World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Research Article)

Check for updates

Style Matcher: A deep learning framework for visual fashion matching using ResNet50

Sai Tharun Jami *, Ganesh Reddy Karri, Habeeb Ur Rahman, Surya Saharsha Merupo and Vikram Sri Durganand Batchu

School of Computer Science and Engineering, VIT-AP University Amaravati, India.

World Journal of Advanced Engineering Technology and Sciences, 2025, 15(01), 1710-1721

Publication history: Received on 06 March 2025; revised on 17 April 2025; accepted on 19 April 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.15.1.0390

Abstract

During present times when digital transformation and personalization shape the market the fashion industry actively seeks powerful solutions for pushing user interest and style production. This research presents "Style Matcher" which operates as a fashion recommendation system through CNNs for generating tailored suggestions of clothing items and accessories. Users can input preferences about t shirts and glasses then the system retrieves fashion items with visually matching characteristics. The proposed system uses three core components which include data processing followed by CNN model training and recommendation score generation. The extracted features from fashion images by CNN models facilitate quick style matching operations. The model uses complex fashion item data to learn detailed style patterns while identifying visually similar fashion products. Through the new ranking system the system can compute precise and appropriate style recommendations by evaluating user preference similarities to suggested items. (Abstract)

Keywords: Fashion recommendation; Convolutional Neural Networks; Visual Similarity Matching; Personalized style; Visual similarity; User engagement

1. Introduction

Rapid industry developments together with rising online shopping popularity created new difficulties which consumers and businesses now face. Users encounter a primary challenge in their attempt to find fashion products that appear visually similar. The traditional use of words as product description methods proves inadequate to fully represent distinctive fashion elements which define individual items. As a result users cannot discover similar items either online or in stores when having inaccurate product descriptions. A documented strategy for connecting vision-based consumption to discovery of available products leads to an organizational requirement for emerging solutions. Solving this issue leads to better user satisfaction and generates substantial impacts on the fashion industry. An e-commerce platform becomes more engaging and converts customers more efficiently after the integration of a fashion recommendation system delivers smooth performance.

Users who access products according to their visual preferences through the system will increase business customer numbers as well as build competitive benefits for companies. Our research stands out because it represents a novel user technology model for digital fashion product interactions. The existing product search and discovery approaches become inadequate because verbal descriptions naturally carry subjective information. The proposed fashion recommendation tool powered by CNN enables users to search for clothing by visual style features then recommends products based on visual cues. Users gain effortless access to various fashion choices that align with their personal preferences because this technology connects perception to action. The applications stemming from our system reach beyond basic fashion suggestions into wider areas.

* Corresponding author: Sai Tharun Jami

Copyright © 2025 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

Figure 1 User Input and Recommendation

By leveraging the capabilities of the CNN-based By developing this method we create a system which can adapt to different recommendation settings. By merging customer purchase data with our model we create personalized recommendations which boost e-commerce cross-sales and user platform usage thus achieving the research focus on creating an efficient fashion recommendation system with CNN functionality. As part of our approach we use ResNet50 architecture to retrieve useful style features from fashion product images.

Our focus centers on building a system which offers visually comparable product recommendations while strengthening recommendations within consolidated systems. Our contributions cover multiple aspects. The first objective involves proving CNN's applicability for fashion recommendations while showing its capability to unite user preferences with visually-comparable products. The system addresses performance needs while generating product discovery demands which require new solutions. The solution of this problem improves both fashion user engagement and has major operational benefits for businesses in fashion.

A fashion recommendation system integrated smoothly into e-commerce systems enhances customer activity by producing higher shopping conversion rates. The system allows businesses to assist customers in discovering related items through visual matching which results as shown in Fig 1. in bringing more clients while gaining market dominance. Our system produces extensive effects which expand past fashion recommendation technologies. Our foundation uses CNN-based capabilities to develop a general system which could serve multiple recommendation purposes. Our model can provide customized product recommendations through purchase history integration because we have implemented Python pickle files to run offline using accessible dataset information. This approach enables the identification of hidden patterns for building additional recommendation systems that exploit extracted features. This paper examines the CNN-based fashion recommendation system architecture as well as methodology alongside experimental findings and implications. The research objective includes providing fashion product discovery solutions at present but also lays the groundwork for superior personalization and broader recommendation systems engagement in the future.

2. Literature Review

Multiple research studies have examined various innovative methods to close the user preference gap against fashion product needs in the fast-growing fashion recommendation space.

The fashion recommendation landscape has evolved from basic collaborative filtering (CF) and content-based filtering (CBF) to hybrid models that address the limitations of traditional techniques. CF methods rely on collective user behavior to suggest items, while CBF leverages item attributes and user profiles. Studies such as those by Venugeetha Y. [1][6] and Ravindra & Patil [2][7] explored the use of deep learning models to improve classification and recommendation accuracy in fashion systems. These approaches paved the way for systems that utilize user interaction and visual content to personalize suggestions.

