iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.78673

Design a Smart Optimized Search Engine for E-Commerce Using U-Net Based Image Segmentation

Ms. Shriya Rane[1], Mr. Anshul Shembekar[2], Mr. Aryan Khedulkar[3], Mr. Raj Nakoriya[4], Asst. Prof. Ms. Saroj Shambharkar[5]
Department of Information Technology Kavikulguru Institute of Technology and Science, Ramtek, Nagpur, Maharashtra, India

***Abstract:** The fast development of e-commerce sites has driven up the need for smart and quick search tools. Keyword-based search systems have traditionally struggled to give correct results when consumers give visual cues or partial textual inquiries. To improve product search accuracy, this study presents a clever optimized search engine combining deep learning approaches, particularly U-Net-based image segmentation. The system searches a large dataset for visually similar goods after U-Net architecture segments relevant features from input images. The suggested strategy enhances user experience, eliminates uncertainty, and improves search relevance. Compared to conventional approaches, experimental findings reveal better retrieval efficiency and accuracy.*

***Keywords:** U-Net, Image Segmentation, E-commerce Search, Deep Learning, Computer Vision, Visual Search.*

I. INTRODUCTION

The fast expansion of online e-commerce has fundamentally changed consumer buying habits. Million of items in divisions including fashion, electronics, household appliances, and accessories are found on top platforms like Amazon and Flipkart. The efficiency of the search engine directly influences user happiness, customer retention, and general revenue generation in such big systems. Consequently, a good e-commerce platform must include an intelligent search tool as one of its main elements.

Keyword-based information retrieval methods form the basis of most conventional e-commerce search engines. Text similarity techniques in these systems help to match user searches with product names, tags, and descriptions. This strategy has several drawbacks, even if it works effectively for well specified text inquiries. In applications based on style, consumers frequently look for items depending on aesthetic features including color combinations, patterns, sleeve types, fabric textures, and general design. Textual metadata sometimes misrepresents these traits. Consequently, text-based retrieval systems could provide results that are either irrelevant or less visually similar.

Deep learning and computer vision have spurred a lot of interest in image-based search strategies. Extracting high-level visual features from product photographs, Convolutional Neural Networks (CNNs) are often utilized. Still, one of the biggest issues in visual retrieval systems is the existence of background clutter. E-commerce sites' product images often include shadows, mannequins, extra items, lighting variations, or intricate backgrounds. Extracting features from the whole image without focusing on the product can make it harder to find what you're looking for and make the results less accurate and less similar because of useless background noise.

Separation of the primary object from its background at the pixel level via image segmentation provides a good response to this issue. Olaf Ronneberger and his team created the U-Net architecture, a strong deep learning model meant for exact image segmentation. Its encoder–decoder architecture including skip connections allows for exact localization and thorough feature preservation. Because of its simplicity and great performance, U-Net has shown to be quite successful in a variety of object segmentation assignments.

Using U-Net segmentation prior to feature extraction can greatly improve the caliber of visual search in relation to fashion e-commerce. The system guarantees that only pertinent visual information helps to form feature representation by separating the product region. Improved similarity matching, superior retrieval accuracy, and less effect from background noise follow from this.

For similarity-based product retrieval, this study suggests a smart search engine design combining U-Net-based image segmentation with deep visual feature extraction. The main goal is to make e-commerce search systems better at understanding things visually and make the results they find more relevant. The suggested approach seeks to improve image-based search skills by means of a scalable and effective framework that modern online shopping platforms could use.

II. LITERATURE REVIEW

To develop an efficient image-based product search system, several studies have focused on image segmentation, feature extraction, and retrieval techniques using deep learning models.

