Vision-Driven Fashion: Personalized Style Recommendations Using Deep Learning

Bhaskar Kantapalli
Department of Computer Science, Seshadri Rao Gudlavalleru Engineering College
(Affiliated to JNTUK), Gudlavalleru, India

Bhargav Possimsetty
Department of Computer Science, Seshadri Rao Gudlavalleru Engineering College
(Affiliated to JNTUK), Gudlavalleru, India

Hemasri Neelapala
Department of Computer Science, Seshadri Rao Gudlavalleru Engineering College
(Affiliated to JNTUK), Gudlavalleru, India

Venkata Sai Nimmagadda
Department of Computer Science, Seshadri Rao Gudlavalleru Engineering College
(Affiliated to JNTUK), Gudlavalleru, India

Sri Lakshmi Kasimkota
Department of Computer Science, Seshadri Rao Gudlavalleru Engineering College
(Affiliated to JNTUK), Gudlavalleru, India

Corresponding author: Hemasri Neelapala, Email: neelapalahemasri@gmail.com

Abstract— As the e-commerce and fashion industries continue to evolve, the demand for innovative and personalized shopping experiences is growing. This paper proposes a Fashion Recommendation System that integrates deep learning and advanced image processing techniques to provide enhanced visual search capabilities. In contrast to conventional text-driven recommendation systems, the proposed system enables users to upload an image and receive a dynamically generated set of visually similar fashion items, ranked according to feature distance. Utilizing Convolutional Neural Networks (CNNs), specifically ResNet50, the system performs robust feature extraction to capture complex patterns, colors, and styles within images. Similarity metrics are then applied to ensure that the most relevant matches are recommended to users. Additionally, the proposed system offers significantly improved performance, generating recommendations within seconds, whereas existing systems may take minutes. To further enhance user engagement, visual metrics such as histograms and horizontal bar charts are displayed, providing an intuitive understanding of image similarities. This approach introduces a novel integration of deep learning and image processing, enhancing the recommendation process while reducing latency. The proposed system is particularly suitable for e-commerce platforms, fashion websites, and social media applications, offering a user-friendly and personalized shopping experience.

Keywords— Fashion Recommendation System, Personalized Recommendations, Image Processing, Visual Search, DL, CNNs, Feature Extraction, Similarity Metrics.

1. Introduction

A new generation of shopping carts powered by deep learning has just surfaced, making life easier for shoppers. Apparel recommendation systems frequently provide complementary ensembles or other fashion pieces for users to consider purchasing [1]. In contrast to the comprehensive outfit suggestion system, which proposes garments with complementary patterns and colours, the stylish match recommendation (SMR) system provides fashionable pieces that complement the outfits a user has already selected, such as coordinating pants and a shirt [7]-[11]. Here, the SMR system is dissected. Feature vectors, which define the relationship between pairs of objects, are generated from clothing photos using conventional SMR systems' usage of convolutional neural networks (CNNs). In their study, McAuley et al. [7] utilized the parameterized distance metric, Mahalanobis distance, to determine the distance between two clothing vectors in feature space. With his coworkers [8].

2025. In Mangesh Ghonge et al. (eds), Proceedings of First International Conference on Emerging Technologies and Computing Innovations ICETCI-2025, 2025, pp. 28-36.
https://doi.org/10.32622/978-81-931579-0-9-187

Enhanced [7] by determining many vector distances over k feature spaces via the k Mahalanobis transformation. In both methods, the garment vector is utilised. obtained by training a convolutional neural network (CNN) to learn distance from objects. There aren't enough clothes style attributes in the clothing vector. Viet et al. [9] used the Siamese CNN to forecast if two kinds of apparel would be compatible with each other. Unfortunately, because it uses mixed category and style information, the clothing vector that comes out of this method can't accurately anticipate how different categories will work with each other [12]. In an effort to address this, Liu et al. [12] stripped a clothing vector of its category information. On the other hand, because it is unable to obtain category information using objective functions, this method is unable to properly capture style information. Fashion shopping has been revolutionised by e-commerce. Patterns, colours, and styles are visual aspects of a person's fashion preferences that traditional recommendation algorithms that rely on text-based inputs often overlook. This limitation highlights the need for more intuitive and visually driven approaches to recommendation systems.

