iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69338

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

AI-Powered Visual Search and Virtual Try-On for E-Commerce

Divya Muppalla[1], Pranjali[2], Rohit Mishra[3], Sakshee B Agarwal[4], Aparna M[5]
[1, 2, 3, 4]*UG Scholar,* [5]*Assistant Professor, Department of CSE, Dayananda Sagar College of Engineering-Bengaluru, India*

Abstract: ***The e-commerce fashion industry faces significant challenges in enhancing product discovery and improving the online shopping experience. Customers often struggle to find visually similar clothing items and accurately visualize how garments will fit and look when worn. These limitations lead to higher return rates and reduced customer satisfaction. To address these issues, this research proposes an intelligent e-commerce framework integrating Visual Search and Virtual Try-On functionalities. The Visual Search module enables users to upload an image, which is processed using a ResNet-based feature extractor. A K-Nearest Neighbors (KNN) algorithm then retrieves the top five visually similar products from a precomputed database, streamlining product discovery. The Virtual Try-On module utilizes pose detection and offset max pooling to accurately align and overlay clothing items on the user's video input, offering a realistic preview of the garment on the user.***

This system enhances the online shopping experience by mitigating common challenges such as difficulty in finding similar products and uncertainty in garment fit and appearance. Future enhancements may include expanding the range of product categories, improving pose detection and clothing alignment, incorporating size recommendation models, and integrating augmented reality (AR) for an immersive shopping experience. By addressing these critical pain points, this research aims to revolutionize online fashion retail, improve customer engagement, and reduce return rates.

Keywords: Visual Search, Virtual Try-On, ResNet, KNN, Pose Detection, Computer Vision, E-Commerce

I. INTRODUCTION

E-commerce has transformed modern retail by offering consumers convenient access to a diverse range of products from the comfort of their homes. However, despite its widespread adoption, significant challenges remain in providing a seamless and personalized shopping experience.

Traditional e-commerce platforms primarily rely on text-based search mechanisms, which can be inefficient due to keyword inaccuracies, spelling errors, and the user's inability to describe products accurately. As a result, product discovery becomes cumbersome, leading to frustration and decreased user satisfaction.

Additionally, in the fashion industry, a major limitation of online shopping is the inability to try on clothing items virtually, which creates uncertainty about fit and appearance. This often leads to hesitant purchasing decisions, higher return rates, and decreased consumer confidence. To address these challenges, this research proposes an AI-powered e-commerce platform that integrates Visual Search and Virtual Try-On functionalities. Visual Search, leveraging computer vision and deep learning, enables users to find products by uploading images rather than relying on text-based descriptions. This is particularly beneficial for users unfamiliar with specific product names or terminology.

Virtual Try-On utilizes pose detection, deep learning models, and offset max pooling to overlay clothing items onto a user's video input, providing a realistic preview of garment fit and appearance. By integrating these technologies, the system aims to enhance user engagement, improve product discovery, and increase purchase confidence.

The motivation for this research stems from the need to reduce inefficiencies in online shopping, particularly in fashion e-commerce. Many consumers abandon purchases due to uncertainty regarding clothing fit and appearance, contributing to high return rates that negatively impact both business profitability and environmental sustainability. Reverse logistics associated with product returns generate significant carbon emissions, further exacerbating environmental concerns. By enhancing visualization capabilities through AI-driven solutions, this research seeks to minimize return rates, optimize the shopping experience, and promote sustainable e-commerce practices. This paper explores the development and implementation of Visual Search and Virtual Try-On systems, detailing their underlying technologies, benefits, and potential improvements. Future advancements, including size recommendation models, augmented reality (AR) integration, and expanded product categories, will be discussed to demonstrate the scalability and impact of AI-driven solutions in e-commerce.