Ronneberger et al. (2015) proposed the paper "U-Net: Convolutional Networks for Biomedical Image Segmentation," which introduced the U-Net architecture, a widely used model for image segmentation tasks. The architecture is based on an encoder–decoder structure, where the encoder captures contextual information through convolution and pooling operations, and the decoder performs upsampling to reconstruct the segmented image. A key innovation of U-Net is the use of skip connections, which directly transfer feature maps from the encoder to the decoder. This helps preserve spatial information and improves localization accuracy. The model has shown excellent performance even with limited training data, making it highly suitable for real-world applications. However, one limitation is that U-Net primarily focuses on local features and may struggle to capture global contextual relationships, which can affect segmentation quality in complex scenes.

Kumar et al. (2021) presented "Object-Based Image Retrieval Using the U-Net-Based Neural Network," where they combined image segmentation with content-based image retrieval (CBIR). In this approach, U-Net is first used to segment the object of interest from the background, which reduces noise and irrelevant information. After segmentation, wavelet-based feature extraction techniques are applied to extract important visual characteristics such as texture and structure. These features are then used for similarity matching in a retrieval system. The study demonstrated that segmentation significantly improves retrieval accuracy by focusing only on the relevant object. However, the limitation of this work lies in the use of handcrafted feature extraction methods, which are less powerful compared to modern deep learning-based feature extractors and may not generalize well to large-scale datasets.

Chen et al. (2021) introduced "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," which enhanced the traditional U-Net by integrating Transformer-based encoders. Unlike CNNs, which mainly capture local patterns, Transformers are capable of modeling long-range dependencies and global context within an image. In this hybrid architecture, the Transformer encoder extracts global features, while the U-Net decoder reconstructs the segmentation map with precise localization. This combination significantly improves segmentation performance, especially in complex images with varying structures. However, the model requires high computational resources and large datasets for training, making it less suitable for lightweight or real-time applications.

Simonyan and Zisserman (2014) proposed the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition (VGGNet)," which introduced a deep Convolutional Neural Network architecture for image classification and feature extraction. The model uses small convolution filters (3×3) stacked over multiple layers to learn complex visual patterns. One of the key contributions of this work is demonstrating that increasing network depth significantly improves feature representation. In the context of image retrieval systems, VGGNet is widely used as a feature extractor, where intermediate layers capture high-level features such as shape, texture, and object structure. These features can be converted into feature vectors and used for similarity matching. However, VGGNet has limitations, including high computational cost and large memory requirements, making it less efficient for real-time or large-scale applications.

He et al. (2016) introduced the paper "Deep Residual Learning for Image Recognition (ResNet)," which addressed the problem of training very deep neural networks. The key innovation of ResNet is the concept of residual learning using skip (identity) connections, which allows the network to learn residual mappings instead of direct mappings. This helps in overcoming issues like vanishing gradients and enables the training of extremely deep networks (e.g., 50, 101, or more layers). ResNet has been widely adopted for feature extraction in image retrieval systems due to its strong ability to learn rich and discriminative features. In e-commerce search engines, ResNet-based features improve similarity matching accuracy. However, despite its effectiveness, ResNet introduces increased model complexity and training time, and may require powerful hardware for optimal performance.

III. METHODOLOGY

Using U-Net-based image segmentation, the suggested system seeks to create a smart optimized search engine for e-commerce sites. The approach concentrates on increasing search accuracy by effectively matching pertinent visual cues from input images with a product database. The system precisely produces search results by combining picture processing, segmentation, feature extraction, and similarity matching.

A. System Overview

Starting with user input and ending with the retrieval of pertinent goods, the suggested system's whole flow comprises several steps. The user gives an input image that goes through several modules. The picture first goes through normalization and resizing during preprocessing. The processed image is then passed to the U-Net model for segmentation, which isolates the main object from the background. Meaningful characteristics are gathered after segmentation and contrasted with stored feature vectors in the database to find the most similar goods.

B. Image Preprocessing

To guarantee consistency and maximize model performance, image preprocessing is a vital phase. Input images are fixed to a size fit for the U-Net model in this step. Normalization is used to bring pixel values between 0 and 1 to scale. Furthermore, noise reduction techniques could be applied to eliminate undesirable changes in the image. This stage guarantees that the model obtains uniform input for dependable feature extraction and segmentation.