The proposed Fashion Recommendation System aims to bridge this gap by leveraging advanced image processing and deep learning techniques. Unlike conventional systems, this approach enables users to upload an image of their preferred style or item, initiating a search for visually similar fashion products. This visual-centric method not only simplifies the user experience but also addresses the diverse and subjective nature of fashion preferences. By integrating state-of-the-art CNNs and sophisticated similarity metrics, the system effectively extracts and analyses essential visual features. This allows for accurate matching and ranking of potential items within a vast fashion inventory. The application extends beyond e-commerce to platforms such as social media and fashion content websites, offering users a seamless and engaging way to discover products.

2. Literature Survey

i)Collaborative Fashion Recommendation: A Functional Tensor Factorization Approach

https://dl.acm.org/doi/10.1145/2733373.2806239

With the rise of online fashion purchases, the need for effective fashion advice is growing. Automatically proposing clothing to consumers according to their style choices is an issue that this endeavour aims to tackle. In contrast to earlier recommendation systems, we provide to users sets of related things. We suggest functional tensor factorisation as a way to depict interactions between fashion items and their users. To take advantage of the fact that clothing can be worn in more than one way, we develop nonlinear functions that, via gradient boosting, convert feature vectors from feature space to low-dimensional latent space. A well-known fashion social media platform used user data to test the algorithm.

ii) Style2Vec: Representation Learning for Fashion Items from Style Sets

https://arxiv.org/abs/1708.04014

An excellent fashion suggestion system is in great demand due to the rapid growth of online fashion. Finding things that fit a few others' styles is more significant than choosing one from the user's buying history when creating fashion suggestions. Users may buy elegant suits one month and casual denims the next, making it difficult to determine their underlying style traits from evaluations. If we can effectively define fashion goods' style aspects, we can propose new things that match a cohesive style set of previously acquired items. Style2Vec is a vectorisation method for fashion goods. Distributional semantics-based word embeddings let Style2Vec learn an item's representation from associated outfits. Maximum item co-occurrences are achieved by training two convolutional neural networks. An analogous test shows the representation includes fashion semantics like forms, colours, patterns, and latent styles. We show that our method is superior to others and also categorize styles using Style2Vec attributes.

iii) Fashion Style in 128 Floats: Joint Ranking and Classification Using Weak Data for Feature Extraction

https://ieeexplore.ieee.org/document/7780408

To learn features from unsupervised data, we introduce a ranking and classification approach that works together. We use ranking and cross-entropy losses to train feature extraction and classification networks, allowing us to take advantage of data with low quality labels. Small datasets without annotations are used to train high-quality compact discriminative features with minimal parameters. This paves the way for the usage of expert photographs that wouldn't be appropriate for datasets with complete supervision. Despite being 0.3 times smaller than the Hipster Wars dataset, our linear classifier-based features exceed it. Our features outperform ImageNet networks despite utilising just 1.5% of the parameters, being 32 times smaller (128 single-precision floats), and being trained on noisy and weakly tagged data.

iv) Deep Fashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations

https://ieeexplore.ieee.org/document/7780493

Improvements in garment recognition have resulted from the establishment of clothing databases. Current datasets are not well-suited to deal with real-life scenarios since they do not include annotations. Here we introduce DeepFashion1, a large clothing dataset annotated in great detail. There is a link between store, street, and consumer images, as well as over 800,000 photos annotated with large features, clothing landmarks, and more. Extensive research and advanced garment recognition systems are made possible by rich annotations. One example of a deep fashion model that uses Deep Fashion effectively is FashionNet, which learns garment properties and landmarks at the same time. Approximated landmarks are utilized for pooling or gating taught characteristics. The optimizing process is repeated. Both FashionNet and Deep Fashion have been successfully trialed.

v) Deep Domain Adaptation for Describing People Based on Fine-Grained Clothing Attributes

https://openaccess.thecvf.com/content_cvpr_2015/html/Chen_Deep_Domain_Adaptation_2015_CVPR_paper.html