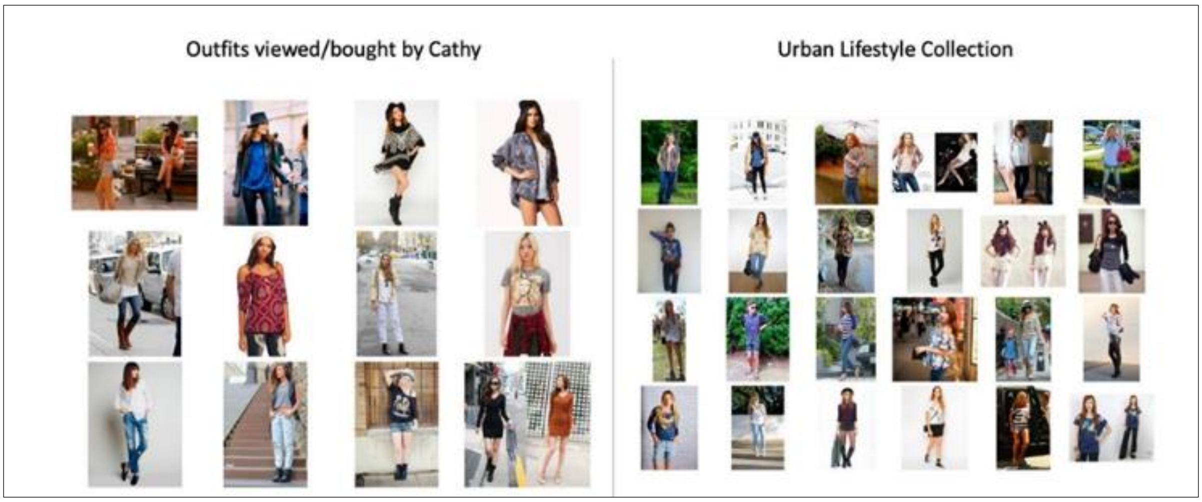


Figure 2 Labelling of different classes

The emergence of deep convolutional neural networks (CNNs) like ResNet50 has transformed visual similarity-based recommendation systems. Notably, the DeepFashion dataset by Li et al. [3][8] and the extended work by Liu et al. [11][21] provided rich annotations that empowered robust clothing recognition and retrieval. These datasets enabled the training and evaluation of CNN models for style matching and visual similarity. Studies like those by Veit et al. [13] and McAuley et al. [14] demonstrated the value of learning visual clothing styles through deep learning techniques, further showing the potential of image-based recommendations fashion.

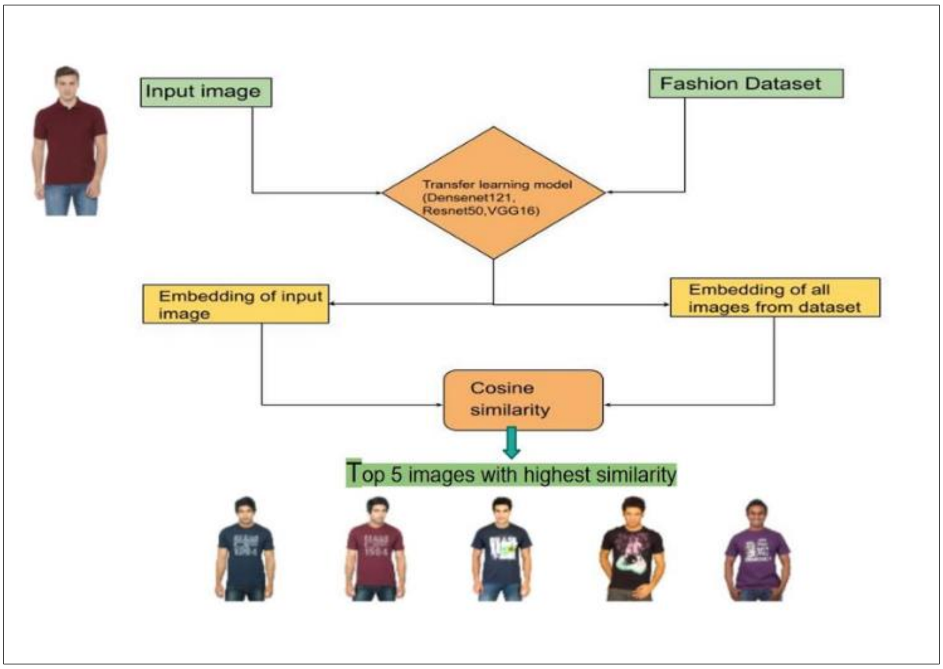


Figure 3 Fashion Recommendation Architecture

In addition to visual content, social interaction has also been leveraged for fashion recommendation. Yan et al. [4][9] examined how social e-commerce platforms could enhance recommendations through user behavior and peer influence. Foundational work in large-scale image recognition, such as the ImageNet dataset by Deng et al. [5][10], supported transfer learning for CNNs used in fashion contexts. Other prominent CNN architectures like VGG [16] and

Inception [17] have also been applied in fashion systems for effective visual feature extraction, with multiple works comparing their performance in identifying relevant fashion features.

