



INFORMATICS INSTITUTE OF TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

**ZENSEARCH: Revolutionizing E-Commerce Search through
Advanced Multimodal Integration and Retrieval Techniques**

A dissertation by

Mr. Ratnalingam Gajeendran

W1867080 | 20210328

Supervised by

Mr. Mithushan Jalangan

Submitted in partial fulfilment of the requirements for the BEng(Hons) Software Engineering
degree at the University of Westminster.

April 2025

Abstract

Problem: The dawn of the multimodal search systems in the e-commerce domain has been promising with an expansion of the variety of online product catalogs. These multimodal systems present forward a challenge for efficient retrieval recommendation systems in a unified space. The existing traditional e-commerce systems struggle to focus the user query to give a relevant product recommendation at the end of the retrieval stage where they mostly rely on unimodal approaches. This project explores this gap through developing an efficient multimodal retrieval system utilizing the ColPali architecture where the product images and captions are mapped effectively into a unified space to ultimately produce accurate and context-aware product recommendations.

Methodology: This research proposes a vision language model to generate contextualized vector embeddings using a product catalog as an input. These embeddings are projected into a 128-dimensional vector space and stored in a vector database for the retrieval where all the above processes occur in an offline stage. When a user is passing a query into the system, it undergoes a similar embedding generation process to maintain the nature of the embeddings. And similarity search systems will be used to fetch the efficient and relevant product suggestion with the integration of the late-interaction mechanism.

Testing: The proposed approach with the above mentioned methodology and problem has not been attempted and the system reaches an accuracy of 89% during the testing and evaluation phase which outperforms the base model trained on. Along with that the proposed approach exhibits better metrics in terms of NDCG, Precision, Recall, F1-Score etc. as well. The model trained shows a better performance in the vidore benchmarking which is a best-fitting benchmark for the colpali model which is quite evident in producing most relevant top-k retrievals by making the proposed approach achieve a greater milestone in the field of e-commerce multimodal search domain.