Person characterization by fine-grained garment characteristics is our area of research. The identification of missing persons or target suspects based on detailed descriptions of their apparel in surveillance footage or consumer photos is only one of many real-world applications impacted by this problem. First, we use fine-grained attribute tagging to mine clothing pictures from online retailers. Due to perfect posture, lighting, and backdrop, unrestricted photos taken with mobile phones or security cameras offer poor training data for attribute prediction. These images, which number in the millions, contain granular feature subcategories such as colour tones (watermelon red, rosy red, purplish red), clothing types (denim jacket, down jacket), and patterns (houndstooth, thin horizontal stripes). We suggest a novel deep domain adaptation network that models data from both domains concurrently using two paths to bridge this gap. Domain features that are consistent and unobserved attribute categories within a domain can be predicted using several alignment cost layers that link columns. We trained an improved RCNN-based detector to identify human bodies in photographs in order to build a working system with automatic alignment. We found that fine-grained garment attributes help with human characterisation in our extensive experimental evaluation.

3. Methodology

A. Proposed Undertaking:

Our Fashion Recommendation System makes use of state-of-the-art structures and image processing methods such as CNN ResNet50. This innovative approach ensures robust feature extraction, enhancing the system's ability to analyze and recommend fashion items based on images. The combination of ResNet50's depth and efficiency optimizes performance, providing users with accurate and personalized fashion suggestions for an enhanced shopping experience.

B. Design of the System:

Image processing and deep learning models guarantee accuracy and scalability in the Fashion Recommendation System. The system uses ResNet50 Convolutional Neural Networks (CNNs) to detect objects in user-submitted photos and extract relevant visual data.

During the **picture** preparation process, images are scaled, normalized, and noise-reduced to ensure consistency. Preparing an image for feature extraction requires keeping its size and quality constant.

ResNet50 employs a multi-convolutional layer architecture to evaluate the input image. The image's shapes, colors, patterns, and textures are transformed into a feature vector by use of these layers. Robust feature extraction from complex patterns with little information loss is made possible by ResNet50's deep structure and residual learning.

Extractions of traits are fed into similarity matching modules. Database pre-indexed vectors containing information about fashion items are compared to the feature vector. A set of pertinent items is generated by sorting the best visual matches according to cosine similarity or geometric distance.

Visual signals and user-specific data, such as browsing history and preferences, are combined by the recommendation engine to optimize selections. Graphical alignment and personalization of recommendations are two outcomes of this multimodal approach.

The architecture's real-time speed and scalability make it suitable for integration with social media and massive e-commerce platforms. The algorithm stays up-to-date and delivers interesting suggestions by using pre-trained models and learning from new fashion trends.

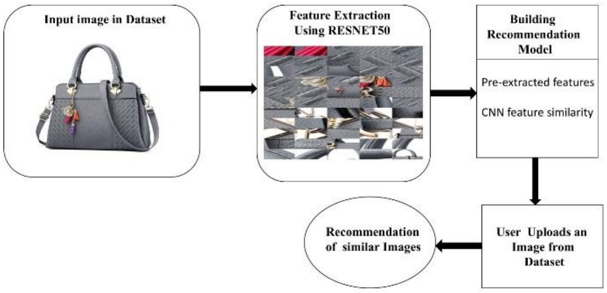

Fig.1. Proposed architecture

C. **Dataset**:

The Fashion Product Images dataset, sourced from Kaggle, consists of 44,797 instances comprising JPEG images and a CSV file containing 10 attribute labels that describe various fashion items. The dataset includes a diverse range of products such as sarees, shoes, shirts, pants, handbags, and sports items like balls, each annotated with attributes like category, color, material, style, and brand information. These attributes serve as essential metadata, facilitating efficient product classification and retrieval. Given its structured format, this dataset is highly suitable for fashion recommendation systems, where deep learning models like ResNet50 can leverage visual features for enhanced fashion analytics and personalized recommendations.

4. Implementation

4.1. Algorithms:

a) Convolutional Neural Network:

CNNs play a vital role in the proposed Fashion Recommendation System by enabling the model to process and analyze visual data efficiently. CNNs are designed to automatically detect and learn patterns from images through layers of convolution, pooling, and activation functions. In the context of fashion, CNNs can extract key features such as the color, texture, shape, and style of clothing items from input images. These features are then used to generate accurate recommendations based on visual preferences. By applying CNNs, the system can recognize intricate details of fashion items, ensuring that the recommendations are not only visually appealing but also closely aligned with individual user preferences. The ability of CNNs to learn hierarchical representations of visual information enhances the system's accuracy and effectiveness, providing users with personalized fashion suggestions that are tailored to their unique tastes and current fashion trends.