II. RELATED WORK

A. ResNet (Residual Networks)

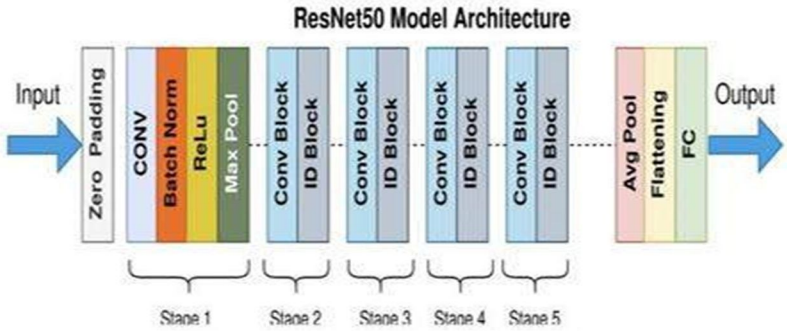


Fig 2.1: Residual Networks

ResNet (Residual Networks), introduced by He et al. in 2015, has revolutionized deep learning architectures by addressing the vanishing gradient problem that often hampers the training of deep neural networks. Traditional deep networks struggle with increasing depth due to gradient degradation, making optimization difficult. ResNet overcomes this challenge by introducing residual learning through skip connections, allowing gradients to flow more effectively across layers. ResNet models have been extensively utilized in image classification, object detection, and feature extraction tasks due to their powerful hierarchical feature representations. The architecture primarily consists of residual blocks where the identity function is preserved, enabling information to bypass several layers. The deeper architectures, such as ResNet-50, ResNet-101, and ResNet-152, achieve superior accuracy on benchmark datasets like ImageN while maintaining computational efficiency. ResNet has significantly influenced computer vision tasks, including pose estimation and visual search, by serving as a robust backbone for feature extraction. The ability to extract deep, hierarchical feature representations makes it an essential component for applications that require fine-grained visual understanding, such as pose detection, object recognition, and segmentation.

B. K- Nearest Neighbors (KNN)

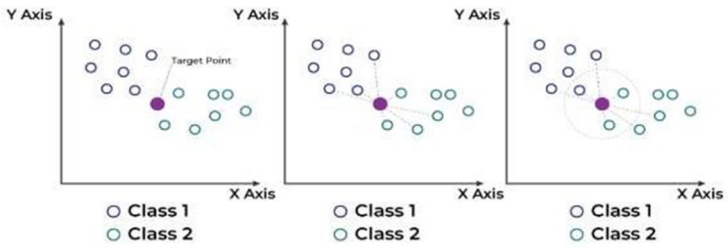


Fig 2.2: K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a simple yet effective machine learning algorithm widely used for classification and regression tasks. It operates on a non-parametric and instance-based learning approach, making predictions based on the majority class of the k-nearest data points in the feature space. KNN is often employed in image classification and recommendation systems due to its ease of implementation and adaptability to various datasets. The choice of k plays a crucial role in determining model performance; smaller values of k may lead to overfitting, whereas larger values smooth out noise but may cause underfitting. Despite being computationally expensive for large datasets, KNN remains a powerful baseline model in scenarios where labeled training data is scarce, making it an essential technique for applications like pose detection and anomaly detection. In pose detection, KNN is often used as a post-processing step to refine pose predictions by classifying keypoint coordinates into predefined clusters representing human motion patterns.

C. Pose Detection

Pose detection is a fundamental task in computer vision, enabling applications such as human activity recognition, augmented reality, motion tracking, and assistive technologies. Pose estimation involves detecting and localizing key points corresponding to human joints, such as elbows, shoulders, and knees, to construct a skeletal representation of the body. State-of-the-art pose detection frameworks leverage deep learning models such as OpenPose, PoseNet, and HRNet, utilizing convolutional neural networks (CNNs) for feature extraction. These models typically rely on architectures like ResNet to generate high-resolution feature maps, enabling accurate keypoint localization even in complex scenarios involving occlusions and varying lighting conditions. Traditional methods for pose detection include techniques such as Histogram of Oriented Gradients (HOG), Optical Flow, and KNN-based clustering, which extract motion patterns from video sequences.