More recent research emphasizes building scalable and end-to-end fashion recommendation pipelines. Elsayed et al. [12] proposed a full image-based system that integrates feature extraction and ranking. Additional datasets like the H&M Personalized Recommendations [15], Fashion Recommender Dataset by Mishra [18], and Pinterest Fashion Dataset by Kolhe [19] provided real-world e-commerce data to test recommendation engines. Agrawal and Garg [20] further explored deep learning-based personalized systems, focusing on enhancing user satisfaction through tailored suggestions. Collectively, these studies laid the foundation for advanced systems like Style Matcher, which combines CNN-based visual similarity, user interaction, and scalable architecture for efficient fashion product recommendations..

3. Dataset Description

The application and assessment of our fashion recommendation system used extensive transaction data provided by H&M that demonstrates actual e-commerce shopping patterns. The database suits tasks requiring image recommendations and metadata processing through its complete retail transactions and extensive product and user details.

The essential part of the database consists of transactions_train.csv which demonstrates every individual customer purchase activity across various dates. A different transaction appears in each row of this file which includes customer_id alongside article_id and date. Two identical records for one product item reveal the quantity of purchases which gives data about user frequency choices. The time-based organization of the data lets the system detect shifting patterns within fashion buying activities including seasonal buying habits and evolving fashion preferences.

Each article available in the store is fully described in the articles.csv data file. The system needs these features to create visual or metadata embeddings for subsequent similarity assessment and personalization methods.

The customers.csv file offers key features which include customer age information together with postal code and membership details. The additional data creates opportunities for custom recommendations when used in multi-featured recommendation systems where visual similarity joins with individual profiles.

The available dataset contains an images/ subfolder which includes product images for most of the article_ids. The product images reside in separate folders corresponding to article ID initial sequences which contain three digits. The existing visual data enables the design of CNN-based models for deep feature extraction despite the fact that all article entries do not have accompanying images. Our system uses these visible features to create the base for identifying similar products.

A sample_submission.csv file provides all customer_id values which need prediction generation for evaluation purposes. The test evaluation solely examines customers who purchased items between 7 days after the training data collection. The system must generate predictions for the article_ids customers will most likely buy within the specified time period. The test evaluation requires scoring all customers even though their purchase history during training is unknown which presents crucial cold-start challenges.

The data collection offers an exhaustive condition for testing and measuring a recommendation platform which unites visual elements with user interaction records to replicate standard e-commerce system functions. The diverse realistic nature of the dataset enables scientists to design adaptable data-based recommendation platforms appropriate for fashion retail needs.

4. Proposed approach

4.1. Convolutional Neural Networks (CNNs) for Feature Extraction:

ResNet50 functions as our chosen architecture for fashion product stylistic detail extraction because it uses ImageNet-dataset-pretrained CNN abilities. Early layers of the model process basic image characteristics including edges and textures which the higher layers identify complex fashion patterns. Our model operates as a feature extractor after removing its classification layers due to which it can represent fashion products in a high-dimensional feature space.