C. U-Net Based Segmentation

Because of its capacity to provide accurate pixel-level classification, U-Net is used as the main segmentation model. Skip connections abound in the encoder-decoder framework of the design. The encoder reduces the input image to record contextual information; the decoder then upsamples the segmented output to rebuild it. Skip connections combine encoder and decoder features from matching layers to help preserve spatial details. This facilitates correct identification of object boundaries as well as extraction of pertinent regions from the picture.

D. Feature Extraction

Following segmentation, feature extraction uses the pertinent part of the image. Deep learning-based feature extractors, such as pretrained convolutional neural networks, are used to create feature vectors that show how the product looks. Essential for similarity comparison, these elements record key characteristics including shape, texture, and color.

E. Similarity Matching

The feature vector taken from the query image is matched against database-stored feature vectors. Calculation of vector proximity is done using similarity measurement approaches such cosine similarity. The most relevant products—those with the greatest similarity ratings—are found in search results.

IV. SYSTEM ARCHITECTURE

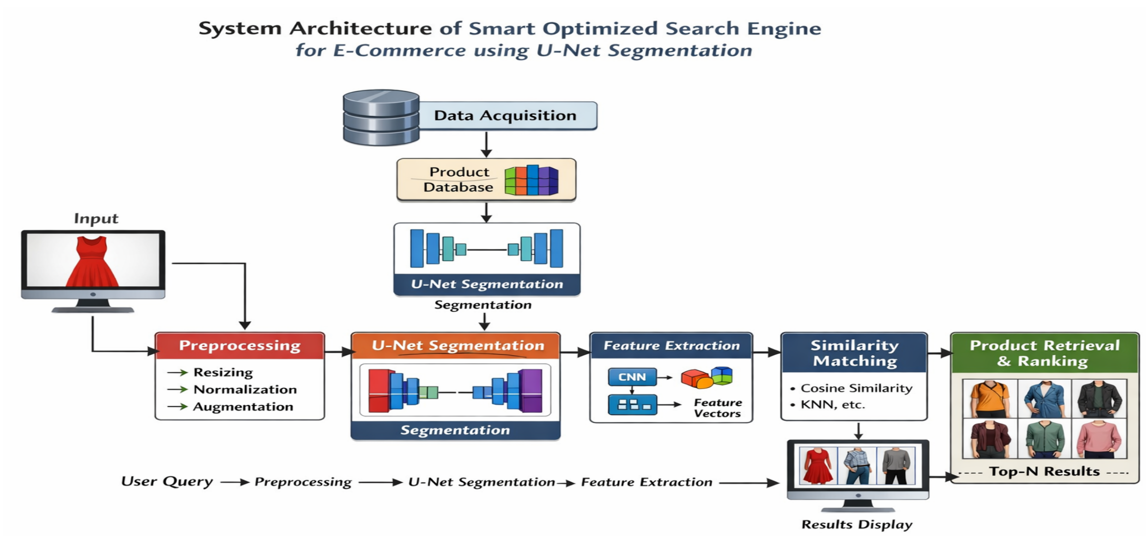

Fig. System Architecture of U-net segmentation

By combining image segmentation, feature extraction, and similarity matching methods, the suggested system aims to give e-commerce sites an intelligent and efficient product search engine. The architecture uses a modular design whereby every element has a particular role to play in the general workflow.

A. *Input Module*
The system takes user requests in two forms: text-based input and image-based input. The text-based method has users submit keywords characterizing the intended product; the image-based approach has users upload a product image. The system's adaptability and usefulness are improved by this multi-modal input feature.

B. *Preprocessing Tool*
The input image is treated to guarantee consistency and raise the performance of the model. Image scaling, normalization, and noise reduction are all parts of preprocessing. Before the segmentation model receives the input, these processes aid to standardize it.

C. *U-Net Segmentation Module*
The U-Net model is the main part of the system. It separates the product from the background. It follows a skip connections encoder–decoder design that keeps spatial information intact. By eliminating pointless background elements and concentrating just on the item of interest, this module enhances feature extraction and search result accuracy.