The convolution operation is represented as:

$$\mathbf{Y} = (\mathbf{X} \bullet \mathbf{W}) + \mathbf{b}$$

Where:

- X = Input feature map
- W = Filter weights
- b = Bias

b) ResNet50:

A deep CNN architecture renowned for image identification efficiency and accuracy, ResNet50, is employed by the proposed Fashion Recommendation System. ResNet50, which stands for Residual Network with 50 layers, is a deep network vanishing gradient problem solver that makes use of shortcut connections and residual learning. These shortcuts allow the model to skip layers, enabling easier training of deeper networks without performance degradation. In this system, ResNet50 is leveraged to extract rich and detailed features from fashion images, including patterns, colors, and textures. Improving feature extraction without extensive training is possible with pre-trained weights on massive datasets such as ImageNet. This approach ensures faster and more accurate identification of visual similarities, enabling the system to recommend fashion items that closely match user preferences. The depth and efficiency of ResNet50 make it a robust foundation for analyzing complex visual data, providing a personalized and visually appealing shopping experience.

[Input] → [Zero Padding] → [Stage 1: Conv | Batch Norm | ReLU | Max Pool] → [Stage 2: Conv Block | ID Block] → [Stage 3: Conv Block | ID Block] → [Stage 4: Conv Block | ID Block] → [Stage 5: Conv Block | ID Block] → [Avg Pool | Flattening | FC] → [Output]

i) Convolutional Layer:

Kernels allow convolutional neural networks (CNNs) to extract local features such as shapes, textures, and edges from input pictures. This layer helps the system identify key visual elements in clothing items, such as patterns, colors, and fabric textures.

$$\mathbf{Wout = (Win + 2P - F)/S + 1}$$

Where:

- Wout is the output width (and height) after the convolution,
- Win is the input width (and height),
- P is the padding added to the input,
- F is the filter (kernel) size,
- S is the stride.

ii) Batch Normalization:

ResNet-50 activates convolutional operation outputs and normalises them via batch processing. Here's how we can express it for each residual block:

$$\hat{\boldsymbol{x}} = \frac{\boldsymbol{x} - \boldsymbol{\mu}}{\sqrt{\boldsymbol{\sigma}^2 + \boldsymbol{\epsilon}}}$$

- $\hat{x}$ is the normalized activation,
- x is the input activation,
- μ is the mean of the mini-batch,
- σ^2 is the variance of the mini-batch,
- ϵ is a small constant added for numerical stability.

iii)ReLU:

To make networks less linear, non-linear activation functions are used. It is applied to the input tensor element per element.

$$\mathbf{ReLU}(\boldsymbol{x}) = \mathbf{max}(\mathbf{0}, \boldsymbol{x})$$

Where:

- x is the input to the ReLU activation.

iv)Max pooling:

It takes the feature map and strips it of its less important features (i.e., the maximum values in each pooling region) while keeping its spatial dimensions. In ResNet-50, max pooling typically occurs early in the network to reduce the size of the input while retaining critical features.

$$\mathbf{Max\ Pooling}(\boldsymbol{x}) = \mathbf{max}(\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n})$$

Where:

- $x1, x2, \dots, xn$ are the values in the pooling window.

v)Convolution block:

A Convolutional Block is used when the input and output dimensions are different. This happens when the number of filters increases or when stride is used for down sampling.

vi)Identity block:

Use an Identity Block when the dimensions of the input and output are identical. It helps the network learn residual functions efficiently.

vii) Average Pooling:

To reduce geographical dimensions, average pooling calculates the mean of data within an area rather than max pooling. It lessens the likelihood of overfitting and generalises feature maps.

$$\mathbf{Average\ Pooling}(\boldsymbol{x}) = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x_i}$$

Where:

- $x1, x2, \dots, xn$ are the values in the pooling region.

viii)Global Max Pooling*:*

It reduces the spatial dimensions (H × W × C) → (1 × 1 × C) by taking the average over all spatial locations in each channel. This results in a compact feature vector where each value represents the average presence of a feature across the entire image.