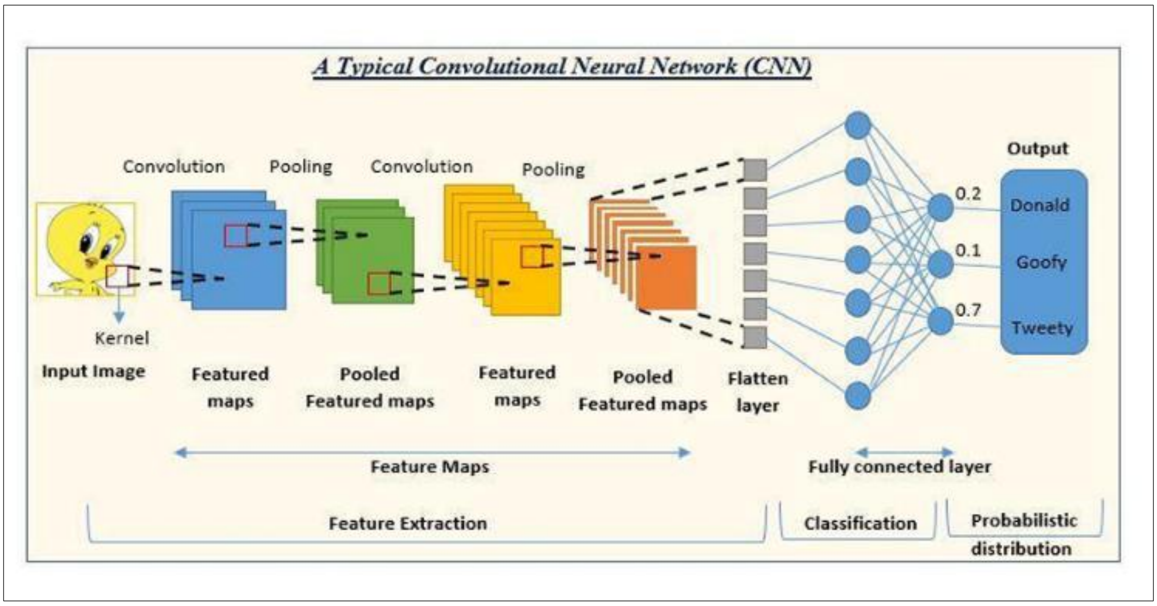


Figure 4 CNN Architecture

4.2. Feature Extraction Workflow

The initial step involves loading ResNet50 model which contains ImageNet weights. The model performs as the main extraction method for the feature extraction section. The model maintains only the trainable layers excluded from its configuration because it needs to sustain its learned features. The next step involves loading fashion product images into the system for pre-processing to match the required ResNet50 input format 224x224x3. The model processes each image once it receives an input to create a feature vector which represents the style details of the image. Feature vectors obtained from the models function as indicators of the stylistic properties within each fashion product. A pickle(.pkl) file contains the stored features that allow for easier utilization in machine learning model development. Our system benefits from the .pkl files because they provide convenient accessibility to these files along with better scalability features. By using pickle files our system can operate smoothly on streamlit hosting without requiring powerful server storage for data and improved execution speed.

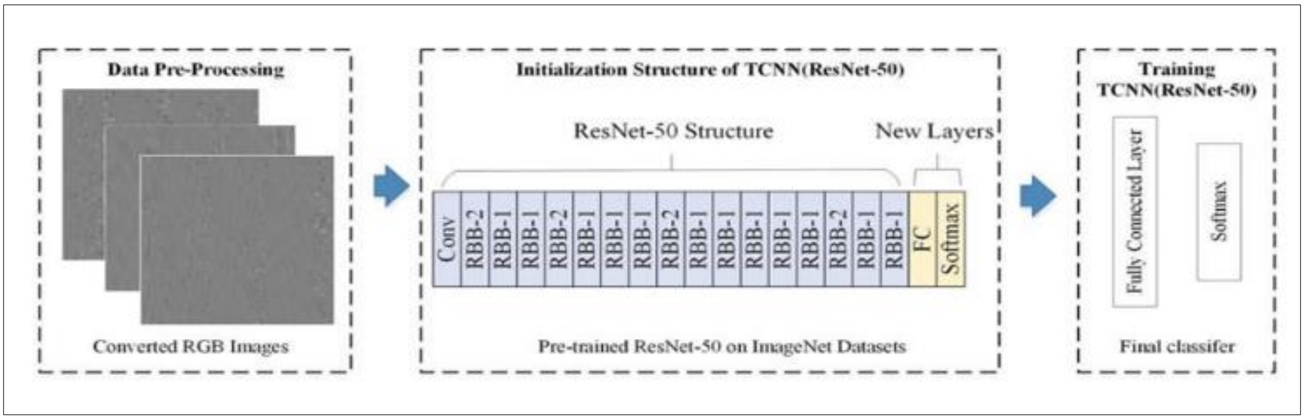


Figure 5 TCNN Architecture and Training Pipeline

5. System Architecture

The Style Matcher system architecture achieves fashion recommendations through precise and speedy visual results by implementing deep learning for feature extraction alongside visual similarity computation. The system constructs its architecture from five distinct modules which include Image Input and Preprocessing and Feature Extraction via CNN (ResNet50) alongside Feature Vector Storage and Similarity Matching Engine and a User Interface Layer made using Streamlit. Each element of the system operates as an individual module which achieves both performance enhancement and scalability.

5.1. Image Input and Preprocessing

The recommendation system begins by allowing users to upload fashion items (t-shirts or sunglasses) through the website interface. Before processing the uploaded image goes through three operations: first it gets resized to 224x224 pixels with RGB normalization and conversion to specified types to match ResNet50 standards. The executed steps maintain input consistency while ensuring compatibility with the pretrained model input specifications.

5.2. CNN-Based Feature Extraction (ResNet50)

A processed image goes through a ResNet50 model which operated as ImageNet pretraining. A GlobalMaxPooling2D layer replaces the upper layers of classification to transform last feature maps into a condensed high-dimensional vector format. The vector contains essential fashion product stylistic elements including texture and color schemes and patterns together with structural details

5.3. Feature Vector Storage and Retrieval

All feature vectors from the product dataset get precomputed and saved as serialized pickle (.pkl) format for faster access. Another pickle file contains all image paths that serve as associations to fashion images. The precomputed data allows real-time recommendations even on machines with minimal computational power because runtime dataset processing becomes unnecessary.

