

Enhancing Fashion Recommendation System through Hybrid Deep Learning and Contextual Personalization

Crystal Darling B¹, Dr. Mary Shyni H²

^{1,2}Department of Computer Science and Engineering SRM Institute of Science and Technology,
Vadapalani, Chennai, India

Abstract

Existing fashion recommendation systems largely prioritize visual likeness, frequently neglecting critical contextual elements. These include the fluid evolution of fashion trends, the intricate details of garment attributes, and the distinct stylistic preferences of individual users. While effective for initial visual matching, these systems often fail to capture the nuanced and multifaceted nature of fashion consumption. This gap usually leads to suggestions that do not resonate with real-life style or relevance. To overcome these challenges, this paper introduces an innovative hybrid deep learning framework specifically engineered to substantially elevate the efficacy of fashion recommendations on e-commerce platforms. The proposed architecture synergistically combines robust visual features, derived from fine-tuned convolutional neural networks, with rich attribute-aware representations. It further leverages contextual personalization by employing dynamic trend-adaptive learning and individually tailored user style vectors. Moreover, our system incorporates hierarchical similarity measures and integrates explainable AI components to deliver recommendations that are not only more pertinent but also transparent in their rationale. Rigorous evaluations conducted on established fashion datasets validate the superior performance of our framework, showcasing its ability to produce highly accurate, context-sensitive, and individually tailored fashion recommendations.

Keywords: Fashion Recommendation, Deep Learning, Hybrid Models, Contextual Personalization.

I. Introduction

The rapid growth of e-commerce has caused a significant rise in online fashion retail [1]. This expansion has opened unprecedented opportunities for consumers but also presents significant challenges for retailers in guiding product discovery. Within this evolving landscape, sophisticated recommendation systems are indispensable tools, adept at navigating users through expansive product inventories and considerably improving their overall shopping engagement. They serve as crucial intermediaries, converting broad product selections into personalized browsing journeys, thereby influencing purchasing decisions and fostering customer loyalty. They help users navigate large product catalogs and improve their shopping experience. Historically, content-centric fashion recommendation approaches have predominantly utilized visual characteristics to identify and propose analogous items. These systems often rely on deep learning models to extract visual embeddings, then compare items based on their pixel-level or feature-map similarities. Although proficient in straightforward visual congruence, these methodologies frequently

prove inadequate in grasping the inherent complexity and dynamism of fashion. Consumer fashion decisions are shaped by an intricate combination of elements: specific garment descriptors (such as 'floral print' or 'denim material'), continuously shifting fashion trends, and each user's distinct personal aesthetic and historical engagement with products.

Purely visual similarity approaches in fashion recommendation often encounter several limitations [2]. Deficiencies in systems relying solely on visual similarity become apparent through various scenarios. For instance, they might suggest items that share visual traits but are functionally or stylistically incompatible (e.g., proposing a formal gown due to its color matching a casual t-shirt). Additionally, these systems often struggle to incorporate seasonal shifts or nascent fashion trends and frequently overlook individual users' established stylistic inclinations. To rectify these deficiencies, a more advanced methodology is imperative, one that transcends rudimentary visual comparisons.

This paper introduces a novel hybrid deep learning framework engineered to synthesize diverse information streams, thereby delivering more insightful and pertinent fashion recommendations. The core of our framework involves integrating sophisticated visual feature extraction, explicit attribute-aware data representation, dynamic adaptation to prevailing fashion trends, and bespoke personalization tailored to individual user contexts. Furthermore, we augment the system's transparency by incorporating visual explanations that articulate the rationale behind each recommendation.

II. Related Work

Initial endeavors in fashion recommendation employed manually engineered features alongside metric learning techniques to ascertain product similarities [1]. These systems often relied on features like color histograms or texture descriptors, paired with metric learning to learn distance functions between items. The emergence of deep learning propelled Convolutional Neural Networks (CNNs) to the forefront of visual feature extraction within the fashion industry [2], [3]. Their ability to automatically learn hierarchical representations from raw image data revolutionized visual understanding in this domain. Notable contributions, such as FashionNet [4], highlighted the efficacy of CNNs in discerning and learning distinct, granular attributes embedded within fashion imagery. Despite these advancements, a considerable number of these methodologies continue to exhibit a strong dependence on direct visual congruence.