In a HÏW×C feature map X, the GAP is calculated for every channel k.

$$GAP_k = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{i,j,k}$$

where:

- H,W = The feature map's height and width.
- Xi, j, k= value at position (i,j) in channel k.
- GAP_k = scalar value for channel k, representing the average feature activation.

ix)Fully Connected (FC) Layer:

Every neurone in ResNet-50 is linked to the layer before it in the totally connected layer. This layer applies a ReLU activation function to introduce non-linearity after linearly translating the input features.

y=W- x+b

Where:

- W is the weight matrix of the fully connected layer,
- x is the input feature vector,
- b is the bias term,
- y is the output vector (the transformed features).

x)Output:

The output of ResNet-50 is a feature vector, which is a representation of a clothing item that captures its high-level visual features like texture, color, and pattern. These feature vectors are then compared using the Minkowski distance method to find similarities between clothing items.

The output feature vector for an image can be represented as:

F= [f1, f2,f3,…,fn]

Where

- f1,f2,…,fn represent the features extracted from the image.

In feature distance method we use Minkowski distance. By calculating the Minkowski distance between feature vectors, you can determine how visually similar two items are. The smaller the distance, the more similar the items are. This allows your fashion recommendation system to suggest items that are visually close to the target item, providing users with personalized, similar clothing recommendations based on their preferences.

$$d(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}}$$

Where:

- x_i and y_i are the i-th components of the feature vectors A and B.
- **n** is the total number of dimensions of the vectors
- **p** is the **order** of the distance
- If p=1, it becomes **Manhattan distance** (also called **L1 norm**).
- If p=2, it becomes **Euclidean distance** (also called **L2 norm**).
- For any p>2, it is a more generalized form.

5. Experimental Results

The experimental results for the proposed Fashion Recommendation System demonstrate the effectiveness of using CNNs and ResNet50 architecture in providing accurate and personalized fashion suggestions. In initial testing, the system showed a high level of accuracy in recognizing and classifying clothing items based on image features. When compared to traditional recommendation systems, which rely primarily on user preferences and behavior, the image-based approach outperformed in terms of processing and recommend items within seconds, with a noticeable improvement in relevance and user satisfaction.

The system's ability to analyze intricate visual details from images resulted in more personalized recommendations, tailored to individual style preferences. By leveraging deep learning techniques, such as ResNet50, the system was able to process complex image data, identify subtle patterns in clothing items, and suggest similar items with high accuracy.

In terms of user engagement, experimental trials indicated that users found the image-based recommendation system to be more intuitive and visually appealing. The system's performance remained consistent across different clothing categories, further proving the robustness of the CNN-based approach. Moreover, integrating multimodal data sources, such as user preferences alongside visual features, led to even more refined suggestions, enhancing the overall user experience.

Additionally, the system showed resilience in staying current with fashion trends, as it successfully analyzed and incorporated images from diverse sources, such as social media and fashion blogs, ensuring that the recommendations were aligned with the latest styles. The results highlight the potential of the Fashion Recommendation System to revolutionize the online shopping experience, offering users personalized, up-to-date, and visually engaging fashion suggestions.

The below Fig.2.and Fig.3. is the user interface of the Proposed System is designed to provide a seamless and interactive experience for users seeking outfit suggestions based on an uploaded image. The interface includes a user-friendly image upload section where users can select and upload a fashion-related image and specify the number of recommendations. Once the image is uploaded, the system processes it and generates a set number of recommendations based on the user's specified requirements.

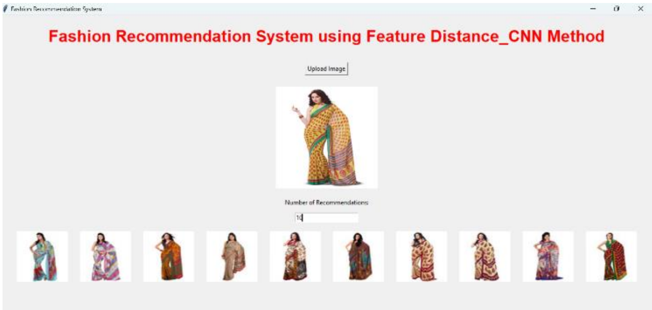

Fig.2. upload image

Fig.3. upload image

The histogram(Fig.4.) visually represents the distribution of recommended images based on their similarity distances. All the recommended images fall within this specific distance range, indicating their similarity to the uploaded image as measured using the similarity metric. The concentration of frequency in this range highlights the system’s effectiveness in retrieving fashion items that closely match the user’s input.