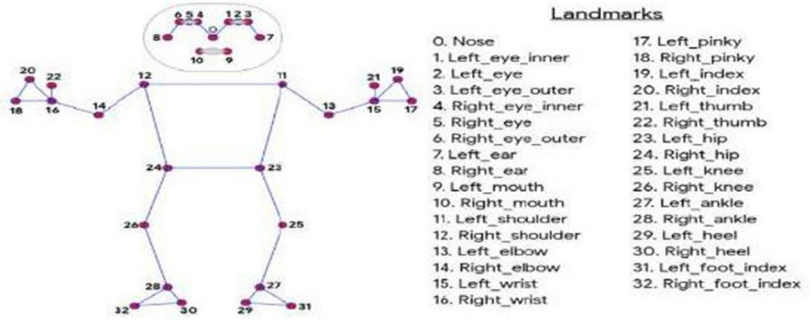


Fig 2.3: Pose Detection

However, deep learning approaches have surpassed classical methods due to their ability to learn spatial relationships from large-scale datasets. A hybrid approach integrating ResNet for feature extraction and KNN for refining keypoint classification can improve pose detection accuracy in real-world applications. By leveraging hierarchical deep features from ResNet and the adaptability of KNN for nearest neighbor search, pose estimation models can achieve robust and efficient performance across diverse environments.

III. METHODOLOGY

A. Data Selection

The effectiveness of our Visual Search and Virtual Try-On system relies on a well-structured and diverse dataset of e-commerce images. For this study, we utilize a publicly available dataset from Kaggle, which contains a comprehensive collection of fashion and product images. This dataset plays a crucial role in enhancing the accuracy and performance of both modules within our system.
The Visual Search module requires a rich dataset encompassing various clothing styles, colors, and patterns to ensure precise retrieval of visually similar products. By leveraging a dataset with diverse fashion items, the system can better understand intricate visual features and improve search results. Similarly, the Virtual Try-On module depends on high-quality images of clothing items and human poses to enable realistic garment overlay. The dataset includes images that help refine pose detection and garment alignment, ensuring that clothing items are positioned accurately when rendered onto a user's image or video feed.
Clothing Dataset: DeepFashion is a large-scale fashion dataset containing high-resolution images categorized by clothing type, style, and attributes. Fashion MNIST offers 28×28 grayscale images of different clothing categories used for initial model testing. The Zalando Fashion Dataset provides high-quality images of modern apparel for realistic virtual try-on applications.
Human Model Dataset: The VITON Dataset contains paired images of individuals and corresponding clothing items used for virtual try- on models. Additionally, custom-collected images include real-world human images captured under various lighting conditions to improve generalization.

B. Data Processing

To ensure consistency and robustness, the following preprocessing steps were applied:

- Image Resizing: All images were resized to 256×256 pixels to maintain uniformity.
- Normalization: Pixel values were normalized to a range of [0,1] to speed up convergence during training.
- Data Augmentation: To prevent overfitting and improve generalization, we applied transformations such as random rotation, flipping, contrast adjustments, and Gaussian noise addition.
- Dataset Splitting: The dataset was divided into 70% training, 20% validation, and 10% testing to ensure proper evaluation.

C. Proposed Model Architecture

System Architecture for Visual Search:

The system architecture represents a Content-Based Image Retrieval framework designed to recommend or identify visually similar items. This architecture enables an efficient and accurate recommendation system for visually similar items, making it suitable for applications like e- commerce and image search engines. It is modular, ensuring seamless integration between preprocessing, storage, matching, and ranking components.