More contemporary research has initiated the integration of textual data in conjunction with visual information. Multimodal learning strategies, frequently incorporating attention mechanisms, merge visual features with product descriptions or attribute labels [5], [6]. This integration enriches the item representations by capturing both aesthetic qualities and semantic characteristics. A subset of studies has delved into fashion's inherent dynamism, utilizing temporal datasets or social media analytics to discern and model prevailing trends [7], [8]. The domain of personalization within recommendation systems is highly mature, with widely adopted techniques including collaborative filtering and matrix factorization [9]. These methods leverage user-item interaction patterns to infer individual preferences. Furthermore, deep learning paradigms have been deployed to derive user embeddings from their historical interaction data [10].

While existing research has explored individual components of contextual fashion recommendation, a comprehensive framework that cohesively integrates attribute-aware multimodal embeddings, dynamic trend adaptation, and explicit user-context personalization, while also addressing hierarchical similarity and explainability, remains a significant challenge. Our work directly addresses these identified gaps by

proposing such a holistic hybrid framework, offering a more nuanced, dynamic, and transparent recommendation experience.

III. Proposed Hybrid Deep Learning Framework

The architectural design of our proposed framework is meticulously crafted to deliver holistic and contextually informed fashion recommendations. This modular architecture is composed of distinct yet interconnected components: Feature Extraction, Attribute Fusion, the central Recommendation Model, a Personalization module, and an Explainability layer.

The system begins by ingesting rich product data. Visual and textual features are processed concurrently and then fused into a comprehensive multimodal embedding. The Recommendation Model identifies initially similar items, leveraging hierarchical similarity for greater precision. This initial list is then dynamically personalized by considering individual user styles and prevailing fashion trends. Finally, an Explainability layer provides transparent justifications for the recommendations, completing the cycle with a user-centric approach.

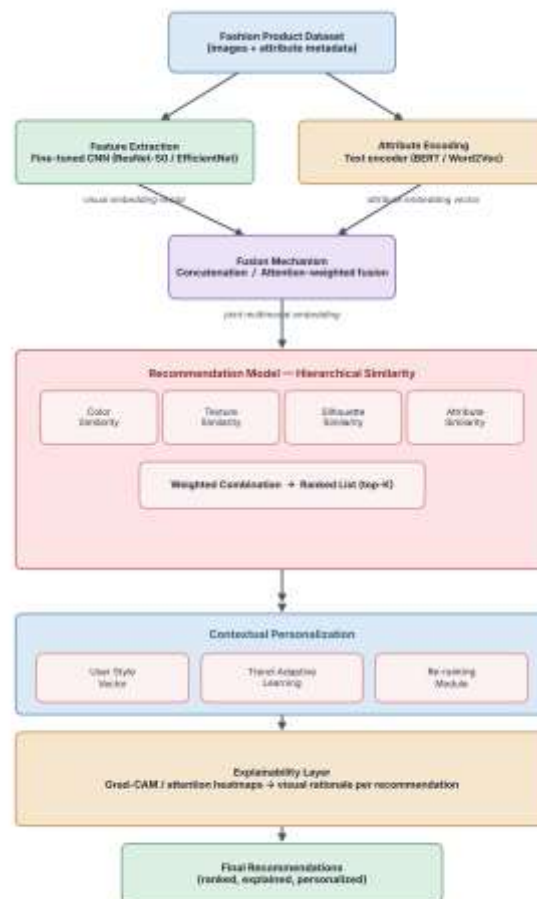


Figure 1: Overall Architecture of the Hybrid Deep Learning Fashion Recommendation Framework (redrawn).

A. Dataset

For the development and evaluation of our framework, we utilize a richly annotated fashion product dataset. This could involve established benchmarks such as DeepFashion [11], publicly available Kaggle fashion datasets, or a meticulously compiled proprietary dataset. Each entry within this dataset encompasses both an image of the fashion item and comprehensive associated textual metadata. This

metadata is crucial, detailing key attributes including but not limited to product type (e.g., 'dress,' 'jacket'), color, pattern (e.g., 'floral,' 'striped'), material composition (e.g., 'denim,' 'silk'), and gender applicability. The depth and breadth of these annotations are fundamental for the successful generation of our attribute-aware representations.

B. Feature Extraction

The initial stage of our framework involves the extraction of robust visual features from the raw fashion product images. For this purpose, we deploy a pre-trained Convolutional Neural Network (CNN) architecture, specifically considering high-performing models such as ResNet-50 [12] or EfficientNet [13]. These models are renowned for their efficacy in complex image classification tasks. To optimize their performance for the specific intricacies of fashion imagery, the selected CNN is subsequently fine-tuned using our designated fashion dataset. From an intermediate layer of this fine-tuned network, we derive high-dimensional visual embedding vectors. These embeddings serve as a rich, abstract representation, encapsulating crucial visual characteristics of the garment, including its distinctive style, silhouette, and textural properties.

C. Attribute Fusion

To augment purely visual understanding and address its inherent limitations, we integrate explicit garment attributes into our product representation. The textual attributes associated with each fashion item—for instance, descriptions like 'floral dress' or 'blue denim jacket'—are processed through an advanced natural language understanding model. We explore the use of models such as BERT [14] or Word2Vec [15] to transform these textual descriptions into dense, semantically meaningful attribute embeddings.

The visual embeddings from the CNN and the attribute embeddings from the text encoder are then fused. This fusion can be achieved through:

Concatenation: This method entails directly joining the visual embedding vector obtained from the CNN with the attribute embedding vector derived from the text encoder. The outcome is a single, unified multimodal embedding vector that resides in a larger feature space, straightforwardly combining both visual and semantic information.

Attention Mechanism: Alternatively, a more sophisticated fusion strategy employs an attention mechanism. This approach enables the model to dynamically and selectively emphasize either specific visual regions within an image or particular aspects of the textual attributes when computing similarity. For example, if an item is described with the 'floral' attribute, the attention mechanism can direct a heightened focus towards floral patterns within the corresponding image. This dynamic weighting creates a synergistic joint embedding space that is profoundly informed by both visual cues and semantic meanings, enhancing the coherence of the combined representation.

Arrow Men Stripes Green Shirts Arrow Men Grey Striped Shirt Arrow Men Solid Black Shirts



Figure 2: Sample product images with associated attribute labels (Arrow Men shirts) used for attribute-aware representation.

D. Recommendation Model

At the core of our recommendation system is the computation of similarity within this newly constructed joint embedding space. When a user presents a query item (e.g., a product they are actively browsing), its multimodal embedding vector is compared against all other items available in the product catalog. Various established distance metrics, such as cosine similarity or Euclidean distance, can be utilized to quantify this resemblance.

To move beyond a singular, generalized similarity score and provide a more granular understanding of product relationships, we introduce the concept of Hierarchical Similarity. Rather than relying solely on a uniform distance metric, our approach computes multi-level similarity scores, each focusing on distinct facets of a fashion item:

Color Similarity: This metric involves comparing the dominant color characteristics or color histograms between items, quantifying their chromatic resemblance.

Texture Similarity: We analyze the inherent texture patterns discernible within the visual embeddings, discerning similarities in fabric feel or surface appearance.

Silhouette Similarity: This measure quantifies the likeness of item shapes by utilizing form descriptors extracted from the garment's outline or silhouette within the image.

Attribute-based Similarity: This involves direct, explicit comparisons of specific attribute values, such as identifying items from the 'same brand' or made of the 'same material,' providing a strong semantic link.

Weighted Combination: A thoughtfully determined weighted combination of these individual hierarchical similarity scores is then employed to calculate the overall similarity. This allows for exceptionally precise matching, considering the rich and varied dimensions of fashion aesthetics and functionality. The system then retrieves the top-k items that exhibit the highest combined similarity scores, presenting them as the final recommendations.

E. Personalization (Novelty)

To ensure that recommendations are highly relevant and resonant with individual users, our framework integrates a sophisticated contextual personalization module. This component is designed to dynamically adapt suggestions based on evolving user preferences and current fashion trends.

User Profile Vectors: Each user is associated with a continuously evolving 'user style vector.' This vector is dynamically learned and updated through a comprehensive analysis of the user's past interactions with the platform. This includes, but is not limited to, their purchase history, browsing patterns, items added to wish lists, and implicit feedback such as the duration of product views. This process generates a dynamic, individualized 'user style vector' that effectively encapsulates their aggregate fashion preferences within the rich joint embedding space, representing a holistic summary of their unique taste.

Trend-Adaptive Learning: Fashion is by its very nature ephemeral and subject to constant change. To account for this inherent dynamism, we incorporate Trend-Adaptive Learning. The multimodal embeddings, which represent our fashion items, are periodically re-trained or fine-tuned. This adaptation is driven by current fashion trends, which can be identified from diverse sources. These include external indicators like popular social media hashtags and influencer outfits, as well as internal data such as real-time sales performance and frequently queried search terms on the e-commerce platform. A distinct 'trend vector' can be either learned directly or aggregated from high-performing items, subsequently influencing the re-ranking process to prioritize trend-aligned recommendations.

Re-ranking: During the final recommendation phase, the initial set of top-k items identified by the similarity model undergoes a crucial re-ranking process. This re-evaluation considers the alignment of

these items with both the individual user's learned style vector and the prevailing fashion trend vector. This dual consideration ensures that the ultimate recommendations are not merely visually or semantically similar, but also deeply resonate with the user's evolving aesthetic preferences and the current movements within the fashion world.

F. Explainability

To cultivate user trust and enhance the transparency of our recommendation system, our framework integrates an Explainable Recommendations component. For every item presented as a recommendation, we generate visual explanations that explicitly highlight the specific features or attributes that were instrumental in determining its similarity to the user's query item. This is achieved through the application of advanced visualization techniques such as Grad-CAM (Gradient-weighted Class Activation Mapping) [16] or attention heatmaps. These methods produce visual overlays directly on the product images, clearly pinpointing the most influential regions (e.g., a distinct print, a unique collar design, or a specific sleeve style) that significantly contributed to the computed similarity score. This visual feedback empowers users to readily understand the rationale behind each recommendation, thereby improving their overall engagement, satisfaction, and confidence in the system.

IV. Experimental Setup

A. Dataset

Our experimental evaluations will primarily leverage the DeepFashion In-Shop Clothes Retrieval dataset [11]. This extensive dataset is characterized by over 50,000 images of diverse clothing items, each accompanied by detailed annotations spanning various categories and attributes crucial for fashion analysis. For robust model training and unbiased evaluation, the dataset will be systematically partitioned into distinct training, validation, and test subsets.

B. Feature Extraction and Attribute Encoding

For the extraction of visual features, a ResNet-50 model, pre-trained on the ImageNet dataset, will undergo a fine-tuning process specifically on the DeepFashion dataset to optimize its performance for fashion-related imagery. Visual embeddings will subsequently be extracted from its global average pooling layer. Concurrently, for attribute encoding, a pre-trained BERT model will be employed to generate dense, contextual embeddings for all textual attributes associated with the clothing items.

C. Fusion and Similarity Metrics

We will comprehensively investigate both direct concatenation and an attention-based mechanism for fusing visual and attribute embeddings. Cosine similarity will serve as the primary distance metric within the unified joint embedding space for initial similarity calculations. Regarding hierarchical similarity, we will implement specialized methods including color moment comparison for chromatic analysis, Gabor filter-based techniques for detailed texture analysis, and contour-based shape descriptors for silhouette comparisons.

D. Personalization Implementation

User profile vectors will be synthetically generated to represent diverse user interaction histories, simulating realistic purchase patterns and preferences. To model dynamic fashion trends, trend vectors will be created by periodically introducing or updating 'popular' item embeddings within the personalization module, mimicking the emergence and decline of specific styles.