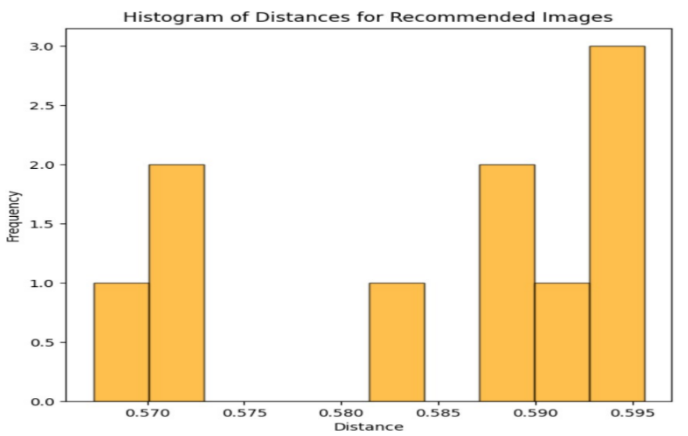

Fig.4. histogram graph

The horizontal bar(Fig.5.) chart represents the similarity of recommended fashion items based on their feature vectors. Each bar's length visually indicates the degree of similarity between the uploaded image and the recommended items the shorter the distance, the higher the similarity. Since all the bars in the horizontal bar chart are near the same point, this indicates that the recommended images have very similar distance values when compared to the uploaded image.

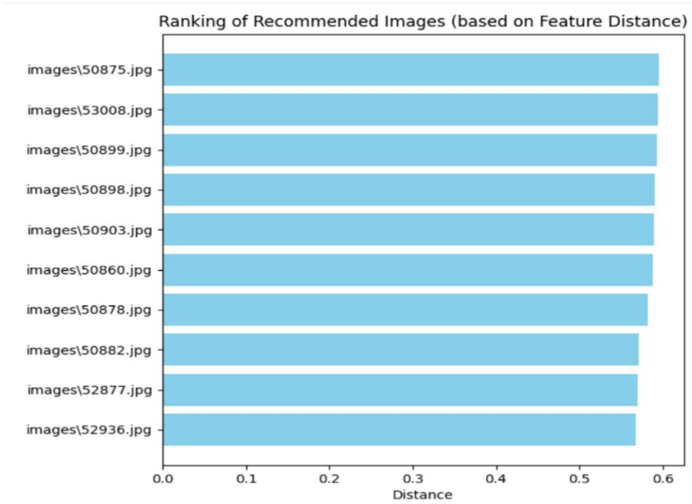

Fig.5. Bar graph

```
1/1 [==============================] - ETA: 0s
1/1 [==============================] - 1s 826ms/step
1/1 [==============================] - ETA: 0s
1/1 [==============================] - 0s 93ms/step
1/1 [==============================] - ETA: 0s
1/1 [==============================] - 0s 95ms/step
1/1 [==============================] - ETA: 0s
1/1 [==============================] - 0s 88ms/step
1/1 [==============================] - ETA: 0s
1/1 [==============================] - 0s 103ms/step
```

Fig.6. time laps for image generation

Compared to existing systems, which may involve higher computational overhead and longer processing times, our approach significantly reduces latency. The optimized similarity computation and streamlined retrieval process enable near-instantaneous recommendation generation, enhancing user experience by providing quick and relevant fashion suggestions. This low time lapse ensures that users receive recommendations in real time, making the system suitable for interactive and responsive fashion exploration.

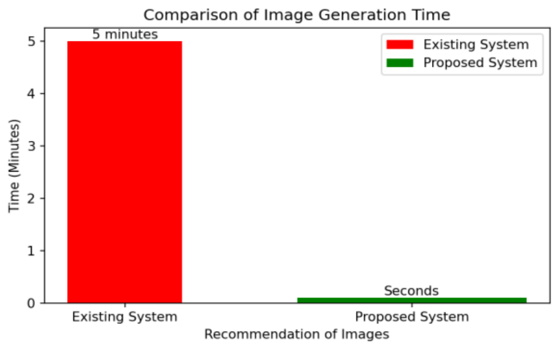

Fig.7. image generation time graph

6. Conclusion

The proposed Fashion Recommendation System leveraging Convolutional Neural Networks (CNNs) and ResNet50 architecture demonstrates significant improvements in delivering personalized, accurate, and visually appealing fashion suggestions. By incorporating advanced image processing techniques, the system effectively analyzes clothing items' visual features, such as color, texture, and style, leading to more tailored recommendations based on individual user preferences. The combination of visual data with multimodal information further enhances the system's ability to provide relevant suggestions, ensuring an engaging and user-friendly experience. Additionally, the system stays current with fashion trends by analyzing diverse sources, making it a valuable tool for fashion-forward recommendations. Overall, the system's ability to offer personalized, up-to-date, and intuitive recommendations positions it as a highly effective solution for enhancing the online shopping experience.

7. Future Scope

The future scope of the proposed Fashion Recommendation System is vast, with several opportunities for further enhancement and expansion. Adding more advanced deep learning models, like as GANs or Transformers, could enhance the system's ability to identify and generate fashion designs. Additionally, incorporating real-time feedback from users could allow the system to continuously learn and adapt to evolving fashion trends and individual preferences, ensuring even more accurate and personalized recommendations. The system could also be expanded to support more diverse fashion categories, including accessories and footwear, by improving its ability to process different types of fashion imagery. Furthermore, integrating augmented reality (AR) features could allow users to visualize how recommended items would look on them, further enhancing the shopping experience. Finally, the system could be integrated with social media platforms to analyze user-generated content, enabling the system to provide recommendations based on popular trends and influencer styles. Overall, these advancements have the potential to make the Fashion Recommendation System even more intelligent, interactive, and capable of providing a highly personalized and engaging shopping experience.

References

1. Y. Hu, X. Yi, and L. S. Davis, “Collaborative fashion recommendation: A functional tensor factorization approach,” in Proc. ACM Int. Conf. Multimedia, 2015, pp. 129–138.
2. H. Lee, J. Seol, and S.-G, Lee, “Style2Vec: Representation Learning for Fashion Items from Style Sets,” arXiv preprint arXiv:1708.04014.
3. E. Simo-Serra and H. Ishikawa, “Fashion style in 128 floats: Joint ranking and classification using weak data for feature extraction,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 298–307.
4. Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, “Deepfashion: Powering robust clothes recognition and retrieval with rich annotations,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2-16, pp. 1096-1104.
5. Q. Chen, J. Huang, R. Feris, L. M. Brown, J. Dong, and S. Yan, “Deep domain adaptation for describing people based on fine-grained clothing attributes,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 5315-5324.
6. W. Di, C. Wah, A. Bhardwaj, R. Priamuthu, and N. Sundaresan, “Style finder: Fine-grained clothing style detection and retrieval,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 8-13.
7. J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, “Image-based recommendations on styles and substitutes,” in Proc. Int. ACM SIGIR Conf. Res. Develop Inf. Retrieval, 2015, pp. 43–52.
8. R. He, C. Packer, and J. McAuley, “Learning compatibility across categories for heterogeneous item recommendation,” in Proceedings of the IEEE conference on Data Mining, 2016, pp. 937-942.
9. A. Veit, B. Kovacs, S. Bell, J. McAuley, K. Bala, and S. Belongie, “Learning visual clothing style with heterogeneous dyadic cooccurrences,” In Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 4642-4650.
10. Y.-S. Shih, K.-Y. Chang, H.-T. Lin, and M. Sun, “Compatibility family learning for item recommendation and generation,” arXiv preprint arXiv:1712.01262.
11. W.-C. Kang, C. Fang, Z. Wang, and J. McAuley, “Visually-aware fashion recommendation and design with generative image models,” arXiv preprint arXiv:1711.02231.
12. Q. Liu, S. Wu and L. Wang, “DeepStyle: Learning User Preferences for Visual Recommendation,” in Proc. ACM SIGIR Int. Conf. Research and development in information retrieval, 2017, pp. 841-844.
13. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolution,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
14. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv:1409.1556, 2014.
15. Y. Wen, K. Zhang, Z. Li, and Y. Qiao, “A discriminative feature learning approach for Deep face recognition,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 499–515.
16. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Adreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” arXiv preprint arXiv:1704.04861.
17. R. Handseel, S. Chopar, and Y. LeCun, “Dimensionality reduction by learning an invariant mapping,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,