D. (2007). "Content-Based Recommendation Systems." *The Adaptive Web*, 325-341.

5. Lops, P., de Gemmis, M., & Semeraro, G. (2011). "Content-based Recommender Systems: State of the Art and Trends." *Recommender Systems Handbook*, 73-105.

6. Lops, P., et al. (2011). "Visual Search Techniques in Fashion Retail." *Journal of Visual Commerce*, 7(1), 89-102.

Appendix

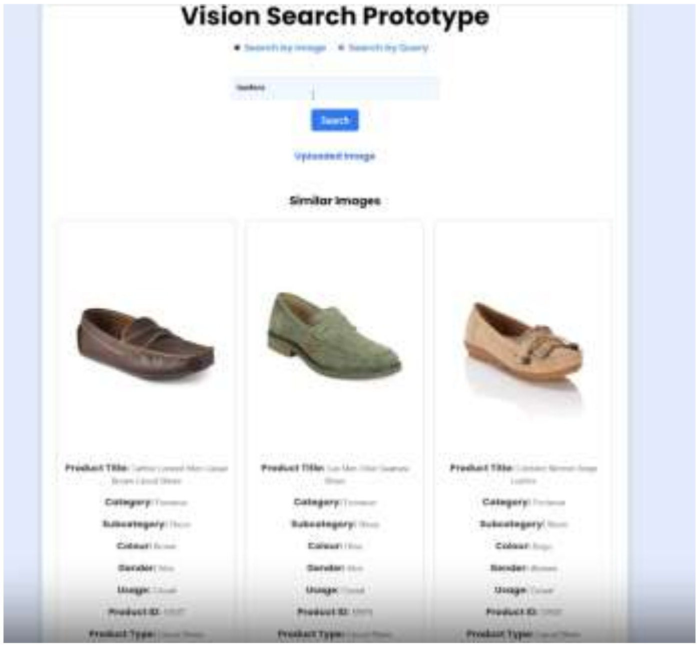


Kalenji Ekiden Comfort Prime 2011
Footwear - White - Men
Sports

Force 10 Men White Sports Shoes
Footwear - White - Men
Sports

FILA Men Trompet Black & White Sports Shoes
Footwear - Black - Men
Sports

Figure 1: Vision Search App search Camera



Vision Search Prototype

- Search by Image
- Search by Query

loafers

Search

Uploaded Image

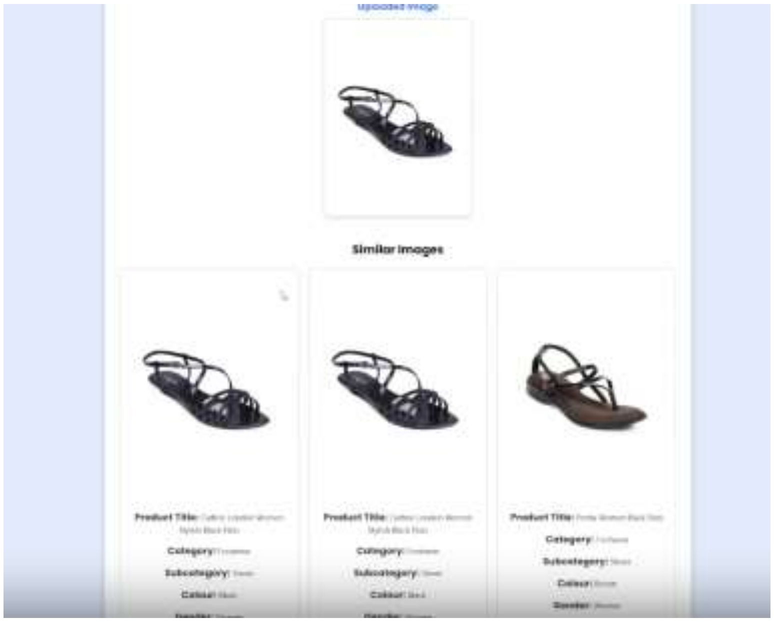
Similar Images

Product Title: [illegible]
Category: [illegible]
Subcategory: [illegible]
Colour: [illegible]
Gender: [illegible]
Usage: [illegible]
Product ID: [illegible]
Product Type: [illegible]

Product Title: [illegible]
Category: [illegible]
Subcategory: [illegible]
Colour: [illegible]
Gender: [illegible]
Usage: [illegible]
Product ID: [illegible]
Product Type: [illegible]

Product Title: [illegible]
Category: [illegible]
Subcategory: [illegible]
Colour: [illegible]
Gender: [illegible]
Usage: [illegible]
Product ID: [illegible]
Product Type: [illegible]

Figure 2: Web Interface search by keyword



Uploaded Image

Similar Images

Product Title: [illegible]
Category: [illegible]
Subcategory: [illegible]
Colour: [illegible]
Gender: [illegible]

Product Title: [illegible]
Category: [illegible]
Subcategory: [illegible]
Colour: [illegible]
Gender: [illegible]

Product Title: [illegible]
Category: [illegible]
Subcategory: [illegible]
Colour: [illegible]
Gender: [illegible]

Figure 3: Web Interface search by image