5.4. Similarity Matching Engine

The system computes similarity measurements by comparing the extracted vector from user images to the saved vectors through Cosine Similarity. The high-dimensional vectors undergo angle computation to assess visual similarity between them. The system utilizes similar scores to sort products and returns to the user the Top-K most related fashion items regarding their initial selection.

5.5. User Interface and Visualization (Streamlit)

Streamlit serves as the front-end development tool for creating an interface that operates interactively. An interface enables users to load images which generates recommended products and shows visualization results through flexible grid display. Users can interact with the system in real time through an interface which provides speed without delays.

6. Workflow Summary

A complete workflow included in this system operates in the following sequence:

- The application receives user images through the Streamlit interface.
- Image is resized and preprocessed.
- ResNet50 model extracts the feature vector.
- A comparison of cosine similarity exists between the entry vector with all stored vectors from the dataset.
- The system shows a selection of Top-K recommended items to the user.

The architectural structure makes Style Matcher both efficient and scalable because it can process extensive datasets quickly. The system design integrates modular architecture which simplifies the process of adding multi-modal inputs (such as text and images) or user feedback-driven personalization.

7. Methodology

7.1. Dataset Description

Fashion Product Images functions as a diverse database featuring multiple types of fashion products such as t-shirts together with dresses and sunglasses and additional accessories. The collection of images stands as an excellent tool for evaluating our recommendation system because they display different style elements in addition to color schemes and design motifs and material details. This dataset exhibits a diverse range of fashion items which lets us study visual similarity problems across multiple fashion domains just like what occurs in real-world fashion businesses. The dataset can be found.

7.2. Evaluation Metrics

Performance evaluation of our system utilized three appropriate metrics which included Mean Average Precision (MAP) for assessing precision across user inputs. The method proves useful when ranking recommendations because it factors in how items position themselves within generated recommendation lists.

$$MAE = (1/n) \times \Sigma_{(i=1)}^{n} |y_i - \hat{y}_i|$$

7.3. Experimental Findings:

Using different machine learning models and products, our system has an accuracy score reliably ranging between 85-95%. So I extracted features of the features using different model including VGG16, ResNet50, Tensorflow recommender model and InceptionNET. Additionally, the use of the MAP metric also showed that our system always produced meaningful and relevant recommendations.

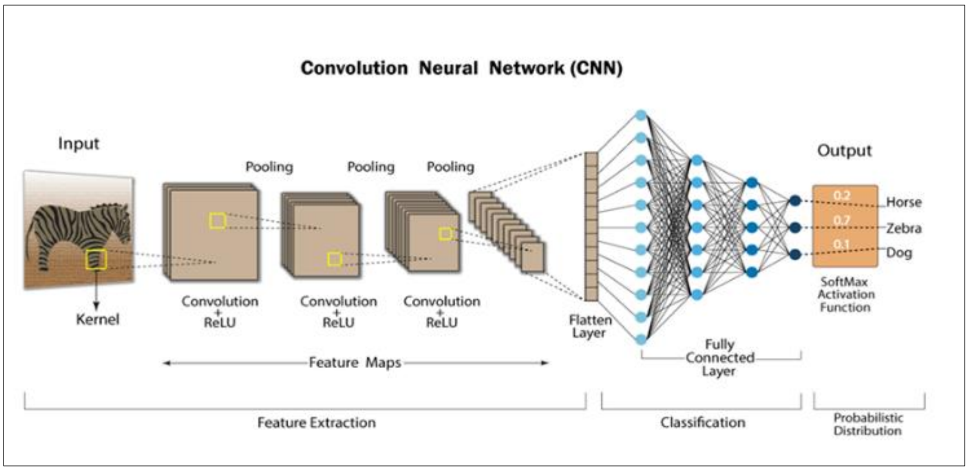


Figure 6 CNN Architecture

7.4. Feature Extraction from Dataset Images:

Import necessary libraries and modules. Build a ResNet50 model and set it up to be fed with extracted features. Here you have to stack ResNet50 with a GlobalMaxPooling2D layer to create a Sequential model. The function to extract features from an image and the image file paths from 'fashion_small/images' directory. Using the defined function, iterate through the image file paths and extract features for each. Save the extracted image features to a pickle file named "image_features_embedding3.pkl". Save the list of image file paths to another pickle file named "img_files3.pkl"

Generating Recommendations from Input Image: Import necessary libraries. Load image features and image file paths loaded from disk. Make a ResNet50 model of feature extraction. Set up the Streamlit app and the title. File upload and processing, feature extraction and recommendation can be defined as functions. Enable user to upload an image and process it. Apply some extraction functions to the uploaded image. Generate the fashion items recommendations by the extracted features to recommend visually similar fashion items. Show the recommended images in a grid packaging. If we are facing file upload issues, then we need to handle errors.

8. Implementation

The structure of the implementation of Style Matcher system is to run it through a pipeline of data preprocessing, deep learning based features extraction, similarity computation, and making of the interactive user interface. The core programming language used is python and libraries used like TensorFlow and Keras for deep learning, OpenCV and PIL for image processing and lastly Streamlit for making the web based interface.