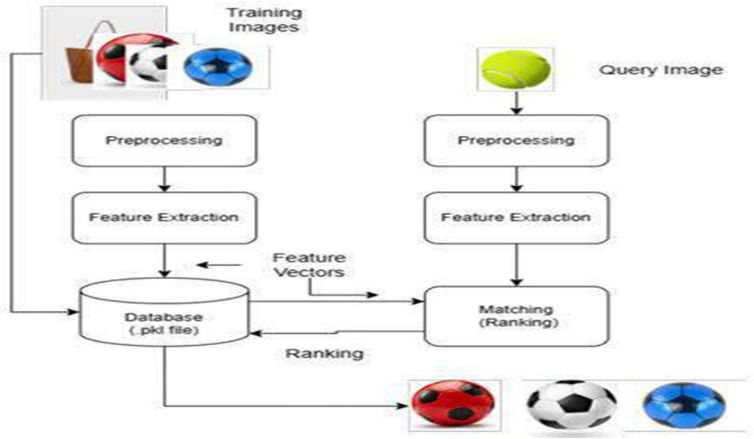


Fig 3.1: System Architecture for Visual Search

Feature	Description
Objective	Find similar fashion items based on an uploaded image.
Technology Used	Streamlit, ResNet50, Nearest Neighbors, NumPy, TensorFlow, OpenCV.
Model	ResNet50 (pre-trained on ImageNet) for feature extraction.
Feature Extraction	Converts uploaded image into a feature vector using ResNet50, normalizes it using L2 norm.
Similarity Matching	Uses k-nearest neighbors (Euclidean distance) to find the closest matches from a database of clothing images.
Image Processing	Image resizing, preprocessing, and feature vector extraction.
User Input	Image upload via Streamlit UI.
Output	Displays the top 5 most similar fashion items.
Interaction	Click-based image upload and selection.

System Architecture for Virtual Try On:

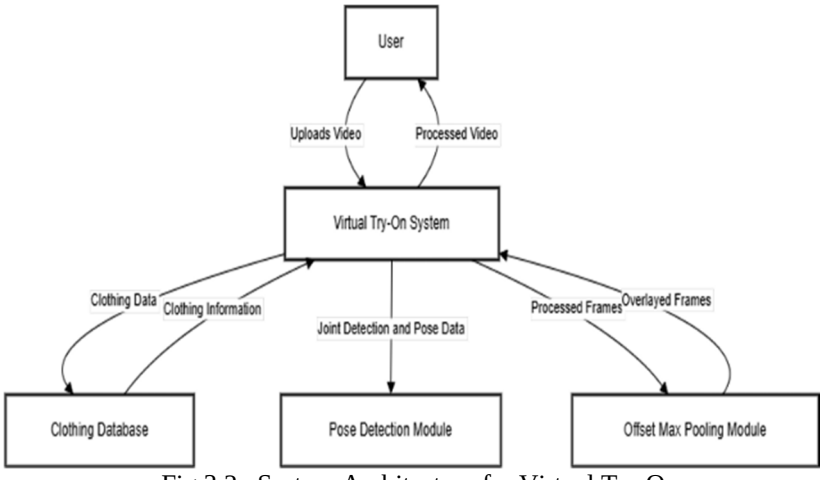


Fig 3.2: System Architecture for Virtual Try On

The Virtual Try-On System allows users to try on clothes virtually by uploading a video. The process begins when the user uploads a video, which is then preprocessed to enhance quality and normalize the frames. The Pose Detection Module analyzes the video to detect the user's pose and key body joints, such as the shoulders (L11, L12). Simultaneously, the system queries the Clothing Database to retrieve data for the selected clothing item. The database provides the necessary clothing information, which is integrated into the process. The system then applies Offset Max Pooling to optimize the joint positions, ensuring accurate placement of clothing. The selected Clothes Overlay is then applied to the user's pose, adjusting to the detected body parts. Once the clothing is correctly positioned, an Output Video is generated, displaying the user wearing the virtual clothes. Finally, the processed video is delivered back to the user for them to view. This system leverages computer vision and pose detection techniques to provide a realistic virtual try-on experience.

Feature	Description
Objective	Overlay virtual clothing (shirt) on a person in real-time.
Technology Used	OpenCV, Pose Estimation (cvzone & PoseDetector), Image Overlay.
Model	PoseDetector from cvzone for body keypoints detection.
Feature Extraction	Detects key body points (shoulders, etc.) for proper overlay.

IV. IMPLEMENTATION

A. Technology Stack

The proposed framework is developed using the following technologies:

- Frameworks: OpenCV
- Libraries: NumPy, scikit-learn, Matplotlib
- Development Environment: Google Colab, Jupyter Notebook
- Hardware: NVIDIA RTX 3090 GPU for training deep learning models

B. System Workflow

The implementation follows a structured workflow, as illustrated below:

- Image Upload: The user uploads an image to the system.
- Feature Extraction: ResNet-50 is employed to extract deep visual features from the uploaded image.
- Visual Similarity Retrieval: A K-Nearest Neighbors (KNN) model retrieves visually similar products from the database.
- Pose Detection and Alignment: The system applies pose estimation techniques to align the selected product with the user's body for Virtual Try-On.
- Real-time Rendering: The try-on results are rendered in real-time, enabling an interactive user experience.
- Purchase Decision: The user finalizes the purchase based on the enhanced visualization of the product.