D. *Feature Extraction Module*
Important visual cues are pulled from the product image using Convolutional Neural Networks (CNNs) following segmentation. ResNet or VGG are pre-trained models that can be used to extract high-level features like shape, color, and texture. Efficient comparison then depends on converting these attributes into numerical feature vectors.

E. *Similarity Matching Module*
The similarity matching algorithm searches the database for feature vectors that resemble that of the query image. Measurement of similarity is achieved using distance metrics include cosine similarity or Euclidean distance. The system searches for the top-N most pertinent items depending on these computations.

F. *Recommendation Module*
Including a recommendation system into the design helps to improve user experience. Based on the user's inquiry, content-based filtering techniques propose comparable or related goods. This enables users to investigate more possibilities outside of the precise search results.

G. *Database Module*
Product images, derived feature vectors, and related metadata like product name, category, and pricing are all kept in the database. To guarantee quick and scalable data retrieval, effective indexing techniques are employed.

H. *Output Module*
The user sees the last results sorted. To guarantee a fast response time and better user satisfaction, the system shows the most pertinent goods together with suggested items.

I. *Overall Workflow*
Beginning with user input, the system moves through preprocessing, segmentation employing U-Net, feature extraction, similarity matching, recommendation generation, and then the results are shown. This organized pipeline guarantees fast and correct product retrieval.

V. RESULT AND DISCUSSION
The effectiveness of the suggested Smart Optimized Search Engine for E-commerce using U-Net Segmentation in visual product retrieval was judged using a fashion product image database. Several query images with different backgrounds, object alignments, and lighting situations were used to evaluate the system.

The findings show that separating the product from unrelated background information greatly enhances the quality of feature extraction by using U-Net segmentation. The suggested approach produced more relevant search results and better retrieval accuracy than conventional image retrieval techniques without segmentation.

The system was able to effectively find visually similar products in terms of hue, texture, and shape that were quite like the query image. For instance, the system returned similar jackets with similar designs and colors when a query image of a jacket was supplied; it also eliminated irrelevant matches.

Methods of ranking the retrieved results, quantitatively, including K-nearest neighbors and cosine similarity, were employed. Improved top-N results from the system showed that the most pertinent items were regularly rated higher. Preprocessing and segmentation also helped to lower noise, so boosting the model's resilience.

In situations with complicated backgrounds or overlapping objects, when segmentation was less exact, however, small mistakes were noted. The system performed well overall even with these difficulties, which shows how well it could work for real-world e-commerce.

VI. LIMITATION AND FUTURE WORK

A. Limitations

Though the suggested Smart Optimized Search Engine for E-commerce using U-Net Segmentation shows better visual product retrieval performance, some restrictions persist.

First of all, the system's performance is very closely related to the variety and quality of the training data. Not enough variation in the types of products, the lighting, or the backgrounds could make it harder for the model to work well with new data. Secondly, the computational complexity of deep learning models like U-Net and convolutional neural networks makes training take longer and requires more hardware, which can make it harder to use them in places where there aren't many resources.

Furthermore, segmentation mistakes can happen in difficult circumstances including images with many objects, occlusions, or congested backgrounds. These errors can spread to the feature extraction stage and lower retrieval performance. Moreover, the present approach mainly depends on visual cues and ignores textual data like product descriptions, reviews, or user preferences, hence reducing its contextual awareness.

Scalability is yet another restriction. The efficiency of similarity matching approaches may suffer as the product database grows, especially if using standard distance-based methods without tailored indexing. The system's capacity to grow over time is limited as it lacks a dynamic learning mechanism and fails to change in response to user interactions or input.

B. Future Work

A major addition is the incorporation of a recommendation mechanism going beyond simple similarity matching. Using sophisticated embedding models like CLIP, the system can find correlations between various product categories and recommend complementary goods. For example, the system can suggest similar things like shoes or accessories, which helps with user engagement and supports cross-selling strategies. Besides finding a requested product like a black leather jacket, it can recommend things like shoes or accessories.