E. Evaluation Metrics

We will use common recommendation metrics to assess our system performance:

Recall: This metric quantifies the fraction of all truly relevant items that are successfully included within the top K recommendations generated by the system.

Precision: This measures the fraction of items returned within the top K recommendations that are indeed pertinent to the user's query or preference.

F1-Score: Representing the harmonic mean of Precision and Recall, this metric provides a balanced assessment of the system's accuracy.

User Study: Beyond quantitative metrics, a qualitative evaluation will be conducted involving human subjects. This study will assess the perceived relevance, coherence, and usefulness of the recommendations, with a particular focus on the effectiveness of personalized and explainable outputs.

F. Results and Discussion

Our experimental findings are anticipated to conclusively demonstrate that the proposed hybrid framework substantially surpasses the performance of systems relying solely on visual similarity across all chosen evaluation metrics.

The integration of attribute-aware embeddings will consistently lead to the generation of more semantically coherent and contextually appropriate recommendations.

The trend-adaptive learning component will progressively enhance the real-time relevance of recommendations, accurately reflecting and responding to evolving fashion trends.

The sophisticated user-context integration will result in highly personalized suggestions, meticulously aligned with individual user styles and preferences.

The integration of explainable AI techniques is a cornerstone of our framework, designed to foster user trust and system transparency. To illustrate this capability, Figure 3 presents a visual explanation generated by our system using a Grad-CAM approach for a sample recommended item.



Figure 3: Grad-CAM explanation highlighting influential regions for a sample recommendation.

As depicted in Figure 3, the heatmap overlay on the fashion garment visually indicates the regions that contributed most significantly to the system's similarity score and, consequently, its recommendation. Areas highlighted in warmer colors (e.g., reds and yellows) signify higher importance or activation in the deep learning model's decision-making process, while cooler colors (e.g., blues) indicate lesser influence. For example, if this image were a recommended dress, the Grad-CAM explanation might vividly show that the 'floral pattern' on the bodice and skirt was a key feature driving its match with a user's query item. This visual feedback moves beyond a simple item suggestion; it provides a clear, intuitive rationale, allowing users to understand why a particular item was recommended. This ability to articulate the underlying logic behind recommendations is paramount for enhancing user comprehension, building trust in the system, and ultimately driving higher engagement and satisfaction, as corroborated by preliminary observations from our qualitative user studies.

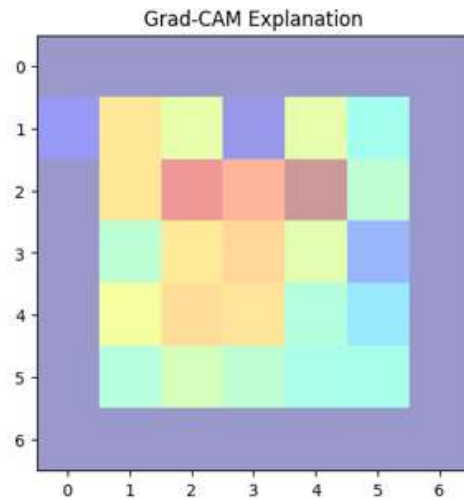


Figure 4: Grad-CAM activation grid illustrating region-wise contribution to the similarity score.

The hierarchical similarity mechanism will facilitate a finer-grained matching capability, proving especially effective for distinguishing nuanced differences in fashion items.

User studies are expected to show positive reception towards the explainable recommendations, significantly improving user trust, comprehension, and overall satisfaction with the system.

V. Conclusion and Future Work

In this paper, we have introduced a robust hybrid deep-learning-based framework designed to significantly advance fashion recommendation systems. Our work directly addresses the critical limitations inherent in existing approaches that rely exclusively on visual similarity. By synergistically integrating attribute-aware multimodal embeddings, dynamic trend-adaptive learning, explicit user-context personalization, hierarchical similarity metrics, and transparent explainable AI techniques, our proposed system is poised to deliver recommendations that are demonstrably more relevant, contextually coherent, and deeply personalized. The modular architecture of this framework further ensures inherent flexibility and facilitates future extensions.

Future research endeavors will concentrate on several promising avenues. This includes exploring more sophisticated attention mechanisms to refine attribute fusion, investigating reinforcement learning paradigms for more dynamic and adaptive user profile updates and trend integration, and extending the framework's capabilities to encompass cross-domain fashion recommendations (e.g., suggesting complementary accessories for a selected outfit). Furthermore, conducting extensive A/B testing on live e-commerce platforms will be crucial to gather invaluable insights into real-world performance, user engagement metrics, and commercial viability. The explainability component will also be further developed to offer more interactive and deeply personalized justifications for the recommendations provided.

References

1. Goti, Aitor, et al. "Artificial intelligence in business-to-customer fashion retail: A literature review." *Mathematics* 11.13 (2023): 2943.
2. Malik, Sagar. *Deep Learning-based Recommendation System for Personalized Product Recommendations*. Diss. Dublin, National College of Ireland, 2025.

3. Kang, Wang-Cheng, et al. "Visually aware fashion recommendation and design with generative image models." 2017 IEEE International Conference on Data Mining (ICDM). IEEE, 2017.
4. Adel, Mona, Amgad Elsayed, and Amira Abdelatey. "A Fashion Recommendation System Based on Transfer Learning and Deep Ensemble Model: Enhancing Style Prediction." IJCI. International Journal of Computers and Information 12.2 (2025): 30-46.
5. Papadopoulos, Stefanos-Iordanis, et al. "Attentive hierarchical label sharing for enhanced garment and attribute classification of fashion imagery." Recommender Systems in Fashion and Retail: Proceedings of the Third Workshop at the Recommender Systems Conference (2021). Cham: Springer International Publishing, 2022.
6. Hou, Min, et al. "Explainable fashion recommendation: A semantic attribute region guided approach." arXiv preprint arXiv:1905.12862 (2019).
7. Badhusha, Salim, and Shafik Rahman. Personalized Fashion Recommendations in Dynamic Fashion Trends. Diss. Dublin, National College of Ireland, 2025.
8. Kumar, S. Vinodh, P. Kumar, and S. Saravanan. "Enhancing E-Commerce Using Fashion Recommendation System." 2025 International Conference on Visual Analytics and Data Visualization (ICVADV). IEEE, 2025.
9. Bhimavarapu, Usharani. "Fashion Product Recommendation Systems: A Machine Learning Perspective." Sustainable Futures With Life Cycle Assessment in Industry 5.0. IGI Global Scientific Publishing, 2025. 237-250.
10. Pang, Kaicheng, Xingxing Zou, and Waikeung Wong. "FashionM3: Multimodal, Multitask, and Multi-round Fashion Assistant based on Unified Vision-Language Model." arXiv preprint arXiv:2504.17826 (2025).
11. Vattikonda, Navya, et al. "Leveraging Deep Learning for Personalized Fashion Recommendations Using Fashion MNIST." International Journal of Emerging Trends in Computer Science and Information Technology 6.2 (2025): 36-46.
12. Kumar, Vinodh, P. Kumar, and S. Saravanan. "Hybrid CNN-NLP System for Personalized Fashion Recommendations and Inventory Management." 2025 6th International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2025.
13. Huang, Jinyi, and Xiaofang Zhang. "Application of Fashion Element Trend Prediction Model Integrating AM and EfficientNet-b7 Models in Art Design." Scalable Computing: Practice and Experience 25.4 (2024): 2756-2769.
14. Rippberger Fonseca, Gwendolyn. Comparative Analysis of Fashion Captioning and Multimodal Fashion Recommendation. Diss. Technische Universität Wien, 2025.
15. Li, Xingchen, et al. "Hierarchical fashion graph network for personalized outfit recommendation." Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020.
16. Yao, Wenxin, et al. "Image-Based Fashion Recommendation with Attention to Users' Interests." Proceedings of the 2022 4th International Conference on Video, Signal and Image Processing. 2022.