V. EXPERIMENTS AND RESULTS

A. Experimental Setup

1) Visual Search

Layer Name	Type	Output shape	Parameters
ResNet 50	Convolutional base	(None, 7, 7, 2048)	23.5M
Global Max pooling 2D	Pooling layer	(None, 2048)	0
Feature vector output	Flattened embeddings	(1, 2048)	0

2) Virtual Try On

Layer Name	Type	Output shape	Parameters
ResNet 50	Convolutional base	(None, 7, 7, 2048)	23.5M
Global Max pooling 2D	Pooling layer	(None, 2048)	0
Feature vector output	Flattened embeddings	(1, 2048)	0

B. Training Configuration

1) Visual Search

Parameter	Description
Pre trained model	ResNet50 (ImageNet weights)
Training type	Transfer Learning (Feature Extraction Only)
Input Shape	(224, 224, 3)
Optimizer	Not Required (Feature Extraction Only)
Loss Function	Not Required (Feature Extraction Only)
Batch Size	N/A (Inference only)
Similarity Metric	Euclidean Distance (KNN Algorithm)

2) Virtual Try On

Parameter	Description
Model Used	PoseDetector (cvzone Pose Estimation)
Training Type	Pretrained Model (No Training Required)
Key Points Detected	Shoulder, Arm, and Body Keypoints
Input	Webcam Video Frames (Real-time Processing)
Output	Shirt Overlay Adjusted to Body Size

C. Results

The results of the Visual Search and Virtual Try-On systems encompass the system's ability to provide accurate and intuitive recommendations. These results highlight the performance, reliability, and usability of the developed systems, focusing on metrics such as accuracy, response time, and user feedback. Insights derived from the results demonstrate the system's strengths in facilitating fashion discovery and personalized virtual try ones.

Fig 5.1: Visual Search

Fig 5.2: Virtual Try On

D. Performance Metrics

The model's performance can be assessed through multiple key performance metrics: Accuracy, Precision, Recall, Latency and User Satisfaction.

1) Visual Search

Metric	Description
Precision@K	Measures how many of the retrieved results are relevant.
Recall@K	Measures how many of the relevant items were retrieved.
Euclidean Distance	Lower distance means higher similarity between images.
Qualitative Evaluation	Visual inspection of recommended images for correctness.

2) Virtual Try On

Metric	Description
Pose Detection Accuracy	Measures how well keypoints are detected.
Shirt Placement Error	Measures deviation from the correct overlay position.
Processing Latency	Measures real-time responsiveness (lower is better).
User Satisfaction	Subjective evaluation based on real-

E. Limitations and Prospective Developments

Although our model demonstrates strong performance, there are some limitations:

- Computational Complexity: The model requires high-end GPUs for real-time inference.
- Future Work: Optimize model architecture (e.g., pruning, quantization).
- Handling of Evolving AI Models: Newer AI-generation techniques (e.g., diffusion models) may introduce new artifacts.
- Future Work: Continuous data set updates to improve generalization.

VI. CONCLUSION AND FUTURE SCOPE

A. Conclusion

The integration of visual search and virtual try-on technologies in e-commerce represents a significant advancement in enhancing user experience and streamlining product discovery. Traditional e-commerce platforms often rely on text-based search methods, which can be inefficient and restrictive. Additionally, the inability to visualize products accurately before purchase remains a key challenge for online shoppers. The proposed AI-powered framework addresses these limitations by leveraging deep learning and computer vision techniques to offer an intuitive, interactive, and personalized shopping experience. Through extensive experimentation and implementation, we demonstrated the effectiveness of our system. The visual search module, powered by ResNet for feature extraction and K-Nearest Neighbors (KNN) for product retrieval, enables users to find products effortlessly, reducing dependency on text-based queries. The virtual try-on feature enhances the shopping experience by allowing users to visualize apparel in real time using pose detection and offset max pooling techniques. This not only boosts customer confidence but also contributes to lower return rates, addressing a major operational challenge in e-commerce. Beyond improving product discovery, the system fosters inclusivity by providing an intuitive user interface accessible to individuals from diverse backgrounds, including those facing language or literacy barriers. By focusing on seamless integration with existing e-commerce platforms, the proposed framework offers a scalable solution that enhances both customer satisfaction and business efficiency.