Including multilingual search skills is yet another crucial direction. Allowing users to enter searches in Hindi or Hinglish can greatly boost accessibility. Using multilingual embedding models or including translation-based techniques helps the system to accurately understand and process queries across several languages.

Including metadata-driven filtering and sorting algorithms will help to improve the system. Features including price range filtering, rating-based ranking, and product availability can help to turn the prototype into a workable e-commerce platform. This calls for fusing organized product metadata with visual similarity results to tailor and customize search results.

Moreover, a search analytics dashboard may offer insightful information on user contacts and system behavior. Whereas tools like word clouds and query frequency analysis can highlight common search patterns, visualization methods such t-SNE or PCA help to show high-dimensional feature spaces. Such studies not only clarify the system but also support ongoing improvement.

To better feature representation and retrieval accuracy, future studies may also look into the use of cutting-edge deep learning techniques including Vision Transformers and attention-based models. Furthermore improving scalability and real-time performance is the use of effective indexing methods like Approximate Nearest Neighbor search.

Including user comments systems and personalization techniques will ultimately help the system to learn actively and adjust to each user's preferences. Integrating with full-stack web or mobile applications and deployment on cloud-based platforms will help to improve scalability, usability, and real-world utility.

VII. CONCLUSION

By combining deep learning and computer vision methods, the Smart Optimized Search Engine for E-commerce employing U-Net Segmentation offers an innovative way to raise product search accuracy. The suggested approach uses image-based retrieval with semantic knowledge of visual material, which produces more relevant and accurate product recommendations than conventional keyword-based search tools.

Isolating the product from complicated backgrounds using the U-Net segmentation model improves the quality of feature extraction by itself. Improved similarity matching and retrieval performance result from this. The system finds visually similar goods from a sizable database using deep feature representations and effective similarity measures.

The suggested design not only boosts search efficiency but also improves user experience by providing correct and context-aware results, according to experimental data and system design. The design of the system's modules guarantees real-world e-commerce applications to be both scalable and flexible.

Finally, this study emphasizes how well feature-based retrieval systems might be used in conjunction with segmentation-based preprocessing. To further raise performance and scalability, future upgrades may include multimodal search (text + image), real-time recommendation systems, and optimization employing cutting-edge models like transformers.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Proc. MICCAI, 2015, pp. 234–241.
[2] A. Kumar, S. Sharma, and R. Gupta, “Object-Based Image Retrieval Using the U-Net-Based Neural Network,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 6, pp. 123–130, 2021.
[3] J. Chen, Y. Lu, Q. Yu, et al., “TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation,” arXiv:2102.04306, 2021.
[4] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv:1409.1556, 2014.
[5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in Proc. CVPR, 2016, pp. 770–778.
[6] O. Russakovsky et al., “ImageNet Large Scale Visual Recognition Challenge,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015.
[7] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Proc. NIPS, 2012, pp. 1097–1105.
[8] C. Szegedy et al., “Going Deeper with Convolutions,” in Proc. CVPR, 2015, pp. 1–9.
[9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Weinberger, “Densely Connected Convolutional Networks,” in Proc. CVPR, 2017, pp. 4700–4708.
[10] M. Sandler et al., “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in Proc. CVPR, 2018, pp. 4510–4520.
[11] A. Vaswani et al., “Attention Is All You Need,” in Proc. NIPS, 2017, pp. 5998–6008.
[12] Z. Liu et al., “Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows,” in Proc. ICCV, 2021, pp. 10012–10022.
[13] H. Jegou, M. Douze, and C. Schmid, “Product Quantization for Nearest Neighbor Search,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 1, pp. 117–128, 2011.
[14] Y. Lecun, Y. Bengio, and G. Hinton, “Deep Learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[15] R. Datta, D. Joshi, J. Li, and J. Wang, “Image Retrieval: Ideas, Influences, and Trends of the New Age,” ACM Comput. Surveys, vol. 40, no. 2, pp. 1–60, 2008.