The first is loading in the fashion product images and resizing them to 224x224x3 pixels which are the dimensions the ResNet50 model takes in for feeding. Then the images are normalized and RGB if needed. An ImageNet pre-trained ResNet50 model with the top classification layers removed is used. Finally, we convert this model to a feature extractor by adding Global MaxPooling2D layer which allows us to get the high dimensional embeddings that capture the stylisitcs of each fashion items.

The modified ResNet50 model will process each fashion image to extract.feature vectors. We also save these vectors into a serialized file using Python's Pickle module for obtaining them efficiently. This design ensures no retraining of the model or reprocessing of the dataset is necessary upon every user interaction in order to speed up and scale up it. Additionally, paths to image file are also saved from which recommendation can be displayed.

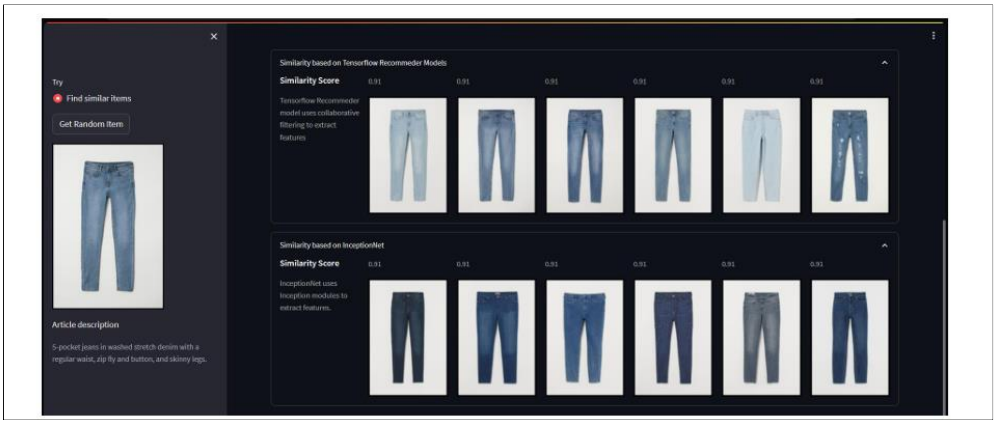
When generating recommendations, the system loads the pre computed feature vectors prior to computing cosine similarity of the uploaded user image to all stored images. Then it provides ranking and return of most visually similar images according to their similarity scores. With this approach, we have fast and accurate recommendation without an active database nor heavy computation while doing inference.

First, the user interface is made with Streamlit since it offers a lightweight and straightforward front end. The web application lets users upload a fashion image when it launches them and shows a fashion image once it is uploaded onto the interface. Then the image is uploaded, processed in real time, its features extracted, and fashion items most visually similar are presented in a grid layout. Thus users can easily view and compare styles to uploaded item. The goal of the complete system is to be deployable on these platforms that provide a responsive and interactive recommendation experience without developing the server infrastructure.

9. Results and Discussion

Next, we present the two sets of experimental details such as the dataset used, the evaluation metrics, and system results obtained using Convolutional Neural Networks (CNNs) in this section. The “Fashion Product Images” dataset from Kaggle, presenting 44,000 images of diverse fashion products were used by us. Moreover, the dataset consists of a wide variety of apparel items to allow training and testing of our system as per certain evaluation. As can be seen from highest accuracy of 94%, the ResNet50 model over the other had a high ability in dealing with classification task. It followed up with the accuracy of 87% on the TensorFlow Recommender Model which had a good performance. However, both models obtained an accuracy of 85%, which implies that all the other models performed better but not too far from VGG16 and InceptionNET. All in all, ResNet50 performed best of the models that were tested.