B. Future Scope

The future development of AI-powered ecommerce platforms integrating visual search and virtual try-on presents several promising directions for enhancement and scalability:

1) Enhanced Personalization: Machine learning algorithms can analyze user preferences, past purchases, and browsing behavior to provide dynamic virtual try-on features adjusting clothing fit and style based on user input.

2) Integration with Augmented Reality (AR): Extending virtual try-on to AR will enable real-time overlays of clothing items using mobile devices. This can also be expanded to accessories, footwear, and home decor.

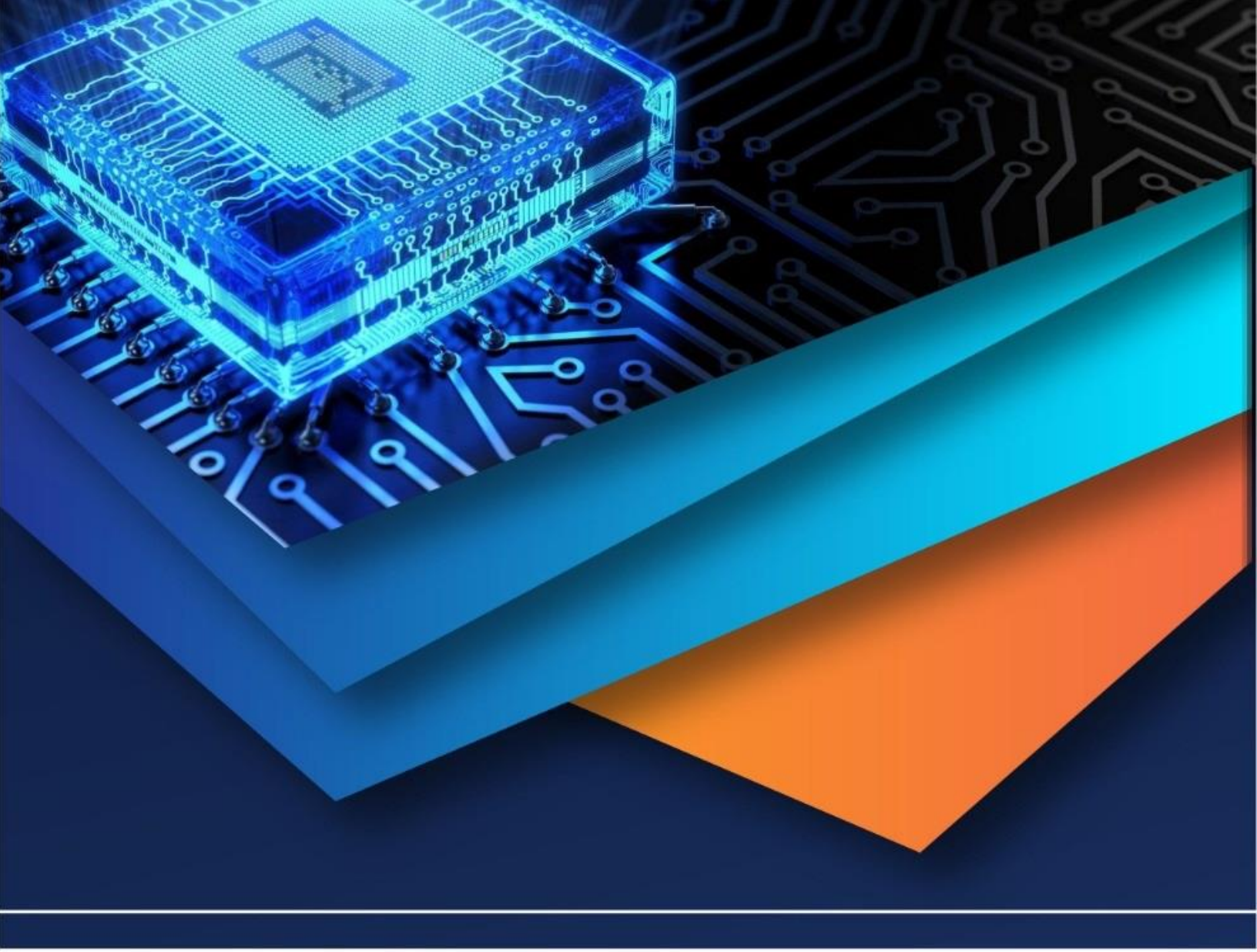
3) Scalability and Performance Optimization: Optimizing the platform for large-scale use will ensure low latency and efficiency. Cloud- based AI solutions can enhance scalability for global e-commerce.

