Deep Learning Techniques for Apparel recommendation

Darsh Vaishnani[1], Rohan Vaghela[2][0009-0004-8195-4590], Vraj Patel[3], Jigar Sarda [4][0000-0002-4530-9904], Akash Bhoi[5], Amit Thakkar[6][0000-0002-7517-0552], Srushti Gajjar[7][0000-0003-1433-9091]

[1, 2, 3,4,6,7] CSPIT, CHARUSAT University, Anand, Gujarat, India
[5]Directorate of Research Sikkim Manipal University Gangtok, India
22cs089@charusat.edu.in, 22cs088@charusat.edu.in, vrajp1706@gmail.com.in,jigarsarda.ee@charusat.ac.in, akashkrbhoi@gmail.com, amitthakkar.it@charusat.ac.in, srushtigajjar.cse@charusat.ac.in.

Abstract. In the era of digital globalization and burgeoning online shopping trends, the clothing E-commerce sector is experiencing exponential growth. However, in the midst of an abundance of options available, consumers often find themselves overwhelmed when selecting the perfect outfit. Current websites typically adopt a generalized approach, predominantly promoting best-selling or popular items, thereby neglecting the essence of customer-centricity and individuality. Consequently, there is a pressing demand for a streamlined and dependable system that aids users in discovering suitable products efficiently. To address this need, a hybrid recommender system (RS) has been proposed in this work. This system leverages both auto-encoder and K-nearest neighbor algorithms to enhance the recommendation process. By integrating user input data, the system can effectively tailor suggestions to match individual preferences and styles. Such personalized recommendations not only save time for users but also contribute to increased sales by fostering a more engaging and satisfying shopping experience. Ultimately, this innovative approach bridges the gap between the vast array of choices available and the discerning needs of consumers, facilitating informed decision-making and fostering a deeper connection between shoppers and their desired apparel.

Keywords Apparel recommendation, Auto-encoder, K-nearest neighbor (KNN), Deep Learning, Recommender System.

1 Introduction

RS may be beneficial for those seeking novel fashion concepts and inspiration, or for those who are uncertain about where to begin their search for certain things [1]. Prior to making a purchase, it is advisable to thoroughly evaluate the suggestions provided by a fashion guidance system. Fashion recommendation systems aim to assist customers in discovering novel apparel and accessories that align with their own style and preferences [2]. RS assists users in efficiently exploring extensive assortments of merchandise to discover items that align with their preferences , utilizing substantial quantities of product data and user indicators such as product views, interactions with items (such as ignoring or following), purchases, and webpage visits to ascertain the optimal timing, manner, and content of recommendations for their clientele. Recommendation systems (RSs) have become integral to the operations of major online merchants, contributing to as much as 35% of Amazon's sales and over 80% of the material consumed on Netflix [3]. In the realm of online clothing shopping, personalized recommendations serve as more than mere convenient suggestions; they serve as the vital link between consumers and the digital market. With the clothing e-commerce sector offering an overwhelming array of choices spanning brands, styles, sizes, colors, and price points, navigating through this multitude of options can prove daunting. Personalized recommendations alleviate this challenge, ensuring consumers are not lost in a maze of choices, but rather guided to items that resonate with their unique preferences and style. However, there has not been any research on RSs based on high involvement product which is characterized as involving low purchasing frequency, high price, and many evaluation criteria and much time in making a purchasing decision [4].

The study [5] presents a Memory-Augmented Neural Network (MANN) framework for suggesting appropriate clothing items based on features such as form and color. This is accomplished by separating the different aspects of fashion items and storing them in separate memory modules. These modules then provide guidance for making suggestions throughout the inference process. This method greatly enhances performance and provides interpretable latent spaces, hence improving the comprehension of recommendation reasons. By using the co-occurrence of qualities, the system efficiently combines different elements to create ensembles that are customized and harmonious, offering customers personalized and coherent fashion recommendations. The article [6] provides a comprehensive analysis of recommender systems enhanced using Language Model (LLM) techniques. It specifically examines the methods of pre-training, fine-tuning, and prompting. This paper explores the use of Language Models (LLMs) as feature encoders for representing users and items in recommender systems. It also investigates the latest developments in improving recommender systems using these approaches. The paper examines fundamental methodologies and explores potential advancements in this dynamic domain, providing valuable perspectives on the incorporation of LLMs to improve recommendation accuracy and user satisfaction. The Outfit Transformer framework, described in [7], employs task-specific tokens and self-attention to acquire outfit-level representations for the purposes of compatibility prediction and retrieval of complimentary items. The system utilizes outfit and target item tokens, as well as a set-wise outfit rating loss, pre-training, and curriculum learning to enhance retrieval performance. It surpasses state-of-the-art approaches in many tasks. In their paper [8], the authors provide a deep learning system called FashionNet that combines ResNet50 and KNN models for recommendation purposes. The system utilizes a two-step training method, where a broad compatibility model is first used to include individual preferences, resulting in personalized suggestions.

The study [9] introduces a two-stage deep learning model for recommending clothing fashion styles, leveraging convolutional neural networks (CNNs) to extract attributes from images and learn user preferences. The extracted attributes inform a correspondence model, facilitating the retrieval of visually related images for recommendation. By employing data-driven techniques, particularly CNNs as visual extractors, the experimental model outperforms previous schemes, demonstrating improved recommendation accuracy. This approach represents a significant advancement in leveraging deep learning for personalized fashion recommendation, enhancing user satisfaction and engagement in the ever-evolving realm of fashion e-commerce. The study [10] provides a comprehensive review of datasets and deep learning (DL) methodologies applied in the fashion domain, aimed at assisting new researchers. It covers five key tasks: Object Detection (including Clothes Landmark Detection, Clothes Parsing, and Product Retrieval), Fashion Classification, Clothes Generation, Automatic Fashion Knowledge Extraction, and Clothes Recommendation. By exploring these tasks, the paper highlights the diverse applications of DL techniques in the fashion industry. Its overarching goal is to underscore the breadth of opportunities for leveraging DL in addressing various challenges and advancing innovations within the realm of fashion.

Many current website approaches in the clothing e-commerce sector struggle to adequately address the challenge of catering to diverse consumer tastes. Instead, they often prioritize showcasing popular items or trends, assuming that what appeals to the masses will also resonate with individual preferences. However, this approach overlooks the unique and varied tastes of consumers, resulting in frustration. Users may find themselves disappointed with irrelevant recommendations or may miss out on discovering hidden gems that perfectly align with their personal style.

The main contributions of this work are as follows:

1. A hybrid approach for the apparel recommendation system has been proposed in this work.
2. The Dataset consist of various apparel and footwear images for both male as well as female.
3. Two well-known ML algorithms namely auto-encoder and KNN has been used for the task of recommendation.

2 Methodology

2.1 Dataset

This study utilizes a dataset sourced from Kaggle, comprising approximately 3000 images of apparel and footwear products [11], with the aim of developing a recommender system grounded in the visual representation of these items. Figure-1 shows some referral images which are used in dataset. Although dataset is a collection of Apparel and Footwear categories in which two gender types Boys (759 images) and Girls (569 images) are under Apparel category and Men (811 images) and Women (769 images) are under Footwear category.

Figure 1. Referral images from dataset

2.2 Implementation

The proposed model uses Convolutional Auto encoders (CAEs) to remove noise and extract visual features from images. It begins with specifying the input shape based on dimensions of the input images which is 256x256 pixels. Figure-2 represents the architecture of CAEs. It is divided into three parts: Encoder, Bottleneck and Decoder. The Encoder compress the raw input data by using dimensionality reduction and passes it to the Bottleneck layer. The Bottleneck contains the most compressed knowledge representation to pass through to the decoder. The decoder gets the compressed information and reconstruct it by using up sampling and convolutions. The design of CAEs is in such a way that it can be used to discover the minimum number of important features. Therefore, Training has been done on CAE model.

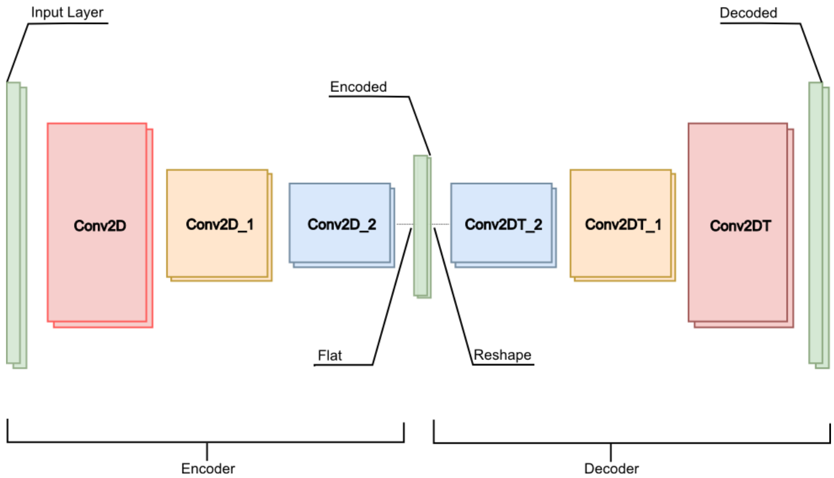


Figure-2. Architecture of CAE

The study involves such a model which is compiled with Mean Squared Error (MSE) loss function and Adam optimizer. Also, the encoder component of trained images is isolated to derive latent space representations (features) of the training images. The resultant features, along with their respective indices denoting the position of each image within the dataset, are aggregated into a dictionary format to serialize and store in pickle file to facilitate future retrieval and usage. Consequently, an array of latent features is generated for each image, which serves as the basis for identifying similar images. Utilizing these latent features, the process of finding most similar images is satisfied which involves Euclidean distance where a function computes the distance between the feature vectors of the query image and other images within the dataset. Subsequently, Function returns the indices of the N most similar images based on the computed distances. Equations-1 shows the mathematical formulation for Euclidean distance calculation.

$$ED(P,Q) = \sqrt{\sum_{i=1}^{n}(Q_i - P_i)^2} \quad (1)$$

Here, Pi and Qi are n components of the feature vectors corresponding to images P and Q.

Overall, this methodology, involving Auto-encoders, a neural network variant, demonstrate exceptional prowess in capturing intricate patterns and designs within images. Through the analysis of product designs, auto-encoders can uncover the underlying preferences and styles specific to individual users. In tandem, the K-nearest neighbour algorithm exploits similarities between users to generate recommendations. By recognizing input images provided by users, the K-nearest neighbour algorithm suggests items that align with similar preferences and styles. The synergy of these two algorithms empowers the hybrid recommender system to provide personalized recommendations that surpass the constraints of conventional approaches.

2.3 System Configuration

A GPU embedded device was used for all-different model training. The device is configured with 2 x Intel Xeon 16 core CPU with total 128 GB DDR4 RAM and a 2 TB SATA Enterprise HDD with NVIDIA RTX A5000 Graphics card. Jupyter Notebook version 6.4.3 is used for code implementation and model training.

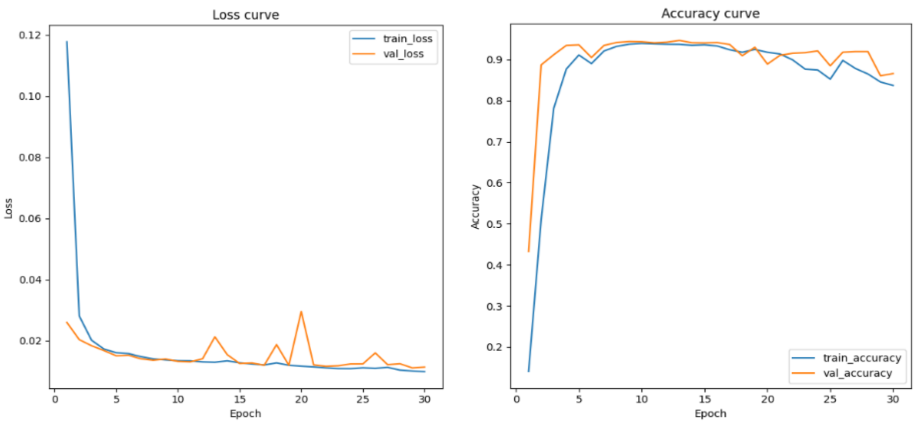
3 Results Analysis

The RS training was consisting of two phases. One with the training of the CAE model and then generate the related the feature and use the generated feature vectors as input to the KNN model. First, the CAE model is trained on various epochs and learning rate value. Table-1 represents values of loss an accuracy at various values of epochs and learning rate used for the training of model. From the table one can state that the highest validation accuracy of 91.90% and validation loss value of 0.0121 is achieved at epoch value of 30 and learning rate 0.001. Also, the highest training accuracy of 88.40% and training loss of 0.0112 is achieved at the same epoch and learning rate. All the combination were trained using Adam optimizer.

Table 1. Results with different parameter

Epochs	Learning Rate	Loss		Accuracy (in %)	
		Training	Validation	Training	Validation
30	0.01	0.254	0.153	70.91	87.67
30	0.001	0.0112	0.0121	88.40	91.90
50	0.01	0.0998	0.0507	66.84	73.78
50	0.001	0.0794	0.0288	85.63	81.45
100	0.01	0.0101	0.013	62.56	63.73
100	0.001	0.0164	0.0152	87.45	83.94

Figure-3 (a) shows the graph plotted between various values of loss and epochs values for the highest validation accuracy combinations. Initially, the loss values were too high but with the increase in the epochs value loss starts to decrease highly which shows good convergence. There some spikes in the validation loss while a smooth curve is obtained in the training. Figure-3 (b) shows the graph plotted between various values of epochs and accuracy for the same combination. The graph seems to rising with the staring value of epochs and become convergent at higher epochs value. Some disturbances are seen in both training as well as validation loss. Some large disturbances are seen in both the graphs but they can be ignored on the accuracy value.



(a) (b)

Figure 3. (a) Loss Curve and (b) Accuracy Curve

Figure-4 shows some inference of the recommender system proposed in the work. From the figure, one can see that three different images were recommended based on the user inputted images. The number of recommended images can be increases in the system as the output is given on basis of Euclidean distance from the inputted feature vector.

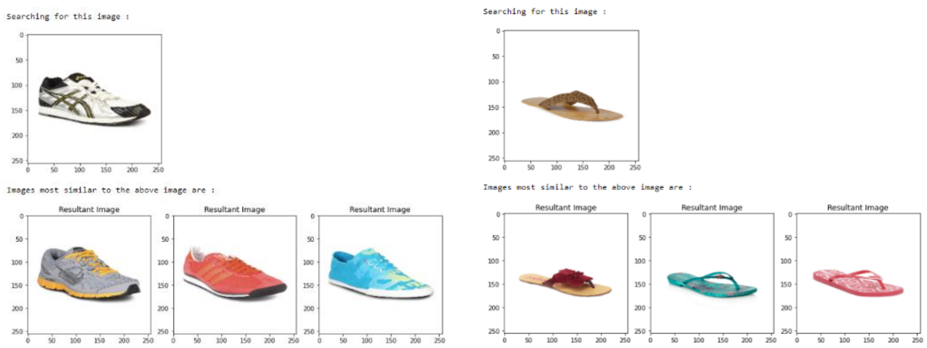


Figure 4. Some inference results

4 Conclusion

The study introduces a recommendation system (RS) tailored for fashion retail, utilizing feature vectors derived from the encoder section of a convolutional auto-encoder (CAE). Recommendations are generated based on the Euclidean distance between these feature vectors and the input vector, facilitating effective image suggestions for users. Moreover, the proposed solutions effectively address and rectify deficiencies found in existing state-of-the-art tools. With its deep understanding of product design and patterns, the system can curate recommendations that complement other products. Rather than suggesting trending items that may not match users' tastes, the system highlights items highly likely to resonate with the product images provided by users. This hybrid recommender system serves as a crucial link between consumer needs and available choices in today's digital platforms. By enhancing the shopping experience, the system boosts consumer satisfaction and loyalty, consequently driving business growth for e-commerce platforms. As technology advances, the significance of hybrid recommender systems is bound to become even more pronounced.

References

1. Pereira, A.M., de Barros Costa, E., Vieira, T., Landim, A.R. and Moura, J.A.B., 2023. Helping Online Fashion Customers Help Themselves: Personalised Recommender Systems. In Reinventing Fashion Retailing: Digitalising, Gamifying, Entrepreneuring (pp. 17-33). Cham: Springer International Publishing.
2. Mohamed Elleuch, Anis Mezghani, Mariem Khemakhem, Monji Kherallah “Clothing Classification using Deep CNN Architecture based on Transfer Learning” ,2021.

3. Ian mackenzie, Chris Meyer, and Steve Noble. 2013. How retailers can keep up with consumers. Mckinsey & Company. Retrieved September 28, 2023.
4. Sang Hyun Choi, Sungmin Kang, Young Jun Jeon, Personalized recommendation system based on product specification values, Expert Systems with Applications, Volume 31, Issue 3, 2006, Pages 607-616, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2005.09.074.

5. Lavinia De Divitiis, Federico Becattini, Claudio Baecchi, and Alberto Del Bimbo.2023. Disentangling features for fashion recommendation. ACM Transactions on Multimedia Computing, Communications and Applications 19, 1s, 1–21.
6. Wenqi Fan, Zihuai Zhao, Jiatong Li, Yunqing Liu, Xiaowei Mei, Yiqi Wang, Jiliang Tang, and Qing Li. 2023. Recommender systems in the era of large language models (LLMs). arXiv preprint arXiv:2307.02046.
7. Rohan Sarkar, Navaneeth Bodla, Mariya I. Vasileva, Yen-Liang Lin, Anurag Beniwal, Alan Lu, and Gerard Medioni. 2023. Outfit Transformer: Learning outfit representations for fashion recommendation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 3601–3609.
8. Panini Challa, Nagendra & Abbaraju, Sai Sathwik & Jinka, Chandra Kiran & Lokesh, Kokkula & Ch, Venkata & Naseeba, Beebi. (2023). Smart Fashion Recommendation System using FashionNet. ICST Transactions on Scalable Information Systems.
9. Muhammad KHALID, Mao KEMING, Tariq HUSSAIN, "Design and implementation of clothing fashion style recommendation system using deep learning", Romanian Journal of Information Technology and Automatic Control, ISSN 1220-1758, vol. 31(4), pp. 123-136, 2021.
10. M. Mameli, M. Paolanti, R. Pietrini, G. Pazzaglia, E. Frontoni and P. Zingaretti, "Deep Learning Approaches for Fashion Knowledge Extraction from Social Media: A Review," in IEEE Access, vol. 10, pp. 1545-1576, 2022.
11. E-commerce Product Images https://www.kaggle.com/datasets/vikashrajluhaniwal/fashion-images.