Figure 7 Result 1



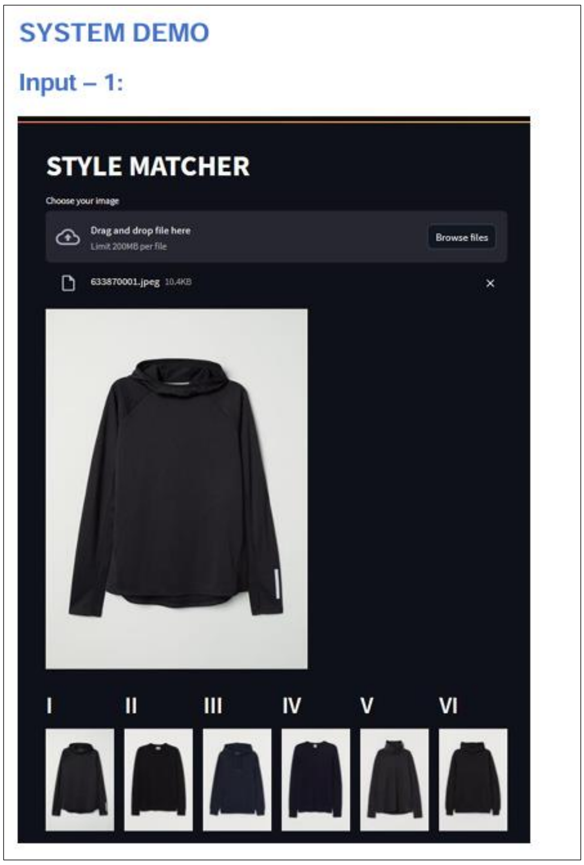


Figure 8 Result 2

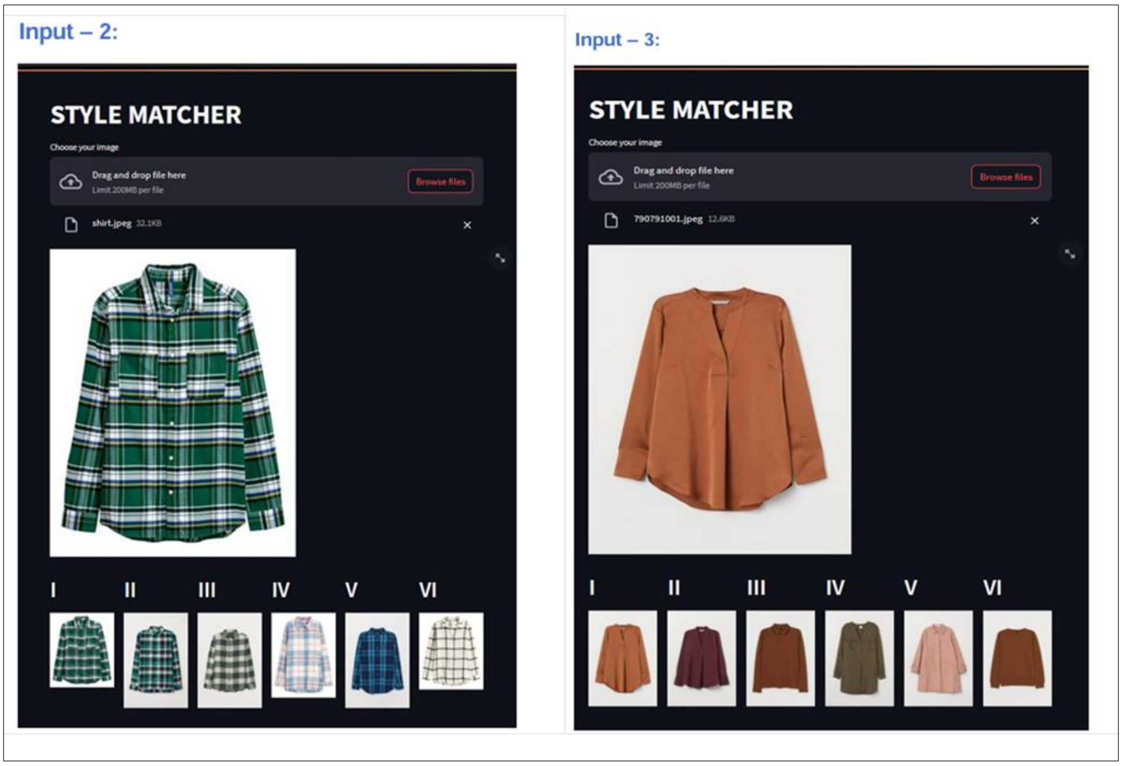


Figure 9 Result 3

10. Conclusion

Over the course of this study, we have touched on a novel fashion recommendation system built on a Convolutional Neural Network (CNN) to the issue of fashion visual similarity. Testing on our research yielded promising results as it confirms the system is able to provide personalized, visually coherent product suggestions to users. Our primary finding highlight the system's capability to correctly present subtle style features in fashion items. Dealing with fashion images is not an easy job, but leveraging on the deep learning techniques especially ResNet50 architecture has been very helpful as it has been able to extract meaningful representation from fashion images. The visual

We conduct evaluations on our solution using a similarity score, top-K recommendations, and mean average precision (MAP) metrics and show that our approach has a high robustness in delivering relevant and interesting fashion recommendations. The implication that this work has for the fashion industry and e-commerce platforms is substantial. Our system bridges user preference to product suggestion gap, with the aim of better user experience which leads to better customer engagement and satisfaction. So importing capacity to cater to individual style preference has the potential to attract and retain the customers in the end benefitting the business.

10.1. Future Work

Building upon the achievements of its existing implementation the Style Matcher system offers room for expansion through several improvements and expansions of capabilities. The system needs to integrate multiple data modals by uniting visual elements with textual product descriptions and user comments and product metadata records. Such system updates would result in advanced recommendation capabilities which would both fit users' style aesthetics while matching their material choices and price level and brand tastes.

The system should implement a mechanism to continuously gather feedback from users so recommendations can automatically modify based on the collected input. The recommendation engine can learn through feedback loops which implement reinforcement learning techniques that modify its behavior by using user interactions including clicks, likes and purchase data. The system would gradually become more accurate at generating appropriate recommendations through this process even when historical datasets are minimal.

Extension of the system with demographic information and social influence models helps solve cold-start issues effectively. No matter how new a user may be to the platform they will receive targeted recommendations that use their age, location or fashion trends observed in user profiles of similar characteristics. Fashion items that recently entered the catalog could receive recommendation matches through their processing from the CNN or textual metadata extraction system.

The existing method uses static embeddings together with offline feature extraction methods that work efficiently yet restrict scalability when inventory conditions change dynamically. Future system versions will examine real-time inference techniques together with embedding compression features to improve operational efficiency in industrial-scale commercial environments.

A mobile application integration system would allow users to make recommendations through instant photo capturing and submission. Augmented reality (AR) technology with this solution permits users to see potential purchases on their bodies before checkout thus enhancing both customer interaction and purchasing success.

The proposed system expansion includes outfit generation capabilities alongside compatibility modeling that recommends suitable combined fashion items such as tops and bottoms and shoes based on fashion principles. The system evolution would lead from a standalone recommendation mode to a full-style advising functionality which provides comprehensive shopping experiences to users.

Compliance with ethical standards

Acknowledgement

The authors would like to thank VIT-AP University for providing the resources and academic support needed to carry out this research work. Special thanks to Dr. Ganesh Reddy Karri for his valuable guidance throughout the development of this project.

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Fashion recommendation system using CNN2 - This paper proposes a content-based recommendation system using ResNet-50 convolutional neural network. [Venugeetha Y] Don Bosco Institute of Technology, Bengaluru, India

[2] An Efficient Fashion Recommendation System using a Deep CNN Model3 - This paper proposes an efficient Deep CNN model for classifying the given product.

[3] Li, M., Liu, H., Zhang, Y., & Hua, X. (2018). DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. In Proceedings of the 26th ACM International Conference on Multimedia (pp. 1656-1664).

[4] Yan, J., Zhang, X., & Wang, X. (2017). Fashion Recommendation in Social E-Commerce. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 1337-1346).

[5] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A LargeScale Hierarchical Image Database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 248-255).

[6] Venugeetha Y. Fashion Recommendation System Using CNN2. Don Bosco Institute of Technology, Bengaluru, India. This paper proposes a content-based recommendation system using the ResNet-50 convolutional neural network.

[7] K. Ravindra, and D. Patil, An Efficient Fashion Recommendation System using a Deep CNN Model. International Journal of Innovative Research in Computer and Communication Engineering, Vol. 9, Issue 5, May 2021.

[8] Li, M., Liu, H., Zhang, Y., & Hua, X. (2018). DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. In Proceedings of the 26th ACM International Conference on Multimedia (pp. 1656-1664).

[9] Yan, J., Zhang, X., & Wang, X. (2017). Fashion Recommendation in Social E-Commerce. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 1337-1346).

[10] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 248-255).

[11] Liu, Z., Luo, P., Qiu, S., Wang, X., & Tang, X. (2016). DeepFashion: Towards Bridging Semantic Gap in Fashion Image Understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2395-2404).

[12] Elsayed, S., Brinkmeyer, L., & Schmidt-Thieme, L. (2022). End-to-End Image-Based Fashion Recommendation. arXiv preprint arXiv:2205.02923.

[13] Veit, A., Kovacs, B., Bell, S., McAuley, J., Bala, K., & Belongie, S. (2015). Learning Visual Clothing Style with Heterogeneous Dyadic Co-occurrences. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4642-4650).

[14] McAuley, J., Targett, C., Shi, Q., & van den Hengel, A. (2015). Image-Based Recommendations on Styles and Substitutes. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 43–52).

[15] H&M Group. (2021). H&M Personalized Fashion Recommendations Dataset. Kaggle. Available: https://www.kaggle.com/competitions/h-and-m-personalized-fashion-recommendations

[16] Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In International Conference on Learning Representations (ICLR).

[17] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1–9.

[18] Mishra, S. (2021). Fashion Recommender System Dataset. Kaggle. Available: https://www.kaggle.com/datasets/shivammishra21/fashion-recommender-system

[19] Kolhe, S. (2021). Pinterest Fashion Dataset. Kaggle. Available: https://www.kaggle.com/datasets/samikshakolhe/pinterest-fashion-dataset

[20] Agrawal, P., and Garg, K. (2021). Fashion Recommendation System Using Deep Learning. In 2021 6th International Conference on Communication and Electronics Systems (ICCES), IEEE.

[21] Liu, Z., Luo, P., Qiu, S., Wang, X., & Tang, X. (2016). DeepFashion: Towards Bridging Semantic Gap in Fashion Image Understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2395-2404).