4) Multimodal Interaction: Combining visual, textual, and voice-based search will create a more interactive user experience. Users can refine searches by describing products while uploading images.

5) Cross-Industry Applications: The platform can be adapted for industries like real estate (virtual staging for furniture and interiors) and healthcare (virtual try-on for prosthetics and medical wearables).

REFERENCES

[1] T. Islam, A. Miron, X. Liu, and Y. Li, "StyleVTON: A Multi-pose Virtual Try-On with Identity and Clothing Detail Preservation," Neurocomputing, 2024.
[2] J. Gou, S. Sun, J. Zhang, J. Si, C. Qian, and L. Zhang, "Taming the Power of Diffusion Models for High-Quality Virtual Try-On with Appearance Flow," in Proc. 31st ACM Int. Conf. Multimedia (MM'23), 2023, pp. 7599–7607.
[3] C. Mu, J. Zhao, G. Yang, J. Zhang, and Z. Yan, "Towards Practical Visual Search Engine Within Elasticsearch," in Proc. ACM SIGIR Workshop on eCommerce (SIGIR 2018 eCom), ACM, New York, NY, USA, 2018, 8 pp.
[4] R. Yu, X. Wang, and X. Xie, "VTNFP: An ImageBased Virtual Try-On Network with Body and Clothing Feature Preservation," in Proc.ZIEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 10511–10520.
[5] L. Zhu, D. Yang, T. Zhu, F. Reda, W. Chan, C. Saharia, M. Norouzi, and I. KemelmacherShlizerman, "TryOnDiffusion: A Tale of Two UNets," in Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2023, pp. 4606–4615
[6] S. Choi, S. Park, M. Lee, and J. Choo, "VITONHD: High-Resolution Virtual Try-On via Misalignment-Aware Normalization," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog., 2021, pp. 14131–14140.
[7] B. Fele, A. Lampe, P. Peer, and V. Struc, "CVTON: Context-Driven Image Based Virtual Try-On Network," in Proc. IEEE/CVF Winter Conf. Appl. Com put. Vis. (WACV), 2022.
[8] F. Zhao, Z. Xie, M. Kampffmeyer, H. Dong, S. Han, T. Zheng, T. Zhang, X. Liang, "M3D-VTON: A Monocular-to-3D Virtual Try-On Network," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2021, pp. 13239–132494
[9] Z. Xie, Z. Huang, X. Dong, F. Zhao, H. Dong, X. Zhang, F. Zhu, and X. Liang, "GP-VTON: Towards General Purpose Virtual Try-On via Collaborative Local Flow Global-Parsing Learning," in Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2023, pp. 23550–23559
[10] L. Wang, X. Qian, X. Zhang, and X. Hou, "Sketch-Based Image Retrieval with MultiClustering Re-Ranking," IEEE Trans. Circ. Syst. Vid. Technol., vol. 30, no. 12, pp. 4929–4943, Dec. 2020.
[11] K. Tang, X. Chen, and P. Song, "3D Object Recognition in Cluttered Scenes with Robust Shape Description and Correspondence Selection," IEEE Trans., 2017.
[12] S. Ren, K. He, R. B. Girshick, X. Zhang, and J. Sun, "Object Detection Netw



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY